Ein stochastisches Modell zur Beschreibung stündlichen Niederschlages

Dissertation
zur Erlangung des wirtschaftswissenschaftlichen Doktorgrades
der wirtschaftswissenschaftlichen Fakultät
der Universität Göttingen

vorgelegt von
Ulrike Ohlmer

aus Hildesheim

Göttingen, 1998

Erstgutachter:	Prof. Dr. W. Zucchini
Zweitgutachter:	Prof. Dr. F. Böker
Tag der mündlichen Prüfung:	11.12.1998

Inhaltsverzeichnis

Abbildungsverzeichnis vii

Tabellenverzeichnis ix

Bezeichnungen x

Abstract **1**

1 Einleitung **3**

2 Stochastische Modelle **7**

1 Stochastische Modellierung von Niederschlag 7

2 Poissonprozesse 10

3 Markovketten 12

4 Hidden–Markov–Modelle 15

5 Alternierende Erneuerungsprozesse 17

6 Erneuerungs–Cox–Markov–Prozesse 19

7 Clustermodelle 19

8 ARMA– und DARMA–Modelle 25

9 Stündlicher Niederschlag - Acremans einfaches Modell 26

10 Punktprozesse in stetiger und Beobachtungen in diskreter Zeit 27

3 Der Datensatz **31**

1 Auswahl eines Schwellenwertes 34

2 Auswahl homogener Perioden 38

3 Der Niederschlagsmengenprozeß 41

4 Täglicher Niederschlag **49**

1 Der Ereignisprozeß 49

2 Die Niederschlagsmengen an nassen Tagen 56

3 Simulation 60

4 Modellvalidation 63

5 Ergebnisse 64

5 Das Neyman–Scott–Modell **69**

1 Definition und Eigenschaften 70

2 Modellanpassung für eine homogene Periode 80

3 Saisonale und innertägliche Variation 85

4 Glättung der Modellparameter – Set B4 89

5 Glättung der Modellparameter – Set B1 111

6 Berechnung der Modellparameter für alle Stunden 129

6 Modellvalidation **131**

1 Simulation stündlichen Niederschlages 131

2 Momente der simulierten Daten 137

3 Extremer Niederschlag 145

4 Verteilung der Niederschlagsmengen in nassen Stunden 160

5 Verteilung der Runlängen 163

6 Verteilung des Niederschlages an nassen Tagen 168

7 Zusammenfassung 173

7 Zusammenfassung **175**

A Die Rohdaten **177**

B Täglicher Niederschlag **183**

1 Der Ereignisprozeß 183

2 Die Niederschlagsmengen an nassen Tagen 185

3 Modellvalidation 188

4 S-PLUS-Routinen 191

C Das Neyman–Scott–Modell **197**

1 Modellanpassung 197

2 Simulation stündlichen Niederschlages 203

3 Algorithmus zum Zählen der Runlängen 206

Literaturverzeichnis **211**

Abbildungsverzeichnis

2.1 Aufbau des Neyman–Scott–Clustermodelles 22

3.1 Jahresniederschlagsmenge und - dauer 32

3.2 Ereignismenge 32

3.3 Ereignisdauer 32

3.4 Cusum Charts 34

3.5 Niederschlagsmengen an nassen Tagen 37

3.6 Stationarität der Beobachtungen 39

3.7 Monatsniederschlag 40

3.8 Regenereignisse 41

3.9 Empirisches Ein-Stunden-Mittel 44

3.10 Empirische Varianzen 45

3.11 Empirische Varianzen 45

3.12 Empirische Lag(1)-Autokorrelationen 47

3.13 Empirische Lag(1)-Autokorrelationen 47

3.14 Empirische Übergangswahrscheinlichkeiten 47

3.15 Empirische Übergangswahrscheinlichkeiten 47

3.16 Empirische Wahrscheinlichkeiten trockener Intervalle 48

4.1 ACF des Ereignisprozesses 50

4.2 Niederschlagsmengen an nassen Tagen 58

4.3 Jahresniederschlag 65

4.4 Monatsniederschlag 66

4.5 Niederschlagsmenge an nassen Tagen 66

4.6 Anzahl nasser Tage pro Monat 66

4.7 Quantile der Niederschlagsmenge an nassen Tagen 67

4.8 Täglicher Niederschlag - Clusterverhalten 67

4.9 Täglicher Niederschlag - Clusterverhalten 68

4.10 Täglicher Niederschlag - Clusterverhalten 68

5.1 **Skizze zur Kovarianzstruktur** 72

5.2 Lag(1)-Autokorrelation als Funktion der Intervalldauer 77

5.3 Lag(1)-Autokorrelation als Funktion von δ und ν 78

5.4 $\phi(h)$ und $\phi_{WW}(h)$ als Funktionen der Modellparameter 78

5.5 $\phi(h)$ und $\phi_{WW}(h)$ als Funktionen der Modellparameter 79

5.6 Empirische Momente als Funktion der Intervalllänge 81

5.7 Empirische Wahrscheinlichkeit $\phi_D(h = 1)$ 86

5.8 Periodogramm von λ — Set B4 91

5.9 Modellierung von λ — Set B4 92

5.10 Modellierung von ν — Set B4 93

5.11 Modellierung von β — Set B4 100

5.12 Modellierung von δ — Set B4 100

5.13 Modellierung von ξ — Set B4 102

5.14 Modellierung von λ — Set B1 117

5.15 Modellierung von ν — Set B1 118

5.16 Modellierung von β — Set B1 119

5.17 Modellierung von δ — Set B1 119

5.18 Modellierung von ξ — Set B1 119

6.1 Beobachtete und simulierte Momente: $\mu(1)$, $\gamma(1)$ 139

6.2 Beobachtete und simulierte Momente: $\gamma(3)$, $\gamma(6)$ 140

6.3 Beobachtete und simulierte Momente: $\gamma(12)$, $\gamma(24)$ 141

6.4 Beobachtete und simulierte Momente: $\rho(1,1)$, $\rho(3,1)$ 141

6.5 Beobachtete und simulierte Momente: $\rho(6,1)$, $\rho(12,1)$ 141

6.6 Beobachtete und simulierte Momente: $\rho(24,1)$, $\phi_{WW}(1)$ 142

6.7 Beobachtete und simulierte Momente: $\phi_{WW}(3)$, $\phi_{WW}(6)$ 142

6.8 Beobachtete und simulierte Momente: $\phi_{WW}(12)$, $\phi_{WW}(24)$ 142

6.9 Beobachtete und simulierte Momente: $\phi_{DD}(24)$, $\phi(24)$ 143

6.10 Jahresniederschlag 143

6.11 Monatssniederschlag: Januar 144

6.12 Monatssniederschlag: Juli 144

6.13 1-Stunden-Maxima 148

6.14 1-Stunden-Maxima 149

6.15 6-Stunden-Maxima 151

6.16 6-Stunden-Maxima 153

6.17 24-Stunden-Maxima 156

6.18 24-Stunden-Maxima 157

6.19 24-Stunden-Maxima: Ausgewählte Monate 158

6.20 Niederschlagsmengen in nassen Stunden 161

6.21 Niederschlagsmengen in nassen Stunden 162

6.22 Niederschlagsmengen in nassen Stunden 162

6.23 Nasse Runs 165

6.24 Nasse Runs – Maximallängen 166

6.25 Nasse Runs – Maximallängen 167

6.26 Trockene Runs - Mindestlängen 168

6.27 Trockene Runs - Mindestlängen 169

6.28 Trockene Runs 170

6.29 Trockene Runs - Klasseneinteilung 171

6.30 Innertägliche Verteilung der Niederschläge an nassen Tagen 172

B.1 Mittlerer Niederschlag an nassen Tagen 187

Tabellenverzeichnis

3.1 Eigenschaften des Jahresniederschlages 33

3.2 ACF der Niederschlagsmengen an nassen Tagen 35

3.3 Auswirkung des Schwellenwertes 38

3.4 Monatsniederschlag 42

3.5 Niederschlagsereignisprozeß 43

3.6 Empirische Momente des aggregierten Niederschlagsprozesses 46

4.1 ACF des Ereignisprozesses 50

4.2 Jahresniederschlag 63

5.1 30-Tages-Perioden in der NSRP-Modellanpassung 86

5.2 Parameterglättung: Set B4, λ 98

5.3 Parameterglättung: ν, β, δ 99

5.4 Parameterglättung: ξ, Set B4 101

5.5 Parameterglättung: Set B1, λ 117

5.6 Parameterglättung: ν, β, δ 118

5.7 Parameterglättung: ξ, Set B1 120

6.1 Erwartungswert und Varianzen 138

6.2 Korrelationen 139

6.3 Wahrscheinlichkeiten nasser und trockener Intervalle 140

6.4 Jahres- und Monatsniederschläge 143

6.5 1-Stunden-Maxima 146

6.6 1-Stunden-Maxima pro Set 147

6.7 6-Stunden-Maxima 150

6.8 24-Stunden-Maxima 154

6.9 24-Stunden-Maxima pro Set 155

6.10 Quantile der Niederschlages in nassen Stunden 161

B.1 Modellierung der Übergangswahrscheinlichkeiten 184

B.2 Monats- und Jahresniederschlag 188

B.3 Anzahl nasser Tage 189

B.4 Niederschlag an nassen Tagen 189

B.5 Niederschlag an nassen Tagen - Quantile 190

Bezeichnungen

Alle verwendeten Bezeichnungen orientieren sich an der in Mathematik und Statistik üblichen Standardnotation. Abweichungen davon sind im Text definiert. In der üblichen Notation bezeichnet $\hat{a}$ den Schätzer eines Parameters a.

Modellüberblick, Beschreibung des Datensatzes, täglicher Niederschlag

$1(\cdot)$	Indikatorfunktion.
$[x]$	größte ganza Zahl, die kleiner oder gleich x ist: $[x] = \max\{z \in \boldsymbol{Z} \mid z \leq x\}$
C_k	Zustand des klimatologischen Prozesses an Tag k in einem Hidden-Markov-Modell.
$CV(X)$	Variationskoeffizient einer Zufallsvariablen X: $CV(X) = \sqrt{\mathrm{Var}X}/\,\mathrm{E}X$.
D_E	Dauer eines Regenereignisses
$\bar{D}_E, s(D_E)$	Mittel bzw. emp. Standardabweichung von D_E.
D_J	Gesamtniederschlagsdauer in einem Jahr
$I(t)$	Dispersionsindex eines Punktprozesses: $I(t) = Var(N(t))/\,\mathrm{E}N(t)$.
$I_E = Y_E/D_E$	Ereignisintensität
L_0, L_1	Länge eines trockenen bzw. nassen Intervalles in einem alternierenden Erneuerungsprozeß.
$N(t)$	Anzahl der Punkte im Intervall $(0,t]$ in stetiger Zeit:$N(t) = N((0,t])$.
N_k	Anzahl der Punkte bis inklusive des k-ten Zeitpunktes (diskrete Zeit)
$NR(t)$	Anzahl der Jahre $n = 1,\ldots,N$, in denen Tag t naß ist: $NR(t) = \sum_{n=1}^{N} 1(Z_{t,n} = 1)$
$ND(t)$	Anzahl der Jahre $n = 1,\ldots,N$, in denen Tag t trocken ist: $ND(t) = \sum_{n=1}^{N} 1(Z_{t,n} = 0)$
$NRR(t)$	$= \sum_{n=1}^{N} 1(Z_{t-1,n} = 1) \cdot 1(Z_{t,n} = 1)$
$NRRR(t)$	$= \sum_{n=1}^{N} 1(Z_{t-2,n} = 1) \cdot 1(Z_{t-1,n} = 1) \cdot 1(Z_{t,n} = 1)$
$NDRR(t)$	$= \sum_{n=1}^{N} 1(Z_{t-2,n} = 0) \cdot 1(Z_{t-1,n} = 1) \cdot 1(Z_{t,n} = 1)$
$NRDR(t)$	$= \sum_{n=1}^{N} 1(Z_{t-2,n} = 1) \cdot 1(Z_{t-1,n} = 0) \cdot 1(Z_{t,n} = 1)$
R_j	Amplitude der j-ten Fourierfrequenz bei Anpassung einer Fouriersumme an Beob. $x_1,\ldots,x_t$.
S	Residuenquadratsumme
T_n	Wartezeit ab dem Zeitpunkt 0 auf den n-ten Punkt.
U_n	Verweildauer zwischen dem $n-1$-ten und dem n-ten Punkt.
$W_{t,n}(d)$	Niederschlagsmenge an Tag t in Jahr n, gegeben an Tag t fällt mehr als $d$mm Niederschlag.
Y_t, $Y_{t,n}$	Niederschlagsintensität zur Zeit t (in Jahr n).
Y_k	Niederschlagsmenge an Tag k.
Y_E	Niederschlagsmenge eines Regenereignisses
$\bar{Y}_E, s(Y_E)$	Mittel bzw. emp. Standardabweichung von Y_E.
Y_J	Jahresniederschlag
$Z_k(y)$	$0-1$-Prozeß täglichen Niederschlages: $Z_k(y) = 1(Y_k > y)$.
$Z_k^{(n)}$	Zustand des Ereignisprozesses an Tag k an Station n (Hidden-Markov-Modell).

d	Niederschlagsschwellenwert. Tag t ist naß wenn $Y_t \geq d$
$m(t)$	$= \mathrm{E}W_{t,n}$ bedingt erwartete Niederschlagsmenge an Tag t, gegeben, Tag t ist naß.
$\hat{m}_t$	$= \sum_{n=1}^{N} w_{t,n}/N_t$ Mittel der Niederschlagsmengen an nassen Tagen t.
θ_j	Phase der j-ten Fourierfrequenz ω_j.
$\mu_t = \mathrm{E}Y_{t,n}$	Erwartungswert der Niederschlagsmenge an Tag t in Jahr n.
π_{ij}	Übergangswahrscheinlichkeiten einer Markovkette: $\pi_{ij} = \mathrm{P}(Z_k = j \mid Z_{k-1} = i)$.
$\Pi = (\pi_{ij})$	Matrix der Übergangswahrscheinlichkeiten einer Markovkette.
ω_j	j-te Fourierfrequenz bei Anpassung einer Fouriersumme an Beob. $x_1 \ldots, x_T$: $\omega_j = 2\pi\, j/T$.

Bei Betrachtung eines stationären Prozesses sind

$\gamma(h)$	$:= \mathrm{Var}\left(Y_i^{(h)}\right)$
$\gamma(h,k)$	$:= \mathrm{Cov}\left(Y_i^{(h)}, Y_{i+k}^{(h)}\right)$
$\mu(h)$	$= E\left(Y_i^{(h)}\right)$
$\rho(h,k)$	$:= \mathrm{Cor}\left(Y_i^{(h)}, Y_{i+k}^{(h)}\right)$
$\phi_{WW}(h)$	$:= P\left(Y_i^{(h)} > 0 \mid Y_{i-1}^{(h)} > 0\right)$
$\phi_{DD}(h)$	$:= P\left(Y_i^{(h)} = 0 \mid Y_{i-1}^{(h)} = 0\right)$
$\phi_D(h)$	$:= \phi(h) := P(Y_i^{(h)} = 0)$

Das Neyman–Scott–Modell

C:	Clustergröße
D:	Abstand zwischen Clusterzentrum und Zellstart
L:	Zelldauer
X:	Zellintensität
β:	$1/\beta$ ist der Erwartungswert des Abstandes zwischen einem Clusterzentrum und dem Start einer Zelle des Clusters
δ:	$:= \eta - \beta$
η :	$1/\eta$ ist der Erwartungswert der Zelldauer
λ:	Intensitätsrate des Prozesses der Clusterzentren
ν:	Erwartungswert der Clustergröße
ξ:	$1/\xi$ ist die erwartete Intensität einer Zelle
$\tilde{\beta}$	$:= \log \beta$
$\tilde{\delta}$	$:= \log \delta$
$\tilde{\lambda}$	$:= \log \lambda$
$\tilde{\nu}$	$:= \log(\nu - 1)$
$\tilde{\xi}$	$:= \log \xi$
f_i	$= f_i(\lambda, \nu, \beta, \eta, \xi)$ Momente des Niederschlagsmengenprozesses unter dem Modell
$\hat{f}_i$:	empirischer Schätzer von f_i
$\tilde{f}_i$:	Wert von f_i unter einem geglätteten Parameterset (B1 bzw. B4)
$\mathrm{P}^{(D)}(i)$, $\mathrm{P}^{(W)}(i)$:	Wahrscheinlichkeiten eines trockenen bzw. nassen i-Runs
$\mathrm{P}^{(D)}(k+)$:	Wahrscheinlichkeit eines trockenen Runs der Mindestlänge k
$\mathrm{P}^{(W)}(k-)$:	Wahrscheinlichkeit eines nassen Runs der Maximallänge k
$S_{f,k}$	$:= \frac{1}{T} \sum_t \frac{\sqrt{(\hat{f}_t - f_{t,k})^2}}{\lvert \hat{f}_t \rvert}$, in der Parameterglättung verwendete Residuenquadratsumme.
S_k	in der Parameterglättung verwendete Res.summe, z.B. $S_k := \sum_t (\hat{\lambda}_t - \lambda_{t,k})^2$
$X_n^{(h)}, \tilde{X}_n^{(h)}$	Maximum des h-Stunden-Niederschlages in Jahr n der beobachteten bzw. der simulierten Daten.
$\mathcal{F}$:	zur Modellüberprüfung verwendete Second-Order-Eigenschaften
$\mathcal{P}$:	zur Modellanpassung verwendete Second-Order-Eigenschaften
Set B1	Parameter des Neyman–Scott–Modelles, durch Modellanpassung an alle im Abstand von 6 Stunden beginnenden 30-Tages-Perioden entstanden
Set B4	Parameter des Neyman–Scott–Modelles, durch separate Modellanpassung an alle 30-Tages-Perioden mit Start um 0h, 6h, 12h, 18h und anschließendes Zusammensetzen der Parametersets im Reißverschlußverfahren entstanden

Kapitel 1

Einleitung

Viele menschliche Aktivitäten werden durch Niederschläge beeinflußt. Niederschläge besitzen eine fundamentale Bedeutung in der Landwirtschaft, sie müssen in der Planung urbaner Kanalisationssysteme berücksichtigt werden, ebenso in der Auswahl geeigneter Deponiestandorte für Hausmüll oder für radioaktiven Müll. Aber auch neuere Modelle für das Pflanzenwachstum berücksichtigen in stärkerem Maße Art und Auftreten von Niederschlägen. Als Input-Variable wird Niederschlag in Niederschlags–Abfluß–Modellen für Wasserstände von Flüssen verwendet, die wiederum für die Konstruktion von Hochwasserschutzeinrichtungen benötigt werden. Die verheerenden Hochwasser in diesem Jahr in Osteuropa oder in Bangladesh verdeutlichen die Notwendigkeit verläßlicher Modelle für hydrologische Zwecke.

Die Modellierung von Niederschlägen ist ein Grenzgebiet, auf dem sich Hydrologie, Meteorologie und Physik überschneiden. Da die Beschreibung eines so komplexen Phänomenes durch deterministische Gesetze nahezu unmöglich ist, ist der Versuch einer stochastischen Modellierung naheliegend. Klimatische Daten, insbesondere solche die Niederschläge betreffende, werden schon lange auf der ganzen Welt gesammelt. Entsprechend wurde schon früh versucht, die stochastische Natur des Phänomenes Regen zu beschreiben. Quetélet untersucht in einer 1852 veröffentlichten Arbeit die Folge trockener und nasser Tage in Brüssel und bemerkt dabei das Phänomen der Persistenz: Die Wahrscheinlichkeit, an einem Tag Regen zu erhalten, ist nach einem vorangegangenen Regentag größer als an einem beliebigen Tag (Katz [1984]). Wünschenswert in der stochastischen Beschreibung von Niederschlägen ist, grundlegende physikalische Strukturen zu identifizieren und diese mit Hilfe eines Modelles zu beschreiben.

Ein Modell in diesem Sinne ist ein Mechanismus, mit dem künstliche Niederschlagsdaten generiert werden können, deren Eigenschaften denen der beobachteten Serie gleichen. Die Komplexität des Phänomenes 'Regen' verhindert im allgemeinen, daß alle Eigenschaften der Daten korrekt reproduziert werden, wenn das Modell nur eine begrenzte Anzahl von Parametern verwendet. Hier wird der in der statistischen Modellierung stets notwendige Kompromiß zwischen einer geringen Anzahl an Modellparametern und einer geringen Abweichung des Modelles von den Beobachtungen evident. Unter Verwendung einer hinreichend großen Anzahl an Modellparametern

ist natürlich eine exakte Beschreibung des beobachteten Phänomenes möglich — dieses Modell wird aber kaum einen anderen Satz an Beobachtungen, beispielsweise Niederschlagsmessungen eines anderen Zeitraumes oder einer benachbarten Meßstation, angemessen beschreiben können. Eine wichtige Aufgabe eines Modelles besteht also in der Betonung relevanter Eigenschaften und damit einhergehend der Glättung der Beobachtungen. Folglich legt der Anwendungszweck des Modelles fest, welche der Eigenschaften des beobachteten Phänomenes beonders exakt wiederzugeben sind. Mögliche Schwerpunkte sind die Niederschlagsmengen auf monatlicher Basis, die zeitliche Verteilung der Niederschläge oder die stündlichen Niederschlagsmengen. Mit einem geeigneten Modell lassen sich dann auch Fragen beantworten, die für Aufgaben des Hoch- und Tiefbaus und für versicherungstechnische Zwecke relevant, aber theoretisch nur schwierig zu klären sind, zum Beispiel: ,,Wie groß ist die Wahrscheinlichkeit, daß ein Sturm mit mehr als 40mm/Std. Niederschlag einmal im Jahr auftritt?" Zur Beantwortung werden mit dem angepaßten Modell hinreichend viele Daten generiert und das Vorkommen der relevanten Ereignisse gezählt. Da Beobachtungen beliebig vieler Jahre erzeugt werden können, sind die mit einem guten Modell erhaltenen Antworten verläßlicher als die aus einer möglicherweise nur geringen Anzahl beobachteter Jahre abgeleiteten.

Niederschläge sind meistens flüssig und werden dann gewöhnlich als Regen bezeichnet; sie können aber als Schnee oder Hagel auch fest sein. Tau und Nebel werden als Spurenniederschlag bezeichnet. Wir werden 'Niederschlag' und 'Regen' synonym verwenden — wodurch offenbar wird, daß die vorliegende Arbeit an einem schneearmen Ort erstellt worden ist. Niederschlagsmessungen sind oft fehlerbehaftet. Beispielsweise können Inhomogenitäten der Daten auftreten, wenn eine Meßstation örtlich verlagert wird oder in ihrer näheren Umgebung Gebäude entstehen. Wird die Ablesezeit verlegt, können systematische Fehler ebenso entstehen wie bei der Verwendung eines anderen Meßgerätes. Bei stark windigem Wetter wird die Niederschlagsmenge häufig unterschätzt, wenn die Meßstation nicht extra geschützt wird (YANG et al. [1998]). Weitere systematische Fehler können durch die Art des Aufzeichnens oder des Ablesens der Niederschlagsmengen entstehen: Die Beobachtungen werden in digitaler oder analoger Form aufgezeichnet. Meßgeräte mit analogem Aufzeichnen der Beobachtungen auf einer sich drehenden Karte können die Anzahl der Tage mit geringer Niederschlagsmenge unterschätzen, falls die Spur des Bleistiftes geringe Veränderungen verbirgt. Müssen die Niederschlagsmengen von einem Menschen abgelesen werden, läßt sich zeigen, daß Vielfache von 0.05 Inch (1.25 mm) wesentlich häufiger notiert werden als andere Werte (WOOLHISER [1992]). YANG et al. [1998] zeigen, daß eine Verringerung des Abstandes des Meßgerätes vom Boden zu signifikant höheren Niederschlagsmeßwerten führt. Ebenso führt Schnee zu verzerrten Messungen, wenn das Meßgerät nicht mit einer Heizmechanik und einem Schneekreuz ausgestattet ist. Diese praktischen Probleme beeinflussen auch die Modellauswahl: Es ist nicht sinnvoll, mit einem komplizierten Modell Niederschläge exakt zu beschreiben, wenn die Schätzer der Modellparameter sensibel auf geringe Änderungen der Beobachtungen reagieren.

Niederschlag ist sowohl ein zeitliches als auch ein räumliches Phänomen, daher sind zur Beschreibung beider Strukturen Konzepte aus dem Bereich der mehrdimensionalen Zufallsvariablen notwendig. Einige stochastische Prozesse, wie der Poissonpro-

zeß, lassen sich auf höhere Dimensionen erweitern, doch generell ist diese Verallgemeinerung nicht leicht. Beispielsweise basieren die oft zur Beschreibung zeitlicher Phänomene verwendeten Markovprozesse auf der Existenz einer natürlichen Ordnungsrelation der Indexmenge, die in höherdimensionalen Räumen nicht gegeben ist. Glücklicherweise lassen sich die räumlich-zeitlichen Strukturen von Niederschlägen besser durch Clusterprozesse als durch mehrdimensionale Markovprozesse beschreiben. In dieser Arbeit werden wir zwar nur die zeitliche Verteilung von Niederschlägen modellieren, das ausgewählte Modell sollte sich aber in zukünftigen Untersuchungen zu einem räumlich–zeitlichen Modell erweitern lassen. Diese Bemerkungen sollen die sich bei der Modellierung ergebenden Schwierigkeiten nur andeuten.

Es gibt zwar eine breite Vielfalt stochastischer Modelle zur Beschreibung von Niederschlag, die meisten sind aber auf Datensätze aus Amerika, Afrika oder Asien angewendet worden. Die meisten Publikationen über Niederschlag in Europa beziehen sich auf Datensätze aus Großbritannien. Messungen aus Deutschland werden bisher selten betrachtet. Anzahlsmäßig dominieren die Modelle für täglich gemessenen Niederschlag über die für stündlich gemessenen, was sich aus der längeren Tradition täglich notierter Niederschläge erklärt. Die technischen Fortschritte in der Entwicklung von Regenmessern, die in stündlichem Abstand die gefallene Regenmenge notieren oder sogar die exakten Zeitpunkte des Anfangs und des Endes eines Regenschauers sowie dessen Niederschlagsmenge, und die daraus resultierenden Datensätze, geben der Erforschung von Modellen für hochaufgelösten Niederschlag neue Impulse. Die in der letzten Zeit am häufigsten verwendeten Modelle zur Beschreibung stündlichen Niederschlages sind zweifelsohne die auf Clusterprozessen basierenden. Sie sind elegante und physikalisch interpretierbare Modelle, die außerdem flexibel sind und mit denen leicht künstliche Daten erzeugt werden können. Allerdings sind ihre theoretischen Eigenschaften nur unter der Annahme von Stationarität bekannt, so daß sie im allgemeinen separat an die Beobachtungen unterschiedlicher Saisonen angepaßt werden, innerhalb derer die jahreszeitlichen Veränderungen als vernachlässigbar gering betrachtet werden.

In dieser Arbeit wird der in der Niederschlagsmodellierung am häufigsten verwendete Clusterprozeß, nämlich der Neyman–Scott–Clusterprozeß mit Rechteckzellen, zu einem Prozeß verallgemeinert, der Saisonschwankungen erklärt. Die Vorgehensweise wird an einem 27 Jahre umfassenden Datensatz hochaufgelöster Niederschlagsmessungen aus Stuttgart demonstriert. Die Modellparameter werden als stetige Funktionen der Zeit betrachtet. Das Ziel besteht darin, die saisonale Variation der ersten und zweiten Momente des Niederschlagsmengenprozesses, also insbesondere des Erwartungswertes, der Varianz und der Autokorrelationsstruktur, korrekt wiederzugeben. Dabei muß die Struktur der Momente bei einer Veränderung der Länge des Zeitintervalles, über das der Niederschlag aggregiert wird, erhalten bleiben.

In Kapitel 2 geben wir einführend einen Überblick über häufig in der Modellierung von Niederschlag verwendete Punktprozesse, der nicht den Anspruch erhebt, eine vollständige Einführung in die Theorie der stochastischen Prozesse zu sein. Dazu konsultiere man Bauer [1991] oder Cox und Isham [1979]. In Kapitel 3 wird der vorliegende Datensatz beschrieben, bevor in Kapitel 4 an die täglichen Niederschlagsmengen ein Markovkettenmodell angepaßt. In Kapitel 5 schließlich wird

das verwendete Neyman–Scott–Clustermodell definiert und an die stündlichen Niederschlagsmengen angepaßt. Im abschließenden Kapitel 6 wird die Validation des angepaßten Modelles auf der Basis künstlich erzeugter Daten durchgeführt.

Die in dieser Arbeit verwendete Notation orientiert sich an der in der gängigen Literatur üblichen. Gelegentlich wird auf mehrfache Indizes oder neue Variablenbezeichnungen verzichtet, wenn der Sinn eindeutig aus dem Kontext hervorgeht, um die Notation nicht unnötig kompliziert zu gestalten. Literatur wird durch den Namen des/der VerfasserIn und die Jahreszahl der Veröffentlichung in eckigen Klammern zitiert, so daß eine Zuordnung des Themas der Referenz erleichtert wird. Die statistische Auswertung ist mit S-PLUS (Version 3.3) durchgeführt worden. Programmiert wurde in C unter Verwendung des Borland- und des gnu-Compilers. S-PLUS ist ein umfangreiches Software-Paket, so daß beispielsweise Mittelwertberechnungen, Modellanpasungen oder Fast-Fourier-Transformationen durch einfache Befehle durchgeführt werden. Zur Vereinfachung der Datenanalyse habe ich zahlreiche S-PLUS-Funktionen erstellt, die ich nicht aufgelistet habe, da sie aus der Abfolge S-PLUS-eigener Befehle bestehen. Für einige zentrale Punkte der Arbeit wie die Modellanpassung oder die Simulation künstlicher Daten erstellte Routinen werden mit ihrem Quellcode gegeben. Der `typewriter`-Font markiert Computerbefehle.

Kapitel 2

Stochastische Modelle der zeitlichen Struktur von Niederschlag

1 Stochastische Modellierung von Niederschlag

Die meisten der nachfolgenden Modelle dienen der Beschreibung der zeitlichen Struktur von Niederschlag an einem Ort. Niederschlag ist eine teils diskrete, teils stetige Zufallsvariable. Entweder es regnet oder es regnet nicht, dieser sogenannte Ereignisprozeß wird durch eine 0–1–Zufallsvariable beschrieben. Regnet es, so ist die Niederschlagsmenge eine stetige Größe. Die Folge der Niederschlagsmengen wird als Mengenprozeß bezeichnet.
Damit treten in natürlicher Weise zwei Folgen von Zufallsvariablen auf: Die eine zählt, in welchen und wievielen der Beobachtungsintervalle es geregnet hat. Die andere verbindet mit jedem nassen Intervall eine bestimmte Niederschlagsmenge. Sind die Niederschlagsmengen in nassen Zeitintervallen unabhängig, so können beide Prozesse separat modelliert werden. Bei täglich akkumulierten Messungen ist das meistens möglich, bei einer stündlichen Zeitskala allerdings nicht.
Es bietet sich an, die erste Folge, nämlich den Ereignisprozeß, durch einen Punktprozeß zu beschreiben, weil Punktprozesse stochastische Zählmaße sind. Sie zählen die Häufigkeit zufälliger Ereignisse Z. Im folgenden werden wir Punktprozesse kurz einführen.

Sei $(E, \mathcal{B})$ ein meßbarer Raum, auf dem sich das zufällige Phänomen ereignet. Wir denken zur Darstellung temporaler Niederschläge an $\mathbb{R}$ oder räumlich-zeitlichen Niederschlages an $\mathbb{R}^3$ jeweils mit der borelschen σ-Algebra $\mathcal{B}$ der halboffenen Intervalle.

(1.1) Definition. Ein stochastischer Prozeß N ist ein **Punktprozeß** auf $(E, \mathcal{B})$ genau dann, wenn er die folgenden Eigenschaften besitzt:

1. $N(\emptyset) = 0$.

2. $N(A) < \infty$ für alle abgeschlossenen $A \in \mathcal{B}$.
3. $N\left(\cup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} N(A_i)$ für je endlich viele, disjunkte Mengen $A_1, \ldots, A_n \in \mathcal{B}$.

In der Modellierung von Niederschlag wird der Punktprozeß die Anzahl der Regenereignisse beschreiben. Ein Ereignis muß noch definiert werden. Einerseits kann der Anfangszeitpunkt t eines Regenschauers ein Punkt des Prozesses N sein. Andererseits ist dieser bei zeitdiskreter Messung nicht beobachtbar. Deshalb kann bei täglicher Messung ein nasser Tag als Punkt des Prozesses N gezählt werden, unabhängig von der tatsächlichen Anzahl der Regenschauer an diesem Tag. Betrachten wir zunächst täglichen Regen; $Y_k, k \in I\!N$, sei die an Tag k beobachtete Niederschlagsmenge. Wir definieren eine neue Folge $\{Z_k\}_{k \in I\!N}$ von Zufallsvariablen

$$\textbf{(1.2)} \quad Z_k(d) = 1(Y_k > d)\ , k \in I\!N\ , d \in I\!R_+\ ,$$

$1(\cdot)$ bezeichne die Indikatorfunktion. Ist $Z_k(d) = 1$, wird Tag k als nasser Tag bezeichnet, ist $Z_k(d) = 0$, wird von einem trockenen Tag gesprochen. Der Schwellenwert d bestimmt, ob ein Tag als naß oder trocken bezeichnet wird. Er sollte so gewählt werden, daß Inhomogenitäten der Daten verschwinden, ohne daß ein zu großer Anteil der Niederschlagsmenge vernachlässigt wird. BUISHAND [1977] verwendet zur Modellierung von Beobachtungen aus den Niederlanden einen Schwellenwert von 0.3mm Niederschlag pro Tag, um Tau oder Nebel auszuschließen. GREEN [1964] verwendet für Beobachtungen aus Tel Aviv zunächst einen Schwellenwert von 0.1mm, merkt aber an, daß die Differenz in der Anzahl der Tage, an denen mehr als 0.0mm, aber weniger als 0.1mm Niederschlag gefallen sei, vernachlässigbar gering sei, so daß er schließlich einen Schwellenwert von 0.0mm verwendet. Der Prozeß extremer Ereignisse, das sind Niederschlagsereignisse mit einer hohen Intensität, wird ebenso aus dem Naß-Trocken-Prozeß durch Ausdünnen (thinning) erhalten, also auch durch einen geeigneten Schwellenwert. ZUCCHINI UND ADAMSON [1989] betrachten beispielsweise einen Tag, an dem mehr als 30 mm Niederschlag fallen, als einen Sturm. Wir werden auch Z_k statt $Z_k(d)$ schreiben, wenn Eindeutigkeit gegeben ist.

Mit $N(t) := N((0,t])$ wird die Anzahl der Punkte im halboffenen Intervall $(0\,,t]$ bezeichnet. Damit ist für beliebige Interalle $(a,b] \in I\!R_+$

$$\textbf{(1.3)} \quad N((a,b]) = \int_a^b dN(z)\ .$$

Die zweite Bedingung in (1.1) stellt sicher, daß $N((a,b])$ stets wohldefiniert ist. Von einem ordentlichen Punktprozeß sprechen wir, wenn multiple Punkte mit Wahrscheinlichkeit null auftreten. Unter Verwendung des Landauschen Symbols $o(\cdot)$ bedeutet das:

(1.4) Definition. Ein Punktprozeß heißt **ordentlich**, wenn

$$\mathrm{P}\Big(N((t,t+\delta]) > 1\Big) = o(\delta) \quad \forall t\ , \delta \in I\!R_+\ . \qquad \square$$

Im folgenden bezeichnen wir den Punktprozeß in stetiger Zeit mit $N(t)$ und den Punktprozeß in diskreter zeit mit N_K. Die Anzahl der nassen Tage bis zum n-ten Tag erhalten wir durch einfaches Aufsummieren:

$$\textbf{(1.5)} \quad N_n := N((0,n]) = \sum_{k=1}^{n} Z_k(d)\,, k \in I\!N\,.$$

Bei der Modellierung des Ereignisprozesses sind unterschiedliche Betrachtungsweisen möglich:

1. Der beobachtete Punktprozeß repräsentiert alle Punkte des Ereignisprozesses.
2. Der beobachtete Punktprozeß repräsentiert eine gefilterte Stichprobe des tatsächlichen Ereignisprozesses, in dem mehrere Punkte innerhalb eines Intervalles auftreten können, von denen aber nur einer registriert wird.

Gemäß der ersten Interpretation erhält ein Punkt die Bedeutung eines nassen Tages oder allgemein eines nassen Intervalles. Punkte werden damit nur an Zeitpunkten notiert, die ganzzahlige Vielfache der Länge des Beobachtungsintervalles sind. Damit liegt ein Punktprozeß in diskreter Zeit vor. Wird trotzdem ein Prozeß in stetiger Zeit angepaßt, und werden die Parameter direkt aus den Beobachtungen geschätzt, so sind die Parameterschätzer nicht erwartungstreu, vergl. Abschnitt 10. Insbesondere bei der Beurteilung des Clusterverhaltens des Prozesses werden Fehler begangen. Mit den Auswirkungen der Diskretisierung haben sich FOUFOULA-GEORGIOU UND GUTTORP [1986] und FOUFOULA-GEORGIOU UND LETTENMAIER [1986] befaßt. Sie haben gezeigt, daß sogar bei Aggregation zu 15-minütigen Beobachtungen Information verloren wird.

Wird die zweite Interpretation gewählt, können die Parameter des vermuteten Punktprozesses nicht direkt aus dem beobachteten Ereignisprozeß geschätzt werden. In diesem Fall wird aus dem aggregierten Mengenprozeß auf den nicht beobachteten Ereignisprozeß in stetiger Zeit geschlossen, vergl. RODRIGUEZ-ITURBE, WAYMIRE UND GUPTA [1984]. Wird den Modellparametern physikalische Bedeutung beigemessen, sollte das resultierende Modell des Ereignisprozesses mit unterschiedlichen Zeiteinteilungen kompatibel sein.

Über 24 Stunden aggregierte Beobachtungen werden als täglicher Regen bezeichnet. Die Wahl des Zeitpunktes, an dem ein neuer Tag in diesem Sinne beginnt, ist nicht unerheblich: Regnet es in einer Gegend gewöhnlich am frühen Morgen und wird die Niederschlagsmenge der letzten 24 Stunden erst am Abend abgelesen, kann ein Teil des Niederschlages schon verdunstet sein. Liegt hingegen der Ablesezeitpunkt vor dem Einsetzen der größten Hitze, aber nachdem es geregnet hat, so wird nur wenig verdunstet sein und die beobachteten Mengen werden regelmäßig größer als im ersten Fall sein. Im folgenden werden zunächst täglich, also diskret gemessene Daten betrachtet, Clusterprozesse und ARMA-Modelle werden auch auf stündliche Daten angewendet. Die Niederschlagsmenge an Tag k wird mit Y_k bezeichnet, die 0-1-Variable $Z_k(d)$ beschreibt, ob an Tag k mindestens einmal geregnet hat und dabei

mindestens $d$mm Niederschlag registriert worden sind Der zugehörige Punktprozeß ist $N_n = \sum_{k=1}^{n} Z_k$. Die Wartezeit ab einem Zeitpunkt $t = 0$ oder $k = 0$ auf den n-ten Punkt sei T_n, die Verweildauer U_n zwischen dem $(n-1)$-ten und dem n-ten Punkt ist $U_n = T_n - T_{n-1}$, $U_1 = T_1$ beschreibt die Wartezeit auf den ersten Punkt.

2 Poissonprozesse

In einem stationären Prozeß wird man sinnvollerweise zunächst untersuchen, ob die Inkremente $N_n - N_{n-1}$ des Ereignisprozesses unabhängig sind. Sind die Zufallsvariablen Z_k unabhängig und identisch verteilt mit $\mathrm{P}(Z_k = 1) = p$ und $\mathrm{P}(Z_k = 0) = q = 1 - p$, so ist N_n binomialverteilt mit Parametern n und p. Die Wahrscheinlichkeitsfunktion lautet also

$$\textbf{(2.1)} \qquad \mathrm{P}(N_n = k) = \binom{n}{k} p^k q^{n-k} \ , 0 \leq k \leq n \ .$$

Der Binomialprozeß ist in der Modellierung von Regenfall nur wenig verwendet wurden (außer von Foufoula-Georgiou und Lettenmaier [1986]), sein Gegenstück in stetiger Zeit, der Poissonprozeß, dafür um so häufiger. Zur Beschreibung des Ereignisprozesses täglichen Regens werden hauptsächlich Prozesse in stetiger Zeit, also mit stetigem Parameterraum, angepaßt, vermutlich wegen der größeren Auswahl an Prozessen mit weit entwickelter Theorie. Der Übergang von einem diskreten zu einem stetigen Parameterraum erfolgt in kanonischer Weise durch Einbettung von $Z \subset \mathbb{R}$. Liegen über einen bestimmten Zeitraum aggregierte Daten vor, so entstehen keine Schwierigkeiten: Ein Ereignis an Tag k ist weiterhin ein Punkt $Z_k = 1$. Allgemein aber ist Regen kein nur einen Moment dauerndes Ereignis; stetige Messungen verursachen deshalb überabzählbar viele Punkte in einem abgeschlossenen Intervall, so daß nicht mehr von einem Punktprozeß gesprochen werden kann. Diese Schwierigkeit wird häufig umgangen, indem den Regenereignissen eine insignifikante Länge zugewiesen wird. Es ist auch möglich, den Anfangszeitpunkt des Regens als Punkt des Zählprozesses zu betrachten. Um keine Information zu verlieren, sollte in diesem Fall mit jedem Punkt ein Zeiger assoziiert werden, der die Dauer des Regenschauers beschreibt.

Der Übergang von einem diskreten Parameterraum zu einem stetigen Parameterraum führt bekanntlich vom Binomialprozeß zum Poissonprozeß. Der Poissonprozeß ist durch zufällige Ereignisse charakterisiert, die sich unabhängig voneinander mit einer mittleren Rate λ ereignen. Wir definieren gemäß Bauer [1991]:

(2.2) Definition. Ein Punktprozeß N ist ein homogener **Poissonprozeß** zum Parameter $\lambda > 0$ genau dann, wenn gilt:

(a) Der Prozeß besitzt stationäre und unabhängige Zuwächse $(N(t) - N(s))$, wobei für $0 \leq s \leq t,\ s, t \in \mathbb{R}_+$ gilt:

$$(N(t) - N(s)) \sim \mathrm{Po}(\lambda\,(t-s))\ .$$

(b) Fast alle Pfade $\omega \mapsto N(t)(\omega)$ sind rechtsseitig stetige, isotone Funktionen mit Sprüngen der Größe 1.

(2.3) Definition. Ein Punktprozeß $\{N_t\}_{t\in I\!R_+}$ wird als homogen bezeichnet, wenn die Parameter des Prozesses konstant sind.

(2.4) Bemerkung. Ein homogener Poissonprozeß N mit Rate λ besitzt die folgenden grundlegenden Eigenschaften:

(a) Die Erwartungswert-Zeit-Funktion $\mathrm{E}N(t)$ erfüllt $\mathrm{E}N(t) = \lambda t$.

(b) Der Dispersionsindex $I(t) = \mathrm{Var}N(t)/\,\mathrm{E}N(t)$ erfüllt: $I(t) = 1$.

(c) Das Nullwahrscheinlichkeitsfunktional $\mathrm{P}(N(A) = 0)$ bzw. $\mathrm{P}(N(t) = 0)$ für beliebige Intervalle A bzw. für $(0, t]$ erfüllt $\mathrm{P}(N(A) = 0) = \exp(-\lambda|A|)$ bzw. $\mathrm{P}(N(t) = 0) = \exp(-\lambda t)$. Dabei bezeichnet $|A|$ das Lebesgue-Maß von A.

(d) Die Verweildauern U_n sind unabhängig und identisch $\exp(\lambda)$-verteilt mit Erwartungswert $\mathrm{E}U_n = 1/\lambda$, $n \in I\!N$. Insbesondere ist die Wartezeit auf den ersten Punkt genauso verteilt wie die Wartezeit zwischen zwei aufeinanderfolgenden Punkten.

(e) Der Variationskoeffizient der Verweildauern $CV(U_n) = \sqrt{\mathrm{Var}U_n}/\,\mathrm{E}U_n$ erfüllt $CV(U_n) = 1$, $n \in I\!N$.

Eine Erweiterung des Poissonprozesses erhalten wir, wenn wir die Ereignisrate $\lambda = \lambda(s)$ als zeitabhängige, deterministische Funktion betrachten:

$$\mathrm{P}(N(t) = n) = e^{-\Lambda(t)}\Lambda(t)^n/n!\ ,\ n \in I\!N_0$$

mit

$$\Lambda(t) = \int_0^t \lambda(s)ds\ .$$

Dieser Prozeß wird als nichthomogener Poissonprozeß bezeichnet. Er ist ein Spezialfall eines Punktprozesses, dessen Ereignisrate (Intensitätsrate) $\lambda(t)$ nicht nur eine deterministische Funktion, sondern ein stochastischer Prozeß ist. Letztere Prozesse führte Cox 1955 unter der Bezeichnung doppelt stochastische Poissonprozesse ein, heute werden sie jedoch meist nach ihrem Erfinder als Cox-Prozesse bezeichnet. Sei $\lambda(t)$ ein nichtnegativer stochastischer Prozeß, $t \geq 0$. Sei $\Lambda(A) = \int_A \lambda(u)\,du$ für meßbare Mengen A sowie $\Lambda(t) = \int_0^t \lambda(u)\,du$, du bezeichne das Lebesgue-Maß auf $I\!R$.

(2.5) Definition. Ein Punktprozeß $N(t)$ ist ein **Cox-Prozeß** mit Intensitätsprozeß $\lambda(t) > 0$, wenn gilt:

(a) Für je endlichviele, disjunkte meßbare Mengen $A_1, \ldots, A_n$ sind $N(A_1), \ldots,$ $N(A_n)$ bedingt unabhängig bei gegebenem Λ.

(b) Die bedingte Verteilung von $N(A)$, gegeben Λ, ist für alle meßbaren Mengen A poissonverteilt mit Parameter $\Lambda(A)$: $\forall\, 0 \leq k \in \boldsymbol{Z}$

$$\mathrm{P}(N(A) = k|\Lambda) = \exp(-\Lambda(A))\,\Lambda(A)^k/k!\,.$$

Der Intensitätsprozeß kann als stochastischer klimatologischer Prozeß interpretiert werden, der die Neigung zu Regen beschreibt (SMITH [1987]).

Bei einem Poissonprozeß mit Parameter λ ist die Folge der Wartezeiten zwischen zwei Ereignissen unabhängig identisch exponentialverteilt mit Erwartungswert $1/\lambda$. Wird nur angenommen, die Folge der Wartezeiten sei unabhängig und identisch verteilt, erhalten wir einen Erneuerungsprozeß und damit eine Verallgemeinerung des Poissonprozesses.

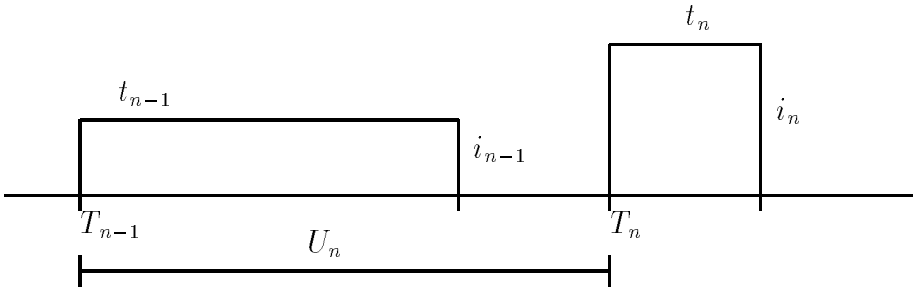
Wird der Ereignisprozeß täglichen Regens als Poissonprozeß modelliert, wird den Punkten eine insignifikante Länge zugewiesen. Tatsächlich beträgt die durchschnittliche Dauer einer nassen Phase in dem dieser Arbeit zugrundeliegenden Datensatz aber etwa 100 Minuten, so daß diese Annahme nicht sinnvoll ist.

Werden Niederschläge in stetiger Zeit aufgezeichnet und die Anfangszeitpunkte eines Regenschauers als Poissonprozeß modelliert, so kann leicht gleichzeitig die Dauer der Regenereignisse beschrieben werden. Wir gelangen zu **Poisson-Rechteckimpulsmodellen**. Sie sind markierte Punktprozesse, deren Punkte gemäß einem Poissonprozeß mit Rate λ auftreten. Ein Punkt zur Zeit T_n repräsentiert die Anfangszeit des n-ten Regenereignisses. Mit ihm wird ein Rechteckimpuls (t_n, i_n) verbunden, t_n mißt die Dauer und i_n die Intensität des Ereignisses. Intensität bedeutet Regen pro Zeiteinheit und wird in [mm/Tag] oder [mm/Stunde] gemessen (vergl. RODRIGUEZ-ITURBE et al. [1984], BACCHI et al. [1994]). Das einfachste Modell nimmt an, alle (t_n, i_n) seien unabhängig identisch verteilte Zufallsvariablen und t_n und i_n seien ebenfalls unabhängig. Als Marginalverteilungen der t_n und i_n werden häufig aus Gründen der mathematischen Einfachheit Exponentialverteilungen gewählt. Regenereignisse mit hoher Intensität werden so nicht angemessen beschrieben, da bei ihnen Intensität und Dauer oft negativ korreliert sind. BACCHI et al. [1994] verwenden zur Beschreibung von Beobachtungen aus Norditalien (Valtellina, Milano, Firenze) eine bivariate Exponentialverteilung, bei der t_n und i_n negativ korreliert sind. Die geschätzte Korrelation beträgt $r = -0.4$.
Insbesondere bei extremen Ereignissen, die durch eine hohe Gesamtniederschlagsmenge definiert werden, scheint die Annahme der Unabhängigkeit der Ereignisse voneinander gerechtfertigt zu sein, vergl. BACCHI. et al. [1994].

3 Markovketten

Die Annahme unabhängiger Inkremente des Zählprozesses bei täglichem Niederschlag ist zur Beschreibung von Niederschlagsprozessen meistens unzutreffend. Da-

Skizze zum Poisson-Rechteckimpulsmodell



T_n: Startzeit des n-ten Ereignisses; $U_n := T_n - T_{n-1}$ Wartezeit

---

mit werden wir zu Prozessen geführt, in denen die Inkremente stochastisch abhängig sind. Die beliebteste Modifikation ist die Annahme, die Folge der Z_k bilde einen Markovprozeß mit diskretem Parameterraum, also eine Markovkette, und zwei möglichen Zuständen 0 und 1. Wir betrachten ausschließlich Markovketten mit zwei Zuständen.

In einer Markovkette hängt die Wahrscheinlichkeit, an einem Tag k Regen zu beobachten, nur vom Zustand des vorhergehenden Tages ab:

$$\mathrm{P}(Z_k = i | Z_{k-1}, Z_{k-2}, \ldots) = \mathrm{P}(Z_k = i | Z_{k-1}) \, .$$

Die Struktur einer Markovkette ist eindeutig durch die Matrix Π der Übergangswahrscheinlichkeiten bestimmt,

$$\Pi := \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix}$$

mit $\pi_{ij} := \mathrm{P}(Z_k = j | Z_{k-1} = i)$. Wir müssen π_{01} und π_{11} schätzen. Offenbar gilt $\pi_{00} + \pi_{01} = 1$ und $\pi_{10} + \pi_{11} = 1$, da mit Sicherheit einer der beiden Zustände 0 oder 1 auf einen vorhergehenden Zustand folgt.

Diese Darstellung impliziert, daß die Übergangswahrscheinlichkeiten zeitinvariant sind, also ein homogener Prozeß im Sinne von (2.3) vorliegt. Sie impliziert außerdem Abhängigkeit nur vom letzten vorhergehenden Zeitpunkt. Eine Verallgemeinerung ist leicht möglich: Mit $\Pi = \Pi(k)$, $\pi_{ij}(k) = \mathrm{P}(Z_k = j | Z_{k-1} = i), k = 1, \ldots, K$ können saisonale Schwankungen berücksichtigt werden. Es ist$k = 1, \ldots, K$, wobei $K = 365$, wenn ein gesamtes Jahr beobachtet worden ist. Eine Markovkette höherer Ordnung $m \geq 2$ wird durch

$$\pi_{i_m \ldots i_1 i_0}(k) := \mathrm{P}(Z_k = i_0 | Z_{k-1} = i_1, \ldots, Z_{k-m} = i_m)$$

beschrieben. Ist die Markovkette ein Punktprozeß, wie in unserem Fall täglichen Regens, so gilt $i_0, i_1, \ldots i_m \in \{0, 1\}$. Die Wahrscheinlichkeiten $\pi_{i_m \ldots i_1 i_0}(k)$ müssen offensichtliche Eigenschaften erfüllen wie z.B.

$$\pi_{i_m \ldots i_1 0}(k) + \pi_{i_m \ldots i_1 1}(k) = 1 \quad \text{für alle} \quad k = 1, \ldots, K, \;\; i_j, \in \{0, 1\} \, .$$

Die empirischen Parameterschätzer einer Markovkette 1. Ordnung sind hier

$$\hat{\pi}_{ij}(k) := \frac{\#\{\text{Beobachtete Tage } k \text{ mit } Z_{k-1} = i\ , Z_k = j\}}{\#\{\text{Beobachtete Tage } (k-1) \text{ mit } Z_{k-1} = i\}}\ ,$$

sofern der Nenner ungleich null ist, sonst ist dieser Schätzer nicht definiert. Damit ist die Anzahl der Ereignisse $(Z_{k-1} = i,\ Z_k = j)$ in N beobachteten Jahren eine binomialverteilte Zufallsvariable mit Erfolgswahrscheinlichkeit $\pi_{ij}(k)$. Die Anzahl der Versuche ist die Anzahl der Ereignisse $(Z_{k-1} = i)$. Ohne zusätzliche Restriktionen ist die Schätzung der Übergangswahrscheinlichkeiten sinnlos, da zu jedem Datum zwei Parameter geschätzt werden müssen. Hier muß der Raum der Abbildungen

$$\begin{aligned} \pi_{01} &: \{1, 2, \ldots, K\} \longrightarrow [0, 1]\ , \quad k \longmapsto \pi_{01}(k)\ , \\ \pi_{11} &: \{1, 2, \ldots, K\} \longrightarrow [0, 1]\ , \quad k \longmapsto \pi_{01}(k). \end{aligned}$$

eingeschränkt werden. Dazu bieten sich vielfältige Möglichkeiten, wie beispielsweise Glättung unter apriori bekannter Eigenschaften. Physikalisch plausibel sind nur Parameter, die sich im Zeitverlauf stetig und zyklisch ändern mit einer Periode von einem Jahr. Daher ist eine Beschreibung durch eine Fouriersumme naheliegend. Wegen der Restriktion $0 < \pi_{ij}(k) < 1$ bietet es sich an, nicht $\pi_{ij}(k), k = 1, \ldots, K$, sondern $logit(\pi_{ij}(k))$ durch eine Fouriersumme mit $L < T$ Frequenzen zu modellieren, wobei

$$logit(\pi_{ij}(k)) = \log\left(\frac{\pi_{ij}(k)}{1 - \pi_{ij}(k)}\right)\ .$$

Weil die Binomialverteilung zur Exponentialfamilie gehört, stehen bei diesem Vorgehen die Schätzverfahren der verallgemeinerten linearen Modelle (vergl. McCullagh und Nelder [1989]) zur Verfügung. Die Anzahl K der anzupassenden Fourierfrequenzen wird mit einem Modellauswahlkriterium bestimmt. Diese Vorgehensweise ist in Kapitel 4.1 ausführlich beschrieben. Zucchini und Adamson [1984] haben täglichen Regen in Südafrika auf diese Art modelliert, Stern und Coe [1984] täglichen Regen in Jordanien und Tansania. Katz und Parlange [1995] modellieren auch stündlichen Regen mit einer Markovkette 1. Ordnung. Mit der Auswahl der Ordnung der Markovkette haben sich u.a. Tong [1975] und Katz [1981] befaßt.

Eine Variante des Markovkettenmodelles haben Roldan und Woolhiser [1982] im Rahmen alternierender Erneuerungsprozesse (vergl. Abschnitt 5) betrachtet. Eine weitere Möglichkeit, markovsche Abhängigkeitsstrukturen zu integrieren, besteht darin, die Folge der nassen und trockenen Tage als Bernoulli-Prozeß mit randomisierter Erfolgswahrscheinlichkeit darzustellen, wobei die Folge der Erfolgswahrscheinlichkeiten eine Markovkette bildet. Ein solcher Prozeß ist ein Hidden-Markov-Modell.

4 Hidden–Markov–Modelle

Das Analogon eines Cox-Prozesses in diskreter Zeit ist ein Bernoulli–Prozeß mit randomisierten Erfolgswahrscheinlichkeiten. In einem Hidden–Markov–Modell folgt ein nicht beobachtbarer Prozeß einer Markovkette mit endlichem Zustandsraum, während die Verteilung der Beobachtungen an einem beliebigem Zeitpunkt nur vom Zustand der Markovkette an diesem Zeitpunkt abhängt. Solche Modelle werden als Hidden–Markov–Modell bezeichnet. ZUCCHINI UND GUTTORP [1991] haben sie zur Beschreibung der zeitlich–räumlichen Verteilung von täglichem Niederschlag in den USA verwendet. In diesem Modell folgen die Zustände eines nicht beobachtbaren klimatologischen Prozesses einer Markovkette. Sie bestimmen die Wahrscheinlichkeit, an einem Tag Niederschlag zu erhalten. Die räumlich–zeitliche Verteilung des Niederschlages ist bedingt unabhängig bei gegebenem klimatologischem Prozeß.

Der klimatologische Prozeß $\{C_k\}_{k\in\mathbb{N}}$ sei eine stationäre Markovkette mit M möglichen Zuständen und Übergangswahrscheinlichkeiten

$$\gamma_{ij} = \mathrm{P}(\, C_k = j \,|\, C_{k-1} = i \,) \,,\ i,j = 1,\ldots,M.$$

C_k ist der Zustand des klimatologischen Prozesses an Tag k, der aber nicht beobachtet werden kann. Der Ereignisprozeß sei $\{Z_k^{(n)}\}_{k,n\in\mathbb{N}}$, $k = 1,\ldots,K$, $n = 1,\ldots,N$, wobei

$$Z_k^{(n)} = \begin{cases} 1, & \text{es regnet an Tag } k \text{ an Station } n, \\ 0 & \text{sonst.} \end{cases}$$

Die Wahrscheinlichkeit, an einem Tag Regen zu erhalten, ist damit für alle Stationen gleich bei gegebenem C_k. Um die Wahrscheinlichkeiten der möglichen Werte $Z_k := (Z_k^{(1)},\ldots,Z_k^{(N)})$ auf kompakte Weise zu beschreiben, verwenden Zucchini und Guttorp die Dezimaldarstellung $\tilde{Z}_k$ der binären Zahl Z_k, welche die Werte $0,1,2,\ldots,L = 2^N - 1$ annehmen kann. Die bedingte Wahrscheinlichkeiten

$$\mathrm{P}(\tilde{Z}_k = l | C_k = m) = \pi_{lm}$$

können jetzt als $((L+1)\times M)$-Matrix dargestellt werden.

Zucchini und Guttorp leiten Eigenschaften des so definierten Hidden–Markov–Prozesses her, unter anderem die Likelihood-Funktion von $\tilde{Z}_k$, so daß die Modellparameter mit der Maximum-Likelihood-Methode geschätzt werden können, was ein Vorzug gegenüber vielen anderen Modellen ist. Hidden–Markov–Modelle und ihre Eigenschaften werden ausführlich in MACDONALD UND ZUCCHINI [1997] beschrieben.

Das Hidden–Markov–Modell kann physikalisch interpretiert werden, indem die Zustände von C_k als „Tendenz zu Regen" betrachtet werden. Es schließt als Spezialfälle Markovkettenprozesse und Bernoulliprozesse ein. Wird nur eine Station betrachtet

und ist die Wahrscheinlichkeit von Regen für alle Zustände des klimatologischen Prozesses gleich, so liegt ein Bernoulliprozeß vor. Kann der klimatologische Prozeß zwei Zustände annehmen, von denen der eine mit einem trockenen Tag, der andere mit einem nassen Tag assoziiert wird, liegt eine Markovkette vor. Letzteres bedeutet $\mathrm{P}(Z_k = 1|C_k = 1) = 1$, $\mathrm{P}(Z_k = 0|C_k = 0) = 1$. Eine Anpassung eines Hidden-Markov-Modelles an den in dieser Arbeit analysierten Datensatz führte zu einem Hidden-Markov-Modell, das einer Markovkette sehr ähnlich war. Der zusätzliche kompliziertere theoretische Modellaufbau sowie die aufwendigere Modellanpassung waren für diesen Datensatz also nicht gerechtfertigt, weshalb die Ergebnisse der Modellanpassung nicht dargestellt werden.

Ein Nachteil des räumlich-zeitlichen Modelles für $n > 1$ Stationen, auf den Zucchini und Guttorp [1991] hinweisen, besteht in der geforderten Unabhängigkeit der Stationen. Für weit entfernte Stationen sei es einerseits nicht plausibel, eine einheitliche „Großwetterlage" anzunehmen. Andererseits sei es unplausibel, für nahe gelegene Stationen Unabhängigkeit anzunehmen. Die Interpretation von $\{C_k\}_{k \in I\!N}$ als klimatologischen Prozeß ist daher fragwürdig. Zucchini und Guttorp [1991] geben Ausblicke auf mögliche Erweiterungen des Modelles, die Kovariablen wie den Abstand einer Station zu anderen Stationen berücksichtigen können.

Smith [1987] entwickelt ein Modell zur Beschreibung täglichen Regenfalles, das auf einer nicht beobachteten Markovkette mit zwei Zuständen basiert, von denen der eine trockenes Wetter impliziert. Dieses Modell berücksichtigt saisonale Schwankungen vermöge zeitabhängiger Parameter. Smith bezeichnet es als **Markov-Bernoulli-Modell**. Im Markov–Bernoulli–Modell ist eine (nicht-homogene) Markovkette C_k, $k \in I\!N$, mit Zustandsraum $\{0, 1\}$ gegeben, die einen nicht beobachtbaren, klimatologischen Prozeß beschreibt. Die Punkte $Z_k, k \in I\!N$, des Prozesses sind Bernoulli–Variablen mit randomisierter Erfolgswahrscheinlichkeit X_k, wobei

$$X_k := p(k)C_k \ , \ k \in I\!N \ ,$$

mit einer Abbildung $p : I\!N \longrightarrow [0, 1]$. Die Erfolgswahrscheinlichkeiten sind also 0 oder $p(k)$. Das führt zur folgenden Definition:

(4.1) Definition. Der Punktprozeß $\{Z_k\}_{k \in I\!N}$ ist ein **Markov-Bernoulli-Modell** mit Parametern p, q_0 und q_1, wenn für alle $n \in I\!N$ und alle $(z_1, \ldots, z_n)$ mit $z_i \in \{0, 1\}$ gilt:

$$\begin{aligned} &\mathrm{P}(Z_K = z_k, \ldots, Z_1 = z_1 |\ X_k, k = 1, \ldots, K) \\ &= \prod_{k=1}^{K} \mathrm{P}(Z_k = z_k |\ X_k) \\ &= \prod_{k=1}^{K} [\ z_k X_k + (1 - z_k)(1 - X_k)\ ] \ . \end{aligned}$$

Der Wert von Z_k hängt also nur von X_k ab.

Sind die Parameter konstant im Zeitverlauf, ist dieses Modell der Spezialfall des Hidden-Markov-Modelles für eine Station, in dem $C_k \in \{0,1\}$ für alle k und $\mathrm{P}(Z_k = 1 | C_k = 0) = 0$. In der Notation aus ZUCCHINI UND GUTTORP [1991] ist

$$\Pi = \begin{pmatrix} 1 & 0 \\ p(k) & 1-p(k) \end{pmatrix} .$$

Die Parameter des Markov-Bernoulli-Modelles können ebenfalls mit der Maximum-Likelihood-Methode geschätzt werden (SMITH [1987]). Eine wichtige Eigenschaft der Klasse der Markov-Bernoulli-Prozesse ist, daß sie abgeschlossen unter zufälligem Ausdünnen ist.

(4.2) Definition. Eine **p-Ausdünnung** eines Punktprozesses $\{Z_k\}$ ist ein Prozeß

$$\tilde{Z}_k := X_k \cdot Z_k,$$

wobei $\{X_k\}_{k \in \mathbb{N}}$ ein Bernoulliprozeß mit Erfolgswahrscheinlichkeit p ist. Die Punkte von Z_k werden also mit Wahrscheinlichkeit p in den Prozeß $\tilde{Z}_k$ übernommen und mit Wahrscheinlichkeit $1-p$ ausgelassen. Die Einführung eines Mindestniederschlagswertes, ab dessen Überschreitung ein Tag als naß betrachtet wird oder durch die ein Sturm definiert wird, ist eine p-Ausdünnung.

SMITH [1987] entwickelt aufbauend auf dem Markov-Bernoulli-Prozeß auch einen Clusterprozeß mit diskretem Parameterraum, vergl. Abschnitt 7.

5 Alternierende Erneuerungsprozesse

Erneuerungsprozesse legen das Augenmerk auf die Wartezeit zwischen zwei Ereignissen. Sind diese Wartezeiten unabhängig und identisch verteilte Zufallsvariablen, liegt ein Erneuerungsprozeß vor.

Eine Verallgemeinerung erhalten wir, wenn in einem stationären Prozeß zwei Arten von Ereignissen betrachtet werden, die abwechselnd aufeinanderfolgen. Die Wartezeit von einem Ereignis bis zu einem Ereignis vom Typ 1 sei eine Zufallsvariable U, die bis zu einem Ereignis vom Typ 2 sei eine Zufallsvariable $\tilde{U}$. Sind die Folgen $(U_n)_{n \in \mathbb{N}}$ und $(\tilde{U}_n)_{n \in \mathbb{N}}$ unabhängig voneinander und sind die Elemente beider Folgen unabhängig und identisch gemäß F bzw. $\tilde{F}$ verteilt, so sprechen wir von einem alternierenden Erneuerungsprozeß.

Bezogen auf die Modellierung täglichen Regens denken wir uns ein Ereignis vom Typ 1 als den Beginn eines nassen Intervalles und ein Ereignis vom Typ 2 als den Beginn eines trockenen Intervalles. Bei täglichem Regen wird ein trockenes Intervall der Länge n definiert als n aufeinanderfolgende trockene Tage, die an jeder Seite durch mindestens einen nassen Tag begrenzt werden. Ein nasses Intervall ist analog definiert.

Alternierende Erneuerungsprozesse sind von GREEN [1964], BUISHAND [1977] und von ROLDAN UND WOOLHISER [1982] zur Beschreibung der Länge trockener und nasser Intervalle verwendet worden. Die Arbeit Buishands ist die erste, in der Erneuerungsprozesse erfolgreich an Beobachtungen aus klimatisch so unterschiedlichen Gegenden wie den Niederlanden, Ägypten, Indien und dem Sudan angepaßt worden sind. Seine Arbeit ist zudem die ausführlichste der drei genannten Arbeiten.

Die Länge der Intervalle modelliert Buishand mittels einer negativen Binomialverteilung, die eine Verallgemeinerung der geometrischen Verteilung ist. Geometrisch verteilte Verweildauern treten beispielsweise in stationären Markovketten mit zwei möglichen Zuständen $\{0,1\}$ auf. Bezeichnet $\pi_{ij} = \mathrm{P}(Z_k = j|Z_{k-1} = i)$ die Übergangswahrscheinlichkeit von i zu j, so ist die Länge L_0 eines trockenen Intervalles bedingt geometrisch verteilt mit Parameter π_{00} bei gegebenem erstem trockenem Tag $Z_1 = 0$:

$$\begin{aligned}
&\mathrm{P}(L_0 = n) \\
&= \mathrm{P}(Z_{k+1} = 1, Z_k = 0, \ldots, Z_1 = 0) \\
&= \mathrm{P}(Z_{k+1} = 1|Z_k = 0, \ldots, Z_1 = 0) \cdot \mathrm{P}(Z_k = 0, \ldots, Z_1 = 0) \\
&= \mathrm{P}(Z_{k+1} = 1|Z_k = 0) \cdot \mathrm{P}(Z_k = 0, \ldots, Z_1 = 0) \\
&= \cdots \\
&= \mathrm{P}(Z_{k+1} = 1|Z_k = 0) \cdot \mathrm{P}(Z_k = 0|Z_{k-1} = 0) \cdot \ldots \cdot \mathrm{P}(Z_2 = 0|Z_1 = 0) \\
&= \pi_{00}^{k-1}\pi_{01} \\
&= \pi_{00}^{k-1}(1 - \pi_{00})
\end{aligned}$$

für $n \in I\!N$. Analog ist die Länge L_1 eines nassen Intervalles geometrisch verteilt mit Parameter π_{11}.

Da negativ binomialverteilte Zufallsvariable den Wert 0 annehmen können, Intervallängen L aber stets strikt positiv sind, muß die negative Binomialverteilung in dieser Anwendung verschoben oder abgeschnitten werden. Mit $\tilde{L}$ werde die neue Zufallsvariable bezeichnet. Es ergeben sich folgende Verteilungen:

1) Die verschobene negative Binomialverteilung:

$$\mathrm{P}(L_0 = n) = \mathrm{P}(\tilde{L}_0 = n + 1) = \binom{n + r}{n + 1}\pi_{00}^{r-1}(1 - \pi_{00})^{n+1} ,$$

wobei $\tilde{L}_0$ negativ binomialverteilt ist und $n > 0$.

2) Die abgeschnittene (truncated) negative Binomialverteilung

$$\mathrm{P}(L_0 = n) = \mathrm{P}(\tilde{L}_0 = n|\tilde{L}_0 \geq 1) = \binom{n + r - 1}{n}\frac{\pi_{00}^r(1 - \pi_{00})^n}{1 - \pi_{00}^r} , \ r > 1 .$$

3) Mit $r = 1$ erhalten wir die abgeschnittene geometrische Verteilung:

$$\mathrm{P}(L_0 = n) = \pi_{00}(1 - \pi_{00})^{n-1} \ , n \in I\!N \, .$$

ROLDAN UND WOOLHISER [1982] zeigen, daß im Fall der ihnen vorliegenden Daten aus unterschiedlichen klimatischen Gegenden der USA die Länge der nassen Intervalle jeweils besser durch eine abgeschnittene geometrische Verteilung modelliert wird, die Länge der trockenen Intervalle hingegen durch eine abgeschnittene negative Binomialverteilung.
Zur Berücksichtigung saisonaler Schwankungen im Jahresverlauf kann der Niederschlagsprozeß als stückweise stationär betrachtet werden, so daß die Parameter für jeden Abschnitt separat geschätzt werden. ACREMAN [1990] entwickelte ein auf einem Erneuerungsprozeß basierendes Modell zur Beschreibung stündlichen Regenfalles, das in Abschnitt 9 beschrieben wird.

6 Erneuerungs–Cox–Markov–Prozesse

SMITH UND KARR [1983] haben auf eine weitere Weise markovsche Abhängigkeitsstrukturen in den Ereignisprozeß integriert. Zur Beschreibung eines Datensatzes täglicher Niederschlagsmessungen der Sommermonate im Potomacdelta, USA, entwickeln sie einen Punktprozeß, der gleichzeitig sowohl ein Erneuerungsprozeß als auch ein Cox-Prozeß mit markovscher Intensitätsrate ist. Diesen Prozeß bezeichnen sie als Erneuerungs–Cox–Markov–Prozeß.

KINGMAN [1964] (zitiert in SMITH UND KARR [1983]) zeigt, daß die Klasse der Prozesse, die gleichzeitig Erneuerungsprozesse und Cox-Prozesse sind, echt größer als die Klasse der Poissonprozesse ist.

(6.1) Definition. Ein **Erneuerungs-Cox-Prozeß** besitzt den Intensitätsprozeß $\lambda(t), \quad t \in I\!R$, wobei $\lambda(t) = \lambda > 0$ und $\lambda(t) = 0$ abwechselnd auf Intervallen ist, deren Längen unabhängige Zufallsvariablen sind. Die Länge der Intervalle, auf denen $\lambda(t) = \lambda$ ist, ist exponentialverteilt und die Länge der Intervalle, auf denen $\lambda(t) = 0$ ist, besitzt eine beliebige Verteilung. □

Wird die Verteilung der Intervalle mit $\lambda(t) = 0$ ebenfalls durch eine Exponentialverteilung mit beliebigem, positivem Parameter beschrieben, so ist $\lambda(t)$ ein Markovprozeß:

(6.2) Definition. Ein Cox-Prozeß N, dessen Intensitätsprozeß ein Markovprozeß mit Zuständen $\lambda > 0$ und 0 ist, heißt ein **Erneuerungs-Cox-Markov-Prozeß** . □

7 Clustermodelle

Die stochastische Modellierung von Regenfall durch Clusterprozesse befindet sich am ehesten unter allen Modellen im Einklang mit meteorologischen Erkenntnissen.

Man geht davon aus, daß Regenschauer häufig auftreten, nachdem eine Regenfront einen Ort passiert hat. Damit liegt die Beschreibung durch ein hierarchisches Modell nahe: Die Regenfronten sind nicht beobachtbare Primärereignisse, die Regenschauer als Sekundärereignisse nach sich ziehen. LE CAM ([1961], zitiert nach KAVVAS UND DELLEUR [1981]) hat die räumliche Verteilung von Regenfall mit einem dreistufigen Modell beschrieben: Die Regenschauer treten an einem Ort im Zeitverlauf in Clustern als Folge einer Regenfront auf, die Regenfronten bilden wiederum Cluster, die als Sturm bezeichnet werden. Sind die Regenschauer innerhalb einer Front unabhängig und die Fronten innerhalb eines Sturmes, so liegt ein räumliches Neyman-Scott-Clustermodell vor. Zur Beschreibung der zeitlichen Verteilung von Regenfall genügt meistens ein zweistufiges Modell. Da die Regenfronten nicht beobachtet werden, ist es plausibel, Primär- und Sekundärereignisse nicht zusammenfallen zu lassen. Ohne diese letzte Annahme ist ein anderes Clustermodell plausibel: Die Wartezeit bis zu dem ersten Regenschauer innerhalb eines Clusters besitzt eine andere Verteilung als die Wartezeiten zwischen den folgenden Elementen des Clusters. In diesem Modell fällt ein Primärereignis stets mit einem Sekundärereignis zusammen; das Augenmerk liegt auf den Wartezeiten. Diese Sichtweise wird in dem Bartlett-Lewis-Clusterprozeß eingenommen.

In einem Punktprozeß wird von Clusterverhalten gesprochen, wenn die Wahrscheinlichkeit, einen Punkt in der Nähe eines anderen Punktes anzutreffen, größer als an beliebiger Stelle ist. Die Wartezeiten zwischen Punkten sind damit insbesondere nicht identisch verteilt.

Clusterprozesse bestehen aus unabhängigen Primärereignissen, sogenannten Clusterzentren, von denen jedes eine zufällige Anzahl an Sekundärereignissen, sogenannten Zellen, nach sich zieht. Cluster überschneiden sich, wenn nach einem neuen Primärereignis Sekundärereignisse eintreffen, die einem früheren Primärereignis zugehören. Diese Eigenschaft, auch als Gedächtnis von Clustern bezeichnet, führt dazu, daß sich die beobachteten Zellen nicht eindeutig einem Clusterzentrum zuordnen lassen. Neyman und Scott konzipierten Clusterprozesse zur Beschreibung der räumlichen Verteilung von Galaxien (NEYMAN UND SCOTT [1958]; zitert nach KAVVAS UND DELLEUR [1981]). Die ersten zur Modellierung von Regenfall verwendeten Neyman-Scott-Clustermodelle sind sogenannte Neyman-Scott-White-Noise-Modelle gewesen, in denen einer Zelle die Dauer null zugewiesen wird (KAVVAS UND DELLEUR [1975], zitiert nach WAYMIRE UND GUPTA [1981c]). FOUFOULA-GEORGIOU UND GUTTORP [1987] haben Widersprüchlichkeiten in diesem Modell entdeckt, was vermutlich RODRIGUEZ-ITURBE et al. [1987] veranlaßt hat, sowohl das Neyman–Scott- als auch das Bartlett–Lewis–Clustermodell zu markierten Punktprozessen zu verallgemeinern, in denen mit jedem Sekundärereignis ein Rechteckimpuls assoziiert wird, welcher die Dauer und die Intensität der Zelle beschreibt. Es gibt inzwischen mehrere Varianten von Clustermodellen, die sich vor allem in der Art der Konstruktion des Rechteckimpulses unterscheiden. Intensität bezeichnet hier die Niederschlagsmenge pro Zeiteinheit, die gewöhnlich in [mm/h] gemessen wird. Clusterprozesse werden zur Modellierung des regenerzeugenden Mechanismus' sowohl für tägliche wie auch für stündliche Beobachtungen verwendet. Abbildung 2.1 skizziert den Aufbau des Neyman–Scott–Modelles.

(7.1) Definition. Ein **Neyman–Scott–Clusterprozeß** läßt sich wie folgt charakterisieren:

1. Der Ereignisprozeß der Clusterzentren bildet einen Poissonprozeß mit Rate λ.
2. An jedes Clusterzentrum schließt sich eine zufällige Anzahl $C > 1$ von Zellen an, die unabhängig vom Clusterzentrum ist und ebenso von der Anzahl der Zellen in anderen Clustern. Als Verteilung für C sind die geometrische Verteilung (KAVVAS UND DELLEUR [1981]) und die Poissonverteilung (RODRIGUEZ-ITURBE et al., [1984]) verwendet worden. KHALIQ et al. [1996] haben den Prozeß der Clusterursprünge nach einer exponentialverteilten Dauer enden lassen.
3. Die Abstände der Zellstarts vom Clusterzentrum sind unabhängig und identisch verteilt. Häufig wird hierfür eine Exponentialverteilung verwendet, so daß nahe am Clusterzentrum mehr Zellen auftreten als weiter entfernt.
4. Mit jedem Zellstart wird ein Rechteckimpuls assoziiert, der die Dauer und die Intensität des Regenereignisses beschreibt.
5. Die Impulsdauern sind unabhängig identisch verteilte Zufallsvariablen, die auch unabhängig von allen anderen Variablen des Clusterprozesses sind. Sie werden oft als exponentialverteilt modelliert.
6. Die Impulsintensitäten sind ebenfalls unabhängig, identisch verteilte Zufallsvariablen, die auch von allen anderen Größen des Prozesses unabhängig sind. Sie werden ebenfalls häufig als exponentialverteilt angenommen.

Mit jedem Zellstart wird also ein Tupel oder Zeiger, bestehend aus der Zelldauer und der Zellintensität assoziiert. Es liegt damit weiterhin ein Punktprozeß vor. Clusterzentren können als Sturmfronten interpretiert werden, Zellen hingegen als Regenschauer, aber auch als ein Moment, in dem sich die Intensität eines Regenschauers ändert. Damit werden in dem Modell die tatsächlichen Regenschauer, deren Intensität sich im Zeitverlauf ändern kann, durch Treppenfunktionen approximiert.

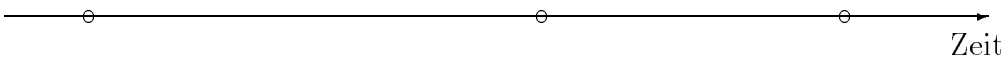
Im Unterschied zum Neyman-Scott-Prozeß werden im Bartlett–Lewis–Clusterprozeß nicht die Abstände der Zellen vom Clusterzentrum, sondern die Abstände zwischen aufeinanderfolgenden Zellstarts modelliert.

(7.2) Definition. Der **Bartlett–Lewis–Clusterprozeß** ist folgendermaßen definiert:

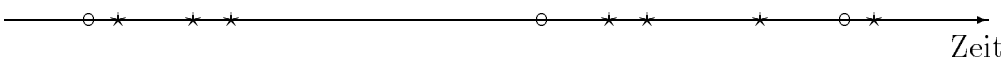
1. Die Primärereignisse ereignen sich gemäß einem Poissonprozeß mit Rate λ.
2. Jedes Primärereignis zieht eine zufällige Anzahl C von Zellen nach sich.
3. Die Abstände aufeinanderfolgender Zellstarts sind unabhängig und identisch verteilte Zufallsvariablen.

Abbildung 2.1: Aufbau des Neyman–Scott–Clustermodelles

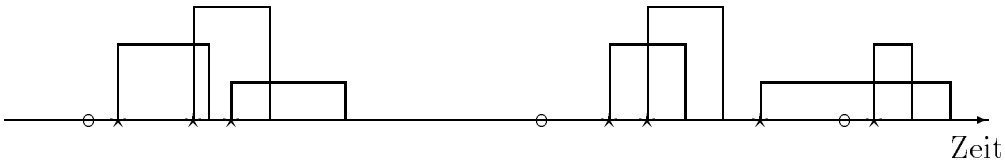
Stürme passieren einen Ort gemäß einem Poissonprozeß.



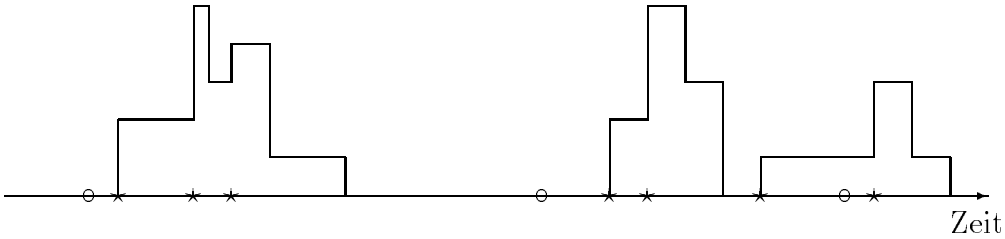
Jeder Sturm zieht eine zufällige Anzahl an (Regen-) Zellen nach sich.



Mit jeder Zelle ist ein Rechteckimpuls verbunden.



Die Gesamtintensität zur Zeit t ist die Summe der Intensitäten aller zur Zeit t aktiven Zellen.



---

4. Mit jedem Zellstart ist ein Rechteckimpuls verbunden, der die Dauer und Intensität der Zelle beschreibt. Beide Zufallsvariablen sind unabhängig voneinander, weiterhin sind die Zelldauern und Zellintensitäten aller Zellen jeweils identisch verteilt.

Rodriguez-Iturbe et al. [1987] beschreiben den Spezialfall, in dem die Zellen eines Clusters einen Poissonprozeß bilden, der nach einer $\exp(\gamma)$-verteilten Zeit endet. Die Auswahl zwischen diesen beiden Typen von Clusterprozessen ist im allgemeinen nicht aufgrund des Datenmaterials möglich, weil die Unterschiede zwischen beiden Modellen in bezug auf den beobachtbaren Prozeß sehr gering sind (Rodriguez-Iturbe et al. [1987]). Variationen dieser Rechteck–Clustermodelle zu Prozessen, in denen die mit einer Zelle assoziierten Impulse nicht rechteckig sind, sind möglich. Aber auch hier dürfte eine auf den Beobachtungen beruhende Auswahl eines Modelles nicht möglich sein. Häufig wird aus Gründen der mathematischen Einfachheit die Exponentialverteilung zur Beschreibung der Zufallsvariablen innerhalb des Clustermodelles gewählt.

In einem Neyman–Scott–Prozeß mit exponentialverteilten Abständen zwischen dem Clusterzentrum und den Zellstarts treten die Zellen am Sturmanfang geclustert auf, was gemäß der physikalischen Interpretation als Regenfronten und -schauer plausibler ist als die im Bartlett-Lewis-Prozeß angenommene gleichmäßige zeitliche Verteilung der Zellen innerhalb des Clusters. Während der Bartlett-Lewis-Prozeß die Ordnungsrelation der Indexmenge zur Beschreibung der Verteilung der Zellen benötigt, besitzt der Neyman-Scott-Clusterprozeß den Vorteil, daß er sich leicht zu einem zeitlich-räumlichen Prozeß erweitern läßt.

Die hier definierten Clusterprozesse eignen sich nur zur Beschreibung eines stationären Prozesses. Da Niederschlag im Jahresverlauf kein stationärer Prozeß ist, wird er üblicherweise durch einen stückweise stationären Prozeß approximiert, z.B. indem die Parameter für jeden Monat separat geschätzt werden. In Kapitel 5 werden wir einen Neyman–Scott–Prozeß betrachten, dessen Parameter Funktionen der Zeit sind.

Clusterprozesse mit Hidden–Markov–Modellen zu vergleichen, ist naheliegend, weil beide Prozesse von einem nichtbeobachtbaren klimatologischen Prozeß ausgehen. In Clusterprozessen treten die Sturmursprünge jedoch unabhängig voneinander auf, während der klimatologische Prozeß C_k in einem Hidden–Markov–Modell eine Markovkette bildet und damit nicht einmal ein Punktprozeß sein muß. Es existieren also keine direkten Analogien. Ein Bartlett–Lewis–Clusterprozeß mit $\exp(\beta)$-verteilten Abständen zwischen den Clusterzellen kann jedoch als Analogon eines Hidden–Markov–Modelles in stetiger Zeit betrachtet werden, in dem der klimatologische Prozeß C_k zwei Zustände annimmt, von denen einer mit Wahrscheinlichkeit $1/\beta$ zu Niederschlag führt und der andere Niederschlag ausschließt ($P(Z_k = 1|C_k = 1) = 1/\beta$, $\mathrm{P}(Z_k = 0|C_k = 0) = 1$).

