

Alle reden vom Klima – wir analysieren es mit R

Helmut Pruscha

Mathematisches Institut der LMU München

Einleitung

Die Auswertung von Klimareihen gewinnt eine immer größer werdende Bedeutung. Das Thema dieses Beitrags ist die statistische Analyse von Temperatur- und Niederschlags-Daten, die in Form von monatlichen Mittelwerten vorliegen (gelegentlich werden auch Tagesmittel verwendet). An drei Stationen, Hohenpeißenberg, Karlsruhe und Potsdam, sind diese Daten über viele Jahre hinweg erhoben worden. Den Schwerpunkt der statistischen Verfahren bildet die Zeitreihenanalyse, mit ihren Methoden der Deskription (Plots), der Modellbildung und Prädiktion, sowie der Spektralanalyse. Sämtliche Auswertungen sind mittels des open-source Programmpakets R durchgeführt worden, vgl.

`cran.r-project.org`.

Die vom Autor benutzten Programmcodes sind in den Text integriert worden. Kommentare, die auch im Output erscheinen, stehen innerhalb von “...” Zeichen; Kommentare, die vom Programm ignoriert werden, erkennt man am vorangestellten # Zeichen. Zusammen mit dem `read.table` Befehl aus Abschn. 1.1 und –optional– dem `sink()` Befehl aus Abschn. 2.1 sind die Programme lauffähig.

Nach der Präsentation der Klimareihen und den Zeitreihenplots der Jahres- und Winter-Mittelwerte studieren wir langfristige Temperatur-Trends und Temperatur-Anstiege als auch Verläufe von Temperatur und Niederschlag über die 12 Monate des Jahres (Jahresgänge). Die Korrelationen zeitlich aufeinanderfolgender Werte (Jahreswerte, Quartals-, Monats-, Tageswerte) führt zur Frage der Binnenstruktur einer Klimareihe, d. h. zum Problem der Modellierung und der Prädiktion. Dies wird an Hand der Jahres-Mittelwerte mit Hilfe von ARMA-Modellen und der Monats-Mittelwerte mit Hilfe kombinierter Modelle (Trend+ARMA) durchgeführt. Bei der Frequenzanalyse geht es um die „Suche nach verborgenen Periodizitäten“.

Im Anhang A finden sich Ergänzungen zu den statistischen Methoden, insbesondere zu solchen der Zeitreihenanalyse. Im Anhang B werden Auszüge aus den – bei der Analyse benutzten – Datensätzen präsentiert.

Helmut Pruscha
Mathematisches Institut der Universität München
`pruscha@math.lmu.de`

11. November 2009

Inhaltsverzeichnis

1	Klima-Reihen	4
1.1	Stationen	4
1.2	Temperaturreihen	5
1.3	Niederschlagsreihen	7
2	Temperatur-Trend und Temperatur-Anstieg	14
2.1	Vergleich der beiden letzten Jahrhunderte	14
2.2	Historische Temperaturschwankungen	17
3	Monatswerte	18
4	Korrelationen: Jahr, Jahreszeit, Monat, Tag	21
4.1	Auto-Korrelationskoeffizient	21
4.2	Korrelation und die Prädiktion von Extremen	24
4.3	Anwendung auf Klimadaten	25
4.4	Wetterregeln	25
5	Modell und Prädiktion: Jahreswerte	28
5.1	Differenzbildung, Prädiktionen	28
5.2	Mittlere Jahrestemperaturen	29
5.2.1	Zum Vergleich: Gleitende Durchschnitte	29
5.2.2	ARIMA-Residuen, GARCH-Modell	32
5.3	Jährliche Niederschlagsmengen	35
6	Modell und Prädiktion: Monatliche Werte	43
6.1	Doppelte ARMA-Methode	43
6.2	Zum Vergleich: Gleitende Durchschnitte und Lag-12 Differenzen	46
6.3	Residuenanalyse	47
7	Periodogramm- und Spektraldichte-Analyse	50
7.1	Periodogramm, auch geglättet	50
7.2	Temperatur: Jahres- und Winterwerte	51
7.3	Niederschlag: Winterwerte und Summarium	51
8	Anhang A: Statistische Methoden	59

8.1	Univariate Stichproben	59
8.2	Univariate 2-Stichproben Situation	60
8.3	Bivariate Stichprobe: Scattergramm und Korrelationskoeffizient	61
8.4	Bivariate Stichprobe: Einfache Lineare Regression	62
8.5	Bedingte Wahrscheinlichkeiten	65
8.6	Zeitreihen: Autokorrelation	66
8.7	Zeitreihen: Frequenzanalyse	68
8.8	Zeitreihen: Modelle	71
9	Anhang B: Datenauszüge	74

1 Klima-Reihen

1.1 Stationen

Die drei Stationen, deren Daten wir verwendet haben, sollen kurz vorgestellt werden.

