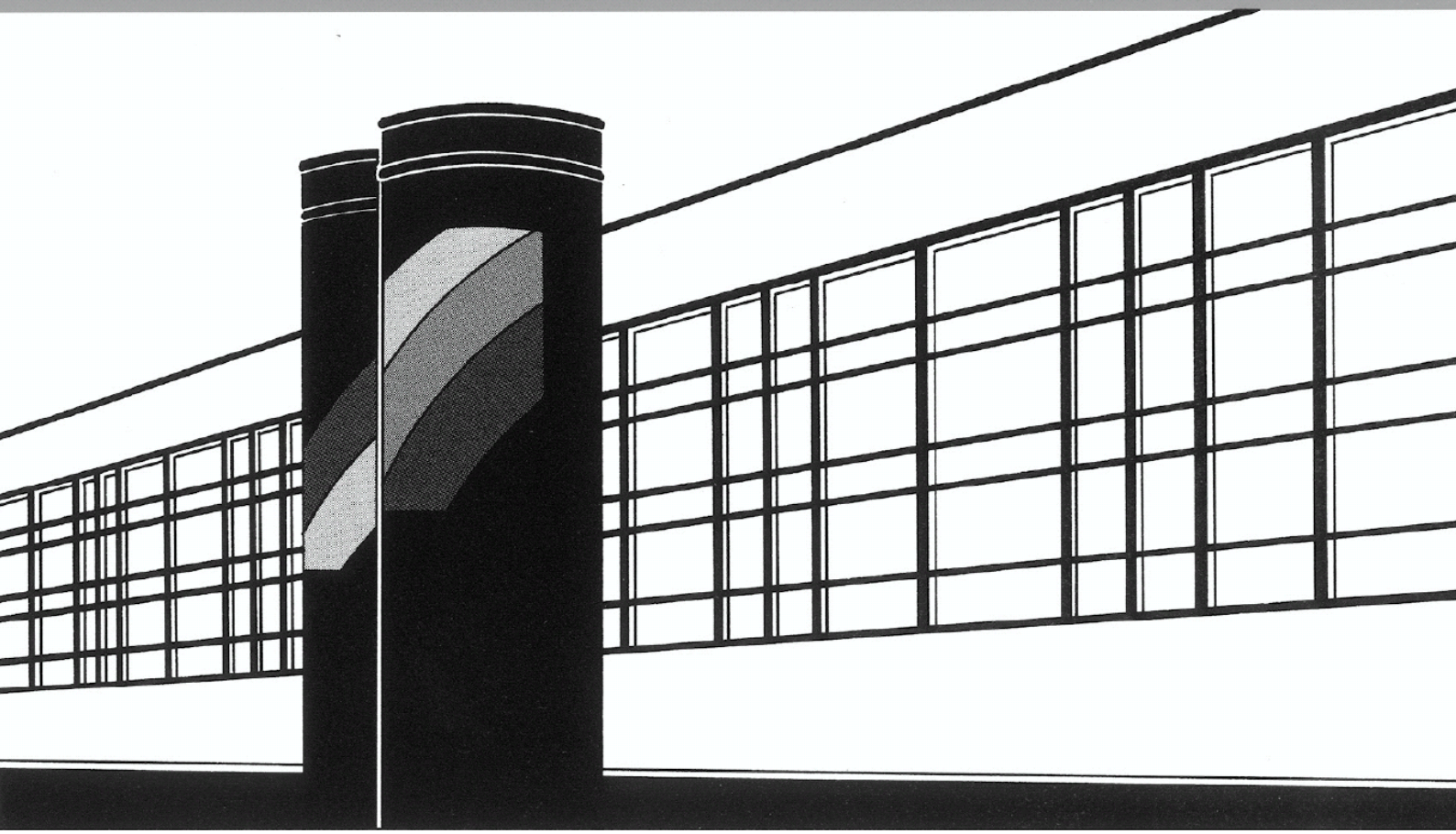
Universität Stuttgart

Institut für Wasser- und Umweltsystemmodellierung

Mitteilungen



Heft 219 Ferdinand Beck

Generation of Spatially Correlated Synthetic Rainfall Time Series in High Temporal Resolution - A Data Driven Approach

Generation of Spatially Correlated Synthetic Rainfall Time Series in High Temporal Resolution
A Data Driven Approach

Von der Fakultät Bau- und Umweltingenieurwissenschaften der Universität Stuttgart zur Erlangung der Würde eines Doktor-Ingenieurs (Dr.-Ing.) genehmigte Abhandlung

Vorgelegt von

Ferdinand Beck

aus Backnang, Deutschland

Hauptberichter:	Prof. Dr. rer.nat. Dr.-Ing. András Bárdossy
Mitberichter:	Prof. Dr.-Ing. Uwe Haberlandt

Tag der mündlichen Prüfung: 12. November 2012

Institut für Wasser- und Umweltsystemmodellierung
der Universität Stuttgart
2013

Heft 219

Generation of Spatially Correlated Synthetic Rainfall Time Series in High Temporal Resolution - A Data Driven Approach

von
Dr.-Ing.
Ferdinand Beck

Eigenverlag des Instituts für Wasser- und Umweltsystemmodellierung
der Universität Stuttgart

D93 Generation of Spatially Correlated Synthetic Rainfall Time Series in High Temporal Resolution - A Data Driven Approach

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://www.d-nb.de abrufbar

Beck, Ferdinand
Generation of Spatially Correlated Synthetic Rainfall Time Series in High Temporal Resolution - A Data Driven Approach von Ferdinand Beck. Institut für Wasser- und Umweltsystemmodellierung, Universität Stuttgart. - Stuttgart: Institut für Wasser- und Umweltsystemmodellierung, 2013

(Mitteilungen Institut für Wasser- und Umweltsystemmodellierung, Universität Stuttgart: H. 219)
Zugl.: Stuttgart, Univ., Diss., 2013
ISBN 978-3-942036-23-8

NE: Institut für Wasser- und Umweltsystemmodellierung <Stuttgart>: Mitteilungen

Herausgegeben 2013 vom Eigenverlag des Instituts für Wasser- und Umwelt-systemmodellierung

Druck: Document Center S. Kästl, Ostfildern

Danksagung

Mein herzlichster Dank gilt Herrn Professor Bárdossy; nicht nur für die Möglichkeit an seinem Lehrstuhl mitzuarbeiten, sondern vor allem für die vielen Ideen und sein großes hydrologisches und statistisches Wissen, an dem er mich über die Jahre hat teilhaben lassen. Er hat mir die Faszination der Statistik näher gebracht, was meine Sicht auf die Welt mit prägen wird.

Genauso möchte ich Herrn Professor Haberlandt herzlich dafür danken, dass er den Mitbericht übernommen hat; für seine sehr freundliche, hilfsbereite und unkomplizierte Betreuung.

Vielen Dank an Herrn Professor Helmig für seine umgängliche Art als Leiter der Prüfungskommission und die sehr angenehme Atmosphäre während der Prüfung.

In meinen Jahren am Institut für Wasserbau habe ich das Arbeitsklima dort immer als sehr freundschaftlich und fröhlich empfunden. Ich bin jeden Tag gerne dort hingegangen und dafür möchte ich den Kollegen in der Abteilung für Hydrologie und Geohydrologie herzlich danken, besonders den Mitstatistikern Dirk und Philipp für interessante und fordernde Diskussionen und, ganz wichtig, Thomas dafür, dass er mir "richtig" programmieren beigebracht hat.

Vielen Dank meinen Eltern, dafür, dass es sie gibt, dass es mich gibt und dass sie mir meine schulische Ausbildung und mein Studium ermöglicht haben. Nicht zuletzt danke ich meinen Brüdern und meinen Freunden, für die Zerstreuung, den Spaß und manchmal auch das offene Ohr, wenn es doch mal zu viel wurde mit der Statistik.

Contents

List of Figures V

List of Tables IX

List of Abbreviations XI

Notation XII

Zusammenfassung XV

1. Introduction 1
1.1. Context and Motivation . . . 1
1.2. Outline . . . 3

2. Characterization of the Hydrological Conditions in Baden-Württemberg 5
2.1. Context in the Global Atmospheric Circulation . . . 5
2.2. Regional Conditions in Baden-Württemberg . . . 10

3. Acquisition of Precipitation Data 12
3.1. Point Measurements . . . 12
3.1.1. Rain Gauge after Hellmann . . . 12
3.1.2. Disdrometer . . . 13
3.1.3. Error Sources . . . 14
3.1.3.1. Examples of Recording and Data Transmission Errors . . . 14
3.1.3.2. Some Aspects of Error Identification . . . 16
3.2. Remote Sensing . . . 18
3.2.1. Radar Measurements . . . 18
3.2.2. Satellite Remote Sensing . . . 22
3.3. The Data Set of Precipitation Measurements . . . 23
3.3.1. Subsets . . . 23

4. Statistical Properties of Precipitation 27
4.1. Spatial and Temporal Heterogeneity . . . 27
4.2. Intermittency and Scaling . . . 30
4.3. Consequences . . . 31

5. Modeling of Precipitation 33
5.1. Purpose of Precipitation Modeling . . . 34
5.2. Spatial Interpolation of Precipitation Data . . . 35

5.3. Markov Chain Models . . . 38
5.4. Point Process Models . . . 42
5.5. Alternating Renewal Models . . . 47
5.6. Scaling Models . . . 48
5.6.1. Autoregressive Models . . . 53

6. Generation of Point Rainfall **55**
6.1. The Synthetic Rainfall Time Series Generator NiedSim . . . 55
6.1.1. Necessary Adaptations to NiedSim . . . 56
6.2. Statistical Parameters of Hourly Precipitation Time Series . . . 58
6.3. Methods for the Statistical Description of Distributions . . . 59
6.3.1. Weibull Distribution . . . 60
6.3.2. Log-Normal Distribution . . . 61
6.3.3. Pareto Distribution . . . 62
6.3.4. Beta Distribution . . . 63
6.3.5. Comparison of Theoretical Distributions . . . 63
6.3.6. Maximum Likelihood Method for the Fit of Distribution Functions . . . 64
6.4. Methods for the Description of Multivariate Dependencies . . . 64
6.4.1. Covariance and Correlation . . . 64
6.4.2. Spearman's Rank Correlation . . . 65
6.4.3. Copulas as a Measure of Dependence . . . 66
6.4.3.1. Empirical Copulas . . . 67
6.4.3.2. Theoretical Copulas . . . 68
6.4.3.3. The Gaussian Copula . . . 69
6.5. Analysis of Observed Precipitation Data . . . 70
6.5.1. Multivariate Dependence in Observed Data . . . 70
6.5.1.1. Gaussian Copula Dependence used for Rainfall Generation . . . 75
6.5.2. Observed Marginal Distributions . . . 75
6.5.2.1. Marginal Distributions used for Rainfall Generation . . . 81
6.6. The Generation Algorithm for 1h Precipitation Time Series . . . 83
6.6.1. Parameter Estimation for the Initial Time Series . . . 83
6.6.2. Rainfall Generation . . . 84
6.6.2.1. Statistical Parameters in the Objective Function . . . 86
6.7. Comparison of Simulated and Observed Rainfall Time Series . . . 88
6.8. Conclusions . . . 93

7. Spatial Generation **94**
7.1. A High Dimensional Problem . . . 94
7.2. Generation Principle . . . 95
7.3. Observed Multivariate Dependencies in Daily Precipitation Data . . . 96
7.4. Marginal Distributions of Daily Precipitation Data . . . 99
7.5. Representation of Spatial Dependencies . . . 101
7.5.1. Observed Interdependencies of Statistical Parameters at Different Measurement Locations . . . 101
7.5.1.1. Analysis of Anisotropy . . . 104

7.5.1.2. Spatial Rank Correlations Used for the Generation of Simultaneous Daily Time Series . . . 107
7.5.2. Spatial Correlation According to Atmospheric Circulation Patterns . . 109
7.5.2.1. Definition of the Circulation Pattern Classification System . . 109
7.5.2.2. CP Dependent Precipitation Statistics . . . 111
7.5.2.3. Analysis of Spatial Correlation . . . 111
7.5.2.4. CP Dependent Correlations Used For the Generation of Simultaneous Time Series . . . 121
7.6. Generation Algorithm for Simultaneous 1h Precipitation Time Series . . . 122
7.6.1. Parameter Estimation for the Initial Daily Time Series . . . 122
7.6.2. Generation of Simultaneous Daily Precipitation Time Series . . . 124
7.6.3. Disaggregation to hourly values . . . 126
7.7. Comparison of Simulated and Observed Simultaneous Time Series . . . 127
7.7.1. Comparison of Observed and Simulated Distributions . . . 128
7.7.2. Representation of the Spatial Dependence . . . 138
7.8. Conclusion . . . 143

8. Some Aspects of Climate Change 146
8.1. Trend Signals in Observed Precipitation Time Series . . . 148
8.1.1. Regional Trend in Extreme Precipitation . . . 149
8.1.2. Change in Scaling . . . 153
8.2. Global and Regional Circulation Models . . . 157
8.2.1. Emission Scenarios . . . 158
8.2.2. Model Errors in Global and Regional Circulation Models . . . 159
8.2.3. Correction and Downscaling of GCM Precipitatioin Data . . . 160
8.2.3.1. Direct Correction of GCM Precipitation Data . . . 160
8.2.3.2. Indirect Correction by Additional Information . . . 163
8.3. Prediction of Future Precipitation by GCM Defined Circulation Patterns . . . 164
8.3.1. Changes in CP-Frequencies According to GCM and Reanalysis Data . 167
8.3.2. Temperature Sensitivity of CP-related Precipitation . . . 171
8.3.3. Expected Trend in 1h Precipitation Amounts due to Changes in Atmospheric Circulation . . . 175
8.4. Conclusions . . . 180
8.4.1. Reliability of Estimated Trend Signals . . . 180
8.4.2. Potential for Climate Change Adaptations in Rainfall Generation . . . 181
8.4.3. Future Predictions as an Ensembles . . . 182

9. Conclusions and Outlook 184

Bibliography 187

A. Generation of Point Rainfall 197
A.1. Marginal Distributions of All 292 Stations . . . 197
A.1.1. Monthly Precipitation Sum . . . 197
A.1.2. Average Precipitation in Wet Hours . . . 199

A.1.3. Standard Deviation of Precipitation Amounts in Wet Hours 201
A.1.4. Hourly Rainfall Probability 203
A.2. Drawing Conditioned Values from a Multivariate Normal Distribution 205

B. Spatial Generation 206
B.1. Observed Multivariate Dependencies in Daily Precipitation Data 206
B.2. Marginal Distribution in Daily Resolution 209
B.2.1. Average Precipitation in Wet Hours 209
B.2.2. Standard Deviation of Precipitation Amounts in Wet Hours 211
B.2.3. Daily Rainfall Probability 213
B.3. Representation of Spatial Dependencies 215

List of Figures

1. Beispiel empirischer Copula XIX
2. Empirische Copula Monatssumme/Regenwahrscheinlichkeit XIX
3. Jährliche Serie Holzgerlingen XX
4. Isobarenkarte, CP2 und CP3 XXII
5. Räumliche Korrelation CP2 XXIII
6. Räumliche Korrelation CP3 XXIV

2.1. Average number of rain days, montly sum Freiburg 8
2.2. Average number of rain days, montly sum Freiburg 8
2.3. Elevation and Yearly Precipitation Baden-Württemberg 9

3.1. Example of an error in a precipitation time series, probably occurred during data transmission 15
3.2. Example of an error in a precipitation time series, probably due to unreported calibration and maintenance 16
3.3. Example of an error in a precipitation time series, probably due to unreported calibration and maintenance (continued) 17
3.4. Hourly and daily precipitation stations 24
3.5. Subset of 137 high resolution rain gauges with continuous measurements from 1997 to 2003 25
3.6. Subset of 30 high resolution rain gauges with continuous measurements from 1991 to 2003 and low number of missing values. 26

4.1. Observed precipitation data: correlation vs. distance 27

5.1. Schematic for Point Process Rectangular Pulse Model 43

6.1. Principle of the permutation scheme in NiedSim 56
6.2. Empirical histogram Neckarsulm 57
6.3. Effect of wrong fit - Log-Normal Distribution 63
6.4. Gauss copula density 70
6.5. Monthwise empirical bivariate copula, monthly sum - 1 h rainfall probability 71
6.6. Monthwise empirical bivariate copula, monthly sum - average 1 h rainfall amount 73
6.7. Monthwise empirical bivariate copula, monthly sum - standard deviation of 1 h rainfall amount 73
6.8. Monthwise empirical bivariate copula, rainfall probability - average of 1 h rainfall amounts 74

6.9. Monthwise empirical bivariate copula, monthly sum - standard deviation of 1 h rainfall amount . . . 74
6.10. Monthwise empirical bivariate copula, average - standard deviation of 1 h rainfall amount . . . 75
6.11. Marginal distribution of monthly precipitation sum, Holzgerlingen . . . 77
6.12. Marginal distribution of average precipitation of wet hours, Holzgerlingen . 78
6.13. Log-Normal and Lomax Distribution moof nthly precipitation sum, Geislingen Waldhof . . . 79
6.14. Marginal distribution of standard deviation of wet hours, Holzgerlingen . . . 80
6.15. Marginal distribution of hourly rainfall frequency, Holzgerlingen . . . 81
6.16. Simulated monthly sum, Holzgerlingen . . . 88
6.17. Annual series Holzgerlingen . . . 90
6.18. Annual series - 1h January an July, Holzgerlingen . . . 91
6.19. Wet spells, Holzgerlingen . . . 91
6.20. 100 longest wet spells, Holzgerlingen . . . 92
6.21. Dry spells, Holzgerlingen . . . 92
6.22. 100 longest dry spells, Holzgerlingen . . . 93

7.1. Marginal distribution of average wet day precipitation, Schönaich . . . 99
7.2. Marginal distribution of standard deviation of wet day precipitation, Schönaich 99
7.3. Marginal distribution of rain day frequency, Schönaich . . . 100
7.4. Absolute frequency of the distances between 575 daily precipitation stations . 101
7.5. Correlation vs. distance – rain day frequency . . . 102
7.6. Correlation vs. distance – average vs. standard deviation of wet day precipitation . . . 103
7.7. Sketch of the sector division for rank correlation analysis . . . 105
7.8. Correlation vs. distance in different orientated sectors . . . 106
7.9. Principle of the fuzzy rule based CP classification . . . 110
7.10. Average pressure anomalies for CPs . . . 112
7.11. Direction dependent correlations – CP2 . . . 114
7.12. Direction dependent correlations – CP3 . . . 115
7.13. Direction dependent correlations – CP15 . . . 116
7.14. Standard deviation of correlation values CP2 . . . 118
7.15. Standard deviation of correlation values CP3 . . . 119
7.16. Standard deviation of correlation values CP15 . . . 120
7.17. Principle of the Kriging interpolation of CP dependent correlations . . . 121
7.18. Matrix of the parameter correlations in the standard normal transformed space for the generation of simultaneous time series . . . 123
7.19. Overview of the target location for the generation of simultaneous time series 128
7.20. Annual Series – multi generation Holzgerlingen 1 . . . 131
7.21. Annual Series – multi generation Holzgerlingen 1 . . . 132
7.22. Annual Series – multi-generation, Holzgerlingen January and July . . . 133
7.23. Wet spells, multi-generation Holzgerlingen . . . 134
7.24. 100 longest wet spells, multi-generation Holzgerlingen . . . 135
7.25. Dry spells, multi-generation Holzgerlingen . . . 136

7.26. 100 longest dry spells, multi-generation Holzgerlingen 137

8.1. Regional trend in extreme precipitation, Baden-Württemberg 1991 to 2003 . . 150
8.2. Regional trend in extreme precipitation, Bavaria 1990 to 2006 152
8.3. Average scaling 1985-1980 to 1981-2003 154
8.4. Principle of the change in scaling exponent 155
8.5. Map of change in scale exponent 156
8.6. Scenarios for Green House Gas (GHG) emissions from 2000 to 2100 (Nakicenovic and Swart, 2000) 158
8.7. Example of QQ-Transformation 162
8.8. Principle of the objective function for the CP definition in the simulated annealing scheme. 165
8.9. MSLP of CPs leading to high precipitation intensites 166
8.10. Exceedance frequency of the 95% and 99% quantile for two seasons. Quantiles and frequencies are calculated for all hours with $H \geq 0.1\,\mathrm{mm}$ 167
8.11. Absolute frequency of anticyclonic CPs (CP2, CP3, CP5 and CP8) according to NCEP/NCAR reanalysis during two seasons 168
8.12. Absolute frequency of cyclonic CPs (CP7, CP10, CP11 and CP12) according to NCEP/NCAR reanalysis during two seasons 168
8.13. Trends in frequency of anticyclonic CPs 169
8.14. Trends in frequency of cyclonic CPs 170
8.15. MSLP of CPs leading to high precipitation intensities - ECHAM5 171
8.16. Precipitation frequency and extreme value frequency vs. temperature – CP2 . 172
8.17. Precipitation frequency and extreme value frequency vs. temperature – CP7 . 173
8.18. Precipitation frequency and extreme value frequency vs. temperature – CP11 174
8.19. Expected statistics 1961 to 2060 – ECHAM5 176
8.20. Linear trend in precipitation statistics –ECHAM5 178
8.21. Linear trend in precipitation statistics –ECHAM5 May to August 179

A.1. Histogram and Weibull PDF - monthly precipitation sum 197
A.2. Linearized Weibull fit - monthly precipitation sum 198
A.3. Histogram and Log-Normal PDF - average wet hour precipitation 199
A.4. Linearized Log-Normal fit - average wet hour precipitation 200
A.5. Histogram and Log-Normal PDF - standard deviation of wet hour precipitation 201
A.6. Linearized Log-Normal fit - standard deviation of wet hour precipitation . . . 202
A.7. Histogram and Beta PDF - hourly rainfall probability 203
A.8. Beta CDF - hourly rainfall probability 204

B.1. Monthwise empirical bivariate copula, monthly sum - rainfall probability; daily values 206
B.2. Monthwise empirical bivariate copula, monthly sum - average wet day precipitation 206
B.3. Monthwise empirical bivariate copula, monthly sum - standard deviation of wet day precipitation 207

B.4. Monthwise empirical bivariate copula, rainfall frequency - average wet day precipitation . . . 207
B.5. Monthwise empirical bivariate copula, rainfall frequency - standard deviation of wet day precipitation . . . 208
B.6. Monthwise empirical bivariate copula, average - standard deviation of wet day precipitation . . . 208
B.7. histogram and Log-Normal PDF - average wet day precipitation . . . 209
B.8. linearized Log-Normal fit -average wet day precipitation . . . 210
B.9. histogram and Log-Normal PDF - standard deviation of wet day precipitation 211
B.10. linearized Log-Normal fit - standard deviation of wet day precipitation . . . 212
B.11. histogram and Beta PDF - daily rainfall frequency . . . 213
B.12. Beta CDF - daily rainfall frequency . . . 214
B.13. Correlation vs. distance – monthly sum . . . 215
B.14. Correlation vs. distance – average wet day precipitation . . . 215
B.15. Correlation vs. distance – standard deviation of wet day precipitation . . . 216
B.16. Correlation vs. distance – monthly sum vs. rain day frequency . . . 216
B.17. Correlation vs. distance – monthly sum vs. average precipitation . . . 217
B.18. Correlation vs. distance – monthly sum vs. standard deviation of wet day precipitation . . . 217
B.19. Correlation vs. distance – rain day frequency vs. average wet day precipitation 218
B.20. Correlation vs. distance – rain day frequency vs. average wet day precipitation 218

List of Tables

6.1. Rank correlation and symmetry monthwise 72
6.2. Correlation hourly precipitation statistics 76
6.3. Distribution functions used in the parameter estimation for the rainfall generation 81
6.4. Basic statistics simulation for Holzgerlingen 89

7.1. Rank correlation and symmetry monthwise - daily values 97
7.2. Correlations in Standard Normal transformed daily precipitation statistics . . 98
7.3. Distribution functions used in the parameter estimation for the generation of simultaneous daily rainfall time series 101
7.4. Parameters of the regression functions according to Eq. (7.5) for the estimation of parameter rank correlation over distance - same parameter at both stations 107
7.5. Parameters of the regression functions according to Eq. (7.5) for the estimation of parameter rank correlation over distance - combination of different parameters 108
7.6. Basic daily precipitation statistic for all anticyclonic, dry CPs 113
7.7. Basic daily precipitation statistic for all moderate humid CPs 113
7.8. Basic daily precipitation statistic for all cyclonic humid CPs 113
7.9. Coordinates - simultaneous simulation 127
7.10. Basic hourly statistics Holzgerlingen - multi site generation 129
7.11. Basic daily statistics Holzgerlingen - multi site generation 129
7.12. Daily correlation – simulations 138
7.13. Daily correlation – target values 138
7.14. Daily correlation – simulations 139
7.15. Daily correlation – target values 139
7.16. hourly correlation - simulations 139
7.17. hourly correlation w. time shift - simulations 140
7.18. hourly correlation w. time shift - observed 140
7.19. hourly correlation with time shift - simulations with high distance 140
7.20. CP-dependent correlation, hourly values 141
7.21. CP-dependent correlation, target values 142
7.22. CP-dependent correlation, hourly values with time shift, Oct-Mar 143
7.23. CP-dependent correlation, hourly values with time shift, Sep-Apr 144

8.1. Significance test by Monte Carlo simulation of the regional trend signals in Baden-Württemberg from 1991 to 2003. 151

8.2. Significance test by Monte Carlo simulation of the regional trend signals in Bavaria from 1990 to 2006. 152
8.3. Significance test by Monte Carlo simulation of the regional trend signals in Bavaria from 1991 to 2003. 153

List of Abbreviations

APREGE	French Global Circulation Model
AR(n) model	Auto Regressive Model of order n
CDF	Cumulative Distribution Function
CP	Circulation Pattern for weather type classification
DDF	Depth-Duration-Frequency Curve
DOF	Degree of Fulfillment (of Fuzzy Logic rule)
DWD	Deutscher Wetterdienst, German meteorological service
ECHAM4 / EHAM5	Global Circulation Model of the Max Planck Institute Hamburg, Germany
GCM	Global Circulation Model
GK X	Easting in Gauss Krüger 3 Coordinates
GK Y	Northing in Gauss Krüger 3 Coordinates
HadCM3	Global Circulation Model of the Hadley Center for Climate Prediction and Research, UK
IPCC	Intergovernmental Panel on Climate Change of the UN
ITCZ	Intertropical Convergence Zone
KOSIM	KOntinuierliches Langzeit- SIMulationsmodell, hydraulic model for sewage system simulations
KOSTRA	KOordinierte STarkniederschlags-Regionalisierungs Auswertungen, official German statistics of extreme precipitation
Lat	Latitude
Lon	Longitude
LUBW	Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg, environmental authority
MLM	Maximum Likelihood Method
MSLP	Mean Sea Level Pressure
NCAR	National Center for Atmospheric Research of the United States
NCEP	National Centers for Environmental Prediction of the United States
QQ-Transformation	Quantile-Quantile Transformation
RCM	Regional Circulation Model
SST	Sea Surface Temperature
TMPA	TRMM Multiple Precipitation Analysis
TRMM	Tropical Rainfall Measurement Mission
WMO	World Meteorological Organization

Notation

The following table shows the significant symbols used in this work. Local notations are explained in the text.

Greek Letters:

α	significance of hypothesis test
α, β	parameters of Beta-distribution
γ	variogram
Δ	scale, aggregation interval
κ, λ	parameters of Weibull Distribution
λ	factor of scale change
λ_i	weight of the measurement at location i in interpolation
μ, σ	average and standard deviation of Normal distributed random variable
ρ_{xy}	Spearman's rank correlation between x and y
$\mathbf{\Sigma}$	covariance matrix
Φ	distribution function of the Standard Normal Distribution
Ω	support of a distribution function

Latin Letters:

h	scaling exponent
C	copula
c	copula density
$D\left(X\right)$	expectation of standard deviation of X
d_{ij}	distance between station pair i, j
d_0	reference distance in exponential regression
$E\left(X\right)$	expectation of X
$F(X)$	theoretical distribution function
$f(X)$	probability density
$F_n\left(x\right)$	observed non-exceedance frequency of x in a sample of size n
g	skewness
H	precipitation depth
H_{wet}	precipitation depth of wet time steps
h	relative frequency
K_q	statistical moment of order q
n	number of elements, e. g. in a time series
O	objective function
P	probability

$P_{1\text{h}}$	rainfall probability on hourly temporal resolution
$P_{24\text{h}}$	rain day probability
p	observed frequency
R, Z	radar reflectivity and calculated rain rate
$\mathbf{r}$	correlation matrix
r_{xy}	Pearson correlation between x and y
r_l	observed autocorrelation of lag l
$r\left(x\right), q\left(y\right)$	ranks of x, y
s	standard deviation
s_{wet}	observed standard deviation of wet time steps
s_{xy}	covariance
sym	symmetry of a bivariate copula
T	Simulated Annealing Temperature
T	time period
U, V	uniformly distributed random variables
$\mathbf{V}$	CP defining Fuzzy rule set
v_{max}	normalized maximum of x
w	cascade generator in fractal scaling models
X, Y	random variables
$\hat{X}$	best estimate for x, e. g. by regression
x	observed data (realization of X)
$\bar{x}$	average of x
Z	standard normal transformed random variable
z	realisation of Z

Zusammenfassung

Einleitung

Eine ausreichende Versorgung mit Trinkwasser ist die Grundlage alles menschlichen Lebens. Die Hauptquellen für Trinkwasser sind Grundwasserleiter oder Oberflächengewässer, die durch Niederschläge gespeist werden. Darum ist die natürliche Variabilität des Niederschlags eine ständige Bedrohung: Ein Mangel an Niederschlag führt zu Dürren und kann im schlimmsten Fall Hungersnöte auslösen. Ein Zuviel an Niederschlag führt zu Überschwemmungen, die nicht nur hohe Infrastruktur Schäden bringen, sondern regelmäßig zu Todesopfern führen. In den USA beispielsweise sterben jedes Jahr rund 100 Menschen bei Überschwemmungen (Ashley and Ashley, 2008). Darüber hinaus können intensive Niederschläge Ernten vernichten und führen zu Bodenerosion. Weitere Risiken sind wasserbürtige Krankheiten oder gefährliche Infektionen durch kontaminiertes Trinkwasser. Die zivilisatorische Entwicklung ist stark davon abhängig, diese Risiken beherrschbar zu machen. Größere Siedlungen sind ohne wasserbauliche Strukturen für eine stetige Versorgung mit Trinkwasser, für den Hochwasserschutz und für die Entsorgung von Abwasser nicht möglich.

Solche Strukturen, wie z. B. städtische Kanalnetze sind sehr langfristige und kostenintensive Investitionen. Die richtige Dimensionierung ist darum entscheidend. Überdimensionierung führt zu unnötigem Ressourceneinsatz, Unterdimensionierung erhöht das Versagensrisiko des Systems. Bei der Auslegung hydraulischer Strukturen wird deshalb versucht, den optimalen Kompromiss zwischen den Konstruktions- und Unterhaltskosten einerseits und den Schadenskosten durch nicht beherrschte Extremereignisse andererseits zu finden.

Heutzutage werden viele Dimensionierungsprobleme mit dem Einsatz hydraulischer oder hydrologischer Modelle, wie z. B. Niederschlagsabflussmodelle für Flusseinzugsgebiete gelöst. Die eingesetzten Modelle werden dabei kontinuierlich verbessert, indem immer mehr Prozesse des natürlichen Abflussregimes numerisch abgebildet werden. Daraus entsteht jedoch ein neues Problem: Je detaillierter das numerische Modell ist, desto höher ist dessen Bedarf an beobachteten Messdaten als Eingangsgrößen. Limitierender Faktor ist darum häufig das Datenangebot, so dass die existierenden Modelle gar nicht in voller Komplexität genutzt werden können.

Besonders deutlich wird diese Einschränkung in der Simulation von Kanalnetzsystemen. Kanalnetzsysteme reagieren sehr schnell auf Niederschlagereignisse mit Konzentrationszeiten von teilweise unter einer Stunde. Eine adäquate Simulation bedarf entsprechend hochaufgelöster Niederschlagszeitreihen als Eingangsdaten. Da die Reaktion eines Kanalnetzes von der Vorfüllung der Kanäle abhängt, reicht die Simulation eines einzelnen Niederschlagsereignisses nicht aus. Für eine korrekte Abschätzung von Rückstauwahrscheinlichkeiten, ist eine Langzweitsimulation von idealerweise mehreren Jahrzehnten notwendig.

Betrachtet man Niederschlag in hoher zeitlicher Auflösung, so ist die räumliche Variabilität sehr hoch. Eine beobachtete Zeitreihe von 1 Stundenwerten ist nur für eine sehr begrenzte Fläche wirklich repräsentativ. Selbst ein engmaschiges Niederschlagsmessnetz wie das in Baden-Württemberg kann deshalb die benötigten langen und hochaufgelösten Niederschlagszeitreihen nicht flächendeckend liefern.

Als Ersatz können statistische Modelle eingesetzt werden, um synthetische Zeitreihen zu erzeugen, die tatsächlich beobachtetem Niederschlag möglichst genau entsprechen.

Statistische Modelle zur Erzeugung von synthetischen Niederschlagszeitreihen

Es existiert eine Vielzahl statistischer Konzepte zur Erzeugung synthetischer Niederschlagszeitreihen. Die Konzeption der meisten Modelle basiert darauf, dass die als grundlegend angesehenen statistischen Eigenschaften des Niederschlags im Modell nachgebildet werden. Erfassen diese die (unbekannten) physikalischen Prozesse des Niederschlagsgeschehens möglichst genau, so die Annahme, werden auch nicht in der Modellkonzeption verwendete Eigenschaften des Niederschlags richtig abgebildet.

Markov-Ketten Modelle beispielsweise gehen vom Gedächtnis der Niederschlagszeitreihe als grundlegende Eigenschaft aus. Markov Ketten simulieren den Niederschlag in diskreten Zeitschritten. Das Auftreten von Regen zu einem Zeitschritt wird als Zufallsprozess verstanden, dessen Wahrscheinlichkeit vom Zustand (Regen / kein Regen) vorheriger Zeitschritte bestimmt wird. Die Niederschlagshöhe für die nassen Zeitschritte wird dabei meist separat modelliert (z. B. Katz, 1977). Die Anwendung von Markov Ketten Modellen auf zeitlich hochaufgelöste Zeitreihen ist sehr komplex, weil hierbei sehr viele vergangene Zeitschritte miteinbezogen werden müssen (Katz and Parlange, 1995). Dies führt zu einer hohen Anzahl an Parametern und einer sehr datenintensiven Kalibrierung.

Punktprozess-Modelle simulieren den Niederschlag als eine Abfolge von zufällig eintreffenden niederschlagsaktiven Einheiten, welche hierarchisch organisiert sind: Die größte Einheit bilden “Stürme”, in denen eine zufällige Anzahl von niederschlagsaktiven “Zellen“ eingelagert sind (Rodriguez-Iturbe et al., 1987). In der Simulierung hochaufgelöster Zeitreihen können die Zellen noch in “Pulse” unterteilt werden (Cowpertwait et al., 2007). Da sich die Grundeinheiten des Modells überlagern können, sind sie in gemessenen Zeitreihen nicht direkt beobachtbar. Die Kalibrierung von Punktprozessmodellen erfolgt daher meist über numerische Methoden. Eine korrekte Parametrisierung ist noch immer Gegenstand aktueller Forschung (Vanhaute et al., 2012).

In Alternating-Renewal Modellen wird die Überlappung der Grundeinheiten nicht zugelassen, was die Kalibrierung vereinfacht. Die Zeitreihe besteht aus sich abwechselnden Nass- und Trockenzeiten, wobei in den Nasszeiten weiter in Regenimpulse unterteilt werden, zwischen denen kurze Regenunterbrechungen eingelagert sind (Haberlandt et al., 2008). Alternating-Renewal Modelle hängen stark von der präzisen Definition der Regenereignisse ab, die zur Kalibrierung herangezogen werden. Entscheidend ist vor allem die Frage, ab welcher Dauer der Unterbrechung zwei Regenereignisse als unabhängig angenommen werden können.

Auf hoher zeitlicher Auflösung gemessene Niederschlagszeitreihen haben, statistisch gesehen, eine sehr komplexe Struktur. Sie sind hochvariabel, jedoch nicht komplett zufällig. Stündliche Niederschlagszeitreihen beispielweise haben eine deutliche Persistenz. Wenn es in einer Stunde regnet, ist die Wahrscheinlichkeit für Regen auch in der nächsten Stunde erhöht. Gleichzeitig treten aber abrupte Wechsel auf zwischen längeren nassen und trockenen Phasen. Dieses Verhalten ist mit den üblichen statistischen Maßzahlen, wie z. B. der Autokorrelation, schwer zu beschreiben. Die verschiedenen niederschlagserzeugenden Prozesse treten alle in typischen räumlichen und zeitlichen Skalen auf. Gewitterzellen beispielsweise sind generell kleinräumige Phänomene und in ihrer Lebensdauer auf wenige Stunden begrenzt, während sich die Fronten eines Tiefdrucksystems über mehrere Tausend Quadratkilometer erstrecken und entsprechend länger anhaltenden und gleichförmigeren Niederschlag erzeugen können. Durch das Zusammenspiel der unterschiedlichen Prozesse zeigen Niederschlagszeitreihen charakteristische Skalierungseigenschaften zwischen den zeitlichen Aggregierungen, z. B. zwischen den statistischen Momenten von Stunden-, 6h- und Tageswerten.

Konzeptionelle Modelle können die komplexe statistische Struktur nur begrenzt nachbilden. Die korrekte Abbildung der Skalierungseigenschaften ist in den meisten Modellen nur mit sehr hohem Parametrisierungsaufwand möglich (Koutsoyiannis, 2006). Es besteht die Gefahr, dass ein Modell auf die zeitliche Skala angepasst ist, die zur Kalibrierung verwendet wurde und die Variabilität auf anderen Aggregierungen unterschätzt (Koutsoyiannis and Foufoula-Georgiou, 1993; Katz and Parlange, 1995).

Ein datenbasierter Ansatz zur Erzeugung von synthetischen Niederschlagszeitreihen in stündlicher Auflösung

In dieser Arbeit wurde ein datenbasierter Ansatz gewählt, wie er in Bárdossy (1998) beschrieben ist. Dieser Ansatz macht möglichst wenig konzeptionelle Einschränkungen, lediglich die Verteilung der stündlichen Niederschlagswerte wird vorgegeben. Außerdem werden statistische Zielvorgaben an eine Zeitreihe formuliert (z. B. der Autokorrelation oder der Skalierungseigenschaften auf unterschiedlichen Dauerstufen), die aus beobachteten Zeitreihen abgeleitet werden. Anhand der Verteilung wird eine zufällige Zeitreihe initialisiert, die anschließend durch Tauschen von Werten so lange optimiert wird, bis sie die statistischen Zielvorgaben möglichst weitgehend erfüllt.

Der Ansatz ist sehr datenintensiv, weil die statistischen Vorgaben von beobachteten Niederschlagszeitreihen abgeleitet werden müssen. Umgekehrt ist er für die Simulation in Regionen mit einem engmaschigen Messnetz, wie das Projektgebiet von Baden-Württemberg, besonders geeignet, denn damit wird die Information aus den vorhandenen Messwerten optimal genutzt. Ein weiterer Vorteil liegt in der einfachen Regionalisierung. Mit mit geostatistischen Methoden können die Zielstatistiken auf jeden beliebigen Raumpunkt des Projektgebiets übertragen werden. Genauso kann eine Konditionierung auf höher aggregierte Werte erfolgen, wenn diese als Zielwerte der Optimierung vorgegeben werden. Deshalb eignet sich die Methode auch für Downscaling, z. B. von Klimamodelldaten.

Bezüglich des Ansatzes von Bárdossy (1998) wurde der Algorithmus vor allem in Bezug auf die Generierung der Initialreihe modifiziert, die an das Optimierungsschema überge-

ben wird. Die Initialreihe wird durch die Niederschlagswahrscheinlichkeit definiert, welche die Anzahl an auftretenden Regenstunden steuert, sowie durch die Parameter der theoretischen Verteilungsfunktion, welcher die stündlichen Niederschlagssummen folgen. Als Verteilungsfunktion wurde in dieser Arbeit die Weibullverteilung gewählt. Die Wahrscheinlichkeitsdichte f und die theoretische Verteilungsfunktion F sind durch folgende Gleichungen gegeben:

$$f(X) = \frac{\kappa}{\lambda}\left(\frac{X}{\lambda}\right)^{\kappa-1} e^{-(x/\lambda)^{\kappa}} \quad (1)$$

$$F(X) = 1 - e^{-(x/\lambda)^{\kappa}} \quad (2)$$

Die richtige Anpassung der Verteilungsfunktion ist entscheidend für die Güte der generierten Zeitreihen. Aufgrund der hohen Schiefe stündlicher Niederschlagssummen wird die Anpassung der Verteilungsparameter λ und κ durch die hohe Anzahl sehr niedriger Werte dominiert. Dadurch besteht die Gefahr, dass die Wahrscheinlichkeit extrem hoher Werte unterschätzt wird. Um die Variabilität in den Extremwerten zu erhöhen und um jahreszeitliche Effekte mit einzubeziehen, wurden die 1h-Regenwahrscheinlichkeit und die Parameter λ und κ monatsweise bestimmt.

Wie in dieser Arbeit gezeigt werden konnte, hängen die für die Erzeugung der Initialreihe maßgeblichen Statistiken der stündlichen Regenwahrscheinlichkeit $P_{1\mathrm{h}}$, der mittleren Niederschlagssumme aller nassen Stunden $E(H_{wet})$ und der Standardabweichung der nassen Stunden $D(H_{wet})$ alle von der Monatssumme H_{mon} ab. Der Zusammenhang ist universal für das gesamte Projektgebiet und kann mit einer vierdimensionalen Gaussschen Copula beschrieben werden.

Die Copula misst den Zusammenhang der Variablen unabhängig von den absoluten Werten. Es ist die multivariate Verteilungsfunktion im $[0,1]$-Raum der gleichverteilt transformierten Variablen U_i:

$$C : [0,1]^{n_{dim}} \rightarrow [0,1]$$

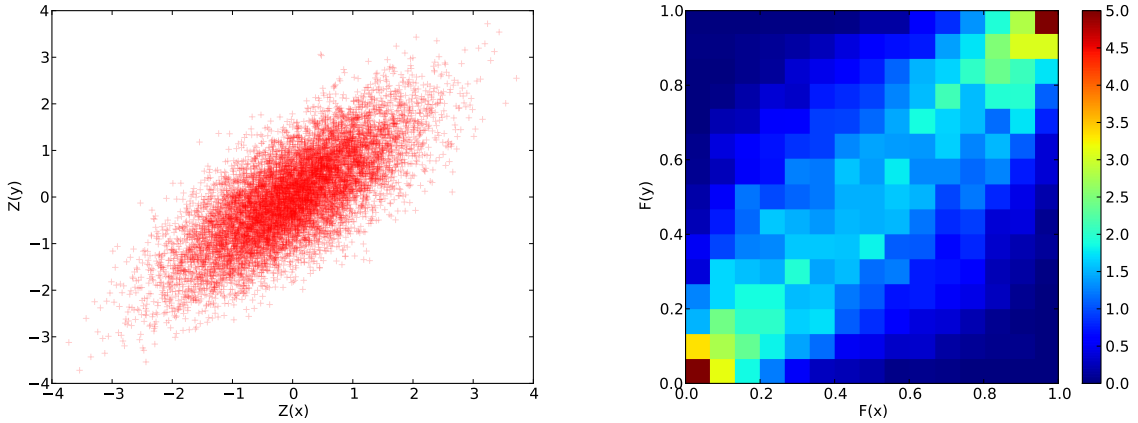
Die Copula beschreibt die multivariate Nicht-Überschreitungswahrscheinlichkeit aller n_{dim} Variablen:

$$C\left(U_1^*, U_2^*, \ldots, U_{n_{dim}}^*\right) = F\left(U_1 \leq U_1^*, U_2 \leq U_2^*, \ldots, U_{n_{dim}} \leq U_{n_{dim}}^*\right) \quad (3)$$

Die Gaussche Copula ist identisch mit einer multivariaten Normalverteilung, wenn alle Variablen U_i in standardnormalverteilte Zufallsvariablen transformiert werden:

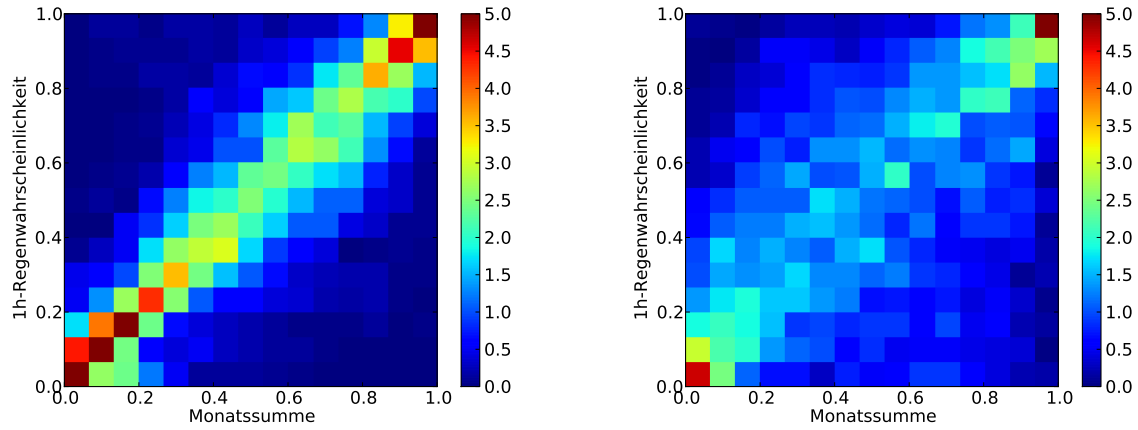
$$C(U_1, U2, \ldots, U_{n_{dim}}) = \Phi_{\mathbf{r}}\left(\Phi^{-1}(U_1), \Phi^{-1}(U_2), \ldots, \Phi^{-1}(U_{n_{dim}})\right) \quad (4)$$

Hierbei is Φ^{-1} die Umkehrfunktion der Standardnormalverteilung und $\Phi_{\mathbf{r}}$ die multivariate Standardnormalverteilung. Der einzige Parameter einer Gausschen Copula ist die Matrix der paarweisen Korrelationen $\mathbf{r}$ zwischen allen Variablenkombinationen im standardnormal-transformierten Raum. Abb. 1 zeigt die empirische Copuladichte zwischen 10000 Zufallszahlen, die einer bivariaten Gaussschen Copula mit einer Korrelation r im standardnormal-transformierten Raum von $r = 0.75$ folgen.



(a) Scatterplot (b) empirische Copuladichte

Abb. 1.: Beispiel einer bivariaten empirischen Copula von 10000 standardnormalverteilten Zufallszahlen mit $r = 0.75$. Links: Skatterplot im standardnormalen Raum, rechts: Empiriche Copuladichte



(a) Januar (b) July

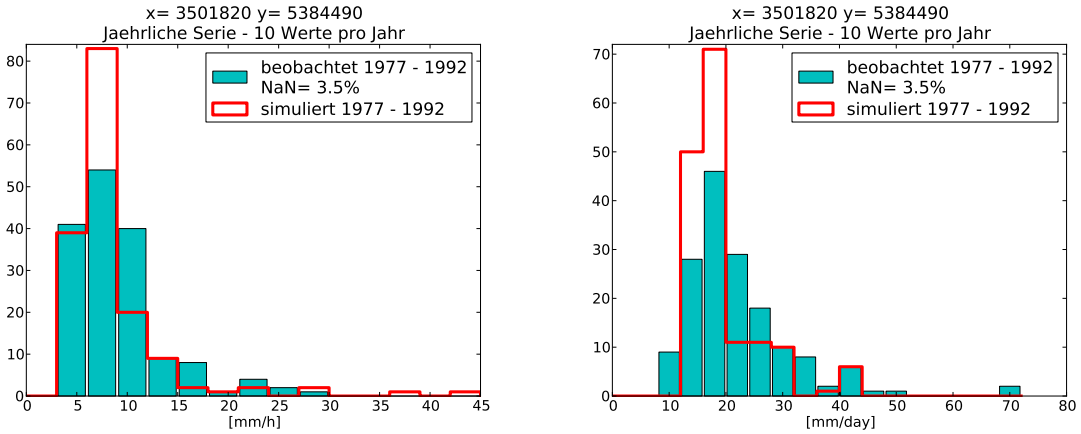
Abb. 2.: Empirische Copuladichte zwischen der Monatssumme und der 1h-Niederschlagswahrscheinlichkeit

Als Beispiel für den vierdimensionalen Zusammenhang zwischen den Parametern ist in Abb. 2 die empirische Copuladichte zwischen der Monatssumme des Niederschlags und der Niederschlagswahrscheinlichkeit dargestellt. Der Zusammenhang zeigt die für die Gaussche Copula charakteristische elliptische Form der Copuladichte. Generell ist der Zusammenhang im Sommer niedriger, weil kleinräumige konvektive Niederschlagsereignisse auftreten.

Anhand der vierdimensionalen Copula können zufällige Werte für P_0, $E\left(H_{wet}\right)$ und $D\left(H_{wet}\right)$ gezogen werden, die sowohl untereinander, als auch zur Monatssumme die rich-

tige Abhängigkeitsstruktur zeigen. Dafür wird der jeweils aktuelle Messwert der Monatssumme über die Randverteilung in einen standardnormalverteilten Wert Z_{mon} umgerechnet und die vierdimensionale Korrelationsmatrix $\mathbf{r}$ auf diesen Wert konditioniert. Über die Standardnormalverteilung und die Umkehrung der Randverteilungen für P_0, $E\left(H_{wet}\right)$ und $D\left(H_{wet}\right)$ werden die im standardnormal-transformierten Raum gezogenen Werte Z_i wieder in absolute Werte X_i zurück gerechnet. Um jahreszeitliche Effekte zu berücksichtigen, wurden die vierdimensionale Copula und die Randverteilungen jeweils für dreimonatige Zeiträume aufgestellt. Als Randverteilung für die 1h-Niederschlagswahrscheinlichkeit wurde die Beta-Verteilung gewählt. Für die anderen Größen wurden mehrere rechtsschiefe Verteilungen getestet. Die Monatssumme wird am besten mit der Weibullverteilung beschrieben, die Momente der nassen Stunden $E\left(H_{wet}\right)$ und $D\left(H_{wet}\right)$ jeweils mit der Log-Normalverteilung. Alle Verteilungen wurden mit der Maximum-Likelihood-Methode angepasst.

Aus den gezogenen Werten $E\left(H_{wet}\right)$ und $D\left(H_{wet}\right)$ für jeden Monat werden die Parameter λ und κ der Weibullverteilung nach der Momentenmethode berechnet. Anschließend wird anhand von P_0 und den Weibullparametern die Initialzeitreihe Monat für Monat generiert. Die vollständig initialisierte Zeitreihe für ein Jahr wird dann dem Optimierungsschema übergeben.



(a) Stundensummen

(b) Tagessummen

Abb. 3.: Empirische Verteilung der zehn höchsten Niederschlagssummen in der beobachteten und simulierten Zeitreihe von Holzgerlingen

Als Test der Modellgüte wurden die erzeugten Zeitreihen mit beobachteten Messreihen in Holzgerlingen verglichen, wo lange Datenreihen mit geringer Fehlwertquote vorliegen. Es zeigte sich, dass die monatsweise Anpassung der Verteilungsfunktion zu guten Ergebnissen führt. Sowohl in den grundlegenden Statistiken (Regenwahrscheinlichkeit, Mittelwert und Schiefe) als auch in den Extremwerten auf unterschiedlichen Aggregierungen (Abb. 3) stimmen die generierten Zeitreihen gut mit den beobachteten überein. Gleiches gilt für die Verteilung der Längen von Nass- und Trockenphasen. Eine Stärke des Generierungsalgorithmus im Hinblick auf die Regionalisierung ist die Konditionierung der Zeitreihen auf die Monatssumme des Niederschlags. Die Monatssumme ist räumlich weit weniger varia-

bel als Statistiken auf höherer zeitlicher Auflösung. Aus diesem Grund ist die Interpolation auf Raumpunkte, an denen keine Messstationen vorliegen, mit einer geringen Unsicherheit verknüpft.

Die Erzeugung von simultanen, synthetischen Niederschlagszeitreihen

Für die hydraulische Simulation größerer Kanalnetze ist die Verwendung einer Niederschlagszeitreihe als Eingangsdaten nicht mehr ausreichend, da die räumliche Heterogenität über dem Einzugsgebiet zu groß wird. Sollen mehrere synthetische Zeitreihen verwendet werden, so muss die räumlich-zeitliche Abhängigkeit zwischen den Zeitreihen bei der Generierung berücksichtigt werden.

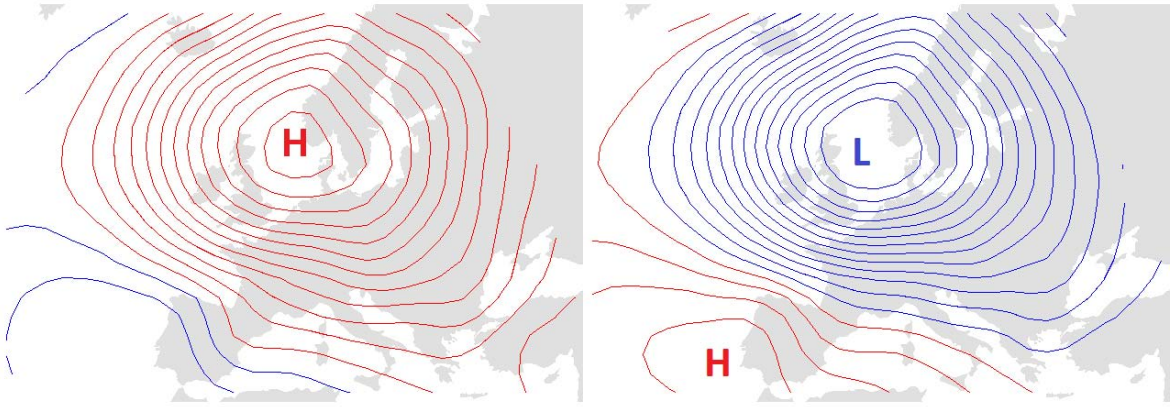
Der für die Generierung von Einzelzeitreihen entwickelte Algorithmus kann für die Erzeugung simultaner Zeitreihen nicht übernommen werden. Der Rechenaufwand des Optimierungsschemas ist abhängig von der Anzahl an möglichen Kombinationen in der zu optimierenden Zeitreihe. Bei der Erzeugung simultaner Zeitreihen steigt dieser überproportional an, da die möglichen Kombinationen der 1h-Niederschlagswerte aller Zeitreihen berücksichtigt werden müssen. Dadurch wird das ursprüngliche Generierungsschema schon bei wenigen Stationen impraktikabel. Es wurde deshalb ein neuer Ansatz entwickelt, bei dem zuerst simultane, räumlich-zeitlich abhängige Zeitreihen von Tageswerten erzeugt werden. Die Tageswerte werden dann in einem zweiten Schritt auf 24 Stundenwerte verteilt. Diese Disaggregierung funktioniert ebenfalls über eine Optimierung, bei der kleine Inkremente der Niederschlagssummen zwischen den Stunden eines Tages ausgetauscht werden. Wie bisher auch wird der Zustand der zu optimierenden Zeitreihen mit einer Zielfunktion bewertet. Neben den statistischen Zielwerten für die Einzelreihen werden dabei auch Vorgaben bezüglich des räumlichen Zusammenhangs gemacht.

Analog zur Generierung von Einzelreihen wird die Initialreihe auf die Monatssumme konditioniert. Im Unterschied zur Einzelgenerierung hat die Zeitreihe nun jedoch eine zeitliche Auflösung von 24 h. Die Parameter zur Generierung der Initialreihe sind somit die tägliche Niederschlagswahrscheinlichkeit $P_{24\text{h}}$, sowie der Mittelwert $E\left(H_{wet}\right)$ und die Standardabweichung $D\left(H_{wet}\right)$ des Niederschlags an nassen Tagen. Neben der Abhängigkeit der Generierungsparameter an jeder Station muss für eine simultane Generierung ebenso berücksichtigt werden, wie die Generierungsparameter verschiedener Stationen miteinander zusammen hängen, z. B. wie die Monatssumme an Station i mit der Regenwahrscheinlichkeit an Station j zusammen hängt. Unter Annahme eines Gausschen Copula Zusammenhangs ergibt sich dadurch eine Korrelationsmatrix $\mathbf{r}$ im standardnormal-transformierten Raum mit $4n_{stat} \times 4n_{stat}$ Einträgen, wobei n_{stat} für die Anzahl an Stationen in der simultanen Generierung steht. Die Korrelationswerte zwischen Parametern von verschiedenen Stationen $i \neq j$ werden mittels einer exponentiellen Regressionsfunktion über den Abstand zwischen den Stationen geschätzt.

Wie bei der Generierung von Einzelzeitreihen beschrieben, wird die Korrelationsmatrix $\mathbf{r}$ auf den Vektor der standardnormal-transformierten Monatssummen an allen Stationen konditioniert und es werden für jede Station i und jeden Monat zufällige Werte für die Niederschlagswahrscheinlichkeit $P^i_{24\text{h}}$ sowie die Momente $E^i\left(H_{wet}\right)$ und $D^i\left(H_{wet}\right)$ gezogen. Über

die Momente wird für jede Station und jeden Monat eine Weibullverteilung angepasst und die tägliche Initialreihe gezogen. Die Initialreihen werden wiederum über ein Permutationsschema optimiert, wobei neben den statistischen Zielwerten für die einzelnen Zeitreihen die Korrelation der Zeitreihen untereinander in die Zielfunktion eingeht.

Um die räumliche Abhängigkeit zwischen den Zeitreihen auf einer kürzeren zeitlichen Skala zu berücksichtigen, wird eine Wetterlagenklassifikation ("Circulation Pattern", im folgenden CP) herangezogen. Für die Definition der Wetterlagen wird auf Bárdossy (2010) verwiesen. Der Zielwert der Korrelation zu jeder Wetterlage für die Disaggregierung wird anhand des Abstandsvektors zwischen den Zielpunkten der Generierung festgelegt. Dabei wird nicht nur die Korrelation gleichzeitiger Werte betrachtet, sondern auch die Korrelation bei einstündigem Zeitversatz zwischen den Zeitreihen.

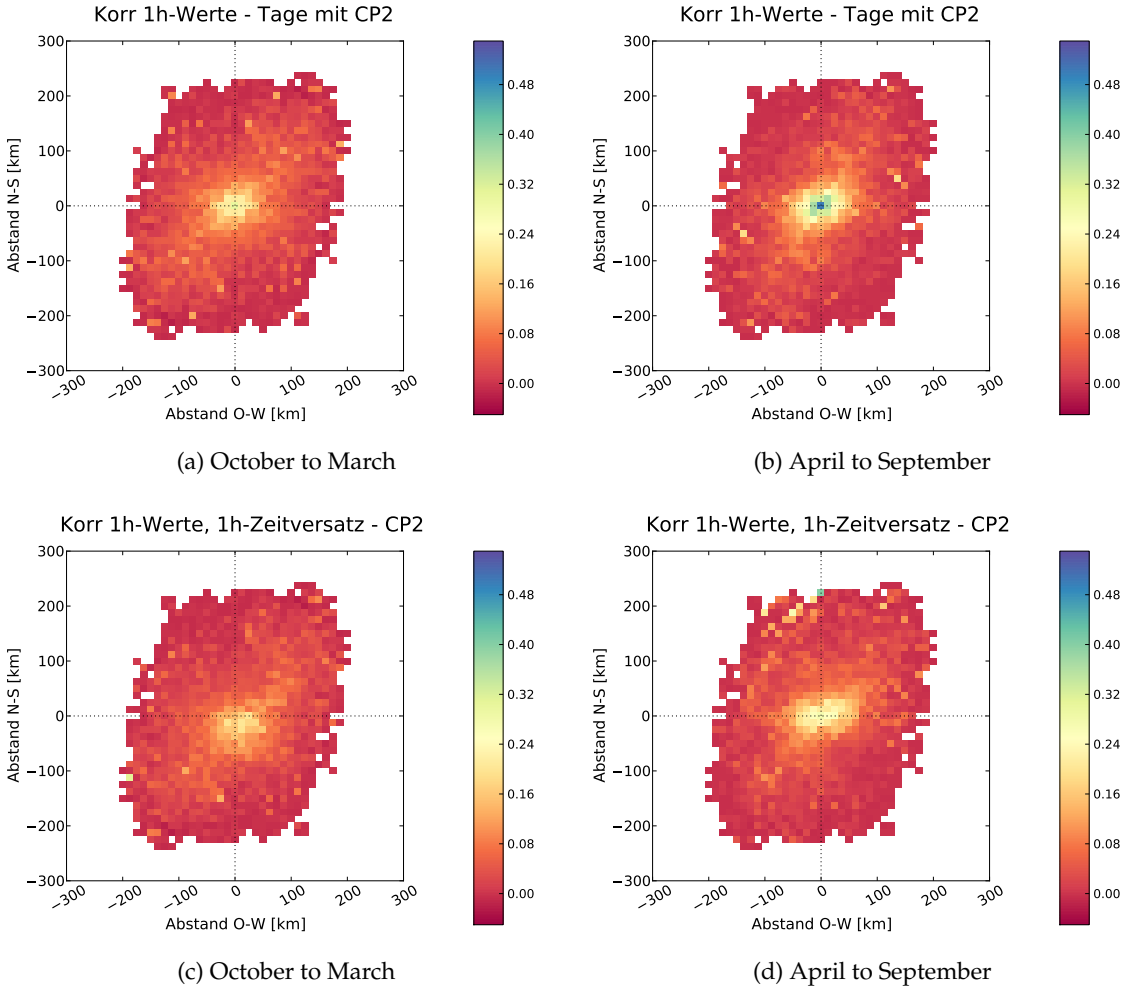


(a) CP2 (b) CP3

Abb. 4.: Durchschnittliche Luftdruckanomalie aller Tage mit CP2 (links) und CP3 (rechts) als 1 hPa Isobaren

Abb. 4 zeigt die mittlere Luftdruckanomalie als Isobarenkarte für zwei der 17 CPs. In Abb. 5 und in Abb. 6 sind für diese CPs die Korrelationen stündlicher Niederschlagswerte in Abhängigkeit des Abstandsvektors zwischen den Messstationen dargestellt. Jede Rasterzelle in den Diagrammen repräsentiert die durchschnittliche Korrelation aller Stationspaare, deren Abstandsvektor in die Rasterzelle fällt. Die absolute Lage der Stationen zueinander wird also ignoriert, entscheidend ist lediglich die räumliche Konfiguration der Stationen zueinander. Die Korrelation wurden auf Basis von 137 Niederschlagszeitreihen aus Baden-Württemberg berechnet.

Es zeigt sich, dass die räumliche Abhängigkeit stark von der CP beeinflusst wird. Während Hochdruckwetterlagen wie CP2, mit schwachem atmosphärischem Austausch, entstehen die meisten Niederschlagsereignisse lokal und die Korrelationen sind generell niedrig (Abb. 5). Während Tiefdrucklagen sind die Korrelationen viel höher und zeigen eine ausgeprägte Anisotropie. Die Korrelationen sind parallel zur atmosphärischen Strömungsrichtung, bei CP3 von West nach Ost, deutlich höher (Abb. 6). Wenn man die Zeitreihe am Ende des Abstandsvektors für die Berechnung der Korrelation um eine Stunde verschiebt (zweite



(a) October to March

(b) April to September

(c) October to March

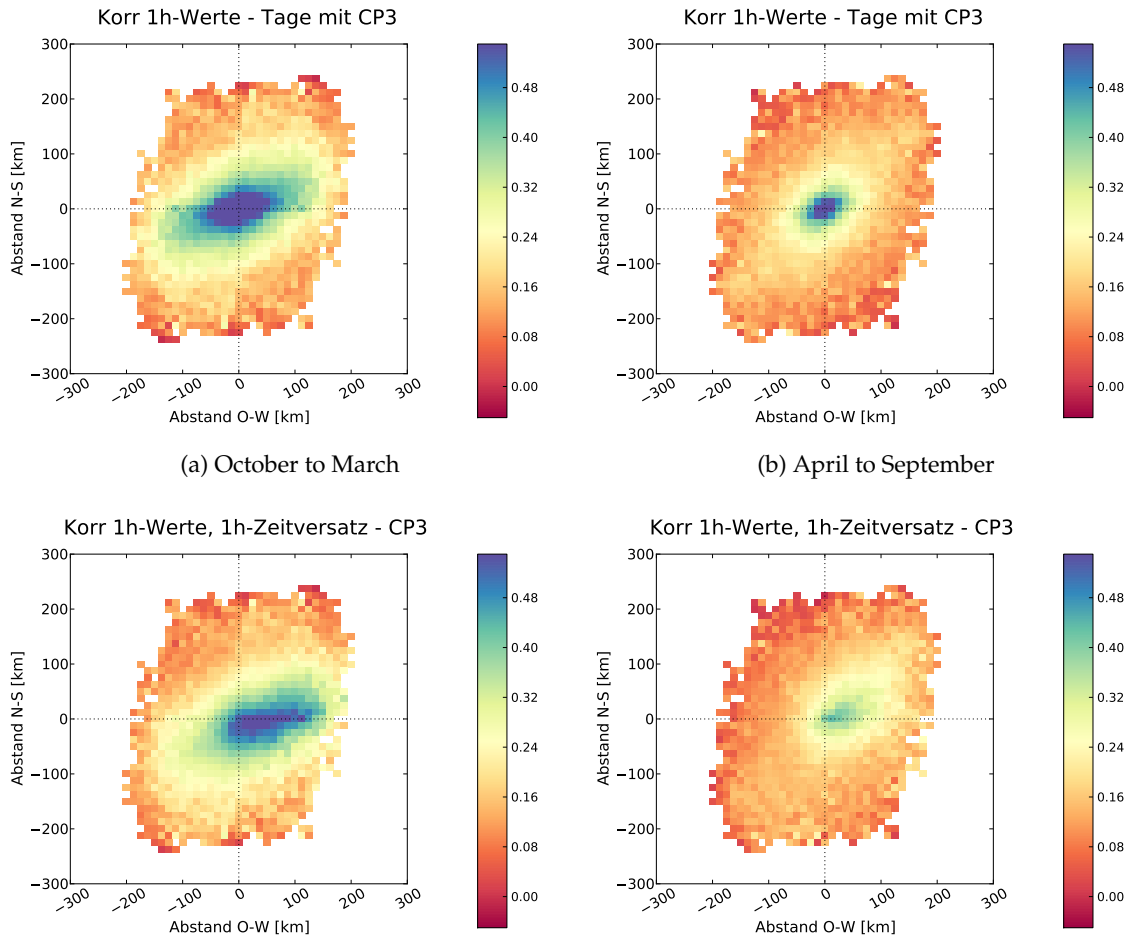
(d) April to September

Abb. 5.: Richtungsabhängige paarweise Korrelation von 137 1h-Niederschlagsstationen in Baden-Württemberg für alle Tage mit CP2; erste Reihe: Korrelation der gleichzeitigen Werte, zweite Reihe: Korrelation mit 1h Zeitversatz

Zeile von Abb. 5 und Abb. 6), zeigt sich bei den Tiefdrucklagen, dass die Korrelation in Strömungsrichtung höher ist als entgegen der Strömungsrichtung. Gleichzeitig gibt es einen deutlichen jahreszeitlichen Effekt. Durch die konvektiven Niederschlagsereignisse in den Sommermonaten sind die Korrelationen zwischen April und September generell niedriger als im advektiv geprägten Winterhalbjahr.

Die paarweisen Korrelationen der gleichzeitigen und versetzten Zeitreihen dienen neben den statistischen Vorgaben für die Einzelreihen als Zielwerte im Optimierungsalgorithmus der Disaggregierung.

Im Vergleich mit beobachteten Zeitreihen zeigt das Generierungsschema für simultane Zeitreihen vergleichbar gute Ergebnisse wie die Einzelreihengenerierung. Die Möglichkeit räumliche Zusammenhänge zu berücksichtigen, ist hingegen noch begrenzt. Den Copula-Zusammenhang zwischen den Generierungsparametern und die Korrelation der 1h-Werte



(a) October to March

(b) April to September

(c) October to March

(d) April to September

Abb. 6.: Richtungsabhängige paarweise Korrelation von 137 1h-Niederschlagsstationen in Baden-Württemberg für alle Tage mit CP3; erste Reihe: Korrelation der gleichzeitigen Werte, zweite Reihe: Korrelation mit 1h Zeitversatz

über den Abstand zwischen den Stationen zu schätzen, ist mit einer hohen Streuung verbunden. Es gibt jedoch keine Anzeichen, dass die Streuung auf Anistropien oder Instationaritäten zurückzuführen ist. Eine weitere Einschränkung liegt darin, dass nicht-lineare Abhängigkeiten, wie z. B. eine erhöhte Korrelation zwischen den extremsten Niederschlagsereignissen nicht in der Generierung berücksichtigt sind.

Auswirkungen des Klimawandels auf Niederschlag in hoher zeitlicher Auflösung

Werden die Daten der entwickelten Niederschlagsgeneratoren in der Planung wasserbaulicher Strukturen eingesetzt, so hängen davon sehr langfristige Investitionen ab. Bei der Dimensionierung städtischer Kanalnetze beispielsweise geht man davon aus, dass die generierten Zeitreihen repräsentativ für den Niederschlag während der gesamten Lebensdauer

von mehreren Jahrzehnten sind. Implizit steckt darin die Annahme, dass die klimatischen Bedingungen aus dem Kalibrierungszeitraum auch für die Zukunft gelten. Unter sich wandelndem Klima ist diese Annahme mehr und mehr fraglich, je weiter man in die Zukunft blickt.

Im zwanzigsten Jahrhundert stieg die jährliche Durchschnittstemperatur weltweit um etwa 0.8° C an. Ergebnisse von Klimamodellläufen lassen darauf schließen, dass sich dieser Temperaturanstieg in diesem Jahrhundert fortsetzen wird (IPCC, 2007). Da sowohl die Verdunstung, als auch die Kapazität der Atmosphäre Wasserdampf aufzunehmen temperaturabhängig sind, kann man davon ausgehen, dass dieser Temperaturanstieg weltweit auch den Niederschlag verändert (z. B. Wentz et al., 2007). Im regionalen Maßstab lassen sich solche Veränderungen schon beobachten. Hundecha and Bárdossy (2005) konnten anhand täglicher Niederschlagsmessungen der letzten Jahrzehnte im Rheineinzugsgebiet einen Anstieg der Niederschlagssumme im Winterhalbjahr nachweisen, gleichzeitig nahm der Niederschlag in den Sommermonaten an einigen Stationen leicht ab. Außerdem stieg der Anteil an der Niederschlagssumme, der während Extremereignissen fällt, deutlich an. Es zeigt sich auch, dass die Wirkung des klimatischen Wandels auf den Niederschlag von der betrachteten zeitlichen Skala abhängt. Bei höherer atmosphärischer Temperatur muss mit weniger häufigen, aber intensiveren Niederschlagsereignissen gerechnet werden (Trenberth, 1999). So kann es sein, dass es trotz allgemein trockeneren Sommern zu mehr Extremereignissen kommt (Christensen and Christensen, 2004).

In dieser Arbeit wurden deshalb Methoden entwickelt, mit denen mögliche Auswirkungen des Klimawandels auf den Niederschlag in den für viele wasserbauliche Fragestellungen entscheidenden kurzen Dauerstufen abgeschätzt werden können.

Eine Abschätzung der zukünftigen klimatischen Verhältnisse kann prinzipiell auf zwei Arten erfolgen, entweder durch eine Extrapolation beobachteter Niederschlagstrends oder durch den Einsatz globaler oder regionaler Klimamodelle. Die Trendfortschreibung basiert auf der Annahme, dass ein beobachteter Trend sich in der Zukunft auf die gleiche Weise fortsetzen wird. Es ist eine rein statistische Methode, die keinerlei atmosphärenphysikalische Zusammenhänge berücksichtigt, sondern lediglich davon ausgeht, dass Änderungen in den physikalischen Prozessen einer gewissen Trägheit unterliegen. Aufgrund dieser Annahme ist der Zeithorizont von Trendabschätzungen generell begrenzt. Eine Prognose für das Jahr 2100, wie sie von den Klimamodellen geliefert wird, ist damit nicht möglich. Dafür kann die hohe Detailliertheit beobachteter Niederschlagsdaten voll ausgenützt werden, welche von globalen oder regionalen Klimamodellen so nicht abgebildet werden kann.

Für eine regionale Trendanalyse wurden die jährlichen Extremwerte an jeder verfügbaren Niederschlagsstation in Baden-Württemberg auf das durchschnittliche jährliche Maximum normiert. Damit können alle Zeitreihen vergleichbar gemacht und gemeinsam analysiert werden, soweit die Daten für denselben Zeitraum vollständig vorliegen. Auf diese Weise wurden 40 stündliche Niederschlagsstationen und 104 Tagesstationen im Zeitraum von 1991 bis 2003 in einer regionalen linearen Regression zusammengefasst. In Baden-Württemberg konnte damit kein signifikanter Trend nachgewiesen werden. Eventuell sind die verfügbaren Zeitreihen dafür zu kurz. In Bayern jedoch, wo die Datenbasis mit 54 stündlichen Stationen und 293 Tagesstationen größer und der mögliche Untersuchungszeitraum von 1991

bis 2006 länger ist, zeigte sich eine signifikante Zunahme in den 1h-Extremwerten während des Sommerhalbjahrs.

Der Anstieg in den Niederschlagsintensitäten kurzer Dauer wurde durch eine Analyse der Skalierungseigenschaften zwischen Niederschlagssummen unterschiedlicher Dauerstufe bestätigt. Zwischen 1958 und 1980 betrug das durchschnittliche jährliche 1h-Niederschlagsmaximum in Baden-Württemberg 28% des täglichen Maximums. Im Zeitraum 1981 bis 2003 stieg dieser Anteil auf 31% an. Das 4h-Maximum stieg im gleichen Zeitraum von 50% des Tagesmaximums auf 55% an. Die Zunahme der kurzzeitigen Intensitäten betrifft nicht nur die Extremwerte, sondern wurde flächendeckend auch für den Mittelwert der Niederschlagssumme nachgewiesen. Die Ergebnisse der Untersuchungen entsprechen den aus der Atmosphärenphysik abgeleiteten Voraussagen von Trenberth (1999), dass bei einer höheren Atmosphärentemperatur mit weniger, aber intensiveren Niederschlägen zu rechnen ist.

Mit Klimamodellen, welche die physikalischen Prozesse in der Atmosphäre numerisch abbilden, ist eine Prognose für längere Zeithorizonte möglich als mit rein statistischen Methoden. Klimamodelle können jedoch den globalen Wasserkreislauf nicht fehlerfrei abbilden (Habemann et al., 2006) und es gibt Indizien, dass der Fehler in den prognostizierten Niederschlagssummen von der betrachteten räumlichen und zeitlichen Skala abhängt (Roeckner et al., 2006). Extreme Niederschlagsereignisse werden generell unterschätzt, wobei die Unterschätzung auf kürzerer zeitlicher Skala stärker wird (Fowler et al., 2007). Doch selbst wenn die Klimamodelle perfekte Vorhersagen liefern würden, könnten die Daten nicht direkt für wasserbauliche Aufgabenstellungen verwendet werden. Zwischen den Niederschlagswerten der Modelle und beobachteten Werten von Niederschlagsmessstationen besteht ein grundsätzlicher Unterschied. Klimamodelle berechnen den Gebietsniederschlag über einer Rasterzelle, während Messstationen Information von einem Raumpunkt liefern. Die beiden Größen würden nur dann übereinstimmen, wenn der Niederschlag in der Rasterzelle vollständig räumlich homogen wäre.

Klimamodelldaten müssen darum korrigiert und durch sogenanntes Downscaling auf die lokale Skala übertragen werden. Eine Möglichkeit ist die Quantil-Quantil-Transformation, mit der die Verteilung der beobachteten lokalen Niederschlagssummen auf die Klimamodelldaten aufgeprägt wird. In dieser Arbeit wurde jedoch gezeigt, dass eine solche direkte Korrektur der Niederschlagsdaten keine zufriedenstellenden Ergebnisse liefert. Selbst nach einer Korrektur ist die Bandbreite zwischen verschiedenen Klimamodellen höher als die Bandbreite unterschiedlicher Szenarios, die ja eigentlich die gesamte Unsicherheit der zukünftigen Entwicklung abdecken sollten. Zudem sind die meisten direkten Korrekturen nur für aggregierte Daten, wie z. B. die Monatssummen, anwendbar, weil zeitliche Abhängigkeiten zwischen den Werten nicht berücksichtigt werden können.

In dieser Arbeit wurde deshalb eine indirekte Korrekturmethode entwickelt, bei der die Nutzung von Niederschlagsdaten aus dem Klimamodell vermieden wird. Es werden lediglich Luftdruck- und Temperaturfelder verwendet, die als weitaus verlässlicher angesehen werden. Anhand von Luftrduckdaten aus den NCEP/NCAR Reanalysen (Kistler et al., 2001) wurde eine CP-Klassifikation vorgenommen, die bezüglich der durchschnittlichen Atmosphärentemperatur weiter unterteilt wurde. Mit den Daten der Messstationen in Baden-

Württemberg wurde die beobachtete Verteilung der 1h-Niederschlagssummen zu jeder CP-Temperatur-Kombination ermittelt. Im Zeitraum von 1958 bis 2003 haben diejenigen CPs, die für Hochdruckwetterlagen stehen, in den Sommermonaten zugenommen, während Tiefdruckwetterlagen seltener wurden. Gleichzeitig reagieren die CPs auf die atmosphärische Durchschnittstemperatur. Vor allem in den Sommermonaten führt eine höhere Temperatur zu einer niedrigeren Niederschlagswahrscheinlickeit. Wenn es aber regnet, steigt die Wahrscheinlichkeit von Extremereignissen deutlich an.

Unter der Annahme, dass die Verteilung der 1h-Niederschlagssummen zu jeder CP-Temperaturklasse in der Zukunft konstant ist, wurde die vom Klimamodell ECHAM5 (Roeckner et al., 2003) prognostizierte Verteilung der 1h-Niederschlagswerte für jedes Jahr bis 2060 ermittelt. Es zeigt sich, dass der beobachtete Trend zu höheren Kurzzeitintensitäten auch vom Klimamodell bestätigt wird. Bis 2060 ist mit einem deutlichen Anstieg der 1h-Extremwerte zu rechnen. Da dieses Signal des Klimawandels sowohl aus beobachteten Messwerten, als auch aus Klimamodelldaten isoliert werden konnte und darüber hinaus aufgrund von atmosphärenphysikalischen Überlegungen vorausgesagt wurde (Trenberth, 1999), muss es als sehr verlässlicher Trend eingestuft werden.

1. Introduction

1.1. Context and Motivation

Live depends on water. A sufficient supply with drinking water is the basis for all human activity. Drinkable freshwater is mainly found in ground or surface water bodies that are fed by precipitation. Hence, the natural variability in precipitation has always been a thread. A lack of precipitation leads to failed harvests and can provoke starvation. An excess of precipitation causes floods which are not only responsible for high damage to goods and infrastructure, but also for a high number of casualties. In the United States, for example, about 100 people are killed each year in flood events (Ashley and Ashley, 2008). Violent precipitation events can destroy crops and degrade soils. Furthermore, there is the risk of water borne diseases and dangerous infections from contaminated drinking water.

The development of human society is highly linked to the ability to manage these water risks. Bigger human settlements are not possible without hydraulic constructions that ensure a steady supply with drinking water, protect from flooding and discharge waste water. Roman aqueducts, as the "Pont Du Gard" in South of France are spectacular early examples of these efforts.

Hydraulic structures like dikes or sewage channel are expensive and long lasting investments. Hence, the correct dimensioning has always been an issue. Overdimensioning consumes a lot of resources, too small structures bear the risk of system failure. The design of hydraulic structures tries to find the optimal trade-off between the construction costs on the one hand and the avoided potential damage on the other hand. The German law states that flood protection measurements at river banks should be designed to retain all flood events with a higher average frequency than once in one hundred years (WHG, 2009).

Without any statistical or physical information, dimensioning was trial and error: if a town was flooded, the dikes were reinforced. Learning from experience, qualitative rules were established, e. g. to avoid construction on areas that are prone to flooding. Since systematic measurements of river discharge and precipitation are available, empirical rules have been developed that are based on statistical properties. Using extreme value theory, distributions, for example of river discharge measurements, can be extrapolated beyond the observed range. The probability of extreme values, that have not yet been measured, can be quantified. These rules, however, do not consider any physical knowledge. They are governed merely by observed distributions.

Today, it is possible to attack many dimensioning problems by sophisticated numerical modeling. Such computer models represent the natural variability of the flow regime. Hydrological models are able to simulate the discharge of a river based on the precipitation over

the catchment area. Hydraulic models of urban sewage systems calculate the discharges in a channel network in dependence of channel diameters and reservoir sizes. Numerical models are continuously improved. More and more processes of the hydrological environment are incorporated in the model systems. This however, raises a new issue: The more detailed a numerical model is, the higher is its demand of real world observations as input data. More often, the performance of numerical simulations is limited by the offer of available input data and not by the limited complexity of the applied model.

The restriction due to data availability is most remarkable in hydraulic modeling of urban sewage channel systems. Sewage channel networks are very fast reacting systems with short concentration times of typically less than one hour. Therefore, accurate hydraulic modeling requires precipitation input on high temporal resolution of one hour or shorter. Since a sewage network reacts differently if the channels and reservoirs are still filled from former precipitation events, it is not sufficient to model one single event. Best practice for a correct estimation of flood and backwater risks is a hydraulic simulation based on long rainfall time series of several decades.

Temporally high resolution precipitation data show pronounced spatial variability. Therefore the representative area of a time series in hourly resolution is very limited. Germany is covered by several thousand rain gauges, most of them run by the German weather service DWD. Additionally, the area is scanned by 17 rain radar devices. Nevertheless, the offer of available data is not sufficient to provide adequate measurements for every target location. Besides, the observed time series are rarely long enough for correct risk assessment.

In 2000 the synthetic rainfall generator "NiedSim" was developed by Prof. András Bárdossy. In a data driven approach NiedSim assembles stochastic time series that mimic the statistical characteristics of observed high resolution precipitation measurements (Bárdossy, 1998). At any locations in Baden-Württemberg, where no measurements are available, NiedSim provides a simulated time series that can be used as substitute.

Like any model, NiedSim represents the characteristics of real precipitation only to a certain extend. It suffers from a limited ability to represent seasonal variability and inconsistencies in the extreme value statistics. Besides, NiedSim does not consider spatial temporal dependence. Two rain gauges at nearby locations record very similar rainfall time series, whereas NiedSim generated time series for the same locations are very different. In hydraulic modeling of larger sewage systems, however, the application of one single time series leads to errors in the modeled channel discharges, due to high spatial temporal variability of precipitation. The application of several simultaneous time series on the other hand is only valid if the spatial temporal interdependencies are taken into account.

This work presents an approach how NiedSim can be modified to overcome these limitations. The spatial temporal dependence of observed precipitation is measured on different scales and incorporated in the generation scheme.

The calibration of stochastic precipitation models is based on past time observations. Therefore, the application of simulated time series in the dimensioning of sewage system networks implies the assumption that the measurements during the calibration period are representative for the future. Urban sewage channels have a life span of 80 to 100 years. Under

changing climatic conditions the assumption, that the past is representative for the future, is less and less true, the further the projection goes into the future. The future climatic conditions already have to be considered during planning and construction. Therefore, this work concludes by presenting several methods for the estimation of climate trends in high resolution precipitation time series. These methods are based on extrapolation of observed regional trends and the correction and downscaling of climate model output.

1.2. Outline

In the next chapter (**Chapter 2**) this work starts with a description of the climatic conditions of Baden-Württemberg. The precipitation regime is explained in the context of global atmospheric circulation. Regional differences within the study area, mostly due to orographic effects of mountain ranges, are pointed out.

Chapter 3 is dedicated to the acquisition of precipitation data. Different measurement methods are explained, from direct measurements on the ground to remote sensing techniques like rain radar and satellite measurements. The focus is on the accuracy and the potential error sources of the different measurement methods. Some examples of identified errors are presented. Next, the precipitation data is described which is used in this work .

In **Chapter 4** some of the important statistical properties of precipitation are explained: the high variability on different spatial and temporal scales, as well as the intermittency in occurrence and the scaling behavior of measurements on different spatial or temporal aggregations.

If no measurement data is available, stochastic models are applied to generate synthetic precipitation data. **Chapter 5** gives an overview of the existing modeling concepts. The basic ideas of each concept are explained, strong points and limitations are pointed out.

In **Chapter 6** the stochastic time series generator NiedSim is presented. The necessary modifications to are explained. A new version of the generation algorithm is set up and tested by a comparison of an observed and a simulated rainfall time series of the same location. The comparison is based on extreme value statistics and the length of wet and dry spells as a measure of temporal persistence.

In **Chapter 7** the spatial dependence structure of precipitation is analyzed and it is explained how the dependence is represented by statistical measures that can be incorporated in the rainfall generation algorithm. A multi-site generator is set up that is able to simulate several spatial correlated precipitation time series in hourly temporal resolution. The performance of the multi-site generator is evaluated by the same statistics as in Chapter 6.

Chapter 8 deals with climate change. The first part of Chapter 8 gives a literature overview of the climatic changes that are expected for the study area. Then, climate trends that are already present in the observed data are identified by statistical analysis. Predictions for the precipitation regime in the future are developed by an indirect method that exploits prognosis for the atmospheric circulation from Global Circulation Models to classify precipitation

measurements from the past. The predicted trend signals of all the different analysis are compared, and the most reliable trends are pointed out. The chapter concludes with some ideas how the identified trends could be incorporated in the statistical precipitation models developed in this work.

2. Characterization of the Hydrological Conditions in Baden-Württemberg

In Central Europe, people often talk about the weather. The topic is always up to date because the weather changes frequently enough to keep it interesting. Sometimes they regard the atmospheric conditions as comfortable ("nice weather"), other times as uncomfortable ("bad weather"). People are accustomed to the fact that the weather keeps changing in irregular patterns. They might judge it as normal but globally seen, this is a very special case, restricted to mid-latitudes and there mostly to the west side of the continents.

In other parts of the world precipitation activity exhibits higher regularity. It is either dominated by daily or by seasonal cycles or completely constant as over the subtropical desserts where it constantly does not rain, sometimes for years. The inner tropics exhibit daytime climate. In Singapore for example, most of the precipitation falls in the afternoon, some in the early morning, but it generally does not rain between 20:00 pm and midnight. In comparison, the changes over the year are of minor importance. The outer tropics and the subtropics on the other hand are dominated by a yearly cycle of distinct rainy and dry seasons with totally different conditions. Examples are the monsoon climates of southern Asia on the one hand where almost the complete yearly rainfall comes at the end of the summer, or the Mediterranean type climates on the other hand were the long summers are as dry as in the dessert, but the rest of the year is humid (Weischet and Endlicher, 2008).

Central Europe exhibits a pronounced annual cycle in temperature, but less in precipitation. In the study region of Baden-Württemberg, the highest monthly sum is observed in June or July, the lowest in February. In general, however, precipitation can occur at any time of the year. At measurement stations in Baden-Württemberg the average number of rain days exceeds ten in any month (Fig. 2.1, Fig. 2.2).

The general atmospheric conditions in Baden Württemberg change in patterns of several days to few weeks. German language has a special word for this phenomenon called "Witterung". It does not describe the actual weather, which is called "Wetter" in German, but the average conditions over several days. The distinction between "Wetter" and "Witterung" shows that climatic variations in Baden-Württemberg take place in different temporal scales, from subdaily to several days, which is a result of the global atmospheric circulation.

2.1. Context in the Global Atmospheric Circulation

Due to the different exposition angles to the sun, the solar radiation income is higher at the equator than at the poles. At the equator, the solar heating provokes a massive convective

uplift of air that results in a high pressure zone in the upper troposphere at about $10\,\text{km}$ of altitude (Weischet and Endlicher, 2008). The cooling over the poles, especially in winter when the poles are permanently shaded from the sun, leads to setting of air masses and so low pressure in the upper troposphere.

The pressure gradient between Equator and Poles drives a strong compensation movement, on the northern hemisphere from south to north. Coriolis force caused by the earth's rotation, deviates the south-north flux. The result is zonal west wind in the upper atmosphere. This wind field is strongest in the mid latitude around $45°$ where the Coriolis force is strong and the energy gradient is the highest.

The west wind field blocks the exchange between tropic and northern latitudes. If the west wind field was stable, the energy gradient between Equator and Pole could never be attenuated. The tropics would continously heat up, the energy gradient would increase and the wind field permanently become stronger. However, if the Energy gradient in mid latitudes exceeds $6°$ Kelvin per $1000\,\text{km}$, the west wind zone becomes instable and starts to undulate. Wide waves are formed that, embedded in the west wind field, move eastwards. Where the waves peak to the north, subtropic warm air can move northwards, where they peak to the south, cold arctic air is transported southwards. In this state, there is a real energy exchange and the driving temperature gradient is reduced. When it drops below $3.5°$ Kelvin per $1000\,\text{km}$, the zonal west wind can recover and the cycle starts anew (Weischet and Endlicher, 2008).

At the end of the undulating phase, when the waves break down, drops of cold are can be cut off from the northern side of the interface between tropical warm and cool arctic air and move southeastwards. The same can happen with warm air from the south. Both can block the westerly wind field until the gradient builds up again and becomes strong enough to reestablish the main flow direction ("blocking action"). This is especially frequent in winter, when cold drops are supported by a cold ground layer over Siberia.

This permanently ongoing cycle governs the atmospheric conditions of Baden-Württemberg:

- Averaged over the year, the main flow direction is from west-southwest. The energy and moisture that determine the weather in Baden-Württemberg mostly come from the Atlantic Ocean.
- The average atmospheric conditions are condemned to change in irregular intervals of several days. Blocking situations sometimes can last longer up to few weeks or, if the west wind zone is especially weak, reestablish after short interruptions.
- The state of the west wind field governs the average conditions. If the west wind field forms a northern wave over Europe, the flux comes from the southwest and the arriving air masses are rather warm and humid. During a southern wave the flux comes from the northwest resulting in cool conditions. In blocking situations, the energy and moisture fluxes are cut and stable dry conditions can establish.
- Longer dry periods are always related to blocking action. If in the same time a high pressure zone is present further east over Asia, easterly wind fields can establish that bring dry continental air that is cold in winter and hot in summer.

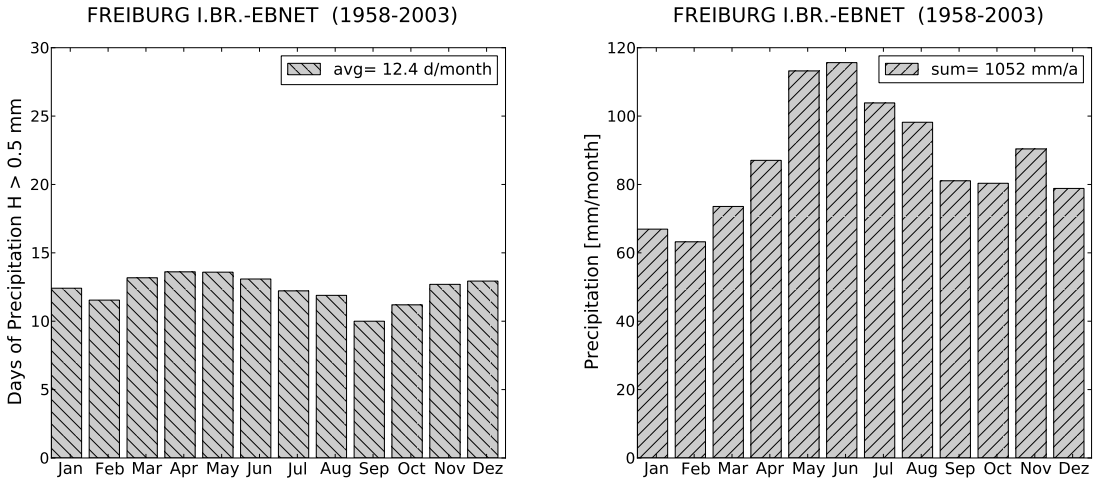
- The energy and moisture of the weather that is experienced in Central Europe is "imported", mostly from the Atlantic Ocean, sometimes from the polar Sea or the Mediterranean. Autochthonal conditions that reflect the local radiation and energy balance can only establish for short periods during blocking events. In winter, when the solar heating is low, it often leads to inversion with cold ground air and low hanging clouds that bring no precipitation. In summer it causes warm and sunny conditions.

The described large scale circulation in the upper troposphere triggers characteristic mesoscale circulation patterns in lower levels. These patterns have a smaller spatial temporal scale and are often referred to as "synoptic". Over the west Atlantic, offshore of Maine, the temperature gradient between cold polar and warm subtropical air masses is the most pronounced. This zone, called the polar front, is the griddle of cyclonic depressions that characterize Central European weather conditions. The undulations in the higher troposphere weaken the stability of the polar front which reacts by a wave movement. In contrast to the upper troposphere waves, the spatial scale is much smaller with the wavelengths limited to 1000 km (Weischet and Endlicher, 2008). The waves move eastwards at the polar front and start to turn and form an eddy, a cyclonic depression. The polar front recovers and a new wave can start off within the next few days. In this way, families of mostly four, sometimes up to six cyclonic depressions are formed (Weischet and Endlicher, 2008) until changes in the west wind zone of the upper troposphere moves the polar front and triggers the next series of depression from a new location.

At the polar front, the subtropical warm air has a higher momentum as the polar cold air, since it comes from lower latitudes were the absolute speed of the earth's rotation is higher. When a wave starts to form at the polar front, the warm air, being lighter, glides upon the cold air at the front side (the east side) of the wave. The forced uplift leads to cooling and thus to condensation. A warm front is forming, which is characterized by compact stratiform cloud bands of several tens of kilometers width and up to a few hundred kilometers length. On the west side of the wave, the cold air from the north moves southwards. Being heavier, it pushes underneath the warm air and lifts it up. The forced uplift leads to cooling and high turbulent mixing of the two air masses. A cold front is formed which is, due to the turbulent mixing, characterized by a scattered band of high convective clouds. The warm and cold air masses behind the fronts try to fill up the low pressure center of the eddy but are deviated to the right by Coriolis forcing. Therefore the eddy starts to rotate anti-clockwise (on the northern hemisphere) forming a cyclonic depression. The cyclone gains speed as long it is over the sea and then looses its momentum over the continent due to the increase in ground friction. Over Eastern Europe, the low pressure center fills up and the cyclone dies.