Zur Modellierung täglichen Regens hat SMITH [1987], von einem Markov–Bernoulli–Modell ausgehend, einen Clusterprozeß mit diskretem Parameterraum entwickelt. In einem Markov–Bernoulli–Prozeß ist die Folge der randomisierten Erfolgswahrscheinlichkeiten $\{X_k\}_{k=1,\ldots,K}$ eine Markovkette erster Ordnung. Das Clustermodell in diskreter Zeit basiert auf Erfolgswahrscheinlichkeiten $\{X_k\}_{k=1,\ldots,K}$, die auf der Intensitätsrate des Neyman–Scott–Prozesses basieren. Zur Konstruktion wird folgendes Ergebnis über Clusterprozesse in stetiger Zeit verwendet:

(7.3) Bemerkung. Ein Neyman-Scott–Clusterprozeß ist ein Cox-Prozeß genau dann, wenn die folgenden drei Bedingungen gelten:

(a) Der Prozeß $\tilde{N}_t$ der Clusterzentren folgt einem Poissonprozeß mit Rate λ.

(b) Die Anzahl der Clusterelemente ist poissonverteilt mit Parameter a.

(c) Die Abstände der Ursprünge der Clusterzellen vom Clusterzentrum sind unabhängig identisch exponentialverteilt mit Erwartungswert $1/b$.

Unter diesen Bedingungen ist der Punktprozeß der Zellursprünge ein Cox–Prozeß

mit stochastischer Intensität

$$X_t = ab \int_0^t e^{-b(t-u)} d\tilde{N}_u \ ,$$

Außerdem ist $\{X_t\}_{t \in I\!R}$ ein Markovprozeß (KARR [1986, Seite 307]). □

Die Konstruktion des Clusterprozesses in diskreter Zeit erfolgt nun in direkter Analogie:

(7.4) Bemerkung. (SMITH [1987]). Sei $\{C_k\}_{k \in I\!N}$ ein stationärer Bernoulliprozeß auf $I\!N$ mit Erfolgswahrscheinlichkeit p, seien $\{\tilde{T}_n\}$ der zugehörige Prozeß der Ankunftszeiten und $\{\tilde{N}_k\}$ der Zählprozeß. Sei

$$X_k = b^k X_0 + ab \sum_{n=1}^{\tilde{N}_k} b^{k-\tilde{T}_n} \ ,$$

mit $a, b > 0$, $a \sum_{k=1}^{\infty} b^k \leq 1$. X_0 habe die Gleichgewichtsverteilung der Markovkette $\{X_k\}$. Der Punktprozeß $\{Z_k\}$ auf $I\!N$ ist ein diskreter Neyman-Scott-Clusterprozeß mit Parametern a, b und p, wenn für alle $n > 0$ und alle $(z_1, \ldots, z_n)$ gilt:

$$\begin{aligned} & \mathrm{P}(Z_K = z_k, \ldots, Z_1 = z_1 | X_k, \, k \geq 0) \\ = & \prod_{k=1}^{K} \mathrm{P}(Z_k = z_k | \, X_k) \\ = & \prod_{k=1}^{K} [z_k X_k + (1 - z_k)(1 - X_k)] \ . \end{aligned}$$

□

Clusterprozesse eignen sich auch zur Modellierung stündlichen Regens, vergl. beispielsweise RODRIGUEZ-ITURBE et al. [1987, 1988], KHALIQ et al. [1996], COWPERTWAIT [1991] und COWPERTWAIT et al. [1994, 1996a]. Teilweise wurden Verbesserungen vorgenommen, zum Beispiel haben RODRIGUEZ-ITURBE et al. [1988] die Verteilung der Zelllängen im Bartlett-Lewis-Prozeß durch eine Exponentialverteilung mit gammaverteiltem Paramter ersetzt, während COWPERTWAIT [1994] ein Modell mit verschiedenen Zelltypen verwendet. Dieses Modell wird als GNSRP(n)-Modell bezeichnet, Generalized Neyman-Scott Rectangular Pulses, wobei der Index n die berücksichtigte Anzahl verschiedener Zelltypen angibt. Damit wird eine in der Meteorologie verwendete Klassifizierung von Niederschlagsgebieten nach ihrer horizontalen Ausdehnung berücksichtigt, denn innerhalb ausgedehnter Niederschlagsgebiete treten typischerweise langandauernde Zellen mit geringer Intensität auf. Eine grobe Unterteilung der Zellen ist die in konvektive Zellen, die als kurze, heftige Regenschauer bemerkt werden, und in stratiforme Zellen, das ist langandauernder, seichter Regen.
Eine Verallgemeinerung des Neyman-Scott-Modelles in ein zeitlich-räumliches Modell ist von COWPERTWAIT [1994] und COWPERTWAIT et al. [1996b] vorgenommen worden.

8 ARMA– und DARMA–Modelle

Die Niederschlagsmengen in nicht überlappenden Intervallen bilden eine Zeitreihe, so daß deren Modellierung als ARMA-Modell erwägenswert ist. Der Nachteil der ARMA-Modelle besteht darin, daß sie keine meteorologische Interpretation bieten. Da die Autokorrelationsstrukturen der über Intervalle bestimmter Länge aggregierten Niederschlagsmengen in verschiedenen meteorologisch motivierten Modellen äquivalent zur Autokorrelationsstruktur einiger ARMA-Prozesses ist, ist die Anpassung eines ARMA-Modelles dann sinnvoll, wenn für Anwendungen des Modelles gerade die Autokorrelation von Bedeutung ist.

OBEYSKARA et al. ([1978], zitiert nach BURLANDO et al., [1993]) haben gezeigt, daß die Autokorrelationsfunktion des Poisson-Rechteckimpulsmodelles und des Neyman-Scott-White-Noise-Modelles mit der eines ARMA(1,1)-Modelles übereinstimmen, und die des Neyman-Scott-Rechteckimpulsmodelles mit der eines ARMA(2,2)-Prozesses. Von diesen Ergebnissen ausgehend, haben BURLANDO et al. [1993] Verfahren zur kurzfristigen Vorhersage (Nowcasting) von Niederschlagsmengen entwickelt, die beispielsweise als Input in Streamflow-Modellen zur Hochwasservorhersage verwendet werden können.

Das ARMA(2,2)-Modell lautet

$$\tilde{Y}_k = \sum_{j=1}^{2} \phi_j \tilde{Y}_{k-j} + \epsilon_k - \sum_{j=1}^{2} \theta_j \epsilon_{k-j} \,,$$

wobei $\tilde{Y}_k := Y_k - EY$, Y_k die Niederschlagsmenge in Stunde k ist und $\epsilon_k \sim \mathrm{N}(0, \sigma_\epsilon^2)$, $k \in I\!N$ unabhängig identisch normalverteilt sind. ϕ_j und θ_j sind die Autoregressions- bzw. Moving Average Koeffizienten des Prozesses $\{\tilde{Y}_k\}$. Die Modellparameter werden unter Verwendung nur der Daten des gerade aktuellen Sturmes geschätzt, also für jeden Sturm separat. Dieser Ansatz ist ein Beispiel der Verwendung eines anderen Modelles für kurzfristige Vorhersagen als zur Beschreibung des gesamten Regenprozesses.

CHANG et al. [1984] haben berücksichtigt, daß Niederschlagsmeßgeräte nur bestimmte Niederschlagsmengen aufzeichnen, z.B. Vielfache von 0.1 mm, und deshalb diskrete ARMA-Modelle (DARMA-Modelle) zur Beschreibung des Prozesses der Niederschlagsmengen verwendet. In diesen Modellen ist der mögliche Wertebereich diskret, dabei kann eine Einteilung der möglichen Niederschlagsmengen in nur drei Klassen ausreichend sein.

Ist eine korrekte Wiedergabe der Autokorrelationsstruktur der Daten wichtig, zeigt der DARMA-Ansatz gute Ergebnisse. Die komplizierte Modellanpassung ist allerdings ein Nachteil. Unbefriedigend ist ferner, daß weder ARMA- noch DARMA-Modelle eine meteorologische Interpretation erlauben. Insbesondere ist eine eindeutige Spezifikation des zugrundeliegenden Ereignisprozesses nicht möglich.

9 Stündlicher Niederschlag - Acremans einfaches Modell

In den bisher betrachteten Modellen wurden mit Regenzellen Rechteckimpulse assoziiert, die eine gleichmäßige Regenintensität über die gesamte Zelldauer implizieren. Werden Zellen mit beobachtbaren Regenschauern gleichgesetzt, ist diese Annahme nicht zutreffend: Die Aufzeichnungen von Hyetographen (Trommelschreibern) sind typischerweise gezackt. Ferner ist die Niederschlagsintensität in den ersten und letzten Stunden einer Regenfront oft niedriger als die mittlere Intensität. Diese Formen werden in Modellen mit Rechteckzellen durch sich überlagernde Zellen wiedergegeben. ACREMAN [1990] hat einen anderen Ansatz gewählt, nämlich die Modellierung von Niederschlagsprofilen, das ist die Aufteilung der Gesamtniederschlagsmenge eines Regenereignisses auf dessen Dauer.

Acreman beschreibt stündlichen Niederschlag aus Farnborough, England, durch einen alternierenden Erneuerungsprozeß. Die Dauern D nasser Perioden, die auch als Regenereignis bezeichnet werden, werden als exponentialverteilt modelliert, die Dauern trockener Perioden hingegen durch eine Pareto-Verteilung. Eine nasse Periode bezeichnet hier eine Zeitspanne, in der während jeder Stunde mindestens 0.1 mm Niederschlag beobachtet werden. Da zur Simulation stündlichen Niederschlages die Dauern nasser und trockener Perioden ganzzahlig sein müssen, werden die Dauern nasser simulierter Perioden jeweils aufgerundet, die Dauern trockener simulierter Perioden jeweils abgerundet. Das bedeutet, daß im Mittel zu den Dauern nasser Perioden 0.5 Stunden hinzuaddiert werden, von den Dauern trockener Perioden entsprechend im Mittel 0.5 Stunden subtrahiert werden. Das berücksichtigt Acreman beim Schätzen der Verteilungsparameter, indem er von den registrierten Dauern nasser Perioden 0.5 subtrahiert. Im Unterschied zu anderen Modellen beschreibt Acreman die Gesamtniederschlagsmenge eines Ereignisses in Abhängigkeit von der Ereignisdauer. Erstere wird durch eine Gammaverteilung beschrieben, deren Formparameter α und Skalenparamter β von der Ereignisdauer D abhängen:

$$\alpha = D^a - c \quad , \quad \beta = \begin{cases} b\,D\,, & D \geq 5\,, \\ c & D \leq 4\,, \end{cases}$$

wobei D die Ereignisdauer in Stunden bezeichnet und a, b, c zu schätzende Parameter sind.

Zur Beschreibung der Ereignisprofile wird der Anteil der Gesamtniederschlagsmenge, der in einer Stunde fällt, durch eine Betaverteilung beschrieben. Diese ist hierfür durch die Beschränkung ihres Definitionsbereiches auf $[0, 1]$ ein natürlicher Kandidat. Die Dichte ist

$$f(t) = t^{\alpha-1}(1-t)^{\beta-1}/B(\alpha,\beta) \quad ,0 \leq t \leq 1,\ \alpha,\beta > 0\ ,$$

wobei $B(\alpha, \beta)$ die Betafunktion beschreibt und t den Anteil der schon vergangenen Zeit an der Gesamtzeit eines Regenereignisses bezeichnet. Die Ereignismenge der

letzten Stunde eines Ereignisses muß separat modelliert werden. Um zu gewährleisten, daß alle Anteile zu eins aufaddieren, ist der Anteil a_K der Niederschlagsmenge in der letzten Stunde K eines Regenereignisses $1 - \sum_{k=1}^{K-1} a_k$. In der Analyse zeigte sich, daß die Profile von Ereignissen unterschiedlicher Länge annähernd übereinstimmen. Auch dieses Modell legt homogene Perioden zugrunde, aber eine Einteilung in eine Sommer- und eine Wintersaison erweist sich für die betrachteten Orte als ausreichend.

Acremans Modellvalidation fällt positiv aus, insbesondere auch im Hinblick auf Extremwertanalysen. Ein positiver Aspekt dieses Modells ist, daß die Modellparameter direkt aus den Daten geschätzt werden können, teils als Maximum-Likelihood-Schätzer, teils über die Methode der Momente. Weiterhin ist der einfache Aufbau des Modelles bestechend, das allerdings nicht von weiteren Autoren aufgegriffen worden zu sein scheint.

10 Punktprozesse in stetiger und Beobachtungen in diskreter Zeit

Wie in Abschnitt 1 bereits erwähnt worden ist, liegen Niederschlagsmessungen häufig als Serie der Gesamtniederschläge in aufeinanderfolgenden Intervallen fester Länge vor, als sogenannte aggregierte Niederschläge. Der aus diesen Daten abgeleitete Punktprozeß beschreibt damit, ob es in einem Intervall mindestens einmal geregnet hat, während der tatsächliche Prozeß der Regenereignisse nicht beobachtet wird. Punkte werden also nur an Vielfachen der Länge des Beobachtungsintervalls notiert, so daß der beobachtete 0–1–Prozeß ein Punktprozeß in diskreter Zeit ist, der auf zwei Weisen interpretiert werden kann:

1. Die beobachteten Punkte repräsentieren den gesamten Prozeß der tatsächlich aufgetretenen Regenereignisse.
2. Die beobachteten Punkte repräsentieren eine gefilterte Stichprobe des tatsächlichen Prozesses.

Gemäß der ersten Interpretation entspricht ein Punkt einem nassen Tag, einer nassen Stunde etc. In diesem Fall müßte ein Punktprozeß mit diskretem Parameterraum angepaßt werden, um Bias zu vermeiden und um insbesondere das Clusterverhalten des beobachteten Prozesses richtig zu beurteilen. Das wollen wir nachfolgend kurz erläutern. Der Referenzprozeß für unabhängige Ereignisse ist im Fall diskreter Punktprozesse nicht der Poissonprozeß, sondern der Bernoulliprozeß. Foufoula-Georgiou und Guttorp [1986], Guttorp [1986] und Foufoula-Georgiou und Lettenmaier [1987] haben sich mit dieser Problematik befaßt und Diskretisierungen der häufig zur Modellierung von Regenfall verwendeten Prozesse in stetiger Zeit untersucht.

Statistischer Hintergrund

Die Anpassung eines Punktprozesses an Beobachtungen basiert häufig auf den ersten und zweiten Momenten des Prozesses, die sich bei einem Prozeß in stetiger Zeit und seinem Äquivalent in diskreter Zeit häufig unterscheiden. Wir betrachten beispielhaft den Poissonprozeß und dessen Äquivalent in diskreter Zeit, den Bernoulliprozeß.

Die zu einem Prozeß in stetiger Zeit gehörende Ereignisserie in diskreter Zeit wird durch zwei Operationen erhalten: Diskretisieren und Clipping. Unter Diskretisieren wird die Zusammenfassung von Ereignissen in Intervallen gleicher Länge verstanden. Das Zuweisen der Werte 0 oder 1 zu jedem Intervall, wenn kein bzw. mindestens ein Ereignis in ihm auftritt, wird als Clipping bezeichnet. Oft werden beide Vorgänge zusammen als Diskretisierung bezeichnet.

Wir betrachten einen homogenen Punktprozeß mit Ereignisrate m einmal in stetiger Zeit und einmal in diskreter Zeit. Es ist

$$m = \lim_{t \to 0} \frac{1}{t} \, \mathrm{P}(N(\,(0,t]\,) > 0) \quad \text{bzw.} \quad m = \mathrm{P}(N((k,k+1]) > 0)\ .$$

Dann sind die Erwartungswert-Zeit-Funktionen im stetigen und im diskreten Fall

$$\mathrm{E}N(t) = m\,t \quad \text{bzw.} \quad \mathrm{E}N_k = (k-1)\,m\ .$$

Die Varianz-Zeit-Funktionen $\mathrm{Var}(N(t))$ bzw. $\mathrm{Var}(N_k)$ des Prozesses und zugehörigen Serie können nicht übereinstimmen, da bei Beobachtungen in stetiger Zeit mehr Punkte gemessen werden können. Folglich sind auch die Dispersionsindizes $I(t) = \mathrm{Var}(N(t))/\mathrm{E}(N(t))$ bzw. $I_k = \mathrm{Var}(N_k)/\mathrm{E}(N_k)$, die ein Maß für das Clusterverhalten sind, bei Prozeß und Serie unterschiedlich.
Ebenso ist die bedingte Intensität ein Maß für das Clusterverhalten. Sie beschreibt die Wahrscheinlichkeit, einen Punkt zur Zeit t zu erhalten, bei einem gegebenen Punkt zur Zeit 0 (Cox und Isham [1979]):

$$h(t) = \lim_{\delta_1,\delta_2 \to 0} \frac{1}{\delta_2} \, \mathrm{P}\Big(N((t,t+\delta_2] > 01 | N((-\delta_1,0]) > 0\Big)\ .$$

Ist die bedingte Intensität größer als die Ereignisrate, so ist die Wahrscheinlichkeit, einen Punkt in der Nähe eines anderen Punktes zu finden, größer als an beliebiger Stelle. Wir sprechen von Überdispersion und bezeichnen den Prozeß als geclustert. Umgekehrt sprechen wir von Unterdispersion oder regelmäßigem Clustern, wenn die bedingte Intensität kleiner als die Ereignisrate ist. Bei einem Poissonprozeß ist die bedingte Intensität gleich der Intensität.

Abschließend betrachten wir die Survivor-Funktion $G(t)$, die ebenfalls ein Maß für das Clusterverhalten eines Prozesses ist,

$$G(t) = \mathrm{P}(N(t) = 0 |\ \text{Punkt zur Zeit } 0)\ .$$

Die folgende Tabelle stellt diese Eigenschaften des Poissonprozesses denen des Bernoulliprozesses gegenüber, die Herleitung gibt GUTTORP [1986]. Weitere Eigenschaften nennen FOUFOULA-GEORGIOU UND LETTENMAIER [1987]. GUTTORP [1986] zeigt ferner anhand einiger exemplarisch ausgewählter Prozesse (Poisson-, Neyman-Scott-, RCM-Prozeß) die Unterschiede zu den korrespondierenden Serien auf.

Parameter	Poissonprozeß	Bernoulliserie
Ereignisrate	λ	π
Varianz-Zeit-Funktion	$\mathrm{Var}N(t) = \lambda t$	$\mathrm{Var}N_k = k\pi(1-\pi)$
Dispersionsindex	$I(t) = 1$	$I_k = 1 - \pi$
Bedingte Intensität	$h(t) = \lambda$	$h_k = \pi$
Log-Survivor-Funktion	$\log G(t) = -\lambda t$	$\log G_k = -k \ln(1-\pi)$

Die Maximum-Likelihood-Schätzer der Ereignisraten λ bzw. π stimmen überein; sie sind NR/N, wenn Beobachtungen von N Tagen vorliegen und NR die Anzahl nasser Tage bezeichnet.

Die weiteren Eigenschaften der Prozesse unterscheiden sich jedoch. Der Dispersionsindex des Poissonprozesses ist konstant 1, während der des Bernoulliprozesses stets kleiner als 1 ist. Der Bernoulliprozeß ist also im Vergleich zum Poissonprozeß unterdispersiert. Wäre täglicher Regen ein Bernoulli–Prozeß, so erschiene er bei einer Modellierung als stetiger Prozeß unterdispersiert. Aber auch eine Ereignisserie, die relativ zum Bernoulliprozeß geclustert ist, kann im Vergleich zum Poissonprozeß unterdispersiert sein. Damit führen Vergleiche der diskreten Ereignisserie mit dem Poissonprozeß zu falschen Folgerungen über die Unabhängigkeit der Ereignisse, wenn sie auf den ersten und zweiten Momenten basieren.

Je stärker der regenerzeugende Prozeß geclustert ist, desto größer ist der Informationsverlust, der mit der Diskretisierung zu täglichem oder stündlichem Regen einhergeht. Er führt zu verfälschten Schätzern. Diskrete Beobachtungen sollten also durch einen Prozeß mit diskretem Parameterraum modelliert werden.

Anmerkungen zur Zeiteinteilung

Gilt das Augenmerk dem zugrundeliegenden regenerzeugenden Mechanismus, so sollte das angepaßte Modell mit unterschiedlichen Zeiteinteilungen kompatibel sein. Führen über verschiedene Zeitniveaus aggregierte Niederschlagsmessungen zu unterschiedlichen Parameterschätzern, kann den Parametern keine physikalische Bedeutung beigemessen werden.

Wir betrachten jetzt also die zweite der oben genannten Interpretationen des Prozesses der Niederschlagsereignisse. Wenn aus den über Intervalle konstanter Dauer, zum Beispiel Stunden oder Tage, gemessenen Niederschlagsmengen auf den zugrundeliegenden regenerzeugenden Mechanismus geschlossen werden soll, werden die Modellparameter meistens auf der Methode der Momente basierend geschätzt. Die so erhaltenen Parameterschätzer sollten mit unterschiedlichen Zeiteinteilungen kompatibel sein, so daß stündliche und tägliche Messungen zum gleichen Modell führen. RODRIGUEZ-ITURBE et al. [1984] haben Untersuchungen dieser Art für das Poisson-White-Noise-Modell, das Rechteckimpulsmodell sowie für eine Version des Neyman-Scott-White-Noise-Clustermodelles anhand zweier Datensätze (Denver, USA und Agua

Fria, Venezuela) durchgeführt. Die Modellparameter wurden unter Verwendung der Momente der aggregierten Niederschlagsmengen in einem Kleinste-Quadrate-Ansatz geschätzt, wobei jeweils nur ein Aggregationsniveau Δ, $\Delta = 1, 4, 6, 12, 24$ Stunden verwendet wurde. Die resultierenden Parameterschätzer wurden dann miteinander verglichen. Nach dieser Analyse scheint nur das Neyman-Scott-Clustermodell mit unterschiedlichen Zeiteinteilungen verträglich zu sein und deshalb als Modell für den regenerzeugenden Mechanismus in Frage zu kommen.
Diese Kompatibilität des Neyman-Scott-Modells mit unterschiedlichen Zeitskalen haben FOUFOULA-GEORGIOU UND GUTTORP [1986] im Rahmen der Diskretisierung stochastischer Prozesse in stetiger Zeit erneut betrachtet. Im Unterschied zu Rodriguez-Iturbe et al. haben sie nicht die theoretischen Momente der aggregierten Niederschlagsmengen unter den Prozessen in stetiger Zeit, sondern die unter den durch Diskretisierung entstandenen Ereignisserien verwendet. Ihre Analyse der gleichen Datensätze wie die von Rodriguez-Iturbe et al. verwendeten führen zu Parameterschätzern, die von der gewählten Zeitskala abhängen. Foufoula-Georgiou und Guttorp merken an, daß sie die Güte ihres Schätzverfahrens zuvor an künstlich erzeugten Daten untersucht haben und dabei die Input-Parameter aus allen verwendeten Aggregationsniveaus mit befriedigender Genauigkeit zurückgewinnen konnten.

Eine Ursache für diese unterschiedlichen Ergebnisse kann auch darin liegen, daß Foufoula-Georgiou und Guttorp die Clustergrößen als geometrisch verteilt annehmen, während Rodriguez-Iturbe et al. von poissonverteilten Clustergrößen ausgehen. Diese Entscheidung wird aber nicht nur nach physikalischen, sondern vor allem auch aus mathematischen Gründen getroffen. Zum Beispiel konstatieren Foufoula-Georgiou und Guttorp, daß sie nur unter der Annahme geometrisch verteilter Clustergrößen geschlossene Ausdrücke für die Momente der Ereignisserie angeben können. Da aber im Neyman-Scott-White-Noise-Modell die Autokorrelationsfunktion der aggregierten Niederschläge nur von der Clustergröße C und den Abständen der Zellen vom Clusterzentrum abhängt, beeinflußt die gewählte Verteilung das Gesamtmodell, wenn die Autokorrelationen in der Modellanpasung verwendet werden. Die geometrische Verteilung wurde u.a. auch von KAVVAS UND DELLEUR [1981] verwendet, die Poissonverteilung u.a. auch von SMITH UND KARR [1985].
Foufoula-Georgiou und Guttorp erhalten bei Anpassung zweier Clustermodelle, die sich nur in der Verteilung der Clustergröße unterscheiden (Poisson- bzw. geometrische Verteilung), an den Denver-Datensatz unterschiedliche Schätzer der Rate λ des Prozesses der Sturmzentren und der erwarteten Niederschlagsmenge $\mathrm{E}(X)$. Auch aus diesem Grund sollte den Parametern keine physikalische Interpretation beigemessen werden.
Die Anpassung von Clustermodellen haben RODRIGUEZ-ITURBE et al. [1987, 1988] und COWPERTWAIT [1994, 1996a] verbessert, indem sie in einem Kleinste-Quadrate-Ansatz zum einen mehr als die minimal benötigte Anzahl von Eigenschaften der Serie der aggregierten Niederschlagsmengen verwenden und diese Eigenschaften sich zum anderen auf mehrere Zeiteinteilungsniveaus beziehen.
Insgesamt läßt sich also schließen, daß Modellparameter, die aus Beobachtungen nur eines Aggregationsniveaus geschätzt werden, nicht mit anderen Aggregationsniveaus kompatibel sind und deshalb dort nicht für Vorhersagen verwendet werden sollten.

Kapitel 3

Beschreibung des Datensatzes

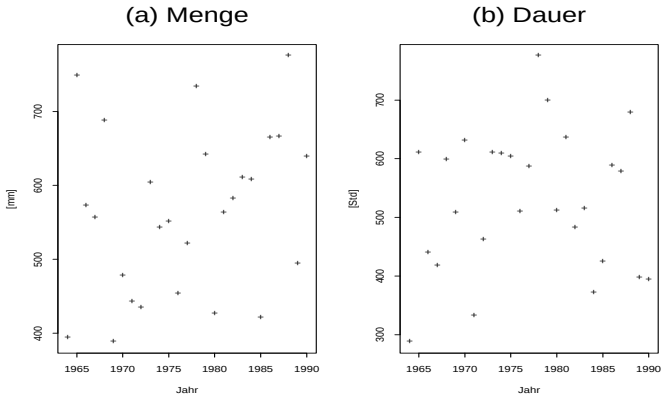
Der verwendete Datensatz besteht aus den Niederschlagsmessungen des Zeitraumes 01.01.1964 - 31.12.1990 der Meßstation ,,Max-Eyth-Schule" in Stuttgart. Die Meßstation wird vom städtischen Tiefbauamt betrieben. Es handelt sich um hochaufgelöste Niederschlagsmessungen mit einem Regenmesser nach Hellmann der Firma Lambrecht, Klimatologische Meßtechnik Göttingen, bei denen Anfang und Ende eines Regenereignisses mit einer Genauigkeit von $\pm$ 5 min sowie der in dieser Zeit gefallene Niederschlag mit einer Auflösung von 0.1mm aufgezeichnet werden. Als ,,Regenereignis" oder kurz ,,Ereignis" wird im folgenden die Periode bezeichnet, innerhalb derer das Meßgerät Niederschlag registriert hat.

Der Niederschlagsprozeß in unterschiedlichen Jahren kann als Realisation unabhängiger stochastischer Prozesse betrachtet werden, wenn die Korrelationen am Ende eines Jahres und dem Anfang des folgenden Jahres vernachlässigt werden. Ob diese Realisationen aber dem gleichen Entstehungsgesetz gehorchen, ist zunächst ungewiß. Klimaänderungen oder - naheliegender - Änderungen in der Meßmethode oder ein Wechsel des Standortes des Meßgerätes können bewirken, daß die Niederschlagsmessungen in unterschiedlichen Jahren verschiedenen Entstehungsgesetzen gehorchen. Für eine erste Analyse des Datenmateriales sind daher folgende Eigenschaften des Niederschlagsprozesses betrachtet worden:

1. Der Jahresniederschlag Y_J,
2. die mittlere Niederschlagsmenge eines Regenereignisses (Ereignismenge) $\bar{Y}_E$,
3. die empirische Standardabweichung der Ereignismenge $s(Y_E)$,
4. die jährliche Gesamtniederschlagsdauer D_J,
5. die mittlere Dauer eines Ereignisses (Ereignisdauer) $\bar{D}_E$,
6. die empirische Standardabweichung der Ereignisdauer $s(D_E)$.

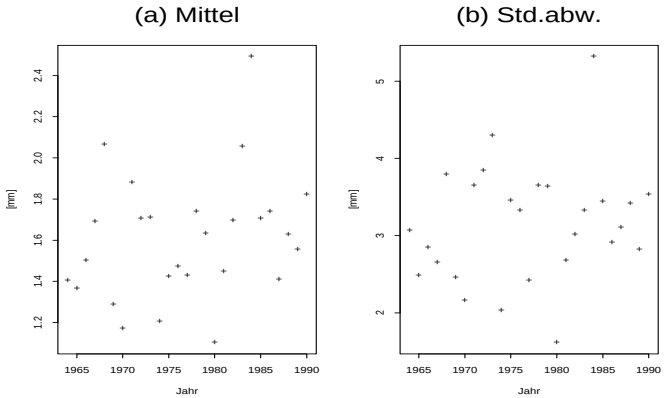
Die Werte dieser Eigenschaften sind in Tabelle 3.1 für jedes Jahr aufgelistet bzw. in den Abbildungen 3.1, 3.2 und 3.3 zu sehen.

Abbildung 3.1: Jahresniederschlagsmenge und - dauer



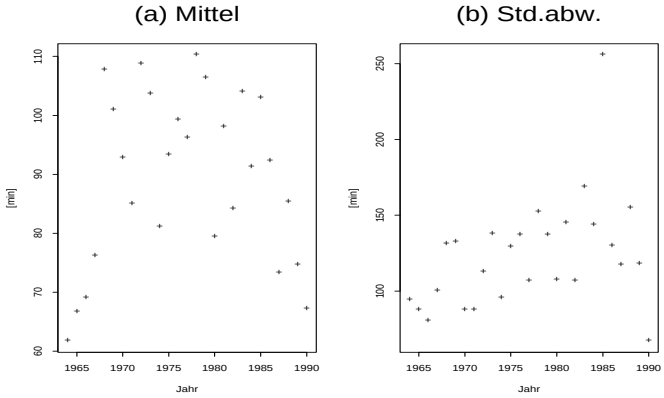
Jahresniederschlag in den Jahren 1964 - 1990. (a) Menge [mm]. (b) Dauer [Std].

Abbildung 3.2: Ereignismenge



Niederschlagsmenge Y_E pro Regenereignis in den Jahren 1964 - 1990 [mm].
(a) Jahresmittel $\bar{Y}_E$. (b) Standardabweichung $s(Y_E)$ pro Jahr.

Abbildung 3.3: Ereignisdauer



Niederschlagsdauer D_E pro Regenereignis in den Jahren 1964 - 1990, [min].
(a): Mittel $\bar{D}_E$. (b): Standardabweichung $s(D_E)$.

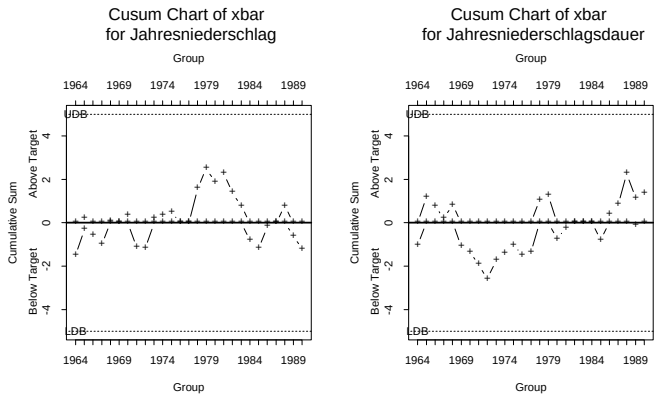
Tabelle 3.1: Eigenschaften des Jahresniederschlages

	Y_J	Y_E	$s(Y_E)$	D_J [Tage:Std:Min]	D_E [Std:Min]	$s(D_E)$[Std:Min]	N
1964	393.4	1.50	3.35	12:03:40	1:06	1:43	263
1965	748.9	1.43	2.69	25:13:50	1:10	1:39	524
1966	572.7	1.63	3.06	18:14:35	1:16	1:26	352
1967	543.1	1.84	2.90	17:19:53	1:27	1:50	295
1968	687.1	2.29	4.41	25:16:30	2:03	2:34	300
1969	388.5	1.39	2.60	21:13:55	1:51	2:25	280
1970	465.9	1.26	2.62	26:24:40	1:45	1:38	369
1971	441.9	2.15	3.92	14:04:15	1:39	1:39	206
1972	434.8	1.80	4.02	19:06:25	1:55	2:13	241
1973	603.6	1.78	4.39	25:10:49	1:48	2:23	340
1974	542.3	1.44	2.38	25:21:46	1:39	1:45	378
1975	552.3	1.77	4.27	25:19:45	1:59	2:42	313
1976	453.4	1.90	4.19	21:22:35	2:12	2:50	239
1977	521.5	1.76	2.88	25:06:53	2:03	2:00	296
1978	733.6	2.03	5.07	32:22:50	2:11	3:10	362
1979	642.1	1.83	4.05	29:10:50	2:01	2:47	351
1980	425.7	1.33	2.17	21:19:20	1:38	2:12	321
1981	562.9	1.63	3.25	26:18:55	1:51	2:50	346
1982	582.0	2.11	3.79	20:14:20	1:47	2:02	276
1983	610.1	2.20	3.42	21:14:05	1:52	2:52	278
1984	607.9	3.42	6.15	15:19:50	2:08	2:35	178
1985	420.3	1.91	3.69	17:20:30	1:57	5:02	220
1986	664.4	1.99	3.23	24:19:20	1:47	2:15	334
1987	666.0	1.50	3.27	24:06:55	1:18	2:00	445
1988	775.1	1.63	3.41	28:07:55	1:25	2:35	477
1989	494.8	1.71	3.13	16:16:25	1:23	2:02	290
1990	640.1	1.95	3.80	16:18:15	1:13	1:14	329

Unbereinigte Daten
Y_J: Jahresniederschlag [mm], $\bar{Y}_E$: mittlere Ereignismenge [mm], $s(Y_E)$: Standardabweichung der Ereignismenge [mm], D_J: Jahresniederschlagsdauer [Tage:Std:Min], $\bar{D}_E$: mittlere Ereignisdauer [Std:Min], $s(D_E)$: Standardabweichung der Ereignisdauer [Std:Min], N: Anzahl an Regenereignissen.

Sowohl die Jahresniederschlagsmenge Y_J als auch die Jahresniederschlagsdauer D_J sind gleichmäßig gestreut. Die mittlere Ereignisdauer $\bar{D}_E$ ist während der Jahre 1964 - 1968 konstant ansteigend von 66 auf 126 Minuten, liegt während der Jahre 1969 - 1985 zwischen 100 und 130 Minuten, bevor sie während der Jahre 1986 - 1990 bei etwa 80 Minuten liegt. Die mittlere Ereignismenge $\bar{Y}_E$, deren Standardabweichung $s(Y_E)$ sowie die Standardabweichung $s(D_E)$ der Ereignisdauer sind während des gesamten Beobachtungszeitraumes gleichmäßig gestreut. Im Jahr 1984 nehmen $\bar{Y}_E$, $s(Y_E)$ und $s(D_E)$ außergewöhnlich hohe Werte an. Es konnte kein Indiz auf veränderte äußere Umstände der Niederschlagsmessung ermittelt werden, das einen Ausschluß der Beobachtungen dieses Jahres gerechtfertigt hätte. Es handelt sich offensichtlich um einen Ausreißer, der in der Analyse der Daten aber beibehalten wird, um die tatsächlichen Schwankungen des Niederschlagsprozesses nicht zu verzerren. Der Jahresniederschlag läßt keine Außergewöhnlichkeiten im Jahr 1984 erkennen. Keine der betrachteten Statistiken läßt offensichtliche Trends erkennen.

Abbildung 3.4: Cusum Charts



Cusum Chart für (a) Jahresniederschlagsmenge, (b) Jahresniederschlagsdauer.

Mit Hilfe von Cusum Charts[1] werden die Jahresniederschlagsmenge und -dauer Y_J bzw. D_J auf Homogenität überprüft, vergl. Abb. 3.4. Die Cusum Charts weisen nicht auf eine Veränderung von Y_J oder D_J hin, so daß der Niederschlagsprozeß in den Jahren des Beobachtungszeitraumes als Realisation des gleichen stochastischen Prozesses betrachtet werden kann. Zusätzlich werden Cusum Charts für die monatliche Niederschlagsmenge sowie für die mittlere Ereignisniederschlagsmenge für alle Monate des Beobachtungszeitraumes gesondert betrachtet. Auch sie lassen keine Inhomogenitäten erkennen. Damit kann insbesondere der Niederschlagsprozeß in allen beobachteten Monaten Januar als eine Kopie des gleichen stochastischen Prozesses betrachtet werden, analog für die übrigen Monate Februar, März etc., vergl. die Graphiken in Anhang A. Im folgenden werden wir stündlichen und täglichen Niederschlag untersuchen. Dazu aggregieren wir die registrierten hochaufgelösten (in fünfminütigem Zeitabstand vorgenommenen) Niederschlagsmessungen unter der Annahme, die Intensität eines Ereignisses sei während dessen Dauer konstant. Sind zum Beispiel an einem beliebigen Tag 6mm Niederschlag zwischen 11:50h und 12:20h registriert worden, so werden der Stunde 11:00 - 11:59h 2mm und der Stunde 12:00 - 12:59h 4mm zugeordnet.

1 Auswahl eines Schwellenwertes

In der Modellierung von Niederschlägen kann es sinnvoll sein, eine Mindestniederschlagsmenge einzuführen, ab deren Überschreitung ein Zeitintervall als naß bezeichnet wird. Für einen solchen Schwellenwert sprechen mehrere Gründe. Die Auflösung vieler Meßgeräte beträgt 0.1mm, so daß eine geringere Niederschlagsmenge nicht registriert wird. Geringe Niederschlagsmengen können auch durch Tau oder Nebel verursacht sein, die nur registriert werden, wenn das Meßgerät in den frühen Tagesstunden abgelesen wird, weil sie im weiteren Tagesverlauf verdunsten. Die registrierte

[1] Zur Erläuterung von Cusum Charts vergl. Anhang A.

Tabelle 3.2: ACF der Niederschlagsmengen an nassen Tagen

$r(j) \setminus d$	0.1	0.2	0.3	0.4	0.5	0.6
r(1)	0.13	0.13	0.14	0.15	0.14	0.14
r(2)	0.05	0.05	0.06	0.07	0.06	0.06
r(3)	0.06	0.06	0.05	0.05	0.06	0.04
r(4)	0.03	0.04	0.06	0.06	0.05	0.06
r(5)	0.01	0.00	0.00	0.01	0.01	0.01
r(6)	0.05	0.05	0.05	0.05	0.05	0.05
r(7)	0.09	0.10	0.10	0.10	0.09	0.08
r(8)	0.03	0.03	0.04	0.03	0.03	0.02
r(9)	0.03	0.03	0.03	0.03	0.04	0.04
r(10)	0.04	0.04	0.04	0.03	0.02	0.03

Empirische Autokorrelationskoeffizienten r(j) der Niederschlagsmengen an nassen Tagen bei unterschiedlichen Schwellenwerten d [mm]. Tag t ist naß, wenn $Y_t \geq d$.

Niederschlagsmenge sollte bei längeren Intervallen (z.B. 12, 24 Stunden) unabhängig von der Uhrzeit sein, zu der das Meßgerät abgelesen wird, damit die Aufzeichnungen unterschiedlicher Meßstationen kompatibel sind. Durch die Einführung eines geeigneten Schwellenwertes wird dieses Ziel erreicht.
Zusätzlich wird die statistische Modellierung des Niederschlagsprozesses erleichtert, wenn die Niederschlagsmengen in nassen Intervallen unabhängig sind. In diesem Fall kann der Niederschlagsprozeß in einen Ereignisprozeß und einen Mengenprozeß aufgegliedert werden, die stochastisch unabhängig voneinander sind und daher separat analysiert werden können. Durch die Verwendung eines moderaten Schwellenwertes wird im allgemeinen nur ein geringer Prozentsatz des Gesamtniederschlages vernachlässigt, so daß dieses Vorgehen akzeptabel ist. Üblich sind Schwellenwerte d für 24-Stunden-Niederschläge. Gebräuchliche Werte sind $d = 0.1, 0.3, 0.5$mm (vergl. BUISHAND [1977]). Die Einführung eines Schwellenwertes für kürzere Intervalle wie $d = 0.1$mm ist nicht sinnvoll, weil damit ein zu großer Anteil des Gesamtniederschlages außer acht gelassen wird. Allerdings kann auch für den 1-Stunden-Niederschlagsprozeß durch die Meßgenauigkeit de facto ein Schwellenwert $d = 0.1$mm gegeben sein.

Im folgenden sei $Y_{t,n}$ die Niederschlagsmenge an Tag $t = 1, \ldots, T$ (z.B. $T = 365$) in Jahr $n = 1, \ldots, N$. Ferner sei $NR(t)$ die Anzahl der Jahre, an denen Tag t naß ist. Wir untersuchen, ob der Schwellenwert d die Autokorrelationsstruktur der Niederschlagsmengen $W_t(d)$ an nassen Tagen t beeinflußt. Bezeichne NA fehlende Werte. Dann sei

$$\textbf{(1.1)} \quad W_{t,n}(d) = \begin{cases} Y_{t,n}, & \text{wenn} \quad Y_{t,n} \geq d, \\ NA & \text{sonst.} \end{cases}$$

Der Lag(j)-Autokorrelationskoeffizient $\rho(j)$ der $W_{t,n}(d)$ wird durch

$$\textbf{(1.2)} \quad r(j) = \hat{\rho}(j) = \frac{\widehat{\mathrm{Cov}}(W_t(d),\ W_{t+j}(d))}{\widehat{\mathrm{Var}}W_t(d)} \ ,$$

geschätzt, wobei

$$\textbf{(1.3)} \quad \widehat{\mathrm{Cov}}(W_t(d), W_{t+j}(d)) =$$

$$\frac{1}{\tilde{N}-j}\sum_{t=1}^{T}\sum_{n=1}^{NR(t)}(W_{t,n}(d) - W_{t,\cdot}(d))((W_{t+j,n}(d) - W_{t+j,\cdot}(d))$$

mit

$$\tilde{N} = \sum_{t=1}^{T}\sum_{n=1}^{N} 1(Y_{t,n} \geq d) \cdot 1(Y_{t+j,n} \geq d)$$

$$\textbf{(1.4)} \quad \widehat{\mathrm{Var}}(W_t(d)) = \frac{1}{(\sum_{t=1}^{T} NR(t)) - 1}\sum_{t=1}^{T}(W_{t,n}(d) - W_{t,\cdot}(d))^2$$

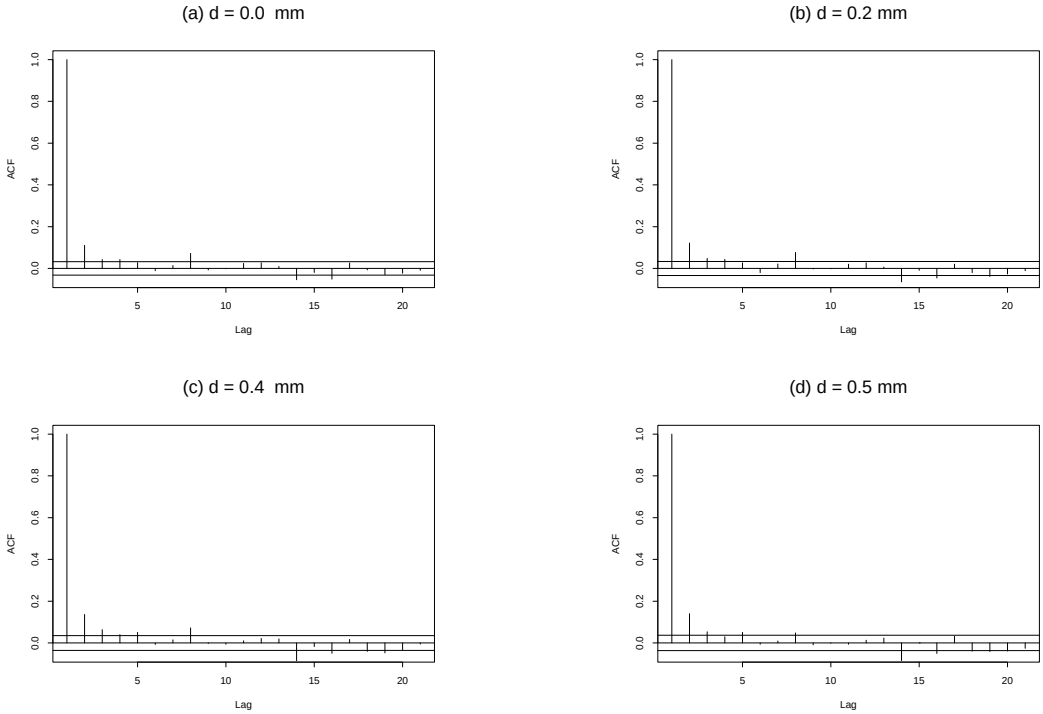
mit $W_{t,\cdot}(d) = \sum_{n=1}^{N_t} W_{t,n}(d)$ die mittlere Niederschlagsmenge an einem nassen Tag t ist. Fehlende Werte werden in der Berechnung der Summe ausgelassen. Saisonale Schwankungen der erwarteten Niederschlagsmenge an einem Tag werden in dieser Formel berücksichtigt.

Die Autokorrelationsfunktionen r besitzen bei allen Schwellenwerten d ähnliche Verläufe. Die Niederschlagsmengen an nassen Tagen sind bei keinem der betrachteten Werte d unabhängig; Der Lag(1)-Koeffizient $r(1)$ ist stets signifikant von null verschieden gemäß der Normalapproximation (vergl. CHATFIELD [1996])

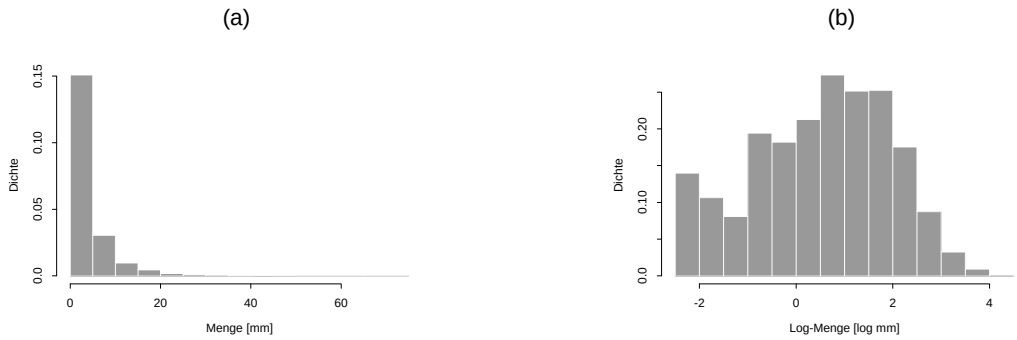
$$r(j) \sim \mathrm{N}(-\frac{1}{n}, \frac{1}{\sqrt{n}}) \ ,$$

wobei n die Anzahl der zum Schätzen von $r(j)$ verwendeten Beobachtungen ist, vergl. Abb. 3.5 und Tab. 3.2. Ein noch höherer Schwellenwert ist aber im Hinblick auf das Ziel der Modellierung stündlichen Niederschlages nicht zu rechtfertigen. Da hier etwa 3500 Beobachtungen verwendet worden sind, sind die so bestimmten Konfidenzintervalle sehr schmal. Als Faustregel gelten nur Korrelationen $r(j) \geq 0.05$ als signifikant von null verschieden, so daß bei allen Werten von d die ersten drei bzw. ersten vier Lags als ungleich null betrachtet werden müssen. Eine sinnvolle Auswahl eines Schwellenwertes kann also nicht aufgrund der ACF getroffen werden.
Der Schwellenwert beeinflußt aber die Verteilung der Niederschlagsmengen an nassen Tagen. An dieser dieser Stelle der Datenanalyse soll weder eine Entscheidung über die Verteilung der Niederschlagsmengen an nassen Tagen noch über eine mögliche Transformation der Niederschlagsmengen getroffen werden. Histogramme der Niederschlagsmengen an nassen Tagen zeigen, daß deren Verteilung rechtsschief ist und

Abbildung 3.5: Niederschlagsmengen an nassen Tagen



ACF der Niederschlagsmengen an nassen Tagen. Tag t ist naß wenn $Y_t \geq d$.
Die obere und die untere horizontale Linie markieren geschätzte 95%-Konfidenzintervalle um Null: [-0.035; 0.035]. (a) d=0.0 mm, (b) d=0.2 mm. (c) d=0.4 mm, (d) d=0.5 mm.



(a) Niederschlagsmengen an nassen Tagen, kein Schwellenwert.
(b) Log-Niederschlagsmengen an nassen Tagen, kein Schwellenwert.

---

eine Logarithmus-Transformation sinnvoll sein könnte, vergl. Abb. 3.5. Um nach einer Log-Transformation nicht die meiste Wahrscheinlichkeitsmasse in der am weitesten links liegenden Klasse und dennoch den Schwellenwert so niedrig wie möglich zu wählen, wählen wir als Schwellenwert $d = 0.2$mm: Tag t ist naß , wenn $Y_t \geq 0.2$mm. Damit wird nur ein geringer Anteil der Gesamtniederschlagsmenge vernachlässigt, vergl. Tab. 3.3.

Bemerkung: In der weiteren Analyse werden also Tage t mit weniger als 0.2mm Niederschlag als trocken betrachtet und wir setzen $Y_t = 0.0$mm.

Wir werden die Begriffe täglicher und stündlicher Niederschlag in dem folgenden Sinne verwenden:

Bemerkung: Als h-**Stunden-Niederschlag** wird der über Zeitintervalle der Länge h aggregierte Niederschlag bezeichnet. Als **Täglichen Niederschlag** oder **24-Stunden-Niederschlag** bezeichnen wir die von 0:00 - 23:59h registrierte Niederschlagsmenge. **Stündlicher Niederschlag** bezeichnet die während einstündiger, sich nicht überlappender Zeitintervalle (0:00 - 0:59h, 1:00 - 1:59h etc.) registrierte Niederschlagmenge. **3-Stunden-Niederschlag** bezeichnet die in den Intervallen 0:00 - 2:59h, 3:00 - 5:59h, 6:00 - 8:59h etc. registrierte Niederschlagsmenge. **6-Stunden-Niederschlag** bezeichnet die in den Intervallen 0:00 - 5:59h, 6:00 - 11:59h, 12:00 - 17:59h, 18:00 - 23:59h registrierte Niederschlagsmenge. **12-Stunden-Niederschlag** bnezeichnet die in den Intervallen 0:00 - 11:59h, 12:00 - 23:59h registrierte Niederschlagsmenge. Abweichungen von dieser Begriffsbildung, insbesondere der in sich überlappenden Intervallen registrierte Niederschlag, werden explizit angemerkt werden.

Bemerkung: Ein Intervall ist naß, wenn in ihm Niederschlag registriert worden ist und an dem entsprechenden Tag zwischen 0:00 und 23:59h mindestens 0.2mm Niederschlag gemessen worden sind.

Tabelle 3.3: Auswirkung des Schwellenwertes

d	Anteil Menge	Anteil Tage
0.2	0.17	7.04
0.3	0.45	12.04
0.4	0.76	16.40
0.5	1.13	20.14
0.6	1.55	23.45

1. Spalte: Mittlerer Anteil [%] der Niederschlagsmenge an Tagen t mit mit $0 < Y_t \leq d$ an der Jahresniederschlagsmenge.
2. Spalte: Mittlerer Anteil [%] der Tage t mit $0 < Y_t \leq d$ an allen nassen Tagen im Jahr.

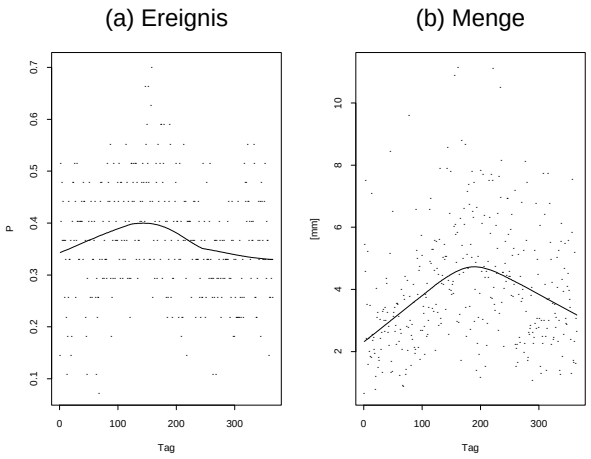
---

2 Auswahl homogener Perioden

Aus physikalischen Gründen sollte der Niederschlagsprozeß im Jahresverlauf zyklisch sein und die Beobachtungen unterschiedlicher Jahre sollten als Realisationen des

gleichen stochastischen Phänomenes betrachtet werden können. Wir vernachlässigen also eventuelle langfristige Trends oder Zyklen, die durch Veränderungen des Klimas hervorgerufen werden könnten ebenso wie die Verschiebungen, die durch das Einfügen des 29. Februars in Schaltjahren ausgeglichen werden. Wir werden also für alle Beobachtungen $Y_{t,n}$, des Tages $t = 1, \ldots, T$ in den Jahren $n = 1, \ldots, N$ annehmen, daß $\mathrm{E}Y_{t,n} = \mu_t$ gilt bzw. $\mathrm{E}W_{t,n}(d) = m_t$. Ebenfalls aus physikalischen Gründen sollten sich die Parameter eines klimatologischen Prozesses im Jahresverlauf stetig ändern. Der Niederschlagsprozeß innerhalb eines Jahres ist kein stationärer Prozeß, beispielsweise verändern sich im Jahresverlauf die empirische Wahrscheinlichkeit $\hat{\pi}_1(t)$ eines nassen Tages t ebenso wie die mittlere Niederschlagsmenge μ_t an einem nassen Tag t, vergl. Abb. 3.6.

Abbildung 3.6: Stationarität der Beobachtungen



(a) Empirische Wahrscheinlichkeit $\hat{\pi}_1(t)$ eines nassen Tages t.

(b) Empirisches Niederschlagsmittel $m(t)$ an Tag t, falls Tag t naß ist.

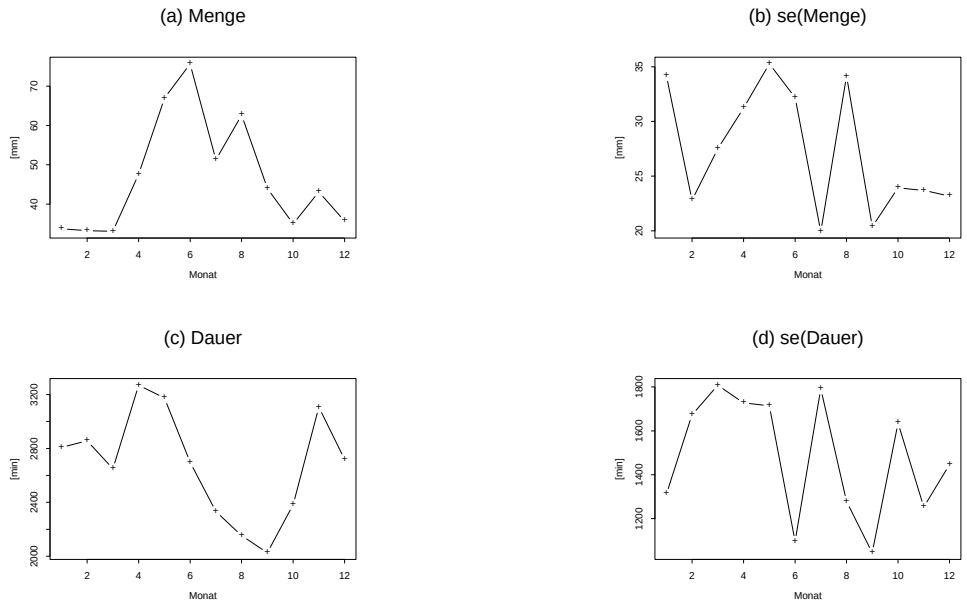
Die eingezeichneten Linien sind robuste Glättungen der Daten (S-PLUS-Befehl: `lowess`).

Theoretische Resultate über stochastische Prozesses basieren häufig auf der Annahme, der betrachtete Prozeß sei stationär. Unter dieser Annahme werden Parameterschätzer hergeleitet, die eine Glättung der Beobachtungen sind, wenn letztere nicht stationär sind. Wird beispielsweise die mittlere Niederschlagsmenge an einem nassen Tag t als konstant im Jahresverlauf angenommen, so stellt das bei den Stuttgarter Daten eine starke Glättung dar.

Für einige Schritte der nachfolgenden Datenanalyse ist die Annahme notwendig, der Niederschlagsprozeß sei stationär. In diesen Fällen werden die Beobachtungen in Perioden eingeteilt, welche separat analysiert werden. Solche Perioden werden als homogene Perioden bezeichnet (abgeleitet vom Begriff eines homogenen Punktprozesses). Wir werden einen 30 Tage umfassenden Zeitraum als homogene Periode betrachten und dabei im Gedächtnis behalten, daß diese Annahme eine Glättung der Realität darstellt.

Zur Überprüfung, ob eine Einteilung des Jahres in längere homogene Perioden sinnvoll ist, berechnen wir für jeden Monat:

Abbildung 3.7: Monatsniederschlag



(a), (b): Mittel und Standardabweichung des Monatsniederschlages.
(c), (d): Mittel und Standardabweichung der monatlichen Niederschlagsdauer.

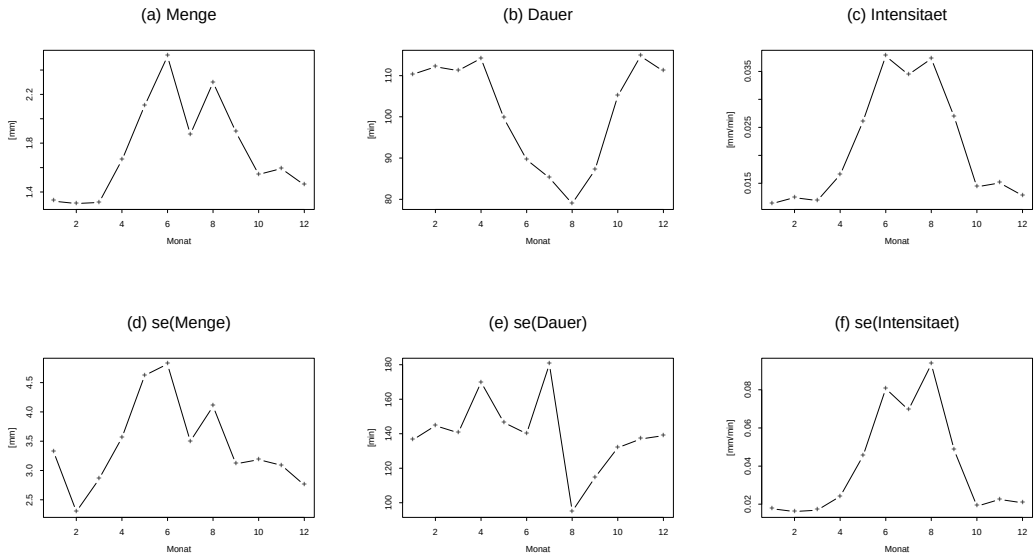
---

1. Mittel $\bar{Y}_M$ und emp. Standardabweichung $s(Y_M)$ des Monatsniederschlages,
2. Mittel $\bar{Y}_D$ und emp. Standardabweichung $s(Y_D)$ der Monatsniederschlagsdauer,
3. Mittel $\bar{Y}_E$ und emp. Standardabweichung $s(Y_E)$ der Ereignismenge,
4. Mittel $\bar{Y}_D$ und emp. Standardabweichung $s(D_E)$ der Ereignisdauer,
5. Mittel $\bar{I}_E$ und emp. Standardabweichung $s(I_E)$ der Ereignisintensität.

Die Ereignisintensität I_E ist der Quotient aus Ereignismenge und Ereignisdauer, $I_E = Y_E/D_E$. Sie ist ein korrekt wiederzugebendes Charakteristikum stündlichen und hochaufgelösten Niederschlags. Die Schätzer dieser Eigenschaften sind in Tab. 3.4, 3.5 und in Abb. 3.7, 3.8 dargestellt.

Eine Einteilung des Beobachtungszeitraumes in Perioden, die mehr als einen Kalendermonat umfassen, ist schwierig, da jede der betrachteten Eigenschaften zu einer anderen Einteilung führt. Im Hinblick auf Monatsniederschlag, Ereignisdauer und -intensität bilden die Monate Oktober - März eine homogene Saison, in bezug auf die Ereignismenge, -dauer und -intensität die Monate Juni - August. Typisch für die Sommersaison sind kurze Regenereignisse, die mit einer hohen Niederschlagsintensität einhergehen. April und Mai sind Übergangsperioden, die in einer Saison zusammengefaßt werden können, auch wenn sich die betrachteten Merkmale in beiden Monaten unterscheiden. Die Eigenschaften des Niederschlagsprozesses im September ähneln denen im Oktober, ohne daß der September der Wintersaison zugeordnet werden kann. Damit ergibt sich die folgende Einteilung:

Abbildung 3.8: Regenereignisse



(a), (d): Mittel und Standardabweichung der Niederschlagsmenge Y_E pro Ereignis [mm].
(b), (e): Mittel und Standardabweichung der Niederschlagsdauer D_E pro Ereignis [min].
(c), (f): Mittel und Standardabweichung der Niederschlagsintensität I_E pro Ereignis [mm/min].

---

1. Periode 1 (Frühling): April, Mai,
2. Periode 2 (Sommer): Juni, Juli, August,
3. Periode 3 (Herbst): September, Oktober,
4. Periode 4 (Winter): November bis März.

Insgesamt ist das Konzept einer Einteilung des Jahres in homogene Perioden unbefriedigend, weil es zu Unstetigkeitsstellen an den Übergängen der Perioden führt, die meteorologisch nicht plausibel sind. Deshalb bevorzugen wir ein Modell zur Beschreibung des Niederschlagsprozesses mit sich stetig im Jahresverlauf ändernden Parametern.

3 Der Niederschlagsmengenprozeß

Das Ziel unserer Arbeit ist die Beschreibung stündlichen Niederschlages durch ein saisonales Neyman-Scott-Modell. Das Modell soll insbesondere Eigenschaften des Prozesses der aggregierten Niederschlagsmengen bei Aggregation über unterschiedlich lange Zeitintervalle adäquat wiedergeben.

Sei $Y_i^{(h)}$ die Niederschlagsmenge im i-ten Intervall der Länge h, beispielsweise $h = 1, 3, 6, 12, 24$ Stunden. Innerhalb eines Zeitraumes, in dem der Niederschlagsprozeß

Tabelle 3.4: Monatsniederschlag

	Y [mm]	$s(Y)$ [mm]	D [Std:min]	$s(D)$ [Std:min]
Jan.	33.70	34.21	46:47	21:52
Feb.	33.24	22.81	47:37	27:53
März	33.05	27.51	44:09	30:08
April	47.57	31.23	54:29	28:47
Mai	66.89	35.26	52:55	28:34
Juni	75.67	32.19	44:58	18:16
Juli	51.26	19.95	38:52	29:51
Aug.	62.63	34.13	35:52	21:19
Sep.	43.99	20.38	33:46	17:25
Okt.	34.99	23.92	39:42	27:16
Nov.	43.09	23.70	51:46	20:53
Dez.	35.67	23.18	45:20	24:06

Y Monatsniederschlag, D Monatsniederschlagsdauer

homogen ist, sind die $Y_i^{(h)}$ identisch verteilte Zufallsvariable. Wir betrachten alle Intervalle, damit ist $Y_i^{(h)} = 0$ möglich.
Abkürzend verwenden wir in Anlehnung an COWPERTWAIT et al. [1996a] die folgenden Bezeichnungen:

$$\begin{aligned} \mu(h) &:= \mathrm{E}(Y_i^{(h)}) & \gamma(h) &:= \mathrm{Var}(Y_i^{(h)}) \\ \gamma(h,j) &:= \mathrm{Cov}(Y_i^{(h)}, Y_{i+j}^{(h)}) & \rho(h,j) &:= \mathrm{Cor}(Y_i^{(h)}, Y_{i+j}^{(h)}) \\ \phi(h) &:= P(Y_i^{(h)} = 0) & \phi_{DD}(h) &:= P(Y_i^{(h)} = 0 | Y_{i-1}^{(h)} = 0) \\ \phi_{WW}(h) &:= P(Y_i^{(h)} > 0 | Y_{i-1}^{(h)} > 0) \end{aligned}$$

Um Saisonalität zu berücksichtigen, schätzen wir diese Eigenschaften für jeden Monat separat. Der Index k (z.B. $\mu_k(h)$) bezeichnet den k-ten Kalendermonat und wird ausgelassen, wenn Verwechslungen ausgeschlossen sind. Mit $Y_{i,k,n}^{(h)}$ wird die Niederschlagsmenge im i-ten Intervall der Länge h Stunden in Monat $k = 1, \ldots, 12$ des Jahres $n = 1, \ldots, N$ bezeichnet. Wir verwenden die folgenden Schätzer:

$$\textbf{(3.1)} \quad \hat{\mu}_k(h) = \frac{1}{\tilde{n}} \sum_{n=1}^{N} \sum_{i=1}^{n_k(h)} Y_{i,k,n}^{(h)} \ ,$$

wobei $n_k(h)$ die Anzahl der h-Stunden-Intervalle im Kalendermonat k ist und $\tilde{n} = n_k(h) \cdot N$. Fehlende Werte gibt es in diesem Datensatz nicht. Die Intervalle werden fortlaufend und ohne Überschneidungen aggregiert, beginnend mit dem 1. Tag des Beobachtungszeitraumes um 0:00 Uhr. Bei Aggregation zu 3-Stunden-Intervallen werden die Stunden 0 - 3h, 3 - 6h, 6 - 9h, ... zusammengefaßt. Aggregation zu 12-Stunden-Intervallen faßt die Messungen Intervalle 0 - 12h und 12 - 24h zusammen, für die Aggregation zu 24-Stunden-Intervallen werden die Intervalle von jeweils 0 - 24h zusammengefaßt. In dieser Notation bezeichnet beispielsweise $Y_{3,1,1}^{(12)}$ die Niederschlagsmenge im 3. 12-Stunden-Intervall in Monat $k = 1$ und Jahr $n = 1$, das ist der 2. Januar, 0:00 - 11:59 h.

Tabelle 3.5: Niederschlagsereignisprozeß

	Y_E	$s(Y_E)$	D_E	$s(D_E)$	I_E	$s(I_E)$	N
Jan.	1.32	3.32	1:50	2:16	0.01	0.02	25.44
Feb.	1.31	2.30	1:52	2:25	0.01	0.02	25.48
März	1.32	2.86	1:51	2:21	0.01	0.02	50.93
April	1.66	3.56	1:54	2:50	0.02	0.02	28.63
Mai	2.11	4.62	1:40	2:27	0.03	0.05	31.85
Juni	2.51	4.82	1:30	2:20	0.04	0.08	30.11
Juli	1.87	3.50	1:25	3:00	0.03	0.07	27.37
Aug.	2.30	4.10	1:19	1:35	0.04	0.09	27.26
Sep.	1.89	3.12	1:27	1:55	0.03	0.05	23.26
Okt.	1.54	3.18	1:45	2:12	0.01	0.02	22.67
Nov.	1.59	3.09	1:55	2:17	0.02	0.02	27.07
Dez.	1.46	2.75	1:51	2:19	0.01	0.02	24.48

Spalten 1 und 2: Mittel und Standardabweichung der Ereignismenge Y_E [mm].
Spalten 3 und 4: Mittel und Standardabweichung der Ereignisdauer D_E [Std:min].
Spalten 5 und 6: Mittel und Standardabweichung der Ereignisintensität I_E [mm/min].
Spalte 7: Mittlere Ereignisanzahl N.

Der Schätzer der empirischen Varianzen ist

$$\textbf{(3.2)} \quad \hat{\gamma}_k(h) = \frac{}{\tilde{n}-1} \sum_{n=1}^{N} \sum_{i=1}^{n_k(h)} \left(Y_{i,k,n}^{(h)} - \hat{\mu}_k(h) \right)^2 .$$

Die empirischen Kovarianzen und Korrelationen werden gemäß den folgenden Formeln berechnet:

$$\textbf{(3.3)} \quad \hat{\gamma}_k(h,j) = \frac{1}{\tilde{n}-j-1} \sum_{n=1}^{N} \sum_{i=j+1}^{n_k(h)} \left(Y_{i,k,n}^{(h)} - \hat{\mu}_k^{(1)}(h) \right) \left(Y_{i+j,k,n}^{(h)} - \hat{\mu}_k^{(2)}(h) \right)$$

mit

$$\hat{\mu}_k^{(1)}(h) = \frac{1}{\tilde{n}-j} \sum_{n=1}^{N} \sum_{i=1}^{n_k(h)-j} Y_{i,k,n}^{(h)} , \quad \hat{\mu}_k^{(2)}(h) = \frac{1}{\tilde{n}-j} \sum_{n=1}^{N} \sum_{i=j+1}^{n_k(h)-j} Y_{i,k,n}^{(h)}$$

$$\textbf{(3.4)} \quad \hat{\rho}_k(h,j) = \frac{\hat{\gamma}_k(h,j)}{\sqrt{\hat{\gamma}_k^{(1)}(h)\, \hat{\gamma}_k^{(2)}(h)}} ,$$

wobei $\hat{\gamma}_k^{(1)}(h)$ der Varianzschätzer aus den ersten $\tilde{n}-j$ und $\hat{\gamma}_k^{(1)}(h)$ der Varianzschätzer aus den letzten $\tilde{n}-j$ beobachteten Intervallen ist.

Die empirischen Wahrscheinlichkeiten $\hat{\phi}(h), \hat{\phi}_{WW}(h)$ und $\hat{\phi}_{DD}(h)$ berechnen sich gemäß:

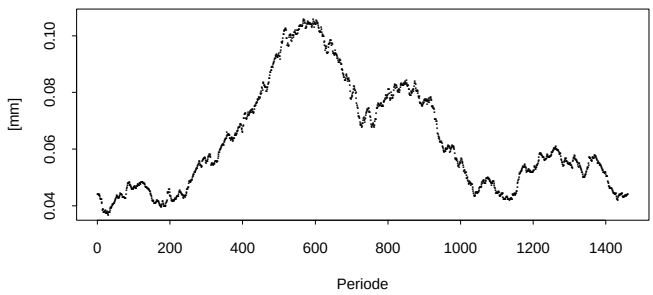
$$\textbf{(3.5)} \quad \hat{\phi}(h) = \frac{1}{\tilde{n}} \sum_{n=1}^{N} \sum_{i=1}^{n_k(h)} 1(Y_{i,k,n}^{(h)} = 0) \ ,$$

$$\textbf{(3.6)} \quad \hat{\phi}_{WW}(h) = \sum_{n=1}^{N} \sum_{i=2}^{n_k(h)} 1(Y_{i,k,n}^{(h)} > 0) \cdot 1(Y_{i-1,k,n}^{(h)} > 0) \Big/ \sum_{n=1}^{N} \sum_{i=1}^{n_k(h)-1} 1(Y_{i,k,n}^{(h)} > 0) \ ,$$

$$\textbf{(3.7)} \quad \hat{\phi}_{DD}(h) = \sum_{n=1}^{N} \sum_{i=2}^{n_k(h)} 1(Y_{i,k,n}^{(h)} = 0) \cdot 1(Y_{i-1,k,n}^{(h)} = 0) \Big/ \sum_{n=1}^{N} \sum_{i=1}^{n_k(h)-1} 1(Y_{i,k,n}^{(h)} = 0) \ ,$$

wobei $1(\cdot)$ die Indikatorfunktion ist. Die Werte dieser Schätzer sind für die Kalendermonate und die Aggregationsniveaus $h = 1, 3, 6, 12, 24$ in Tab. 3.6 aufgelistet. Sie sind unter Beachtung des Schwellenwertes von $d = 0.2$mm für täglichen Niederschlag berechnet worden.

Abbildung 3.9: Empirisches Ein-Stunden-Mittel



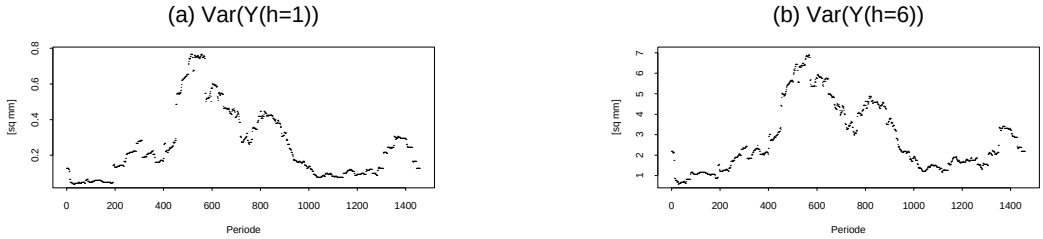
Schätzer $\hat{\mu}(1)$ für alle im Abstand von 6 Stunden beginnenden 30-Tages-Perioden des Jahres.

In der Entwicklung eines Modelles für stündlichen Niederschlag werden wir die empirischen Momente verwenden, die aus allen 30 Tage umfassenden und im Abstand von 6 Stunden jeweils um 0:00h, 6:00h, 12:00h und 18:00h beginnenden Perioden geschätzt werden. Auf diesen Schätzern wird die Modellierung stündlichen Niederschlages in Kapitel 5 basieren. Die Schätzer sind in den Abb. 3.9 bis 3.16 dargestellt. Um die Darstellung griechischer Buchstaben mit S-PLUS zu vermeiden, weichen die Bezeichnungen in den S-PLUS-Bildüberschriften von den Bilduntertiteln ab. Das Verständnis wird dadurch - so hoffen wir - nicht erschwert.
Das h-Stunden Mittel ist nach Konstruktion des Schätzers proportional zur Länge des Intervalles: $\hat{\mu}(j\,h) = j\,\hat{\mu}(h)$, $j \in I\!R_+$. Daher wird nur $\hat{\mu}(1)$ betrachtet. Die empirischen Varianzen nehmen überproportional, aber nicht quadratisch zu, in etwa gemäß der Relation $\hat{\gamma}(j\,h) = \sqrt{2}\,j\,\hat{\gamma}(h)$. Die Lag(1)-Autokorrelationen $\hat{\rho}(h,1)$ des h-Stunden-Niederschlages sind in den meisten Perioden für $h = 1$ am größten und fallen monoton mit zunehmender Länge h des Aggregationsintervalles ab. Allerdings deuten der geringe Rückgang sowie der in einigen Monaten auch zu beobachtende Anstieg von $\hat{\rho}(12,1)$ auf $\hat{\rho}(24,1)$ auf innertägliche Zyklen hin, vergl. Tab. 3.6.

Die Übergangswahrscheinlichkeiten $\hat{\phi}_{WW}(h)$ nasser Intervalle sind erwartungsgemäß mit zunehmender Länge h ansteigend. Etwa 90% aller beobachteten Stunden sind trocken, aber nur etwa 65% aller beobachteten Tage, was auf ein Überwiegen kurzer Regenschauer hindeutet.

Im folgenden Kapitel 4 werden wir an die Niederschlagsmessungen ein Modell zur Beschreibung des 24-Stunden-Niederschlages anpassen, bevor wir in Kapitel 5 ein Modell betrachten, das Eigenschaften des über unterschiedlich lange Intervalle aggregierten Niederschlages unter Berücksichtigung saisonaler Variation beschreibt.

Abbildung 3.10: Empirische Varianzen



Varianz des 1-Stunden- bzw. des 6-Stunden-Niederschlages. (a) $\hat{\gamma}(1)$, (b) $\hat{\gamma}(6)$.

Abbildung 3.11: Empirische Varianzen

Varianz des 12-Stunden- bzw. des 24-Stunden-Niederschlages. (a) $\hat{\gamma}(12)$, (b) $\hat{\gamma}(12)$.

Tabelle 3.6: Empirische Momente des aggregierten Niederschlagsprozesses

Erwartungswert und Varianzen

	$\hat{\mu}(1)$	$\hat{\gamma}(1)$	$\hat{\gamma}(3)$	$\hat{\gamma}(6)$	$\hat{\gamma}(12)$	$\hat{\gamma}(24)$
Jan.	0.045	0.134	0.92	2.20	5.05	12.97
Feb.	0.049	0.068	0.45	1.30	3.50	9.46
März	0.045	0.170	0.64	1.45	3.54	8.62
April	0.066	0.213	0.92	2.39	6.05	14.98
Mai	0.090	0.622	2.38	5.43	13.22	27.24
Juni	0.106	0.610	2.58	6.02	14.06	30.87
Juli	0.069	0.276	1.36	3.31	8.23	17.55
Aug.	0.085	0.427	1.79	4.59	11.10	23.35
Sep.	0.061	0.169	0.80	2.22	5.64	13.19
Okt.	0.047	0.010	0.55	1.53	4.35	10.18
Nov.	0.059	0.107	0.70	1.86	5.25	13.15
Dez.	0.051	0.245	0.92	2.10	5.05	11.27

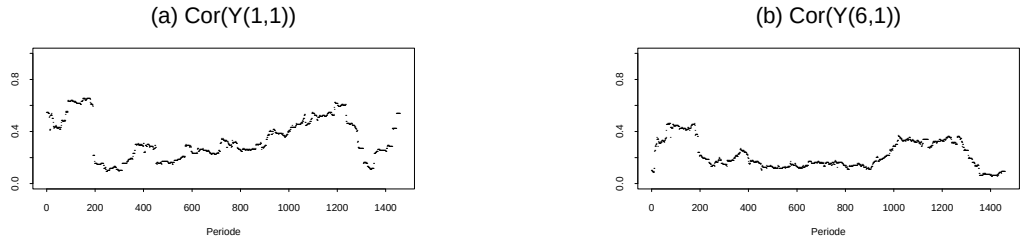
Korrelationen

	$\hat{\rho}(1,1)$	$\hat{\rho}(3,1)$	$\hat{\rho}(6,1)$	$\hat{\rho}(12,1)$	$\hat{\rho}(24,1)$
Jan.	0.56	0.18	0.11	0.40	0.32
Feb.	0.66	0.52	0.44	0.36	0.23
März	0.14	0.15	0.17	0.14	0.10
April	0.25	0.27	0.24	0.24	0.19
Mai	0.18	0.18	0.15	0.11	0.21
Juni	0.25	0.17	0.14	0.09	0.07
Juli	0.36	0.20	0.18	0.06	0.05
Aug.	0.28	0.23	0.16	0.07	0.13
Sep.	0.40	0.34	0.23	0.16	0.16
Okt.	0.51	0.41	0.34	0.18	0.12
Nov.	0.62	0.40	0.36	0.23	0.19
Dez.	0.13	0.16	0.15	0.10	0.11

Übergangswahrscheinlichkeiten nasser Intervalle

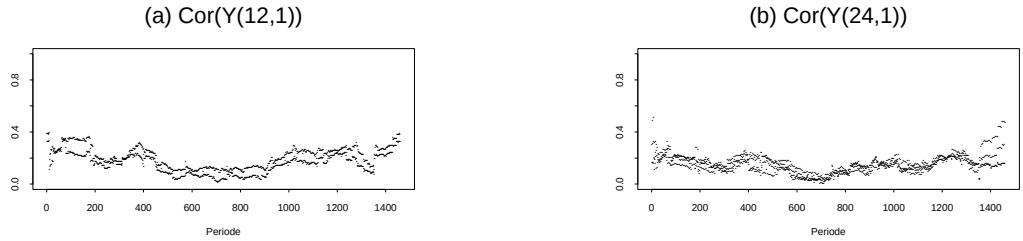
	$\hat{\phi}_{WW}(1)$	$\hat{\phi}_{WW}(3)$	$\hat{\phi}_{WW}(6)$	$\hat{\phi}_{WW}(12)$	$\hat{\phi}_{WW}(24)$
Jan.	0.16	0.24	0.33	0.44	0.58
Feb.	0.18	0.27	0.35	0.43	0.56
März	0.13	0.21	0.27	0.39	0.51
April	0.15	0.23	0.33	0.44	0.60
Mai	0.14	0.21	0.29	0.39	0.58
Juni	0.12	0.19	0.26	0.38	0.55
Juli	0.14	0.18	0.25	0.34	0.53
Aug.	0.14	0.21	0.29	0.37	0.52
Sep.	0.10	0.17	0.24	0.33	0.47
Okt.	0.13	0.21	0.29	0.43	0.57
Nov.	0.16	0.24	0.31	0.44	0.58
Dez.	0.12	0.19	0.28	0.39	0.52

Abbildung 3.12: Empirische Lag(1)-Autokorrelationen



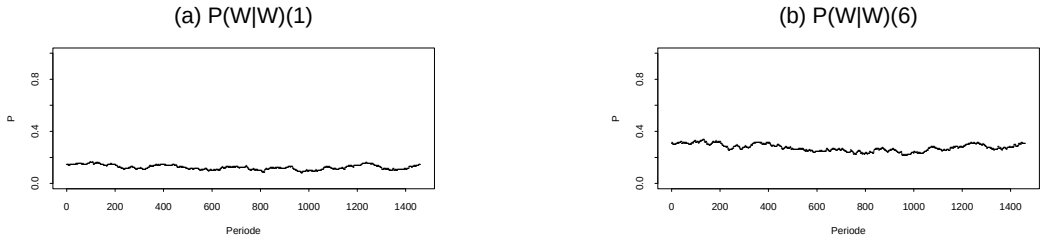
Lag(1)-Autokorrelation des 1-Stunden-bzw. 6-Stunden-Niederschlages. (a) $\hat{\rho}(1,1)$, (b) $\hat{\rho}(6,1)$.

Abbildung 3.13: Empirische Lag(1)-Autokorrelationen



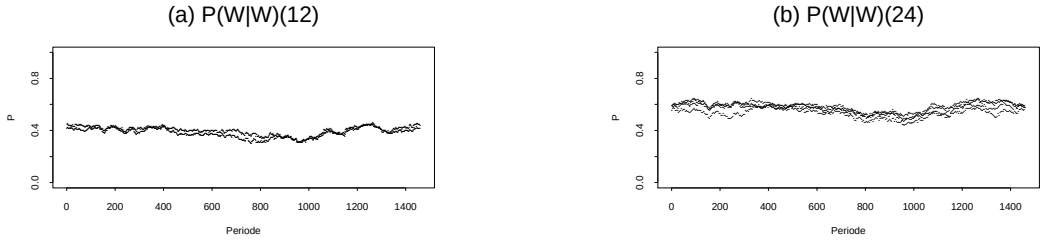
Lag(1)-Autokorrelation des 12-Stunden- bzw. 24-Stunden-Niederschlages. (a) $\hat{\rho}(12,1)$, (b) $\hat{\rho}(24,1)$.

Abbildung 3.14: Empirische Übergangswahrscheinlichkeiten



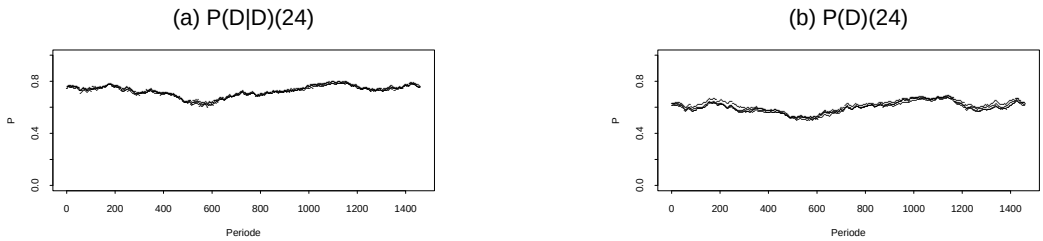
Übergangswahrscheinlichkeiten $\hat{\phi}_{WW}(h)$ für (a) 1-Stunden- bzw. (b) 6-Stunden-Intervalle.

Abbildung 3.15: Empirische Übergangswahrscheinlichkeiten



Übergangswahrscheinlichkeiten $\hat{\phi}_{WW}(h)$ für (a) 12-Stunden- bzw. (b) 24-Stunden-Intervalle.