Station Hohenpeißenberg:

Der zwischen Weilheim und Schongau (Bav.) gelegene Hohe Peißenberg (989 m ü. M.) liegt im Lee-Bereich der nördlichen Alpen. Auf ihm befindet sich ein Pfarrhaus, das seit 1781 der Ort von Wetteraufzeichnungen ist. Im Jahre 1940 zog diese Bergwetterstation in das neugebaute Observatorium um.

Wir haben die Temperatur-Originaldaten von Attmannspacher übernommen. Ausnahmen bilden die Jahre 1811, 1812 (Aufzeichnungslücke) und 1879–1900 (geänderter Ablesemodus). Hier haben wir auf die homogenisierten Werte von Grebe zurückgegriffen, ferner auf Korrekturmitteilungen von W. Fricke, DWD.

Quelle: Deutscher Wetterdienst, Offenbach/m, Klimadaten ausgewählter deutscher Stationen: www.dwd.de/ (*Klima+Umwelt, Klimadaten*).

Ferner: Grebe (1957), Attmannspacher (1981).

Station Karlsruhe:

Hier werden seit 1799 Temperatur-Aufzeichnungen durchgeführt. Sie enden leider mit dem Oktober 2008. Seitdem befindet sich die Messstationen in Rheinstetten, ca. 20 km außerhalb von Karlsruhe; die Homogenität der Datenreihe ist nicht mehr gegeben.

Wir haben die fehlenden Monatswerte Nov. und Dez. 2008 aus der Quelle DWD ergänzt (Temperatur) bzw. extrapoliert (Niederschlag).

Quelle: www.klimadiagramme.de

Station Potsdam:

Die „Säkularstation“ liegt im Südwesten der Stadt Potsdam ca. 600 m außerhalb der bebauten Zone.

Quelle: <http://saekular.pik-potsdam.de>

Übersicht:

Stationsname	Stationshöhe	geogr. Breite	geogr. Länge	Beginn der Temp.reihe	Beginn der Niederschl.reihe
Hohenpeißenberg	977 m	47° 48'	11° 00'	1781	1879
Karlsruhe	112 m	49° 02'	08° 21'	1799	1876
Potsdam	81 m	52° 23'	13° 03'	1893	1893

[R] Die Klimadaten HohenT, HohenN etc. sind im Ordner C:/KLIMA als Textfiles gespeichert, und zwar in der Form, wie sie in Anhang B wiedergegeben sind (ohne die Trennzeilen - - - - -). Die Kopfzeile (header) besteht aus den Variablennamen

Jahr dzvj jan feb mae apr mai jun jul aug sep okt nov dez Tjahr

bzw. Njahr. Das Einlesen in das R Programm erfolgt - je nach Anwendung - durch

```
hohenTp<- read.table("C:/KLIMA/HohenT.txt",header=T)
hohenNs<- read.table("C:/KLIMA/HohenN.txt",header=T)
karlsTp<- read.table("C:/KLIMA/KarlsT.txt",header=T)
```

```
karlsNs<- read.table("C:/KLIMA/KarlsN.txt",header=T)
potsdTp<- read.table("C:/KLIMA/PotsdT.txt",header=T)
potsdNs<- read.table("C:/KLIMA/PotsdN.txt",header=T)
```

1.2 Temperaturreihen

Es werden für jede Station zwei Zeitreihen präsentiert: die mittleren Jahrestemperaturen (oben) und die mittleren Wintertemperaturen (unten). Der meteorologische Winter erstreckt sich über Dezember (des Vorjahrs), Januar und Februar. Die Temperaturwerte sind in [°C] angegeben.