If the study region of Baden-Württemberg is within the zone of influence of a cyclonic depression, depends on its birth place and the pathway it has taken over the Atlantic ocean. Due to asymmetric Coriolis force, cyclonic depressions tend to move northwards, which is the reason why west-southwest (and not west) is the main flow direction in Central Europe (Weischet and Endlicher, 2008). The storms arriving at one point in Central Europe are clus-

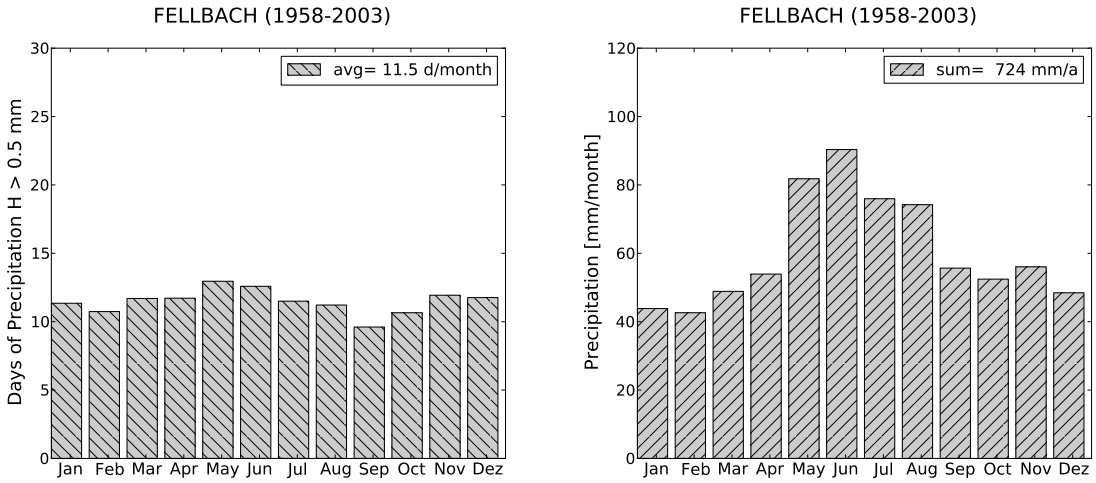
tering in time (Vitolo et al., 2009). It can be assumed that the first storm of the family defines the preferential pathway for its successors. So cyclonic depression frequently pass in series of four to five with intervals of one to two days.



(a) Number of rain days

(b) Monthly precipitation sum

Figure 2.1.: Average number of rain days with more than 0.5 mm and average monthly rainfall sum in Freiburg-Ebnet



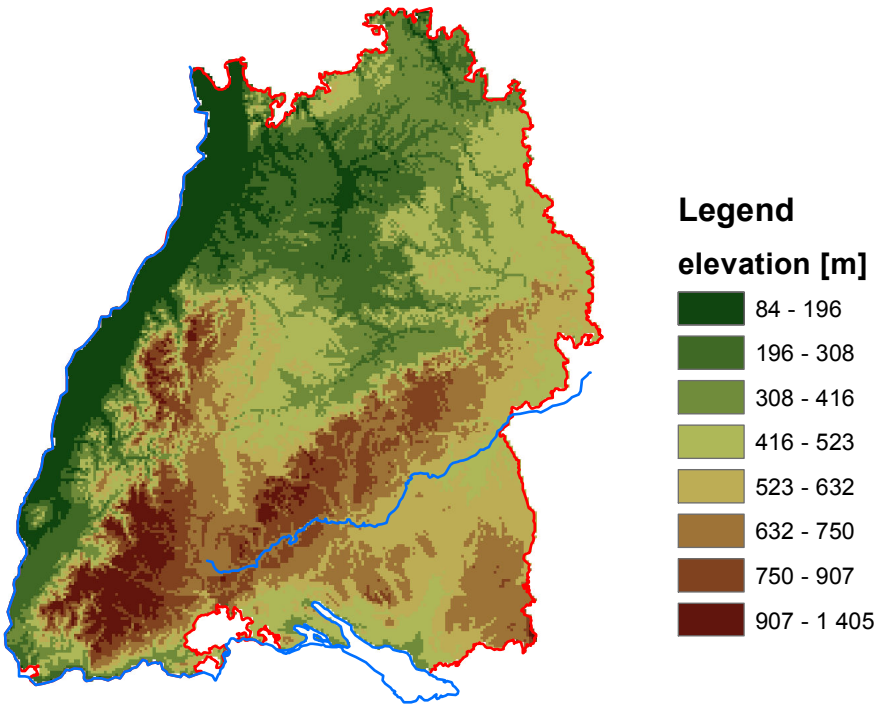
(a) Number of rain days

(b) Monthly precipitation sum

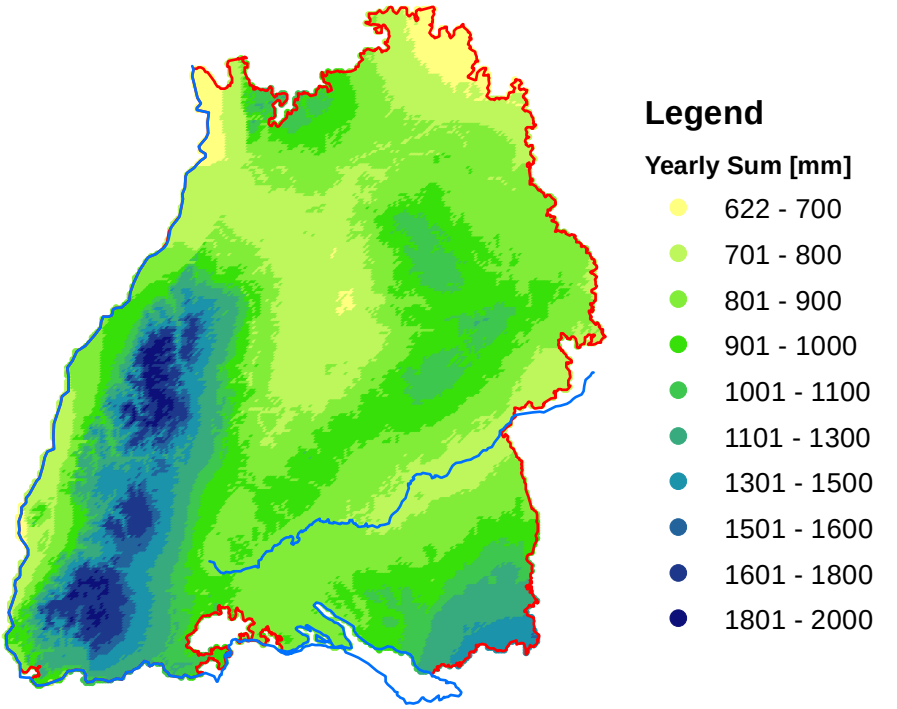
Figure 2.2.: Average number of rain days with more than 0.5 mm and average monthly rainfall sum in Fellbach

The weather in the study area is to a high extend characterized by the cyclonic activity over the northern Atlantic. This has the following consequences:

- Most of the precipitation is bound to warm and cold fronts of cyclonic depressions.



(a)



(b)

Figure 2.3.: Elevation in 1 km × 1 km averaged raster and average yearly precipitation sum for the study area of Baden-Württemberg

- There are two main types of precipitation: advective precipitation at warm fronts and convective precipitation at cold fronts with different characteristics in duration, variability and intensity.
- The warm front of a cyclonic depression always arrives first. The warm front is characterized by stratiform clouds that at first cover the sky at high altitudes and then gain in thickness by growing downwards. The resulting rainfall is rarely very intensive but generally long lasting and spatially homogeneous. The following cold front is characterized by turbulent mixing. It only rains under convective channels where the air is forced upwards by the mixing. Therefore, the precipitation field is highly variable and scattered. The intensities are generally higher, but the duration is shorter than at warm fronts. It sometimes rains in violent showers or even thunder storms. Behind the cold front, the colder polar air masses are in general instable and bring some more, but sparse and less intensive showers.
- Cyclonic activity is highest in winter when the temperature gradient between the Equator (constantly warm) and the North Pole (much colder in winter) is the steepest. The evaporation over the Atlantic Ocean depends on the water temperature. In combination of the two factors, the precipitation due to cyclonic activity reaches its maximum in late autumn.
- In the study region however, the highest monthly precipitation sums are measured in June (Fig. 2.1 and Fig. 2.2). The additional precipitation in summer is the result of enhanced convection. The increased solar heating is an additional driving force for the turbulent mixing at cold fronts resulting in higher precipitation intensities. The solar heating can become so strong that it even triggers convective precipitation events which are not bound to an approaching front. For this reason, the highest precipitation intensities are measured during summer, although the average number of rain days is not higher.
- Passing fronts react to orographic effects. Mountain ranges are obstacles in the flow field that force the approaching air masses to move upwards which induces cooling and thus condensation. The release of moisture results in enhanced precipitation (Weischet and Endlicher, 2008). With the main flow direction of west-south west, orographic enhancement is most perceptible at the western slopes of mountain ranges. The regions east of mountain ranges the other hand are shaded from the main flow direction and exhibit lower precipitation sums.

2.2. Regional Conditions in Baden-Württemberg

Baden-Württemberg has an area of $35752\,\mathrm{km}^2$, the elevation is between $84\,\mathrm{m}$ above sea level in the Rhine valley in the North West and $1493\,\mathrm{m}$ at the peak of the Feldberg in the southern Black Forest. The Rhine valley covers the complete Western Boarder to France and Rhineland-Palatinate. East of it is the Black Forest which is the highest and also the widest mountain range in the study area. Other Mountain ranges are the Swabian Alb with peaks

of more than 1000 m located in the East of Baden-Württemberg, as well as the Odenwald in the very north and the Swabian-Franconian Forest in the northeast whose highest peaks are around 600 m.

The highly structured landscape of Baden-Württemberg finds its expression in the spatial variability of precipitation. Fig. 2.3 compares a $1\,\mathrm{km}^2$ average elevation raster with an interpolated field of the average yearly precipitation sum between 1958 and 2003. The Black Forest is a pronounced obstacle for the flow in west-east direction. Orographic effects are the strongest here and therefore it receives the highest yearly precipitation. The areas east of it are shaded by the Black Forest, thus the Neckar valley is the driest region in the study area. The range of average yearly precipitation reaches from less than 700 mm, e. g. in Stuttgart to about 2000 mm, e. g. at the Hornisgrinde, the highest peak in the northern Black Forest.

To a lesser extent, orographic effects can also be found at all the other mountain ranges. At the Swabian Alb however, due to shading, the average yearly precipitation sums reach only about 60% of the values from the Black Forest at the same altitude. The Danube valley which is, seen from the main flow direction, behind Black Forest and Swabian Alb is again rather dry. Then South of the Danube, the average yearly precipitation sum is gradually increasing. This is the increasing orographic influence of the nearby Alps – especially affecting fluxes from the Northwest. The Odenwald, although much lower than the Swabian Alb, is more humid anyway because it is not shaded by any high mountain range west of it.

The orographic effect of the Black Forest can also be seen in the distribution of precipitation over the year. The deficit of precipitation in Fellbach (Fig. 2.2), downstream of the Black Forest, compared to Freiburg (Fig. 2.1), upstream, is of more than 300 mm per year. The deficit is highest in the months of November and December when the cyclonic activity is highest. The orographic shading is the most effective during that time of the year. During summer months, when precipitation is dominated by convective events, the difference between Fellbach and Freiburg is lower. As a result, the ratio between the rainfall sum in the driest month of February and the wettest month of June is lower in Fellbach. This can be extended to a rule of thumb: the further east a station is located in Baden-Württemberg, the higher is the part of the yearly precipitation sum that falls during the summer months.

3. Acquisition of Precipitation Data

The essential basis on any statistical research on precipitation are measured precipitation data. In principle measuring precipitation is easy. It is sufficient to collect all the precipitation that falls during a defined time interval on a surface of defined area and measure the volume or the mass of the collected water. Precipitation is generally measured in volume per area which is equivalent to a depth. The most common unit is $\frac{\ell}{\mathrm{m}^2} = \mathrm{mm}$. Referring the depth to a time interval, results in rainfall intensity, measured for example in $\frac{\mathrm{mm}}{\mathrm{h}}$.

In reality however, accurate precipitation measurement is challenging. It is difficult to catch all drops and in the same time not to disturb the process, for example not to alter the wind field by the measurement device. Measurements on high temporal resolution are especially difficult as the measured volume becomes very small on short time intervals. Therefore, small losses can induce a high relative error.

Direct measurements only give information about the precipitation at one point which is a limitation if for example the water balance over a river catchment is analyzed. For information on areal precipitation remote sensing techniques are applied.

3.1. Point Measurements

3.1.1. Rain Gauge after Hellmann

The most common point measurement device in Germany are rain gauges that follow the design principle developed by Hellmann in 1886. It features a $200\,\mathrm{cm}^2$ circular collection funnel (diameter of $15.95\,\mathrm{cm}$) made from steal or other non-corroding metal. The collection funnel is limited by a sharp edged ring to avoid side effects at the fringe and it is equipped with a metal cross on top that prevents fallen snow from being blown out.

The funnel conducts the precipitation to a container that is collecting the water. In the simplest case, the container is emptied once in $24\,\mathrm{h}$ either by hand or automatically and the volume or mass of the collected water is measured. Daily precipitation sums are generally measured from 7AM to 7AM.

For higher time resolutions instantaneous rainfall intensities have to be measured. There are several different types of devices:

- **Measurement by Water Level**

  In such devices the water trickling down the funnel is collected in a small cylindrical container equipped with a floater. The floater registers the rise of the water level.

Knowing the base area of the cylinder, the rising speed can be transferred into rainfall intensity. When the cylinder is full, it is automatically emptied into a bigger container that collects the total precipitation.

- **Tipping-Buckets Measurements**

  In a Tipping Bucket device the precipitation from the funnel drips into a small cup of defined volume. When the cup is full, it tips over and releases the collected water. The number of tips is counted and transformed into a rainfall intensity. In general, the Tipping Bucket has two chambers. When one chamber is emptying, the other one is in place so that no water gets lost during the emptying. Due to the discrete nature of Tipping Bucket Measurements, the temporal resolution depends on the rain rate. No signal is emitted before the bucket tips over. Hence, if it rains less intensive, the measurement interval is longer.

- **Gravimetric Measurements**

  In gravimetric devices the container collecting the water is mounted on a balance. The precipitation intensity is calculated by the change in weight over time. Gravimetric Measurements are principally more accurate than Tipping Buckets Measurements. Being fully continuous, they show no time delay except the time it takes for the water to trickle down into the container.

Compared to daily stations, high resolution rain gauges require much effort in maintenance and calibration. The maintenance and calibration standard of the German Weather Service (DWD) recommends to equip each station with an network connected analysis and alarm system for remote technical controls (Bartels et al., 1999). To prevent the Tipping Buckets or weighted containers from freezing, the housing has to be heated. Bartels et al. (1999) demand thermostat controlled heating to avoid excessive evaporation losses.

3.1.2. Disdrometer

Disdrometer do not measure rainfall intensity but the full spectrum of rain drop sizes. ("Disdro" stands for "distribution".) This is especially important for the calibration of rain radar devices (Section 3.2.1). The first disdrometer has been developed by Joss and Waldvogel in 1967. It consists of a styrofoam body that is kept in position by two magnetic coils. If a rain drop hits the styrofoam body, it transfers its momentum, provoking a small deviation of the styrofoam body. By electromagnetic induction an electric current is generated at the magnetic devices. As the momentum of a raindrop depends on its mass, the current is related to the size of the raindrop.

Malvern Particle Sizers measure the distribution optically by a laser beam. The beam is emitted at one side of the device and received at the other side by light sensitive diodes. Each raindrop that falls through the scanned volume leads to a short interruption of the laser beam. The drop size is calculated from the extinction amplitude. The time length of the extinction is a measure for the falling speed of the drop.

While the disdrometer of Joss and Waldvogel is suited for the detection of big drops, the Malvern Particle Sizer is especially useful for the detection of small droplets with a diameter of less than 0.1 mm (Emeis, 2010).

3.1.3. Error Sources

To attain high accuracy in short interval precipitation measurements is an ambitious task. The calibration outline of the German Weather Service defines a spread of 5% in three measurements as acceptable for Tipping Bucket devices. In gravimetric measurements the spread in three measurements must not exceed 0.4 mm (Bartels et al., 1999). Groisman and Legates (1994) estimates that the total error in accumulated annual precipitation of US rain gauges is between 5% and 25% with larger errors in higher elevation and in winter time.

Main error source in rain gauge measurements is wind. It prevents the rain drops from landing in the funnel. The World Meteorological Organization (WMO) defined a standard wind protected reference rain gauge. It is either buried into the ground or protected by a double wind shield (Sevruk et al., 2009). Comparing measurements from ordinary, unshielded gauges to the reference, Sevruk et al. (2009) estimate a wind induced error of about 3% to 6%. Wind errors can be corrected by empirical formulas based on numerical simulation.

Another important source of error is snow. According to Sevruk et al. (2009),at a wind speed of 6 m/s an ordinary rain gauge catches only about 20% of the precipitation when it is snowing and only 50% of mixed precipitation of rain and snow. The loss due to evaporation on the other hand is comparably small (Groisman and Legates, 1994). However, it affects especially the smallest values which are important in the estimation of precipitation probability. Most susceptible for evaporation losses are graviemtric devices since the water remains in the container for a long time.

Other errors concern the mechanics. The rain gauge funnel can get clogged by debris or the reed switches of Tipping Buckets can be blocked (Ciach and Krajewski, 2006). To detect and fix such errors it is required that rain gauges are controlled in short maintenance intervals. Recording and data transmission are further potential error sources. When for example the baseline signal of a recording station is slightly drifting, the rain gauge gives out small rainfall increments instead of zero values. This is not important in terms of the overall precipitation volume but can be a severe problem in the estimation of rainfall probability.

Finally, for spatial correlation analysis, it is important that all precipitation stations are synchronized. If the time stamp at one station is wrong (for example shifted by some minutes) the observed spatial correlations to all other stations is altered.

3.1.3.1. Examples of Recording and Data Transmission Errors

Fig. 3.1 shows an error that is probably caused by data transmission. It was found at the station in Wiesloch (Gauss-Krüger 3 of $x = 3478250$ and $y = 5462670$). The picture is an edit of the time series delivered from DWD. The seven digit code at the beginning of each record

```
St-Nr. X Y Hoehe Jahr 30min Nieder in mm)(Kalenderjahr Jan.-Dez.!)
       18288     3478250     5462670         160        1989
 Tag/Std/Min Nieder (I3,I2,I2,F8.3)
10000 0.000
10030 0.000
10100 0.000
10130 0.000
10200 0.000
10230 0.000
...

...
0492330 0.000
0490000 0.000
0490030 10.000
0490100 0.000
0490130 0.000
...

...
0512330 0.000
0520000 0.000
0520030 10.000
0520100 0.000
0520130 0.000
...
```

Figure 3.1.: Example of an error in a precipitation time series, probably occurred during data transmission

is the time stamp. It states the day in the year (first three digits) and the daytime (next four digits). What is particular about this series are the isolated high values that always occur at half past midnight. It is assumed that these are errors, due to their regularity, the crisp value of exactly $10.0\,$mm and the periodic behavior. The reasons are less clear. Perhaps it was caused by a misinterpretation of an end of line character when daily files were combined to longer time series or a signal error in data transmission that altered the code for "new day".

The data presented in Fig. 3.2 and Fig. 3.3 are from the station of Craislheim ($X = 3577650$; $Y = 5446865$). The time series attracted attention due to the highly monthly sum in July 2003, which was a rather dry month at other stations. The values on the first day, the 14^{th} of July, all sum up to the same amount of $15\,$mm (Fig. 3.2) the values of the second day to about $30\,$mm Fig. 3.3. Most probably a service team was at the station and checked the accuracy. Later, it was forgotten to cancel out the values or the report got lost. It is very unlikely that these values are really rainfall because of the isolated, unrealistic high intensities. The hypothesis of maintenance is more likely. The values on the next morning, now the 16^{th} of July, are less suspicious. However, in combination with the unrealistic rain depths $24\,$h and $48\,$h before, they were sorted out.

```
 St-Nr. X Y Hoehe Jahr (5Min Nieder in mm)(KALENDER Jahr Jan-Dez!)
       61368     3577650     5446865         395        2003
 Tag/Std/Min Nieder (I3,I2,I2,F8.3)
0010000   0.030
0010005   0.000
0010010   0.030
0010015   0.000
0010020   0.000
0010025   0.000
0010030   0.000
...

...
1950650 0.000
1950655 0.000
1950700 0.000
1950705 0.000
1950710 0.000
1950715 15.000
1950720 0.000
1950725 0.000
1950730 0.000
1950735 0.000
1950740 15.000
1950745 0.000
1950750 4.860
1950755 10.140
1950800 0.000
1950805 0.000
1950810 0.000
1950815 15.000
1950820 0.000
1950825 15.000
1950830 0.000
1950835 0.000
1950840 0.000
...
```

Figure 3.2.: Example of an error in a precipitation time series, probably due to unreported calibration and maintenance

3.1.3.2. Some Aspects of Error Identification

Many errors can be detected by simple cross checking. Therefore, Ciach and Krajewski (2006) recommend to always run two rain gauges at the same location. If the stations in Crailsheim and Wiesloch featured two independent measurement devices, the error could have been canceled out immediately: One of the devices would have reported precipitation and the other would not. To control for measurement errors all new DWD stations have to be equipped with two measurement devices, either with Tipping Bucket and gravimetric devices or with water level and gravimetric devices (Bartels et al., 1999).

It is hard to set up a fully automatic algorithm for error detection. Many errors are obvious for a human being looking at the data but difficult to translate in purely mathematical logic. A good strategy is a semi-automated method. Many errors can be detected by unrealistic high yearly, monthly or daily precipitation sums. However, not all extreme sums are due to measurement errors. The search for suspicious values can be performed by a computer algorithm. Then this sample of potential data errors has to be checked by eye. In this manner, the errors reported in Section 3.1.3.1 were found.

The person in charge with error correction is in a dilemma. As Emil Julius Gumbel, one

```
 St-Nr. X Y Hoehe Jahr (5Min Nieder in mm)(KALENDER Jahr Jan-Dez!)
       61368     3577650     5446865         395        2003
 Tag/Std/Min Nieder (I3,I2,I2,F8.3)

...
1960530 0.000
1960535 0.000
1960540 0.230
1960545 0.000
1960550 0.180
1960555 17.830
1960600 15.000
1960605 0.000
1960610 0.000
1960615 0.000
1960620 27.570
1960625 2.430
1960630 0.000
1960635 0.000
1960640 0.000
1960645 0.000
1960650 0.000
...

...
1970530 0.000
1970535 0.490
1970540 2.150
1970545 0.000
1970550 0.000
1970555 9.820
1970600 0.000
1970605 0.000
1970610 0.000
1970615 0.000
1970620 0.000
1970625 7.600
1970630 0.000
1970635 0.020
1970640 12.050
1970645 9.970
1970650 0.000
...
```

Figure 3.3.: Example of an error in a precipitation time series, probably due to unreported calibration and maintenance (continued)

of the founders of extreme value statistics, stated: "It's impossible that the improbable will never happen." (Gumbel, 1958). So it is not a good strategy to sort out *all the values* that seem unrealistic. In this way one would systematically eliminate all *real* extremes events too, which is problematic because it leads to a sever underestimation of potentially dangerous precipitation events.

However, not all the errors are as obvious as the two examples stated in Section 3.1.3.1. Sometimes error detection demands for a little fantasy. It often helps to check on the internet. If an extreme rainfall event occurred, then maybe there was an article in a local newspaper or a note in the chronicle of the fire brigade that can be found on the web.

3.2. Remote Sensing

Remote sensing can be defined as a technique that allows to measure "without direct contact between instrument and object" (Emeis, 2010). For precipitation, this is mainly the measurement of precipitation related radiation. There are two different measurement principles: passive and active measurement. Passive measurement devices only detect naturally emitted radiation, while active measurement devices send out a signal and measure the reflection (e. g. the fraction of reflected energy).

The two main types of operational remote sensing devices are terrestrial rain radar and satellite remote sensing. Rain radar measurements can give detailed images of spatial heterogeneous rain fields in temporal resolutions up to few minutes. Terrestrial rain radar measurements are active measurements. Satellite remote sensing combines different actively and passively measured signals as microwave radiation and infrared radiation in several wavelengths. Satellite images are limited in spatial-temporal resolution, but are the most important source of information about precipitation intensities in sparsely gauged regions and over the ocean.

3.2.1. Radar Measurements

Electromagnetic radiation gets scattered by any particles that are much smaller than the wavelength. The fraction of scattered energy depends on the 6^{th} power of the droplet diameter (Emeis, 2010) and on the inverse of the 4^{th} power of the wavelength. This phenomena, called Rayleigh scattering, is the measurement principle of rain radar devices. Water droplets and ice crystals scatter the energy of a radar beam that was emitted into the atmosphere. The scattered energy dissipates in every direction. Part of it is reflected back to where the beam was sent out from. A functional dependence can be set up between the reflected fraction of energy Z and the rainfall intensity R in the scanned atmospheric volume. It is called the Z-R relation.

A rain radar consists of an emitter sending out a distinct beam of known duration and defined amount of energy and an receiver that measures the energy reflected back to the radar station. Emitter and receiver are synchronized. From the travel-time between emission and reception the reflection distance is calculated. Applying the Z-R-relationship, transforms the reflectivity profile into a rainfall intensity profile along the beam axis. The intensity is an average over the diameter of the radar beam. In general, radar beams of a certain aperture angle, typically 1° are sent out. This means that the beam measures the intensity in a conic volume whose diameter is increasing with the distance from the radar station. Thus, the spatial resolution depends on the distance to the radar station.

There are different modes of rain radar scans. The simplest application is a one dimensional scan along the axis of a radar with fixed orientation. This is mostly used for vertical profiles of the atmosphere, e. g. to identify height levels on which precipitation is formed. Two dimensional scans can be performed if the radar emitter sends out horizontal beams according to a fixed pattern, e. g. every degree of angle, while rotating on its vertical axis, e. g.

once in five minutes. The profile of all beams from one rotation can be composed to two dimensional images. For a volumetric scan the rotations are repeated with varying elevation angle.

Different radar bands are used for rain radar, namely the X, C and S band with wavelengths of $3.2\,\text{cm}$, $5\,\text{cm}$ and $10\,\text{cm}$. X band radar achieves the highest spatial resolution but, compared to the other two bands, the range is lowest since the energy dissipation is higher for shorter wavelengths. The S band radar signal becomes the least attenuated when traveling through the atmosphere but needs longer antennas and more energy to give a reasonable spatial resolution (Delrieu et al., 2000).

Radar images are the only source of information that gives a high resolution continuous picture of the spatial patterns in precipitation fields. Detailed investigations of spatial structure and dependence in precipitation could not start before the first radar images were available. Quantitative precipitation measurement, however, is difficult since rain radar measurements are affected by many different error sources:

- **Errors due to variations in the Z-R relation.**
  A priori, the relation between reflectivity and rain rate is a physical law Since the reflectivity depends on the rain drop surface and not directly on the volume and due to the fact that the falling velocity increases with the drop size, the Z-R relation is highly non-linear. Generally it is described by a power law function of the form

  $$Z = a \cdot R^b \tag{3.1}$$

  (e. g. Klazura et al. (1999), Delrieu et al. (2000) or Lee and Zawadski (2005)).
  The non-linear function can only be determined if the drop size distribution is fully identified. Since the drop size distribution is changing in time, the Z-R relation is non-stationary.

  The real drop size distribution within the volume that is scanned by the radar beam cannot be observed. It can only be measured at the ground and only point-wise. The Z-R relation is usually set up for the climatological average of the drop size distribution (e.g. Klazura et al., 1999). This purely statistical function averages out any variability in the drop size distribution, for example due to differences in rainfall formation. Rainfall from melted snow for example leads to different drop sizes than warm precipitation as snowflakes aggregate in the formation and break up in the melting (Mitra et al., 1990). Precipitation from stratiform clouds consists of smaller drops than rainfall from convective events (Lee and Zawadski, 2005).

  For Montreal, Canada, the average relative error that is induced by using a climatological average Z-R relation was quantified by 41% (Lee and Zawadski, 2005). Daily averaged $Z - R$ relations induced a relative error of 28%. The drop size distribution can change even within one rainfall event, e. g. upstream or downstream of a passing front (Lee and Zawadski, 2005). Therefore, recent publications investigate statistical tools for a better identification of the drop size distribution, by classifying events according to the rainfall formation (Ignaccolo and Michele, 2012a) or by the presence of orographic effects (Ignaccolo and Michele, 2012b).

- **Bright band**
  Related to the errors in the Z-R relation, however more specific, is the problem of bright band occurrence. The water on melting snow flakes or hail leads to stronger reflection than expected by their water content. The radar "sees" the diameter of the snow flake and assumes a droplet of this size. The result is a strong overestimation of the precipitation rate if the radar beam hits the melting layer in a precipitation field (Germann and Joss, 2002) or in the presence of hail (Austin, 1987). Since hail is correlated with very high precipitation intensities, bright band correction can lead to a loss of information on most extreme events (Baeck and Smith, 1998).

- **Ground clutter and beam blockage.**
  Raindrops are not the only objects that reflect the radar beams. Beside other flying objects as birds or airplanes, the reflection at high towers, trees or mountain peaks is a real issue. If such objects block the beam completely, the rain radar is blind for what is located behind. If the beam is only blocked partly, the radar assumes a high rain rate at the clutter point.

  Clutter identification can be done by pixel-wise accumulation of rainfall sum (e. g. during one month), which is highly overestimated at clutter points, or geographically with the help of a precise digital elevation model. More sophisticated data driven clutter correction algorithms are based on Fuzzy logic or Bayesian framework (Rico-Ramirez and Cluckie, 2008) or on aggregated statistics of the radar signal (Hubbert et al., 2009).

- **Errors due to vertical variability**
  The issue of clutter and beam blockage is especially severe in mountainous regions. To avoid excessive losses in scanning range, rain radar stations are often built on mountain peaks. However, the higher elevation comes at the price of reinforced errors due to the vertical variability in the rainfall intensity (Germann and Joss, 2002). If the rain clouds are (partly) at a lower altitude than the lowest scanning level, the radar beam does not measure the full rainfall intensity since part of the precipitation is forming below the scanned volume ("overshooting"). According to Germann and Joss (2002), this is the most severe error source in alpine regions where the radar beams can reach altitudes of 2000 or $3000\,\mathrm{m}$. Andrieu and Creutin (1995) developed a method to correct for vertical variability errors, but it requires at least data from two different elevation angles. Germann and Joss (2002) propose a pixel-wise correction factor relating the radar measured intensities to gauge measurements at the ground.

- **Attenuation**
  The raindrops falling through the radar beam do not only reflect the energy back to the radar station but dissipate it in all possible directions. Each detected raindrop diminishes the amount of energy remaining for the following measurement farther out on the radar beam. The higher the rainfall intensity, the higher the attenuation. Since the rain rate is calculated from the fraction of radar energy reflected back to the receiver, attenuation leads to underestimation of the rainfall intensity in areas that are hidden behind intensive rainfall events. S-band radar is much less subject to attenuation than C-band radar. In heavy rainfall events C-band rain rates can be more than 50% lower

(Delrieu et al., 2000).

The correction for attenuation errors is very difficult since the error is proportional to the integral over the reflectivity measured along the beam. There is an analytical solution to calculate the path integrated attenuation, but, due to the dependence on the measurement signal, it is very sensitive to errors and can become instable (Delrieu et al., 2000).

Radial fields from two dimensional radar scans are stored in a polar coordinate system that is centered on the radar stations. In overlapping regions, the segments of the fields from two different stations are not identical. For the use of these measurements in hydrological applications, e.g. flash flood warning, the measurements are interpolated on a regular, Cartesian grid. Besides, errors are detected and corrected. Many national meteorological services give out radar data in interpolation products (e. g. RADOLAN for Germany, Bartels et al., 2004). RADOLAN consists of five minute sums of the areal precipitation on a $1\,\mathrm{km} \times 1\,\mathrm{km}$ grid. It is calculated from the two dimensional scans of 17 C band radar stations. Clutter is eliminated and the rainfall amounts are adjusted with data of several hundred precipitation stations from all over Germany. Such products, however, are still subject to errors. Baeck and Smith (1998) for example find a general underestimation in the WSR-88D-III product emitted by the National Weather Service of the United States of America.

Radar and rain gauge measurements, however, are not directly comparable. Even if both methods produced perfect error free measurements, there would be deviations. Rain gauges measure rainfall intensities or accumulations at one point at ground level. Radar measures instantaneous rainfall intensities in a scanned atmospheric volume that are accumulated to precipitation sums of a certain time interval (e. g. five minute values) over a defined area (e. g. a square of $1\,\mathrm{km} \times 1\,\mathrm{km}$). In the absence of measurement errors, both would only give identical values if the precipitation was spatially homogeneous over the area of the radar cell. Besides, radar measures the precipitation at a certain altitude and it is unknown where or when the same precipitation can be measured on the ground because the raindrops need some time to fall and drift in the wind during their travel.

Rain gauge and radar measurement can only be compared in accumulations over long time intervals during which spatial variability is assumed to cancel out.

The difference between point and areal precipitation makes it difficult to correct for errors in measured radar fields. The true areal precipitation, which would be needed as a reference to develop error correction algorithms, is unknown. Rain gauge measurements are generally considered as more accurate but only give point values. The radar exhibits more scatter but gives a detailed picture of the spatial heterogeneity. Merging algorithms (Ehret et al., 2008) try to benefit from both features by embossing the variability of the radar measurements on the station values. A test of different merging algorithms can be found in Goudenhoofdt and Delobbe (2009).

Germann et al. (2009) point out that the correlation structure in the error is very important for any hydrological application that uses radar data as input. Hydrological models to some extend integrate over the catchment area, and therefore uncorrelated errors are mostly canceling out in the hydrological response whereas highly correlated errors are adding up.

When radar measurement came up, there was a hope that radar measurements could one day substitute for the expensive ground station network. Collier (1986) reports: "There has been considerable debate in recent years as to whether measurement techniques based upon remote sensing technology, using radar and/or satellite systems, can complement, or indeed replace, rain gauge measurements." Despite of the advances in radar technology and radar correction during the last 25 year, a replacement of station measurements by radar data did not take place. In reality the opposite is the case; rain radars depend on ground measurements. Full benefit from radar data can only be achieved if it is combined with rain gauge data. Firstly, as all indirect measurement methods, rain radar measurements depend on calibration that is only possible with directly measured ground data. Secondly, radar correction algorithms depend on the additional information of rain gauge data. The denser the station network is, the better is already the spatial representation in the point measurement network and the higher is the accuracy of the spatial-temporal precipitation fields combining the two sources of information.

3.2.2. Satellite Remote Sensing

Satellites used for precipitation estimation are equipped with different devices that measure precipitation related signals in radiation. The satellites feature sensors for detection of infrared radiation of different wavelengths and for microwave radiation. The newest precipitation satellite, the Tropical Rainfall Measurement Mission (TRMM) satellite, is also equipped with a rain radar device (JAXA, 2012).

The detection of precipitation by infrared and microwave measurements is very indirect. Primarily, the radiation scans give information on the shape and thickness of clouds which in turn is used for the estimation of precipitation rates (Kidd, 2001). As the Earth's background radiation is different over sea than over land and as the rainfall formation processes vary with latitude, the signal proceeding has to be adapted on the geographic location.

Precipitation detection satellites are on Low Earth Orbits (LEO) of several hundred kilometers above ground. They surround the Earth in on or two hours (Kidd, 2001). Seen from the Earth, they are constantly moving. The satellites scan the surface in swaths of few hundred kilometers width which is depending on the measurement sensor. As a consequence, each point on the Earth is visited once in a certain satellite specific interval.

As a result of the radar movement the nature of radar measurements is discontinuous. The radar images for one location are instantaneous snapshots with long regular gaps in between. For continuous time series, radar data have to be aggregated and interpolated. The interpolation is a trade off between temporal resolution and spatial coverage. Products, as for example the TRMM Multisite Precipitation Analysis (TMPA) combine data from different satellites and also from different sensor types. The temporal resolution of TMPA data is $3\,$h, the spatial resolution of $0.25° \times 0.25°$, which is equivalent to $27.8\,$km at the equator (Huffman et al., 2007).

Radar measurements are of great benefit for previously ungauged regions as remote parts of the continents and the ocean. Radar data help to complete the picture of global precipitation and are therefore essential in research on global atmospheric circulation and climate.

For many hydrological applications however, for example rainfall runoff simulations for small scale river catchments or urban sewage systems, radar data is sill too coarse. Experts state that a spatial-temporal resolution of $4\,\text{km}$ and $30\,\text{min}$ "can be achieved in the near future" (Sorooshian et al., 2011) but accuracy issues remain. Sorooshian et al. (2011) point out that there is still demand for research in validation, uncertainty analysis and bias removal. AghaKouchak et al. (2011) analyzed several satellite precipitation products for the Central United States. They state that "no single procuct can be considered ideal for detecting extreme events". They all tend to "miss a significant volume of rainfall". A limitation that up to now prevents the design of radar based flood warning systems (Sorooshian et al., 2011).

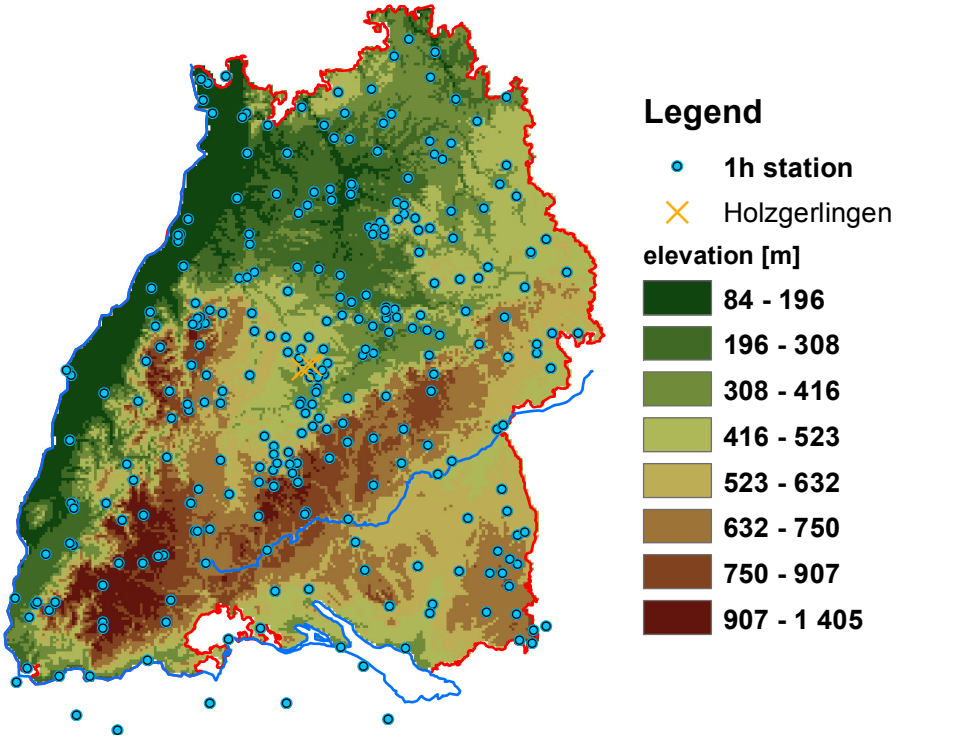
3.3. The Data Set of Precipitation Measurements

The research of this work is based on precipitation measurements from ground stations. The precipitation time series were delivered by the environmental authority of Baden-Württemberg ("Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg" – LUBW), except for eight stations that were delivered by the Swiss meteorological service, MeteoSwiss. The data was provided in numeric form. The network of rain gauges in Baden-Württemberg includes stations maintained by the LUBW and the German meteorological service ("Deutscher Wetterdienst", DWD). According to LUBW the data was verified, but a close manual inspection revealed some errors in the time series. Some examples can be seen in Section 3.1.3. The data set was already used in former studies and originally consisted of 295 high resolution rainfall stations (Brommundt, 2008). Three stations however, had to be sorted out completely due to data problems that were detectable in aggregated statistics as yearly precipitation sum but could not be located in the time series (see Section 3.1.3). These were the stations in Oberburken with DWD-Number 61347, Buchen with DWD-Number 116558 and Oberrotweil-Vogstburg with DWD-Number 70293.

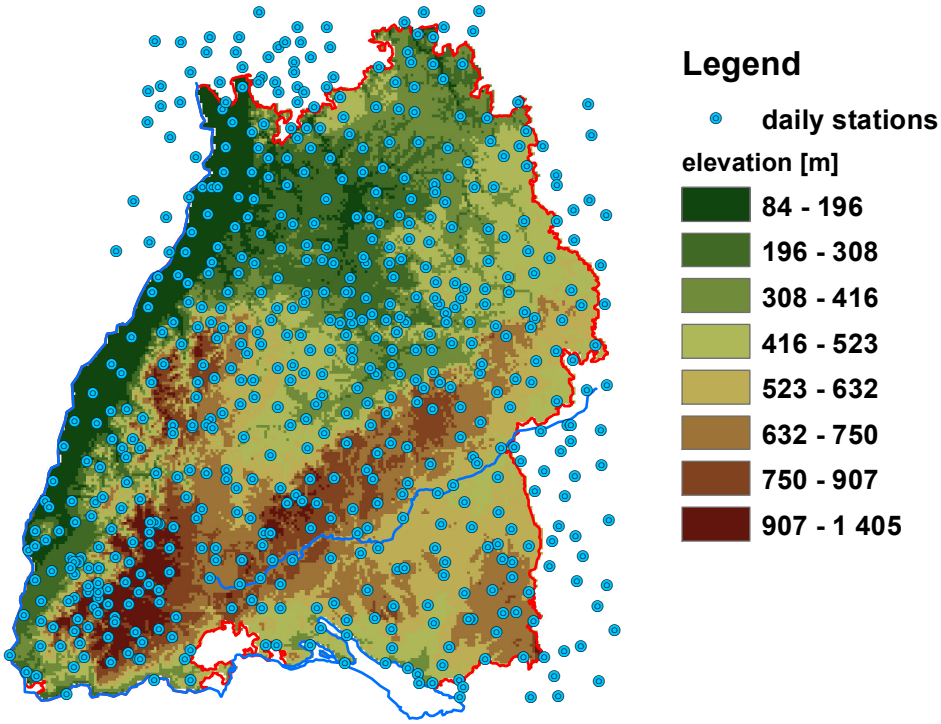
The rainfall stations are rain gauges after Hellmann (Section 3.1.1). Most high resolution stations are tipping bucket rain gauges. The stations inaugurated after 1997 are equipped with two measurement devices for cross checking, either a gravimetric and a water level device or a gravimetric and a tipping bucket measurement device. The average spatial density is one rain gauge in $70\,\text{km}^2$ for the daily stations and one station in $122\,\text{km}^2$ for the high resolution rain gauges. The average length in the daily time series is 28.9 years (Brommundt, 2008) in the hourly time series of about 9 years. The longest time series in hourly resolution is the station in Karlsruhe (DWD-Nb. 70621) with 33 years of available data from 1958 to 1990 but with a high fraction of missing values. Due to the simpler measurement procedure for daily precipitation values and the lower effort in maintenance, the daily time series are not only longer and more frequent but also exhibit a better quality with less missing values.

3.3.1. Subsets

The number of available stations varies over the years. Many of the longer high resolution time series end in the late 1980s or early 1990s. The situation improves again by the year



(a)



(b)

Figure 3.4.: Location of the daily and hourly precipitation stations used in this work

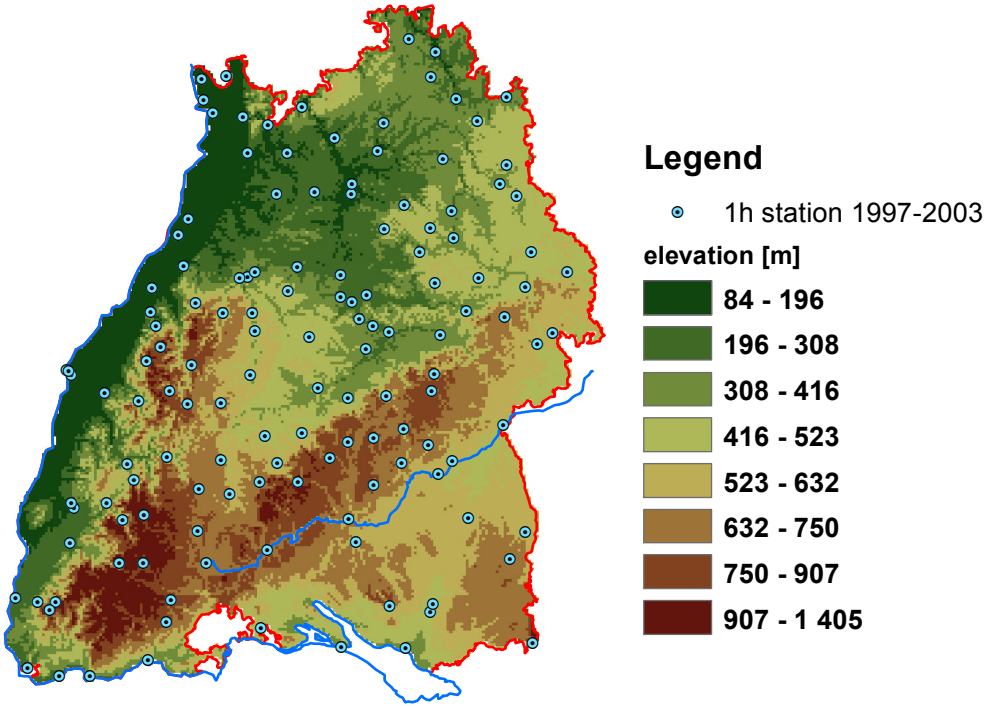


Figure 3.5.: Subset of 137 high resolution rain gauges with continuous measurements from 1997 to 2003

of 1997 when the DWD inaugurated many new rain gauges. For analysis of spatial dependencies it is required that all measurement stations have the same measurement period. For such tasks, only the sub-period from 1997 to 2003 is considered. During this time, data from 137 of the 292 high resolution rain gauges is available. The location of all stations in the subset can be found in Fig. 3.5.

Another station subset of special importance is the subset of precipitation stations with 30 min measurement interval where the fraction of missing values is particularly low. It consists of 30 stations. For for the measurement period from 1991 to 2003 the data availability is close to 100% (Fig. 3.6).

The red crosses in Fig. 3.4a indicate the two stations in Holzgerlingen. These two stations were in operation from 1977 to 1992. With only about 4% percent of missing values, the data quality is high. The two stations are located near the center of the study region in a distance of 3 km to each other in approximately east west direction (Gaus-Krüger 3 coordinates of: X=3498810, Y=5383440 and X=3501820, Y=5384490). Due to the high data quality, the vicinity and the long common measurement period, the two stations are regarded as an representative example for spatial correlated precipitation time series and used for verification purposes throughout this work.

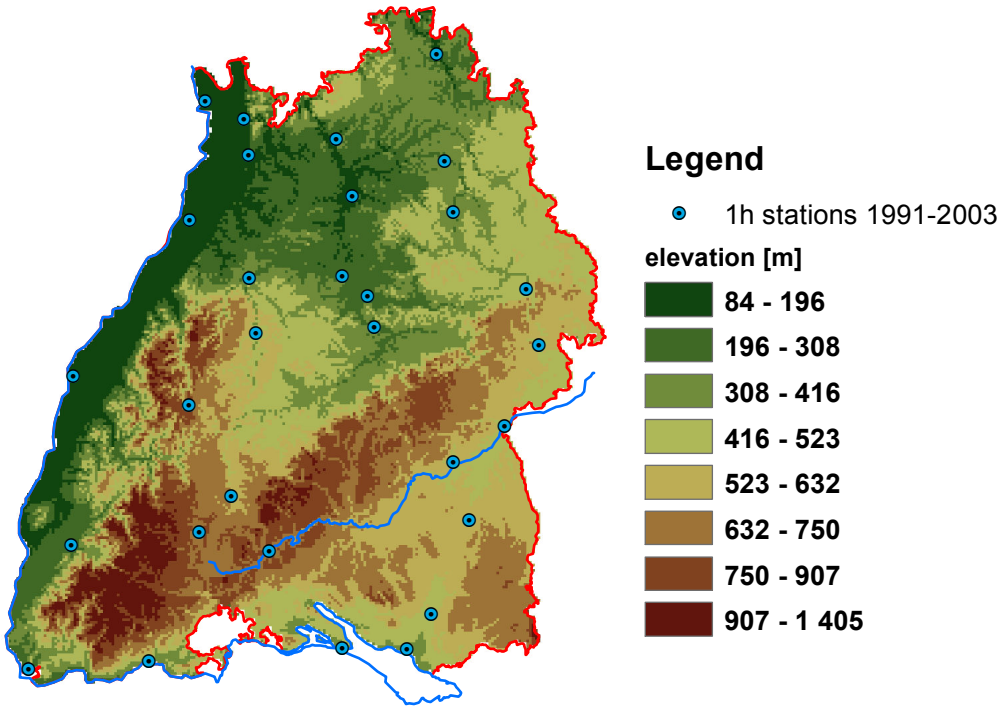
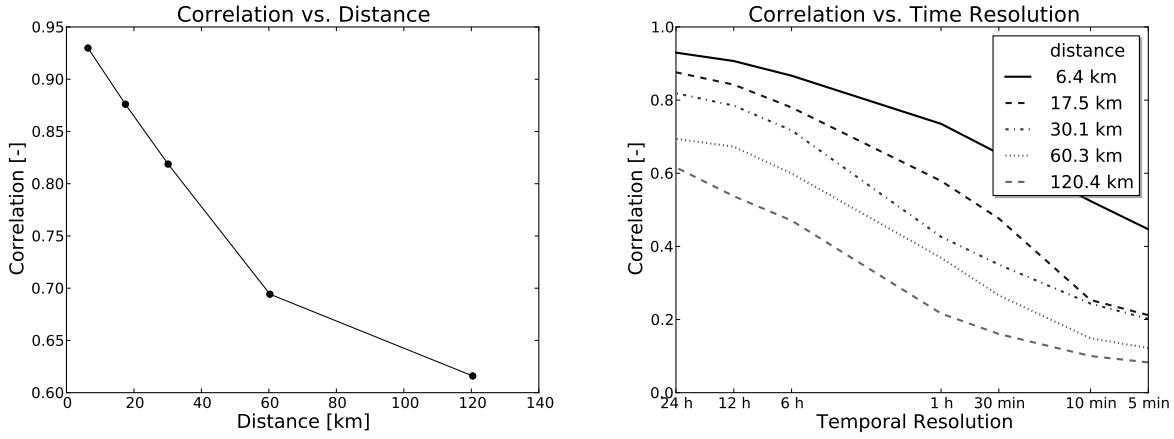


Figure 3.6.: Subset of 30 high resolution rain gauges with continuous measurements from 1991 to 2003 and low number of missing values.

4. Statistical Properties of Precipitation

The German weather service DWD operates several thousand rain gauges in Germany. Their temporal resolution ranges from daily scale to five minute values. Additionally, the whole area of Germany is covered by a network of 17 precipitation radar devices. Despite these efforts, there is a lack between the demand for precipitation data as input for hydrological and hydraulic simulation and the offer of available measurement data. This lack of data is due to statistical properties of precipitation: spatial temporal heterogeneity, intermittency and scaling.

4.1. Spatial and Temporal Heterogeneity



(a) (b)

Figure 4.1.: Dependence of correlation between rainfall time series on the station distance. Example for Westernheim (GK3: $x = 3546150$; $y = 5375070$) and five surrounding rainfall stations. Data from 1998.
a) correlation of the daily time series versus distance.
b) correlation depending on the temporal aggregation from five minutes to 24 hours

One of the statistical characteristics of precipitation is a high spatial and temporal variability and the fact that variation in space and in time are connected. A typical example is displayed in Fig. 4.1. The figure shows the correlation of the time series for 1998 at five rainfall stations (Wiesensteig, Muensingen-Apfelstetten, Baltmannsweiler-Hohengehren, Schwäbisch Hall

and Durbach-Ebersweier) to the station of Westerheim on the Swabian Alb (Gauss-Krüger 3 coordinates of $x = 3546150$; $y = 5375070$). The stations and the target year were chosen arbitrarily. The only criteria were a big range of distances (between 6 km and 120 km) and 100% data availability. Fig. 4.1a displays the correlation of the daily time series at these stations to the time series at Westerheim. The correlation can be seen as a measure of their similarity. From Fig. 4.1a it becomes clear that the similarity is higher if the distance between the stations is low. The correlation values are within a typical range for Middle Europe. Osborn and Hulme (1997) analyzed 185 stations pairs in Middle and Western Europe and found an average distance of 150 km to 200 km for a correlation of 0.5.

However, independently of the distance between the rainfall stations, the correlation also depends on the temporal resolution of the records (Fig. 4.1b). At least at short distances, the daily values are strongly related. The hourly values show more freedom in their variations while five minute values have a very low correlation if the distance between the stations is higher than a few kilometers. If one thinks of a horizontal line through Fig. 4.1b at the 0.6 correlation level one can say (for this particular set of stations - but similarly for all other stations in Baden-Württemberg) that two five minute time series of a few kilometer distance differ as much from each other as two time series of daily values in more than one hundred kilometer distance.

This permits the reversal conclusion that the area for which a rainfall record is representative is more and more limited, the shorter the considered temporal resolution is. For a complete spatial representation, the network of five minute rain gauges should therefore be much denser than the network of daily rain gauges. Unfortunately, the opposite is the case in Germany due to the much higher maintenance requirement (Section 3.1) for high resolution measurement stations.

Two stations at a longer distance have a lower overall level of correlation but show the same decline towards shorter aggregation time intervals. The different curves in Fig. 4.1b are more or less parallel. Different scales in rainfall statistics are influenced by different meteorologic factors. Whether one month is wet or dry depends on the large scale atmospheric conditions during longer periods. If for example one rain gauge records an especially high monthly rainfall sum, it is very likely that all surrounding stations do the same, since they were influenced by the same atmospheric conditions during that month.

On daily scale it is less true. Daily rainfall values are more sensitive to local effects at the rain gauge as monthly sums. One example for such local effects is orographic uplift at hill slopes enhancing rainfall intensities. Weston and Roy (1994) proofed for the Scottish Highlands that orographic effects depend on the flow direction. Daily rainfall sums can vary by a factor of two or more depending on whether the rain gauge is situated at the upwind or downwind side of the mountain range (Weston and Roy, 1994). Since the flow direction can vary considerably between one day and another, it can provoke differences in the daily precipitation sums at nearby station with different exposition.

Another phenomena which induces high heterogeneity, especially in daily and subdaily temporal scales, is the presence of convective rainfall events. In middle European climate estival thunder storms can provoke very high intensities over very small areas. The flooding

at the Starzel river in South West Germany, 2008, is an extreme example (Ruiz-Villanueva et al., 2012). It is possible that one rainfall stations is hit by such an event while the the surrounding stations in few kilometer distance do not even record any precipitation. Over a longer time range this effect averages out to some extent. Therefore convection induced heterogeneity is hardly visible in monthly sums but effects records of daily or subdaily timescales.

Free convection is caused by heat induced uplift of air and moisture and occurs mostly in summer months. Due to the presence of convection, the spatial correlation in precipitation on daily or subdaily scale is always lower during summer.

In hourly rainfall data local effects are more dominant than in daily data. In high temporal resolution, the location of the stations relative to the current flow direction becomes important. For example, if one station is located ten kilometers upwind of another station, it will be hit earlier by an approaching rainfall event. As a result the hourly records of the two stations will differ considerably but show similar daily sums since it is the same rainfall that passed over both stations. However, if the distance vector of the two stations is parallel to an approaching front, both will be attained by the approaching rainfall approximately at the same time and the hourly records will be similar (Brommundt, 2008).

Since rainfall radar images have been available, the spatial and temporal heterogeneity could be examined more closely. Austin and Houze (1972) examined the spatial patterns in rain fields that were recorded during heavy rain storms in Massachusetts. They distinguish different active zones with different characteristic lifespan: “Large Meso Scale” areas as the largest structure, like for example the whole rainfall zone of an approaching front containing “Small Meso Scale Areas” which are especially active. In each of these one can find one or several “Rain Cells”, which exhibit the highest rainfall intensities. The distinction is more conceptual than physical but gives a qualitative picture of the variability, both in time and space. Since all rain events are moving, the spatial heterogeneity leads to temporal variations when the rain event passes over a certain point. The structures are not stable and the lifespan of spatially smaller structures is generally shorter (Austin and Houze, 1972) which is also contributing to temporal heterogeneity.

The airport of Oklahoma City features a dense and regular rain gauge network located in a flat and open area in which all stations are doubled for cross checking. The network is used for experimental analysis of small scale variability. The spatial correlations between the network stations are neither isotropic nor stationary (Ciach and Krajewski, 2006). Furthermore, the correlations depend on the rain rate. Storms with higher peak intensities show more spatial coherence. The highest influence on the correlations, however, comes from the width of the aggregation interval (Ciach and Krajewski, 2006). The shorter the time span is, over which the rainfall amount is summed up (e. g. hourly values compared to ten minute values), the lower the correlation is at a certain distance between the stations.

Berne et al. (2004) examine the consequences of the spatial and temporal variability on urban hydrology taking the city of Marseilles, France, as an example where a dense rain gauge network is available, as well as vertical and horizontal rain radar. Berne et al. (2004) compare the recorded intensities with discharges measured in the urban sewage system by 100

Telemeter stations. They conclude that the spatial density of the rain gauge network was not high enough to capture the spatial variability in the six minute rainfall sums. An empirical formula is given to approximate the representative range of a rainfall station depending on its temporal resolution. By comparing the measurement resolution with the concentration time of a catchment, the spatial and temporal resolution of rainfall data can be calculated that is necessary for hydraulic modeling of the discharges in the catchment. The smaller the catchment is, the higher is the required spatial and temporal resolution (Berne et al., 2004). It is found that most networks of recording rainfall stations in the Southern France are spatially too coarse (Berne et al., 2004). Rain radar on the other hand, according to Berne et al. (2004), delivers a sufficient spatial resolution but is limited temporally by the scanning interval (usually some minutes) to be of full benefit in hydrological modeling of small scale catchments. The findings, however, are limited to areas with similar Mediterranean type climate.

4.2. Intermittency and Scaling

If only one temporal scale is regarded, the occurrence and amount of precipitation is temporally persistent. On the scale of very short time intervals, e. g. five minute values, it very likely that it will be raining in the next time step if it is raining now. The same is true on daily scale: To assume that the average weather tomorrow will be "about the same as today" is a forecast that is more likely to be true than false for the temperate climate of Baden-Württemberg – even if the weather can be very variable during one day with periods of sunny and rainy conditions.

On the other hand, there are sudden changes. Within a few hours the weather can switch to a completely new situation that then again is stable for a longer time. The changes are triggered by processes on different spatial-temporal scales, e. g. shifts in the large scale westerly flow field of the upper troposphere or the synoptic scale of cyclonic depression (see Section 2.1). As a result, there is a high variety of different events from estival thunder storms to the passing of warm and cold fronts that all have a typical range of temporal scales. A thunder storm for example is physically bound to thermal uplift by solar heating and therefore will not last more than a few hours. Drizzle from a dying front can last for one or two days. The same holds true for dry periods. The information of "no rain" can mean very different things depending on the meteorological situation. An hour with zero rainfall in a time series from Central Europe can either be just a gap within a longer precipitation event, the start of a dry period lasting for several hours, e. g. between the passing of warm and cold front of a cyclonic depression, or the start of a high pressure situation that in some cases prevents rainfall during several weeks.

It is difficult to attack this behavior by statistical tools. Autocorrelation for example is an approbate measure to describe the persistence within a precipitation event, but it is not able to account for the sudden change when the event is ending. The fact that the occurrence of precipitation is at the same time temporally persistent and highly variable is statistically paradoxical. It is named by the specific term of *intermittency*. Koutsoyiannis (2006) shows

how difficult it is, to describe this behavior with standard statistical methods. The fit of a time series model for example, that describes the sequence of wet and dry time steps, is only possible with high effort in parametrization.

Since spatial and temporal variations are linked, the same is true for the spatial structure of precipitation. Rainfall fields are highly variable but not arbitrarily scattered. Rainfall clusters in active zones on different spatial scales. The spatial variability is bound by physical processes in precipitation formation: Stratiform precipitation arrives in wide bands due to large scale frontal uplift; convective precipitation occurs in scattered showers because the convective uplift channel that generates the condensation is fed by the moisture from a wider surrounding area (Weischet and Endlicher, 2008).

As a direct consequence of intermittency precipitation exhibits a very characteristic scaling behavior. Scaling means how statistical properties change when the data is aggregated for longer time intervals or over wider areas. Extreme precipitation on different temporal scales for example is caused by different type of events. In Baden-Württemberg, extreme hourly intensities are generally caused by convective precipitation events, whereas extreme daily precipitation sums are mainly due to cyclonic depressions (Section 2.1). During one hour, the rainfall intensity can be much higher than during a whole day. As a result, the relation between maximum precipitation and the aggregation time is a concave function between $1\,\mathrm{h}$ and $24\,\mathrm{h}$. Each statistical parameter, e. g. the rainfall intensity, the average and the standard deviation of precipitation sums or the rainfall probability has its own scaling properties. It depends on the spatial and temporal variability and so on the typical scales of different rainfall producing elements present in the time series.

Closely related to scaling is the question whether precipitation exhibits long term memory. The presence of long time memory means that extremely high values attract other extremely high values, and therefore, it has a strong effect on extreme value probability. (The same for extremely low values.) Long term memory can be measured by the Hurst exponent which is the exponent of a power function describing the increase in range when aggregating the values of a time series. Bunde et al. (2005) find long term memory in reconstructed precipitation records of New Mexico and relate it to the clustering of extreme events. Arnaud et al. (2007) find clustering in the most extreme precipitation events of many regions of France. However, strictly speaking, the presence of long term memory could only be proved by an infinite long time series (Lovejoy and Mandelbrot, 1985). With limited data, the Hurst exponent can only be approximated. The question whether precipitation exhibits long term memory is still subject to ongoing discussion.

4.3. Consequences

Due to the spatial-temporal heterogeneity, none of the three principle precipitation measurement tools rain gauges, rain radar and satellite remote sensing can give a complete high resolution picture of areal precipitation. Rain gauges are principally seen as the most accurate, but they can only give point values. The surrounding area for which the measurements are representative is very limited, especially for high temporal resolution. Rain radar on the

other hand can give a spatially detailed picture but are highly affected by measurement errors. Most problematic in the application of radar data is the risk of underestimating extreme events. Satellite remote sensing too is subject to many error sources and it does not deliver the spatial and temporal resolution required, e. g. for applications in urban sewage system simulation.

There is a demand for high resolution precipitation data, e, g. for urban sewage system modeling, that cannot be met by any observed precipitation data. The only solution is to bridge the gap by synthetic precipitation data generated by either physical or statistical precipitation models.

Any statistical model has to account for the high statistical complexity of precipitation. Intermittency and scaling require an adequate response in the model structure. Simple models are not able to catch all characteristic statistical features (Koutsoyiannis, 2002). On the other hand, a more complex model exhibits more parameters that have to be fitted by observed data, which makes the application to sparsely gauged regions more and more difficult. A trade-off has to be found between adequate model complexity and data demand for calibration.

Due to the scaling properties of precipitation, it is not sufficient if an applied model reproduces the statistical characteristics of precipitation at one scale. The extreme value probabilities on different temporal or spatial aggregations depend highly on the scaling behavior. Besides, the scale of the events that are crucial in terms of flood risk assessment and damage protection depends on the concentration time of the catchment. In small catchments, high local rainfall intensities during few hours are the most dangerous (Ruiz-Villanueva et al., 2012). The lower courses of big continental streams are threatened by flooding if it rains for several days over a big part of the catchment area (see for example the description of the Elbe flooding of 2002 in Haberlandt (2007)). The flood producing events in both cases are not the same. The events that bring the highest intensities are different from the events that bring the highest total volume. A model that is only adapted to one scale cannot represent all of these events at once. Hence, there is a high risk that it will fail to estimate the probabilities of the crucial extreme events correctly.

5. Modeling of Precipitation

Modeling of precipitation is the attempt to find a numerical representation of the natural precipitation process. This representation, as in every model, necessarily implies simplification. The physical processes in the formation of precipitation are way too complex, and partly still unknown, to be completely described by a model. And even if a complete physical description was possible, it would not be feasible to provide all necessary measurement data to define the initial state of all variables.

There are generally two types of precipitation models. Physical models, also called dynamical models and conceptual / statistical models.

Physical models try to capture the physical laws that govern precipitation as completely as possible. The major simplification concerns the spatial resolution of such models, restricted to a spectral grid. Since the formation of precipitation is linked to many other atmospheric variables as well, e. g. temperature, humidity or the wind field, a full meteorological model is required for this task. Due to the high complexity, physical precipitation modeling can only be provided by public or private weather services that have both, the required physical knowledge to represent the atmospheric processes in the model and the necessary CPU power to run it. Regarding the effort in data acquisition by ground stations and satellites for accurate precipitation forecast and that forecasts are still limited to several hours or few days (Bliefernicht, 2010), one can estimate the complexity of this task.

The main application of physical models is in (quantitative) precipitation forecast for weather reports, reservoir management or as a service for farmers. In reanalysis projects, measurements from different data sources are integrated in a physical model for spatial extrapolation from point sources to areal precipitation on a homogeneous regular grid. Global and Regional Circulation models (GCM and RCM) used in climate research are also physical meteorological models but, due to the long forecasting time of up to 100 years, on a reduced spatial-temporal resolution. It should be mentioned that physical models are not purely physical. They also depend on conceptual parameters for quantities that cannot directly be measured. Examples are the leaf area index and fraction of green vegetation in GCM that are necessary to describe the average albedo due to plants (Roeckner et al., 2003, for the model ECHAM5).

Conceptual precipitation models do not consider any physical laws. They are based on the statistical characteristics of precipitation. Most models try to mimic certain statistical properties of precipitation that are assumed to be the most essential. This can be for example the memory in precipitation time series (Section 5.3) or the scaling between precipitation statistics on different spatial-temporal resolutions (Section 5.6). The idea is that if the considered statistical characteristic reflect the (unknown) physical nature of precipitation as close as

possible also other properties that were not used in the model set up will be reproduced correctly by the model. Conceptual models require parameter calibration for the actual climatic conditions that can only be done by observed precipitation data.

As physical models are not applied in the generation of long synthetic precipitation time series, which is the aim of this work, the following review of precipitation models will only consider conceptual models.

5.1. Purpose of Precipitation Modeling

The modeling of precipitation is always performed in the focus of a hydrological or hydraulic application that uses the synthetic precipitation data, e. g. for dimensioning of hydraulic structures or flood frequency calculations.

According to the complex scaling of precipitation, the performance of precipitation models is scale dependent. Since they can only give an incomplete picture of the precipitation process, precipitation models usually perform best on the scales that were used in calibration to observed precipitation and worse on scales of very different spatial or temporal aggregation. As a consequence, the applied precipitation model has to be chosen carefully in regard of the hydrological or hydraulic application that it is used for. Many models are even designed for one specific purpose.

The application also defines how closely the precipitation model can stick to measured precipitation data. In the order of increasing abstraction from observed data, these tasks are the following:

1. **extension of time series, ensembles of different realizations**

   In some cases it is only necessary to increase the data volume of a given precipitation time series. Either the time series is extended, e. g. for a better extreme value estimation, or an ensemble of different possible time series is created, that can be used for example in probabilistic precipitation discharge models.

   From the precipitation modeler's point of view, this is the easiest task since the parameters of the precipitation model can be derived from the time series itself. No indirect parameter estimation, neither the transfer from other locations nor from higher aggregated measurements is needed.

2. **disaggregation and downscaling**

   Precipitation measurements on coarser scales are by far better available than measurements on finer scales. Reanalysis for example provide long time series of gridded areal precipitation that cover wide regions or even the whole globe. An example are the NCEP/NCAR reanalysis which provide a continuous time series of global precipitation fields from 1948 till today for $24\,\mathrm{h}$ or $6\,\mathrm{h}$ intervals on a $2.5° \times 2.5°$ grid (Kistler et al., 2001). The precipitation time series from climate models are available on similar scales. For most hydrological or hydraulic applications however, the resolution of these data sets is not sufficient.

It is a common situation that coarse scale data has to be brought to a finer scale for the use in a hydrological or hydraulic application. A spatial scale change is called "downscaling", a temporal scale change "disaggregation". A precipitation model used for disaggregation or downscaling depends on parameters that link the coarser scale statistics to the finer scale. It is often the case that fine scale measurements are only available for a short time period or for one locations. Thus, the model parameters are estimated by means of the available data and considered constant in time and/or transferable to new locations.

Not all precipitation models can directly be used for downscaling. Some models, generate free simulations that cannot be conditioned on coarse scale observations. A possible work-around, e. g. in temporal disaggregation, is to run long simulations and cut out pieces of the time series that by chance fit to the coarse scale values.

3. **generation for ungauged areas**

   This is the most ambitious task in precipitation modeling. It consists not only of the setup of a suitable precipitation model but also in the transfer of all necessary parameters of the model to the new, ungauged location. Unlike the task of downscaling and disaggregation, no further information on the precipitation regime at the new location is available that guides the simulation. The modeler might enjoy the freedom of simulating precipitation time series without any constraints, but he or she should keep in mind that the risk of generating inadequate simulations is highest in this task.

   It becomes even more challenging for spatial-temporal simulation, for example of several, spatially correlated time series. In this case not only the model parameters have to be transfered, but also the spatial dependence between the time series has to be estimated at the new, ungauged locations.

As a consequence of the spatial-temporal heterogeneity of precipitation and the scaling behavior, modeling becomes the more challenging, the higher the required spatial-temporal resolution is. This is common to all models and all modeling tasks. For most model concepts, the number of parameters increases dramatically with increasing temporal resolution. As a result, many publications can be found on the generation of daily precipitation time series. Models for hourly or even five minute time series are much less frequent.

5.2. Spatial Interpolation of Precipitation Data

Being confronted with a lack of precipitation data, the first idea one would probably come up with is interpolation: If a location is ungauged, it might be possible to approximate the missing values by taking a weighted average of the values from surrounding stations. Doing this for each time step, one could set up an interpolated time series.

An appropriate interpolation method is Kriging, which has been developed for the mining industry. A typical problem in mining is that there are some point measurements from test drillings, e. g. of the thickness of an metal ore deposit and one has to decide from these if

the deposit is worth exploiting. With Kriging the point measurements can be interpolated to locations that have not been probed.

Kriging is a linear estimator. The interpolated value $\bar{z}_0$ at the target location is a linear combination of the measurement values at surrounding measurement points:

$$\hat{z}(x_0) = \sum_{i=1}^{n} \lambda_i \cdot z(x_i) \tag{5.1}$$

with $\hat{z}(x_0)$ estimation for location x_0
$z(x_i)$ measurement at location x_i
λ_i weight of the measurement at location x_i

Eq. (5.1) holds true for many linear interpolation methods. The difference between them, is how the weights λ are defined.

Kriging can be applied if the underlying field Z from which the samples z are drawn exhibits intrinsic stationarity. It means that the expectation of Z is independent of the location and that the expectation of the variance between two measurements at locations x_i and x_j depends only on their Euclidean separation distance h_{ij}:

$$E[Z(x_i) - Z(x_j)] = 0 \tag{5.2}$$

$$\frac{1}{2}E\left[(Z(x_i) - Z(x_j))^2\right] = \gamma(h_{ij}) \tag{5.3}$$

where γ is called the variogram.Under the condition that intrinsic stationarity in Z is fulfilled, the weights λ are chosen so that the estimation is unbiased and that the expected variance between the true value $Z(x_0)$ and the estimated value $\hat{Z}_0$ is minimal:

$$E\left[\hat{Z}_0 - Z_0\right] \stackrel{!}{=} 0 \tag{5.4}$$

$$E\left[\left(\hat{Z}_0 - Z_0\right)^2\right] \stackrel{!}{=} \min \tag{5.5}$$

It follows from the unbiasedness condition Eq. (5.4) that the weights λ have to sum up to one.

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{5.6}$$

Inserting Eq. (5.1) and exploiting the definition of the variogram γ (Eq. (5.3)), Eq. (5.6) can be transformed to

$$E\left[\left(\hat{Z}_0 - Z_0\right)^2\right] = -\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma(h_{ij}) + 2\sum_{i=1}^{n} \lambda_i \gamma(h_{i0}) \tag{5.7}$$

where n is the number of measurements that are used in the interpolation.

Minimizing Eq. (5.7) as a function of the weights λ_i under the unbiasedness condition Eq. (5.4) leads to the linear equation system of Ordinary Kriging

$$-\sum_{j=1}^{n} \lambda_j \gamma\left(h_{ij}\right) + \nu = -\gamma\left(h_{i0}\right),\ i = 1, 2, \ldots, n \tag{5.8}$$

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{5.9}$$

The Kriging System consists of $n + 1$ equations with $n + 1$ unknowns. The variable ν is the Lagrange multiplier which is necessary to link the unbiasedness condition. For the variogram γ empirical values can be used or it can be estimated from a theoretical function that is fit to the observed data.

Minimizing the sample variance, Kriging is the best unbiased linear estimator ("BLUE") for the unknown value at x_0 under the assumption that intrinsic stationarity is fulfilled and that the error variance (Eq. (5.5)) is a meaningful estimator. The latter condition implies that the distribution of Z should be symmetric and approximately Gaussian.

To apply Kriging on precipitation fields is difficult since precipitation amounts exhibit right skewed distributions. The possibilities to account for the skewness are limited, e. g. by logarithmic transformation of the data. Therefore, Kriging works best for higher aggregated data like monthly sums where the skewness is lower. On short aggregation intervals, the occurrence of zero values becomes a problem. Like any interpolation the Kriging estimations $\hat{z}$ have a lower variance than the underlying field Z. Hence, the Kriging estimations underrate the fraction of zero values.

A possible solution to the skewness problem is Indicator Kriging. Instead of the value of Z, the indicator $I_C\left(x_i\right)$ is interpolated that indicates if the value $z\left(x_i\right)$ is within the class C or not. If "yes", $I_C\left(x_i\right) = 1$; else $I_C\left(x_i\right) = 0$. The interpolation result at x_0 is the probability that $z_0 \in C$. If Indicator Kriging is performed for several rainfall amount classes, the distribution of Z at x_0 can be approximated (Haberlandt, 2007, , for the interpolation of the precipitation field causing the Elbe Flooding in 2002).

Kriging can also be improved by incorporating the additional information of an external variable Y that is linked to Z and available at any measurement location x_i as well as the target location x_0. In External Drift Kriging it is assumed that Y are linked linearly:

$$E\left[Z\left(x_0\right)|Y\left(x_0\right)\right] \tag{5.10}$$

where $E\left[Z\left(x_0\right)|Y\left(x_0\right)\right]$ is the expectation of $Z\left(x_0\right)$ given that the value $Y\left(x_0\right)$ is known.

The linear equation system for External Drift Kriging is

$$-\sum_{j=1}^{n} \lambda_j \gamma\left(h_{ij}\right) + \nu_1 + \nu_2 Y\left(x_i\right) = -\gamma\left(h_{i0}\right),\ i = 1, 2, \ldots, n \tag{5.11}$$

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{5.12}$$

$$\sum_{i=1}^{n} \lambda_i Y\left(x_i\right) = Y\left(x_0\right) \tag{5.13}$$

where ν_1 and ν_2 are Lagrange multipliers.

Haberlandt (2007) uses the external information of rain radar fields for External Drift Kriging of the precpitation amounts measured during the Elbe flooding in 2002. During this event, the rain radar fields showed severe underestimation up to a factor of six (Haberlandt, 2007). Nevertheless, even if the absolute values of the radar data are wrong, the information on spatial variability improves the interpolation.

Indicator Kriging and External Drift Kriging can be combined: Whether it is raining at x_0 or not, is decided by Indicator Kriging. If yes, the amount of precipitation is interpolated by External Drift Kriging (Verworn and Haberlandt, 2011).

It should be noted that the cited investigations dealt with the interpolation of single precipitation fields. For the generation of rainfall time series interpolation is not appropriate. Kriging is not able to take temporal persistence into account. The generation of a time series by stepwise interpolation results in a time series that overestimates the autocorrelation. At the same time, the variance reduction leads to an underestimation of extreme events (Bárdossy et al., 2000). Due to the scaling behavior, the deviations are highest for short aggregation intervals (e. g. five minute values).

The variance reduction can be compensated by stochastic simulation that randomly disturbs the estimated values to restore the original variability. However, the simulation has to take the skewness of precipitation into account, as well as the temporal persistence.

5.3. Markov Chain Models

Markov Chain models do not represent the continuous nature of precipitation but discrete time series of equal measurement intervals (e. g. daily values or hourly values). The Markov Chain produces a binary sequence of rainfall occurrences expressed as the states “rain” and “no rain” at each time step. For a complete simulated precipitation time series the rainfall amounts for the wet intervals have to be modeled, e. g. by drawing from a distribution function.

Let X be the random variable of rainfall occurrence that takes two values 0 and 1:

$$X_t = \begin{cases} 1 & \text{if it rains at time step } t \\ 0 & \text{if not} \end{cases} \tag{5.14}$$

Then a Markov chain is defined by the following equation:

$$P_r(X_t = i) = \sum_{i,j,k \ldots w} P\left(X_t = i \middle| X_{t-1} = j, X_{t-2} = k, \ldots X_{t-n} = w\right) \tag{5.15}$$

P_r	probability for $X = i$ ("rain" or "no rain") at time step t
$i, j, k \ldots w$	states of X at the n previous time-steps

$P_r(X_t = i)$ is a conditional probability which depends on the state of X at n previous time steps. n is the order of the Markov Chain model. More descriptive n is the number of previous time steps that have to be known to determine the rainfall probability at time step t. The conditional probabilities are assumed to be constant over time. The Markov Chain Model for the occurrence of precipitation is a two-state Markov chain since the random variable X can only take two values, 0 and 1. In this case, the sum in Eq. (5.15) consists of 2^{n+1} elements that one has to estimate to set up a Markov Chain model. They are called the transition probabilities (e. g. Chin, 1977) because they describe the transition from one state ("rain" or "no rain") to another state at the next time step. The transition probabilities can be approximated by counting the transition frequencies in observed precipitation time series. Thereby, some of the transition probabilities can be deduced by complementary probabilities which reduces the number of independent parameters so 2^{n-1} where n is the order of the Markov Chain (Koutsoyiannis, 2006). For a first order Markov Chain Model, the transition probabilities can also be deduced from the unconditional probabilities for the states of "rain" and "no rain".

The Markov Chain Model determines many statistical properties of the generated bivariate time series of the rainfall occurrence. For a first oder Markov Chain the expected number of consecutive wet (or dry) time steps follows an geometric distribution (Gabriel and Neumann, 1962). The autocorrelation on different time lags follows an exponentially declining function. These characteristics can be used to verify a Markov Chain Model. For more details on the properties of first order Markov Chain Models one might refer to Katz (1974) or Lloyd (1974).

The computation of first order Markov Chain time series is straight forward and fast. These qualities made the models very popular when numerical calculations were not yet as available as today. Gabriel and Neumann (1962) applied a first order Markov Chain model to the daily rainfall occurrence in Tel Aviv. They used a threshold of $0.1\,\mathrm{mm}$ for the limit between "rain" and "no rain". Since the model assumes constant transition probabilities, it is restricted to the rainy season in winter months.

Todorovic and Woolhiser (1975) present a full model for daily precipitations including a simulation of the rainfall amounts which are drawn from an exponential distribution. The amounts of consecutive days are assumed as independent. The model is verified by the distribution of the maximum rainfall amount in n consecutive days, as well as the mean and variance of the total precipitation sum during n days for which they deduced analytical solutions. At the test location in Austin, Texas, the model shows deviations in the statistics for longer aggregation intervals. The average 10 day precipitation sum is overestimated, the

distribution of the 20 day and 30 day sums have lack of variance compared to the observations. Katz (1977) modified the approach by assuming precipitation amounts as Gamma distributed with two different sets of parameters: one for the first day in a row of consecutive wet days and one for the following days.

Katz and Parlange (1993) refined the Markov Chain Model by using information on atmospheric circulation to improve the adjustment of the transition probabilities. The approach was tested with the rainfall data from several daily stations in California. Two amtmospheric Circulation Patterns (CPs) are defined based on the mean sea level pressure map. The transition probabilities π of the occurrence model as well as the parameters of distribution describing the rainfall amounts were set up for each CP separately. The model was verified by the simulated and observed variance of monthly rainfall sums. Using the CP classification, reduces the lack of variance in higher aggregation sums remarkably compared to a Markov Chain model which is not conditioned by the CP.

The model proposed by Haan et al. (1976) avoids the division into an occurrence and an intensity model part. The advantage is that the memory of the time series is directly expressed in the rainfall amounts and not only in the occurrences. It is realized by defining six rainfall amount classes. Adding the dry state, a first order Markov Chain with seven different states is set up leading to 7×7 transition probabilities. The disadvantage of this model is the high sensitivity to sampling errors. To avoid seasonal effects the model is set up for each month separately which leads to $12 \times 7 \times 7 = 588$ transition probabilities that have to be estimated from the observed time series. Even with time series of several decades as data input the estimation of the rarest transition probabilities is not stable (Haan et al., 1976).

If the statistical properties of observed rainfall differ significantly from first order Markovian properties, a Markov Chain of higher order has to be applied. Higher order Markov Chains are able to represent more complex temporal persistence. The autocorrelation over different time lags is more flexible than in the first order chain were it is bound to exponential decline. The higher flexibility helps to reduce deviation in aggregated properties over n consecutive time steps as in Todorovic and Woolhiser (1975).

In an extensive study with more than one hundred 25 year long time series from rainfall stations all across the United States of America Chin (1977) proofed that a first order Markov Chain is not alway sufficient to describe daily rain fall occurrence. The required Markov Chain order depends on the region and most of all on the seasons. In summer months a first order chain is sufficient in most of the United States. In winter months however, the use of a second or third order model is required. In summer months rainfall comes primarily from local thunder storms with low temporal persistence. During winter season, the rainfall in the US is mostly due to polar fronts and cyclonic depressions which have a characteristic life cycle of about four days which leads to higher order autocorrelation in the rainfall occurrence (Chin, 1977).

Precipitation on the sub-daily temporal scale generally shows higher autocorrelation than daily values and a more complex autocorrelation structure. Therefore, modeling requires sophisticated higher order Markov Chain Models. Katz and Parlange (1995) construct a third order Markov Chain Model to simulate time series of hourly precipitation. Daily cycles in the precipitation activity, e. g. during summer months, can be taking into account

by applying a sinus-cosinus-function with a period of $24\,\mathrm{h}$ to the transition probabilities. The amounts in consecutive hours are assumed as autocorrelated. Due to the higher order Markov Chain and the additional parameters for daily cycle and persistence, the model is heavy in the number of parameters. The estimation is burdensome since the parameters cannot directly be derived from the observations. Katz and Parlange (1995) verify the model by the rainfall probability and rainfall sum on daily aggregation. While the daily averages are modeled correctly, the simulated time series exhibit a lack of variance in daily values. Katz and Parlange (1995) conclude that "virtually all conceptual models tend to underestimate the variance of daily total precipitation", a phenomena that is often referred to as "overdispersion" (e. g. in Katz and Parlange (1998) or Koutsoyiannis (2006)).

Katz and Parlange (1998) test different methods to reduce overdispersion in daily Markov Chain Models. One possibility is to draw autocorrelated rainfall values, another to apply different distributions at the first and the following rain days (as in Katz, 1977) or by applying a higher order Markov Chain occurrences model. The overdispersion is tested by different statistics on longer (e. g. monthly) aggregation as the distribution of the number of wet days as well as the mean and variance of monthly rainfall sums. The lack of variance in the number of wet days can be reduced by applying a higher Order Markov Chain. However, the lack of variance in aggregated sums, does persist (Katz and Parlange, 1998).

All models presented so far simulate single time series for one location. There are only few approaches published that extent the Markov Chain concept to simulate simultaneous, spatially and temporally correlated time series (space-time models). Wilks (1998) sets up a Markov-Chain Model for spatially correlated daily precipitation values. The model is restricted to precipitation stations whose time series can be described by a first order Markov Chain, which is the basis of his approach. Spatially correlated random numbers are drawn which decide over the state of rainfall occurrence at all sites simultaneously. The rainfall amounts are simulated by a mixed exponential distribution. The mixing weight is linked to the rainfall occurrence. Thus, for stations at the edge of a rainfall field the expected amount is lower than for a station in the middle of a rainfall field, which is consistent with real world observations. However, the model has the issue that some of the parameters cannot be calibrated directly from observations which necessitates empirical estimation functions. Furthermore, the model assumes that the rainfall amounts are temporal uncorrelated, which leads to an underestimation of the spatial correlations with time lag between the stations.

Markov Chain Models are conceptually simple. In the same time, the model structure determines many statistical properties that can be used for verification. It is a strong point that the model can be extended by additional information as diurnal cycles (Katz and Parlange, 1995) or weather type information (Katz and Parlange, 1993). The fact that the model produces discrete time series is an advantage in parameter fitting. The model can directly be linked to discretely measured precipitation time series by the observed transition frequencies. The parameters are not "hidden" as it is the case for Point Process Models, which are discussed in the next section (Section 5.4).

Markov Chain Models are most suitable for the generation of point rainfall. The extension to spatial-temporal modeling is difficult. In areal precipitation the amount of rainfall is dependent on the size and shape of the rainfall field. The separate simulation of rainfall occurrence

and rainfall amounts that is applied in most Markov Chain Models causes problems in this case.

Most applications of Markov Chain Models deal with daily rainfall. In most climates the sequence of wet and dry days can well be described by a Markov Chain of first or second order. On the sub-daily scale however, the temporal pattern of rainy and dry time steps becomes more complex. The occurrence of intermittency and long term persistence (Section 4.2) requires higher order Markov Chain Models. However, the fit of a Markov Chain Model of order n involves the estimation of 2^{n-1} parameters (Koutsoyiannis, 2006). The demand for observed data to fit all these parameters increases over-linearly with the model order. Since the model assumes stable conditions, it has to be fit to each month separately if the climate exhibits an annual cycle. The high demand of observation data is a weak point of Markov Chain Models. Geng et al. (1986) state "it usually is believed that at least twenty years of data should be used to provide reliable estimates." Chin (1977) states that the model order is underrated if the data series is too short for an adequate fit. In long time series, on the other hand, it becomes doubtful if the assumption of constant transition frequencies can still be accepted.

Additionally, the high number of parameters makes it difficult to use Markov Chain Models in ungauged areas. None of the cited publications describes a method to transfer the transition probabilities found at one rain gauge to a new location.

5.4. Point Process Models

In contrast to Markov Chain Models, Point Process Models describe rainfall (as a temporal process at one site or as a spatial-temporal process over an area) in its continuous nature. Rainfall is seen as an overlay of several elementary rain "cells". The arrival of the cells, as well as the duration and intensity of each cell are seen as random variables and stochastically modeled. To account for the intermittency of precipitation, Point Process Models exhibit a hierarchical structure of "storms" and "cells". Storms arrive with a certain Poisson distributed arrival rate. Within each storm a random number of rainfall "cells" is active. (A zero-rainfall interval in this model can either be a break within a storm with a short expected duration or a possibly longer break between two storms.) The instantaneous rainfall rate at any time point t_0 is the sum of the intensities of all cells that are active (Rodriguez-Iturbe et al., 1987). If the cells are assumed to have a constant intensity over their whole life span, the model is called a "Rectangular Pulse Model" (Rodriguez-Iturbe et al., 1987).

Two very similar models of this kind are the Bartlett-Lewis Rectangular Pulse Model and the Neyman-Scott Rectangular Pulse Model (Rodriguez-Iturbe et al., 1987). The difference between the Bartlett-Lewis Rectangular Pulse Model and the Neyman-Scott Rectangular Pulse Model is only in the definition of the rain cell arrival. In the Bartlett-Lewis model cells arrive according to a Poisson process as long as the current storm is active. In the Neyman-Scott Model at first the number of cells belonging to each storm is drawn from a random distribution. Then the location, symmetric around the storm center, is modeled (Rodriguez-Iturbe et al., 1987).

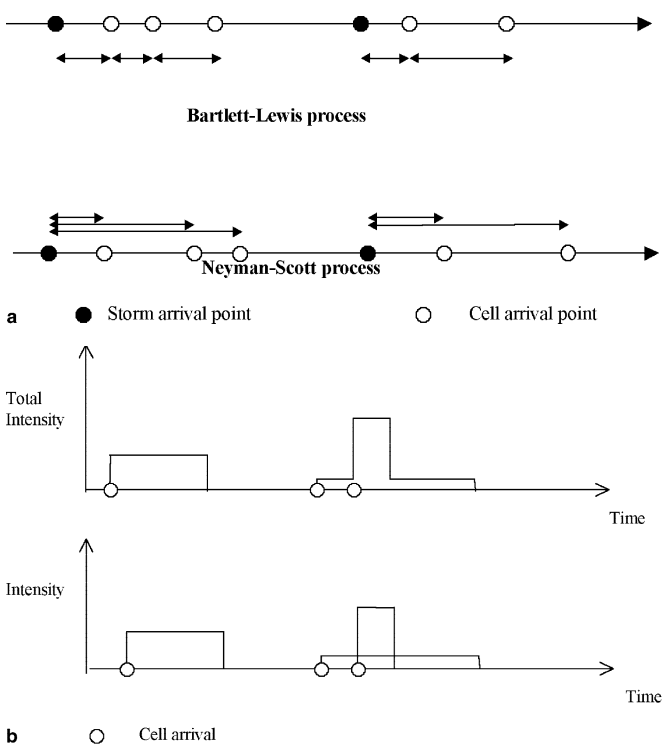


Figure 5.1.: Schematic for Point Process Rectangular Pulse Model.
On top: storm and cell arrival in the Bartlett-Lewis and the Neyman-Scott Process. Below: intensity superposition in Point Process Models.
Scheme by Onof et al. (2000)

Since the rainfall generating cells are overlaying, they cannot be identified in observed precipitation time series. Therefore, the parameters of Point Process Models (e. g. of the distributions for the rain cell intensity and the duration) cannot be derived from observations directly. Generally, Point Process Models are fitted by linking the parameters to observable statistical properties. Rodriguez-Iturbe et al. (1987) for example set up a Neyman-Scott Model assuming the number of cells per storm as geometrically distributed, cell intensity and duration as exponentially distributed. In this form of the model the mean, the variance and the autocorrelation of the generated time series can be expressed as a function of the generation parameters and fitted to observed statistics. Another possibility is to vary the model parameters in a numerical optimization scheme until the properties of the simulated time series are in agreement with the observations (Onof and Wheater, 1993).

However, the model of Rodriguez-Iturbe et al. (1987) differs to real world observations in the scaling properties. If the model is fitted repeatedly with the same data but aggregated on different scales (from 1 h to 1 day), it results in different model parameters (Foufoula-Georgiou and Guttorp, 1986). They conclude that "there is not a unique Neyman-Scott Model" which, when fitted to one time scale, reproduces the observed statistical moments on other temporal scales.

To improve the consistency between modeled statistics at different temporal aggregations,

Calenda and Napolitano (1999) propose an alternative parameter fitting procedure. The model parameters are expressed as a function of the scaling in the variance of the rainfall intensity over different aggregations. Parameter fit is done by non-linear regression over the scales.

Onof and Wheater (1993) modify the generation algorithm of the Neyman-Scott Model by replacing the exponential parameter in the distribution of the rain cell duration by another random function. As a result, the cells of each storm follow a different cell duration distribution. The increased variability improves the representation on scales that are not used in the fitting. Cowpertwait (1994) enhanced the variability of the model by introducing different cell types. He distinguishes between "heavy short-duration convective cells" and "light long-duration stratiform cells" with different distribution parameters for duration and intensity.

Evin and Favre (2008) link the intensity and duration of the rain cells, which are assumed to be independent in the original set-up, by cubic copulas. The deduction of method of moment estimators for the model parameters is extremely tedious and only possible if it is assumed that both, duration and intensity follow an exponential distribution. However, the benefit of the model concept extension, e. g. on the simulation of extreme values on different scales, is small.

For a better representation of the intemittency of five minute values Cowpertwait et al. (2007) extended the model concept to a three level structure: The rainfall time series is modeled as a succession of "storms" that contain a random number of "cells" which consist of several precipitation "pulses". The model in this form has eleven parameters that are fit by the coefficient of variation, the autocorrelation, the skewness on four different aggregation levels.

Arnaud and Lavabre (1999) enlarged the concept of a Point Process Model in a less conceptual but more heuristic sense. Instead of modeling complete time series they concentrated on extreme rainfall events. The presented model, named "SHYPRE"´, is designed for the use in hydraulic dimensioning and flood protection measures. The aim is to simulate very long time series (e. g. 10000 years) and to calculate extreme value statistics based on the empirical frequencies in these simulations. All rainfall events (defined as a period of consecutive rain days) with less than 20 mm are ignored. Compared to former Neyman-Scott approaches, the model has several new parameters. Most important, it does not work with rectangular but triangular pulses. Therefore the ratio of peak intensity to average intensity and the relative peak position in each rain cell have to be modeled as random variables. Furthermore, it distinguishes between "major storms" and "minor storms" which have different parameter distributions. This is necessary to adjust the model to observed temporal persistence.

In 2002 SHYPRE was combined with a discharge model to investigate extremes in river discharges and to analyze errors in the whole modeling chain from rainfall to runoff (Arnaud, 2002). Arnaud et al. (2007) proved the performance of the generator in all available climate zones of France and its oversea territories. Arnaud et al. (2007) point out that the model performance is highly dependent on the right parametrization of storm persistence. They show that major storms in one event are often connected, the presence of the first major

storm enhances the probability for another one. Ignoring this connection, leads to a (potentially dangerous) underestimation of extremes. Muller et al. (2009) made an extensive study of SHYPRE's sensitivity to parametrization and fitting data. Uncertainties in the parameters were modeled in a Bayesian framework by applying a prior distribution to each parameter and checking the range of the extremes in the resulting time series. Lack of data was simulated by leaving out some of the calibration data and checking for differences in the simulation results. It could be shown that SHYPRE give reasonable results in higher annuities up to 100 years with 20 years of data.

However, SHYPRE suffers from some minor drawbacks. Firstly, it does not model a complete time series but only the most extreme events in the year. This limits the possible applications, e. g. it is not suitable for sewage system design where more frequent events are more important than extremes of high annuity. Secondly, its heuristic character can be seen as a conceptual weakness. Especially the definition of "major" and "minor" storms and their persistence parameters are rather arbitrarily and not justified by any physically based or phenomenological concept of rainfall.