Abbildung 3.16: Empirische Wahrscheinlichkeiten trockener Intervalle



(a) Übergangswahrscheinlichkeit $\hat{\phi}_{DD}(24)$ sowie (b) Wahrscheinlichkeit $\hat{\phi}(24)$

Fortsetzung Tabelle 3.6

(Übergangs-) Wahrscheinlichkeiten trockener Intervalle

	$\hat{\phi}_{DD}(1)$	$\hat{\phi}_{DD}(6)$	$\hat{\phi}_{DD}(12)$	$\hat{\phi}_{DD}(24)$	$\hat{\phi}_{D}(1)$	$\hat{\phi}_{D}(24)$
Jan.	0.93	0.86	0.81	0.77	0.92	0.64
Feb.	0.92	0.85	0.80	0.75	0.91	0.64
März	0.93	0.85	0.80	0.75	0.93	0.66
April	0.92	0.83	0.79	0.74	0.91	0.60
Mai	0.92	0.81	0.73	0.66	0.91	0.56
Juni	0.92	0.80	0.73	0.63	0.92	0.55
Juli	0.93	0.85	0.79	0.74	0.93	0.64
Aug.	0.94	0.86	0.79	0.73	0.93	0.64
Sep.	0.94	0.86	0.82	0.75	0.94	0.68
Okt.	0.94	0.88	0.84	0.80	0.93	0.69
Nov.	0.92	0.83	0.79	0.75	0.91	0.63
Dez.	0.93	0.85	0.80	0.76	0.93	0.67

Kapitel 4

Modellierung täglichen Niederschlages

Zur Modellierung täglichen und stündlichen Niederschlages wurde ursprünglich der Ansatz verfolgt, zunächst den Ereignis- und Mengenprozeß täglichen Regens zu beschreiben und darauf basierend ein Modell für die bedingte Verteilung der Niederschläge innerhalb eines nassen Tages zu entwickeln. Dieser Ansatz wurde nur bis zur Modellierung täglichen Regens umgesetzt, weil die Beschreibung sowohl täglichen als auch stündlichen Regens durch das gleiche Modell, nämlich das in Kapitel 5 beschriebene Neyman-Scott-Modell, überzeugender erschien. Vorgehensweise sowie Ergebnisse der Modellierung täglichen Regens werden hier dennoch dargestellt.
Als täglichen Regen bezeichnen wir die Serie der von 0:00 - 23:59h aggregierten Niederschlagsmengen. Die Niederschlagsmengen an nassen Tagen sind schwach korreliert mit $r(1) = 0.14$, vergl. Abschnitt 3.1. Das ist so gering, daß wir den Ereignis- und den Mengenprozeß separat anpassen.

In Abschnitt 1 wird der Ereignisprozeß modelliert, in Abschnitt 2 die bedingte Verteilung der Niederschlagsmengen an nassen Tagen. In Abschnitt 3 wird der Simulationsalgorithmus beschrieben, bevor in Abschnitt 4 im Rahmen der Validation ein Modell ausgewählt wird.

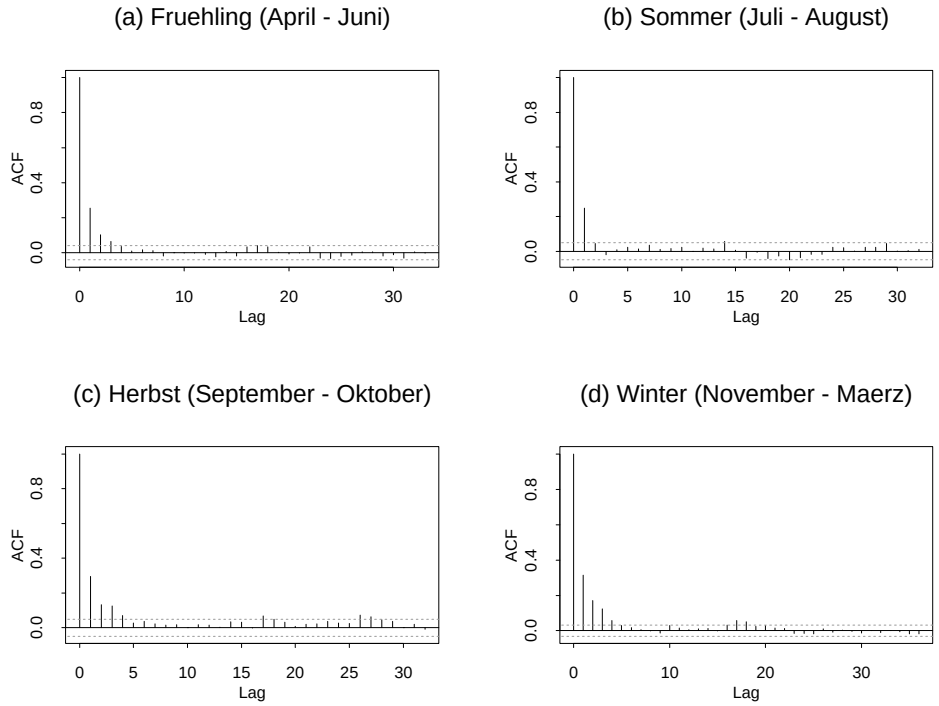
1 Der Ereignisprozeß

Der Ereignisprozeß täglichen Niederschlages ist kein Bernoulliprozeß, was aus dessen empirischer Autokorrelationsfunktion deutlich wird: Der empirische Lag(1)-Autokorrelationskoeffizient $r(1)$ ist in allen Saisonen signifikant von Null verschieden, je nach Jahreszeit auch gilt das auch für $r(2)$ und $r(3)$, vergl. Tab. 4.1 und Abb. 4.1. Markovketten sind ein naheliegendes Modell zur Beschreibung des Ereignisprozesses, weil sie leicht auf Markovketten höherer Ordnung verallgemeinert werden können, die Modellparameter leicht zu schätzen sind und weil Saisonalität in das Modell integriert werden kann. Die Anpassung des Markovkettenmodelles folgt der Vorgehensweise in Zucchini und Adamson [1984].

Tabelle 4.1: ACF des Ereignisprozesses

Lag	Frühling April - Juni	Sommer Juli - Aug.	Herbst Sept. - Okt.	Winter Nov. - März
1	0.26	0.25	0.30	0.32
2	0.10	0.05	0.13	0.17
3	0.06	-0.02	0.13	0.12
4	0.04	0.01	0.07	0.06
5	0.01	0.02	0.03	0.03

Abbildung 4.1: ACF des Ereignisprozesses



Die horizontalen Linien markieren 95%-Konfidenzintervalle für $\hat{\rho}(j)$ um Null.
Frühling: [-0.04; 0.04]; Sommer: [-0.05; 0.05]; Herbst: [-0.05; 0.05]; Winter:[-0.03; 0.03].

Es sei $Z_{t,n}$ der Zustand des Ereignisprozesses täglichen Regens an Tag t in Jahr n,

$$Z_{t,n} = \left\{ \begin{array}{ll} 1, & \text{Tag } t \text{ in Jahr } n \text{ ist naß,} \\ 0, & \text{sonst,} \end{array} \right\} \; t = 1, \ldots T, \; n = 1, \ldots, N.$$

Bei uns ist $T = 365$. Nehmen wir zunächst an, die Anzahl $NR(t)$ der nassen Tage t innerhalb von N Jahren sei unabhängig für $t \neq s$. Dann ist $NR(t)$ binomialverteilt mit Erfolgswahrscheinlichkeit $\pi_1(t) = \mathrm{P}(Z_t = 1)$. Der empirische Schätzer $\hat{\pi}_1(t)$ ist

$$\hat{\pi}_1(t) := \frac{1}{N} \sum_{n=1}^{N} Z_{t,n} \; .$$

Genauso ist die Anzahl $NRR(t)$ der in N Jahren aufgetretenen Folgen $(Z_{t-1,n} = 1, \; Z_{t,n} = 1)$ binomialverteilt mit Erfolgswahrscheinlichkeit $\pi_{11}(t) = \mathrm{P}(Z_{t,n} = 1 | Z_{t-1,n} = 1)$ und Anzahl der Versuche $NR(t-1) = \sum_{n=1}^{N} 1(Z_{t-1,n}$:

$$NRR(t) \sim B\Big(\pi_{11}(t), NR(t-1))\Big) \; ,$$

wobei $1(\cdot)$ die Indikatorfunktion ist. Dieser Datensatz enthält keine fehlenden Werte, daher existiert $Z_{t,n}$ für alle t und für alle n. Die $NRR(t)$ sind bedingt unabhängig, gegeben $NR(t-1)$. Der empirische Schätzer $\hat{\pi}_{11}(t)$ der Übergangswahrscheinlichkeit $\pi 11(t)$ ist

$$\hat{\pi}_{11}(t) := \frac{NRR(t)}{NR(t-1)}$$

Die Übergangswahrscheinlichkeiten $\pi_{00}(t) = \mathrm{P}(Z_{t,n} = 0 | \; Z_{t-1,n} = 0)$ oder die Übergangswahrscheinlichkeiten einer Markovkette zweiter Ordnung werden analog geschätzt.

Wir übernehmen die Bezeichnungen aus ZUCCHINI UND ADAMSON [1984]:

$$\begin{array}{lcll} NR(t) & = & \sum_{n=1}^{N} 1(Z_{t,n}), & t = 1, \ldots, T \\ ND(t) & = & \sum_{n=1}^{N} 1(Z_{t,n} = 0), & t = 1, \ldots, T \\ NRR(t) & = & \sum_{n=1}^{N} 1(Z_{t-1,n} = 0, \; Z_{t,n} = 1), & \\ NDR(t) & = & \sum_{n=1}^{N} 1(Z_{t-1,n} = 0, \; Z_{t,n} = 1), & \\ NRRR(t) & = & \sum_{n=1}^{N} 1(Z_{t-2,n} = 1, \; Z_{t-1,n} = 1, \; Z_{t,n} = 1), & \\ NDRR(t) & = & \sum_{n=1}^{N} 1(Z_{t-2,n} = 0, \; Z_{t-1,n} = 1, \; Z_{t,n} = 1), & \\ NRDR(t) & = & \sum_{n=1}^{N} 1(Z_{t-2,n} = 1, \; Z_{t-1,n} = 0, \; Z_{t,n} = 1), & \\ \ldots & & & \end{array}$$

wobei T die Länge des Beobachtungszeitraumes (z.B. T=365) und N die Anzahl der beobachteten Jahre bezeichnen.

Entsprechend seien

$$\begin{aligned}
\pi_1(t) &= \mathrm{P}(Z_{t,n}=1), \qquad t=1,\ldots,T,\ n=1,\ldots,N \\
\pi_0(t) &= \mathrm{P}(Z_{t,n}=0), \\
\pi_{11}(t) &= \mathrm{P}(Z_{t,n}=1|Z_{t-1,n}=1), \\
\pi_{01}(t) &= \mathrm{P}(Z_{t,n}=1|Z_{t-1,n}=0), \\
\pi_{111}(t) &= \mathrm{P}(Z_{t,n}=1|Z_{t-1,n}=1, Z_{t-2,n}=1), \\
\pi_{011}(t) &= \mathrm{P}(Z_{t,n}=1|Z_{t-1,n}=1, Z_{t-2,n}=0), \\
\pi_{101}(t) &= \mathrm{P}(Z_{t,n}=1|Z_{t-1,n}=0, Z_{t-2,n}=1), \\
\pi_{001}(t) &= \mathrm{P}(Z_{t,n}=1|Z_{t-1,n}=0, Z_{t-2,n}=0), \\
&\text{etc.}
\end{aligned}$$

Die empirischen Schätzer dieser Wahrscheinlichkeiten werden zwar aus unterschiedlichen Daten, aber auf analoge Weise bestimmt, zum Beispiel:

$$\hat{\pi}_{01}(t) = \frac{NDR(t)}{ND(t-1)} \ .$$

Deshalb verwenden wir vereinfachend die folgende Notation:

$M(t)$: Wert der binomialverteilten Zufallsvariablen an Tag t (z.B. $M(t) = NRR(t)$),
$MM(t)$: Anzahl der Versuche für $M(t)$,
$\pi(t)$: Erfolgswahrscheinlichkeit von $M(t)$,
$\hat{\pi}(t)$: empirischer Schätzer von $\pi(t)$.

Es ist

$$M(t) \sim B\Big(\pi(t), MM(t)\Big),$$

und

$$\hat{\pi}(t) = M(t)/MM(t) \ ,$$

Die $M(t)$ sind bedingt unabhängig, gegeben $MM(t)$, $t=1,\ldots,T$. Ferner gelten die Relationen:

$$\begin{aligned}
\pi_1(t) + \pi_0(t) &= 1\,, \\
\pi_{11}(t) + \pi_{10}(t) &= 1\,, \\
\pi_{111}(t) + \pi_{110}(t) &= 1\,.
\end{aligned}$$

und allgemein $\pi_{mm1}(t) + \pi_{mm0)}(t) = 1$, wobei mm eine beliebige (auch beliebig lange) Folge von 0 und 1 ist, die bei beiden Summanden gleich ist.

Wird an den Ereignisprozeß in einem T Tage umfassenden Zeitraum auf diese Weise eine Markovkette 1. Ordnung angepaßt, so müssen $2T$ Parameter geschätzt werden, nämlich $\pi_{11}(t)$ und $\pi_{01}(t)$ für $t = 1, \ldots, T$. Für eine Markovkette 2. Ordnung sind es bereits $2^2 T$ Parameter. Das führt einerseits zu einem großen Fehler durch Schätzung, zum anderen würden sich die Übergangswahrscheinlichkeiten $\hat{\pi}(t)$ im Zeitverlauf nicht stetig ändern, was meteorologisch nicht plausibel ist. Die Parameter eines klimatologischen Prozesses sollten stetige, zyklische Funktionen der Zeit mit der Periode eines Jahres sein, damit die beobachtbaren Eigenschaften wiedergegeben werden. Außerdem müssen die Wahrscheinlichkeiten $\pi(t)$ für alle Folgen $M(t)$ strikt zwischen 0 und 1 liegen, was die empirischen Schätzer der Parameter von Markovketten insbesondere höherer Ordnung nicht gewährleisten: für sie treten häufig die Werte 0 oder 1 auf. Diese Eigenschaften legen eine Transformation der Übergangswahrscheinlichkeiten und ansschließende Anpassung von Fouriersummen nahe. Zur Glättung zyklischer Daten sind Fouriersummen naheliegend, da sie stetig sind und die gewünschte Periodizität ermöglichen. Aus der Fourieranalyse ist bekannt, daß T Werte $x(1), \ldots, x(T)$ durch

$$\textbf{(1.1)} \quad x(t) = R_0 + \sum_{0<k<T/2} R_k \cos(\omega_k t + \theta_k) + (-1)^t R_{T/2}\,, t = 1, \ldots, T,$$

dargestellt werden können, wobei $\omega_k = 2\pi k/T$. Der letzte Summand tritt nur auf, wenn T gerade ist. Bei uns ist $T = 365$, so daß wir den letzten Summanden im folgenden nicht mehr anführen. R_k und θ_k sind die Amplitude bzw. Phase der Frequenz ω_k (BLOOMFIELD [1976]). Enthalten die $x(t)$ nur endlichviele Frequenzen $\omega_{k_j}, j = 1, \ldots, J$ und $k_j \in \{1, \ldots, T/2\}$, so genügt eine Beschreibung

$$\textbf{(1.2)} \quad x(t) = R_0 + \sum_{j=1}^{J} R_{k_j} \cos(\omega_{k_j} t + \theta_{k_j})$$

mit $0 \leq J < T/2$ und geeigneten Frequenzen ω_{k_j}. In klimatologischen, einmal täglich registrierten Daten dominieren die kleinen Frequenzen $\omega_k, k = 1, 2, \ldots$, so daß eine Approximation

$$\textbf{(1.3)} \quad x(t; L) = R_0 + \sum_{k=1}^{L} R_k \cos(\omega_k t + \theta_k)\,,$$

mit geeignetem L möglich ist. Diese Annahme muß durch eine Spektralanalyse bestätigt werden. Die Parameter eines linearen Gleichungssystemes können leichter geschätzt werden, daher formen wir (1.3) unter Anwendung der Additionstheoreme um:

$$\textbf{(1.4)} \quad x(t) = A_0 + \sum_{k=1}^{L} A_k \cos(\omega_k t) + B_k \sin(\omega_k t)\,, t = 1, \ldots, T.$$

Um zu gewährleisten, daß die angepaßten Wahrscheinlichkeiten strikt zwischen 0

und 1 liegen, werden statt der empirischen Wahrscheinlichkeitsschätzer $\hat{\pi}(t)$ deren Logits durch eine Fouriersumme approximiert. Die Logit-Transformation ist:

$$\eta(t) := logit(\pi(t)) := \log\left(\frac{\pi(t)}{1-\pi(t)}\right) .$$

Sie erhält Periodizitäten. Da $\pi(t) = \exp(\eta(t))/\Big(1+\exp(\eta(t))\Big)$, folgt für die Schätzer $0 < \hat{\pi}(t) < 1$. Wir erhalten

$$\textbf{(1.5)} \quad logit(\pi(t)) \approx R_0 + \sum_{k=1}^{L} R_k \cos\left(\omega_k t + \theta_k\right) , \; t = 1, \ldots, T,$$

bzw.

$$\textbf{(1.6)} \quad logit(\pi(t) \approx A_0 + \sum_{k=1}^{L} A_k \cos(\omega_k t) + B_k \sin(\omega_k t) \; , t = 1, \ldots, T$$

Damit liegt ein verallgemeinertes lineares Modell vor (McCullagh und Nelder [1983]), dessen Parameter $\hat{A}_0, \hat{A}_j, \hat{B}_j, j = 1, \ldots, L$, mit der Methode der iterativ gewichteten kleinsten Quadrate aus den empirischen Schätzern $\hat{\pi}(t)$ geschätzt werden. Als Startwerte verwenden wir die gewöhnlichen Kleinste-Quadrate-Schätzer:

$$\textbf{(1.7)} \qquad \hat{A}_0 = \frac{1}{T}\sum_{t=1}^{T} logit(\hat{\pi}(t)) \; ,$$

$$\begin{aligned} \hat{A}_k &= \frac{2}{T}\sum_{t=1}^{T} \Big(logit(\hat{\pi}(t)) - \hat{A}_0\Big) \cos(\omega_k t), \; 0 < k < T/2 \; , \\ \hat{B}_k &= \frac{2}{T}\sum_{t=1}^{T} \Big(logit(\hat{\pi}(t)) - \hat{A}_0\Big) \sin(\omega_k t), \; 0 < k < T/2 \; . \end{aligned}$$

Die Anpassung erfolgt in S-PLUS mit dem Befehl `glm`. Aus den geschätzten Werten $\hat{A}_j, \hat{B}_j$ werden die Amplitude $\hat{R}_j$ und die Phase $\hat{\theta}_j$ berechnet. Wir bestimmen zunächst die optimale Parameteranzahl L für jede der (Übergangs-)Wahrscheinlichkeiten eines Markovmodelles der Ordnungen 0, 1, 2 mit der Deviance.

(1.8) Definition. Die **Deviance** oder **Log-Likelihood-Ratio-Statistik** eines Modelles mit Parametervektor $\theta_k \in I\!R^k$ ist (vergl. McCullagh und Nelder [1989])

$$D_k := 2(l(\theta_{max}; y) - l(\theta_k; y)) \; ,$$

wobei θ_{max} das Modell mit maximaler Parameteranzahl, das ist hier T, bezeichnet.

□

Zur Auswahl zwischen ineinander geschachtelten Modellen mit q bzw. p Parametern, $q < p$, wird die Prüfgröße $D_q - D_p$ verwendet. Unter der Annahme, beide Modelle beschreiben die Daten angemessen, gilt approximativ $D_p \sim \chi^2_{T-p}$ bzw. $D_q - D_p \sim \chi^2_{p-q}$. Damit kann die Hypothese getestet werden, das Modell mit q Parametern beschreibe die Beobachtungen ebenso angemessen wie das Modell mit p Parametern.

(1.9) Parameteranzahl der Markovketten. Die Ergebnisse der Anpassung von Fouriersummen an die Übergangswahrscheinlichkeiten befinden sich in Anhang B.1. Die Anzahlen der zur Beschreibung des Ereignisprozesses durch eine Markovkette mit geglätteten Übergangswahrscheinlichkeiten insgesamt benötigten Parameter sind der folgenden Tabelle zu entnehmen:

Ordnung	0	1	2	3
Parameter	5	8	10	12

Die Ordnung der Markovkette wird unter Anwendung des Bayes'schen Informationskriteriums BIC und Akaikes' Informationskriteriums AIC ausgewählt. KATZ [1981] zeigt in einem Vergleich des BIC mit dem AIC, daß in Simulationen das BIC häufiger die wahre Ordnung einer Markovkette wählt, während das AIC zum Überschätzen der Ordnung neigt; daher betrachten wir beide Kriterien.

(1.10) Definition. Seien $z \in I\!R^n$ der Vektor der Beobachtungen, $\theta \in I\!R^k$ der Parametervektor eines angenommenen Modelles, $\hat{\theta}$ ein Schätzer von θ und $l(\hat{\theta}; z)$ die zugehörige Log-Likelihood-Funktion. **Akaikes Informationskriterium (AIC)** eines Modelles mit Parameter $\theta \in I\!R^k$ ist:

$$AIC(\theta) := -2\, l(\hat{\theta}; z) + 2k \ .$$

□

Diese Definition ist PARZEN et al. [1998] entnommen. Das BIC gewichtet die Anzahl der Modellparameter mit der Anzahl n der berücksichtigten Beobachtungen (KATZ [1981]).

(1.11) Definition. Das **Bayes'sche Informationskriterium (BIC)** eines Modelles mit Parameter $\theta \in I\!R^k$ bei n Beobachtungen $z \in I\!R^n$ ist:

$$BIC(\theta) := -2l(\hat{\theta}; z) + k \cdot \ln n \ .$$

□

(1.12) Bemerkung. Mit dem AIC wird unter Modellen mit $k, k \in Na$ Parametern das Modell mit Parameter θ_{AIC} ausgewählt, dessen AIC-Wert minimal ist, also:

$$AIC(\theta_{AIC}) = \min_{\theta} AIC(\theta) \ .$$

Mit dem BIC wird das Modell mit Parameter θ_{BIC} ausgewählt, wobei

$$BIC(\theta_{BIC}) = \min_{\theta} BIC(\theta) \; .$$

□

(1.13) Bemerkung. Seien $z_{t,n}, t = 1, \ldots, T, \; n = 1, \ldots, N$, die Beobachtungen des Ereignisprozesses an Tag t in Jahr n. Sei $n_{i_1 \ldots i_{h+1}}(t)$ die Anzahl der in allen Jahren beobachteten $(h+1)$-Tupel $(z_{t-h,n} = i_1, \ldots, z_{t,n} = i_{h+1})$. Die bedingte Log-Likelihood-Funktion l einer Markovkette der Ordnung h, gegeben die ersten h Beobachtungen, ist:

$$l = \sum_{t=h+1}^{T} \sum_{(i_1,\ldots,i_{h+1})} \left(n_{i_1 \ldots i_{h+1}}(t) \; \ln \hat{\pi}_{i_1,\ldots,i_h,i_{h+1}} \right) \; .$$

Die innere Summe wird über alle $(h+1)$-Tupel $(i_1, \ldots, i_{h+1}) \in \{0,1\}^{h+1}$ gebildet. Die Werte der Kriterien in der nachfolgenden Tabelle basieren auf der bedingten Log-Likelihood-Funktion, bei gegebenen Beobachtungen $z_{1,1}, z_{2,1}, z_{3,1}$ der ersten drei Tage.

	Markovkette			
Kriterium	0. Ordnung	1. Ordnung	2. Ordnung	3. Ordnung
AIC	12944	12095	12064	12153
BIC	12980	12152	12136	12239

Sowohl mit dem AIC als auch mit dem BIC wird die Markovkette 2. Ordnung angepaßt. Da der Unterschied zwischen den Werten der Kriterien für Markovketten erster und zweiter Ordnung gering ist, wird erst im Rahmen der Modellvalidation die Ordnung der Markovkette zusammen mit der Verteilung der Niederschlagsmengen an nassen Tagen ausgewählt.

2 Die Niederschlagsmengen an nassen Tagen

Zur Modellierung der Niederschlagsmengen an nassen Tagen gehen wir wie Zucchini und Adamson [1984] vor, da die gleichen Voraussetzungen gegeben sind: Die bedingte Verteilung der Niederschlagsmengen an nassen Tagen ist breit und rechtsschief, der Mittelwert verändert sich im Jahresverlauf. Er beträgt im Januar 3.72mm, im Juli 4.6mm und im Dezember 3.7mm, vergl. Abb. 4.2.
Die Weibullverteilung ist rechtsschief und flexibel und damit eine geeignete Kandidatin zur Modellierung der Niederschlagsmengen an nassen Tagen. Die Anpassung wird mit der Methode der Momente durchgeführt.

Die Zufallsvariablen W_t bzw. $W_{t,n}$ seien die Niederschlagsmengen an einem nassen Tag $t, t = 1, \ldots T$, bzw. an einem nassen Tag t in Jahr $n, n = 1, \ldots, N$. Ist Tag t

trocken, sind W_t bzw. $W_{t,n}$ nicht definiert. Wir betrachten $W_{t,n}$ als identisch verteilt für alle n bei konstantem t und als unabhängig für alle t und alle n. Die Annahme der Unabhängigkeit ist trotz der in Abschnitt 3.1, Tab. 3.2 geschätzten Lag(1)-Korrelation von $r(1) \simeq 0.14$ sinnvoll.
Zunächst schätzen wir die erwartete Niederschlagsmenge $\mathrm{E}W_{t,n} = m_t$ an einem nassen Tag $t, t = 1, \ldots, T$, durch die empirischen Mittel $\hat{m}_t$ an Tag t und glätten diese Schätzer anschließend mit einer Fouriersumme, da wie bei den Parametern einer Markovkette auch hier nur stetig sich ändernde Parameter physikalisch plausibel sind. Als zweites Moment schätzen wir den Variationskoeffizienten $\mathrm{CV}(W_t)$. Aus $\hat{m}_t$ und $\widehat{\mathrm{CV}}(W_t)$ berechnen wir Schätzer der Parameter der Weibullverteilung. Die genaue Vorgehensweise wird nachfolgend beschrieben. Der Erwartungswert m_t sollte eine stetige und sinusartige Funktion der Zeit sein:

$$\textbf{(2.1)} \quad m_t \approx R_0 + \sum_{j=1}^{L} R_j \ \cos(\omega_j t + \theta_t) \ , t = 1, \ldots, T,$$

mit $\omega_j = (2\ \pi/T)$ und einem geeigneten L. Die Parameter R_j, θ_j werden aus den empirischen Mittelwerten $\hat{m}_t$ geschätzt,

$$\hat{m}_t = \frac{1}{NR(t)} \sum_{t=1}^{NR(t)} W_{t,n} \ \ , t = 1, \ldots, T,$$

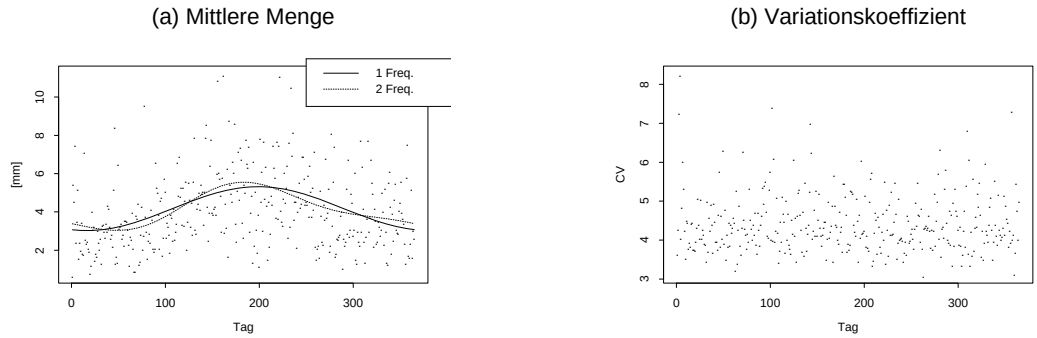
wobei $NR(t)$ die Anzahl nasser Tage t innerhalb der beobachteten N Jahre ist. Wie im vorigen Abschnitt wird (2.1) in

$$\textbf{(2.2)} \quad m_t \approx A_0 + \sum_{j=1}^{L} \Big(A_j \cos(\omega_j t) + B_j \sin(\omega_j t) \Big)$$

umgeformt. Damit liegt ein lineares Modell vor, dessen Parameter A_j, B_j mit dem Verfahren der iterativ gewichteten kleinsten Quadrate aus den Mittelwerten $\hat{m}_t$ geschätzt werden. Als Gewichtungsfaktoren dienen die Anzahlen $NR(t)$.
Alternativ hätten, wie bei Zucchini und Adamson [1984], die Beobachtungen $W_{t,n}$ mit konstanten Gewichtsfaktoren zum Schätzen verwendet werden können. Das besäße den Vorteil, daß die tatsächlichen Beobachtungen approximiert würden, aber den Nachteil größerer Datenmatrizen. Berechnungen zeigten, daß sich die Schätzer aus beiden Verfahren bei diesem Datensatz nur geringfügig unterscheiden. Die erste Vorgehensweise entspricht der bei der Approximation der Parameter der Markovkette verwendeten und wird hier bevorzugt.
Die Anzahl L der benötigten Fourierfrequenzen wird ohne Verteilungsannahme ausgewählt, und zwar unter Verwendung der Summe S der kleinsten Quadrate:

$$S = \sum_{t=1}^{T} \left(\hat{m}_t(L) - \hat{m}_t \right)^2 \ ,$$

Abbildung 4.2: Niederschlagsmengen an nassen Tagen



(a) Mittel $\hat{m}_t$ der Niederschlagsmenge an nassen Tagen t und Fourierapproximation.
(b) Empirischer Variationskoeffizient $c_t = s(W_t)/\hat{m}_t$.

---

wobei $\hat{m}_t(L), t = 1, \ldots, T$, die mit L Frequenzen angepaßten Werte der m_t sind. Ein deutlicher Rückgang von S ist beim Übergang von null zu einer Fourierfrequenz zu sehen, so daß wir $L = 1$ wählen und m_t durch $\hat{m}_t(L = 1)$ approximieren, vergl. Abb. 4.2 und Anhang B. Unter Annahme einer Weibullverteilung mit konstantem Formparameter und zeitabhängigem Skalenparameter kann die Auswahl der Anzahl L an Frequenzen mit dem AIC getroffen werden; dieses Vorgehen führt ebenfalls zu $L = 1$.

Die Schätzer der Varianz $\mathrm{Var}(W_t)$ und des Variationskoeffizienten $CV(W_t)$ gemäß der Methode der Momente sind

$$\textbf{(2.3)} \qquad \widehat{\mathrm{Var}}(W_t) = \frac{1}{NR(t)} \sum_{n=1}^{N_t} \left(W_{t,n}\right)^2 \; - \; \hat{m}_t^2 \; ,$$

$$\textbf{(2.4)} \qquad \widehat{CV}(W_t) = \sqrt{\widehat{\mathrm{Var}}(W_t)} / \; \hat{m}_t \; .$$

Die Schätzer $\widehat{CV}(W_t)$ sind im Zeitverlauf annähernd konstant, vergl. Abb. 4.2. Ist der Variationskoeffizient konstant, $CV(W_t) = CV, t = 1, \ldots, T$, so folgt aus (2.4)

$$\left(\sum_{t=1}^{T} \sum_{n=1}^{NR(t)} \left(W_{t,n} - \hat{m}_t\right)^2 \right) = \widehat{CV}^2 \sum_{t=1}^{T} NR(t) \; \hat{m}_t^2 \; .$$

Wenn statt des empirischen Mittels $\hat{m}(t)$ die mit (2.1) bestimmten Schätzer $\hat{m}_t(L)$ verwendet werden, erhalten wir

$$\textbf{(2.5)} \qquad \widehat{CV} := \sqrt{\frac{\sum_{t=1}^{T} \sum_{n=1}^{NR(t)} \left(W_{t,n} - \hat{m}_t(L)\right)^2}{\sum_{t=1}^{T} N_t \; \hat{m}_t^2(L)}} \; .$$

Dieser Schätzer ist im allgemeinen nicht erwartungstreu. Aus den Schätzern der Erwartungswert-Zeit-Funktion m_t und des Variationskoeffizienten CV lassen sich die Parameter der Weibullverteilung bestimmen. Die Dichte der Weibullverteilung mit Skalenparameter α und Formparameter β lautet

$$\textbf{(2.6)} \quad f(x) = \frac{\beta}{\alpha^\beta}\, x^{\beta-1} \exp((-x/\alpha)^\beta) \ .$$

Es sind

$$\textbf{(2.7)} \quad \mathrm{E}X = \alpha\,\Gamma(1+1/\beta), \quad \mathrm{Var}X = \alpha^2 \left[\Gamma(1+2/\beta) - \Gamma^2(1+1/\beta)\right] \ .$$

Werden die W_t durch Weibullverteilungen mit Parametern α_t und β_t beschrieben, so ist der Formparameter $\beta_t = \beta$ als Funktion des Variationskoeffizienten nicht zeitabhängig:

$$\textbf{(2.8)} \quad \mathrm{CV} = \sqrt{\left(\frac{\Gamma(1+2/\beta)}{\Gamma^2(1+1/\beta)} - 1\right)} \ .$$

Es existiert keine geschlossene Form einer Umkehrfunktion zu (2.8), aber ZUCCHINI UND ADAMSON [1984] geben eine Approximation durch eine rationale Funktion:

$$\textbf{(2.9)} \quad \hat{\beta} = \frac{339.5410 + 148.4445\ \widehat{CV} + 192.7492\ \widehat{CV}^2 + 22.4401\ \widehat{CV}^3}{1 + 257.1162\ \widehat{CV} + 287.8362\ \widehat{CV}^2 + 157.2230\ \widehat{CV}^3} \ .$$

Da

$$\textbf{(2.10)} \quad \alpha_t = m_t/\Gamma(1+1/\beta)\ , t = 1, \ldots, T,$$

wird $\hat{\alpha}$ durch Einsetzen von $\hat{\beta}$ und $\hat{m}_t(L)$ bestimmt.

Wir passen also eine Weibullverteilung mit konstantem Formparameter β und zeitabhängigem Skalenparameter $\alpha_t(L=1)$ an. Ein zeitabhängiger Skalenparameter α_t führt zum Modell $m_t = m_t(L=1)$. Die Werte der Schätzer $\hat{\alpha}_t$ und $\hat{\beta}$ sind

$$\hat{\beta} = 0.7561838$$

und

$$\hat{\alpha}_t(1) = \hat{m}_t(1)/\Gamma(1+1/\hat{\beta})$$

mit

$$\hat{m}_t(1) = 4.173 + 1.143\cos(\omega_1 t - 2.860), \quad t = 1, \ldots, T.$$

Die Werte der Parameterschätzer und der Residuenquadratsummen weiterer Approximationen $\hat{m}_t(L)$ befinden sich im Anhang B.2.

3 Simulation

Der Ereignisprozeß kann nach Abschnitt 1 durch eine Markovkette erster oder zweiter Ordnung beschrieben werden, die Niederschlagsmengen an nassen Tagen gemäß Abschnitt 2 durch eine Weibullverteilung mit konstantem Form- und zeitabhängigem Skalenparameter. Mit diesen beiden Modellen werden jeweils 1000 Jahre künstlicher täglicher Niederschlagsmessungen erzeugt. Die Werte der Parameter der Modelle befinden sich in Anhang B. Der Ereignisprozeß und die Niederschlagsmengen an nassen Tagen werden in zwei Schritten aus den geschätzten Modellparametern simuliert. Die Zustände des Ereignisprozesses seien $Z_{t,n}, t = 1, \ldots T = 365,\ n = 1, \ldots, N = 1000$.

(3.1) Simulation einer Markovkette 1. Ordnung.

1. Tag 1:
   Sind in einer Markovkette die Übergangswahrscheinlichkeiten stetige Funktionen der Zeit, so auch die Gleichgewichtswahrscheinlichkeiten, und es gilt approximativ $\pi_1(t) \simeq \pi_1(t-1)$ bzw. $\pi_0(t) \simeq \pi_0(t-1)$, $t = 1, \ldots, T$, wobei $\pi_1(0) = \pi_1(T)$ und $\pi_0(0) = \pi_0(T)$. Damit berechnen sich die Gleichgewichtswahrscheinlichkeiten $\delta = (\pi_0, \pi_1)$ einer Markovkette 1. Ordnung aus der für stationäre Ketten geltenden Gleichung

   $$\delta = \delta^T \Pi \ ,$$

   wobei Π die Matrix der Übergangswahrscheinlichkeiten ist,

   $$\Pi = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix}$$

   mit $\mathrm{P}_{ij} = \mathrm{P}(Z_{t,n} = j | Z_{t-1,n} = i)$. Lösen des Gleichungssystemes unter Beachtung von $\pi_0 + \pi_1 = 1$ führt zu

   $$\pi_1(t) = \frac{\pi_{01}(t)}{1 - \pi_{11}(t) + \pi_{01}(t)} \ , \quad \pi_0(t) = 1 - \pi_1(t) \ .$$

   Für nicht-stationäre Ketten gilt diese Gleichung approximativ. Als Startwahrscheinlichkeiten verwenden wir die Gleichgewichtswahrscheinlichkeiten $\pi_1(1)$, $\pi_0(1)$. Es wird eine $U(0,1)$-verteilte Zufallszahl u_1 erzeugt (in S-PLUS mit dem Zufallszahlengenerator `runif`). Der Zustand $Z_{1,1}$ des Ereignisprozesses an Tag 1 in Jahr 1 ist

   $$Z_{1,1} = \begin{cases} 1, & u_1 \leq \pi_1(1) \ , \\ 0, & u_1 > \pi_1(1) \ . \end{cases}$$

2. Tage $2, 3 \ldots$:
   Basierend auf $Z_{1,1}$ werden die weiteren Werte des Ereignisprozesses simuliert. Für jeden zu simulierenden Zustand $Z_{t,n}$ wird eine $U(0,1)$-verteilte Zufallszahl u erzeugt.

$$Z_{t,n} = \left\{ \begin{array}{ll} 1, & u_1 \leq \pi_{11}(t), \\ 0, & u_1 > \pi_{11}(t), \end{array} \right\} \quad \text{falls } Z_{t-1,n} = 1,$$

$$Z_{t,n} = \left\{ \begin{array}{ll} 1, & u_1 \leq \pi_{01}(t)\ , \\ 0, & u_1 > \pi_{01}(t)\ , \end{array} \right\} \quad \text{falls } Z_{t-1,n} = 0$$

   für $t = 1, \ldots, T$, $n = 1, \ldots, N$, und $Z_{0,n} = Z_{T,n-1}$ für $n \geq 2$.

(3.2) Simulation einer Markovkette 2. Ordnung.
Ist $Z_1, Z_2, \ldots$ mit $Z_t \in \{0,1\}$, eine stationäre Markovkette 2. Ordnung mit Übergangswahrscheinlichkeiten $\pi_{ijk} = \mathrm{P}(Z_t = k | Z_{t-1} = j, Z_{t-2} = i)$, so berechnen sich die Gleichgewichtswahrscheinlichkeiten (π_0, π_1), indem die Markovkette in eine Markovkette 1. Ordnung transformiert wird. Wir definieren $Y_t := (Z_t, Z_{t+1})$. Dann ist $Y_1, Y_2, \ldots$ eine Markovkette 1. Ordnung mit Zustandsraum $\{0,1\}^2$ und Übergangswahrscheinlichkeiten

$$\begin{aligned} \pi_{ijkl} &= \mathrm{P}(Z_{t+1} = l, Z_t = k | Z_t = j, Z_{t-1} = i) \\ &= \mathrm{P}(Y_t = (k,l) | Y_{t-1} = (i,j)) \\ &= \left\{ \begin{array}{ll} \pi_{ijjl} & \text{für } j = k, \\ 0 & \text{sonst.} \end{array} \right. \end{aligned}$$

Die Übergangsmatrix der $\{Y_t\}_{t \in \boldsymbol{Z}}$ ist damit $\Pi = \Big(\mathrm{P}(Y_{t+1} = j | Y_t = i) \Big)_{ij}$,

$$\Pi = \begin{array}{c} \text{Übergang} \\ \text{Ausgang} \\ 00 \\ 10 \\ 01 \\ 11 \end{array} \begin{array}{c} \begin{array}{cccc} 00 & 10 & 01 & 11 \end{array} \\ \begin{pmatrix} \pi_{0000} & 0 & \pi_{0001} & 0 \\ \pi_{1000} & 0 & \pi_{0001} & 0 \\ 0 & \pi_{0110} & 0 & \pi_{0111} \\ 0 & \pi_{1110} & 0 & \pi_{1111} \end{pmatrix} \end{array} .$$

Die Gleichgewichtswahrscheinlichkeiten für $\{Y_t\}_{t \in \boldsymbol{Z}}$ berechnen sich unter Beachtung von $\sum_{i=1}^{4} \delta_i = 1$ aus

$$\delta = \delta^T \Pi$$

mit $\delta^T = (\pi_{00}, \pi_{10}, \pi_{01}, \pi_{11})$ und $\pi_{ij} = \mathrm{P}(Y_t = (i,j)) = \mathrm{P}(Z_t = i, Z_{t+1} = j)$.

Daraus berechnen sich die Gleichgewichtswahrscheinlichkeiten für Z_t:

$$\begin{aligned}
\pi_0 &= \mathrm{P}(Z_t = 0) = \mathrm{P}(Z_{t-1} = 0, Z_t = 0) + \mathrm{P}(Z_{t-1} = 1, Z_t = 0) \\
&= \pi_{00} + \pi_{10} \\
\pi_1 &= \mathrm{P}(Z_t = 1) = \mathrm{P}(Z_{t-1} = 0, Z_t = 1) + \mathrm{P}(Z_{t-1} = 1, Z_t = 1) \\
&= \pi_{01} + \pi_{11} \ .
\end{aligned}$$

Die letzten Gleichheiten in jeder Zeile gelten aufgrund der Stationarität. Unter Annahme stetig sich ändernder Übergangswahrscheinlichkeiten lassen sich Startwahrscheinlichkeiten $\pi_1(1), \pi_0(1)$ und daraus die bedingten Wahrscheinlichkeiten $\pi_{11}(2) = \mathrm{P}(Z_2 = 1|Z_1 = 1)$ und $\pi_{01}(2) = \mathrm{P}(Z_2 = 1|Z_1 = 0)$ berechnen. Der Ereignisprozeß $Z_{t,n}, t = 1, \ldots, T,\ n = 1, \ldots, N$ wird analog zu dem oben beschriebenen simuliert:

1. Tag 1:
   Es wird eine $U(0,1)$-verteilte Zufallszahl u_1 erzeugt. Dann ist

   $$Z_{1,1} = \begin{cases} 1, & u \leq \pi_1(1) \ , \\ 0, & u > \pi_1(1) \ . \end{cases}$$

2. Tag 2:
   Sei $u_2 \sim U(0,1)$-verteilt. Dann setzen wir

   $$\begin{aligned}
   Z_{2,1} &= \left\{ \begin{array}{ll} 1, & u \leq \pi_{11}(2), \\ 0, & u > \pi_{11}(2), \end{array} \right\} \quad \text{falls } Z_{1,1} = 1 \ , \\
   Z_{2,1} &= \left\{ \begin{array}{ll} 1, & u \leq \pi_{01}(2), \\ 0, & u > \pi_{01}(1), \end{array} \right\} \quad \text{falls } Z_{1,1} = 0 \ .
   \end{aligned}$$

3. Tage $3, 4, \ldots,$:
   Es werden sukzessive die Daten $Z_{t,n}$ erzeugt, indem für jeden Tag eine $U(0,1)$-verteilte Zufallszahl u erzeugt wird und

   $$Z_{t,n} = \begin{cases} 1, & u \leq \pi_{ij1}(t), \\ 0, & u > \pi_{ij1}(t), \end{cases}$$

   wobei $i, j \in \{0,1\}$ und $t = 1, \ldots, T,\ n = 1, \ldots, N$, und $Z_{0,n} = Z_{T,n-1}$, $Z_{-1,n} = Z_{T-1,n-1}$ gilt.
   Die Berechnung der Gleichgewichtswahrscheinlichkeiten kann der S-PLUS-Funktionen `markov2.eq.fun` in Anhang B.4 entnommen werden.

(3.3) Simulation der Niederschlagsmengen an nassen Tagen.
Zur Simulation der Niederschlagsmengen an nassen Tagen werden die Tage $t, t = 1, \ldots T$, in den Jahren $n, n = 1, \ldots, N$, durchlaufen. Ist $Z_{t,n} = 1$, wird eine weibullverteilte Zufallszahl mit Skalenparameter $\hat{\alpha}_t$ und Formparameter $\hat{\beta}$ erzeugt (in S-PLUS mit dem Zufallszahlengenerator `rweibull`).

Die verwendeten S-PLUS-Routinen befinden sich in Anhang B.4.

4 Modellvalidation

Zur Modellauswahl wird mit einer Markovkette 1. Ordnung und mit einer 2. Ordnung jeweils ein Datensatz, bestehend aus 1000 Jahren täglicher Niederschlagsmessungen, simuliert. Die Eigenschaften der simulierten Daten werden mit denen der Beobachtungen verglichen.

Wir verwenden die folgenden Bezeichnungen für die betrachteten Datensätze: OBS seien die Beobachtungen, M1 sind die mit einer Markovkette 1. Ordnung und M2 die mit einer Markovkette 2. Ordnung erzeugten Datensätze. Die Niederschlagsmengen an nassen Tagen sind jeweils mit einer Weibullverteilung mit den in Abschnitt 2 gegebenen Parametern erzeugt, also mit konstantem Form- und zeitabhängigem Skalenparameter.

Zur Modellauswahl betrachten wir

1. Mittel und Standardabweichung des Jahresniederschlages,
2. Mittel und Standardabweichung des Monatsniederschlages,
3. Mittel und Standardabweichung der Anzahl nasser Tage,
4. Mittel und Standardabweichung des Niederschlages an einem nassen Tag,
5. Clustereigenschaften nasser Tage.

Die Werte dieser Statistiken befinden sich in Anhang B.3. Die Verteilung der Jahresniederschlagsmengen beider simulierter Sets ähneln der beobachteten Verteilung, vergl. Abb. 4.3. Das Mittel $\bar{Y}_J$ des Jahresniederschlages wird unter M1 besser wiedergegeben, unter beiden Modellen ist dessen Standardabweichung $s(Y_J)$ zu klein. Die empirischen Quantile des Jahresniederschlages sind in beiden Simulationssets größer als die der Beobachtungen, vergl. Tab. 4.2. Der mittlere Monatsniederschlag wird von M1 besser wiedergegeben als von M2, die Standardabweichung ist unter beiden Modellen zu klein, vergl. Anhang B.3. Insgesamt ist die Performance beider Modelle bezüglich des Monatsniederschlages ähnlich, vergl. Abb. 4.4.

Tabelle 4.2: Jahresniederschlag

Empirische Quantile

	OBS	M1	M2
99%	387.6	389.4	395.1
98%	387.6	407.7	419.6
95%	392.9	431.7	451.0
90%	419.8	454.6	477.5
50%	561.5	562.7	588.3

Jahresniederschlag

	OBS	M1	M2
Y_J	563.7	565.6	589.4
$\hat{s}(Y_J)$	111.0	85.0	86.7

links: Empirische Quantile des Jahresniederschlages aus den Beobachtungen OBS bzw. den simulierten Daten M1 und M2. rechts: Mittel und Standardabweichung des Jahresniederschlages.

Beide Modelle besitzen annähend identische Eigenschaften bezüglich der Niederschlagsmenge W_t an nassen Tagen. Das Mittel $\hat{m}_t$ ist unter beiden Modellen um etwa 1mm zu groß. Die Standardabweichung $s(W_t)$ ist ebenfalls unter beiden Modellen etwas zu groß. Modell M2 ist geringfügig besser, Abb. 4.5. Extremer täglicher Niederschlag besitzt unter beiden Modellen eine fast identische Verteilung. Die 90%-, 95%-, 98%-Quantile sind unter den Modellen etwas zu groß. Das 99%-Quantil ist unter den Modellen in den Monaten Mai, Juni und Dezember um etwa 5mm bzw. um 15% zu klein, was für Anwendungen des Modelles von Bedeutung ist.
Abschließend betrachten wir das Clusterverhalten nasser Tage. Die Wahrscheinlichkeit $\hat{P}_{1-Run}(k)$ eines nassen Runs der Länge k ist die Wahrscheinlichkeit, k nasse Tage in Folge zu erhalten, bei gegebenem ersten nassen Tag:

$$\hat{P}_{1-Run}(k) = P(Z_{t+1} = 1, \ldots, Z_{t+k-1} = 1, Z_{t+k} = 0 | Z_t = 1, Z_{t-1} = 0)$$

Der empirische Schätzer $\hat{P}_{1-Run}(k)$ ist

$$\textbf{(4.1)} \qquad \hat{\mathrm{P}}_{1-Run}(k) = \frac{\text{Anzahl an 1-Runs der Länge } k}{\text{Gesamtanzahl an 1-Runs}} .$$

Um saisonale Schwankungen zu berücksichtigen, werden die Schätzer $\hat{\mathrm{P}}_{1-Run}(k)$ für jeden Monat separat berechnet. Die Runlängen werden hier jeweils ab einem Monatsanfang und bis zum Monatsende gezählt. Ist beispielsweise der 1. Januar naß, zählt er als erster Tag eines nassen Runs, ein am 31. Januar beginnender Run ist hingegen automatisch ein Run der Länge 1. Die beobachteten Clusterwahrscheinlichkeiten werden von beiden Modellen sehr gut wiedergegeben, vergl. Abb. 4.8 ff.

5 Ergebnisse

Täglicher Niederschlag läßt sich gut modellieren, indem der Ereignisprozeß durch eine Markovkette erster oder zweiter Ordnung beschrieben wird und die Niederschlagsmengen an nassen Tagen durch eine Weibullverteilung mit konstantem Form- und zeitabhängigem Skalenparameter.
Werden die Übergangswahrscheinlichkeiten der Markovketten durch Fouriersummen approximiert, besitzt die angepaßte Markovkette 1. Ordnung 8 Parameter, die Markovkette 2. Ordnung 10 Parameter. Sowohl mit dem AIC als auch mit dem BIC wird eine Markovkette 2. Ordnung zur Beschreibung des vorliegenden Datensatzes ausgewählt. Wir haben zur Modellvalidation zwei Datensätze täglicher Niederschlagsmessungen aus 1000 Jahren simuliert, und zwar im Datensatz M1 basierend auf einer Markovkette 1. Ordnung und im Datensatz M2 basierend auf einer Markovkette 2. Ordnung. Die Niederschlagsmengen an nassen Tagen sind in beiden Datensätzen mit der gleichen Verteilung, nämlich einer Weibullverteilung mit zeitabhängigem Skalenparameter simuliert worden. Alle in diesem Abschnitt untersuchten Eigenschaften der Beobachtungen werden von beiden Datensätzen sehr gut wiedergegeben.

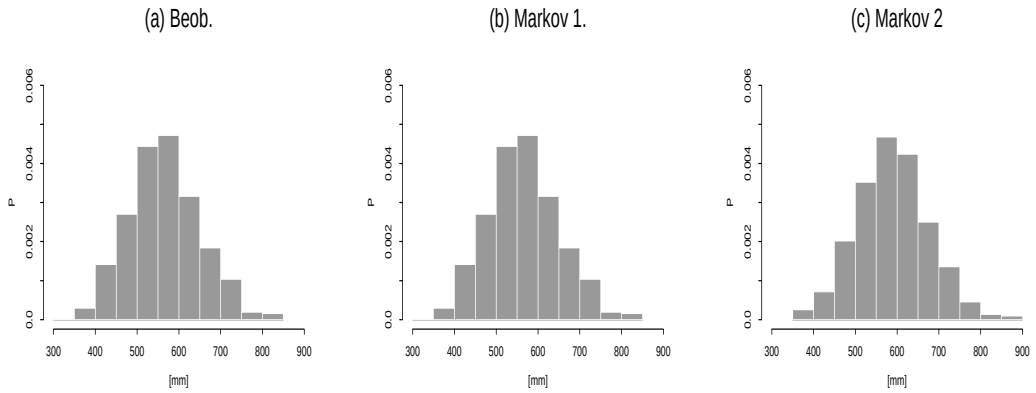
Lediglich der Jahresniederschlag weicht unter M2 in stärkerem Maße von dem beobachteten ab. Die Anzahl nasser Tage pro Monat wird hingegen von Modell M2 besser wiedergegeben als von M1. Das 99%-Quantil des Tagesniederschlages ist unter beiden Modellen zu klein. Sollte eines der Modelle also zur Beschreibung extremer Niederschläge angewendet werden, müßten gegebenfalls Korrekturen vorgenommen werden.
Angesichts der geringeren Parameterzahl und der fast identischen Performance beider Modelle entscheiden wir uns für Modell M1, der Ereignisprozeß wird also durch eine Markovkette 1. Ordnung beschrieben. Die geschätzten Werte der Parameter der Markovkette sind

$$\pi_{11}(t) = 0.22 + 0.14\cos(\omega_1 t - 1.95)\ , t = 1, \ldots, T,$$

$$\pi_{01}(t) = -1.07 + 0.17\cos(\omega_1 t - 0.41) + 0.2\cos(\omega_2 t - 3.13), t = 1, \ldots, T,$$

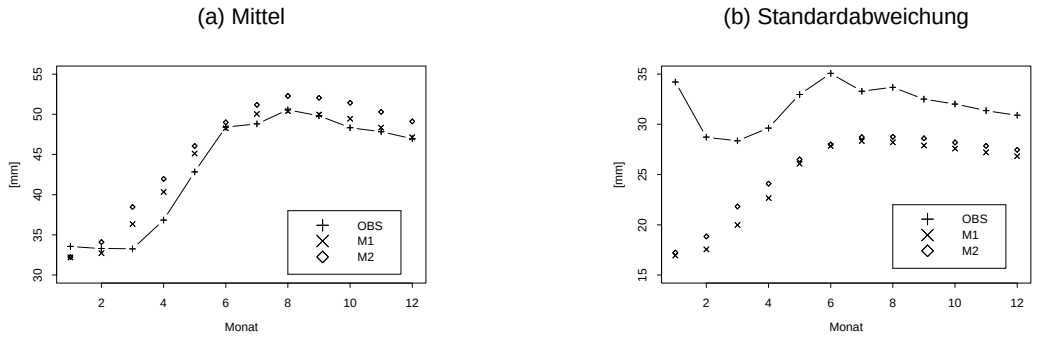
mit $\omega_j = 2\pi j/T, T = 365$. Die geschätzten Werte der Parameter der Weibullverteilung sind auf Seite 59 gegeben.

Abbildung 4.3: Jahresniederschlag



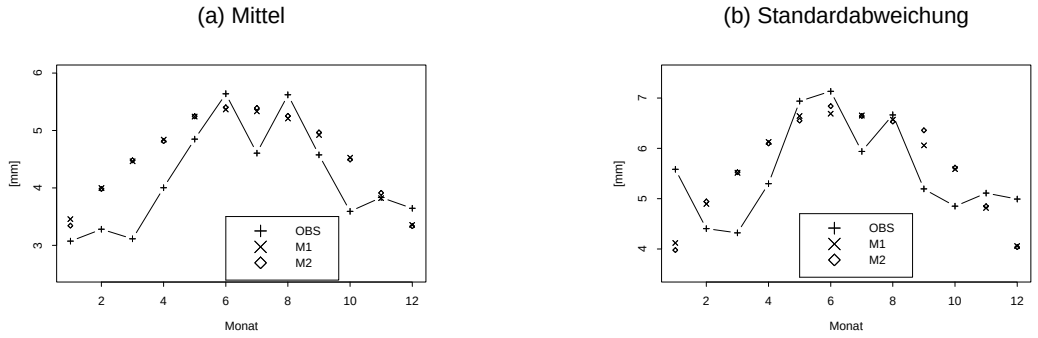
(a) Beobachtungen (27 Jahre), (b) und (c) Simulationen (1000 Jahre) mit
(b) Markovkette 1. Ordnung, (c) Markovkette 2. Ordnung;
weibullverteilte Niederschlagsmengen an nassen Tagen in (b) und (c).

Abbildung 4.4: Monatsniederschlag



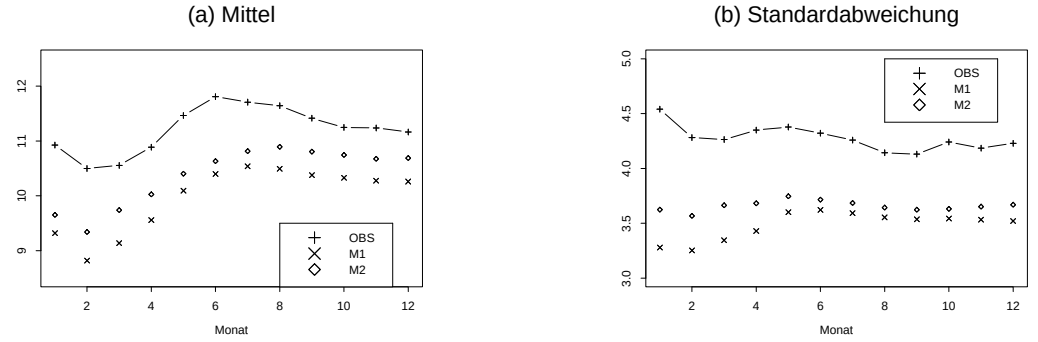
Monatsniederschlag: (a) Mittel. (b) Standardabweichung.

Abbildung 4.5: Niederschlagsmenge an nassen Tagen



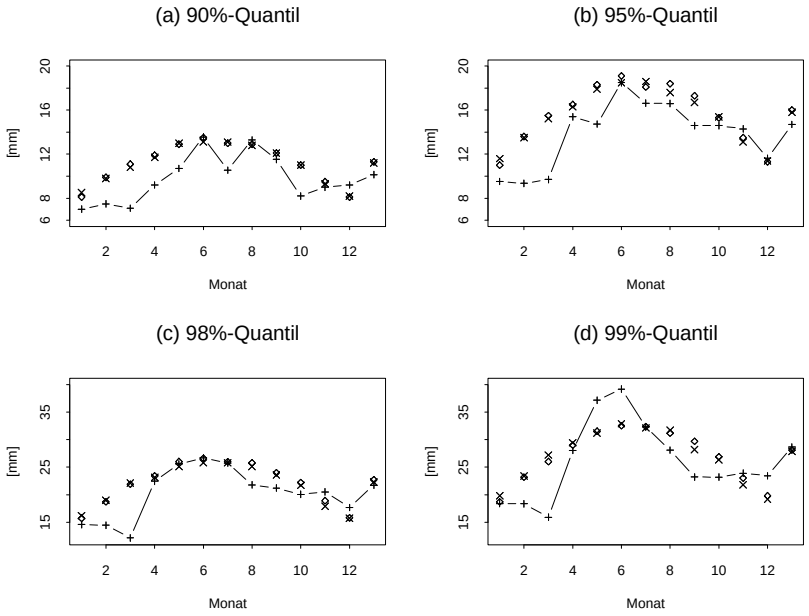
Niederschlagsmenge an nassen Tagen: (a) Mittel. (b) Standardabweichung.

Abbildung 4.6: Anzahl nasser Tage pro Monat



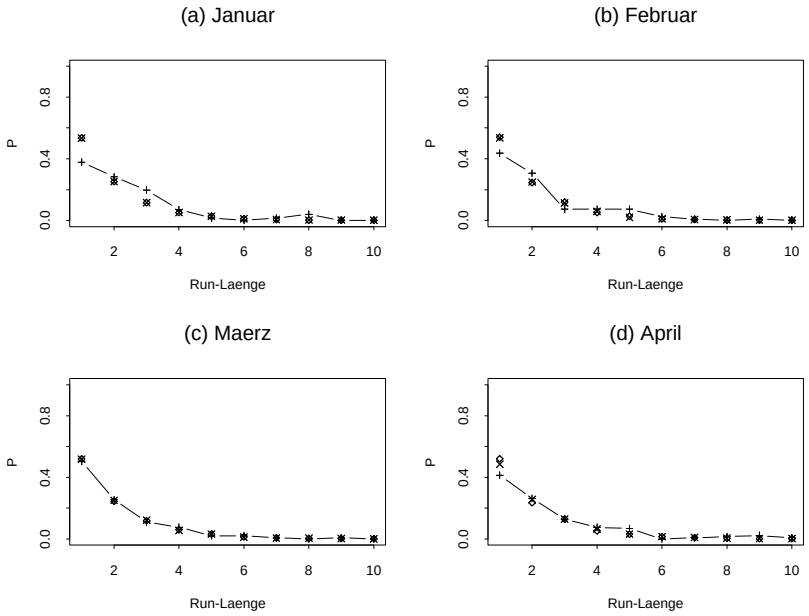
Anzahl nasser Tage pro Monat: (a) Mittel. (b) Standardabweichung.

Abbildung 4.7: Quantile der Niederschlagsmenge an nassen Tagen



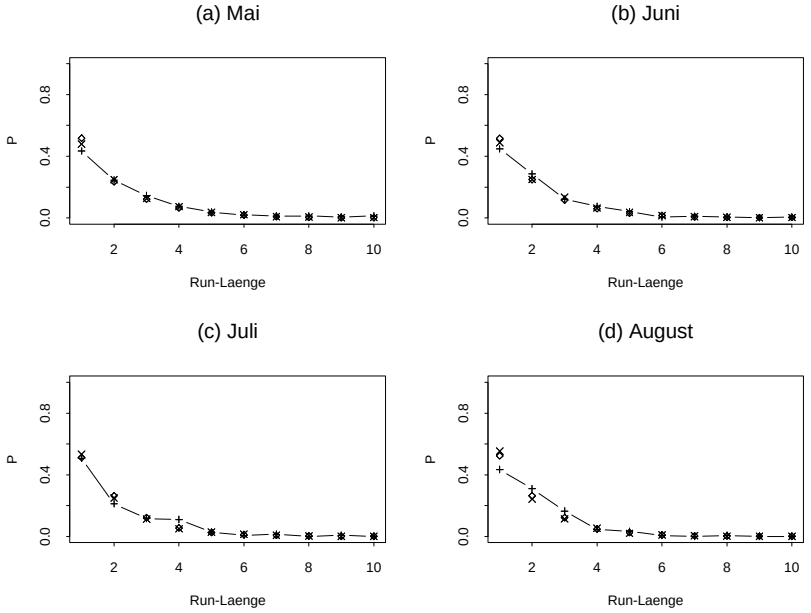
Empirische Quantile des Niederschlages an nassen Tagen pro Monat. Der 13. Eintrag auf der horizontalen Achse ist jeweils der Jahreswert. (a) 90%-Quantil, (b) 95%-Quantil, (c) 98%-Quantil, (d) 99%-Quantil. + OBS, × M1, ◇ M2.

Abbildung 4.8: Täglicher Niederschlag - Clusterverhalten



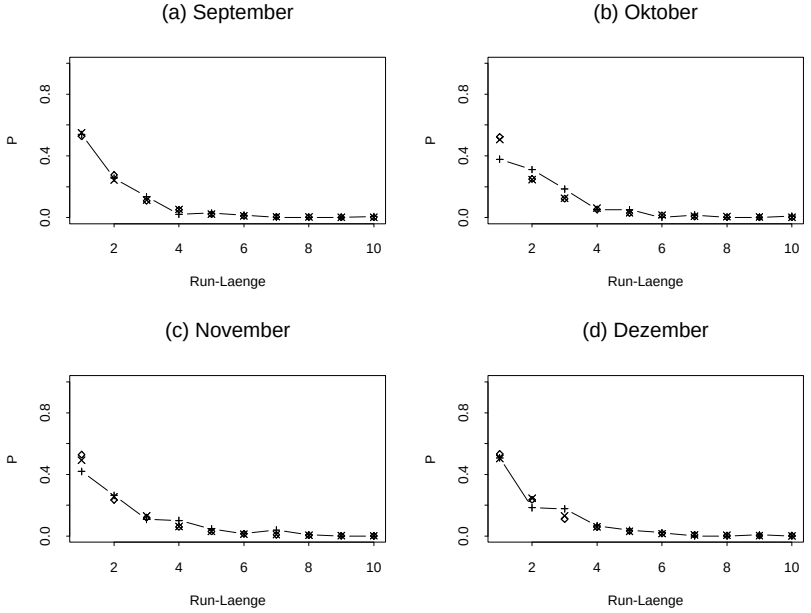
Gemäß (5.1) geschätzte Wahrscheinlichkeiten der Runlängen der Monate Januar - April. horizontale Achse: Run-Längen $k = 1, \ldots, 10$. + OBS, × M1, ◇ M2.

Abbildung 4.9: Täglicher Niederschlag - Clusterverhalten



Gemäß (5.1) geschätzte Wahrscheinlichkeiten der Runlängen der Monate Mai - August. horizontale Achse: Run-Längen $k = 1, \ldots, 10$. + OBS, $\times$ M1, $\diamond$ M2.

Abbildung 4.10: Täglicher Niederschlag - Clusterverhalten



Gemäß (5.1) geschätzte Wahrscheinlichkeiten der Runlängen der Monate September - Dezember. horizontale Achse: Run-Längen $k = 1, \ldots, 10$. + OBS, $\times$ M1, $\diamond$ M2.

Kapitel 5

Das Neyman–Scott–Modell

An den vorliegenden Stuttgarter Datensatz ist zur Modellierung stündlichen Regens ein Neyman-Scott-Clusterprozeß angepaßt worden. Die Arbeiten von RODRIGUEZ-ITURBE, COX UND ISHAM [1984] und insbesondere die Arbeit von COWPERTWAIT et al. [1996a], in denen Niederschlagsmessungen aus unterschiedlichen klimatischen Regionen wie Colorado, Venezuela oder Großbritannien erfolgreich mit einem Neyman-Scott-Clusterprozeß beschrieben werden, lassen vermuten, daß dieses flexible Modell auch zur Beschreibung des vorliegenden Datensatzes geeignet ist. Insbesondere die Eleganz des Modelles und die Möglichkeit, die Parameter physikalisch zu interpretieren, haben zur Wahl dieses Modelles geführt. Ein vorab unternommener Vergleich der empirischen Eigenschaften der Folge der stündlichen Niederschlagsmessungen mit den unter dem Modell zu erwartenden Eigenschaften haben diese Auswahl gerechtfertigt. Clustermodelle sind bereits in Kapitel 2.7 beschrieben worden. In den uns bekannten Arbeiten sind Clustermodelle allerdings stets zur Beschreibung stationärer Prozesse verwendet worden. Saisonalität wird dann berücksichtigt, indem die Beobachtungen in Perioden eingeteilt werden, in denen der Niederschlagsprozeß als stationär betrachtet werden kann und das Modell für jede dieser Perioden separat angepaßt wird. Bei dieser Vorgehensweise weisen die Modellparameter Unstetigkeiten beim Übergang von einer Periode zur folgenden auf. Würde dieses Vorgehen auf die Stuttgarter Daten angewendet, müßten die Beobachtungen in mindestens fünf, besser in zwölf den Kalendermonaten entsprechende Perioden eingeteilt werden, vergl. Kapitel 3.2. Diese unelegante Vorgehensweise steht im Widerspruch zu dem Anspruch des Modelles, physikalisch interpretierbar zu sein, da ein klimatischer Prozeß sich im Zeitverlauf stetig entwickelt und verändert. Das Ziel dieser Arbeit besteht deshalb darin, das Neyman-Scott-Modell mit Rechteckimpulsen (NSRP-Modell) zu verallgemeinern, so daß die Modellparameter stetige Funktionen der Zeit sind.

In diesem Kapitel leiten wir zunächst Eigenschaften des stationären Neyman–Scott–Modelles her und beschreiben die Anpassung für einen stationären Prozeß. Anschließend verallgemeinern wir das Modell, so daß es Saisonalität erfaßt und passen es an die Stuttgarter Daten an. Die Modellvalidation wird im folgenden Kapitel beschrieben.

1 Definition und Eigenschaften

(1.1) Definition. Das in der vorliegenden Arbeit verwendete **saisonale Neyman-Scott-Modell** ist wie folgt aufgebaut:

1. Stürme als Primärereignisse oder Clusterzentren des Punktprozesses treten gemäß einem nicht-homogenen Poissonprozeß mit Rate $\lambda(t) > 0,\ t \in \mathbb{R}$, auf. Sie können nicht beobachtet werden.

2. Jeder Sturm zieht eine zufällige Anzahl C an Zellen nach sich; sie repräsentieren die beobachtbaren Regenereignisse. Die Clustergrößen unterschiedlicher Stürme sind unabhängig. Die Größe des Clusters mit Zentrum in t ist C_t. Um die Existenz mindestens einer Zelle pro Sturm zu gewährleisten, modellieren wir $C_t - 1$ als poissonverteilt: $(C_t - 1) \sim \mathrm{Po}(\nu(t) - 1), \nu(t) > 1$. Damit ist $\mathrm{E}C_t = \nu(t)$.

3. Die Zellstarts sind zufällig und unabhängig vom Clusterzentrum entfernt. Die Abstände $D_{t,j}$ zwischen einem Clusterzentrum in t und den zugehörigen Zellen $j = 1, \ldots, C_t$ sind exponentialverteilt mit Erwartungswert $1/\beta(t)$. Die Zellstarts repräsentieren die Punkte des Zählprozesses.

4. Mit einem Zellstart in $t + D_{t,j}$ wird ein Rechteckimpuls assoziiert, der die Länge L und die Intensität X der Zelle beschreibt. Beide Zufallsvariablen sind unabhängig voneinander exponentialverteilt mit Erwartungswerten $1/\eta(t + D_{t,j})$ bzw. $1/\xi(t + D_{t,j})$, wenn der Zellstart in $t + D_{t,j}, j = 1, \ldots, C_t$, erfolgt.

Die Modellparameter $\lambda,\ \nu,\ \beta,\ \eta,\ \xi : \mathbb{R} \longrightarrow \mathbb{R}_+$ sind stetige und strikt positive (bzw. $\nu(t) > 1$) reelle Funktionen. Alle Zufallsvariablen des Modelles sind unabhängig voneinander. □

Die Intensität einer Zelle ist konstant, aber Zellen können sich überlagern, was sich in einer veränderten Intensität des beobachteten Regenereignisses äußert. Bei der unten erfolgenden Herleitung der Autokovarianzstruktur wird die Annahme notwendig werden, daß $\beta(t) < \eta(t + D_{t,j}),\quad j = 1, \ldots, C_t$. Die Abstände einer Zelle vom Sturmanfang sind damit im Mittel größer als die Zelldauer. Unterschiedliche Zellen sind in diesem Modell nicht mehr identisch verteilt wie in dem in Kapitel 2.7 beschriebenen und in der Literatur verwendeten Neyman-Scott-Modell.

Da Sturmanfänge nicht beobachtet werden und Zellen sich überlagern können, ist es nicht möglich, die Modellparameter direkt aus den Beobachtungen zu schätzen. Stattdessen werden sie mit der Methode der Momente oder einem Kleinste-Quadrate-Ansatz aus der beobachteten Serie stündlicher Niederschlagsmengen geschätzt. Dazu werden Eigenschaften des aggregierten Niederschlages wie der Erwartungswert oder die Varianz mit den entsprechenden Eigenschaften unter dem Modell verglichen.

Wir werden die Parameter des in (1.1) definierten Prozesses schätzen, indem der Prozeß an sich überlagernde 30-tägige Perioden unter der Annahme angepaßt wird,

der Prozeß sei während dieser Periode stationär. Die resultierenden Parameter werden anschließend geglättet. Diese Vorgehensweise wird in Abschnitt 3 detailliert erläutert.

Wir beschreiben zunächst die Anpassung des Modelles unter der Annahme von Stationarität. Unter dieser Annahme werden Eigenschaften des aggregierten Niederschlagsprozesses von RODRIGUEZ-ITURBE et al. [1987] und von COWPERTWAIT [1996a] genannt, aber nicht hergeleitet. In diesem Abschnitt leiten wir die Momente erster und zweiter Ordnung der Niederschlagsmengen in Intervallen der Länge h her.

Herleitung von Erwartungswert und Kovarianzstruktur

Sei $Y(t)$ die Gesamtintensität zur Zeit t, also die Summe der Intensitäten aller zur Zeit t aktiven Zellen. $N(t)$ bezeichne den Zählprozeß, also die Summe der Punkte (Zellstarts) im Intervall $[0,t]$. Es ist

$$\textbf{(1.2)} \quad Y(t) = \int_0^\infty X_{t-u}(u)\, dN(t-u) = \int_0^\infty X \cdot I(L>u)\, dN(t-u)\,,$$

wobei $X_\tau(u)$ die Intensität der zur Zeit τ beginnenden Zelle zur Zeit $u+\tau$ bezeichnet und $I(L>u)$ die Indikatorfunktion ist:

$$I(L>u) = \begin{cases} 1, & L \geq u, \\ 0, & L < u. \end{cases}$$

Die Intensität X einer Zelle habe die Verteilungsfunktion $F_X(u) = 1 - e^{-\eta u}$, die Survivor-Funktion $\mathcal{F}_X(u) = 1 - F_X(u)$. Da die Intensität während der Dauer einer Zelle konstant ist, folgt:

$$\textbf{(1.3)} \quad X_{t-u}(u) = \begin{cases} X & \text{mit Wahrscheinlichkeit} \quad \mathcal{F}_L(u) = e^{-\eta u}, \\ 0 & \text{mit Wahrscheinlichkeit} \quad 1 - \mathcal{F}_L(u) = 1 - e^{-\eta u}. \end{cases}$$

Die erwartete Intensität $\mathrm{E}Y(t)$ zur Zeit t berechnet sich wegen der Unabhängigkeit der Zufallsvariablen als

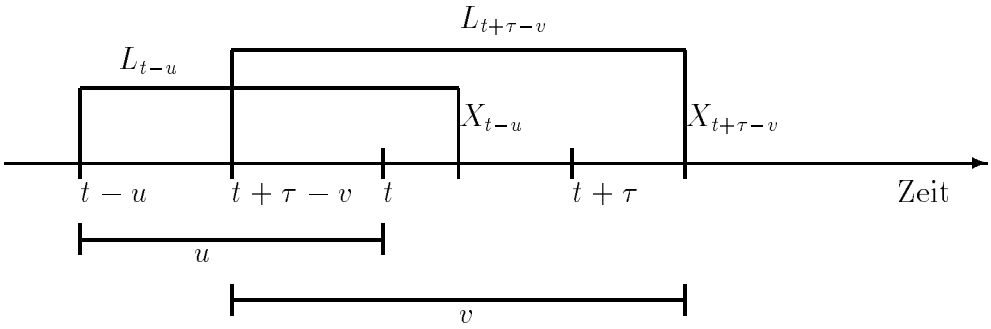
$$\textbf{(1.4)} \quad EY(t) = EX\, E(dN(t)) \int_0^\infty \mathcal{F}_L(u) du = \frac{\lambda\,\nu}{\eta\,\xi}\,.$$

Daraus kann die Autokovarianzstruktur von $Y(t)$ berechnet werden. Skizze 5.1 veranschaulicht die Rechnungen.

Für $\tau > 0$ sei

$$\textbf{(1.5)} \quad c_Y(\tau) := \mathrm{Cov}(Y(t), Y(t+\tau))$$

Abbildung 5.1: **Skizze zur Kovarianzstruktur**



Zur Zeit $t-u$ beginnt die linke Zelle mit Länge L_{t-u} und Intensität X_{t-u}. Zur Zeit $t+\tau-v$ beginnt eine Zelle mit Länge $L_{t+\tau-v}$ und Intensität $X_{t+\tau-v}$.

$$= \mathrm{E}\Big(Y(t)Y(t+\tau)\Big) - \mathrm{E}Y(t)\ \mathrm{E}Y(t+\tau)\ .$$

Der erste Summand berechnet sich als

$$\textbf{(1.6)} \qquad \mathrm{E}\Big(Y(t)Y(t+\tau)\Big) =$$

$$\underbrace{\int_0^\infty \int_0^\infty}_{u\neq v} \mathrm{E}\Big(X_{t-u}(u)\, X_{t+\tau-v}(v)\ \cdot I(L_{t-u} > u \cap L_{t+\tau-v} > v)\Big)$$

$$\cdot\ \mathrm{P}(dN(t+\tau-v) = 1 |\, dN(t-u) = 1)\, \lambda\, \nu\, dv\, du$$

$$+ \int_0^\infty \mathrm{E}\Big(X^2\, I(L_{t-u} > \tau + u)\Big) \lambda\, \nu\, du\ .$$

Der erste Summand bezieht sich auf zwei unterschiedliche Zellen, von denen eine zum Zeitpunkt $t-u$ beginnt und die andere zum Zeitpunkt $t+\tau-v$, bei gegebener Zelle mit Start in $t-u$. Die Dauern der zur Zeit $t-u$ bzw. $t-v$ startenden Zellen werden mit L_{t-u} bzw. L_{t-v} bezeichnet. Der zweite Summand beschreibt die Erwartung, daß eine Zelle in $t-u$ beginnt und sowohl zur Zeit t als auch zur Zeit $t+\tau$ noch aktiv ist.

Die bedingte Intensitätsfunktion $h(t)$ eines stationären Punktprozesses wird in Cox und Isham [1979, S. 32] gegeben,

$$\textbf{(1.7)} \qquad h(t) := \lim_{\delta_1,\delta_2\to 0+} \frac{1}{\delta_2}\, \mathrm{P}(N(t, t+\delta_2) > 0 | N(-\delta_1, 0) > 0)\ t > 0\ .$$

Für einen Neyman-Scott-Clusterprozess wird $h(t)$ in COX UND ISHAM [1979, S. 78] hergeleitet:

$$\textbf{(1.8)} \quad h(t) = \lambda\,\nu + \frac{E(C(C-1))}{\nu}\int_{-\infty}^{\infty} f_D(x)\, f_D(x+t)dx$$

$$= \lambda\,\nu + \frac{1}{2}\frac{E(C(C-1))}{\nu}\beta e^{-\beta t}\,, \quad t > 0\,,$$

wobei $f_D(x) = \beta\, e^{-\beta x}$ die Dichte der Abstände D zwischen dem Sturmursprung und den Zellstarts ist.

Damit berechnet sich der erste Summand in (1.6) mittels der Fallunterscheidung $u > \tau - v$ und $u < \tau - v$ und einiger Umformungen als

$$\begin{aligned}
& \lambda\frac{\nu}{\xi^2}\Big[\int_{u=0}^{\infty} e^{-\eta\,\mu}\int_{v=0}^{\tau+u} e^{-\eta v}h(\tau+u-v)dv\,du \\
& + \int_{u=0}^{\infty} e^{-\eta\,\mu}\int_{v=\tau+u}^{\infty} e^{-\eta v}h(v-\tau-u)dv\,du\Big] \\
= \; & \frac{\lambda^2\,\nu^2}{\eta^2\xi^2} - \frac{1}{2}\frac{\lambda\,\beta}{\xi^2}\,E(C^2-C)\,e^{-\beta\tau}/(\beta^2-\eta^2) \\
& + \frac{1}{2}\frac{\lambda\,\beta^2}{\eta\,\xi^2}\,E(C^2-C)\,e^{-\eta\tau}/(\beta^2-\eta^2)\,.
\end{aligned}$$

Der zweite Summand in (1.6) ergibt

$$\int_0^{\infty} E(X^2\,I(L>\tau+u))\lambda\,\nu\,du = \lambda\,\nu\,E(X^2)e^{-\eta\tau}/\eta\,.$$

Da die Zellintensitäten unterschiedlicher Zellen nach Annahme unabhängig und identisch verteilt sind, ist

$$\mathrm{E}Y(t)\;\mathrm{E}Y(t+\tau) = \frac{\lambda^2\,\nu^2}{\eta^2\,\xi^2}\,,$$

so daß die Autokovarianz bei Lag τ insgesamt

$$\textbf{(1.9)} \quad c_Y(\tau) = \lambda\,\nu E(X^2)e^{-\eta\tau}/\eta \;+\; \frac{1}{2}\,\lambda\,\frac{\beta^2}{\eta\,\xi^2}\,E(C^2-C)\,e^{-\eta\tau}/(\beta^2-\eta^2)$$

$$-\frac{1}{2}\,\lambda\,\frac{\beta}{\xi^2}\,E(C^2-C)\,e^{-\beta\tau}/(\beta^2-\eta^2)$$

beträgt. Insbesondere berechnet sich die Varianz des Intensitätsprozesses, indem in (1.9) $\tau = 0$ gesetzt wird:

$$\textbf{(1.10)} \qquad \mathrm{Var}(Y(t)) = \frac{\lambda}{\eta}\,\nu E(X^2) + \frac{1}{2}\frac{\lambda}{\eta}\,\beta\, E(C^2 - C)/(\xi(\beta+\eta))\ .$$

Die Eigenschaften des Prozesses der aggregierten Niederschlagsmengen lassen sich nun berechnen. Es bezeichne $Y_i^{(h)}$ die Gesamtniederschlagsmenge im i-ten Intervall der Länge h. Aus der Definition des Modelles folgt $\mathrm{E}(X^2) = 2/\xi^2$ sowie $\mathrm{E}(C^2 - C) = \nu^2 - 1$.

Damit folgt für den Prozeß der aggregierten Niederschlagsmengen $\{Y_i^{(h)}\}_{i\in\mathbb{N}}$:

$$\textbf{(1.11)} \qquad \mathrm{E}(Y_i^{(h)}) = \frac{h\lambda\nu}{\eta\xi}\int_{(i-1)h}^{i\,h} Y(t)\,dt$$

Die Kovarianz berechnet sich als

$$\textbf{(1.12)} \qquad \mathrm{Cov}(Y_i^{(h)}, Y_{i+k}^{(h)}) = \int_{k\,h}^{(k+1)h}\int_0^h \mathrm{Cov}(Y(t), Y(s))\,dt\,ds\ .$$

Wird (1.9) in (1.12) eingesetzt und der Wert des Integrales berechnet, folgt für $k \geq 1$ das Ergebnis

$$\begin{aligned}\textbf{(1.13)} \qquad \mathrm{Cov}(Y_i^{(h)}, Y_{i+k}^{(h)}) &= \frac{\lambda}{\eta^3}\,(1-e^{-\eta h})^2\ e^{-\eta(k-1)\,h} \\ &\cdot\Big(\nu\ \mathrm{E}(X^2) + \frac{1}{2}\,\mathrm{E}(C^2 - C)\ \beta^2/(\xi^2(\beta^2-\eta^2))\Big) \\ &- \frac{1}{2}\,\lambda\,(1-e^{-\beta h})^2\ e^{-\beta(k-1)\,h}\ \ \mathrm{E}(C^2-C)/\Big(\xi^2\beta(\beta^2-\eta^2)\Big)\ .\end{aligned}$$

Insbesondere ist die Lag(1)-Autokovarianz

$$\begin{aligned}\textbf{(1.14)} \qquad \mathrm{Cov}(Y_i^{(h)}, Y_{i+1}^{(h)}) = \frac{\lambda}{\xi^2}\Big\{&(1-e^{-\eta h})^2\cdot\Big[2\nu + \frac{1}{2}\,\beta^2\,\frac{\nu^2-1}{\beta^2-\eta^2}\Big]/\eta^3 \\ &- \frac{1}{2}(1-e^{-\beta h})^2\ \frac{\nu^2-1}{\beta^2-\eta^2}/\beta\Big\}\ .\end{aligned}$$

Wird in (1.12) $k = 0$ gesetzt, errechnet sich die Varianz unter Berücksichtigung der Fallunterscheidung $t > s$ und $t < s$ als

$$\begin{aligned}\mathrm{Var}(Y_i^{(h)}) &= \int_{t=0}^h\int_{s=0}^h \mathrm{Cov}(Y(t), Y(s))ds\,dt \\ &= 2\int_{t=0}^h\int_{s=0}^t \mathrm{Cov}(Y(t), Y(s))ds\,dt\ .\end{aligned}$$

Wir erhalten

$$\textbf{(1.15)} \quad \mathrm{Var}(Y_i^{(h)}) = \frac{\lambda}{\eta^3}(\eta h - 1 + e^{-\eta h})\Big[2\,\nu\,\mathrm{E}(X^2) + \mathrm{E}(C^2 - C)\,\beta^2/(\xi^2(\beta^2 - \eta^2))\Big]$$
$$- \lambda(\beta h - 1 + e^{-\beta h})\ \mathrm{E}(C^2 - C)/(\xi^2\beta\,(\beta^2 - \eta^2)\,)$$
$$= \frac{\lambda}{\xi^2} \cdot \Big[(\eta h - 1 + e^{-\eta h}) \cdot \Big(4\,\nu + \beta^2 \frac{\nu^2 - 1}{\beta^2 - \eta^2}\Big)/\eta^3$$
$$-(\beta h - 1 + e^{-\beta h})\ \frac{\nu^2 - 1}{\beta^2 - \eta^2}/\beta\Big]\ .$$

Der Lag(1)-Autokorrelationskoeffizient ist damit

$$\textbf{(1.16)} \quad \mathrm{Cor}(Y_i^{(h)}, Y_{i+1}^{(h)}) =$$

$$\frac{(1 - e^{-\eta h})^2 \cdot \Big[2\nu + \frac{1}{2}\,\beta^2\,\frac{\nu^2-1}{\beta^2-\eta^2}\Big]/\eta^3 - \frac{1}{2}(1 - e^{-\beta h})^2\ \frac{\nu^2-1}{\beta^2-\eta^2}/\beta}{(\eta h - 1 + e^{-\eta h}) \cdot \Big(4\,\nu + \beta^2\frac{\nu^2-1}{\beta^2-\eta^2}\Big)/\eta^3 - (\beta h - 1 + e^{-\beta h})\ \frac{\nu^2-1}{\beta^2-\eta^2}/\beta}.$$

Herleitung der Wahrscheinlichkeit trockener Intervalle beliebiger Länge

Die Wahrscheinlichkeit, daß ein beliebiges Intervall der Länge h trocken ist, wird von COWPERTWAIT [1991] hergeleitet. Sie beträgt:

$$\textbf{(1.17)} \quad \mathrm{P}(Y_i^{(h)} = 0) = \exp\Big[-\lambda h + \frac{\lambda}{\beta(\nu - 1)}\Big(1 - \exp[1 - \nu + (\nu - 1)\ e^{-\beta h}]\Big)$$
$$-\lambda \int_0^\infty (1 - p_h(t))dt\Big]$$

mit

$$p_h(t) = \Big(e^{-\beta(t+h)} + 1 - \frac{\eta e^{-\beta t} - \beta e^{-\eta t}}{\eta - \beta}\Big)$$
$$\cdot \exp\Big\{(\nu - 1)\,[e^{-\beta(t+h)} - e^{-\beta t} - \beta\,\frac{e^{-\beta t} - e^{-\eta t}}{\eta - \beta}]\Big\}\ .$$

Zur Herleitung dieser Wahrscheinlichkeit wird zunächst ein einziger Sturm betrachtet, ohne die Effekte der Zellen anderer Stürme zu berücksichtigen. Mit der Annahme, $C-1$ sei poissonverteilt mit Erwartungswert $\nu-1$, wird die Existenz mindestens einer Zelle pro Cluster gesichert. Unter der Annahme, daß sich eine Zelle am Clusterzentrum bei $t = 0$ befindet, wird die Wahrscheinlichkeit bestimmt, daß ein Intervall

$[t,\, t+h]$ trocken ist. Der Faktor $(1-e^{-\eta\, t})$ in Cowpertwaits Formel (5) beschreibt die Wahrscheinlichkeit, daß die Zelle am Clusterzentrum vor der Zeit t endet. Anschließend wird auf den Fall verallgemeinert, daß sich nicht notwendigerweise eine Zelle am Clusterzentrum befindet. Bei der Berechnung der Wahrscheinlichkeit $\mathrm{P}(k$ Zellanfänge in $[a,\, b] |\, C = n)$ müssen dabei die drei Fälle $k = 0,\ 0 < k \leq n,\ k > n$ unterschieden werden. Mit dieser Fallunterscheidung folgt Cowpertwaits Formel (6b). Abschließend wird auf mehrere Stürme verallgemeinert, die sich überlagern können. Die von Cowpertwait aufgeführten Folgerungen führen zu Formel (1.17).