Hohenpeißenberg (1781–2008): Abb 1

Karlsruhe (1799–2008): Abb 2

Potsdam (1893–2008): Abb 3

R Berechnung elementarer statistischer Kenngrößen; einmal einfach für *Jahr*, einmal dreifach mittels Funktion `printL` für *Jahr*, *Winter*, *Sommer*.

```
attach(hohenTp)

Y<- Tjahr/100;          #Temperaturmittel Jahr in Celsius
N<- length(Y);meanY<- mean(Y);sdY<- sqrt(var(Y))
rhoY<- cor(Y[1:(N-1)],Y[2:N])
c("N Jahre"=N, "Mean"=meanY, "StDev"=sdY, "Autokor(1)"=rhoY)

#-----

printL<- function(Y){
N<- length(Y);meanY<- mean(Y);sdY<- sqrt(var(Y))
rhoY<- cor(Y[1:(N-1)],Y[2:N])
list("N Jahre"=N, "Mean"=meanY, "StDev"=sdY, "Autokor(1)"=rhoY)
}

Y<- Tjahr/100;          "Temperaturmittel Jahr in Celsius"
unlist(printL(Y))

Y<- (dzvj+jan+feb)/30;  "Temperaturmittel Winter in Celsius"
unlist(printL(Y))

Y<- (jun+jul+aug)/30;   "Temperaturmittel Sommer in Celsius"
unlist(printL(Y))

detach(hohenTp)
```

Eine Übersicht über die statistischen Kenngrößen Mittelwert (m), Standardabweichung (s) und Autokorrelationskoeffizient $r(1)$ (d. i. der Ordnung 1) bietet die Tabelle; vgl. auch Anhang A, 8.1, und bez. $r(1)$ den Vorgriff auf 4.1.

	Hohenp. $n = 228$			Karlsruhe $n = 210$			Potsdam $n = 116$		
	m	s	$r(1)$	m	s	$r(1)$	m	s	$r(1)$
Winter	-1.36	1.74	0.08	1.76	1.89	0.11	0.19	2.09	0.13
Frühling	5.56	1.32	0.16	10.18	1.08	0.24	8.45	1.15	0.19
Sommer	14.25	1.08	0.20	18.72	1.07	0.25	17.40	1.01	0.15
Herbst	6.96	1.30	0.01	10.19	1.03	0.05	8.91	1.08	0.08
Jahr	6.35	0.84	0.29	10.22	0.80	0.33	8.78	0.81	0.36

Die jährlichen Temperaturmittel sind größenmäßig deutlich abgestuft, Karlsruhe $>$ Potsdam $>$ Hohenpeißenberg, ihre Streuungen s (≈ 0.8) sowie auch noch ihre Autokorrelationen $r(1)$ sind auffallend gleich groß. Die Korrelation aufeinanderfolgender Jahres-Temperaturmittel beträgt demnach 0.29...0.36. Wir werden weiter unten aufdecken, welchen Anteil an diesen Werten der Trend innerhalb der Zeitreihen hat.

Zu den Quartalswerten: Die größte Streuung weisen die Wintertemperaturen auf, die geringsten Autokorrelationen die Winter- und vor allem die Herbstwerte. Die Korrelation aufeinanderfolgender Herbst-Temperaturmittel beträgt demnach 0.01...0.08, was praktisch Unkorreliertheit bedeutet. Die Zeitreihenplots der Wintertemperaturen, das sind die unteren Abb. 1 – 3, reflektieren die Werte der Kenngrößen s und $r(1)$. Sie zeigen eine hohe Fluktuation bei einem (im Vergleich zu den oberen Abb. der Jahrestemperaturen) geringer ausgeprägten Trendverlauf und ähneln phasenweise den Plots von reinen Zufallsreihen (vgl. Anhang A, 8.6).

R Plot der Temperatur-Jahresmittel, mit einem Ausgleichspolynom 4ten Grades und mit zentrierten Gleitenden 10-Jahres Durchschnitten.

Das Postscript-File wird unter C:/KLIMA/HohenTj.ps gespeichert.

```
attach(hohenTp)
postscript(file="C:/KLIMA/HohenTj.ps",height=6,width=16,horizontal=F)

Y<- Tjahr/100 # Jahresmittel in Celsius
Ja<- Jahr-1800 #Werte verkleinern
#Ausgleichsploynom 4.ten Grades
J2<- Ja*Ja; J3<- J2*Ja; J4<- J3*Ja
tppol<- lm(Y~Ja+J2+J3+J4)

#Zentrierte Gleitende 10-Jahres Durchschnitte
N<- length(Y); p<- 10; m<- p/2
sequ<- 1:N
glD<- sequ; su<- sequ
for (t in (m+1):(N-m))
{su[t]<- 0
{ for (k in -(m-1):(m-1))
su[t]<- su[t]+ Y[t+k] }}
for (t in (m+1):(N-m))
{glD[t]<- 0
glD[t]<- glD[t]+((Y[t-m]+Y[t+m])/2+su[t])/p}

ytext<- "Temperatur-Mittel Jahr [C]"
ttext<- "Mittlere Jahrestemperatur"
```

```

cabl<- c(4,5,6,7,8)                                #Horizontale Hilfslinien

plot(Jahr,Y,type="l",lty=1,xlim=c(1780,2008),ylim=c(4.0,8.5),
     xlab="Jahr",ylab=ytext,cex=1.3)
title(main="Hohenpeissenberg, Temperatur 1781-2008",cex=1.0)
abline(h=cabl,lty=2)
abline(h=mean(Y),lty=3)
text(2008,mean(Y),trunc(mean(Y)*1000)/1000,cex=0.8)
text(1880,8.4,ttext,cex=1.2)

lines(Jahr,predict(tppol),lty=2)                    #Ausgleichpol. plotten
lines(Jahr[(m+1):(N-m)],glD[(m+1):(N-m)],lty=1)    #Gltde Durchschn

dev.off()
detach(hohenTp)