The extension of the Point Process Model into spatial domain is conceptually easy. For spatial modeling it is sufficient to consider storms and cells as two dimensional features (Rodriguez-Iturbe et al., 1986). Storms arrival is modeled by a spatial two dimensional Poisson Process. The cells cluster around the storm centers according to the Neyman-Scott Process with a random number of cells per storm. The cells are seen as circular discs with a maximum intensity at the cell center and declining intensity towards the edge. The decline can be modeled as a random function as well (Rodriguez-Iturbe et al., 1986). For a fully temporal-spatial model the movement and the evolution of storms and cells has to be taken into account. Rodriguez-Iturbe and Eagleson (1987) extend the model in this sense by adding another exponential decline function describing the evolution of rain cell intensity in time.

Cowpertwait (2002) follow the same idea to set up a space-time rainfall model for the UK. They assume the rainfall intensity in each cell to be constant and to follow a Weibull distribution. The fit of the model is done by observed statistics like the variance, the coefficient of variation, the skewness, the autocorrelation and the cross correlation in observed rainfall records. It is not straight forward and involves an iteration scheme that passes by the fit of single site models which are, in a second step, linked by means of spatial statistics.

Cowpertwait (2006) applied the model in a multipurpose scheme to the Thames Catchment in the UK. It is designed for disaggregation, gap-filling in measured data series and the generation of time series at ungauged locations. However, the model cannot directly be used for gap-filling and disaggregation because it is not possible to condition the model on observed values (e. g. for the daily rainfall sum). Therefore an indirect method is applied in which 200 years of data are generated in a free simulation and then slices are cut out according to similarity measures (Cowpertwait, 2006). This technique might become difficult for extreme events where only few similar situations are available in the generated time series. The model is available under the name of "RAINSIM", commercially distributed in a software package called "STORMPAC". It is widely used by the UK water industry but also applied in alpine climate in Italy or Switzerland, in the Dutch Lowlands and the arid climate

of Spain (Burton et al., 2008).

Willems (2001) proposes a spatial-temporal generation scheme that takes the movement of cells and storms explicitly into account. The rain field is represented as Small Meso Scale Areas (the cells) clustered in Large Meso Scale areas (the storms) within a Large Synoptic Scale Areas (e. g. the front of a cyclonic depression). The movement of the cells is processed similarly to mass-conservative diffusion: As a cell moves, it becomes broader and its maximum intensity decreases. For the model fit the rain cells have to be identified in observed precipitation time series. The propagation velocity is determined by tracking the rain cell movement from gauge to gauge (Willems, 2001). Therefore, the model is very data demanding. Correct tracking requires a very dense rain gauge network that is at the same time large enough to represent the Large Synoptic Scale Areas completely.

Point Process Models are frequently applied and well tested in a wide range of different climates (see Onof et al., 2000, for an overview). Their strong point is the continuous nature of the simulated rainfall process. The model can be fit with aggregated data (e.g. in daily resolution) and still deliver time series on high temporal resolution (e. g. hourly data). Another quality is that the extension to spatial-temporal modeling is conceptually easy. The same holds true for the regionalization of the model. It is sufficient to transfer the statistics of the aggregated data that were used in the parameter fitting to the ungauged locations. Cowpertwait et al. (1996) for example regionalized a Neyman-Scott Model over Great Britain by linking the parameters to geographic related variables like altitude, distance to the distance to the coast or a location index for the west or the east side of Britain.

A weak point of Point Process Models is the parameter estimation, which is burdensome since the model parameter values are not directly observable. At the beginning, much effort was made to find analytical equations (or equation systems) that link the model parameters to observable statistical properties (e. g. Rodríguez-Iturbe et al. (1987). With the improvement in computer power in recent year the focus shifted towards parameter fitting by numerical optimization (Burton et al., 2008). Vanhaute et al. (2012) compared different optimization methods and their performance in the parameter fit of a Bartlett-Lewis model. It shows that 25 years after the first use of Point Process Models for rainfall time series, the right calibration is still an issue. Katz and Parlange (1995) critisize that Point Process Models assume climatic stationarity and cannot account for additional information as djurnal cycles. They state that it would "make already difficult parameter estimation problems more complex, if not infeasible".

Another issue is, that the concept of "storms" and "cells" gives a rather rough picture of the real physical rainfall process and that the modeling results, especially in extreme rainfall, are sensitive to the assumed persistence in storm and cell arrival (Arnaud et al., 2007). Over the Eastern Atlantic Ocean and most of Europe, the storms are clustering (Mailier et al., 2006). The assumption of Poisson distributed storm arrival rates is doubtful. Mailier et al. (2006) tracked the paths of all cyclonic storms over the North Atlantic and over Europe during the winter seasons of 53 years. From the emerging storms over the Western Atlantic, not all storms are able to reach the European continent. However, if the first storm manages to pass through, it enhances the probability for its successors. Vitolo et al. (2009) prove that the clustering is even more pronounced in the most intensive storm events.

There are approaches that take the persistence or the movement of events into account. However, they rely on methods and parameters that are more heuristic than physically justified. An example is the division into "major" and "minor storm" in the SHYPRE model family (Arnaud and Lavabre, 1999) or the rain cell identification algorithm proposed by Willems (2001).

5.5. Alternating Renewal Models

In contrast to Point Process Models, Alternating Renewal Models do not simulate the arrival rate of overlaying storms and cells but explicitly the duration of "wet" and "dry spells". Each wet spell is followed by a dry spell. The length of both are drawn randomly from a theoretic distribution function. Similar to the hierarchic concept of storms and cells in Point Process Models, Alternating Renewal Model exhibit an external and an internal structure. The external structure is the sequence of wet and dry spells. While the dry spells are completely dry, the wet spells consist of different rain pulses that can be separated by short dry periods in between. The internal structure is characterized by the rainfall frequency within the wet spell, the shape of the pulses, as well as the distribution of intensity and duration of the pulses.

The model of Acreman (1990) for an 1 h precipitation time series in Farnborough, UK is an early example of an Alternating Renewal Model. In his concept dry spell lengths are modeled by a Generalized Pareto Distribution, wet spells lengths by an Exponential Distribution. The internal structure is kept simple as dry intervals within the wet spells are not considered. The wet spells are assumed to be bell shaped, the total precipitation sum to follow a Gamma distribution conditioned on the duration. The model was validated in terms of the extreme values on different aggregation levels ranging from 1 h to 48 h. On any aggregation level, the model captures the shape of the extreme value distribution but systematically underestimates the extremes of high annuity.

Bernardara et al. (2007) set up Alternating Renewal Models for different 5 min and 1 h precipitation stations in Italy. They model the external structure by two Generalized Pareto Distributions for the length of wet and dry spells, and the additional parameter of a threshold length that determines if a zero rain interval is regarded as a gap in a wet spell or as its own dry spell. The internal structure is consists of a mixed distribution of Generalized Pareto type representing the intensities and the rainfall probability within the wet spell. The distribution of the rainfall pulse intensities is modeled by its spectrum. The spectrum loads are assumed to be power-law dependent on the spectrum frequency. In validation by the observed time series the model exhibits a slight overestimation of the annual precipitation volume and the expected average total volume per event. The standard deviation of the total event volume is highly overestimated. On the other hand, the distribution of wet fractions is well represented as well as the extreme value behavior in terms of depth duration frequency curves over a wide range of scales from 1 h to 36 h.

The model of Haberlandt et al. (2008) uses the Alternating Renewal Approach to simulate 1 h precipitation time series for the Bode catchment in the Harz mountains in Middle Germany.

The wet spell length is modeled by a General Extreme Value Distribution, the length of dry spells by a Weibull distribution. The intensity of the rainfall pulses is modeled by a Kappa Distribution. A particular feature of this model is that the rain pulse duration and its average intensity are assumed to be dependent, modeled by a two dimensional Frank Copula.

Compared to Point Process Models, Alternating Renewal Model have the clear advantage that there is no overlap between the model elements. The length of wet and dry spells as well as the properties of the rainfall pulses within the wet spells can be observed in recorded precipitation time series. Therefore, the model calibration is straight forward allowing a higher flexibility in the model concept, for example in the shape of the rainfall pulses. Regionalization of the generation parameters works principally in the same way as for Point Process Models by spatial interpolation of the model parameters (Haberlandt et al., 2008).

The model structure considers no memory in the sequence of alternating wet and dry spells. Per definition two following wet spells are independent events. However, whether this condition is fulfilled, depends on the length of the dry spell in between. Therefore, Alternating Renewal Model react very sensitively on the threshold of the minimum dry spell length. In reality, the time period after which two rain events can be assumed as independent may vary with the weather situation (Acreman, 1990), wich is not considered in the model. The empirical distributions of wet and dry spell length in observed time series are very vulnerable to measurement errors. A slight shift in the zero-rainfall baseline for example, can strongly affect the derived model parameters.

One reason why Alternating Renewal Models are not used as frequently as Point Process Models is that they are restricted to the simulation of point rainfall. An extension of the model concept to spatial-temporal simulation is difficult. One among the few existing approaches of spatial-temporal modeling is the "string of beads" model by Pegram and Clothier (2001). Here, the alternating renewal process simulates the sequence of two dimensional radar fields. The radar pictures are classified into three categories: dry, scattered rain and rain according to the fraction of wet pixels in the image. The external structure is the sequence of "rain", "scattered rain" or "dry" radar blocks. The Alternating Renewal Process merely refers to the temporal sequence. The spatial structure within the rain fields is simulated separately as the internal structure of the uninterrupted "rain" and "scattered rain" events.

Haberlandt et al. (2008) simulates several spatial correlated time series by permutation of the order of wet ad dry spells in an optimization scheme. The objectives of the optimization are the joint rainfall probability between the stations, the Pearson correlation of the intensity and the expectation of the rainfall amount conditioned on whether it rains on surrounding stations or not. The optimization is performed sequentially, every new station is optimized in relation to the stations that already have been optimized before.

5.6. Scaling Models

As it was pointed out in Section 4.1 and in Section 4.2, the statistical characteristics of precipitation (e. g. mean, variance or autocorrelation) depend highly on the spatial and tem-

poral aggregation that is analyzed. One of the main challenges for Markov-Chain as well as Point-Process Models is to reproduce these characteristics equally well on different spatial or temporal scales (see e. g. Rodriguez-Iturbe et al. (1986), Foufoula-Georgiou and Guttorp (1986), Katz and Parlange (1995)). This fact let to the idea to create models directly on the spatial and/or temporal scaling behavior of rainfall. The assumption is that scaling is one of the main features of precipitation and that an adequate representation of the scaling behavior in a precipitation model also leads to an adequate representation of the main physical processes.

Scaling models follow the basic idea of "fractal scaling" which was extensively researched by Mandelbrot (for more details see Mandelbrot, 1983). Fractal scaling objects or quantities are "self-similar". It means that they exhibit the same characteristics on any regarded scale. (In most publications the word "fractal" is omitted and the therm "scaling" already stands for fractal scaling. This section will adapt to this conduct.) Mathematically, scaling can be infinite. Regarding real world objects and variables, scaling is always limited to a certain range of scales with a lower and upper scale limit Δ_{min} and Δ_{max}.

Concerning precipitation, scaling is assumed in the distribution of precipitation sums or intensities on different spacial or temporal aggregations. For the intensity this can be expressed as:

$$I(\lambda\Delta) \overset{\mathrm{d}}{=} \lambda^h I(\Delta) \tag{5.16}$$

where	$I(\Delta)$	rainfall intensity averaged over scale Δ
	$I(\lambda\Delta)$	rainfall intensity averaged over scale $\lambda\Delta$
	λ	factor of scale change
	h	scaling exponent
	$\overset{\mathrm{d}}{=}$	equality in distribution

The factor λ takes the value of any integer or the reciprocal of any integer as long as the two dimensions Δ and $\lambda\Delta$ do not exceed the limit scales Δ_{min} and Δ_{max}. Δ can be a temporal or spacial scale. In the latter case the intensity becomes the average areal precipitation over a certain time interval scaling with the area.

If precipitation is scaling, it can only be described by self-similar distribution functions that fulfill Eq. (5.16). Since precipitation measurements generally exhibit positively skewed distributions, possible choices are for example the Pareto Distribution (Lovejoy and Mandelbrot, 1985) or Log-Normal Distribution (Burlando and Rosso, 1996).

Lovejoy and Mandelbrot (1985) present several empirical arguments to support that rainfall fields exhibit fractal scaling (over a certain range in space and time): The energy fluxes and velocities in a fully turbulent flow field are described by the Navier-Stokes equation which is a fractal scaling function. Since the atmosphere is a turbulent flow field, it is fractal scaling and rainfall should be scaling too since it is mainly governed by the atmospheric flow. Another indication is that the boundaries of cloud fields derived from satellite images are fractal scaling (Lovejoy and Mandelbrot, 1985).

If the distribution of the rainfall intensity is scaling, then the statistical moments of the distribution are scaling too. There is a distinction between simple and multi fractal scaling. In a simple scaling system the statistical moments of order q scale all according to the scaling exponent h of the underlying distribution. In the calculation of the q^{th}-moment every value of I is multiplied q-times with itself. Changing the scale from Δ to $\lambda\Delta$ the factor of scale change λ is raised q-times by the scaling exponent h. This leads to the following equation:

$$E\left(K_q\left(\lambda\Delta\right)\right) = \lambda^{hq} \cdot E\left(K_q\left(\Delta\right)\right) \tag{5.17}$$

In multi scaling systems the scaling exponents of moments $q > 1$ are lower:

$$E\left(K_q\left(\lambda\Delta\right)\right) = \lambda^{\phi(q)h} \cdot E\left(K_q\left(\Delta\right)\right) \tag{5.18}$$

where K_q statistical moment of order q on scale Δ
h scaling exponent of the underlying distribution
$\phi(q)$ non-linear, concave function

According to Harris et al. (1997) "multifractality is associated with the presence of spatial correlation of order higher than two". This means that not all peaks show the same spatial correlation, but "more intense peaks cluster more" (Harris et al., 1997). Therefore, multifractality has severe consequences in hydrology: it means that one extreme event can attract the next, which increases flood risks. Gupta and Waymire (1990) find multi scaling in the statistical moments of river discharges and rain radar images.

If the scaling assumption is fulfilled, the statistical properties of all scales between Δ_{min} and Δ_{max} are linked by the scaling exponent h. The relation can be exploited for statistical modeling. Koutsoyiannis and Foufoula-Georgiou (1993) for example use the scaling properties of duration and intensity of rain storms in observed rainfall time series. A stochastic model is setup for the instantaneous rainfall intensity given the storm duration. In their definition a storm is a rainfall event that includes all periods of zero intensity shorter than seven hours. The model can be used to disaggregate rainfall data from point measurements if the total storm duration in known. It can be easily extended to a full rainfall model for point measurements by combining it with a distribution for the storm duration and with a stochastic model of the storm arrival rate (e.g. as a Poisson process).

Burlando and Rosso (1996) use the concept of scaling to derive depth-duration-frequency (short: DDF) curves, which describe the dependence between rainfall duration and rainfall depth for different exceedance frequencies (e. g. once a year, once in ten years). If the analyzed rainfall records exhibits scaling, the exceedance frequencies can be estimated for one duration and then be transfered to all the other duration by the scaling relation.

By a similar approach Menabde et al. (1999) showed that under the assumption of simple scaling the rainfall depth to any extreme event is simply the product of a power-law function of duration and a scale invariant function of exceedance frequency. This relation holds true for any distribution function that can be normalized by its parameters in the form $F\left(I\right) = F^*\left(\frac{I-\mu}{\sigma}\right)$, e.g. the Gumbel distribution. This means that extremes of any duration within the scaling limits Δ_{min} and Δ_{max} can be calculated if the extreme value distribution at one

scale and the scaling exponent are known. Menabde et al. (1999) tested the approach on time series from South Africa and Australia. The observed scaling limits were $\Delta_{min} = 30\,\text{min}$ and $\Delta_{max} = 24\,\text{h}$ which means precipitation extremes on high temporal resolution can be calculated from daily data with generally much better availability.

The Sum of Fractal Pulses Model (Lovejoy and Mandelbrot, 1985) is a deviation of the Point Process Model discussed in Section 5.4. The pulses correspond to what is called rain cells in the Point Process Model. In the Fractal Pulses Model the intensity of a pulse is fully dependent on its duration and calculated by the scaling relation. In one dimension this model can be used to simulate a time series of point rainfall at one measurement location. For spatial modeling the pulses are seen as two dimensional discs with a certain range. The intensity is modeled as function of the disc size. For spatial heterogeneity the discs can be modeled as ellipses. Although the model is strikingly versatile in producing different spatial patterns, it cannot account for all features in real world rainfall fields like fronts and bands (Lovejoy and Mandelbrot, 1985). More details on the mathematic properties of the Sum of Fractal Pulses Model can be found in Mandelbrot (1995).

Multi fractal scaling can ge generated by a Multiplicative Cascade Model (Gupta and Waymire, 1990). This means that the modeled quantity X - this can be the the total precipitation volume - is entered into the system on the upper limit scale Δ_{max} and then consecutively passed down to finer scales. In each step one coarser parent cell is linked to a number of b subcells and the amount of X received from the parent cell is divided among these subcells. This procedure is repeated until the lower limiting scale Δ_{min} is reached. The function that controls which fraction of X is distributed to which subcell is called the cascade generator w. The number b of subcells at each scale jump is called the branch number.

In contrast to the additive Point Process Model the cascade is a multiplicative process. To calculate the rainfall depth of any cell i on the finest scale, one has to track the path from Δ_{max} down to Δ_{min} and at each scale jump multiply X_{sum} with the cascade generator for the respective subcell. Two separated cells are correlated to each other as long as they share one of the coarser cells on the higher resolution (see for example Marsan et al., 1996)). In this way, correlations of higher order are introduced between the values on the finest scale. Higher order correlation means that the value at a time step t cannot be explained sufficiently only by the preceding time step $t - 1$.

The performance of a Mutliplicative Cascade Model depends strongly on the distribution function of the cascade generator w (Molnar and Burlando, 2005). For some distributions of w an analytical formulation can be given for the moment term $\phi(q)$ in the multiple scaling equation Eq. (5.18) (Gupta and Waymire, 1993), e. g. for the frequently used log-normal distributed cascade generator (Molnar and Burlando, 2005). This relation can be exploited in the parameter fit of w, as for example in Harris et al. (1997).

Multiplicative Cascade Models are probably the most frequently used among scaling models. They are especially popular for spatial or temporal downscaling purposes if the rainfall sum on the coarsest scale is known. Lovejoy and Schertzer (1986) propose a Cascade Model for spatial rainfall fields. The rainfall sum is introduced on the coarsest resolution in a rectangular grid and then cascaded down to a sub-grid of n times m new rectangles (in

North-South and East-West direction) in each step. Choosing $n \neq m$ one can account for anisotropy. Intermittency is considered by allowing the cascade generator w to become zero with a certain probability P_0.

Marsan et al. (1996) extend this approach to spatial-temporal modeling. The temporal scale is just added as a new dimension. To account for differences in the spatial-temporal scaling properties, the branch number b in space and time is different, e. g. $b = 3$ in space and $b = 2$ in time. One consequence of this approach is that spatially extended structures are also temporally long lasting (Marsan et al., 1996). This corresponds well to the observations of Austin and Houze (1972) discussed in Section 4.1.

Onof et al. (2005) developed a cascade model for temporal disaggregation of 1 h-values into 5 min minute values as a possible extension to the RAINSIM-model (Cowpertwait, 2006, see Section 5.4). The model can be calibrated without any 5 min-data just by extrapolating the scaling behavior of 1 h to 3 h data, which is important since RAINSIM is applied to ungauged locations.

Scaling models are versatile and flexible, e. g. in their ability to model higher order autocorrelations, which are difficult to represent in Point Process or Markov Chain Models. Nevertheless, Scaling Models are hardly used in operational rainfall generation systems. One reason may be that the estimation of the scaling exponent h (or $\phi(q)\, h$ in multifractal scaling) is very sensitive to bias for several reasons.

Scaling models are based on statistical moments of different orders of magnitudes. In an observed time series of given length the sampling error in the estimation depends on the aggregation. To adjust a multiplicative cascade generator w based on seven scale levels, data from ten minute scale to $10 \cdot 2^7 = 1280$ min scale (almost daily scale) is needed. A one year long record consists of more than 50000 ten minute values but only of about 400 values on the 1280 min scale. Therefore, the calculated statistical moments on the finest scale are much less subject of sampling errors than in the coarsest scales. At a certain aggregation level Δ_{max} the sampling errors become predominant due to the limited number of observations and the scaling "breaks". A correct estimation of the scaling range Δ_{min} to Δ_{max} is essential for an unbiased estimation of the scaling exponent. Hence, the scaling exponent h is severely affected by the systematic sampling error asymmetry (Harris et al., 1997).

The sampling error regarding different statistical moments q show a similar systematic. The higher the moment, the more weight is given to the most extreme values in the time series. If extreme events are randomly "missing" due to the limited number of records in the time series, the effect on the moment increases non-linearly with the moment order q (Harris et al., 1997). As a result the scaling in higher order moment q of a limited sample from a perfectly scaling distribution will flatten out and become increasingly linear instead of power law shaped. In very low moment orders $q < 1$ on the other hand, measurement errors gain more and more weight because they especially affect very low values (high relative error). Therefore, scaling is not only limited to the aggregation between Δ_{min} and Δ_{max}, but there is also range in the moment order between q_{min} and q_{max} for which the scaling behavior is observable (Harris et al., 1997). q_{min} and q_{max} depend on the length of the time series.

The revealed systematic in sampling errors lead to the question if scaling behavior can be detected at all by the usually applied methods as spectrum analysis (Marsan et al., 1996) or box counting (Mandelbrot, 1983) in the presence of sampling errors. In a theoretical experiment Harris et al. (1997) additively mix data from processes with different scaling exponents. The *additive* (instead of multiplicative) mixing should destroy the scaling all together. Instead, a scaling relationship is detected that is closer to the more intermittent of the two data sets – which is a problem regarding precipitation fields. Passing cold fronts for example exhibit higher intermittency in the postfrontal than in the prefrontal rainfall. Therefore, the parameter estimation has a dilemma. On the one hand, the scaling estimation should be restricted to physically homogeneous precipitation processes, on the other hand the time series should be as long as possible to avoid sampling errors, especially in the calculation of higher moments and in the higher aggregated data.

Ferraris et al. (2002) have doubts if observed fractal scaling is a proof for a multiplicative process in precipitation formation and higher order dependencies in rainfall field. They build random fields from from Fourier transformed linearly correlated data that show similar scaling behavior as rain radar fields. This means that the scaling can be mimicked by non-linear filtered random fields that do not exhibit higher order correlation structures.

5.6.1. Autoregressive Models

Autoregressive models assume a memory between the elements in a rainfall time series of discrete time steps. In this respect, they are related to Markov-Chain models (Section 5.3). The memory, however, is described differently. It is not expressed as the transition probability between defined discrete states (e. g. "rain" / "no rain") but as a fraction of information in each new value that is herited from the last n preceding values. The number n of preceding values that have to be considered is called the order of the autoregressive model. In short term, it is called an AR(n) model.

In a temporal AR(1) model the autocorrelation r_l declines exponentially with the lag l between two considered time steps, since it is always the same part of information that is passed. A deviation of the observed autocorrelation from exponential decline indicates that a higher order autoregressive model is required. The parameters for the set-up of higher order models cannot be accessed directly from an observed time series because the autocorrelations parameters of different lags l all influence the time series at the same time. They have to be derived from the observed autocorrelations by a linear equation system.

AR(1) models can be extended to spatial-temporal modeling if the autocorrelation parameter is seen as a $n \times n$ matrix that describes how the values at each station depend on the previous values of all n stations. A second matrix is needed for the description of the cross correlations at simultaneous time steps.

Bárdossy and Plate (1992) present a spatial-temporal model for daily precipitation sums based on an autoregressive model. The daily rainfall sums are described by a power transform of the truncated normal distribution whereas the truncation is based on the rainfall probability. The daily rainfall sums are assumed to follow a spatial-temporal AR(1) model

in the normal transformed space. The interdependencies between all modeled locations are considered in a crosscorrelation matrix and a lag-one crosscorrelation matrix, linking the values at the different locations to the subsequent time step. The rainfall probability and spatial-temporal correlations are estimated separately for 11 different CP groups to account for persistence and long term memory in the time series. The model is able to represent the statistical moments, the autocorrelations and the distribution of wet and dry spell lengths (Bárdossy and Plate, 1992). The spatial representation of intermittency is checked by calculating the statistical moments conditioned on whether the precipitation sums at surrounding stations exceed a certain threshold or not. Bárdossy and Plate (1992) also describe how the model can be transformed to previously ungauged locations.

Koutsoyiannis et al. (2003) proposes a model for spatially correlated disaggregation e. g. of several spatially correlated daily precipitation series to hourly time series. The model needs at least one hourly time series as a starting point, which can for example be generated by a purely temporal (and though simpler) disaggregation model. The spatial-temporal correlation between the hourly time series are described by a non-Gaussian transform of an AR(1) model. The correlations on the hourly scale are estimated by a matrix multiplication that is coupling the correlation between the existing hourly time series to the daily time series at the same location and the correlation between the daily time series of the different locations. The model has the shortcoming that it can produce negative values that have to be set to zero which disturbs the aggregated statistics (the moments at the daily scales).

6. Generation of Point Rainfall

6.1. The Synthetic Rainfall Time Series Generator NiedSim

The conceptual rainfall generation models presented in Chapter 5 rebuild the statistical characteristics of precipitation only to a limited extend. Markov Chain Models work very well on daily scale, but the application on time series of shorter aggregation is difficult due to the persistence and intermittency of precipitation. A regionalization to ungauged areas is hardly possible (Section 5.3). Point Process Models have problems in simulating the statistical characteristic on all temporal aggregations equally well. A correct assessment of the scaling is very challenging and requires a high number of model parameters, as well as burdensome parameter fitting routines (Section 5.4). Scaling models, have a high risk of bias if the amount of available data for the model fit is limited (Section 5.6).

Data driven time series generators try to avoid these problems by making as few conceptual specifications as possible. Instead, precipitation is merely described by its statistical appearance.

Since the year 2000 the data driven model NiedSim (Bárdossy, 1998) is operational. It is used by the authorities of several federal states in South Germany for the generation of rainfall time series in hourly or five minute temporal resolution. The only conceptual specification in NiedSim concerns the theoretical distribution of precipitation values. An initial time series is set up that has the right intensity distribution but differs from real precipitation in statistics related to the temporal sequence of values. In a second step, the temporal sequence is optimized by subsequently swapping arbitrarily chosen pairs of values (Fig. 6.1). The performed swaps are evaluated by an objective function O that measures the similarity between the simulated time series and statistical target values derived from real world observations, for example the autocorrelation on different lags, the extreme value behavior or the daily rainfall probability. The optimization stops when the different between the simulated statistics and the target values is minimized.

This generation principle is very data demanding because the statistical target values have to be derived from data of observed measurement station. On the other hand, the method takes maximum benefit from the high density of the German rain gauge network. In principle, the detailedness of the generated precipitation time series is only limited by the accuracy of the applied theoretical distribution and the number of statistical target values that are considered in the objective function O.

As another strong point NiedSim is well suited for the generation at ungauged locations. For a regionalization it is sufficient to transfer the statistical target values from the observed rain

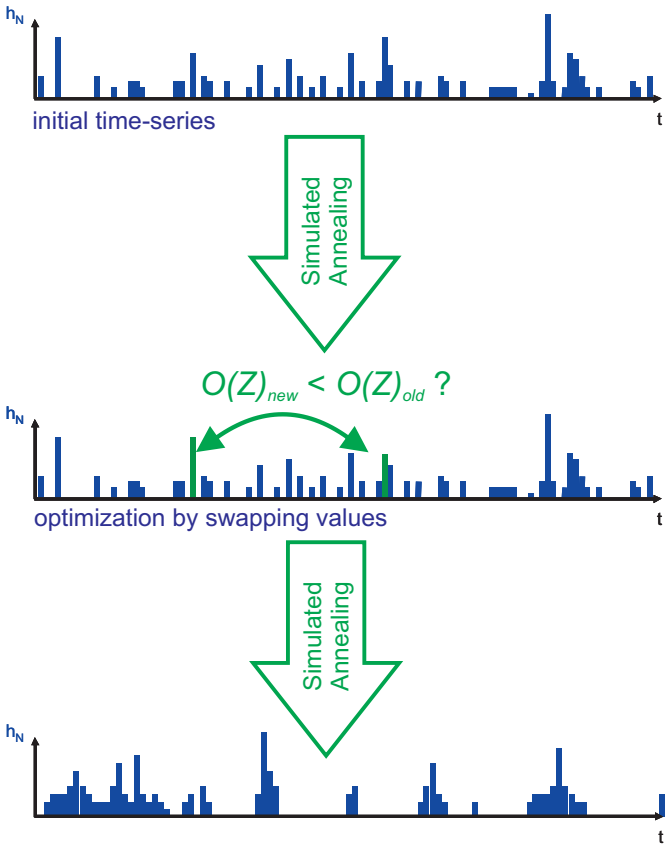


Figure 6.1.: Principle of the permutation scheme in NiedSim

gauges by geostatistical methods. The same is true for downscaling. The generation can be easily conditioned on higher aggregated values, e. g. monthly sums by implementing them as target values in the objective function.

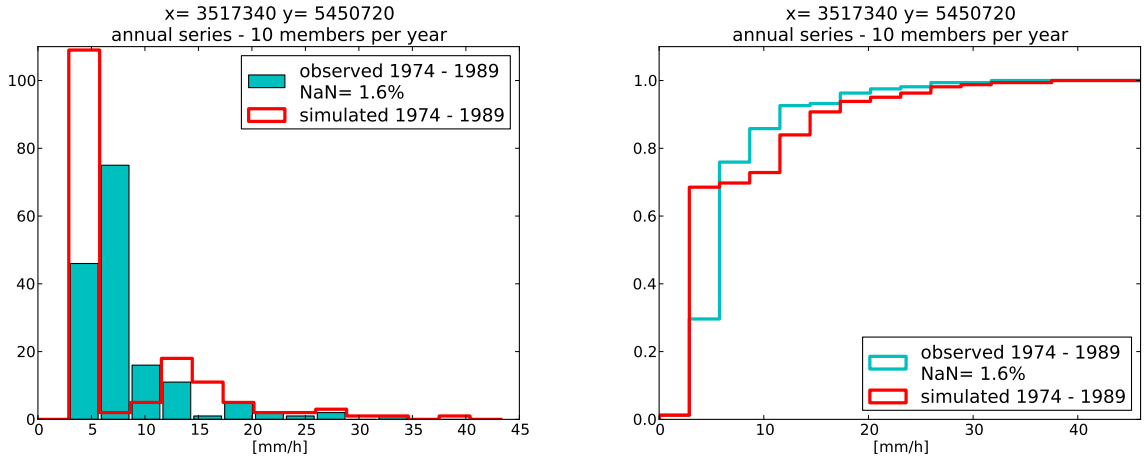
The temporal resolution of NiedSim is one hour. In higher temporal resolution, the number of possible combination of the values from one year soon becomes unfeasible in the optimization. For a further refinement to 5 min values, NiedSim is combined with a disaggregation scheme in which the hourly values are distributed among the twelve five minute intervals of the hour. The disaggregation is again performed by an optimization that exchanges small increments of precipitation between the five minute intervals.

The generation scheme of NiedSim is the basis for this work. The experiences in operation has shown the need for several adaptations.

6.1.1. Necessary Adaptations to NiedSim

1. **Mixing of different distribution functions**
   Hourly precipitation values in Baden-Württemberg exhibit very skewed distributions. Most hourly precipitation values are smaller than 1 mm while on the other hand ex-

tremes can go up to 50 mm. The fit of a distribution function that represents the low values and the extremes equally well is difficult. In Niedsim, three different distribution functions are used. The maximum is drawn from a Gumbel distribution the three next highest values of the year from an Exponential distribution. Both distributions are fitted to the theoretical one yearly and one hundred yearly extreme values according to the official extreme value statistics KOSTRA of the German Weather Service DWD (Bartels et al., 2005). The remaining values are drawn from a Gamma distribution whose parameters are fit to observed precipitation time series of 292 rain gauges (Section 3.3) and then interpolated on the target location by Kriging.



(a) empirical absolute frequencies

(b) empirical cumulative distribution

Figure 6.2.: Comparison of the observed and NiedSim-simulated empirical distribution of the ten highest 1 h precipitation values per year at the rain gauge in Neckarsulm

Since the distributions are derived from different data sets, they do not correspond very well at every location. A generation for the rain gauge in Neckarsulm is presented as an example (Fig. 6.2). The histogram of the simulated time series (red line in Fig. 6.2a) shows a remarkable gap in the frequency of values between six and twelve millimeters per hour that cannot be observed in the measurements. This leads to a particularly flat part in the cumulative distribution between six and twelve millimeters of hourly rainfall (Fig. 6.2b). The bimodal distribution is not realistic but a statistical artifact of the mixing. It could be avoided if the three distributions functions were replaced by one single distribution. However, applying only one distribution would reduce the variability. Either the highest values would be underestimated or the expectation of the mean 1 h precipitation amount would be overestimated.

2. **Variability between wet and dry periods**
   In Niedsim every month of the year is generated with the same distribution parameters. Whether the month is wet or dry, the distribution of 1 h precipitation values with non-zero precipitation $F(H \mid H > 0\,\text{mm})$ is constant. In real world observations however, the difference between wet and dry periods can have a direct effect on the distribution of the 1 h precipitation values. Besides, the distribution of hourly rainfall

sums depends on the season. Such dependencies of the distribution parameters are not yet considered in the operational version of NiedSim.

3. **Uncertainties in the regionalization of the simulation parameters**
   Most parameters of NiedSim are derived for observed rain gauges and regionalized by External Drift Kriging, (Section 5.2) with the square root of the elevation used as drift variable. The fact that parameter interdependencies are not considered can lead to inconsistencies. For example, if there is a hidden, non-linear interdependence between the shape and the scale parameter of the Gamma distribution, it will be altered by the interpolation and the distribution at the target location will be biased.

To overcome these short comings, the generation scheme will be modified in this work. The generation of the initial time series will be restricted to only one distribution function, namely the Weibull distribution. To enhance the variability, the distribution parameters will be conditioned on the monthly precipitation sum. In this way, wet months will exhibit a different distribution than dry months. At the same time, the interdependencies of the distribution parameters are taken into account. The interdependencies are described by a Gaussian copula. To account for annual cycles, the description of the interdependencies is done for each season separately.

6.2. Statistical Parameters of Hourly Precipitation Time Series

The two parametric Weibull distribution can be fit by the first two statistical moments, average and standard deviation (method of moments). Assuming that the 1 h precipitation values follow the Weibull distribution, it can be defined by the following statistics:

- rainfall probability, that controls the fraction of wet time steps

$$P_{1\mathrm{h}} = 1 - F\left(H = 0\,\mathrm{mm}\right) \tag{6.1}$$

  where H is the 1 h rainfall sum

- the expected average precipitation of wet hours $E\left(H_{wet}\right)$

$$E\left(H_{wet}\right) = E\left(H \mid H > 0\,\mathrm{mm}\right) \tag{6.2}$$

- the expectation of the standard deviation

$$D\left(H_{wet}\right) = \left[E\left(H_{wet} - E\left(H_{wet}\right)\right)^2\right]^{\frac{1}{2}} \tag{6.3}$$

  of the rainfall amounts of wet hours

The rainfall probability $P_{1\mathrm{h}}$, as well as the expectations $E\left(H_{wet}\right)$ and $D\left(H_{wet}\right)$ are theoretical values that are strictly speaking only observable in a sample of infinite size. With data from real world measurements, where the number of observations is limited, one can only give estimates of these theoretical quantities.

In the following the variables $X, Y, \ldots$ ($U, V, \ldots$ for uniform distributions) in capital letters are dedicated to random variables in the theoretical world of infinite samples. The use of lower case letter $x, y \ldots$, $(u, v, \ldots)$ indicates observed real world data sets of limited length n.

The values of $P_{1\mathrm{h}}$, $E\left(H_{wet}\right)$ and $D\left(H_{wet}\right)$ can be approximated by the rainfall frequency $p_{1\mathrm{h}}$, as well as the average $\bar{H}_{wet}$ and the standard deviation s_{wet} of wet hour precipitation amounts calculated from an observed precipitation time series. For the estimation of the statistical parameters of one month they are calculated in the following manner:

$$p_{1\mathrm{h}} = \frac{1}{n_{mon}} \sum_{i=1}^{n_{mon}} 1_{\{H(i)>0\,\mathrm{mm}\}} \tag{6.4}$$

$$= \frac{n_{wet}}{n_{mon}} \tag{6.5}$$

$$\bar{H}_{wet} = \frac{1}{n_{wet}} \sum_{i=1}^{n_{mon}} H\left(i\right) \tag{6.6}$$

$$s_{wet} = \frac{1}{n_{wet}-1} \sum_{i=1}^{n_{mon}} 1_{\{H(i)>0\}} \left(H\left(i\right) - \bar{H}_{wet}\right)^2 \tag{6.7}$$

$$\tag{6.8}$$

where n_{mon} number of hours in current month

$1_{\{H(i)>0\,\mathrm{mm}\}}$ $\begin{cases} 1 & \text{if} \quad H\left(i\right) > 0 \\ 0 & \text{if} \quad H\left(i\right) = 0 \end{cases}$

n_{wet} number of wet hours in current month

The observed values of $p_{1\mathrm{h}}$, $\bar{H}_{wet}$ and s_{wet} vary from month to month. Even if the underlying theoretical expectations were constant, there would be spread due to sampling effects. In addition, there is variability resulting from the different rainfall characteristic of each month and due to seasonal effects. The resulting variability can be described by theoretical distribution functions $F\left(X\right)$.

6.3. Methods for the Statistical Description of Distributions

For a quantity X defined on the domain Ω_X the distribution function $F(X)$ assigns the non-exceedance probability $P\left(X \leq X_0\right)$ to each value X_0 within the domain. It can be seen as the limit of the observed non-exceedance frequency $F\left(x\right)$ if it was possible to draw a sample of infinite size. Ω_X is also called the "event space" of X.

$$F(X_0) = P\left(X \leq X_0\right), X \in \Omega_X \tag{6.9}$$

$$= \lim_{n\to\infty} F\left(x_0\right) \tag{6.10}$$

There are different families of theoretical distribution functions that can be fit to an observed distribution x. The choice of the distribution depends on the domain Ω_X and on the shape of the observed distribution. For some configurations of x the choice of the distribution function can be justified by probability theory.

6.3.1. Weibull Distribution

The Weibull Distribution was first described in the context of life spans in material control. The probability density function (PDF) is given by:

$$f(X) = \frac{\kappa}{\lambda}\left(\frac{X}{\lambda}\right)^{\kappa-1} e^{-(X/\lambda)^{\kappa}} \tag{6.11}$$

Its cumulative distribution function (CDF) is given by:

$$F(X) = 1 - e^{-(X/\lambda)^{\kappa}} \tag{6.12}$$

The support of the Weibull Distributions are the positive rational numbers $X \in [0, \infty]$.

The parameter κ defines the shape of the distribution, λ defines the scale. For $\kappa = 1$ the Weibull Distribution is equivalent to the Exponential Distribution. For $\kappa \leq 1$ the density is highest at $X = 0$, for $\kappa > 1$ the PDF shows a maximum at $X > 0$.

The first two moments of the Weibull Distribution are given by the following equations:

$$E(X) = \lambda\Gamma\left(1 + \frac{1}{\kappa}\right) \tag{6.13}$$

$$D^2(X) = \frac{1}{\lambda^2}\left[\Gamma\left(1 + \frac{2}{\kappa}\right) - \Gamma^2\left(1 + \frac{1}{\kappa}\right)\right] \tag{6.14}$$

where the Γ-function is:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \tag{6.15}$$

The Weibull Distribution can be linearized by a double logarithmic transformation:

$$\ln(-\ln(1 - F(x))) = -\kappa \ln\lambda + \kappa \ln X \tag{6.16}$$

$$W = a^* + b^* V \tag{6.17}$$

This relation is used to check if the data of the precipitation stations follow a Weibull Distribution or not. If the observed data pairs $(x|F_n(x))$ from an observed precipitation time series spread symmetrically around a straight line in the v, w-plot where

$$v = \ln x \tag{6.18}$$

$$w = \ln(-\ln(1 - F_n(x))) \tag{6.19}$$

the assumption of Weibull distributed data is valid.
$F_n(x)$ is the observed non-exceedance frequency that is calculated by the ranks $r(x) = [1, 2, \ldots n]$ of the observed data:

$$F_n(x) = \frac{r(x)}{n} \tag{6.20}$$

To account for sample affects in the transformed plot, $F_n(x)$ will not be calculated according to Eq. (6.20) if the data volume n is lower than 50 but with the following formula:

$$F_n(x) = \frac{r - 0.3}{n + 0.4} \text{ if } n < 50 \tag{6.21}$$

where r rank of x according to Eq. (6.46)
n number of observations

6.3.2. Log-Normal Distribution

A quantity X exhibits a Log-Normal Distribution if the transformed quantity $V = \ln X$ is normally distributed.

The central limit theorem states that any mean or sum of a large number of identically distributed random variables approximately follow a Normal Distribution (see for example Hartung, 2009). In the logarithmic transformed space any product is transformed into a sum. The sum of logarithms is equal to the logarithm of the product. Applying the central limit theorem, it follows that any product of a large number of identically distributed random variables is approximately Log-Normal distributed.

On that account the Log-Normal Distribution is a reasonable choice for random variables that arise from multiplicative processes. According to the theory of multifractal scaling (see Section 5.6), this is an argument that the Log-Normal distribution could be an appropriate choice for precipitation related statistics.

The Log-Normal Distribution is characterized by the following equations:

$$f(X) = \frac{1}{X\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln X - \mu)^2}{2\sigma^2}} \tag{6.22}$$

$$F(X) = \int f(X)\, dX \tag{6.23}$$

The parameters μ and σ can be interpreted as the average and standard deviation in the logarithmic transformed space. The integral $F(X)$ can only be estimated numerically.

The fit of the Log-Normal Distribution can be checked graphically in the linearized space of v, w:

$$v = \ln(x) \tag{6.24}$$

$$w = \phi^{-1}(F_n(x)) \cdot \sigma + \mu \tag{6.25}$$

$$\tag{6.26}$$

where ϕ^{-1} inverse of the Standard Normal distribution
μ, σ parameters of the fitted Log-Normal distribution

6.3.3. Pareto Distribution

The highest extremes in hydrological variables (e. g. precipitation amounts or discharges) are responsible for the most dangerous situations. Therefore, it is important that the applied theoretical distribution exhibits a sufficiently high skewness to avoid potentially dangerous underestimations of the extremes. Distributions that attribute a high probability to extreme values are often called "thick tailed".

An especially "thick tailed" theoretical distribution function is the Pareto Distribution. Koutsoyiannis (2005) argues that it was the principal distribution to describe extreme precipitation frequencies. Like the Log-Normal Distribution it is self-similar and in accordance with the theory of fractal scaling.

The PDF and the CDF are given by:

$$f(X) = \frac{\alpha X_m^\alpha}{X^{\alpha+1}}\ ,\ X \geq X_m \tag{6.27}$$

$$F(X) = 1 - \left(\frac{X_m}{X}\right)^\alpha\ ,\ X \geq X_m \tag{6.28}$$

The support of the Pareto distribution is $X \in [X_m; \infty]$. X_m is the basic scale of the distribution α is the shape parameter. All probabilities are expressed in relation to X_m. This is a problem, when X can become zero. For $X_m = 0$ the distribution is not defined. In that case it is necessary to shift the distribution so that the support starts at zero. The resulting distribution is called the Lomax Distribution. PDF and CDF are given by:

$$f(X) = \frac{\alpha}{\lambda}\left(1 + \frac{X}{\lambda}\right)^{-(\alpha+1)}\ ,\ X \geq 0 \tag{6.29}$$

Its CDF is given by:

$$F(X) = 1 - \left(1 + \frac{X}{\lambda}\right)^{-\alpha}\ ,\ X \geq 0 \tag{6.30}$$

The equations 6.29 and 6.30 can be derived from the CDF and PDF of the Pareto Distribution by setting $\lambda = X_m$ and substituting X by $X + X_m$.

The Lomax distribution can be linearized if the scale parameter λ is known:

$$\ln(1 - F(X)) = -\ln(\lambda) - \alpha \ln(X + \lambda) \tag{6.31}$$

$$W = a^* + b^* V \tag{6.32}$$

where the transformed observed values are

$$v = \ln(x + \lambda) \tag{6.33}$$

$$w = \ln(1 - F_n(x)) \tag{6.34}$$

$F_n(x)$ is calculated by Eq. (6.20).

6.3.4. Beta Distribution

The Beta Distribution is given by the following equations for the PDF and CDF:

$$f\left(X,\alpha,\beta\right) = \frac{X^{\alpha-1}\left(1-X\right)^{\beta-1}}{B\left(\alpha,\beta\right)}, X \in (0,1) \tag{6.35}$$

$$F\left(X,\alpha,\beta\right) = \frac{B_X\left(\alpha,\beta\right)}{B\left(\alpha,\beta\right)}, \; X \in (0,1) \tag{6.36}$$

where $B\left(\alpha,\beta\right)$ is the Beta Function and $B_X\left(\alpha,\beta\right)$ the incomplete Beta Function:

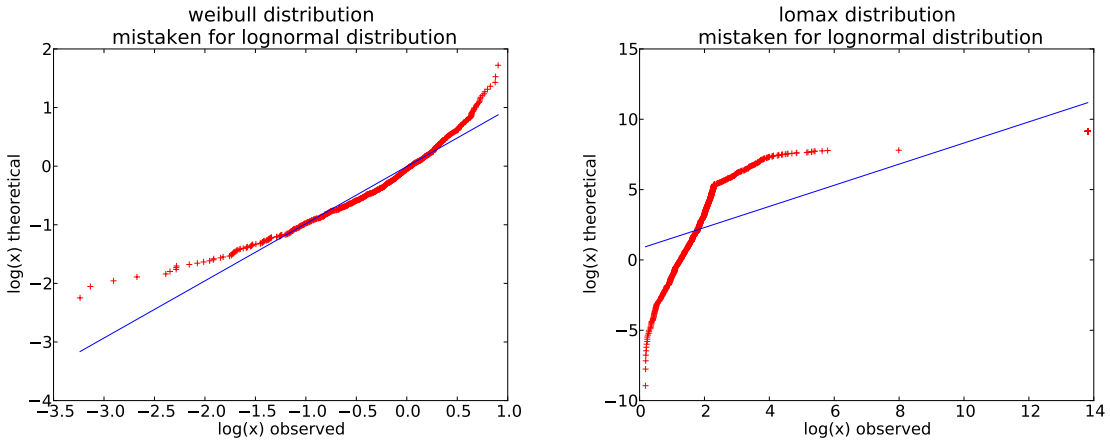
$$B\left(\alpha,\beta\right) = \int_0^1 t^{\alpha-1}\left(1-t\right)^{\beta-1} dt \tag{6.37}$$

$$B_X\left(\alpha,\beta\right) = \int_0^X t^{\alpha-1}\left(1-t\right)^{\beta-1} dt \tag{6.38}$$

Since there are no analytical solutions to the integrals, both equations can only be solved numerically.

The support for the beta distribution is $\Omega = (0,1)$ while the limits of the interval are excluded. For $\alpha = 1$ and $\beta = 1$ the density is constant over Ω leading to a uniform distribution. If $\alpha < 1$ and $\beta < 1$ the PDF is highest for $X \to 0$ and $X \to 1$. If $\alpha > 1$ and $\beta > 1$ the density is unimodal. For $\alpha = \beta$ the distribution is symmetric to $X = 0.5$.

6.3.5. Comparison of Theoretical Distributions



(a) Weibulll Distribution

(b) Lomax Distribution

Figure 6.3.: Effect of fitting a Log-Normal Distribution by Maximum Likelihood Method to a distribution of 1000 random numbers drawn from a different theoretical distribution

The three distributions Weibull Distribution, Log-Normal Distribution and Lomax Distribution differ the most in the probabilities attributed to extreme precipitation. Fig. 6.3 illustrates

the differences. 1000 random values are drawn from a Weibull Distribution (Fig. 6.3a) or a Lomax Distribution (Fig. 6.3b) and are plotted in the linearized Log-Normal Plot. For Log-Normal distributed data, the red crosses of the observation would spread symmetrically around the blue line.

Data from the Weibull distribution, leads to a convexly bending point cloud. The Log-Normal distribution overestimated very low and very high values. Therefore the Log-Normal distribution exhibits higher extreme value probabilities while the Weibull distribution allocates more density in the vicinity of the median. The concave curbing of the Lomax distributed observations indicates the opposite effect (Fig. 6.3b). The Lomax distribution has a higher skewness. In this case, the application of the Log-Normal Distribution leads to an underestimation in the extremes. The values are generally very low, except for the extremes. The density around the median is low compared to the Normal Distribution.

6.3.6. Maximum Likelihood Method for the Fit of Distribution Functions

Theoretical distribution functions exhibit parameter that have to be adjusted to the observed distribution. On possible estimation is the Maximum Likelihood Method (MLM). It consists of finding the configurations of the parameters $\alpha, \beta \ldots$ of the tested distribution functions for which the observed values of x are associated with the highest PDF-values. In other words: One searches the configuration of the distribution function for which the observed distribution x is most likely to occur. (More on MLM in Wilks, 2011)

$$L\left(x|\alpha,\beta\ldots\right) = \prod_{i=1}^{n} f\left(x_i|\alpha,\beta,\ldots\right) \rightarrow max! \tag{6.39}$$

Transforming Eq. (6.39) into logarithmic space does not change the location of the maximum. In logarithmic form, however, the product can be transferred to a sum:

$$\ln L\left(x|\alpha,\beta,\ldots\right) = \sum_{i=1}^{n} \ln f\left(x_i|\alpha,\beta,\ldots\right) \rightarrow max! \tag{6.40}$$

This sum is maximized by numerical optimization.

6.4. Methods for the Description of Multivariate Dependencies

6.4.1. Covariance and Correlation

A standard statistical tool to measure the linear dependence between two observed quantities x and y is the correlation r_{xy}

$$r_{xy} = \frac{s_{xy}}{s_x s_y} \tag{6.41}$$

calculated by covariance between x and y

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) \tag{6.42}$$

and the standard deviations in x and y

$$s_x = \left(\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right)^{\frac{1}{2}} \tag{6.43}$$

where $\bar{x}$ the average of the observations in x
n number of data pairs

The covariance and the correlation can easily be extended to the multivariate case:

$$\mathbf{\Sigma} = [s_{ij}] \text{ where } i, j \in [1, n_{var}] \tag{6.44}$$

$$\mathbf{r} = [r_{ij}] \text{ where } i, j \in [1, n_{var}] \tag{6.45}$$

where n_{var} the number of correlated variables. The bold typing of $\mathbf{r}$ and $\mathbf{\Sigma}$ indicates that these are matrices. The diagonal elements r_{ii} $i \in [1, n_{var}]$ are all equal to one. The diagonal elements Σ_{ii} are equal to the respective variances s_i^2.(for more details, see for example Hartung, 2009).

Despite the wide spread use of multivariate covariance and correlation, it has considerable draw backs. The data pairs that are far away from the point of the average $(\bar{x} \mid \bar{y})$ have a higher weight in the calculation of the covariance. The correlation therefore is especially governed by the highest values which can be a problem in skewed distributions like many precipitation related statistics. Besides, the restriction to linear dependence is somehow contradictory. Any non-linear transformation of x or y changes the value of the correlation r_{xy}. However, if x or y are multiplied by a deterministic non-linear function, why should this affect the "real" dependence between the two quantities? If x and y follow a functional relation $y = f(x)$ the correlation can take any value $-1 \leq r_{xy} \leq 1$ although y is completely determined by x. As a consequence, the correlation r_{xy} is highly sensitive to non-linear measurement errors in the correlated quantities.

6.4.2. Spearman's Rank Correlation

A more robust measure of dependence is the Spearman's rank correlation ρ_{xy}. The rank of an observed distribution is calculated by sorting the data in ascenting order according to their value and then replacing each value by its position number

$$\begin{aligned} x &= [x_1, x_2, x_3 \ldots x_n] \text{ where } x_1 \leq x_2 \leq x_3 \ldots \leq x_n \\ Rank(x) &= [1, 2, 3 \ldots n] \end{aligned} \tag{6.46}$$

ρ_{xy} is then calculated as the correlation from the position numbers $r = Rank(x)$ and $q = Rank(y)$.

$$\rho_{xy} = \frac{s_{rq}}{s_r s_q} \tag{6.47}$$

where the covariance s_{rq} and the standard deviations s_r and s_q are calculated by Eq. (6.42) and Eq. (6.43) just be replacing x and y by their respective ranks r and q.

Spearman's ρ only depends on the configuration of the data and not on the distribution. As long as the order of values is kept unchanged, the rank correlation is constant. It is not affected by any monotonic transformation of the data. For this reason Spearman's ρ is insensitive to any measurement errors, as long as the order of the data is not affected. $|\rho|$ becomes 1 for *any* functional dependence $y = f(x)$ between x and y.

6.4.3. Copulas as a Measure of Dependence

Spearman's ρ cannot measure all aspects of multivariate dependence. Environmental variables may exhibit asymmetry in the correlation structure, so that a subsample of higher-ranked data pairs show a different rank correlation than a subsample of lower ranked pairs. While the rank correlation reduces the whole data range to one specific number and cannot take such effects into account, copulas and copula density are tools to measure the dependence structure over the whole range of the distributions.

Due to its ability to measure complex dependence structures, copulas have been used more and more frequently in precipitation analysis for the last ten years. Theoretical copulas of the Archimedean family for example are applied to relate volume, duration and intensity of rain storm events and to calculate joint probabilities for simultaneous extremes in these three quantities (Zhang and Singh, 2007). It is shown that the estimated probabilities are more realistic than joint probabilities calculated from multivariate correlation (Zhang and Singh, 2007).

AghaKouchak et al. (2010) use a V-transformed copula to describe the dependence between $1\,\mathrm{h}$ precipitation amounts from rain gauge measurements and areal precipitation amounts derived from rain radar. The relationship is used to generate ensembles of rainfall fields that are conditioned on the rain gauge measurements and exhibit the spatial variability and the error structure of the radar fields.

van den Berg et al. (2011) exploit empirical, non parametric copulas for the downscaling of areal $1\,\mathrm{h}$ precipitation sums from a coarse scale of $19.1\,\mathrm{km} \times 19.1\,\mathrm{km}$ to a fine scale of $600\,\mathrm{m} \times 600\,\mathrm{m}$. It is found that the distribution on the finer scale is not constant but depends on the coarse scale rainfall sum. The effect can be taken into account be the copulas.

Salvadori and De Michele (2010) give an overview of how copulas can be applied in the estimation of joint extreme value return periods. As an example they model joint return periods of extreme discharge events at four river gauging stations in the Spey catchment, UK.

A copula represents the cumulative distribution function of n_{dim} joint variables if the variables all follow a uniform distribution in the interval $[0, 1]$

$$C : [0, 1]^{n_{dim}} \rightarrow [0, 1] \tag{6.48}$$

where the hypercube $[0, 1]^{n_{dim}}$ is the support for the copula. The copula describes the joint non-exceedance probability in the n_{dim} variables.

$$C\left(U_1^*, U_2^*, \ldots, U_{n_{dim}}^*\right) = F\left(U_1 \leq U_1^*, U_2 \leq U_2^*, \ldots, U_{n_{dim}} \leq U_{n_{dim}}^*\right) \tag{6.49}$$

The copula becomes zero if one of the n_{dim} variables is zero. It is equal to U if one of the variables has a value of U and the others are one.

$$C\left(U_1, \ldots, U_{i-1}, 0, U_{i+1}, \ldots, U_{n_{dim}}\right) = 0 \tag{6.50}$$

$$C\left(1, \ldots, 1, U_i, 1, \ldots, 1\right) = U_i \tag{6.51}$$

For any interval B of dimension n_{dim} in the hypercube $[0,1]^n_{dim}$ the probability that $U = [U_1, U_2, \ldots, U_n]$ is within B is non-negative.

$$P\left(U \in B\right) = \int_B dC\left(U\right) \geq 0 \tag{6.52}$$

The derivative $c = dC$ is called the copula density. It is a multivariate extension of the probability density $f\left(X\right)$ for joint uniformly distributed variables U.

According to the Sklar's Theorem (see for example Nelsen, 2006) the joined distribution of n_{dim} random variables $X_1, X_2, \ldots X_{n_{dim}}$ can be calculated by a copula if the distribution functions $F_1\left(X_1\right), F_2\left(X_2\right), \ldots F_{n_{dim}}\left(X_{n_{dim}}\right)$ are known:

$$F\left(X_1, X_2, \ldots X_{n_{dim}}\right) = C\left(F_1\left(X_1\right), F_2\left(X_2\right), \ldots F_{n_{dim}}\left(X_{n_{dim}}\right)\right) \tag{6.53}$$

In this way, the analysis of the dependence between the variables can be separated from the analysis of their distributions, which are from now on called the marginal distributions.

6.4.3.1. Empirical Copulas

Since copulas are describing probabilities they can, strictly speaking, only be observed for an infinite number of data tuples. Nevertheless, a copula can be easily approximated by real world data since the joined non-exceedance frequency $F_n\left(x\right)$ of a data set x with n values is a uniform distribution in the interval $[0,1]$ for any possible distribution.

The estimation of the empirical copula density follows a five step proceeding. For the bivariate case of x and y data with n data pairs:

1. Replacing the data by its rank according to Eq. (6.46): $x \rightarrow r\left(x\right)$, $y \rightarrow q\left(x\right)$
2. Transforming the ranks into the uniform distribution of non-exceedance frequencies u and v:
$$u = F_n\left(x\right) = \frac{r\left(x\right) - 0.5}{n} \tag{6.54}$$
(v respectively by transforming y)
3. Mapping the pairs of of u_i and v_i in the unit square $[0,1]^2$.
4. Dividing the square into $n_{bins} \times n_{bins}$ sub-squares. By counting the number of pairs $(u_i \mid v_i)$ in each square and by dividing the number by the size of the sub-square area, the empirical copula density is calculated.
$$c_{ij} = n\left\{u \in \delta_i; v \in \delta_j\right\} \cdot n^2_{bins} \tag{6.55}$$

where	c_{ij}	empirical average copula density in cell ij
	i, j	cell number in u and v-direction, $i, j \in [1, n_{bins}]$
	$\delta_i = [\frac{i-1}{n}, \frac{i}{n}]$	borders of the i^{th} sub-square in u-direction
	$\delta_j = [\frac{j-1}{n}, \frac{j}{n}]$	borders of the j^{th} sub-square in v-direction
	$1/n_{bins}^2$	area of one sub-square

5. Finally, the copula can be calculated by summation over the sub-squares in u and v direction:

$$C(u,v) = \sum_{i=1}^{n_{bins}} \sum_{j=1}^{n_{bins}} 1_{\{i/n_{bins} \leq u\}} 1_{\{j/n_{bins} \leq v\}} c_{ij} \tag{6.56}$$

where $1_{\{i/n_{bins} \leq u\}} \begin{cases} 1 & \text{if} \quad i/n_{bins} \leq u \\ 0 & \text{if} \quad i/n_{bins} > u \end{cases}$

6.4.3.2. Theoretical Copulas

In probability calculations observed non-exceedance frequencies $F(x)$ of finite samples are used to fit theoretical distribution functions $F(X)$. It is assumed that $F(X)$ is the limiting case for the observed distribution $F(x)$ if the sample size could be extended to infinity:

$$F(X) = 1 \lim_{n \to \infty} F(x) \tag{6.57}$$

Assuming a certain shape of the theoretical distribution, e. g. Exponential, Normal or Gamma distribution, it can be used to estimate non-exceedance probabilities for values that have not yet been observed in the fitting sample or that even exceed the range of the observed distribution.

Applying the same concept, a theoretical copula is fit to a given data set. Instead of the observation x, n_{dim} dimensional tuples of uniformly transformed observations u are used.

$$C(U_1, U_2, \ldots, U_{n_{dims}}) = \lim_{n \to \infty} C(u_1, u_2, \ldots, u_{n_{dims}}) \tag{6.58}$$

There are different choices of theoretical copulas that can be fit to the data. If the marginal distributions are known, it is common to perform a maximum likelihood fit based on the copula densities (Schoelzel and Friederichs, 2008). For some copula families predefined maximum likelihood estimators exist.

One of the main difference between the copula families is the tail dependence structure (AghaKouchak et al., 2010). Tail dependence means that the highest or lowest extreme values show a different dependence structure than the other, non-extreme values. In this work the Gaussian copula is exploited. The Gaussian copula assumes absence of tail dependence. The dependence structure is constant over the whole range of all uniformly transformed random variables $U_1, U_2, \ldots U_{n_{dim}}$.

6.4.3.3. The Gaussian Copula

The Gaussian copula can be described by the following equation:

$$C\left(U_1, U_2, \ldots, U_{n_{dim}}\right) = \Phi_{\mathbf{r}}\left(\Phi^{-1}\left(U_1\right), \Phi^{-1}\left(U_2\right), \ldots, \Phi^{-1}\left(U_{n_{dim}}\right)\right) \tag{6.59}$$

where $\Phi_{\mathbf{r}}$ Multivariate normal distribution of dimension n_{dim} with multivariate correlation $\mathbf{r}$
Φ^{-1} inverse of the Standard Normal distribution

The copula density is given by:

$$c\left(U_1, U_2, \ldots, U_{n_{dim}}\right) = \frac{1}{\sqrt{|\mathbf{r}|}} \cdot \left(-\frac{1}{2} Z^T \left(\mathbf{r}^{-1} - \mathbf{I}\right) Z\right) \tag{6.60}$$

where $\mathbf{r}^{-1}$ the inverse of the correlation matrix
$\mathbf{I}$ the identity matrix
$Z = \phi^{-1}(U)$ Standard Normal transform of U

The Gaussian copula is constructed from any tuples of random variables $X_1, X_2, \ldots X_{n_{dim}}$ by first transforming the random variables in the unit space using their inverse distribution $U_i = F^{-1}(X_i)$ and further transforming them to Standard Normally distributed values $Z_i = \Phi^{-1}(U_i)$ and by finally applying a multivariate normal distribution.

If the variables X and Y are transformed into Standard Normal space, the fit of a Gaussian copula is reduced to the calculation of the multivariate correlation $\mathbf{r}$ according to Eq. (6.41) (Renard and Lang, 2007). It can be estimated by the Spearman's rank correlation ρ_{ij} (Eq. (6.47)) between any combination of variables i and $j \in n_{dim}$:

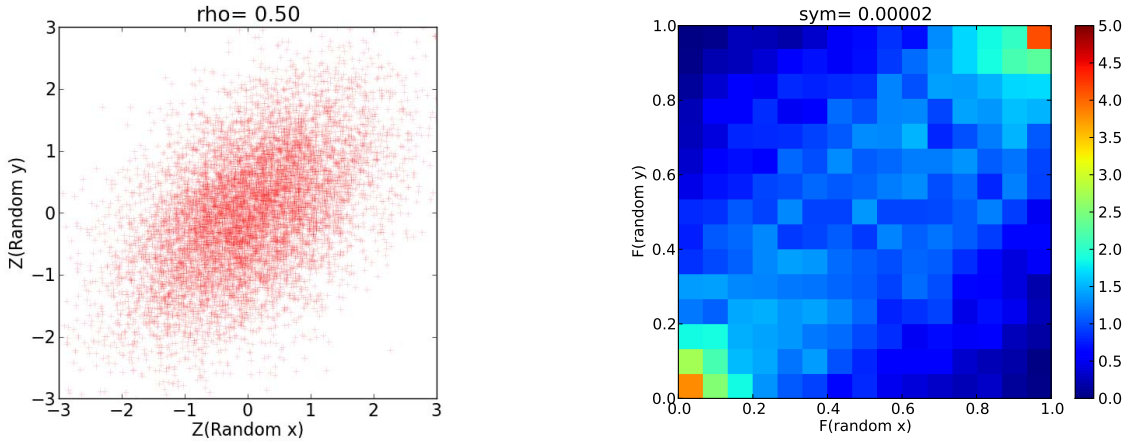
$$r_{i,j} = 2 \sin\left(\frac{\pi}{6} \rho_{i,j}\right) \tag{6.61}$$

The departures of $r_{i,j}$ from $\rho_{i,j}$ are in the range of a few percent.

Fig. 6.4 shows an example of a bivariate Gaussian dependence structure. The scatter plot Fig. 6.4a exhibits the Standard Normal transformed values of 10000 random pairs drawn from Gaussian copula with a rank correlation of $\rho = 0.5$. Fig. 6.4b displays the empirical copula density for a grid with 16×16 sub-squares. The distribution of Y does not depend on X except for the average. The probability density on any vertical (or horizontal) line through Fig. 6.4a follows the Standard Normal distribution. Transformed to the copula space, it results in the typical elliptic shaped density of the Gauss copula.

The Gauss copula is symmetric. The symmetry of a bivariate Copula can be checked by:

$$Sym = \frac{1}{n} \sum_{i=1}^{n} \left(\left(F_n\left(x_i\right) - 0.5\right)^2 \left(F_n\left(y_i\right) - 0.5\right) + \left(F_n\left(x_i\right) - 0.5\right) \left(F_n\left(y_i\right) - 0.5\right)^2\right) \tag{6.62}$$



(a) Standard Normal transformed

(b) copula density

Figure 6.4.: Scatter plot in the Standard Normal transformed space and empirical copula density of 10000 random pairs with rank correlation $\rho = 0.5$ and Gaussian dependence structure

where $F_n(x_i)$ observed non-exceedance frequency of x according to Eq. (6.54)
$F_n(y_i)$ observed non-exceedance frequency of x according to Eq. (6.54)
n number of data pairs

(See Haslauer (2011))
The name symmetry is somehow misleading as a perfectly symmetric copula would exhibit $Sym = 0$. Generally the values of sym are small. A value exceeding 0.01 would indicate a significantly skewed copula.

6.5. Analysis of Observed Precipitation Data

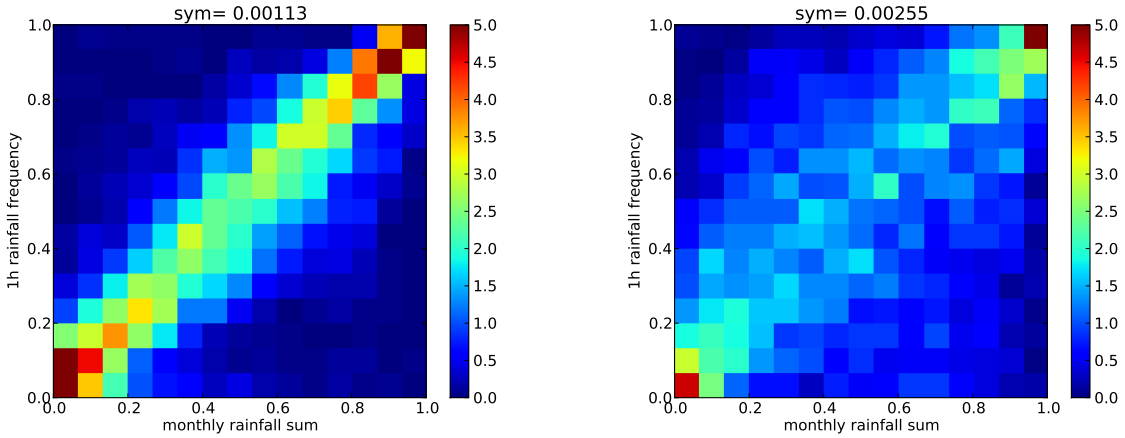
6.5.1. Multivariate Dependence in Observed Data

This section will explore the interdependencies of monthly rainfall sum H_{mon}, rainfall frequency $p_{1\text{h}}$, average $\bar{H}_{wet}$ and standard deviation s_{wet} of wet hour precipitation amounts in the available observed data from 292 rain gauges (Section 3.3). It is assumed that the dependence structure between the statistical characteristics is independent of the location throughout the whole study region and that it is an expression of the overall climatic conditions of the region.

The dependence structure is described by a copula (Section 6.4.3) in the four dimensional space of the $[0, 1]$-transformed ranks from H_{mon}, $P_{1\text{h}}$, $E(H_{wet})$ and $D(H_{wet})$. Since copulas exploit the ranks, they do not depend on absolute values. The absolute values, seen as an expression of the local effects, are modeled by the marginal distribution, which is explored in the next section (Section 6.5.2).

A four dimensional Copula cannot be displayed on two dimensional paper. Instead, the dependence between H_{mon}, P_{1h}, $E(H_{wet})$ and $D(H_{wet})$ is analyzed pairwise. The main question is whether the bivariate Copulas are Gaussian or not. Only if all bivariate Copulas are of Gaussian type, it can be assumed that the four dimensional Copula might also be Gaussian. This, however, is only a necessary condition but not a sufficient condition.

All 292 stations in the study area are examined at once. The empirical statistics H_{mon}, p_{1h}, $\bar{H}_{wet}$ and s_{wet} are calculated for each month in the whole measurement period from 1958 to 2003. Each month at each station contributes with one data pair to the bivariate copula densities. The overall data volume is of $n = 27788$ station-months which is an average of about 95 months per station. The empirical copulas are calculated according to the five point scheme in Section 6.4.3.1. To account for seasonal effects, the copulas are calculated for each month separately. Table 6.1 gives an overview of the rank correlations values ρ (Eq. (6.46)) and the values of the symmetry measure sym (Eq. (6.62)). In the following the empirical bivariate copula densities for all parameter combinations are displayed (Fig. 6.5 to Fig. 6.10). The two plots of each figure feature the months that differ the most in terms of the empirical copula density.



(a) March (b) July

Figure 6.5.: Empirical bivariate copula between monthly rainfall sum and 1 h rainfall frequency calculated from month-wise censored data

There is a significant dependence between the monthly rainfall sum H_{mon} and the 1 h rainfall frequency p_{1h} (Fig. 6.5). The rank correlation is highest in March and October. The cyclonic activity over the Atlantic Ocean is also highest in spring and autumn and lowest during summer (Weischet and Endlicher, 2008). It can be assumed that advective precipitation from cyclonic fronts enhances the dependence (Fig. 6.5a), whereas free convective precipitation with high intensity and short duration in Summer months decreases the dependence (Fig. 6.5b). In summer, the empirical copula density is slightly asymmetric. The dependence between pairs with high monthly sums and high average precipitation is weaker.

The dependence of the average rainfall amount of wet hours $\bar{H}_{wet}$ on the monthly sum H_{mon} in comparison is lower. A higher monthly sum is to a greater part generated by more rainy

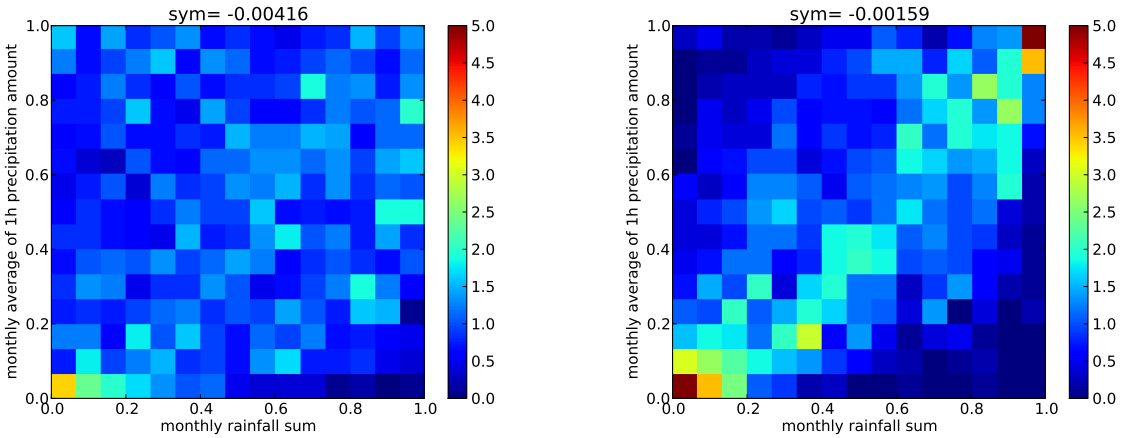
Table 6.1.: Rank correlation and symmetry of the empirical copula from monthly data for bivariate combinations of all parameters

Month	$H_{mon} - p_{1\text{h}}$		$H_{mon} - \bar{H}_{wet}$		$H_{mon} - s_{wet}$	
	ρ	Sym	ρ	Sym	ρ	Sym
January	0.85	-0.0017	0.56	-0.0017	0.55	-0.0028
February	0.81	-0.0023	0.60	0.0016	0.59	0.0008
March	0.86	0.0011	0.57	-0.0019	0.48	-0.0043
April	0.80	0.0023	0.49	-0.0004	0.41	0.0001
May	0.83	-0.0010	0.33	-0.0063	0.20	-0.0075
June	0.73	-0.0027	0.21	-0.0041	0.10	-0.0049
July	0.55	0.0026	0.40	-0.0046	0.37	-0.0068
August	0.75	-0.0020	0.46	-0.0021	0.45	-0.0034
September	0.80	0.0004	0.16	-0.0042	0.10	-0.0054
October	0.86	0.0001	0.57	-0.0054	0.48	-0.0062
November	0.84	-0.0002	0.63	-0.0016	0.59	-0.0017
December	0.77	0.0021	0.64	-0.0002	0.57	-0.0003

Month	$p_{1\text{h}} - \bar{H}_{wet}$		$p_{1\text{h}} - s_{wet}$		$\bar{H}_{wet} - s_{wet}$	
	ρ	Sym	ρ	Sym	ρ	Sym
January	0.40	-0.0015	0.44	-0.0019	0.86	-0.0019
February	0.37	-0.0009	0.40	-0.0014	0.90	0.0004
March	0.35	-0.0033	0.30	-0.0063	0.84	-0.0011
April	0.18	0.0008	0.18	0.0010	0.85	-0.0008
May	0.04	-0.0073	-0.02	-0.0081	0.85	-0.0003
June	-0.21	-0.0054	-0.22	-0.0055	0.82	-0.0002
July	-0.18	-0.0030	-0.11	-0.0040	0.88	-0.0012
August	0.10	-0.0037	0.19	-0.0042	0.84	-0.0009
September	-0.22	-0.0028	-0.20	-0.0036	0.85	-0.0001
October	0.31	-0.0068	0.25	-0.0068	0.86	-0.0020
November	0.41	-0.0014	0.40	-0.0005	0.85	-0.0013
December	0.32	0.0033	0.29	0.0028	0.90	-0.0004

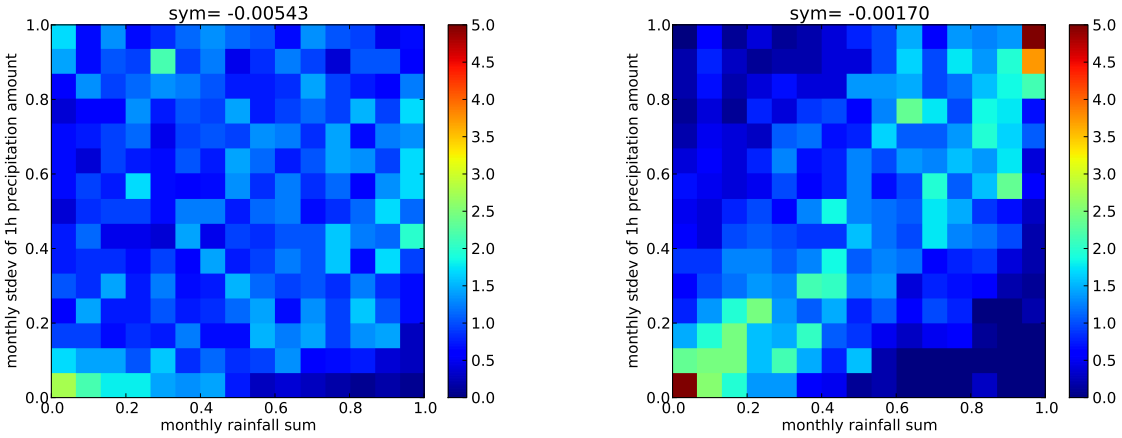
hours than by higher 1 h precipitation amounts. The correlations are lowest in the beginning of the convective summer season from Mai to June and in September. They are highest in November, December and February. Despite the high difference in rank correlation ($\rho = 0.16$ in September, $\rho = 0.59$ in November) the shape of the copula density does not look very different (Fig. 6.6a and Fig. 6.6b). Both exhibit the typical lens shape of the Gaussian copula.

The rank correlation between H_{mon} and $\bar{H}_{wet}$ on the one hand and between H_{mon} and s_{wet} on the other hand shows the same yearly cycle (Table 6.1). It is the same months that exhibit the highest and lowest rank correlations. The absolute values, however, are slightly lower for the combination of H_{mon} and s_{wet}. In both cases, the shape of the empirical copulas between H_{mon} and s_{wet} is close to Gaussian shape (Fig. 6.7).



(a) September (b) November

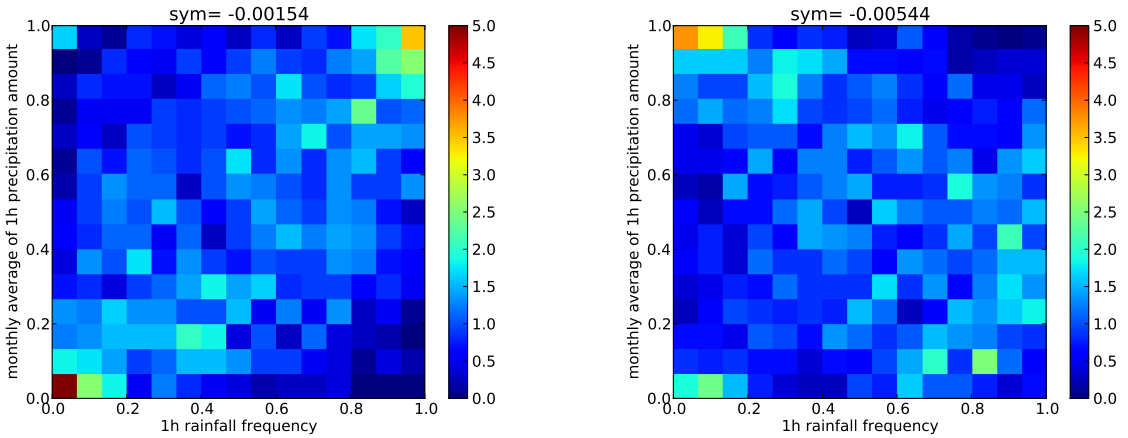
Figure 6.6.: Empirical bivariate copula between monthly rainfall sum and average 1 h rainfall amount of wet hours calculated from month-wise censored data



(a) September (b) November

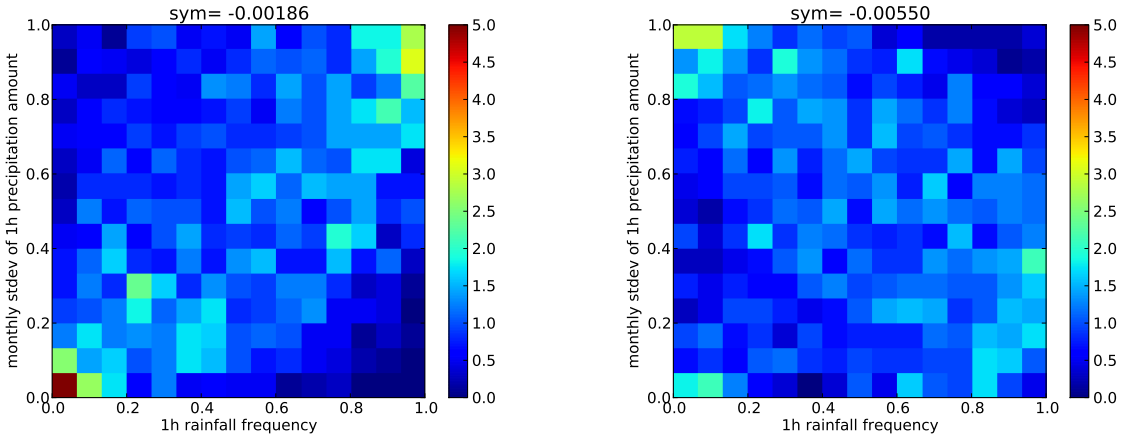
Figure 6.7.: Empirical bivariate copula between monthly rainfall sum and standard deviation of 1 h rainfall amounts of wet hours calculated from month-wise censored data

The average $\bar{H}_{wet}$ and the standard deviation s_{wet} also react very similar in their dependence to the 1 h rainfall probability. In the convective summer season the correlations are negative (except for August). It is not during the rainiest summers (with the most rainy hours) that the highest intensities are expected. Generally, long lasting advective events have lower rainfall intensities than convective events of short duration. In absence of convective events, between September and April, there is a clear positive correlation between $P_{1\mathrm{h}}$ and $\bar{H}_{wet}$ (and between $P_{1\mathrm{h}}$ and s_{wet}). In the empirical copula density Fig. 6.8b and Fig. 6.9b one can see the clear concentration of points in the upper left corner. These points represent months



(a) January (b) June

Figure 6.8.: Empirical bivariate copula between 1 h rainfall probability and average 1 h rainfall amount of wet hours calculated from month-wise censored data

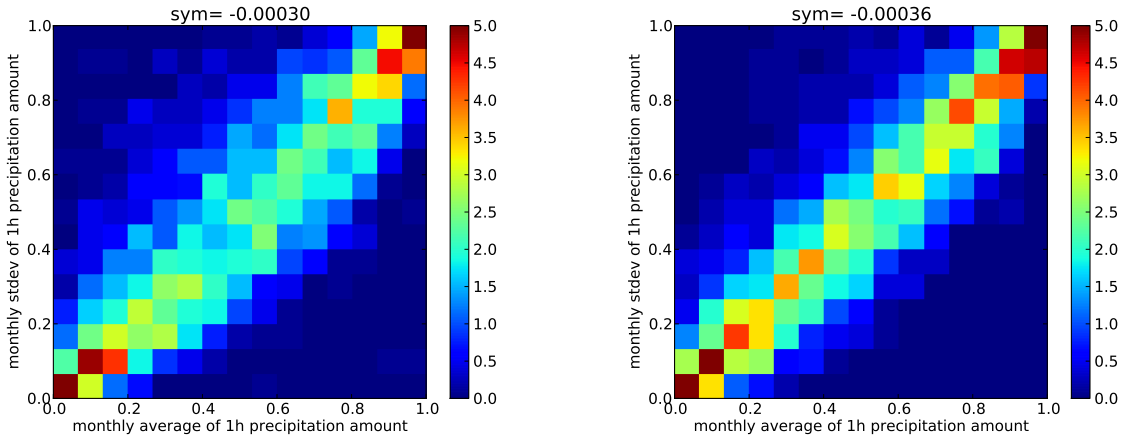


(a) January (b) June

Figure 6.9.: Empirical bivariate copula between 1 h rainfall probability and standard deviation of 1 h rainfall amounts of wet hours calculated from month-wise censored data

with very few but intensive rainfall events. A combination that is much less frequent in the plots of the advective seasons (Fig. 6.8a and Fig. 6.9a).

The rank correlation between the average $\bar{H}_{wet}$ and the standard deviation s_{wet} of wet hour precipitation shows only minor variations between the months (Fig. 6.1). The rank correlation is lowest during May, June and July and highest in December. The empirical copula plots look all very similar (Fig. 6.10). They exhibit a high symmetry (a low value of sym) and the elliptical shape of the Gaussian copula.



(a) May (b) December

Figure 6.10.: Empirical bivariate copula between average and standard deviation of 1 h rainfall amounts of wet hours calculated from month-wise censored data

With the exception of the dependence between the average and the standard deviation of precipitation amounts of wet hours, distinct differences between the seasons can be found. The cyclonic activity which is highest in spring and autumn enhance the dependence of $P_{1\mathrm{h}}, \bar{H}_{wet}, s_{wet}$ on the monthly sum, the occurrence of convective events from May to September lead to a decrease in rank correlation. The strongest yearly cycle is observed in the correlation of the monthly sum H_{mon} to the average $\bar{H}_{wet}$ and standard deviation s_{wet} of precipitation amounts of wet hours.

The statistics of p_{1h}, $\bar{H}_{wet}$, s_{wet} all exhibit a positive rank correlation to the monthly rainfall sum. These findings are in favor of the idea that the rainfall generation can be conditioned on the monthly sum. Assuming a Gaussian copula leads to some minor distortions of the real dependence structure, most of all in the dependence of the average amount and the standard deviation on the rainfall frequency. However, the dependence is anyway weak in these combinations and so the departures from a Gaussian copula are of minor importance.

6.5.1.1. Gaussian Copula Dependence used for Rainfall Generation

The described dependencies will be exploited in the rainfall simulations to draw values of $P_{1\mathrm{h}}$, $E\left(H_{wet}\right)$ and $D\left(H_{wet}\right)$ that are conditioned on the observed monthly precipitation sum H_{mon}. The dependence structure will be calculated for three monthly seasons (December to February, March to Mai, June to August and September to November). Table 6.2 displays the respective correlations r in the Standard Normal transformed space.

6.5.2. Observed Marginal Distributions

According to Sklar's theorem, the description of the dependence structure and of the distributions can be separated. In the last section the copula dependence between the rainfall

Table 6.2.: Correlation in the Standard Normal transformed space of monthly precipitation sum, hourly rainfall probability and monthly average and standard deviation of wet-hour precipitation amounts

December to February				
	H_{mon}	P_{1h}	$E(H_{wet})$	$D(H_{wet})$
H_{mon}	1.000	0.873	0.623	0.589
P_{1h}	0.873	1.000	0.387	0.405
$E(H_{wet})$	0.623	0.387	1.000	0.895
$D(H_{wet})$	0.589	0.405	0.895	1.000

March to May				
	H_{mon}	P_{1h}	$E(H_{wet})$	$D(H_{wet})$
H_{mon}	1.000	0.838	0.503	0.423
P_{1h}	0.838	1.000	0.113	0.092
$E(H_{wet})$	0.503	0.113	1.000	0.898
$D(H_{wet})$	0.423	0.092	0.898	1.000

June to August				
	H_{mon}	P_{1h}	$E(H_{wet})$	$D(H_{wet})$
H_{mon}	1.000	0.786	0.329	0.296
P_{1h}	0.786	1.000	-0.118	-0.054
$E(H_{wet})$	0.329	-0.118	1.000	0.861
$D(H_{wet})$	0.296	-0.054	0.861	1.000

September to November				
	H_{mon}	P_{1h}	$E(H_{wet})$	$D(H_{wet})$
H_{mon}	1.000	0.854	0.418	0.352
P_{1h}	0.854	1.000	0.029	0.014
$E(H_{wet})$	0.418	0.029	1.000	0.898
$D(H_{wet})$	0.352	0.014	0.898	1.000

governing parameters H_{mon}, P_{1h}, $E(H_{wet})$ and $D(H_{wet})$ was explored. Besides, it is necessary to estimate the marginal distribution of the four statistics for the transformation of real world values into the copula space (and back).

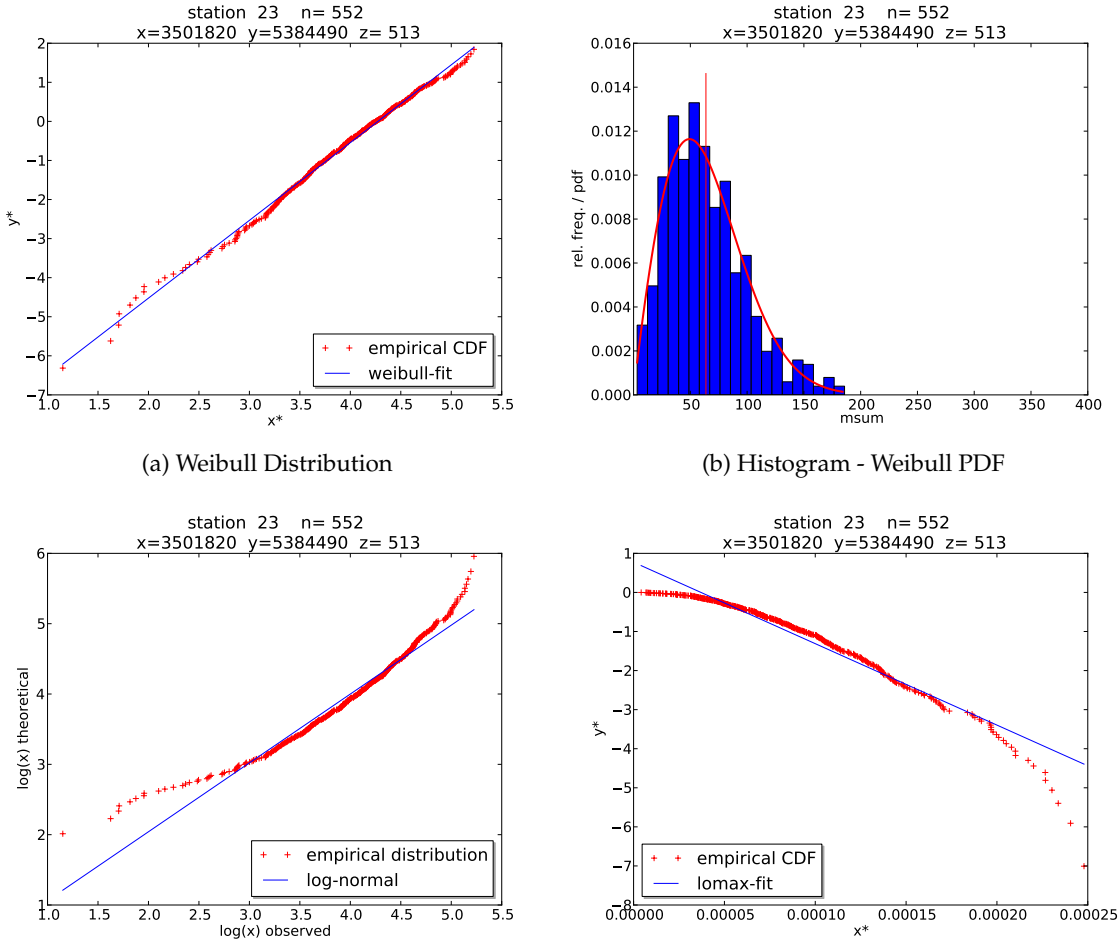
Since the dependence structure was estimated for all stations at once, it is required that one unique marginal distribution can be found to each of the statistics.

The analyzed observed statistics H_{mon}, $\bar{H}_{wet}$ and s_{wet} exhibit right skewed distributions. The values by their nature cannot become smaller than zero. Therefore, the Weibull, Log-Normal and the Pareto Distribution will be tested. To the monthly $1\,\mathrm{h}$ precipitation frequencies the

Beta Distribution will be fit since it is a distribution with support $X \in (0, 1)$.

For each statistical parameter it is tested if all 292 station agree on one distribution function. If so, the distribution is fit to the whole data set of all 27788 station months between 1958 to 2003. In the following the distributions at the station in Holzgerlingen (Gaus-Krüger 3 coordinates of X=3501820, Y=5384490, see Section 3.3) are shown as examples of the single site fit (Fig. 6.12 to 6.15).

- **Monthly Precipitation Sum**



(a) Weibull Distribution

(b) Histogram - Weibull PDF

(c) Log-Normal Distribution

(d) Lomax Distribution

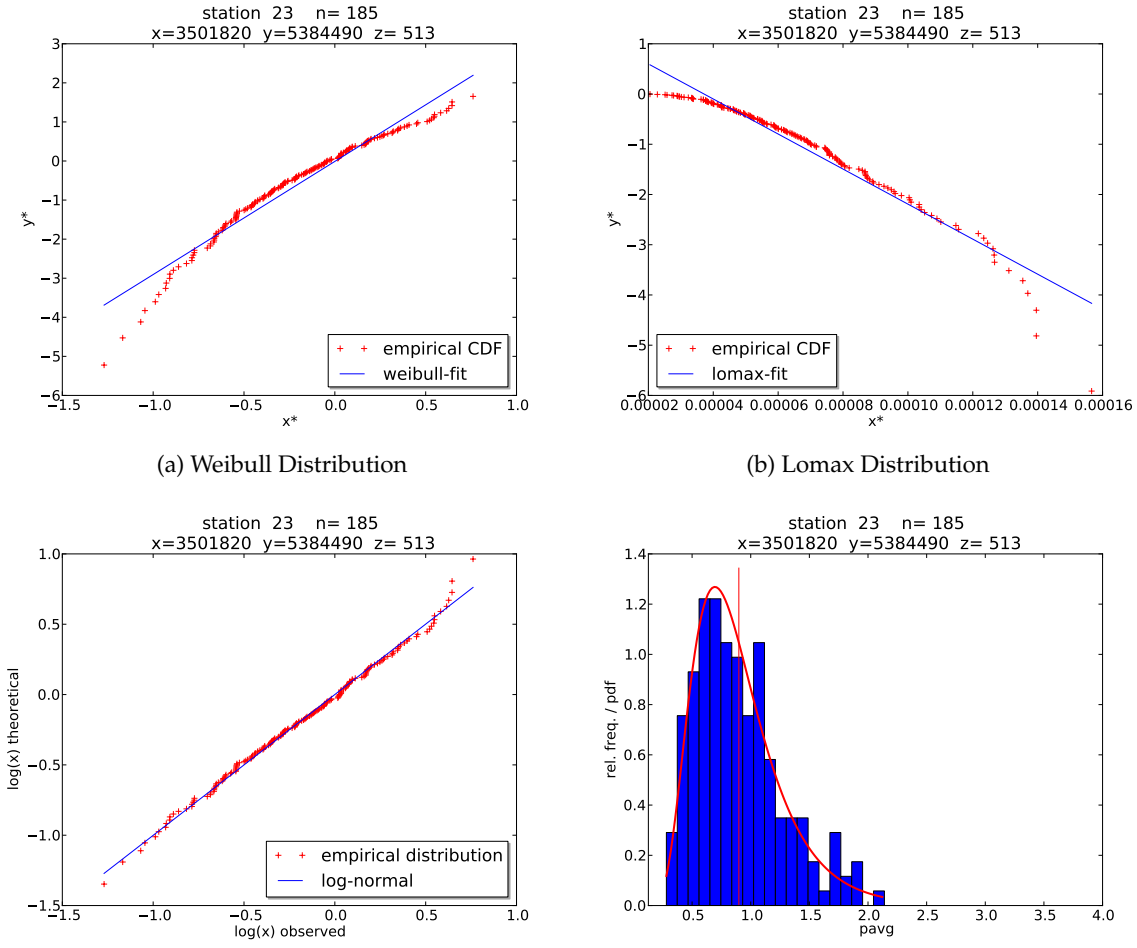
Figure 6.11.: Observed and theoretical distribution in the linearized space of the monthly precipitation sum in Holzgerlingen. Fit of three different theoretical distributions. Histogram and PDF of the best fitting Weibull Distribution.

In Fig. 6.11 the MLH-fit of the three tested distribution functions Weibull Distribution, Log-Normal Distribution and Lomax Distribution as well as the observed histogram and the PDF of the best fitting Weibull Distribution is shown. The vertical line in Fig. 6.11b indicates the average value of the distribution. The fit is performed to the

monthly sums kriged from the daily stations. In this way, a complete time series is available between 1958 and 2003. Using kriged values is a reasonable choice because the developed simulation scheme will also be based on kriged values of the monthly sum.