Abkürzend werden wir folgende Bezeichnungen verwenden:

$$\mu(h) := \mathrm{E}(Y_i^{(h)}), \quad \gamma(h) := \mathrm{Var}(Y_i^{(h)}), \quad \gamma(h,k) := \mathrm{Cov}(Y_i^{(h)}, Y_{i+k}^{(h)}).$$

Die Autokorrelationsfunktion sei

$$\rho(h,k) := \mathrm{Cor}(Y_i^{(h)}, Y_{i+k}^{(h)}) = \gamma(h,k)/\gamma(h)\ .$$

Die Wahrscheinlichkeiten trockener bzw. nasser Intervalle seien

$$\phi(h) := \mathrm{P}(Y_i^{(h)} = 0)\ ,$$

$$\phi_{DD}(h) := \mathrm{P}(Y_{i+1}^{(h)} = 0 | Y_i^{(h)} = 0)\ , \quad \phi_{WW}(h) := \mathrm{P}(Y_{i+1}^{(h)} > 0 | Y_i^{(h)} > 0)$$

Die Übergangswahrscheinlichkeiten berechnen sich folgendermaßen (COWPERTWAIT [1996a]):

(1.18) $\phi_{DD}(h) = \phi(2h)/\phi(h)$,

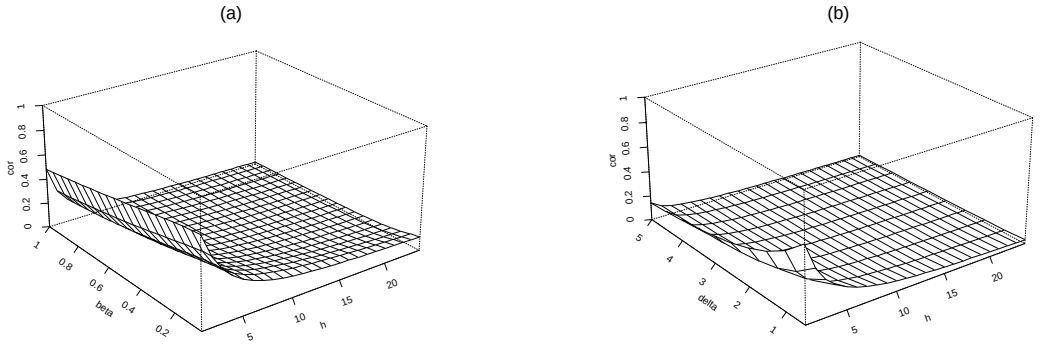
(1.19) $\phi(h) = \phi_{DD}(h)\ \phi(h) + (1-\phi_{WW}(h))(1-\phi(h))$ bzw.

(1.20) $\phi_{WW}(h) = (1 - 2\phi(h) + \phi(2h))/(1-\phi(h))$.

Vor einer Anpassung des Modelles an Beobachtungen sollte überprüft werden, ob die Struktur der Beobachtungen mit der Modellstruktur übereinstimmen kann.

Deshalb schätzen wir das Verhalten der Second-Order-Eigenschaften als Funktion der Intervalldauer h ab, ohne die formalen Ableitungen von (1.15) - (1.20) zu nennen, die leicht durch formales Differenzieren erhalten werden. Der Erwartungswert erfüllt $\mu(k \cdot h) = k \cdot \mu(h), k \in \mathbb{R}_+$, für den in 3.(3.1) definierten Schätzer gilt das ebenfalls, so daß wir nur $\mu(1)$ betrachten werden. Aus (1.15) ist ersichtlich, daß $\gamma(h)$ bei zunehmender Intervalldauer h zunimmt. Das Verhalten von $\rho(h,1)$ bei Änderung von h hängt hingegen von den Werten der übrigen Parameter ab, insbesondere von β und η. Es wird analytisch aus Gleichung (1.14) und (1.15) mittels der ersten und

Abbildung 5.2: Lag(1)-Autokorrelation als Funktion der Intervalldauer



(a) $\rho(h,1)$ als Funktion von β und h mit $\lambda = 0.01, \nu = 2, \delta = 0.5,\ \xi = 1$.
(b) $\rho(h,1)$ als Funktion von δ und h mit $\lambda = 0.01, \nu = 2, \beta = 0.5, \xi = 1$.
h: Länge des Aggregationsintervalles, $\delta := \eta - \beta$.

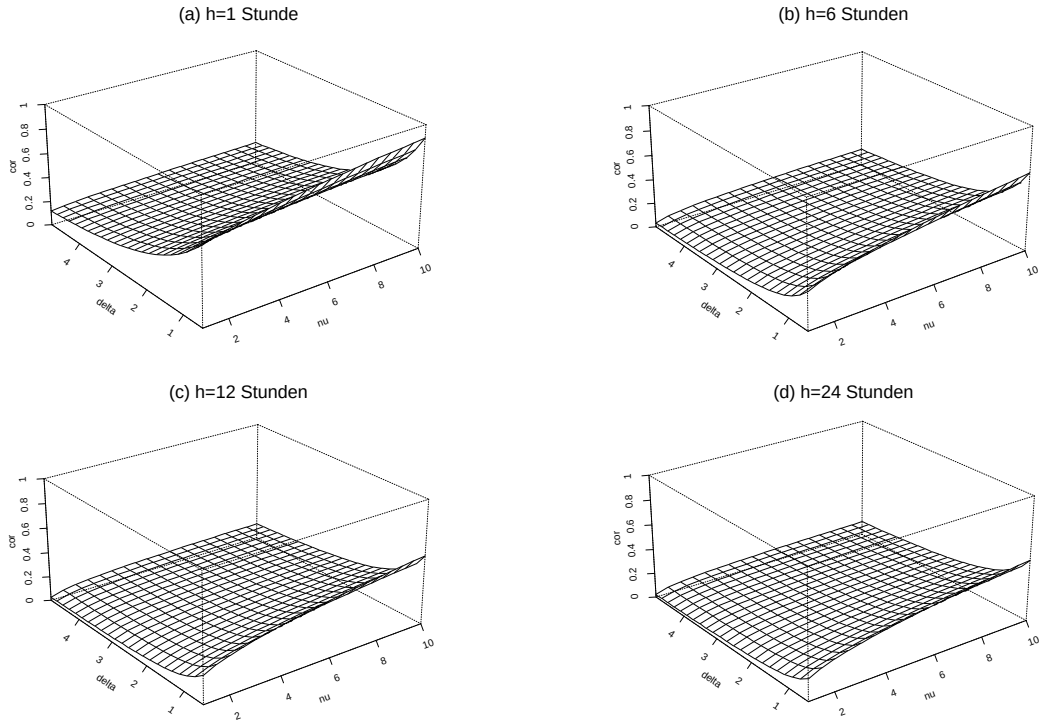
---

zweiten Ableitungen hergeleitet. Anschaulicher sind dreidimensionale Graphiken, in denen jeweils zwei Modellparameter variieren. Als Funktion von zwei Parametern ist der Lag(1)-Autokorrelationskoeffizient $\rho(h,1)$ stetig und monoton. Gemäß (1.16) haben λ und ξ keinen Einfluß auf die Autokorrelation. Der Einfluß einer Änderung von ν ist ebenfalls gering, weil ν im Zähler und im Nenner in den gleichen Potenzen auftritt. Aus dem gleichen Grund besitzen Änderungen von β und η (jeweils ceteris paribus) keinen großen Einfluß (kleinere Werte dieser Parameter führen zu etwas höheren Korrelationen, vergl. Abb. 5.2. Wichtiger ist die Differenz $\delta := \eta - \beta$. Kleine Werte von δ führen zu höherer Korrelation. In Abbildung 5.2 sind exemplarisch zwei Schnitte durch die Korrelationslandschaft dargestellt. Das Verhalten von $\rho(h,1)$ in Abhängigkeit von δ und ν ist für Aggregationsintervalle der Längen $h = 1, 6, 12, 24$ in Abbildung 5.3 dargestellt. Die Ebenen besitzen ähnliche Verläufe, sind aber in der Höhe verschoben. Für größere Werte von h wird $\rho(h,1)$ kleiner.

Aus (1.17) und (1.20) folgt, daß die Wahrscheinlichkeit $\phi(h)$ eines trockenen Intervalles der Länge h streng monoton fallend ist und $\phi_{WW}(h)$ streng monoton wachsend. Die Werte der Modellparameter $\lambda, \nu, \beta, \delta$ beeinflussen die Lage dieser Funktionen. Die Steigung von $\phi_{WW}(h)$ wird hingegen nur von den Werten von λ und β beeinflußt: Größere Werte von λ bzw. kleinere Werte von β bewirken größere Werte von $|\partial\,\phi_{WW}(h)/\partial\,h|$. Um das abzuschätzen, ohne die Ableitung formal zu nennen, wird beachtet, daß in der Formel für $\phi(h)$ nur Produkte von h mit λ und β auftreten. Größere Werte von β bewirken, daß $|\partial\,p_h(t)/\partial\,h\,|$ größer ist. Der Effekt im zweiten Summanden von $\phi(h)$ ist zwar gegenläufig, aber er wird durch größere Werte von β mehr abgeschwächt als durch kleinere Werte (β auch im Nenner). Das bedeutet, daß größere Werte von β zu größeren Änderungen von $|\partial\,\phi(h)/\partial\,h\,|$ und damit auch von $|\partial\,\phi_{WW}(h)/\partial\,h\,|$ führen. Größere Werte von λ bewirken eindeutig größere Werte von $|\partial\,\phi(h)/\partial\,h\,|$ und damit auch von $|\partial\,\phi_{WW}(h)/\partial\,h\,|$. Das Verhalten von $\phi(h)$ und $\phi_{WW}(h)$ als Funktion von λ und h ist beispielhaft in Abbildung 5.4 zu sehen, als Funktion von β und h in Abbildung 5.5.

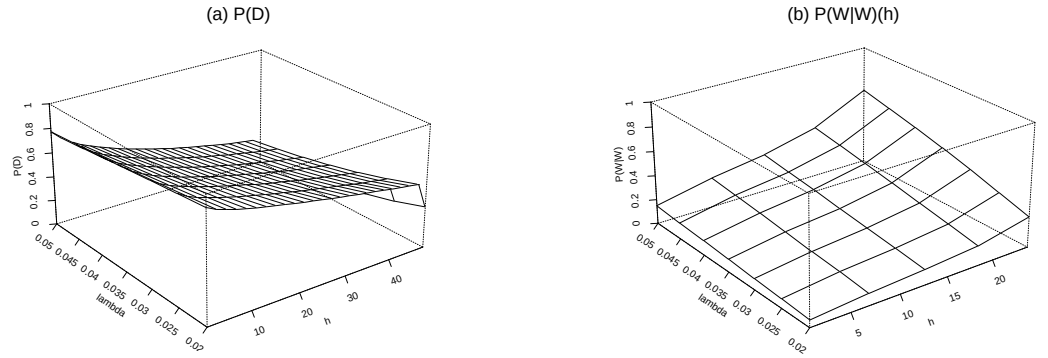
Der NSRP-Prozeß ist also einerseits hinreichend flexibel, um unterschiedliche Struk-

Abbildung 5.3: Lag(1)-Autokorrelation als Funktion von δ und ν



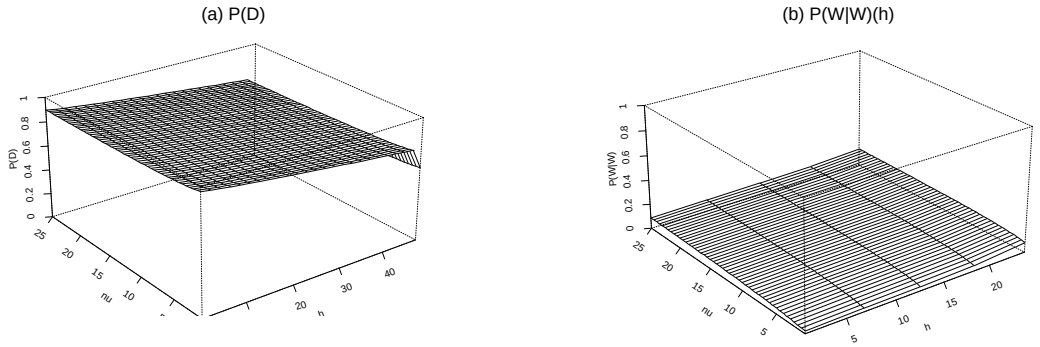
$\rho(h,1)$ als Funktion von δ und ν, wobei h: Länge des Aggregationsintervalles.
In allen Abbildungen: $\lambda = 0.0063, \beta = 0.225, \xi = 0.413$, linke Achse: $\delta = \eta - \beta$, rechte Achse: ν.
(a) $h = 1$. (b) $h = 6$. (c) $h = 12$. (d) $h = 24$.

Abbildung 5.4: $\phi(h)$ und $\phi_{WW}(h)$ als Funktionen der Modellparameter

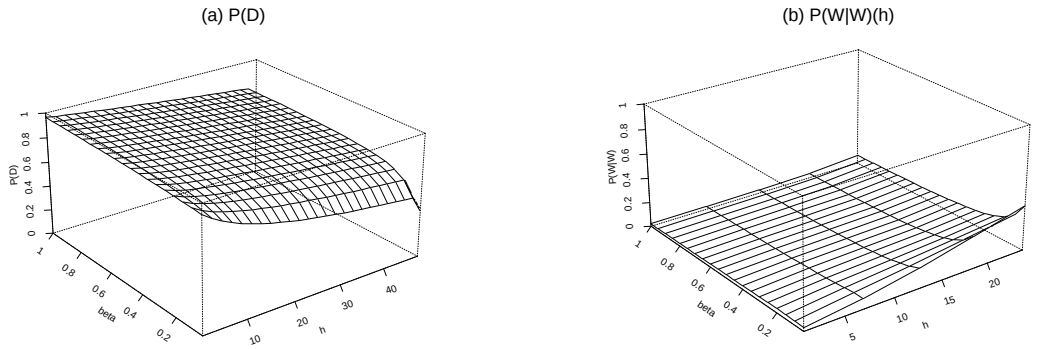


(a) $\phi(h)$, (b) $\phi_{WW}(h)$ als Funktionen von h und λ.
In beiden Abbildungen: $\nu = 8, \beta = 0.4, \delta = 5, \xi = 1$, linke Achse: λ, rechte Achse: h.

Abbildung 5.5: $\phi(h)$ und $\phi_{WW}(h)$ als Funktionen der Modellparameter

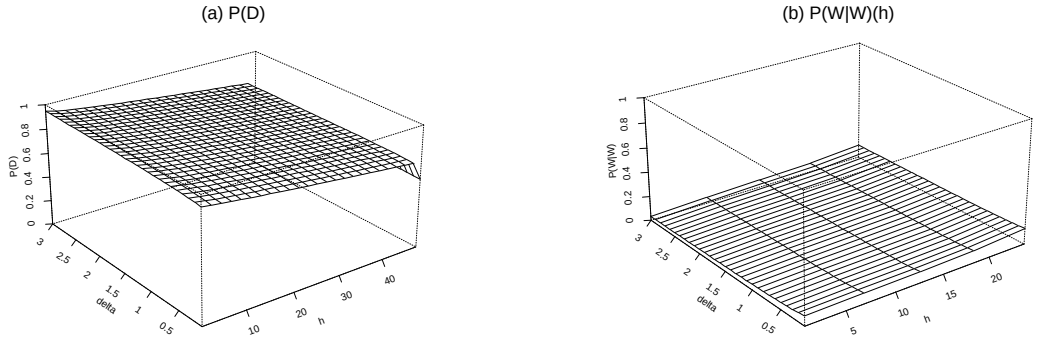


(a) $\phi(h)$, (b) $\phi_{WW}(h)$ als Funktionen von h und ν.
Linke Achse: ν, rechte Achse: $h.\lambda = 0.01, \beta = 0.4, \delta = 0.5, \xi = 1$.



(a) $\phi(h)$, (b) $\phi_{WW}(h)$ als Funktionen von h und β.
Linke Achse: β, rechte Achse: h. $\lambda = 0.01, \nu = 8, \delta = 5, \xi = 1$,

---



(a) $\phi(h)$, (b) $\phi_{WW}(h)$ als Funktinen von h und δ.
In beiden Abbildungen: $\lambda = 0.01, \nu = 8, \beta = 0.4, \xi = 1$, linke Achse: δ, rechte Achse: h.

---

turen von Niederschlägen beschreiben zu können, andererseits sind aber die Auswirkungen von Parameteränderungen schwierig abzuschätzen.

2 Modellanpassung für eine homogene Periode

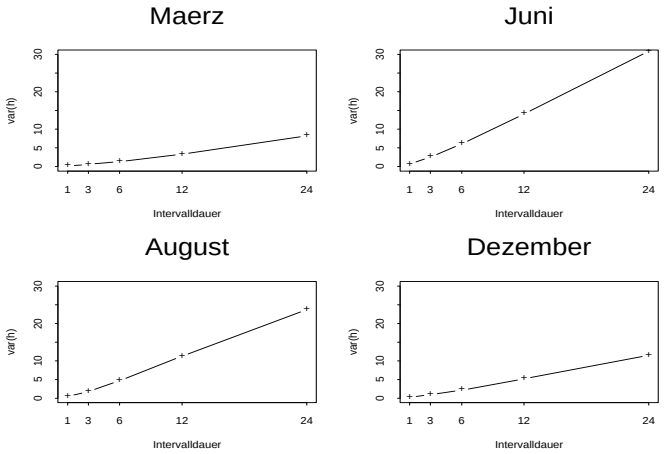
Um vor der eigentlichen Modellanpassung abzuschätzen, ob die Beobachtungen mit einem NSRP-Modell kompatibel sind, betrachten wir in Abbildung 5.6 die Struktur einiger empirischer Momente als Funktion der Länge h des Aggregationsintervalles für ausgewählte Monate. Die Momente werden für die zwölf Kalendermonate separat geschätzt und während eines Kalendermonates wird der Prozeß als stationär betrachtet. Damit werden einerseits saisonale Schwankungen berücksichtigt, andererseits wird innerhalb der Kalendermonate eine Glättung der Beobachtungen vorgenommen, vergl. Abschnitt 3.2.
Die empirischen Varianzen nehmen mit zunehmender Länge des Aggregationsintervalles h in allen Monaten zu, was mit der Varianzstruktur des NSRP-Modelles übereinstimmt, vergl. Abb. 5.6. Die Lag(1)-Autokorrelationskoeffizienten weisen in den dargestellten Monaten unterschiedliche Verläufe auf. Unter dem NSRP-Modell ist $\rho(h,1)$ eine monoton abfallende Funktion von h, wie im vorigen Abschnitt beschrieben wurde. Der Kurvenverlauf im August mit $\rho(12,1) < \rho(24,1)$ kann als leichtes Indiz auf innertägliche Zyklen interpretiert werden. Die Kurvenverläufe im März und im Dezember, deren Maxima bei Aggregationsintervallen der Länge $h = 6$ bzw. $h = 3$ Stunden liegen, stimmen mit Beobachtungen überein, daß in dieser Jahreszeit langandauernde Regenfelder auftreten, die kurze trockene Phasen eingebettet haben. Eventuell hätte dieses Phänomen beseitigt werden können, wenn auch nichtmeßbarer Niederschlag (Niederschlag, dessen Intensität so gering ist, daß er die Gesamtmenge an einem Tag nicht erhöht, sog. Niederschlagsspuren) gesondert von den Meßgeräten erfaßt worden wäre. Die empirischen Übergangswahrscheinlichkeiten $\hat{\phi}_{WW}(h)$ sind als Funktion von h monoton wachsend, was mit dem NSRP-Modell übereinstimmt.

Die Werte der Modellparameter müssen aus den beobachteten aggregierten Niederschlagsmengen geschätzt werden, da weder die Clusterzentren noch die einzelnen Zellen beobachtet werden können. Ein geschlossener Ausdruck für die Verteilung der aggregierten Niederschlagsmengen $Y_i^{(h)}$ scheint noch nicht gefunden worden zu sein, müßte aber auch wegen der zu berücksichtigenden Autokorrelation der aggregierten Niederschlagsmengen kompliziert sein. Die Modellanpassung kann deshalb nicht mit der Maximum-Likelihood-Methode durchgeführt werden, sondern wird auf den Momenten der aggregierten Niederschlagsmengen basieren.

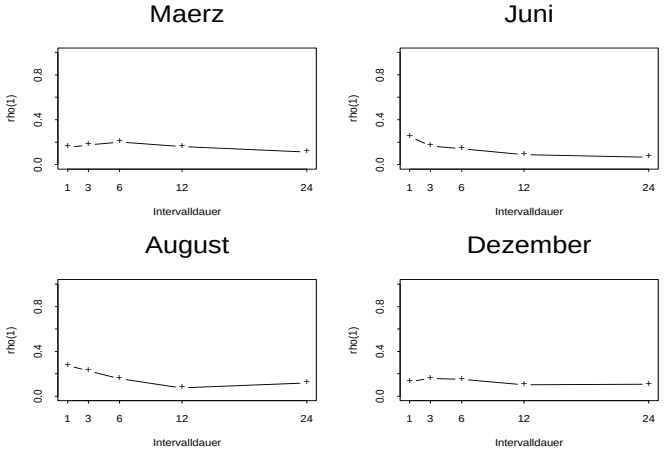
Mit der Methode der Momente werden die fünf Modellparameter geschätzt, indem exakt fünf Momente des Mengenprozesses mit ihren empirischen Äquivalenten verglichen werden, also ein nicht-lineares Gleichungssystem gelöst wird. Damit würden die verwendeten Momente exakt angepaßt. Eine andere Möglichkeit besteht in einem Kleinste-Quadrate-Ansatz, in dem eine beliebige Anzahl m an Momenten berücksichtigt werden kann. Aus Gründen der Modellauswahl ist es sinnvoller, mehr Second-Order-Eigenschaften des Mengenprozesses approximativ anzupassen als nur

Abbildung 5.6: Empirische Momente als Funktion der Intervalllänge

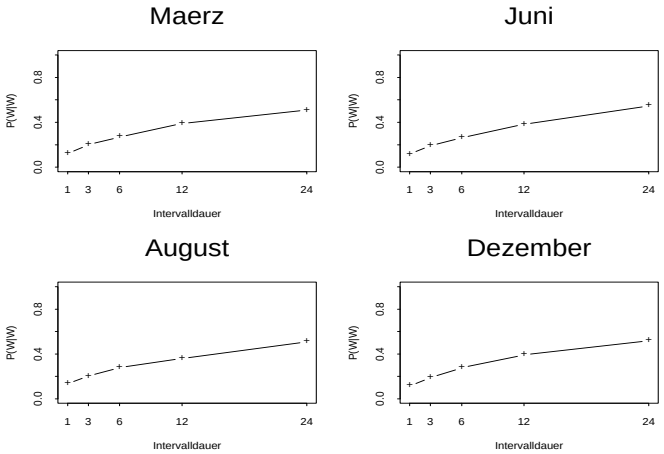
Empirische Varianz $\hat{\gamma}(h)$



Empirische Lag(1)-Autokorrelation $\hat{\rho}(h,1)$



Empirische Übergangwahrscheinlichkeiten $\hat{\phi}_{WW}(h)$



Empirische Momente der Niederschlagsmenge $Y_i^{(h)}$ in Intervallen der Längen $h = 1, 3, 6, 12, 24$.

einige Eigenschaften exakt, um den Fehler durch Schätzung zu reduzieren. Ferner soll die Anpassungsmethode ohne Ableitungen auskommen, weil diese aufgrund der Form der Gleichungen (1.13) - (1.17) kompliziert sind. Deshalb entscheiden wir uns für einen Kleinste-Quadrate-Ansatz. Minimiert wird

$$\textbf{(2.1)} \quad S = \sum_{i=1}^{m} w_i \left(1 - \frac{f_i}{\hat{f}_i}\right)^2 ,$$

wobei die $f_i = f_i(\lambda, \nu, \beta, \eta, \xi)$ Momente des Prozesses der aggregierten Mengen unter dem Modell sind, wie beispielsweise $\gamma(h$. Die $\hat{f}_i$ sind die entsprechenden, aus den Beobachtungen geschätzten Werte dieser Eigenschaften. Der Term in den Klammern ist eine Standardisierung der Residuen, damit große numerische Werte das Schätzverfahren nicht dominieren. Die Gewichtungsfaktoren w_i legen fest, welche Eigenschaften des Modelles besonders gut angepaßt werden.

Weil die partiellen Ableitungen der Second-Order-Momente unhandliche Ausdrücke sind, sollte die Minimierungsroutine ohne Ableitungen auskommen. Wir haben uns deshalb für die vielseitige und robuste Downhill-Simplex-Methode nach Nelder und Mead entschieden. Der verwendete Minimierungsalgorithmus ist der Sammlung numerischer Algorithmen von PRESS et al. [1992] entnommen und leicht modifiziert worden, vergl. Anhang C.1. Die Downhill-Simplex-Methode bewegt zur Minimierung einer Funktion $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ ein nicht-entartetes Simplex im $\mathbb{R}^n$, also $n+1$ Punkte des $\mathbb{R}^n$, in ein Tal der Funktionslandschaft und zieht es schließlich um ein (lokales) Minimum zusammen. Die verwendeten Bewegungen sind Verschiebungen, Spiegelungen und Kontraktionen. Wenn sich die Funktionswerte an den Simplexecken nur noch um einen geringen Anteil des Funktionswertes im Zielbereich unterscheiden, bricht die Routine ab. Da unsere Zielfunktion S als Minimum den Wert null besitzt, unterscheiden sich die Funktionswerte an den Ecken des Zielsimplizes nur noch minimal von null. Ist die Zielfunktion nicht monoton wie in unserem Fall, so läuft die Minimierungsroutine von unterschiedlichen Startparametern in unterschiedliche Täler, was in unterschiedlichen Zielsimplizes resultiert. Instabilität der Lösung entsteht ebenfalls, wenn die Zielfunktion S in der Nähe des Minimums flach ist (was auch vom verwendeten Parameterset abhängt). Die Downhill-Simplex-Methode führt zu besseren Ergebnissen, wenn alle Parameter annähernd die gleiche Größenordnung haben.
Die Ergebnisse der Minimierung hängen also von folgenden Faktoren ab:

1. dem verwendeten Parameterset $\mathcal{P}$,
2. den Startwerten der Parameter in der Minimierung,
3. den Gewichten w_i,
4. der Mengeneinheit der Daten ([0.1mm] bzw. [1mm]),
5. einer Transformation der Modellparameter.

Die Mengeneinheit der Daten beeinflußt die Einheit des Parameters ξ und damit dessen Zahlenwert. Die Mengeneinheit [1mm] führt zu kleineren Zahlenwerten des

Erwartungswertes und der Varianzen als die Einheit [0.1mm] und damit zu größeren Werten von ξ (da $\mathrm{E}X = 1/\xi$), was sich als numerisch vorteilhafter herausstellt.
Um in den Gleichungen (1.14) und (1.15) Division durch null zu vermeiden, führen wir als weitere Restriktion $\eta > \beta$ ein und definieren $\delta := \eta - \beta > 0$, so daß im Mittel der Abstand eines Zellstartes vom Clusterzentrum größer als die Zelldauer ist – eine physikalisch plausible Annahme. Die zu schätzenden Modellparameter sind im folgenden $\psi := (\lambda, \nu, \beta, \delta, \xi)$. Werden alle Modellparameter vor der Minimierung logarithmiert (und hinterher zurücktransformiert), müssen die Parameterrestriktionen $\lambda, \beta, \delta, \xi > 0$, $\nu > 1$ nicht beachtet werden, was Ecklösungen in der numerischen Minimierung verhindert.
Der Wert des in der Formel (1.17) für $\phi(h)$ auftretenden Integrales

$$\int_0^\infty (1 - p_h(t))\, dt$$

wird numerisch mit der Trapezregel unter Verwendung der Algorithmen aus PRESS et al. [1992] bestimmt. Analytische Abschätzungen und Ausprobieren verschiedener Parameterkombinationen zeigen, daß der Wert des Integrales von 40 bis ∞ null ist, so daß die Berechnung des uneigentlichen Integrales, z.B. unter Verwendung einer Transformation der Integrationsgrenzen, unterbleiben kann. Alle Programmcodes sind in Anhang B.1 gegeben und dort ausführlich beschrieben.

Mit der Auswahl der zum Parameterschätzen verwendeten Statistiken haben sich RODRIGUEZ-ITURBE et al. [1988] und ausführlicher KHALIQ et al. [1996] beschäftigt. Die Auswahl des Sets der verwendeten Eigenschaften ist schwierig. Sie sollten geringe Meßfehler aufweisen und unkorreliert sein. Die zitierten Untersuchungen verschiedener Sets zeigen, daß die Berücksichtigung einiger Eigenschaften die exakte Wiedergabe anderer Eigenschaften nahezu ausschließt und daß eine Verwendung von mehr als maximal sieben Eigenschaften in einer insgesamt schlechten Anpassung resultiert. Das weist darauf hin, daß der Niederschlagsprozeß nicht perfekt durch das NSRP-Modell beschrieben wird. Damit das resultierende NSRP-Modell den Niederschlagsprozeß bei unterschiedlichen Zeiteinteilungen korrekt wiedergibt, müssen Momente in der Modellanpassung berücksichtigt werden, die sich auf verschieden lange Aggregationsniveaus beziehen. Die Ergebnisse von RODRIGUEZ-ITURBE et al. [1984] für das Neyman-Scott-White-Noise-Modell sind also nicht auf das Neyman-Scott-Rechteckimpuls-Modell übertragbar.

Bevor das Modell an die Beobachtungen des gesamten Jahres angepaßt wird, untersuchen wir anhand der Perioden Januar und Juli Eigenschaften unterschiedlicher Parametersets und Gewichtungsfaktoren.

Auswahl eines Parametersets

Das in der Modellanpassung verwendete Set $\mathcal{P}$ von Momenten beeinflußt die Ergebnisse der Minimierungsroutine und damit die Qualität der Anpassung. Um ein geeignetes Set auszuwählen, passen wir das NSRP-Modell mit verschiedenen Parameterkombinationen an die Januar- und Juli-Beobachtungen an. Für eine erste

Überprüfung der Qualität der Anpassung vergleichen wir die Werte der Eigenschaften $f \in \mathcal{F}$ unter dem Modell mit ihren empirischen Gegenstücken, wobei

$$\textbf{(2.2)} \quad \mathcal{F} = \{\mu(1),\ \gamma(1),\ \gamma(3),\ \gamma(6),\ \gamma(12),\ \gamma(24),$$
$$\rho(1,1),\ \rho(3,1),\ \rho(6,1),\ \rho(12,1),\ \rho(24,1),$$
$$\phi_{WW}(1),\ \phi_{WW}(3),\ \phi_{WW}(6),\ \phi_{WW}(12),\ \phi_{WW}(24),$$
$$\phi_{DD}(24),\ \phi(24)\}\ .$$

Wir betrachten in $\mathcal{F}$ die Lag(1)-Autokorrelationen für längere Intervalle und nicht die Lag(k)-Autokorrelationen für $k > 1$, weil erstere für die kurzfristige Vorhersage wichtiger sind, analog für die Übergangswahrscheinlichkeiten aufeinanderfolgender Intervalle. Damit berücksichtigen wir, daß Niederschlagsmessungen in unterschiedlichen zeitlichen Aggregationsniveaus vorliegen können. Auf die Modellierung der Übergangswahrscheinlichkeiten nasser Intervalle legen wir größeren Wert als auf die trockener Intervalle, weil erstere für die Wiedergabe des tatsächlichen Niederschlagsprozesses in Verbindung mit der Modellierung der Niederschlagsmenge wichtiger sind. Die Formeln (1.18) bis (1.20) zeigen, daß die Wahrscheinlichkeit $\phi(h)$ eines trockenen Intervalles in $\mathcal{P}$ enthalten sein sollte, damit $\phi_{DD}(h)$ und $\phi_{WW}(h)$ adäquat angepaßt werden.

Bei der Verwendung unterschiedlicher Sets zeigt sich:

1. Enthält $\mathcal{P}$ zwar Varianzen $\gamma(h)$, aber keine Korrelationen $\rho(h,1)$, so werden letztere unterschätzt.

2. Eine korrekte Wiedergabe der Autokorrelationsstruktur $\rho(h,1)$ für $h = 1,3,6,$ $12,24$ ist nur möglich, wenn $\rho(h,1) \in \mathcal{P}$ für alle h.

3. Enthält $\mathcal{P}$ zwar $\rho(h,1)$ für mehrere Werte von h, aber keine Varianzen $\gamma(h)$, so werden letztere unterschätzt.

4. Ist keine der Übergangswahrscheinlichkeiten $\phi_{WW}(h)$ in $\mathcal{P}$ enthalten, so werden alle $\phi_{WW}(h)$, $h = 1,3,6,12,24$ überschätzt.

5. Gilt $\gamma(h)$, $\rho(h,1)$, $\phi_{WW}(h) \in \mathcal{P}$ für $h = 1,6,24$, so werden diese korrekt wiedergegeben, aber die Übergangswahrscheinlichkeiten $\phi_{WW}(h)$ werden unterschätzt und steigen als Funktion von h zu langsam an.

6. Der Erwartungswert $\mu(1)$ sollte in (2.1) stärker gewichtet werden als die übrigen Parameter, da er sonst unterschätzt wird. Einheitliches Gewichten der verwendeten Inputparameter mit $w = 1$ bei einer stärkeren Gewichtung von $\mu(1)$ mit $w = 10$ ergibt die besten Resultate.

7. Gilt $\mathcal{P} = \mathcal{F}$, so werden in Abhängigkeit von den Startparametern entweder $\gamma(h)$ und $\phi_{WW}(h)$ oder $\gamma(h)$ und $\rho(h,1)$ stark unterschätzt. Insbesondere werden die Veränderungen der Momente in Abhängigkeit vom Aggregationsniveau

h nicht wiedergegeben. Das gilt für unterschiedliche Werte der Gewichtsfaktoren w_i.

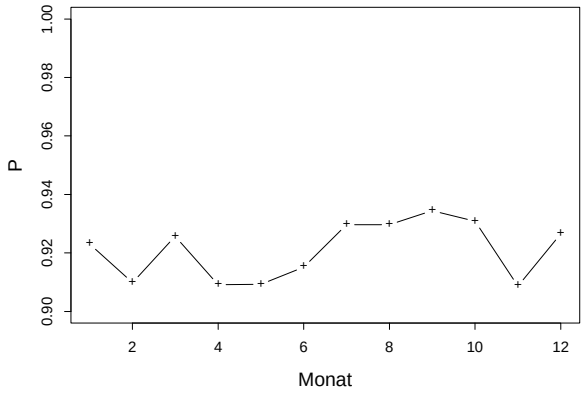
Fazit:
Eine optimale Wiedergabe aller in $\mathcal{F}$ enthaltenen Second-Order-Eigenschaften mit dem Neyman-Scott-Modell ist nicht möglich. Eine gute Anpassung einiger (mehr als fünf) Eigenschaften ist jedoch möglich, so daß eine Beschreibung stündlichen Regens im Jahresverlauf mit diesem Modell aussichtsreich erscheint.
Es zeigt sich, daß unabhängig von den Startparametern der Minimierungsroutine Parametersets P mit mehr als 6 Momenten schlechtere Ergebnisse ergeben, in denen keines der Momente, außer stark gewichteten, adäquat wiedergeben wird. Das Parameterset $\mathcal{P}$ sollte also $\mu(1)$ und $\phi(24)$ enthalten, außerdem eine Varianz $\gamma(h)$, einen Autokorrelationskoeffizient $\rho(h,1)$ und eine Übergangswahrscheinlichkeit $\phi_{WW}(h)$. Die letzten drei Eigenschaften sollten sich auf unterschiedliche Zeitniveaus beziehen. Bemerkenswert ist, daß die mit einem solchen Set ermittelten Modellparameter Werte im physikalisch plausiblen Bereich besitzen, wie beispielsweise $\lambda = 0.001013[Std^{-1}]$, $\nu = 4.503519$, $\beta = 0.010292[Std^{-1}]$, $\eta = 2.468206[Std^{-1}]$, $\xi = 0.084206[Std\ mm^{-1}]$.

3 Saisonale und innertägliche Variation

Der in (1.1) definierte Neyman-Scott-Prozeß ist ein Modell für stündlichen Niederschlag, das saisonale Variation integriert, weil die Modellparameter stetige Funktionen der Zeit sind. Daher bestünde ein Schätzverfahren darin, die Werte der Modellparameter $\psi_t = (\lambda_t, \nu_t, \beta_t, \eta_t, \xi_t)$ für jede Stunde t, $t = 1, \ldots, 24 \cdot 365 = 8760$, des Jahres separat zu schätzen und die resultierenden Parameterwerte unter Verwendung von a-priori Informationen wie der geforderten Stetigkeit oder jahreszyklischem Verhalten zu glätten, analog zur Vorgehensweise bei der Beschreibung täglichen Regens. Da aber je nach Monat 90% bis 93% aller Stunden trocken sind, vergl. Abb. 5.7, können die zum Schätzen der zu einer Stunde t gehörenden Modellparameter benötigten empirischen Momente nicht nur aus den in allen Jahren beobachteten Stunden t geschätzt werden. Ein analoges Vorgehen zu dem in der Analyse von täglichem Regen ist also nicht möglich. Stattdessen schätzen wir die Parameter aus den Beobachtungen eines Zeitintervalles, das t als Mittelpunkt hat und das so lang ist, daß einerseits genügend nasse Stunden auftreten, andererseits aber der Niederschlagsprozeß während der Intervalldauer als stationär angenommen werden kann. Wir entscheiden uns für Intervalle einer Länge von 30 Tagen, innerhalb derer der Niederschlagsprozeß, wie in Abschnitt 3.2 gezeigt, als stationär angenommen werden kann. Beobachtungen des 29. Februars werden nicht berücksichtigt.

Das NSRP-Modell wird auf die in Abschnitt 2 beschriebene Weise an alle 30 Tage (= 720 Stunden) umfassenden Perioden eines Jahres angepaßt. Idealerweise beginnen die Perioden im Abstand von einer Stunde. Der dafür benötigte Rechen- und Speicheraufwand ist jedoch sehr groß, so daß wir das Modell nur an alle im Abstand von sechs Stunden beginnenden, nämlich um 0h, 6h, 12h, 18h, und 30 Tage

Abbildung 5.7: Empirische Wahrscheinlichkeit einer trockenen Stunde $\hat{\phi}(h = 1)$



$$\hat{\mathrm{P}}(Y_t^{(h=1)} = 0) = \frac{\#\{\text{beob. trockenen Stunden im Monat}\}}{\#\{\text{ Std. im Monat}\}}$$

Tabelle 5.1: 30-Tages-Perioden in der NSRP-Modellanpassung

Nr.	Start		Ende		Mittelpunkt	
	Tag	Uhrzeit	Tag	Uhrzeit	Tag	Uhrzeit
1	1.1.,	0:00h	30.1.,	23:59h	15.1.,	12:00h
2	1.1.,	6:00h	30.1.,	5:59h	15.1.,	18:00h
3	1.1.,	12:00h	30.1.,	11:59h	16.1.,	0:00h
4	1.1.,	18:00h	30.1.,	17:59h	16.1.,	6:00h
5	2.1.,	0:00h	1.2.,	23:59h	16.1.,	12:00h
etc.						

umfassenden Perioden anpassen, vergl. Tabelle 5.1. Wir erhalten damit Vektoren $\psi_t = (\lambda_t, \nu_t, \beta_t, \eta_t, \xi_t) \in I\!R^5$, $t = 1, \ldots, 4 \cdot 365 = 1460$. Die empirischen Momente sind in Kapitel 3, Abb. 3.9 - 3.16 dargestellt. Innertägliche Zyklen werden nicht vollständig ausgeglättet, z.B. lassen die Varianzen $\gamma_t(24), t = 1, \ldots, 1460$, vier unterschiedliche Kurven erkennen, die mit den vier Startzeiten der Perioden korrespondieren.

Die so erhaltenen Schätzer ψ_t werden wir durch Fouriersummen glätten, aus denen Werte für die übrigen Stunden berechnet werden können. Da sich die Perioden überlappen, sind die empirischen Momente der stündlichen Niederschlagsmengen aufeinanderfolgender Perioden korreliert und deshalb auch die Schätzer der Modellparameter. Weil die Parameter stetige Funktionen der Zeit sein sollen, verwenden wir als Startsimplex zum Schätzen der Parameter einer Periode die Parameterschätzer der 6 vorhergehenden Perioden. Dadurch variieren die Schätzer $\hat{\psi}_t \in I\!R^5$, $t = 1, \ldots, 1460$, als Funktionen der Zeit annähernd stetig. Der Algorithmus ist in Anhang C detailliert beschrieben. Die in der Modellanpassung verwendeten Second-Order-Momente $\mathcal{P}$ wählen wir unter Berücksichtigung der Ergebnisse des vorigen Abschnittes aus. Um zu untersuchen, ob die Werte der Modellparameter davon abhängen, mit welcher Periode die Anpassung beginnt, starten wir die zyklische Anpassung für alle betrach-

teten Sets $\mathcal{P}$ einmal in Periode 1 (Beginn am 1. Januar, 0:00h) und einmal in Periode 91 (Beginn am 1. April, 0:00h). Nach jeder Anpassung werden Restarts durchgeführt, die als Startparameter die leicht variierten Ergebnisparameter des vorherigen Programmdurchlaufes verwenden. Bei der zyklischen Anpassung werden häufig physikalisch unplausible Parameterwerte erhalten, wie Werte von $\delta := \eta - \beta > 100$, die bei der Anpassung an nur eine Periode nicht aufgetreten sind. Dieses Problem ließ sich durch Einführung einer Straffunktion verhindern; wir bevorzugten jedoch häufigere Restarts mit veränderten Startparametern, um keine weiteren Restriktionen des Lösungsraumes einzuführen. Wir passen das Modell unter Verwendung folgender Sets $\mathcal{P}$ an:

1. $\mathcal{P} = \{\mu(1), \sqrt{\gamma(1)}, \rho(24), \phi_{WW}(1), \phi_D(24)\}$,
2. $\mathcal{P} = \{\mu(1), \sqrt{\gamma(24)}, \rho(1), \phi_{WW}(1), \phi_D(24)\}$,
3. $\mathcal{P} = \{\mu(1), \sqrt{\gamma(1)}, \rho(24), \phi_{WW}(24), \phi_D(24)\}$,
4. $\mathcal{P} = \{\mu(1), \sqrt{\gamma(1)}, \rho(24), \phi_{WW}(1), \phi_{WW}(24), \phi_D(24)\}$,
5. $\mathcal{P} = \{\mu(1), \sqrt{\gamma(1)}, \sqrt{\gamma(24)}, \phi_{WW}(24), \phi_D(24)\}$, (Set A)
6. $\mathcal{P} = \{\mu(1), \gamma(1), \gamma(24), \phi_{WW}(24), \phi_D(24)\}$,
7. $\mathcal{P} = \{\mu(1), \gamma(1), \rho(1), \phi_{WW}(24), \phi_D(24)\}$,
8. $\mathcal{P} = \{\mu(1), \gamma(1), \rho(24), \phi_{WW}(24), \phi_D(24)\}$. (Set B)

Das Ein-Stunden-Mittel wird jeweils mit $w = 10$ gewichtet, alle anderen Momente mit $w = 1$. Die besten Resultate werden unabhängig von der verwendeten Startperiode mit dem 5. und dem 8. Set erzielt, die wir als Set A bzw. Set B bezeichnen. Da die Startperiode die Qualität der Anpassung nicht entscheidend beeinflußt, starten wir die zyklische Anpassung im folgenden stets mit Periode 1 (1.1., 0:00h - 30.1., 23:59h). Unter Verwendung der Sets A und B führen wir die Modellanpassung mit unterschiedlichen Startsimplizes durch. Als Startsimplex werden

(a) die Modellparameter von COWPERTWAIT et al. [1996a],

(b) die Modellparameter von RODRIGUEZ-ITURBE et al. [1984],

(c) unabhängige gleichverteilte Zufallszahlen aus $[0, 1]^5$

gewählt und um fünf weitere Vektoren $\psi \in I\!R^5$ ergänzt.

Die Qualität einer Anpassung an T Perioden messen wir mit den Residuenquadratsummen S_F in der Anpassung und S_E in der vorläufigen Evaluation,

$$S_F = \sum_{t=1}^{T} S_{F,t} = \sum_{t=1}^{T} \sum_{f \in \mathcal{P}} w_f \left(1 - \frac{f_t}{\hat{f}_t}\right)^2 ,$$

$$S_E = \sum_{t=1}^{T} S_{E,t} = \sum_{t=1}^{T} \sum_{f \in \mathcal{F}} w_f \left(1 - \frac{f_t}{\hat{f}_t}\right)^2 ,$$

wobei weiterhin f den Wert eines Moments unter dem Modell und $\hat{f}$ dessen empirischen Schätzer bezeichnen.

Da alle empirischen Momente, die sich auf das Aggregationsniveau $h = 24$ Stunden beziehen, innertägliche Zyklen aufweisen, führen wir die zyklische Modellanpassung mit diesen Parametersets zusätzlich für jede der vier Periodenstartzeiten separat durch. Daraus resultieren für jeden der 5 Modellparameter vier getrennte Kurvenverläufe, die sich zum Teil beträchtlich unterscheiden. In der folgenden Tabelle sind die Residuenquadratsummen der jeweils besten Anpassungen bei gesonderter Anpassung für die vier Startzeiten (Zeilen 1 - 5) sowie bei einer Anpassung (Zeile 6) für alle 1460 Perioden zu sehen.

	Set A		Set B	
Startzeit	S_F	S_E	S_F	S_E
0 h	12.37	3898.75	11.91	3852.13
6 h	5.37	3598.06	12.18	2460.89
12 h	6.05	7667.52	8.64	3560.21
18 h	4.25	3720.02	15.43	2865.47
Summe	28.05	31085.03	48.17	12738.70
Alle Perioden	114.55	20648.93	135.61	19565.83

Die Unterschiede zwischen S_E und S_F verdeutlichen den einzugehenden Kompromiß zwischen der exakten Anpassung einiger Eigenschaften und der approximativen Anpassung einer größeren Anzahl von Eigenschaften. Eine gesonderte Modellanpassung für jede der Startzeiten ist flexibler und resultiert deshalb in einer kleineren Residuenquadratsumme. Obwohl Set A die bessere Performance in bezug auf die Modellanpassung besitzt, wird mit Set B die bessere Anpassung über alle Second--Order-Eigenschaften aus Set $\mathcal{F}$ erreicht.
Weil wir möglichst viele der Eigenschaften des Niederschlagsprozesses erfassen wollen, verwenden wir Set B in der weiteren Modellanpassung. Im nächsten Abschnitt werden wir die mit Set B (unter Verwendung von vier Startsimplizes bzw. einem Startsimplex) erhaltenen Modellparameter $\psi_t, t = 1, \ldots, T$, glätten und anschließend die Parameterwerte für alle Stunden t eines Jahres berechnen, $t = 1, \ldots, 24 \cdot 365 = 8760$. Damit ist das in (1.1) definierte Modell angepaßt.

4 Glättung der Modellparameter – Set B4

Auf die im vorigen Abschnitt beschriebene Weise haben wir für jede der Perioden t, $t = 1, \ldots, T$ mit $T = 1460$, Parameterschätzer $\hat{\psi}_t = (\hat{\lambda}_t, \hat{\nu}_t, \hat{\beta}_t, \hat{\delta}_t, \hat{\xi}_t) \in I\!R^5$ erhalten. Jede der Komponenten von $\psi = (\psi_t) \in I\!R^{5\times T}$ soll eine stetige und im Jahresverlauf zyklische Funktion der Zeit sein. Deshalb ist es naheliegend, jede der Komponenten durch eine Fouriersumme zu glätten. Da die Ergebnisse der Modellanpassung in starkem Maße von dem verwendeten Startsimplex abhängen und die aus den geschätzten Modellparametern resultierenden Second-Order-Eigenschaften nicht in allen Perioden mit ihren empirischen Gegenstücken übereinstimmen, kann eine Glättung der Modellparameter auch zu einer besseren Anpassung führen. Insbesondere verringert eine Glättung der Komponenten von $\hat{\psi}$ die Anzahl der Modellparameter und damit den Fehler durch Schätzung.
Die Vorgehensweise ist für beide Parametersets (1 Startsimplex bzw. 4 Startsimplizes) und für jede der Komponenten von $\hat{\psi}$ gleich und wird deshalb nur für λ detailliert beschrieben, während für die anderen Parameter nur die Ergebnisse dargestellt werden.

Wird das Neyman-Scott-Modell jeweils an alle 30-Tages-Perioden mit Start um 0h, 6h, 12h, 18h auf die in Abschnitt 3 beschriebene Weise separat angepaßt, erhalten wir vier Sets von Modellparametern $\psi_{t,i}, t = 1, \ldots, 365, i = 0, 6, 12, 18$h. Diese Sets lassen sich im Reißverschlußverfahren zusammensetzen, so daß ein Parametervektor $\psi = (\psi_1, \ldots, \psi_T) \in I\!R^{5\times T}$ entsteht, $T = 1460$. ψ_t korrespondiert mit der an Tag $d = [t/4]$ in Stunde $s = ((t-1) \bmod 4) + 1$ beginnenden Stunde, wobei $s = 1, 2, 3, 4$ die Startuhrzeiten 0, 6, 12, 18h bezeichnet. $[x]$ ist die größte ganze Zahl, die kleiner oder gleich x ist. Dieses Parameterset B4 enthält nach Konstruktion innertägliche Zyklen. Zusätzlich weisen alle Modellparameter saisonale Zyklen auf, vergl. die Graphiken am Ende dieses Abschnittes.

Die Komponenten von ψ weisen einige (lokale) Ausreißer auf. Weil der Schätzalgorithmus nicht stabil ist, werden sie vor der weiteren Analyse durch das Mittel der benachbarten Werte ersetzt. Aus den dann angepaßten Fouriersummen berechnen sich Eigenschaften des aggregierten Niederschlagsprozesses mit geringerem mittlerem quadratischen Fehler als aus den Fouriersummen, die unter Einbeziehung der Ausreißer ermittelt werden.

Beschreibung von λ

Die Werte von $\hat{\lambda}_t, t = 1, \ldots, T$, schwanken zwischen einem Minimum von 0.0020 [Std^{-1}] im Januar und einem Maximum von 0.0311[Std^{-1}] im Juni, bei einem Median von 0.0088[Std^{-1}] und einer Standardabweichung von 0.0039[Std^{-1}], vergl. Abb. 5.9. Es sind vier relativ parallel verlaufende Kurven erkennbar, die innertäglichen Zyklen entsprechen, und außerdem saisonale Schwankungen.

Die Fouriersummendarstellung der $\lambda_t, t = 1, \ldots, T$, ist

$$\lambda_t = R_0 + \sum_{0<j<T/2} R_j \cos(\omega_j t + \theta_j) + R_{T/2} \cos(\pi t), \quad t = 1, \ldots, T,$$

mit $\omega_j = 2\pi j/T$. Die Werte der Amplituden R_j und der Phasen θ_j werden mit einer Fouriertransformation in einer harmonischen Analyse berechnet. Zur Vorgehensweise vergleiche man BLOOMFIELD [1976, Kap. 3 bis 5] oder CHATFIELD [1996, Kap. 6 und 7]. Die Parameter R_j und θ_j können mit einer Fast–Fourier–Transformation bestimmt werden, in S-PLUS mit `fft`. Dieser Befehl liefert die Koeffizienten in der komplexen Form $z_j = a_j + ib_j \in \mathbb{C}$, wobei $R_j = |z_j|$ (S-PLUS: `Mod`) und $\theta_j = \arg(z_j)$ (S-PLUS: `Arg`).

Parsevals Theorem

$$\textbf{(4.1)} \quad \frac{1}{T} \sum_{t=1}^{T} (x_t - \bar{x})^2 = \sum_{1<j<T/2} R_j^2/2 + R_{T/2}^2$$

zeigt, daß die Frequenzen ω_j mit den größten Amplituden $|R_j|$ den größten Beitrag zur Stichprobenvarianz liefern. Das Spektrum von λ besitzt keine singulären Peaks an einzelnen Frequenzen, vergl. Abb. 5.8 (a). Beispielsweise sind die Peaks an den Fourierkomponenten 364 und 365 annähernd gleichgroß und die ersten fünf harmonischen Komponenten besitzen annähernd die gleiche Stärke. Offensichtlich ist nicht eine einzige Frequenz dominant und es wäre deshalb irreführend, die Frequenzen gemäß der Größe ihrer Amplituden zu berücksichtigen. Das geglättete Spektrum zeigt deutlicher, welche Frequenzen berücksichtigt werden sollten. Die Frequenz ω_{730} ist die Nyquist-Frequenz, also die größte in den Daten erkennbare Frequenz, die einer Periode von 12 Stunden entspricht. Die Frequenz ω_{365} entspricht einer Periode von 24 Stunden, ω_1 beschreibt den Jahreseffekt. Der Peak bei $\omega_j, j > 650$, resultiert aus dem Peak bei ω_{365} und zeigt außerdem, daß die Modellparameter in kürzeren Abständen hätten geschätzt werden müssen, um die gesamte in den Daten enthaltene Variation zu erfassen.

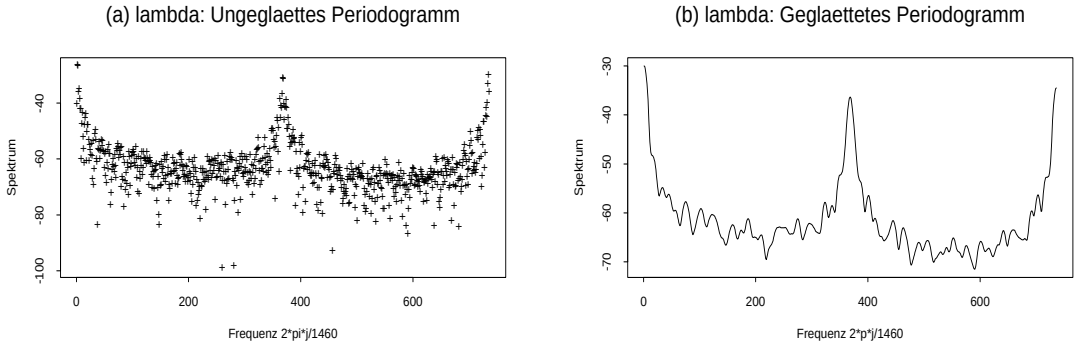
Wir bestimmen deshalb zunächst Frequenzen ω_{j_l}, $l = 1, \ldots, L$, mit den L größten Amplituden (z.B. L=20) und berechnen sukzessive die angepaßten Werte $\tilde{\lambda}_t, t = 1, \ldots, T$, unter Modellen $M_k, k = 1, \ldots, L$:

$$\textbf{(4.2)} \quad M_k : \tilde{\lambda}_{t,k} = R_0 + \sum_{l=1}^{k} R_{j_l} \cos(\omega_{j_l} t + \theta_{j_l}), \quad k = 1, \ldots, L, \; t = 1, \ldots, T,$$

wobei zunächst die niedrigen Frequenzen $\omega_j, j = 1, 2, 3, \ldots$, dann die hohen Frequenzen ω_j, z.B. $j = 365, 730$, einbezogen werden. Es wird das Modell M_k ausgewählt, bei dem die Residuenquadratsumme

$$\textbf{(4.3)} \quad S_k = \sum_{t=1}^{T} (\hat{\lambda}_t - \tilde{\lambda}_{t,k})^2$$

Abbildung 5.8: Periodogramm von λ



Set B4: (a) ungeglättetes und (b) geglättetes Periodogramm der Schätzer $\hat{\lambda}_t, t = 1, \ldots, 1460$.

---

als Funktion von k einen Wendepunkt besitzt, also weniger steil abfällt. Ab dieser Parameteranzahl bringt die Verwendung weiterer Fourierfrequenzen nur noch eine geringe Verbesserung der Anpassung mit sich.

Ein weiteres Kriterium zur Auswahl der einzubeziehenden Frequenzen sind die mittleren relativen Abweichungen $S_{f,k}$ zwischen dem Wert einer Second-Order-Eigenschaft $f_{t,k} \in \mathcal{F}$ unter Modell M_k für Periode t und dem empirischen Schätzer $\hat{f}_t$, also

$$\textbf{(4.4)} \qquad S_{f,k} = \frac{1}{T} \sum_{t=1}^{T} \frac{\sqrt{(\hat{f}_t - f_{t,k})^2}}{|\hat{f}_t|},$$

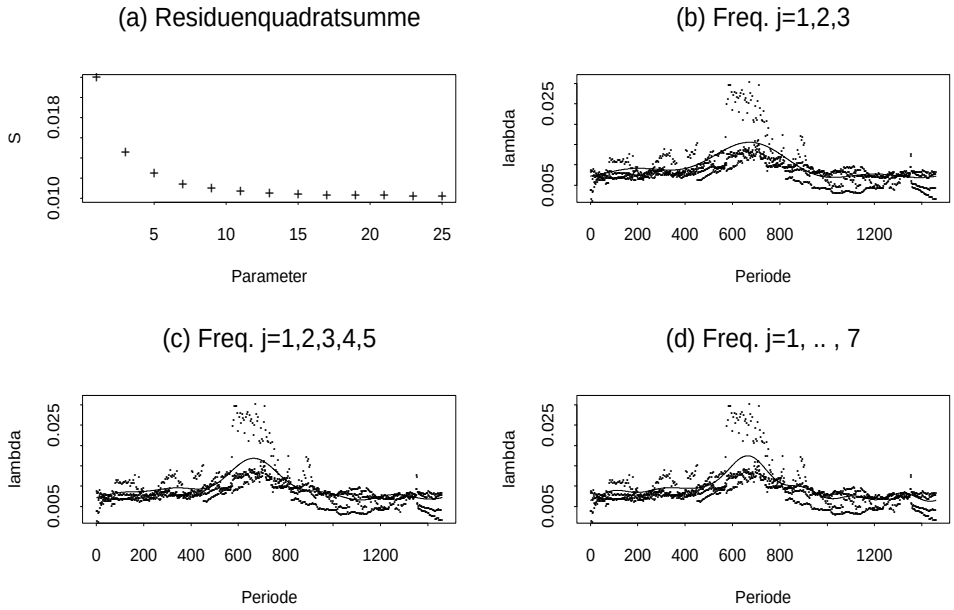
wobei $f_{t,k}$ unter dem Parametersatz $\hat{\psi} = (\hat{\psi}_t)$ berechnet wird, in dem $\hat{\lambda}_t$ durch den geglätteten Wert $\tilde{\lambda}_{t,k}, t = 1, \ldots, T$, unter Modell M_k ersetzt wird. Wäre f_t ein stationärer Prozeß, wäre (4.4) ein Schätzer des Variationskoeffizienten und damit ein übliches Maß zur Beurteilung der Qualität einer Schätzung.

Die Qualität der Anpassung könnte auch auf andere Weise beurteilt werden, beipielsweise mit der mittleren Residuenquadratsumme

$$\frac{1}{T} \sum_{t=1}^{T} (\hat{f}_t - f_{t,k})^2 \ ,$$

die aber nicht unabhängig von der Größenskala der Momente f ist und zudem sensibel auf Ausreißer reagiert. Weil das Modell aber sensibel auf kleine Parameteränderungen reagiert, führen die geglätteten Modellparameter zu Ausreißern in den Second-Order-Eigenschaften. Deshalb könnte die Anpassung auch mit einer robusten Statistik beurteilt werden, beispielsweise mit der mittleren Residuenquadratsumme der mittleren 90% der Residuen $|\hat{f}_t - f_t|$. Durch kleinere Zahlenwerte würde dann eine bessere Anpassung suggeriert. Da aber Ausreißer in einer Anpassung nicht wünschenswert sind, sollten sie nicht von vornherein aus der Beurteilung ausgeschlossen werden. Hier ist (4.4) am besten geeignet.

Abbildung 5.9: Modellierung von λ



(a): Res.summe (4.3) $S(j)$ bei sukzessiver Anpassung der Freq. $\omega_j, j = 0, 1, 2, 3, \ldots, 12$.
(b): Schätzer $\hat{\lambda}$ und Glättung mit $\omega_j, j = 0, 1, 3$.
(c): Schätzer $\hat{\lambda}$ und Glättung mit $\omega_j, j = 0, 1, 3, 4, 5$.
(d): Schätzer $\hat{\lambda}$ und Glättung mit $\omega_j, j = 0, 1, 3, 4, 5, 6, 7$.

Aus den Residuenquadratsummen (4.3) der Glättungen mit 3 bzw. 5 Frequenzen (entsprechend 7 bzw. 11 Parametern) ist nicht ersichtlich, welche Anpassung optimal im Sinne der kleinsten Gesamtdiskrepanz ist, vergl. Abb. 5.9 (a). Die Anpassungen unterscheiden sich nur geringfügig, vergl. Abb. 5.9 (b) - (d).

Wir berechnen deshalb die Second-Order-Eigenschaften unter dem Modellparameterset B4, wobei $\hat{\lambda}_t$ durch die angepaßten Werte $\tilde{\lambda}_t$ der Glättungen G1, G2 und G3 mit Frequenzen ω_j ersetzt wird mit G1: $j = 1, 2, 3$, G2: $j = 1, 2, 3, 4, 5$ bzw. G3: $j = 1, 2, 3, 365, 730$ sowie die zugehörigen mittleren quadratischen Abweichungen (4.4), vergl. Tab. 5.2. Die Berücksichtigung von mehr als drei Frequenzen führt nur zu einer geringfügigen Verbesserung in der Anpassung der Momente. Die nicht in der Schätzung verwendeten Momente werden nicht gut angepaßt, insbesondere die Übergangswahrscheinlichkeiten $\phi_{WW}(h)$ nicht. Große Werte der Residuenquadratsummen S werden durch einzelne Ausreißer in den theoretischen Momenten verursacht, die ein Indiz für das sensible Verhalten des Modelles auf Parameteränderungen sind. Aber auch unter den ungeglätteten Parametern stimmen die theoretischen Momente nicht exakt mit den empirischen überein, vergl. Tab. 5.2 auf Seite 98. Damit glätten wir $\hat{\lambda}$ durch die Fouriersumme mit drei Frequenzen $\omega_j, j = 1, 2, 3$,

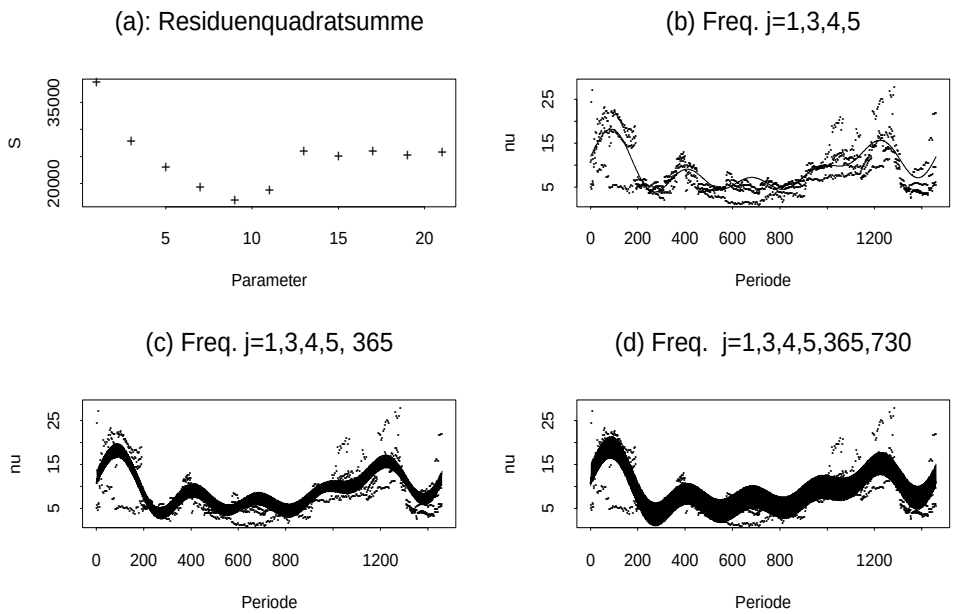
$$\tilde{\lambda}_t = \hat{R}_0 + \sum_{j=1}^{3} \hat{R}_j \cos(\omega_j t + \hat{\theta}_j)$$

j	$\hat{R}_j$	$\hat{\theta}_j$
0	$95.43 \cdot 10^{-4}$	
1	$31.82 \cdot 10^{-4}$	-2.6533
2	$16.83 \cdot 10^{-4}$	0.4418
3	$12.51 \cdot 10^{-4}$	-2.5383

Beschreibung von ν

Die Schätzer $\hat{\nu}_t, t = 1, \ldots, T$, zeigen saisonale und in geringerem Umfang auch tägliche Zyklen. Das Maximum von 28.60 wird im November angenommen, das Minimum von 1.62 im Juni. Median und Standardabweichung betragen 6.23 bzw. 5.13, vergl. Abb. 5.10(b).

Abbildung 5.10: Modellierung von ν



(a): Res.summe (4.3) $S(j)$ bei sukzessiver Anpassung der Freq. $\omega_j, j = 0, 1, 5, 4, 3, 365, 730, 2, 366, 7, 367$.
(b): Schätzer $\hat{\nu}$ und Glättung mit $\omega_j, j = 0, 1, 3, 4, 5$.
(c): Schätzer $\hat{\nu}$ und Glättung mit $\omega_j, j = 0, 1, 3, 4, 5, 365$.
(d): Schätzer $\hat{\nu}$ und Glättung mit $\omega_j, j = 0, 1, 3, 4, 5, 365, 730$.

Den größten Beitrag zum Spektrum liefern die Frequenzen $\omega_j, j = 1, 5, 4, 3, 365, 730$ in der Reihenfolge der Größe der Amplitude. Insbesondere tritt ω_2 nicht auf. Wir betrachten die Glättungen $G1$ mit den Frequenzen $\omega_j, j = 1, 3, 4, 5$ (9 Parameter), und $G2$ mit $\omega_j, j = 1, 3, 4, 5, 365, 730$ (13 Parameter). Anpassung $G1$ berücksichtigt damit keine Tageszyklen. Zur Auswahl zwischen $G1$ und $G2$ berechnen wir wieder die mittleren quadratischen Abweichungen (4.4) zwischen empirischen Second-Order-Eigenschaften und denen unter dem Modell mit den geglätteten $\tilde{\nu}_t$, vergl. Tabelle 5.3. Die Anpassung unter Set $G2$ ist trotz der größeren Anzahl an Fourierfrequenzen

schlechter als die unter Set $G1$. Wie auch im vorigen Abschnitt ist die Wiedergabe der nicht in der Anpassung verwendeten Übergangswahrscheinlichkeiten $\phi_{WW}(h)$ nicht gelungen, sie sind unter dem Modell zu groß. Wir entscheiden uns für die Glättung mit den Fourierfrequenzen $\omega_j, j = 1, 3, 4, 5$:

$$\tilde{\nu}_t = \hat{R}_0 + \sum_{j \in \mathcal{J}} \hat{R}_j \cos(\omega_j t + \hat{\theta}_j) \ , \quad \mathcal{J} = \{1, 3, 4, 5\} \ ,$$

mit

j	$\hat{R}_j$	$\hat{\theta}_j$
0	9.1498	
1	3.8510	0.4824
3	2.5832	−1.9375
4	2.2097	−1.5637
5	1.8512	−1.4119

Modellierung von β

Die ungeglätteten Schätzer $\hat{\beta}_t$, $t = 1, \ldots, T$, besitzen sowohl saisonale als auch tägliche Zyklen. Das Minimum von 0.0070 [Std^{-1}] wird im Juni und das Maximum von 0.6693 [Std^{-1}] im September angenommen. Median und Standardfehler betragen 0.0895 [Std^{-1}] bzw. 0.0646 [Std^{-1}]. Die Jahreszyklen dominieren, geringe Tageszyklen sind erkennbar, vergl. Abb. 5.11 auf Seite 100.

Die beste Anpassung der ungeglätteten Schätzer im Sinne der Residuenquadrate (4.3) zwischen $\hat{\beta}$ und einer Glättung $\tilde{\beta}$ wird erreicht, wenn sukzessive die kleinen Frequenzen $\omega_j, j = 1, 2, 3, 4, 5, \ldots$ verwendet werden. Modell $G2$ mit sieben Frequenzen ist im Hinblick auf die Second-Order-Eigenschaften unwesentlich besser als Modell $G1$ mit 5 Frequenzen, vergl. Tab. 5.3. Wir entscheiden uns deshalb für Modell $G1$. An $\hat{\beta}_t$ wird also die Summe

$$\tilde{\beta}_t = \hat{R}_0 + \sum_{j=1}^{5} \hat{R}_j \cos(\omega_j t + \hat{\theta}_j)$$

angepaßt mit

j	$\hat{R}_j$	$\hat{\theta}_j$
0	$9.391 \cdot 10^{-2}$	
1	$3.452 \cdot 10^{-2}$	3.1134
2	$1.443 \cdot 10^{-2}$	0.3813
3	$3.008 \cdot 10^{-2}$	−2.4907
4	$2.045 \cdot 10^{-2}$	0.5616
5	$1.903 \cdot 10^{-2}$	−2.3084

Beschreibung von δ

Die ungeglätteten Schätzer $\hat{\delta}_t, t = 1, \ldots, T$, variieren zwischen 0.84[Std^{-1}] im September und 135.30[Std^{-1}] im Juni bei einem Median von 13.08[Std^{-1}] und einer Standardabweichung von 10.98[Std^{-1}], vergl. Abb. 5.12. Diese Schwankungsbreite weist darauf hin, daß die Modellierung von δ schwierig sein wird. Aus den Formeln der Second-Order-Eigenschaften ist ersichtlich, daß Varianz, Autokorrelationskoeffizienten sowie Wahrscheinlichkeiten trockener Intervalle sensibel auf Änderungen von δ reagieren.

Eine gute Anpassung der Schätzer $\hat{\delta}_t$ ergibt sich bei sukzessiver Anpassung der Frequenzen $\omega_j, j = 1, 2, \ldots$. Die Kurve der Residuenquadrate verringert sich ab der Verwendung von vier Frequenzen weniger stark. Die separate Modellierung der zu den einzelnen Startzeiten 0h, 6h, 12h, 18h gehörenden Werte $\hat{\delta}_t$ durch Fouriersummen bewirkt keine nennenswerte Verbesserung. Die mit den Startzeiten korrespondierenden Teilsets würden durch 4 (0h), 5 (6h), 3 (12h) bzw. 5 (18h) Frequenzen beschrieben, also durch insgesamt 38 Parameter.
Die geglätteten Werte $\tilde{\delta}_t$ führen zu Ausreißern in den nicht in der Modellanpassung verwendeten Varianzen $\gamma(h)$ und den Übergangswahrscheinlichkeiten $\phi_{WW}(h)$, was wiederum zu großen Werten der Residuenquadrate $S_{f,k}$ führt, vergl. Tab. 5.3. Die Glättung mit 5 Frequenzen bewirkt im Vergleich zur Glättung mit 4 Frequenzen keine wesentliche Verbesserung, so daß wir die Fouriersumme

$$\tilde{\delta}_t = \hat{R}_0 + \sum_{j=1}^{4} \hat{R}_j \cos(\omega_j\, t + \hat{\theta}_j)$$

anpassen mit

j	$\hat{R}_j$	$\hat{\theta}_j$
0	13.0804	
1	4.3953	−3.1028
2	2.9747	0.0542
3	4.8002	−2.7770
4	4.4846	0.7779

Beschreibung von ξ

Die ungeglätteten Werte $\hat{\xi}_t, t = 1, \ldots, T$, nehmen ihr Minimum von 0.0075 [Std/mm] im Mai und ihr Maximumm von 1.0590 [Std/mm] im Januar an, bei einem Median von 0.1185[Std/mm] und einem Standardfehler von 0.1963[Std/mm], vergl. Abb. 5.13. Sie weisen saisonale Schwankungen auf, die je nach Periodenstartzeit unterschiedlich stark ausfallen. Innertägliche Zyklen fallen besonders im Januar und im September auf. Im Januar liegen die 18h-Schätzer mit Werten von $\hat{\xi}_t \sim 0.9$ deutlich über den Schätzern der anderen Kurven, die Werte von $\hat{\xi}_t \sim 0.4$ annehmen. Im

September liegen die 6h-Schätzer mit Werten von $\hat{\xi}_t \sim 0.8$ über denen der anderen Kurven mit $\hat{\xi}_t \sim 0.3$.

Eine befriedigende Glättung der $\hat{\xi}_t$ aller Startzeiten wird mit den Frequenzen $\omega_j, j = 1, \ldots, 5$, erzielt (11 Parameter), allerdings werden Tageszyklen dann nicht geglättet. Werden die $\hat{\xi}_t$ für die verschiedenen Periodenstartzeiten separat geglättet, werden jeweils 6 Frequenzen verwendet. Die zusätzliche Anzahl an Parametern führt aber nicht zu einer verbesserten Beschreibung der Momente der Niederschlagsmengen. Einbeziehung der Frequenzen $\omega_j, j = 365, 730$, führt zu einer schlechteren Anpassung der $\hat{\xi}_t$ und insbesondere der Second-Order-Eigenschaften des Niederschlagsprozesses. Damit werden die $\hat{\xi}_t, t = 1, \ldots, T$, geglättet durch

$$\tilde{\xi}_t = \hat{R}_0 + \sum_{j=1}^{5} \hat{R}_j \cos(\omega_j t + \hat{\theta}_j)$$

mit

j	$\hat{R}_j$	$\hat{\theta}_j$
0	0.1820	
1	0.1124	0.3494
2	0.0555	−1.8661
3	0.0838	−1.2738
4	0.0781	−1.6669
5	0.0643	−2.0194

In Tabelle 5.4 ist zu sehen, daß insbesondere die Varianzen unter den Modellen viel zu groß sind, was sich aus den Formeln (1.15) erklärt: Die Varianzen $\gamma(h)$ reagieren wegen des Faktors $1/\xi^2$ und der Größenordnung der Werte der $\hat{\xi}_t$ sensibel auf Änderungen von ξ_t. Auf die Autokorrelationsstruktur sowie die Wahrscheinlichkeiten trockener Intervalle besitzt ξ keinen Einfluß.

Das geglättete Set

Die in den vorigen Abschnitten geglätteten Komponenten der Modellparameter aus Set B4 ergeben ein geglättetes Parameterset $\tilde{\psi} = (\tilde{\lambda}, \tilde{\nu}, \tilde{\beta}, \tilde{\delta}, \tilde{\xi})$. Das so konstruierte saisonale NSRP-Modell besitzt 49 Parameter:

Parameter	Parameteranzahl	Frequenzen
λ	7	1,2,3
ν	9	1,3,4,5
β	11	1,2,3,4,5
δ	9	1,2,3,4
ξ	11	1,2,3,4,5
Gesamt	47	

Die Second-Order-Eigenschaften des Niederschlagsprozesses unter diesem Set sind in den Abbildungen auf den Seiten 103 - 110 zu sehen.

Zusammenfassung

Der Parametersatz $B4$ ist aus vier Teilsets $\hat{\psi}_d(h) := (\hat{\lambda}_d(h), \hat{\nu}_d(h), \hat{\beta}_d(h), \hat{\delta}_d(h), \hat{\xi}_d(h))$ im 'Reißverschlußverfahren' zu einem Parameterset $\hat{\psi}_t$ zusammengesetzt worden, wobei d den Starttag, $d = 1, \ldots, 365$, und h=0, 6, 12 18h die Startuhrzeit der Periode bezeichnen. Diese Konstruktion führt zwangsläufig zu Tageszyklen in den Schätzern $\hat{\psi}_t$. Dennoch besitzen die einzelnen Kurven ähnliche Verläufe, so daß Glätten des Gesamtsets möglich ist. Die Parameter nehmen physikalisch plausible Werte an.

In der Anpassung haben wir die Momente $\mu(1), \gamma(1), \rho(24), \phi_{WW}(24), \phi_D(24)$ verwendet, die unter dem Modell insgesamt gut wiedergegeben werden. Das 1-Stunden-Mittel und die Varianzen sind unter dem Modell etwas zu klein. Die Übergangswahrscheinlichkeiten $\phi_{WW}(h)$ sind unter dem Modell für Intervalle der Länge $h = 1, 3, 6, 12$ wesentlich zu groß, $\phi_{WW}(24)$ wird gut wiedergegeben, ebenso die Wahrscheinlichkeiten $\phi_{DD}(24)$ und $\phi(24)$. Die Lag(1)-Autokorrelationen werden sehr gut modelliert. Die Residuenquadrate der Momente unter dem geglätteten Gesamtset sind in Tabelle 5.4 zu sehen, Graphiken der Momente unter dem Modell und der Modellparameter nachfolgend.

Tabelle 5.2: Parameterglättung: Set B4, λ

Parameter		$\hat{f}$	ungeglättet	λ		
			S_{B4}	$S_{\lambda,1}$	$S_{\lambda,2}$	$S_{\lambda,3}$
$\mu(1)$	$(*)$	0.06	0.01	0.21	0.21	0.21
$\gamma(1)$	$(*)$	0.26	0.03	0.21	0.21	0.20
$\gamma(3)$		1.17	0.20	0.26	0.26	0.26
$\gamma(6)$		2.88	0.24	0.28	0.28	0.28
$\gamma(12)$		6.86	0.25	0.29	0.28	0.28
$\gamma(24)$		16.05	0.24	0.29	0.29	0.29
$\rho(1,1)$		0.35	0.48	0.48	0.48	0.48
$\rho(3,1)$		0.28	0.30	0.30	0.30	0.30
$\rho(6,1)$		0.23	0.27	0.27	0.27	0.27
$\rho(12,1)$		0.20	0.39	0.39	0.39	0.39
$\rho(24,1)$	$(*)$	0.17	0.04	0.04	0.04	0.04
$\phi_{WW}(1)$		0.14	1.86	1.86	1.86	1.86
$\phi_{WW}(3)$		0.21	1.42	1.43	1.43	1.43
$\phi_{WW}(6)$		0.29	1.01	1.01	1.01	1.01
$\phi_{WW}(12)$		0.41	0.51	0.52	0.52	0.52
$\phi_{WW}(24)$	$(*)$	0.58	0.09	0.10	0.10	0.10
$\phi_{DD}(24)$		0.74	0.09	0.07	0.07	0.07
$\phi_{D}(24)$	$(*)$	0.62	0.09	0.07	0.07	0.07

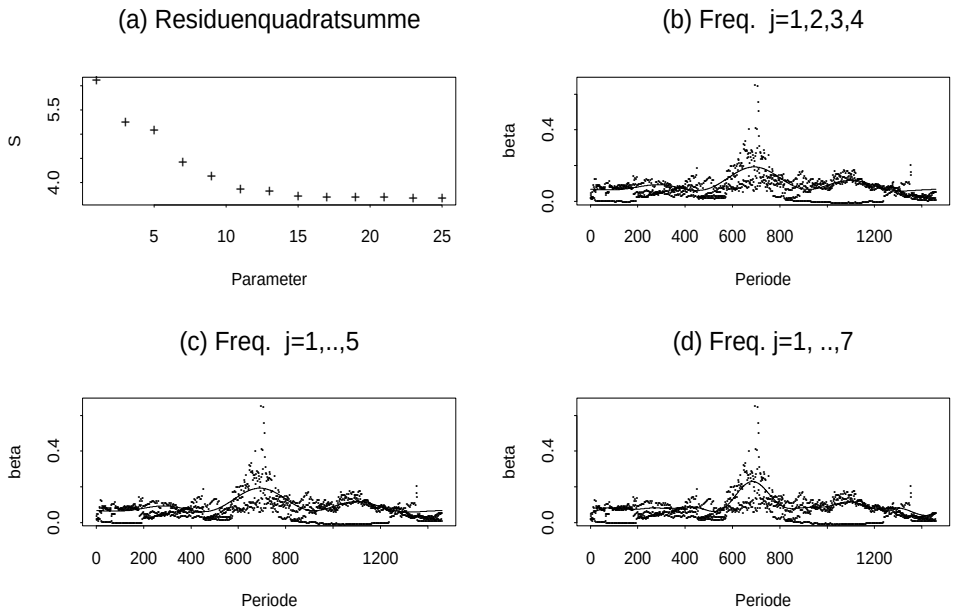
Werte der Abweichungen (4.4) $S_{f,k}$ unter Parameterset B4.
1. Spalte: Mittelwert $\bar{\hat{f}} = \sum_{t=1}^{T} \hat{f}_t / T$ des empirischen Schätzers.
2. Spalte: Ungeglättetes Set B4.
3. Spalte: $\hat{\lambda}_t$ durch Glättung mit Frequenzen $\omega_j, j = 1, 2, 3$ ersetzt.
4. Spalte: $\hat{\lambda}_t$ durch Glättung mit Freq. $\omega_j, j = 1, 2, 3, 4, 5$ ersetzt.
5. Spalte: $\hat{\lambda}_t$ durch Glättung mit Freq. $\omega_j, j = 1, 2, 3, 4, 5, 365, 730$ ersetzt.
$(*)$: Eigenschaft ist in der Modellanpassung verwendet worden.

Tabelle 5.3: Parameterglättung: ν, β, δ

Parameter		ν		β		δ	
		$S_{\nu,1}$	$S_{\nu,2}$	$S_{\beta,1}$	$S_{\beta,2}$	$S_{\delta,1}$	$S_{\beta,2}$
$\mu(1)$	$(*)$	0.33	0.47	0.01	0.01	0.48	0.47
$\gamma(1)$	$(*)$	0.37	0.51	0.04	0.04	1.07	1.04
$\gamma(3)$		0.38	0.51	0.21	0.22	0.94	0.91
$\gamma(6)$		0.39	0.51	0.25	0.25	0.93	0.91
$\gamma(12)$		0.40	0.54	0.25	0.25	0.92	0.89
$\gamma(24)$		0.39	0.54	0.24	0.24	0.92	0.90
$\rho(1,1)$		0.48	0.50	0.50	0.50	0.58	0.58
$\rho(3,1)$		0.29	0.32	0.29	0.29	0.35	0.35
$\rho(6,1)$		0.28	0.31	0.24	0.24	0.30	0.30
$\rho(12,1)$		0.37	0.34	0.44	0.45	0.40	0.40
$\rho(24,1)$	$(*)$	0.21	0.29	0.22	0.21	0.04	0.04
$\phi_{WW}(1)$		1.88	1.82	1.89	1.89	1.58	1.58
$\phi_{WW}(3)$		1.51	1.48	1.55	1.56	1.36	1.36
$\phi_{WW}(6)$		1.09	1.08	1.13	1.13	0.99	0.99
$\phi_{WW}(12)$		0.56	0.56	0.58	0.59	0.51	0.51
$\phi_{WW}(24)$	$(*)$	0.10	0.12	0.10	0.10	0.09	0.09
$\phi_{DD}(24)$		0.09	0.09	0.09	0.09	0.09	0.09
$\phi_{D}(24)$	$(*)$	0.10	0.11	0.09	0.09	0.09	0.09

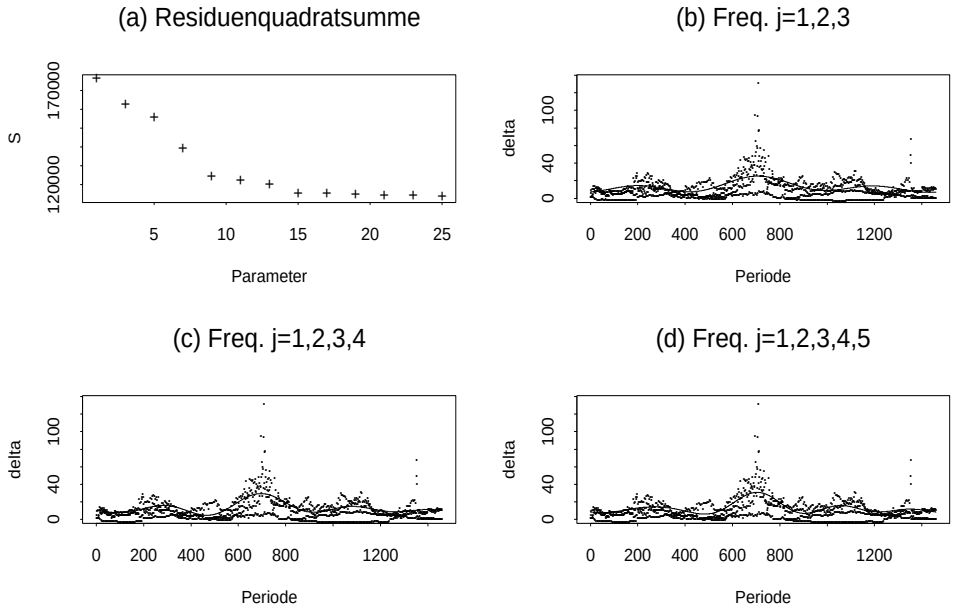
Werte der Abweichungen (4.4) $S_{f,k}$ unter Parameterset B4.
1. Spalte: $\hat{\nu}_t$ durch Glättung mit Frequenzen $\omega_j, j = 1, 3, 4, 5$ ersetzt.
2. Spalte: $\hat{\nu}_t$ durch Glättung mit Frequenzen $\omega_j, j = 1, 3, 4, 5, 365, 730$ ersetzt.
3. Spalte: $\hat{\beta}_t$ durch Glättung mit Frequenzen $\omega_j, j = 1, \ldots, 5$ ersetzt.
4. Spalte: $\hat{\beta}_t$ durch Glättung mit Frequenzen $\omega_j, j = 1, \ldots, 7$ ersetzt.
5. Spalte: $\hat{\delta}_t$ durch Glättung mit Frequenzen $\omega_j, j = 1, 2, 3, 4$ ersetzt.
6. Spalte: $\hat{\delta}_t$ durch Glättung mit Frequenzen $\omega_j, j = 1, \ldots, 5$ ersetzt.
$(*)$: Eigenschaft ist in der Modellanpassung verwendet worden.