```

1.3 Niederschlagsreihen

Es werden für jede Station wiederum zwei Zeitreihen präsentiert: die Jahres-Niederschlagshöhen (oben) und die Winter-Niederschlagshöhen (unten), beides in [dm].

Hohenpeißenberg (1879–2008): Abb 4

Karlsruhe (1876–2008): Abb 5

Potsdam (1893–2008): Abb 6

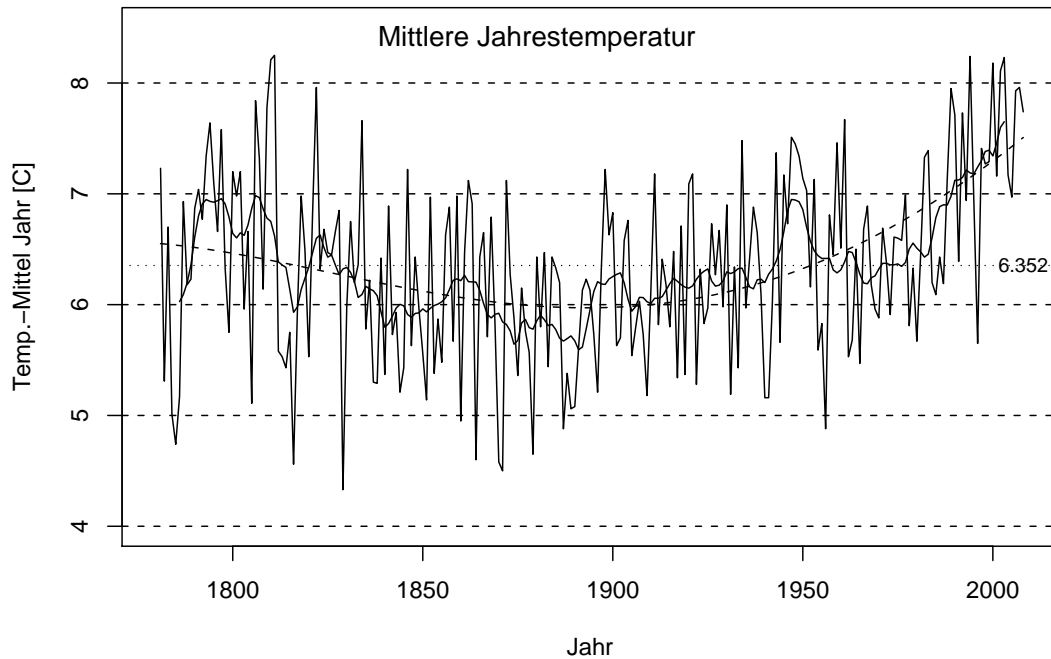
Eine Übersicht über die statistischen Kenngrößen Mittelwert (m), Standardabweichung (s) und Autokorrelationskoeffizient $r(1)$ bietet die Tabelle (Niederschlagshöhen in [mm]).

	Hohenp. $n = 130$			Karlsruhe $n = 133$			Potsdam $n = 116$		
	m	s	$r(1)$	m	s	$r(1)$	m	s	$r(1)$
Winter	166	54	0.14	168	54	-0.04	130	37	0.06
Frühl	264	73	0.23	178	56	0.11	131	42	-0.02
Sommer	453	94	-0.12	228	70	-0.20	196	60	-0.03
Herbst	246	79	0.04	189	64	-0.01	133	43	-0.22
Jahr	1130	173	0.27	762	135	0.01	590	97	-0.08

Jahreswerte: Am fast 1000 m hohen Hohenpeißenberg fällt doppelt so viel Niederschlag wie in Potsdam. Anders als bei der Temperatur (die auf einer Skala ohne einen –für uns relevanten– Nullpunkt gemessen wird) liegen hier beim Niederschlag (mit echtem Nullpunkt) die Streuungswerte s größtmäßig in der gleichen Reihenfolge vor wie die Mittelwerte m . Der Koeffizient $r(1)$ für Karlsruhe und Potsdam ist praktisch Null.