In Holzgerlingen, the Weibull Distribution fits best. The other two systematically overestimate the probabilities of very high monthly sums. The distributions from all 291 other stations agree on this result. At only a handful of stations the Weibull Distribution slightly underestimates the probabilities of the wettest months but remains the best choice among the three distributions.

- **Average Precipitation Amount of Wet Hours**



(a) Weibull Distribution

(b) Lomax Distribution

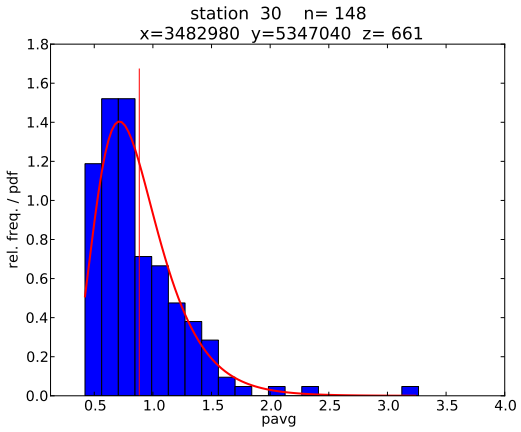
(c) Log-Normal Distribution

(d) Histogram - Log-Normal PDF

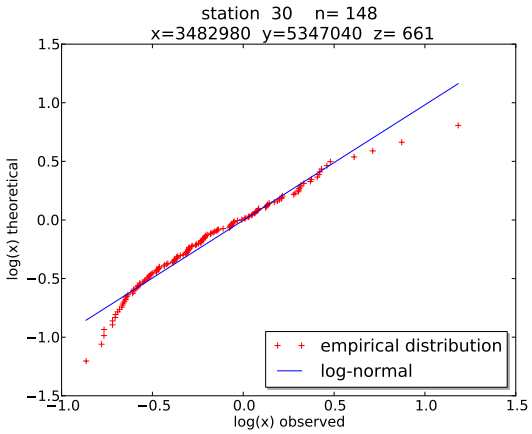
Figure 6.12.: Observed and theoretical distribution in the linearized space of the average precipitation amount of wet hours in Holzgerlingen. Fit of three different theoretical distributions. Histogram and PDF of the best fitting Log-Normal Distribution.

The distribution of the average precipitation amount of wet hours is explained best

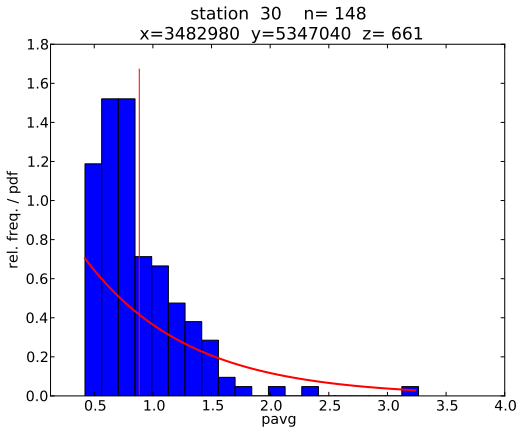
by the Log-Normal distribution (Fig. 6.12). The Weibull distribution underestimates the frequencies in the highest events. The Lomax distribution slightly overshoots in the highest events and also overestimates the probabilities of very low precipitation amounts.



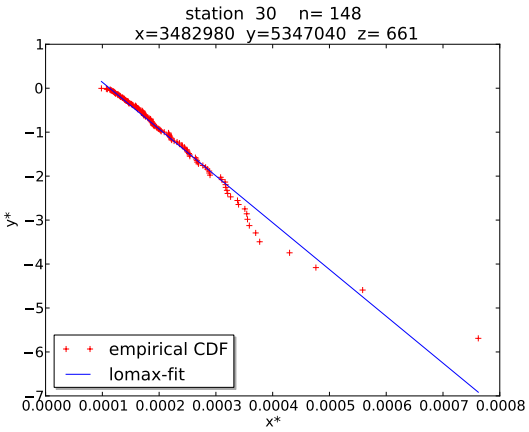
(a) Histogram - Log-Normal PDF



(b) Log-Normal Distribution



(c) Histogram - Lomax PDF

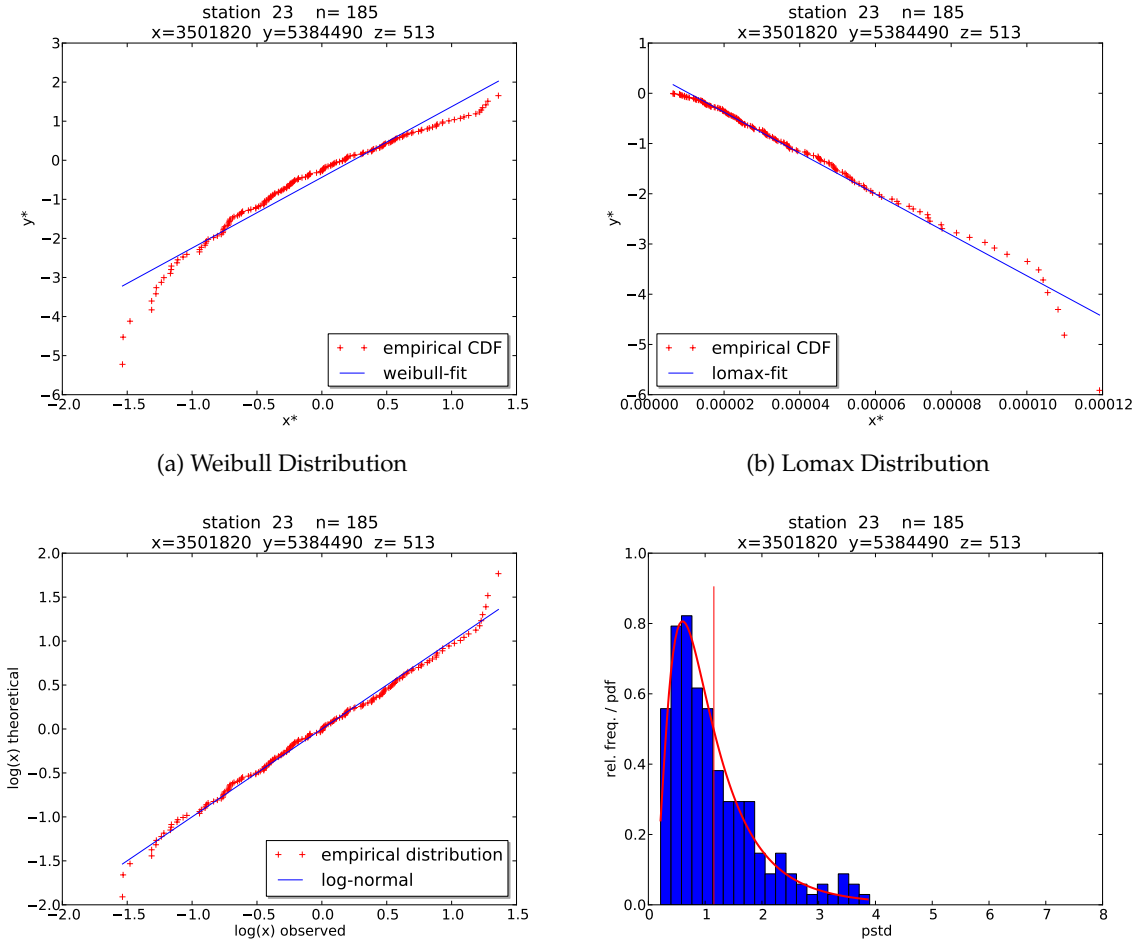


(d) Lomax Distribution

Figure 6.13.: Comparison of Weibull Distribution and Log-Normal Distribution fitted by Maximum Likelihood Method to the observed average precipitation of wet hours from Geislingen Waldhof

Holzgerlingen is again exemplary for the whole of the study region. Only few stations show minor deviations to the Log-Normal distribution, most of them an underestimation in the highest amounts. Fig. 6.13 displays a comparison of the histograms and the linearized distributions of Log-Normal and Lomax Distribution at station of Geislingen Waldhof where the fit of the Log-Normal Distribution performs worst. However, the Log-normal Distribution remains the best choice. The Lomax distribution puts more weight on the highest precipitation amounts but turns the underestimation into an overestimation. At the same time, there is a clear deviation in the observed and theoretical frequencies of the lower values.

- **Standard Deviation of Precipitation Amounts of Wet Hours**



(a) Weibull Distribution

(b) Lomax Distribution

(c) Log-Normal Distribution

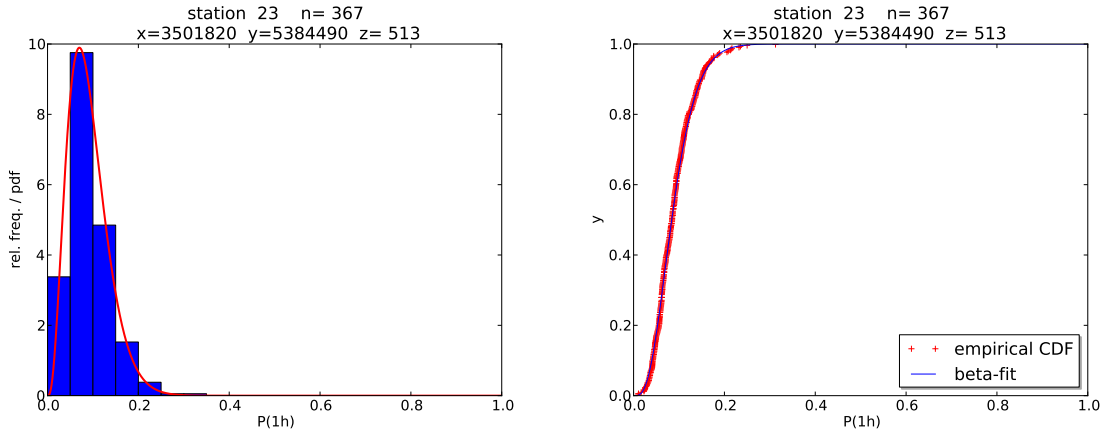
(d) Histogram - Log-Normal PDF

Figure 6.14.: Observed and theoretical distribution in the linearized space of the standard deviation of precipitation amounts of wet hours. Fit of three different theoretical distributions. Histogram and PDF of the best fitting Log-Normal Distribution.

One outcome of the multivariate dependence analysis in Section 6.5.1 is the close relation between the average and standard deviation of the hourly precipitation amounts of wet hours. They are very similar in terms of the marginal distribution, too: The best suited theoretical distribution function for the standard deviation is the Log-Normal Distribution as well (Fig. 6.14).

- **Hourly Precipitation Probability**

The monthly frequency of 1 h precipitation can be very well represented by a Beta-Distribution. The fit for the station in Holzgerlingen shown in Fig. 6.15 is representative for the whole study area. None of the 292 stations exhibits visible deviation between the CDF $F(X)$ and the observed frequency $F_n(x)$.



(a) Histogram + PDF (b) CDF

Figure 6.15.: Observed distribution and fitted Beta Distribution of the monthly 1 h precipitation frequency of wet hours at the station in Holzgerlingen

6.5.2.1. Marginal Distributions used for Rainfall Generation

In all of the four parameters the distributions at all 292 rain gauges can be described by the same theoretical distribution functions. Minor deviations, mostly in the distribution of the highest values, only occur at a few stations. It is, therefore, reasonable to apply the respective distribution to all station months at once. Fig. A.1 to Fig. A.8 in the appendix to this chapter show the histogram and the (linearized) theoretical and observed distributions. For consistency with other generation parameters the fit is performed season wise.

Table 6.3.: Distribution functions used in the parameter estimation for the rainfall generation

Parameter	Distribution Function	Parameter	December -February		March-May	
H_{mon}	Weibull Distribution	λ, κ	81.366	1.456	88.616	1.767
P_{1h}	Beta-Distribution	α, β	2.938	20.919	2.826	22.507
$E(H_{wet})$	Log-Normal Distribution	μ, σ	-0.427	0.340	-0.187	0.346
$D(H_{wet})$	Log-Normal Distribution	μ, σ	-0.436	0.465	-0.093	0.502
Parameter	Distribution Function	Parameter	June-August		September -November	
H_{mon}	Weibull Distribution	λ, κ	106.643	2.263	83.733	1.665
P_{1h}	Beta-Distribution	α, β	3.451	33.722	3.026	23.958
$E(H_{wet})$	Log-Normal Distribution	μ, σ	0.214	0.329	-0.145	0.338
$D(H_{wet})$	Log-Normal Distribution	μ, σ	0.526	0.474	-0.025	0.487

Fit to the data of all 292 stations, the distributions exhibit about the same behavior as in the fit to single stations. Fig. A.1 and Fig. A.2 show the histogram, the empirical and theoretical distributions of the monthly sum for all four seasons, which are used in the gen-

eration scheme. The fit of the Weibull distribution is best during the winter months from December to February. In spring season the scatter cloud bends concavely, indicating an underestimation of the highest observed monthly sums. The overall marginal distribution of the standard deviation and the monthly average (Fig. A.3 to Fig. A.6) show a slight underestimation in the highest values, especially during the winter months from December to February. However, the linearized plots overemphasis these deviations since the density of points in the middle of the point cloud is several times higher than near the edges.

The fit of the Beta Distribution to the 1h rainfall frequencies is good in all four seasons. No systematic deviations can be found neither in the histogram (Fig. A.7) nor in the CDF (Fig. A.8).

6.6. The Generation Algorithm for 1h Precipitation Time Series

In Section 6.5.1 it was shown that the dependence structure between the observed monthly sum, the 1h precipitation frequency, the monthly average and standard deviation of wet hours can be approximated by a Gaussian Copula. In the last section a universal marginal distribution for the whole study area was deduced for each of the parameters. Based on these findings, the parameter estimation of the NiedSim algorithm will be set up anew. The generation parameters governing the distribution of 1h precipitation values will be conditioned on the time series of observed monthly sums.

6.6.1. Parameter Estimation for the Initial Time Series

It is assumed that the 1h rainfall amounts follow a Weibull Distributions. Its parameters $\lambda(H_{wet})$ and $\kappa(H_{wet})$ are calculated for each month from stochastic values of $E(H_{wet})$ and $D(H_{wet})$ according to Eq. (6.13) and Eq. (6.14). Additionally, the value of the rainfall probability P_{1h} has to be estimated which governs the number of wet hours per month.

1. Starting point is the time series of monthly rainfall sums H_{mon} at the target location of the simulation. If the monthly sum is not available, H_{mon} is estimated from surrounding stations by External Drift Kriging (Eq. (5.11)) with the smoothed altitude as external information.

2. This target value H_{mon} is transformed into the copula space U of uniformly distributed values between 0 and 1 by its marginal distribution F:

$$u_{mon} = F(H_{mon}) \tag{6.63}$$

   u_{mon} can be seen as the non-exceedance probability of H_{mon} in reference to the whole study region and simulation time period. The parameters of the marginal distribution for the monthly sum can be found in Table 6.3.

3. The non-exceedance probabilities of 1h rainfall probability U_P, expectation of the mean U_H and expectation of the standard deviation U_s are drawn conditioned on u_{mon} from the four dimensional Gaussian copula describing the parameter interdependencies. Drawing from a Gaussian copula is identical to drawing values from a correlation matrix $\mathbf{\Sigma}$ in the Standard Normal transformed space. In the following, the index convention for the parameters is $1\hat{=}H_{mon}$, $2\hat{=}P_{1h}$, $3\hat{=}E(H_{wet})$ and $4\hat{=}D(H_{wet})$.

$$z_{|a} \sim \mathcal{N}\left(\mu_{|a}, \mathbf{\Sigma}_{|a}\right) \tag{6.64}$$

$$a = \Phi^{-1}(u_{mon}) \tag{6.65}$$

where
- $z_{|a} = \left[z_{i|a}, i = 2,3,4\right]^T$ — correlated conditioned random numbers in the Standard Normal transformed space
- $\mathcal{N}(\mu, \mathbf{\Sigma})$ — multivariate normal distribution with mean vector μ and covariance matrix $\mathbf{\Sigma}$
- $\mu_{|a} = \left[\mu_{i|a}, i = 2,3,4\right]$ — mean vector conditioned on $x_1 = a$
- $\Sigma_{|a}$ — covariance matrix conditioned on $x_1 = a$

The conditioned mean vector $\mu_{|a}$ is calculated by multiplying the values r_{1j} , $j = 2, 3, 4$ of the first column of the multivariate correlation matrix (Table 6.2) with the conditioning value a. The conditioned correlation matrix $\mathbf{\Sigma}_{|a}$ can be calculated by inverting the overall correlation matrix $\mathbf{\Sigma}$ according to Table 6.2, dropping the row $i = 1$ and column $j = 1$ belonging to H_{mon} and inverting back. (For more mathematical details, see Section A.2 in the appendix).

The conditioned, Standard Normal transformed random numbers $z_{|a}$ are drawn by means of a vector $z =$ of uncorrelated random numbers and lower triangle matrix $\mathbf{L}$ of the Cholesky decomposition of $\mathbf{\Sigma}_{|a}$

$$z_{|a} = \mu_{|a} + \mathbf{L} \cdot z \tag{6.66}$$

$$\mathbf{\Sigma_{|a}} = \mathbf{L}\mathbf{L}^T \tag{6.67}$$

4. The estimated Standard Normal transformed values x_i^* representing the rainfall probability, as well as the average and the standard deviation of precipitation amounts of wet hours are back-transformed to absolute values in two steps.

$$u_i = \Phi\left(z_{i|a}\right) \tag{6.68}$$

$$x_i^* = F_i^{-1}\left(u_i\right) \tag{6.69}$$

where
- x_i^* estimated target value of parameter $i = 2, 3, 4$
- Φ Standard Normal Distribution Function
- $z_{i|a}$ conditioned random number in the Standard Normal transformed space drawn for parameter i
- u_i uniform transformed of $z_{i|a}$
- F_i^{-1} inverse of the theoretical distribution function of parameter i according to Table 6.3

5. The parameters $\lambda\left(H_{wet}\right)$ and $\kappa\left(H_{wet}\right)$ of the Weibull Distribution for the 1 h rainfall amounts of wet hours are calculated from the drawn statistics $E^*\left(H_{wet}\right)$ and $D^*\left(H_{wet}\right)$ by the method of moment (Eq. (6.13) and Eq. (6.14)).

To account for seasonal effects, the parameters of the marginal distribution (Table 6.3) and the Gaussian copula (Table 6.2) are depending on the season (December to February, March to May, June to August, September to November)

6.6.2. Rainfall Generation

The generation is performed year by year.

1. Given the monthly series of rainfall probabilities $P_{1\text{h}}$ and the Weibull Distribution parameters $\lambda\left(H_{wet}\right)$ and $\kappa\left(H_{wet}\right)$ according to the proceeding described in the last section, an initial time series is generated. For each time-step in the year

a) a random number u is drawn from a uniform distribution in the interval $[0, 1]$. If u is not greater than the rainfall probability of the current month ($u \leq P_{1\text{h}}$), the current time step is assumed to be a rainy hour.

b) If the current hour is wet, another random number u is drawn. By the inverse of the Weibull Distribution with the parameters $\lambda\left(H_{wet}\right)$ and $\kappa\left(H_{wet}\right)$ of the current month, u is transformed into a 1 h precipitation amount $H\left(t\right)$.

$$H\left(t\right) = F^{-1}\left(u, \lambda, \kappa\right) \tag{6.70}$$

c) 1a to 1b are repeated for the next time step.

The initial time series exhibits the right distribution of the 1 h precipitation amounts and follows the expectation of the monthly precipitation sums. It is passed to the permutation scheme for the optimization of the temporal sequence of values.

2. The optimization is performed by a Simulated Annealing algorithm (Aarts and van Laarhoven, 1989). It is based on an objective function O that measures the deviations of the current state from the statistical target state for a perfect simulation. The better the simulated time series reproduces the statistical features expected for a measured time series at the same location, the smaller the value of O.

$$O = w_1 O_1 + w_2 O_2 + w_3 O_3 + \ldots + w_n O_n \rightarrow \text{min!} \tag{6.71}$$

$$O_i = \left(\rho_{i,sim} - \rho_{i,tar}\right)^2 \tag{6.72}$$

where $\rho_{i,sim}$ value of the simulated series for the i^{th} statistical feature
$\rho_{i,tar}$ target value for the i^{th} statistical feature

The weights $w\left(i\right)$ of the objective function parts are chosen depending on possible range in O_i. By the weights $w\left(i\right)$ it is assured that the different statistics i have more or less the same influence on the overall objective function O.

The Simulated Annealing algorithm consists of the following steps:

a) Randomly choose one value of the time series i and a second value j within the same month so that

$$\max\left(x_i, x_j\right) > 0$$

(If both values are zero, the exchange has no effect.)
Exchange the two values and calculate the state of the objective function for the altered series O^*.

b) If the exchange improves the state of the objective function

$$O^* < O_{old}$$

keep the changes.

$$O_{old} = O^*$$

Go to 2a and perform the next exchange.

c) If the swap worsens the time series ($O^* > O_{old}$), it is kept anyhow with a certain probability. This feature is implemented to avoid that the optimization gets stuck in a local optimum. The probability of keeping a "bad" swap depends on the difference in the objective function and an optimization parameter T called the annealing temperature.

$$P_{keep} = \exp\left(\frac{O^* - O_{old}}{T}\right) \tag{6.73}$$

For each bad swap, a uniform random number $U \in [0, 1]$ is drawn. If $U \leq P_{keep}$, the swap is kept. The algorithm continues at 2a by choosing the next two time steps for a swap.

d) The steps 2a to 2c are repeated n_T times. Then the temperature is decreased

$$T_{new} = d_T \cdot T_{old} \tag{6.74}$$

where $d_T < 1$.

e) Each temperature decrease reduces the probability of accepting a "bad" swap. At the beginning of the simulation the system has much freedom in performing different swaps. Later mostly "good" swaps are accepted. The generation stops when the number of performed swaps drops below a pre-defined value (e. g. one swap in 10000 trials).

6.6.2.1. Statistical Parameters in the Objective Function

The different parts O_i of the objective function evaluating the simulated time series are:

- **24 h extreme value according to KOSTRA statistics**
  The official German extreme value statistics KOSTRA (Bartels et al., 2005) assumes that the extreme precipitations of the annual series (the highest precipitations of each year) follows a Gumbel Distribution. The parameters of this Gumbel distribution can be re-established by the yearly and 100 yearly extreme value from KOSTRA. The restored distribution is used to draw the target value for the highest 24h precipitation sum in the simulation year.
- **scaling of the statistical moments between different aggregation levels**
  The k^{th} statistical moment m_k of an empirical distribution is defined as:

$$m_k = \frac{1}{n} \sum_{i=1}^{n} x\,(i) \tag{6.75}$$

where n number of values in the distribution (here: number of time steps in the simulated time series for the current year)

The characteristics in the scaling of the k^{th} statistical moments over different aggregation times Δ_t are regarded. Δ_t ranges from $1\,\text{h}$ to $24\,\text{h}$. The scaling of the k^{th} moments is described by a non-linear regression function:

$$m_k\left(\Delta_t\right) = a \cdot \Delta_t^{b_k} \tag{6.76}$$

If the simulated time series exhibit the right scaling behavior, the exponent b_k of the simulated data will correspond to the target value. The focus is not on the absolute values but on the relation of the values from different aggregations. Therefore, the parameter a is ignored and only b is considered as a target value in the objective function.

- **autocorrelation on different aggregation levels**

$$r\left(l\right) = \frac{\sum_{i=1}^{n-l} w\left(i\right) \cdot \left(x_i - \bar{x}\right)\left(x_{i+l} - \bar{x}\right)}{\sum_{i=1}^{n} \cdot \left(x_i - \bar{x}\right)} \tag{6.77}$$

where $w\left(i\right)$ a seasonal weighting factor
n number of values in the time series (here the number of time steps in the simulated time series of the current year).
k lag of the autocorrelation.

The autocorrelation is a measure for the memory of the time series. In the generation the autocorrelation to the preceding time step is considered, therefore $k = 1$. It is calculated for six different aggregations of the yearly time series $x\left(\Delta_t\right)$ with $\Delta_t = 1\,\text{h}, 2\,\text{h}, 3\,\text{h}, 6\,\text{h}, 12\,\text{h}$ and $24\,\text{h}$. The weighting factor is set to $w = 1$.

Additionally, a seasonally weighted autocorrelation is calculated for the time series of the $1\,\text{h}$ rainfall amounts, the weighting factor $w\left(i\right)$ depends on the relative position of the time step in the year:

$$w\left(i\right) = 1 + \sin\left(\frac{i}{n_{year}} \cdot 2\pi + \frac{1}{6}\right) \tag{6.78}$$

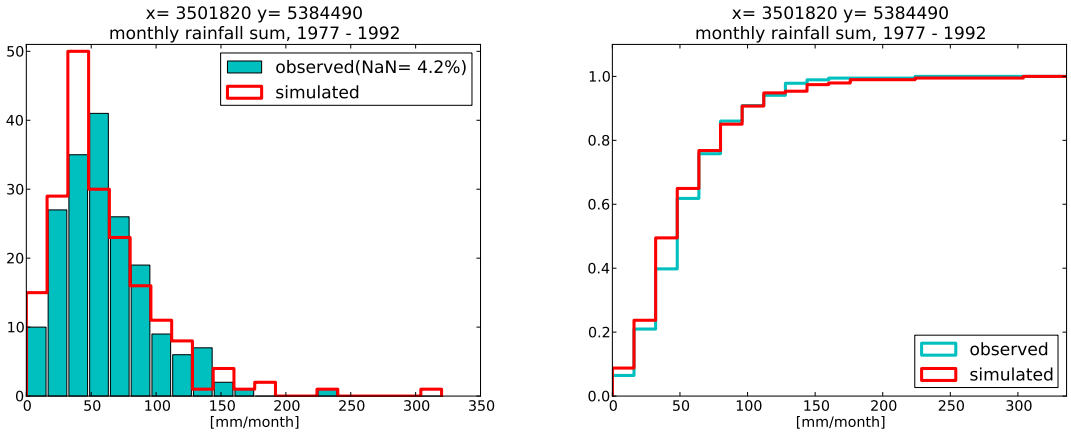
The target values of the autocorrelation depend on the generation year.

- **average daily precipitation amount and frequency for three different weather type classes**
The weather classes are found by a Fuzzy Logic Based classification algorithm that divides the days by their mean sea level pressure patterns (see Section 7.5.2). The originally 17 different weather types are regrouped into three classes: dry, medium and humid. The average of daily precipitation amount and frequency for the three groups can be found in Table 7.6 to Table 7.8. The statistics are calculated from 575 daily rain gauges (Section 3.3).

Since the monthly precipitation sums do not correspond to the observed values but are drawn randomly, the average precipitation is not considered as absolute values. It is normed by the simulated monthly rainfall sum.

6.7. Comparison of Simulated and Observed Rainfall Time Series

The performance of the generator is tested by a comparison of the statistical characteristics of observed and simulated time series at the station in Holzgerlingen (Gaus-Krüger 3 coordinates of X=3501820, Y=5384490, see Section 3.3). The time period of the comparison ranges from 1977 to 1992.



(a) Histogram (b) CDF

Figure 6.16.: Empirical Distribution of the simulated and observed monthly precipitation sum.

Since the generation is conditioned on the monthly precipitation sums, the generation restores the distribution of the monthly precipitation sums (Fig. 6.16) as well as the average yearly sum (727 mm in the simulations, 735 mm observed). The small fluctuations are due to sampling variability.

The comparison of hourly statistics presented Table 6.4 reflects the accuracy in the distribution of the 1 h time series. The daily statistic reflect not only the distribution but also the temporal sequence in the generated time series. The calculation of hourly mean, standard deviation and skewness is based on all values exceeding a threshold of 0.1 mm. The threshold for daily values is 0.5 mm.

Over all, the statistics of the hourly time series are modeled correctly Table 6.4. Significant deviations concern the rainfall frequency $p_{>0.1mm}$ and the average precipitation of wet hours. Since the generation respects the yearly sum, the underestimation of the rainfall frequency results in an overestimation of the average hourly rainfall. The deviations, however, only concern the smallest values, the frequencies with higher threshold are all modeled correctly.

In daily aggregation too the main difference between the observed and simulated time series is found in the rainfall frequency. The fewer number of rain days indicates that the simulated rainfall events are more concentrated than in reality. Another deviation concerns the

Table 6.4.: Basic statistics of the simulated and observed time series in Holzgerlingen, 1977 to 1992

	hourly values		daily precipitation sum	
	simulated	observed	simulated	observed
mean [mm]	1.108	0.879	8.655	5.432
stdev. [mm]	1.498	1.382	7.485	6.403
skewness [-]	7.643	6.605	0.870	3.199
$p_{>0.1mm}$ [-]	7.34%	9.04%	32.3%	43.4%
$p_{>0.5mm}$ [-]	4.34%	4.32%	22.6%	36.8%
$p_{>2mm}$ [-]	1.07%	0.85%	16.3%	24.4%
$p_{>5mm}$ [-]	0.164%	0.153%	12.6%	12.6%
$p_{>10mm}$ [-]	0.017%	0.033%	9.22%	5.39%
$p_{>20mm}$ [-]	0.0004%	0.0005%	0.84%	1.37%
H_{max} [mm]	42.9	27.2	42.9	69.2

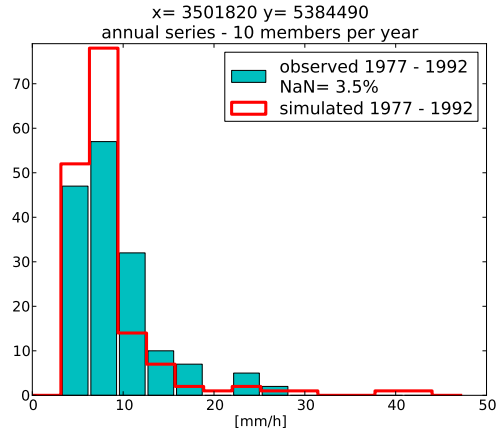
skewness. The frequency of small daily sums is underestimated, the frequency of $p_{>10mm}$ is overestimated.

As a representation of the extreme value behavior the empirical distribution of the ten highest precipitation amounts is calculated (Fig. 6.17). On hourly aggregation the distribution of the simulated time series is reasonable, although moderate extremes (between 10 mm and 20 mm) are slightly underestimated. The same is true for the six hourly extremes. On daily scale, the new generation scheme exhibits a slight lack of variability. The frequency of values around 15 mm is overestimated. The most extreme events, on the other hand, are lower than the observed. The lack of very high values is possibly caused by the KOSTRA statistics, which restricts the depth of the 24 h precipitation sums (see Section 6.6.2.1). This is supported by the fact that the simulated maximum hourly and daily value are the same.

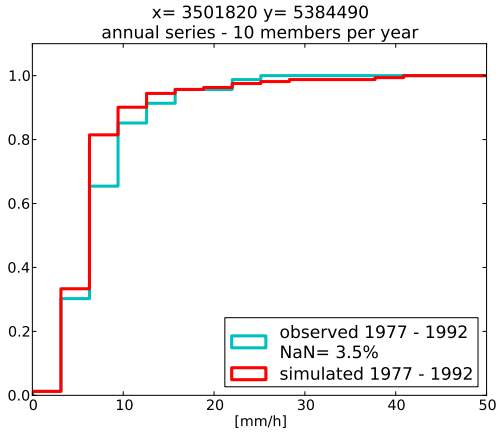
Calculating the annual series of 1 h values for each month separately reveals the seasonal variability in the simulated time series. Fig. 6.18 displays the distributions for January and July. The generation is able to distinguish between the seasons. In both months, the distribution is close to the observed. The representation of seasonal variability is a new feature. In NiedSim the same distribution is applied during the whole year.

Fig. 6.19 presents the distribution of the wet spell lengths. Gaps of one hour are not considered as dry spells and therefore do not interrupt a wet spell. This threshold is introduced to avoid that longer, dependent precipitation events are interrupted due to a short break in rainfall. To avoid a high influence of measurement errors, only hours with no less than 0.1 mm rainfall are considered as wet.

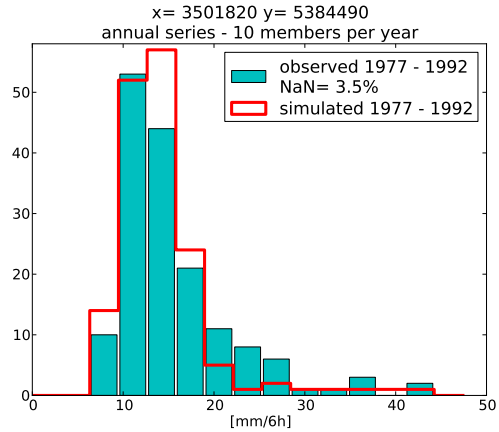
The generated time series counts 2515 wet spells which is in the same range as the observed time series with 2380 wet spells. The frequencies are generally captured well,except of very short wet spells that are too frequent. The simulated time series slightly underestimates the duration of the longest wet spells (Fig. 6.20a). However, the distribution is very sensitive to the definition of a wet spell. If gaps up to three hours are accepted, the duration is



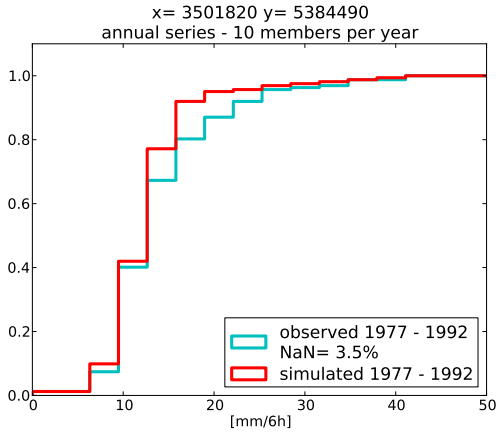
(a) Histogram - 1 h



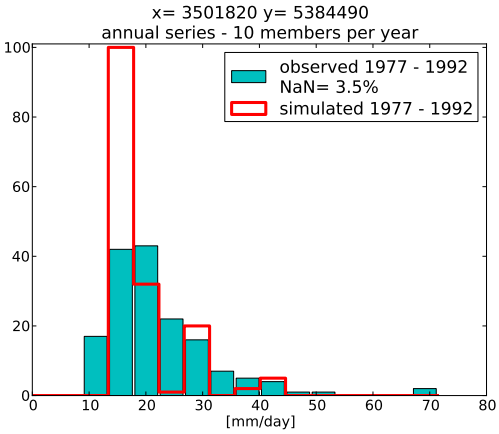
(b) CDF - 1 h



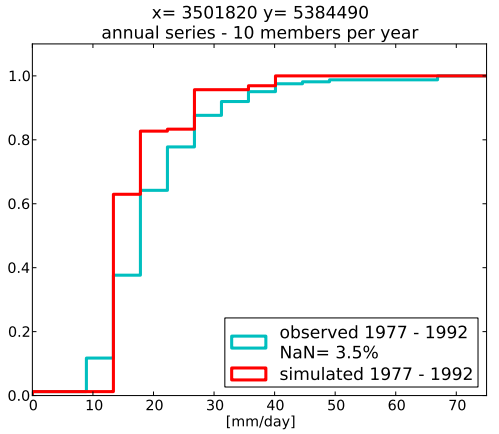
(c) Histogram - 6 h



(d) CDF - 6 h

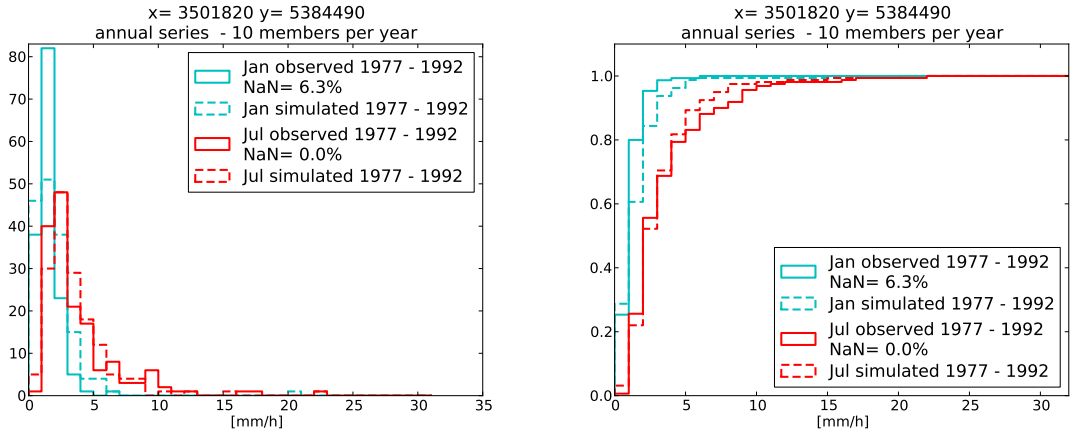


(e) Histogram - 24 h



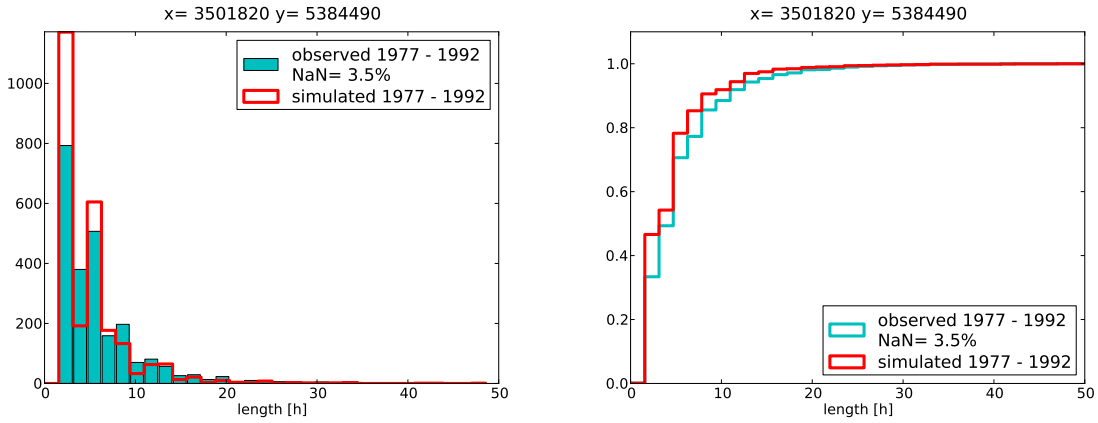
(f) CDF - 24 h

Figure 6.17.: Empirical distribution of the ten highest rainfall amounts in the observed and the simulated time series



(a) Histogram (b) CDF

Figure 6.18.: Empirical distribution of the ten highest 1 h rainfall amounts in the observed and simulated time series during January and July for the station in Holzgerlingen

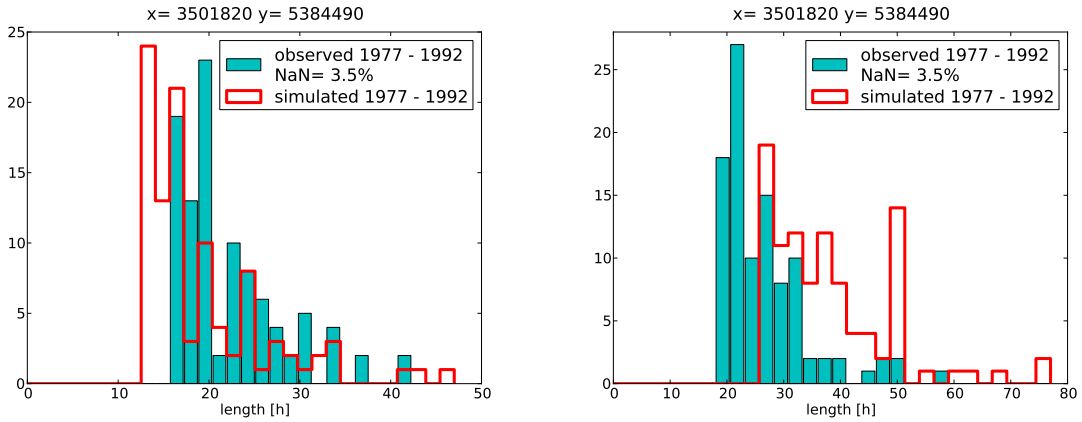


(a) Histogram (b) CDF

Figure 6.19.: Empirical distribution of the wet spell length in hourly resolution. Gaps up to a length of 1 h do not interrupt a wet spell.

overestimated (Fig. 6.20b).

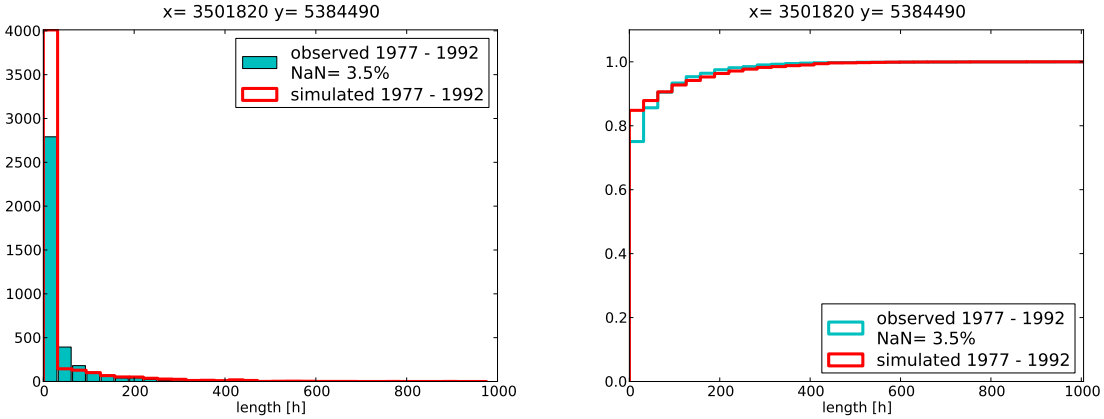
A dry spell is defined as a consecutive sequence of dry hours that is not interrupted by a precipitation event with more than 0.5 mm. The simulated distribution of dry spell lengths is displayed in Fig. 6.21 as histogram, logarithmic histogram and CDF. The simulated distribution of dry spell lengths is more skewed than the observed. The shortest rain breaks of only one to two hours are overestimated, as well as long dry spells of more than seven days. The distribution of the longest 100 dry spells is shifted by about four to five days (Fig. 6.22). Whereas the length of the longest reported dry spell is comparable in the observations and the simulations.



(a) Histogram - 1 h gabs

(b) Histogram - 3 h gabs

Figure 6.20.: Empirical distribution of the 100 longest wet spells in hourly resolution. Gaps up to a length of 1 h (left) / 3 h (right) do not interrupt a wet spell.

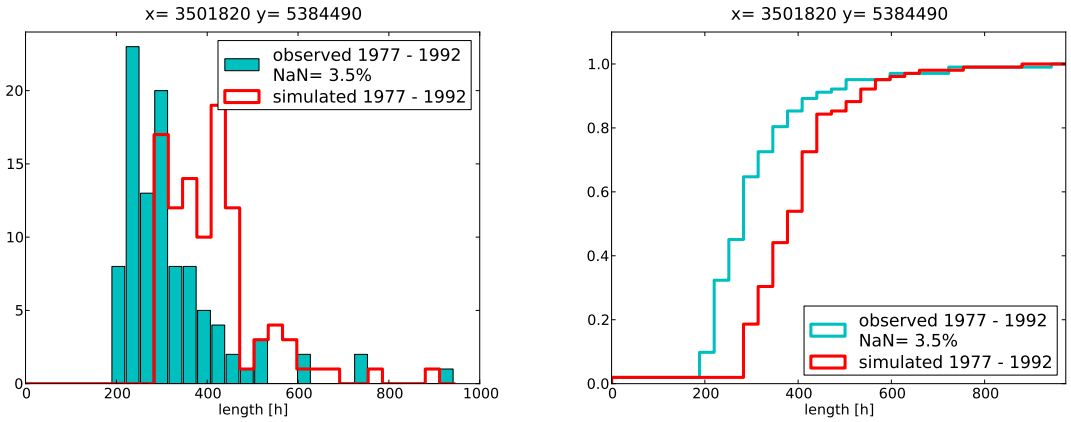


(a) Histogram

(b) CDF

Figure 6.21.: Empirical distribution of the dry spell length in hourly resolution. Rainfall events with less ore equal 0.5 mm cumulated precipitation sum do not interrupt a dry spell.

It has to be noticed that the definitions of wet and dry spells are not complementary. If it was so, the number of wet and dry spells would differ at most by one. In this analysis an overlap is accepted. There are indeed wet hours that do not interrupt a dry spell but are counted within a wet spell.



(a) Histogram (b) CDF

Figure 6.22.: Empirical distribution of the 100 longest dry spells in hourly resolution. Rainfall events with less ore equal 0.5 mm cumulated precipitation sum do not interrupt a dry spell.

6.8. Conclusions

The new generation algorithm is based on the assumption that the dependence between monthly precipitation sum, rainfall probability and the statistical moments of wet hours is constant over the study regions. It is modeled by a Gaussian Copula. The local effects are expressed by the marginal distributions of the parameters. Since all statistics are related to the monthly precipitation sum, the generation is conditioned on observed monthly sums. While the dependence structure shows some minor deviation to a Gaussian copula, the fit of the marginal distributions is very close.

The developed generation algorithm is able to represent the statistical characteristics of precipitation reasonably well. The restriction to one single distribution function, compared to NiedSim, does not lead to reduced variability. On all tested aggregations the extreme values are represented reasonably well. The strongest deviation to observed precipitation is found in the rainfall frequency. It shows that a correct estimation of the rainfall probability is crucial for the performance of the model. Unfortunately, the assessment of observed rainfall frequencies is strongly effected by measurement errors. It depends on a correct assessment of the smallest values, which exhibit the highest relative measurement errors.

The analysis of wet and dry spell lengths proves that the generation is able to reproduce the temporal persistence of observed rainfall time series. Deviations concern an overestimation in the frequency of the shortest rainfall events and at the same time of very short dry spells. However, longer wet and dry spells are modeled correctly. Hence, it is not subject to "overdispersion" unlike most conceptual models (see Section 4.2 in Chapter 4 and Section 5.3). The opposite is the case. The rainfall events cluster more than in reality, which leads to an underestimation of the number of rain day.

7. Spatial Generation

Hydraulic sewage system models require a correct estimation of the total storm water inflow, which is the sum of precipitation volume that falls on sealed areas. An exact estimation would only be possible if the areal precipitation was known. However, time series of rain gauge measurements, as well as synthetic time series generated by NiedSim are point information. In most sewage system models, like for example in KOSIM (ITWH, 2009), the assigned time series is transformed into aerial precipitation of the same intensity. This approximation is only valid for small catchments. Spatial effects, e. g. that only a part of the network is affected by a storm event or a time shift in the peak of rainfall intensity over different subareas, are not considered. Modeling larger areas, it will lead to errors, typically an overestimation of the storm water discharges.

If more than one precipitation time series is required and no rain gauge data is available, the only alternative is the application of several simulated time series. The operational NiedSim generator, however, is not able to consider spatial correlation. Therefore, NiedSim time series of two nearby locations exhibit very similar statistical characteristics (for example the extreme value statistics) but independent temporal sequences. For example, they do not feature simultaneous extreme events. For realistic hydraulic simulations of larger sewage channel networks simultaneous, spatially correlated synthetic rainfall time series are required.

7.1. A High Dimensional Problem

The NiedSim approach cannot be applied without modifications to the generation of several, spatially correlated time series. A permutation based generation reaches its limits if it is used for the generation of simultaneous time series. The reason will be explained in a small example.

If it is supposed that there is a time series with three entries that has to be optimized by ressampling, the number of possible sequences is factorial three:

$$n_{comb} = 3! \tag{7.1}$$

$$= 3 \cdot 2 \cdot 1 = 6 \tag{7.2}$$

One can choose among three values for the first position in the time series, then among the two remaining for the second position. The optimization algorithm has to choose the best of these six different sequences.

If the task is to generate simultaneous time series, the intercorrelations have to be taken into account. Therefore, it is not sufficient to search for the best permutations individually. Instead, the optimum hast to be found in the combination of them all. This increases the number of possible combinations to

$$n_{comb} = (3!)^{n_{stat}} \tag{7.3}$$

$$= 6^{n_{stat}} \tag{7.4}$$

where n_{stat} is the number of stations. Thus, for the simulation of two stations, this leads to 36 possible combinations, for three stations 216, for four stations as much as 1296. If it is assumed that optimization algorithm has to try a constant (very small) portion of all possible combinations to find the optimum sequence, the required CPU time increases by the power of the number of stations.

The factorial is an extremely fast growing function. For the simulation of an hourly time series of one day, assuming no identical values (e.g. zero values), one gets $24! = 6.2 \cdot 10^{22}$ combinations, for the simulation of a one week long time series in hourly data $2.5 \cdot 10^{302}$ different combinations. The simulation algorithm that was set up for the generation of one single time series (Chapter 6) is based on month-wise resampling. The number of possible combinations for one month in hourly values is $720!$ (assuming no identical values). This number is so huge that it cannot be represented in double precision floating point. For the simulation of simultaneous time series, it is raised to the power of n_{stat}. As a result, the number of possible combination becomes unfeasible when the number of simultaneous stations is increasing.

7.2. Generation Principle

It is obvious that the number of possible combinations has to be drastically restricted. This is realized by a two step approach. At first, daily precipitation sums are simulated and optimized by Simulated Annealing. Then the daily time series are disaggregated into 24 one hour values.

1. **Generation of Spatial Correlated Daily Time Series**

   The generation of daily time series follows the same principle as the single site simulation (see Section 6.6 in Chapter 6). Initial time series for the n_{stat} locations are defined by the rainfall probability as well as the average and the standard deviation of precipitation amounts during wet time steps – but now on a 24 h temporal resolution. The statistical moments define the parameters of a Weibull Distribution where the daily precipitation values are drawn from.

   The parameter interdependencies at the n_{stat} target locations are represented by a Gaussian copula. Compared to the simulation at a single location, the correlation matrix is extended by the interdependencies of the generation parameters at the different locations. The question is for example "How does the average daily precipitation at station A depend on the monthly sum at station B?".

The optimization of the time series by simulated Annealing is performed for all stations simultaneously. The objective function consists of the statistical target values at each location and information on the spatial dependence among the stations.

2. **Disaggregation into Spatial Correlated Hourly Values**

   The optimized daily time series is disaggregated by randomly distributing the daily precipitation sum of each day among the 24 hourly values of that day. In the next step, an optimization is performed. Small precipitation increments are exchanged by subtracting them from the value of one time step and adding them to another. To avoid that the optimized daily sums are altered, the chosen time steps are restricted to be from the same day in the year. The objective function of the optimization evaluates the statistical characteristics at each location, as well as the spatial dependence on the hourly time scale.

The statistics that are used to represent the spatial dependencies in the generation are described in Section 7.5. Next, the developed generation algorithm is explained (Section 7.6). Finally, the statistical characteristics of simulated and observed time series are compared in terms of the statistical characteristics (Section 7.7).

7.3. Observed Multivariate Dependencies in Daily Precipitation Data

Like in the single site generation, the simulation will be conditioned on the monthly precipitation sums at the chosen generation locations. Therefore, an estimation of the bivariate Gaussian dependencies of the monthly sum H_{mon}, the rainfall probability $P_{24\text{h}}$, the average rainfall $E\left(H_{wet}\right)$ and the standard deviation $D\left(H_{wet}\right)$ on daily time scale is required. The statistics are estimated by the observed frequency of rain days $p_{24\text{h}}$, as well as the average precipitation sum $\bar{H}_{wet}$ and the standard deviation s_{wet} of wet day precipitation. In the coarser daily resolution one can benefit from the broader data base of the 575 daily stations (Section 3.3).

The interdependencies are analyzed for monthly censored data. The values are ranked (Eq. (6.46)) and transformed into the $[0,1]$-space of uniform distributions by the non-exceedance frequencies (Eq. (6.54)). Then the bivariate copula densities are regarded and compared to the Gaussian copula.

Table 7.1 gives an overview of the rank correlation ρ according to Eq. (6.46) and the symmetry measure sym according to Eq. (6.62). For all parameter combinations the monthly differences in rank correlations and symmetry are smaller than in the hourly data. It means that daily precipitation values are less influenced by seasonal effects than hourly values. Nevertheless, the rank correlations ρ are generally lower in the summer months with convective precipitation, but not all parameter combinations follow a clear seasonal cycle.

Compared to hourly resolution, the correlation between monthly sum H_{mon} and average precipitation $\bar{H}_{wet}$ increases. it is an effect of the temporal resolution. The monthly sum

Table 7.1.: Rank correlation and symmetry of the empirical copula from monthly data for bivariate combinations of all parameters; data in daily resolution

Month	$H_{mon} - p_{24\text{h}}$		$H_{mon} - \bar{H}_{wet}$		$H_{mon} - s_{wet}$	
	ρ	Sym	ρ	Sym	ρ	Sym
January	0.71	-0.00562	0.87	0.00101	0.83	0.00027
February	0.70	-0.00498	0.84	0.00189	0.77	0.00173
March	0.69	-0.00039	0.88	0.00072	0.82	-0.00007
April	0.73	0.00410	0.80	0.00022	0.72	-0.00083
May	0.71	-0.00331	0.85	-0.00012	0.76	-0.00064
June	0.61	-0.00408	0.78	-0.00020	0.67	-0.00115
July	0.59	-0.00431	0.78	-0.00081	0.69	-0.00186
August	0.69	-0.00215	0.84	-0.00011	0.74	-0.00162
September	0.71	-0.00058	0.72	-0.00118	0.68	-0.00180
October	0.67	-0.00033	0.81	-0.00179	0.79	-0.00198
November	0.60	-0.00268	0.84	0.00006	0.76	-0.00080
December	0.65	-0.00128	0.87	0.00105	0.82	0.00062

Month	$p_{24\text{h}} - \bar{H}_{wet}$		$p_{24\text{h}} - s_{wet}$		$\bar{H}_{wet} - s_{wet}$	
	ρ	Sym	ρ	Sym	ρ	Sym
January	0.33	-0.00386	0.37	-0.00501	0.90	0.00030
February	0.26	-0.00182	0.25	-0.00216	0.90	0.00062
March	0.31	0.00133	0.32	-0.00017	0.90	0.00025
April	0.22	0.00217	0.21	0.00055	0.88	0.00011
May	0.26	-0.00328	0.27	-0.00248	0.86	-0.00103
June	0.04	-0.00368	0.06	-0.00341	0.84	-0.00102
July	0.02	-0.00543	0.04	-0.00621	0.85	-0.00064
August	0.23	-0.00241	0.26	-0.00418	0.82	-0.00043
September	0.07	-0.00132	0.16	-0.00291	0.84	0.00039
October	0.17	-0.00265	0.27	-0.00280	0.88	-0.00041
November	0.13	-0.00184	0.14	-0.00247	0.88	0.00044
December	0.24	0.00041	0.30	-0.00026	0.90	-0.00008

consists of only 28 to 31 daily values and thus the relation exhibits a lower freedom than in hourly resolution with 672 to 744 values per month. On the hourly scale precipitation frequency and the statistical moments were almost uncorrelated. On daily scale there is a significant positive correlation, except for May and June.

The respective copula densities are presented in Fig. B.1 to Fig. B.6 in the appendix to this chapter. Each figure features the plots of the two most different months in terms of the respective parameter combination. Generally, the deviations to bivariate Gaussian copulas are low. The combinations $p_{24\text{h}}$ and $\bar{H}_{wet}$, as well as $p_{24\text{h}}$ and s_{wet} , however, are asymmetric with a high density in the lower left corner (Fig. B.4 and Fig. B.5): In a month with only one or two wet days, the variations in the daily rainfall amount are very limited.

Table 7.2.: Correlation in the Standard Normal transformed space of monthly precipitation sum, daily rainfall probability as well as monthly average and standard deviation of wet day precipitation amounts applied in the simulation of daily precipitation amounts.

December to February				
	H_{mon}	P_{24h}	$E(H_{wet})$	$D(H_{wet})$
H_{mon}	1.000	0.710	0.875	0.831
P_{24h}	0.710	1.000	0.310	0.343
$E(H_{wet})$	0.875	0.310	1.000	0.911
$D(H_{wet})$	0.831	0.343	0.911	1.000

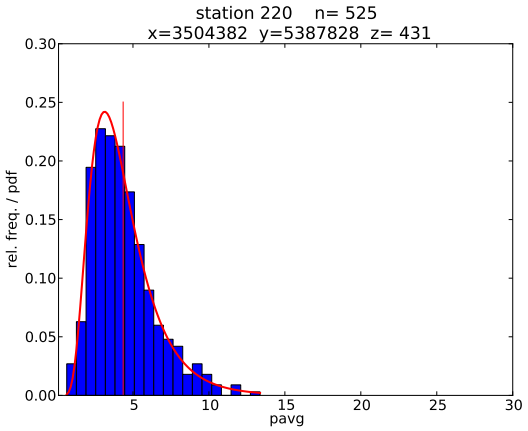
March to May				
	H_{mon}	P_{24h}	$E(H_{wet})$	$D(H_{wet})$
H_{mon}	1.000	0.696	0.867	0.797
P_{24h}	0.696	1.000	0.255	0.262
$E(H_{wet})$	0.867	0.255	1.000	0.900
$D(H_{wet})$	0.797	0.262	0.900	1.000

June to August				
	H_{mon}	P_{24h}	$E(H_{wet})$	$D(H_{wet})$
H_{mon}	1.000	0.669	0.813	0.716
P_{24h}	0.669	1.000	0.132	0.154
$E(H_{wet})$	0.813	0.132	1.000	0.852
$D(H_{wet})$	0.716	0.154	0.852	1.000

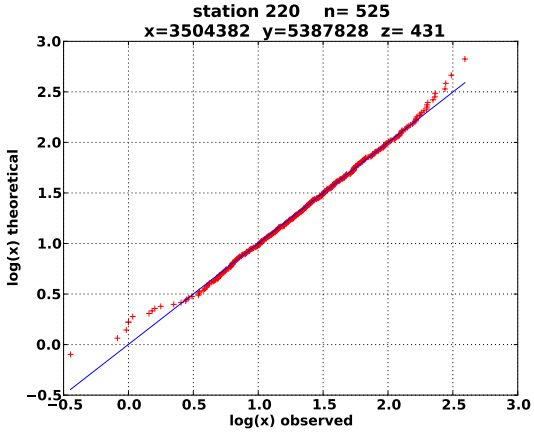
September to November				
	H_{mon}	P_{24h}	$E(H_{wet})$	$D(H_{wet})$
H_{mon}	1.000	0.672	0.796	0.761
P_{24h}	0.672	1.000	0.103	0.183
$E(H_{wet})$	0.796	0.103	1.000	0.884
$D(H_{wet})$	0.761	0.183	0.884	1.000

As in the hourly data, the ranks in all parameters are positively correlated to the ranks in the monthly precipitation sum. Therefore, it is possible to condition the generation parameters on the monthly sum. Table 7.2 displays the bivariate correlation coefficients in the Standard Normal transformed space that are used to model the Gaussian copula in the rainfall generation. The deviations to Gaussian copula density in the parameter combination p_{24h} versus $\bar{H}_{wet}$ and p_{24h} versus s_{wet} are of minor importance. If the rainfall probability is anyhow almost zero, the expectation of the average and standard deviation used for the simulation have a very low influence on the result.

7.4. Marginal Distributions of Daily Precipitation Data

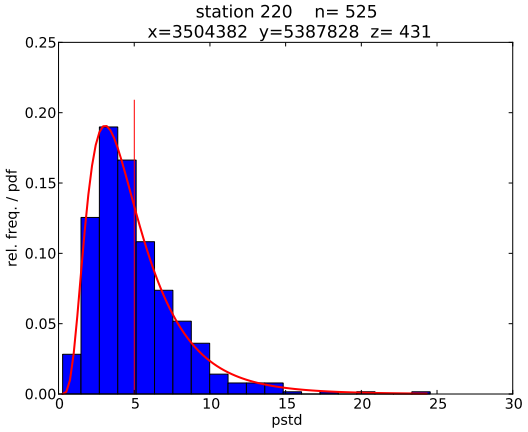


(a) Histogram - Log-Normal PDF

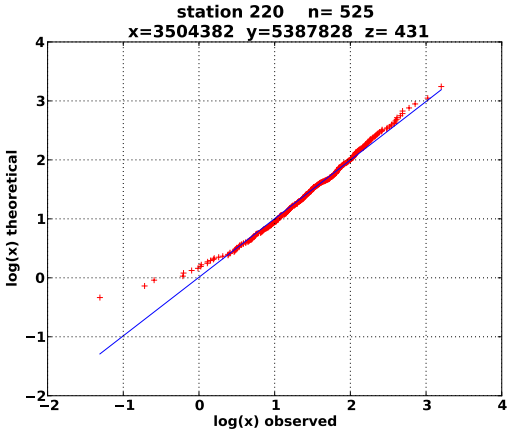


(b) Log-Normal Distribution

Figure 7.1.: Observed and theoretical Log-Normal Distribution in the linearized space of the average wet day precipitation amount in Schönaich.



(a) Histogram - Log-Normal PDF

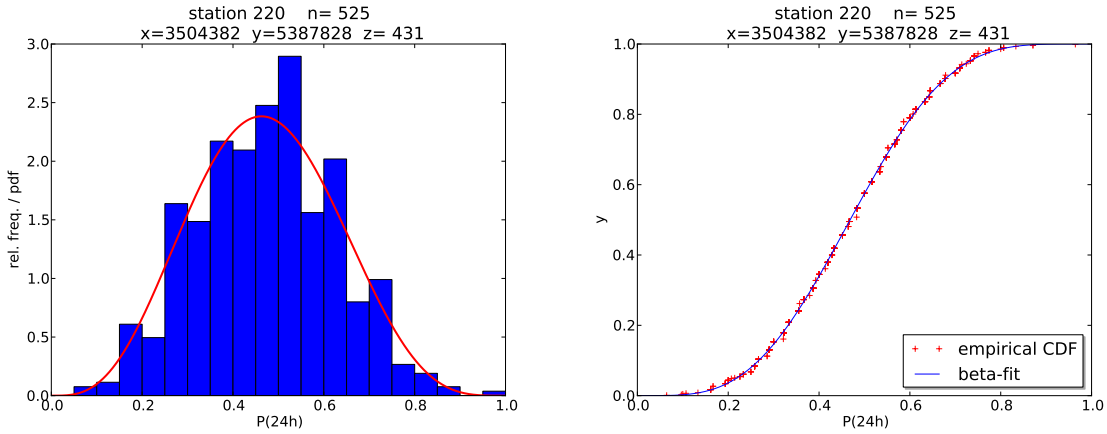


(b) Log-Normal Distribution

Figure 7.2.: Observed and theoretical Log-Normal Distribution in the linearized space of the standard deviation of the wet day precipitation amount in Schönaich.

The marginal distributions of the new statistical parameters $P_{24\text{h}}$, $E\left(H_{wet}\right)$ and $D\left(H_{wet}\right)$ have to be estimated. They are used for the transformation into the copuloa space and for the back transformation into real world values.

Like in hourly resolution the average and standard deviation of wet day precipitation show right skewed empirical distributions at all 575 stations. Therefore, the same theoretical distributions functions as in Section 6.5.2 are tested: Weibull Distribution, Log-Normal Distribution and Lomax Distribution (Section 6.3). It is found that the Log-Normal distribution



(a) Histogram

(b) Log-Normal Distribution

Figure 7.3.: Observed and theoretical Beta Distribution of the rainfall frequency in daily resolution in Schönaich.

is suited best for the description of the average wet day precipitation as well as for the standard deviation of wet day precipitation. The rainfall probability is described by a Beta-distribution (Section 6.3.4).

Fig. 7.1, Fig. 7.2 and Fig. 7.3 show the empirical and theoretical distribution in Schönaich as an example. This station is situated four kilometers Northeast of Holzgerlingen, so that the results of the hourly and daily fit can be compared.

The histograms and CDFs of the seasonal fit to all daily precipitation stations can be found in the appendix to this chapter (Section B.2). The fit for the average and standard deviation of wet day precipitation is closer to the observed than the fit for the hourly data, especially in the highest values (compare Fig. A.4 with Fig. B.8 and Fig. A.6 with Fig. B.10). Deviations only concern the lowest values.

The fit of the Beta distribution, on the other hand, is better for the hourly data. This, however, is a sampling effect due to the limited number of possible values in the observed frequencies. E. g. for a month with 31 days and a rainfall probability of $P_{24\mathrm{h}} = 0.95$ there is a probability of 20% that all days are wet and that the observed rainfall frequency $p_{24\mathrm{h}}$ is one.

Table 7.3 resumes the parameters of the theoretical distribution function used to describe the marginal distributions in the generation of simultaneous daily time series. The distribution of the monthly sum is unchanged as it does not depend on the temporal resolution of the observed data.

Table 7.3.: Distribution functions used in the parameter estimation for the generation of simultaneous daily rainfall time series

Parameter	Distribution Function	Parameter	December -February		March-May	
H_{mon}	Weibull Distribution	λ, κ	81.366	1.456	88.616	1.767
P_{24h}	Beta-Distribution	α, β	3.700	3.418	4.807	4.759
$E(H_{wet})$	Log-Normal Distribution	μ, σ	1.373	0.573	1.486	0.483
$D(H_{wet})$	Log-Normal Distribution	μ, σ	1.448	0.628	1.541	0.563
Parameter	Distribution Function	Parameter	June-August		September -November	
H_{mon}	Weibull Distribution	λ, κ	106.643	2.263	83.733	1.665
P_{24h}	Beta-Distribution	α, β	5.525	6.440	3.585	4.186
$E(H_{wet})$	Log-Normal Distribution	μ, σ	1.795	0.413	1.577	0.513
$D(H_{wet})$	Log-Normal Distribution	μ, σ	1.866	0.496	1.657	0.560

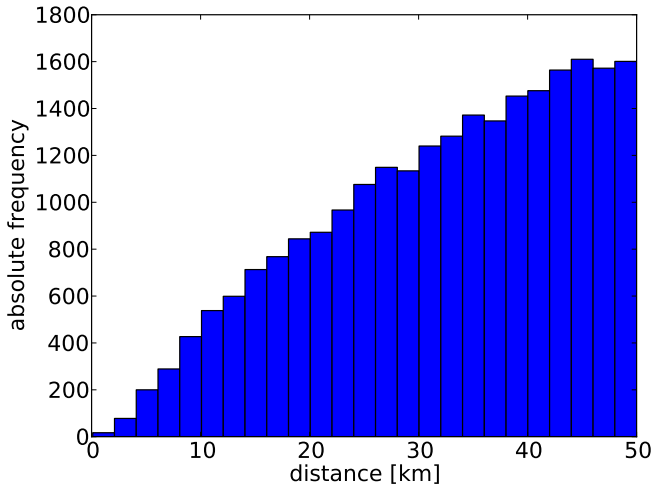


Figure 7.4.: Absolute frequency of the distances between 575 daily precipitation stations

7.5. Representation of Spatial Dependencies

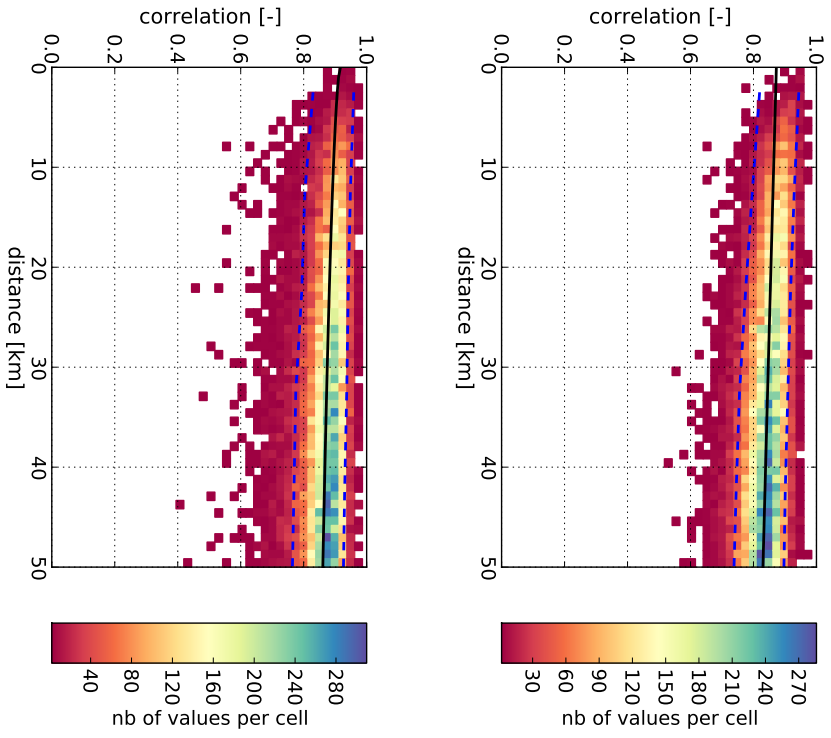
7.5.1. Observed Interdependencies of Statistical Parameters at Different Measurement Locations

Exploiting the results of the last section, it is possible to draw interdependent values of daily rainfall probability P_{24h}, of expected average wet day precipitation $E(H_{wet})$ and of expected standard deviation of wet day precipitation $D(H_{wet})$ that are conditioned on the monthly sum H_{mon}. With these information, it is possible to set up a generation scheme for daily precipitation time series analogous to the scheme for hourly precipitation described in the

last chapter. For the generation of simultaneous time series, however, the information is not sufficient yet. The missing link is, how the statistical parameters at different stations are influencing each other.

The four statistical parameters H_{mon}, P_{24h}, $E\left(H_{wet}\right)$ and $D\left(H_{wet}\right)$ at different locations are interdependent. If for example the rainfall frequency is high at one station, it will be high at any neighboring station too. Besides, the information of a high rain day frequency does not only condition the rainfall frequency at other station. Due to the interdependencies of the statistics, the distribution of the average or standard deviation at the surrounding stations will react too.

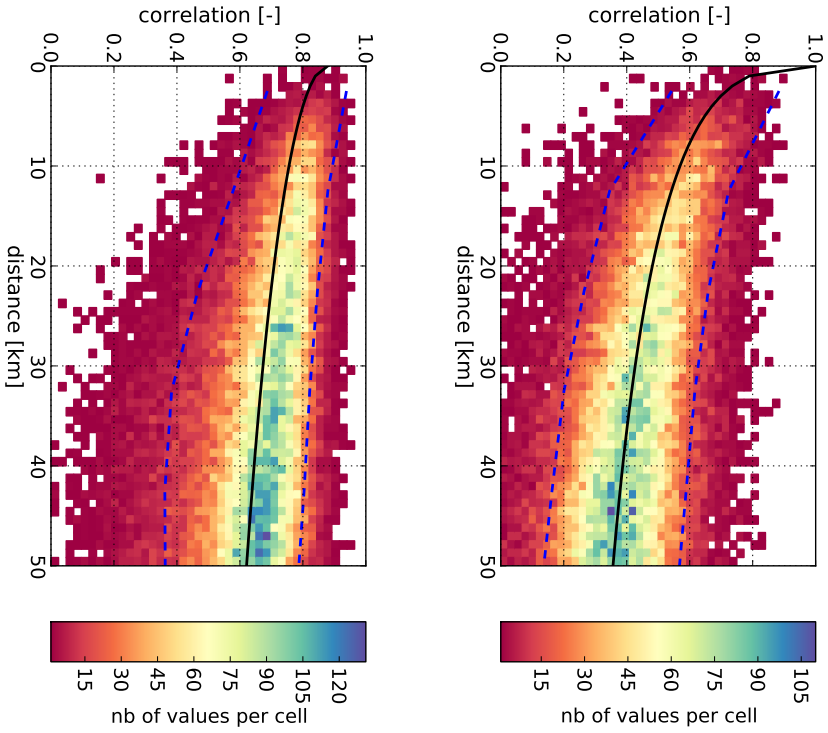
The strength of the rank correlation between two stations depends on the parameter combination considered, the distance between the two stations and the season. Fig. 7.5 and Fig. 7.6 display the rank correlation of two of parameter combinations as a function of the distance. The graphs of the remaining parameter combinations can be found in the appendix to this chapter (Fig. B.13 to Fig. B.20). Each station combination in the data set of daily stations attributes one correlation value to each of the diagrams. In total there are $575 \times 575 = 330625$



(a) December to February

(b) June to August

Figure 7.5.: Observed rank correlation between the rain day frequency at two stations depending on the separation distance; for all station pairs closer than $50\,\text{km}$. Black line: best exponential regression fit; blue dashed lines: empirical 90% confidence interval.



(a) December to February (b) June to August

Figure 7.6.: Observed rank correlation between the average wet day precipitation at one station and the standard deviation of wet day precipitation at another station depending on the separation distance for all station pairs closer than 50 km. Black line: best exponential regression fit; blue dashed lines: empirical 90% confidence interval.

combinations. All station pairs with less than 12 overlapping values in the respective season are excluded, which leads to about 290000 remaining pairs. In respect to the generation of simulation time series the focus is on small distances. For this reason, the following analysis are restricted to all station pairs of less than 50 km distance.

For a better readability, the plots are not presented as scatter plots but as rasters of point densities. The black line in Fig. 7.5 and Fig. 7.6 indicates an exponential regression function that gives the best estimation of the correlation depending on the distance:

$$\hat{\rho}\left(d\right) = \rho_0 \cdot \exp\left(-\left(\frac{d}{d_0}\right)^{s_0}\right) \tag{7.5}$$

where
- $\hat{\rho}\left(d\right)$ best estimation for the rank correlation at distance d
- ρ_0 rank correlation between parameters at zero distance
- d_0 reference distance
- s_0 scale parameter

The parameters ρ_0, d_0 and s_0 of the regression are estimated based on numerical minimization of the Mean Square Error between the observed and estimated rank correlations:

$$E = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\rho_{ij} - \hat{\rho}\left(d_{ij}\right)\right)^2 \rightarrow \min! \tag{7.6}$$

where ρ_{ij} observed rank correlation between the station pair i, j
$\hat{\rho}\left(d_{ij}\right)$ estimated rank correlation for the station pair i, j
d_{ij} distance between the stations i and j

As a robust measure of the spread, the 90% confidence interval is indicated by two dashed blue lines. It is calculated empirically for classes of $5\,$km distance. The total range, which is most visible in the scatter plots, is highly affected by the number of available data pairs.

Fig. 7.5 and Fig. 7.6 are chosen as examples because the graphs illustrate the range of different spatial behavior in the parameter interdependence. The correlation of the rainfall frequency $p_{24\text{h}}$ at two stations can be seen as a regional characteristic. It does not depend on the distance between the station pair. In winter months the best estimation for the rank correlation at $5\,$km distance is $\hat{\rho} = 0.89$, at $50\,$km distance it is still $\hat{\rho} = 0.86$. In the summer months, the correlation is only slightly lower. $\hat{\rho} = 0.86$ at $5\,$km and $\hat{\rho} = 0.83$ at $50\,$km distance. Seasonal effects can hardly be seen.

The monthly sum (Fig. B.13), as well as the combination of monthly sum and daily rainfall frequency (Fig. B.16) show similar characteristics. The rank correlation of p_{24} versus $\bar{H}_{wet}$ and of p_{24} versus $\bar{s}_{wet}$ do not depend on the distance either. The correlations, however, are much lower and the spread is higher (Fig. B.19 and Fig. B.20). In any distance uncorrelated stations can be found, as well as station pairs with $\rho > 0.7$

The rank correlation of $\bar{H}_{wet}$ versus s_{wet}, on the other hand, shows a strong dependence of the correlation on the distance as well as a strong seasonal difference in correlation (Fig. 7.6). The average and the standard deviations are both statistics that are related to local effects and therefore the dependence weakens rapidly with growing station distance. In winter the best estimation for the correlation is $\hat{\rho} = 0.79$ at $5\,$km distance, at $50\,$km distance it is $\hat{\rho} = 0.64$. In summer it drops from $\hat{\rho} = 0.66$ to $\hat{\rho} = 0.35$. Similar characteristics are found in the rank correlations of the average and the standard deviation of wet day precipitation (Fig. B.14 and Fig. B.15), as well as in the combination of monthly sum versus average (Fig. B.17) and monthly sum versus standard deviation (Fig. B.18).

In all parameter combinations, due to the presence of small scale convective events, the rank correlation are lower in summer.

7.5.1.1. Analysis of Anisotropy

For the configuration of several target locations of the time series generation the rank correlations ρ between the statistical parameters are unknown. The true correlation between

two locations is substituted by the regression estimation $\hat{rho}(d_{ij})$. The spread of the correlation around the regression functions is discarded. For some of the relations, this is not a problem since the spread is low. These are principally the relation of the monthly sum and the rain day probability to themselves as well as the combination of these two parameters. The rank correlations involving the average or standard deviation of wet day precipitation on the other hand show high spread, the deviation to the regression function can go up to $\Delta\rho = 0.2$ or more.

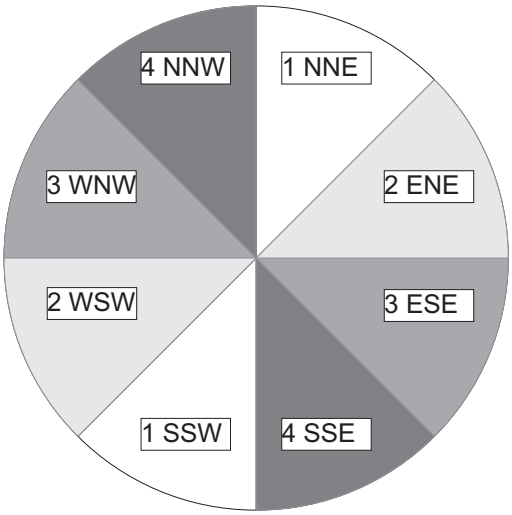
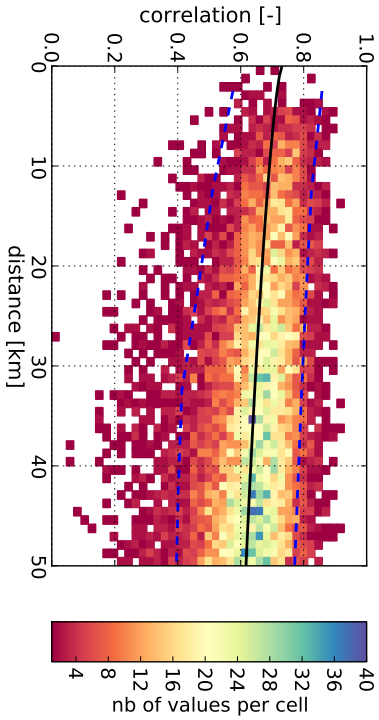


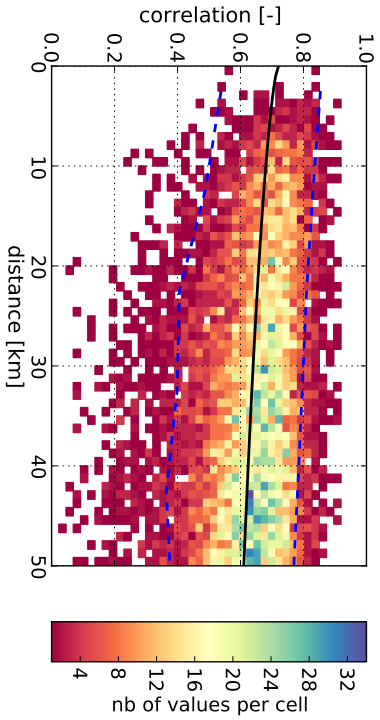
Figure 7.7.: Sketch of the sector division for rank correlation analysis

It is checked if the spread is due to anisotropy in the relation between distance and correlation. For this analysis the station pairs are classified into 4 sectors according to their geographic orientation (Fig. 7.7). The results are presented for the parameter combination of H_{mon} and s_{wet} as an example. It is chosen because this combination exhibits the widest spread in the correlation over distance relation. The results of other parameter combinations lead to the same conclusions.

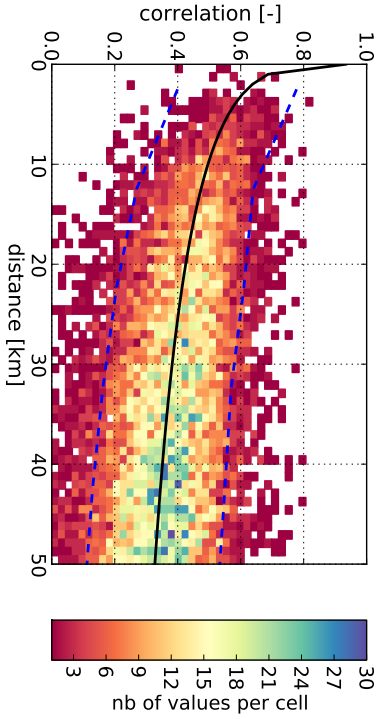
The main atmospheric flow direction in the study area is from west-southwest to east-northeast (Section 2.1). It was found in a former study that two stations a more related, regarding absolute precipitation values, if the vector joining the two stations is parallel to the main atmospheric flow direction than if it is perpendicular to the flow direction (Brommundt, 2008). In the presence of anisotropy a classification of the station pairs according their to geographic orientation should alter the shape of the point clouds and the regression functions. According to the work of Brommundt (2008) the correlations should be highest if the vector linking the two stations falls in sector 2 of Fig. 7.7 and lowest for sector 4. However, the results of the classification (Fig. 7.8) do not support the idea. The graphs for the four sectors are hardly distinguishable. The regression line for the best estimation in rank correlation $\hat{\rho}$ takes the same course as in Fig. B.18 for any of the sectors. Hence, the spread in the correlations does not come from an anisotropic dependence. The confidence intervals for the sector classified data are by no means smaller than for the whole data set with all station combinations. Hence, the anisotropy found in the correlation of absolute daily



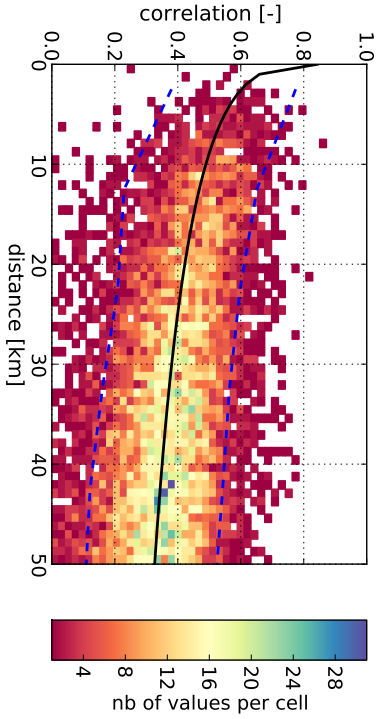
(a) December to February – sector 2



(b) December to February – sector 4



(c) June to August sector 2



(d) June to August sector 4

Figure 7.8.: Observed rank correlation during winter months between the monthly precipitation sum at one station and the standard deviation of wet day precipitation at another station depending on the separation distance and the sector

values (Brommundt, 2008) cannot be identified in the rank correlation of deduced statistical parameters.

7.5.1.2. Spatial Rank Correlations Used for the Generation of Simultaneous Daily Time Series

The rank correlation between the parameters at the locations chosen for simultaneous time series generation will be estimated by the respective regression functions according to Eq. (7.5). The parameters of this function for each combination of H_{mon}, P_{24h}, $E(H_{wet})$ and $D(H_{wet})$ can be found in Table 7.4 and Table 7.5.

Table 7.4.: Parameters of the regression functions according to Eq. (7.5) for the estimation of parameter rank correlation over distance - same parameter at both stations

Parameter at		December to February			March to May		
Station 1	Station 2	r_0	d_0	s_0	r_0	d_0	s_0
H_{mon}	H_{mon}	0.956	1503.4	0.751	0.982	753.3	0.614
P_{24h}	P_{24h}	0.917	5724.3	0.584	0.912	1613.0	0.770
$E(H_{wet})$	$E(H_{wet})$	0.937	433.3	0.524	1.000	277.2	0.468
$D(H_{wet})$	$D(H_{wet})$	1.000	262.1	0.490	1.000	185.2	0.489
Parameter at		June to August			September to November		
Station 1	Station 2	r_0	d_0	s_0	r_0	d_0	s_0
H_{mon}	H_{mon}	1.000	337.4	0.483	0.982	918.4	0.644
P_{24h}	P_{24h}	0.873	990.2	1.003	0.908	2527.9	0.672
$E(H_{wet})$	$E(H_{wet})$	1.000	64.9	0.454	0.972	261.9	0.488
$D(H_{wet})$	$D(H_{wet})$	1.000	50.1	0.421	1.000	175.4	0.505

Table 7.5.: Parameters of the regression functions according to Eq. (7.5) for the estimation of parameter rank correlation over distance - combination of different parameters

Parameter at		December to February			March to May		
Station 1	Station 2	r_0	d_0	s_0	r_0	d_0	s_0
H_{mon}	P_{24h}	0.766	1042.7	1.521	0.695	1330.1	0.830
H_{mon}	$E\,(H_{wet})$	0.776	695.2	0.772	0.876	421.1	0.487
H_{mon}	$D\,(H_{wet})$	0.735	729.7	0.637	0.784	395.2	0.484
P_{24h}	$E\,(H_{wet})$	0.420	528.9	2.372	0.339	18366.0	0.264
P_{24h}	$D\,(H_{wet})$	0.367	642.6	1.986	0.406	176400.0	0.064
$E\,(H_{wet})$	$D\,(H_{wet})$	0.879	384.7	0.519	0.913	264.2	0.483
Parameter at		June to August			September to November		
Station 1	Station 2	r_0	d_0	s_0	r_0	d_0	s_0
H_{mon}	P_{24h}	0.661	1451.2	0.713	0.718	2307.6	0.767
H_{mon}	$E\,(H_{wet})$	1.000	60.1	0.355	0.781	430.9	0.496
H_{mon}	$D\,(H_{wet})$	0.907	47.1	0.324	0.676	374.8	0.675
P_{24h}	$E\,(H_{wet})$	0.350	11842.3	0.074	0.211	5743.0	10.536
P_{24h}	$D\,(H_{wet})$	0.166	1091.2	0.907	0.222	6334.9	0.599
$E\,(H_{wet})$	$D\,(H_{wet})$	1.000	46.0	0.376	0.877	249.3	0.513

7.5.2. Spatial Correlation According to Atmospheric Circulation Patterns

The effect of spatial dependencies in precipitation depends on the regarded temporal scale (Section 4.1). On monthly scale, the dependence is described by the parameter interdependencies derived in the last section. The investigation of the spatial dependence on subdaily temporal scale is based on weather type classification.

The dependence of two rain gauges is affected by the current atmospheric flow direction. Brommundt (2008) uses radar data to define principal flow directions depending on the Circulation Pattern (CP) as a measurement of the overall atmospheric condition. The correlation of two 1 h precipitation time series is different if the two stations are located parallel to the flow direction or orthogonal to the flow direction (Brommundt, 2008). In this section it is analyzed how the CP effects the spatial correlations in all possible orientations.

7.5.2.1. Definition of the Circulation Pattern Classification System

The Circulation Pattern used in this examination were taken from Bárdossy (2010). They result from an automated fuzzy rule based weather type classification system. The setup of the classification is described in Bárdossy and Filiz (2005). Data basis of the classification are mean sea level pressure fields (MSLP) from NCEP/NCAR reanalysis that provide an archive of homogenized and gridded atmospheric variables (Kistler et al., 2001). The temporal resolution is $2.5°$ by $2.5°$. The classification is based on normalized anomalies. At each grid point, first, the mean value for this point is subtracted. Then, the difference is divided by the standard deviation of the grid point resulting in a dimensionless number. Positive numbers indicate high pressure, negative numbers low pressure zones.

Based on the pressure anomalies, five fuzzy states are defined:

1. large positive anomalies
2. medium positive anomalies
3. medium negative anomalies
4. large negative anomalies
5. arbitrary

The fuzzy rule set defining the k CPs specifies the required fuzzy state for each of the grid points. It can be described by a matrix:

$$\mathbf{V} = v\,(i,j)\,;\; i = 1,\ldots n_{pts};\; j = 1,\ldots k \tag{7.7}$$

The definition of one CP class j consists of the state index $v = [1,5]$ to each grid point i of the n_{pts} grid point in the NCEP/NCAR anomalie field of the Europe and the northern Atlantic Ocean. The state "arbitrary" is assigned to all points that are indifferent to the current weather type definition. The principle of the weather classification is explained

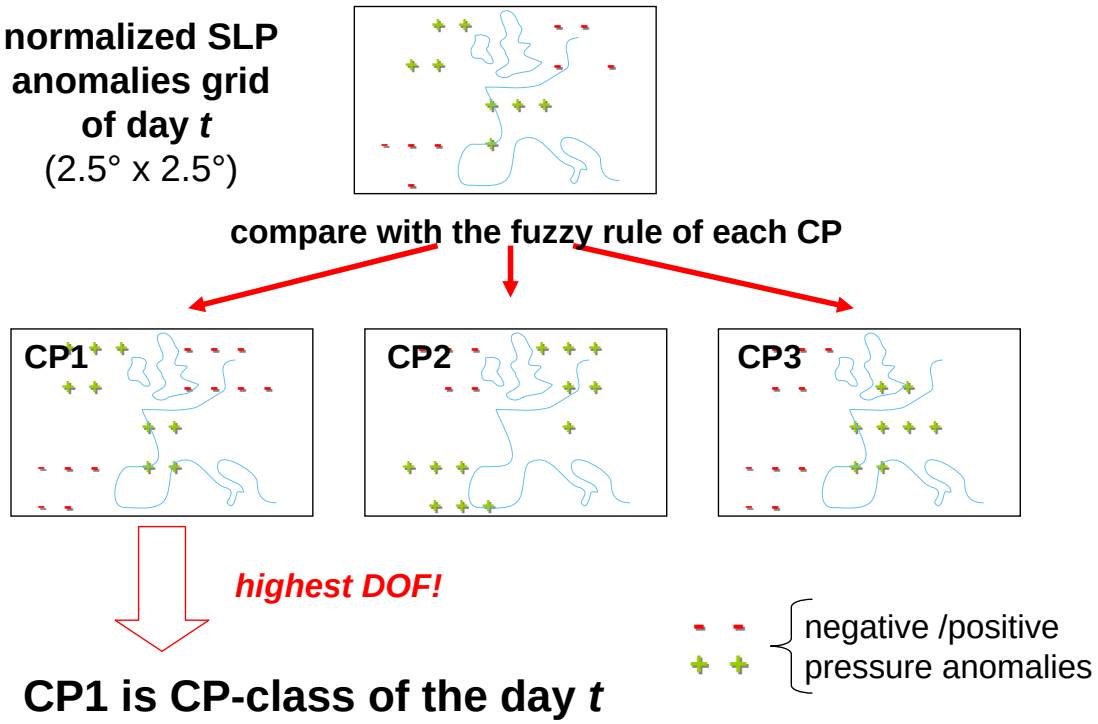


Figure 7.9.: Principle of the fuzzy rule based CP classification

graphically by the sketch in Fig. 7.9. For each day, the anomaly field is compared with the CP-defining fuzzy rule set. The CP with the highest resemblance, resulting in the highest degree of fulfillment of the fuzzy rule, is chosen as the CP of the day.

The CP definition, which is the choice of the states v to each grid point i, is done by an automated system based on Simulated Annealing (for a description of the Annealing algorithm, see Section 6.6.2). The Annealing algorithm arbitrarily chooses grid points and assigns fuzzy anomaly states. If the resulting classification is of any sense, the resulting CPs should be different in their precipitation characteristics. This is measured by the objective function O of the Annealing algorithm. The optimization is performed by maximizing the objective function O.

The classification used in this study was set up by Bárdossy (2010). O is defined by the fraction of stations that exceed a certain precipitation threshold:

$$O(\eta) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(p_c\left(CP\left(t\right)\right)_i - \bar{p}_{c_i}\right)^2} \tag{7.8}$$

η	spatial coverage taken as a threshold, expressed as the number of observing stations with non-zero rainfall; $\eta \in [0,1]$
$p_c\left(CP\left(t\right)\right)_i$	relative frequency of the coverage exceeding the threshold η in the neighborhood of station i at days with a given CP

$\bar{p}_{c^i}$	relative frequency of coverage threshold exceedance for all days without classification
T	time period of the calibration.

See Bárdossy (2010) for further explanations.
To identify wet and dry situations equally well, the objective function O for the classification considers two thresholds:

$$O = O(\eta = 0.0) + O(\eta = 0.999) \rightarrow \max! \tag{7.9}$$

(Bárdossy, 2010)

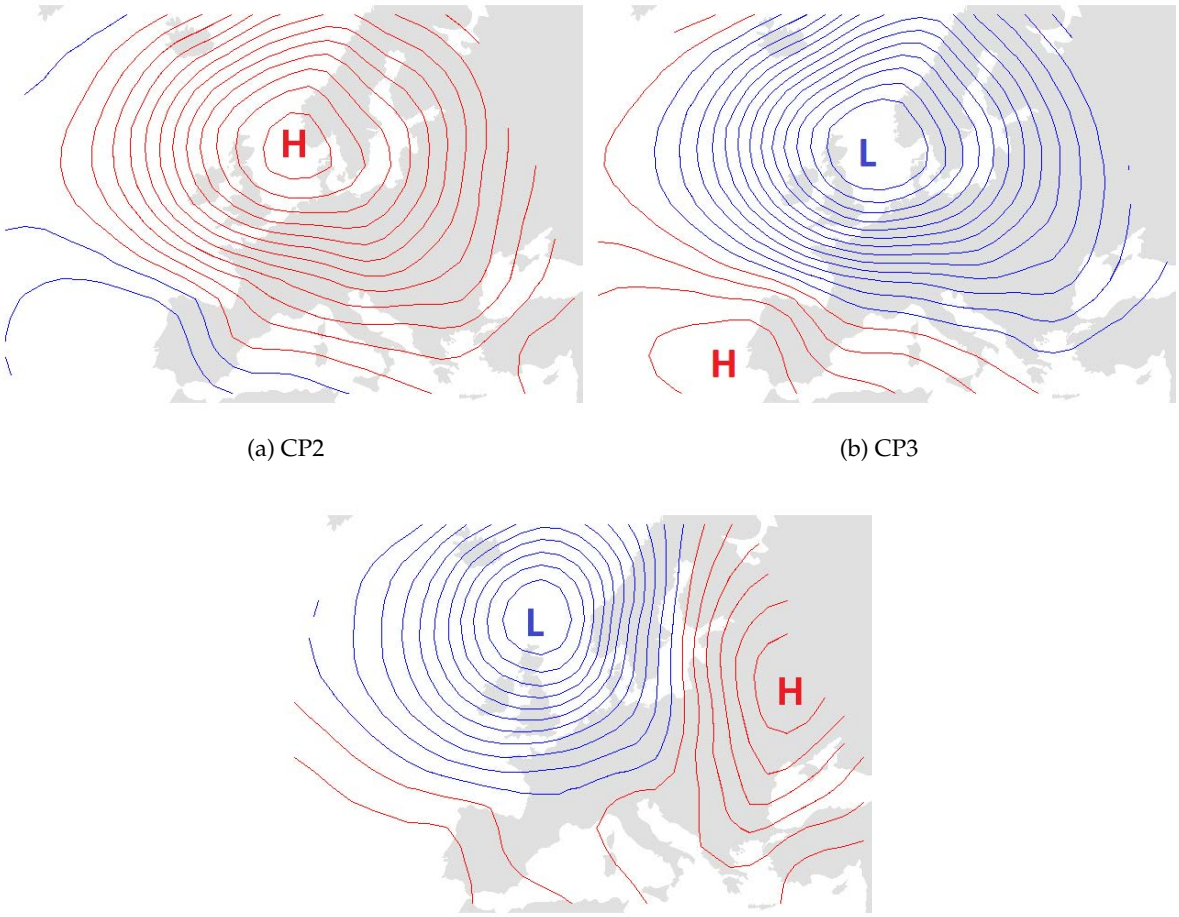
The classification system optimizes the CP definitions but not *the number* of different CPs. It has to be chosen by the user, e. g. by fitting classification systems with different numbers of CPs and choosing the best by a performance measure. For the cited objective function Bárdossy (2010) estimated $17 + 1$ as the optimal CP number. "$+1$" stands for "CP99" which is the CP class for all days that cannot be assigned to any other class.