Abbildung 5.11: Modellierung von β



(a): Res.quadratsumme (4.3) S_k bei sukzessiver Anpassung der Freq. $\omega_j, j = 0, 1, 2, \ldots, 10$.
(b): Schätzer $\hat{\beta}$ und Glättung mit $\omega_j, j = 0, 1, 3, 4$.
(c): Schätzer $\hat{\beta}$ und Glättung mit $\omega_j, j = 0, 1, 3, 4, 5$.
(d): Schätzer $\hat{\beta}$ und Glättung mit $\omega_j, j = 0, 1, 3, 4, 5, 6, 7$.

Abbildung 5.12: Modellierung von δ



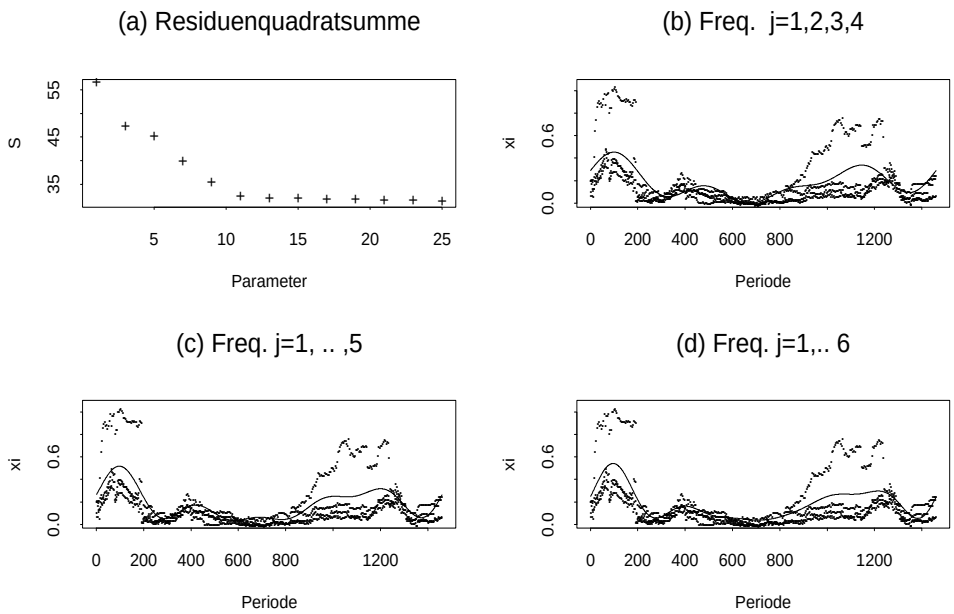
(a): Res.summe (4.3) S_k bei sukzessiver Anpassung der Freq. $\omega_j, j = 0, 1, 2, \ldots, 10$.
(b): Schätzer $\hat{\delta}$ und Glättung mit $\omega_j, j = 0, 1, 3, 4$.
(c): Schätzer $\hat{\delta}$ und Glättung mit $\omega_j, j = 0, 1, 3, 4, 5$.
(d): Schätzer $\hat{\delta}$ und Glättung mit $\omega_j, j = 0, 1, 3, 4, 5, 6, 7$.

Tabelle 5.4: Parameterglättung: ξ, Set B4

Parameter		ξ		Gesamtset
		$S_{\xi,1}$	$S_{\xi,2}$	S_{B4}
$\mu(1)$	$(*)$	0.52	6.02	0.28
$\gamma(1)$	$(*)$	1.27	1674.19	0.43
$\gamma(3)$		1.19	1514.18	0.52
$\gamma(6)$		1.10	1318.70	0.54
$\gamma(12)$		1.06	1309.59	0.54
$\gamma(24)$		1.03	1290.46	0.53
$\rho(1,1)$		0.48	0.48	0.57
$\rho(3,1)$		0.30	0.30	0.30
$\rho(6,1)$		0.27	0.27	0.26
$\rho(12,1)$		0.39	0.39	0.42
$\rho(24,1)$	$(*)$	0.04	0.04	0.26
$\phi_{WW}(1)$		1.86	1.86	1.72
$\phi_{WW}(3)$		1.42	1.42	1.58
$\phi_{WW}(6)$		1.01	1.01	1.17
$\phi_{WW}(12)$		0.51	0.51	0.61
$\phi_{WW}(24)$	$(*)$	0.09	0.09	0.11
$\phi_{DD}(24)$		0.09	0.09	0.07
$\phi_{D}(24)$	$(*)$	0.09	0.09	0.05

Werte der Abweichungen (4.4) $S_{f,k}$ unter Parameterset B4.
1. Spalte: $\hat{\xi}_t$ durch Glättung mit Freq. $\omega_j, j = 1, \ldots, 5$ ersetzt.
2. Spalte: $\hat{\xi}_t$ separat je nach Startzeit geglättet.
3. Spalte: Geglättetes Gesamtset.
$(*)$: Eigenschaft ist in der Modellanpassung verwendet worden.

Abbildung 5.13: Modellierung von ξ



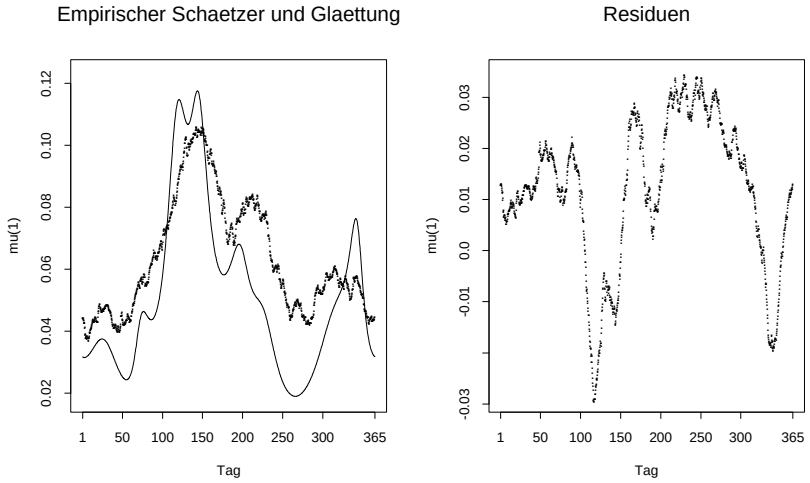
(a): Res.summe (4.3) S_k bei sukzessiver Anpassung der Freq. $\omega_j, j = 0, 1, 2, \ldots, 10$.
(b): Schätzer $\hat{\xi}$ und Glättung mit $\omega_j, j = 0, 1, 3, 4$.
(c): Schätzer $\hat{\xi}$ und Glättung mit $\omega_j, j = 0, 1, 3, 4, 5$.
(d): Schätzer $\hat{\xi}$ und Glättung mit $\omega_j, j = 0, 1, 3, 4, 5, 6, 7$.

Der Niederschlagsprozeß unter Set B4

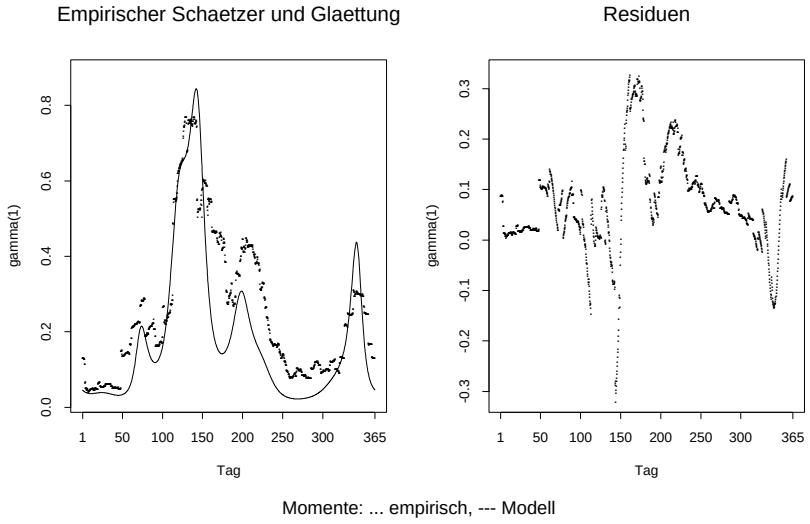
Nachfolgend sind zunächst Eigenschaften stündlichen Regens zu sehen, die aus allen im Abstand von sechs Stunden beginnenden 30-Tages-Perioden geschätzt worden sind sowie die entsprechenden Eigenschaften unter dem Modell, die aus dem geglätteten Parameterset B4 berechnet worden sind. Anschließend sind die ungeglätteten und die geglätteten Parameterschätzer aus Set B4 abgebildet, die in Kapitel 5.4 beschrieben worden sind.

Die empirischen Momente und die ungeglätteten Schätzer sind mit ····· markiert, — markiert jeweils die Werte unter dem Modell.

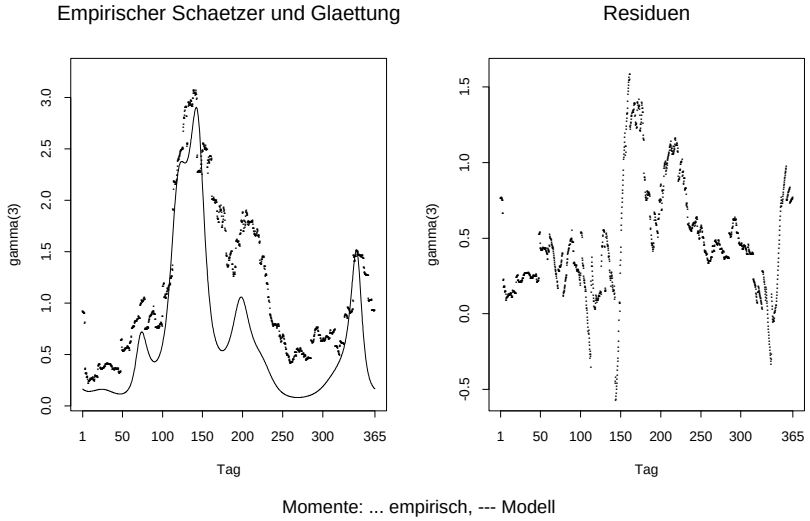
Mittel des 1-Stunden-Niederschlages



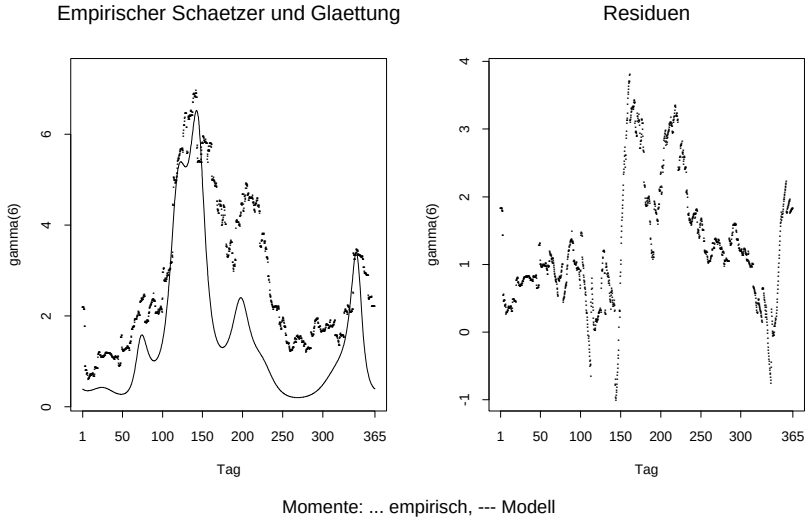
Varianz des 1-Stunden-Niederschlages



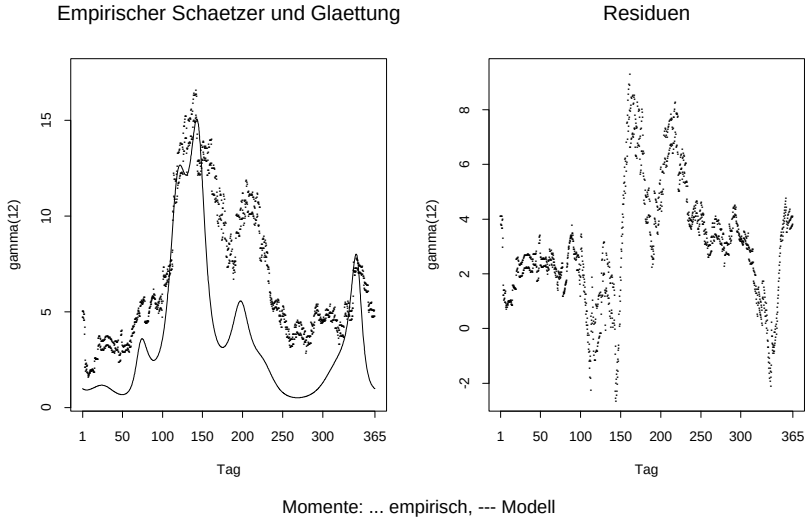
Varianz des 3-Stunden-Niederschlages



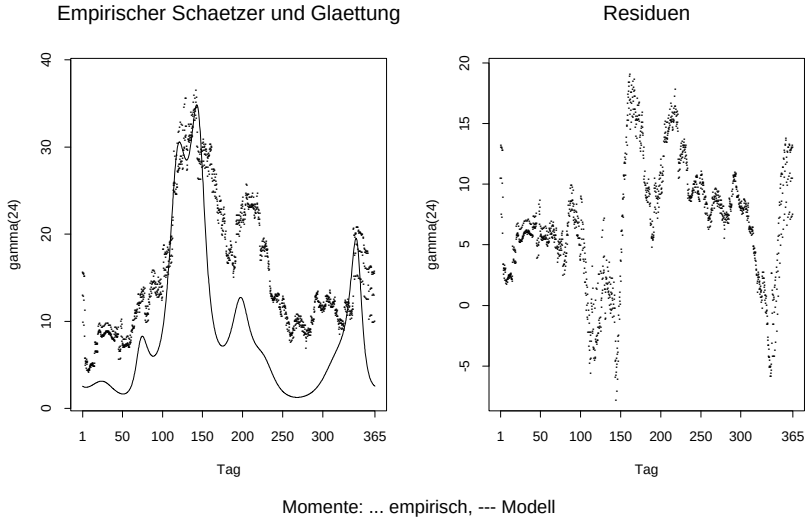
Varianz des 6-Stunden-Niederschlages



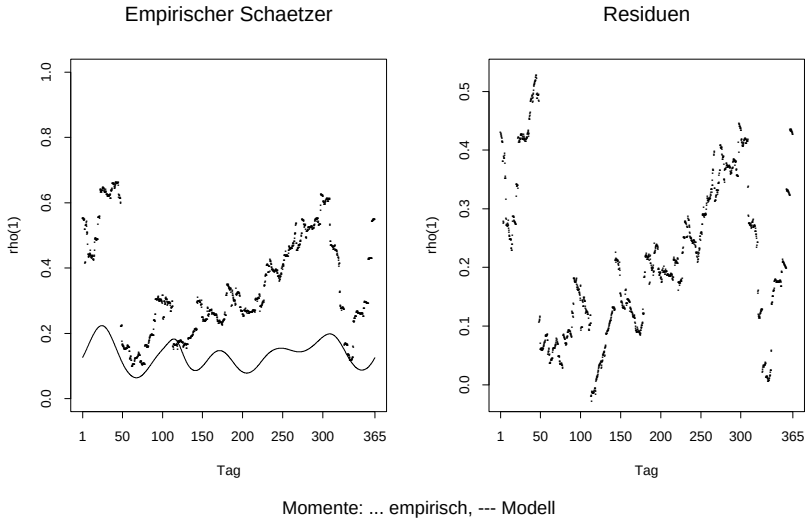
Varianz des 12-Stunden-Niederschlages



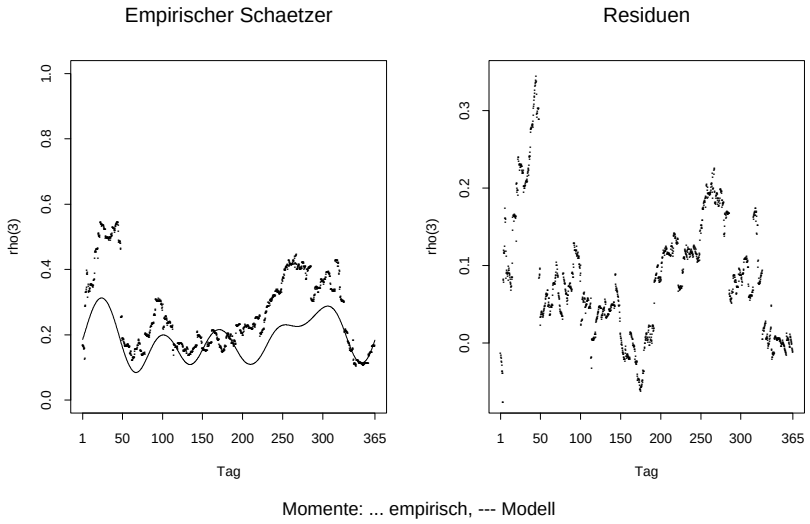
Varianz des 24-Stunden-Niederschlages



Lag(1)-Autokorrelation des 1-Stunden-Niederschlages



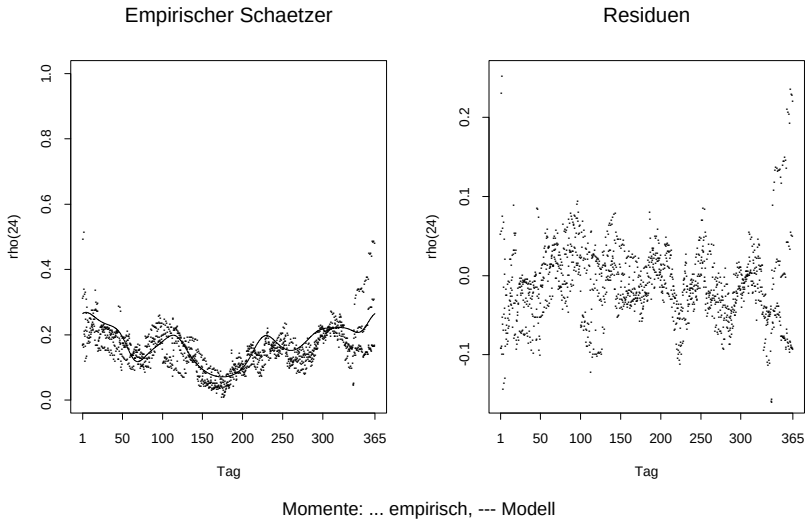
Lag(1)-Autokorrelation des 3-Stunden-Niederschlages



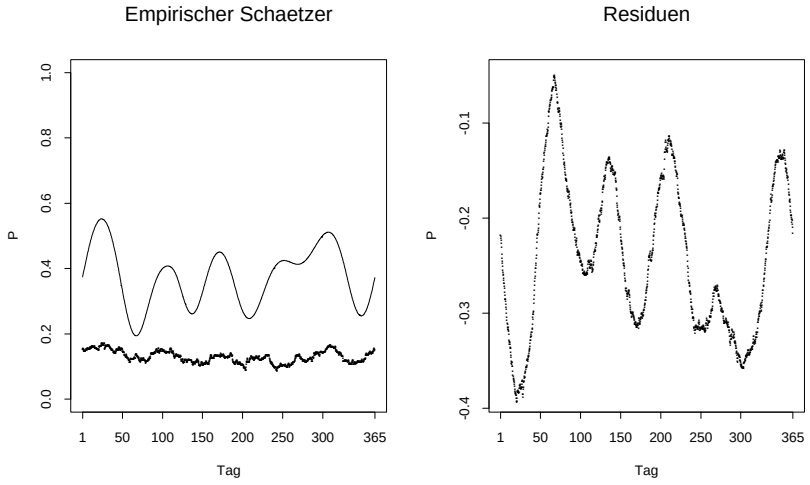
Lag(1)-Autokorrelation des 6-Stunden-Niederschlages

Lag(1)-Autokorrelation des 12-Stunden-Niederschlages

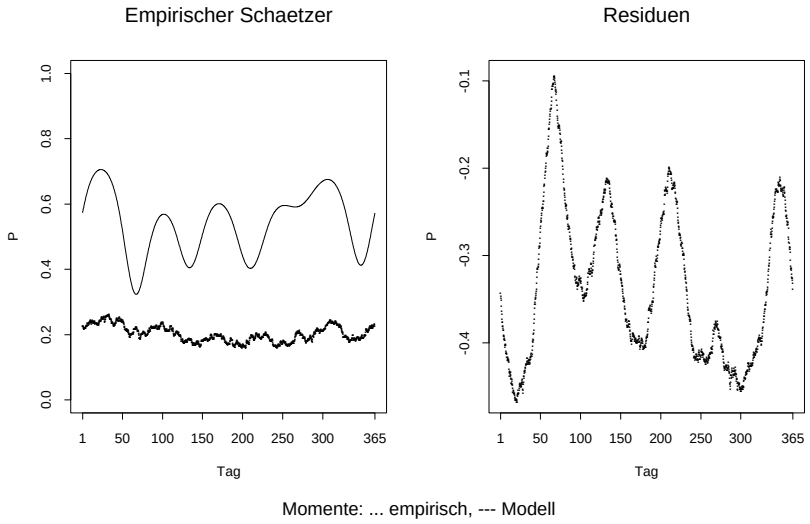
Lag(1)-Autokorrelation des 24-Stunden-Niederschlages



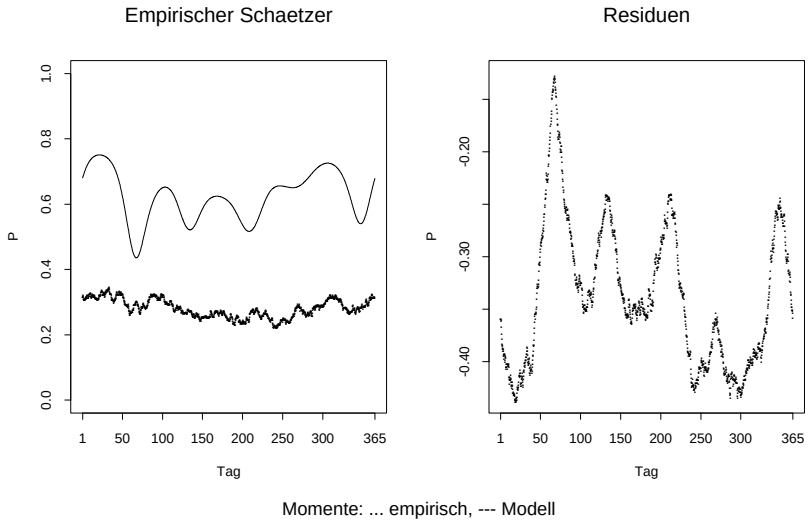
Übergangswahrscheinlichkeit $\phi_{WW}(1)$



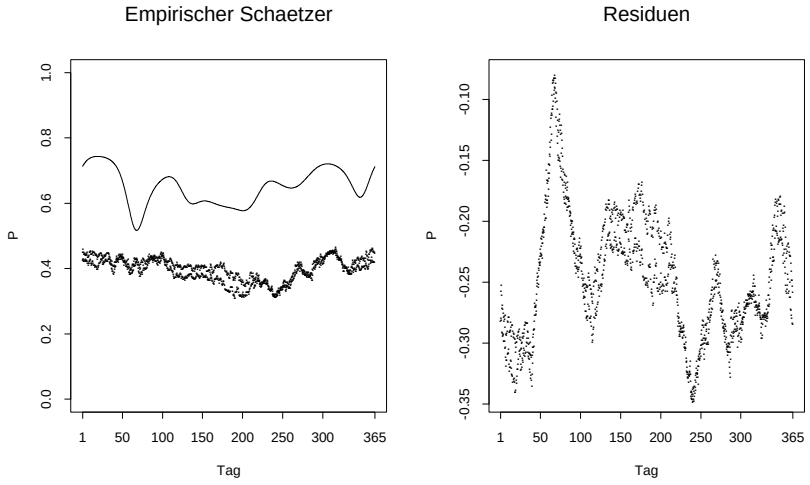
Übergangswahrscheinlichkeit $\phi_{WW}(3)$



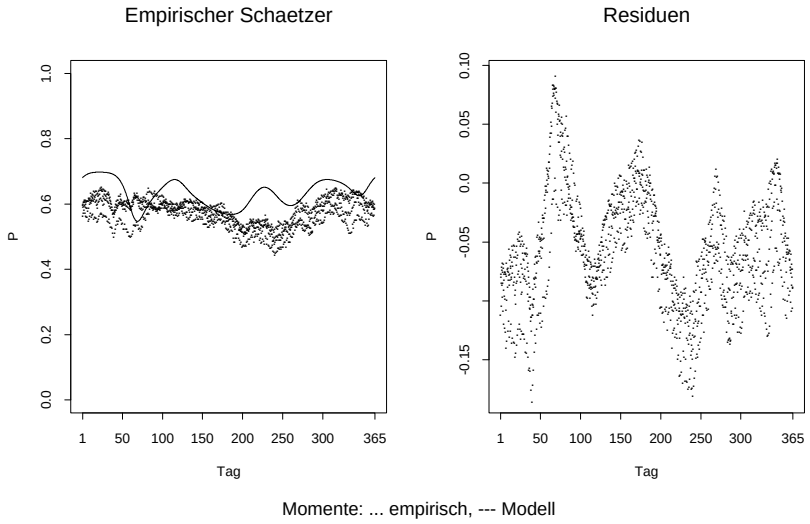
Übergangswahrscheinlichkeit $\phi_{WW}(6)$



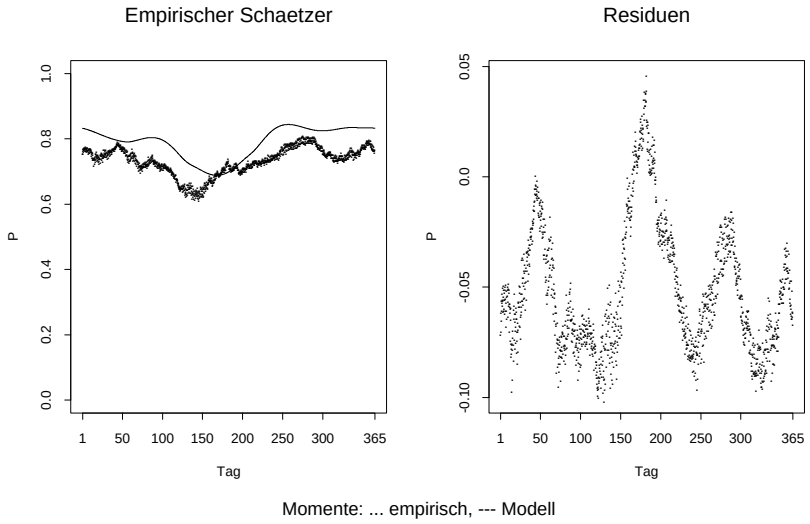
Übergangswahrscheinlichkeit $\phi_{WW}(12)$



Übergangswahrscheinlichkeit $\phi_{WW}(24)$

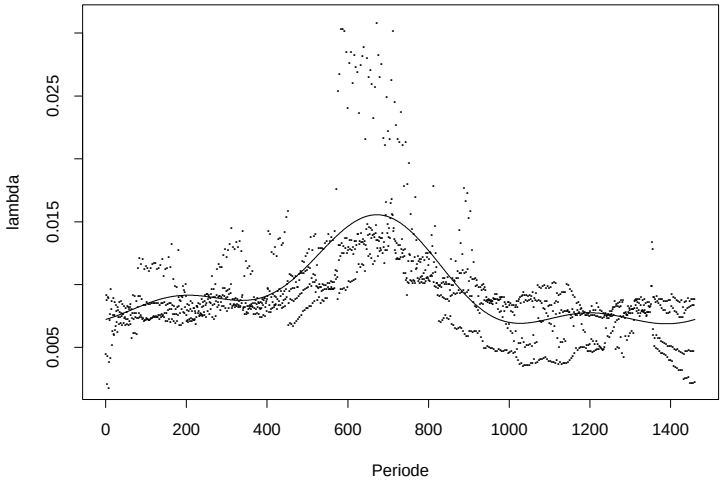


Übergangswahrscheinlichkeit $\phi_{DD}(24)$

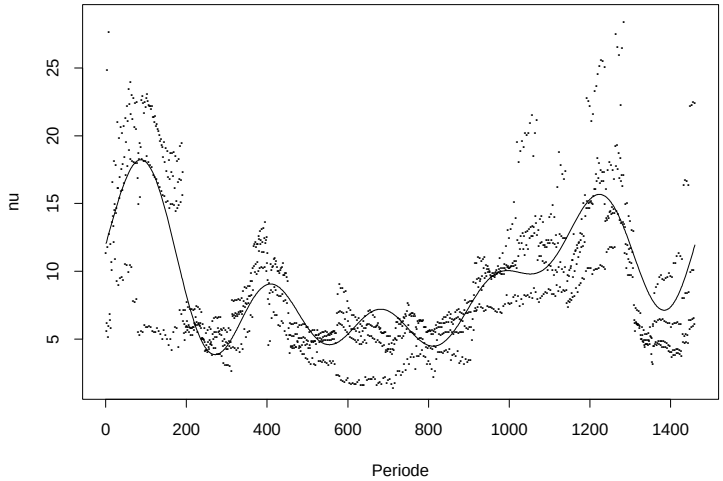


Wahrscheinlichkeit $\phi(24)$

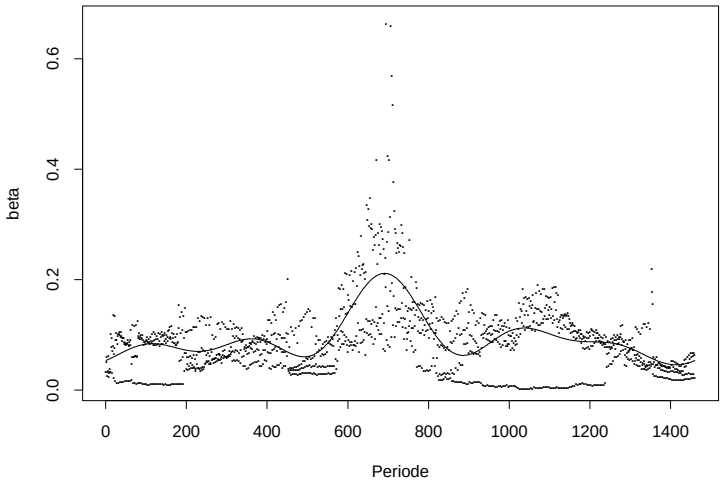
$\hat{\lambda}$ und Glättung durch $\omega_j, j = 1, 2, 3$



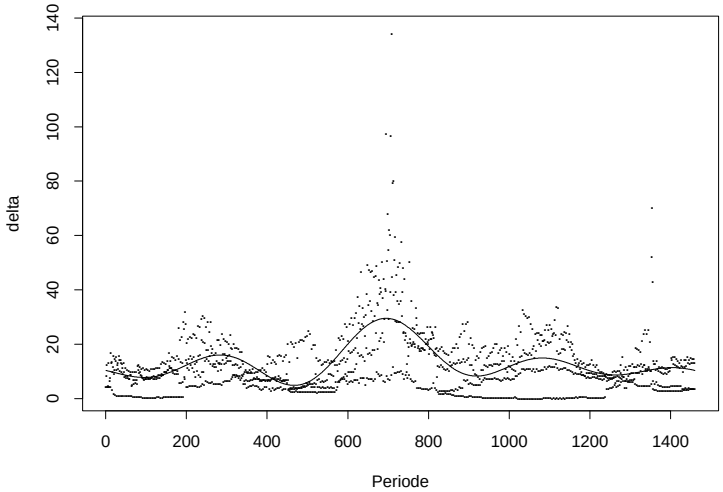
$\hat{\nu}$ und Glättung durch $\omega_j, j = 1, 2, 3, 365, 730$



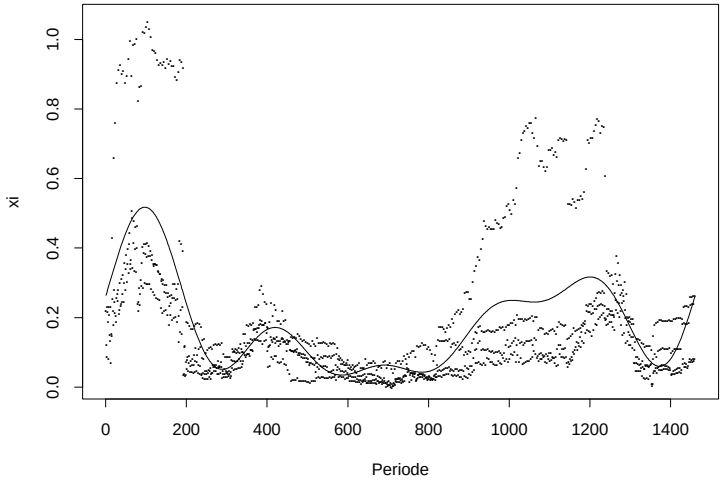
$\hat{\beta}$ und Glättung durch $\omega_j, j = 1, 2, 3, 4, 5$



$\hat{\delta}$ und Glättung durch $\omega_j, j = 1, 2, 3, 4$



$\hat{\xi}$ und Glättung durch $\omega_j, j = 1, 2, 3, 4, 5$



5 Glättung der Modellparameter – Set B1

Das Parameterset B1 resultiert aus der Anpassung des Neyman-Scott-Modelles an alle Perioden $t = 1, \ldots, T$ mit $T = 1460$ unter Verwendung der Second-Order-Eigenschaften aus Set B und eines Startsimplizes und wird deshalb als Set B1 bezeichnet. Die Schätzer $\hat{\psi}_t = (\hat{\lambda}_t, \hat{\nu}_t, \hat{\beta}_t, \hat{\delta}_t, \hat{\xi}_t)$, $t = 1, \ldots, 1460$, weisen saisonale und in geringem Maße auch tägliche Zyklen auf. Wie im vorigen Abschnitt werden die Modellparameter durch Fouriersummen approximiert, um die Gesamtanzahl an Parametern zu reduzieren und die aus physikalischen Gründen wünschenswerten Eigenschaften der Parameter wie Stetigkeit und Periodizität zu garantieren. Ausreißer in den Komponenten von $\hat{\psi}$ werden wieder durch das Mittel der benachbarten Werte ersetzt.

Beschreibung von λ

In den Werten der $\hat{\lambda}_t, t = 1, \ldots, T$, sind saisonale Zyklen erkennbar, in den Tagen 1 bis 150 (Januar - Mai) auch tägliche Zyklen, vergl. Abb. 5.14. Das Minimum von 0.001 wird im Dezember angenommen, das Maximum von 0.036 im Juni. Median und Standardfehler betragen 0.0075 bzw. 0.0059.

Die Frequenzen $\omega_j, j = 1, 2, 3, 4, 365, 7, 6, 5, 370, 730$, liefern in dieser Reihenfolge den größten Beitrag zum Spektrum, Verwendung der Frequenzen $\omega_j, j = 365, 370, 730$ (innertägliche Variation), in der Fouriersummendarstellung verbessert die Anpassung aber in geringerem Maße als die sukzessive Einbeziehung der Frequenzen $\omega_j, j = 1, 2, \ldots$. Die Summe der Residuenquadrate weist bei 5 und bei 7 Frequenzen Stufen auf. Um eine dieser beiden Anpassungen auszuwählen, ersetzen wir in Set B1 $\hat{\lambda}_t$ durch die angepaßten Werte $\tilde{\lambda}_{t,k}, t = 1, \ldots, T,\ k = 5, 7$, und berechnen die zugehörigen Second-Order-Eigenschaften des Mengenprozesses sowie die Residuenquadrate (4.4) $S_{f,k}$, vergl. Tab. 5.5 auf Seite 117.

Unter der Glättung mit 7 Frequenzen werden die Varianzen etwas besser wiedergegeben, ebenso der Erwartungswert $\hat{\mu}(1)$. Die Übergangswahrscheinlichkeiten $\phi_{WW}(h)$ sind unter dem Modell offensichtlich zu groß. Da λ die Rate der Stürme bestimmt und damit der zentrale Parameter des NSRP-Modelles ist, entscheiden wir uns für die exaktere Anpassung mit 7 Frequenzen, vergl. Abb. 5.14. Die geglätten Parameter sind

$$\tilde{\lambda}_t = \hat{R}_0 + \sum_{j=1}^{7} \hat{R}_j \cos(\omega_j t + \hat{\theta}_j)$$

mit

j	$\hat{R}_j$	$\hat{\theta}_j$
0	$92.90 \cdot 10^{-4}$	
1	$44.85 \cdot 10^{-4}$	-2.564
2	$34.01 \cdot 10^{-4}$	0.025
3	$28.95 \cdot 10^{-4}$	-2.488
4	$20.05 \cdot 10^{-4}$	1.018
5	$10.73 \cdot 10^{-4}$	-1.452
6	$11.20 \cdot 10^{-4}$	2.407
7	$12.26 \cdot 10^{-4}$	-1.646

Beschreibung von ν

Die Schätzer $\hat{\nu}_t,\ t = 1, \ldots, T$, zeigen saisonale, aber kaum innertägliche Variation. Das Minimum von 0.3368 wird im Juni angenommen, das Maximum von 19.95 im April, Median und Standardabweichung betragen 4.283 bzw. 3.2198.

Die Residuenquadratsumme besitzt Stufen bei fünf und sieben Frequenzen (11 bzw. 15 Parameter), vergl. Abb. 5.15. Zur Auswahl zwischen diesen Anpassungen berechnen wir die Second-Order-Eigenschaften unter Parameterset B1 mit geglätteten $\hat{\nu}_t$ sowie die zugehörigen Residuenquadratsummen, vergl. Tab. 5.6. Die Anpassung mit 7 Frequenzen führt nur zu geringfügig kleineren Abweichungen bei den Varianzen, während die Anpasung der Korrelationen $\rho(3,1)$ und $\rho(12,1)$ sowie der Übergangswahrscheinlichkeiten $\phi_{WW}(6)$ und $\phi_{WW}(12)$ sogar verschlechtert wird. Wir beschreiben daher $\hat{\nu}_t$ durch die Fouriersumme $\tilde{\nu}_t$ mit 5 Frequenzen,

$$\tilde{\nu}_t = \hat{R}_0 + \sum_{j=1}^{5} \hat{R}_j \cos(\omega_j\, t + \hat{\theta}_j)$$

mit

j	$\hat{R}_j$	$\hat{\theta}_j$
0	5.904	
1	1.656	-0.571
2	1.791	2.782
3	1.519	0.626
4	1.534	-1.507
5	1.514	-2.008

Beschreibung von β

Die ungeglätteten Schätzer $\hat{\beta}_t, t = 1, \ldots, T$, nehmen das Maximum von 0.1972[Std^{-1}] im Mai an, das Minimum von 0.0021[Std^{-1}] im November, bei einem Median von 0.0244[Std^{-1}] und einer Standardabweichung von 0.0323[Std^{-1}].

Die Summe der Residuenquadrate S_k weist bei $k = 7$ und $k = 11$ Parametern Stufen auf vergl. Abb. 5.16. Wir untersuchen deshalb die Performance dieser Glättungen $\tilde{\beta}_t$ in bezug auf die Momente des Niederschlagsmengenprozesses, vergl. Tab. 5.6.

Die Berücksichtiging von ω_j, $j = 4, 5$, bewirkt keine nennenswerte Verbesserung in der Anpassung der Momente. Damit beschreiben wir $\hat{\beta}$ durch die Fouriersumme

$$\tilde{\beta}_t = \hat{R}_0 + \sum_{j=1}^{3} \hat{R}_j \cos(\omega_j t + \hat{\theta}_j)$$

mit

j	$\hat{R}_j$	$\hat{\theta}_j$
0	$36.73 \cdot 10^{-4}$	
1	$36.77 \cdot 10^{-4}$	-2.047
2	$13.33 \cdot 10^{-4}$	2.267
3	$10.43 \cdot 10^{-4}$	0.651

Modellierung von δ

Die Schätzer $\hat{\delta}_t$, $t = 1, \ldots, T$, weisen saisonale und geringfügige tägliche Zyklen auf. Maximum und Minimum werden zu den gleichen Zeitpunkten wie von den $\hat{\beta}_t$ angenommen, nämlich 34.36[Std^{-1}] im Mai bzw. 0.0905[Std^{-1}] im November. Median und Standardabweichung betragen 1.755[Std^{-1}] bzw. 5.2208[Std^{-1}].

Werden die $\hat{\delta}_t$ durch eine Fouriersumme approximiert, entstehen negative angepaßte Werte $\tilde{\delta}_t$, was zu negativen Varianzen und unzulässigen Werten der Autokorrelationsfunktion der Niederschlagsmengen führt. Deshalb werden die logarithmierten Schätzer $\log(\hat{\delta}_t)$ durch eine Fouriersumme approximiert. Die Summe der Residuenquadrate S_k bei Anpassung der $\log(\hat{\delta}_t)$ besitzt bei Verwendung von $k = 5, 6, 7$ Frequenzen Stufen. Zur Auswahl einer Anpassung werden wie in den vorigen Abschnitten die Momente des Niederschlagsmengenprozesses unter Parameterset B1 berechnet, wobei die $\hat{\delta}_t$ durch die geglätteten Werte $\tilde{\delta}_t$ ersetzt werden, sowie die Residuenquadratsummen. Die Fouriersumme mit 7 Frequenzen führt zu einer verbesserten Anpassung aller Varianzen $\gamma(h)$, vergl. Tab. 5.6. Deshalb glätten wir $\hat{\delta}_t$ durch $\tilde{\delta}_t$ mit

$$\log(\tilde{\delta}_t) = \hat{R}_0 + \sum_{j=1}^{7} \hat{R}_j \cos(\omega_j t + \hat{\theta}_j) \ ,$$

wobei

j	$\hat{R}_j$	$\hat{\theta}_j$
0	$7.666 \cdot 10^{-1}$	
1	$14.398 \cdot 10^{-1}$	-2.015
2	$1.380 \cdot 10^{-1}$	2.185
3	$4.179 \cdot 10^{-1}$	1.081
4	$1.932 \cdot 10^{-1}$	-2.542
5	$2.205 \cdot 10^{-1}$	0.399
6	$0.893 \cdot 10^{-1}$	-2.646
7	$1.331 \cdot 10^{-1}$	-0.397

Beschreibung von ξ

Die Schätzer $\hat{\xi}_t, t = 1, \ldots, T$, weisen geringe saisonale Schwankungen auf. Im Januar und Februar sind leichte tägliche Zyklen erkennbar. Im Mai wird das Minimum von 0.0228[Std/mm], im Oktober das Maximum von 1.6769[Std/mm] angenommen, bei einem Median von 0.3226[Std/mm] und einem Standardfehler von 0.4207[Std/mm].

Wie im vorigen Abschnitt entstehen beim Glätten der $\hat{\xi}_t$ durch eine Fouriersumme negative Schätzer $\tilde{\xi}_t$, so daß wieder die logarithmierten Schätzer $\log(\hat{\xi}_t)$ geglättet werden. Aus der Residuenquadratsumme S_k ist ersichtlich, daß $k = 5$ Frequenzen ausreichen, vergl. Abb. 5.18. Die angepaßten Werte sind $\tilde{\xi}_t$ mit

$$\log(\tilde{\xi}_t) = \hat{R}_0 + \sum_{j=1}^{5} \hat{R}_j \cos(\omega_j t + \hat{\theta}_j) ,$$

wobei

j	$\hat{R}_j$	$\hat{\theta}_j$
0	1.207	
1	1.343	0.824
2	0.331	-1.521
3	0.446	-1.939
4	0.238	-0.211
5	0.371	-2.417

Die Übergangswahrscheinlichkeiten $\phi_{ww}(h), h = 1, 3$ sind unter dem Modell zu groß, während die große Residuenquadratsumme der Varianzen $\gamma(24)$ durch einzelne sehr große Werte wie $\gamma(24) \approx 500$ verursacht wird, vergl. Tab. 5.7.

Das geglättete Parameterset

Das Paramterset B1 wird also durch $\tilde{\psi} = (\tilde{\lambda}, \tilde{\nu}, \tilde{\beta}, \tilde{\delta}, \tilde{\xi})$ geglättet, das sind insgesamt 59 Parameter:

Parameter	Parameteranzahl	Frequenzen
λ	15	1, ..., 7
ν	11	1, ..., 5
β	7	1, 2, 3
δ	15	1, ..., 7
ξ	11	1, ..., 5
Gesamt	59	

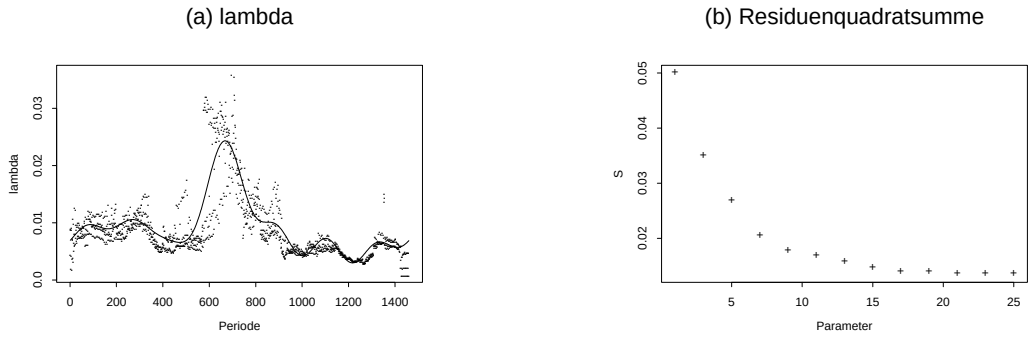
Zusammenfassung

Das Parameterset B1 entsteht, indem an alle im Abstand von 6 Stunden beginnenden 30-Tages-Perioden das Neyman-Scott-Modell auf die in Abschnitt 1 beschriebene Weise angepaßt wird. Die angepaßten Parameter einer Periode dienen dabei als Startwerte in der Anpassung der folgenden Periode. Die Momente unter dem Modell sowie die Modellparameter sind in den Abbildungen auf den Seiten 121 - 128 dargestellt, die Residuenquadrate des geglätteten Gesamtsets in Tabelle 5.7.

Keines der betrachteten Momente des Sets $\mathcal{F}$ des Prozesses der aggregierten Niederschlagsmengen entspricht exakt seinem empirischen Äquivalent. Die Abweichungen sind bei dem 1-Stunden-Mittel $\mu(1)$ und den h-Stunden-Varianzen $\gamma(h)$ gering — man beachte in den Graphiken die Größenskala. Das empirische 1-Stunden-Mittel weist unter Set B1 allerdings deutlich andere saisonale Schwankungen auf als unter den Beobachtungen. Die h-Stunden-Varianzen sind im Mai (Tag 150) zwischen 50% ($h = 1$) und 25% ($h = 24$) zu groß. Zusätzlich treten mit größer werdender Intervalldauer h saisonale Schwankungen im Juli und August (Tage 180 - 230) auf, die nur in geringerem Maß zu beobachten sind. Die saisonalen Schwankungen der Varianzen unter dem Modell treten im Vergleich zu den beobachteten zeitverzögert auf. Die Lag(1)-Autokorrelationen $\rho(h, 1)$ weisen für $h = 1$ saisonale Zyklen auf, die so nicht zu beobachten sind. Für $h = 3$ und $h = 6$ ist die Anpassung durch das Modell angemessen, für $h = 12$ und $h = 24$ sogar sehr gut. Dabei ist zu bedenken, daß $\rho(24, 1)$ in der Modellanpassung verwendet worden ist.
Die Übergangswahrscheinlichkeiten $\phi_{WW}(h), h = 1, 3, 6$, sind unter dem Modell zu groß und weisen zusätzlich starke saisonale Schwankungen auf, die empirisch nicht aufgetreten sind. In den Modellparametern findet sich keine offensichtliche Erklärung der Senke in $\phi_{WW}(h)$, $h = 12, 24$, um Tag $t = 290$, was wieder auf die schwer erfaßbaren Wechselwirkungen gleichzeitiger Änderungen mehrerer Modellparameter hinweist. Die Wahrscheinlichkeiten trockener Intervalle $\phi_{DD}(24)$ und $\phi_D(24)$ weisen ebenfalls saisonale Schwankungen auf, die nicht den Beobachtungen entsprechen. Die Anpassung von $\phi_{DD}(24)$ ist besser als die von $\phi_D(24)$, was erstaunlich ist, da in der Anpassung nur $\phi_D(24)$ verwendet worden ist.

Insgesamt ist die Modellanpassung akzeptabel, insbesondere angesichts des Versuches, eine große Anzahl, nämlich $26280 = 18 \cdot 1460$, an Second–Order–Momenten mit einer wesentlich kleineren Anzahl, nämlich 59, an Modellparametern zu beschreiben. Im folgenden Kapitel überprüfen wir die Performance des Modelles anhand weiterer Eigenschaften. Da die Residuenquadrate unter dem geglätteten Parameterset B1 geringer als unter Set B4 sind und insbesondere die Wiedergabe der Autokorrelationsstruktur und der Varianzen unter Set B1 wesentlich besser ist, wird die nachfolgende Modellvalidation ausschließlich auf Set B1 basieren.

Abbildung 5.14: Modellierung von λ



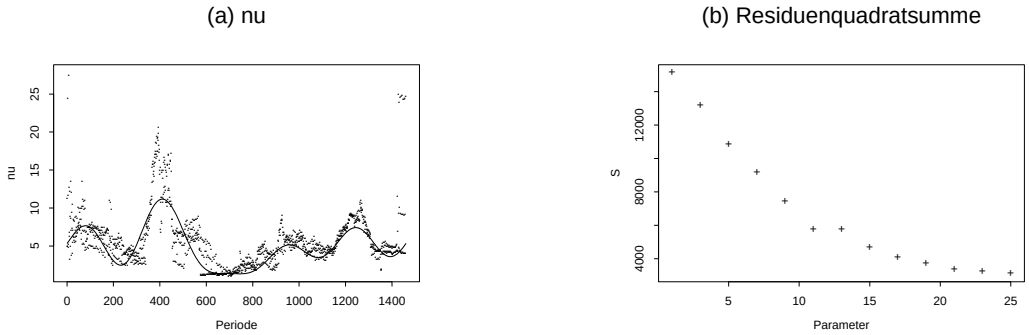
(a) $\hat{\lambda}_t$ und angepaßte Fouriersumme mit Freq. $\omega_j, j = 1, \ldots, 7$.
(b) Residuenquadratsumme (4.3) S_k.

Tabelle 5.5: Parameterglättung: Set B1, λ

Parameter		ungeglättet	λ	
		S_{B1}	$S_{\lambda,1}$	$S_{\lambda,2}$
$\mu(1)$	$(*)$	0.06	0.29	0.25
$\gamma(1)$	$(*)$	0.09	0.34	0.30
$\gamma(3)$		0.11	0.34	0.29
$\gamma(6)$		0.15	0.37	0.32
$\gamma(12)$		0.17	0.42	0.37
$\gamma(24)$		0.19	0.48	0.43
$\rho(1,1)$		0.45	0.45	0.45
$\rho(3,1)$		0.40	0.40	0.40
$\rho(6,1)$		0.47	0.47	0.47
$\rho(12,1)$		0.36	0.36	0.36
$\rho(24,1)$	$(*)$	0.08	0.08	0.08
$\phi_{WW}(1)$		2.34	2.35	2.35
$\phi_{WW}(3)$		0.89	0.90	0.90
$\phi_{WW}(6)$		0.46	0.48	0.48
$\phi_{WW}(12)$		0.27	0.30	0.29
$\phi_{WW}(24)$	$(*)$	0.15	0.18	0.18
$\phi_{DD}(24)$		0.12	0.10	0.10
$\phi_{D}(24)$	$(*)$	0.17	0.15	0.15

Werte der Abweichungen (4.4) $S_{f,k}$ unter Parameterset B1.
1. Spalte: Ungeglättetes Set B1.
2. Spalte: $\hat{\lambda}_t$ durch Glättung mit Freq. $\omega_j, j = 1, \ldots, 5$ ersetzt.
3. Spalte: $\hat{\lambda}_t$ durch Glättung mit Freq. $\omega_j, j = 1, \ldots, 7$ ersetzt.
$(*)$: Eigenschaft ist in der Modellanpassung verwendet worden.

Abbildung 5.15: Modellierung von ν



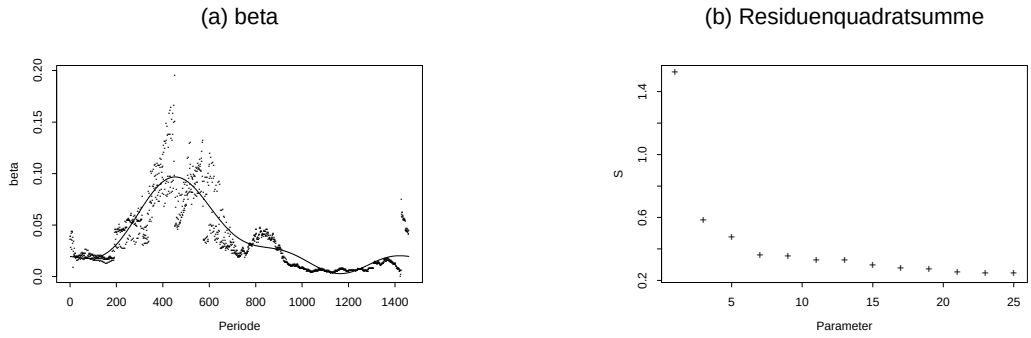
(a) $\hat{\nu}_t$ und angepaßte Fouriersumme mit Freq. $\omega_j, j = 1, \ldots, 5$.
(b) Residuenquadratsumme (4.3) S_k.

Tabelle 5.6: Parameterglättung: ν, β, δ

Parameter		ν		β		δ		
		$S_{\nu,1}$	$S_{\nu,2}$	$S_{\beta,1}$	$S_{\beta,2}$	$S_{\delta,1}$	$S_{\delta,2}$	$S_{\delta,3}$
$\mu(1)$	$(*)$	0.3	0.27	0.06	0.06	0.31	0.3	0.29
$\gamma(1)$	$(*)$	0.31	0.31	0.09	0.09	0.65	0.64	0.58
$\gamma(3)$		0.34	0.32	0.12	0.12	0.64	0.63	0.57
$\gamma(6)$		0.34	0.32	0.15	0.15	0.66	0.64	0.59
$\gamma(12)$		0.34	0.32	0.17	0.17	0.68	0.67	0.62
$\gamma(24)$		0.35	0.32	0.20	0.19	0.70	0.68	0.63
$\rho(1,1)$		0.45	0.45	0.46	0.45	0.45	0.46	0.46
$\rho(3,1)$		0.38	0.39	0.39	0.39	0.39	0.38	0.39
$\rho(6,1)$		0.44	0.44	0.47	0.46	0.45	0.45	0.45
$\rho(12,1)$		0.36	0.37	0.36	0.36	0.37	0.37	0.37
$\rho(24,1)$	$(*)$	0.26	0.24	0.15	0.13	0.09	0.09	0.09
$\phi_{WW}(1)$		2.35	2.34	2.32	2.32	2.30	2.30	2.31
$\phi_{WW}(3)$		0.93	0.93	0.89	0.89	0.88	0.88	0.88
$\phi_{WW}(6)$		0.49	0.51	0.48	0.47	0.47	0.47	0.47
$\phi_{WW}(12)$		0.30	0.32	0.31	0.29	0.27	0.27	0.28
$\phi_{WW}(24)$	$(*)$	0.18	0.18	0.16	0.16	0.15	0.15	0.15
$\phi_{DD}(24)$		0.12	0.12	0.12	0.12	0.12	0.12	0.12
$\phi_D(24)$	$(*)$	0.20	0.19	0.18	0.17	0.17	0.17	0.17

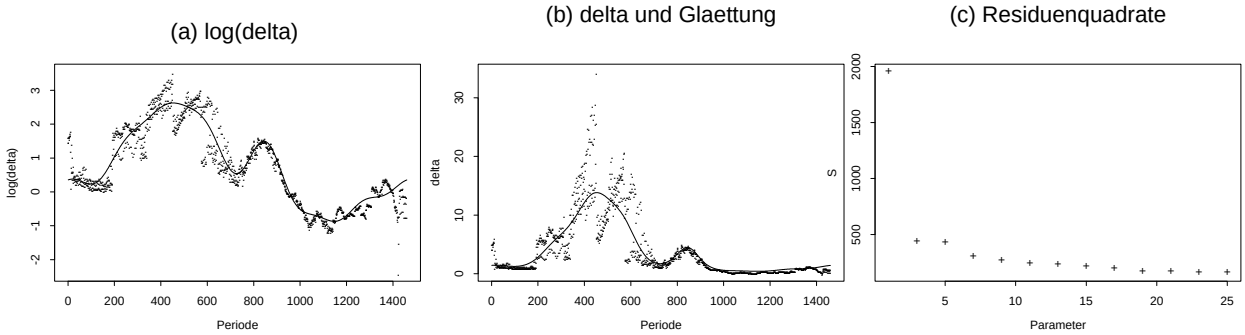
Werte der Abweichungen (4.4) $S_{f,k}$ unter Parameterset B1.
1. Spalte: $\hat{\nu}_t$ durch Glättung mit Freq. $\omega_j, j = 1, \ldots, 5$ ersetzt.
2. Spalte: $\hat{\nu}_t$ durch Glättung mit Freq. $\omega_j, j = 1, \ldots, 7$ ersetzt.
3. Spalte: $\hat{\beta}_t$ durch Glättung mit Freq. $\omega_j, j = 1, 2, 3$ ersetzt.
4. Spalte: $\hat{\beta}_t$ durch Glättung mit Freq. $\omega_j, j = 1, \ldots, 5$ ersetzt.
5. Spalte: $\hat{\delta}_t$ durch Glättung mit Freq. $\omega_j, j = 1, \ldots, 5$ ersetzt.
6. Spalte: $\hat{\delta}_t$ durch Glättung mit Freq. $\omega_j, j = 1, \ldots, 6$ ersetzt.
7. Spalte: $\hat{\delta}_t$ durch Glättung mit Freq. $\omega_j, j = 1, \ldots, 7$ ersetzt.
$(*)$: Eigenschaft ist in der Modellanpassung verwendet worden.

Abbildung 5.16: Modellierung von β



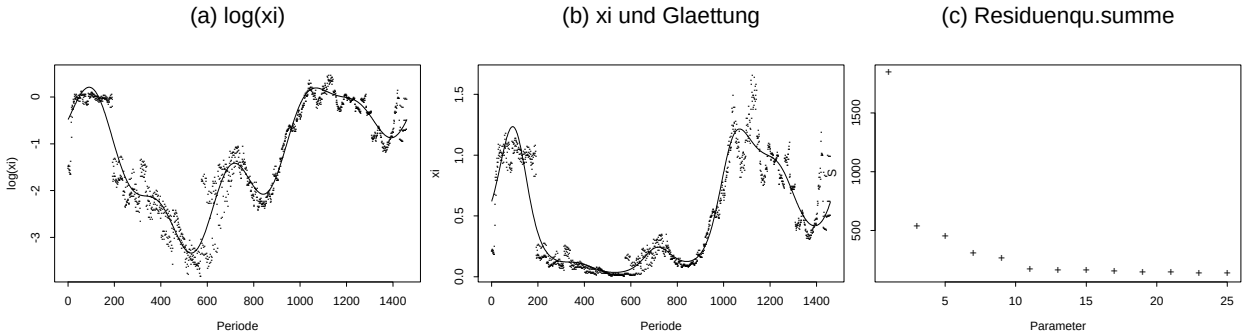
(a) $\hat{\beta}_t$ und angepaßte Fouriersumme mit Freq. $\omega_j, j = 1, 2, 3$.
(b) Residuenquadratsumme (4.3) S_k.

Abbildung 5.17: Modellierung von δ



(a) $\log(\hat{\delta}_t)$ und angepaßte Fouriersumme mit Freq. $\omega_j, j = 1, \ldots, 7$.
(b) $\hat{\delta}_t$ und Gättung.
(c) Residuenquadratsumme (4.3) S_k.

Abbildung 5.18: Modellierung von ξ



(a) $\log(\hat{\xi})$ und angepaßte Fouriersumme mit Freq. $\omega_j, j = 1, \ldots, 5$.
(b) $\hat{\xi}_t$ und Glättung.
(c) Residuenquadratsumme (4.3) S_k.

Tabelle 5.7: Parameterglättung: ξ, Set B1

Parameter		ξ	Gesamtset
		S_ξ	S_{B1}
$\mu(1)$	$(*)$	0.28	0.20
$\gamma(1)$	$(*)$	0.57	0.29
$\gamma(3)$		0.58	0.27
$\gamma(6)$		0.59	0.28
$\gamma(12)$		0.60	0.29
$\gamma(24)$		0.62	0.31
$\rho(1,1)$		0.45	0.46
$\rho(3,1)$		0.40	0.37
$\rho(6,1)$		0.47	0.42
$\rho(12,1)$		0.36	0.37
$\rho(24,1)$	$(*)$	0.08	0.28
$\phi_{WW}(1)$		2.34	2.32
$\phi_{WW}(3)$		0.89	0.91
$\phi_{WW}(6)$		0.46	0.49
$\phi_{WW}(12)$		0.27	0.33
$\phi_{WW}(24)$	$(*)$	0.15	0.17
$\phi_{DD}(24)$		0.12	0.11
$\phi_D(24)$	$(*)$	0.17	0.16

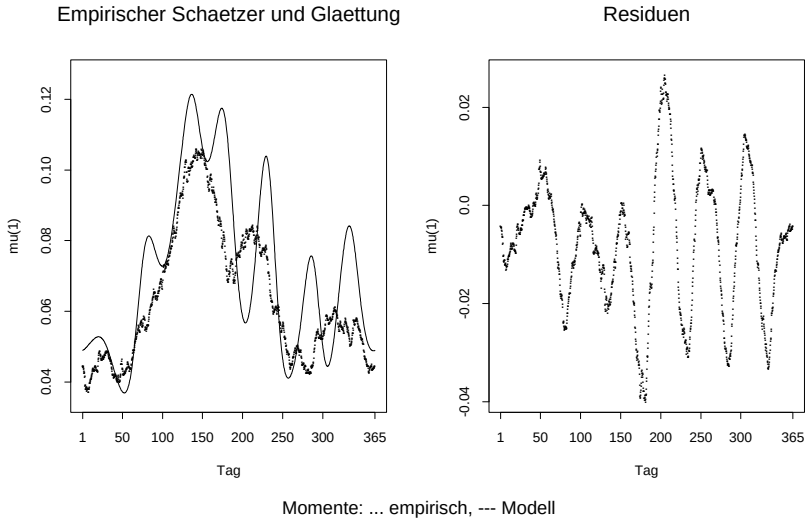
Werte der Abweichungen (4.4) $S_{f,k}$ unter Parameterset B1.
1. Spalte: $\hat{\xi}_t$ durch Glättung mit Freq. $\omega_j, j = 1, \ldots, 5$ ersetzt.
2. Spalte: Geglättetes Gesamtset B1.
$(*)$: Eigenschaft ist in der Modellanpassung verwendet worden.

Der Niederschlagsprozeß unter Set B1

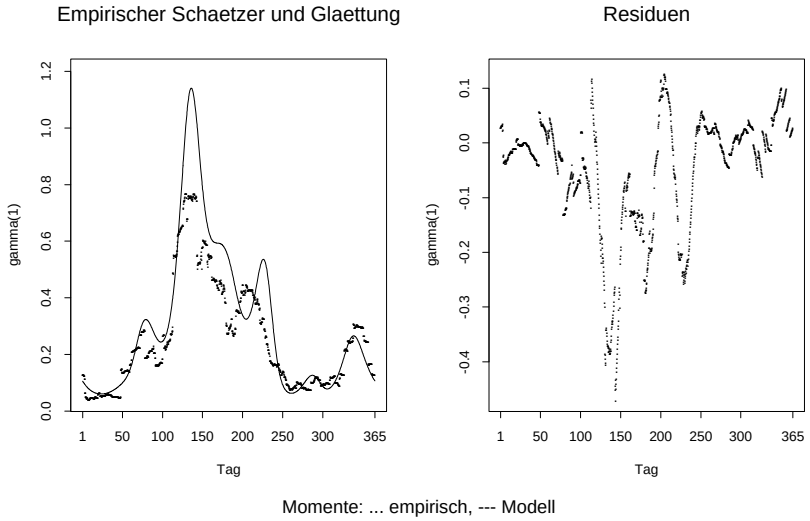
Im folgenden Abschnitt werden zuerst die die empirischen, aus den Beobachtungen geschätzten Second-Order-Eigenschaften des Niederschlagsprozesses und deren Werte unter dem geglätteten Set B1 dargetellt. Anschließend sind die Parameter des ungeglätteten Sets und die angepaßten Fouriersummen abgebildet.

Die empirischen Momente sowie die ungeglätteten Schätzer sind mit ····· markiert, — markiert jeweils die Werte unter dem Modell.

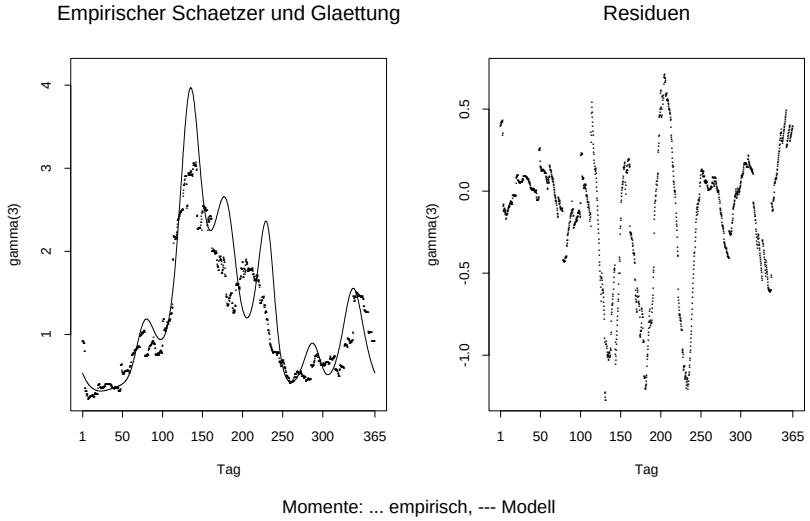
Mittel des 1-Stunden-Niederschlages



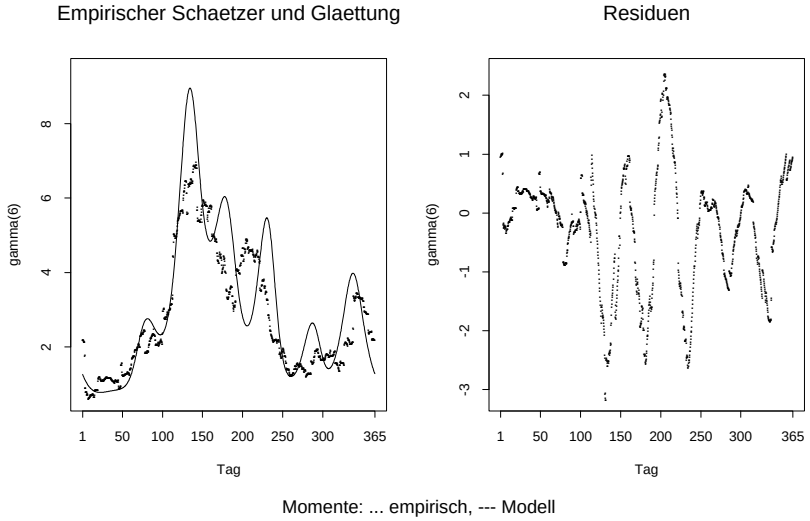
Varianz des 1-Stunden-Niederschlages



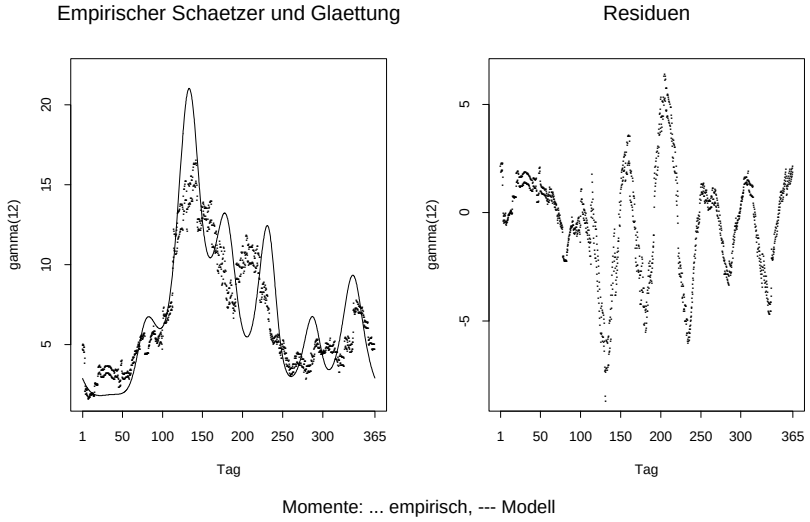
Varianz des 3-Stunden-Niederschlages



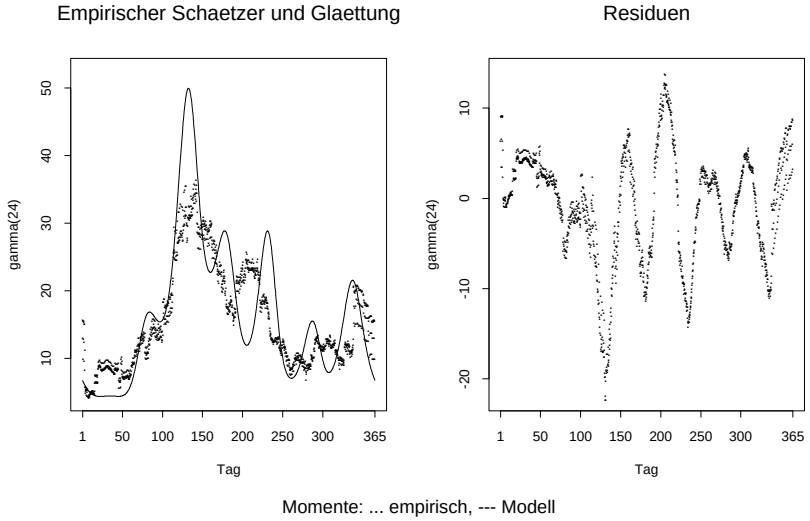
Varianz des 6-Stunden-Niederschlages



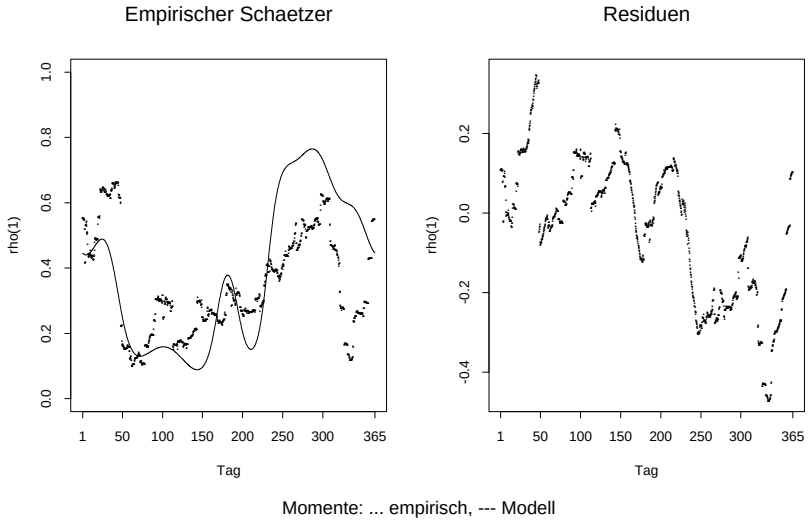
Varianz des 12-Stunden-Niederschlages



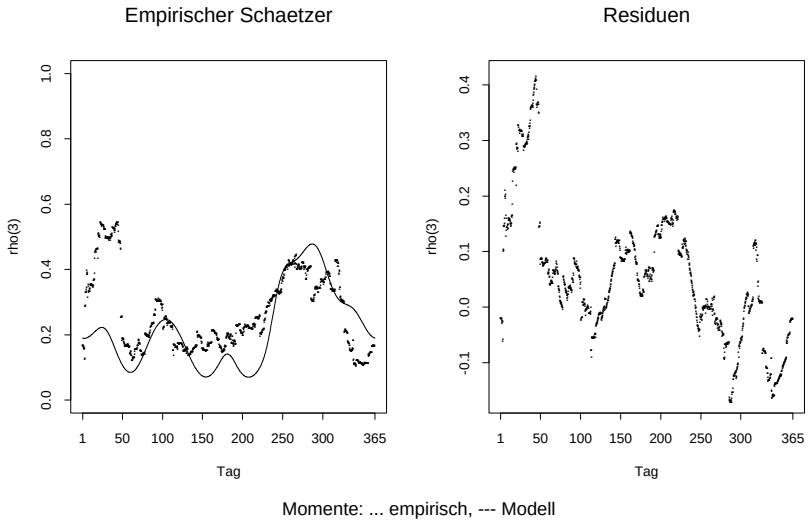
Varianz des 24-Stunden-Niederschlages



Lag(1)-Autokorrelation des 1-Stunden-Niederschlages

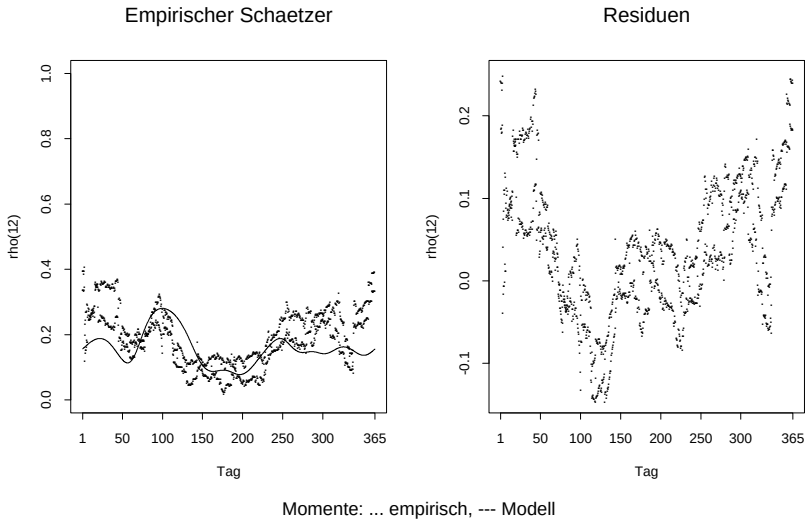


Lag(1)-Autokorrelation des 3-Stunden-Niederschlages

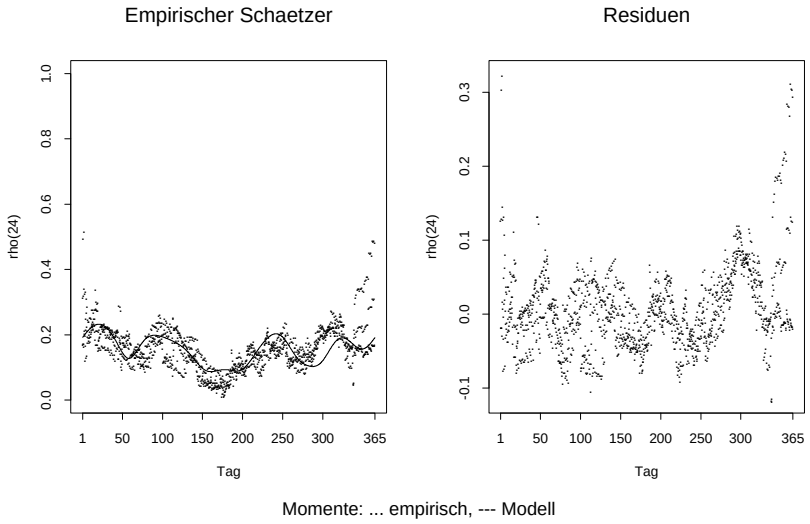


Lag(1)-Autokorrelation des 6-Stunden-Niederschlages

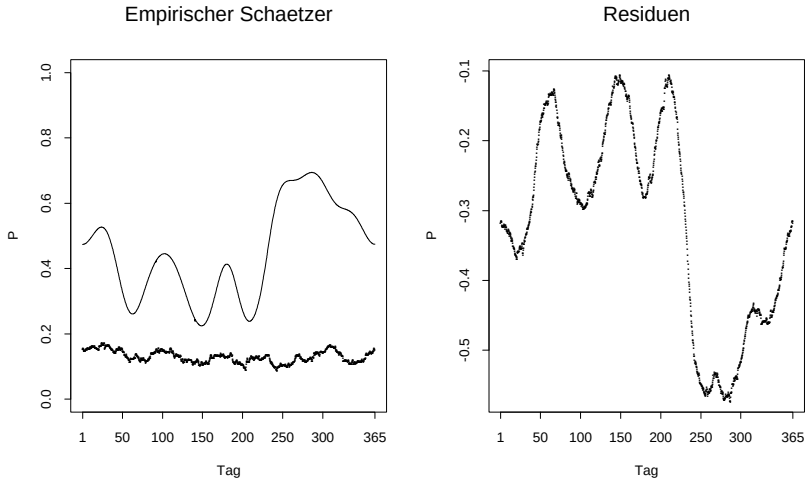
Lag(1)-Autokorrelation des 12-Stunden-Niederschlages



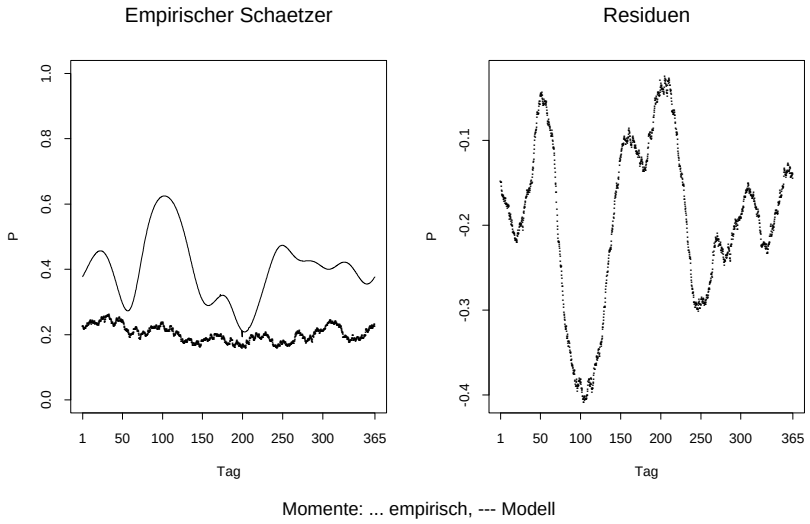
Lag(1)-Autokorrelation des 24-Stunden-Niederschlages



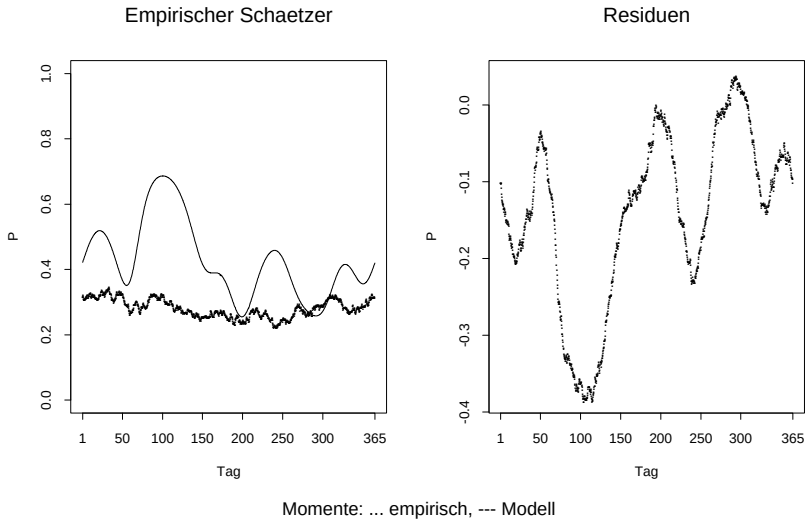
Übergangswahrscheinlichkeit $\phi_{WW}(1)$



Übergangswahrscheinlichkeit $\phi_{WW}(3)$

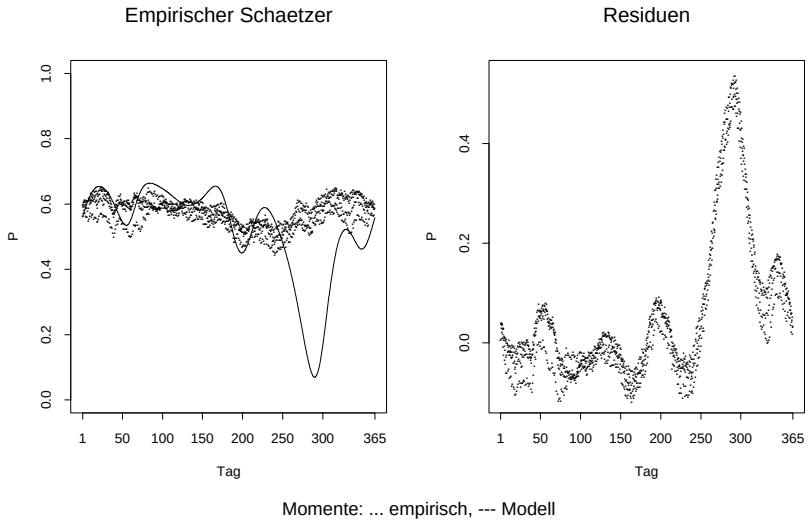


Übergangswahrscheinlichkeit $\phi_{WW}(6)$

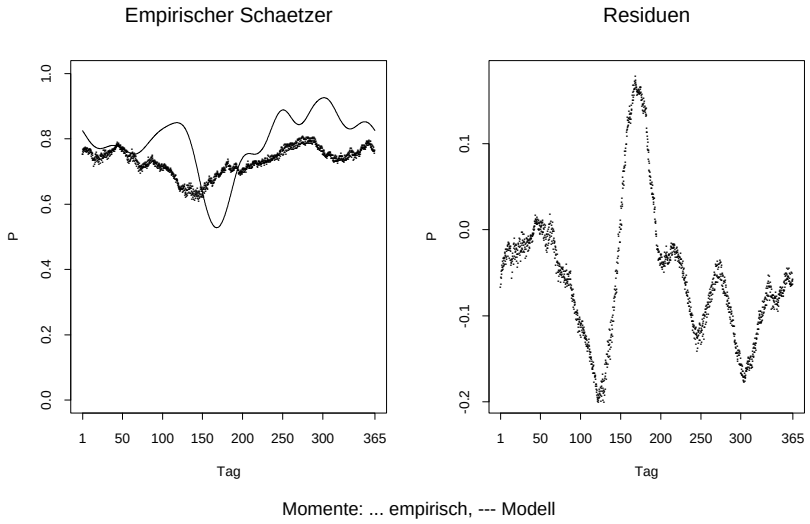


Übergangswahrscheinlichkeit $\phi_{WW}(12)$

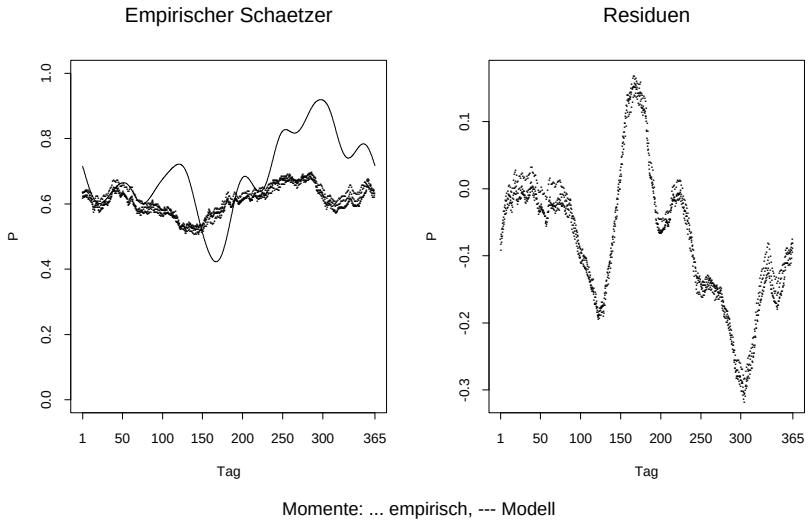
Übergangswahrscheinlichkeit $\phi_{WW}(24)$



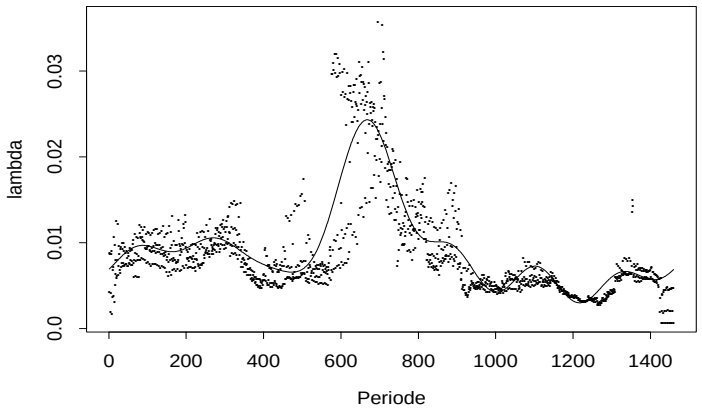
Übergangswahrscheinlichkeit $\phi_{DD}(24)$



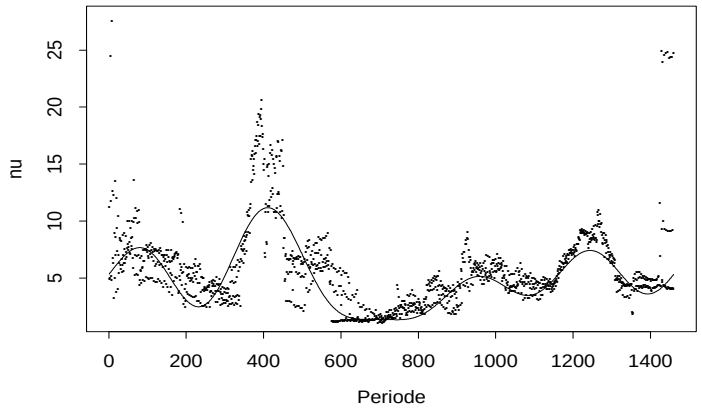
Wahrscheinlichkeit $\phi(24)$



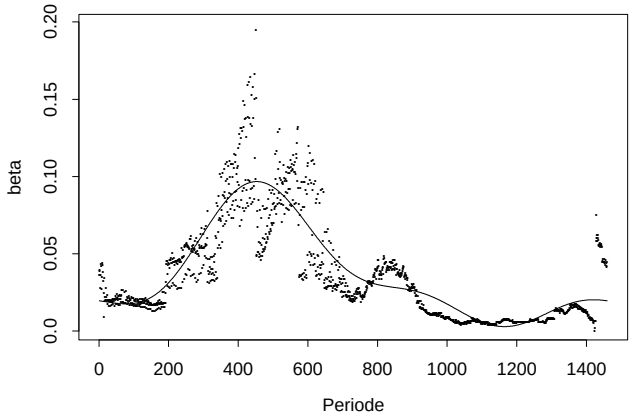
$\hat{\lambda}$ und Glättung durch $\omega_j, j = 1, \ldots, 7$



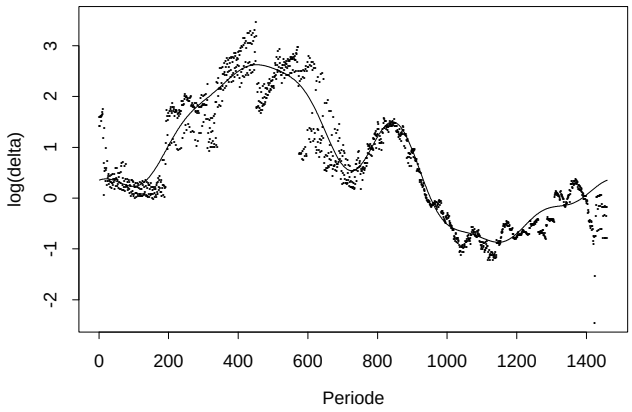
$\hat{\nu}$ und Glättung durch $\omega_j, j = 1, \ldots, 5$



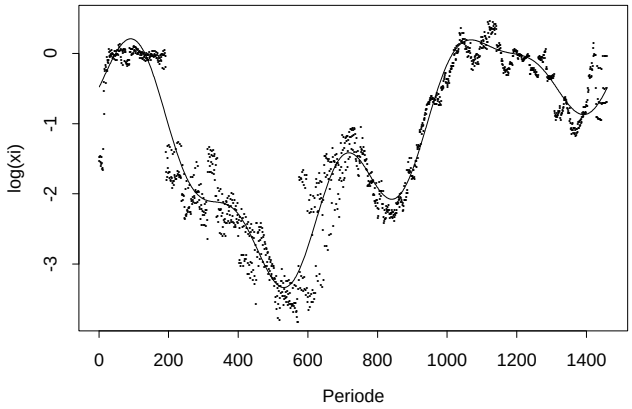
$\hat{\beta}$ und Glättung durch $\omega_j, j = 1, 2, 3$



$\log(\hat{\delta})$ und Glättung durch $\omega_j, j = 1, \ldots, 7$



$\log(\hat{\xi})$ und Glättung durch $\omega_j, j = 1, \ldots, 5$



6 Berechnung der Modellparameter für alle Stunden

In den vorigen Abschnitten sind die Parameter ψ_t des Neyman-Scott-Modelles aus allen im Abstand von sechs Stunden beginnenden 30-Tages-Perioden geschätzt und anschließend durch Fouriersummen geglättet worden.