Quartalswerte: Die winterlichen Niederschlagswerte weisen die geringsten Streuungen auf (bei der Temperatur waren es die größten). Außerdem sind (mit Ausnahme von Hohenp.) ihre Autokorrelationen $r(1)$ sehr gering.

Hohenpeissenberg, Temperatur 1781–2008



Hohenpeissenberg, Temperatur 1781–2008

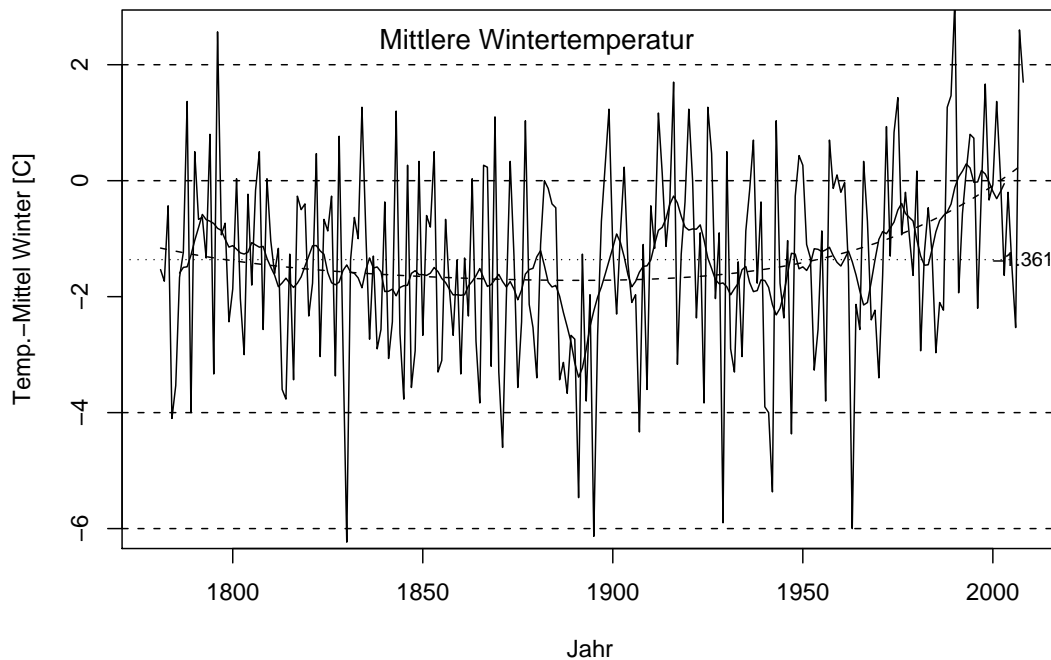
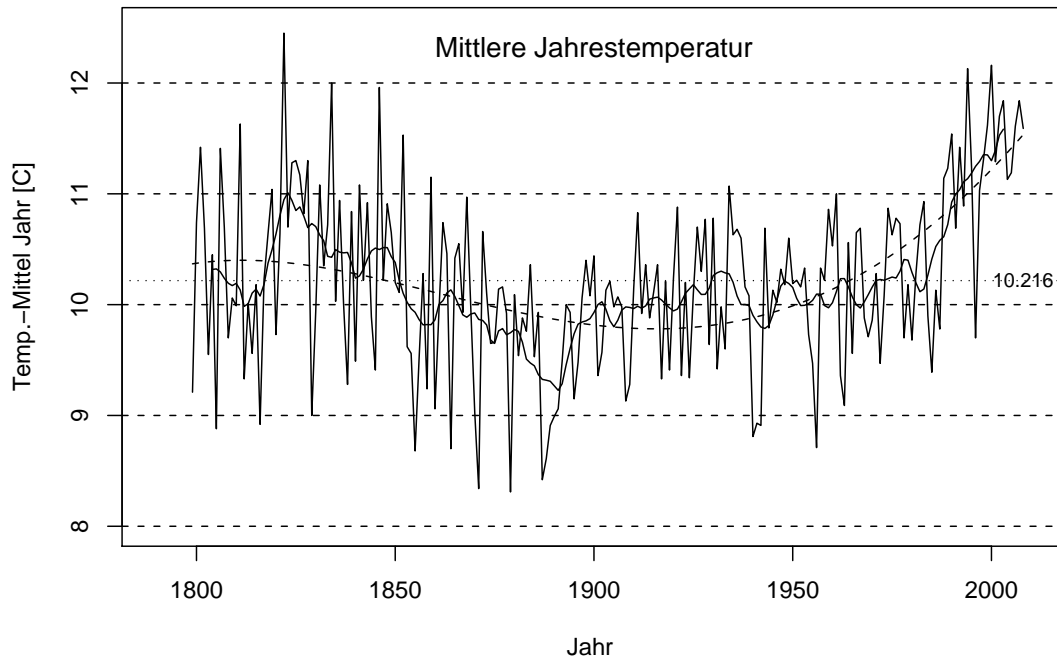