7.5.2.2. CP Dependent Precipitation Statistics

The objective function O in the CP class definition evaluates the hydrological response and not the weather. However, similar rainfall characteristics are the result of similar atmospheric conditions. If the mean sea level pressure anomalies of all days belonging to the same CP are averaged, the result represents common weather situations. Fig. 7.10 shows three examples. CP2 is the driest among the 18 CPs. The rain day frequency averaged over 575 daily stations is about 12% and the average daily precipitation 0.3 mm. The average pressure anomalies of all days belonging to CP2 explain why. CP2 exhibits anticyclonic conditions over Baden-Württemberg. In average, the center of the high pressure zone is over Denmark. CP3 is exactly the inverse situation and therefore, the wettest CP in the classification system. Southwest Germany is under the influence of a cyclonic depression whose center is close to Baden-Württemberg, over the North sea. The average rain day frequency is 85% and the average amount about 7 mm per day. CP15 is somehow inbetween. It is a typical situation that sometimes is refered to as "Shottland Low". The distance of the center of the depression to the study region is higher than for CP3, and the cyclonic depression is pushed on a track in northern direction by the high pressure zone over eastern Europe. Over Baden-Württemberg the approaching warm and cold fronts are generally less intensive due to the higher distance to the center of the low. Therefore, the rain day frequency of 55% and the average daily precipitation of 5.2 mm are lower than for CP3. Table 7.6 to Table 7.8 list the basic daily statistics of all CPs in the classification.

7.5.2.3. Analysis of Spatial Correlation

Considering the gradients resulting from the average pressure anomalies and the effect of Coriolis force, the main flow directions can be identified. During CP3 precipitation arrives



(a) CP2

(b) CP3

(c) CP15

Figure 7.10.: Average pressure anomalies of all days belonging to the same CP, lines are 1 hPa isobars; red – positive anomalies; blue – negative anomalies

by western flux and during CP15 by southwestern flux. The anticyclonic conditions of CP2 lead to a calm situation with low atmospheric exchange. An analysis of spatial correlation reveals anisotropies caused by the flow direction.

The analysis is based on 575 daily stations and 137 stations sharing the same measurement period from 1997 to 2003 (Section 3.3). The stations are paired and the correlations Eq. (6.41) between the time series of the station pairs are calculated. A pair of daily stations is considered if it shares at least 30 daily values in the respective CP class. Hourly station pairs are considered if they share at least 100 hourly values.

The resulting correlations are examined in perspective of the spatial configuration of the station pairs. Fig. 7.11 to Fig. 7.13 display the correlation in daily precipitation time series according to the distance vector (x kilometers to the east and y kilometers to the north) separating the two stations. Each raster cell in these plots represents the average correlation of all station pairs that have approximately the same distance vector. The absolute location

Table 7.6.: Relative frequency, rainfall frequency, average and standard deviation of wet day precipitation for all anticyclonic dry CPs

CP	October to March				April to September			
	rel. Freq.	p_{24h}	$\bar{H}_{24h}$	s_{24h}	rel. Freq.	p_{24h}	$\bar{H}_{24h}$	s_{24h}
1	0.0454	0.2559	0.8736	2.7229	0.0581	0.3132	1.6683	4.5967
2	0.0617	0.1141	0.173	0.9379	0.0608	0.1328	0.5507	2.7702
7	0.052	0.2704	0.8775	2.9705	0.0688	0.2953	1.4761	4.4296
8	0.0511	0.1944	0.5137	1.9051	0.0445	0.1932	0.9964	3.8954
10	0.0315	0.2618	0.6121	2.0257	0.0396	0.2149	0.8654	3.1074
11	0.0822	0.2022	0.4929	1.8411	0.0639	0.2042	1.2645	4.7771
12	0.0753	0.2348	0.7793	2.8574	0.0811	0.2565	1.3752	4.5016
13	0.0657	0.2796	0.8924	2.6953	0.0638	0.2324	0.9634	3.2042
gr1	*0.4649*	*0.2228*	*0.6439*	*2.2387*	*0.4806*	*0.2338*	*1.1743*	*3.9772*

Table 7.7.: Relative frequency, rainfall frequency, average and standard deviation of wet day precipitation for all cyclonic moderate CPs

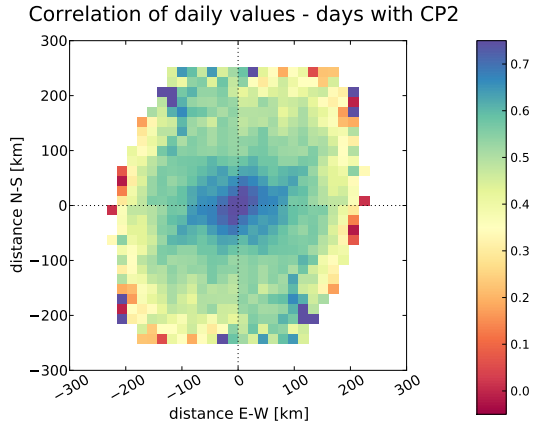
CP	October to March				April to September			
	rel. Freq.	p_{24h}	$\bar{H}_{24h}$	s_{24h}	rel. Freq.	p_{24h}	$\bar{H}_{24h}$	s_{24h}
5	0.0914	0.6671	2.6319	4.3994	0.1038	0.5819	2.8309	5.2148
6	0.045	0.6626	3.5101	5.7622	0.0575	0.5976	3.7232	6.3764
14	0.0401	0.3938	1.1529	2.7727	0.0448	0.3951	2.203	5.2943
15	0.0192	0.5376	2.2298	4.5033	0.0349	0.5885	3.9771	6.9979
99	0.0274	0.4573	2.5435	5.4805	0.022	0.4952	3.459	7.3874
gr2	*0.2231*	*0.5801*	*2.4977*	*4.5238*	*0.263*	*0.5471*	*3.1237*	*5.9006*

Table 7.8.: Relative frequency, rainfall frequency, average and standard deviation of wet day precipitation for all cyclonic humid CPs

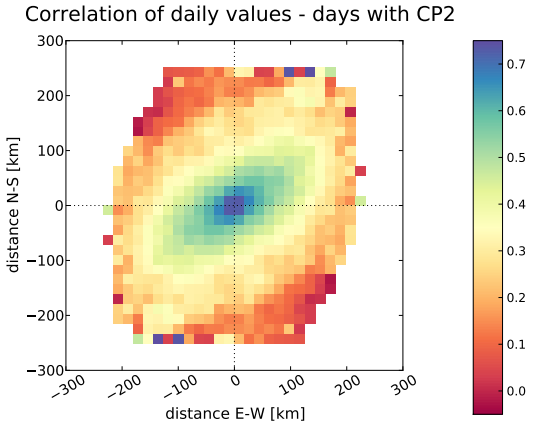
CP	October to March				April to September			
	rel. Freq.	p_{24h}	$\bar{H}_{24h}$	s_{24h}	rel. Freq.	p_{24h}	$\bar{H}_{24h}$	s_{24h}
3	0.0866	0.8954	7.8709	9.623	0.0633	0.804	6.1722	7.9476
4	0.071	0.8186	5.5872	7.8459	0.0356	0.7341	4.7911	7.3703
9	0.0719	0.6321	3.2564	5.7177	0.0665	0.7233	5.1617	8.1486
16	0.0276	0.7558	3.1647	4.6589	0.0344	0.7158	5.8871	9.0076
17	0.055	0.8169	4.85	6.3503	0.0564	0.7932	5.9318	7.8464
gr3	*0.3121*	*0.7911*	*5.3399*	*7.3036*	*0.2564*	*0.7591*	*5.6268*	*8.0397*

of the station pairs is ignored.

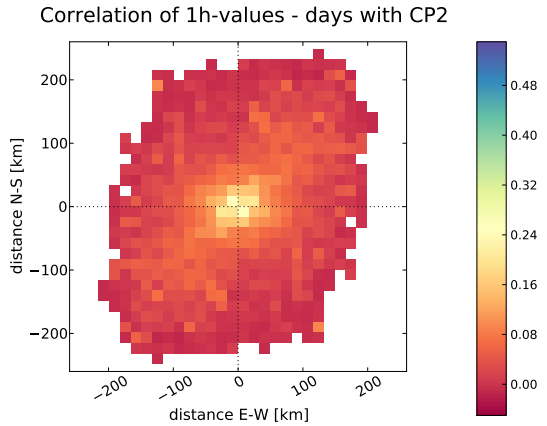
The results of the correlation analysis for all days belonging to CP2 are presented in Fig. 7.11.



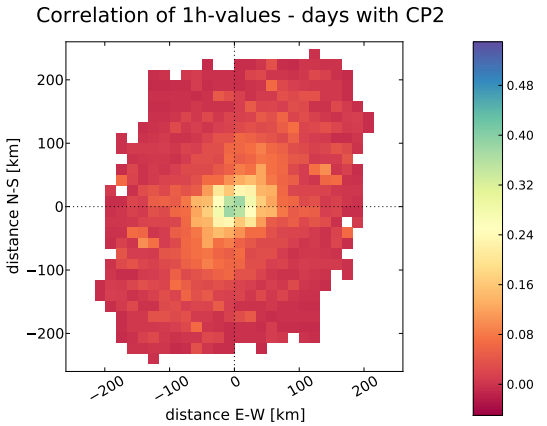
(a) October to March



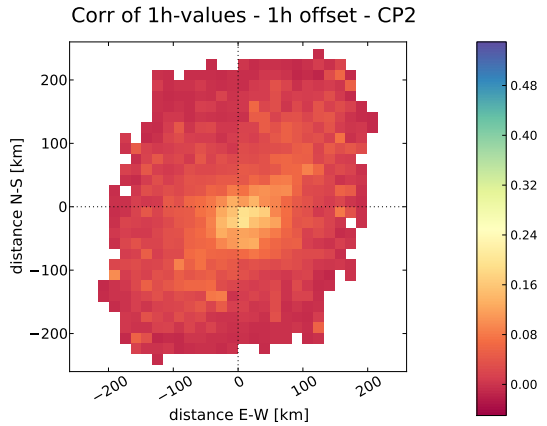
(b) April to September



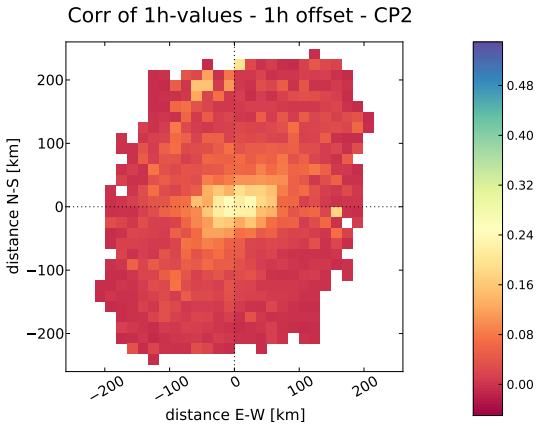
(c) October to March



(d) April to September

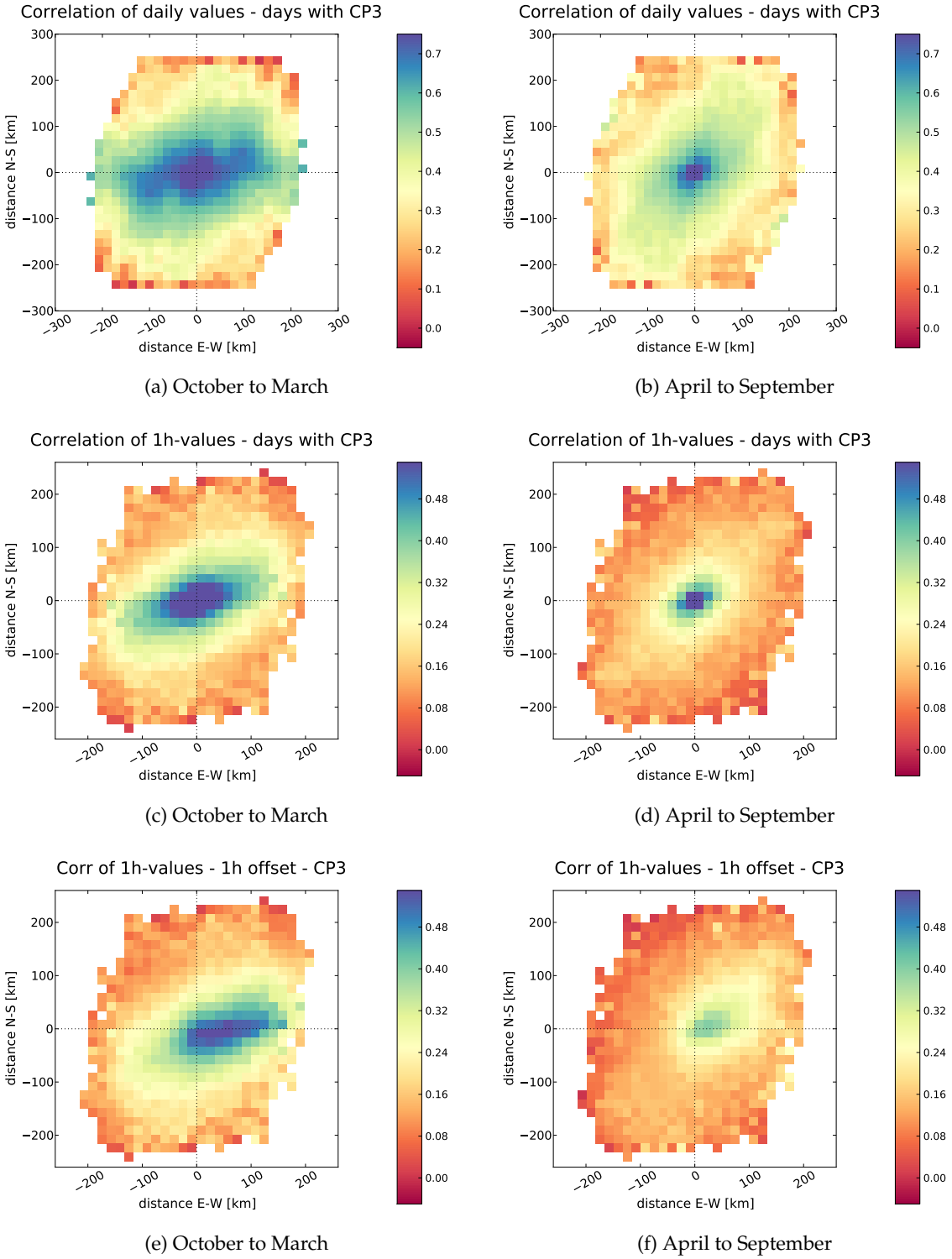


(e) October to March



(f) April to September

Figure 7.11.: Direction dependent correlation of daily and hourly precipitation stations for two seasons and all days belonging to CP2



(a) October to March

(b) April to September

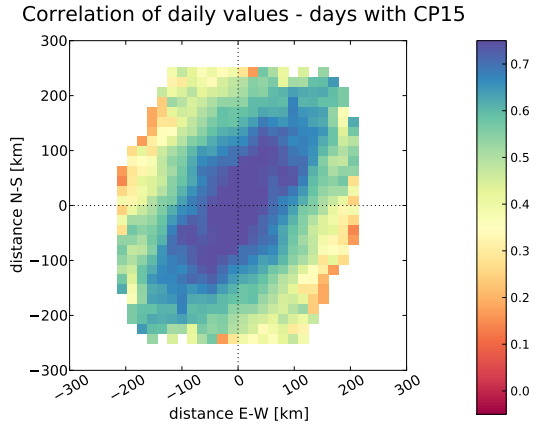
(c) October to March

(d) April to September

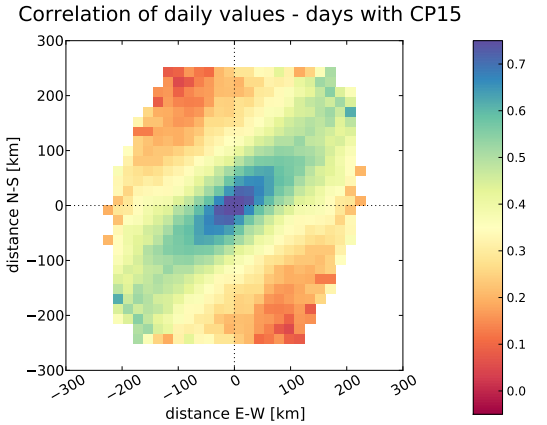
(e) October to March

(f) April to September

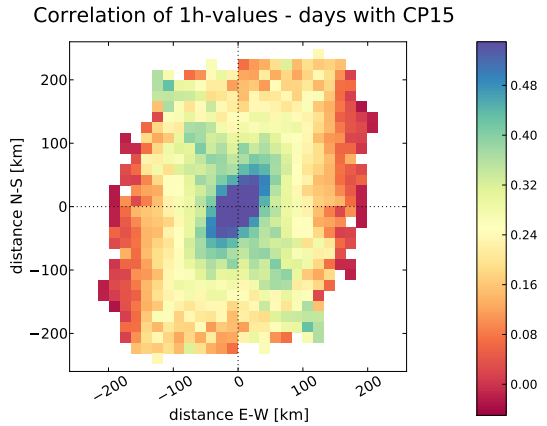
Figure 7.12.: Direction dependent correlation of daily and hourly precipitation stations for two seasons and all days belonging to CP3



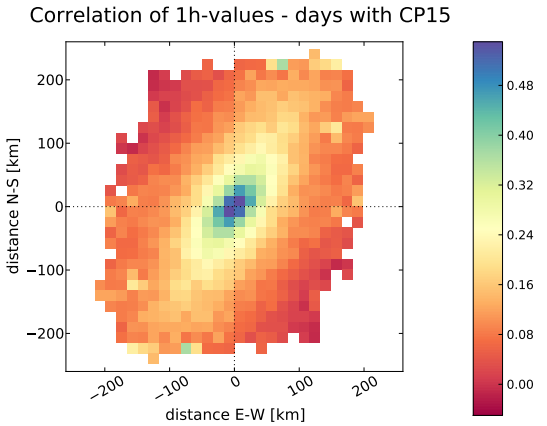
(a) October to March



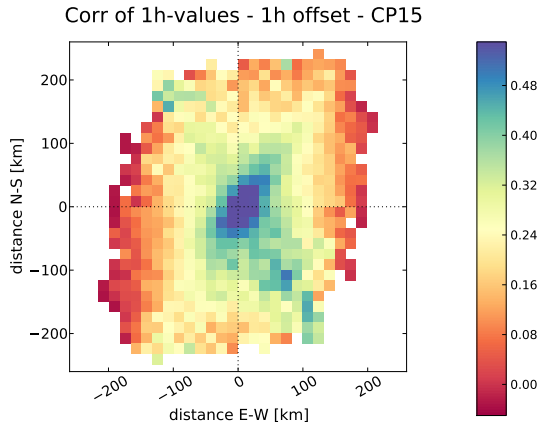
(b) April to September



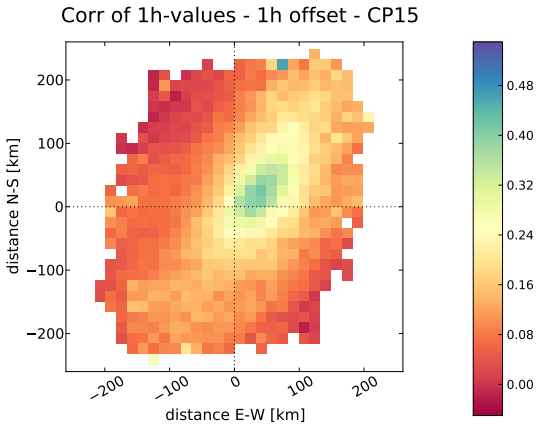
(c) October to March



(d) April to September



(e) October to March



(f) April to September

Figure 7.13.: Direction dependent correlation of daily and hourly precipitation stations for two seasons and all days belonging to CP15

In the left column only days during the six months period from October to March are considered, in the right column from April to September. The graphs in the first row display the correlations of daily values. The plots in the second row show the correlations of the $1\,\mathrm{h}$ rainfall time series. The plots are point-symmetrical to the origin. (The correlation between the stations i and j is the same as between the stations j and i.) The last row shows the hourly correlation if the time-series at the second station is shifted by $\Delta_t = 1\,\mathrm{h}$. Due to the shift, the fields in the last row are not symmetric. For the shifted correlation it makes a difference if station i in the pair i, j is chosen as starting or end point of the linking vector. The correlation is higher in flow direction than against it.

CP2 does not contribute much precipitation and rather inhibits atmospheric exchange. Therefore, the spatial correlations are low, as well as the anisotropy. A preferential flow direction can only be seen in daily values, from west in winter and from west-southwest in summer months. The correlation for stations in high distance is slightly higher in winter (Fig. 7.11a and Fig. 7.11b).

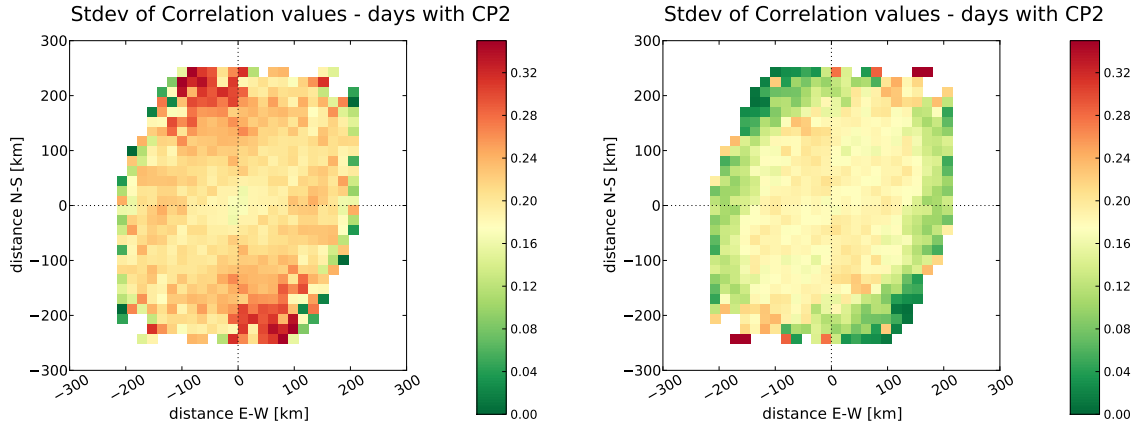
The days belonging to CP3 exhibit a very different spatial structure. The correlations are overall much higher, especially in the winter half-year. A clear preferential flow direction can be seen, between October and March from west-southwest and between September and April from southwest. The spatial correlation is stronger in winter months, at the same distance about twice as high. The shifted correlation (Fig. 7.12e and Fig. 7.12f) exhibit strong asymmetry. In winter months the correlation against the flow direction drops by 50% within the first 20 kilometers.

The highest correlation in daily precipitation values occur at days belonging to CP15 (Fig. 7.13a and Fig. 7.13b). Besides, CP15 exhibits the particularity of having two preferential flow directions during the winter half-year (Fig. 7.13c and Fig. 7.13e). What might seem paradox in the first moment, can be explained by the typical wind field of a cyclonic depression. The passing of a warm or cold front is generally accompanied by a turn in the wind direction. An approaching cyclonic depression is often preceded by winds from southwest. During the passage of the warm front, the wind direction turns to west, during the passage of the cold front to northwest (Weischet and Endlicher, 2008).

The generally higher range in daily correlations is related to the average atmospheric flow velocity. During $24\,\mathrm{h}$, passing fronts of the cyclonic depression have much time to travel and can generally reach both stations of a pair, even if the separation distance is $100\,\mathrm{km}$ or more. In one hour however, an approaching front can only propagate by several tens of kilometers, depending on the wind speed. Therefore, station pairs at higher distances remain uncorrelated in the hourly time series.

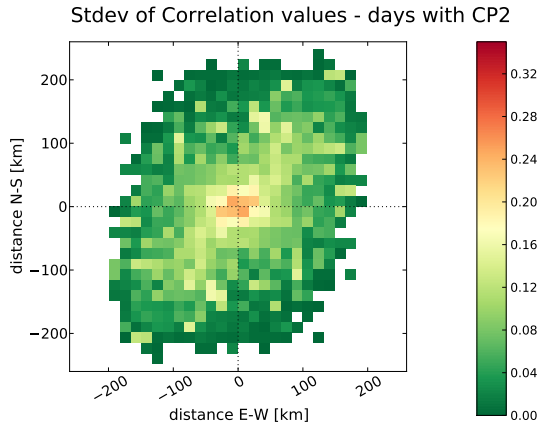
The number of available data pairs is very different depending on the configuration. Generally, the number is higher for closer distances. However, there are hardly any stations at distances of less than $10\,\mathrm{km}$. The low number of values at high distances explains some of the variations at fringes of the correlation plots.

Fig. 7.14 to Fig. 7.16 display the standard deviation of the correlation values in each grid cell. It measures the spread in the correlation distance relation. For anti-cyclonic CPs, as CP2 (Fig. 7.14), the standard deviation depends on the correlation value. At short distances,

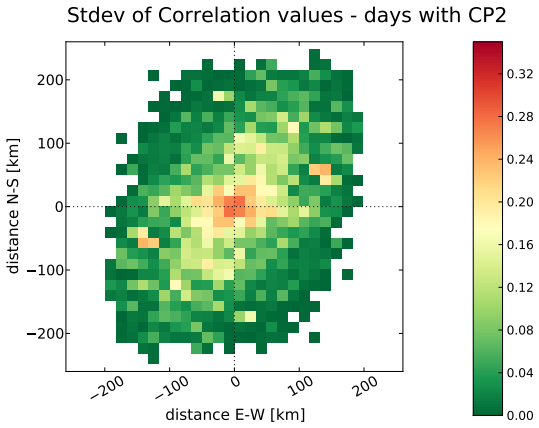


(a) October to March

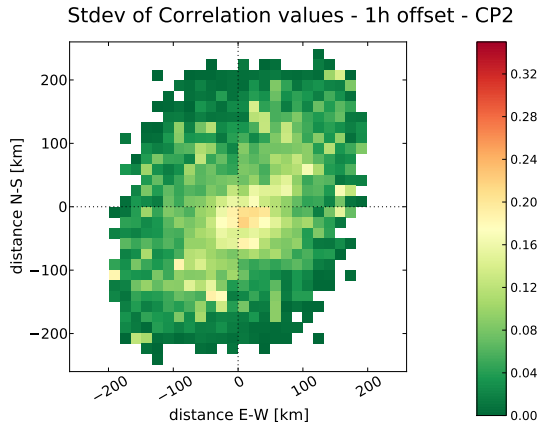
(b) April to September



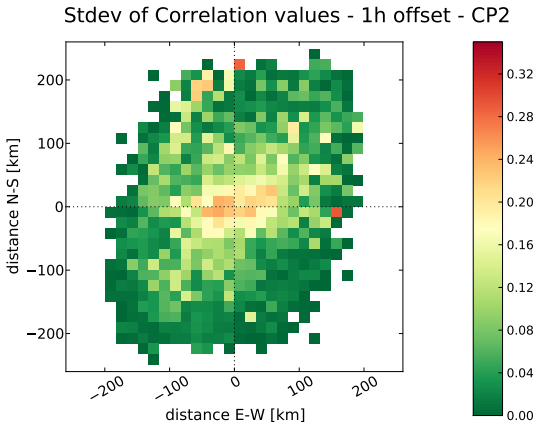
(c) October to March



(d) April to September

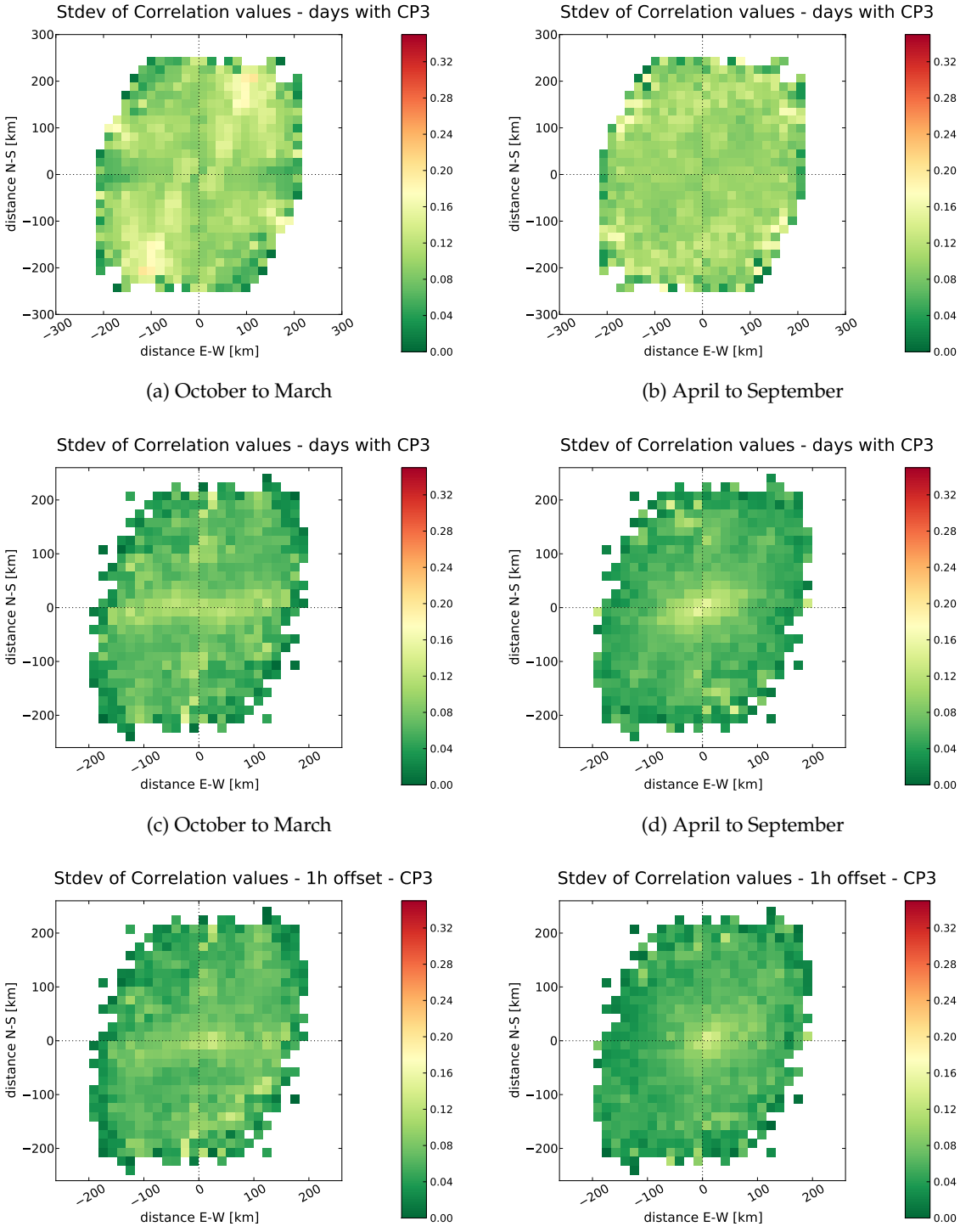


(e) October to March



(f) April to September

Figure 7.14.: Standard deviation of all correlation values in the same grid cell for two seasons and all days belonging to CP2



(a) October to March

(b) April to September

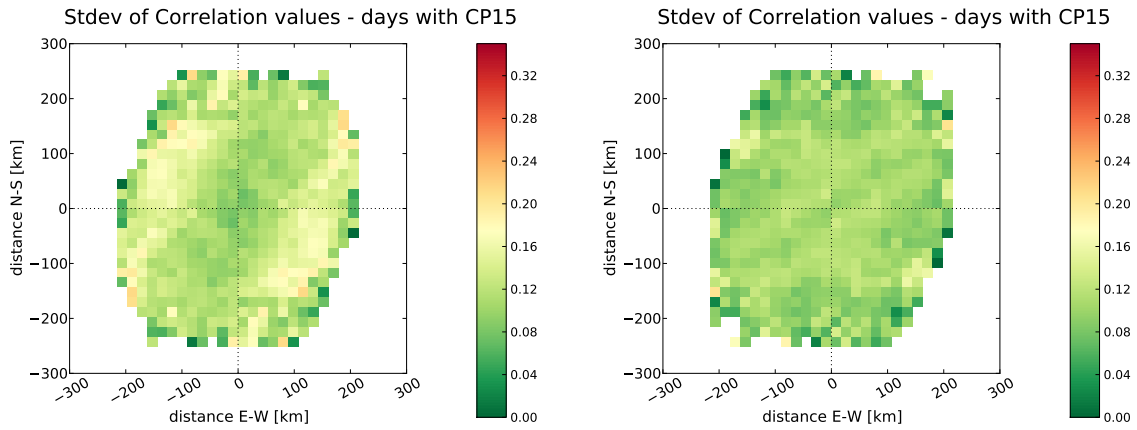
(c) October to March

(d) April to September

(e) October to March

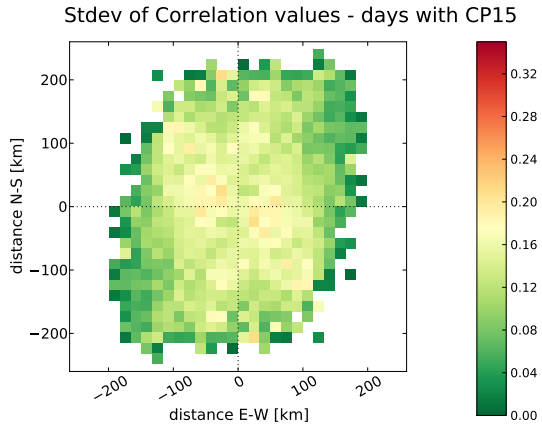
(f) April to September

Figure 7.15.: Standard deviation of all correlation values in the same grid cell for two seasons and all days belonging to CP3

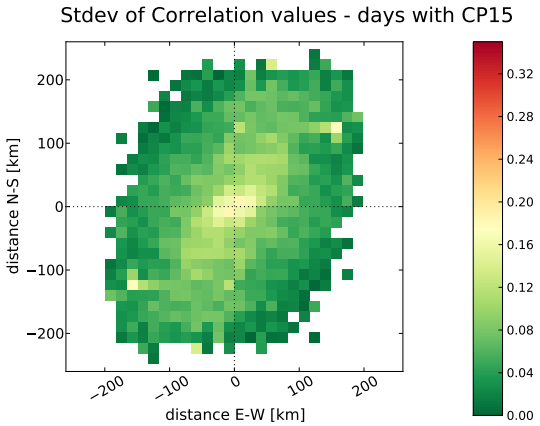


(a) October to March

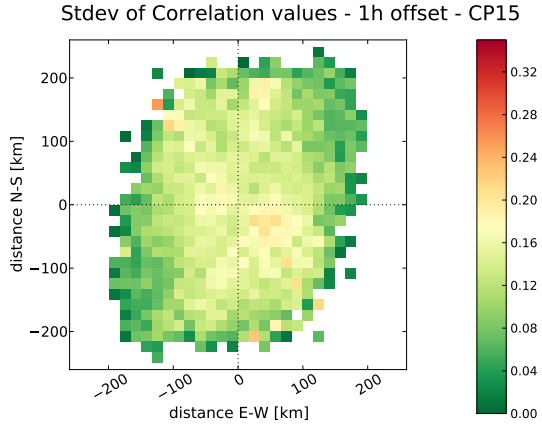
(b) April to September



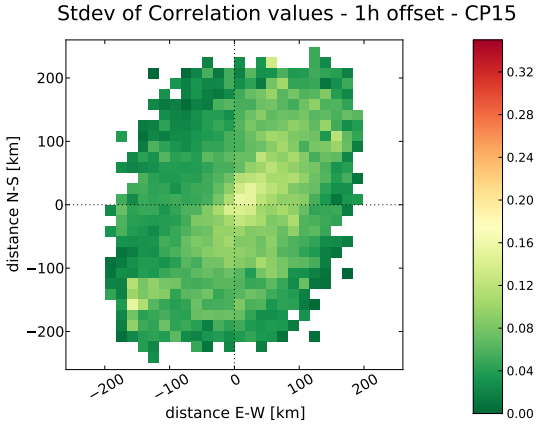
(c) October to March



(d) April to September



(e) October to March



(f) April to September

Figure 7.16.: Standard deviation of all correlation values in the same grid cell for two seasons and all days belonging to CP15

where the correlation is high, the spread is high too. The pictures of correlation and standard deviation look very similar. The relation between the distance vector and the correlation is heteroscedastic.

For the cyclonic CPs (Fig. 7.15 and Fig. 7.16), the standard deviation does not depend on the value. It shows only marginal spatial structure and the values are generally low. An exception is the secondary direction in CP15 that causes some spread. At very low separation distances the cyclonic CPs show a slight increase in standard deviation during the summer months (the second columns of (Fig. 7.15 and Fig. 7.16). At a lower level, it is a similar structure as for anticyclonic CPs. It is caused by convective, small scale precipitation events that are the main source of precipitation during anticyclonic CPs but can also occur during cyclonic CPs.

7.5.2.4. CP Dependent Correlations Used For the Generation of Simultaneous Time Series

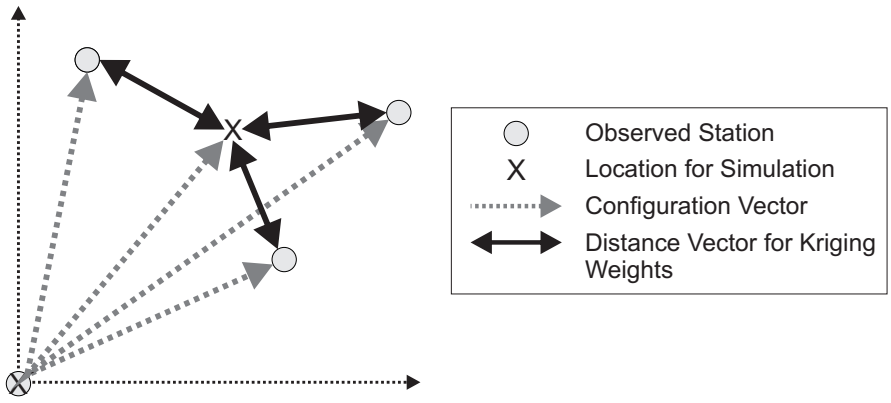


Figure 7.17.: Principle of the Kriging interpolation of CP dependent correlations

The generation of simultaneous precipitation time series requires the estimation of the CP dependent spatial correlations between ungauged locations. It is performed by ordinary kriging (Eq. (5.8)) according the the scheme in Fig. 7.17. The correlation values are interpolated from observed station pairs according to the configuration of the simulation locations. This means the x and y component of the vector linking each station pair (gray dashed arrows in Fig. 7.17). Like in the correlation analysis the starting point of the vector linking the generation locations is projected onto the origin $(0|0)$. Thus, the absolute location within the study area is ignored. The configuration of the simulation locations (black crosses) is compared with the configurations of all station pairs of the observed data set. The twenty pairs with the most similar configuration are chosen. Then the values for each CP are interpolated by ordinary Kriging. The Kriging weights are assigned by an empirical variogram $\gamma(h)$ (Eq. (5.3)) according to the distances h between the configuration vector end points (black arrows).

7.6. Generation Algorithm for Simultaneous 1h Precipitation Time Series with Spatial Temporal Dependence

The generation of n_{stat} simultaneous, spatial interdependent time series is realized by a three step simulation passing by the intermediate of simultaneous daily time series. In the following, the superscript $i \in [1, n_{stat}]$ is used to identify the generation location. It is not an exponent, but an identifier.)

1. First an initial time series in daily temporal resolution is set up for each of the n_{stat} generation locations. The required parameters are shape $\kappa^i\left(H_{wet}\right)$ and scale $\lambda^i\left(H_{wet}\right)$ of the Weibull Distribution for daily precipitation values and daily rainfall probability $P^i_{24\text{h}}$. $\kappa^i\left(H_{wet}\right)$ and $\lambda^i\left(H_{wet}\right)$ are calculated according to the equations Eq. (6.13) and Eq. (6.14) from the expected values for average $E\left(H^i_{wet}\right)$ and standard deviation $D\left(H^i_{wet}\right)$ of wet day precipitation.

   In the same way as for the single site simulation, the rain day probability $P^i_{24\text{h}}$, as well as the expectations $E\left(H^i_{wet}\right)$ and $D\left(H^i_{wet}\right)$ are conditioned on the monthly precipitation sum H^i_{mon} at each location i. Additionally, the parameter interdependencies at the different stations are considered.

2. The temporal order of the daily time series is optimized by Simulated Annealing for all n_{stat} stations simultaneously.

3. The daily values are disaggregated into hourly values. The daily precipitation sum of each day is randomly distributed among the 24 associated hours. Next, the hourly series is optimized by exchanging small precipitation increments among the hours of the same day. The optimization is performed by Simulated Annealing.

7.6.1. Parameter Estimation for the Initial Daily Time Series

The estimation of the parameters defining the initial daily time series is performed month by month.

1. Starting point are the time series of monthly rainfall sums H^i_{mon} at the n_{stat} target locations of the simulation. If the monthly sum at a station i is not available, H^i_{mon} is estimated from surrounding stations by External Drift Kriging Eq. (5.11) with smoothed altitude as external information.

2. The target values H^i_{mon} are transformed into the copula space U of uniformly distributed values between 0 and 1 by the marginal distribution F. F is depending on the season (December to February, March to May, June to August, September to November):

$$u^i_{mon} = F\left(H^i_{mon}\right) \tag{7.10}$$

3. To condition the generation parameters on the monthly sum, the matrix $\mathbf{\Sigma}$ of parameter interdependencies in the standard transformed space has to be set up. The correlations that have to be estimated are ρ_{ijkl} with $i, j \in [1, n_{stat}]$ defining the station

pair and $k, l \in [1, 4]$ defining the parameter combination of parameter k at station i and parameter l at station j. Therefore, the matrix $\mathbf{\Sigma}$ of parameter interdependencies has $4n_{stat} \times 4n_{stat}$ entries. The convention for the order of the parameters k and l is: $1 \hat{=} H_{mon}$; $2 \hat{=} P_{24\text{h}}$; $3 \hat{=} E\left(H_{wet}\right)$; $4 \hat{=} D\left(H_{wet}\right)$

$$\Sigma = \begin{array}{c}
\begin{array}{ccc}
\overbrace{\hspace{12em}}^{\text{Station 1}} & \overbrace{\hspace{12em}}^{\text{Station 2}} & \overbrace{\hspace{12em}}^{\text{Station } n}
\end{array} \\
\left(\begin{array}{cccccccccccccc}
r_{1111} & r_{1112} & r_{1113} & r_{1114} & \hat{r}_{1211} & \hat{r}_{1212} & \hat{r}_{1213} & \hat{r}_{1214} & & \hat{r}_{1n11} & \hat{r}_{1n12} & \hat{r}_{1n13} & \hat{r}_{1n14} \\
r_{1121} & r_{1122} & r_{1123} & r_{1124} & \hat{r}_{1221} & \hat{r}_{1222} & \hat{r}_{1223} & \hat{r}_{1224} & \ldots & \hat{r}_{1n21} & \hat{r}_{1n22} & \hat{r}_{1n23} & \hat{r}_{1n24} \\
r_{1131} & r_{1132} & r_{1133} & r_{1134} & \hat{r}_{1231} & \hat{r}_{1232} & \hat{r}_{1233} & \hat{r}_{1234} & & \hat{r}_{1n31} & \hat{r}_{1n32} & \hat{r}_{1n33} & \hat{r}_{1n34} \\
r_{1141} & r_{1142} & r_{1143} & r_{1144} & \hat{r}_{1241} & \hat{r}_{1242} & \hat{r}_{1243} & \hat{r}_{1244} & & \hat{r}_{1n41} & \hat{r}_{1n42} & \hat{r}_{1n43} & \hat{r}_{1n44} \\
\\
\hat{r}_{2111} & \hat{r}_{2112} & \hat{r}_{2113} & \hat{r}_{2114} & r_{2211} & r_{2212} & r_{2213} & r_{2214} & & \hat{r}_{2n11} & \hat{r}_{2n12} & \hat{r}_{2n13} & \hat{r}_{2n14} \\
\hat{r}_{2121} & \hat{r}_{2122} & \hat{r}_{2123} & \hat{r}_{2124} & r_{2221} & r_{2222} & r_{2223} & r_{2224} & \ldots & \hat{r}_{2n21} & \hat{r}_{2n22} & \hat{r}_{2n23} & \hat{r}_{2n24} \\
\hat{r}_{2131} & \hat{r}_{2132} & \hat{r}_{2133} & \hat{r}_{2134} & r_{2231} & r_{2232} & r_{2233} & r_{2234} & & \hat{r}_{2n31} & \hat{r}_{2n32} & \hat{r}_{2n33} & \hat{r}_{2n34} \\
\hat{r}_{2141} & \hat{r}_{2142} & \hat{r}_{2143} & \hat{r}_{2144} & r_{2241} & r_{2242} & r_{2243} & r_{2244} & & \hat{r}_{2n41} & \hat{r}_{2n42} & \hat{r}_{2n43} & \hat{r}_{2n44} \\
 & \ldots & & & & \ldots & & & \ldots & & \ldots & & \\
\hat{r}_{n111} & \hat{r}_{n112} & \hat{r}_{n113} & \hat{r}_{n114} & \hat{r}_{n211} & \hat{r}_{n212} & \hat{r}_{n213} & \hat{r}_{n214} & & r_{nn11} & r_{nn12} & r_{nn13} & r_{nn14} \\
\hat{r}_{n121} & \hat{r}_{n122} & \hat{r}_{n123} & \hat{r}_{n124} & \hat{r}_{n221} & \hat{r}_{n222} & \hat{r}_{n223} & \hat{r}_{n224} & \ldots & r_{nn21} & r_{nn22} & r_{nn23} & r_{nn24} \\
\hat{r}_{n131} & \hat{r}_{n132} & \hat{r}_{n133} & \hat{r}_{n134} & \hat{r}_{n231} & \hat{r}_{n232} & \hat{r}_{n233} & \hat{r}_{n234} & & r_{nn31} & r_{nn32} & r_{nn33} & r_{nn34} \\
\hat{r}_{n141} & \hat{r}_{n142} & \hat{r}_{n143} & \hat{r}_{n144} & \hat{r}_{n241} & \hat{r}_{n242} & \hat{r}_{n243} & \hat{r}_{n244} & & r_{nn41} & r_{nn42} & r_{nn43} & r_{nn44}
\end{array}\right)
\begin{array}{l}
\left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} \text{Station 1} \\
\\
\left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} \text{Station 2} \\
\\
\left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} \text{Station } n
\end{array}
\end{array}$$

Figure 7.18.: Matrix of the parameter correlations in the standard normal transformed space for the generation of simultaneous time series

Fig. 7.18 shows a sketch of this matrix for the generation of n stations. The parameter interdependencies at the same site ($i = j$), which are the four times four blocks around the diagonal of $\mathbf{\Sigma}$, are assumed to be constant. In the same way as for the single site generation, the correlations $r_{iikl},\ i \in [1, n_{stat}]$ are taken from Table 6.2 (according to the current season). The parameter dependencies between the different locations $\hat{r}_{ijkl}$ (where $i \neq j$) are calculated by the distance d_{ij} using the regression function (see Table 7.4 and Table 7.5) for the respective parameter combination k, l and Eq. (6.61) for the transformation of the rank correlations ρ_{ijkl} into Standard Normal correlations r_{ijkl}.

4. The matrix $\mathbf{\Sigma}$ is being conditioned on the vector of Standard Normal transformed monthly sums.

$$a = \left[\Phi\left(u_{mon}^{i}\right), i = 1 \ldots n_{stat}\right]^{T} \tag{7.11}$$

where Φ is the Standard Normal Distribution Function.

This leads to a normal distribution $X \sim \mathcal{N}\left(\mu_{|a}, \mathbf{\Sigma}_{|a}\right)$ with conditioned mean vector $\mu_{|a}$ and conditioned covariance matrix $\mathbf{\Sigma}_{|a}$. $\mu_{|a}$ and $\mathbf{\Sigma}_{|a}$ are calculated according to Eq. (A.5) and Eq. (A.6) in Section A.2 of the appendix. The conditioned correlation matrix $\mathbf{\Sigma}_{|a}$ can be calculated by inverting the overall correlation matrix $\mathbf{\Sigma}$ (Fig. 7.18), dropping all the rows and columns belonging to H_{mon}^{i} and inverting back.

The conditioned, Standard Normal transformed random number vector $z_{|a}$ representing the parameters $P_{24\text{h}}$, $E\left(H_{wet}\right)$ and $D\left(H_{wet}\right)$ at all stations n_{stat} are drawn by means of a vector z of uncorrelated random numbers and lower triangle matrix $\mathbf{L}$ of the

Cholesky decomposition of $\mathbf{\Sigma}_a$

$$z_{|a} = \mu_{|a} + \mathbf{L} \cdot z \tag{7.12}$$

$$\mathbf{\Sigma_{|a}} = \mathbf{L}\mathbf{L}^T \tag{7.13}$$

$z_{|a}$ and z have $3 \cdot n_{stat}$ elements.

It is a necessary condition for the applied Cholesky decomposition that the matrix Σ is positive definite. In theory, any correlation matrix is positive definite. If not, the pairwise correlations cannot be valid all at once without contradiction. However, due to estimation errors in the regression, it can happen that the estimated correlation matrix Σ is not positive definite. In this case, it is corrected to the nearest positive definite matrix according to the algorithm described in Higham (2002).

It has been tested how much the correction according to Higham (2002) deforms the estimated correlation matrix. In none of the tested configurations, the correlations were changed by more than 5%.

5. The drawn values, x^* representing the rainfall probability, the average and the standard deviation of wet day precipitation are now Standard Normally distributed. They are back-transformed to absolute values in two steps.

$$u\left(P^i_{24\text{h}}\right) = \Phi\left(z^i_{|a}\left(P^i_{24\text{h}}\right)\right) \tag{7.14}$$

$$P^{*i}_{24\text{h}} = F^{-1}\left(u\left(P^i_{24\text{h}}\right)\right) \tag{7.15}$$

where

$z_{\|a}\left(P_{24\text{h}}\right)$	conditioned random number in the Standard Normal transformed space for the $24\,\text{h}$ rainfall probability in the current month
$u\left(P^i_{24\text{h}}\right)$	uniform transform of the $24\,\text{h}$ rainfall probability
Φ	standard normal distribution function
$F^{-1}\left(u\left(P_{24\text{h}}\right)\right)$	inverse of the theoretical distribution function of the $24\,\text{h}$ rainfall probability according to Table 7.3
$P^{*i}_{24\text{h}}$	target value for the $24\,\text{h}$ rainfall probability of the current month at station i

The proceeding for the target values of the average rain day precipitation sum $E^*\left(H^i_{wet}\right)$ and the standard deviation $D^*\left(H^i_{wet}\right)$ it the same. The parameters of the distribution functions applied for back-transformation can be found in Table 7.3.

6. The parameters $\lambda\left(H_{wet}\right)$ and $\kappa\left(H_{wet}\right)$ of the Weibull Distribution for the $1\,\text{h}$ rainfall amounts in wet hours are calculated from the drawn statistics $E^*\left(X\right)$ and $D*\left(X\right)$ by the method of moment (Eq. (6.13) and Eq. (6.14)).

7.6.2. Generation of Simultaneous Daily Precipitation Time Series

The generation of daily time series follows the same steps as the generation of hourly time series for single locations (see Section 6.6.2):

Given the monthly series of rainfall probabilities $P^i_{24\text{h}}$ and the Weibull Distribution parameters $\lambda\left(H^i_{wet}\right)$ and $\kappa\left(H^i_{wet}\right)$, an initial daily time series is generated at each location i separately.

For each day, a uniform random number u between 0 and 1 is drawn. It is compared with the rainfall probability of the current month $P^i_{24\text{h}}$. If the day is chosen to be a rain day, the daily precipitation amount is drawn according to Eq. (6.70) from the inverse of the Weibull Distribution.

The n_{stat} initial time series exhibit the right distribution of the daily precipitation amounts and, at each location i, they follow the expectation of the monthly precipitation sums. They are passed to the resampling scheme for optimization of the temporal sequence of values.

The optimization is performed by a Simulated Annealing algorithm similar to the single site simulation scheme. In each step, one of the stations is randomly chosen. Then two time steps of the same month are selected for a swap. All n_{stat} time series are evaluated simultaneously by one global objective function O according to Eq. (6.71). The consequences of the swap for all n_{stat} stations is evaluated by O.

The statistical parameter that are incorporated in the objective function O are the following:

- **Average daily precipitation amount and precipitation frequency of the three CP classes defined in Section 7.5.2**
  The average values of the three groups can be found in Table 7.6 to Table 7.8. The statistics are calculated from 575 daily stations.

  Since the monthly precipitation sums do not correspond to the observed values but are drawn randomly, the average precipitation is not considered in absolute values. It is normed at each station by the simulated monthly rainfall sum.

- **Cross correlation between the daily time series of all station pairs**
  The target value of the cross correlations between the station pair ij, where $i, j \in [1, n_{stat}]$ is estimated in the four dimensional space (x_s, x_e, y_s, y_e) of the start and endpoint of the vectors linking the station pair ij.

  The applied method is proposed by Brommundt (2008). The estimated correlation $\hat{r}_{ij}$ is the sum of a regression estimation based on the four dimensional distance $d_{4\text{D}}$ of the stations and a residual estimated by ordinary kriging of the observed residuals in the four dimensional space. If $\vec{ij}$ and $\vec{kl}$ are two vectors linking the stations pairs ij and kl, four dimensional distance $d_{4\text{D}}$ is defined as:

$$d_{4\text{D}} = \sqrt{(x_i - x_k)^2 + (x_j - x_l)^2 + (y_i - y_k)^2 + (y_j - y_l)^2} \tag{7.16}$$

  The x component is the easting in Gauss-Krüger 3 coordinates of the station location. y is the northing. More details on the method can be found in Brommundt (2008).

- **The autocorrelation of $24\,$h precipitaton values according to Eq. (6.77)**

7.6.3. Disaggregation to hourly values

The initial hourly time series is set up by distributing the amount of each day randomly among the 24 associated hours. To achieve a time series with realistic scaling behavior as a starting point, the values are distributed in the following manner: One hour of the day is randomly chosen. It receives 40% of the daily rainfall but at least 0.1 mm. The next hour is chosen and receives again 40% of the remaining daily sum but at least 0.1 mm. One hour can be picked several times. As soon as the daily sum is completely distributed, the procedure continues on the next day.

The Simulated Annealing algorithm in the disaggregation scheme exchanges small precipitation increments between the time steps of each day. For each swap at first the day is chosen followed by the two hours of the day. At the beginning, the precipitation increments are in a range between 0.75 mm and 0.05 mm. An alternation between series of big and small increments is applied to avoid that the structure is dominated by big rainfall amounts. Towards the end of the optimization small increments of 0.01 mm are used.

The objective function O consists of two parts. One part is evaluating the performance at each side individually and the other part measures the spatial dependence. At the beginning, the Simulated Annealing algorithm alternates between evaluating each station separately and evaluating all station simultaneously. The number of single site evaluation at each Annealing Temperature T drops proportional to the temperature T. The alternation is required to give the individual time series enough freedom in resampling. Without single site optimizations, one runs the risk that a positive change at a single site is prevented due to the relation with the other stations whose sequence is not yet perfect either.

The objective function parts O_i evaluating the single site performance are mostly the same as in Section 6.6.2.1:

- **1 h extreme value according to KOSTRA statistics**
  In the new setup the daily precipitation values are completely defined by the Weibull Distribution. Therefore, the objective function according to the 24 h precipitation extreme cannot be applied any more. It is replaced by the 1 h precipitation extreme according to the official German extreme value statistics KOSTRA (Bartels et al., 2005). The Gumbel Distribution of KOSTRA is re-established at each simulation location by the yearly and 100 yearly extreme value from KOSTRA. The restored distribution is used to draw the target value for the highest 1 h precipitation amount in the simulation year. To consider the high spatial correlation in the potential extremes, the 1 h extreme values at all stations are drawn with one single random number. (This means that the maximum value at nearby simulated stations is similar. However, it does *not* mean that maximum is reached at all stations in the same time step.)
- **scaling of the statistical moments between different aggregation levels**
  The scaling exponent b (Eq. (6.76)) is evaluated for the first three statistical moments of the time series (Eq. (6.75)). The aggregation intervals Δ_t are ranging from 1 h to 24 h.
- **autocorrelation on different aggregation levels**
  The autocorrelation according to Eq. (6.77) is evaluated for the same aggregation levels

as in the single site simulation. The aggregations are $\Delta_t = 1\,\mathrm{h}, 2\,\mathrm{h}, 3\,\mathrm{h}, 6\,\mathrm{h}$ and $12\,\mathrm{h}$. The $24\,\mathrm{h}$ autocorrelation has already been exploited in the generation of the intermediate daily time series (Section 7.6.2).

Additionally, the seasonally weighted autocorrelation of hourly values is evaluated. The target values of the autocorrelation depend on the generation year.

- **average hourly precipitation frequency for three different weather type classes**
  The CP classes are the same as in the optimization of the intermediate daily time series (Section 7.6.2). In the daily generation not only the precipitation frequency, but also the average precipitation is used to evaluate the time series. For the disaggregation, however, the average is of no information any more. Since no rainfall volume is exchanged between the days and thus between the CPs, the average on any regarded temporal aggregation is constant and completely defined by the daily generation.

The objective functions evaluating the spatial dependence are the CP-dependent correlations estimated in Section 7.5.2. The correlations are measured pairwise and for two seasons from March to October and April to September, accounting for differences due to convective events in the summer months.

As a representation of the flow direction, two types of correlations estimated for each station pair i, j: the correlation of simultaneously measured time steps $x^i(t)$ and $x^j(t)$ as well as the correlation between $x^i(t)$ and $x^j(t + 1\,\mathrm{h})$ with one hour time shift.

The result is a series of 72 matrices with $n_{stat} \times n_{stat}$ elements; one matrix per season, CP and for simultaneous or shifted correlation. The target values of the correlations are determined by interpolation (see Fig. 7.17).

7.7. Comparison of Simulated and Observed Simultaneous Time Series

The performance of the algorithm will be tested by a comparison of simulated and observed time series. First, the same statistics as for the single site generation (Chapter 6) are investitaged. Then, the representation of spatial dependence is tested.

Table 7.9.: Coordinates of target locations for the generation of simultaneous time series

Station	x-GK3 [m]	y-GK3 [m]
1	3498810	5383440
2	3501820	5384490
3	3499810	5386440
4	3500810	5381940

Since the generator is primarily designed for urban sewage systems, the stations should lie in a spatial configuration that is of realistic scale for sewage system simulations. Even

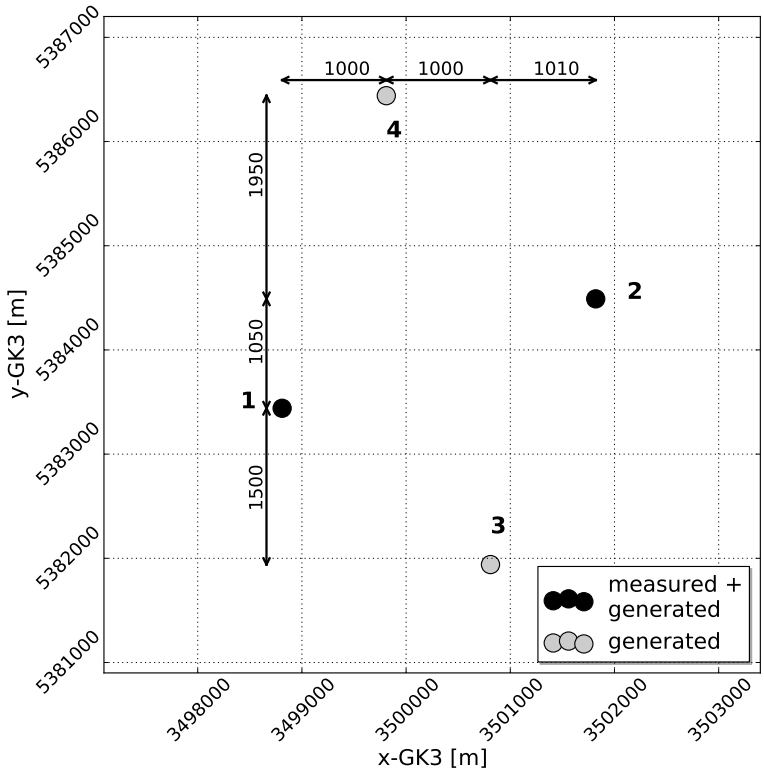


Figure 7.19.: Overview of the target location for the generation of simultaneous time series

in big cities of Baden-Württemberg, the sewage system rarely exceeds a dimension of one hundred square kilometers. Therefore, the maximum distance between the stations should not exceed several kilometers.

The verification is done for two stations in Holzgerlingen that are three kilometers apart. The two stations share a long common data period from 1977 to 1992 and exhibit both a very low number of missing values (less than 10%). To make the simulation more realistic for a real sewage system setup and to add more constraints, the generation is not only performed for the two measurement stations but for a total of four stations. The coordinates of all four stations can be found in Table 7.9. Observed data is available at stations 1 and 2. Station 2 is the one that was used for the verification of the single site generation (Section 6.7 in Chapter 6). The spatial configurations of the stations is displayed in Fig. 7.19.

7.7.1. Comparison of Observed and Simulated Distributions

In this section the same statistics as in Section 6.7 for the single site generator are tested. The statistical moment, the exceedance frequencies to different threshold in hourly and in daily resolution, the empirical annual series as a representation of the extreme value behavior on different aggregations and the length of wet and dry spells as a measure of the temporal persistence in the generated time series.

Table 7.10.: basic hourly statistics of the simultaneous simulated and observed time series in Holzgerlingen, 1977 to 1992

	Station 1		Station 2		Station 3	Station 4
	simulated	observed	simulated	observed	simulated	simulated
mean [mm]	0.532	0.857	0.526	0.879	0.572	0.537
stdev. [mm]	0.864	1.288	0.907	1.382	0.950	0.902
skewness [-]	12.119	6.133	12.465	6.605	10.006	10.613
$p_{>0.0mm}$	27.74%	25.29%	28.78%	27.16%	27.98%	29.91%
$p_{>0.1mm}$	15.78%	8.20%	16.31%	9.04%	16.19%	16.47%
$p_{>0.5mm}$	4.638%	3.812%	4.662%	4.313%	4.850%	4.777%
$p_{>2mm}$	0.499%	0.796%	0.544%	0.859%	0.588%	0.591%
$p_{>5mm}$	0.073%	0.130%	0.086%	0.153%	0.102%	0.074%
$p_{>10mm}$	0.015%	0.021%	0.016%	0.033%	0.021%	0.012%
$p_{>20mm}$	0.0028%	0.0030%	0.0050%	0.0052%	0.0028%	0.0028%
H_{max} [mm]	30.3	27.8	29.0	27.2	29.650	29.035

Table 7.11.: basic daily statistics of the simultaneous simulated and observed time series in Holzgerlingen, 1977 to 1992

	Station 1		Station 2		Station 3	Station 4
	simulated	observed	simulated	observed	simulated	simulated
mean [mm]	5.654	5.125	5.641	5.432	5.899	5.668
stdev. [mm]	6.283	6.006	6.804	6.403	6.910	6.655
skewness [-]	2.822	3.219	4.905	3.199	3.835	3.668
$p_{>0.1mm}$	41.6%	41.4%	42.5%	43.4%	42.2%	43.3%
$p_{>0.5mm}$	36.7%	34.6%	38.0%	36.8%	37.8%	37.8%
$p_{>2mm}$	25.9%	22.3%	25.9%	24.5%	26.6%	26.2%
$p_{>5mm}$	13.6%	11.7%	14.0%	12.6%	14.8%	14.3%
$p_{>10mm}$	5.3%	4.6%	5.5%	5.4%	6.2%	6.0%
$p_{>20mm}$	1.5%	1.0%	1.3%	1.4%	1.5%	1.4%

In the comparison between Table 6.4 on one hand and Table 7.10 and Table 7.11 on the other hand, some effects of the different generation algorithms become apparent. In single site generation the hourly statistics are closer to the statistics of observed time series whereas in multi site generation the daily values show less differences to observed time series.

The representation of the observed time series by the generation is generally good. On hourly aggregation (Table 7.10), the rainfall probability is modeled well if 0.0 mm is taken as threshold, however, overestimated if 0.1 mm is taken. The observed time series exhibits a high fraction of very small rainfall amounts that cannot be reproduced by the multi site generation. Since the deviation concerns only the smallest values, it does not have a serious impact. The probabilities of more intensive precipitation are all very well modeled. The skewness of hourly values is overestimated.

On daily scale all statistics are modeled very well. The only exception is the fraction of very small values with less than 0.5 mm per day that are underestimated. The very different generated daily precipitation maxima reveal the high possible variability in the generated statistics. While the observed maxima are very close to each other, the simulated differ by a factor of more than two. From this data, it seems that the simulation overestimates the variability in the highest 24 h extremes. However, referring to only one single values, the sampling variance in the maximum is high. With the available data it cannot be judged if the deviation is random or systematic. From the practical point of view, an overestimation in the extreme value variability is at least better than an underestimation because it prevents from "bad surprises" of catastrophically high events.

Fig. 7.20 and Fig. 7.21 compare the observed and simulated empirical distributions at the two observing stations in Holzgerlingen. The annual series are calculated from the ten highest hourly precipitation amounts of each year, so 120 values in total. In all aggregation the range is realistic. On hourly scale the simulated frequency of the lowest values between two and four millimeters is too high. Medium values between eight and twelve millimeters are less frequent than in the observations. For longer aggregation intervals, coming closer to the daily values that are explicitly modeled by the distribution function, the variability in the annual series increases. Therefore, the distributions for 6 h and 12 h extremes are closer to the observed.

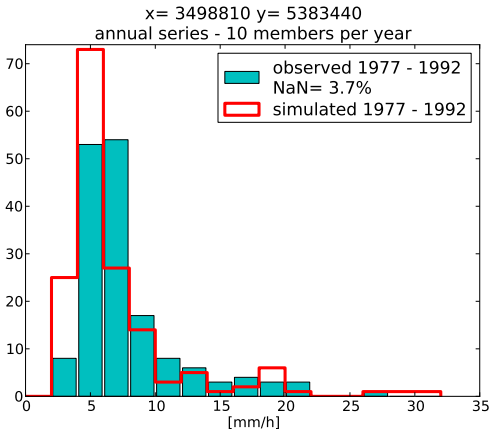
Although the two stations are only three kilometers apart, the observed annual series differ considerably. At station 2 the average extreme is significantly higher. Only on the 24 h aggregation level (last line of Fig. 7.20 and Fig. 7.21) the differences between the simulated annual series reach the same extend.

The analyzes of the annual series is repeated monthwise (Fig. 7.22). Compared to the single site generator, the generation algorithm for multiple time series has a lower ability to distinguish between the seasons The actual differences in the distributions for January and Juli are not completely represented in the simulated time series.

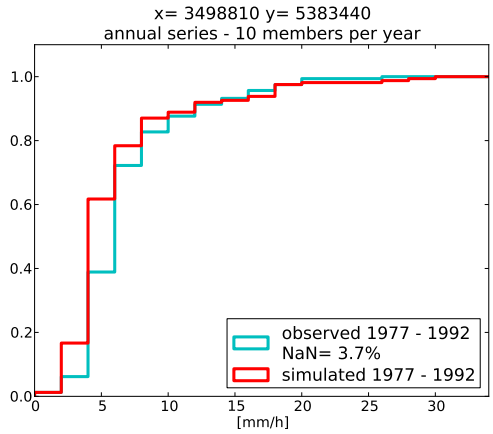
The overestimation in the rainfall probability on hourly scale (Table 7.10) leads to an overestimation in the length and the absolute number of wet spells (Fig. 7.23). At Station 1, the overestimation is of 3146 simulated wet spells instead of 2213 spells observed in the sixteen years from 1977 to 1992, at Station 2 of 3246 instead of 2380. The shape of the distribution on the other hand is realistic. Nevertheless, the frequency of very long wet spells over two days is overestimated at both stations.

The distribution of wet spell lengths is very sensitive to the definition of the end of a spell. Changing the maximum length of gaps that are ignored in the wet spell length estimation from 1 h to 3 h, the overestimation of the total number of wet spells can be reduced to realistic values. This means that the the simulated occurrence of precipitation is too scattered. Compared to the observations, short gaps of one or two hours are more frequent.

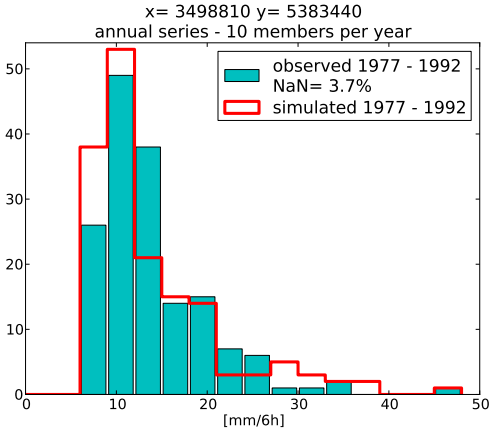
Fig. 7.24 shows the histogram and the empirical CDF of the 100 longest observed and simulated wet spells at both stations. Gaps up to 1 h are ignored. In contrast to the single site generator that slightly underestimates the lengths, the simulated distribution is shifted to



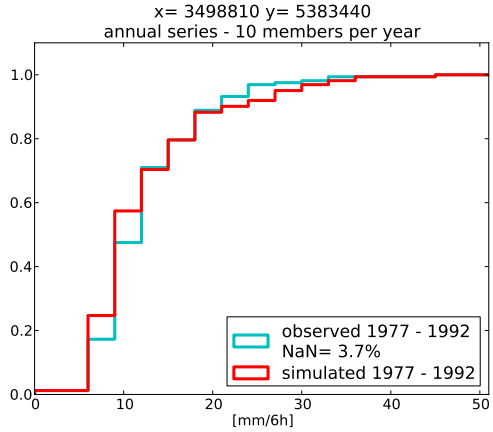
(a) Histogram – 1 h



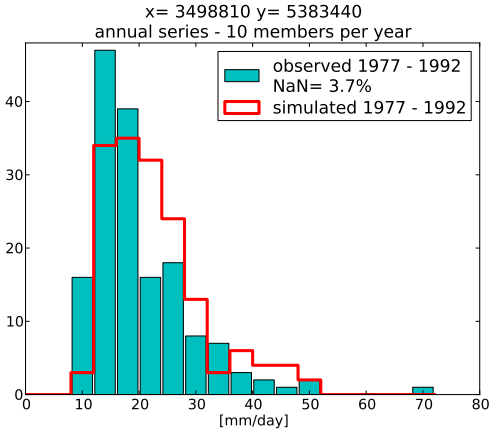
(b) CDF – 1 h



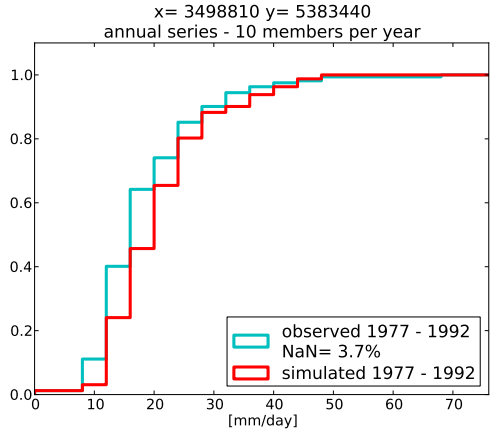
(c) Histogram – 6 h



(d) CDF - 6 h

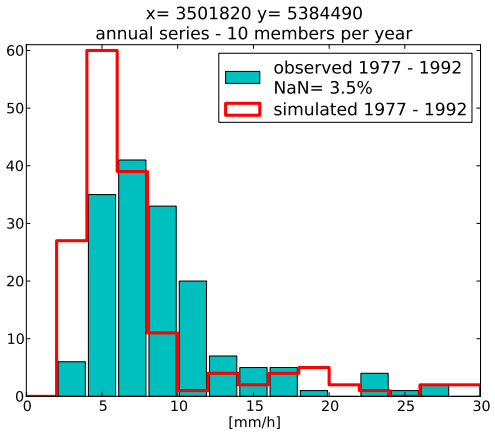


(e) Histogram – 24 h

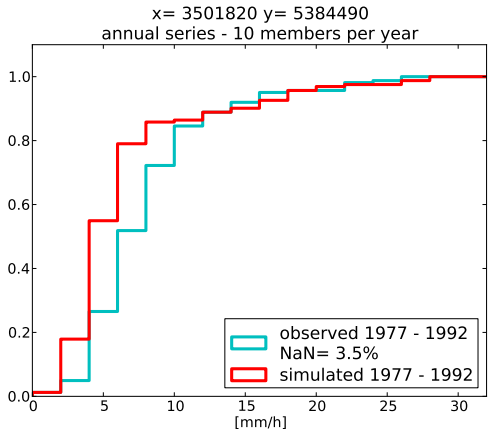


(f) CDF – 24 h

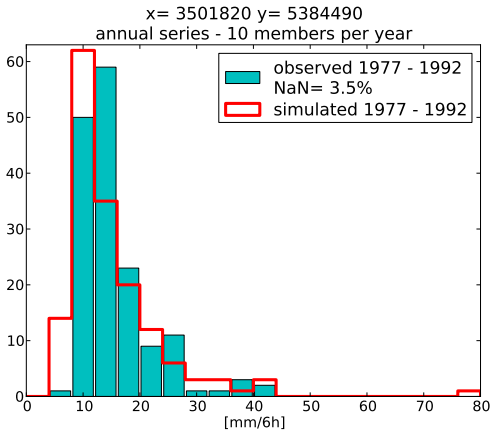
Figure 7.20.: Empirical Distribution of the ten highest 1 h rainfall amounts in the observed and simulated time series in Holzgerlingen, Station 1, as part of a spatial temporal simulation with four simultaneous time series



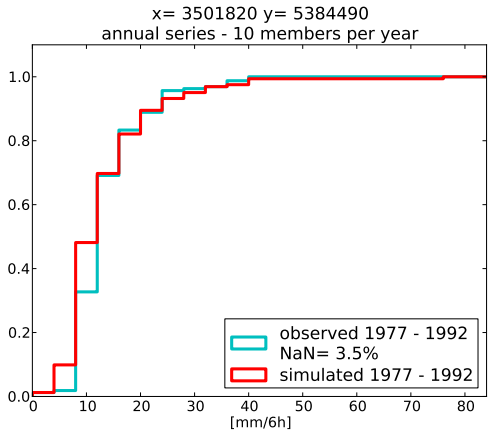
(a) Histogram – 1 h



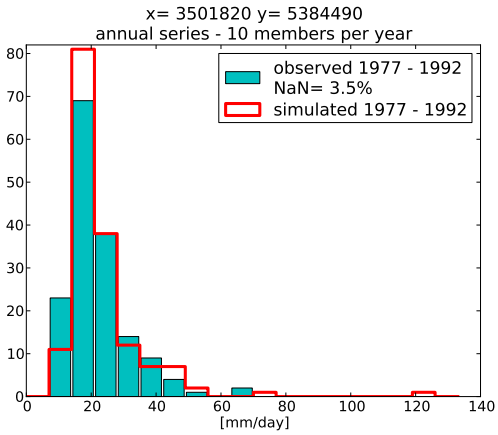
(b) CDF – 1 h



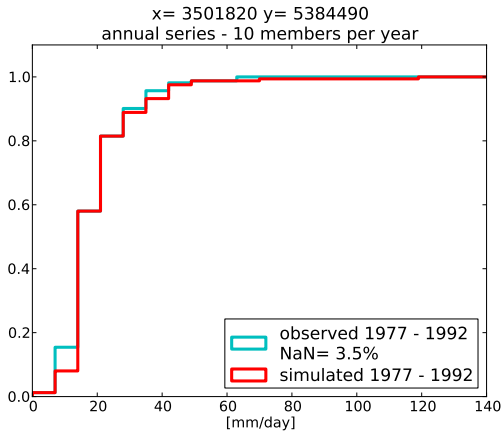
(c) Histogram – 6 h



(d) CDF – 6 h

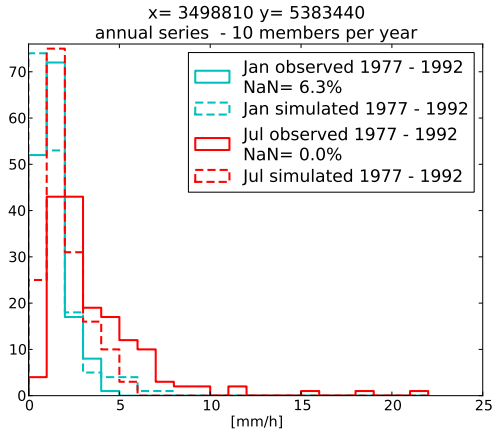


(e) Histogram – 24 h

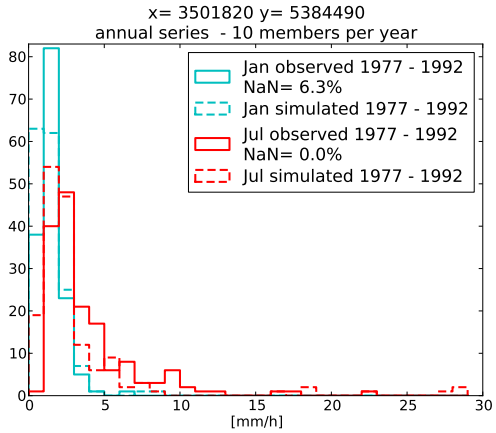


(f) CDF – 24 h

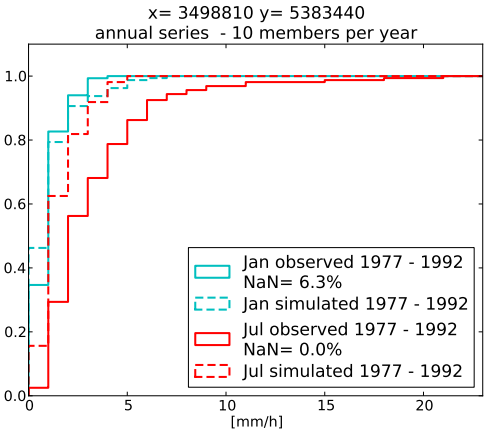
Figure 7.21.: Empirical Distribution of the ten highest 1 h rainfall amounts in the observed and simulated time series in Holzgerlingen, Station 2, as part of a spatial temporal simulation with four simultaneous time series



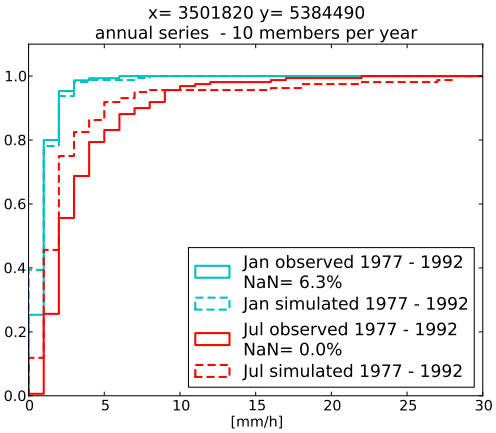
(a) Histogram - Station 1



(b) Histogram - Station 2



(c) CDF - Station 1

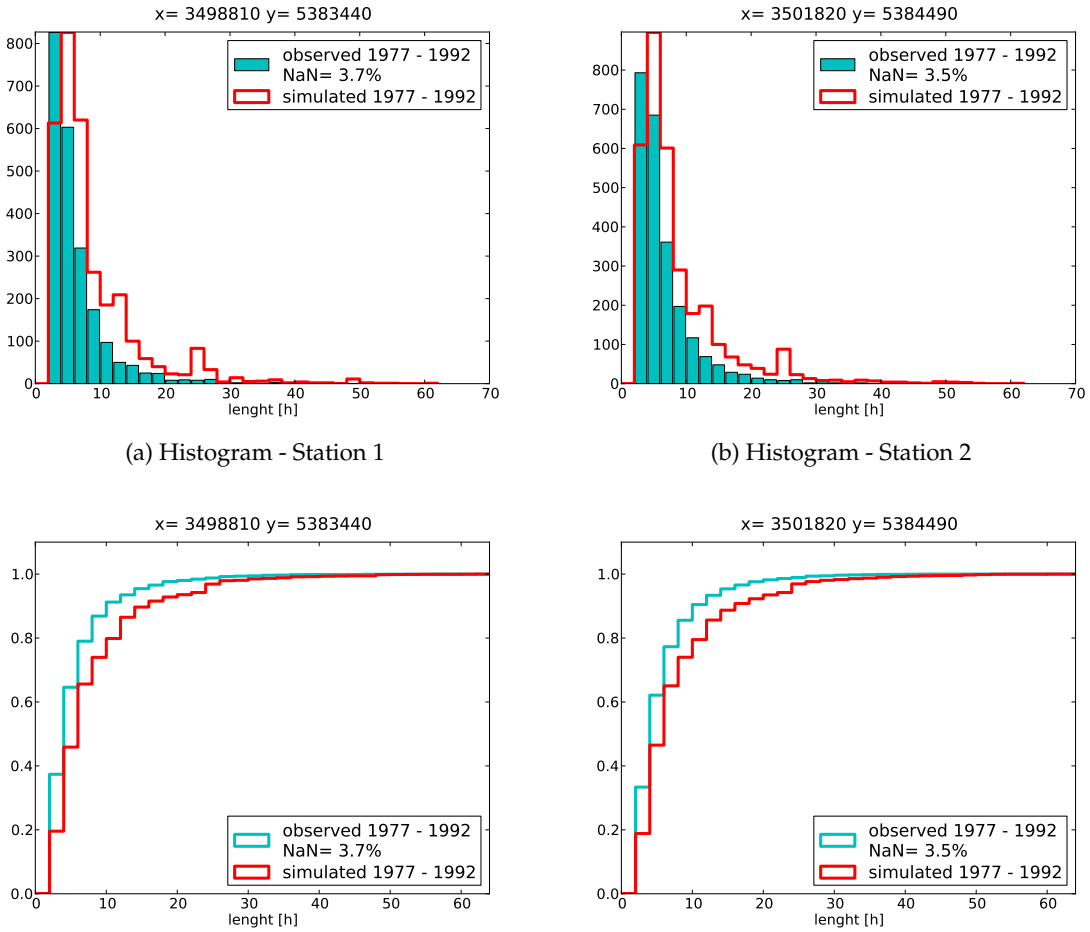


(d) CDF - Station 2

Figure 7.22.: Empirical distribution of the ten highest 1 h rainfall amounts in the observed and simulated time series during January and July in Holzgerlingen as part of a spatial temporal simulation with four simultaneous time series

higher lengths. In average, the 100 longest wet spells are about ten hours longer in the simulation. Extremes go up to more than four days of rainfall without a gap that is longer than one hour – which is not unrealistic but almost twice as high as the observed extremes.

Regarding the dry spell lengths (Fig. 7.25), the overestimation of very short dry spells becomes obvious. The simulated number of short gaps up to a few hours is more than three times as high as the observed number. This is another indication that the occurrence of precipitation in the simulated time series is too scattered. The shape of the distribution, on the other hand, is realistic. The distribution of the longest dry spells (Fig. 7.26) is shifted to the right compared to the observations. The 100^{th} longest dry spell is about eight days instead of nine at Station 1 and seven days instead of eight at Station 2. Overall, the dry spell length are realisticly modeled. The highest difference in modeled and observed exceedance frequency is around 300 h (Fig. 7.26c and Fig. 7.26d).



(a) Histogram - Station 1

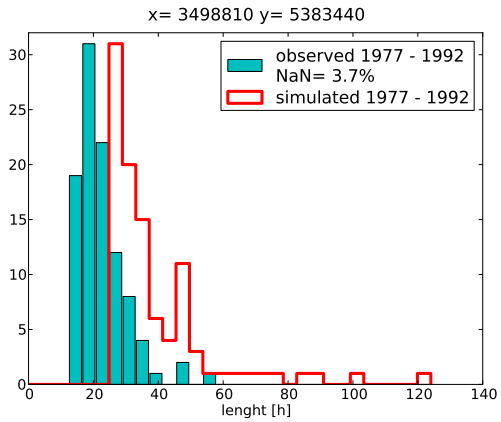
(b) Histogram - Station 2

(c) CDF - Station 1

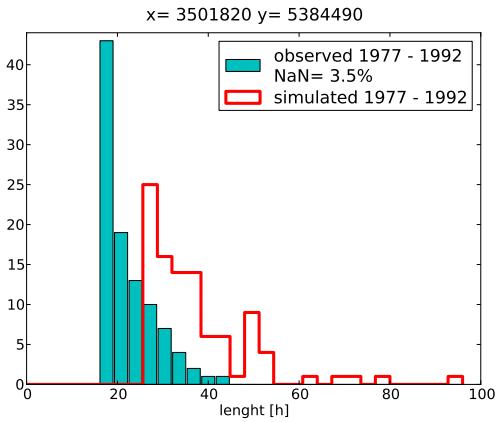
(d) CDF - Station 2

Figure 7.23.: Empirical Distribution of the wet spell length in hourly resolution; time series of Holzgerlingen as part of a spatial temporal simulation with four simultaneous time series. Gaps up to a length of 1 h do not interrupt a wet spell.

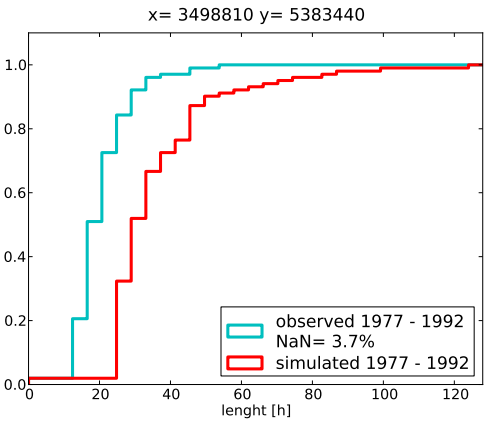
(The length of wet and dry spells is not the same as the definitions of wet and dry spell are not complementary.)



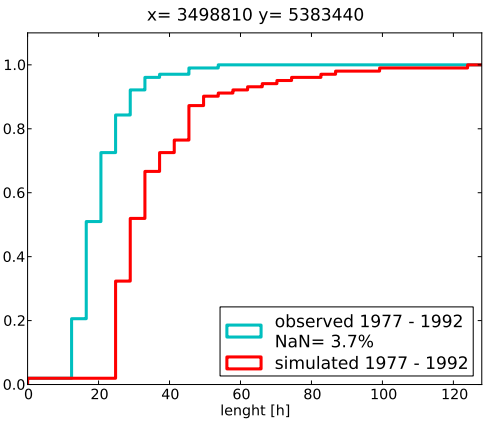
(a) Histogram - Station 1



(b) Histogram - Station 2

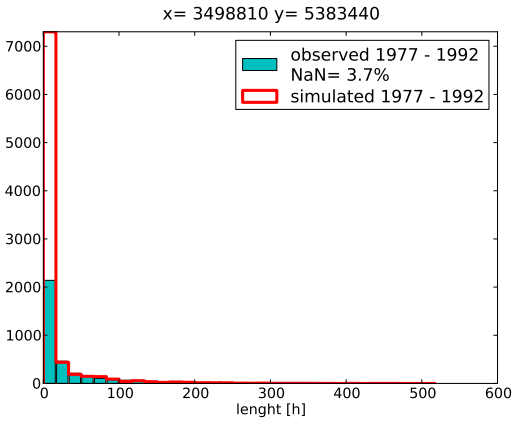


(c) CDF - Station 1

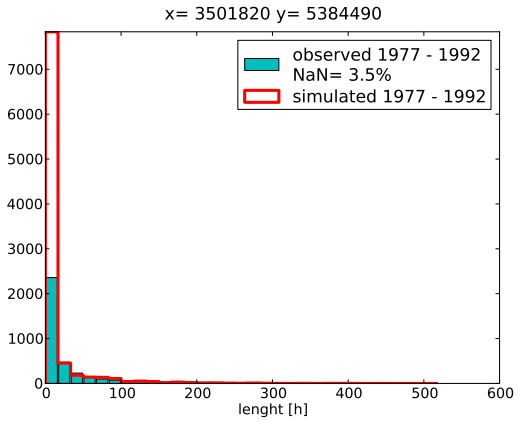


(d) CDF - Station 2

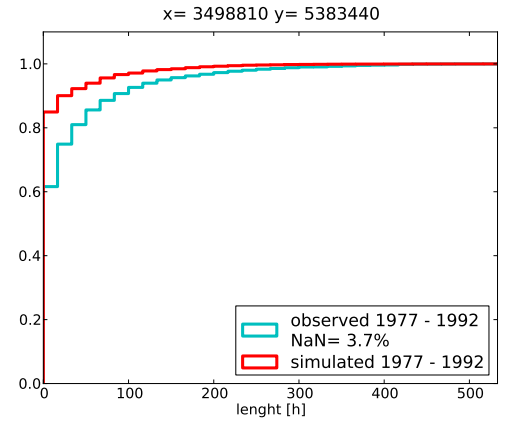
Figure 7.24.: Empirical Distribution of the 100 longest wet spells in hourly resolution; time series of Holzgerlingen as part of a spatial temporal simulation with four simultaneous time series. Gaps up to a length of 1 h do not interrupt a wet spell.



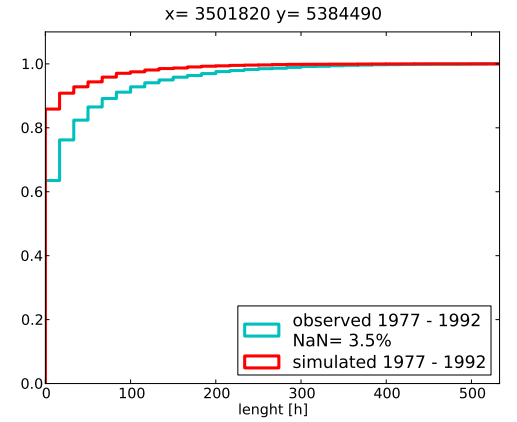
(a) Histogram - Station 1



(b) Histogram - Station 2

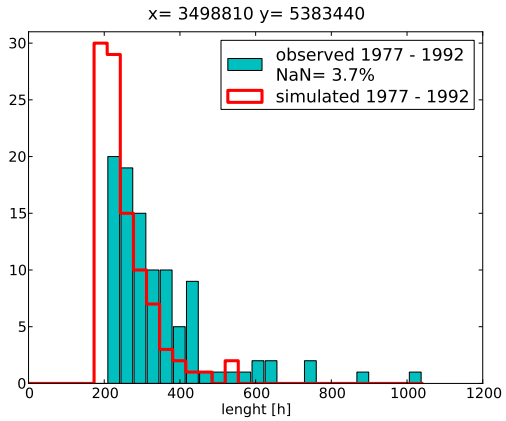


(c) CDF - Station 1

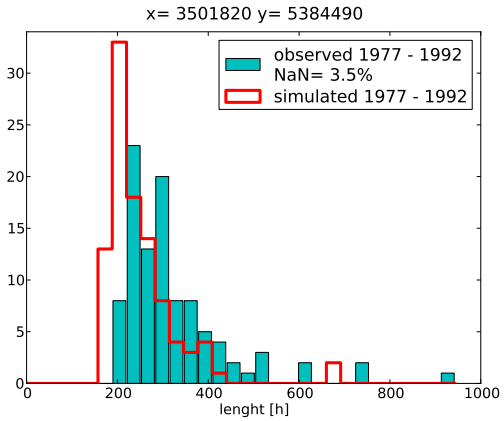


(d) CDF - Station 2

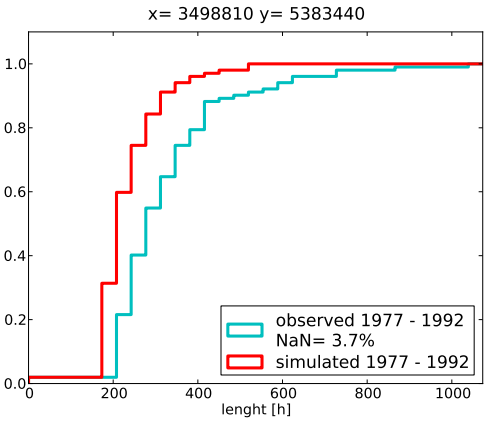
Figure 7.25.: Empirical distribution of the dry spell length in hourly resolution; time series of Holzgerlingen as part of a spatial temporal simulation with four simultaneous time series. Rainfall events with less or equal $0.5\,\mathrm{mm}$ cumulated precipitation sum do not interrupt a dry spell.



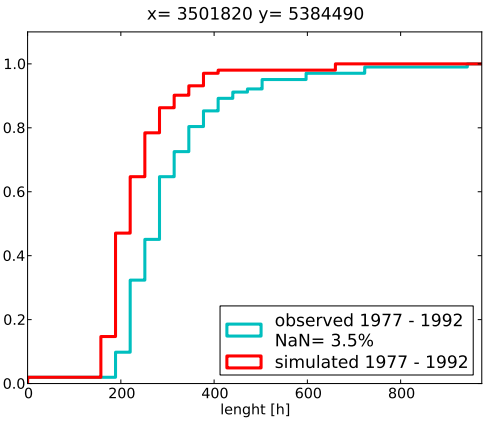
(a) Histogram - Station 1



(b) Histogram - Station 2



(c) CDF - Station 1



(d) CDF - Station 2

Figure 7.26.: Empirical Distribution of the 100 longest dry spells in hourly resolution; time series of Holzgerlingen as part of a spatial temporal simulation with four simultaneous time series. Rainfall events with less or equal 0.5 mm cumulated precipitation sum do not interrupt a dry spell.