Die Parameter ψ_t sind nur für jede sechste Stunde geschätzt worden, also für $365 \cdot 4 = 1460$ Perioden. Die angepaßten Fouriersummen ermöglichen aber, die Modellparameter für beliebige Werte $t \in I\!R$ zu berechnen. Um Parameterwerte, beispielsweise λ_t, für alle Stunden $t = 1, \ldots, 24 \cdot 365 = 8760$ zu erhalten, berechnet man

$$\lambda_{t+\,r/6} = R_0 + \sum_j R_j \cos(\omega_j(t + r/6) + \theta_j, \; r = 1, \ldots, 5, \; t = 1, \ldots, 1460,$$

und transformiert den Indexraum:

$$\begin{aligned} \{1, 5/6, 7/6, \ldots, \ldots, 1460 + 5/6\} &\longrightarrow \{1, \ldots 8760\}, \\ t &\longmapsto 6 \cdot (t - 1) + 1. \end{aligned}$$

Damit bezeichnet ψ_t jetzt die zu der in Stunde t beginnenden 30-Tages-Periode gehörenden Modellparameter, $t = 1, \ldots, 8760$. Da sich ψ_t auf den Periodenmittelpunkt bezieht, verschieben wir die Indizes, so daß sich ψ_t auf Stunde t bzw. die in Stunde t zentrierte Periode bezieht. Der Mittelpunkt einer in Stunde h beginnenden Periode liegt in Stunde $h + 12$. Damit liegt der Mittelpunkt der in Stunde t beginnenden 30-Tages-Periode in Stunde $t + 12 + 15 \cdot 24$, ψ_t ist also der zu Stunde $\tilde{t} = t + 15 \cdot 24 + 12$ gehörende Parameter für $t = 1, \ldots, 8388$ bzw. der zu Stunde $(t + 15 \cdot 24 + 12) \bmod 8760$ gehörende Parameter für $t = 8389, \ldots, 8760$. Wir behalten die Bezeichnung ψ_t auch nach der Verschiebung der Indizes bei.

Kapitel 6

Modellvalidation

Zur Überprüfung der Anpassung des saisonalen Neyman-Scott-Modelles wird untersucht, wie gut das Modell Eigenschaften des Prozesses der aggregierten Niederschlagsmengen wiedergibt, die nicht in der Parameterschätzung verwendet worden sind.
Die Momente erster und zweiter Ordnung des Niederschlagsmengenprozesses sowie die Wahrscheinlichkeit, daß ein Intervall beliebiger Länge trocken ist, lassen sich für ein stationäres Modell berechnen. Für das saisonale Modell könnten diese Eigenschaften analytisch hergeleitet oder unter der vereinfachenden Annahme berechnet werden, der Prozeß sei innerhalb eines um Stunde t zentrierten Intervalles stationär. Eine weitere Möglichkeit besteht darin, auf dem angepaßten Modell basierend stündliche Niederschlagsmessungen zu simulieren und die Eigenschaften des künstlichen Datensatzes mit denen des beobachteten zu vergleichen. Diese Vorgehensweise besitzt den Vorteil, daß auch Eigenschaften betrachtet werden können, die analytisch nur schwer herzuleiten sind, wie beispielsweise die Verteilung der sogenannten h-Stunden-Extrema, das sind die jährlichen Maxima der in einem h-Stunden-Intervall beobachteten Niederschlagsmenge, oder die Verteilung der Längen trockener und nasser Runs. Aus einer großen Anzahl, z.B. $N = 1000$, simulierter Jahre kann dann verläßlich auf die tatsächlichen Eigenschaften des Prozesses geschlossen werden.

Wir werden deshalb die Validation anhand eines simulierten Datensatzes durchführen. In Abschnitt 1 wird beschrieben, wie stündliche Niederschlagsmessungen simuliert werden. Anschließend werden die simulierten Daten ausgewertet, und zwar im Hinblick auf die Momente der aggregierten Niederschlagsmengen (Abschnitt 2), die h-Stunden-Extrema (Abschnitt 3), die Niederschlagsmengen in nassen Stunden (Abschnitt 4) und die Runlängen (Abschnitt 5).

1 Simulation stündlichen Niederschlages

Zur Simulation von stündlichem Regen mit dem saisonalen Neyman-Scott-Prozeß werden einzelne Cluster gemäß der Modelldefiniton in Kapitel 5, (1.1) erzeugt. Der

durch die Zellen verursachte Niederschlag wird anschließend zu stündlichem Regen aggregiert.

Die Clusterzentren treten gemäß einem nichthomogenen Poissonprozeß mit Rate $\lambda_t,\ t \in I\!R$, auf. Leichter als ein nichthomogener Poissonprozeß ist dessen Äquivalent in diskreter Zeit zu simulieren, der Bernoulliprozeß mit Erfolgswahrscheinlichkeit $\lambda_t,\ t \in \boldsymbol{Z}$. Daraus resultierende Ungenauigkeiten sind vertretbar, da zum einen aufgrund der kleinen Werte von λ ($\bar{\lambda} = 9.29 \cdot 10^{-3}[\text{Std}^{-1}]$, entsprechend einer mittleren Wartezeit zwischen Stürmen von 4.9 Tagen) keine multiplen Punkte zu erwarten sind, und wir zum anderen an dem beobachtbaren Resultat in diskreter Zeit, nämlich der Serie stündlicher Niederschlagsmengen, interessiert sind. Das Ereignis, daß in Stunde t ein Sturmursprung (Clusterzentrum) auftritt, ist damit bernoulliverteilt mit Wahrscheinlichkeit λ_t. Clusterzentren werden in den Mittelpunkt einer Stunde gelegt, den erwarteten Eintrittszeitpunkt.

Die übrigen Komponenten des Clusterprozesses lassen sich gemäß der Modelldefinition simulieren. Sobald ein Cluster erzeugt ist, wird die mit ihm einhergehende Niederschlagsmenge zu stündlichem Regen aggregiert. Die Simulation wird mit S-PLUS durchgeführt; die Funktion ist im Anhang C.2 aufgeführt.
Mit diesem Algorithmus werden 10 Sets, bestehend aus je 100 Jahren stündlicher Niederschlagsmessungen simuliert, die auf den in Abschnitt 5 beschriebenen Modellparametern basieren. Damit kann die Variation zwischen den Sets erfaßt werden.
Der Simulationsalgorithmus wird in S-PLUS umgesetzt und nutzt deshalb die Objektorientiertheit von S-PLUS aus. In einer anderen Programmiersprache wären gegebenfalls mehr Schleifen notwendig. Die benötigten Zufallszahlen werden mit den in S-PLUS integrierten Zufallszahlengeneratoren wie `rbinom, rpois, rexp` erzeugt.

Das Minimum der beobachteten Niederschlagsmengen in nassen Stunden beträgt aus technischen Gründen 0.1 mm; zusätzlich haben wir in Kapitel 3 einen Mindestschwellenwert für täglichen Regen von $d = 0.2$ mm eingeführt. Die gleichen Schwellenwerte werden folglich für den simulierten Datensatz verwendet, indem kleinere Niederschlagsmengen gleich null gesetzt werden.

Unser Algorithmus zur Simulation stündlichen Niederschlages benötigt als Input die Koeffizienten der Fouriersummenapproximation der Modellparameter $(\lambda, \nu, \beta, \delta, \xi)$, wie sie in Abschnitt 5 gegeben wurden. Die Werte der Modellparameter für jede Stunde eines Jahres werden innerhalb der Simulationsroutine gemäß Abschnitt 6 berechnet. Weiterhin sind die Anzahl $Year$ der zu simulierenden Jahre, die ersten Buchstaben der Ausgabedateinamen `file` und das Verzeichnis der Ausgabedateien `dir` zu übergeben. Der Quellcode ist in Anhang C.2 gegeben.

Simulationsalgorithmus

Schritt 1: Konstruiere die Ausgabedateinamen, `file_1.txt`, ..., `file_`$Year$`.txt`.

Schritt 2: Berechne die Werte der Modellparameter $(\lambda_t, \nu_t, \beta_t, \delta_t, \xi_t)$ für jede Stunde t eines Jahres, $t = 1, \ldots, 8760$ $(365 \cdot 24 = 8760)$, innerhalb der Unterroutine

`ShiftParameter.fun`.
Berechnen von $\eta_t = \beta_t + \delta_t$.

Schritt 3: Anlegen des Vektors $Rain$ der stündlichen Niederschlagsmengen, mit Länge $(365 + 30) \cdot 24 = 9480$ und Einträgen $Rain_t = 0, t = 1, \ldots, 9480$.

Schritt 4: Schleife über alle Jahre $year = 1, \ldots, Year$.

Schritt 5: Erzeugen der Clusterzentren CZ in Jahr y durch bernoulliverteilte Zufallszahlen. Die Erfolgswahrscheinlichkeit in Stunde t beträgt λ_t, $t = 1, \ldots, 8760$: P(Clusterzentrum in Stunde t) $= \lambda_t$.

Schritt 6: Notiere die Stunden mit Clusterstarts als CS.

Schritt 7: Erzeuge die zu den Clusterzentren gehörenden Zellgrößen C, wobei $C_t - 1 \sim \text{Po}(\nu_t - 1)$ mit $t \in CS$.

Schritt 8: Schleife über alle Clusterzentren $CZ_t, t \in ZS$.

Schritt 9: Erzeuge die Distanzen D_c der Zellen vom Clusterzentrum in CS_t:
$D_c \sim \exp(\beta_t), \quad j = 1, \ldots, C_t$.

Schritt 10: Berechne die Startzeiten ZS sowie der Startstunden ZSH der Zellen, wobei das Clusterzentrum CZ_t in die Stundenmitte plaziert wird:

$$ZS_c = t - D_c - 0.5, \; j = 1, \ldots, C_t.$$

$$ZSH_c = [ZS_c] + 1, \quad \text{wobei} \quad [x] = \max\{z \in \boldsymbol{Z} | z \leq x\} \; .$$

Die Addition +1 bewirkt, daß z.B. die Stunde 0:00h - 0:59h als erste Stunde gezählt wird.

Schritt 11: Erzeuge die Zelllängen L_c, $L_c \sim \exp(\eta_{ZSH})$, $j = 1, \ldots, C_t$.

Schritt 12: Erzeuge die Zellintensitäten X_c, $X_c \sim \exp(\xi_{ZSH_c})$, $j = 1, \ldots, C_t$.

Schritt 13: Berechne die Zellendzeitpunkte ZE_c und zugehörigen Stunden ZEH_c: $ZE_c = ZS_c + L_c$ und $ZEH_c = [ZE_c]$.

Schritt 14: Aggregiere den mit Cluster CZ_t verbundenen Niederschlag:
Schleife über die Zellen $c = 1, \ldots, C_t$.

Schritt 15: Fallunterscheidung der Zellen nach Zelllänge:
Fall a: Zellstart und Zellende befinden sich in der gleichen Stunde ($ZSH_c = ZEH_c$) $\Longrightarrow$ Schritt 15a.
Fall b: Zellstart und Zellende befinden sich nicht in der gleichen Stunde ($ZSH_c < ZEH_c$) $\Longrightarrow$ Schritt 15b.

Schritt 15a: $ZSH_c = ZEH_c$ in Stunde s:
$Rain_s = Rain_s + (ZE_c - ZS_c) \cdot X_c$.

Schritt 15b: $ZSH_c < ZEH_c$.

Schritt 15b,i: 1. Stunde der Zelldauer:

$$Rain_{ZSH_c} = \begin{cases} Rain_{ZSH_c} + X_c(ZSH_c - ZS_c), & \text{falls } ZS_c \neq ZSH_c, \\ Rain_{ZSH_c} + X_c, & \text{falls } ZS_c = ZSH_c. \end{cases}$$

Anmerkung: $ZS_c = ZSH_c$ tritt bei Zellstarts zur vollen Stunde, z.B. um 0:00h, auf.

Schritt 15b,ii: Mittlere Stunden der Zelldauer:
Schleife über die Stunden $s = ZSH_c + 1, \ldots, ZEH_c - 1$: $Rain_s = Rain_s + X_c$.

Schritt 15b,iii: letzte Stunde der Zelldauer:

$$Rain_{ZEH_c} = \begin{cases} Rain_{ZEH_c} + (ZE_c - ZEH_c - 1)X_c, & \text{falls } ZEH_c \neq ZE_c, \\ Rain_{ZEH_c} + \frac{1}{60}X_c & \text{falls } ZEH_c = ZE_c. \end{cases}$$

Anmerkung: $ZE_c = ZEH_c$ tritt bei Zellende zur vollen Stunde, z.B. um 0:00h, auf.

Schritt 16: Ende der Schleife über die Zellen *cell* (Schritt 14).

Schritt 17: Ende der Schleife über die Clusterzentren $CZ_t, t \in ZS$ (Schritt 8).

Schritt 18: Speichere den Niederschlag eines Jahres als Matrix ($Rain.Matrix_{(i,j)}$) mit $i = 1, \ldots, 365,\ j = 1, \ldots, 24$.

Schritt 18a: Runde $Rain.Matrix_{(i,j)}$ zu einer Nachkommastelle und beachte den Schwellenwert stündlichen Niederschlages:

$$Rain.Matrix_{(i,j)} = 0, \text{ falls } Rain.Matrix_{(i,j)} < 0.1.$$

Schritt 18b: Schwellenwert für täglichen Niederschlag:

$$Rain.Matrix_{(i,j)} = 0, \text{ falls } \sum_{j=1}^{24} Rain.Matrix_{(i,j)} < 0.2, i = 1, \ldots, 365.$$

Schritt 18c: Speichere ($Rain.Matrix_{(i,j)}$) in Datei `filename_y.txt` in Verzeichnis `dir`.

Schritt 19: Speichere Regen der Tage $d = 1, \ldots, 30$ bzw. der Stunden $s = 1, \ldots, 720$ im neuen Jahr $year + 1$, die durch Clusterzentren in Jahr $year$ verursacht werden:

Schritt 19a: $Dummy.Vektor = (Rain_{8761}, \ldots, Rain_{9480})$.

Schritt 19b: Setze $Rain_t = 0,\ t = 1, \ldots, 9480$.

Schritt 19c: Setze $Rain_t = Dummy.Vektor_t$, $t = 1, \ldots 720$.

Schritt 20: Ende der Schleife über die Jahre $year$ (Schritt 4).

Programmende

Algorithmus zur Berechnung der Werte der Modellparameter

Die Modellparameter sind für alle im Abstand von sechs Stunden beginnenden perioden geschätzt worden. Zur Simulation werden Parameterwerte für jede Stunde benötigt und zusätzlich müssen die Parameter gemäß Kapitel 5.6 verschoben werden. Das geschieht mit Hilfe des folgenden Algorithmus'.
Input der S-PLUS-Routine ist ein S-PLUS-Objekt `PL` vom Typ `list` der Länge 5. Komponente k von PL, das ist $PL[[k]]$, enthält die an Modellparameter k angepaßte Fouriersumme und zwar als Objekt vom Typ `list`. Komponenten von $PL[[k]]$ sind $PL[[k]]\$Freq$, die Indizes j der verwendeten Frequenzen ω_j sowie $PL[[k]]\$Koef$, die zugehörigen Koeffizienten in der Form (a_j, b_j) der an Parameter k angepaßten Fouriersummen, $k = 1, \ldots, 5$. Der Quellcode befindet sich ebenfalls in Anhang C.2.

Schritt 1: Erzeuge die (8760,5)-Matrix $Fitted$ der Parameterwerte:
$Fitted_{(i,k)} = 0, i = 1, \ldots 8760,\ k = 1, \ldots, 5.$ $(8760 = 365 \cdot 24)$

Schritt 2: Schleife über die Komponenten $k = 1, \ldots 5$ von $\psi = (\lambda, \nu, \beta, \delta, \xi)$.

Schritt 3: Bestimme die Anzahl f an Fourierfrequenzen der Fouriersummenapproximation von ψ_k: $f =$ Länge von $PL[[k]]\$Freq$.

Schritt 4: Initialisiere den Vektor ZP der Zeitpunkte:
$ZP = (1, 7/6, 8/6, 9/6, \ldots, 1459 + 5/6)$. $(1460 = 365 \cdot 4$, der Vektor hat also die Länge 1460).

Schritt 3: Setze
$Fitted_{(i,k)} = |\, PL[[k]]\$Koef[1] \,|$ für $i = 1, \ldots, 8760$, das ist der Mittelwert.

Schritt 5: Schleife über die Komponenten $j = 2, \ldots, f + 1$ von $PL[[k]]$.

Schritt 6: Für $i = 1, \ldots, 8760$:

$$\begin{aligned} Fitted_{(i,k)} &= Fitted_{(i,j)} + |PL[[k]]\$Koef[j]| \\ &\quad \cdot \cos\left(\frac{2\pi PL[[k]]\$Freq[j] \cdot ZP_i}{1460} + \arg(PL[[k]]\$Koef[j])\right). \end{aligned}$$

Schritt 7: Ende der Schleife über die Frequenzen j (Schritt 5).

Schritt 8: Ende der Schleife über die Komponenten k (Schritt 2).

Schritt 9: Translation der Werte der Parameter. Sei $Fitted_i^t$ die i-te Zeile der Parametermatrix (vergl. Abschnitt 5.6):

$$Fitted = (Fitted_{8389}^t, \ldots, Fitted_{8760}^t, Fitted_1^t, \ldots, Fitted_{8388}^t)\ .$$

Schritt 10: Ausgabe von *Fitted*.

Programmende

2 Momente der simulierten Daten

Grundlegend für ein Modell ist dessen Fähigkeit, die Momente erster und zweiter Ordnung, also den Erwartungswert, die Varianz und die Autokorrelationsstruktur, des aggregierten Niederschlages korrekt zu reproduzieren, und zwar möglichst für unterschiedliche Längen h des Aggregationsintervalles. Diese Eigenschaften werden aus den simulierten Daten auf die gleiche Weise wie aus den beobachteten geschätzt, nämlich separat für jeden Monat. Wir betrachten deshalb die Werte der Eigenschaften $f \in \mathcal{F}$, die sich aus den simulierten Datensätzen berechnen, wobei $\mathcal{F}$ wie in Kapitel 5. (2.2) ist. Diese Werte werden für jedes der 10 simulierten Datensets berechnet und ermöglichen so Aufschluß darüber, wie stark die Eigenschaften der unterschiedlichen Sets variieren und ob schon 100 Jahre simulierten stündlichen Niederschlages ausreichen, um verläßliche Aussagen über den tatsächlichen Prozeß treffen zu können.

Sei $\hat{f}_k$ der aus den Beobachtungen in Monat $k = 1, \ldots, 12$ geschätzte Wert der Eigenschaft f, $\tilde{f}_{j,k}$ der aus Simulationsset $j, j = 1, \ldots, 10$, geschätzte Wert für Monat k und $\bar{f} := \sum_{j=1}^{10} \tilde{f}_{j,k}/10$ deren Mittelwert. Die Mittel $\bar{f}$ sind zusammen mit den entsprechenden beobachteten Werten in den Tabellen 6.1 bis 6.3 zu sehen. Die resultierenden Kurven der $\tilde{f}_{j,k}$ sowie die entsprechenden beobachteten Eigenschaften sind in Abb. 6.1 - 6.9 dargestellt.

Das 1-Stunden-Mittel sowie die Varianzen des h-Stunden-Niederschlages variieren zwischen den Sets deutlich mehr als die Autokorrelationen und die Übergangswahrscheinlichkeiten, insbesondere in den Sommermonaten, in denen diese Eigenschaften große Werte besitzen. Die Variation zwischen den Simulationssets weist darauf hin, daß 100 simulierte Jahre nicht ausreichen, um verläßlich auf Eigenschaften des Modelles schließen zu können.
Die Varianzen $\tilde{\gamma}_{j,k}(h)$ der simulierten Niederschlagsmengen verändern sich im Verlauf der Monate $k = 1, \ldots, 12$ glatter als die der Beobachtungen. Die empirischen h-Stunden-Varianzen $\hat{\gamma}_k(h)$ werden saisonal unterschiedlich gut wiedergegeben: In den Sommermonaten Mai - August ist $\hat{\gamma}_k(h)$ unter dem Modell tendenziell etwas zu groß, während in den Monaten Oktober - März die h-Stunden Varianzen $\hat{\gamma}_k(h)$ gut wiedergegeben werden, vergl. Abb. 6.1 - 6.3. Die Lag(1)-Autokorrelationen werden unter dem Modell von allen einzelnen Sets exakt wiedergegeben; zwischen den Simulationssets ist nur geringe Variation erkennbar, vergl. Abb. 6.4 - 6.6.
Die Übergangswahrscheinlichkeiten $\tilde{\phi}_{WW,j,k}(h)$ variieren zwischen den Simulationssets j nur wenig und entsprechen für $h = 1, 3, 6$ ihren empirischen Äquivalenten, $\hat{\phi}_{WW}(h)$ ist für $h = 12, 24$ etwas zu groß, vergl. Abb. 6.6 - 6.8.
Die Wahrscheinlichkeiten $\tilde{\phi}_{j,k}(24)$ und $\tilde{\phi}_{DD,j,k}(24)$ variieren saisonal stärker als die empirischen Schätzer. Das ist erstaunlich, weil $\phi_{WW}(24)$ und $\phi(24)$ in der Modellanpassung verwendet worden sind. Die aus den simulierten Niederschlagsdaten berechneten Momente entsprechen den empirischen sogar wesentlich stärker als die analytisch berechneten. Offensichtlich unterscheiden sich die Second-Order-Eigenschaften des saisonalen NSRP-Modelles von denen des stationären Modelles. Eine analytische Untersuchung der Eigenschaften des saisonalen Modelles wäre hier notwendig.

Tabelle 6.1: Erwartungswert und Varianzen

		$\hat{\mu}(1)$	$\hat{\gamma}(1)$	$\hat{\gamma}(3)$	$\hat{\gamma}(6)$	$\hat{\gamma}(12)$	$\hat{\gamma}(24)$
Jan.	beob.	0.045	0.13	0.92	2.11	5.05	12.97
	sim.	0.048	0.11	0.58	1.39	3.19	7.41
Feb.	beob.	0.049	0.07	0.45	1.30	3.50	9.46
	sim.	0.052	0.07	0.36	0.87	2.02	4.70
März	beob.	0.045	0.17	0.64	1.45	3.54	8.62
	sim.	0.049	0.19	0.70	1.54	3.42	7.74
April	beob.	0.066	0.21	0.92	2.39	6.05	14.98
	sim.	0.078	0.28	1.07	2.58	6.43	16.26
Mai	beob.	0.090	0.62	2.38	5.43	13.22	27.24
	sim.	0.098	0.67	2.42	5.72	14.08	35.00
Juni	beob.	0.106	0.61	2.58	6.02	14.06	30.87
	sim.	0.109	0.82	2.86	6.23	13.86	30.74
Juli	beob.	0.069	0.28	1.36	3.31	8.23	17.55
	sim.	0.101	0.53	2.32	5.15	11.25	24.48
Aug.	beob.	0.085	0.43	1.79	4.59	11.10	23.35
	sim.	0.072	0.39	1.51	3.30	7.24	15.97
Sept.	beob.	0.061	0.17	0.80	2.22	5.64	13.19
	sim.	0.070	0.24	1.30	3.26	7.84	18.44
Okt.	beob.	0.047	0.01	0.55	1.53	4.35	10.18
	sim.	0.045	0.07	0.49	1.42	3.53	8.13
Nov.	beob.	0.059	0.11	0.70	1.86	5.24	13.15
	sim.	0.054	0.01	0.67	1.87	4.72	10.59
Dez.	beob.	0.051	0.25	0.92	2.10	5.05	11.28
	sim.	0.088	0.26	1.48	3.80	8.96	20.40

Obere Zeile: Originaldaten.
Untere Zeile: Mittel $\bar{f}_{sim} = \frac{1}{10} \sum_{j=1}^{10} \hat{f}_{sim,j}/10$ über die simulierten Sets $j = 1, \ldots, 10$.

Jahresniederschlag und Monatsniederschlag

Ein Modell zur Beschreibung stündlichen Niederschlages sollte natürlich auch grundlegende Eigenschaften des Monats- und des Jahresniederschlages korrekt wiedergeben, wie den Erwartungswert und die Varianz. Die aus allen 1000 simulierten Jahren berechneten Werte sind in Tabelle 6.4 genannt; Histogramme der simulierten und der beobachteten Niederschläge des Jahres sowie ausgewählter Monate sind in den Abbildungen 6.10 - 6.12 zu sehen.

Das Jahresmittel ist unter dem Modell um 12%, die Varianz des Jahresniederschlages um 8% zu groß. Die simulierten und beobachteten Werte unterscheiden sich insbesondere in den Sommermonaten. Die Unterschiede in den Histogrammen können auf die unterschiedliche Anzahl der den Beobachtungen bzw. den künstlichen Daten zugrundeliegenden Jahren zurückgeführt werden. Da der Hauptaspekt des Neyman-Scott-Modelles in der Beschreibung stündlichen Niederschlages liegt, ist die nur grobe Anpassung akzeptabel. Eine bessere Beschreibung von niedrigaufgelöstem Niederschlag, wie Tages-, Monats- und Jahresniederschlag, ist mit einem weniger aufwendigem Modell, wie dem in Kapitel 4 beschriebenem Markovkettenmodell, möglich.

Tabelle 6.2: Korrelationen

		$\hat{\rho}(1,1)$	$\hat{\rho}(1,3)$	$\hat{\rho}(1,6)$	$\hat{\rho}(1,12)$	$\hat{\rho}(1,24)$
Jan.	beob.	0.56	0.18	0.11	0.40	0.32
	sim.	0.47	0.20	0.14	0.15	0.18
Feb.	beob.	0.66	0.52	0.44	0.36	0.23
	sim.	0.44	0.20	0.16	0.16	0.21
März	beob.	0.14	0.15	0.17	0.14	0.10
	sim.	0.16	0.10	0.11	0.13	0.13
April	beob.	0.25	0.27	0.24	0.24	0.19
	sim.	0.15	0.20	0.25	0.26	0.20
Mai	beob.	0.18	0.18	0.15	0.11	0.21
	sim.	0.13	0.19	0.23	0.23	0.16
Juni	beob.	0.25	0.17	0.14	0.09	0.07
	sim.	0.11	0.09	0.11	0.12	0.11
Juli	beob.	0.36	0.21	0.18	0.06	0.05
	sim.	0.32	0.12	0.09	0.09	0.09
Aug.	beob.	0.28	0.23	0.16	0.07	0.13
	sim.	0.21	0.09	0.09	0.10	0.13
Sept.	beob.	0.40	0.34	0.23	0.16	0.16
	sim.	0.51	0.26	0.18	0.17	0.19
Okt.	beob.	0.51	0.41	0.34	0.18	0.12
	sim.	0.75	0.45	0.26	0.17	0.14
Nov.	beob.	0.62	0.40	0.36	0.23	0.19
	sim.	0.71	0.41	0.24	0.13	0.13
Dez.	beob.	0.13	0.16	0.15	0.10	0.11
	sim.	0.58	0.28	0.17	0.14	0.16

Obere Zeile: Originaldaten.
Untere Zeile: Mittel $\bar{f}_{sim} = \frac{1}{10} \sum_{j=1}^{10} \hat{f}_{sim,j}/10$ über die simulierten Sets $j = 1, \ldots, 10$.

Abbildung 6.1: Beobachtete und simulierte Momente: $\mu(1)$, $\gamma(1)$

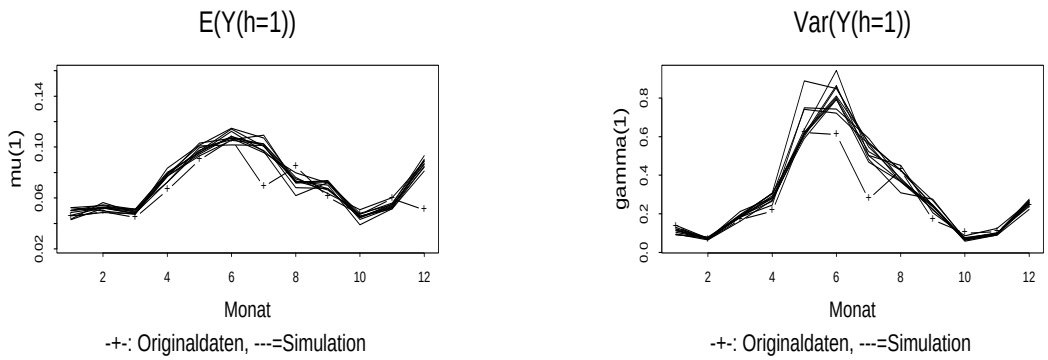


Tabelle 6.3: Wahrscheinlichkeiten nasser und trockener Intervalle

		$\hat{\phi}_{ww}(1)$	$\hat{\phi}_{ww}(3)$	$\hat{\phi}_{ww}(6)$	$\hat{\phi}_{ww}(12)$	$\hat{\phi}_{ww}(24)$	$\hat{\phi}_{DD}(24)$	$\hat{\phi}_{D}(24)$
Jan.	beob.	0.16	0.24	0.33	0.44	0.58	0.77	0.64
	sim.	0.13	0.25	0.37	0.53	0.69	0.73	0.54
Feb.	beob.	0.18	0.27	0.35	0.43	0.56	0.75	0.64
	sim.	0.18	0.33	0.47	0.64	0.79	0.63	0.37
März	beob.	0.13	0.21	0.27	0.39	0.51	0.75	0.66
	sim.	0.08	0.19	0.31	0.46	0.65	0.71	0.55
April	beob.	0.15	0.23	0.33	0.44	0.60	0.74	0.60
	sim.	0.12	0.26	0.38	0.52	0.67	0.80	0.62
Mai	beob.	0.14	0.21	0.29	0.39	0.58	0.66	0.56
	sim.	0.10	0.21	0.31	0.46	0.62	0.83	0.69
Juni	beob.	0.12	0.19	0.26	0.38	0.55	0.63	0.55
	sim.	0.07	0.18	0.29	0.45	0.64	0.65	0.50
Juli	beob.	0.14	0.18	0.25	0.34	0.53	0.74	0.64
	sim.	0.09	0.18	0.30	0.46	0.66	0.57	0.44
Aug.	beob.	0.14	0.21	0.29	0.37	0.52	0.73	0.64
	sim.	0.07	0.17	0.29	0.44	0.62	0.69	0.55
Sept.	beob.	0.10	0.17	0.24	0.33	0.47	0.75	0.68
	sim.	0.13	0.22	0.33	0.48	0.66	0.78	0.61
Okt.	beob.	0.13	0.21	0.29	0.43	0.57	0.80	0.69
	sim.	0.11	0.17	0.26	0.39	0.58	0.71	0.59
Nov.	beob.	0.16	0.24	0.31	0.44	0.58	0.75	0.63
	sim.	0.11	0.18	0.27	0.41	0.60	0.66	0.54
Dez.	beob.	0.12	0.19	0.28	0.39	0.52	0.76	0.67
	sim.	0.13	0.23	0.35	0.50	0.67	0.65	0.48

Obere Zeile: Originaldaten.
Untere Zeile: Mittel $\bar{f}_{sim} = \frac{1}{10} \sum_{j=1}^{10} \hat{f}_{sim,j}/10$ über die simulierten Sets $j = 1, \ldots, 10$.

Abbildung 6.2: Beobachtete und simulierte Momente: $\gamma(3)$, $\gamma(6)$

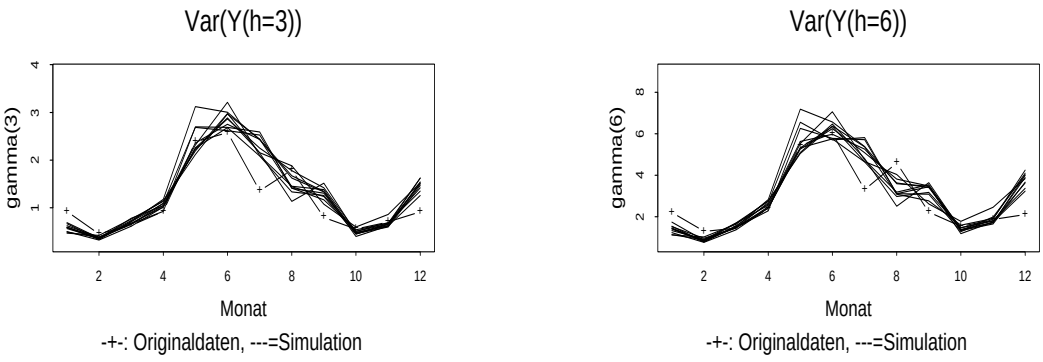


Abbildung 6.3: Beobachtete und simulierte Momente: $\gamma(12)$, $\gamma(24)$

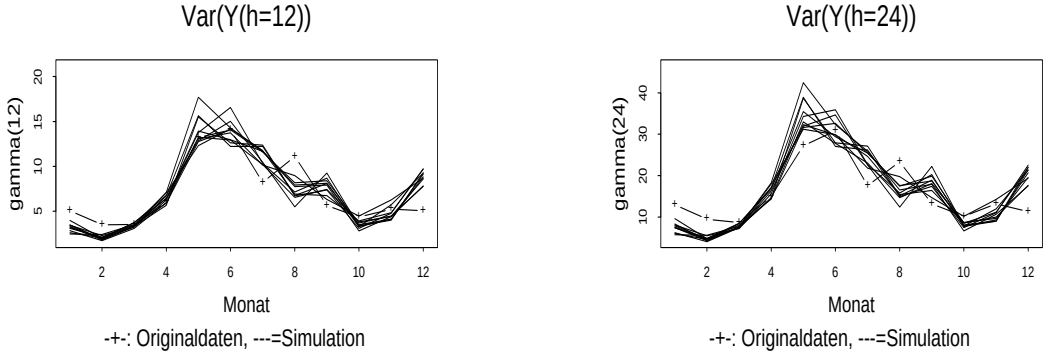


Abbildung 6.4: Beobachtete und simulierte Momente: $\rho(1,1)$, $\rho(3,1)$

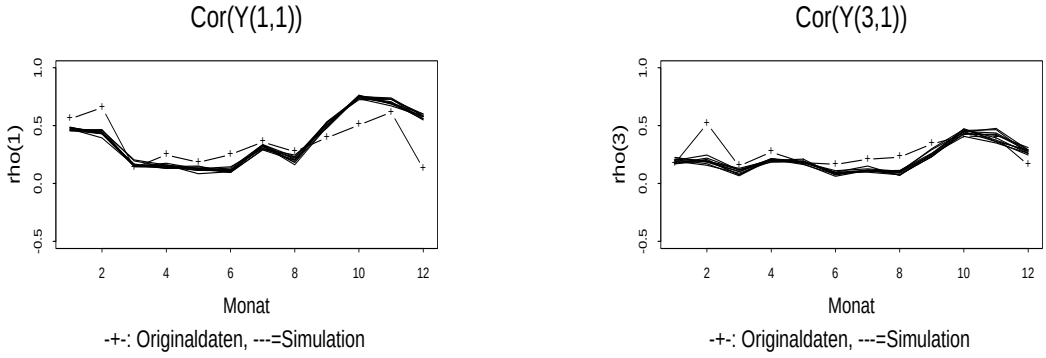


Abbildung 6.5: Beobachtete und simulierte Momente: $\rho(6,1)$, $\rho(12,1)$

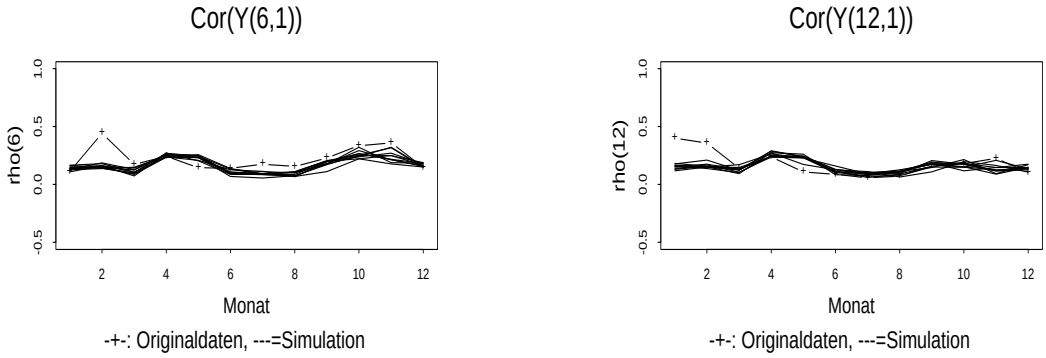


Abbildung 6.6: Beobachtete und simulierte Momente: $\rho(24,1)$, $\phi_{WW}(1)$

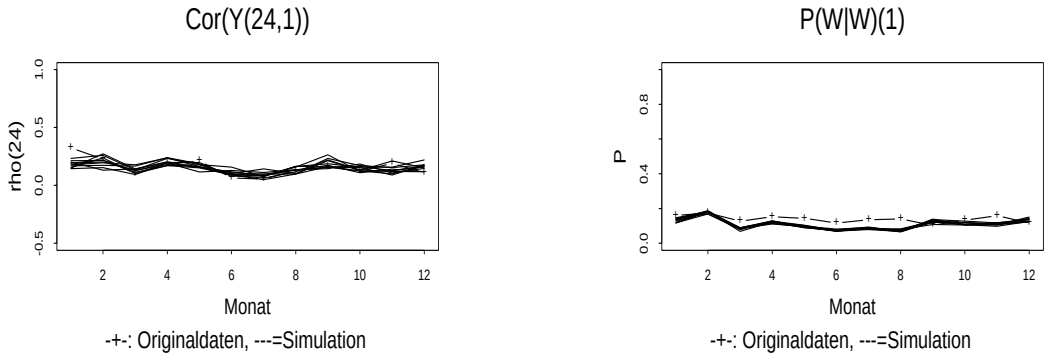


Abbildung 6.7: Beobachtete und simulierte Momente: $\phi_{WW}(3)$, $\phi_{WW}(6)$

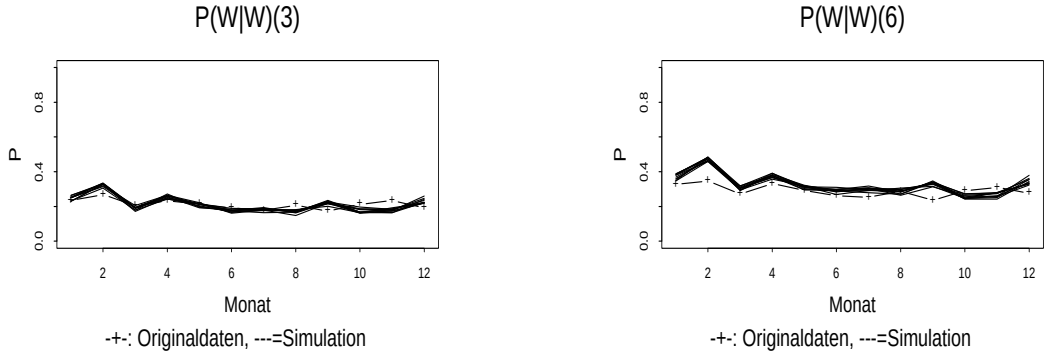


Abbildung 6.8: Beobachtete und simulierte Momente: $\phi_{WW}(12)$, $\phi_{WW}(24)$

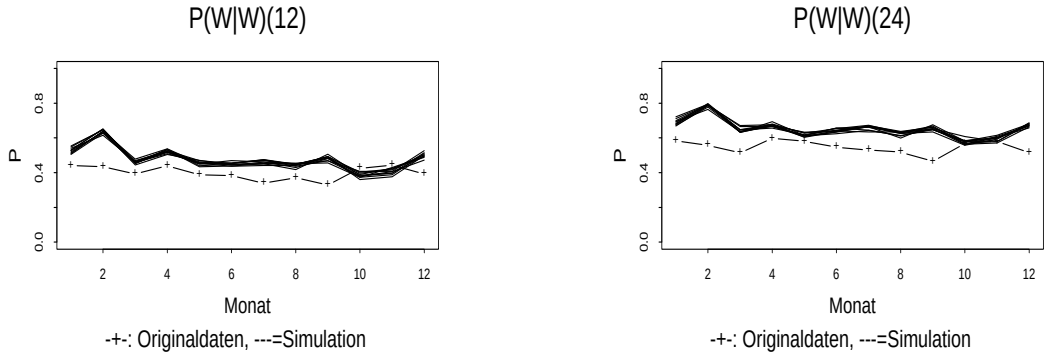


Abbildung 6.9: Beobachtete und simulierte Momente: $\phi_{DD}(24)$, $\phi(24)$

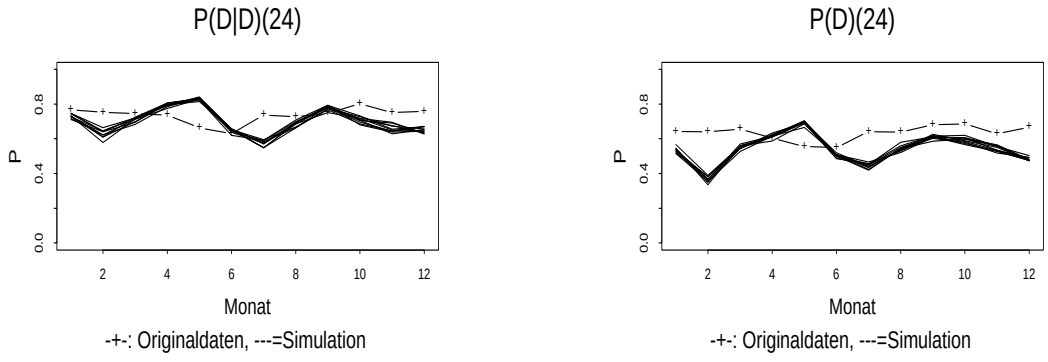


Tabelle 6.4: Jahres- und Monatsniederschläge

Monat	Mittel [mm]		Varianz [mm^2]	
	Beobachtet	Simuliert	Beobachtet	Simuliert
Jan.	33.57	35.92	34.21	20.25
Feb.	33.04	35.24	22.60	17.03
März	33.20	36.54	28.18	18.34
April	47.59	55.87	31.26	26.72
Mai	66.81	72.77	35.30	37.71
Juni	76.23	78.09	32.34	33.51
Juli	51.17	75.43	19.86	30.87
Aug.	62.85	53.85	34.40	27.47
Sept.	43.87	50.22	20.39	31.40
Okt.	34.98	33.75	23.83	21.97
Nov.	42.73	38.97	23.56	25.86
Dez.	37.67	65.09	23.82	33.86
Jahr	563.71	631.75	110.99	101.46

Abbildung 6.10: Jahresniederschlag

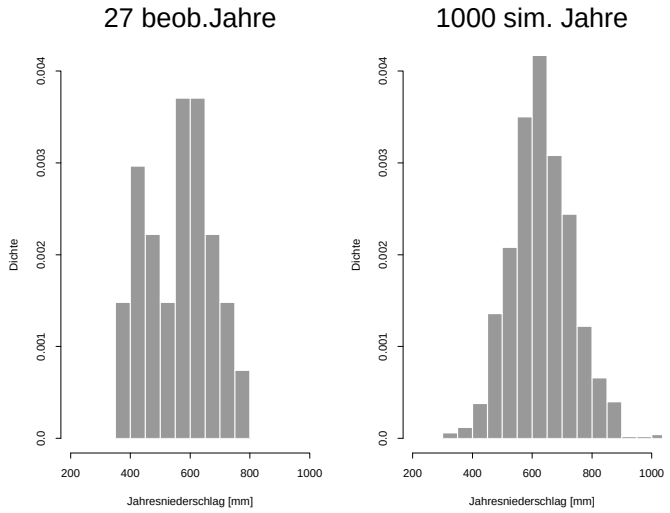


Abbildung 6.11: Monatssniederschlag: Januar

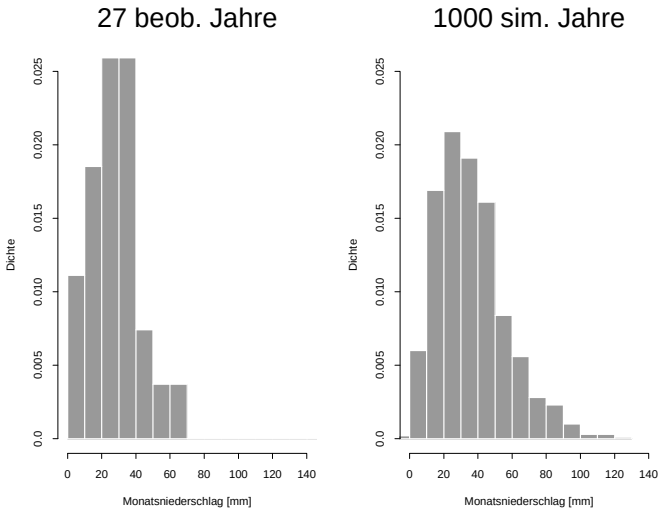
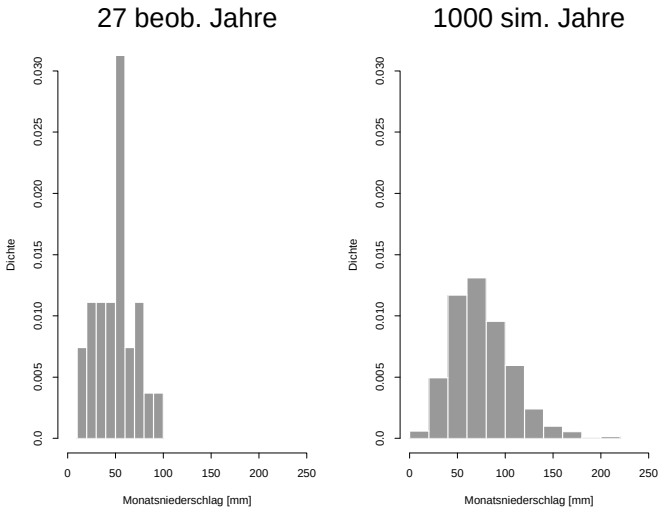


Abbildung 6.12: Monatssniederschlag: Juli



3 Extremer Niederschlag

Einige praktische Anwendungen eines Niederschlagsmodelles erfordern die Beschreibung der maximal in einem bestimmten Zeitraum zu erwartenden Niederschlagsmengen. Beispiele hierfür sind die Konstruktion von urbanen Kanalisationssystemen oder von Hochwasserschutzeinrichtungen, aber auch die Kalkulation entsprechender Versicherungstarife. Dieses Gebiet wird als Extremwertanalyse bezeichnet. Man möchte Aussagen wie diese treffen: „Eine h-Stunden-Niederschlagsmenge y wird mit Wahrscheinlichkeit p in jedem Jahr mindestens einmal beobachtet werden." Unter der plausiblen Annahme, daß die jährlichen Maxima $X^{(h)}$ der h-Stunden-Niederschläge unabhängig voneinander sind, lassen sie sich asymptotisch durch eine Extremwertverteilung beschreiben (Benjamin und Cornell [1977]). Die Auswahl des Typus' der Extremwertverteilung (Typ I, II oder III) kann schwierig sein, zusätzlich existieren Verteilungen, deren Extrema gegen keinen Typ der Extremwertverteilung konvergieren. Die Konvergenzgeschwindigkeit hängt von der exakten Verteilung ab und kann langsam sein, so daß die aus der Approximation durch eine Extremwertverteilung gewonnenen Schätzer ungenau sein können. Wenn das Neyman-Scott-Modell die Beobachtungen angemessen beschreibt, lassen sich Aussagen über extreme Niederschlagsmengen leicht aus simulierten Daten ableiten. Dazu verwenden wir die simulierten 1000 Jahre stündlicher Niederschlagsmessungen.

Unter der Annahme, daß die jährlichen h-Stunden-Maxima $(\tilde{X}_1^{(h)}, \ldots, \tilde{X}_N^{(h)})$ der simulierten stündlichen Niederschlagsmengen der Jahre $n, n = 1, \ldots, N$, die tatsächliche Verteilung approximieren, können die oben gestellten Fragen entweder unter direkter Verwendung der empirischen Quantile der simulierten Maxima $\tilde{X}_n^{(h)}$ beantwortet werden oder, indem an letztere eine Verteilung angepaßt wird. Da es möglich ist, beliebig viele, beispielsweise $N = 1000$, Jahre zu simulieren, ermöglicht die Anpassung einer Verteilung an die simulierten Maxima verläßlichere Aussagen als die Anpassung einer Verteilung an die wesentlich geringere Anzahl von $N = 27$ beobachteten Jahren.
Zunächst muß jedoch die Fähigkeit des Neyman–Scott–Modelles, extreme Niederschlagsmengen angemessen zu reproduzieren, überprüft werden. Dazu werden wir die Verteilungen der 1–, 6– und 24–Stunden–Extrema betrachten.
Im folgenden seien $X_n^{(h)}$ und $\tilde{X}_n^{(h)}$ das h-Stunden-Maximum der Beobachtungen bzw. der Simulationen in Jahr $n, n = 1, \ldots, N$. Die p-Quantile der $X_n^{(h)}$ bzw. der $\tilde{X}_n^{(h)}$, $n = 1, \ldots, N$, seien $Q_e(p)$ bzw. $Q_s(p)$. Die Quantile $Q_e(p)$ werden durch lineare Interpolation zwischen den Ordnungsstatistiken $X_{(1)}^{(h)} \leq X_{(2)}^{(h)} \leq \ldots \leq X_{(N)}^{(h)}$ unter der Annahme berechnet, die i-te Ordnungsstatistik sei das $(i-1)/(n-1)$–Quantil. Das bewirkt, daß die größte Beobachtung das 100%-Quantil ist und die extremen Quantile wie $Q(0.99)$ und $Q(0.98)$ der 27 Beobachtungen durch lineare Interpolation zwischen dem größten und dem zweitgrößten beobachteten Wert der $X_n^{(h)}$ berechnet werden, das 95%-Quantil durch Interpolation zwischen dem zweit- und dem drittgrößten Wert der Jahresmaxima. Die Quantile $Q_s(p)$ werden analog berechnet.

Ein-Stunden-Extrema

Als 1-Stunden-Maximum oder 1-Stunden-Extremum $X_n^{(1)}, n = 1, \ldots, N$, bezeichnen wir das Maximum der während eines Jahres registrierten 1-Stunden-Niederschlagsmengen. Stunden sind hier die sich nicht überlappenden Intervalle 0:00 - 0:59h, 1:00 - 1:59h etc. Ebenso überlappen sich die Jahre nicht und dauern vom 1. Januar bis zum 31. Dezember.

Tabelle 6.5: 1-Stunden-Maxima

Monat		Mittel	50%	75%	90%	95%	98%	99%
Jahr	beob.	23.95	20.30	29.40	39.68	42.26	52.74	58.12
	sim.	20.14	18.31	23.16	30.42	36.82	44.29	47.89
Jan.	beob.	3.15	2.00	2.90	5.26	11.26	15.95	17.23
	sim.	4.48	4.04	5.53	7.42	8.93	11.20	12.89
Feb.	beob.	2.68	1.80	3.05	5.90	6.94	7.72	8.11
	sim.	3.18	3.00	3.90	4.90	5.63	6.82	7.37
März	beob.	5.04	2.20	4.10	7.36	18.26	32.35	37.57
	sim.	6.94	5.91	8.32	11.70	15.19	20.62	23.62
Apr.	beob.	6.50	3.40	7.75	15.94	21.26	26.02	27.66
	sim.	7.94	7.06	9.64	12.78	15.73	20.23	24.19
Mai	beob.	11.62	7.10	10.55	27.00	36.34	50.66	57.08
	sim.	12.99	11.21	15.81	22.34	28.24	36.01	41.26
Juni	beob.	12.41	10.80	14.40	22.20	30.13	35.61	36.80
	sim.	14.44	12.56	17.46	23.54	29.92	37.31	41.43
Juli	beob.	8.68	7.20	9.55	15.06	18.54	24.46	26.98
	sim.	10.82	9.61	13.31	17.57	20.41	24.02	25.85
Aug.	beob.	10.64	10.1	15.25	16.94	19.31	23.90	25.85
	sim.	9.78	8.74	12.21	16.85	20.10	22.94	25.42
Sep.	beob.	6.38	5.80	8.00	12.46	13.44	13.60	13.65
	sim.	6.33	5.62	8.09	10.85	13.46	16.67	18.29
Okt.	beob.	4.00	2.70	4.55	9.32	12.97	14.12	14.56
	sim.	2.65	2.50	3.28	4.30	4.96	5.82	6.40
Nov.	beob.	3.86	3.20	4.60	6.80	7.45	11.87	14.19
	sim.	3.17	3.04	4.00	5.06	6.08	7.14	8.08
Dez.	beob.	7.39	2.90	7.00	19.96	25.26	33.41	37.20
	sim.	6.00	5.56	7.39	9.32	10.47	13.09	14.16

Mittel und Quantile der jährlichen 1-Stunden-Maxima aus 27 beob. und 1000 sim. Jahren.

Das 99%-Quantil $Q_e(0.99)$ der Beobachtungen ist wesentlich größer als das der Simulationen, $Q_s(0.99)$. Wie eben beschrieben, folgt daraus $Q_e(0.98) > Q_s(0.98)$. Die Quantile $Q_e(p)$ und $Q_s(p)$ stimmen für $p \leq 0.9$ recht gut überein, ebenso die Mittelwerte, vergl. Tab. 6.5. Aber auch bei Berechnung der Quantile auf Basis von 100 simulierten Jahren ergeben sich die gleichen Resultate, vergl. Tab. 6.6. Der QQ-Plot in Abbildung 6.14 (b) zeigt, daß die Verteilung der $X_n^{(1)}$ mehr Masse im Bereich von 30 - 45mm besitzt als die Verteilung der $\tilde{X}_n^{(1)}$. Die kleineren Quantile stimmen gut überein. Das Maximum der $\tilde{X}_n^{(1)}$ ist größer als das der $X_n^{(1)}$, so daß unter dem Modell auch genügend große Werte erhalten werden können. Die gleichen Aussagen treffen auf die einzelnen Monate zu, vergl. Abb. 6.13. Ein Vergleich der Histogramme der beobachteten und der simulierten 1-Stunden-Maxima besitzt aufgrund der

Tabelle 6.6: 1-Stunden-Maxima pro Set

Monat	Mittel	50%	75%	90%	95%	98%	99%
Set 1	20.42	18.09	22.67	31.62	37.27	38.59	41.27
Set 2	19.89	18.39	22.10	33.53	40.85	44.31	45.32
Set 3	20.49	19.22	23.14	30.72	32.54	42.76	45.01
Set 4	20.57	18.91	23.85	29.76	34.25	41.48	44.65
Set 5	22.49	19.40	25.33	35.11	36.98	56.31	75.00
Set 6	19.75	17.64	21.74	30.15	36.88	47.31	47.91
Set 7	20.11	18.13	23.43	28.05	32.28	37.38	38.79
Set 8	18.35	16.21	21.02	27.18	35.00	38.47	39.43
Set 9	19.07	17.46	21.80	26.64	34.78	45.03	45.61
Set 10	20.25	17.58	23.24	32.17	36.82	41.38	48.36
Mittel	20.14	18.10	22.83	30.49	35.77	43.30	47.14
Beob.	23.95	20.30	29.40	39.68	42.26	52.74	58.12

Zeilen 1 - 10: Mittel und Quantile der sim. 1-Stunden-Maxima aus Sets á 100 Jahren.
Zeile 11: Spaltenmittel der Zeilen 1 - 10.
Zeile 12: Mittel und Quantile der beob. 1-Stunden-Maxima aus 27 Jahren.

unterschiedlichen Anzahl an zugrundeliegenden Jahren keine Aussagekraft und wird deshalb hier nicht gegeben.
Wir folgern, daß das Modell fähig ist, die Verteilung der 1-Stunden-Extrema angemessen wiederzugeben und basieren weitere Aussagen auf die simulierten Werte $\tilde{X}_n^{(1)}$.

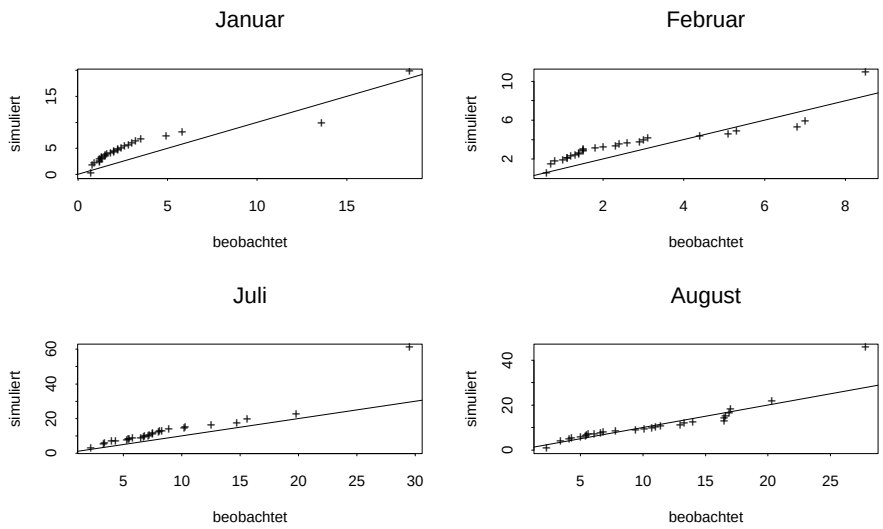
Die Jahresmaxima der 1-Stunden-Maxima $\tilde{X}_n^{(1)}$ können durch eine Lognormalverteilung beschrieben werden, vergl. Abb. 6.14 (a). Zusätzlich haben wir eine Weibull- und eine Gammaverteilung angepaßt, die beide offensichtlich nicht angemessen sind. Ebenso können die 1-Stunden-Maxima $X_{n,k}^{(1)}$ für jeden Monat k, $k = 1, \ldots, 12$, durch eine Lognormalverteilung beschrieben werden, vergl. Abb. 6.14. Die Veränderung der Lageparameter $\mu_k = \mathrm{E}(\log(X_{n,k}^{(1)}))$ der angepaßten Lognormalverteilung im Jahresverlauf folgt einem Cosinus, während der Formparameter $\sigma_k = \sqrt{\mathrm{var}(\log(X_{n,k}^{(1)}))}$ annähernd konstant ist, vergl. Abb. 6.13:

$$\hat{\mu}_k = 1.73 + 0.65\ \cos(2\pi\, k/12\ + 3.06), \quad k = 1, \ldots 12.$$

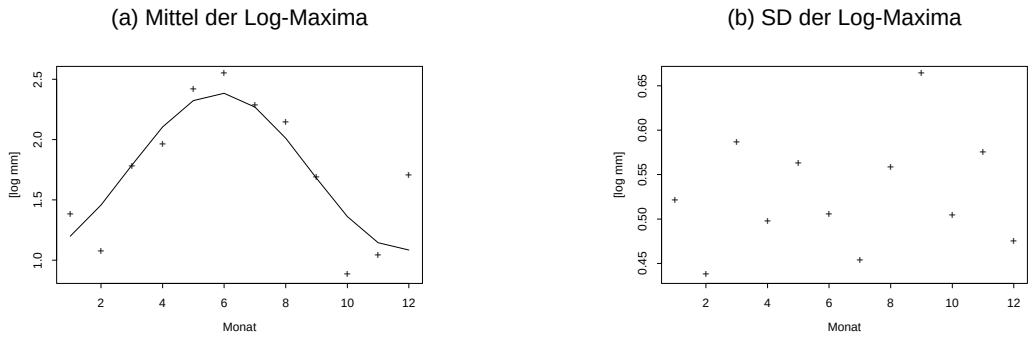
Die Koeffizienten sind mit der Methode der kleinsten Quadrate geschätzt worden, die angepaßten Werte sind als $\tilde{\mu}$ in der Tabelle auf Seite 148 bzw. in Abbildung 6.13 zu sehen. Die Standardabweichung der Log-Maxima wird als konstant angenommen und durch den Mittelwert $\tilde{\sigma} = \sum_{k=1}^{12} \hat{\sigma}_k/12 = 0.53$, $k = 1, \ldots, 12$, geschätzt.
Die Glättung führt zwar zu einer Verringerung der Parameteranzahl, aber das rechte Ende der angepaßten Lognormalverteilungen ist in den meisten Monaten zu kurz, vergl. Abb. 6.14, so daß eine eigene Modellierung der 1-Stunden-Extrema für jeden Monat zu bevorzugen ist.

Wir stellen also zusammenfassend fest, daß die Verteilungen sowohl der Jahresmaxima als auch die der Monatsmaxima der Monate $k = 1, \ldots, 12$ der 1-Stunden-

Abbildung 6.13: 1-Stunden-Maxima



QQ-Plots der sim. und der beob. 1-Stunden-Maxima sowie eingezeichnete Winkelhalbierende.



Schätzer der Parameter der an jeden Monat $k = 1, \ldots, 12$ angepaßten Lognormalverteilung: (a) Lageparameter $\hat{\mu}_k$ und Glättung durch Fouriersumme. (b) Formparameter $\hat{\mu}_k$.

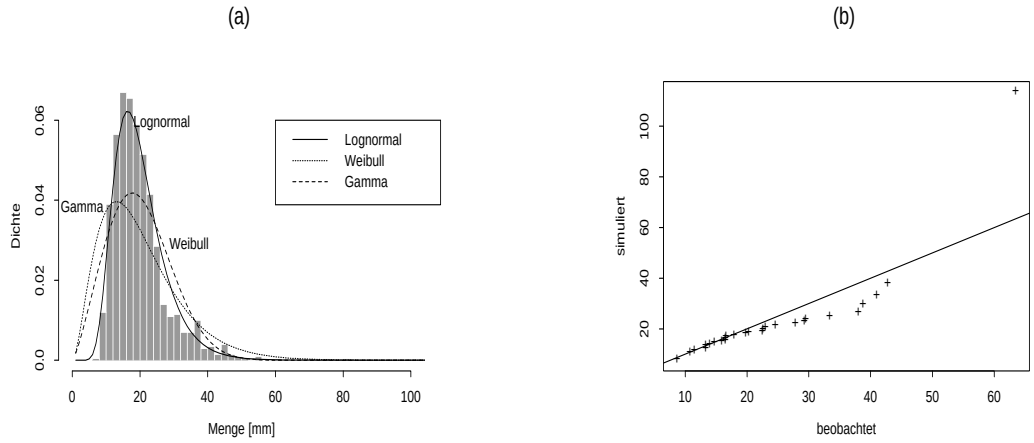
Niederschlagsmenge durch eine Lognormalverteilung beschrieben werden können, deren Parameter für das gesamte Jahr bzw. jeden Monat aus den Simulationen geschätzt werden.

1-Stunden-Maxima: Parameter der Lognormalverteilung

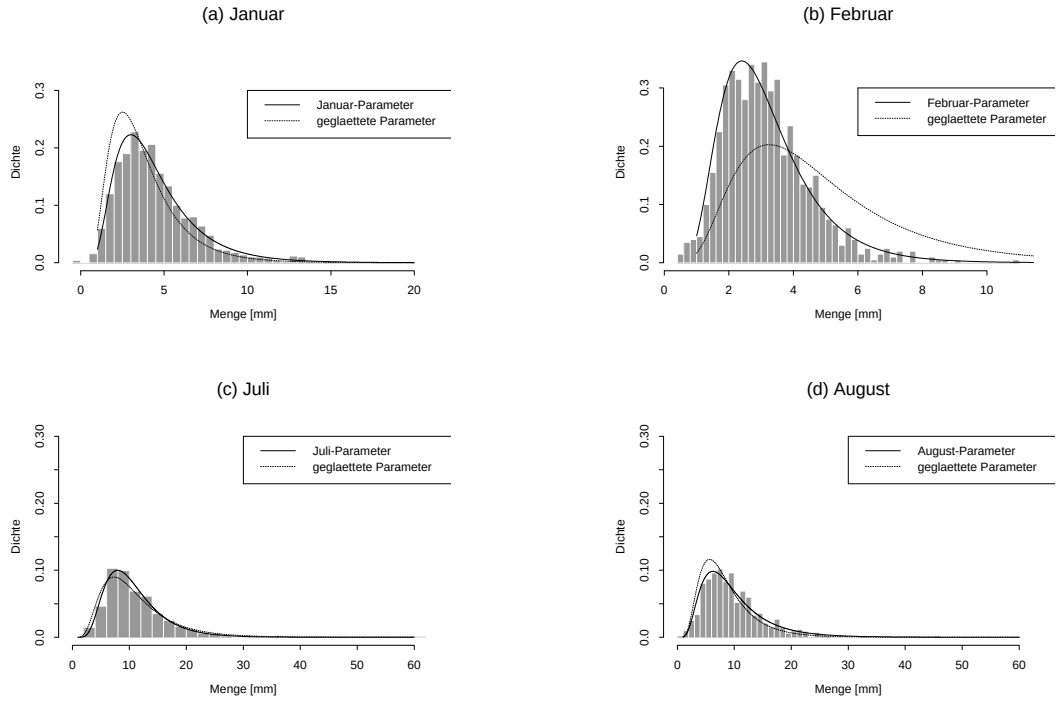
	Jahr	Jan.	Feb.	März	April	Mai	Juni	Juli	Aug.	Sep.	Okt.	Nov.	Dez.
$\hat{\mu}$	2.93	1.37	1.06	1.77	1.95	2.41	2.54	2.28	2.14	1.68	0.87	1.03	1.69
$\tilde{\mu}$		1.19	1.46	1.79	2.11	2.32	2.38	2.27	2.01	1.68	1.36	1.14	1.09
$\hat{\sigma}$	0.37	0.52	0.44	0.59	0.49	0.56	0.50	0.45	0.56	0.66	0.50	0.58	0.47
$\tilde{\sigma}$		0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53

Mittel und Standardabweichung der Log-Maxima.
1. Zeile: $\hat{\mu}_k$ für Monat k. 2. Zeile: mit Fouriersumme geglättete $\tilde{\mu}_k$.
3. Zeile: Formparameter $\hat{\sigma}_k$ für Monat k. 4. Zeile: $\tilde{\sigma} = \sum_{k=1}^{12} \hat{\sigma}_k / 12$.

Abbildung 6.14: 1-Stunden-Maxima



(a) Histogramm der 1-Stunden-Maxima aus 1000 sim. Jahren und angepaßte Verteilungen.
(b) QQ-Plot der sim. und der beob. 1-Stunden-Maxima sowie eingezeichnete Winkelhalbierende.



Histogramme der 1-Stunden-Maxima aus 1000 sim. Jahren und Lognormalverteilung:
—: Parameter $\hat{\mu}_k, \hat{\sigma}_k$ des Monats k. - - -: geglättete Parameter $\tilde{\mu}, \tilde{\sigma}$.
(a) Januar. (b) Februar. (c) Juli. (d) August.

6-Stunden-Maxima

Als 6-Stunden-Maxima $X_n^{(6)}, n = 1, \ldots, N$, bezeichnen wir die Maxima aller gleitenden, sich also überlagernden, 6-Stunden-Intervalle eines Jahres. Das sind die Intervalle 0:00 - 5:59h, 1:00 - 6:59h etc. Die 6-Stunden-Maxima in Monat $k, k = 1, \ldots, 12$ seien $X_{n,k}^{(6)}$.
Die 99%- und 98%-Quantile der beobachteten und der simulierten 6-Stunden-Jahresmaxima stimmen gut überein, ebenso die weiteren berechneten Quantile $Q(p)$, vergl. Tab. 6.7. Die Anpassung auf Basis monatlicher Maxima fällt für die einzelnen Monate unterschiedlich gut aus. Zum Teil bestehen erhebliche Abweichungen, beispielsweise im Januar und im September. In einigen Monaten sind die Quantile der simulierten Maxima größer als die der beobachteten, in anderen Monaten verhält es sich umgekehrt. Die Übereinstimmungen sind für das 90%-Quantil und kleinere Quantile sehr gut, so daß wir die Abweichung der extremen Quantile durch die unterschiedliche Anzahl an Daten erklären. Diese Aussage wird durch den QQ-Plot in Abbildung 6.15 verdeutlicht. Wir schließen also, daß die Verteilung der simulierten 6-Stunden-Extrema der wahren Verteilung entspricht und basieren wie oben weitere Analysen auf die mit dem Neyman–Scott–Modell simulierten Daten.

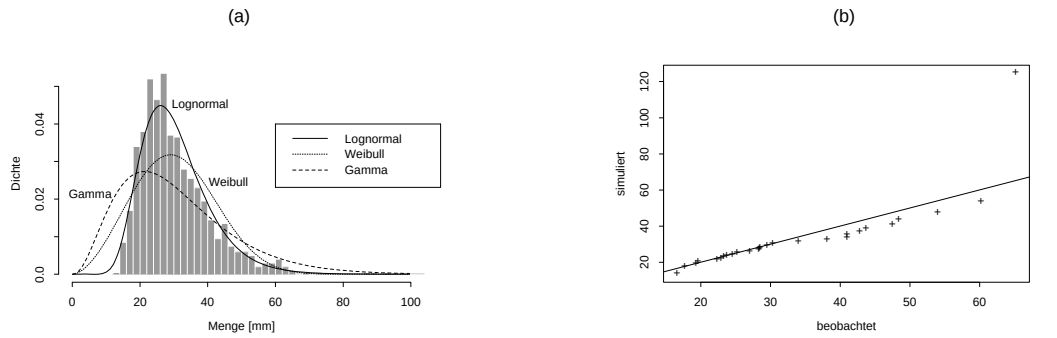
Tabelle 6.7: 6-Stunden-Maxima

Monat		Mittel	50%	75%	90%	95%	98%	99%
Jahr	beob.	33.44	28.50	41.90	50.64	58.34	62.60	63.90
	sim.	30.64	27.99	35.73	44.83	51.12	60.11	63.25
Jan.	beob.	7.93	6.00	8.50	10.46	13.56	33.41	43.70
	sim.	9.17	7.60	11.81	16.26	20.81	26.45	28.01
Feb.	beob.	7.66	5.80	7.55	12.80	16.88	27.81	32.86
	sim.	6.61	6.04	8.05	10.79	13.20	14.95	17.39
März	beob.	7.81	5.90	7.70	9.60	18.86	32.35	37.57
	sim.	9.43	8.06	11.44	16.36	19.96	26.56	30.64
Apr.	beob.	10.69	9.00	11.90	21.32	23.37	26.71	28.51
	sim.	11.91	10.91	14.44	18.45	22.47	27.22	30.81
Mai	beob.	18.02	11.40	17.40	42.08	46.36	56.00	60.60
	sim.	18.65	16.64	23.00	30.85	38.18	50.11	57.37
Juni	beob.	18.91	16.50	22.03	28.12	35.97	43.04	45.72
	sim.	18.87	16.43	22.44	31.46	39.12	47.85	54.55
Juli	beob.	16.01	13.20	19.15	28.84	30.13	32.03	32.92
	sim.	17.53	15.58	21.70	29.14	34.94	41.30	47.79
Aug.	beob.	15.81	13.20	17.00	25.10	28.15	43.61	51.91
	sim.	13.74	11.91	17.24	23.52	28.94	34.39	41.46
Sep.	beob.	10.84	9.90	12.20	18.38	21.52	22.97	23.34
	sim.	14.05	12.56	18.15	25.74	31.02	39.59	45.67
Okt.	beob.	9.33	9.40	13.40	18.42	21.81	22.94	23.02
	sim.	9.28	8.35	12.20	16.82	19.93	25.92	26.90
Nov.	beob.	10.08	8.40	11.90	17.64	24.50	29.84	31.92
	sim.	10.31	8.93	13.39	18.66	22.54	28.36	32.24
Dez.	beob.	10.86	8.20	14.55	25.26	26.82	33.72	37.36
	sim.	15.15	13.80	19.31	25.03	31.16	37.56	44.40

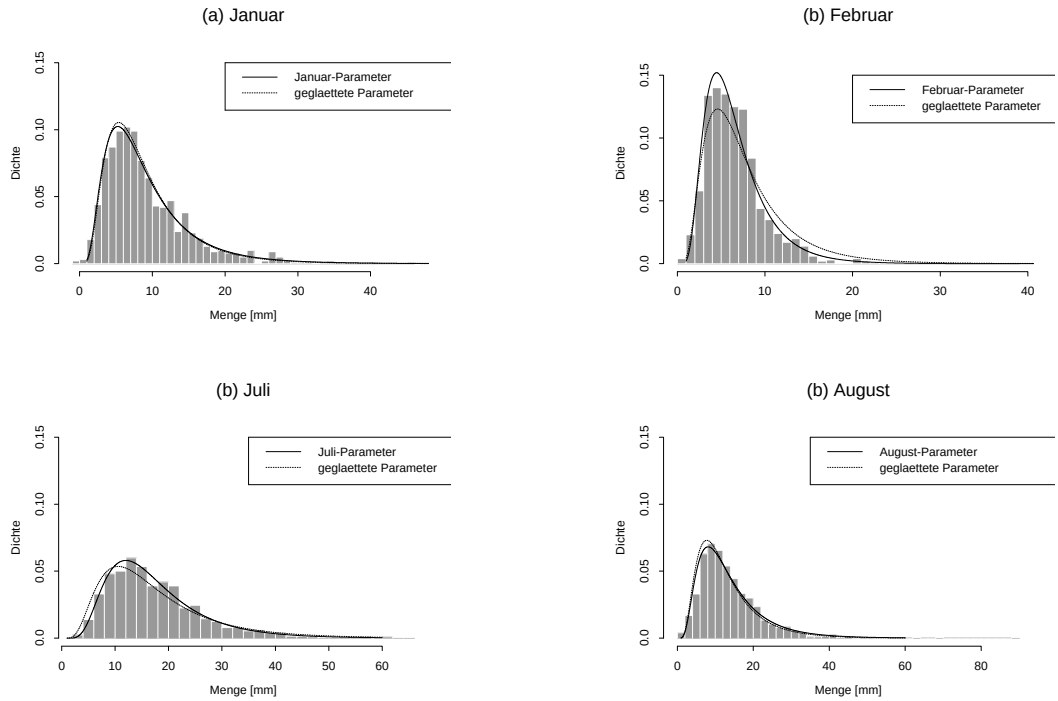
Mittelwert und Quantile der beob. und der sim. 6-Stunden-Maxima (gleitende Intervalle).

An die simulierten 6-Stunden-Maxima passen wir wie oben eine Lognormal-, ei-

Abbildung 6.15: 6-Stunden-Maxima



(a) Histogramm der 6-Stunden-Maxima der Simulationen und angepaßte Verteilungen.
(b) QQ-Plot der beob. und der sim. 6-Stunden-Maxima sowie eingezeichnete Winkelhalbierende.



Histogramme der 6-Stunden-Maxima aus 1000 sim. Jahren und Lognormalverteilung:
—: Parameter $\hat{\mu}_k, \hat{\sigma}_k$ des Monats k. - - -: geglättete Parameter $\tilde{\mu}, \tilde{\sigma}$.
(a) Januar. (b) Februar. (c) Juli, (d) August.

ne Gamma- und eine Weibullverteilung an, von denen die Lognormalverteilung am besten paßt, und zwar sowohl für die Maxima des ganzen Jahres wie auch für die der einzelnen Monate, vergl. Abb. 6.15. Die Veränderung der Lageparameter $\hat{\mu}_k$ im Jahresverlauf kann durch eine Fouriersumme mit zwei Frequenzen geglättet werden, vergl. Abb. 6.16, während der Formparameter für das gesamte Jahr als konstant betrachtet werden kann. Die geglätteten Werte seien wieder $\tilde{\mu}_k$ bzw. $\tilde{\sigma} := \sum_{k=1}^{12} \hat{\sigma}_k/12 = 0.59$,

$$\tilde{\mu}_k = \hat{R}_0 + \sum_{j=1}^{2} \hat{R}_j \cos(\omega_j k + \hat{\theta}_j) \ , \ \text{mit} \ \ \omega_j = 2\pi\, j/12 \ , \ \ k = 1, \ldots, 12 \ ,$$

j	$\hat{R}_j$	$\hat{\theta}_j$
0	2.35	
1	0.35	−2.81
2	0.24	−0.21

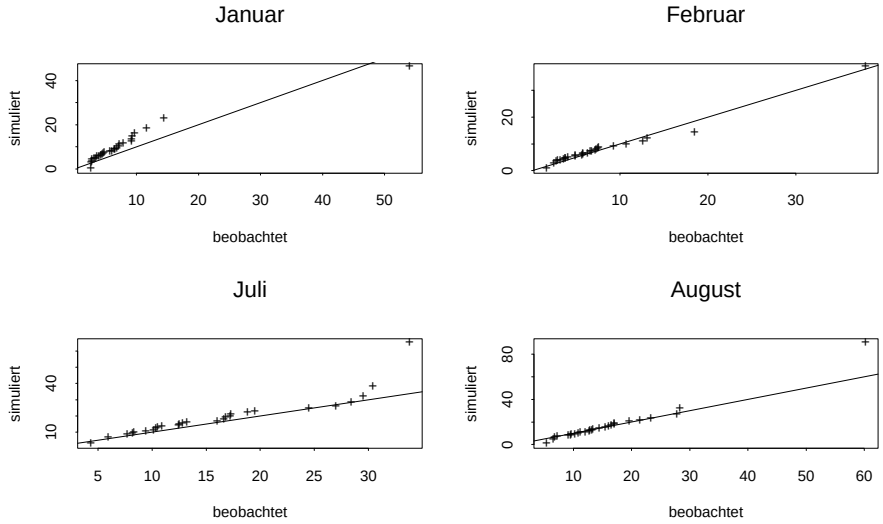
Da die 6-Stunden-Maxima in allen Monaten auch durch eine Lognormalverteilung mit den geglätteten Parametern angemessen beschrieben werden, führt die Modellierung der $\hat{\mu}_k$ zu einer Reduktion der Parameteranzahl. Aussagen über die Verteilung der 6-Stunden-Maxima können also auf eine Lognormalverteilung basiert werden, deren Parameter aus künstlichen, mit dem Neyman-Scott-Modell erzeugten Daten geschätzt werden.

6-Stunden-Maxima: Parameter der Lognormalverteilung

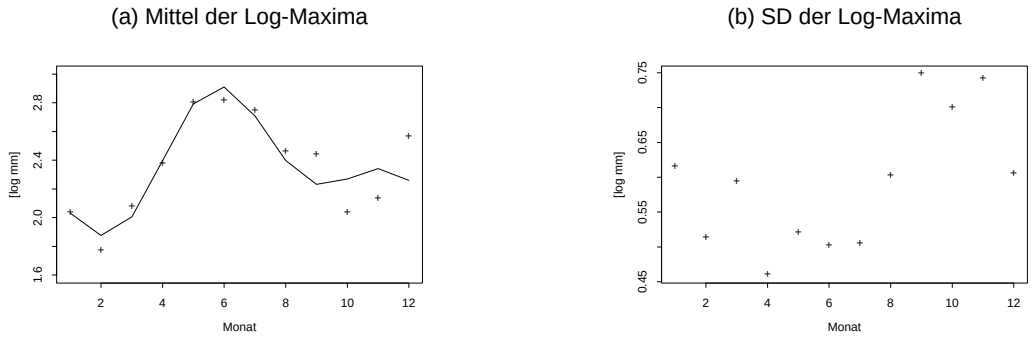
	Jahr	Jan.	Feb.	März	April	Mai	Juni	Juli	Aug.	Sep.	Okt.	Nov.	Dez.
$\hat{\mu}$	3.37	2.04	1.76	2.08	2.38	2.80	2.81	2.74	2.46	2.44	2.03	2.13	2.56
$\tilde{\mu}$		2.03	1.88	2.01	2.40	2.79	2.91	2.71	2.40	2.23	2.27	2.34	2.26
$\hat{\sigma}$	0.32	0.62	0.51	0.59	0.46	0.52	0.50	0.50	0.60	0.75	0.70	0.74	0.60
$\tilde{\sigma}$		0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59

Mittel und Standardabweichung der Log-Maxima.
1. Zeile: $\hat{\mu}_k$ für Monat k. 2. Zeile: mit Fouriersumme geglättete $\tilde{\mu}_k$.
3. Zeile: Formparameter $\hat{\sigma}_k$ für Monat k. 4. Zeile: $\tilde{\sigma} = \sum_{k=1}^{12} \hat{\sigma}_k/12$.