Abbildung 1: Mittlere Jahrestemperaturen und mittlere Wintertemperaturen [$^{\circ}C$], Hohenpeissenberg, 1781–2008, jeweils mit Ausgleichspolynom vierten Grades (---), mit zentrierten gleitenden (10-Jahres) Durchschnitten (innere durchgezogene Linie) und mit dem Mittelwert über alle 228 Jahre (···)

Karlsruhe, Temperatur 1799–2008



Karlsruhe, Temperatur 1799–2008

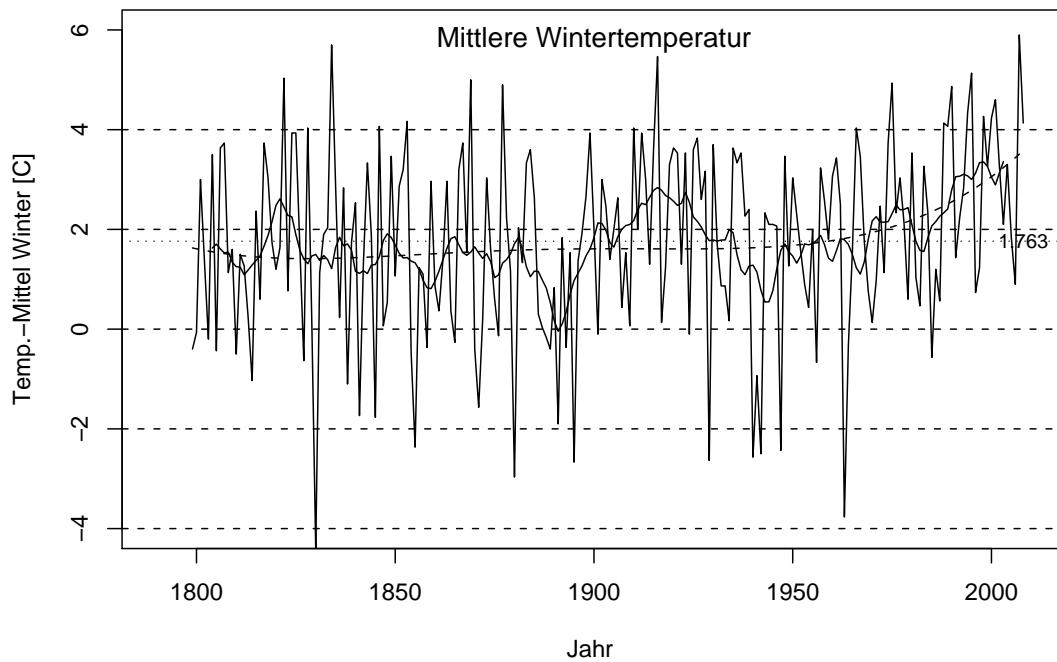
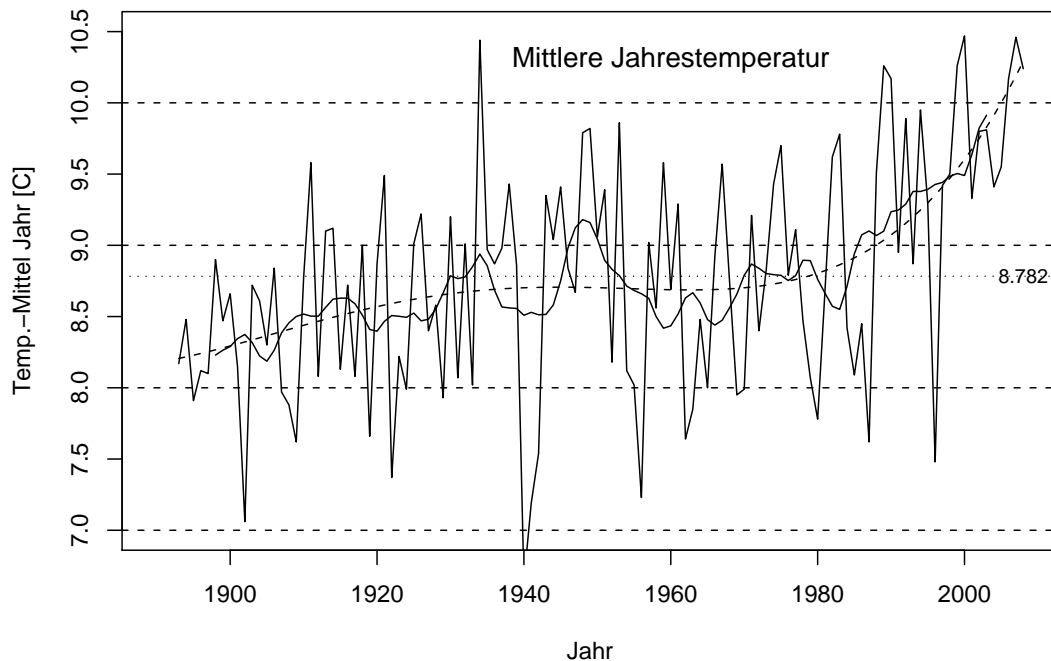