7.7.2. Representation of the Spatial Dependence

The two stations in Holzgerlingen that are analyzed in this section are not used in the estimation of the generation parameters describing the spatial dependence structure. The parameter interdependencies (Section 7.5.1) are estimated from the set of daily stations (Fig. 3.4b in Section 3.3). The CP dependent correlation were calculated for the subset of 137 stations with homogeneous time periods (Fig. 3.5 in Section 3.3), in which the stations in Holzgerlingen are not included. The two stations are, therefore, suitable to test if the generation is able to model the spatial dependence in a realistic manner.

Table 7.12.: Correlation of daily values between the four simulated time series

Station	1	2	3	4
1	1	0.7887	0.793	0.8046
2	0.7887	1	0.7798	0.7714
3	0.793	0.7798	1	0.7891
4	0.8046	0.7714	0.7891	1

Table 7.13.: Target values for the simulation of daily correlation

Station	1	2	3	4
1	1	0.903	0.904	0.911
2	0.903	1	0.911	0.913
3	0.904	0.911	1	0.886
4	0.911	0.913	0.886	1

The correlation (Eq. (6.41)) on daily temporal resolution (Table 7.12) is a measure of the spatial dependence between the daily time series at the four simulation locations displayed in Fig. 7.19. As a reverence the target values of daily correlation are given in Fig. 7.13. Measurements are only available at location 1 and 2, thus only the correlation r_{12} can be compared with an observed value. In the measurements $r_{12} = 0.93$ which is very close to the target values of the simulation (Table 7.13). The correlations that are actually modeled, however, are lower than the target values. The optimization is not able to reproduce such high correlations by resampling. The individual daily time series are already too different. The target values are missed by about 12 to 18%.

This is a special problem of very highly related time series where the random drawing of monthly generation parameters (Section 7.6.1) induces too much freedom in the distribution of the daily precipitation values. Table 7.14 and Table 7.15 show the daily correlations for a simulation with ten times higher distances between the four locations (but the same configuration). In this setup the target values for daily correlation are fulfilled for most station combinations. Only the combinations involving station 4 are underestimated by 9 to 18%. However, the correlations remain biased. Deviations between the simulated correlations and the target values always concern underestimations. The correlations are never

overestimated. (Unfortunately, there are no observed stations in in this configuration that could be used to verify the simulated correlations for the higher distances.)

Table 7.14.: Simulated spatial correlation of daily values for a simulation with 10 times higher distances between the station locations.

Station	1	2	3	4
1	1	0.721	0.725	0.639
2	0.721	1	0.716	0.635
3	0.725	0.716	1	0.602
4	0.639	0.635	0.602	1

Table 7.15.: Target values for the simulation of daily correlation for a simulation with 10 time higher distances between the station locations.

Station	1	2	3	4
1	1	0.735	0.724	0.732
2	0.735	1	0.731	0.775
3	0.724	0.731	1	0.664
4	0.732	0.775	0.664	1

Table 7.16.: Correlation of hourly values between the four simulated time series

Station	1	2	3	4
1	1	0.6185	0.6119	0.6209
2	0.6185	1	0.6155	0.6227
3	0.6119	0.6155	1	0.5941
4	0.6209	0.6227	0.5941	1

The spatial correlations on hourly scale are not defined by any generation parameter. They depend on the CP sequence and the CP dependent correlations. If the correlation of the simulated time series is close to the observed correlation, the average over all CP dependent correlations is realistic. The observed correlation between Station 1 and 2 is $r_{12} = 0.6938$, which is about 10% higher than the modelled correlation of $r_{12} = 0.6185$ (Table 7.16). Compared to daily scale, the underestimation is reduced.

Table 7.17 and Table 7.18 present the hourly correlation values with one hour time shift between the time series $r(x_i(t); x_j(t + 1\,\mathrm{h}))$. The line of the table indicates the first station i, the column the second station j where the time series is shifted. Due to the shift, the matrix is not symmetric. Which of the correlations ij and ji is higher, depends on the preferential flow direction. The diagonal elements represent the autocorrelation (Eq. (6.77)) of hourly values with lag $k = 1\,\mathrm{h}$. The autocorrelation at each station is a target value in the Simulated Annealing optimization, therefore, the simulated values are close to the observed ones. The cross correlation between Station 1 and Station 2 on the other hand are too low in the simu-

Table 7.17.: Correlation of hourly values with one hour time shift between the four simulated time series

Station	1	2	3	4
1	0.449	0.393	0.444	0.336
2	0.348	0.461	0.390	0.356
3	0.369	0.342	0.461	0.351
4	0.389	0.434	0.408	0.456

Table 7.18.: Correlation of hourly values with one hour time shift between the observed time series

Station	1	2	3	4
1	0.444	0.456	–	–
2	0.422	0.425	–	–
3	–	–	–	–
4	–	–	–	–

lated time series. The underestimation of about 10% of the order of magnitude is the same as of the correlation without time lag. The anisotropy, however, is well represented. Both, in the observations and the simulations, the correlation is slightly higher, if the time series of Station 2 is shifted by one hour. This means that the simulation assumes a preferential flow direction from west to east, which is realistic for the climatic context of Baden-Württemberg (see Chapter 2.1).

The anisotropy is not very pronounced, neither in the observed nor in the simulated time series because of the short distances between the stations. Assuming purely advective flow, the highest anisotropy in the correlation of the hourly time series of Station 1 and Station 2 would occur if the rainfield traveled at $3\,\mathrm{km}$ per hour, which is very slow. During most of the time the advection velocities are higher and, therefore, the time shift between the stations is shorter. Hence, it cannot be represented to full extend in the time series of hourly resolution.

Table 7.19.: Correlation of hourly values with one hour time shift for a simulation with ten times higher distances between the station locations.

Station	1	2	3	4
1	0.425	0.336	0.259	0.345
2	0.263	0.452	0.251	0.283
3	0.309	0.341	0.440	0.253
4	0.218	0.273	0.207	0.463

In a simulation with the same station configuration as in Fig. 7.19 but with ten times higher distances a time shift of $1\,\mathrm{h}$ between Station 1 and Station 2 corresponds to an advection velocity of $30\,\mathrm{km/h}$, which is closer to typical rain field travel times. Therefore, the anisotropy

in the simulated correlations with one hour time shift is more pronounced in this case (Table 7.19).

Table 7.20.: Correlation between the hourly time series from 1977 to 1992 depending on the CP type

CP	October to March			April to September		
	frequency [d]	simulated	observed	frequency [d]	simulated	observed
1	150	0.586	0.655	168	0.777	0.590
2	174	0.609	0.801	191	0.388	0.559
3	270	0.817	0.751	194	0.701	0.748
4	217	0.719	0.780	85	0.750	0.794
5	260	0.516	0.702	325	0.505	0.643
6	118	0.693	0.724	182	0.702	0.711
7	141	0.631	0.811	193	0.706	0.578
8	164	0.678	0.851	152	0.363	0.298
9	203	0.800	0.824	203	0.641	0.640
10	91	0.786	0.577	134	0.625	0.771
11	217	0.655	0.805	191	0.468	0.754
12	196	0.641	0.734	212	0.572	0.773
13	221	0.877	0.833	180	0.547	0.494
14	109	0.755	0.731	117	0.745	0.693
15	61	0.764	0.793	89	0.698	0.550
16	82	0.749	0.812	85	0.605	0.602
17	157	0.689	0.774	167	0.726	0.741
99	85	0.849	0.824	60	0.629	0.686
avg		0.712	0.766		0.619	0.646
stdev		0.096	0.070		0.124	0.125

For the examination of the CP dependent correlations only the combination of Station 1 and Station 2 is considered where the simulation results can be compared to the observed data (Table 7.20). In average over all CPs the spatial dependence is modeled correctly. There is only a slight bias between the observed and simulated correlations. Between October and March the simulated correlations are higher, between April and September the observed correlations. This means that the difference between the seasons is slightly higher in the simulations. The variability among the CPs, measured by the standard deviation of the correlation values, is modeled in a realistic manner. The standard deviations are close to the observed in both seasons. The correlation between the pairs of observed and simulated CP dependent correlations is $r = 0.48$ in the summer half-year and $r = 0.24$ in the winter half-year. Hence, there is at least a part of the CP dependent signal in the spatial correlation that is correctly modeled. During summer months, the representation is better.

The differences between the simulated CP dependent correlations and the target values for the Simulated Annealing algorithm are low (Table 7.21). Deviations only occur if a CP is

Table 7.21.: Comparison of the simulated CP dependent correlation between the hourly time series from 1977 to 1992 and the target values

CP	October to March		April to September	
	simulated	target value	simulated	target value
1	0.586	0.687	0.777	0.807
2	0.609	0.682	0.388	0.612
3	0.817	0.850	0.701	0.793
4	0.719	0.900	0.750	0.800
5	0.516	0.666	0.505	0.701
6	0.693	0.773	0.702	0.762
7	0.631	0.708	0.706	0.724
8	0.678	0.742	0.363	0.499
9	0.800	0.874	0.641	0.724
10	0.786	0.862	0.625	0.671
11	0.655	0.746	0.468	0.582
12	0.641	0.697	0.572	0.621
13	0.877	0.863	0.547	0.695
14	0.755	0.795	0.745	0.786
15	0.764	0.894	0.698	0.797
16	0.749	0.79	0.605	0.787
17	0.689	0.832	0.726	0.765
99	0.849	0.883	0.629	0.755

rare during one year so that the correlation is calculated from very few precipitation values, especially if it is a very dry CP. Therefore, the correlation of anticyclonic CPs like CP5 and CP17 is the most underestimated. Since the target values are well represented in the simulation, the deviations between observed and simulated time series comes mainly from the estimation of the target values. Station pairs with very short distances are rare. Therefore, the interpolation of the CP dependent correlations (see Section 7.5.2) is linked to high uncertainty.

In Table 7.22 and Table 7.23 the CP dependent correlation with one hour time shift are listed. The time series at the station that is named last is the one that is shifted by one hour. In average the correlations r_{12} from west to east are higher than the correlations r_{21} from east to west. Overall, the representation of the observed correlations in the simulations is better than for the CP dependent correlations without time shifts. In the winter half-year the simulated correlations are 10% to 20% too low, in the summer half-year they are a little higher than the observed correlation. This means that the simulated seasonal effect on the correlation is lower than the observed. The simulated variability in the CP dependent correlation values, measured by the standard deviation, is in the same range as the observed. In average it is about 5% lower. The correlation between the pairs of simulated and observed correlation of all CPs is $r_{12} = 0.29$ and $r_{21} = 0.29$ in the winter half-year and $r_{12} = 0.60$ and $r_{21} = 0.62$ in the summer half-year. The shifted CP dependent correlation between April

Table 7.22.: Correlation between the hourly time series from 1977 to 1992 with one hour time shift, data from October to March

CP	frequency [d]	Station 1 to Station 2 simulated	Station 1 to Station 2 observed	Station 2 to Station 1 simulated	Station 2 to Station 1 observed
1	150	0.328	0.353	0.317	0.366
2	174	0.264	0.524	0.277	0.635
3	270	0.599	0.556	0.524	0.489
4	217	0.522	0.517	0.487	0.468
5	260	0.327	0.635	0.335	0.545
6	118	0.493	0.489	0.377	0.503
7	141	0.466	0.580	0.386	0.610
8	164	0.502	0.694	0.478	0.653
9	203	0.657	0.746	0.598	0.580
10	91	0.572	0.883	0.473	0.334
11	217	0.535	0.561	0.531	0.751
12	196	0.496	0.559	0.478	0.676
13	221	0.693	0.453	0.647	0.688
14	109	0.444	0.685	0.465	0.440
15	61	0.512	0.731	0.472	0.538
16	82	0.573	0.684	0.482	0.621
17	157	0.459	0.628	0.412	0.557
99	85	0.624	0.492	0.602	0.599
	avg	0.497	0.605	0.455	0.556
	stdev	0.113	0.127	0.097	0.114

and September is the spatial information that is represented best by the simulation.

7.8. Conclusion

For the generation of simultaneous time series the resampling approach of NiedSim cannot be applied without modifications. The high number of possible combinations in the temporal sequence of precipitation values at all stations has to be restricted. Therefore, the generation algorithm is changed, passing by an intermediate time series of daily precipitation values. The modifications, however, affect the capacity of the generator to reproduce the statistical characteristics of observed precipitation time series.

In the single site generation (Section 6.6) the distribution of hourly precipitation amounts is directly defined by the parameters of the Weibull Distribution that are estimated for each month. Higher aggregated values have more freedom because they vary with the sequence of the hourly values. In the multi site generation the opposite is the case. The distribution of daily values is directly defined by the theoretical Weibull Distribution while the values on

Table 7.23.: Correlation between the hourly time series from 1977 to 1992 with one hour time shift, data from September to April

CP	frequency [d]	Station 1 to Station 2		Station 2 to Station 1	
		simulated	observed	simulated	observed
1	168	0.330	0.228	0.377	0.371
2	191	0.407	0.363	0.245	0.171
3	194	0.336	0.401	0.437	0.441
4	85	0.551	0.336	0.497	0.436
5	325	0.393	0.321	0.291	0.347
6	182	0.293	0.412	0.276	0.311
7	193	0.439	0.270	0.393	0.272
8	152	0.294	0.410	0.222	0.281
9	203	0.499	0.437	0.395	0.357
10	134	0.576	0.766	0.453	0.295
11	191	0.695	0.503	0.299	0.342
12	212	0.399	0.338	0.331	0.251
13	180	0.196	0.197	0.287	0.272
14	117	0.394	0.257	0.401	0.310
15	89	0.375	0.344	0.376	0.378
16	85	0.495	0.475	0.418	0.484
17	167	0.462	0.394	0.399	0.323
99	60	0.465	0.441	0.461	0.434
avg		0.420	0.379	0.359	0.332
stdev		0.121	0.130	0.078	0.077

shorter aggregation have more freedom in their variations. Therefore, the statistics on daily scale are represented best by the multi site generator.

The distributions of extreme values on shorter than daily temporal aggregations are mainly governed by the scaling exponents b of the statistical moments (Section 7.6.3). In the analyzed example, the shorter time scale is correctly represented by the simulated time series. The modeled distributions of annual series on hourly, three hourly and six hourly aggregation correspond well to the observed distributions.

On the other hand, the necessary modifications in the generation principle reduce the capacity of the generator to model seasonal differences between summer and winter months. The highest intensities in summer months are generally caused by convective events (see Chapter 2). These event are hardly ever longer than a few hours. Therefore, they are less visible in the average and the standard deviation of daily precipitation amounts than in the statistics of hourly precipitation. In the generation the main seasonal signal is induced by the monthwise estimates of the parameters of the Weibull Distribution for the precipitation amounts. For the multi-site generator this information is only available on daily scale. Therefore, the multi-site generator underestimates the seasonal variability in short aggre-

gation intervals. The new generation scheme with intermediate daily time series comes at the price of reduced seasonal variability because the seasonal information on hourly scale cannot be used.

The distribution of *all* hourly values depends not only on the scaling exponents b but is also highly influenced by the estimates of the hourly rainfall probability for the different CP groups (Section 7.6.3). This can be problematic. The relative error in observed precipitation measurements is highest for the smallest values. Therefore, the estimation of observed rainfall frequencies is subject to high measurement uncertainty. In combination with the modeled daily sums this can induce errors in the distribution of hourly precipitation amounts. In the tested examples the modeled hourly rainfall probability is too high (Table 7.10). Since the daily statistics are correctly modeled, the overestimation in rainfall probability causes an underestimation in the average wet hour precipitation amount.

In Section 7.7 it is found that the simulated precipitation is too scattered. There are too many very short rainfall events and at the same time too many short interruptions in longer precipitation events. Probably, the scatter is caused by a contradiction between the scaling exponent and the erroneous estimate of the hourly rainfall probability.

The spatial dependence in the multi site generation is considered at three temporal scales. On monthly scale, the generation parameters of daily rainfall probability, as well as the average and the standard deviation of wet day precipitation are assumed as spatially correlated in the standard normal transformed space. They are linked to the monthly rainfall sums at the simulation locations. Daily scale correlation is used in the optimization of the temporal sequence of daily precipitation values. On the hourly scale, the spatial dependence is represented by the CP dependent correlation of simultaneous and temporally shifted time series.

If the time series are used in sewage system modeling, the distances between the simulation locations are generally short, in the range of a few kilometers. In the data set that is used for the estimation of the CP dependent correlations the distances between stations are generally higher. For short distances the sampling uncertainty in the CP dependent correlation is high. In the analyzed example the average correlation of hourly precipitation values is modeled in a realistic manner. The characteristics of the different CPs, however, are not represented very well in the simulated time series. It can be assumed that this is an effect of the high estimation uncertainty at short distances.

At very short distances the correlations on daily scale are underestimated by simulated time series. If the target values of daily correlation are higher than about $r = 0.9$, they cannot be reached in the optimization because the differences in the initial daily time series cannot be fully compensated by the optimization of the temporal sequences in the time series. It seems that the generation parameter interdependencies do not capture all the spatial information. The individual values of the different time series exhibit too much freedom. Probably, the problem could be attacked by drawing the daily values of all time series simultaneously with correlated random numbers.

8. Some Aspects of Climate Change

Statistical precipitation models all rely on observed precipitation data. The data is required to estimate conceptual model parameters, for example the transition frequencies in a Markov Chain model. The resampling-based generators developed in this work, are linked to observations by the monthly precipitation sums, the precipitation frequencies, the parameters of the distributions for the $1\,\mathrm{h}$ precipitation values and the target statistics of the Simulated Annealing algorithm like the autocorrelation and the scaling exponent.

Data from synthetic precipitation generators is frequently used in planning and design issues. NiedSim, for example, is dedicated to the dimensioning of urban sewage systems. If the design of a hydraulic structure is based on NiedSim data, the precipitation during the aspired life span of the planned structures is represented by the measurement period of the calibration data. It implies the assumption that the climatic conditions in the future are the same as during the calibration period. Under changing climatic conditions this assumption is less and less true, the further the projection looks into the future.

The fourth assessment report of the International Pannel on Climate Change (IPCC) states that the climate is about to change and that this change is most likely due to anthropogenic effects (IPCC, 2007). From 1900 to 2005 the global mean temperature increased by about $0.8^\circ\,\mathrm{C}$ (IPCC, 2007). The increase is spatially heterogeneous. The alps for example warmed up more rapidly than the average. The temperature rise during the 20^{th} century in the alpine region is of $1.3^\circ\,\mathrm{C}$ (Auer et al., 2007). The faster temperature rise is related to an increase in average atmospheric pressure, probably due to a shift of the subtropic high pressure zone to the north (Auer et al., 2007). It is an indication that climate change does not only concern atmospheric temperature but also affects atmospheric circulation.

The global water cycle is sensitive to the average global temperature. Warmer surface temperatures enhance evaporation, especially over the Oceans. Besides, the capacity of air masses to carry moisture is increasing as a function of temperature according to the Clausius Clapeyron Relation. If more water is evaporated and carried, it has to fall down somewhere so that the mass balance is fulfilled. Wentz et al. (2007) state that an increase in atmospheric temperature by $1^\circ\,\mathrm{C}$ leads to an increase in total global precipitation volume of about 7.4% (with a possible error of $\pm 2.6\%$). Held and Soden (2006) analyze the Global Circulation Models that were used in IPCC's fourth assessment report and estimate that the global precipitation increases by 1 - 3% per degree Celcius of global warming.

However, an increase in total precipitation volume does not say anything about the spatial-temporal distribution. If the average yearly precipitation amount increases, but it falls in fewer and more intensive events, with higher seasonality or higher interannual variability,

flood risks and the probability of water stress can rise considerably – even both at the same time.

On the regional scale Hundecha and Bárdossy (2005) found season specific trends in daily precipitation from the German part of the Rhine catchment. The study is based on records of 632 daily precipitation stations. Hundecha and Bárdossy (2005) investigate several precipitation related statistics as the 10% and 90% quantile of daily precipitation sum or the maximum five day total precipitation. During the recording period from 1958 to 2001, most of the precipitation related indices have been increasing. The increase is strongest in winter months from December to February. The summer months, from June to August are the only season where the indices of daily precipitation show a slightly decreasing trend. The trend signal is most pronounced for extreme events exceeding the 90% quantile (Hundecha and Bárdossy, 2005). Except for the summer season, the percentage of total precipitation volume that falls during extreme events is increasing. That means that the highest daily precipitaton sums rise faster than the average and that there is a shift in the distribution.

Haberlandt et al. (2010) conduct similar investigations for 263 daily rain gauges in Lower Saxony. The trend signals in this regions confirm the findings from the Rhine catchment. Daily precipitation is increasing in winter and decreasing in summer. Daily extreme precipitation shows a positive trend. Additionally, Haberlandt et al. (2010) analyze the length of dry spells and find a significant increase in summer months.

Haylock and Goodess (2004) link the trend signal in daily extreme precipitation to a trend in the North Atlantic Oscillation (NAO) which is a measure of the meridional pressure gradient over the Atlantic Ocean and hence of the activity of the mid-latitude west wind zone (see Chapter 2). Thus, the increase in daily precipitation during the second half of the last century is, to a large extend, caused by stronger atmospheric fluxes from the Atlantic Ocean.

Predictions from Global Circulation Models for Central Europe lead to the conclusion that the ongoing changes in precipitation will continue. Based on model runs of ECHAM4, a global circulation model run by the Max Planck Institute in Hamburg, Germany, Arpe and Roeckner (1999) predict a precipitation increase in Middle Europe that is at the expense of precipitation over the Mediterranean countries. The simulations of the AM3P model, run by the Met Office Hadley Centre in the UK, make similar predictions (Kendon et al., 2010). Déqué et al. (2007) summarize the climate projections from 19 different combinations of Global and Regional Circulation models for the time period from 2071 to 2100. For Central Europe the ensemble of all simulations agrees on a precipitation increase in winter and a decrease in summer months.

Christensen and Christensen (2004) focus on the extremes in climate model projections. They show that time series from Regional Circulation Models can have trend signals of different signs in the average and extreme precipitation. In regions where the estival daily precipitation decreases in average, there can be an increasing signal in the extreme events.

The cited trend analyzes are all on daily or monthly scale. For the target scale of the presented stochastic rainfall generator the information on monthly or daily precipitation is of little use. Since extreme events on different scales are caused by different type of events

(Section 4.2) they exhibit different climate change signals. From regional, daily information it cannot be deduced how local extremes in hourly time series react to global warming.

In Central Europe extreme precipitation sums over longer periods of one day or of several days are caused by cyclonic depressions that carry moisture from the Atlantic Ocean or the Mediterranean. Extreme short time intensities, however, mainly occur during convective precipitation events in summer months. There are meso-scale circulation patterns that are in favor of heavy convective precipitation events, but the events also reflect the local heat flux and the availability of moisture.

The ocean and the land masses of the continents react very differently to global warming. Over the ocean a higher sea surface temperature (SST), in any case, leads to higher evaporation. Over the continents the evaporation may be limited by the available soil moisture, especially during summer. Thus, the portion of incoming solar radiation that is transformed into latent heat is smaller over land. The atmosphere heats up faster than over the ocean. In summer months, the strong increase in storage capacity of the heated air masses over the continent can overcompensate the increase of evaporation over the ocean. As a result the relative humidity may be lower, even if the absolute volume of atmospheric water is higher (Kendon et al., 2010).

If the relative humidity of the atmosphere is lower, fewer occurrence of precipitation can be expected. If at the same time the absolute water content is higher, the intensity of precipitation is expected to increase (Trenberth, 1999). As a result, more water can rain down in shorter time during fewer events. The effect, however, is only detectable on sub daily temporal scale. For this reason, Trenberth (1999) states that "more datasets of hourly precipitation rates and frequency need to be developed and analyzed" for the evaluation of global climate change.

Global Circulation Models are spatially too coarse to represent the regional processes leading to extreme precipitation (Christensen and Christensen, 2004) and the temporal resolution is not able to image the increasing intensity of events. The question is, how the available GCM information can anyhow be used to estimate trends on sub-daily scale and which other sources of information are available.

8.1. Trend Signals in Observed Precipitation Time Series

One possibility to predict future climate conditions is the extrapolation of observed trend signals. Trend estimation is based on the assumption that an observed trend will continue in the same way in the future. It is a purely statistical method that does not use any information of atmospheric physics, but it assumes some sort of inertness in the physical processes governing the trend.

For this assumption, statistical trend estimation does not allow to look far into the future. A time horizon of the year 2100, as in Global Circulation Models, is not possible. On the other hand, it allows to benefit from the full detailedness of observed precipitation data that future predictions from GCMs or RCMs cannot deliver.

As there is indication that the climate trend signal in precipitation is scale dependent, the focus in the following investigations is on hourly precipitation data.

8.1.1. Regional Trend in Extreme Precipitation

If only one representative station was picked out for trend analysis, the results would be highly influenced by local effects and sampling uncertainty. Since precipitation is highly variable in space (Section 4.1), it is possible that the trend at one station is opposite to the regional trend signal. The use of just one station might lead to the wrong conclusion.

To enable the use of data from all available rain gauges in the same analysis, the measurements have to be made comparable. One possibility is to reference all values on the station average of the respective statistics.

$$\bar{H}_{max} = \frac{1}{n_{yr}} \sum_{i=1}^{n_{yr}} H_{max}(i) \tag{8.1}$$

$$v_{max}(i) = \frac{H_{max}(i)}{\bar{H}_{max}} \tag{8.2}$$

$$\tag{8.3}$$

where	$H_{max}(i)$	maximum precipitation observed in year i
	$\bar{H}_{max}$	average yearly maximum precipitation
	$v_{max}(i)$	normalized maximum of year i
	n_{yr}	number of years in reference period

The normalized maxima v_{max} of all stations can be joined together in the same statistics. The average over all years is one. If there is for example a positive trend in the maxima, the average will be below one in the beginning of the analyzed period and above one towards the end. The trend signal can be estimated by linear regression.

$$\hat{v}_{max}(i = x) = a_0 + b_0 x \tag{8.4}$$

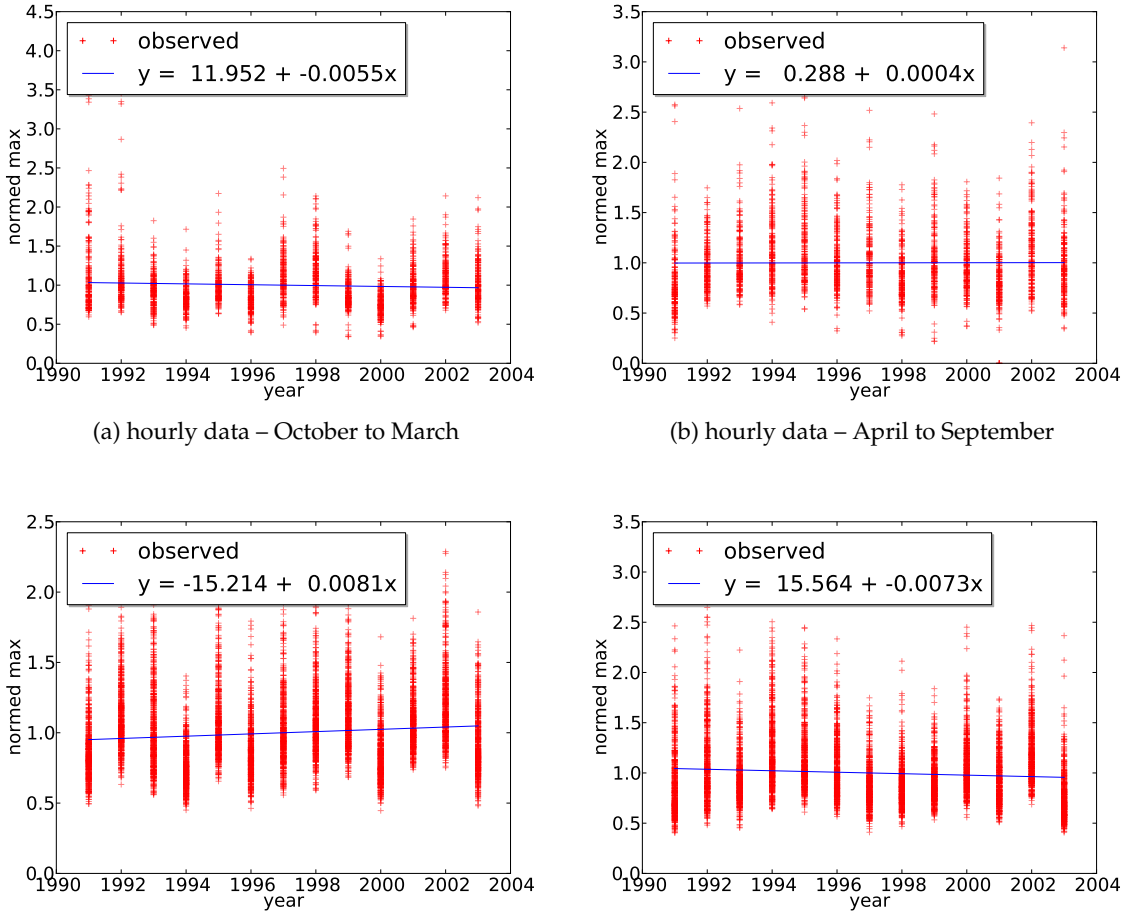
where	$\hat{v}_{max}(i = x)$	best guess for the normalized annual maximum in year x
	a_0	intersect of the linear regression function
	b_0	slope of the linear regression function

The parameters of the regression are estimated by minimizing the least square error between the observed values v_{max} and the best regression estimate $\hat{v}_{max}$.

The normalization is only valid if the used average maximum H_{max} at all stations refers to the same time period. Therefore, not all stations and not the full time period of the data set (Section 3.3) can be used for the regional trend analysis. It is restricted to a homogeneous set in which all stations are available in every year. As a trade off between the longest possible time period and the highest possible number of available stations the time period from 1991

to 2003 was used. In this period 40 stations with hourly time resolution and 104 stations with daily resolution are available.

For a broader data basis the analysis is repeated twice, at first with the annual maximum and then again with the five highest values per year. To account for seasonal effects, it was performed with censored data from two seasons, October to March and April to September.



(a) hourly data – October to March

(b) hourly data – April to September

(c) daily data – October to March

(d) daily data – April to September

Figure 8.1.: Regional trend in the five highest hourly and daily precipitation sums at 40 hourly and 104 daily rain gauges in Baden-Württemberg from 1991 to 2003

Due to the broader data basis, the trend signal in the data of the five highest values are seen as more reliable. They are presented in Fig. 8.1. (The results for the annual maximum are comparable.) In both seasons the extreme values do not change much over the years. There is a slightly decreasing trend in the hourly extremes from October to March (Fig. 8.1a) and in the daily extreme values from April to September. The daily extremes in the winter half-year are increasing.

The significance of the trend signal is tested by a non-parametric Monte Carlo method. If the data exhibits a significant temporal trend, it should be destroyed, if the data from the

different years is randomly mixed. If the trend signal is just a sampling effect due to inter-annual variability, it is possible that the random trend signal is more pronounced after the mixing.

The portion of linear regression slopes calculated from mixed data that have a steeper slope (of the same sign) than the regression calculated from the observed data can be seen as the level of significance in the trend signal. The only restriction of this significance testing is that the number of trials in the mixing should be much lower than the number of possible combinations. With 13 years of available data and hence $13!$ different combinations, which is more than six billion, the Monte Carlo method is performed with 10000 trials.

The significance test is performed as a single sided test. The tested hypothesis H_0 and H_1 depend on the slope b_0 of regression line in the observed data:

- H_0: $\beta_0 \leq 0$
  H_1: $\beta_0 > 0$ if $b_0 > 0$
- H_0: $\beta_0 \geq 0$
  H_1: $\beta_0 < 0$ if $b_0 < 0$

where β_0 is the "true" slope wich would be calculated if the regression was based on an infinite number of data pairs. The significance α is the probability of making a mistake when H_0 is rejected in favor of H_1. It can be approximated by the fraction of random slopes b that are steeper than b_0.

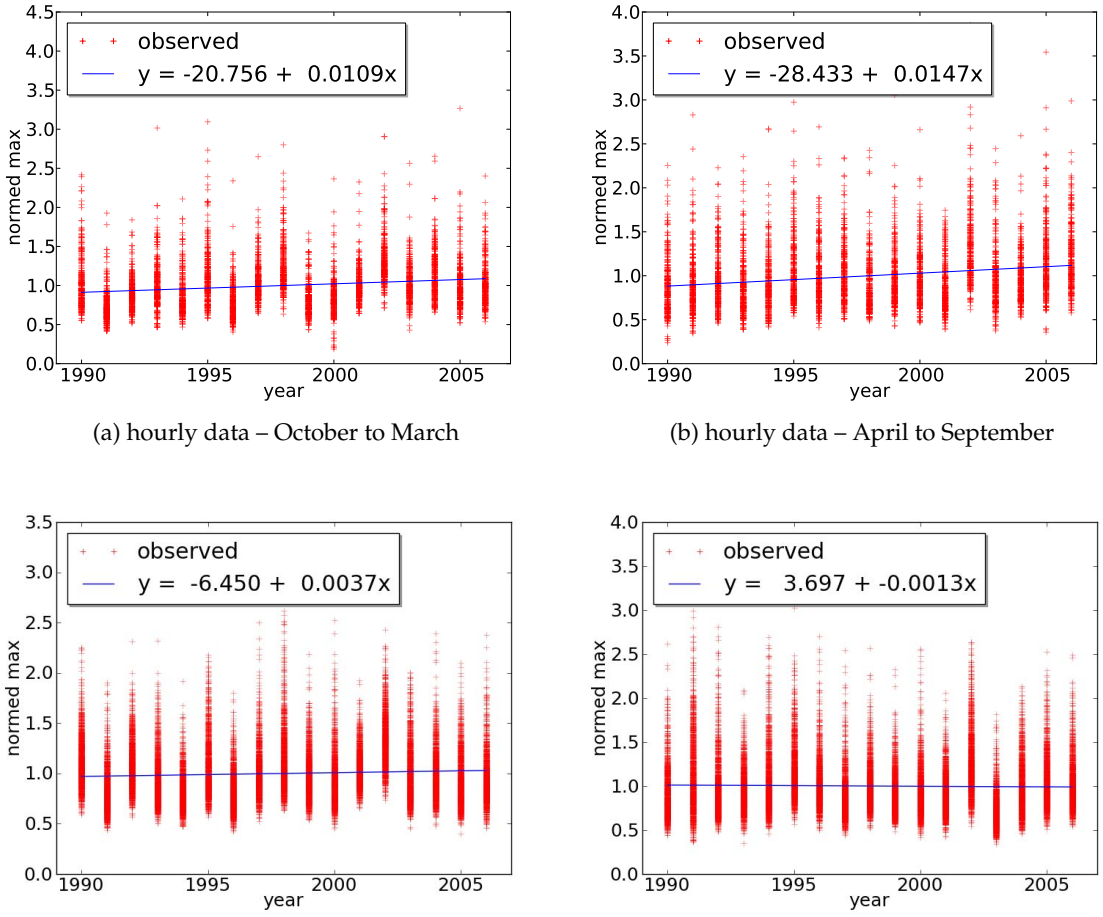
Table 8.1.: Significance test by Monte Carlo simulation of the regional trend signals in Baden-Württemberg from 1991 to 2003.

	Hourly Data		Daily Data	
	October to March	April to September	October to March	April to September
b_0	-0.0132	0.0004	0.0081	-0.0073
$b_{rand} > b_0$	8283	4855	2147	7434
α	17.17%	48.55%	21.47%	25.66%

The results of the significance testing in Table 8.1 show that none of the trend signals is highly significant. For the most pronounced trend in the hourly data of the winter half-year, one out of six regression lines calculated from randomly mixed data have a steeper slope.

The data period reaching over thirteen years is probably too short for a significant trend detection. As a second trial data from Bavaria is used, which is the direct eastern neighbor of the federal state of Baden-Württemberg. In Bavaria 54 hourly stations and 293 daily stations are available in the time period from 1990 to 2006. If the longer time series in Bavaria exhibit significant trend signals, the same could be expected for Baden-Württemberg if longer time series were available.

In Bavaria the daily extreme values between 1990 and 2006 are constant (Fig. 8.2). The hourly extreme values. on the other hand, are increasing in both seasons. The results presented in



(a) hourly data – October to March

(b) hourly data – April to September

(c) daily data – October to March

(d) daily data – April to September

Figure 8.2.: Regional trend in the five highest hourly and daily precipitation sum at 40 hourly and 104 daily rain gauges in Bayern from 1990 to 2006

Table 8.2.: Significance test by Monte Carlo simulation of the regional trend signals in Bavaria from 1990 to 2006.

	Hourly Data		Daily Data	
	October to March	April to September	October to March	April to September
b_0	0.0109	0.0147	0.0037	-0.0013
$b_{rand} > b_0$	940	52	2868	5842
α	9.40%	0.52%	28.68%	41.58%

Fig. 8.2 refer to the data set of five extremes per year. (For the annual series of only one extreme, the trends go in the same direction but are less pronounced.) The trend towards higher hourly values in the summer season has a high significance (Table 8.2). Only 52 out

of the 10000 randomly mixed data sets exhibit a higher slope. In winter the significance is still over 10%

Table 8.3.: Significance test by Monte Carlo simulation of the regional trend signals in Bavaria from 1991 to 2003.

	Hourly Data		Daily Data	
	October to March	April to September	October to March	April to September
b_0	0.0062	0.0129	0.0057	-0.0096
$b_{rand} > b_0$	2855	769	3476	8202
α	28.55%	7.69%	34.76%	17.98%

To judge the effect of limited data, the analysis for Bavaria is repeated with a data set that is restricted to the time interval from 1991 to 2003 (Table 8.3). Cutting down the data period from 17 to 13 years reduces the significance considerably. The slope of the regression lines is lower too. In winter months it reduces to one half. The overall tendency however is the same in the shorter data set. The hourly extreme values are still increasing with a stronger increase between April and September.

For the same time period the trend signal in Bavaria and Baden-Württemberg show some deviations. Between October and March hourly extreme precipitation is decreasing in Baden-Württemberg but increasing further east in Bavaria. The significance of the trends is however low. It is possible that the difference is only due to sampling effects and not to regional differences in the climatic conditions.

For a final judgment whether there is a trend signal in hourly precipitation extremes, the time period of 17 years is too short. In such a short time period climatic cycles with several year long periods, as for example the “El-Nino” – “La-Ninja” cycle (Weischet and Endlicher, 2008), do not average out and can lead to misinterpretations. The increasing branch of such a periodic signal might be mistaken for an increasing trend. On the other hand, the findings from Bavaria are supported by the results of Trenberth (1999) who states that global atmospheric warming will lead to less but more intensive precipitation events due to the higher moisture capacity of the atmosphere. Thus, there is a physically meaningful interpretation in the observed increase of hourly precipitation extremes.

8.1.2. Change in Scaling

Normalization by the scaling is another possibility to make the precipitation data from different stations comparable. If the hourly extreme values are increasing while the daily extremes are constant, as it is found for the Bavarian data, the scaling of precipitation must change. If it is presumed that an increasing trend in extremes is present in Baden-Württemberg but could not be measured due to lack of data, the related change in scaling might be detectable. Scaling, in this case, means the statistical properties of precipitation

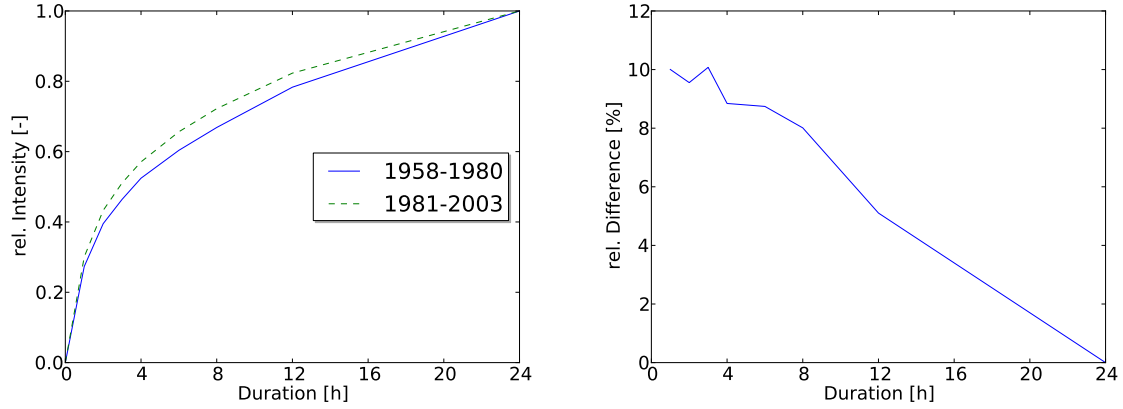
data aggregated over different time intervals (for example, hourly data, daily data, see Section 4.2).

$$\bar{H}_{max}\left(\Delta t\right) = \frac{1}{n_{yr}} \sum_{i=1}^{n_{yr}} H_{max}\left(i, \Delta t\right) \tag{8.5}$$

$$\bar{v}_{max}\left(\Delta t\right) = \frac{\bar{H}_{max}\left(\Delta t\right)}{\bar{H}\left(24\mathrm{h}\right)} \tag{8.6}$$

where $H_{max}\left(\Delta t\right)$ average yearly maximum at aggregation level Δt
$\bar{H}_{max}\left(24\mathrm{h}\right)$ average daily maximum

As in the last subsection, the analysis is based on the five highest extremes of each year. The normed average maximum $\bar{v}_{max}$ at each station is between 0 and 1. It is the portion of the daily extreme that can fall in a shorter aggregation interval Δt. If the scaling behavior is changing over time, v_{max} of all aggregations $\Delta t < 24\mathrm{h}$ is changing too. Trends in daily precipitation, on the other hand, are eliminated by the normalization. Since it requires data from several years for a reliable calculation of the average maximum on every aggregation level, it is not possible to make a trend estimation on yearly basis. Instead, the available data was divided into two periods and the scaling characteristics were compared. Fig. 8.3 shows the average of the normed extremes v_{max} over all rain 292 hourly rain gauges in the data set (Section 3.3) from 1958 to 1980 and from 1981 to 2003.



(a) average scaling

(b) rel. difference 1981-2003 to 1958-1980

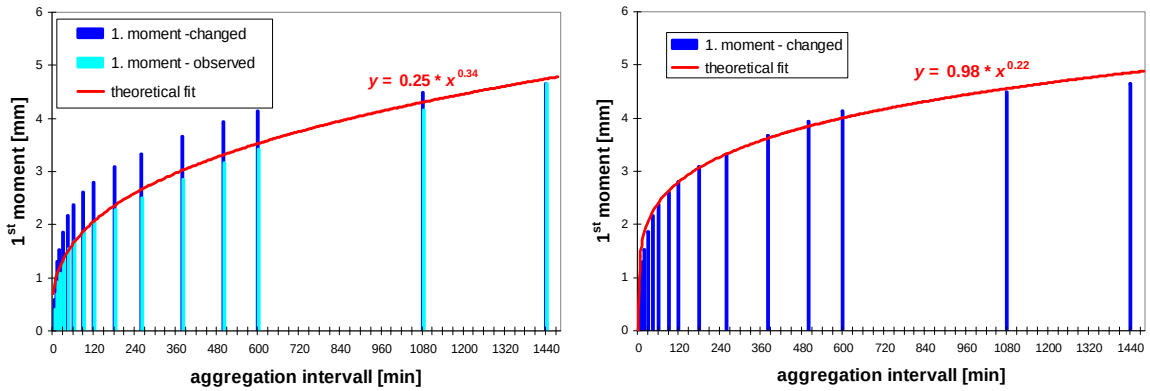
Figure 8.3.: Average of the five highest precipitation events per year depending on the temporal aggregation normed on the daily extreme. Average over all stations in the data set for two time periods.

In average over all stations, about 50% of the daily extreme precipitation can fall within 4 hours, 20% wihin one hour. Comparing the data from the two sub periods, it is noticeable that the dashed line from the later data period is always higher. The right plot in Fig. 8.3

indicates the relative changes at each aggregation interval between the two periods. It is always positive and the relative increase is highest in the shortest aggregation intervals.

The results, again, support the thesis of Trenberth (1999) that the precipitation over short time periods is intensifying with global warming. Compared to the daily extremes, the short time intervals up to six hours increased by about 10% between the two time periods. An increase of 10% does not seem dramatic during a measurement period of 46 years, but one should keep in mind that the trend was revealed by coarsely dividing the data set in two halfs of 23 years. On this coarse scale much of the actual trend will be averaged out within the subperiod. A further division into more intervals, however, is not possible due to the short average time series length of the precipitation stations. If the data is divided any further, the calculation of the average extremes will become uncertain.

The changes in scaling, however, cannot only be detected in extreme events. One of the target statistics of the precipitation generators developed in this study is the exponent of a non-linear regression function through the statistical moments on different aggregation intervals (see Eq. (6.76) in Section 6.6.2.1). Therefore, it is analyzed how the scaling exponent b in Eq. (6.75) is changing between the two time periods from 1958 to 1980 and from 1981 to 2003.



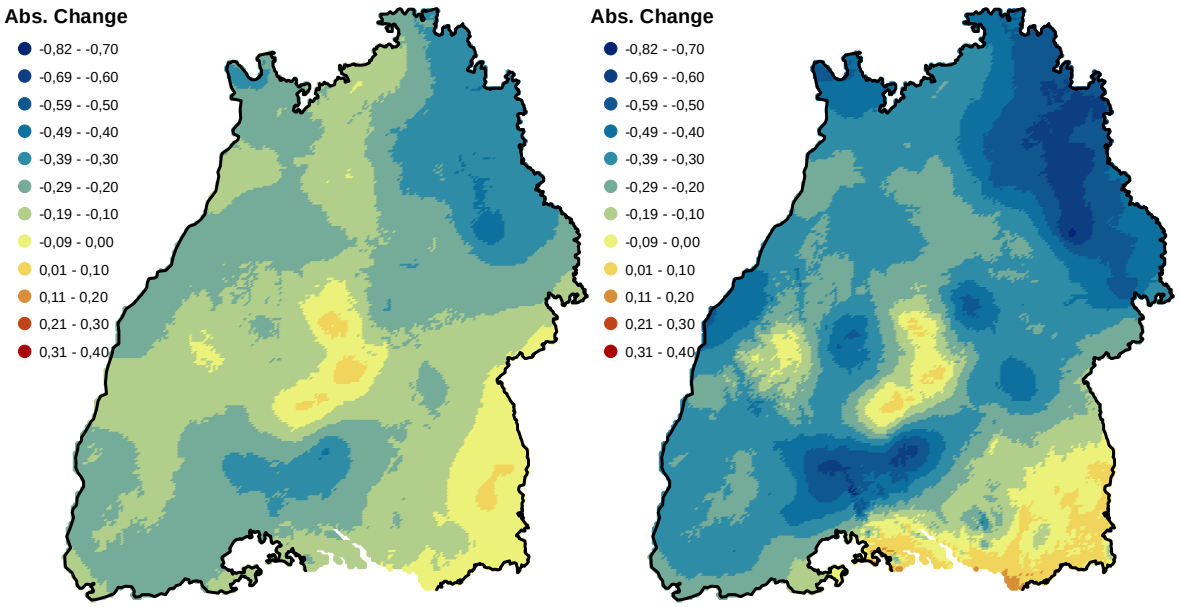
(a) original scaling exponent

(b) changed scaling exponent

Figure 8.4.: Principle of the change in scaling exponent.
By increasing the averages over shorter aggregation intervals, the scaling exponent b is getting smaller.

Fig. 8.4 illustrates how the scaling exponent b for the first moment reacts on a scale change. The light blue bars indicate the original measured average precipitation sum over different aggregation intervals. For the blue bars, the averages of all aggregations shorter than $24\,\mathrm{h}$ were artificially increased. The relative change is higher in the shorter aggregations. The equation in Fig. 8.4a represents the regression function trough the original data, the equation in Fig. 8.4b the regression function through the changed data set. The scale exponent of the regression drops from $b = 0.34$ to $b = 0.22$. The scale exponent is always a positive number between 0 and 1. If $b = 0$, the regression function would be horizontal which would

mean that all daily precipitation would fall within one hour of the day. If $b = 1$, it would permanently rain with constant intensity.



(a) October to March (b) April to September

Figure 8.5.: Absolute change in scaling exponent b between the two time periods from 1958 to 1980 and from 1981 to 2003

Fig. 8.5 illustrates the changes in scale exponent between the two time periods. Colors from dark blue to green-yellow indicate a decrease of the scale exponent b and thus an increase of hourly precipitation values relative to daily values. The colors from orange to red indicate an increase of b. The spatial detailed maps are calculated by the difference of two $1\,\mathrm{km} \times 1\,\mathrm{km}$ raster fields of interpolated scaling exponent values. The intermediate step of interpolation is necessary since not all stations are available in both time periods. Therefore, some of the spatial characteristics can be artifacts due to interpolation uncertainty.

Nevertheless, hardly all of Baden-Württemberg shows a decrease in scale exponent and thus a relative increase in short term intensities. The only exceptions are in the vicinity of some precipitation station in the upper Neckar valley and, during the summer months, the southernmost part of Baden-Württemberg in the vicinity of the Alps (Fig. 8.5b). Generally, the decrease is more pronounced in summer months, which is in accordance with the physical explication by Trenberth (1999).

Both regions with positive scaling exponent exhibit a climatic particularity. The Upper Neckar valley is the driest region in the study area that is the most shaded by the Black Forest. Perhaps heavy convective precipitation events with high intensities are especially rare in this part of Baden-Württemberg. The regions close to the Alps receive the highest precipitation sums when the flux comes from northwest to north due to orographic effects on the high mountains further south. In the rest of Baden-Württemberg, the highest sums are

received from western to south western direction. It is possible that generally cool northern air masses react the most on the increase in evaporation caused by global warming. Most likely, the positive scaling exponent in the pre-alpine region is not caused by decreasing short term intensities but because the intensification is overcompensated by an increase in daily precipitation sum.

8.2. Global and Regional Circulation Models

Global Circulation Models are physically based models that solve equations describing atmospheric processes as for example energy and mass balance. Most models work with discretized degree grid cells in several vertical layers. The physical equations are solved for the interfaces of the grid boxes. There are several GCM by different climatic research groups. Examples are the ECHAM model family developed by the Max Planck Institute for Meteorology in Hamburg, Germany (Roeckner et al., 2003) or the models of the Met Office Hadley Center in the UK (Johns et al., 2003).

In principle, GCM are weather forecast models for the whole globe with a very long temporal horizon. Due to the heavy computational effort, the resolution is limited, spatial to about one to two degrees of latitude and longitude (which is equivalent a few hundred kilometers at the equator), temporally to six hour intervals.

Although the models are physically based, there are quantities that have to be represented by conceptual parameters because they vary on smaller scales than the GCM resolution. Examples are the leaf area index, representing the seasonally changing albedo due to crowing plants or the reflection of solar radiation by clouds that is modeled by a statistical cloud cover scheme in the ECHAM5 model (Roeckner et al., 2003).

Global circulation models can be divided into two groups: atmospheric models and coupled atmospheric oceanic models. In atmospheric models the oceans are represented as huge water reservoirs that interact with the atmosphere on the surface. Lateral energy and mass fluxes within the oceans are ignored. Coupled models explicitly take the ocean circulation into account. This increases considerable the number of grid points. HadCM3, the coupled model of the Hadley Center (Johns et al., 2003), has six ocean grid points per atmospheric grid point in the horizontal. To compensate for the higher computational demand that is necessary for the better process representation, the spatial resolution in coupled models is lower.

There are two possibilities to refine the resolution on certain target areas, either by using models with variable grid sizes or nesting of Regional Climate Models (RCM). Regional models have grid lengths of several tens of kilometers and more complex representation of the physical processes than GCMs. The higher spatial resolution requires a limitation of the modeled area. Typically, RCMs are set up for one continent or one part of a continent. The RCMs are nested in the GCM by the boundary conditions. The RCMs receive values for all state variables, for example pressure, energy or moisture flux at the boundary grid cells from the governing GCM. An example for an RCM is the REMO model by the Max Planck Institute (Elizalde et al., 2011) that has been set up for Europe and West Africa among others.

8.2.1. Emission Scenarios

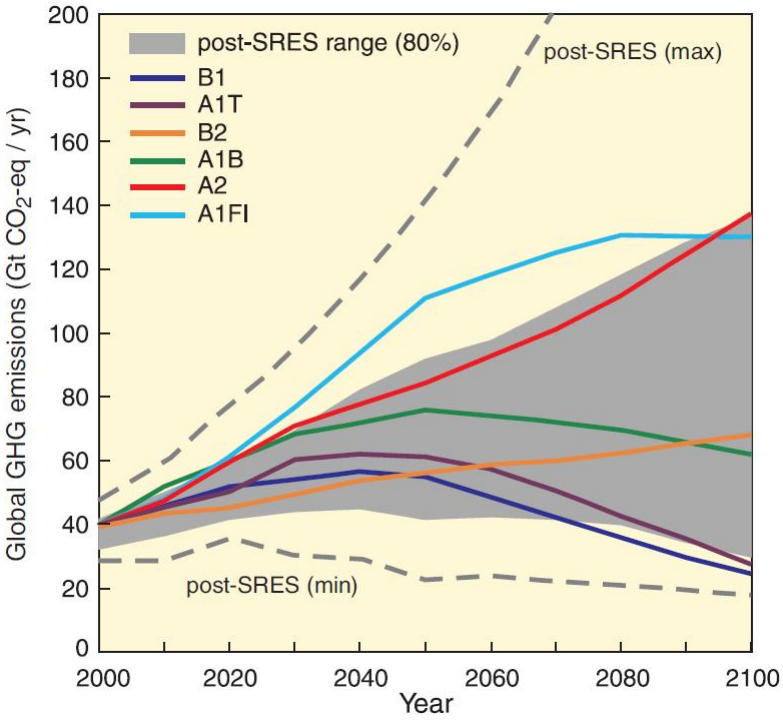


Figure 8.6.: Scenarios for Green House Gas (GHG) emissions from 2000 to 2100 (Nakicenovic and Swart, 2000)

Climatic predictions from Global and Regional Circulation Models depend on state variables that are external to atmospheric simulation. An example is the albedo of land surface that depends mostly on human agricultural activity or, the most important, the concentration of CO_2 in the atmosphere. Before a global circulation model can be run, the future values of the external variables have to be predicted.

The prediction of CO_2 is based on an ensemble of scenarios. Each scenario is defined by a story line that describes a way in which the Earth will probably evolve. The story line includes the evolution of the Earth's population, as well as prognosis about technical progress and how the energy demand of the world population will be met in the Future.

In 2000, the IPCC published a Special Report on Emission Scenarios (SRES) (Nakicenovic and Swart, 2000). The SRES scenarios are composed of four different scenario families (A1, A2, B1, B2) with corresponding story lines. These are subdivided into scenario groups with six marker scenarios (A1FI, A1T, A1B, A2, B1, B2). The first letter indicates if the overall orientation on economic growth (A) or environmental issues (B), the following one digit number indicates if the economic development on the planet is homogeneous and globally linked (1) or regionally diverse (2). The last letters in the "A" scenarios indicates how the energy demand is satisfied, e.g. by fossil energy sources (FI) or a mix of fossil and renewable sources (B).

The ensenmble of scenarios claims to cover the whole range of possible developments in the CO_2 concentration. Among the marker scenarios, B1 is the most "optimistic", with the lowest CO_2 concentration, A2FI the most "pessimistic".

8.2.2. Model Errors in Global and Regional Circulation Models

To cover the full range of possible future climate conditions, GCM models are run for different CO_2 emission scenarios. Additionally, control runs are performed for past time climate, e. g. the 20^{th} century runs of the ECHAM and Hadley Center models. In the control runs the models are forced with historic initial conditions and the observed CO_2 concentration time series. The resulting simulation of 20^{th} century climate can be compared to observed measurements of climatic variables.

GCM and RCM can only give a simplified representation of the atmospheric processes. Besides, the limited resolution leads to model errors. To estimate the model errors, climate model runs are generally compared to reanalysis data, which are gridded precipitation fields in similar spatial-temporal resolution deduced from a worldwide set of observed precipitation records (see also Section 7.5.2).

Habemann et al. (2006) examines the global water cycle simulated by ECHAM5. The comparison of monthly precipitation sums to GPCP reanalysis data from the Max Planck Institute for the control period from 1979 to 1999 reveals some deviations. Seasonal differences are generally overemphasized by the GCM. Over the Oceans an during Monsoon season in Asia the precipitation values are generally too high.

Roeckner et al. (2006) shows that the performance depends highly on the spatial resolution of the model. The simulated precipitation react especially sensitive to the vertical resolution, which means the number of atmospheric levels present in the model. The fact that the spatial resolution has a high influence on simulated precipitation values leads to the conclusion that GCMs induce model errors by the scale difference of GCM and real world precipitation.

Kendon et al. (2010) check for the reliability of precipitation predictions from the Hadley Center Model HadAM3P in Europe by separating different mechanism influencing future climate. For example, the model is run with boundary conditions of the control period except for the soil moisture that is taken from a future scenario run. In this way, she can judge if the different mechanism induce trends in the same direction, which makes estimated trends reliable, or if they work against each other.

Analyzing several precipitation related indices as the 95% quantile, she jugded the trend signal to warmer and more humid winters in Central Europe as reliable. The trend in the extremes is more pronounced than in the average. During summer central Europe is in a transition zone between a increasing trend in Scandinavia and an decreasing trend in the Mediterranian and, therefore, the trend is less reliable (Kendon et al., 2010). The analysis, however, is only qualitative and concentrates merely on the sign of the trend signal. She states that the trend in winter precipitation can be altered by changes in the circulation pattern that are not considered in the model.

In an extensive study, including six different combinations of four RCM nested in three different GCM, Fowler et al. (2007) estimate the possible error range in climate model simulations for the European continent. All GCM-RCM combinations of the ensemble exhibit similar error characteristics. On the one hand, they typically overestimate the number of wet days and, on the other hand, underestimate the average daily precipitation as well as the standard deviation. The reduced standard deviation indicates that the models are not able to represent the full climatic variability in precipitation. Thus, the underestimation is the most pronounced in extreme values, as for example the daily precipitation extreme with five year return period (Fowler et al., 2007). Besides, the error in extremes depends on the scale of the temporal aggregation. The underestimation in five day precipitation sums is less sever than in daily values.

The results of a similar study of the alpine region confirm the findings of Fowler et al.. Frei et al. (2003) compare the simulation results from five RCMs with gridded observed precipitation from more than 6000 rain gauges in the alpine region. All models considerably underestimate the standard devaition of precipitation intensities (by 16% to 42%) as well as the frequency of heavy events. In this study the RCMs are forced by reanalysis data, which means by observations and not by a Global Circulation Model. Therefore, the quantified error signals only concern the RCMs. In a climate simulation the error would add up with the model error from the GCM simulation govering the RCM.

8.2.3. Correction and Downscaling of GCM Precipitatioin Data

Due to the various error sources listed in the last subsection, GCM and RCM data must not be applied as it is in any precipitation modeling or hydraulic or hydrological simulation. Especially the reduced variability is dangerous in terms of the underestimation of extreme value frequencies. However, even if GCM and RCM produced perfect forecasts, there would be the scale difference between the circulation models and rain gauge measurements. GCM and RCM calculate the areal average precipitation for each grid cell, which is not the same as point values simulated by a rainfall generator or observed by a rain gauge.

Willems and Vrac (2011) presents different methods that can be used to correct and downscale precipitation data from GCMs or RCMs. Correction methods are generally based on a comparison of the distributions from the climate model and observations from reanalysis or rain gauges during a control period in the past. Willems and Vrac (2011) distinguishes between two different types of downscaling approaches that either try to directly correct the simulated precipitation time series or use other GCM variables, as the simulated sea surface pressure, for a classification of observed precipitation.

8.2.3.1. Direct Correction of GCM Precipitation Data

A direct correction method is the Delta-Change approach. Knowing that the absolute values of the GCM precipitation is wrong, only the trend in different quantiles is considered. To calculate a future precipitation scenario, the relative trend in each quantile is multiplied

to the observed precipitation value of the same quantile (Willems and Vrac, 2011). This approach is not unique when there are several time steps with identical precipitation values. Values of the same rank have to be ordered according to additional criteria, for example the amount of precipitation in preceding and following time steps (Willems and Vrac, 2011). Another issue is that the delta factor can take a value of infinity for some quantiles if the rainfall probability is increasing in the future. The problem can be avoided if the ranks are calculated only on the basis of wet time steps, but then the future rainfall probability has to be corrected separately.

The QQ-transformation is a direct correction method that is based on a comparison between the simulated distributions of a GCM and the observed distribution of a ground reference. It is able to correct for non-linear and non-proportional errors. It means that an individual correction factor is applied at every quantile of the distribution. Sennikovs and Bethers (2009) use this kind of transformation in a climate change study for the Baltic sea region. Boé et al. (2007) apply a QQ-Transformation to data from the French GCM ARPEGE during a study on the hydrology of the Seine river in Central France.

Mathematically, the QQ-Transformation is described by:

$$H_{corr} = F_{true}^{-1}(F_{err}(H_{err})) \tag{8.7}$$

$$d_{corr} = \frac{H_{corr}}{H_{err}} \tag{8.8}$$

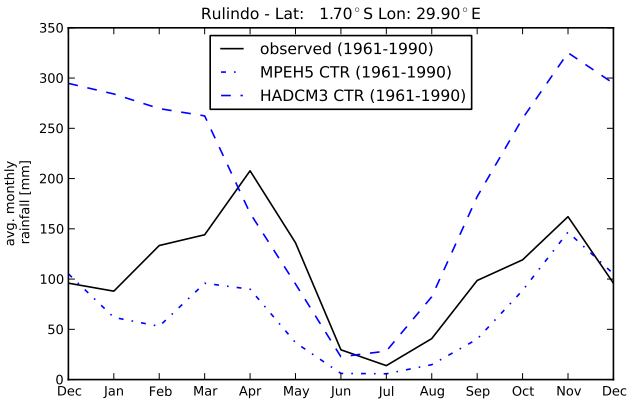
where
- H_{corr} corrected monthly precipiation sum
- H_{err} original value from the erroneous distribution
- F_{err} non-exceedance probability of H_{err} according to the erroneous distribution
- F_{true}^{-1} Inverse of the cumulative distribution function of the 'true' distribution from the ground reference
- d_{corr} correction factor for the respective quantile

QQ-transformation is calibrated by the distributions of GCM and ground reference data during a control period, typically 30 years. The transformation according to Eq. (8.7) is applied to the whole range of the GCM distribution. The GCM distribution is transferred into the distribution of the ground reference and thus the errors in the control period data are eliminated. The transformation factors to each quantile are kept. Assuming that the error characteristic of the GCM is constant over time, the correction factors are applied to data from future runs of the climate model. The result is a corrected distribution that is comparable with past time observations from the ground reference.

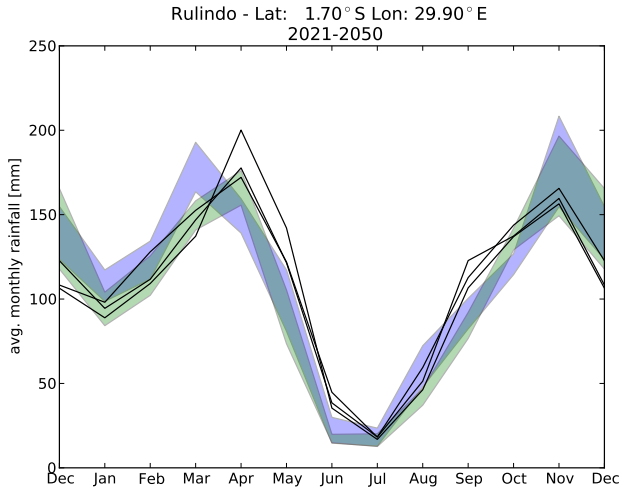
To account for potential seasonal cycles in the error structure, QQ-Transformation can be set up for each month of the year separately.

QQ-Transformation is not only a correction but also a downscaling tool. The data from one GCM grid cell is transformed differently at each point. Within one GCM grid cell, the distribution F_{err} of the GCM data is constant, whereas the distribution on the finer gridded ground reference F_{true} is spatially varying due to local effects. Since the correction factors

d_{corr} depend on both distributions, the spatial resolution of the ground truth is impressed on the coarser GCM data.



(a) Raw data



(b) QQ- corrected predictions

Figure 8.7.: Example of the effect of QQ-Transformation. a) average annual cycle in raw data; b) minimum to maximum range of future predictions according to three scenarios A1B, A2 and B1 in HadCM3 (blue) and ECHAM5 (green) data, black line: ground references used for downscaling.

In a case study for Central Africa, monthly precipitation sums from two GCM, the atmospheric model ECHAM5 and the coupled model HadCM3, are downscaled on the $0.5° \times 0.5°$ latitude-longitude grid of three different reanalysis products that are used as reference for the true observed precipitation.

Fig. 8.7 demonstrate the effect of the QQ-transformation on the average annual cycle for a grid cell near Kigali, Rwanda. Fig. 8.7a compares the average annual cycle of the raw data for the control period from 1961 to 1990. Untransformed, the GCM are not able to model the data correctly. There is a strong bias in both models, HadCM3 shows an overestimation and

ECHAM5 sever underestimation. The solid black lines in Fig. 8.7b represent the average annual cycle of the reanalysis that are used as ground reference in the QQ-Transformation. Fig. 8.7b displays the minimum to maximum range in the ensemble of future predictions for the average annual cycle from 2021 to 2050. The range spans over an ensemble of three scenarios combined with the three reanalysis products. The color indicates the GCM model.

The chosen scenarios (A1B, A2 and B1) cover a wide range of possible CO_2 emissions. A2 is one of the scenarios with the highest CO_2 concentration in the regarded time period from 2021 to 2050, B1 is one of the scenarios with the lowest concentration. Nevertheless, the expected change in the annual cycle does not principally depend on the scenario. The choice of the GCM model has about the same effect on the future prediction, in some months (February, March) the influence of the GCM is higher. As both GCM predict the same future if the same scenario is applied, any deviation can be interpreted as model uncertainty. Due to the huge deviations in the original data (see 8.7a), the QQ-transformation cannot eliminate all errors. The model error in the uncorrelated data is several times higher than the predicted changes. The different characteristics of the GCM are still visible in the QQ-Transformed data.

For precipitation on daily or subdaily scale the presented direct correction methods are not suitable since it cannot consider any information on temporal persistence. Strictly speaking, the application is limited to temporally independent values. In European climate, the values on daily or sub-daily temporal resolution are significantly autocorrelated. As each time step is corrected differently, depending on its quantile, the autocorrelation structure is ignored. The persistence structure of the GCM is altered, errors in the persistence are not corrected.

Besides, the methods are not able to correct the frequency of zero values, which is important on daily or subdaily scale. If the GCM generally underestimates the rainfall frequency, the QQ-Transformation is not unique. It cannot be judged if a certain zero value should be kept or corrected to a positive precipitation value. If a zero value is corrected to non-zero rainfall, a value has to be chosen, e.g. by stochastic simulation – which induces further uncertainty in the predicted GCM time series.

Finally, it should be noted that the assumption of direct methods, that the error in GCM precipitation is constant over time, cannot be checked. During QQ-Transformation, a potential drift in the relation between GCM and ground reference is multiplied by the transformation factors d_{corr}. The induced error, if the relation between GCM and reanalysis is changing over time, is amplified by the strong correction that has to be applied to the original GCM output.

8.2.3.2. Indirect Correction by Additional Information

As an indirect method Willems and Vrac (2011) present the Analogues Method based on mean sea level pressure fields that are judged as a more reliable than GCM precipitation. To each day in the GCM output an analogue day is searched among all days of the same month and the same weather type in an archive of mean sea level pressure fields from reanalysis data. The observations of the analogue day is seen as the future prediction. Additional

criteria, as for example a sub-classification according to temperature can be applied. A shortcoming of this approach is that a refinement in more classes reduces the number of values in each class, hence induces sampling uncertainty and might lead to an underestimation of very rare events (Willems and Vrac, 2011). Another limitation is that the future predictions cannot be extrapolated beyond what was measured in the past. The latter issue can be attacked in a very rough manner by multiplying each value by a correction factor depending on temperature, e. g. 7% per degree of temperature rise(Willems and Vrac, 2011).

Willems and Vrac (2011) test the method with precipitation data from the rain gauge in Uccle, Belgium, which is one of the longest high resolution precipitation time series in Europe. They use an ensemble of 17 ECHAM5 runs as GCM data. The model approach is calibrated on a 30 year period and tested on the next 30 years. During the validation period extreme precipitation is systematically underestimated after the correction. It is found that the error is scale dependent and decreases for higher aggregation intervals. Compared to the Delta-Change approach (Section 8.2.3.1) the Analogues Methods predicts lower changes. Willems and Vrac (2011) suspect that not all of the trend signal can be explained by changes in the weather type sequence (Willems and Vrac, 2011).

Another indirect method that exploits weather type information is to use the distribution of hourly precipitation amounts conditioned on different weather type classes as a prediction for future rainfall. Such a method will be described in the next section.

8.3. Prediction of Future Precipitation by GCM Defined Circulation Patterns

In physical atmospheric modeling the representation of precipitation is the most difficult. It depends on a high number of other variables that have to be calculate before, as for example the available atmospheric moisture from evaporation, the air temperature and the wind field. In weather forecast, the exact predictions of precipitation is the most difficult task (see for example Bliefernicht, 2010). It has been illustrated in the last section that precipitation predictions of GCM are related with high modeling uncertainties.

Many publications indicate that the change in Central European climate is at least partly due to changes in meso-scale atmospheric conditions (Haylock and Goodess (2004), Kendon et al. (2010) or Auer et al. (2007)). By a weather type classification, the changes in atmospheric circulation can be related to changes in precipitation.

To estimate the influence of changes in the atmospheric circulation on short term precipitation intensities, a fuzzy rule based CP-classification system is set up. The response in precipitation of each CP is calculated from observed rain gauge data. The application of uncertain precipitation data from GCM is avoided in this way. From the GCM only predictions of the Mean Sea Level Pressure (MSLP) are considered to calculate a CP sequence for the Future. The changes in precipitation are seen as the result of changes in the CP sequence as well as the reaction of the CPs to changes in air temperature.

The applied CP classification is based on NCEP/NCAR reanalysis (Kistler et al., 2001). The definition of CP classes follows the approach by simulated annealing described in Bárdossy (2010) but applies a different objective function. (See also Section 7.5.2 for a detailed description of the method.) To examine the effects of climate change in hourly precipitation intensities, the objective function is designed to detect differences in the distribution of hourly precipitation amounts.

According to the sketch in Fig. 8.8, the optimal CP classification is found when distribution of 1 h precipitation amounts of all the days belonging to one CP differ the most from the climatic average. The difference is measured by the variable χ^2 which is equal to the squared sum of the residuals in relative frequency between the histogram for all days belonging to one CP and the histogram for the climatic average without classification. (In Fig. 8.8 χ^2 is equal to the sum of the squared lengths of the arrows.)

$$\chi^2 = \sum_{i=1}^{n_{cp}} \sum_{j=1}^{k} \left(h_i\left(j\right) - h_{clim}\left(j\right)\right)^2 \rightarrow \max! \tag{8.9}$$

where
- $h_i\left(j\right)$ observed relative frequency of intensity class j during all days belonging to CP i
- $h_{clim}\left(j\right)$ relative frequency of class j in the calibration period from 1991 to 2003 without CP classification
- n_{cp} number of different Circulation Patterns in the classification
- k number of histogram classes

Data basis for the histograms calculation is a homogeneous and close to complete subset of 30 hourly precipitation stations of the NiedSim data set (see Fig. 3.6 in Section 3.3). The calibration period is from 1991 to 2003. For comparability reasons the histogram classes are not based on absolute values but on the quantiles at each rain gauge location. This

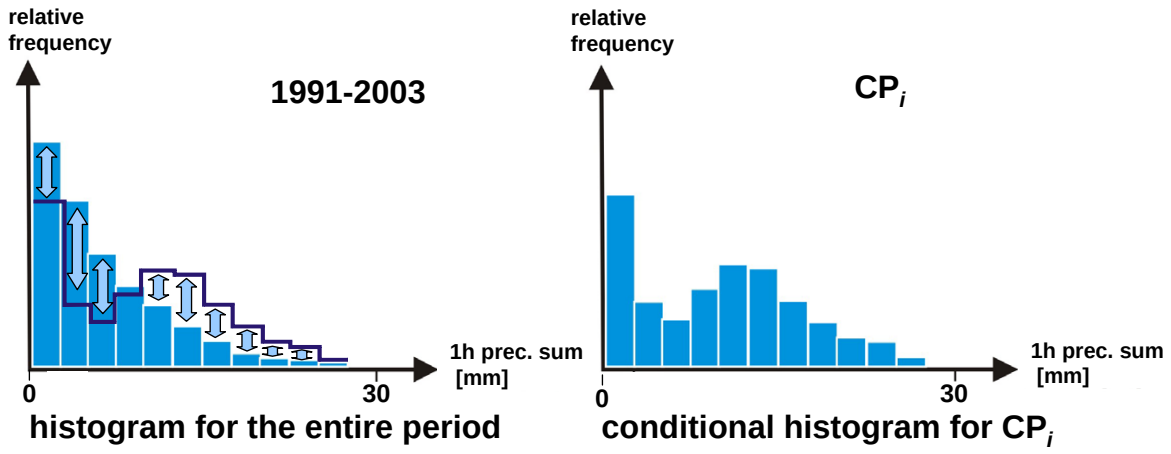


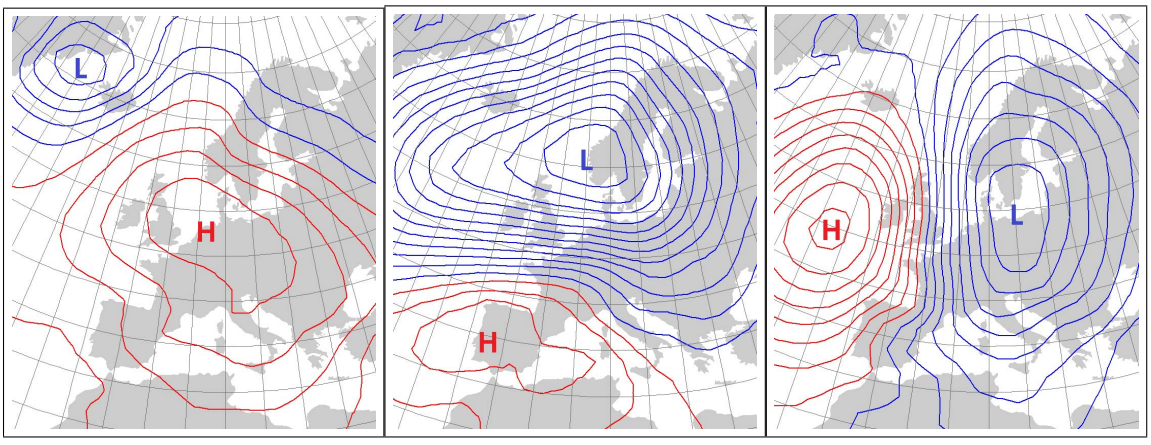
Figure 8.8.: Principle of the objective function for the CP definition in the simulated annealing scheme.

means that the values within one class have different absolute amounts, depending on the location, but lie within the same range of the CDF at their respective location. To shift the focus towards extreme events, the classes are not equidistant but smaller in the higher CDFs. The upper limits of the histogram classes are $F = 0.6$, $F = 0.8$, $F = 0.9$, $F = 0.95$, $F = 0.98$, $F = 0.99$ and $F = 1$. The CDFs are calculated from all wet hours with no less than $0.1\,\mathrm{mm}$ of precipitation.

The histograms are calculated for two seasons, from September to April and from May to August to account for convective precipitation enhancement during summer months. As a trade-off between a good representation of the possible variability in atmospheric circulation and a sufficient number of precipitation events in each CP class, the number of CPs is restricted to $n_{cp} = 12 + 1$ – the last CP reserved for all days that cannot be assigned to any of the other twelve classes.

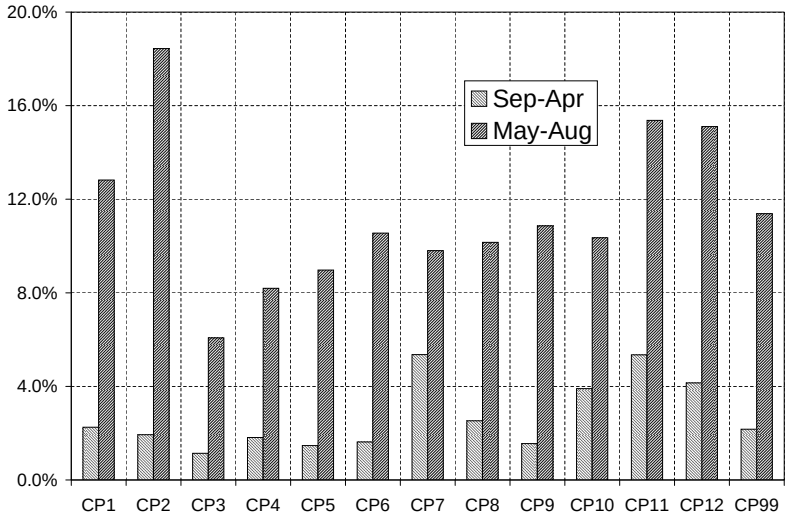
Fig. 8.15 displays the average pressure anomalies in the NCEP-NCAR data set of all days belonging to three CPs with high potential for extreme precipitation exceeding the 95% quantile. It is surprising that CP2 (Fig. 8.9a) is among the extreme CPs. It is a high pressure situation that generally leads to dry condition. The point is that is rarely rains on days with CP2, but if it rains, the precitation can reach high intensities during summer months.

Fig. 8.15 displays the relative exceedance frequencies of the 95% and 99% quantile of all wet hour precipitation amounts. In average over all CPs and both seasons the frequency is 5% respectively 1%. Extreme intensities are generally more likely in summer. The exceedance freqencies from May to August are higher in all CPs. However, the seasonal response of the CPs is varying. CP2 only has an elevated extreme value potential in summer months, while the frequencies of CP11 are among the highest in both seasons. CPs with westerly fluxes as CP7 show the lowest increase in extreme intensity frequency during summer months. It is less influenced by convective intensity enhancement.

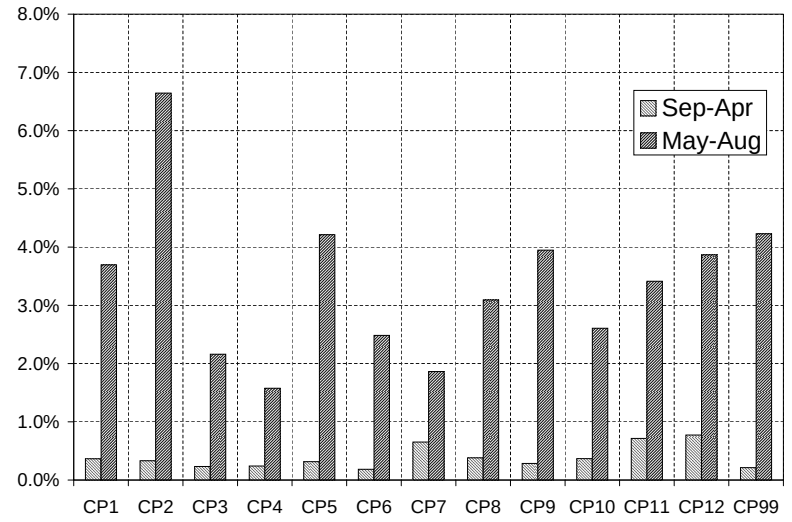


(a) CP2 (b) CP7 (c) CP11

Figure 8.9.: Mean sea level pressure anomalies of CPs related to high precipitation intensities



(a) 95% quantile

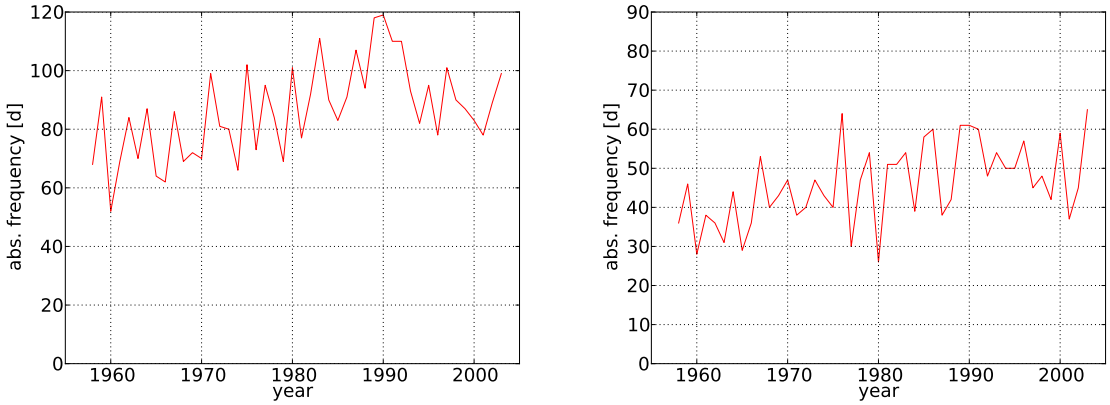


(b) 99% quantile

Figure 8.10.: Exceedance frequency of the 95% and 99% quantile for two seasons. Quantiles and frequencies are calculated for all hours with $H \geq 0.1\,\mathrm{mm}$

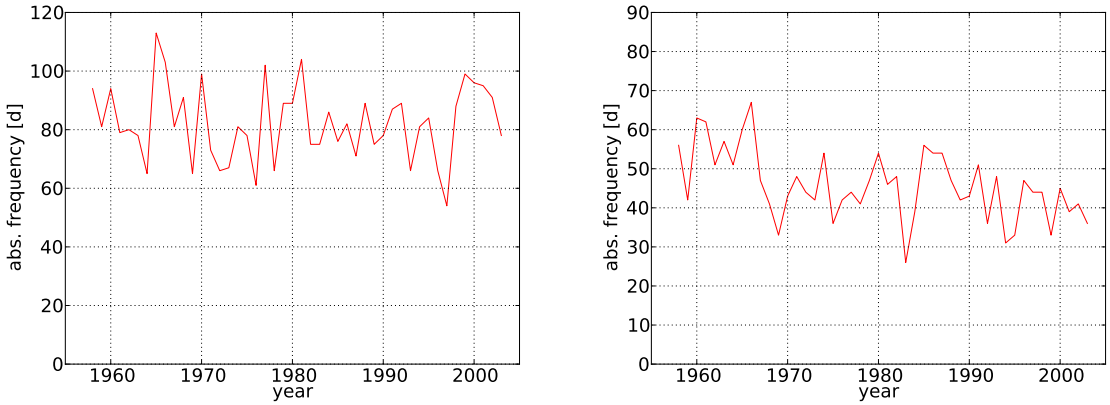
8.3.1. Changes in CP-Frequencies According to GCM and Reanalysis Data

For the analysis of the occurrence frequencies the CPs are grouped according to their meteorological characteristics. Fig. 8.11 shows the evolution of the frequency of anticyclonic CPs. During the calibration period from 1991 to 2003, one of these four CPs occurred on 35.6% of all days, but only 17.5% of all rainy hours at the 30 rain gauges used for calibration were counted during that time. In Fig. 8.12 the absolute frequency of cyclonic CPs is displayed.



(a) September to April (b) May to August

Figure 8.11.: Absolute frequency of anticyclonic CPs (CP2, CP3, CP5 and CP8) according to NCEP/NCAR reanalysis during two seasons



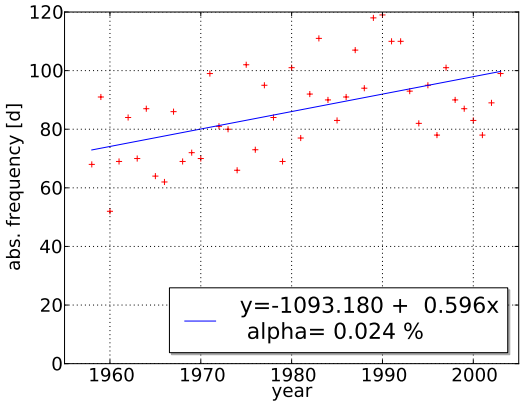
(a) September to April (b) May to August

Figure 8.12.: Absolute frequency of cyclonic CPs (CP7, CP10, CP11 and CP12) according to NCEP/NCAR reanalysis during two seasons

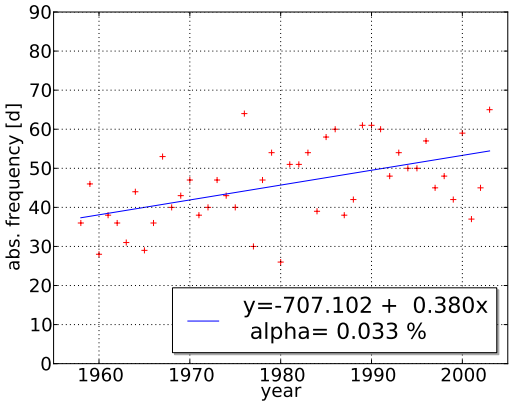
From 1991 to 2003, one of these CPs occurred on 30.0% of all days. They are responsible for 55.4% of all hours with non-zero precipitation.

Fig. 8.11 and Fig. 8.12 suggest that the atmospheric circulation has shifted between 1958 and 2004. Especially in summer months, the number of days with antyclonic CPs has increased, the number of days with cyclonic CPs decreased. In Fig. 8.13 and Fig. 8.14 the trend signal is quantified by linear regression. The value of "alpha" below the equation of the regression function is the p-value of a two sided hypothesis testing that the slope of the true regression line is zero. If alpha is low, it is safe to reject the hypothesis, thus the increasing or decreasing trend is significant.

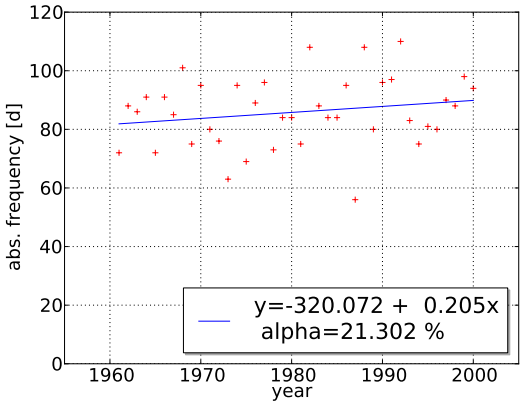
Taking 5% as a limit, there is a significant inccreasing trend in high pressure situations in



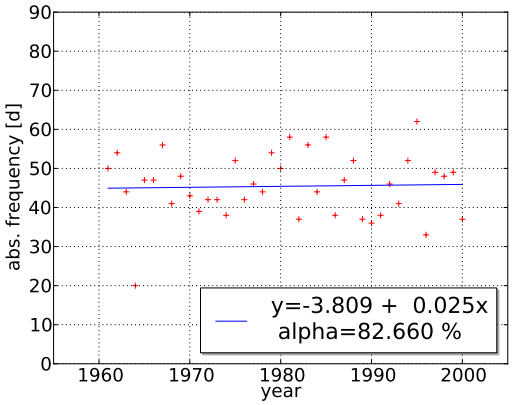
(a) NCEP/NCAR - September to April



(b) NCEP/NCAR - May to August



(c) ECHAM5 - September to April

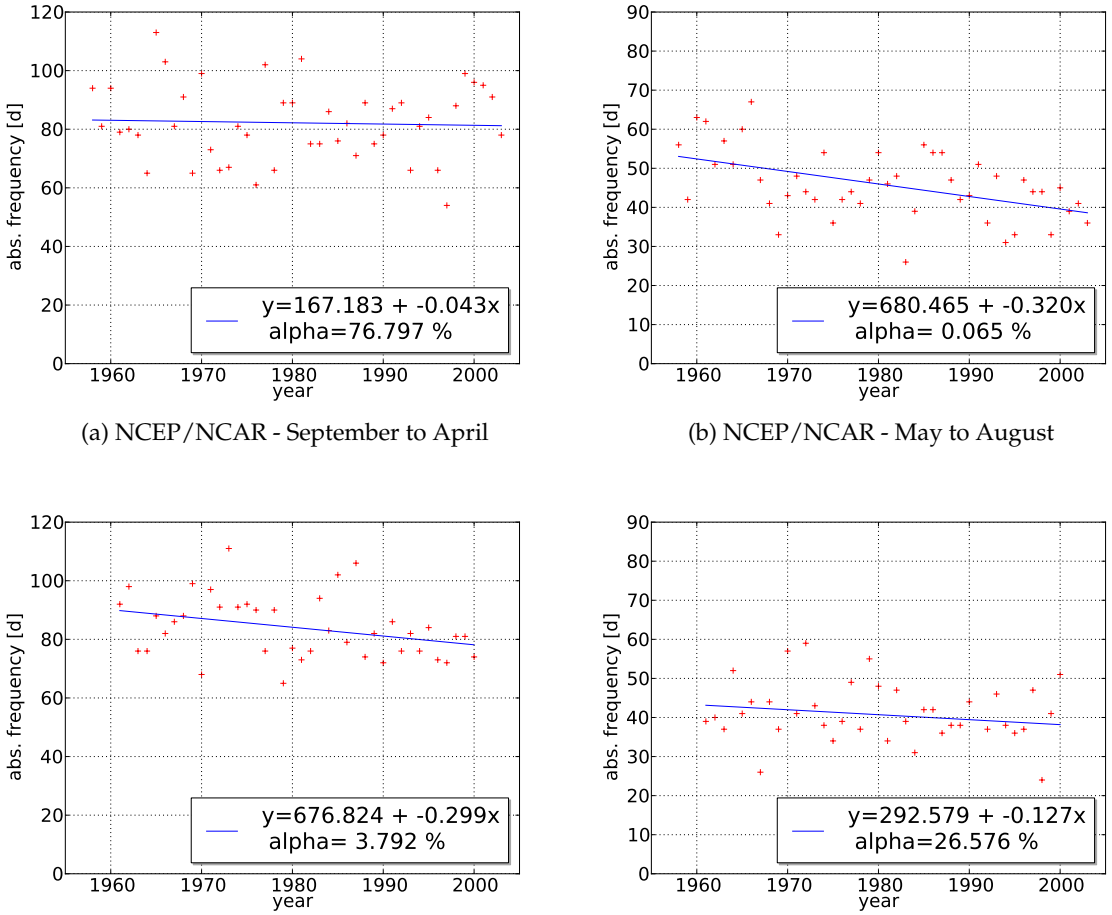


(d) ECHAM5 - May to August

Figure 8.13.: Trends in the frequency of anticyclonic CPs (CP2, CP3, CP5 and CP8) according to NCEP/NCAR reanalysis and the 20th century control run of ECHAM5 during two seasons

both seasons and a decreasing trend in low pressure situations for the non-convective season from September to April. The trends are derived from NCEP/NCAR MSLP data that is seen as the observation reference. (Other reanalysis products were tested too. The results stay the same, for example using ERA40 data.)

The second line of Fig. 8.13 and Fig. 8.14 presents CP-frequencies predicted by the ECHAM5 model. The data is from the 20th century run. This is a free climate simulation but initiated with observed starting values and forced by the real measured land use and CO_2 emissions during the last century. The CPs are defined on the basis of the available MSLP data from 1961 to 2000. Since the GCM is run in a free simulation and not forced by any measurements (except for the external quantities of land use and CO_2) it cannot be expected that the absolute CP-frequencies are the same as in the reanalysis data. In average over several years, however, the frequencies should be comparable if the GCM is able to simulate the atmospheric circulation correctly.



(a) NCEP/NCAR - September to April

(b) NCEP/NCAR - May to August

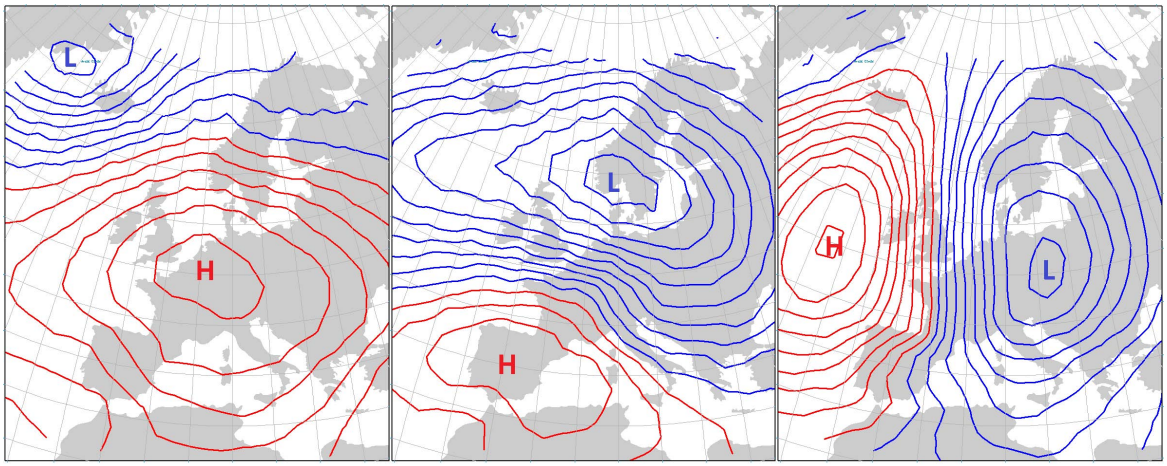
(c) ECHAM5 - September to April

(d) ECHAM5 - May to August

Figure 8.14.: Trends in the frequency of cyclonic CPs (CP7, CP10, CP11 and CP12) according to NCEP/NCAR reanalysis and the 20th century control run of ECHAM5 during two seasons

It can be seen that ECHAM5 has limited capacities in reproducing the atmospheric conditions. Average frequencies of cyclonic CPs are correctly modeled (Fig. 8.13c and Fig. 8.13d), but ECHAM5 misses most of the trend signal. In the non-convective season the trend is underestimated by two thirds (Fig. 8.13c), in summer months it is completely missed (Fig. 8.13d) although anticyclonic CPs increased by almost two days in every five years between 1958 to 2004. For the anticyclonic CPs ECHAM5 models a decreasing trend from September to April (Fig. 8.14c) that cannot be seen in the reanalysis data. The observed decreasing trend in summer, on the other hand is underestimated (Fig. 8.14d).

The CPs are defined on the basis of normalized pressure anomalies and not on absolute pressure values, hence the problems of ECHAM5 in reproducing the CP frequencies do not come from different mean conditions. They express differences in the temporal sequence of tbe MSLP-fields. The deviation between NCEP/NCAR and ECHAM5 data in CP conditioned pressure anomalies are low. The average pressure anomaly maps in Fig. 8.15 derived



(a) CP2 (b) CP7 (c) CP11

Figure 8.15.: Mean sea level pressure anomalies according to ECHAM5 data of CPs related to high precipitation intensities

from the GCM data look very similar to the pressure maps from the reanalysis (Fig. 8.9). The centers of the high and low pressure zones are at the same locations and the shapes are comparable.