Abbildung 6.16: 6-Stunden-Maxima



QQ-Plots der beob. und der sim. 6-Stunden-Maxima sowie eingezeichnete Winkelhalbierende.



Schätzer der Parameter der an jeden Monat $k = 1, \ldots, 12$ angepaßten Lognormalverteilung: (a) Lageparameter $\hat{\mu}_k$ und Glättung durch Fouriersumme. (b) Formparameter $\hat{\sigma}_k$.

24-Stunden-Maxima

Die Serie der Maxima $X_n^{(24)}, n = 1, \ldots, N$, der 24-Stunden-Niederschläge in Jahr n wird als 24-Stunden-Maxima bezeichnet. Wir betrachten die Maxima aus sich nicht überlappenden 24-Stunden-Intervallen von 0:00h bis 23:59h. Die simulierten 24-Stunden-Maxima seien $\tilde{X}_n^{(24)}$.

Tabelle 6.8: 24-Stunden-Maxima

Monat		Mittel	50%	75%	90%	95%	98%	99%
Jahr	beob.	36.48	31.10	43.55	56.16	62.81	67.10	69.05
	sim.	35.64	32.62	41.20	51.15	57.98	67.02	74.42
Jan.	beob.	10.57	7.60	10.30	15.00	16.95	43.13	57.06
	sim.	10.47	8.86	13.20	18.81	23.50	28.25	34.53
Feb.	beob.	10.57	7.70	12.35	17.28	29.92	38.45	40.38
	sim.	7.88	7.23	9.79	13.05	14.98	17.75	18.68
März	beob.	10.03	8.10	9.80	14.06	25.75	38.37	42.68
	sim.	10.82	9.43	13.55	18.79	21.83	29.95	35.93
Apr.	beob.	14.39	11.40	22.30	29.28	30.03	30.83	31.11
	sim.	15.21	13.91	18.53	23.71	27.57	32.60	37.48
Mai	beob.	20.69	13.70	25.30	42.24	50.19	57.94	60.72
	sim.	23.25	21.17	28.99	38.75	45.47	57.15	62.61
Juni	beob.	21.77	18.60	26.10	40.64	45.67	48.09	48.24
	sim.	21.49	18.81	25.56	35.17	44.23	53.99	61.56
Juli	beob.	18.16	16.60	24.20	32.58	33.56	34.86	35.43
	sim.	19.59	17.94	24.29	31.84	38.08	45.87	50.66
Aug.	beob.	18.76	15.20	21.70	28.34	36.68	50.23	55.71
	sim.	15.40	13.79	19.22	26.81	31.34	38.71	43.26
Sep.	beob.	14.32	13.30	18.95	23.18	26.86	28.48	28.74
	sim.	16.28	14.49	21.23	29.02	35.63	45.42	50.90
Okt.	beob.	12.20	10.60	16.95	23.26	25.70	28.45	29.77
	sim.	10.79	9.45	13.95	20.67	24.74	29.90	33.73
Nov.	beob.	13.69	10.10	18.95	24.54	32.64	36.47	36.88
	sim.	11.86	10.23	15.27	21.91	26.59	35.00	39.72
Dez.	beob.	12.35	9.40	15.65	24.12	28.50	35.93	39.46
	sim.	17.29	15.82	21.64	28.61	35.23	43.14	52.90

Mittelwert und Quantile der beob. und der sim. 24-Stunden-Maxima für einzelne Monate und das gesamte Jahr.

Auf Jahresbasis stimmen die 99%- und 98%-Quantile gut überein, vergl. Tab. 6.8. Die 95%- und 90%-Quantile der beobachteten 24-Stunden-Maxima sind etwas größer als die der simulierten, vergl. Tab. 6.8. Bei Berechnung der Quantile auf Basis von 100 simulierten Jahren ergeben sich engere Übereinstimmungen, vergl. Tab. 6.9. Wir folgern daraus, daß Differenzen zwischen den Quantilen der beobachteten und der simulierten 24-Stunden-Maxima vor allem auf der unterschiedlichen Anzahl an beobachteten Jahren basieren. Schlüsse über die tatsächliche Verteilung der 24-Stunden-Extrema können also auf das Neyman-Scott-Modell gegründet werden. Für einige Monate ergeben sich aber größere Differenzen zwischen den beobachteten und den simulierten Quantilen. Aussagen über die Verteilung der 24-Stunden-Extrema auf monatlicher Basis sind also weniger verläßlich.

Tabelle 6.9: 24-Stunden-Maxima pro Set

Monat	Mittel	50%	75%	90%	95%	98%	99%
Set 1	36.18	33.25	40.84	47.39	54.00	73.43	99.32
Set 2	36.20	32.67	40.66	53.06	59.17	72.16	75.62
Set 3	36.53	33.91	42.05	56.89	61.52	66.18	66.88
Set 4	36.49	34.85	42.21	48.91	55.37	66.35	68.64
Set 5	37.00	34.14	42.02	52.22	56.49	63.85	87.80
Set 6	36.69	33.03	42.63	55.04	62.26	65.86	79.23
Set 7	33.95	32.19	39.21	45.85	52.96	56.06	66.84
Set 8	33.08	30.45	37.79	47.98	51.17	59.87	61.18
Set 9	34.32	31.04	39.04	50.24	62.19	65.99	68.53
Set 10	35.99	33.13	41.72	55.04	61.38	71.99	73.75
Mittel	35.64	32.87	40.82	51.26	57.65	66.17	74.78
Beob.	36.48	31.10	43.55	56.16	62.81	67.10	69.05

Zeilen 1 - 10: Mittel und Quantile der sim. 24-Stunden-Maxima aus 10 Sets á 100 Jahren.
Zeile 11: Spaltenmittel der Zeilen 1 - 10.
Zeile 12: Mittel und Quantile der beob. 24-Stunden-Maxima aus 27 Jahren.

Die Verteilungen der simulierten 24-Stunden-Extrema auf jährlicher sowie auf monatlicher Basis können durch eine Lognormalverteilung approximiert werden, vergl. Abb. 6.17. Wird an die 24-Stunden-Maxima $\tilde{X}_{n,k}$ der Monate $k = 1, \ldots, 12$ für jeden Monat separat eine Lognormalverteilung angepaßt, so können die Schätzer $\hat{\mu}_k$ der Lageparameter für die einzelnen Monate durch eine Fouriersumme mit zwei Fourierfrequenzen $\omega_j = 2\pi\, j/12$ geglättet werden,

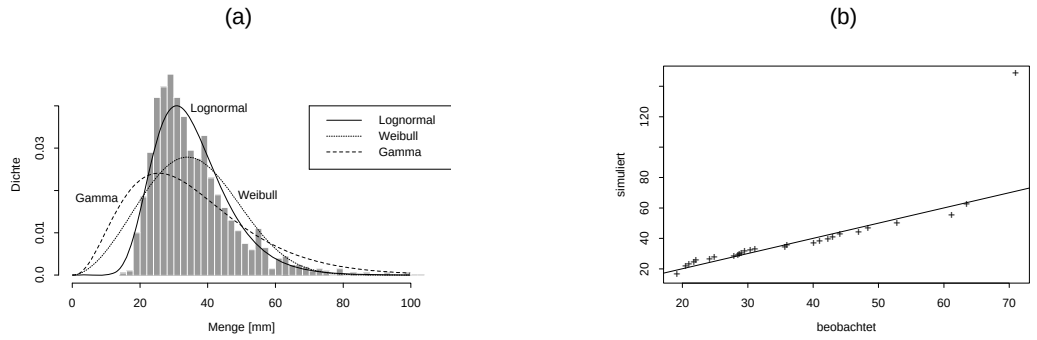
$$\tilde{\mu}_k = \hat{R}_0 + \sum_{j=1}^{2} \hat{R}_j \cos(\omega_j k + \hat{\theta}_j) \;\text{ mit }\; \omega_j = 2\pi\, j/12 \;\;,\; k = 1, \ldots, 12 \;,$$

j	$\hat{R}_j$	$\hat{\theta}_j$
0	2.50	
1	0.35	−2.92
2	0.23	−0.32

Der Formparameter σ kann als konstant im Jahresverlauf betrachet werden und wird durch $\tilde{\sigma}$, das Mittel der geschätzten Formparameter $\tilde{\sigma}_k$ der einzelnen Monate geschätzt, $\tilde{\sigma} := \sum_{k=1}^{12} \hat{\sigma}_k/12 = 0.60$.

Diese Glättung bewirkt in den Monaten k mit $\tilde{\mu}_k < \hat{\mu}_k$, das sind insbesondere September und Dezember, daß das rechte Ende der Verteilung zu kurz ist. In den übrigen Monaten führt die Glättung zu keiner Verschlechterung, vergl. Abb. 6.19 und die Tabelle auf Seite 157. Die Glättung der Parameterschätzer der Lognormalverteilung führt zu einer Verringerung der Parameteranzahl um ein Drittel von 24 zu 6 Parametern und ist gerechtfertigt, weil die Qualität der Anpassung in zehn Monaten sich nur geringfügig verschlechtert.

Abbildung 6.17: 24-Stunden-Maxima

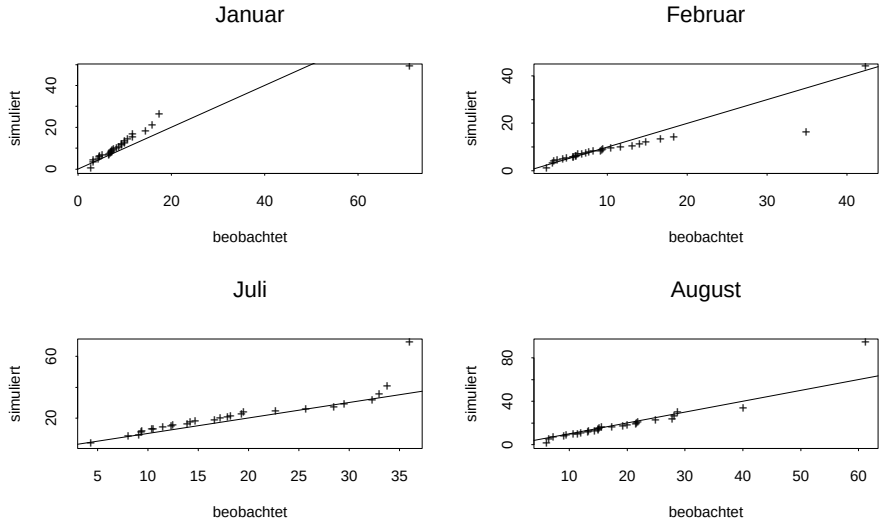


(a) Histogramm der 24-Stunden-Maxima der Simulationen und angepaßte Verteilungen.
(b) QQ-Plot der 24-Stunden-Maxima der Simulationen und der Beobachtungen sowie eingezeichnete Winkelhalbierende.

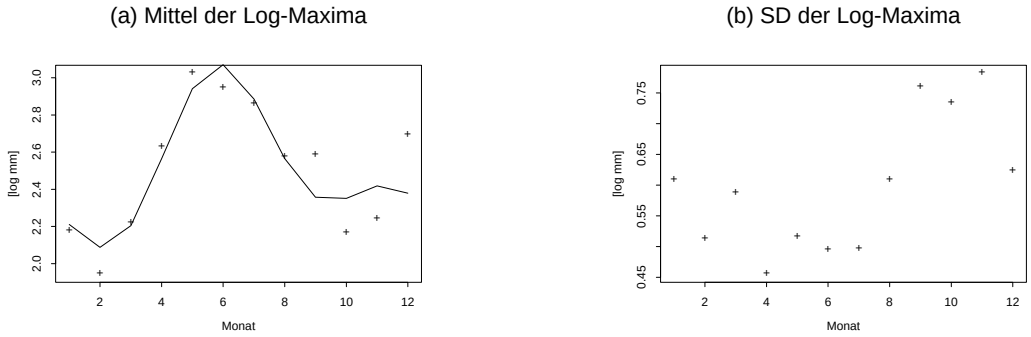
Zusammenfassung

Die Verteilungen der 1-, 6- und 24-Stunden-Maxima werden unter dem Neyman–Scott–Modell unterschiedlich gut wiedergegeben.
Die Verteilung der 1-Stunden-Extrema wird unter dem Modell im rechten Bereich nicht korrekt angepaßt, weil die extremen Quantile wie $Q(0.99)$ und $Q(0.98)$ unter dem Modell zu klein sind. Das kann nicht nur auf die unterschiedliche Anzahl von beobachteten Jahren zurückgeführt werden, da das Modell genügend große 1-Stunden-Maxima zwar durchaus erzeugen kann, sie aber zu selten auftreten, was bei der Analyse der einzelnen, aus jeweils 100 Jahren bestehenden, simulierten Datensets deutlich wird.
Die 6- und die 24-Stunden-Extrema werden unter dem Modell befriedigend wiedergegeben, so daß verläßliche Aussagen über deren Verteilungen anhand von simulierten Daten getroffen werden können. Die Verteilung der simulierten Jahresmaxima kann in beiden Fällen durch eine Lognormalverteilung approximiert werden. Die Verteilungen der monatlichen 6- und 24-Stunden-Extrema können ebenfalls durch Lognormalverteilungen beschrieben werden, deren Skalenparameter sich im Jahresverlauf gemäß einer Fouriersumme verändern. Aussagen auf monatlicher Basis sind allerdings weniger verläßlich als solche auf Jahresbasis, wie oben bereits beschrieben.
Formale Anpassungstests wie der χ^2– oder der Kolmogorov-Smirnov-Test der Hypothese, die h–Stunden–Jahresmaxima seien lognormalverteilt, werden für alle betrachteten Werte von $h = 1, 6, 24$ verworfen. Das kann zwar mit der großen Anzahl von $N = 1000$ Beobachtungen begründet werden, ist aber auch ein Indiz dafür, daß die Lognormalverteilung die simulierten Werte in beiden Randbereichen nicht exakt beschreibt. Für viele Anwendungen der Extremwertanalyse sind aber gerade die großen Quantile $Q(p), p > 0.95$, relevant. Weil die Quantile $Q_e(p)$ der Beobachtungen größer sind als die der simulierten Daten, erscheint die Aussagekraft des Modelles in bezug auf Extremwertanalysen ohnehin eingeschränkt zu sein.
Als alternativer Ansatz könnte die Verteilung der h-Stunden-Niederschlagsmenge betrachtet werden, die einen bestimmten, hohen Schwellenwert y überschreiten, $Y^{(h)} > y$. In der Literatur finden sich dafür Ansätze bei Zucchini und Adamson

Abbildung 6.18: 24-Stunden-Maxima



QQ-Plots der beob. und der sim. 24-Stunden-Maxima sowie eingezeichnete Winkelhalbierende.



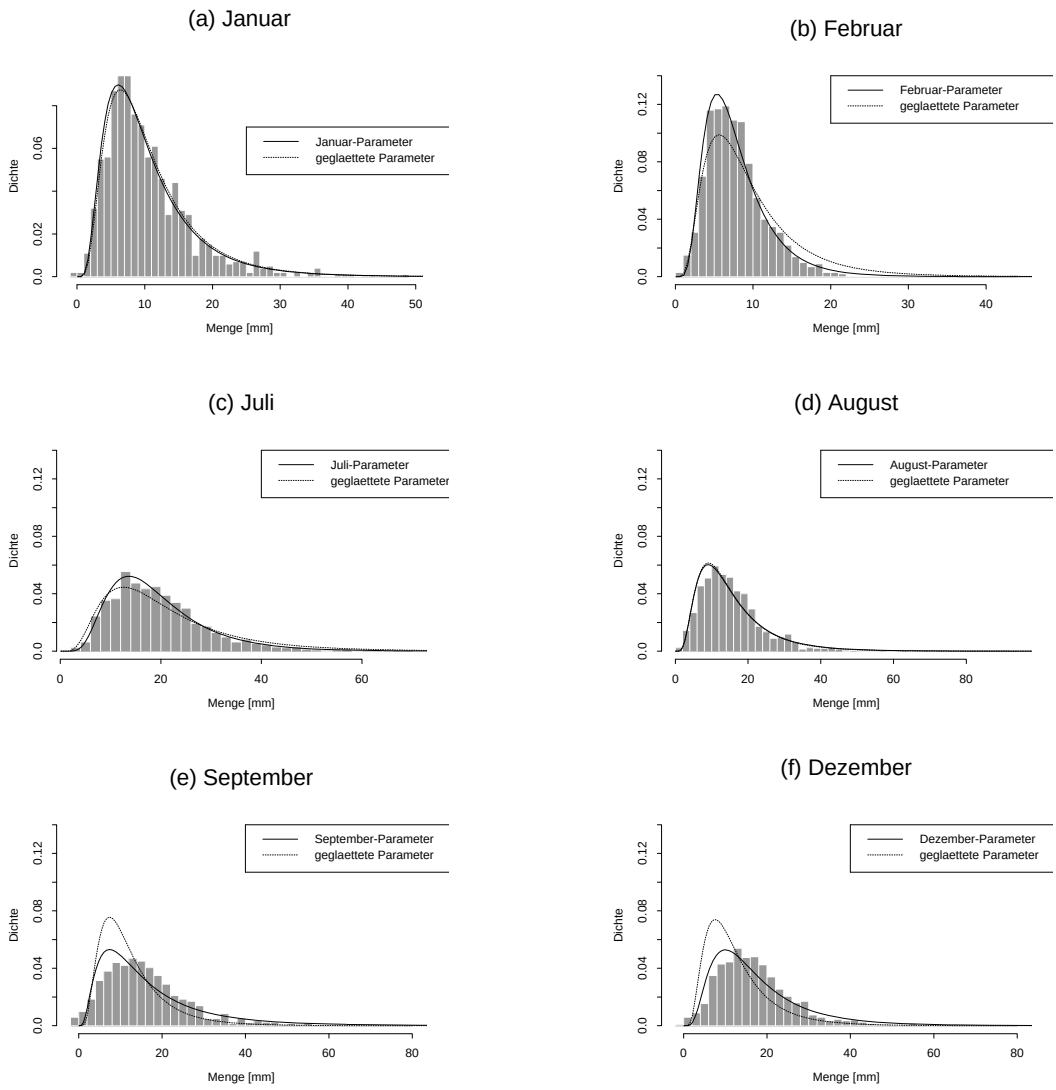
Schätzer der Parameter der an jeden Monat $k = 1, \ldots, 12$ angepaßten Lognormalverteilung:
(a) Lageparameter $\hat{\mu}_k$ und Glättung durch Fouriersumme. (b) Formparameter $\hat{\sigma}_k$.

24-Stunden-Maxima: Parameter der Lognormalverteilung

	Jahr	Jan.	Feb.	März	April	Mai	Juni	Juli	Aug.	Sep.	Okt.	Nov.	Dez.
$\hat{\mu}$	3.52	2.17	1.94	2.22	2.63	3.02	2.95	2.86	2.57	2.58	2.16	2.24	2.69
$\tilde{\mu}$		2.21	2.09	2.20	2.56	2.94	3.07	2.88	2.56	2.36	2.35	2.42	2.38
$\hat{\sigma}$	0.31	0.61	0.51	0.59	0.45	0.52	0.49	0.50	0.61	0.76	0.73	0.78	0.62
$\tilde{\sigma}$		0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59

Mittel und Standardabweichung der Log-Maxima.
1. Zeile: $\hat{\mu}_k$ für Monat k. 2. Zeile: mit Fouriersumme geglättete $\tilde{\mu}_k$.
3. Zeile: Formparameter $\hat{\sigma}_k$ für Monat k. 4. Zeile: $\tilde{\sigma} = \sum_{k=1}^{12} \hat{\sigma}_k / 12$.

Abbildung 6.19: 24-Stunden-Maxima: Ausgewählte Monate



Histogramme der 24-Stunden-Maxima aus 1000 sim. Jahren und Lognormalverteilung:
—: Parameter $\hat{\mu}_k, \hat{\sigma}_k$ des Monats k. - - -: geglättete Parameter $\tilde{\mu}, \tilde{\sigma}$
(a) Januar. (b) Februar. (c) Juli. (d) August. (e) September. (f) Dezember.

[1989] oder bei SMITH, TAWN UND COLES [1990]. Auch eine Kombination beider Ansätze, nämlich die Modellierung der jährlichen h-Stunden-Extrema, die einen bestimmten Schwellenwert überschreiten, könnte ein erfolgversprechender Ansatz sein, der in weiteren Arbeiten zu untersuchen wäre.

In Risikoabschätzungen des Hoch- und Tiefbaus wird manchmal nicht nach der Wahrscheinlichkeit gefragt, daß ein bestimmter Schwellenwert einmal im Jahr überschritten wird, sondern es werden Aussagen über Return-Perioden verlangt. Die Return-Perioden lassen sich leicht aus der Verteilung der jährlichen Maxima berechnen: das 99%-Quantil $Q(0.99)$ hat eine Return-Periode von 100 Jahren, das 98%-Quantil $Q(0.98)$ eine Return-Periode von 50 Jahren usw. Wegen der Äquivalenz der Sachverhalte haben wir Aussagen über die Extremwerte nur in Termen ihrer Verteilungen bzw. Quantile getroffen.

4 Verteilung der Niederschlagsmengen in nassen Stunden

Ein Modell zur Beschreibung stündlichen Niederschlages sollte auch die Verteilung der Niederschlagsmengen in nassen Stunden korrekt wiedergeben. Wir betrachten deshalb die bedingte Verteilung der aggregierten Niederschlagsmengen, gegeben, ein Zeitintervall ist naß. Dieses Vorgehen ist notwendig, da sonst die trockenen Stunden dominierten und damit die meiste Wahrscheinlichkeitsmasse auf der Null läge.

Da sich die Verteilung der Niederschlagsmengen saisonal ändert, wird jeder Monat separat analysiert. Wir werden exemplarisch die Monate Januar und Juli betrachten. Die Verteilung der Niederschlagsmengen in nassen Stunden ist rechtsschief mit einem sehr langen rechten Ende, vergl. Tab. 6.10. Die Verteilung der simulierten Mengen ist breiter als die der beobachteten, vergl. Abb. 6.20 und 6.21 (a). Im Januar stimmen zwar die 99%-Quantile überein, nicht aber die Quantile $Q(p)$ für $p < 0.99$. Die Anpassung im Juli ist schlechter. Das Modell kann offensichtlich die bedingte Verteilung der aggregierten Niederschlagsmengen nicht korrekt reproduzieren. Wir modellieren deshalb die bedingte Verteilung der stündlichen Niederschlagsmengen in nassen Stunden direkt anhand der Beobachtungen.

Es bezeichne in diesem Abschnitt $W > 0$ die Niederschlagsmenge nasser Stunden während einer homogenen Periode, z.B. einem Monat. Natürliche Kandidaten zur Beschreibung von Y durch eine Verteilung sind die Lognormal-, die Weibull- und die Gammaverteilung, deren Parameter jeweils mit der Maximum–Likelihood–Methode geschätzt werden. QQ-Plots der Beobachtungen gegen die Quantile der angepaßten Verteilungen zeigen, daß keine der Verteilungen das rechte Ende der empirischen Verteilung angemessen beschreibt. Ein Ausweg besteht darin, nur den Bereich bis zum 98%- oder 99%-Punkt durch eine Verteilung zu beschreiben und größere Werte gesondert zu betrachten. Eine nur auf den Beobachtungen eines Monates basierende Analyse ist dann jedoch wegen der geringen Anzahl an Beobachtungen schwierig. In allen Monaten beschreibt unter den betrachteten Verteilungen (Lognormal-, Weibull-, Gammaverteilung) die Lognormalverteilung den Bereich der Mengen $W < Q(0.9)$ am besten. Für die übrigen Monate gelten entsprechende Ergebnisse.

Eine weitere Möglichkeit ist, die Niederschlagsmengen W vor dem Anpassen einer Verteilung zu transformieren, beispielsweise

$$\phi: \quad W \longmapsto \tilde{W} = \log(10 \cdot W) \; .$$

Da die minimale Niederschlagsmenge 0.1mm beträgt, gilt $\tilde{W} \geq 0$. In 20% der nassen Stunden wird eine Niederschlagsmenge von 0.1mm registriert, so daß keine der für strikt positive Werte definierten Verteilungen die Verteilung der $\tilde{W}$ angemessen beschreiben kann. Das Problem bleibt bestehen, auch wenn ϕ mit einer Translation verknüpft wird, weil die am weitesten links liegende Klasse stets mehr Masse besitzt als unter einer an die $\tilde{W}$ angepaßten (Weibull-, Gamma- oder Lognormal-)

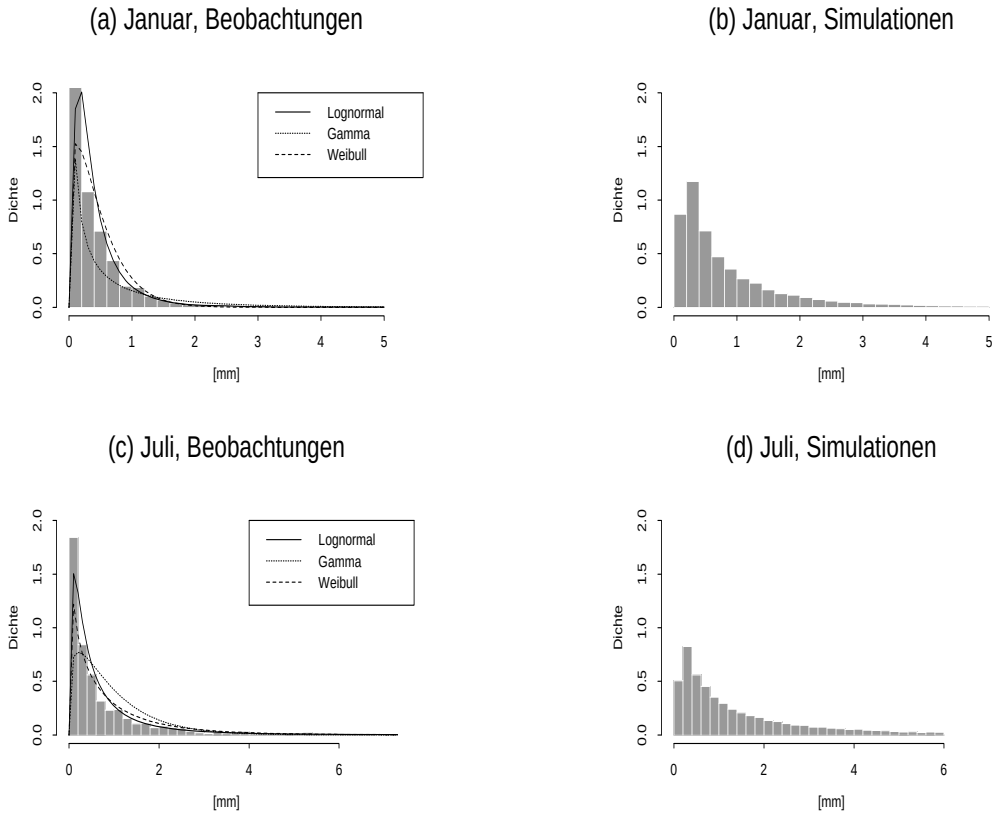
Verteilung. Da die kleinen Niederschlagsmengen auch ohne Transformation geglättet werden können, haben wir hier nur die $\tilde{W} > 0$ betrachtet, was $W > 0.1$ entspricht. Sie können durch eine Weibullverteilung beschrieben werden, vergl. Abb. 6.22. Dieser Ansatz bleibt unbefriedigend, da ein Fünftel der Beobachtungen nicht berücksichtigt wird.

Die besten Resultate in der Modellierung der Niederschlagsmengen W in nassen Stunden werden also erzielt, wenn der Bereich bis zum 99%-Quantil durch eine Lognormalverteilung beschrieben wird und extreme Werte gesondert modelliert werden, zum Beispiel durch eine Extremwertverteilung. In der Literatur finden sich dafür auch zahlreiche neuere Ansätze.

Tabelle 6.10: Quantile der Niederschlages in nassen Stunden

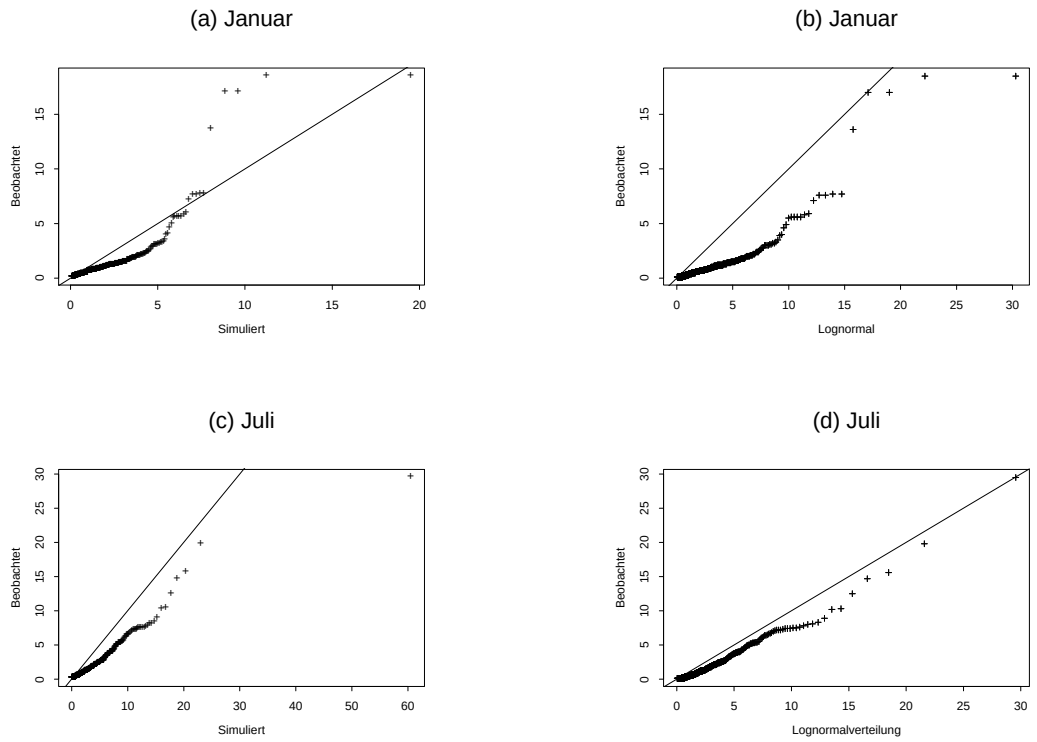
		$Q(0.25)$	$Q(0.5)$	$Q(0.75)$	$Q(0.8)$	$Q(0.9)$	$Q(0.95)$	$Q(0.99)$
Jan.	beob.	0.2	0.3	0.6	0.7	1.1	1.5	5.6
	sim.	0.3	0.5	1.2	1.4	2.2	3.2	5.9
Juli	beob.	0.2	0.4	1.1	1.3	2.4	3.9	7.5
	sim.	0.4	1.0	2.4	2.9	4.7	9.9	12.6

Abbildung 6.20: Niederschlagsmengen in nassen Stunden



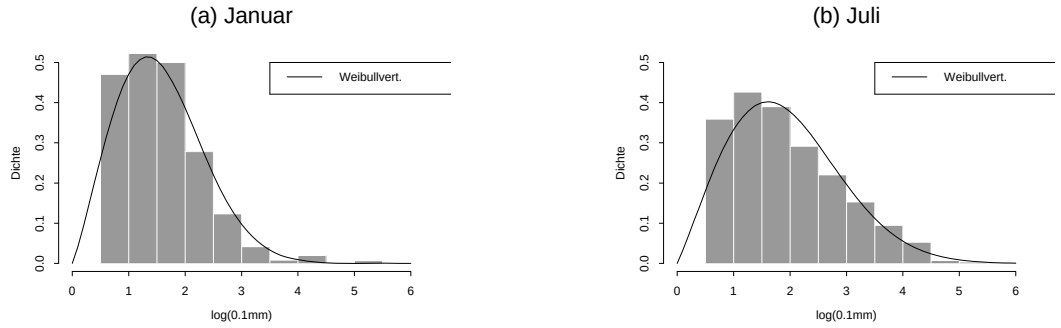
Beobachtete Niederschlagsmengen in nassen Stunden, (a), (b): Januar, (c),(d): Juli.

Abbildung 6.21: Niederschlagsmengen in nassen Stunden



Beobachtete Niederschlagsmengen in nassen Stunden.
(a), (c): QQ-Plot der beobachteten und der simulierten Werte.
(b), (d): QQ-Plot gegen Lognormal-Verteilung.
(a), (b): Januar, (c), (d): Juli.

Abbildung 6.22: Niederschlagsmengen in nassen Stunden



Transformierte beobachtete Niederschlagsmengen $\tilde{W} = \log(10 \cdot W)$ in nassen Stunden und Weibullverteilung. Das Histogramm zeigt nur Mengen $\tilde{W}$ mit $W > 0.1$mm/Std. (a) Januar, (b) Juli.

5 Verteilung der Runlängen

Ein weiterer Aspekt der Modellüberprüfung ist die zeitliche Verteilung der Niederschläge. Wir betrachten dazu die Verteilungen trockener und nasser Runs sowie die Anzahl nasser Stunden an nassen Tagen.

Für einige Anwendungen, beispielsweise im Straßenbau oder in der Landwirtschaft, ist es wichtig, die Dauer einer trockenen oder einer nassen Periode abschätzen zu können. In dem gemäßigten Klima Mitteleuropas ist die Wahrscheinlichkeit langer trockener Perioden von beispielsweise mehr als 30 Tagen, die eine ernstzunehmende Gefährdung des Pflanzenwachstums und damit der Ernteerträge darstellen, sehr gering. Stattdessen sind Fragen wie diese relevant: „Wie groß ist die Wahrscheinlichkeit, daß maximal i nasse Stunden folgen, wenn es jetzt (in dieser Stunde) zu regnen beginnt?", „Wie groß ist die Wahrscheinlichkeit, daß es mindestens weitere i Stunden trocken bleibt, gegeben, jetzt ist die erste trockene Stunde nach einem Regenschauer?".

Ein trockener Run der Länge i ist eine Sequenz von exakt i aufeinanderfolgenden trockenen Stunden. Ein nasser Run der Länge i ist analog dazu definiert. Wir sprechen auch kurz von trockenen und nassen i-Runs. Zur Untersuchung eines Phänomenes, das sich über einen längeren Zeitraum erstreckt, ist die Annahme hilfreich, der Niederschlagsprozeß sei in dem betrachteten Zeitraum stationär. Es bezeichne weiterhin Z_t den Zustand des Ereignisprozesses in Intervall t. Damit können wir definieren:

(5.1) Definition. Ein trockener Run der Länge i ist durch die Sequenz

$$Z_{t-1} = 1,\ Z_t = 0,\ \ldots,\ Z_{t+i-1} = 0,\ Z_{t+i} = 1$$

gegeben. Ein nasser Run der Länge i ist entsprechend als

$$Z_{t-1} = 0,\ Z_t = 1,\ \ldots,\ Z_{t+i-1} = 1,\ Z_{t+i} = 0$$

definiert. □

(5.2) Definition. Die bedingte Wahrscheinlichkeit eines nassen i-Runs bei einer gegebenen ersten nassen Stunde ist

$$P^{(W)}(i) := \mathrm{P}(Z_{t+1} = 1, \ldots, Z_{t+i-1} = 1, Z_{t+i} = 0 | Z_t = 1, Z_{t-1} = 0)\ .$$

Aus der angenommenen Zeitinvarianz folgt

$$P^{(W)}(i) = \mathrm{P}(Z_2 = 1, \ldots, Z_{i-1} = 1, Z_i = 0 | Z_1 = 1, Z_0 = 0)\ .$$

Analog ist die bedingte Wahrscheinlichkeit eines trockenen i-Runs bei einer gegebenen ersten trockenen Stunde

$$P^{(D)}(i) := \mathrm{P}(Z_2 = 0, \ldots, Z_{i-1} = 0, Z_i = 1 | Z_1 = 0, Z_0 = 1) \; .$$

□

Wir verwenden als Schätzer dieser Wahrscheinlichkeiten

$$\textbf{(5.3)} \qquad \hat{\mathrm{P}}^{(W)}(i) = \frac{\#\{\text{nasse } i\text{-Runs}\}}{\#\{\text{alle nassen Runs}\}} \; ,$$

$$\textbf{(5.4)} \qquad \hat{\mathrm{P}}^{(D)}(i) = \frac{\#\{\text{trockene } i\text{-Runs}\}}{\#\{\text{alle trockenen Runs}\}} \; .$$

Aus den Wahrscheinlichkeiten (5.2) lassen sich weitere Wahrscheinlichkeiten berechnen:

(5.5) Definition. Die bedingte Wahrscheinlichkeit $\hat{\mathrm{P}}^{(D)}(k+)$, einen trockenen Run der Länge $i \geq k$ zu erhalten, beträgt, wenn in der gegebenen Stunde ein trockener Run beginnt:

$$\mathrm{P}^{(D)}(k+) := \mathrm{P}(Z_2 = 0, \ldots, Z_{k-1} = 0 | Z_1 = 0, Z_0 = 1) = \sum_{i \geq k} \mathrm{P}^{(D)}(i) \; .$$

Analog ist die bedingte Wahrscheinlichkeit, einen nassen Run der Länge $i \leq k$ zu erhalten, wenn in der gegebenen Stunde ein nasser Run beginnt,

$$\begin{aligned} \mathrm{P}^{(W)}(k-) &:= \mathrm{P}\Big(\max_j \{Z_2 = 1, \ldots, Z_{j-1} = 1, Z_j = 0 | Z_1 = 1, Z_0 = 0\} = k \Big) \\ &= \sum_{i \leq k} \mathrm{P}^{(W)}(i) \; . \end{aligned}$$

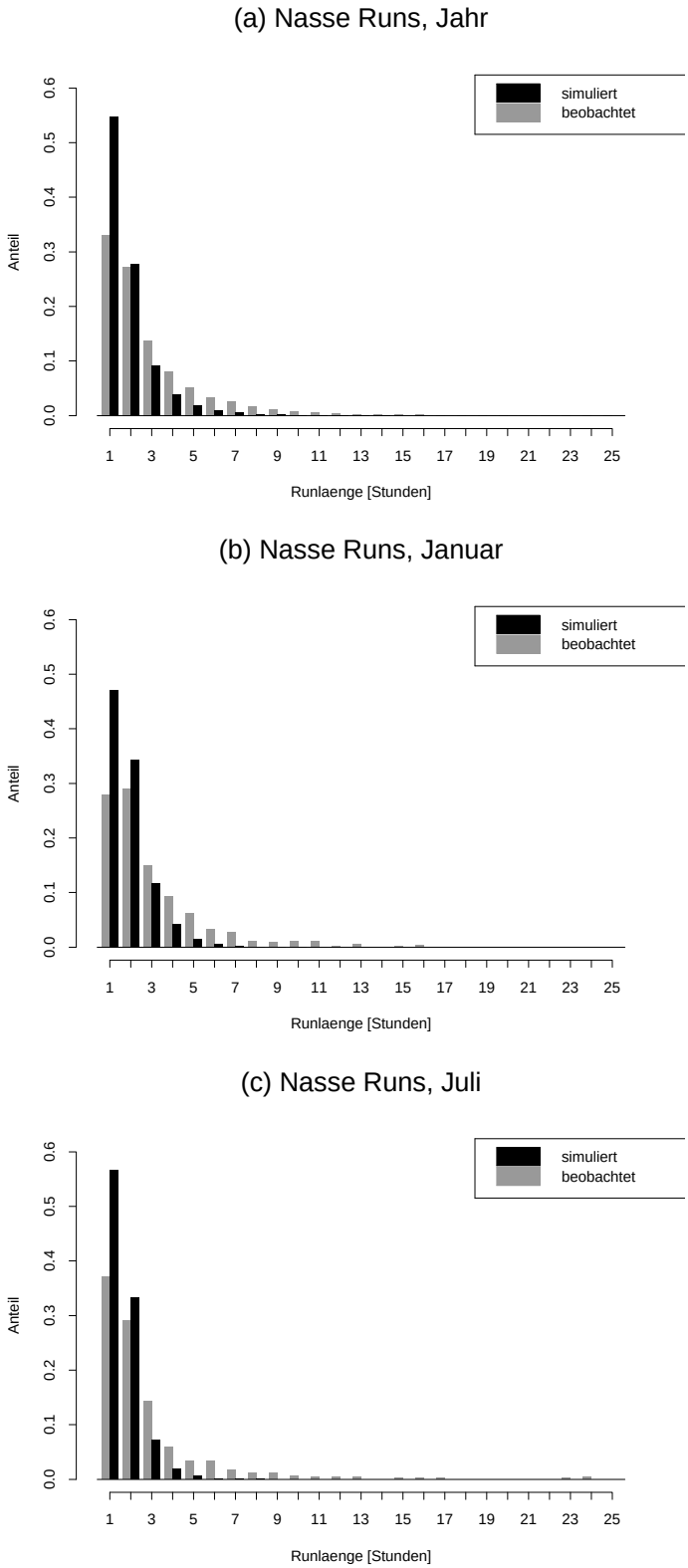
□

Die Wahrscheinlichkeit $\mathrm{P}^{(W)}(k-)$ ist nicht die bedingte Wahrscheinlichkeit, daß Stunde k trocken ist (Stunde k könnte wieder naß sein), sondern die bedingte Wahrscheinlichkeit, daß bis inklusive der Stunde k mindestens eine trockene Stunde aufgetreten ist, bei gegebener Anfangsstunde eines nassen Runs. Die Schätzer sind

$$\hat{\mathrm{P}}^{(D)}(k+) = \sum_{i \geq k} \hat{\mathrm{P}}^{(D)}(i) \quad \text{bzw.} \quad \hat{\mathrm{P}}^{(W)}(k-) := \sum_{i \leq k} \hat{\mathrm{P}}^{(W)}(i) \; .$$

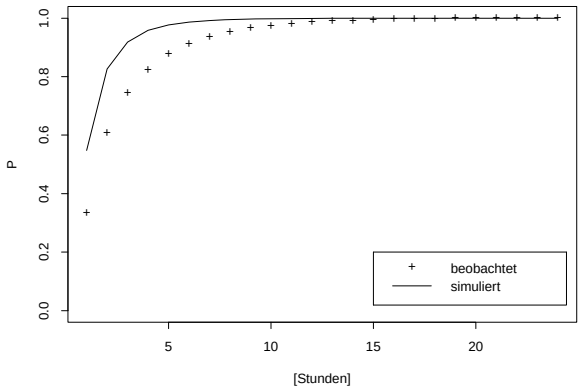
Die Anzahl der beobachteten Runs in einem Monat wird mit einem einfachen Zählalgorithmus als die Anzahl der in einem Monat beginnenden Runs ermittelt. Der erste Run in einem Monat beginnt also mit dem ersten Wechsel eines Zustandes (von naß

Abbildung 6.23: Nasse Runs



Bedingte Wahrscheinlichkeiten nasser Runs mit i Stunden, $i = 1, \ldots, 25$, berechnet (a) auf Jahresbasis, (b) für Januar, (c) für Juli.
Hellgrau: 27 beobachtete Jahre, schwarz: 1000 simulierte Jahre.

Abbildung 6.24: Nasse Runs – Maximallängen



Bedingte Wahrscheinlichkeiten $\hat{\mathrm{P}}^{(W)}(k-)$ nasser Runs der Längen $i \leq k$, $k = 1, \ldots, 24$ Stunden; ganzes Jahr

---

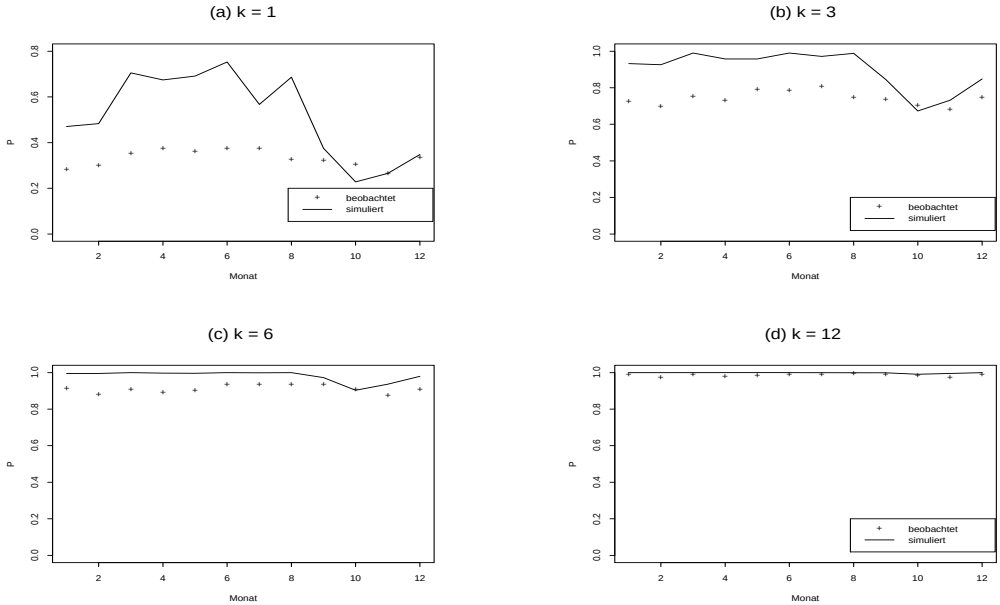
zu trocken oder umgekehrt). Am Monatsende wird stets bis zum Ende des Runs gezählt, auch wenn der Run in einem anderen Monat endet, als er begonnen hat. Der Zählalgorithmus sowie der Quellcode werden in Anhang C.3 gegeben. Wir zählen Runs der Länge $i, i = 1, \ldots, 168, 169+$ Stunden ($168 = 7 \cdot 24$ Stunden sind exakt eine Woche, $169+$ bedeutet $i \geq 169$), und berechnen daraus die empirischen bedingten Wahrscheinlichkeiten nasser und trockener i-Runs gemäß (5.3) und (5.4).

In Abbildung 6.23 sind die bedingten Wahrscheinlichkeiten nasser Runs einer Länge von exakt i Stunden dargestellt. Die Wahrscheinlichkeit $\hat{\mathrm{P}}^{(W)}_{sim}(1)$ unter dem Modell ist deutlich größer als die beobachtete Wahrscheinlichkeit $\hat{\mathrm{P}}^{(W)}_{beob}(1)$ eines nassen 1-Runs, und zwar sowohl auf Jahresbasis wie auch auf monatlicher Basis. Entsprechend werden mehr lange nasse Runs beobachtet als unter dem Modell erzeugt werden.

In Abbildung 6.24 steigt die unter dem Modell erhaltene Kurve der bedingten Wahrscheinlichkeiten $\mathrm{P}^{(W)}(k-)$ nasser Runs der Länge von maximal k Stunden als Funktion von k schneller an als die aus den Beobachtungen erhaltene Kurve. Das bedeutet, daß unter dem Modell im Mittel kürzere nasse Runs auftreten als in den Beobachtungen. Die Wahrscheinlichkeit, einen nassen Run von mehr als $k = 10$ Stunden zu beobachten, ist jedoch sowohl unter dem Modell wie auch in den Beobachtungen fast null. Abbildung 6.25 zeigt ebenfalls, daß unter dem Modell in stärkerem Maße als in den Beobachtungen nasse 1-Runs auftreten. Aber die Wahrscheinlichkeiten $\mathrm{P}^{(W)}(k-)$ für $k \geq 3$ stimmen unter dem Modell und unter den Beobachtungen befriedigend überein.

Abbildung 6.28 zeigt die unter dem Modell und die aus den Beobachtungen geschätzten bedingten Wahrscheinlichkeiten $\mathrm{P}^{(D)}(i)$ trockener Runs der Länge i. Offensichtlich enthalten die Beobachtungen mehr trockene 1-Runs. Die übrigen Schätzer $\hat{\mathrm{P}}^{(D)}(i)$ stimmen jedoch gut überein. Eine weitere Fragestellung ist die nach den bedingten Wahrscheinlichkeiten eines trockenen Runs einer Dauer zwischen k_1 und k_2 Stunden, wenn die Startstunde eines trockenen Runs gegeben ist. Das sind die

Abbildung 6.25: Nasse Runs – Maximallängen



Bedingte Wahrscheinlichkeiten $\hat{\mathrm{P}}^{(W)}(k-)$ nasser Runs der Länge $i \leq k$ Stunden.
(a) $k = 1$, (b) $k = 3$, (c) $k = 6$, (d) $k = 12$.

Wahrscheinlichkeiten

$$\sum_{i=k_1}^{k_2} \mathrm{P}^{(D)}(i) \ ,$$

die durch Aufsummation der Schätzer $\hat{\mathrm{P}}^{(D)}(i)$, $k_1 \leq i \leq k_2$, geschätzt werden. Eine Einteilung der trockenen Runs in Klassen á 12 Stunden, nämlich in Runs der Dauer $1 - 12, 13 - 24, \ldots$ Stunden, ist sinnvoll. Abbildung 6.29 zeigt die gute Übereinstimmung der Beobachtungen mit den simulierten Daten in dieser Hinsicht.

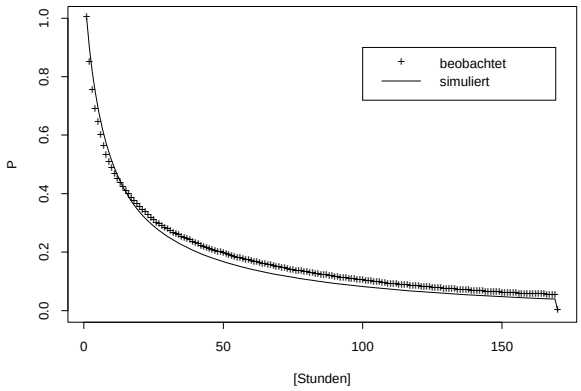
Ist die Mindestlänge eines trockenen Runs von Interesse, so ist die Übereinstimmung zwischen den aus den Beobachtungen und den aus den Simulationen berechneten Schätzern noch besser, und zwar auf Jahresbasis wie auch auf Basis der einzelnen Monate, vergl. Abb. 6.26 und 6.27. Für manche Anwendungen, insbesondere in der Landwirtschaft, ist die Wahrscheinlichkeit von Interesse, mindestens eine komplett trockene Woche zu erhalten. Diese Wahrscheinlichkeit (in %) beträgt in den einzelnen Monaten:

Bedingte Wahrscheinlichkeit [%] eines trockenen 169+-Runs

Monat	J	F	M	A	M	J	J	A	S	O	N	D	Jahr
Beob.	5.6	6.2	5.9	4.1	2.7	2.8	3.8	5.9	6.8	7.5	5.7	5.1	5.0
Sim.	4.1	2.1	4.9	4.5	5.5	1.9	2.4	4.5	6.5	6.5	4.6	3.1	3.9

Das Modell stimmt also in bezug auf die bedingten Verteilungen der trockenen und nassen Runs gut überein. Das ist umso bemerkenswerter, als in der Modellanpassung

Abbildung 6.26: Trockene Runs - Mindestlängen



Bedingte Wahrscheinlichkeiten $\hat{\mathrm{P}}^{(D)}(k+)$ trockener Runs der Länge $i \geq k$, $k = 1, \ldots, 169+$ Stunden; ganzes Jahr.

---

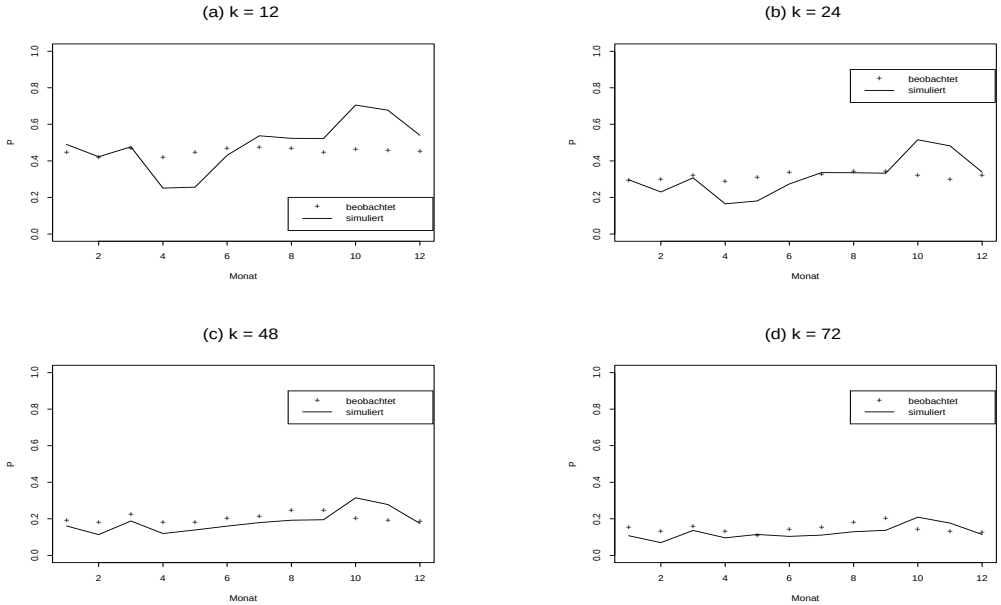
nur die Wahrscheinlichkeiten $\phi(h = 24)$ und $\phi_{WW}(h = 1)$ verwendet worden sind, also sich nur auf zwei Zeiteinteilungen beziehende Wahrscheinlichkeiten. In praktischen Anwendungen wird die Wahrscheinlichkeit, exakt i trockene oder maximal i Stunden ununterbrochenen Niederschlages zu erhalten, eher von untergeordnetem Interesse sein. Man wird sich auch für die bedingten Wahrscheinlichkeiten interessieren, innerhalb der nächsten k Stunden nur eine Stunde mit Niederschlag zu erhalten, wenn in der gegebenen Stunde ein trockener Run beginnt oder umgekehrt für die bedingte Wahrscheinlichkeit, innerhalb der nächsten k Stunden vorwiegend nasse Stunden zu erhalten, wenn in der gegebenen Stunde ein nasser Run beginnt. Aber auch diese Wahrscheinlichkeiten lassen sich durch einfaches Auszählen berechnen und nach den eben durchgeführten Betrachtungen sind unter dem Modell verläßliche Aussagen zu erwarten.

6 Verteilung des Niederschlages an nassen Tagen

Der Ereignisprozeß stündlichen Niederschlages wird durch die Verteilung der Runlängen und die Verteilung der Anzahl nasser Stunden an nassen Tagen charakterisiert, die wir in diesem Abschnitt betrachten werden.

Der Anteil beobachteter nasser Tage mit i nassen Stunden an allen nassen Tagen ist $r_i = N_i/NR$, $i = 1, \ldots, 24$, wenn N_i die Anzahl an Tagen mit i nassen Stunden und NR die Gesamtanzahl nasser Tage sind. Die Kurven der r_i und der aus den simulierten Daten berechneten Anteile $\tilde{r}_i$ besitzen ähnlich Verläufe, vergl. Abb. 6.30. Es ist aber deutlich erkennbar, daß $r_i < \tilde{r}_i$ für $i = 1, \ldots, 5$ und $r_i > \tilde{r}_i$ für $i = 6, \ldots, 24$. In den Beobachtungen und unter dem Modell gibt es am häufigsten nasse Tage mit zwei nassen Stunden, aber die größten Differenzen $r_i - \tilde{r}_i$ finden sich ebenfalls bei $i = 2$ und $i = 3$, nämlich 4.11% bzw. 3.38%. Bei separater Analyse der einzelnen Monate sind die Ergebnisse ähnlich.

Abbildung 6.27: Trockene Runs - Mindestlängen



Bedingte Wahrscheinlichkeiten $\hat{\mathrm{P}}^{(D)}(k+)$ trockener Runs der Länge $i \geq k$ Stunden.
(a) $k = 12$, (b) $k = 24$, (c) $k = 48$, (d) $k = 72$.

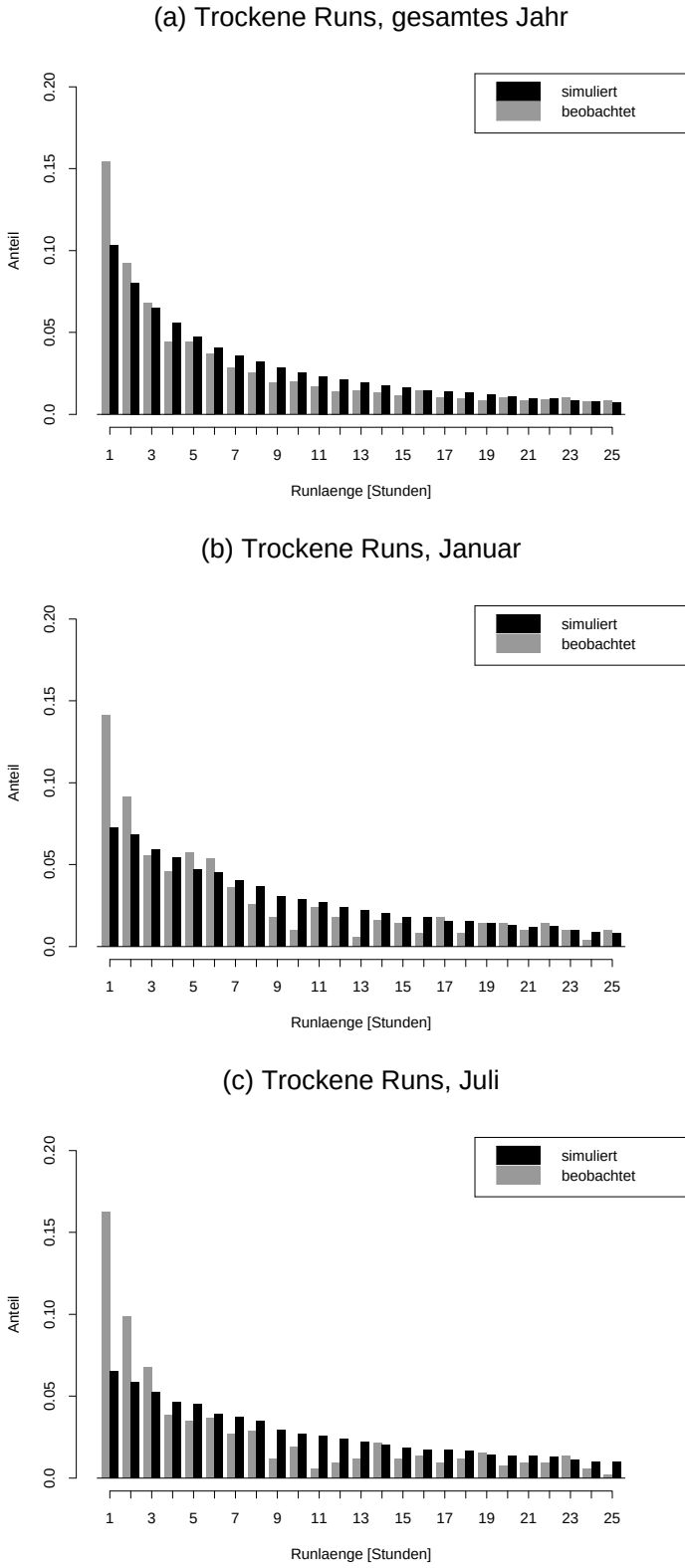
Die auf Jahresbasis aus den Beobachtungen und aus den Simulationen berechneten Anteile r_i (i %) sind in der folgenden Tabelle zu sehen. Evident ist, daß unter dem Modell zu wenige Tage mit mehr als 15 nassen Stunden auftreten.

Anteil r_i [%] der nassen Tage mit i nassen Stunden

i	1	2	3	4	5	6	7	8	9	10	11	12
Beob.	11.8	17.6	14.0	11.8	9.4	7.4	5.8	5.5	3.7	3.0	2.6	1.7
Sim.	14.9	21.7	17.4	13.4	9.9	7.1	5.1	3.7	2.4	1.6	1.0	0.7
i	13	14	15	16	17	18	19	20	21	22	23	24
Beob.	1.0	1.1	1.1	0.6	0.4	0.3	0.5	0.3	0.1	0.1	0.1	0.1
Sim.	0.5	0.3	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

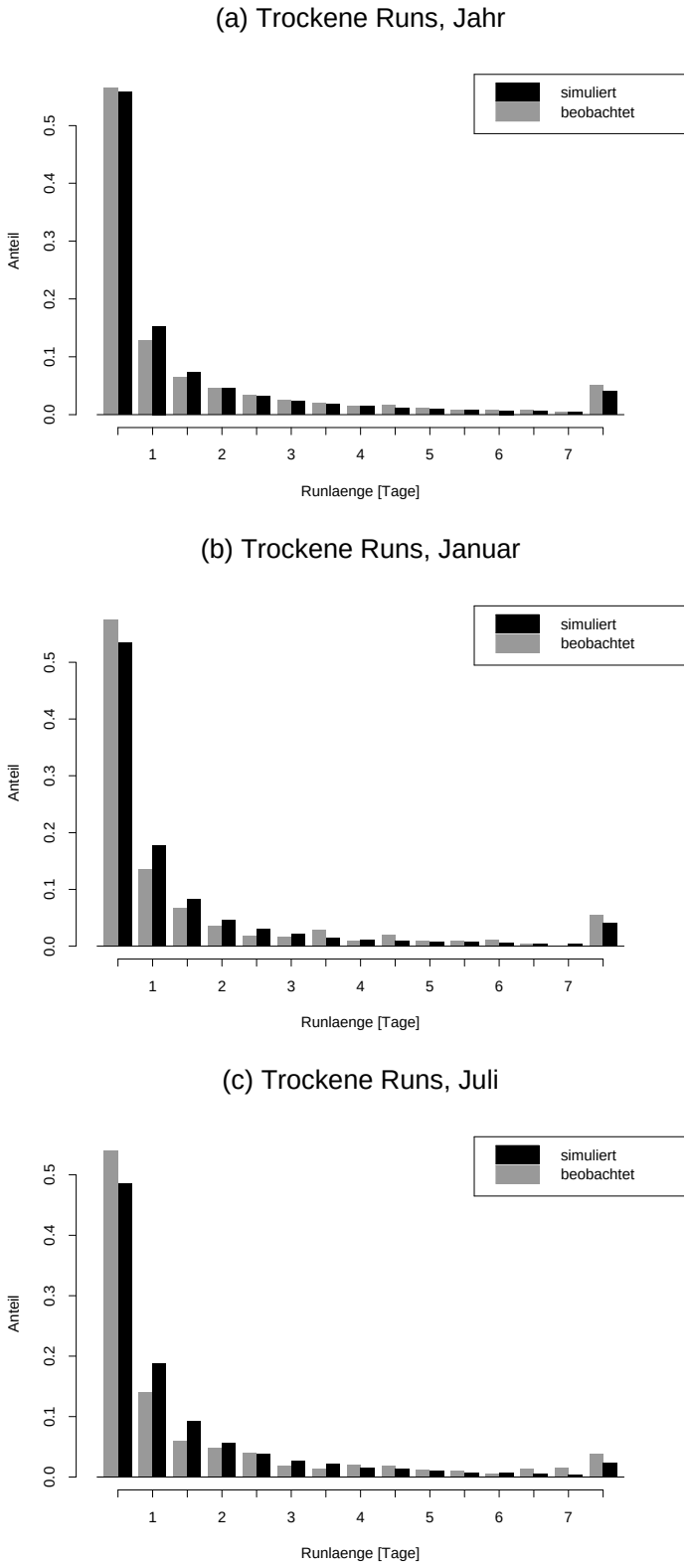
Das Modell scheint nicht in der Lage zu sein, den zu beobachtenden langanhaltenden Regen zu erzeugen. Unter dem Modell treten zu wenige Tage mit mehr als 10 nassen Stunden auf, was mit einem zu hohen Anteil kurzer nasser Runs und zu niedrigem Anteil kurzer trockener Runs korrespondiert. Die Anteile r_i und $\tilde{r}_i$ unterscheiden sich zwar, aber in akzeptablem Maße. Ob Aussagen über die Anzahl nasser Stunden an einem nassen Tag auf das Neyman-Scott-Modell basiert werden können, sollte von der für eine Anwendung erforderlichen Genauigkeit der Vorhersage abhängig gemacht werden.

Abbildung 6.28: Trockene Runs



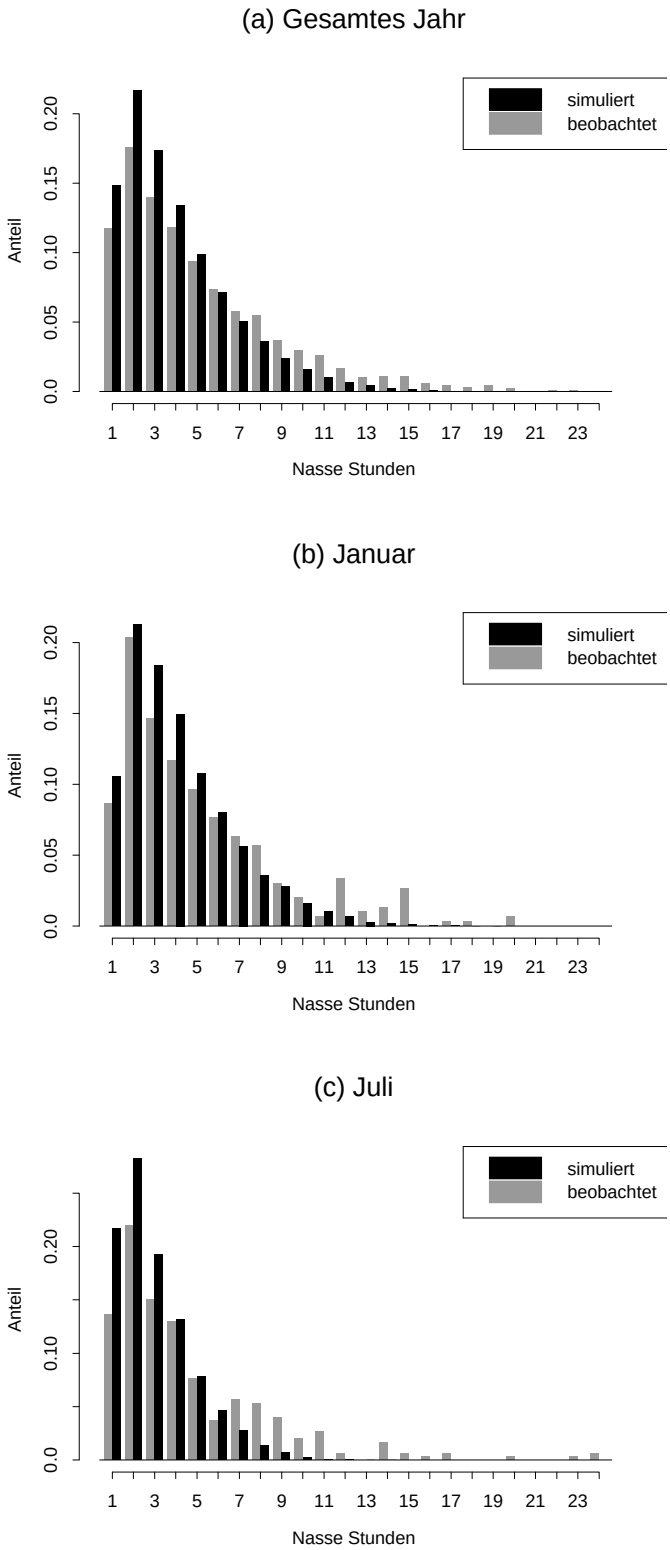
Bedingte Wahrscheinlichkeiten trockener Runs einer Länge von exakt $i = 1, \ldots, 25$ Stunden. Hellgrau: 27 beobachtete Jahre, schwarz: 1000 simulierte Jahre.

Abbildung 6.29: Trockene Runs - Klasseneinteilung



Bedingte Wahrscheinlichkeiten trockener Runs einer Dauer von $1-12, 13-24, \ldots, 145-168, > 168$ Stunden, entsprechend einer Klassifikation in halbe Tage.
Hellgrau: 27 beobachtete Jahre, schwarz: 1000 simulierte Jahre.

Abbildung 6.30: Innertägliche Verteilung der Niederschläge an nassen Tagen



Anteil r_i der nassen Tage mit i, $i = 1, \ldots, 24$ nassen Stunden, berechnet
(a) auf Jahresbasis, (b) für Januar, (c) für Juli.
Hellgrau: 27 beobachtete Jahre, schwarz: 1000 simulierte Jahre.

7 Zusammenfassung

Zur Modellvalidation des angepaßten saisonalen Neyman–Scott–Modelles sind unterschiedliche Eigenschaften des Niederschlagsprozesses unter dem Modell mit ihren empirischen Äquivalenten verglichen worden. Der Vergleich basiert auf 1000 Jahren künstlicher stündlicher Niederschlagsmessungen, die unter Verwendung des in Abschnitt 5.5 geglätteten Parametersets B1 erzeugt worden sind. Die analytisch berechneten Momente des Niederschlagsmengenprozesses sind ebendort bereits mit den korrespondierenden empirischen Schätzern verglichen worden.

In diesem Abschnitt sind die in Kapitel 5.(2.2) als Set $\mathcal{F}$ bezeichneten Momente des Niederschlagsmengenprozesses für jeden Monat aus den simulierten Daten unter der Annahme berechnet worden, ein Monat sei eine homogene Periode. Diese Momente werden mit den empirischen, aus dem Stuttgarter Datensatz berechneten verglichen. Alle der betrachteten Momente werden sehr gut reproduziert. Dabei weisen lediglich die Varianzen $\gamma(h)$ Variationen auf, wenn sie aus unterschiedlichen, jeweils 100 Jahre umfassenden Sets separat berechnet werden. Die übrigen Momente weisen nur geringe Variation von Set zu Set auf.
Die bemerkenswerten Unterschiede zu den in Abschnitt 5.5 analytisch für jede 30-Tages-Periode eines Jahres berechneten Momente zeigen, daß das saisonale Neyman–Scott–Modell andere Eigenschaften besitzt als das Stationarität zugrunde legende Modell. Sie verdeutlichen damit auch, daß die aufgezeigten saisonalen Schwankungen des Modelles stark geglättet werden, wenn mit dem saisonalen Modell simulierte Daten unter der Annahme eines stückweise stationären Prozesses analysiert werden. Die Mittel der Monatsniederschläge und des Jahresniederschlages sind unter dem Modell wesentlich zu groß, der mittlere Jahresniederschlag beispielsweise um 20%. Wird für eine Anwendung Wert auf die korrekte Wiedergabe dieser Eigenschaften gelegt, sollte also ein anderes Modell gewählt werden, etwa das in Kapitel 4 angepaßte Markovkettenmodell.
In einer Extremwertanalyse sind die 1–, 6– und 24–Stunden–Extrema auf Jahres– und auf Monatsbasis betrachtet worden. Die 98%– und 99%–Quantile der 1–Stunden–Extrema sind deutlich zu klein. Für die 6– und 24–Stunden–Niederschläge werden sie angemessen wiedergegeben, auch wenn die tatsächliche saisonale Variation nicht exakt reproduziert wird. Die Verteilung der simulierten Maxima läßt sich durch eine Lognormalverteilung beschreiben. Die auf monatlicher Basis analysierten 6– und 24–Stunden–Maxima können ebenfalls durch Lognormalverteilungen beschrieben werden, deren Skalenparameter μ sich im Jahresverlauf gemäß einer Fouriersumme verändert, während der Formparameter σ konstant ist.
Die Niederschlagsmengen in nassen Stunden sind unter dem Neyman–Scott–Modell im Mittel deutlich zu groß. Die großen Quantile wie $Q(0.99)$ werden wiederum akzeptabel wiedergegeben, was die passable Performance des Modelles in der Extremwertanalyse erklärt. Hier liegt eine Schwäche des angepaßten Modelles, denn gerade eine detaillierte Beschreibung der Niederschlagsmengen in nassen Stunden — und nicht nur der Mittelwerte — ist ein nicht zu vernachlässigender Anspruch an ein Modell für stündlichen Niederschlag.
Abschließend haben wir in diesem Kapitel die zeitliche Verteilung der Niederschläge unter dem Modell mit der beobachteten verglichen. Die Übereinstimmung ist be-

merkenswert gut, und zwar sowohl in bezug auf die Längen trockener und nasser Runs wie auch in bezug auf die mittlere Anzahl nasser Stunden an nassen Tagen. Diese Aussagen treffen auf Jahresbasis wie auch auf monatlicher Basis zu.

Zusammenfassend folgern wir, daß das angepaßte Neyman–Scott–Modell zwar kein perfektes Modell des beobachteten Niederschlagsprozesses ist — insbesondere der zu große Jahresniederschlag ist ein ernsthaftes Manko, ebenso die zu kleinen extremen Quantile der Verteilung der h-Stunden–Niederschlagsmengen —, daß es aber insgesamt dennoch verläßliche Aussagen über die Struktur des Niederschlagsprozesses ermöglicht.

Kapitel 7

Zusammenfassung und Ausblicke

Das Ziel dieser Arbeit bestand in der Beschreibung stündlicher Niederschlagsdaten durch ein verallgemeinertes Neyman–Scott–Clustermodell, dessen Modellparameter stetige Funktionen der Zeit sind, so daß dadurch Saisonschwankungen in den Daten erklärt werden können.

Nach einem Überblick über gängige Modelle zur Beschreibung von Niederschlag ist an die zu täglichem Niederschlag aggregierten Messungen erfolgreich ein Markovkettenmodell mit weibullverteilten Niederschlagsmengen an nassen Tagen angepaßt worden. Das Modell erklärt Saisonschwankungen und kann sowohl täglichen Niederschlag als auch die Verteilung der Monats- und Jahresniederschläge korrekt reproduzieren.

Anschließend ist an die stündlichen Niederschlagsmessungen ein saisonales Neyman–Scott–Clustermodell angepaßt worden. Die Anpassung basierte auf einer Kombination eines Kleinste–Quadrate–Ansatzes mit der Methode der Momente. Die damit geschätzten Werte der Modellparameter hingen in starkem Maße von den in der Minimierung verwendeten Momenten ab und, da die Zielfunktion nicht konkav war, in gleichem Ausmaße von den Startwerten der Minimierungsroutine. Die sich anschließende Glättung der Werte der Modellparameter durch Fouriersummen war arbeitsintensiv, weil die Schätzer zunächst um einzelne Ausreißer bereinigt und geeignete Fourierfrequenzen ausgewählt werden mußten. Diese beiden Schritte mußten sorgsam erfolgen und ließen sich nicht automatisieren. Aus diesen beiden Gründen erwies sich die Modellanpassung als schwierig und nicht im mindesten als so einfach, wie es die gängige Literatur vermuten ließ.

Wie bereits erwähnt, reproduziert das angepaßte Modell die Saisonschwankungen der Momente des Niederschlagsmengenprozesses mindestens akzeptabel, zum Teil sehr gut und erklärt auch deren Verhalten in Abhängigkeit von der Länge des Aggregationsintervalles.
Auf stündlicher Basis werden das Mittel der h-Stunden–Niederschlagsmenge angemessen, deren Varianzen gut und die Lag(1)-Autokorrelationen sehr gut reproduziert. Die Übergangswahrscheinlichkeiten nasser h-Stunden-Intervalle sind unter dem Modell zu groß. Die Übergangswahrscheinlichkeiten und die Wahrscheinlichkeiten trockener h-Stunden-Intervalle werden angemessen wiedergegeben. Insbesondere

werden die Saisonschwankungen der Momente zufriedenstellend reproduziert.
Auf monatlicher Basis werden diese Momente sämtlich sehr gut wiedergegeben. Der mittlere Jahresniederschlag ist unter dem Modell um 20% zu groß. Sollte also eine Anwendung eine exakte Wiedergabe der Eigenschaften desselben erfordern, so müßte ein anderes Modell gewählt werden, beispielsweise eines, das nur auf Jahresdaten basiert. Ebenso ist die Wiedergabe der Verteilung der Niederschlagsmengen in nassen Stunden nicht korrekt. Unter dem Modell ist deren Mittel zu groß und die Verteilung nicht genügend rechtsschief. Die Verteilung der 1-Stunden-Extrema ist unter dem Modell ebenfalls nicht genügend rechtsschief, die der der 6– und 24–Stunden–Extrema wird hingegen adäquat beschrieben. Weiterhin beschreibt das Modell die Verteilung der Runlängen und die Anzahl der nassen Stunden an nassen Tagen korrekt. Detailliert werden die Stärken und Schwächen des Modelles in den Abschnitten 5.5 und 6.7 genannt.

Dieses Modell bietet unterschiedliche Ansatzpunkte für zukünftige Forschungsarbeit. Als Erweiterung des Modelles ist an andere Zellformen zu denken, um die Wiedergabe der Niederschlagsmengen in nassen Stunden zu verbessern. Hier bietet sich zum einen an, den Parameter ξ der Verteilung der Niederschlagsmengen mit einem weißen Rauschen zu überlagern. Zum anderen könnten dreieckige Zellen untersucht werden, mit denen die häufig zu beobachtende erst zunehmende, dann wieder abnehmende Intensität von Regenschauern erklärt werden könnte. Die Zellen könnten als gleichschenklige Dreiecke konstruiert werden, deren Kantenlänge zufällig ist.

Aus mathematischem Interesse wären die theoretischen Eigenschaften des nichtstationären Modelles zu untersuchen, insbesondere die Momente erster und zweiter Ordnung des Mengenprozesses. Aufgrund der Komplexität des Modelles dürfte das aber ein aufwendiges Unterfangen sein. Für Anwendungen lassen sich diese wie auch andere, für spezielle Anwendungen relevante Eigenschaften leicht mit Monte–Carlo–Simulationsmethoden aus einmal geschätzten Modellparametern bestimmen. Damit ist mit diesem Modell insgesamt eine Erklärung der Saisonschwankungen in den analysierten stündlichen Niederschlagsmessungen gelungen.

Literaturverzeichnis

Acreman, M.C. (1990) A simple stochastic model for hourly rainfall for Farnborough, England. *Hydrol. Sci. J.* 35(2), 119 - 147.

Bacchi, B., Becciu, G., Kottegoda, N. (1994) A bivariate exponential model applied to intensities and durations of extreme rainfall. *J. Hydrol.* 155, 225 - 236.

Bauer, H. (1991) *Wahrscheinlichkeitstheorie.* De Gruyter, Berlin.

Benjamin, J., Cornell, C. (1970) *Probability, Statistics and Decision for Civil Engineers.* McGraw-Hill Book Company, New York.

Bloomfield, P. (1976) *Fourier Analysis of Time Series: An Introduction.* Wiley & Sons, New York.

Buishand, T.A. (1977) The stochastic modelling of daily rainfall sequences. *Rep.* 77(3), *Medelingen Landbouwhogeschool*, Wageningen, Die Niederlande.

Burlando, P., Rosso, R., Cadavid, L., Salas, J. (1993) Forecasting of short-term rainfall using ARMA models, *J. Hydrol.* 144, 193 - 211.

Chang, T.J., Kavvas, M.L., Delleur, J.W. (1984) Daily precipitation modelling by discrete autoregressive moving average processes. *Water Resour. Res.* 20(5), 565 - 580.

Chatfield, C. (1978) *Statistics for Technology.* 2nd ed., Chapman & Hall, London.

Chatfield, C. (1996) *The Analysis of Time Series: An Introduction.* 5th edition, Chapman & Hall, London.

Cowpertwait, P. (1991) Further Developments of the Neyman-Scott Clustered Point Process for Modeling Rainfall. *Water Resour. Res.* 27(7), 1431 - 1438.

Cowpertwait, P. (1994) A generalized point process model for rainfall. *Proc. Roy. Soc. London Ser. A* 447, 23 - 37.

Cowpertwait, P., O'Connell, P., Metcalfe, A., Mawdsley, J. (1996a) Stochastic point process modelling of rainfall. I. Single-site fitting and validation. *J. Hydrol.* 175, 17 - 46.

Cowpertwait, P., O'Connell, P., Metcalfe, A., Mawdsley, J. (1996b) Stochastic point process modelling of rainfall. II. Regionalisation and disaggregation. *J. Hydrol.* 175, 47 - 65.

Cox, D., Isham, V. (1979) *Point Processes.* Chapman & Hall, London.

Foufoula-Georgiou, E., Guttorp, P. (1986) Compatibility of continuous rainfall occurrence models with discrete rainfall observations. *Water Resour. Res.* 22(5), 1316 - 1322.

Foufoula-Georgiou, E., Lettenmaier, D. (1986) Continuous-time versus discrete-time point process models for rainfall occurrence series. *Water Resour. Res.* 22(4), 531 - 542.

Foufoula-Georgiou, E., Lettenmaier, D. (1987) A Markov Renewal Model for rainfall occurrences. *Water Resour. Res.* 23(5), 875 - 884.

Green, J.R. (1964) A model for rainfall occurrence. *J. Roy. Stat. Soc. Ser. B* 26, 345 - 353.

Guttorp, P. (1986) On binary time series obtained from continuous time point processes describing rainfall, *Water Resour. Res.* 22(6), 897 - 904.

Haberland, E. (1975): Infinitely divisible stationary recurrent point processes. *Math. Nachr.* 70, 260 - 264.

Karr, A. (1986) *Point Processes and Their Statistical Inference.* Marcel Dekker Verlag, New York.

Katz, R. (1981) On some criteria for estimating the order of a markov chain. *Technometrics* 23(3), 243 - 249.

Katz, R. (1984) Discussion of the paper by Dr. Stern and Mr. Coe. *J. Roy. Stat. Soc. Ser. A* 147, Part 1, 29.

Katz, R., Parlange, M. (1995) Generalizations of chain-dependent processes: Application to hourly precipitation. *Water Resour. Res.* 31(5), 1331 - 1341.

Kavvas, M., Delleur, J. (1981) A stochastic cluster model of daily rainfall sequences. *Water Resour. Res.* 17(4), 1151 - 1160.

Khaliq, M., Cunnane, C. (1996) Modelling point rainfall occurrences with the modified Bartlett-Lewis rectangular pulses model. *J. Hydrol.* 180, 109 - 138.

Le Cam, L. (1961) A stochastic description of precipitation, in Neyman, J. (ed.): *Fourth Berkeley Symposium on Mathematics, Statistics, and Probability Proceedings*, University of California Press, Berkeley.

Linhart, H., Zucchini, W. (1986) *Model Selection.* Wiley & Sons, New York.

MacDonald, I., Zucchini, W. (1997) *Hidden-Markov and other models for discrete-valued time series*, Chapman & Hall, London.

McCullagh, P., Nelder, J. (1989) *Generalized Linear Models*, Chapman & Hall, London.

Neyman, J., Scott, E. (1958) A statistical approach to problems of cosmology, *J. R. Stat. Soc. Ser. B* 20, 1 - 43.

Parzen, E., Tanabe, K., Kitagawa, G. (1988) *Selected Papers of Hirotugu Akaike*. Springer, New York.

Press, W., Teukolsky, S., Vetterling, W., Flannery, B. (1992) *Numerical Recipes in C - The Art of Scientific Computing*. Cambridge University Press, 2nd ed., Cambridge.

Rodriguez-Iturbe, I., Gupta, V., Waymire, E. (1984) Scale considerations in the modeling of temporal rainfall. *Water Resour. Res.* 20(11), 1611 - 1619.

Rodriguez-Iturbe, I., Cox, D., Isham, V. (1987) Some models for rainfall based on stochastic point processes. *J. Roy. Stat. Soc. Ser. A* 410, 269 - 288.

Rodriguez-Iturbe, I., Cox, D., Isham, V. (1988) A point process model for rainfall: further developments. *Proc. R. Soc. Lond. Ser. A* 417, 283 - 298.

Roldan, J., Woolhiser, D. (1982) Stochastic daily precipitation models, 1. A comparison of occurrence models. *Water Resour. Res.* 18(5), 1451 - 1459.

Smith, J., Karr, A. (1983) A point process model of summer season rainfall occurrences. *Water Resour. Res.* 19(1), 95 - 103.

Smith, J., Karr, A. (1985) Statistical inference for point process models of rainfall. *Water Resour. Res.* 21(1), 73 - 79.

Smith, J. (1987) Statistical modeling of daily rainfall occurrences. *Water Resour. Res.* 23(5), 885 - 893.

Smith, R., Tawn, J., Coles, S. (1990) Markov chain models for threshold exceedances. *Biometrika* 84, 249 - 268.

Stern, R., Coe, R. (1984) A model fitting analysis of daily rainfall data. *J. Roy. Soc. Ser. A* 147(1), 1 - 34.

Tong, H. (1975) Estimating the order of a markov chain by Akaike's information criterion. *J. Appl. Prob.* 12, 488 - 497.