Abbildung 2: Mittlere Jahrestemperaturen und mittlere Wintertemperaturen [°C], Karlsruhe, 1799–2008, jeweils mit Ausgleichspolynom vierten Grades (---), mit zentrierten gleitenden (10-Jahres) Durchschnitts (innere durchgezogene Linie) und mit dem Mittelwert über alle 210 Jahre (···)

Potsdam, Temperatur 1893–2008



Potsdam, Temperatur 1893–2008

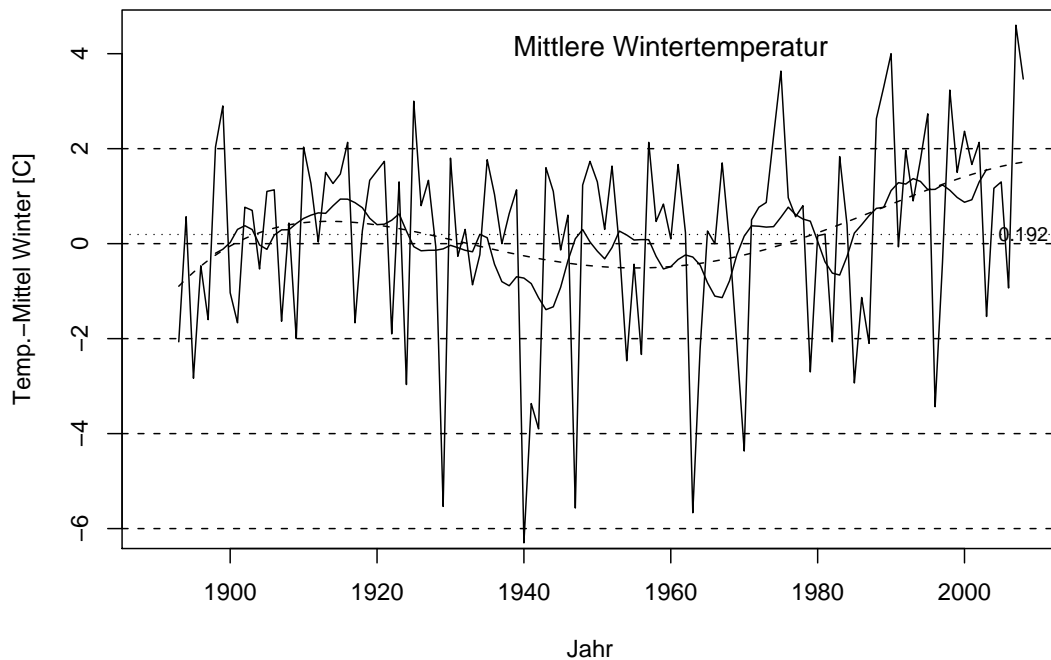
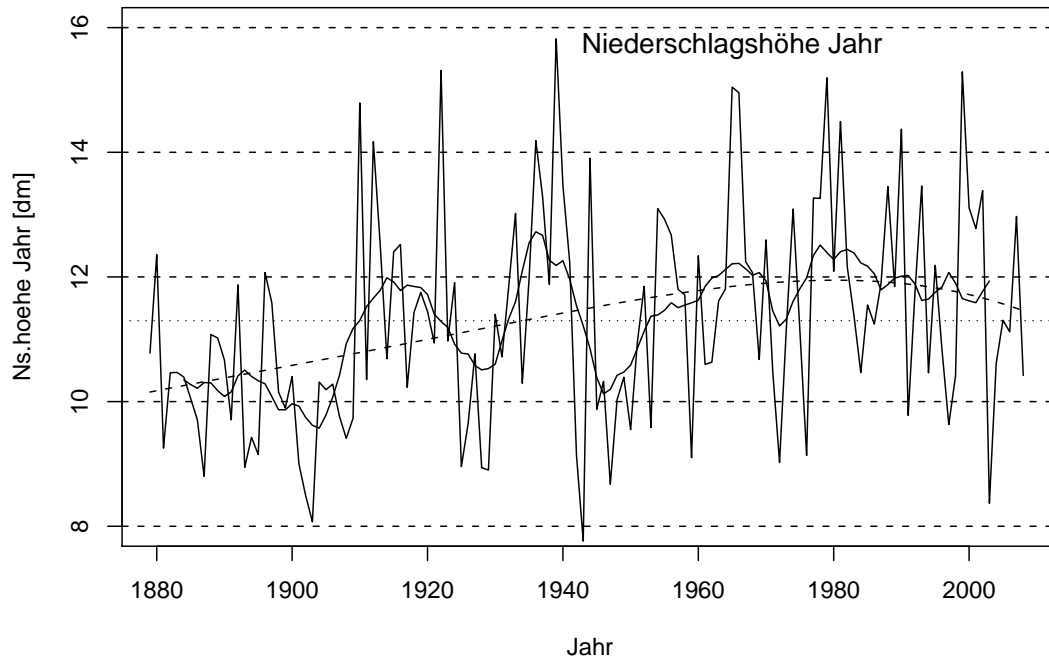