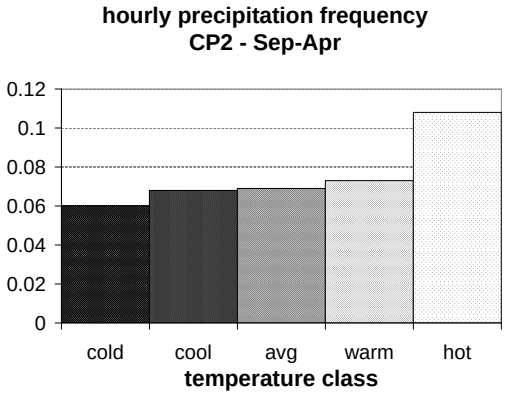
8.3.2. Temperature Sensitivity of CP-related Precipitation

The capacity of the atmosphere to carry moisture is increasing with temperature, the evaporation over the ocean and over land too. It can be expected that the precipitation response of the CPs is temperature dependent. By the average observed daily air temperature at 156 measurement stations in Baden-Württemberg the CP classes were further divided into five temperature classes. Since temperature is a far less varying quantity than precipitation, the set of these point measurements can be seen as representative for the average temperature in the low level atmosphere.

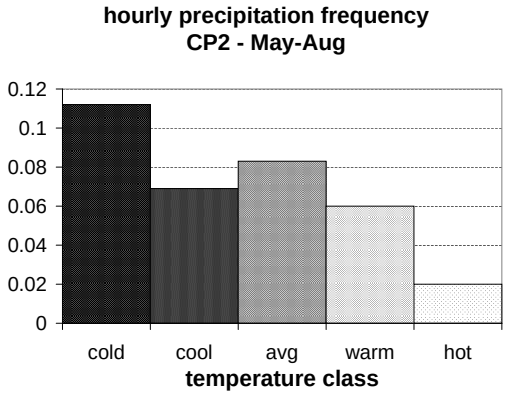
The temperature subdivision is made by the observed quantile. The 20% with coldest average temperature in each CP class are declared "cold", the next 20% cool, followed by "avg", "warm" and "hot" conditions. To avoid seasonal effects, the quantiles are calculated on basis of anomalies from the average yearly cycle.

The temperature reaction is tested by the hourly precipitation frequency with a threshold of $0.1\,\mathrm{mm}$ and the exceedance frequency of the 95% and 99% quantile. The quantiles are calculated based on all values exceeding the treshold. The exceedance frequenccies refer to all wet hours ($H > 0.1\,\mathrm{mm}$) of the respective CP.

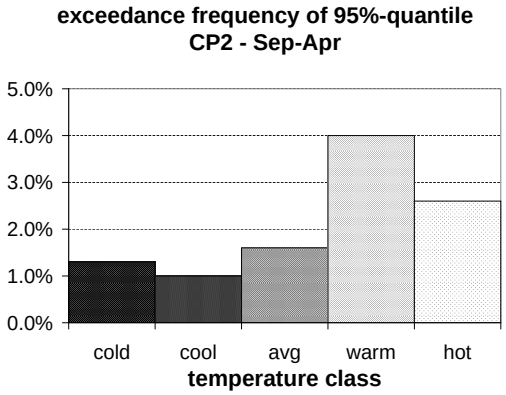
All three tested CPs react on temperature (Fig. 8.16 to Fig. 8.18) and the reaction is in general different for the non-convective and convective season. The results for the summer months support the findings of Trenberth (1999) that the rainfall frequency was decreasing and the peak intensities increasing with global atmospheric temperature. In the summer months, all



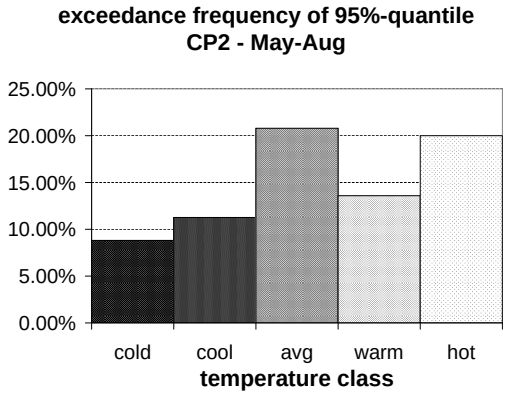
(a) Sep to April



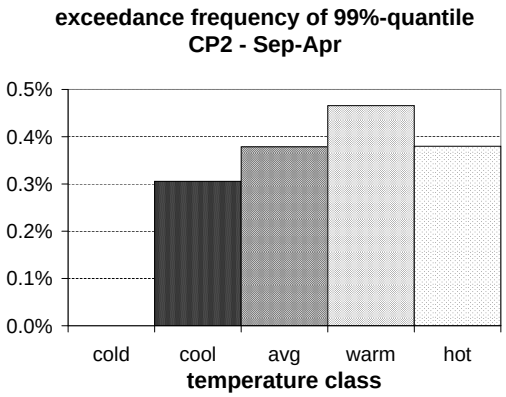
(b) May to August



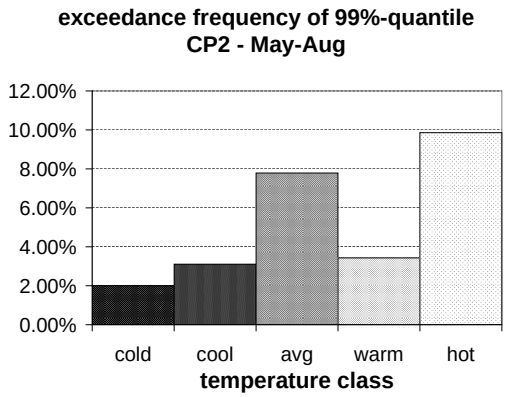
(c) 95% Quantile – Sep to April



(d) 95% Quantile – May to August

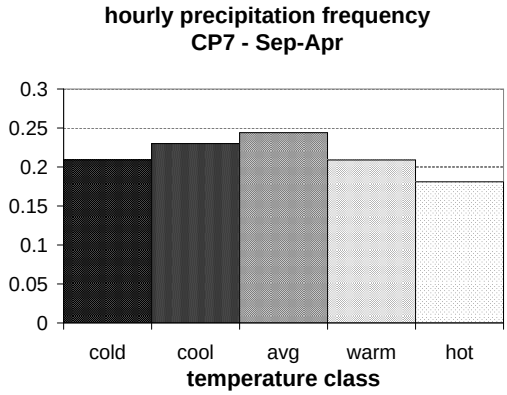


(e) 99% QuantileSep to April

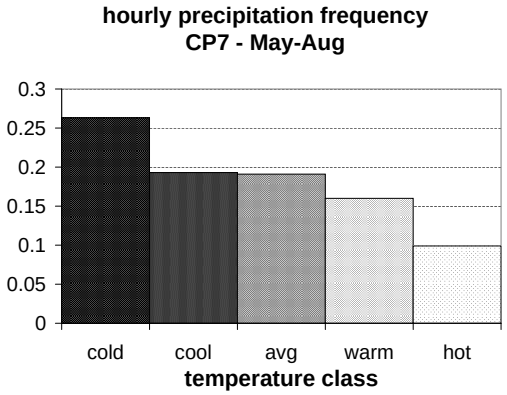


(f) 99% Quantile – May to August

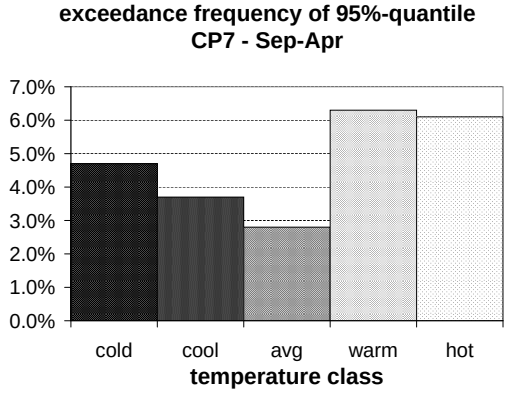
Figure 8.16.: Precipitation frequency and frequency of extreme 1h precipitation amounts during all days belonging to CP2 as a function of temperature.
The extreme value frequency is referred to the number of wet hours ($H > 0.1\,\mathrm{mm}$) during CP2



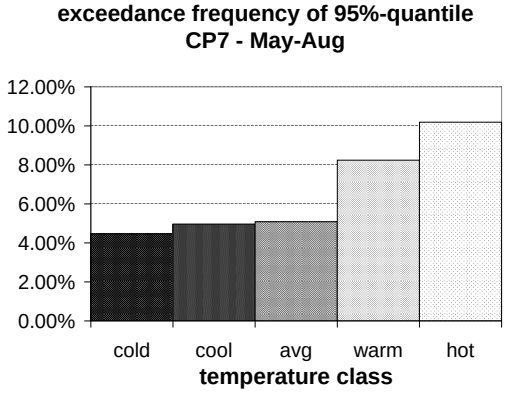
(a) Sep to April



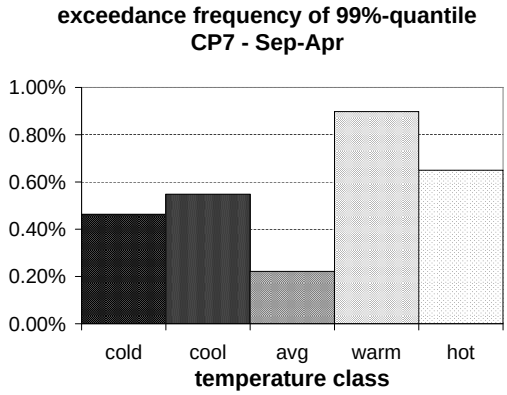
(b) May to August



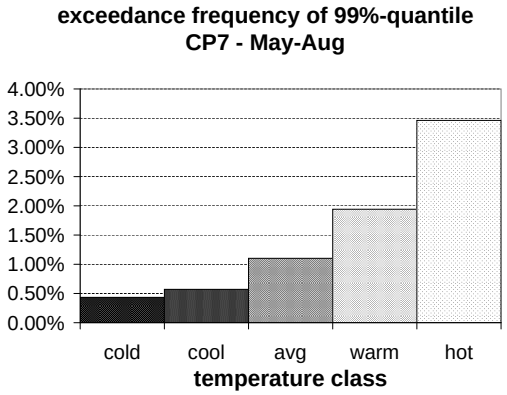
(c) 95% Quantile – Sep to April



(d) 95% Quantile May to August

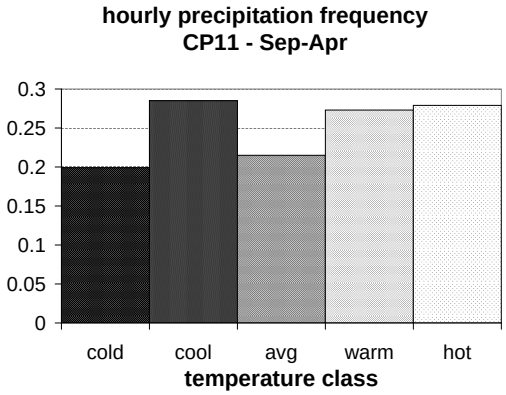


(e) 99% Quantile – Sep to April

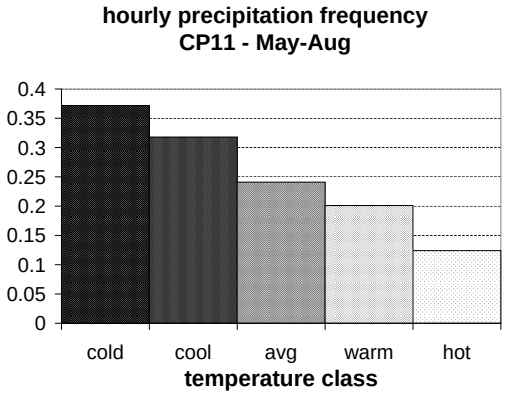


(f) 99% Quantile – May to August

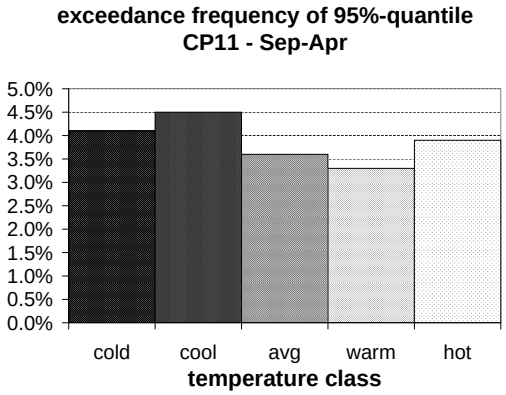
Figure 8.17.: Precipitation frequency and frequency of extreme 1h precipitation amounts during all days belonging to CP7 as a function of temperature.
The extreme value frequency is referred to the number of wet hours ($H > 0.1\,\mathrm{mm}$) during CP7



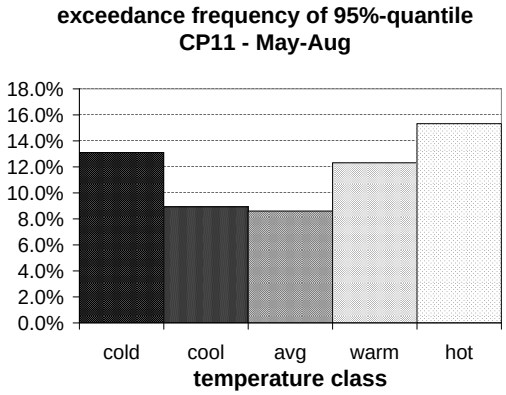
(a) Sep to April



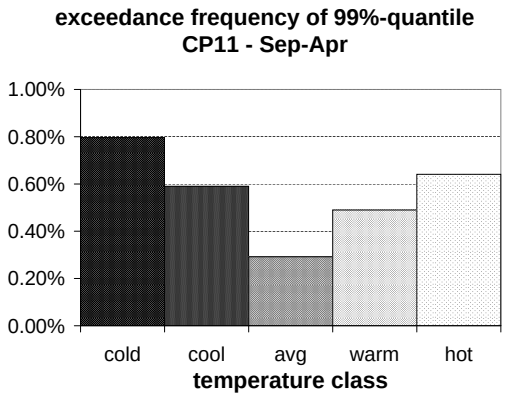
(b) May to August



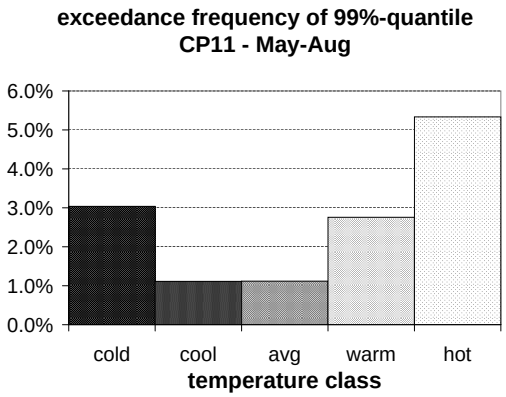
(c) 95% Quantile – Sep to April



(d) 95% Quantile – May to August



(e) 99% Quantile – Sep to April



(f) 99% Quantile – May to August

Figure 8.18.: Precipitation frequency and frequency of extreme 1h precipitation amounts during all days belonging to CP11 as a function of temperature.
The extreme value frequency is referred to the number of wet hours ($H > 0.1\,\mathrm{mm}$) during CP11

CPs show a pronounced drop in precipitation frequency with increasing temperature while during the rest of the year the rainfall frequency is constant (CP7 and CP11) or slightly increasing (CP2).

The extreme intensity reacts more sensitively on temperature than the rainfall frequency, especially in summer months. If it is raining, the exceedance frequency of the 99% quantile on CP7 "hot" days is more than three times higher than the average over all days belonging to CP7 (Fig. 8.17f). And it seems that the highest extremes are affected the most. The differences in the 99% quantile are far more pronounced than in the 95% quantile. During CP7 and CP11 the increase in wet hour extreme frequency overcompensates the decrease in precipitation frequency. Refered to all hours, the frequency of extrem hourly precipitation is higher than average on "hot" days – even if it rains rarely.

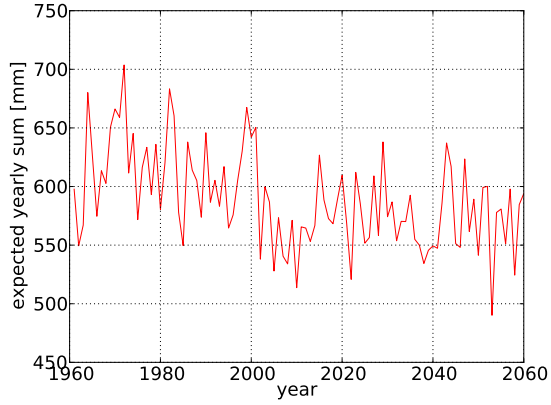
CP2, however, reacts differently. The highest extreme value frequencies, referred to all hours, are measured during days of average temperature. There are several probable explications. Either the increasing moisture storage capacity of hot air inhibits precipitation (as it is described in Kendon et al. (2010)). However, it is not clear why this should especially effect the most extreme events. Or it depends on the available soil moisture. The evaporation consumes energy and thus reduces the air temperature. Hence, the days with the highest energy level might not be the hottest days because part of the energy goes into latent heat. If the water availability to evaporation is limited, the temperature can rise more and precipitation is reduced. The effect is not present in the other CPs since they receive moisture by advective transport from the Atlantic Ocean (CP7) or the North Sea and Polar Sea (CP11).

Besides, intense precipitation is always linked to cooling. Much of the raindrop water in estival thunder storms is directly evaporated before the drop reaches the ground (Weischet and Endlicher, 2008). The energy for evaporation is delivered by the surrounding air leading to a drop in temperature. Since most thunderstorms occur in the afternoon, the cooling affects the average daily air temperature. A more representative measure in this respect would be the average temperature during the time that the thunderstorm builds up, or the maximum just before it begins to rain, which both is not available.

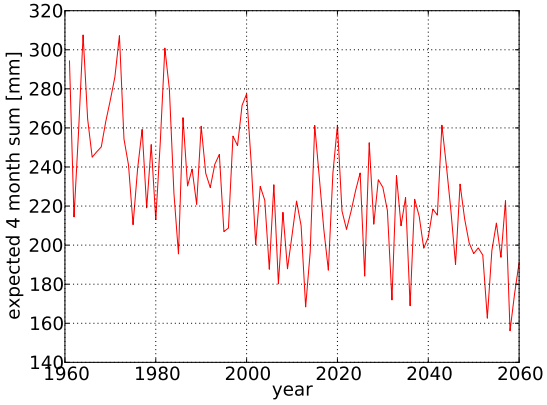
During CP11, the coldest days have particularly high extreme value frequencies (Fig. 8.18d and Fig. 8.18d). This is another indication that the average daily temperature cannot measure all effects of temperature sensitivity. During CP11 the atmospheric flow comes from the north. If the atmosphere was warm before, the drop in temperature when the northern flux sets in releases much water, due to the decreasing storage capacity of the atmosphere. The amount of released atmospheric water is a function of the temperature difference and depends on the amount of available water that is present from the days before. Therefore, it could be beneficial to consider the temperature history before a precipitation event.

8.3.3. Expected Trend in 1h Precipitation Amounts due to Changes in Atmospheric Circulation

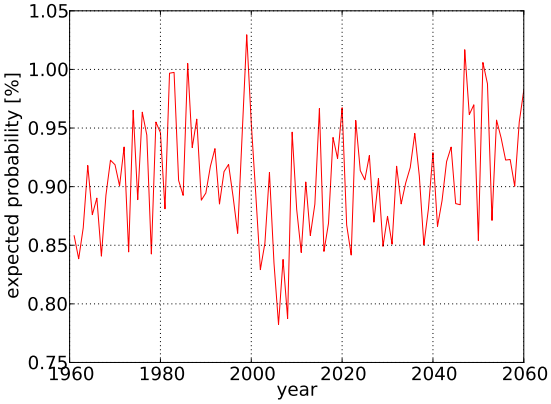
During the last forty years, the CP sequence was significantly changing. The CPs in turn have a high influence on the distribution of $1\,\mathrm{h}$ precipitation amounts. In the same time, CP-



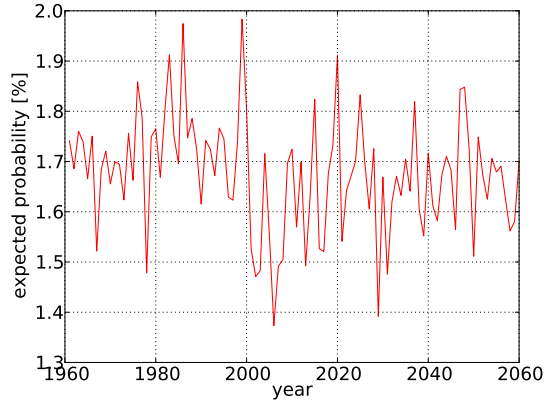
(a) Yearly Sum



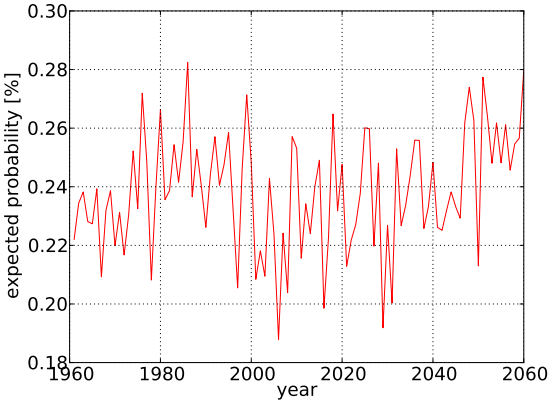
(b) Precipitation Sum – May to August



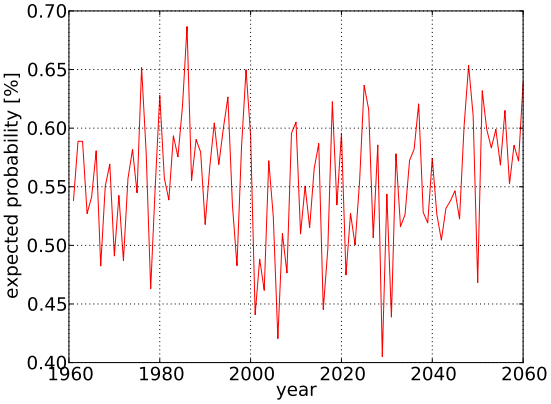
(c) 95% quantile – whole year



(d) 95% quantile – May to August



(e) 99% quantile – whole year



(f) 99% quantile – May to August

Figure 8.19.: Expected precipitation sum and expected extreme value probability estimated by the CP sequence based on ECHAM5 data for 1961 to 2060, CO_2 emission scenario A1B

related precipitation is reacting on temperature. The question is, how the different mechanism are combining in the future.

For a test it is assumed that the precipitation response of each CP in each temperature class is constant over time. The CP sequence of the future can be estimated by applying the fuzzy rule set developed on basis of the NCEP/NCAR reanalysis on MLSP data from GCM runs for the future. The atmospheric temperature level can directly be derived from the GCM. Temperature is seen as the most reliable quantity in GCM output, (see for example Kendon and Clark, 2008).

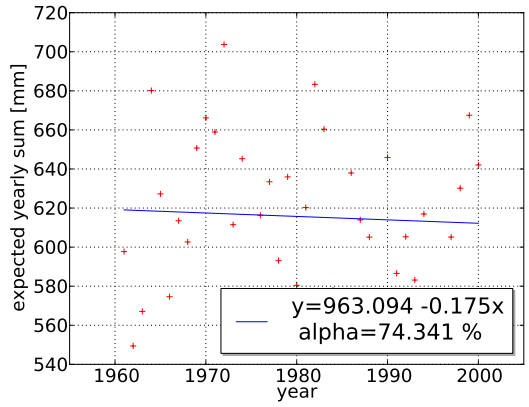
If the time series of the combined CP-temperature classes is known, the expectation of different precipitation related statistics can be calculated for the future. The method is illustrated by the example of the 95% precipitation quantile: It is assumed that a day i between May and August in the year j is assigned to CP2 and the temperature level to "warm". Then the 95% extreme value probability is assumed to be 13.6% and the rainfall probability 18.6% (see Fig. 8.16d and Fig. 8.16b) according to the average observed values in the calibration period. In reference to all hour (wether it is raining or not), the 95% extreme precipitation probability is $P_{95\%} = 0.136 \cdot 0.186 = 2.53\%$ in each hour of day i. Doing the same calculation for all days $i \in [1, 365]$ of one year (with the statistics of the respective season) and averaging over the year, one can calculate the expectation of the 95% extreme value probability for the respective year. (Since the calculation of the 95% quantile is based on wet hour only, the expected exceedance frequencies in respect to all hours in the year is lower than 5%.)

The described method is applied to ECHAM5 data of the time period from 1961 to 2060. Until 2000, data from the 20$^{\text{th}}$ century run is used. The future climate predictions start in the year 2001. The considered scenario is A1B. Yearly expectations of the 95% and 99% extreme probability are calculated and similarly of the yearly preciptiation sum (in this case not by averaging but by adding up the contributions of each day). An additional analysis is performed for the convective summer months from May to August.

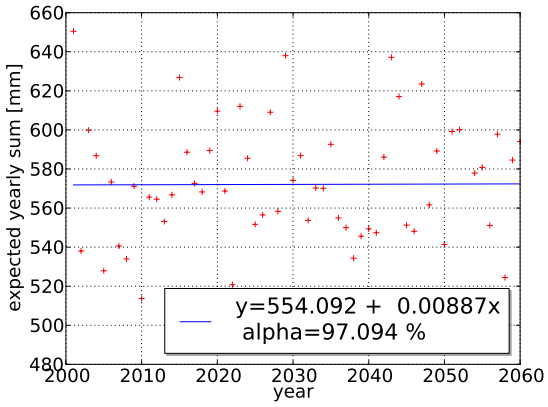
Compared to the reanalysis, the GCM tends to dryer CP-temperature combinations (Fig. 8.19a and Fig. 8.19b). The average precipitation in the control period from 1961 to 2000 is underestimated . It spreads around $600\,$mm per year. The observed average over the 30 precipitation station used for calibration is about $150\,$mm higher. In all the graphs of Fig. 8.19, the step from year 2000 to 2001, when the data basis is changing from control period to scenario run, is marked by a large jump.

Dividing the GCM data into control and scenario run, trend signals can be found that are hidden when the data is combined (Fig. 8.20 and Fig. 8.21). The exceedance frequencies of the 95% and 99% quantile are increasing. The increase is strongest in the exceedance of the 99% quantile in summer months (Fig. 8.21e and Fig. 8.21f), indicating a shift in the distribution towards the highest events. In the same time the precipitation sum during summer months is decreasing (Fig. 8.21a and 8.21b) while the yearly precipitation sum stays about the same (Fig. 8.20a and Fig. 8.20b), indicating a shift in the precipitation activity towards winter.

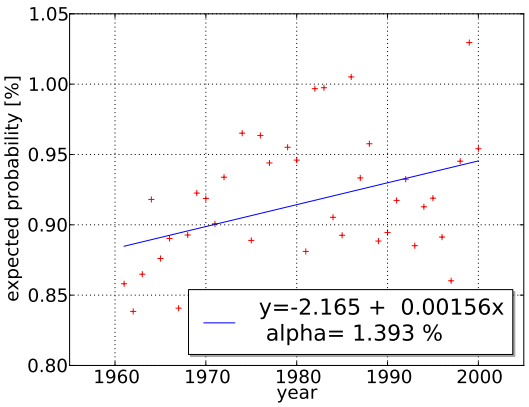
The trend signals in the control period and the scenario run correspond well. Except for the summer precipitation sum, the slopes of the linear regression function in both periods



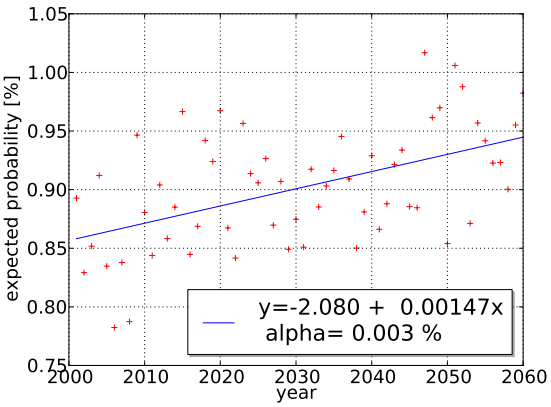
(a) Yearly Sum – Control Period



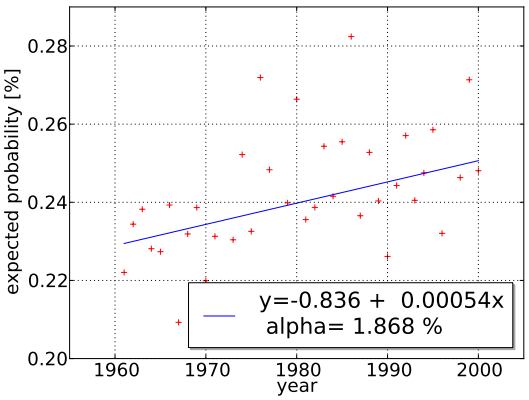
(b) Yearly Sum – Scenario A1B



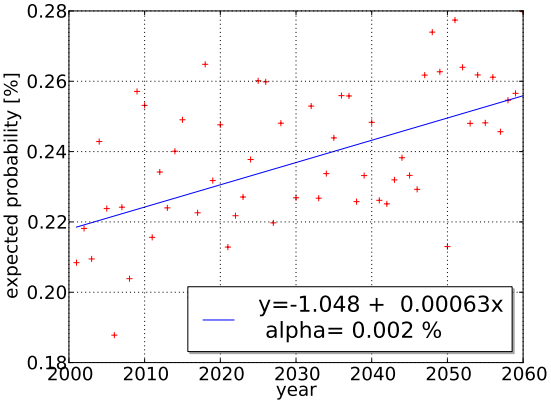
(c) 95% quantile – Control Period



(d) 95% quantile – Scenario A1B

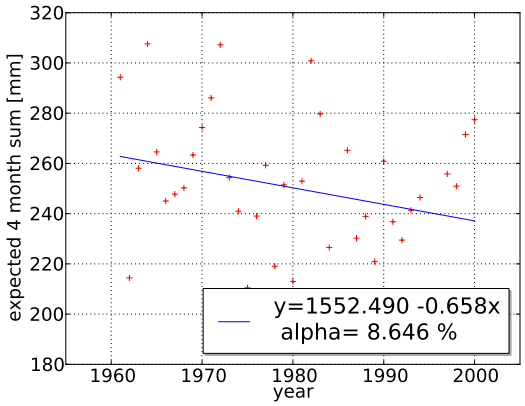


(e) 99% quantile – Control Period

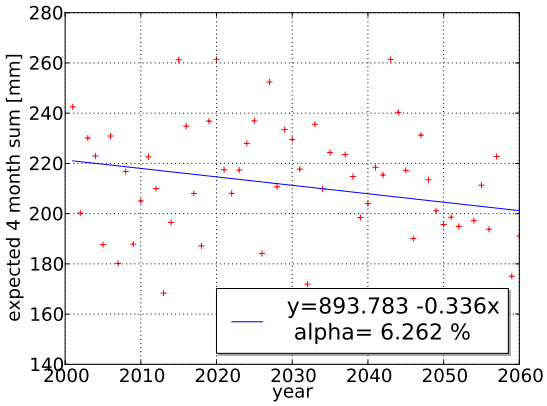


(f) 99% quantile – Scenario A1B

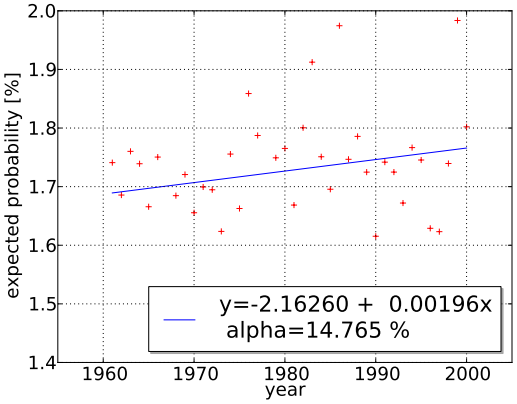
Figure 8.20.: Linear trend in the theoretic expectation of yearly precipitation sum and yearly extreme value probability according to the CP sequence derived from ECHAM5 MSLP field



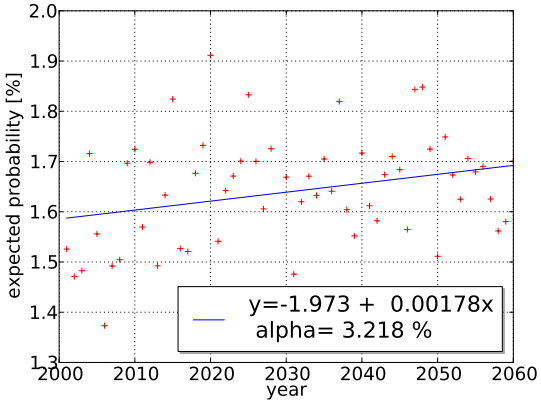
(a) Precipitation Sum – May to August



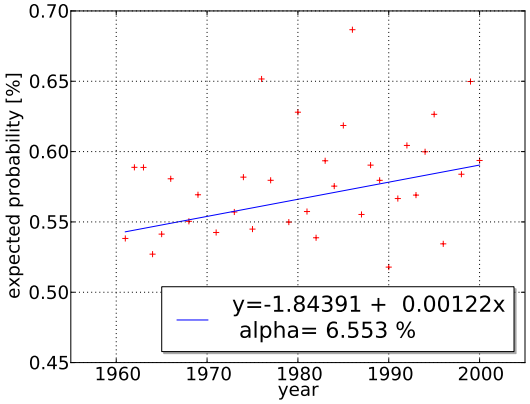
(b) Precipitation Sum – May to August



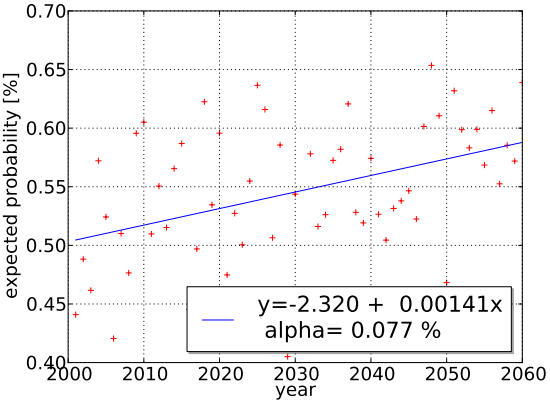
(c) 95% quantile – May to August



(d) 95% quantile – May to August



(e) 99% quantile – May to August



(f) 99% quantile – May to August

Figure 8.21.: Linear trend in the theoretic expectation of precipitation sum and extreme value probability for the summer months according to the CP sequence derived from ECHAM5 MSLP field

do not differ by more than some percent. On the other hand, the absolute values are discontinuous. At the break point between control and scenario run, from 2000 to 2001, the extreme value probabilities drop again to the level of 1960 to start increasing a new. Only in the precipitation sum, the end of the regression function of the control period is close to the starting point for the future period.

8.4. Conclusions

8.4.1. Reliability of Estimated Trend Signals

- The capacity of GCMs to model realistic precipitation time series is very limited. It is reported from literature that the variability and extreme value probability are underestimated. As the model uncertainties are scale dependent, it should be avoided to use GCM precipitation data in high temporal resolution. The use of aggregated data, for example monthly precipitation sums, is preferable.
- By QQ-Transformation, monthly GCM or RCM precipitation sums can be corrected according to observed precipitation values from a control period. However, QQ-Transformation, as most direct correction methods, is only valid on higher temporal aggregations. The correction of daily or subdaily data is not recommended.
- GCM predictions of Mean Sea Level Pressure (MSLP) and temperature are considered as more reliable than precipitation data. An alternative to the direct use of GCM or RCM precipitation is to relate predictions of MSLP and temperature to an expected precipitation distribution by an atmospheric Circulation Pattern (CP) classification system.

  Applying the classification, information on hourly temporal scale is derived from the GCM merely by using daily MLSPE and temperature values. Hence, it can be seen as a temporal downscaling technique whereas the precipitation distribution is estimated indirectly by a classification of past time observations. The underlying assumption is that the future precipitation response of each CP-temperature class is constant. Trends that are not related to CP or temperature or will be missed. The same can happen if the future brings new atmospheric conditions that where not observed during the calibration period and do not fit in the CP classification.
- If a trend in precipitation is found by different methods in different data sets, for example in observed rainfall records as well as in precipitation from GCM simulations or the expected precipitation according to the CP sequence in the GCM, it can be considered as a reliable signal and used for future prognosis.

  The increase in winter precipitation and decrease in summer precipitation, that is predicted by the CP-temperature sequence (Section 8.3), is confirmed by observed trends (Hundecha and Bárdossy, 2005) as well as by direct analysis of RCM precipitation (Déqué et al., 2007). The increase in extreme precipitation and the related shift in scaling toward more intensive short term events that has been deduced from observed

precipitation measurements (see Section 8.1) is also found in the expected 1 h precipitation statistics according to the CP-temperatures classification for the future (Section 8.3). In a physically sound interpretation, these purely statistical results can be linked to higher atmospheric temperature (Trenberth, 1999).

Therefore, these trend signals are considered as reliable predictions of the future. If so, the potential for water stress and catastrophic extreme events in summer are both increasing in the same time. Especially agriculture will suffer from the expected changes in summer. The effort for irrigation will raise, but more crops will be destroyed in extreme events (e. g. hail storms) too.

- Although the general tendencies of the trend signals are confirmed by different studies, high uncertainty remains about the absolute values. The CP sequence modeled by the GCM for example misses some of the trend signal in the observed CP time series. Especially, the decreasing trend in cyclonic situations during summer is ignored. If the trend to less cyclonic CPs continues, it will have two effects. On the one hand, the precipitation sum in summer will decrease faster than predicted by the CP sequence of the GCM. On the other hand, the extreme value probability will be lower than predicted because extreme precipitation is more frequent during cyclonic CPs.

8.4.2. Potential for Climate Change Adaptations in Rainfall Generation

In perspective of the estimated trends, there ares several possibilities for modifications of the resampling based rainfall generators presented in Chapter 6 and Chapter 7.

- The time series of monthly precipitation sum could be replaced by a QQ-Transformed time series of monthly precipitation from a GCM model.
- The daily rainfall frequency and daily average precipitation of the CP-groups that are used as a target function in the resampling scheme (see Section 6.6.2.1 and Section 7.6.2) could be replaced by a CP and temperature dependent classification according to Section 8.3. The CP and temperature time series could be derived from GCM data.
- The target values of the scaling exponent (Eq. (6.76)) in the first three statistical moments of precipitation on different aggregations from 1 h to 24 h, that is part of the objective function in the resampling scheme of the single site generation (Section 6.6.2.1) or the disaggregation from daily to hourly values in generation of simultaneous time series (Section 7.6.3), could be replaced by a time dependent function that extrapolates the trend observed between the two time periods from 1958 to 1980 and from 1981 to 2003 (Section 8.1.2)

Future predictions are not available for all statistics used in the presented resampling based rainfall generator. For example, it has to be assumed that the autocorrelation structure is not altered in the future. Replacing only some parameters by future projections induces heterogeneity in the parameter set which can alter the interactions in the generation algorithm.

The basic idea of the generators presented in Chapter 6 and Chapter 7 is that the interdependencies between the precipitation probability, as well as the average and the standard

deviation of the precipitation amounts in wet time steps can be expressed as a function of the the monthly sum. If the time series of the observed monthly sum was replaced by GCM data, it would imply two assumptions; firstly that the relation between these four statistics is the same in the future as it was in the observed past and secondly that the relation is not altered by errors in the GCM time series. Especially the second assumption is problematic, considering the high errors in uncorrected GCM precipitation. Correction methods, like QQ-Transformation, can only eliminate a part of the errors.

In the observations there is obviously a close relation between the monthly precipitation sum and the CP related precipitation statistics as each month consists of a 28 to 31 days long CP sequence. In the future predictions, the monthly precipitation sum and the CP related precipitation origin from different sources. While the monthly precipitation is a direct output of the GCM, the CP related precipitation arise from observed distributions that are only conditioned on the GCM. This can lead to conflicts in the permutation scheme. If the monthly precipitation sum derived from the CGM is far off the expected sum according to the CP sequence, the objective function part evaluating the CP dependent statistics will demand for unrealistic values and the optimisation will fail.

Finally, and most problematic for the use in rainfall generation, it is doubtful weather the ECHAM5 data set is homogeneous. The jump in most statistics between the control period and the future period indicates that it might not be the case. If there is a break point between the control run and the future scenario run, the error structure in the scenario is different than in the control run. This affects not only the hydrological response of the CP temperature sequence in the future but also breaks the assumption of the QQ-Transformation, Delta-Chance Method or other correction methods that the relation between model and observation of the control period is constant in the future.

8.4.3. Future Predictions as an Ensembles

A remaining problem is the choice of the emission scenario and most of all of the Circulation Model from which the data is taken. For the scenarios, one could argue by the aim of the simulation, for example to always take the A2 emission scenario for a "worst case" study with the most pronounced changes. However, such reasoning is not possible for choice of the Circulation Model as there is no general order from the model with the least changes to the model with the most severe consequences. A priori all models have the same probability of giving a realistic picture of the future. All claim to represent the major physical processes involved in the movement of air and moisture in the atmosphere. Which of the models performs best cannot be judges from today's point of view. If only one model scenario combination is chosen, it is very likely that it *is not* the combination giving the best prediction of future climate conditions.

A good way to deal with the uncertainty is to consider all the different data sets as an ensemble. It means that the rainfall generation is run several times with all available data sets from the different model-scenario combinations. The resulting time series are used as a range of possible future precipitation. All following applications, as for example hydraulic sewage system modeling, have to be run with the whole ensemble too. The result is not one

crisp prognosis of the future conditions but a range of possible outcomes considering the uncertainty in the applied models and future scenarios. For the hydraulic application, for example the sewage system dimensioning, the user can judge the uncertainties and account for it by engineering solutions.

Of course, this is an ambitious task as it multiplies the effort for the computationally demanding rainfall generation and all following hydrological or hydraulic applications. If the same GCM and reanalysis data as in Africa case study was available, it would imply the set up 18 different data bases for the rainfall generator (with 18 different time series of the monthly precipitation sums and of the sequence of CP and temperature classes) and to run it 18 times for each location.

However, even with the ensemble data set, there is no guarantee that the future lies within the predicted range. If all the models share a common error, for example in the representation of the physical processes or the initial conditions, they can all be off target.

9. Conclusions and Outlook

The monthly precipitation sum is identified as one of the main statistical characteristics of precipitation. The rainfall probability, as well as the mean and standard deviation are to some extend determined by the monthly rainfall sum. Therefore, the distribution of hourly rainfall depths can be deduced if the monthly sum is known. The relation is exploited for the generation of hourly rainfall time series that are conditioned on the time series of monthly sums. Since the monthly sum exhibits low spatial variability, the regionalization to ungauged locations is easily conducted by geostatistical methods.Therefore, the developed time series generator can be applied at every location in the study region of Baden-Württemberg.

The interdependencies of the four statistical parameters monthly sum, rainfall probability, mean and standard deviation are described by a Gaussian copula in the uniform $[0, 1]$ space. The interdependencies are a characteristic of the regional climatic conditions and thus regarded as constant over the study region. Local effects find their expression in the absolute values. They are described by the marginal distributions.

The developed time series generator performs a two step simulation. First, an initial time series of hourly values is set up. Then, it is optimized in a permutation scheme according to target values of different statistics concerning the temporal persistence, the extreme value probabilities and the temporal scaling over different aggregations.

A comparison between simulated and observed time series shows that the generator delivers realistic simulations. It was tested by extreme value statistics on different temporal aggregations and by the distribution of wet and dry spell lengths as a measure of the temporal persistence. Since a data driven approach takes maximum benefit from the high density of rain gauges in Baden-Württemberg, it produces realistic results on a wide range of temporal scales from $1\,\mathrm{h}$ to $24\,\mathrm{h}$. This is a clear advantage compared to most conceptual rainfall models that perform best on the scale of the calibration data, but lack of variance in higher temporal aggregations. As another feature, the developed generator is able to model the difference in distribution between summer and winter months. It represents the seasonal variability in a realistic manner.

The generation scheme was extended for the generation of several, spatially interdependent, simultaneous time series. The spatial interdependencies are considered on three different scales: on the monthly scale by a Gaussian copula that describes the interdependencies of the generation parameters at all simulation locations, by the correlation of daily precipitation sums and by the correlation of hourly values conditioned on the atmospheric Circulation Pattern (CP).

The spatial temporal interdependencies of observed precipitation measurements are statistically very complex and cannot be considered without simplifications. The generations for locations of very short distance (few kilometers) are especially difficult. For some configurations of the station locations there are inconsistencies between the parameter interdependencies and the target values of daily correlation. Secondly, the estimation of the CP dependent correlations is subject to high uncertainty due to the low number of observed rain gauges in short distances. Generally, there is high spread in the parameter interdependencies of different locations. Systematic effects, however, that are responsible for the high observed variability, e. g. anisotropy or non-stationarity in the relations, could not be identified.

In recent years, researchers have explored new methods for the description of dependence. Bárdossy and Pegram (2009) analyze the joint exceedance probabilities of different threshold in triplets of rain gauge records by means of the Shannon entropy. Observed triplets exhibit lower Shannon entropy than triplets modeled by correlation matrices or copula. Since the Shannon entropy is a measure of disorder, it means that there is a part of the interdependence information that is ignored by pairwise correlations, e. g. a higher probability of simultaneous extreme events. Incorporating such information, especially of the tail dependence in the highest events, could further improve the generation of simultaneous time series.

For small sewage systems hourly temporal resolution is not yet sufficient and the generated time series have to be further disaggregated to five minute values. Hydraulic sewage system models react very sensitively to persistence on short temporal scale. The way in which the disaggregation is performed has a high influence on the modeling result. Therefore, the development of an accurate disaggregation scheme for simultaneous time series should be the next research task in this context. The representation of spatial dependence in five minute values, however, will be difficult. Probably, the interdependencies are limited to a close range of several kilometers that is very poorly sampled in the existing data set. Furthermore, the simultaneous five minute time series are highly affected by synchronization errors between the rain gauges.

Another important research task is the estimation of rainfall probability, which is one of the main parameters in the developed rainfall generation algorithms. If the estimated rainfall probability is wrong, it affects many of the other statistical parameters too, e. g. the average and the standard deviation of precipitation of wet time steps or the scaling behavior. Unfortunately, the relative error in rain gauge measurements is highest in the smallest values. Therefore, observed rainfall frequencies are subject to high uncertainty. For accurate calibration of the rainfall generators a more robust estimation of rainfall probability would be a clear improvement. A possible approach could be to extrapolate the rainfall probability from the exceedance frequencies of higher thresholds that are less susceptible to measurement errors.

A strong point of the developed rainfall models is the applicability in climate change studies. In the last decade climate model output was evaluated in various studies. It has been shown that GCMs model real world climatic conditions insufficiently, especially in terms of precipitation. Although GCMs are permanently improved, the global climate is probably way too complex to be ever modeled correctly. Therefore, studies are promising that change

the focus from "best guess" for a certain target year in the future to "what if" like the work of Kendon et al. (2010) that was presented in Chapter 8. Such studies contribute to a better understanding of the climate system, independently of the precise outcome of the climate model.

In this context, data driven approaches can be of great benefit for the understanding of climate change effects on precipitation of high temporal resolution. The structure of the developed rainfall generation models allows to modify every generation parameter separately. Since all parameters have clear statistical meanings, they can be adapted to the results of climate change studies. In Section 8.1.2 for example, it was shown that climate change effects the scaling of precipitation on different aggregations. This could be easily modeled by the modification of the respective parameter in the objective functions. In the same way, a higher variability in monthly sums, a change in distribution or a reduced rainfall probability in summer months could be modeled. A comparison of the model results can be used to quantify the effect of each possible change signal as well as potential reinforcement or compensation effects in their combination.

Bibliography

E. Aarts and P. van Laarhoven. Simulated annealing: An introduction. *Statistica Neerlandica*, 43:31–52, 1989.

M. C. Acreman. A simple stochastic model of hourly rainfall for farnborough, england. *Hydrological Sciences Journal*, 35(2):119–148, 1990.

A. AghaKouchak, A. Bárdossy, and E. Habib. Conditional simulation of remotely sensed rainfall data using non-gaussian v-transformed copula. *Advances in Water Resources*, 33: 624–634, 2010.

A. AghaKouchak, A. Behrangi, S. Sorooshian, K. Hsu, and E. Amitai. Evaluation of satellite-retrieved extreme precipitation rates across the central united states. *Journal of Geophysical Research*, 116:D02115, 2011.

H. Andrieu and J. D. Creutin. Identification of vertical profiles of radar reflectivity for hydrological applications using an inverse method. part i: Formulation. *Journal of Applied Meteorology*, 34:225–239, 1995.

P. Arnaud. Coupled rainfall model and discharge model for flood frequency estimation. *Water Resources Research*, 38(6):1075, 2002.

P. Arnaud and J. Lavabre. Using a stochastic model for generating hourly hyetographs to study extreme rainfalls. *Hydrological Sciences Journal*, 44(3):433–446, 1999.

P. Arnaud, J. Lavabre, and J. A. Fine. An hourly rainfall generation model applicable to all types of climate. *Atmospheric Research*, 85:230–242, 2007.

K. Arpe and E. Roeckner. Simulation of the hydrological cycle over europe: Model validation and impacts of increasing greenhouse gases. *Advances in Water Resources*, 23:105–119, 1999.

S. T. Ashley and W. S. Ashley. Flood fatalities in the united states. *Journal of Applied Meteorology and Climatology*, 47(3):808–818, 2008.

I. Auer, R. Böhm, A. Jurkovic, W. Lipa, A. Orlik, R. Potzmann, W. Schöner, M. Ungersböck, C. Matulla, K. Briffa, P. Jones, D. Efthymiadis, M. Brunetti, T. Nanni, M. Maugeri, L. Mercalli, O. Mestre, J.-M. Moisselin, M. Begert, G. Müller-Westermeier, V. Kveton, O. Bochnicek, P. Stastny, M. Lapin, S. Szalai, T. Szentimrey, T. Cegnar, M. Dolinar, M. Gajic-Capka, K. Zaninovic, Z. Majstorovic, and E. Nieplova. Histalp–historical instrumental climatological surface time series of the greater alpine region. *International Journal of Climatology*, 27(1):17–46, 2007. ISSN 1097-0088. doi: 10.1002/joc.1377.

P. M. Austin. Relation between measured radar reflectivity and surface rainfall. *Monthly Weather Review*, 115:1053–1170, 1987.

P. M. Austin and R. A. Houze. Analysis of the structure of precipitation patterns in new england. *Journal of Applied Meteorology*, 11:926–934, 1972.

M. L. Baeck and J. A. Smith. Rainfall estimation by the wsr-88d for heavy rainfall events. *Weather and Forecasting*, 13(2):416–436, 1998.

H. Bartels, N. Demuth, B. Dietzer, C. Kollmann, M. Moser, O. Plörer, V. Stalmann, W. Timmerbrink, and H. Verworn. *MANOB - Mindestanforderungen an automatische Niederschlagsmessgeräte (Ombrometer) und deren Betrieb Richtlinie des Deutschen Wetterdienstes*. Deutscher Wetterdienst, Offenbach am Main, 1999.

H. Bartels, B. Dietzer, G. Malitz, F. M. Albrecht, and J. Guttenberger. Kostra dwd - starkniederschlagshöhen für deutschland (1951 - 2000) - fortschreibungsbericht. Technical report, Deutscher Wetterdienst, 2005.

M. Bartels, E. Weigl, T. Reich, P. Lang, A. Wagner, O. Kohler, and N. Gerlach. Projekt radolan–routineverfahren zur online-aneichung der radarniederschlagsdaten mit hilfe von automatischen bodenniederschlagsstationen (ombrometer). Abschlussbericht, Deutscher Wetterdienst, Offenbach, 2004.

P. Bernardara, C. De Michele, and R. Rosso. A simple model of rain in time: An alternating renewal process of wet and dry states with a fractional (non-gaussian) rain intensity. *Atmospheric Research*, 84:291–301, 2007.

A. Berne, G. Delrieu, J.-D. Creutin, and C. Obled. Temporal and spatial resolution of rainfall measurements required for urban hydrology. *Journal of Hydrology*, 299:166–179, 2004.

J. Bliefernicht. *Probability Forecasts of Daily Areal Precipitation for Small River Basins*. PhD thesis, Institut für Wasserbau, Universität Stuttgart, 2010.

J. Boé, L. Terray, F. Habets, and E. Martin. Statistical and dynamical downscaling of the seine basin climate for hydro-meteorological studies. *International Journal of Climatology*, 27:1643–1655, 2007.

A. Bárdossy. Generating precipitation time series using simulated annealing. *Water Resources Resarch*, 34(7):1737–1744, 1998.

A. Bárdossy. Atmospheric circulation patterns classification for south-west germany using hydrolodical variables. *Physics and Chemistry of the Earth*, 35:498–506, 2010.

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

A. Bárdossy and G. G. S. Pegram. Copula based multisite model for daily precipitation simulation. *Hyrdology and Earth System Sciences*, 13:2299–2314, 2009.

A. Bárdossy and E. J. Plate. Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resources Resarch*, 28(5):1247–1259, 1992.

A. Bárdossy, H. Giese, B. Haller, and J. Ruf. Erzeugung synthetischer niederschlagszeitreihen in hoher zeitlicher auflösung für baden-württemberg. *Wasserwirtschaft*, 90:548–553, 2000.

J. Brommundt. *Stochastische Generierung räumlich zusammenhängender Niederschlagszeitreihen*. PhD thesis, Institut für Wasserbau, Universität Suttgart, 2008.

A. Bunde, J. F. Eichner, J. W. Kantelhardt, and S. Havlin. Long-term memory: A natural mechanism for the clustering of extreme events and anomalous residual times in climate records. *Physical Review Letters*, 94(4):048701, 2005.

P. Burlando and R. Rosso. Scaling and muitiscaling models of depth-duration-frequency curves for storm precipitation. *Journal of Hydrology*, 187:45–64, 1996.

A. Burton, K. C. G., H. J. Fowler, P. S. P. Cowpertwait, and P. O'Connell. Rainsim: A spatial-temporal stochastic rainfall modelling system. *Environmental Modelling & Software*, 23: 1356–1369, 2008.

G. Calenda and F. Napolitano. Parameter estimation of neyman-scott processes for temporal point rainfall simulation. *Journal of Hydrology*, 225:45–66, 1999.

E. H. Chin. Modeling daily precipitation occurrence process with markov chain. *Water Resources Resarch*, 13(6):949–956, 1977.

O. Christensen and J. Christensen. Intensification of extreme european summer precipitation in a warmer climate. *Global and Planetary Change*, 44:107–117, 2004.

G. J. Ciach and W. F. Krajewski. Analysis and modeling of spatial correlation structure in small-scale rainfall in central oklahoma. *Advances in Water Resources*, 29(10):1450–1463, 2006.

C. Collier. Accuracy of rainfall estimates by radar, part i: Calibration by telemetering raingauges. *Journal of Hydrology*, 83:207–223, 1986.

P. S. P. Cowpertwait. A generalized point process model for rainfall. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 447(1929):23–37, 1994.

P. S. P. Cowpertwait. A space-time neyman-scott model of rainfall: Empirical analysis of extremes. *Water Resources Research*, 38(8):1131, 2002.

P. S. P. Cowpertwait. A spatial-temporal point process model of rainfall for the thames catchment, uk. *Journal of Hydrology*, 330:586–595, 2006.

P. S. P. Cowpertwait, P. O'Connell, A. V. Metcalfe, and J. A. Mawdsley. Stochastic point process modelling of rainfall. ii. regionalisation and disaggregation. *Journal of Hydrology*, 175:47–65, 1996.

P. S. P. Cowpertwait, V. Isham, and C. Onof. Point process models of rainfall: developments for fine-scale structure. *Proceedings of the R*, 463:2569–2587, 2007.

G. Delrieu, H. Andrieu, and J. D. Creutin. Quantification of path-integrated attenuation for x- and c-band weather radar systems operating in mediterranean heavy rainfall. *Journal of Applied Meteorology*, 39:840–850, 2000.

M. Déqué, D. Rowell, D. Lüthi, F. Giorgi, J. Christensen, B. Rockel, D. Jacob, E. Kjellström, M. de Castro, and B. van den Hurk. An intercomparison of regional climate simulations for europe: assessing uncertainties in model projections. *Climatic Change*, 81:53–70, 2007. ISSN 0165-0009. 10.1007/s10584-006-9228-x.

U. Ehret, J. Götzinger, A. Bárdossy, and G. G. S. Pegram. Radar-based flood forecasting in small catchments, exemplified by the goldersbach catchment, germany. *International Journal of River Basin Management*, 6:323–329, 2008.

A. Elizalde, D. Sein, U. Mikolajewicz, and D. Jacob. Technical report: Atmosphere-ocean-hydrology coupled regional climate model. Technical report, Max Planck Institute for Meteorology, Hamburg, Germany, 2011.

S. Emeis. *Measurement Methods in Atmospheric Sciences*. Bortraeger Science Publishers, Stuttgart, 2010.

G. Evin and A.-C. Favre. A new rainfall model based on the neyman-scott process using cobic copulas. *Water Resources Resarch*, 44:W03433, 2008.

L. Ferraris, S. Gabellani, U. Parodi, and N. Rebora. Revisiting multifractality in rainfall fields. *Journal of Hydrometeorology*, 4:544–551, 2002.

E. Foufoula-Georgiou and P. Guttorp. Compatibility of continuous rainfall occurrence models with discrete rainfall observations. *Water Resources Resarch*, 22(8):1316–1322, 1986.

H. J. Fowler, M. Ekström, and S. Blenkinsop. Estimating change in extreme european precipitation using a multimodel ensemble. *Journal of Geophysical Research*, 112:D18104, 2007.

C. Frei, J. H. Christensen, M. Déqué, D. Jacob, and P. L. Jones. Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the european alps. *Journal of Geophysical Research*, 108:0148–0227, 2003.

K. R. Gabriel and J. Neumann. A markov chain model for daily rainfall occurrence at tel aviv. *Quarterly Journal of the Royal Meteorological Society*, 88:90–95, 1962.

S. Geng, F. W. P. de Vries, and I. Supit. A simple method for generating daily rainfall data. *Agricultural and Forest Meteorology*, 36(4):363–376, 1986.

U. Germann and J. Joss. Mesobeta profiles to extrapolate radar precipitation measurements above the alps to the ground level. *Journal of Applied Meteorology*, 41:542–557, 2002.

U. Germann, M. Berenguer, D. Sempere-Torres, and M. Zappa. Real - ensemble radar precipitation estimation for hydrology in a mountainous region. *Quarterly Journal of the Royal Meteorological Society*, 135:445–456, 2009.

E. Goudenhoofdt and L. Delobbe. Evaluation of radar-gauge merging methods for quantitative precipitation estimates. *Hydrology and Earth System Sciences*, 13:195–2006, 2009.

P. Y. Groisman and D. R. Legates. The accuracy of united states precipitation data. *Bulletin of the American Meteorological Society*, 75(2):215–227, 1994.

E. Gumbel. *Statistics of extremes*. Columbia University Press, New York, 1958.

V. K. Gupta and E. Waymire. Multiscaling properties of spatial rainfall and river flow distributions. *Journal of Geophysical Research*, 95:1999–2009, 1990.

V. K. Gupta and E. C. Waymire. A statistical analysis of mesoscale rainfall as a random cascade. *Journal of Applied Meteorology*, 32:251–267, 1993.

C. T. Haan, D. M. Allen, and J. O. Street. A markov chain model of daily rainfall. *Water Resources Resarch*, 12(3):443–449, 1976.

S. Habemann, K. Arpe, and E. Roeckner. Evaluation of the hydrological cycle in the echam5 model. *Journal of Climate*, 19(19):3810–3827, 2006.

U. Haberlandt. Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *Journal of Hydrology*, 332(1-2):144–157, Jan. 2007.

U. Haberlandt, A.-D. Ebner von Eschenbach, and I. Buchwald. A space-time hybrid hourly rainfall model for derived flood frequency analysis. *Hydrology and Earth System Sciences*, 12:1353–1367, 2008.

U. Haberlandt, A. Belli, and J. Hölscher. Trends in beobachteten zeitreihen von temperatur und niederschlag in niedersachsen. *Hydrologie und Wasserbewirtschaftung*, 54:28–36, 2010.

D. Harris, A. Seed, M. Menabde, and G. Austin. Factors affecting multiscaling analysis of rainfall time series. *Nonlinear Processes in Geophysics*, 4(3):137–156, 1997.

J. Hartung. *Statistik - Lehr und Handbuch der angewandten Statistik*. Oldenburg, fifteenth edition, 2009.

C. Haslauer. *Analysis of Real-World Spatial Dependence of Subsurface Hydraulic Properties Using Copulas With a Focus on Solute Transport Behaviour*. PhD thesis, Institut für Wasserbau, Universität Stuttgart, 2011.

M. R. Haylock and C. M. Goodess. Interannual variability of european extreme winter rainfall and links with mean large-scale circulation. *International Journal of Climatology*, 24(6): 759–776, 2004. ISSN 1097-0088. doi: 10.1002/joc.1033.

I. M. Held and B. J. Soden. Robust responses of the hydrological cycle to global warming. *Journal of Climate*, 19:5686–5699, 2006.

N. J. Higham. Computing the nearest correlation matrix – a problem from finance. *IMA Journal of Numerical Analysis*, 22(3):329–343, 2002.

J. Hubbert, D. M., and M. Ellis. Weather radar ground clutter. part ii: Real-time identification and filtering. *Journal of Atmospheric and Oceanic Technology*, 26:1181–1197, 2009.

G. J. Huffman, R. F. Adler, D. T. Bolvin, G. Gu, E. J. Nelkin, K. P. Bowman, Y. Hong, E. F. Stockner, and D. B. Wolff. The trmm multisatellite precipitation analysis (tmpa): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8:38–55, 2007.

Y. Hundecha and A. Bárdossy. Trends in daily precipitation and temperature extremes across western germany in the second half of the 20th century. *International Journal of Climatology*, 25(9):1189–1202, 2005. ISSN 1097-0088. doi: 10.1002/joc.1182.

M. Ignaccolo and C. D. Michele. Skewness as measure of the invariance of instantaneous renormalized drop diameter distributions - part 1: Convective vs. stratiform precipitation. *Hydrology and Earth System Sciences*, 16:319–327, 2012a.

M. Ignaccolo and C. D. Michele. Skewness as measure of the invariance of instantaneous renormalized drop diameter distributions - part 2: Orographic precipitation. *Hydrology and Earth System Sciences*, 16:329–343, 2012b.

IPCC. *Fourth Assessment Report: Climate Change 2007: The AR4 Synthesis Report*. Geneva: IPCC, 2007. URL `http://www.ipcc.ch/ipccreports/ar4-wg1.htm`.

ITWH. *KOSIM – das KOntinuierliche Langzeit-SIMulationsmodell für den Nachweis und die Bemessung von Entwässerungsbauwerken.* Institut für technisch-wissenschaftliche Hydrologie GMBH, Hannover, 2009.

JAXA. Trmm page of the japan aerospace exploration agency, 6 2012. URL `http://www.eorc.jaxa.jp/TRMM/index_e.htm`.

T. C. Johns, J. M. Gregory, W. J. Ingram, C. E. Johnson, A. Jones, J. A. Lowe, J. F. B. Mitchell, D. L. Roberts, D. M. H. Sexton, D. S. Stevenson, S. F. B. Tett, and M. J. Woodage. Anthropogenic climate change for 1860 to 2100 simulated with the hadcm3 model under updated emissions scenarios. *Climate Dynamics*, 20:583–612, 2003. ISSN 0930-7575. 10.1007/s00382-002-0296-y.

R. W. Katz. Computing probabilities associated with the markov chain model for precipitation. *Journal of Applied Meteorology*, 13:953–954, 1974.

R. W. Katz. Precipitation as a chain-dependent process. *Journal of Applied Meteorology*, 16(7): 671–676, 1977.

R. W. Katz and M. B. Parlange. Effects of an index of atmospheric circulation on stochastic properties of precipitation. *Water Resources Resarch*, 29:2335–2344, 1993.

R. W. Katz and M. B. Parlange. Generalizations of chain-dependent processes: Application to hourly precipitation. *Water Resources Resarch*, 31(5):1331–1341, 1995.

R. W. Katz and M. B. Parlange. Overdispersion phenomenon in stochastic modeling of precipitation. *Journal of Climate*, 11:591–601, 1998.

E. Kendon and R. Clark. Reliability of future changes in heavy rainfall over the uk. In *BHS 10th National Hydrology Symposium*, Exeter, 2008.

E. Kendon, D. Rowell, and R. Jones. Mechanisms and reliability of future projected changes in daily precipitation. *Climate Dynamics*, 35:489–509, 2010. ISSN 0930-7575. 10.1007/s00382-009-0639-z.

C. Kidd. Satellite rainfall climatology: a review. *International Journal of Climatology*, 21(9): 1041–1066, 2001. ISSN 1097-0088. doi: 10.1002/joc.635.

R. Kistler, E. Kalnay, W. Collins, S. Saha, G. White, J. Woollen, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, H. van den Dool, R. Jenne, and M. Fiorino. The ncep–ncar 50-year reanalysis: Monthly means cd-rom and documentation. *Bulletin of the American Meteorological Society*, 82:247–268, 2001.

G. E. Klazura, D. S. Kelly, and P. Jendrowski. A comparison of nexrad wsr-88d radar estimates of rain accumulation with gauge measurements for high- and low-reflectivity horizontal gradient precipitation events. *Journal of Atmospheric and Oceanic Technology*, 16(11): 1842–1850, 1999.

D. Koutsoyiannis. The hurst phenomenon and fractional gaussian noise made easy. *Hydrological Sciences Journal*, 47(3):573–595, 2002.

D. Koutsoyiannis. Uncertainty, entropy, scaling and hydrological statistics. 1. marginal distributional properties of hydrological processes and state scaling. *Hydrological Sciences Journal*, 50(3):381–404, 2005.

D. Koutsoyiannis. An entropic-stochastic representation of rainfall intermittency: The origin of clustering and persistence. *Water Resources Resarch*, 42:W01401, 2006.

D. Koutsoyiannis and E. Foufoula-Georgiou. A scaling model of a storm hyetograph. *Water*, 29:2345–2361, 1993.

D. Koutsoyiannis, C. Onof, and H. S. Wheater. Multivariate rainfall disaggregation at a fine timescale. *Water Resources Research*, 39:1173, 2003.

G. W. Lee and I. Zawadski. Variability of drop size distributions: Time-scale dependence of the variability and its effects on rain estimation. *Journal of Applied Meteorology*, 44:241–255, 2005.

E. Lloyd. What is, and what is not, a markov chain? *Journal of Hydrology*, 22:1–28, 1974.

S. Lovejoy and B. B. Mandelbrot. Fractal properties of rain, and a fractal model. *Tellus A*, 37A:209–232, 1985.

S. Lovejoy and D. Schertzer. Scale invariance, symmetries, fractals, and stochastic simulations of atmospheric phenomena. *Bulletin of the American Meteorological Society*, 67(1): 21–32, 1986.

P. J. Mailier, D. B. Stephenson, C. A. Ferro, and K. I. Hodges. Serial clustering of extratropical cyclones. *Monthly Weather Review*, 134(8):2224–2240, 2006.

B. Mandelbrot. Introduction to fractal sums of pulses. In M. Shlesinger, G. Zaslavsky, and U. Frisch, editors, *Lévy Flights and Related Topics in Physics*, volume 450 of *Lecture Notes in Physics*, pages 110–123. Springer Berlin / Heidelberg, 1995. ISBN 978-3-540-59222-8.

B. B. Mandelbrot. *The fractal geometry of nature*. Wh Freeman, 1983.

D. Marsan, D. Schertzer, and S. Lovejoy. Causal space-time multifractal processes: Predictability and forecasting of rain fields. *Journal of Geophysical Research*, 101:333–346, 1996.

M. Menabde, A. Seed, and G. G. S. Pegram. A simple scaling model for extreme rainfall. *Water Resources Resarch*, 35:335–339, 1999.

K. S. Mitra, O. Vohl, M. Ahr, and H. R. Pruppacher. A wind tunnel and theoretical study of the melting behaviour of atmospheric ice particles. iv: Experiment and theory for snow flakes. *Journal of the atmospheric sciences*, 47(5):584–591, 1990.

P. Molnar and P. Burlando. Preservation of rainfall properties in sthochastic disaggregation by a simple random cascade model. *Atmospheric Research*, 77:137–151, 2005.

A. Muller, P. Arnaud, M. Lang, and J. Lavabre. Uncertainties of extreme rainfall quantiles estimated by a stochastic rainfall model and by a generalized pareto distribution. *Hydrological Sciences Journal*, 53(3):417–429, 2009.

N. Nakicenovic and R. Swart, editors. *IPCC SPECIAL REPORT EMISSIONS SCENARIOS: Summary for Policymakers: A Special Report of IPCC Working Group III*. Published for the Intergovernmental Panel on Climate Change, 2000. ISBN 92-91691135.

R. B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics, second edition, 2006.

C. Onof and H. S. Wheater. Modelling of british rainfall using a random parameter bartlett-lewis rectangular pulse model. *Journal of Hydrology*, 149:67–95, 1993.

C. Onof, R. E. Chandler, A. Kakou, H. S. Wheater, and V. Isham. Rainfall modelling using poisson-cluster processes: a review of developments. *Stochastic Environmental Research and Risk Assessment*, 14(6):384–411, 2000.

C. Onof, J. Townend, and R. Kee. Comparison of two hourly to 5-min rainfall disaggregators. *Atmospheric Research*, 77:176–187, 2005.

T. Osborn and M. Hulme. Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. *Journal of Climate*, 10:1885–1905, 1997.

G. G. S. Pegram and A. N. Clothier. High resolution space-time modelling of rainfall: the string of beads model. *Journal of Hydrology*, 241:26–41, 2001.

B. Renard and M. Lang. Use of a gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Advances in Water Resources*, 30:897–912, 2007.

M. Rico-Ramirez and I. Cluckie. Classification of ground clutter and anomalous propagation using dual-polarization weather radar. *Geoscience and Remote Sensing, IEEE Transactions on*, 46:1892–1904, 2008.

I. Rodríguez-Iturbe, B. F. D. Power, and J. B. Valdés. Rectangular pulses point process models for rainfall: Analysis of empirical data. *Journal of Geophysical Research*, 92:9646–9656, 1987.

I. Rodriguez-Iturbe and P. S. Eagleson. Mathematical models of rainstorm events in space and time. *Water Resources Resarch*, 23:181–190, 1987.

I. Rodriguez-Iturbe, D. R. Cox, and V. Isham. Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 410:269–288, 1987.

R. Rodriguez-Iturbe, D. R. Cox, and E. P. S. Spatial modelling of total storm rainfall. *Proceedings of the Royal Society, London*, 403:27–50, 1986.

E. Roeckner, G. Bäuml, L. Bonaventura, R. Brokopf, M. Esch, M. Giorgetta, S. Hagemann, I. Kirchner, L. Kornblueh, E. Manzini, A. Rhodin, U. Schlese, U. Schulzweida, and A. Tompkins. The atmospheric general circulation model echam5. Technical report, Max-Planck-Institut für Meteorologie, 2003.

E. Roeckner, R. Brokopf, M. Esch, M. Giorgetta, S. Hagemann, L. Kornblueh, E. Manzini, U. Schlese, and U. Schulzweida. Sensitivity of simulated climate to horizontal and vertical resolution in the echam5 atmosphere model. *Journal of C*, 19(16):3771–3791, 2006.

V. Ruiz-Villanueva, M. Borga, D. Zoccatelli, L. Marchi, E. Gaume, U. Ehret, and E. Zehe. Extreme runoff response to short-duration convective rainfall in south-west germany. *Hydrology and Earth System Sciences*, 16:1543–1559, 2012.

G. Salvadori and C. De Michele. Multivariate multiparameter extreme value models and return periods: A copula approach. *Water Resources Resarch*, 46(10):W10501, 2010.

C. Schoelzel and P. Friederichs. Multivariate non-normally distributed random variables in climate research - introduction to the copula approach. *Nonlinear Processes in Geophysics*, 15(5):761–772, 2008.

J. Sennikovs and U. Bethers. Statistical downscaling method of regional climate model results for hydrological modelling. In *18th World IMACS / MODSIM Congress*, volume 18, 2009.

B. Sevruk, M. Ondrás, and B. Chvíla. The wmo precipitation measurement intercomparisons. *Atmospheric Research*, 92(3):376–380, 2009.

S. Sorooshian, A. AghaKouchak, P. Arkin, J. Eylander, E. Foufoula-Georgiou, R. Harmon, J. M. H. Hendrickx, B. Imam, R. Kuligowskski, B. Skahill, and G. Skofronick-Jackskson. Advanced concepts on remote sensing of precipitation at multiple scales. *Bulletin of the American Meteorological Society*, 92:1353–1357, 2011.

P. Todorovic and D. A. Woolhiser. A stochastic model of n-day precipitation. *Journal of Applied Meteorology*, 14:17–24, 1975.

K. E. Trenberth. Conceptual framework for changes of extremes of the hydrological cycle with climate change. *Climatic Change*, 42:327–339, 1999. ISSN 0165-0009. 10.1023/A:1005488920935.

M. J. van den Berg, S. Vandenberghe, B. De Baets, and N. E. C. Verhoest. Copula-based downscaling of spatial rainfall: a proof of concept. *Hydrology and Earth System Sciences*, 15:1445–1457, 2011.

W. J. Vanhaute, S. Vandenberghe, K. Scheerlinck, B. De Baets, and N. E. C. Verhoest. Calibration of the modified bartlett-lewis model using global optimization techniques and alternative objective functions. *Journal of Hydrology and Earth System Sciences*, 16(3):873–891, 2012.

A. Verworn and U. Haberlandt. Spatial interpolation of hourly rainfall - effect of additional information, variogram inference and storm properties. *Hydrology and Earth System Sciences*, 15:569–584, 2011.

R. Vitolo, D. B. Stephenson, I. M. Cook, and K. Mitchell-Wallace. Serial clustering of intense european storms. *Meteorologische Zeitschrift*, 18(4):411–424, 2009. doi: doi:10.1127/0941-2948/2009/0393.

W. Weischet and W. Endlicher. *Einführung in die allgemeine Klimatologie*. Gebrüder Borntraeger Verlagsbuchhandlung, seventh edition, 2008.

F. J. Wentz, L. Ricciardulli, K. Hilburn, and C. Mears. How much more rain will global warming bring? *Science*, 317:233–235, 2007.

K. J. Weston and M. G. Roy. The directional-dependence of the enhancement of rainfall over complex orography. meteorological applications. *Meteorological Applications*, 1(3):267–275, 1994.

WHG. Wasserhaushaltsgesetz der bundesrepublik deutschland, 2009.

D. Wilks. Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, 210:178–191, 1998.

D. S. Wilks. *Statistical Methods in Atmospheric Sciences*. Academic Press, third edition, 2011.

P. Willems. A spatial rainfall generator form small spatial scales. *Journal of Hydrology*, 252: 126–144, 2001.

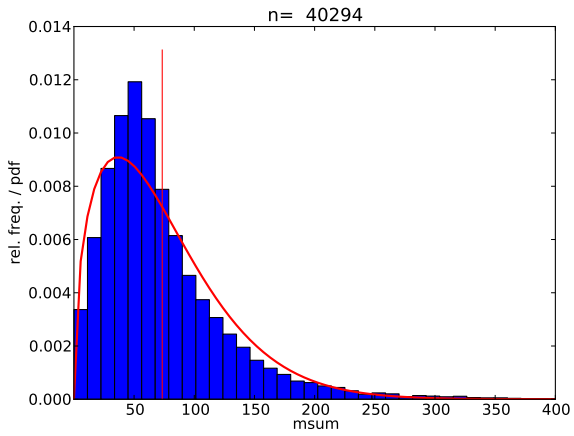
P. Willems and M. Vrac. Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change. *Journal of Hydrology*, 402:193–205, 2011.

L. Zhang and V. P. Singh. Bivariate rainfall frequency distributions using archimedean copulas. *Journal of Hydrology*, 332(1-2):93–109, Jan. 2007.

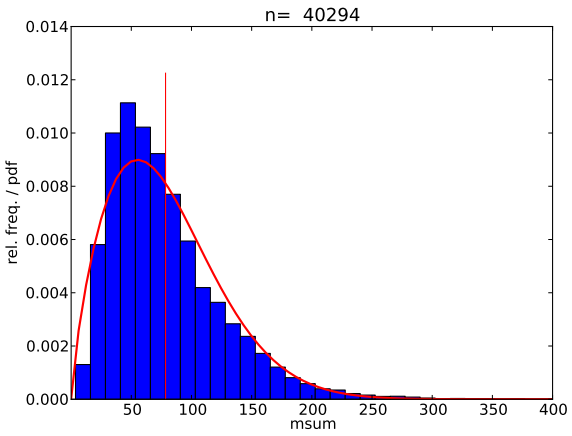
A. Generation of Point Rainfall

A.1. Marginal Distributions of All 292 Stations

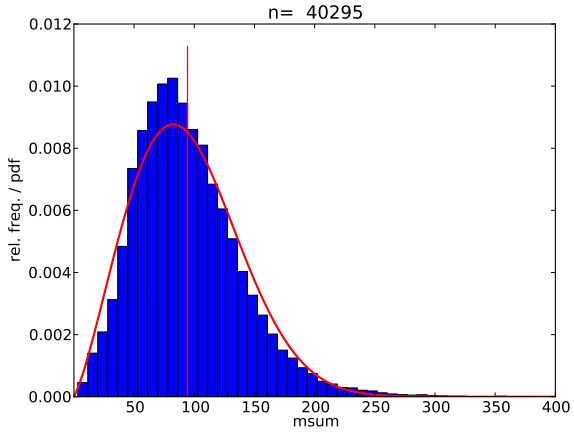
A.1.1. Monthly Precipitation Sum



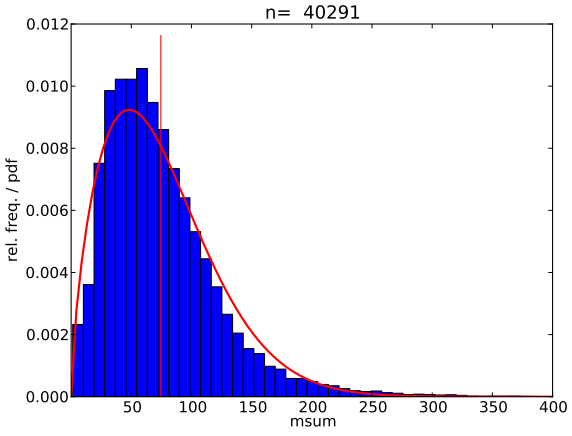
(a) December - February



(b) March - May

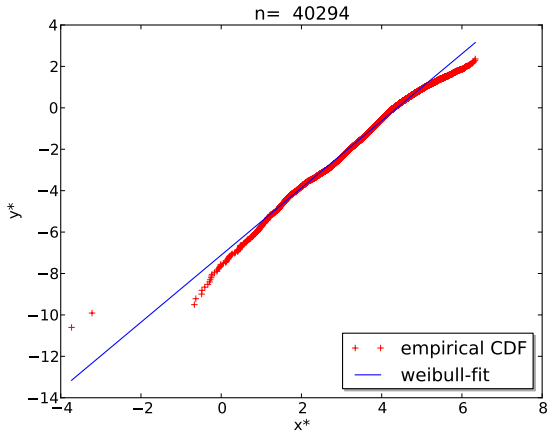


(c) June - August

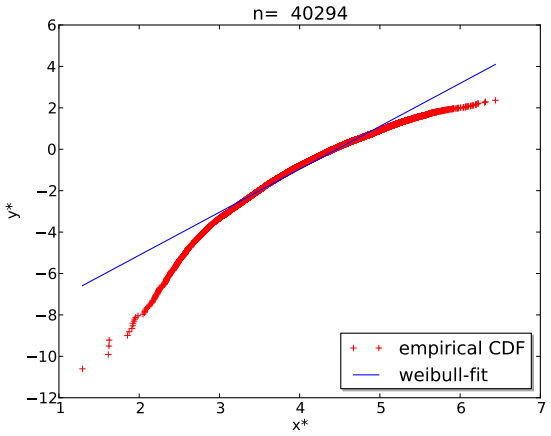


(d) September - November

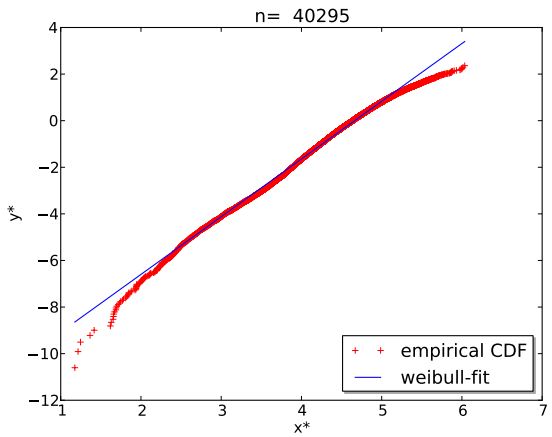
Figure A.1.: Histogram and Weibull PDF of the monthly precipitation sum at 292 stations



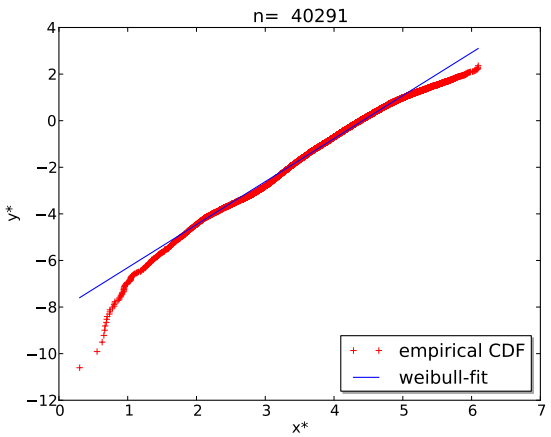
(a) December - February



(b) March - May



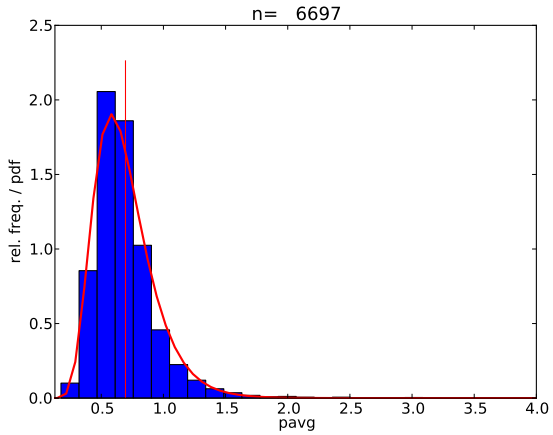
(c) June - August



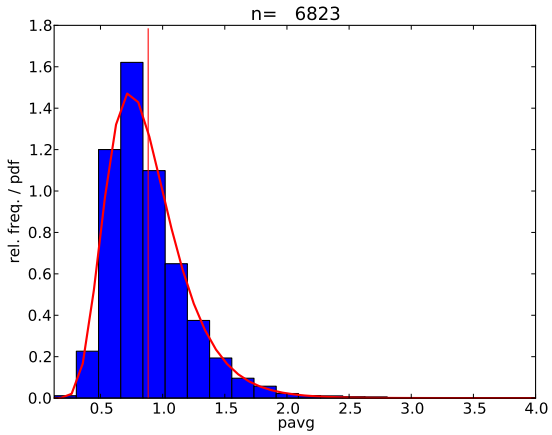
(d) September - November

Figure A.2.: Observed distribution and Weibull fit of the monthly precipitation sum at 292 stations in the linearized space

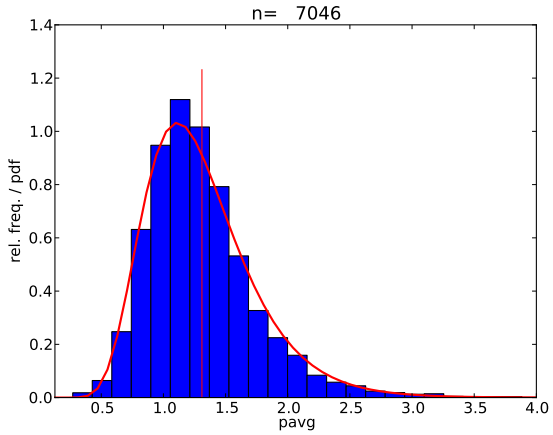
A.1.2. Average Precipitation in Wet Hours



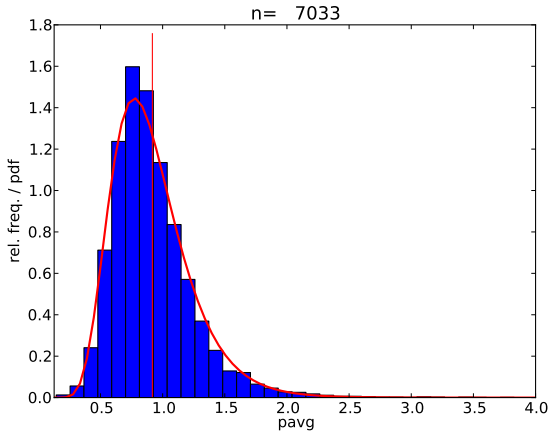
(a) December - February



(b) March - May

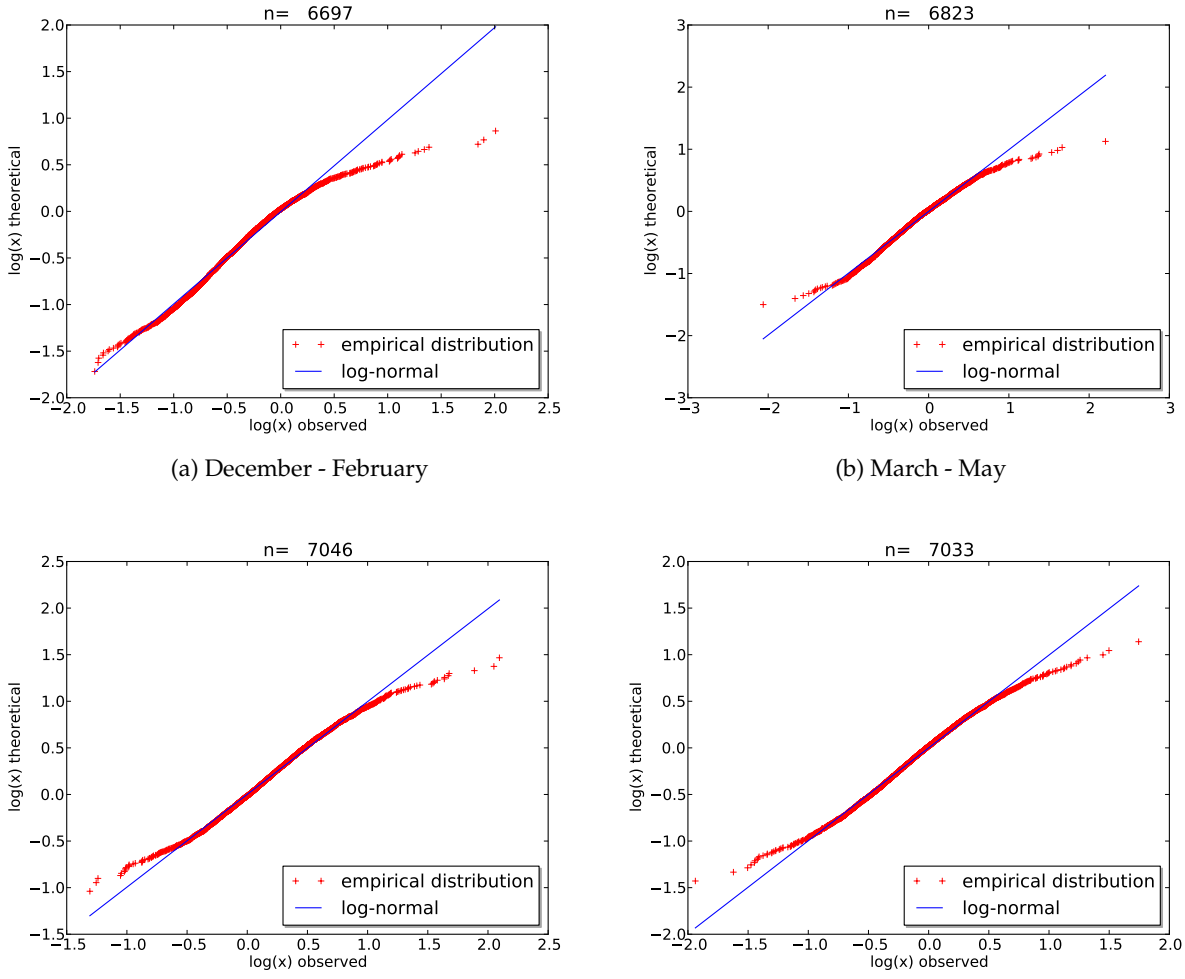


(c) June - August



(d) September - November

Figure A.3.: Histogram and Log-Normal PDF of the average precipitation in wet hours at 292 stations



(a) December - February

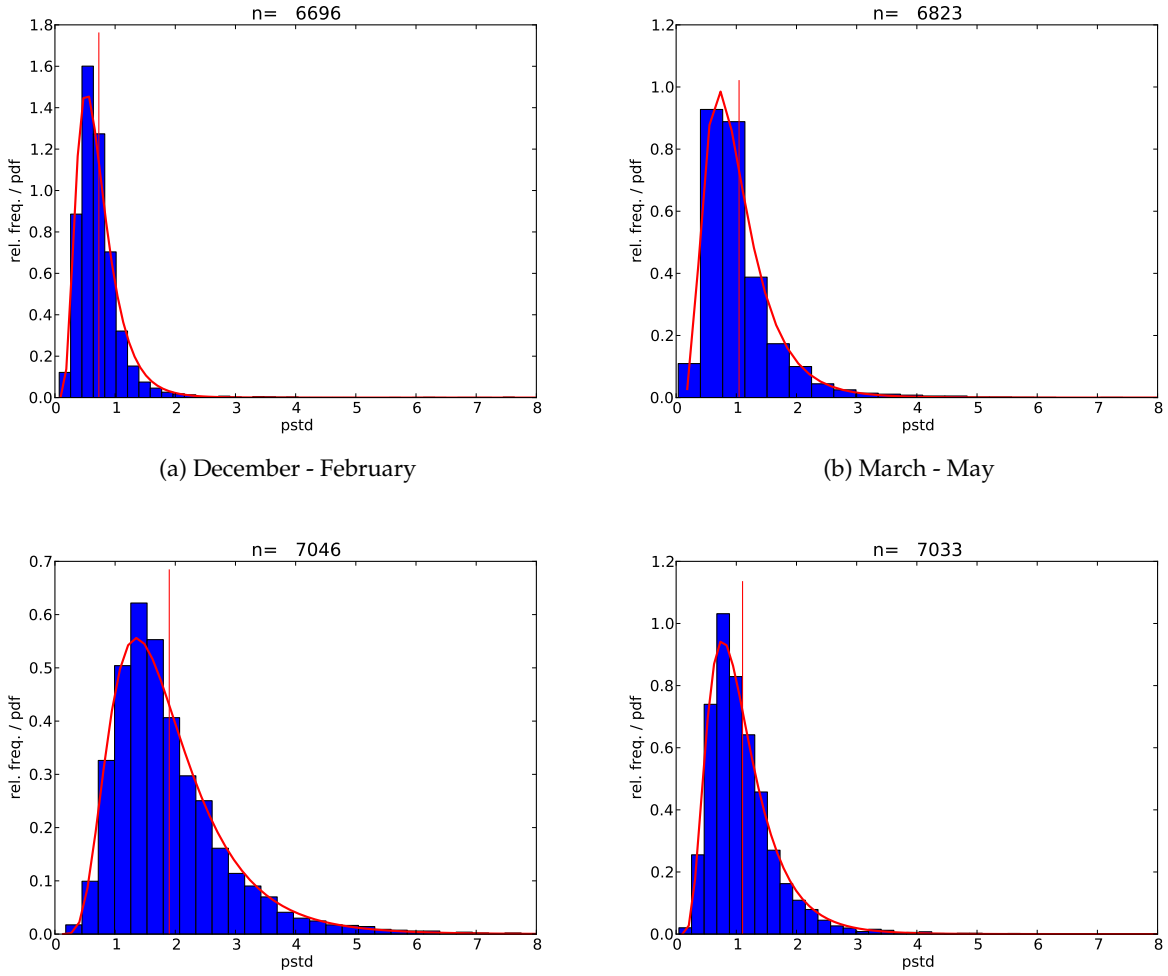
(b) March - May

(c) June - August

(d) September - November

Figure A.4.: Observed distribution and Log-Normal fit of the average precipitation in wet hours at 292 stations in the linearized space

A.1.3. Standard Deviation of Precipitation Amounts in Wet Hours



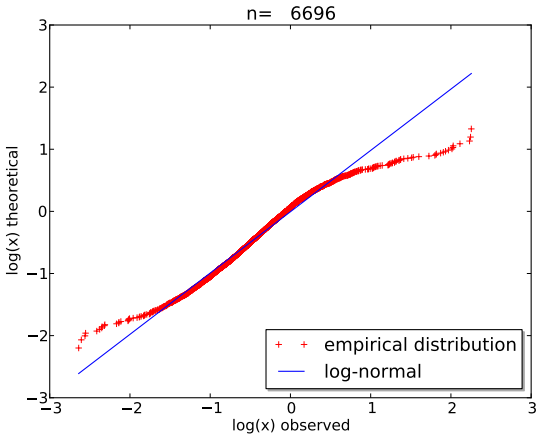
(a) December - February

(b) March - May

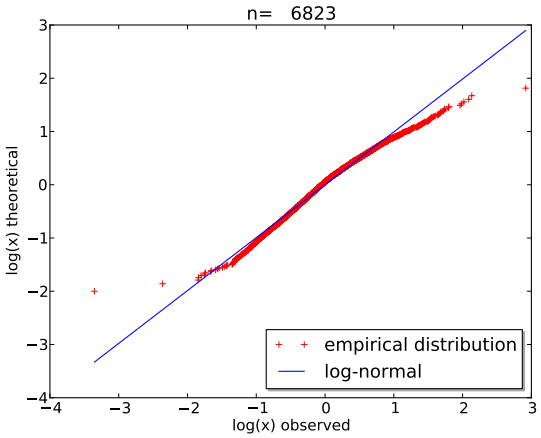
(c) June - August

(d) September - November

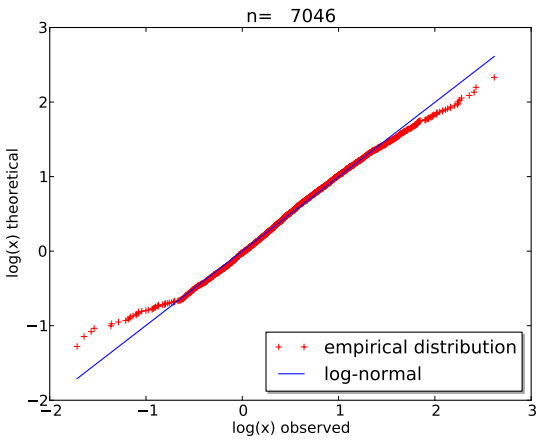
Figure A.5.: Histogram and Log-Normal PDF of the standard deviation of the precipitation amounts in wet hours at 292 stations



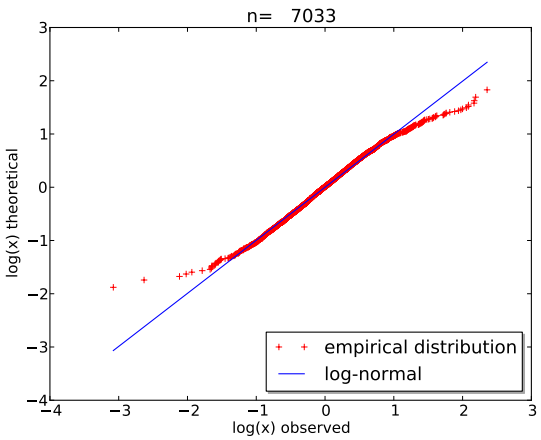
(a) December - February



(b) March - May



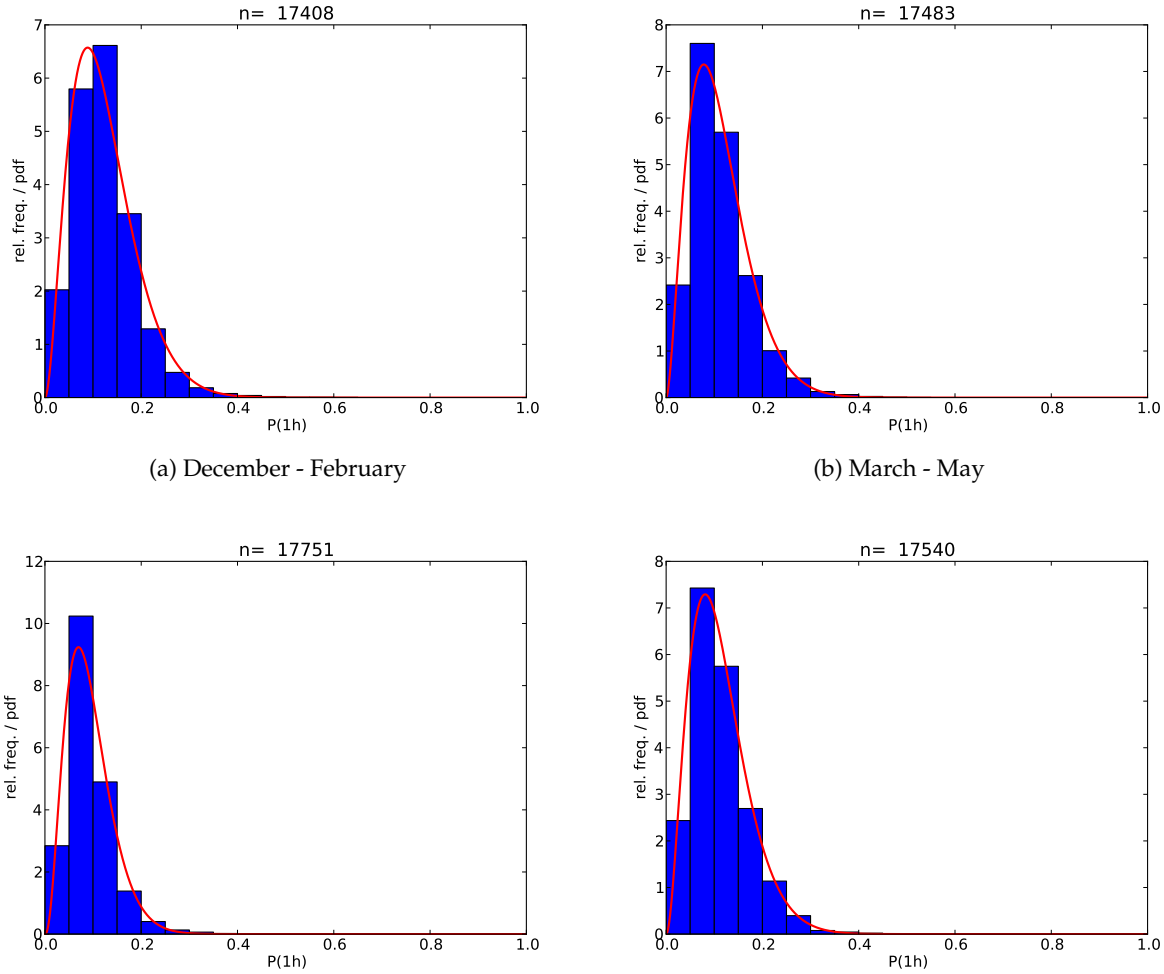
(c) June - August



(d) September - November

Figure A.6.: Observed distribution and Log-Normal fit of the standard deviation of precipitation amounts in wet hours at 292 stations in the linearized space

A.1.4. Hourly Rainfall Probability



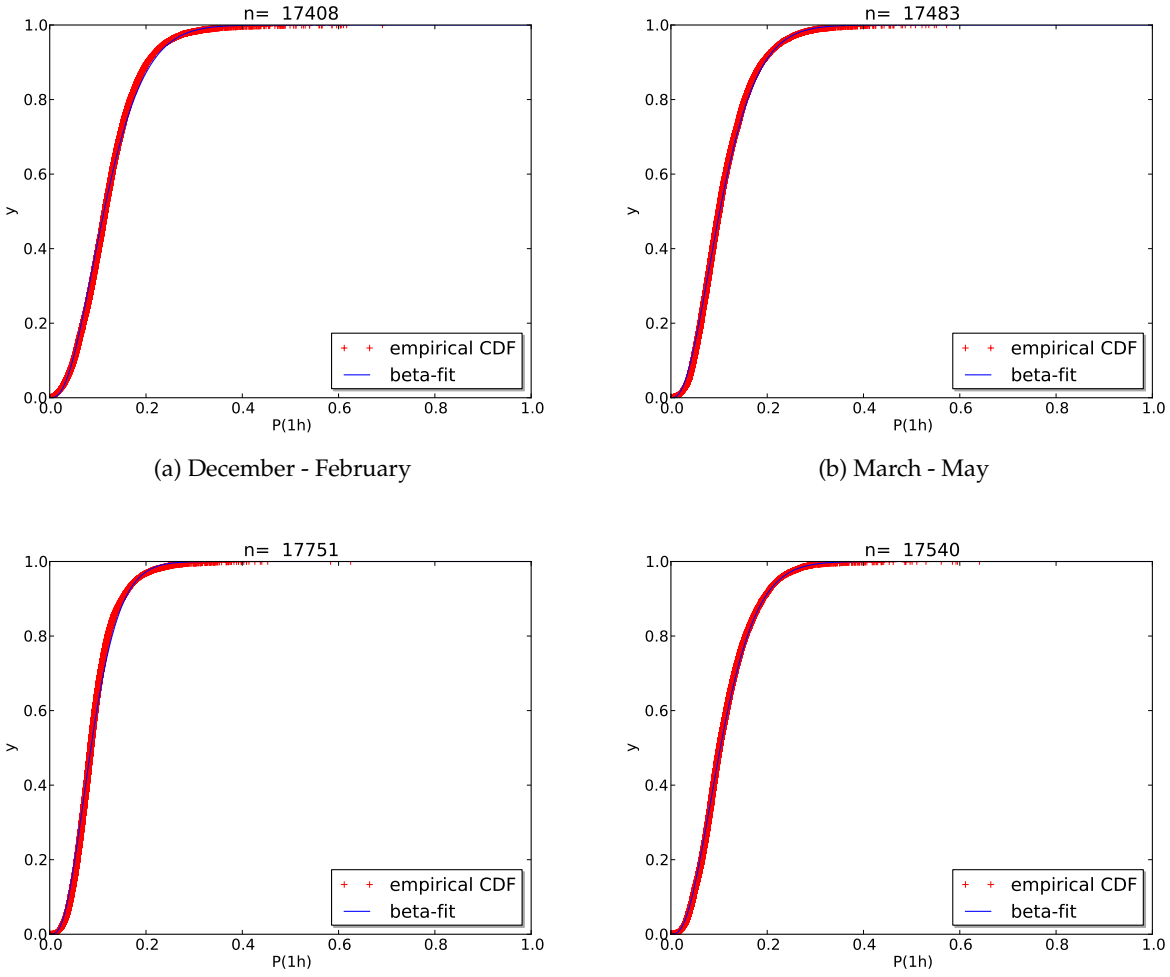
(a) December - February

(b) March - May

(c) June - August

(d) September - November

Figure A.7.: Histogram and Beta PDF of the monthly 1 h rainfall probability at 292 stations



(a) December - February

(b) March - May

(c) June - August

(d) September - November

Figure A.8.: Observed distribution and Beta fit of the monthly 1 h rainfall probability at 292 stations

A.2. Drawing Conditioned Values from a Multivariate Normal Distribution

A multivariate Gaussian random vector X can be generated from a vector of multivariate independent standard normal random numbers if the covariance matrix and the mean vector are known:

$$X \sim \mathcal{N}(\mu + \mathbf{\Sigma}) \tag{A.1}$$

with
- X multivariate Gaussian random variable
- $\mathcal{N}$ Normal distribution function
- μ vector of averages in X
- $\mathbf{\Sigma}$ matrix of covariances in X

Dividing X in two vectors $X = [X_1, X_2]^T$, $\mathbf{\Sigma}$ and μ can be written as:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \text{ with sizes } \begin{bmatrix} q \times 1 \\ (N-q) \times 1 \end{bmatrix} \tag{A.2}$$

$$\mathbf{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \text{with sizes } \begin{bmatrix} q \times q & q \times (N-q) \\ (N-q) \times q & q \times q \end{bmatrix} \tag{A.3}$$

The distribution of X conditioned on a value of $X_2 = a$ is given by:

$$(X_1 | X_2 = a) = \mu_{1|a} + \mathbf{\Sigma}_{1|a} Z_1 \tag{A.4}$$

with
- Z_1 multivariate standard normal random variable of size $q \times 1$
- $\mu_{1|a}$ average vector of X_1 conditioned on $X_2 = a$ of size $q \times 1$
- $\mathbf{\Sigma}_{1|a}$ covariance matrix of X_1 conditioned on $X_2 = a$ of size $q \times q$

Applying the convention of Eq. (A.3) the conditional mean and covariance can be written as:

$$\mu_{1|a} = \mu_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(a - \mu_2) \tag{A.5}$$

$$\mathbf{\Sigma}_{1|a} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} \tag{A.6}$$

The conditioned covariance $\mathbf{\Sigma}_{1|a}$ is independent of the conditioning value $\mathbf{a}$. The shape and spread of the distribution is independent of a, only its location is changing according to Eq. (A.5).