Woolhiser, D. (1992) Modeling daily precipitation - progress and problems. In: Walden, A.T., Guttorp, P. (ed.): *Statistics in the Evironment and Earth Sciences*. Edward Arnold Publishers, London.

Yang, D., Goodison, B., Ishida, S., Benson, C. (1998) Adjustment of daily precipitation data at 10 climate stations in Alaska. *Water Res. Res.* 34(2), 241 - 256.

Zucchini, W., Adamson, P. (1984) The Occurrence and Severity of Droughts in South-Africa. Report to the Water Research Commission by the Department of Civil Engineering, University of Stellenbosch and the Department of Water Affairs, South Africa. *WRC Report No. 91/1/84*.

Zucchini, W., Adamson, P. (1989) Bootstrap confidence intervals for design storms from exceedance series. *Hydrol. Sci. J.* 34(1), 41 - 48.

Zucchini, W., Guttorp, P. (1991) A hidden markov model for space-time precipitation. *Water Resour. Res.* 27(8), 1917 - 1923.

Dank

Diese Arbeit entstand während meiner Tätigkeit als wissenschaftliche Angestellte am Institut für Statistik und Ökonometrie der Universität Göttingen.
Allen, die mir während dieser Zeit hilfreich zur Seite standen, möchte ich ganz herzlich danken.

Herr Prof. Zucchini hat das Thema dieser Arbeit vorgeschlagen und mich damit für das Gebiet der Niederschlagsforschung begeistert. Ihm gilt mein Dank für die Betreuung der Arbeit und insbesondere für seine stete Bereitschaft, mich mit konstruktiven Ratschlägen zu unterstützen. Herrn Prof. Böker danke ich für die Übernahme des Zweitgutachtens.

Frau Monika Dannhauer hat mir wertvolle Hilfe bei der Datenaufbereitung geleistet und geduldig meine ersten Schritte in der Welt der Bits und Bytes begleitet, dafür danke ich ihr von ganzem Herzen.

Meinem Kollegen Herrn Dipl.-Phys. Peter Majer danke ich sehr für die vielen anregenden Gespräche über das Thema dieser Arbeit. Mein aufrichtiger Dank gilt meinem Kollegen Herrn Dipl.-Math. Olaf Dannenberg für seine stets freundliche Hilfe bei diversen Schwierigkeiten im Umgang mit Hard- und Software sowie für zahlreiche wertvolle Ideen auf dem Gebiet der Statistik. Die kooperative und motivierende Arbeitsatmosphäre des Institutes hat ebenfalls wesentlich zu dieser Arbeit beigetragen.

Meinen Eltern danke ich für die stetige Unterstützung, die sie mir während meines Studiums zuteil werden ließen.

Lebenslauf

Ulrike Ohlmer wurde am 10. Oktober 1968 in Hildesheim geboren, wo sie von 1974 bis 1978 die Grundschule am Moritzberg besuchte und von 1978 bis 1987 das Goethegymnasium. Der Schulabschluß mit dem Abitur folgte im Juni 1987. Die Aufnahme des Mahematikstudiums an der Philipps–Universität zu Marburg im Oktober 1987 erschien ihr bald ein ebenso törichter Fehler zu sein, wie der im April 1988 vorgenommene Wechsel zum Studium der Betriebswirtschaftslehre an die Georg–August–Universität zu Göttingen. Diese Phase endete mit der nun definitiven Wiederaufnahme des Mathematikstudiums in Göttingen im Oktober 1988, dessen Schwerpunkt in der Reinen Mathematik lag. Im Juli 1993 erfolgte mit einer Diplomarbeit auf dem Gebiet der Topologie der Studienabschluß als Diplom-Mathematikerin. An ihn schloß sich eine Übergangstätigkeit am Mathematischen Institut der Universität Göttingen an. Als Konsequenz einer allmählichen Interessensverlagerung hin zu angewandten Fragestellungen wechselte sie als wissenschaftliche Angestellte an das Institut für Statistik und Ökonometrie der Universität Göttingen. Im Rahmen dieser Tätigkeit nahm sie nicht nur Einblicke in unterschiedliche Gebiete der Statistik, sondern fertigte auch die vorliegende Dissertation an.

Ich versichere an Eides Statt, daß ich die eingereichte Dissertation

> „Ein stochastisches Modell zur Beschreibung stündlichen Niederschlages"

selbständig verfaßt habe. Anderer als der angegebenen Hilfsmittel und Schriften habe ich mich nicht bedient. Alle wörtlich oder sinngemäß den Schriften anderer Autoren entnommenen Stellen habe ich kenntlich gemacht.

Anhang A

Die Rohdaten

Der Datensatz wird unter anderem mit Cusum-Charts auf Homogenität überprüft, wie in Kapitel 3 beschrieben. Dort sind Cusum-Charts der Jahresniederschlagsmenge und -dauer gegeben. Zur weiteren Überprüfung des Datensatzes sind Cusum-Charts der Niederschlagsmenge und der mittlere Ereignismenge pro Monat während des Beobachtungszeitraumes 1964 - 1990 erstellt worden.

Konstruktion eines Cusum Charts

Cusum Charts (CHATFIELD [1978, S. 306 ff.]) sind eine graphische Methode aus dem Bereich der statistischen Qualitätskontrolle zur Überprüfung, ob sich ein Prozeß, z.B. die Temperatur während eines industriellen Fertigungsprozesses, im zulässigen Bereich befindet oder ob Veränderungen des Prozesses auftreten, die ein Eingreifen erfordern, z.B. ein Abschalten des Prozesses.
Mit Cusum Charts läßt sich ebenso überprüfen, ob ein Datensatz homogen ist. Würde beispielsweise ein neuer, windgeschützterer Standort eines Niederschlagsmeßgerätes zu signifikant höheren Jahresniederschlägen führen, wäre diese Veränderung in einem Cusum Chart sichtbar. In diesem Fall müßten die Daten nachträglich bereinigt oder der Datensatz müßte in zwei homogene Perioden unterteilt werden.

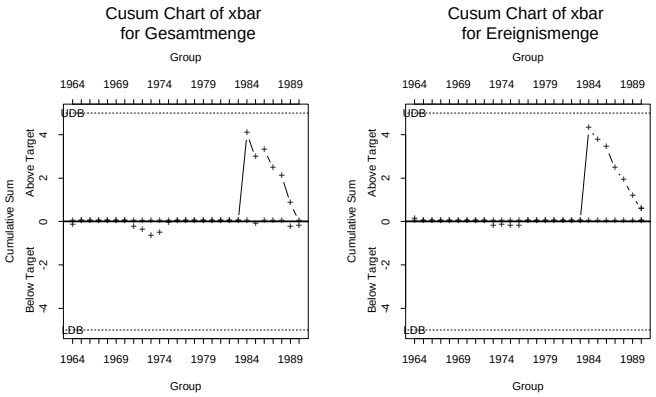
In einem Cusum Chart werden die kumulativen, standardisierten Abweichungen U_n nach oben und L_n nach unten von einem Zielwert Z_n abgetragen, die bei Realisationen x_n einer Zufallsvariablen X_n beobachtet werden. Es sind:

$$U_n = \max(0, (z_n - 0.5) + U_{n-1}) \ , \quad L_n = \max(0, (-z_n - 0.5) + L_{n-1}) \ ,$$

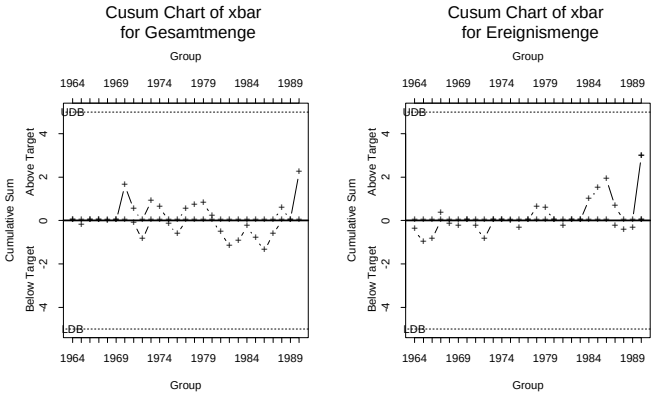
wobei $z_n = (x_n - Z_n)/\sigma \ , n = 1, \ldots, N$. Bei uns sind $Z_n = \sum_{n=1}^{N} x_n/n$ und σ die Standardabweichung $s(x)$ der Stichprobe. Die in den Graphiken eingezeichneten Entscheidungsgrenzen UDE und LDE verlaufen bei $\pm 5s(x)$. Eine Überschreitung der Intervallgrenzen bedeutet eine signifikante Veränderung des Prozesses bzw. eine Inhomogenität der Daten. Bei uns ist x_n beispielsweise der Niederschlag in einem bestimmten Monat, z.B. Januar, des Jahres n. Die Cusum Charts in der vorliegenden Arbeit sind mit den S-PLUS-Befehlen `qcc`, `cusum` und der Option `xbar` erstellt worden.

In keinem Monat wird eine der Entscheidungsgrenzen gekreuzt, so daß es keinen Hinweis auf Inhomogenitäten in den Daten gibt.

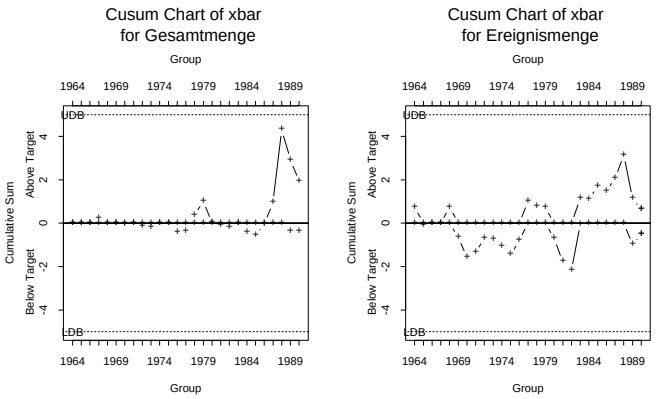
Januar



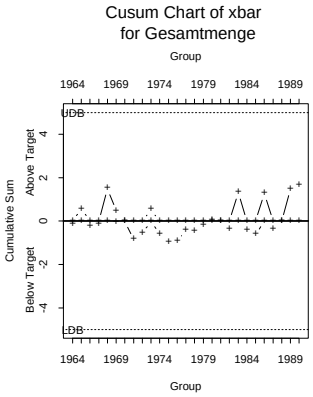
Februar

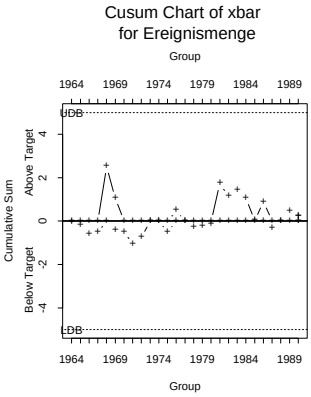


März

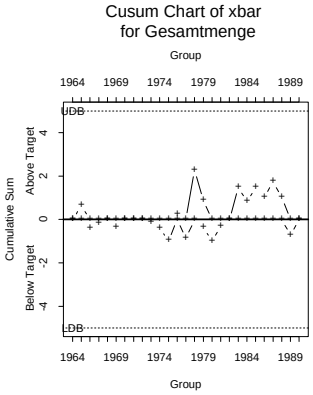


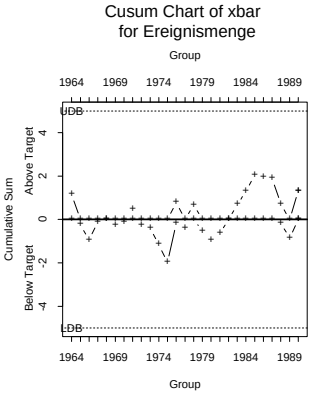
April



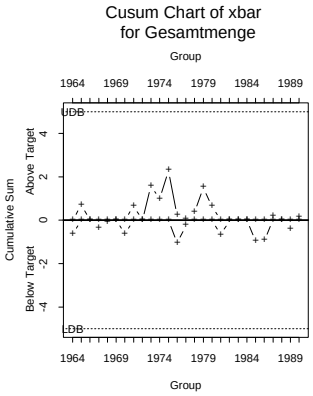


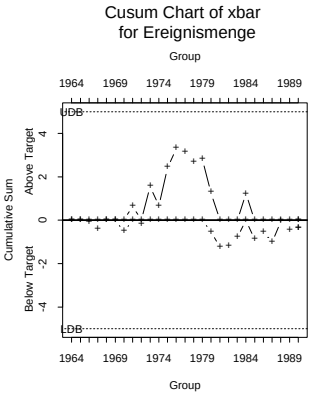
Mai



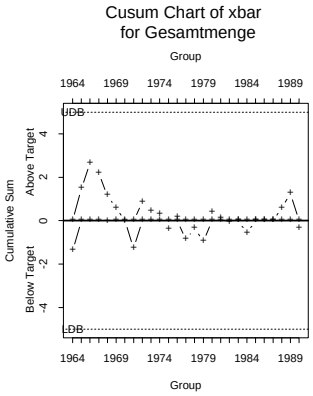


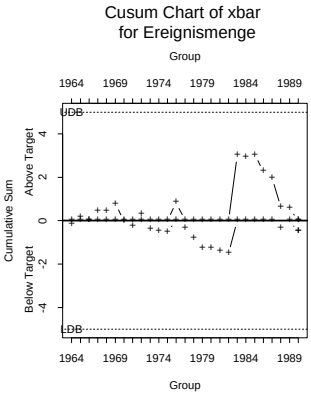
Juni



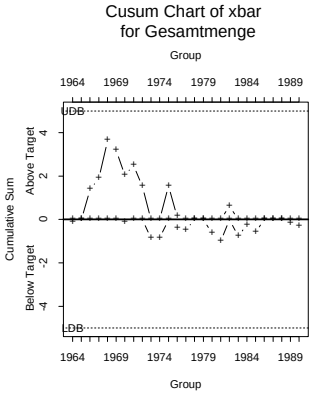


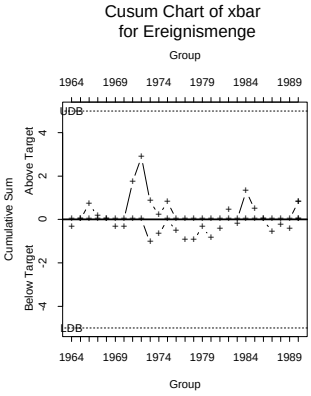
Juli



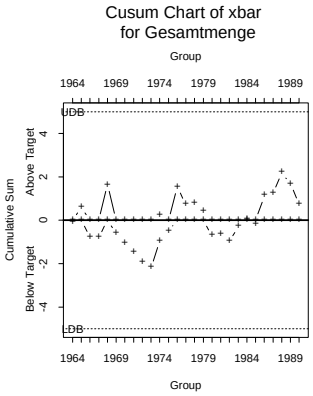


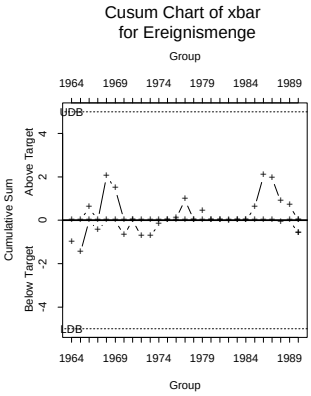
August



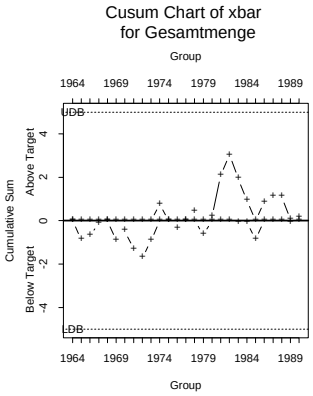


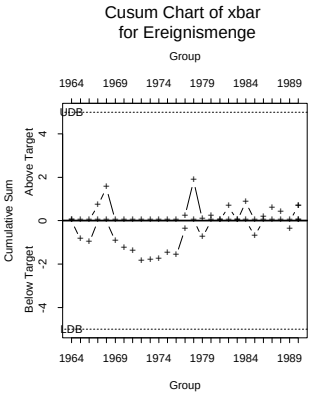
September



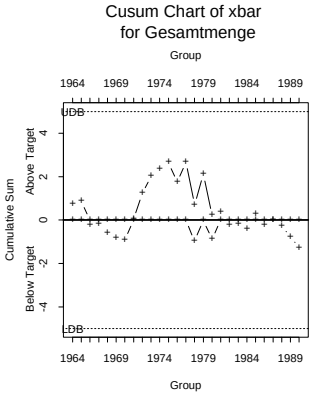


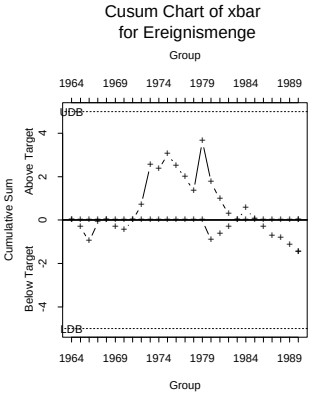
Oktober



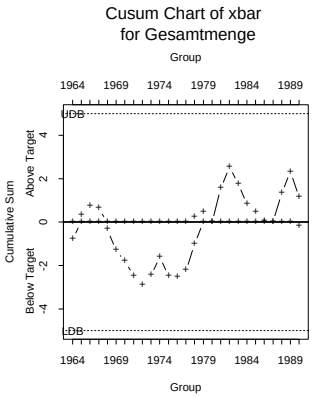


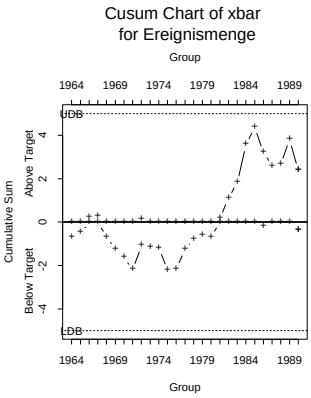
November





Dezember





Anhang B

Täglicher Niederschlag

1 Der Ereignisprozeß

Die partielle Autokorrelationsfunktion der beobachteten Serie trockener und nasser Tage läßt vermuten, daß eine Markovkette erster oder zweiter Ordnung zur Beschreibung des Ereignisprozesses adäquat ist. Um die Entscheidung für eines dieser Modelle eindeutig mit dem AIC zu treffen und die Auswirkung einer niedrigeren oder höheren Ordnung besser beurteilen zu können, werden auch ein Modell mit unabhängigen Ereignissen und eine Markovkette 3. Ordnung angepaßt. Aus dem Ereignisprozeß täglichen Regens werden die Parameter dieser Modelle geschätzt und durch Fouriersummen geglättet.
Wie in Kapitel 4.1 beschrieben, werden die Logits der Übergangswahrscheinlichkeiten $\pi = \pi_t$ der Markovketten durch Fouriersummen approximiert:

$$logit\Big(\pi(t)\Big) \approx R_0 + \sum_{j=1}^{L} R_j \cos(\omega_j t + \theta_j),\ \omega_j = 2\pi j/T\ , t = 1 \ldots, T = 365\ .$$

Es liegen also verallgemeinerte lineare Modelle vor. Das Anpassen von p Fourierfrequenzen führt zu $2p+1$ Parametern. Zwischen Modellen M_q und M_p mit Frequenzen $\omega_j, j = 1, \ldots, q$ bzw. $j = 1, \ldots, p$, $q < p$, und Deviancen D_{2q+1} bzw. D_{2p+1} wird mit Hilfe der Deviance–Differenzen $D_{2q+1} - D_{2p+1}$ ausgewählt. Beschreiben beide Modelle die Beobachtungen angemessen, so gilt approximativ $D_{2q+1} - D_{2p+1} \sim \chi^2_{2p-2q}$. In Tabelle B.1 werden die Werte der Deviance–Differenzen Δ von Modellen mit p und $p+1$ Fourierfrequenzen sowie deren Wahrscheinlichkeit (1−p-Wert) unter der χ^2_2 -Verteilung genannt.

(1.1) Parameteranzahl der Markovkettenmodelle.
Zur Auswahl der Ordnung der Markovkette mit dem AIC oder dem BIC benötigen wir die Anzahl der Modellparameter der jeweiligen Markovkette:

Ordnung	0	1	2	3
Parameteranzahl	5	8	10	12

Tabelle B.1: Modellierung der Übergangswahrscheinlichkeiten

		$D_0 - D_3$	$D_3 - D_5$	$D_5 - D_7$	$D_7 - D_9$	Freq.
$\pi_1(t)$	Δ	38.710	15.497	1.848	9.390	2
	1− p-Wert	1.000	0.999	0.603	0.991	
$\pi_{11}(t)$	Δ	7.308	2.252	1.771	1.057	1
	1− p-Wert	0.974	0.676	0.587	0.410	
$\pi_{01}(t)$	Δ	25.415	10.862	2.052	9.259	2
	1− p-Wert	1.000	0.996	0.642	0.990	
$\pi_{111}(t)$	Δ	8.976	0.163	1.507	1.6296	1
	1− p-Wert	0.989	0.078	0.529	0.557	
$\pi_{011}(t)$	Δ	1.652	0.412	1.100	3.303	0
	1− p-Wert	0.562	0.186	0.423	0.808	
$\pi_{101}(t)$	Δ	1.310	0.719	0.916	2.316	0
	1− p-Wert	0.481	0.302	0.367	0.686	
$\pi_{001}(t)$	Δ	19.268	9.716	0.234	15.714	2
	1− p-Wert	1.000	0.992	0.110	1.000	
$\pi_{1111}(t)$	Δ	3.314	0.939	0.316	1.933	0
	1− p-Wert	0.809	0.375	0.146	0.620	
$\pi_{1101}(t)$	Δ	0.300	0.629	2.289	1.427	0
	1− p-Wert	0.139	0.270	0.682	0.510	
$\pi_{1011}(t)$	Δ	0.860	0.664	2.000	2.367	0
	1− p-Wert	0.349	0.282	0.633	0.694	
$\pi_{0111}(t)$	Δ	3.885	0.778	5.576	0.171	0
	1− p-Wert	0.857	0.322	0.939	0.082	
$\pi_{1001}(t)$	Δ	2.786	1.571	0.818	0.122	0
	1− p-Wert	0.752	0.544	0.336	0.059	
$\pi_{0101}(t)$	Δ	0.549	1.225	0.735	0.453	0
	1− p-Wert	0.240	0.458	0.308	0.203	
$\pi_{0011}(t)$	Δ	0.811	0.621	1.438	5.587	0
	1− p-Wert	0.333	0.267	0.513	0.939	
$\pi_{0001}(t)$	Δ	25.252	11.376	0.708	15.190	2
	1− p-Wert	1.000	1.000	0.298	1.000	

Die Spalte *Freq.* nennt die Anzahl der angepaßten Fourierfrequenzen.

(1.2) Parameterschätzer der Markovkettenmodelle.
Da sich die Werte der Modellauswahlkriterien der Markovketten erster und zweiter Ordnung nur geringfügig unterscheiden, untersuchen wir in der Modellvalidation beide Modelle. Die Werte der Parameterschätzer in den Fouriersummenapproximationen der Übergangswahrscheinlichkeiten sind:

	$\hat{R}_0$	$\hat{R}_1$	$\hat{\theta}_1$	$\hat{R}_2$	$\hat{\theta}_2$	Freq.
π_{11}	0.2233	0.1362	-1.9480			1
π_{01}	-1.0737	0.1719	-0.4140	0.2010	-3.1286	2
π_{111}	0.4396	0.2469	-2.4213			1
π_{011}	0.2291					0
π_{101}	-0.4375					0
π_{001}	-1.2494	0.5014	-0.1315	0.1437	0.9559	2

2 Die Niederschlagsmengen an nassen Tagen

An die Niederschlagsmengen an nassen Tagen wird eine Verteilung mit der Methode der Momente unter Verwendung des 1. Momentes und des Variationskoeffizienten angepaßt. Die Mengeneinheit beträgt stets [1mm].

Wir nehmen an, der erwartete Niederschlag $m = m_t$ an nassen Tagen verhalte sich im Zeitverlauf gemäß einer Fouriersumme mit L Frequenzen:

$$\textbf{(2.1)} \quad m_t = R_0 + \sum_{j=1}^{L} R_j \cos(\omega_j t + \theta_j), \quad t = 1, \ldots, T.$$

Die Parameter $R_j, \theta_j, j = 0, \ldots, L$, werden durch Anpassen einer Fouriersumme an die empirischen Mittel $\hat{m}_t$ geschätzt, die Schätzer seien $\hat{m}_t(L)$.

(2.2) Die Schätzer der Koeffizienten sind:

L	$\hat{R}_0$	$\hat{R}_1$	$\hat{\theta}_1$	$\hat{R}_2$	$\hat{\theta}_2$	$\hat{R}_3$	$\hat{\theta}_3$
0	4.208						
1	4.173	1.143	-2.860				
2	4.160	1.124	-2.862	0.339	-0.412		
3	4.159	1.129	-2.861	0.350	-0.380	0.219	-0.176

Die aus diesen Schätzern resultierende Summen S der Residuenquadrate mit

$$\textbf{(2.3)} \quad S = \sum_{t=1}^{T} \left(\hat{m}_t(L) - \hat{m}_t\right)^2, \quad t = 1, \ldots, T.$$

sind:

Freq.	0	1	2	3
S	1377.234	1156.937	1137.667	1132.238

Die Anzahl L der anzupassenden Frequenzen soll ohne Verteilungsannahme ausgewählt werden. Damit kann keines der Modellauswahlkriterien wie das AIC oder

die Deviance angewendet werden. Durch eine Frequenz wird eine gute Anpassung erreicht, weitere Frequenzen bewirken nur eine marginale Verbesserung. Der optische Eindruck wird durch die Summe der Residuenquadrate bestätigt, die bei der Anpassung einer Frequenz stark abnimmt, beim Anpassen weiterer Frequenzen nur noch wenig, vergl. Abb. B.1. Wir entscheiden uns also für $L = 1$.
Zur Modellauswahl könnte auch ein Auswahlkriterium Υ konstruiert werden, wie z.B. Die Auswahl könnte auch formaler auf ein Modellauswahlkriterium Υ basiert werden, z.B.

$$\Upsilon(L) = \widetilde{SQ}(L) + 2L + 1 \ ,$$

wobei $\widetilde{S}$ die normierte Residuenquadratsumme

$$\widetilde{S}(L) = \sum_{t=1}^{T} \left(1 - \frac{\hat{m}_t(L)}{m_t}\right)^2$$

ist, um den Einfluß der Mengeneinheit der Niederschlagsmessungen zu eliminieren. Beispielsweise wurden die Daten in der Auflösung von [0.1mm] registriert, während üblicherweise die Einheit [1mm] verwendet wird. Gewählt wird das Modell mit dem kleinsten Wert des Kriteriums. Wir erhalten:

L	0	1	2	3	4
Υ	78.78	71.90	72.44	75.41	76.57

Damit wird ebenfalls $L = 1$ gewählt. Der Variationskoeffizient wird aus $\hat{m}_t(1)$, basierend auf 4.(2.5) geschätzt:

(2.4) $\widehat{CV} = 1.331648$.

Mit der in Kapitel 4.2 genannten Approximation für den Formparameter β erhalten wir

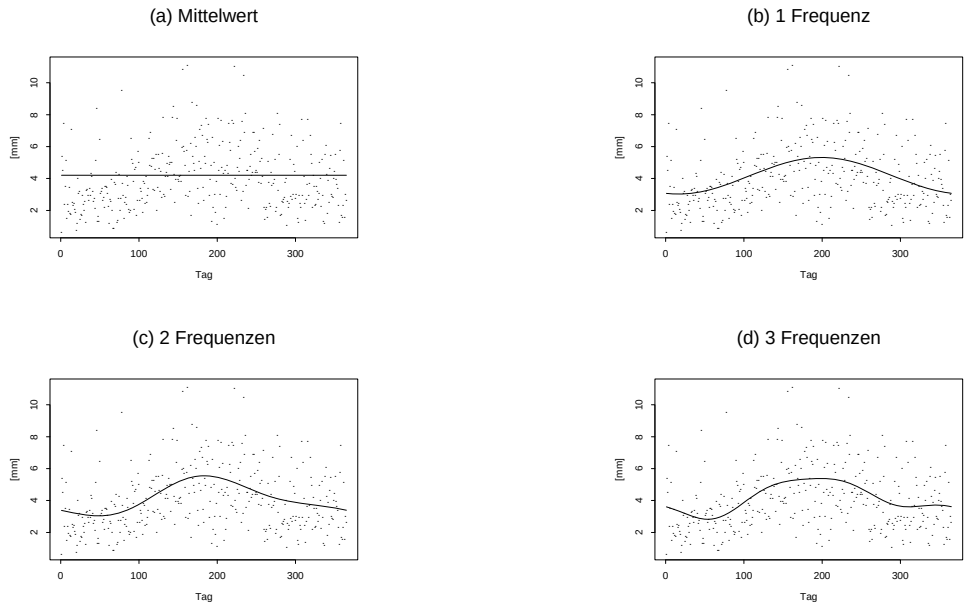
(2.5) $\hat{\beta} = 0.7561838$.

(2.6) Parameterschätzer der Weibullverteilung
Wir passen eine Weibullverteilung mit zeitabhängigem Skalenparameter

$$\hat{\alpha}_t(1) = \hat{m}_t(1)/\Gamma(1 + 1/\hat{\beta})$$

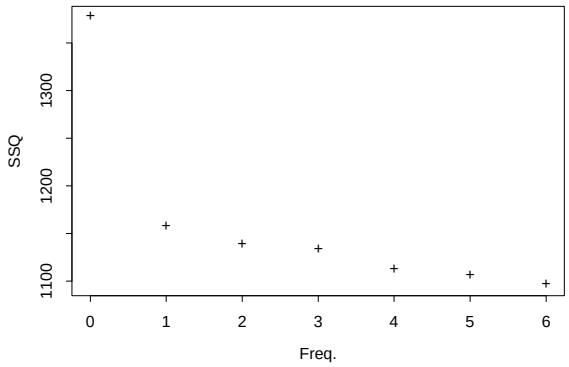
mit $\hat{m}_t(1)$ wie in (2.1) mit den in (2.2) genannten Parametern und konstantem Formparameter $\hat{\beta} = 0.7561838$ an.

Abbildung B.1: Mittlerer Niederschlag an nassen Tagen



Empirische Mittel $\hat{m}_t$ der Niederschlagsmenge an nassen Tagen t und Approximation durch (a) Mittelwert bzw. durch Fouriersummen mit Freq. (b) $\omega_j, j = 1$, (c) $\omega_j, j = 1, 2$, (d) $\omega_j, j = 1, 2, 3$.

Residuenquadrate



Residuenquadratsumme S (2.3) zwischen Mittelwerten $\hat{m}_t$ und Schätzern $\hat{m}_t(L)$ in der saisonalen Modellierung des bedingten Niederschlagsmittels.

---

3 Modellvalidation

Zur Modellvalidation werden 1000 Jahre täglicher Niederschlagsdaten simuliert, wie in Kapitel 4.3 beschrieben. Die zu den hier aufgeführten Tabellen gehörenden Graphiken befinden sich in Kapitel 4.4.

Im folgenden bezeichne OBS die Beobachtungen, M1 die mit einer Markovkette 1. Ordnung und M2 die mit einer Markovkette 2. Ordnung erzeugten Datensätze. Die Niederschlagsmengen an nassen Tagen wurden jeweils mit einer Weibullverteilung mit den im vorigen Abschnitt gegebenen Parametern erzeugt, also mit konstantem Form- und zeitabhängigem Skalenparameter.

Tabelle B.2: Monats- und Jahresniederschlag

Mittel

Monat	OBS	M1	M2
Jahr	563.7	565.6	589.4
Jan.	33.6	32.2	32.3
Feb.	33.3	32.7	34.1
März	33.3	36.3	38.5
April	36.8	40.3	42.0
Mai	42.8	45.1	46.1
Juni	48.4	48.3	49.0
Juli	48.8	50.0	51.2
Aug.	50.6	50.4	52.3
Sept.	49.8	50.0	52.0
Okt.	48.3	49.4	51.4
Nov.	47.8	48.3	50.3
Dez.	47.0	47.1	49.1

Standardabweichung

Monat	OBS	M1	M2
Jahr	111.0	85.0	86.7
Jan.	34.2	16.9	17.2
Feb.	28.7	17.6	18.8
März	28.4	20.0	21.8
April	29.6	22.7	24.1
Mai	33.0	26.1	26.5
Juni	35.1	27.8	28.0
Juli	33.3	28.3	28.7
Aug.	33.7	28.2	28.8
Sept.	32.5	27.9	28.6
Okt.	32.0	27.6	28.2
Nov.	31.4	27.2	27.9
Dez.	30.9	26.8	27.5

Mittel (links) und Standardabweichung (rechts) des Jahres- und Monatsniederschlages in den Beobachtungen sowie den simulierten Daten.

Tabelle B.3: Anzahl nasser Tage

Mittel

Monat	OBS	M1	M2
Jahr	134.0	123.1	128.3
Jan.	10.9	9.3	9.7
Feb.	10.5	8.8	9.3
März	10.6	9.1	9.7
April	10.9	9.6	10.0
Mai	11.5	10.1	10.4
Juni	11.8	10.4	10.6
Juli	11.7	10.5	10.8
Aug.	11.6	10.5	10.9
Sept.	11.4	10.4	10.8
Okt.	11.2	10.3	10.7
Nov.	11.2	10.3	10.7
Dez.	11.2	10.3	10.7

Standardabweichung

Monat	OBS	M1	M2
Jahr	19.8	11.6	12.5
Jan.	4.3	3.3	3.6
Feb.	4.3	3.3	3.6
März	4.3	3.3	3.7
April	4.4	3.4	3.7
Mai	4.3	3.6	3.7
Juni	4.3	3.6	3.7
Juli	4.1	3.6	3.7
Aug.	4.1	3.6	3.6
Sept.	4.2	3.5	3.6
Okt.	4.2	3.5	3.6
Nov.	4.2	3.5	3.7
Dez.	4.2	3.5	3.7

Mittel (links) und Standardabweichung (rechts) der Anzahl nasser Tage im Jahr bzw. pro Monat in den Beobachtungen sowie den simulierten Daten.

Tabelle B.4: Niederschlag an nassen Tagen

Mittel

Monat	OBS	M1	M2
Jahr	3.9	4.6	4.6
Jan.	3.1	3.5	3.3
Feb.	3.3	4.0	4.0
März	3.1	4.5	4.5
April	4.0	4.8	4.8
Mai	4.8	5.2	5.2
Juni	5.6	5.4	5.4
Juli	4.6	5.3	5.4
Aug.	5.6	5.2	5.3
Sept.	4.6	4.9	5.0
Okt.	3.6	4.5	4.5
Nov.	3.8	3.8	3.9
Dez.	3.6	3.4	3.3

Standardabweichung

Monat	OBS	M1	M2
Jahr	5.4	5.8	5.9
Jan.	5.6	4.1	4.0
Feb.	4.4	4.9	4.9
März	4.3	5.5	5.5
April	5.3	6.1	6.1
Mai	6.9	6.6	6.6
Juni	7.1	6.7	6.8
Juli	5.9	6.7	6.6
Aug.	6.7	6.6	6.5
Sept.	5.2	6.1	6.4
Okt.	4.8	5.6	5.6
Nov.	5.1	4.8	4.8
Dez.	5.0	4.1	4.0

Mittel (links) und Standardabweichung (rechts) der Niederschlagsmenge an nassen Tagen in den Beobachtungen sowie den simulierten Daten.

Tabelle B.5: Niederschlag an nassen Tagen - Quantile

		50%	75%	90%	95%	98%	99%
Jahr	OBS	2.2	5.3	10.1	14.7	21.7	28.7
	M1	2.6	5.9	11.2	15.8	22.3	27.9
	M2	2.5	5.9	11.3	16.0	22.7	28.1
Jan.	OBS	1.6	3.5	7.0	9.5	14.6	18.4
	M1	2.0	4.5	8.5	11.6	16.2	19.8
	M2	1.9	4.4	8.1	11.0	15.7	18.7
Feb.	OBS	1.8	4.4	7.5	9.3	14.5	18.4
	M1	2.3	5.2	9.8	13.5	19.0	23.4
	M2	2.2	5.1	9.9	13.6	18.7	23.2
März	OBS	1.8	4.1	7.1	9.7	12.2	15.9
	M1	2.5	5.8	10.8	15.2	22.1	27.2
	M2	2.5	5.8	11.1	15.5	21.9	26.0
April	OBS	2.2	5.1	9.2	15.4	22.5	28.0
	M1	2.7	6.2	11.7	16.3	23.2	29.5
	M2	2.7	6.1	11.9	16.5	23.4	29.0
Mai	OBS	2.6	6.4	10.7	14.7	25.6	37.2
	M1	3.0	6.7	13.0	17.9	25.1	31.2
	M2	2.9	6.8	12.9	18.3	26.0	31.5
Juni	OBS	3.0	7.6	13.6	18.5	26.7	39.2
	M1	3.0	7.0	13.1	18.5	25.8	32.9
	M2	3.0	6.9	13.5	19.1	26.6	32.5
Juli	OBS	2.6	5.6	10.5	16.6	25.8	32.3
	M1	3.0	6.8	13.1	18.6	25.8	32.1
	M2	3.1	7.1	13.0	18.1	26.0	32.4
Aug.	OBS	3.3	7.8	13.3	16.6	21.8	28.1
	M1	2.9	6.8	12.8	17.6	25.2	31.7
	M2	3.0	6.8	12.9	18.4	25.8	31.2
Sept.	OBS	2.6	6.0	11.5	14.6	21.2	23.2
	M1	2.8	6.4	12.1	16.7	23.6	28.2
	M2	2.8	6.4	12.1	17.3	24.0	29.7
Okt.	OBS	1.8	4.0	8.2	14.6	20.1	23.2
	M1	2.6	5.9	11.0	15.4	21.7	26.3
	M2	2.5	5.8	11.0	15.3	22.2	26.9
Nov.	OBS	2.0	4.4	9.0	14.3	20.5	23.9
	M1	2.2	4.9	9.3	13.1	17.9	21.8
	M2	2.2	5.1	9.5	13.5	18.9	23.0
Dez.	OBS	1.8	4.7	9.2	11.6	17.7	23.4
	M1	1.9	4.3	8.2	11.4	15.7	19.2
	M2	1.9	4.3	8.1	11.3	15.8	19.8

4 S-PLUS-Routinen

Simulation einer Markovkette 1. Ordnung

Es werden `N` Jahre mit `Tage` Tagen täglicher Niederschlagsmessungen einer Markovkette 1. Ordnung mit der S-PLUS-Routine `sim.markov1.fun` simuliert. Die Niederschlagsmengen an nassen Tagen sind weibullverteilt. Die Vorgehensweise folgt Kapitel 4.4. Je m Jahre werden in eine Datei geschrieben. Ist $N = Km + R$ mit $R < m$, so werden K Dateien mit den simulierten Daten von m Jahren und eine Datei mit den simulierten Daten von R Jahren erzeugt. Input sind Vektoren der Übergangswahrscheinlichkeiten `pww`$=\pi_{11}(t)$ und `pwd`$=\pi_{01}(t)$, $t = 1, \ldots,$ `Tage`, sowie die Parameter der Weibullverteilung, `shape` und `scale`. Letzterer muß ein Vektor der Länge `Tage` sein.

```
sim.markov1.menge.fun <-
function(N, Tage = 365, pww , pwd , scale, shape,  m = 100, pfad )
{
# generate n years with Tage days of a non-stationary two-state-markov-chain
# pww= P(W|W), pwd=P(W|D)
# weibull distributed rainfall depths with scale=scale (time-dependent) and
# shape=shape(constant)
      if(length(pww) != Tage) {
            cat("Falsche Spaltenanzahl! \n")
            break
      }
      K <- N %/% m
      R <- N %% m
# generate output file names
      nummer <- seq(1:(K + 1))
      datei <- paste("mark1_", nummer, ".txt", sep = "")
      datei <- paste(pfad, datei, sep = "")
# compute start distribution on first day: delta =(P(D),P(W))
      delta <- c(1 - pww[1], pwd[1])/(1 - pww[1] + pwd[1])
      starttag <- ifelse(runif(1) < delta[2], 1, 0)
      if(N >= m) {
# generate files k=1, ... ,K
            for(k in 1:K) {
# generate rainfall occurrences in years 1, .., K*m
                  rain <- rep(0, times = m * Tage)
# special case: first day in first year
                  if(k == 1) {
                        rain[1] <- starttag
                  }
                  else {
                        if(starttag == 1)
                          rain[1] <- ifelse(runif(1) <= pww[1], 1, 0)
                        else rain[1] <- ifelse(runif(1) <= pwd[1], 1, 0)
                  }
# generate all following days i=2, 3, ..
                  for(i in 2:(m * Tage)) {
                        j <- ifelse((i %% Tage) != 0, i %% Tage, Tage)
                        if(rain[i - 1] == 1) {
                          if(runif(1) < pww[j])
                            rain[i] <- 1
                        }
                        if(rain[i - 1] == 0) {
                          if(runif(1) < pwd[j])
                            rain[i] <- 1
                        }
                  }
# Ende i-Schleife
# save state on last day in year K*m
```

```
                starttag <- rain[m * Tage]
                rain <- matrix(rain, ncol = Tage, byrow = T)
# generate rainfall depths
                Nw <- apply(rain, 2, sum)      # number of wet days
                for(i in 1:Tage) {
                      rain[, i][rain[, i] == 1] <- round(rweibull(Nw[i],
                           shape=shape,scale=scale[i]),1)
                }
                rain[rain < 0.2] <- 0               # rainfall threshold
                write.table(rain,file=datei[k],sep ="",dimn=F,append = F)
                cat("Schreibe Datei ", k, "\n")
            } # end k-loop
      } # end if(N>=m)
# generate last R years
      if(R != 0) {
            rain <- rep(0, times = R * Tage) # generate rainfall occurrences
            if(starttag == 1) # first day
                  rain[1] <- ifelse(runif(1) <= pww[1], 1, 0)
            else rain[1] <- ifelse(runif(1) <= pwd[1], 1, 0)
            for(i in 2:(R * Tage)) { # following days
                  j <- ifelse((i %% Tage) != 0, i %% Tage, Tage)
                  if(rain[i - 1] == 1) {
                        if(runif(1) < pww[j])
                          rain[i] <- 1
                  }
                  if(rain[i - 1] == 0) {
                        if(runif(1) < pwd[j])
                          rain[i] <- 1
                  }
            }
            rain <- matrix(rain, ncol = Tage, byrow = T)
# generate rainfall deoths on wet days
            if(R == 1) {
                  Nw <- sum(rain == 1)
                  rain[rain == 1] <- round(rweibull(Nw,shape=shape,
                                                    scale=scale),1)
            }
            else {
                  Nw <- apply(rain, 2, sum)
                  for(i in 1:Tage) {
                        rain[, i][rain[, i] == 1] <- round(rweibull(Nw[i],
                                      shape = shape, scale = scale[i]), 0)
                  }
            }
            rain[rain < 0.2] <- 0
            write.table(rain, file = datei[K + 1], sep = " ", dimn = F, append = F)
      }
# Ende if(R>0)
}
```

Simulation einer Markovkette 2. Ordnung

Die Simulation von N Jahren täglicher Niederschlagsmessungen gemäß einer Markovkette 2. Ordnung erfolgt mit der S-PLUS-Routine `sim.mark2.menge.fun`. Die Berechnung der Gleichgewichtswahrscheinlichkeiten der Markovkette 2. Ordnung geschieht in der S-PLUS-Funktion `markov2.eq.fun`. Die Vorgehensweise folgt Kapitel 4.4.

Gleichgewichtswahrscheinlichkeiten

Die Gleichgewichtswahrscheinlichkeiten einer nichtstationären Markovkette 2. Ordnung mit 2 Zuständen für Tag $t = 1$ werden unter der Annahme bestimmt, die Übergangswahrscheinlichkeiten von Tag $t = 365$ zu Tag $t = 1$ seien konstant. Input der Funktion `markov2.eq.fun` sind die Übergangswahrscheinlichkeiten `pwww` =

$\pi_{111}(t)$, `pwwd` $= \pi_{011}(t)$, `pwdw` $= \pi_{101}(t)$, `pwdd` $= \pi_{001}(t)$. Ausgegeben wird $\delta^T = (\mathrm{P}_{(00)}, \mathrm{P}_{(10)}, \mathrm{P}_{(01)}, \mathrm{P}_{(11)})$ mit $\mathrm{P}_{(ij)} = \mathrm{P}(Y_t = (i,j)) = \mathrm{P}(Z_t = i, Z_{t+1} = j)$.

```
markov2.eq.fun <- function(pwww , pwwd, pwdw,  pwdd )
{
# Compute the equilibrium probabilities of a two-state second-order  markov chain
# transition probabilities:
# pwww=P(W|WW), pwwd =P(W|WD), pwwd=P(W|WD), pwdw=P(W|DW), pwdd=P(W|D),
        delta <- rep(0, times = 4)        # equilibrium prob of Y=(Z_T, Z_1)
        # transition matrix of Y
        P <- matrix(c(1 - pwdd, 0, pwdd, 0, 1 - pwdw, 0, pwdw, 0, 0, 1 - pwwd,
                0, pwwd, 0, 1 - pwww, 0, pwww), nrow = 4, ncol = 4, byrow = T)
        p1 <- 1 - P[1, 1]
        p2 <- p1 - P[2, 2] - P[1, 1] * P[2, 2] + P[1, 2] * P[2, 1]
        del2.3 <- P[3, 2] * p1 + P[1, 2] * P[3, 1]
        del2.4 <- P[4, 2] * p1 + P[1, 2] * P[4, 1]
        D3.zaehler <- (P[2, 3] * p1 + P[1, 3] * P[2, 1]) *
                        (P[4, 2] * p1 + P[1, 2] * P[4, 1]) +
                P[4, 3] * p1 * p2
        D3.nenner <- p1 * p2 * (1 - P[3, 3]) - (P[2, 3] * p1 + P[1, 3] * P[2, 1]) *
                            (P[3, 2] * p1 + P[1, 2] * P[3, 1])
        D3 <- D3.zaehler/D3.nenner
        D2.3 <- P[3, 2] * p1 + P[1, 2] * P[3, 1]
        D2.4 <- P[4, 2] * p1 + P[1, 2] * P[4, 1]
        D1.2 <- P[2, 1]/p1
        D1.3 <- P[3, 1]/p1
        D1.4 <- P[4, 1]/p1
        D4 <- 1 + D3 * (1 + (D2.3 * (1 + D1.2))/p2 + D1.3) +
                            (D2.4 * (1 + D1.2))/p2 + D1.4
        delta[4] <- D4^-1
        delta[3] <- D3 * delta[4]
        delta[2] <- (delta[3] * D2.3 + delta[4] * D2.4)/p2
        delta[1] <- sum((P[-1, 1] * delta[-1])/p1)
        names(delta) <- c("PDD", "PWD", "PDW", "PWW")
        prob.x <- c(delta[1] + delta[2], delta[3] + delta[4])
        # equilibrium prob of X P(X=0), P(X=1)
        names(prob.x) <- NULL
        names(delta) <- NULL
        ergebnis <- list(PDD = delta[1], PWD = delta[2], PDW = delta[3],
                    PWW = delta[4], PD = prob.x[1], PW = prob.x[2])
        return(ergebnis)
}
```

Simulation des Ereignisprozesses und der Mengen

Es werden `N` Jahre mit einer Anzahl `Tage` an Tagen simuliert. Input sind die Vektoren der Übergangswahrscheinlichkeiten, aus denen auch die benötigten Startwahrscheinlichkeiten berechnet werden. Je m simulierte Jahre werden in eine Datei geschrieben. Bei $N = K \cdot m + R$ mit $R < m$ müssen $K + 1$ Dateien angelegt werden, deren Dateinamen `mark2`$_k$`.txt` mit $k = 1, \ldots, K + 1$ sind. Zunächst wird der Ereignisprozeß in den Jahren $k = 1, \ldots K \cdot m$ simuliert, dann die Niederschlagsmenge an den nassen Tagen. Die Niederschlagsmengen an nassen Tagen sind weibullverteilt mit Formparameter `shape` und Skalenparameter `scale`, wobei `scale` ein Vektor der Länge `Tage` sein muß. Die Mengeneinheit der simulierten Daten ist [1mm].

```
sim.mark2.menge.fun  <-
function(N = 1000, Tage = 365, pwww, pwwd,  pwdw, pwdd, scale, shape, m = 100, pfad)
{
## generate N years with Tage days of a non-stationary two-state-second-order
## markov-chain
## pwww= P(W|WW), pwdw=P(W|DW),pwwd=P(W|WD),pwdd=P(W|DD),
## equilibrium prob's: PDW=P(Y(t-1)=D,Y(t)=W)
```

```
## depths on wet days: Weibull-distributed, shape=shape(constant),
## scale=scale(time dependent)
      K <- ℕ %/% m # K+1: number of files to generate
      R <- ℕ %% m  # R: number of years in last file
# generate file names
      nummer <- seq(1:(K + 1))
      datei <- paste("mark2_", nummer, ".txt", sep = "")
      datei <- paste(pfad, datei, sep = "")
      rain <- rep(0, times = ℕ * Tage)  # vector of occurrences
# calculate start distribution
      prob.eq <- markov2.eq.fun(pwww, pwwd, pwdw, pwdd)
# generate rainfall occurrences
# generate days 1 and 2 in first year
      if(runif(1) <= prob.eq$PW) {
            tag1 <- 1
            prob <- prob.eq$PWW/prob.eq$PW
      }
      else {
            tag1 <- 0
            prob <- prob.eq$PDW/prob.eq$PD
      }
      tag2 <- ifelse(runif(1) <= prob, 1, 0)
      for(k in 1:K) {
# generate files 1 to K
# special case: days 1 and 2
            if(k == 1) { # first year
                  rain[1] <- tag1
                  rain[2] <- tag2
            }
            else { # next years
                  if(tag2 == 1) {
# tag2: Tag 365; tag1: Tag 364 Vorjahr,
                      if(tag1 == 1) rain[1] <- ifelse(runif(1) <= pwww[1], 1, 0)
                      else rain[1] <- ifelse(runif(1) <= pwwd[1], 1, 0)
                      if(rain[1] == 1)
                          rain[2] <- ifelse(runif(1) <= pwww[2], 1, 0)
                      else rain[2] <- ifelse(runif(1) <= pwdw[2], 1, 0)
                  }
                  else {
                      if(tag1 == 1)  rain[1] <- ifelse(runif(1) <= pwdw[1], 1, 0)
                      else rain[1] <- ifelse(runif(1) <= pwdd[1], 1, 0)
                      if(rain[1] == 1) rain[2] <- ifelse(runif(1) <= pwwd[2], 1, 0)
                      else rain[2] <- ifelse(runif(1) <= pwdd[2], 1, 0)
                  }
            }
# generate all following days i=3,4, ...
            for(i in 3:(m * Tage)) {
                  j <- ifelse((i %% Tage) == 0, Tage, i %% Tage)
                      if(rain[i - 2] == 1) {
                          if(rain[i - 1] == 1)
                                 rain[i] <- ifelse(runif(1) <= pwww[j], 1, 0)
                          else rain[i] <- ifelse(runif(1) <= pwdw[j], 1, 0)
                      }
                      else { # rain[i-2]==0
                          if(rain[i - 1] == 1) rain[i] <- ifelse(runif(1) <= pwwd[j], 1, 0)
                          else rain[i] <- ifelse(runif(1) <= pwdd[j], 1, 0)
                      }
            }
            tag1 <- rain[m * Tage - 1]        # save states on days T-1, T for next year
            tag2 <- rain[m * Tage]
            rain <- matrix(rain, ncol = Tage, byrow = T)
# generate rainfall depths on wet days
            ℕw <- apply(rain, 2, sum)        # number of wet days
            for(i in 1:Tage) {
                  rain[, i][rain[, i] == 1] <- round(
                          rweibull(Tagew[i], shape = shape, scale = scale[i]), 1)
            }
            rain[rain < 0.2] <- 0 # rainfall threshold used in observations
            write.table(rain, file = datei[k], sep = " ", dimn = F, append = F)
            cat("Schreibe Datei ", k, "\n")
      } # end of k-loop
```

```
# generate last file
      if(R != 0) {
            rain <- rep(0, times = R * Tage)
# special case: days 1 and 2
            if(tag2 == 1) {
                  if(tag1 == 1)
                        rain[1] <- ifelse(runif(1) <= pwww[1], 1, 0)
                  else rain[1] <- ifelse(runif(1) <= pwwd[1], 1, 0)
                        if(rain[1] == 1)
                                rain[2] <- ifelse(runif(1) <= pwww[2], 1, 0)
                        else rain[2] <- ifelse(runif(1) <= pwdw[2], 1, 0)
            }
            else {
                  if(tag1 == 1) rain[1] <- ifelse(runif(1) <= pwdw[1], 1, 0)
                  else rain[1] <- ifelse(runif(1) <= pwdd[1], 1, 0)
                  if(rain[1] == 1) rain[2] <- ifelse(runif(1) <= pwwd[2], 1, 0)
                  else rain[2] <- ifelse(runif(1) <= pwdd[2], 1, 0)
           }
# generate following days i=3,4 ..
           for(i in 3:(R * Tage)) {
                  j <- ifelse((i %% Tage) == 0, Tage, i %% Tage)
                  if(rain[i - 2] == 1) {
                        if(rain[i - 1] == 1)
                              rain[i] <- ifelse(runif(1) <= pwww[j], 1, 0)
                         else rain[i] <- ifelse(runif(1) <= pwdw[j], 1, 0)
                  }
                  else {   # rain[i-2]==0
                         if(rain[i - 1] == 1)
                              rain[i] <- ifelse(runif(1) <= pwwd[j], 1, 0)
                         else rain[i] <- ifelse(runif(1) <= pwdd[j], 1, 0)
                  }
            }
            rain <- matrix(rain, ncol = Tage, byrow = T)
# generate rainfall depths on wet days
            Nw <- apply(rain, 2, sum)  # number of wet days
            if(R == 1) {
# special case: last file contains one line
                  Nw <- sum(rain)
                  rain[rain == 1] <- round(rweibull(Nw,shape=shape,scale=scale), 1)
            }
            else {
                  for(i in 1:Tage) {
                         rain[, i][rain[, i] == 1] <- round(
                                    rweibull(Nw[i],shape=shape,scale=scale[i]), 1)
                  }
           }
           rain[rain < 0.2] <- 0
           write.table(rain, file = datei[K + 1], sep = " ", dimn = F, append = F)
      }# end if(R>0)
}
```

Anhang C

Das Neyman–Scott–Modell

1 Modellanpassung

Die Anpassung des saisonalen Neyman-Scott-Modelles erfolgt mit der Methode der kleinsten Quadrate für alle 30 Tage umfassenden Perioden eines Jahres, die im Abstand von sechs Stunden beginnen. Die Modellanpassung geschieht mit dem C-Programm `zykfit`, das nachfolgend beschrieben wird. Zunächst wird der Programmablauf im Überblick beschrieben, bevor die verwendeten Unterroutinen im einzelnen beschrieben werden.

Minimiert wird die Summe S der Residuenquadrate,

$$S = \sum_{i=1}^{m} w_i \left(1 - \frac{f_i}{\hat{f}_i}\right)^2 ,$$

wobei $f_i = f_i(\lambda, \nu, \beta, \eta, \xi)$ Momente des aggregierten Niederschlagsprozesses, z.B. $\mathrm{Var}(Y_i^{(h)})$), $\hat{f}_i$ die entsprechenden, aus Beobachtungen geschätzten Werte dieser Eigenschaften und w_i Gewichtungsfaktoren sind. Die f_i sind Funktionen der Modellparameter $\psi := (\lambda, \nu, \beta, \eta, \xi)$, damit ist $S = S(\lambda, \nu, \beta, \eta, \xi)$. Das Modell wird für jede der 30 Tage umfassenden, im Abstand von sechs Stunden beginnenden Perioden P_t mit Start in Stunde $t, t = 1, \ldots, 1460$, angepaßt. Die Summe der Residuenquadrate in Periode t ist damit:

$$\textbf{(1.1)} \quad S_t = \sum_{i=1}^{m} w_i \left(1 - \frac{f_{i,t}}{\hat{f}_{i,t}}\right)^2 .$$

Zur Modellanpassung werden nur einige Momente des aggregierten Niederschlagsprozesses verwendet, deren Auswahl in Kapitel 5.3 beschrieben ist. Unterschiedliche Sets an Momenten führen zu unterschiedlichen Versionen von `zykfit`. Zur Überprüfung der Modellanpassung werden für jede Periode t die Momente $\mathcal{F}_t$ des Mengenprozesses unter dem Modell berechnet,

$$\begin{aligned}\mathcal{F}_t \;:=\; & \{\hat{\mu}_t(1), \hat{\gamma}_t(1), \hat{\gamma}_t(3), \hat{\gamma}_t(6), \hat{\gamma}_t(12), \hat{\gamma}_t(24),\\ & \hat{\rho}_t(1,1), \hat{\rho}_t(3,1), \hat{\rho}_t(6,1), \hat{\rho}_t(12,1), \hat{\rho}_t(24,1),\\ & \hat{\phi}_{WW,t}(1), \hat{\phi}_{WW,t}(3), \hat{\phi}_{WW,t}(6), \hat{\phi}_{WW,t}(12), \hat{\phi}_{WW,t}(24),\\ & \hat{\phi}_{DD,t}(24), \hat{\phi}_t(24)\}\end{aligned}$$

Die damit berechnete Residuenquadratsumme sei $\tilde{S}_t$, die Gewichte sind $w_\mu = 10$, $w_i = 1$ sonst.

Die Minimierung der Residuenquadratsumme S_t erfolgt für jede Periode t numerisch mit der Downhill-Simplex-Methode von Nelder und Mead, einem mehrdimensionalen Verfahren zur Minimierung von Funktionen $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, das die Zielfunktion f nur für $x \in \mathbb{R}^n$ auswertet und ohne Ableitungen auskommt. Das Verfahren ist ausführlich in PRESS et al. [1992] beschrieben, deren *Numerical Recipes* die Minimierungsroutinen `amoeba` und `amotry` entnommen und leicht verändert worden sind. Als Startwerte werden $(n+1)$ Punkte $E_i \in \mathbb{R}^n$, die ein nichtentartetes Simplex $E := (E_1, \ldots, E_{n+1})$ bilden, benötigt. Das Simplex E wird mittels Spiegelungen, Streckungen und mehrdimensionaler Kontraktionen im $\mathbb{R}^n$ bewegt, bis es sich um einem Punkt mit minimaler Residuenquadratsumme zusammenzieht. Das Abbruchskriterium ist die relative Differenz der Funktionswerte an den Simplexecken. Zurückgegeben wird ein Simplex $D = (D_1, \ldots, D_6)$ mit $S(D_1) \leq S(D_j), j = 2, \ldots, 6$.
Wie in Kapitel 5.3 beschrieben, müssen die Modellparameter $\lambda, \beta, \eta, \xi > 0$, $\nu > 1, \eta > \beta$ erfüllen. Um keine Restriktionen beachten zu müssen, wird in der Minimierungsroutine `amoeba` mit transformierten Parametern gerechnet, und zwar mit

$$\textbf{(1.2)} \quad \tilde{\lambda} := \log\lambda, \quad \tilde{\nu} := \log(\nu - 1), \quad \tilde{\beta} := \log\beta, \quad \tilde{\delta} := \log\delta, \quad \tilde{\xi} := \log\xi\ ,$$

wobei $\delta = \eta - \beta$. Damit gilt $-\infty < \tilde{\lambda}, \tilde{\nu}, \tilde{\beta}, \tilde{\delta}, \tilde{\xi} < \infty$. Es erwies sich trotzdem als sinnvoll, in `amotry` eine Straffunktion einzubauen, die bei zu kleinen Parameterwerten eine große Zahl zum Wert der zugehörigen Residuenquadratsumme S addiert. Diese Logarithmus-Transformation der Parameter betrifft nur die Schrittweite in der Minimierungsroutine. Zur Berechnung der Momente unter dem Modell und der Residuenquadratsummen S_t und $\tilde{S}_t$ werden die Parameter wieder auf die nichtlogarithmische Größenskala zurücktransformiert.

Das Hauptprogramm
Zur Anpassung des Modelles an die erste Periode P_1 wird ein Startsimplex $E_1 = (E_{1,1}, \ldots, E_{1,6}) \in \mathbb{R}^{5\times 6}$ aus einer Datei `simplex.input` eingelesen. Für nachfolgende Perioden P_t, $t = 7, \ldots, 1460$ werden die ausgegebenen Parameter (Zielparameter) $D_{s,1}$ vorheriger Perioden s verwendet, so daß

$$\textbf{(1.3)} \quad E_t = (D_{t-1,1}, D_{t-2,1} \ldots, D_{t-6,1})\ .$$

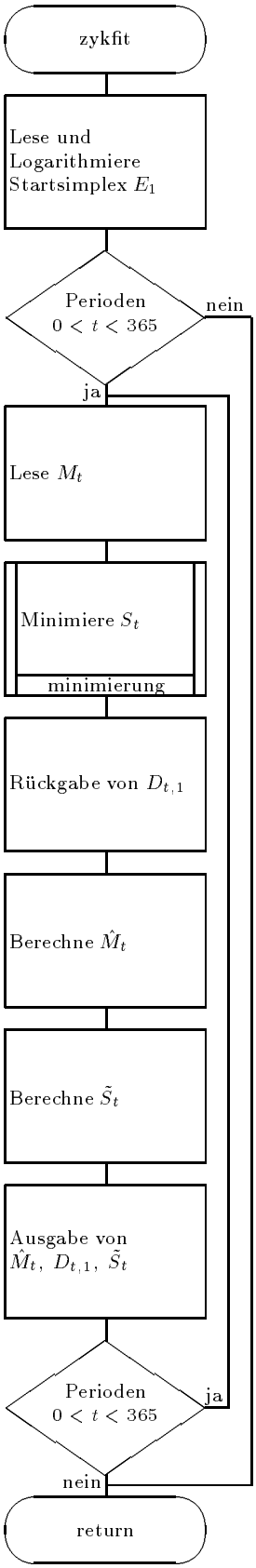
Als weitere Inputdaten werden die empirischen Momente $\hat{M}_t$ für jede Periode t des Niederschlagsprozesses aus einer Datei `momente.input` eingelesen. Sie werden in S-PLUS berechnet. Die Ausgabe der Zielparameter $D_{t,1}$ und der Momente M_t erfolgt in zwei Dateien `parameter.aus` und `momente.aus`.

Minimierung von S_t
Zur Minimierung von S_t ist eine 'Rahmenfunktion' der eigentlichen Minimierungsroutine notwendig, in der die Modellparameter logarithmiert, die Werte der Residuenquadratsumme $S(E_i), i = 1, \ldots, 6$, an den Simplexecken und folglich auch die theoretischen Werte der Momente berechnet werden. Dazu dient die Routine `minimierung`.

Berechnung der theoretischen Momente des Niederschlagsprozesses
Die Berechnung der zu einer Parameterkombination $\psi := (\lambda, \nu, \beta, \eta, \xi)$ gehörenden Werte der Momente erfolgt in einer Unterroutine `BerechneTheoretischeMomente`, die von `minimierung` aufgerufen wird. Das zur Berechnung der Wahrscheinlichkeiten $\phi(h)$ zu bestimmende Integral

$$\int_0^\infty (1 - p_h(t))\, dt \ ,$$

vergl. Kapitel 5.1, (1.17), wird numerisch mit der Trapezregel unter Verwendung der Routinen `trapzd` und `qtrap` bestimmt, die ebenfalls den Numerical Recipes (PRESS et al. [1992]) entnommen sind.



Anpassung eines NSRP-Modells für alle 30-Tage-Perioden

Routine zyklus

Nachfolgend geben wir den Quellcode der Anpassungsroutine `zykfit`, der im Diagramm auf Seite 200 dargestellt ist.

Aufruf der Routine:
`zykfit empmom.dat simplex.dat datei1.aus datei2.aus`
wobei
`simplex.dat`: Input-Datei, das Startsimplex enthaltend,
`datei1.aus`: Ausgabedatei der angepaßten Modellparameter,
`datei2.aus`: Ausgabedatei der Second-Order-Momente unter der 'optimalen' Parameterkombination.

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#define NRANSI
#include "zyklus.h"


FILE *ausgabe_par; // Zeiger auf das Ausgabe-File der Parameter
FILE *ausgabe_mom; // Zeiger auf das Ausgabe-File der Momente
FILE *moment_dat; // Zeiger auf das File mit den emp. Momenten

// zeitintervalle f. probs
double ZeitProb[ANZ_INTERVALL]={1.0, 2.0, 3.0, 6.0, 12.0, 24.0, 48.0};
// zeitintervalle fuer varianzen
double ZeitVarianz[ANZ_VARIANZ]={1.0, 3.0, 6.0, 12.0, 24.0};

int main(int argc, char *argv[])
{

double Simplex[ANZ_PARAMETER+1][ANZ_PARAMETER]; // Startsimplex
double LogSimplex[ANZ_PARAMETER+1][ANZ_PARAMETER]; // LogSimplex
double LSimplexKopie[ANZ_PARAMETER+1][ANZ_PARAMETER]; // Kopie zur Uebergabe an minimierung

int i,j;
int status;  // Status nach Dateioeffnen;
int periode; // Zaehlindex ueber die Perioden

momente schaetzer, *pschaetzer;
pschaetzer=&schaetzer;

// Oeffne Ein-und Ausgabedateien, Einlesen des Startsimplex
status=OeffneDateien(argc, argv,Simplex);
if(status==0){

// Logarithmiere Startsimplex: Simplex |--> LogSimplex
LogarithmiereSimplex(Simplex,LogSimplex);

fprintf(ausgabe_par,"lam \t mynu \t beta \t delta \t xi \t ResSumFit
        \t ResSumEval \n");
fflush(ausgabe_par);

periode=1;
while(!feof(moment_dat)){
printf("Bearbeite Periode %d\n", periode);
fflush(stdout);
// Einlesen einer Zeile emp. Momente
LeseEmpMom(moment_dat, pschaetzer);

// Erstelle Kopie des Log-Startsimplex
        KopiereMatrix(LogSimplex, LSimplexKopie);
```

```
// Minimierung mit gegebenem Log-Startsimplex und emp. Schaetzern
// Ausgabe der optimalen Parameter
fprintf(ausgabe_mom, "%d \t",periode);
fprintf(ausgabe_par, "%d \t",periode);
minimierung(LSimplexKopie, schaetzer);


// Neues Log-Startsimplex konstruieren
// Verwende die  letzten Zielkombinationen als neues Log-Startsimplex
    for(i=ANZ_PARAMETER;i>0;i--)
        for(j=0;j<ANZ_PARAMETER;j++)
              LogSimplex[i][j]=LogSimplex[i-1][j];
    for(j=0;j<ANZ_PARAMETER;j++)
        LogSimplex[0][j]=LSimplexKopie[0][j];

periode++;

}// Ende while( !feof(emp_dat))

fclose(moment_dat);
fclose(ausgabe_mom);
fclose(ausgabe_par);

} // Ende if(status==0)
else{
printf("Bitte Richtige Anzahl von Argumenten eingeben\n");
fflush(stdout);
}

return 0;
}      // Ende von main



#undef NRANSI
```

2 Simulation stündlichen Niederschlages

Die S-PLUS-Funktion `simuliere.NSRP.fun` setzt die in Kapitel 6.1 beschriebene Vorgehensweise zur Simulation von J Jahren stündlicher Niederschlagsmessungen mit dem saisonalen Neyman-Scott-Modell in einer S-PLUS-Routine um.

Die in Kapitel 5.6 beschriebene Translation und Interpolation der aus Periode t geschätzten Modellparameter $\hat{\psi}_t$ wird in der Simulationsroutine mit der Funktion `ShiftParameter.fun` vorgenommen. Deshalb wird als Input eine Liste der Länge 5 übergeben, deren Komponente k die an Modellparameter k angepaßte Fouriersumme enthält ($k = 1, 2, 3, 4, 5$ entsprechend $\lambda, \nu - 1$ (!), β, δ, ξ). Jede Komponente ist also eine Liste mit Komponenten `Freq`, das sind die Indizes j der angepaßten Frequenzen ω_j und `Coef`, das sind die zu Frequenz ω_j gehörenden Koeffizienten in der Form $A_j + iB_j \in \mathbb{C}$, wie sie von der S-PLUS-Funktion `fft` zurückgegeben werden. Aus Speicherplatzgründen wird die mit einem Cluster einhergehende Niederschlagsmenge den entsprechenden Stunden sofort zugeordnet; nach Erzeugung eines Jahres wird die resultierende Datenmatrix als ASCII-Datei gespeichert. Die Simulation wird als Schleife über die zu simulierenden Jahre durchgeführt. Zunächst werden die Clusterzentren mit dem Befehl `rbinom` simuliert. Anschließend werden in einer Schleife über die Stunden eines Jahres für jede Stunde, in der ein Clusterzentrum auftritt, die Zellanzahl und die zugehörigen Zellen erzeugt (Abstand vom Zentrum, Dauer und Intensität der Zellen). Dazu werden die Clusterzentren in der Mitte einer Stunde angeordnet, dem erwarteten Auftrittszeitpunkt. Der mit dem Cluster verbundene Niederschlag wird zu stündlichem Niederschlag aggregiert, wobei der Niederschlag einer Zelle anteilig auf die von der Zelle berührten Stunden aufgeteilt wird. Nach dem Ende eines jeden simulierten Jahres werden die bis dahin erzeugten Niederschlagsmessungen in einer Datei gespeichert. Am Jahreswechsel wird angenommen, daß jeder im vorigen Jahr begonnene Cluster spätestens mit Ende des 30-ten Tages des neuen Jahres inklusive aller zugehörigen Zellen abgeschlossen ist. Jedes simulierte Jahr wird in einer separaten Datei im Verzeichnis `dir` gespeichert. Der Dateiname ist `filenn.txt`, wobei nn die Nummer des simulierten Jahres ist.

Als Besonderheit ist in S-PLUS zu beachten, daß leere Schleifen zu Fehlermeldungen führen und Schleifen rückwärts durchlaufen werden, wenn der Startindex kleiner als der Endindex ist. Der exakte Algorithmus ist nachfolgend aufgeführt.

Simulationsalgorithmus

```
simuliere.NSRP.fun <-
function(par.list, Year = 1000, file = "sim_", dir = "d:\\temp\\")
{
# simulation of a seasonal Neyman-Scott model
# cluster centers: occurrence in period t is Bernoulli distributed
# par.list: list of length 5.
# component i contains the fourier approximation of parameter i,
# stored as a list with components Freq (frequencies) and
# Coef (coefficients as returned by fft)
# Year = number of years to be generated
# file = first characters of filenames
# dir = directory to store simulated dat in
     hour.threshold <- 0.1  # threshold hourly rain
     day.threshold <- 0.2   # threshold daily rain
```

```
    file <- paste(dir, file, 1:Year, ".txt", sep = "") # construct file names
    ModParameter <- ShiftParameter.fun(par.list)     # shift + calculate all model parameters
    lambda <- ModParameter[, 1]
    mynu <- ModParameter[, 2]
    beta <- ModParameter[, 3]
    eta <- ModParameter[, 3] + ModParameter[, 4]
    xi <- ModParameter[, 5]
    Rain <- rep(0, times = 9480)                # hourly rain depths, 24*395 hours
    for(year in 1:Year) {                       # loop over years
            cat("Year: ", year, "\n")
            write.table(file = datei[year], " ") # files start with an empty line
            Z <- rbinom(8760, 1, lambda)   # generate cluster centers
            ZStart <- (1:8760)[Z == 1]          # hours with cluster centers
            for(cluster in ZStart) {    # loop over cluster centers
            # cluster: hour of cluster center occurrence
            # generate cells
                 CSize <- rpois(1, mynu[cluster]) + 1   # cluster size
                 CDistance <- rexp(CSize, beta[cluster]) # distance between cell and center
                 # starts of cells:  place center in the middle of an hour
                 CStart <- cluster - 0.5 + CDistance
                 CStartHour <- ifelse(CStart != round(CStart), ceiling(CStart),
                        CStart + 1)  # hour of cell starts
                 CLength <- rexp(CSize, eta[CIndex])    # cell lengths
                 CIntens <- rexp(CSize, xi[CIndex])    # cell intensities
                 CEnd <- CStart + CLength    # time of cell end
                 CEndHour <- ceiling(CEnd)           # hour of end of cells
                 # aggregate precipitation
                 for(cell in 1:CSize) {                # loop over cells of cluster
                 # case: start + end of cell in the same hour
                       if(CStartHour[cell] == CEndHour[cell]) {
                             Rain[CStartHour[cell]] <- Rain[CStartHour[cell]] +
                                  CIntens[cell] * CLength[cell]
                       }
                 # case: start + end of cell not in the same hour
                       if(CStartHour[cell] < CEndHour[cell]) {
                       # first hour of cell
                       # special case: ceiling(CStart)==CStart e.g. Start at 4:00h
                            Rain[CStartHour[cell]] <- ifelse(CStart[cell] != CStartHour[
                                cell], Rain[CStartHour[cell]] + CIntens[cell] * (ceiling(
                                CStart[cell]) - CStart[cell]), Rain[CStartHour[cell]] +
                                CIntens[cell])
                            # strictly middle hours of cell
                            if(CStartHour[cell] < (CEndHour[cell] - 1)) {
                                for(hour in (CStartHour[cell] + 1):(CEndHour[cell] - 1))
                                       Rain[hour] <- Rain[hour] + CIntens[cell]
                            }
                            # Last hour of cell
                            Rain[CEndHour[cell]] <- ifelse(CEndHour[cell] != CEnd[cell],
                                Rain[CEndHour[cell]] + CIntens[cell] * (CEnd[cell] - floor(
                                CEnd[cell])), Rain[CEndHour[cell]] + CIntens[cell])
                       } # end: if start + end not in the same hour
                 }         # end of loop over cells, next cell
            }                # end of loop over clusters, next cluster
        # round Rain and observe threshold
        RainMatrix <- matrix(round(Rain[1:8760], 0), nrow=365, ncol=24, byrow=T)
        RainMatrix[RainMatrix < hour.threshold] <- 0        # threshold hourly rain
        Threshold <- apply(RainMatrix,1,sum)     # threshold daily rain
        Rain[Threshold < day.threshold, ] <- 0
        write.table(cbind(1:365, RainMatrix), file = datei[year], append=T, sep="\t")
        dummy.vec <- Rain[8761:9480] # save overlapping rain at end of year
        Rain <- rep(0, times = 9480)
        Rain[1:720] <- dummy.vec
    } # next year
}
```

Die Berechnung der Modellparameter und die anschließende Translation werden, wie in Kapitel 5.6 beschrieben, mit der Funktion `ShiftParameter.fun` vorgenommen.

Berechnung der Parameterwerte aller Stunden

```
ShiftParameter.fun <- function(par.list)
{
# par.list: list of length 5.
# component i contains the fourier approximation of parameter i,
# stored as a list with components Freq (frequencies) and
# Coef (coefficients as returned by fft)
# 1) calculate parameters for all 24*365 hours of a year
# 2) shift matrix
     timepoints <- seq(1,by=1/6, length=8760)
     Fitted.Mat <- matrix(0, nrow=8760, ncol=5)
     for(i in 1:5){
         parameter <- par.list[[i]]
         anz.par <- length(parameter$Koef) # number of frequencies used
         # calculate parameter values by successively adding fitted frequencies
         # loop over frequencies
         Fitted <- Mod(parameter$Koef[1])  # mean R_0
         for(k in 2:anz.par){
            Fitted <- Fitted + Mod(parameter$Koef[k])*
                      cos(2*pi*timepoints*parameter$Freq[k]/1460 +
                      Arg(parameter$Koef[k]))
         }
         Fitted.Mat[,i] <- Fitted
     }
     # shift parameter matrix
     Fitted.Mat <- rbind(Fitted.Mat[8389:8760,],Fitted.Mat[1:8388,])
     Fitted.Mat
}
```

3 Algorithmus zum Zählen der Runlängen

In Kapitel 6.5 wird die bedingte Verteilung der trockenen und der nassen Runs bei einer gegebenen Startstunde betrachtet. Zum Schätzen der dort definierten Wahrscheinlichkeiten $\mathrm{P}^{(W)}(i)$ und $\mathrm{P}^{(D)}(i)$ wird die Anzahl der nassen und der trockenen Runs der Länge $i, i = 1, \ldots, 168, 169+$, aus stündlichen Niederschlagsbeobachtungen mit einem einfachen Zählalgorithmus bestimmt. Die Wahrscheinlichkeiten werden aus den Anzahlen in S-PLUS bestimmt.

Vorausgesetzt werden stündliche Niederschlagsmessungen, die für jedes beobachtete oder simulierte Jahr in einer eigenen Datei als Matrix mit 365 Zeilen und 25 Spalten vorliegen, wobei die erste Spalte das Datum enthält und Spalte i die in Stunde $i-1$ beobachtete Niederschlagsmenge. An das Programm werden der Name der Ausgabedatei sowie die Namen der zu bearbeitenden, Niederschlagsmessungen enthaltenden Dateien übergeben.

Schritt 1: Initialisierung der Variablen: $zaehler = 0$ (Laufindex, zählt die Runlänge),
$hour$, day, $monat$, $jahr = 0$ (Laufvariablen),
$index$ (notiert den Zustand trocken oder naß),
$LimitStd$ (Endstunden der Monate $1, \ldots, 12$ zum Zuordnen der Runs),
$RunWet$, $RunDry = 0$ ((12,169)-Matrizen zum Speichern der Anzahl der beobachteten Runlängen),
$Daten$ (Vektor der Länge 365, $Daten_t$ = Zustand an Tag t).
Mit der Anzahl der zu bearbeitenden Dateien ist die Anzahl der zu bearbeitenden Jahre $Jahr$ festgelegt.
Öffnen der Ausgabedatei.

Schritt 2: Schleife über die Jahre $jahr = 1, \ldots, Jahr$.

Schritt 3: Lese die Beobachtungen eines Jahres ein, wandele sie in einen 0–1–Vektor um (entsprechend dem Zustand trocken /naß der Stunde, speichere sie in $Daten$.

Schritt 4: Setze $index = Daten_1$.

Schritt 5: Lese bis zum ersten neuen Run:
Solange $Daten_{hour} = index$: $hour = hour + 1$.
Abbruch in der Startstunde eines neuen Runs.

Schritt 5: Initialisiere $monat = startmonat$ gemäß der Abbruchstunde $hour$, so daß Runs dem korrekten Monat zugeordnet werden.

Schritt 6: Schleife über die Monate: Solange $monat \leq 12$:

Schritt 7: Schleife über die Stunden: Solange $hour \leq LimitStd_{monat}$:

Schritt 8: Setze $zaehler = 0$, $index = Daten_{hour}$.

Schritt 9: Lese bis zum Ende des Runs:
Solange $Daten_{hour} = index$: $hour = hour + 1$, $zaehler = zaehler + 1$.

Schritt 10: Bei Abbruch: Ordne die Runlänge korrekt zu.
Beachte die Fallunterscheidungen:
a) trockener/nasser Run.
b) $zaehler \leq 168$, $zaehler > 169$.

Schritt 11: Nächster Run: Weiter bei Schritt 7.

Schritt 12: Ende der Schleife über die Stunden (Schritt 6).
$monat = monat + 1$.

Schritt 13: Erreichen des Dateiendes (Schritt 2).

Schritt 14: Ausgabe der Zählmatrizen $RunDry$ und $RunWet$.
Schließen der Ausgabedatei.

Programmende

Der zu diesem Algorithmus gehörende C-Quellcode ist nachfolgend gegeben:

```
/*
Berechnung der Runlaengen aus stuendlichem Regen fuer jeden Monat k=1,.. , 12
fuer nasse und trockene Runs.

Bearbeitet werden Dateien mit 25 Spalten (Datum + 24 Stunden)
und 365 Zeilen.
Die Runlaengen werden in Klassen von 1 bis KlassenMax =169 eingeteilt
(169=7*24+1 --> ein Run dauert echt laenger als eine Woche)

An das Programm wird uebergeben
1. Name der Ausgabe-Datei
2. Namen der zu bearbeitenden Dateien
z.B. runs aus.txt `cat datei.lst`

Das Ergebnis wird in eine Ausgabe-Datei geschrieben. Es besteht aus 2 Matrizen
(trockene und nasse Runs) mit je 12 Zeilen (fuer jeden Monat eine).
Spalte i enthaelt die Anzahl der i-Runs.

Zaehlt nur Runs, die innerhalb eines Monats beginnen und zaehlt stets bis zum
Ende des Runs.

Indizes von Vektoren der Laenge L laufen von 0 bis L-1 (C-Konvention)
*/



#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#define NRANSI
#define KlassenMax 169 // 169=7*24+1 > 1 Woche
#define TagMax 24      // Anzahl Eintraege pro Tag: 24 Std

FILE *out_file; // Zeiger auf das Ausgabefile
FILE *in_file;  // Zeiger auf das Eingabefile
int main(int argc, char *argv[])
{

// Monatsendstd
int LimitStd[12]={744,1416,2160,2880,3624,4344,5088,5832,6552,7296,8016,8760};
```

```
int jahr,monat,day,hour,i;   // Laufvariable
int index;                // Index 0-1 = trocken-nass
int zaehler;           // Anzahl an Runs
int RunWet[12][KlassenMax];   // Matrix fuer nasse Runs
int RunDry[12][KlassenMax];   // Matrix fuer trockene Runs

double Beob[TagMax];              // Beob. fuer ein Tag
int Daten[365*TagMax];           // Beob. fuer ein Jahr
double Dummy;
char Monate[12][4]={"Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep",
    "Okt","Nov","Dec"};


// Initialisierung
for(monat=0;monat<12;monat++){
    for(hour=0;hour<KlassenMax;hour++){
         RunWet[monat][hour]=(long)0;
         RunDry[monat][hour]=(long)0;
    }
}


out_file=fopen(argv[1],"w");
printf("Oeffne %s\n",argv[1]);
fflush(stdout);

// Durchlaufe alle Dateien argv hat indizes 0, .. argc-1
for(jahr=2;jahr<argc;jahr++){
    in_file=fopen(argv[jahr],"r");  // Oeffne Datei, argv[1]=ausgabe-datei
    printf("Oeffne %s \n",argv[jahr]);
    fflush(stdout);
    day=0;
    while(!feof(in_file)){ // lese Datei zeilenweise ein
         fscanf(in_file,"%lf ",&Dummy);  // 1. Eintrag = Datum ueberlesen
         for(hour=0;hour<TagMax;hour++){  // Stunden 1 - 24 lesen
              fscanf(in_file,"%lf ",&Beob[hour]);
              // Beob in 0-1-Vektor umwandeln
              Daten[day*24+hour]= (Beob[hour] >(double)0.0) ? (int)1 : (int)0 ;
        }
        day++;
    }
    fclose(in_file);

// Runs zaehlen fuer jeden Monat
    hour=0; // Initialisierung. BeobVektor beginnt mit Index 0
    // 1. Monat: zaehlen bis zum 1. Wechsel + achte auf Monatswechsel
    index=Daten[hour];
    while(Daten[hour] == index)
         hour++;
    index=Daten[hour]; // index neu setzen, Jetzt: Beginn des 1. Runs.
    // moeglicher Sonderfall: 1. Monat ist einheitlich (nur 1 oder nur 0)
    // Monat zuordnen
    if(hour <= LimitStd[0]) monat =0;
    for(i=0; i< 10; i++)
         if(hour > LimitStd[i] && hour <= LimitStd[i+1]) monat = i+1;
    while(monat < 12){
         while(hour<LimitStd[monat]){ // Achte auf korrekten Laufbereich
              index=Daten[hour];
              zaehler=0;
              // Durchlaufe einen Run
              while(Daten[hour] == index){
                   zaehler++;
                   hour++;
              } // RunEnde
              // Runlaenge zuordnen
              if(zaehler<KlassenMax-(int)1){
              // MatrixIndizierung beachten: Beginn bei 0
                   if(index==1)
                        RunWet[monat][zaehler-(int)1]++;  // wet run
                   else
```

```
                    RunDry[monat][zaehler-(int)1]++;// dry run
            } // Ende if duration < 1 week
            else { // run duration >= 1 week
                if(index==1)
                    RunWet[monat][KlassenMax-1]++;  // wet run
                else
                    RunDry[monat][KlassenMax-1]++;// dry run
            } // Ende else duration > 1 week
        } // RunZaehler zuordnen beendet, neuer Run
        monat++; // hier: hour > LimitStd[monat]. also: neuer Monat
    } // Ende while monat < 12, naechste Datei
} // Ende Schleife ueber die Jahre/Dateien

// Ausgabe
fprintf(out_file,"\nNasse Runs\n\n");
for(monat=0;monat<12;monat++){
    fprintf(out_file,"%s \t",Monate[monat]);
    for(hour=0;hour<KlassenMax;hour++)
        fprintf(out_file,"%d \t",RunWet[monat][hour]);
    fprintf(out_file,"\n");
}
fflush(out_file);

fprintf(out_file,"\n\ntrockene Runs\n\n");
for(monat=0;monat<12;monat++){
    fprintf(out_file,"%s \t",Monate[monat]);
    for(hour=0;hour<KlassenMax;hour++)
        fprintf(out_file,"%d \t",RunDry[monat][hour]);
    fprintf(out_file,"\n");
}
fflush(out_file);


fclose(out_file);
printf("Schliesse %s\n",argv[1]);
return 0;
} // Ende main

#undef NRANSI
```