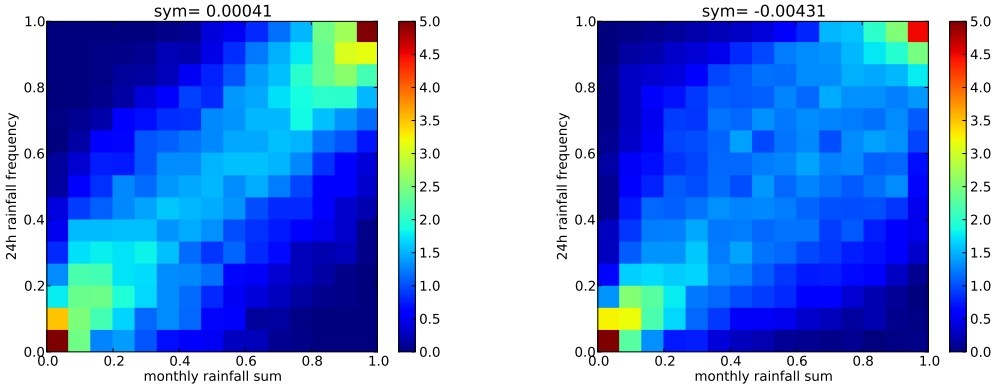
Correlated random numbers from the conditioned normal distribution is are drawn by means of a vector Z of uncorrelated standard normal random variables and the lower triangle matrix $\mathbf{L}$ of the Cholesky decomposition of $\mathbf{\Sigma}_a$

$$x^* = \mu_{1|a} + \mathbf{L} \cdot z \tag{A.7}$$

$$\mathbf{\Sigma_{1|a}} = \mathbf{L}\mathbf{L}^T \tag{A.8}$$

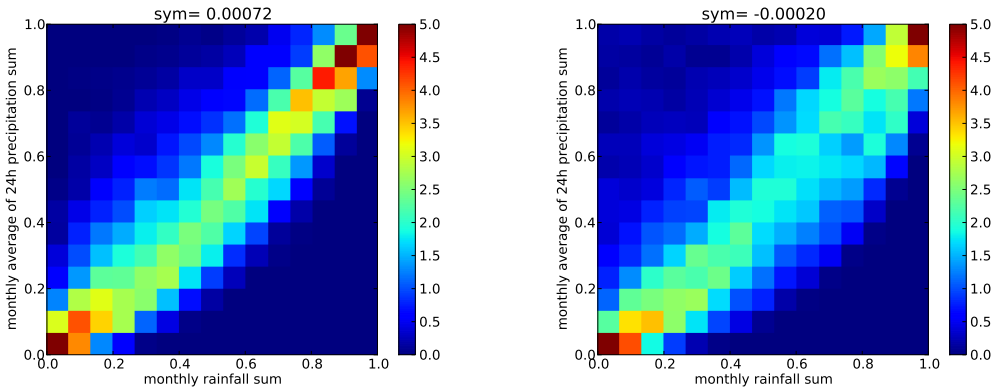
B. Spatial Generation

B.1. Observed Multivariate Dependencies in Daily Precipitation Data



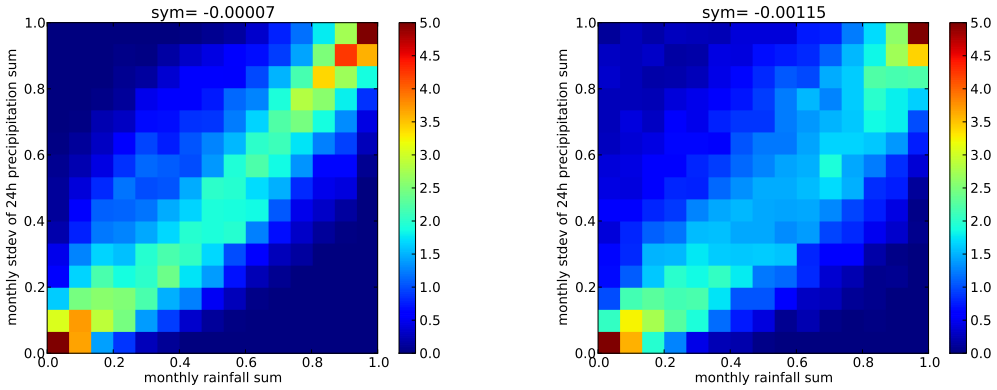
(a) April (b) July

Figure B.1.: Empirical bivariate copula between monthly precipitation sum and rainfall frequency in daily resolution calculated from month-wise censored data



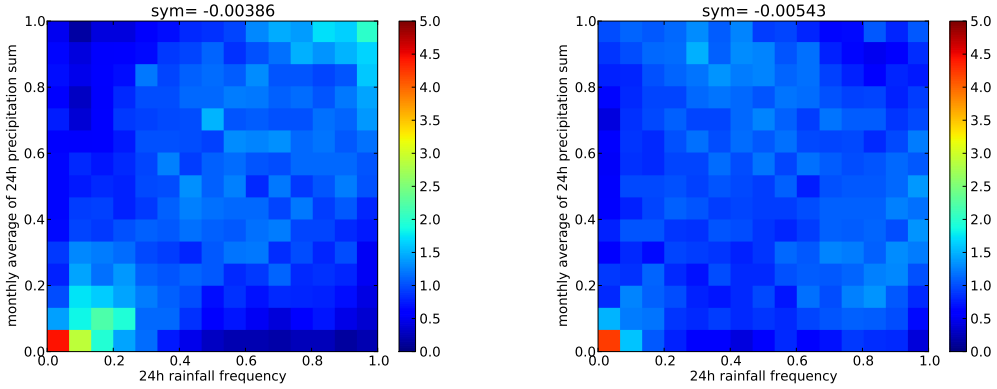
(a) March (b) June

Figure B.2.: Empirical bivariate copula between monthly precipitation sum and average wet day precipitation sum calculated from month-wise censored data



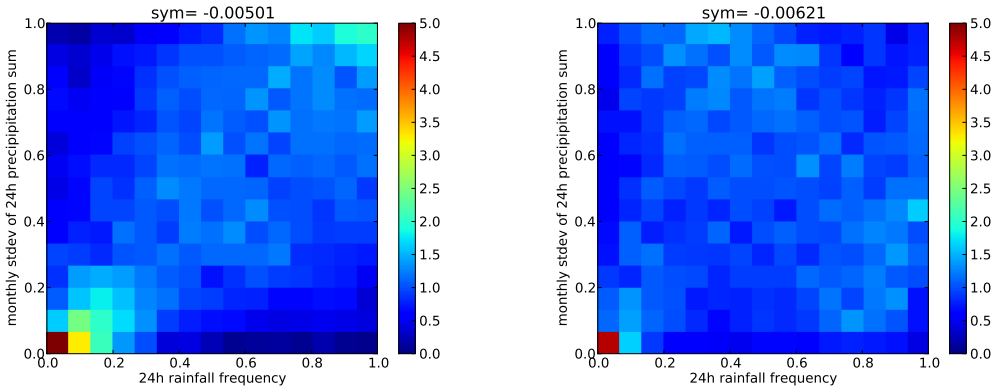
(a) March (b) June

Figure B.3.: Empirical bivariate copula between monthly precipitation sum and standard deviation of wet day precipitation sum calculated from month-wise censored data



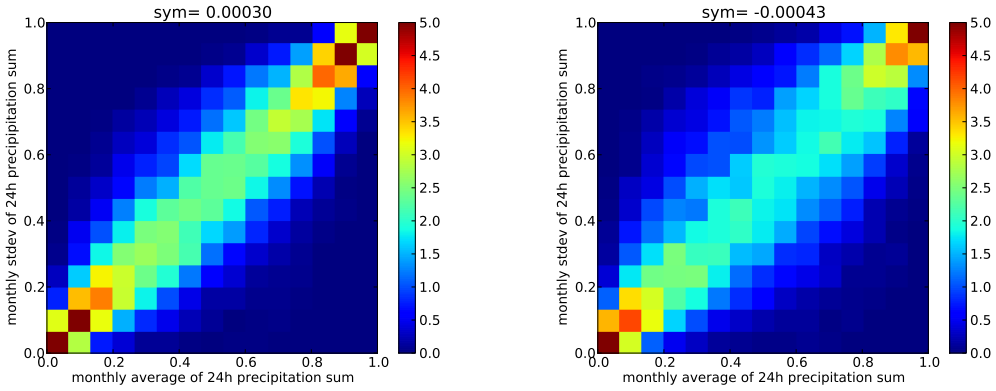
(a) January (b) July

Figure B.4.: Empirical bivariate copula between monthly rainfall frequency in daily resolution average and wet day precipitation sum calculated from month-wise censored data



(a) January (b) July

Figure B.5.: Empirical bivariate copula between monthly rainfall frequency in daily resolution and standard deviation of wet day precipitation sum calculated from month-wise censored data

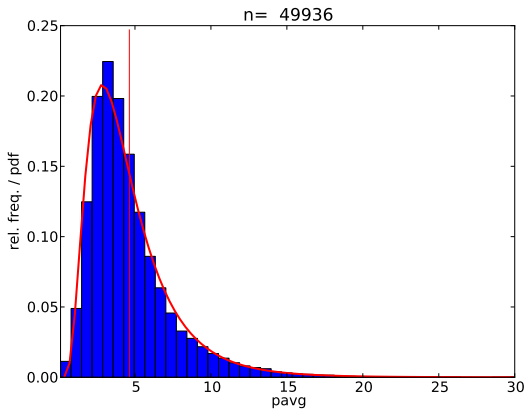


(a) January (b) August

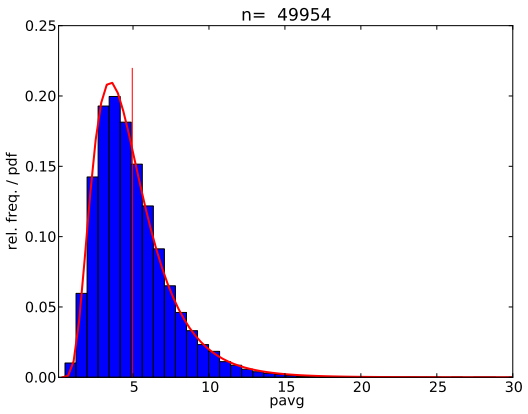
Figure B.6.: Empirical bivariate copula between average and standard deviation of wet day precipitation sum calculated from month-wise censored data

B.2. Marginal Distribution in Daily Resolution

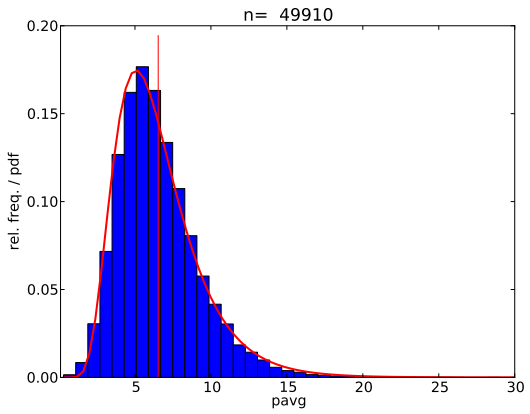
B.2.1. Average Precipitation in Wet Hours



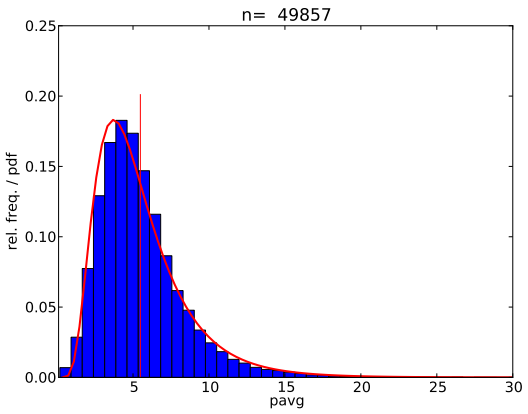
(a) December - February



(b) March - May

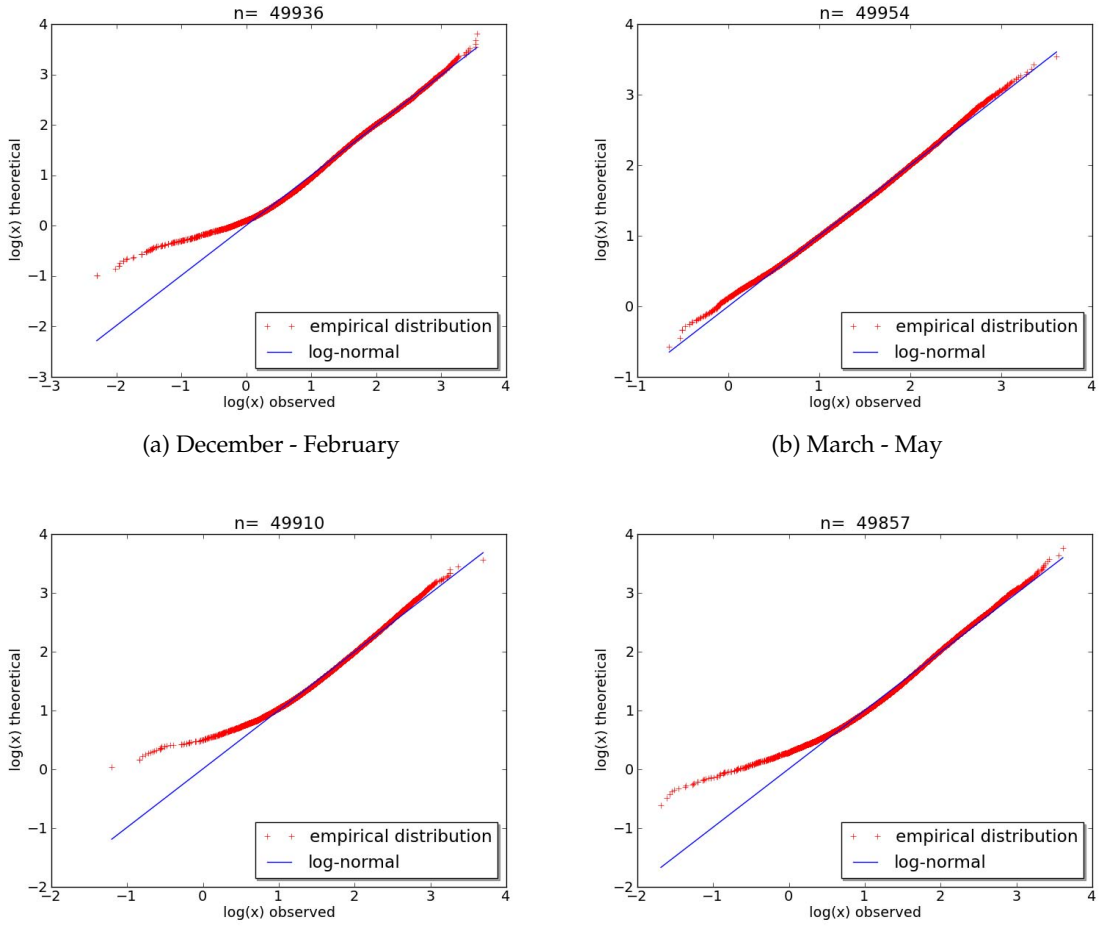


(c) June - August



(d) September - November

Figure B.7.: Histogram and Log-Normal PDF of the average wet day precipitation at 575 stations



(a) December - February

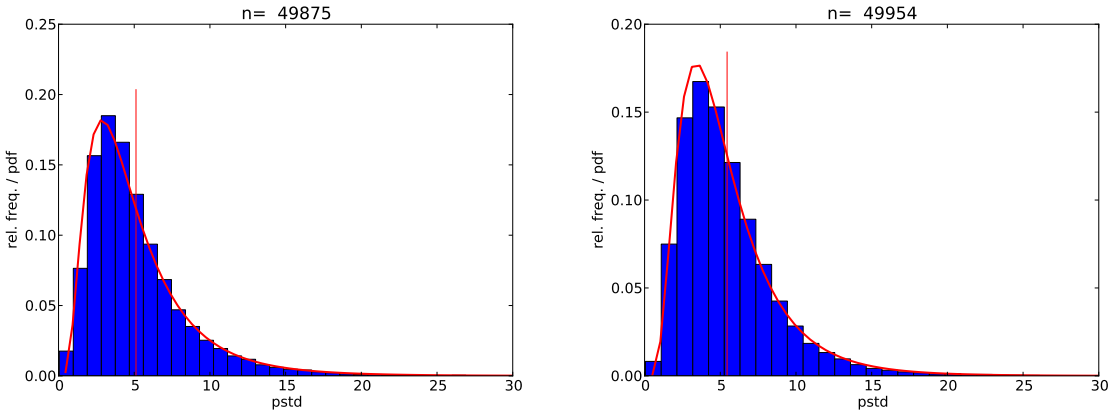
(b) March - May

(c) June - August

(d) September - November

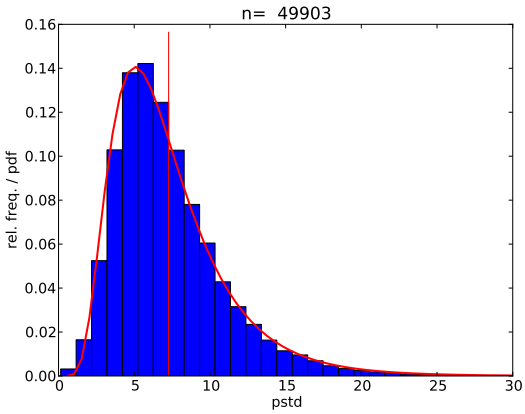
Figure B.8.: Observed distribution and Log-Normal fit of the average wet day precipitation 575 stations in the linearized space

B.2.2. Standard Deviation of Precipitation Amounts in Wet Hours

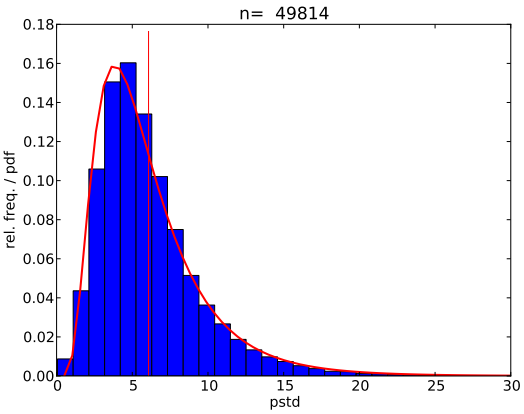


(a) December - February

(b) March - May

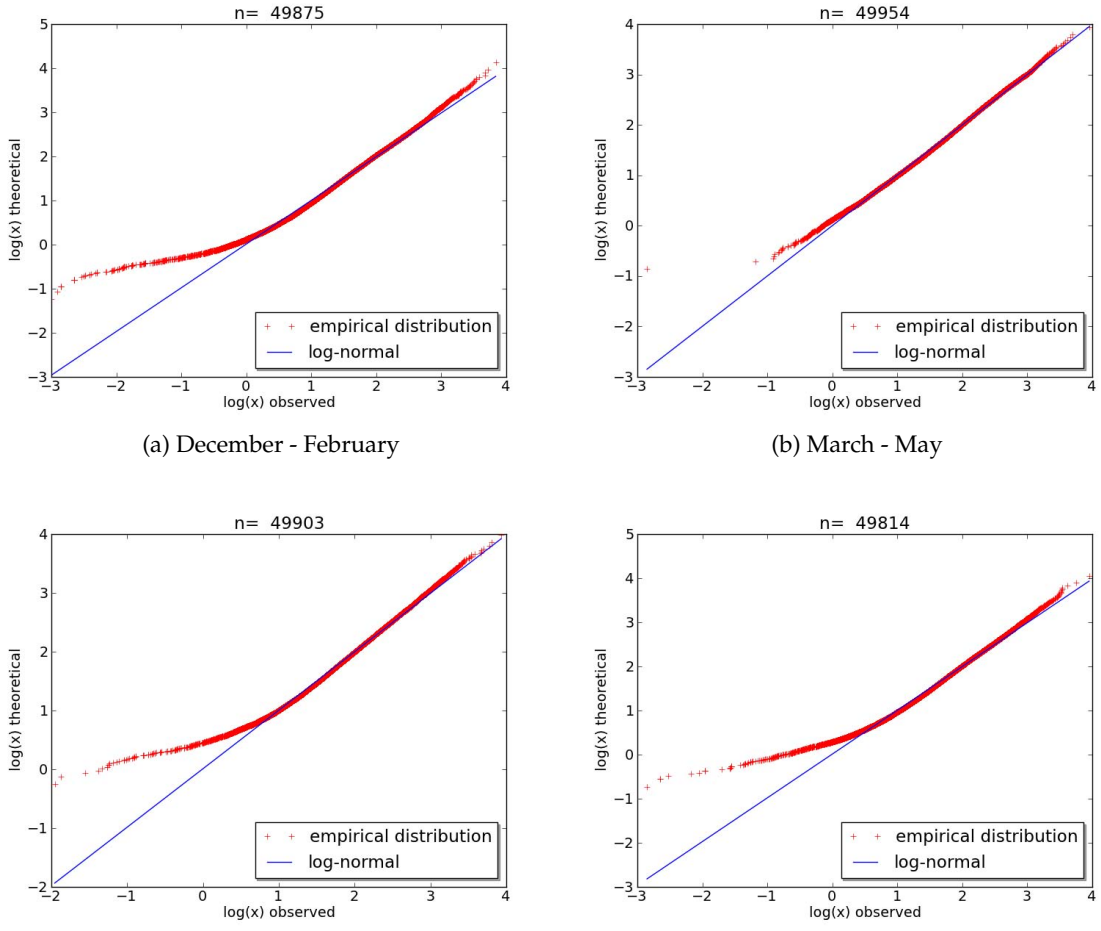


(c) June - August



(d) September - November

Figure B.9.: Histogram and Log-Normal PDF of the standard deviation of wet day precipitation amounts at 575 stations



(a) December - February

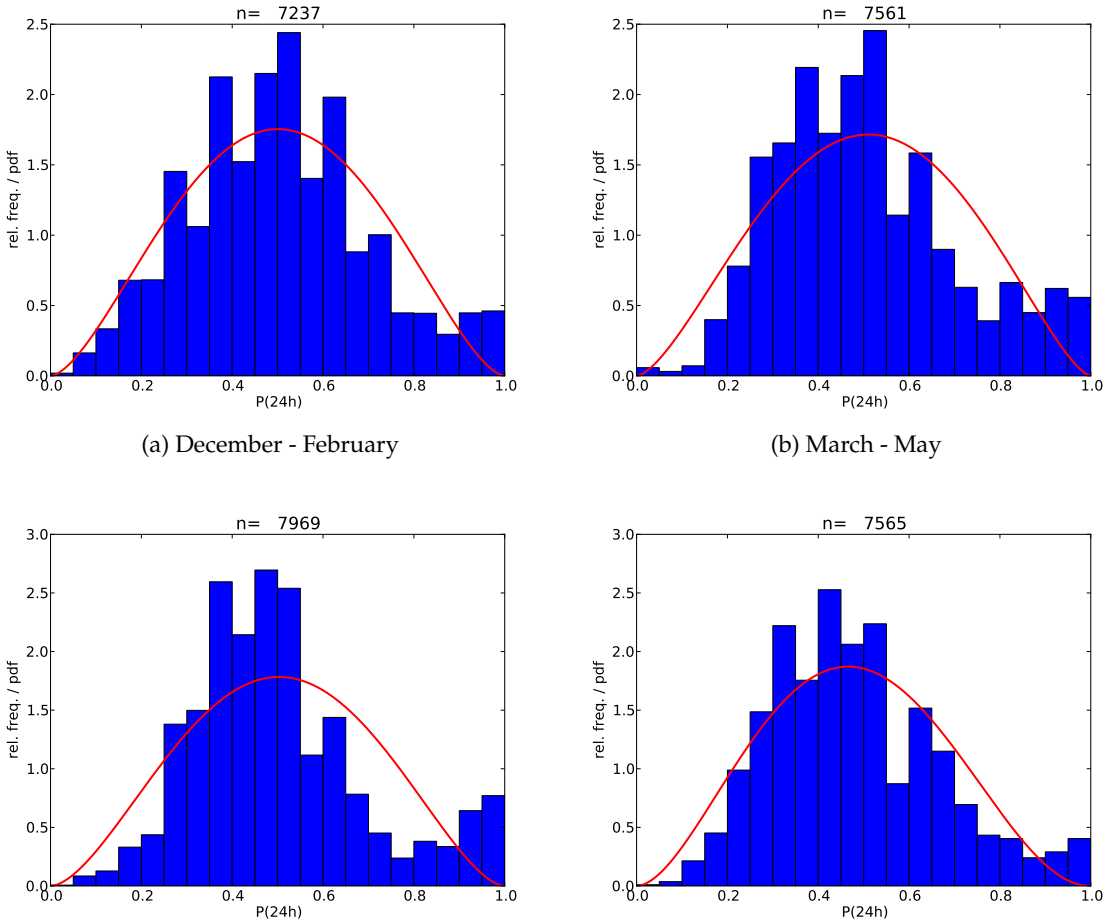
(b) March - May

(c) June - August

(d) September - November

Figure B.10.: Observed distribution and Log-Normal fit of the standard deviation of wet day precipitation amounts at 575 stations in the linearized space

B.2.3. Daily Rainfall Probability



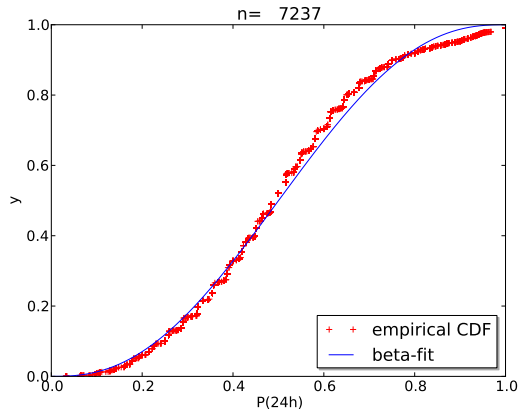
(a) December - February

(b) March - May

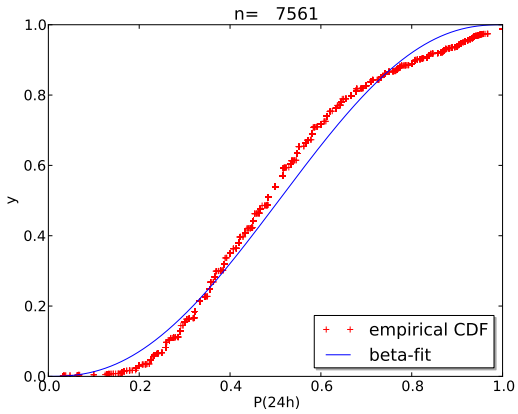
(c) June - August

(d) September - November

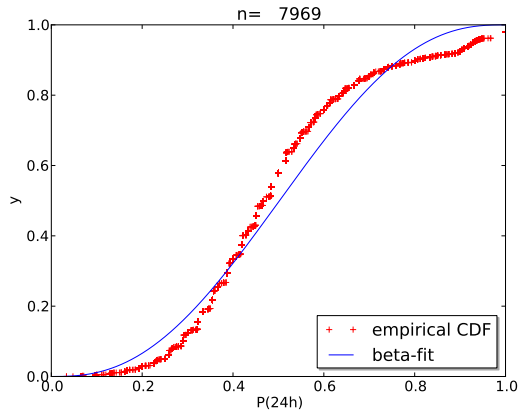
Figure B.11.: Histogram and Beta PDF of the monthly 24 h rainfall frequency at 575 stations



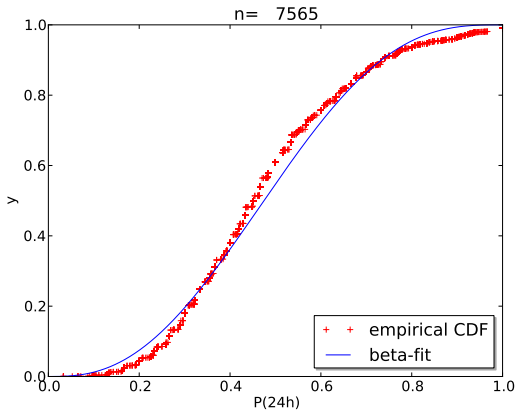
(a) December - February



(b) March - May



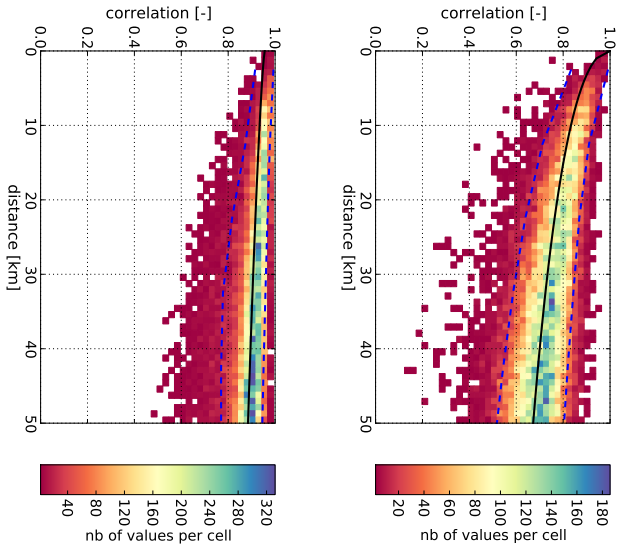
(c) June - August



(d) September - November

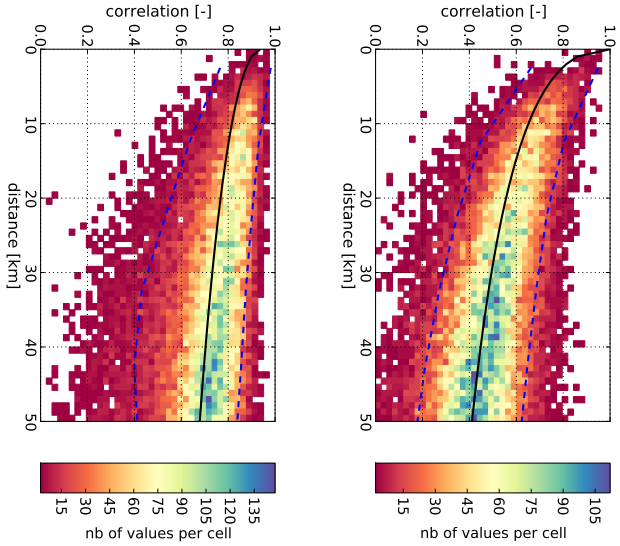
Figure B.12.: Observed distribution and Beta fit of the monthly 24 h rainfall frequency at 575 stations

B.3. Representation of Spatial Dependencies



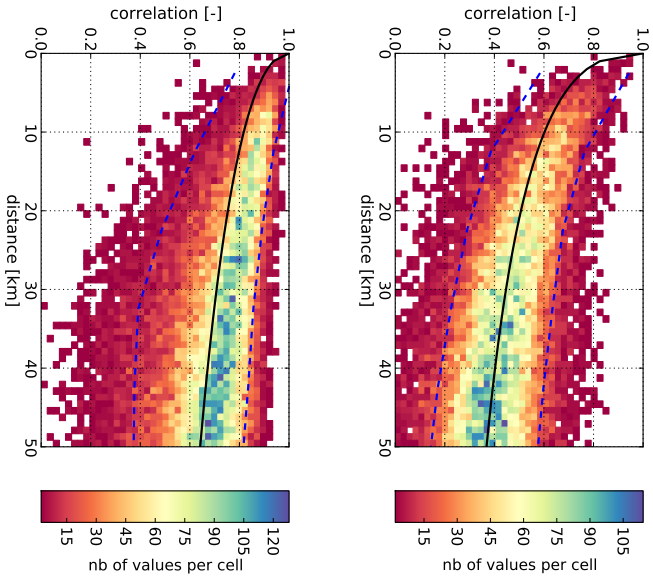
(a) December to February (b) June to August

Figure B.13.: Observed rank correlation between the monthly precipitation sum at two stations depending on the separation distance



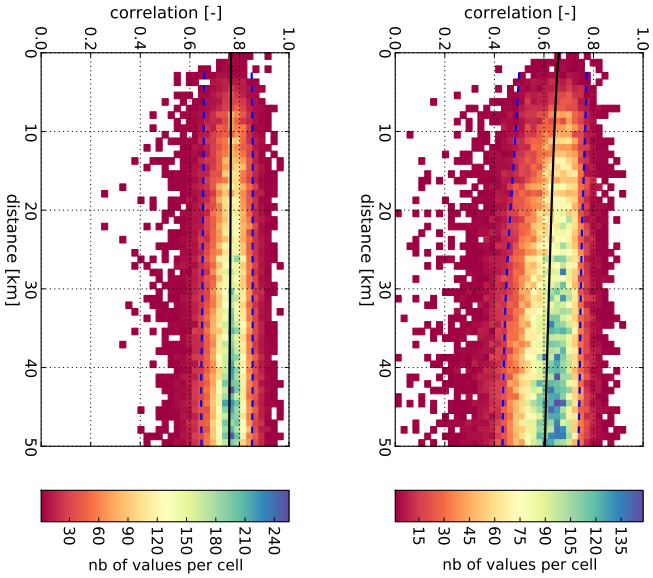
(a) December to February (b) June to August

Figure B.14.: Observed rank correlation between the average wet day precipitation at two stations depending on the separation distance



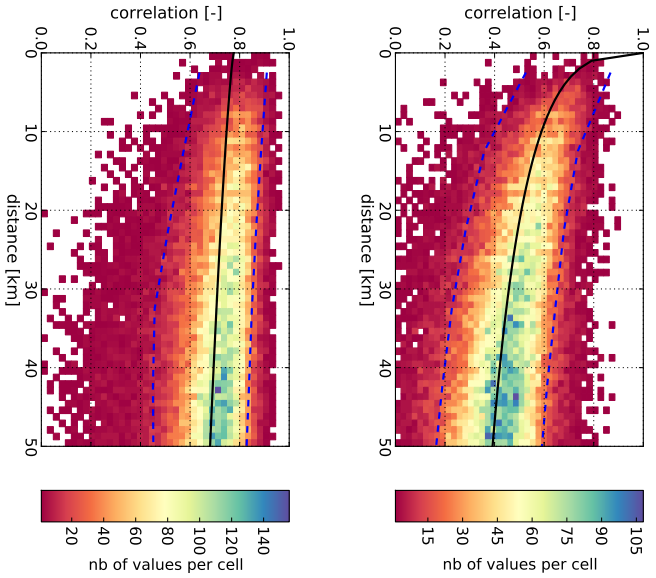
(a) December to February (b) June to August

Figure B.15.: Observed rank correlation between the standard deviation of wet day precipitation at two stations depending on the separation distance



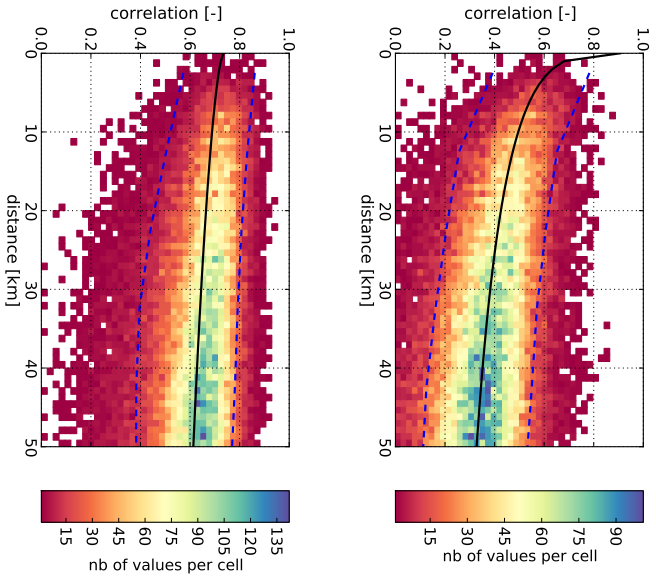
(a) December to February (b) June to August

Figure B.16.: Observed rank correlation between the monthly precipitation sum at one station and the rain day frequency at another station depending on the separation distance



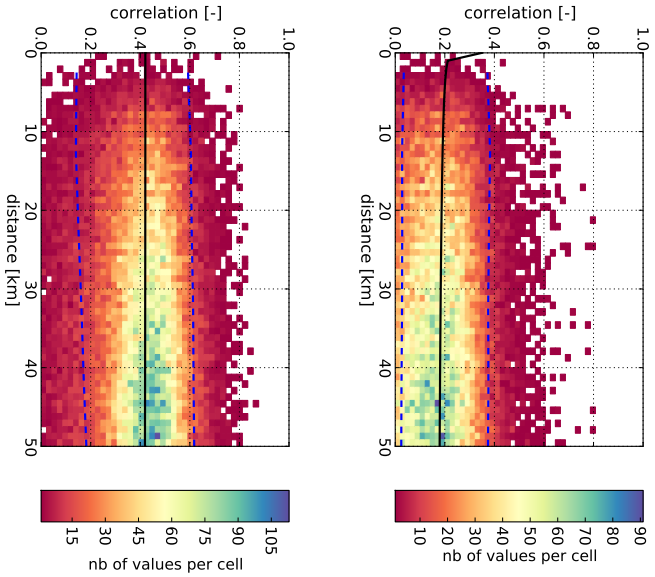
(a) December to February (b) June to August

Figure B.17.: Observed rank correlation between the monthly precipitation sum at one station and the average wet day precipitation at another station depending on the separation distance



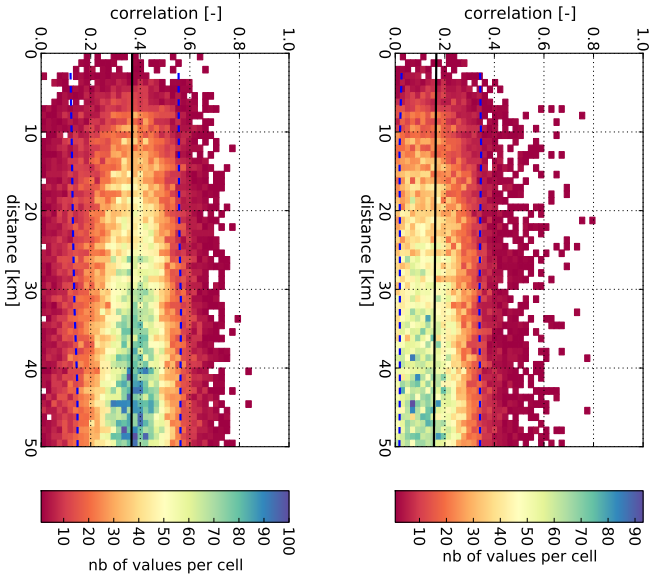
(a) December to February (b) June to August

Figure B.18.: Observed rank correlation between the monthly precipitation sum at one station and the standard deviation of wet day precipitation at another station depending on the separation distance



(a) December to February (b) June to August

Figure B.19.: Observed rank correlation between the rain day frequency at one station and the average wet day precipitation at another station depending on the separation distance



(a) December to February (b) June to August

Figure B.20.: Observed rank correlation between the rain day frequency at one station and the average wet day precipitation at another station depending on the separation distance

Curriculum Vitae

born on May 10, 1979 in Backnang, Germany

Experiences

2006 - 2012	**University of Stuttgart,** **Institute for Modelling Hydraulic and Environmental Systems** Research assistant at the department of Hydrology and Geohydrology Lecturer for statistics in the degree programs of Civil Engineering and Environmental Engineering Assisted the setup of a large scale experiment at the Research Facility for Subsurface Remediation – VEGAS
2004 - 2005	**Institut polytechnique de Grenoble,** **Laboratoire des Transfert en Hydrologie et Environnement** Internship in the domain of probabilistic weather forecast

Research Topics

2010-2012	Generation of simultaneous synthetic rainfall time series of high temporal resolution
2010-2011	Downscaling of Precipitation Data from Global Circulation Models for Equatorial Africa
2006-2009	Generation of synthetic rainfall time series under changing climatic conditions

Education

1998 - 2005	**University of Stuttgart** Master degree of Water Resources Engineering and Management (WAREM) Diplom engineer of Umweltschutztechnik (Environmental Engineering)
2001 - 2002	**Ecole Nationale des Travaux Public de l'Etat** Exchange studies in Vaulx-en-Velin, France

Institut für Wasser- und Umweltsystemmodellierung
Universität Stuttgart

Pfaffenwaldring 61
70569 Stuttgart (Vaihingen)
Telefon (0711) 685 - 64717/64749/64752/64679
Telefax (0711) 685 - 67020 o. 64746 o. 64681
E-Mail: iws@iws.uni-stuttgart.de
http://www.iws.uni-stuttgart.de

Verzeichnis der Mitteilungshefte

1 Röhnisch, Arthur: *Die Bemühungen um eine Wasserbauliche Versuchsanstalt an der Technischen Hochschule Stuttgart*, und
Fattah Abouleid, Abdel: *Beitrag zur Berechnung einer in lockeren Sand gerammten, zweifach verankerten Spundwand*, 1963

2 Marotz, Günter: *Beitrag zur Frage der Standfestigkeit von dichten Asphaltbelägen im Großwasserbau*, 1964

3 Gurr, Siegfried: *Beitrag zur Berechnung zusammengesetzter ebener Flächentragwerke unter besonderer Berücksichtigung ebener Stauwände, mit Hilfe von Randwert- und Lastwertmatrizen*, 1965

4 Plica, Peter: *Ein Beitrag zur Anwendung von Schalenkonstruktionen im Stahlwasserbau*, und Petrikat, Kurt: *Möglichkeiten und Grenzen des wasserbaulichen Versuchswesens*, 1966

5 Plate, Erich: *Beitrag zur Bestimmung der Windgeschwindigkeitsverteilung in der durch eine Wand gestörten bodennahen Luftschicht*,
und
Röhnisch, Arthur; Marotz, Günter: *Neue Baustoffe und Bauausführungen für den Schutz der Böschungen und der Sohle von Kanälen, Flüssen und Häfen; Gestehungskosten und jeweilige Vorteile*, sowie Unny, T.E.: *Schwingungsuntersuchungen am Kegelstrahlschieber*, 1967

6 Seiler, Erich: *Die Ermittlung des Anlagenwertes der bundeseigenen Binnenschiffahrtsstraßen und Talsperren und des Anteils der Binnenschiffahrt an diesem Wert*, 1967

7 *Sonderheft anläßlich des 65. Geburtstages von Prof. Arthur Röhnisch mit Beiträgen von* Benk, Dieter; Breitling, J.; Gurr, Siegfried; Haberhauer, Robert; Honekamp, Hermann; Kuz, Klaus Dieter; Marotz, Günter; Mayer-Vorfelder, Hans-Jörg; Miller, Rudolf; Plate, Erich J.; Radomski, Helge; Schwarz, Helmut; Vollmer, Ernst; Wildenhahn, Eberhard; 1967

8 Jumikis, Alfred: *Beitrag zur experimentellen Untersuchung des Wassernachschubs in einem gefrierenden Boden und die Beurteilung der Ergebnisse*, 1968

9 Marotz, Günter: *Technische Grundlagen einer Wasserspeicherung im natürlichen Untergrund*, 1968

10 Radomski, Helge: *Untersuchungen über den Einfluß der Querschnittsform wellenförmiger Spundwände auf die statischen und rammtechnischen Eigenschaften*, 1968

11 Schwarz, Helmut: *Die Grenztragfähigkeit des Baugrundes bei Einwirkung vertikal gezogener Ankerplatten als zweidimensionales Bruchproblem*, 1969

12 Erbel, Klaus: *Ein Beitrag zur Untersuchung der Metamorphose von Mittelgebirgsschneedecken unter besonderer Berücksichtigung eines Verfahrens zur Bestimmung der thermischen Schneequalität*, 1969

13 Westhaus, Karl-Heinz: *Der Strukturwandel in der Binnenschiffahrt und sein Einfluß auf den Ausbau der Binnenschiffskanäle*, 1969

14 Mayer-Vorfelder, Hans-Jörg: *Ein Beitrag zur Berechnung des Erdwiderstandes unter Ansatz der logarithmischen Spirale als Gleitflächenfunktion*, 1970

15 Schulz, Manfred: *Berechnung des räumlichen Erddruckes auf die Wandung kreiszylindrischer Körper*, 1970

16 Mobasseri, Manoutschehr: *Die Rippenstützmauer. Konstruktion und Grenzen ihrer Standsicherheit*, 1970

17 Benk, Dieter: *Ein Beitrag zum Betrieb und zur Bemessung von Hochwasserrückhaltebecken*, 1970

18 Gàl, Attila: *Bestimmung der mitschwingenden Wassermasse bei überströmten Fischbauchklappen mit kreiszylindrischem Staublech*, 1971, vergriffen

19 Kuz, Klaus Dieter: *Ein Beitrag zur Frage des Einsetzens von Kavitationserscheinungen in einer Düsenströmung bei Berücksichtigung der im Wasser gelösten Gase*, 1971, vergriffen

20 Schaak, Hartmut: *Verteilleitungen von Wasserkraftanlagen*, 1971

21 *Sonderheft zur Eröffnung der neuen Versuchsanstalt des Instituts für Wasserbau der Universität Stuttgart mit Beiträgen von* Brombach, Hansjörg; Dirksen, Wolfram; Gàl, Attila; Gerlach, Reinhard; Giesecke, Jürgen; Holthoff, Franz-Josef; Kuz, Klaus Dieter; Marotz, Günter; Minor, Hans-Erwin; Petrikat, Kurt; Röhnisch, Arthur; Rueff, Helge; Schwarz, Helmut; Vollmer, Ernst; Wildenhahn, Eberhard; 1972

22 Wang, Chung-su: *Ein Beitrag zur Berechnung der Schwingungen an Kegelstrahlschiebern*, 1972

23 Mayer-Vorfelder, Hans-Jörg: *Erdwiderstandsbeiwerte nach dem Ohde-Variationsverfahren*, 1972

24 Minor, Hans-Erwin: *Beitrag zur Bestimmung der Schwingungsanfachungsfunktionen überströmter Stauklappen*, 1972, vergriffen

25 Brombach, Hansjörg: *Untersuchung strömungsmechanischer Elemente (Fluidik) und die Möglichkeit der Anwendung von Wirbelkammerelementen im Wasserbau*, 1972, vergriffen

26 Wildenhahn, Eberhard: *Beitrag zur Berechnung von Horizontalfilterbrunnen*, 1972

27 Steinlein, Helmut: *Die Eliminierung der Schwebstoffe aus Flußwasser zum Zweck der unterirdischen Wasserspeicherung, gezeigt am Beispiel der Iller*, 1972

28 Holthoff, Franz Josef: *Die Überwindung großer Hubhöhen in der Binnenschiffahrt durch Schwimmerhebewerke*, 1973

29 Röder, Karl: *Einwirkungen aus Baugrundbewegungen auf trog- und kastenförmige Konstruktionen des Wasser- und Tunnelbaues*, 1973

30 Kretschmer, Heinz: *Die Bemessung von Bogenstaumauern in Abhängigkeit von der Talform*, 1973

31 Honekamp, Hermann: *Beitrag zur Berechnung der Montage von Unterwasserpipelines*, 1973

32 Giesecke, Jürgen: *Die Wirbelkammertriode als neuartiges Steuerorgan im Wasserbau*, und Brombach, Hansjörg: *Entwicklung, Bauformen, Wirkungsweise und Steuereigenschaften von Wirbelkammerverstärkern*, 1974

33 Rueff, Helge: *Untersuchung der schwingungserregenden Kräfte an zwei hintereinander angeordneten Tiefschützen unter besonderer Berücksichtigung von Kavitation*, 1974

34 Röhnisch, Arthur: *Einpreßversuche mit Zementmörtel für Spannbeton - Vergleich der Ergebnisse von Modellversuchen mit Ausführungen in Hüllwellrohren*, 1975

35 *Sonderheft anläßlich des 65. Geburtstages von Prof. Dr.-Ing. Kurt Petrikat mit Beiträgen von*: Brombach, Hansjörg; Erbel, Klaus; Flinspach, Dieter; Fischer jr., Richard; Gàl, Attila; Gerlach, Reinhard; Giesecke, Jürgen; Haberhauer, Robert; Hafner Edzard; Hausenblas, Bernhard; Horlacher, Hans-Burkhard; Hutarew, Andreas; Knoll, Manfred; Krummet, Ralph; Marotz, Günter; Merkle, Theodor; Miller, Christoph; Minor, Hans-Erwin; Neumayer, Hans; Rao, Syamala; Rath, Paul; Rueff, Helge; Ruppert, Jürgen; Schwarz, Wolfgang; Topal-Gökceli, Mehmet; Vollmer, Ernst; Wang, Chung-su; Weber, Hans-Georg; 1975

36 Berger, Jochum: *Beitrag zur Berechnung des Spannungszustandes in rotationssymmetrisch belasteten Kugelschalen veränderlicher Wandstärke unter Gas- und Flüssigkeitsdruck durch Integration schwach singulärer Differentialgleichungen*, 1975

37 Dirksen, Wolfram: *Berechnung instationärer Abflußvorgänge in gestauten Gerinnen mittels Differenzenverfahren und die Anwendung auf Hochwasserrückhaltebecken*, 1976

38 Horlacher, Hans-Burkhard: *Berechnung instationärer Temperatur- und Wärmespannungsfelder in langen mehrschichtigen Hohlzylindern*, 1976

39 Hafner, Edzard: *Untersuchung der hydrodynamischen Kräfte auf Baukörper im Tiefwasserbereich des Meeres*, 1977, ISBN 3-921694-39-6

40 Ruppert, Jürgen: *Über den Axialwirbelkammerverstärker für den Einsatz im Wasserbau*, 1977, ISBN 3-921694-40-X

41 Hutarew, Andreas: *Beitrag zur Beeinflußbarkeit des Sauerstoffgehalts in Fließgewässern an Abstürzen und Wehren*, 1977, ISBN 3-921694-41-8, vergriffen

42 Miller, Christoph: *Ein Beitrag zur Bestimmung der schwingungserregenden Kräfte an unterströmten Wehren*, 1977, ISBN 3-921694-42-6

43 Schwarz, Wolfgang: *Druckstoßberechnung unter Berücksichtigung der Radial- und Längsverschiebungen der Rohrwandung*, 1978, ISBN 3-921694-43-4

44 Kinzelbach, Wolfgang: *Numerische Untersuchungen über den optimalen Einsatz variabler Kühlsysteme einer Kraftwerkskette am Beispiel Oberrhein*, 1978, ISBN 3-921694-44-2

45 Barczewski, Baldur: *Neue Meßmethoden für Wasser-Luftgemische und deren Anwendung auf zweiphasige Auftriebsstrahlen*, 1979, ISBN 3-921694-45-0

46 Neumayer, Hans: *Untersuchung der Strömungsvorgänge in radialen Wirbelkammerverstärkern*, 1979, ISBN 3-921694-46-9

47 Elalfy, Youssef-Elhassan: *Untersuchung der Strömungsvorgänge in Wirbelkammerdioden und -drosseln*, 1979, ISBN 3-921694-47-7

48 Brombach, Hansjörg: *Automatisierung der Bewirtschaftung von Wasserspeichern*, 1981, ISBN 3-921694-48-5

49 Geldner, Peter: *Deterministische und stochastische Methoden zur Bestimmung der Selbstdichtung von Gewässern*, 1981, ISBN 3-921694-49-3, vergriffen

50 Mehlhorn, Hans: *Temperaturveränderungen im Grundwasser durch Brauchwassereinleitungen*, 1982, ISBN 3-921694-50-7, vergriffen

51 Hafner, Edzard: *Rohrleitungen und Behälter im Meer*, 1983, ISBN 3-921694-51-5

52 Rinnert, Bernd: *Hydrodynamische Dispersion in porösen Medien: Einfluß von Dichteunterschieden auf die Vertikalvermischung in horizontaler Strömung*, 1983, ISBN 3-921694-52-3, vergriffen

53 Lindner, Wulf: *Steuerung von Grundwasserentnahmen unter Einhaltung ökologischer Kriterien*, 1983, ISBN 3-921694-53-1, vergriffen

54 Herr, Michael; Herzer, Jörg; Kinzelbach, Wolfgang; Kobus, Helmut; Rinnert, Bernd: *Methoden zur rechnerischen Erfassung und hydraulischen Sanierung von Grundwasserkontaminationen*, 1983, ISBN 3-921694-54-X

55 Schmitt, Paul: *Wege zur Automatisierung der Niederschlagsermittlung*, 1984, ISBN 3-921694-55-8, vergriffen

56 Müller, Peter: *Transport und selektive Sedimentation von Schwebstoffen bei gestautem Abfluß*, 1985, ISBN 3-921694-56-6

57 El-Qawasmeh, Fuad: *Möglichkeiten und Grenzen der Tropfbewässerung unter besonderer Berücksichtigung der Verstopfungsanfälligkeit der Tropfelemente*, 1985, ISBN 3-921694-57-4, vergriffen

58 Kirchenbaur, Klaus: *Mikroprozessorgesteuerte Erfassung instationärer Druckfelder am Beispiel seegangsbelasteter Baukörper*, 1985, ISBN 3-921694-58-2

59 Kobus, Helmut (Hrsg.): *Modellierung des großräumigen Wärme- und Schadstofftransports im Grundwasser*, Tätigkeitsbericht 1984/85 (DFG-Forschergruppe an den Universitäten Hohenheim, Karlsruhe und Stuttgart), 1985, ISBN 3-921694-59-0, vergriffen

60 Spitz, Karlheinz: *Dispersion in porösen Medien: Einfluß von Inhomogenitäten und Dichteunterschieden*, 1985, ISBN 3-921694-60-4, vergriffen

61 Kobus, Helmut: *An Introduction to Air-Water Flows in Hydraulics*, 1985, ISBN 3-921694-61-2

62 Kaleris, Vassilios: *Erfassung des Austausches von Oberflächen- und Grundwasser in horizontalebenen Grundwassermodellen*, 1986, ISBN 3-921694-62-0

63 Herr, Michael: *Grundlagen der hydraulischen Sanierung verunreinigter Porengrundwasserleiter*, 1987, ISBN 3-921694-63-9

64 Marx, Walter: *Berechnung von Temperatur und Spannung in Massenbeton infolge Hydratation*, 1987, ISBN 3-921694-64-7

65 Koschitzky, Hans-Peter: *Dimensionierungskonzept für Sohlbelüfter in Schußrinnen zur Vermeidung von Kavitationsschäden*, 1987, ISBN 3-921694-65-5

66 Kobus, Helmut (Hrsg.): *Modellierung des großräumigen Wärme- und Schadstofftransports im Grundwasser*, Tätigkeitsbericht 1986/87 (DFG-Forschergruppe an den Universitäten Hohenheim, Karlsruhe und Stuttgart) 1987, ISBN 3-921694-66-3

67 Söll, Thomas: *Berechnungsverfahren zur Abschätzung anthropogener Temperaturanomalien im Grundwasser*, 1988, ISBN 3-921694-67-1

68 Dittrich, Andreas; Westrich, Bernd: *Bodenseeufererosion, Bestandsaufnahme und Bewertung*, 1988, ISBN 3-921694-68-X, vergriffen

69 Huwe, Bernd; van der Ploeg, Rienk R.: *Modelle zur Simulation des Stickstoffhaushaltes von Standorten mit unterschiedlicher landwirtschaftlicher Nutzung*, 1988, ISBN 3-921694-69-8, vergriffen

70 Stephan, Karl: *Integration elliptischer Funktionen*, 1988, ISBN 3-921694-70-1

71 Kobus, Helmut; Zilliox, Lothaire (Hrsg.): *Nitratbelastung des Grundwassers, Auswirkungen der Landwirtschaft auf die Grundwasser- und Rohwasserbeschaffenheit und Maßnahmen zum Schutz des Grundwassers*. Vorträge des deutsch-französischen Kolloquiums am 6. Oktober 1988, Universitäten Stuttgart und Louis Pasteur Strasbourg (Vorträge in deutsch oder französisch, Kurzfassungen zweisprachig), 1988, ISBN 3-921694-71-X

72 Soyeaux, Renald: *Unterströmung von Stauanlagen auf klüftigem Untergrund unter Berücksichtigung laminarer und turbulenter Fließzustände*,1991, ISBN 3-921694-72-8

73 Kohane, Roberto: *Berechnungsmethoden für Hochwasserabfluß in Fließgewässern mit überströmten Vorländern*, 1991, ISBN 3-921694-73-6

74 Hassinger, Reinhard: *Beitrag zur Hydraulik und Bemessung von Blocksteinrampen in flexibler Bauweise*, 1991, ISBN 3-921694-74-4, vergriffen

75 Schäfer, Gerhard: *Einfluß von Schichtenstrukturen und lokalen Einlagerungen auf die Längsdispersion in Porengrundwasserleitern*, 1991, ISBN 3-921694-75-2

76 Giesecke, Jürgen: *Vorträge, Wasserwirtschaft in stark besiedelten Regionen; Umweltforschung mit Schwerpunkt Wasserwirtschaft*, 1991, ISBN 3-921694-76-0

77 Huwe, Bernd: *Deterministische und stochastische Ansätze zur Modellierung des Stickstoffhaushalts landwirtschaftlich genutzter Flächen auf unterschiedlichem Skalenniveau*, 1992, ISBN 3-921694-77-9, vergriffen

78 Rommel, Michael: *Verwendung von Kluftdaten zur realitätsnahen Generierung von Kluftnetzen mit anschließender laminar-turbulenter Strömungsberechnung*, 1993, ISBN 3-92 1694-78-7

79 Marschall, Paul: *Die Ermittlung lokaler Stofffrachten im Grundwasser mit Hilfe von Einbohrloch-Meßverfahren*, 1993, ISBN 3-921694-79-5, vergriffen

80 Ptak, Thomas: *Stofftransport in heterogenen Porenaquiferen: Felduntersuchungen und stochastische Modellierung*, 1993, ISBN 3-921694-80-9, vergriffen

81 Haakh, Frieder: *Transientes Strömungsverhalten in Wirbelkammern,* 1993, ISBN 3-921694-81-7

82 Kobus, Helmut; Cirpka, Olaf; Barczewski, Baldur; Koschitzky, Hans-Peter: *Versucheinrichtung zur Grundwasser und Altlastensanierung VEGAS, Konzeption und Programmrahmen*, 1993, ISBN 3-921694-82-5

83 Zang, Weidong: *Optimaler Echtzeit-Betrieb eines Speichers mit aktueller Abflußregenerierung*, 1994, ISBN 3-921694-83-3, vergriffen

84 Franke, Hans-Jörg: *Stochastische Modellierung eines flächenhaften Stoffeintrages und Transports in Grundwasser am Beispiel der Pflanzenschutzmittelproblematik*, 1995, ISBN 3-921694-84-1

85 Lang, Ulrich: *Simulation regionaler Strömungs- und Transportvorgänge in Karstaquiferen mit Hilfe des Doppelkontinuum-Ansatzes: Methodenentwicklung und Parameteridentifikation*, 1995, ISBN 3-921694-85-X, vergriffen

86 Helmig, Rainer: *Einführung in die Numerischen Methoden der Hydromechanik*, 1996, ISBN 3-921694-86-8, vergriffen

87 Cirpka, Olaf: *CONTRACT: A Numerical Tool for Contaminant Transport and Chemical Transformations - Theory and Program Documentation* -, 1996, ISBN 3-921694-87-6

88 Haberlandt, Uwe: *Stochastische Synthese und Regionalisierung des Niederschlages für Schmutzfrachtberechnungen*, 1996, ISBN 3-921694-88-4

89 Croisé, Jean: *Extraktion von flüchtigen Chemikalien aus natürlichen Lockergesteinen mittels erzwungener Luftströmung*, 1996, ISBN 3-921694-89-2, vergriffen

90 Jorde, Klaus: *Ökologisch begründete, dynamische Mindestwasserregelungen bei Ausleitungskraftwerken*, 1997, ISBN 3-921694-90-6, vergriffen

91 Helmig, Rainer: *Gekoppelte Strömungs- und Transportprozesse im Untergrund - Ein Beitrag zur Hydrosystemmodellierung*-, 1998, ISBN 3-921694-91-4, vergriffen

92 Emmert, Martin: *Numerische Modellierung nichtisothermer Gas-Wasser Systeme in porösen Medien,* 1997, ISBN 3-921694-92-2

93 Kern, Ulrich: *Transport von Schweb- und Schadstoffen in staugeregelten Fließgewässern am Beispiel des Neckars*, 1997, ISBN 3-921694-93-0, vergriffen

94 Förster, Georg: *Druckstoßdämpfung durch große Luftblasen in Hochpunkten von Rohrleitungen 1997*, ISBN 3-921694-94-9

95 Cirpka, Olaf: *Numerische Methoden zur Simulation des reaktiven Mehrkomponententransports im Grundwasser*, 1997, ISBN 3-921694-95-7, vergriffen

96 Färber, Arne: *Wärmetransport in der ungesättigten Bodenzone: Entwicklung einer thermischen In-situ-Sanierungstechnologie*, 1997, ISBN 3-921694-96-5

97 Betz, Christoph: *Wasserdampfdestillation von Schadstoffen im porösen Medium: Entwicklung einer thermischen In-situ-Sanierungstechnologie*, 1998, ISBN 3-921694-97-3

98 Xu, Yichun: *Numerical Modeling of Suspended Sediment Transport in Rivers*, 1998, ISBN 3-921694-98-1, vergriffen

99 Wüst, Wolfgang: *Geochemische Untersuchungen zur Sanierung CKW-kontaminierter Aquifere mit Fe(0)-Reaktionswänden,* 2000, ISBN 3-933761-02-2

100 Sheta, Hussam: *Simulation von Mehrphasenvorgängen in porösen Medien unter Einbeziehung von Hysterese-Effekten*, 2000, ISBN 3-933761-03-4

101 Ayros, Edwin: *Regionalisierung extremer Abflüsse auf der Grundlage statistischer Verfahren*, 2000, ISBN 3-933761-04-2, vergriffen

102 Huber, Ralf: *Compositional Multiphase Flow and Transport in Heterogeneous Porous Media*, 2000, ISBN 3-933761-05-0

103 Braun, Christopherus: *Ein Upscaling-Verfahren für Mehrphasenströmungen in porösen Medien*, 2000, ISBN 3-933761-06-9

104 Hofmann, Bernd: *Entwicklung eines rechnergestützten Managementsystems zur Beurteilung von Grundwasserschadensfällen*, 2000, ISBN 3-933761-07-7

105 Class, Holger: *Theorie und numerische Modellierung nichtisothermer Mehrphasenprozesse in NAPL-kontaminierten porösen Medien,* 2001, ISBN 3-933761-08-5

106 Schmidt, Reinhard: *Wasserdampf- und Heißluftinjektion zur thermischen Sanierung kontaminierter Standorte*, 2001, ISBN 3-933761-09-3

107 Josef, Reinhold:, *Schadstoffextraktion mit hydraulischen Sanierungsverfahren unter Anwendung von grenzflächenaktiven Stoffen*, 2001, ISBN 3-933761-10-7

108 Schneider, Matthias: *Habitat- und Abflussmodellierung für Fließgewässer mit unscharfen Berechnungsansätzen*, 2001, ISBN 3-933761-11-5

109 Rathgeb, Andreas: *Hydrodynamische Bemessungsgrundlagen für Lockerdeckwerke an überströmbaren Erddämmen*, 2001, ISBN 3-933761-12-3

110 Lang, Stefan: *Parallele numerische Simulation instätionärer Probleme mit adaptiven Methoden auf unstrukturierten Gittern*, 2001, ISBN 3-933761-13-1

111 Appt, Jochen; Stumpp Simone: *Die Bodensee-Messkampagne 2001, IWS/CWR Lake Constance Measurement Program 2001*, 2002, ISBN 3-933761-14-X

112 Heimerl, Stephan: *Systematische Beurteilung von Wasserkraftprojekten,* 2002, ISBN 3-933761-15-8, vergriffen

113 Iqbal, Amin: *On the Management and Salinity Control of Drip Irrigation,* 2002, ISBN 3-933761-16-6

114 Silberhorn-Hemminger, Annette: *Modellierung von Kluftaquifersystemen: Geostatistische Analyse und deterministisch-stochastische Kluftgenerierung,* 2002, ISBN 3-933761-17-4

115 Winkler, Angela: *Prozesse des Wärme- und Stofftransports bei der In-situ-Sanierung mit festen Wärmequellen,* 2003, ISBN 3-933761-18-2

116 Marx, Walter: *Wasserkraft, Bewässerung, Umwelt - Planungs- und Bewertungsschwerpunkte der Wasserbewirtschaftung,* 2003, ISBN 3-933761-19-0

117 Hinkelmann, Reinhard: *Efficient Numerical Methods and Information-Processing Techniques in Environment Water,* 2003, ISBN 3-933761-20-4

118 Samaniego-Eguiguren, Luis Eduardo: *Hydrological Consequences of Land Use / Land Cover and Climatic Changes in Mesoscale Catchments*, 2003, ISBN 3-933761-21-2

119 Neunhäuserer, Lina: *Diskretisierungsansätze zur Modellierung von Strömungs- und Transportprozessen in geklüftet-porösen Medien*, 2003, ISBN 3-933761-22-0

120 Paul, Maren: *Simulation of Two-Phase Flow in Heterogeneous Poros Media with Adaptive Methods*, 2003, ISBN 3-933761-23-9

121 Ehret, Uwe: *Rainfall and Flood Nowcasting in Small Catchments using Weather Radar*, 2003, ISBN 3-933761-24-7

122 Haag, Ingo: *Der Sauerstoffhaushalt staugeregelter Flüsse am Beispiel des Neckars - Analysen, Experimente, Simulationen -,* 2003, ISBN 3-933761-25-5

123 Appt, Jochen: *Analysis of Basin-Scale Internal Waves in Upper Lake Constance,* 2003, ISBN 3-933761-26-3

124 Hrsg.: Schrenk, Volker; Batereau, Katrin; Barczewski, Baldur; Weber, Karolin und Koschitzky, Hans-Peter: *Symposium Ressource Fläche und VEGAS - Statuskolloquium 2003, 30. September und 1. Oktober 2003,* 2003, ISBN 3-933761-27-1

125 Omar Khalil Ouda: *Optimisation of Agricultural Water Use: A Decision Support System for the Gaza Strip,* 2003, ISBN 3-933761-28-0

126 Batereau, Katrin: *Sensorbasierte Bodenluftmessung zur Vor-Ort-Erkundung von Schadensherden im Untergrund*, 2004, ISBN 3-933761-29-8

127 Witt, Oliver: *Erosionsstabilität von Gewässersedimenten mit Auswirkung auf den Stofftransport bei Hochwasser am Beispiel ausgewählter Stauhaltungen des Oberrheins*, 2004, ISBN 3-933761-30-1

128 Jakobs, Hartmut: *Simulation nicht-isothermer Gas-Wasser-Prozesse in komplexen Kluft-Matrix-Systemen*, 2004, ISBN 3-933761-31-X

129 Li, Chen-Chien: *Deterministisch-stochastisches Berechnungskonzept zur Beurteilung der Auswirkungen erosiver Hochwasserereignisse in Flussstauhaltungen,* 2004, ISBN 3-933761-32-8

130 Reichenberger, Volker; Helmig, Rainer; Jakobs, Hartmut; Bastian, Peter; Niessner, Jennifer: *Complex Gas-Water Processes in Discrete Fracture-Matrix Systems: Upscaling, Mass-Conservative Discretization and Efficient Multilevel Solution,* 2004, ISBN 3-933761-33-6

131 Hrsg.: Barczewski, Baldur; Koschitzky, Hans-Peter; Weber, Karolin; Wege, Ralf: *VEGAS - Statuskolloquium 2004,* Tagungsband zur Veranstaltung am 05. Oktober 2004 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2004, ISBN 3-933761-34-4

132 Asie, Kemal Jabir: *Finite Volume Models for Multiphase Multicomponent Flow through Porous Media*. 2005, ISBN 3-933761-35-2

133 Jacoub, George: *Development of a 2-D Numerical Module for Particulate Contaminant Transport in Flood Retention Reservoirs and Impounded Rivers,* 2004, ISBN 3-933761-36-0

134 Nowak, Wolfgang: *Geostatistical Methods for the Identification of Flow and Transport Parameters in the Subsurface,* 2005, ISBN 3-933761-37-9

135 Süß, Mia: *Analysis of the influence of structures and boundaries on flow and transport processes in fractured porous media*, 2005, ISBN 3-933761-38-7

136 Jose, Surabhin Chackiath: *Experimental Investigations on Longitudinal Dispersive Mixing in Heterogeneous Aquifers,* 2005, ISBN: 3-933761-39-5

137 Filiz, Fulya: *Linking Large-Scale Meteorological Conditions to Floods in Mesoscale Catchments,* 2005, ISBN 3-933761-40-9

138 Qin, Minghao: *Wirklichkeitsnahe und recheneffiziente Ermittlung von Temperatur und Spannungen bei großen RCC-Staumauern*, 2005, ISBN 3-933761-41-7

139 Kobayashi, Kenichiro: *Optimization Methods for Multiphase Systems in the Subsurface - Application to Methane Migration in Coal Mining Areas*, 2005, ISBN 3-933761-42-5

140 Rahman, Md. Arifur: *Experimental Investigations on Transverse Dispersive Mixing in Heterogeneous Porous Media,* 2005, ISBN 3-933761-43-3

141 Schrenk, Volker: *Ökobilanzen zur Bewertung von Altlastensanierungsmaßnahmen,* 2005, ISBN 3-933761-44-1

142 Hundecha, Hirpa Yeshewatesfa: *Regionalization of Parameters of a Conceptual Rainfall-Runoff Model*, 2005, ISBN: 3-933761-45-X

143 Wege, Ralf: *Untersuchungs- und Überwachungsmethoden für die Beurteilung natürlicher Selbstreinigungsprozesse im Grundwasser,* 2005, ISBN 3-933761-46-8

144 Breiting, Thomas: *Techniken und Methoden der Hydroinformatik - Modellierung von komplexen Hydrosystemen im Untergrund*, 2006, 3-933761-47-6

145 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Müller, Martin: *Ressource Untergrund: 10 Jahre VEGAS: Forschung und Technologieentwicklung zum Schutz von Grundwasser und Boden,* Tagungsband zur Veranstaltung am 28. und 29. September 2005 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2005, ISBN 3-933761-48-4

146 Rojanschi, Vlad: *Abflusskonzentration in mesoskaligen Einzugsgebieten unter Berücksichtigung des Sickerraumes,* 2006, ISBN 3-933761-49-2

147 Winkler, Nina Simone: *Optimierung der Steuerung von Hochwasserrückhaltebecken-systemen,* 2006, ISBN 3-933761-50-6

148 Wolf, Jens: *Räumlich differenzierte Modellierung der Grundwasserströmung alluvialer Aquifere für mesoskalige Einzugsgebiete*, 2006, ISBN: 3-933761-51-4

149 Kohler, Beate: *Externe Effekte der Laufwasserkraftnutzung*, 2006, ISBN 3-933761-52-2

150 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Stuhrmann, Matthias: *VEGAS-Statuskolloquium 2006,* Tagungsband zur Veranstaltung am 28. September 2006 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2006, ISBN 3-933761-53-0

151 Niessner, Jennifer: *Multi-Scale Modeling of Multi-Phase - Multi-Component Processes in Heterogeneous Porous Media*, 2006, ISBN 3-933761-54-9

152 Fischer, Markus: *Beanspruchung eingeerdeter Rohrleitungen infolge Austrocknung bindiger Böden,* 2006, ISBN 3-933761-55-7

153 Schneck, Alexander: *Optimierung der Grundwasserbewirtschaftung unter Berücksichtigung der Belange der Wasserversorgung, der Landwirtschaft und des Naturschutzes*, 2006, ISBN 3-933761-56-5

154 Das, Tapash: *The Impact of Spatial Variability of Precipitation on the Predictive Uncertainty of Hydrological Models,* 2006, ISBN 3-933761-57-3

155 Bielinski, Andreas: *Numerical Simulation of CO_2 sequestration in geological formations*, 2007, ISBN 3-933761-58-1

156 Mödinger, Jens: *Entwicklung eines Bewertungs- und Entscheidungsunterstützungssystems für eine nachhaltige regionale Grundwasserbewirtschaftung*, 2006, ISBN 3-933761-60-3

157 Manthey, Sabine: *Two-phase flow processes with dynamic effects in porous media - parameter estimation and simulation,* 2007, ISBN 3-933761-61-1

158 Pozos Estrada, Oscar: *Investigation on the Effects of Entrained Air in Pipelines,* 2007, ISBN 3-933761-62-X

159 Ochs, Steffen Oliver: *Steam injection into saturated porous media – process analysis including experimental and numerical investigations*, 2007, ISBN 3-933761-63-8

160 Marx, Andreas: *Einsatz gekoppelter Modelle und Wetterradar zur Abschätzung von Niederschlagsintensitäten und zur Abflussvorhersage, 2007,* ISBN 3-933761-64-6

161 Hartmann, Gabriele Maria: *Investigation of Evapotranspiration Concepts in Hydrological Modelling for Climate Change Impact Assessment*, 2007, ISBN 3-933761-65-4

162 Kebede Gurmessa, Tesfaye: *Numerical Investigation on Flow and Transport Characteristics to Improve Long-Term Simulation of Reservoir Sedimentation*, 2007, ISBN 3-933761-66-2

163 Trifković, Aleksandar: *Multi-objective and Risk-based Modelling Methodology for Planning, Design and Operation of Water Supply Systems*, 2007, ISBN 3-933761-67-0

164 Götzinger, Jens: *Distributed Conceptual Hydrological Modelling - Simulation of Climate, Land Use Change Impact and Uncertainty Analysis*, 2007, ISBN 3-933761-68-9

165 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Stuhrmann, Matthias: *VEGAS – Kolloquium 2007,* Tagungsband zur Veranstaltung am 26. September 2007 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2007, ISBN 3-933761-69-7

166 Freeman, Beau: *Modernization Criteria Assessment for Water Resources Planning; Klamath Irrigation Project, U.S.*, 2008, ISBN 3-933761-70-0

167 Dreher, Thomas: *Selektive Sedimentation von Feinstschwebstoffen in Wechselwirkung mit wandnahen turbulenten Strömungsbedingungen*, 2008, ISBN 3-933761-71-9

168 Yang, Wei: *Discrete-Continuous Downscaling Model for Generating Daily Precipitation Time Series*, 2008, ISBN 3-933761-72-7

169 Kopecki, Ianina: *Calculational Approach to FST-Hemispheres for Multiparametrical Benthos Habitat Modelling,* 2008*,* ISBN 3-933761-73-5

170 Brommundt, Jürgen: *Stochastische Generierung räumlich zusammenhängender Niederschlagszeitreihen*, 2008, ISBN 3-933761-74-3

171 Papafotiou, *Alexandros: Numerical Investigations of the Role of Hysteresis in Heterogeneous Two-Phase Flow Systems*, 2008, ISBN 3-933761-75-1

172 He, Yi: *Application of a Non-Parametric Classification Scheme to Catchment Hydrology*, 2008, ISBN 978-3-933761-76-7

173 Wagner, Sven: *Water Balance in a Poorly Gauged Basin in West Africa Using Atmospheric Modelling and Remote Sensing Information*, 2008, ISBN 978-3-933761-77-4

174 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Stuhrmann, Matthias; Schrenk, Volker: *VEGAS-Kolloquium 2008 Ressource Fläche III*, Tagungsband zur Veranstaltung am 01. Oktober 2008 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2008, ISBN 978-3-933761-78-1

175 Patil, Sachin: *Regionalization of an Event Based Nash Cascade Model for Flood Predictions in Ungauged Basins*, 2008, ISBN 978-3-933761-79-8

176 Assteerawatt, Anongnart: *Flow and Transport Modelling of Fractured Aquifers based on a Geostatistical Approach,* 2008, ISBN 978-3-933761-80-4

177 Karnahl, Joachim Alexander: *2D numerische Modellierung von multifraktionalem Schwebstoff- und Schadstofftransport in Flüssen*, 2008, ISBN 978-3-933761-81-1

178 Hiester, Uwe: *Technologieentwicklung zur In-situ-Sanierung der ungesättigten Bodenzone mit festen Wärmequellen,* 2009, ISBN 978-3-933761-82-8

179 Laux, Patrick: *Statistical Modeling of Precipitation for Agricultural Planning in the Volta Basin of West Africa,* 2009, ISBN 978-3-933761-83-5

180 Ehsan, Saqib: *Evaluation of Life Safety Risks Related to Severe Flooding,* 2009, ISBN 978-3-933761-84-2

181 Prohaska, Sandra: *Development and Application of a 1D Multi-Strip Fine Sediment Transport Model for Regulated Rivers,* 2009, ISBN 978-3-933761-85-9

182 Kopp, Andreas: *Evaluation of CO_2 Injection Processes in Geological Formations for Site Screening,* 2009, ISBN 978-3-933761-86-6

183 Ebigbo, Anozie: *Modelling of biofilm growth and its influence on CO_2 and water (two-phase) flow in porous media,* 2009, ISBN 978-3-933761-87-3

184 Freiboth, Sandra: *A phenomenological model for the numerical simulation of multiphase multicomponent processes considering structural alterations of porous media*, 2009, ISBN 978-3-933761-88-0

185 Zöllner, Frank: *Implementierung und Anwendung netzfreier Methoden im Konstruktiven Wasserbau und in der Hydromechanik*, 2009, ISBN 978-3-933761-89-7

186 Vasin, Milos: *Influence of the soil structure and property contrast on flow and transport in the unsaturated zone*, 2010, ISBN 978-3-933761-90-3

187 Li, Jing: *Application of Copulas as a New Geostatistical Tool*, 2010, ISBN 978-3-933761-91-0

188 AghaKouchak, Amir: *Simulation of Remotely Sensed Rainfall Fields Using Copulas*, 2010, ISBN 978-3-933761-92-7

189 Thapa, Pawan Kumar: *Physically-based spatially distributed rainfall runoff modelling for soil erosion estimation*, 2010, ISBN 978-3-933761-93-4

190 Wurms, Sven: *Numerische Modellierung der Sedimentationsprozesse in Retentionsanlagen zur Steuerung von Stoffströmen bei extremen Hochwasserabflussereignissen*, 2011, ISBN 978-3-933761-94-1

191 Merkel, Uwe: *Unsicherheitsanalyse hydraulischer Einwirkungen auf Hochwasserschutzdeiche und Steigerung der Leistungsfähigkeit durch adaptive Strömungsmodellierung*, 2011, ISBN 978-3-933761-95-8

192 Fritz, Jochen: *A Decoupled Model for Compositional Non-Isothermal Multiphase Flow in Porous Media and Multiphysics Approaches for Two-Phase Flow*, 2010, ISBN 978-3-933761-96-5

193 Weber, Karolin (Hrsg.): *12. Treffen junger WissenschaftlerInnen an Wasserbauinstituten*, 2010, ISBN 978-3-933761-97-2

194 Bliefernicht, Jan-Geert: *Probability Forecasts of Daily Areal Precipitation for Small River Basins,* 2011, ISBN 978-3-933761-98-9

195 Hrsg.: Koschitzky, Hans-Peter; Braun, Jürgen: *VEGAS-Kolloquium 2010 In-situ-Sanierung - Stand und Entwicklung Nano und ISCO -,* Tagungsband zur Veranstaltung am 07. Oktober 2010 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2010, ISBN 978-3-933761-99-6

196 Gafurov, Abror: *Water Balance Modeling Using Remote Sensing Information - Focus on Central Asia,* 2010, ISBN 978-3-942036-00-9

197 Mackenberg, Sylvia: *Die Quellstärke in der Sickerwasserprognose: Möglichkeiten und Grenzen von Labor- und Freilanduntersuchungen,* 2010, ISBN 978-3-942036-01-6

198 Singh, Shailesh Kumar: *Robust Parameter Estimation in Gauged and Ungauged Basins,* 2010, ISBN 978-3-942036-02-3

199 Doğan, Mehmet Onur: *Coupling of porous media flow with pipe flow,* 2011, ISBN 978-3-942036-03-0

200 Liu, Min: *Study of Topographic Effects on Hydrological Patterns and the Implication on Hydrological Modeling and Data Interpolation,* 2011, ISBN 978-3-942036-04-7

201 Geleta, Habtamu Itefa: *Watershed Sediment Yield Modeling for Data Scarce Areas,* 2011, ISBN 978-3-942036-05-4

202 Franke, Jörg: *Einfluss der Überwachung auf die Versagenswahrscheinlichkeit von Staustufen,* 2011, ISBN 978-3-942036-06-1

203 Bakimchandra, Oinam: *Integrated Fuzzy-GIS approach for assessing regional soil erosion risks,* 2011, ISBN 978-3-942036-07-8

204 Alam, Muhammad Mahboob: *Statistical Downscaling of Extremes of Precipitation in Mesoscale Catchments from Different RCMs and Their Effects on Local Hydrology,* 2011, ISBN 978-3-942036-08-5

205 Hrsg.: Koschitzky, Hans-Peter; Braun, Jürgen: *VEGAS-Kolloquium 2011 Flache Geothermie - Perspektiven und Risiken,* Tagungsband zur Veranstaltung am 06. Oktober 2011 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2011, ISBN 978-3-933761-09-2

206 Haslauer, Claus: *Analysis of Real-World Spatial Dependence of Subsurface Hydraulic Properties Using Copulas with a Focus on Solute Transport Behaviour,* 2011, ISBN 978-3-942036-10-8

207 Dung, Nguyen Viet: *Multi-objective automatic calibration of hydrodynamic models – development of the concept and an application in the Mekong Delta,* 2011, ISBN 978-3-942036-11-5

208 Hung, Nguyen Nghia: *Sediment dynamics in the floodplain of the Mekong Delta, Vietnam,* 2011, ISBN 978-3-942036-12-2

209 Kuhlmann, Anna: *Influence of soil structure and root water uptake on flow in the unsaturated zone,* 2012, ISBN 978-3-942036-13-9

210 Tuhtan, Jeffrey Andrew: *Including the Second Law Inequality in Aquatic Ecodynamics: A Modeling Approach for Alpine Rivers Impacted by Hydropeaking,* 2012,
ISBN 978-3-942036-14-6

211 Tolossa, Habtamu: *Sediment Transport Computation Using a Data-Driven Adaptive Neuro-Fuzzy Modelling Approach,* 2012, ISBN 978-3-942036-15-3

212 Tatomir, Alexandru-Bodgan: *From Discrete to Continuum Concepts of Flow in Fractured Porous Media,* 2012, ISBN 978-3-942036-16-0

213 Erbertseder, Karin: *A Multi-Scale Model for Describing Cancer-Therapeutic Transport in the Human Lung,* 2012, ISBN 978-3-942036-17-7

214 Noack, Markus: *Modelling Approach for Interstitial Sediment Dynamics and Reproduction of Gravel Spawning Fish,* 2012, ISBN 978-3-942036-18-4

215 De Boer, Cjestmir Volkert: *Transport of Nano Sized Zero Valent Iron Colloids during Injection into the Subsurface*, 2012, ISBN 978-3-942036-19-1

216 Pfaff, Thomas: *Processing and Analysis of Weather Radar Data for Use in Hydrology*, 2013, ISBN 978-3-942036-20-7

217 Lebrenz, Hans-Henning: *Addressing the Input Uncertainty for Hydrological Modeling by a New Geostatistical Method*, 2013, ISBN 978-3-942036-21-4

218 Darcis, Melanie Yvonne: *Coupling Models of Different Complexity for the Simulation of CO_2 Storage in Deep Saline Aquifers*, 2013, ISBN 978-3-942036-22-1

219 Beck, Ferdinand: *Generation of Spatially Correlated Synthetic Rainfall Time Series in High Temporal Resolution - A Data Driven Approach*, 2013,
ISBN 978-3-942036-23-8

Die Mitteilungshefte ab der Nr. 134 (Jg. 2005) stehen als pdf-Datei über die Homepage des Instituts: www.iws.uni-stuttgart.de zur Verfügung.