Abbildung 3: Mittlere Jahrestemperaturen und mittlere Wintertemperaturen [°C], Potsdam, 1893–2008, jeweils mit Ausgleichspolynom vierten Grades (---), mit zentrierten gleitenden (10-Jahres) Durchschnitts (innere durchgezogene Linie) und mit dem Mittelwert über alle 116 Jahre (···)

Hohenpeissenberg, Niederschlag 1879–2008



Hohenpeissenberg, Niederschlag 1879–2008

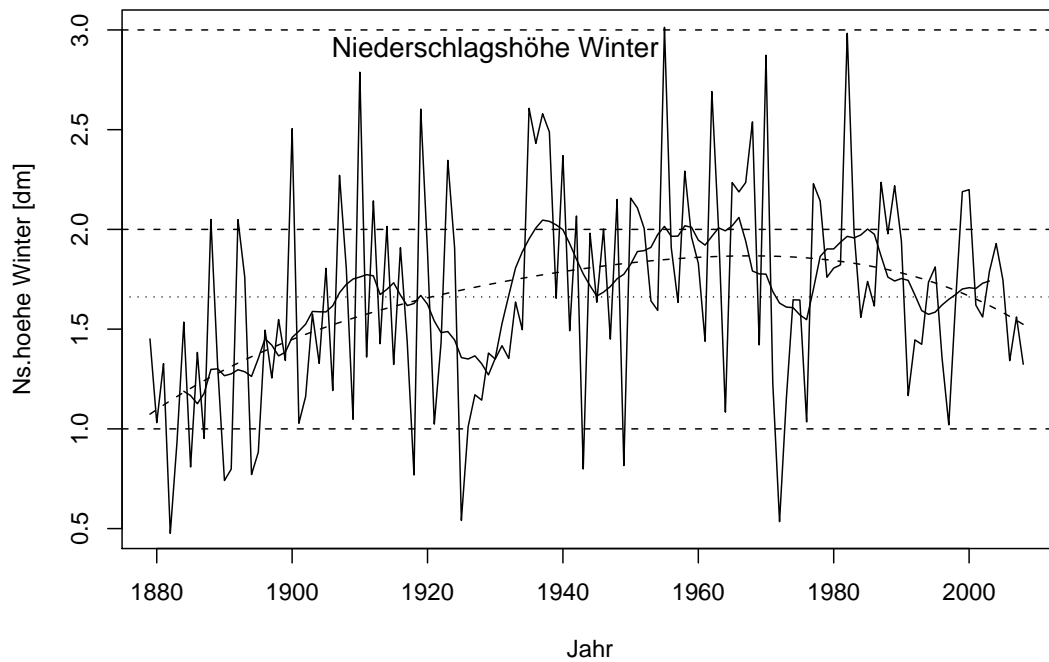
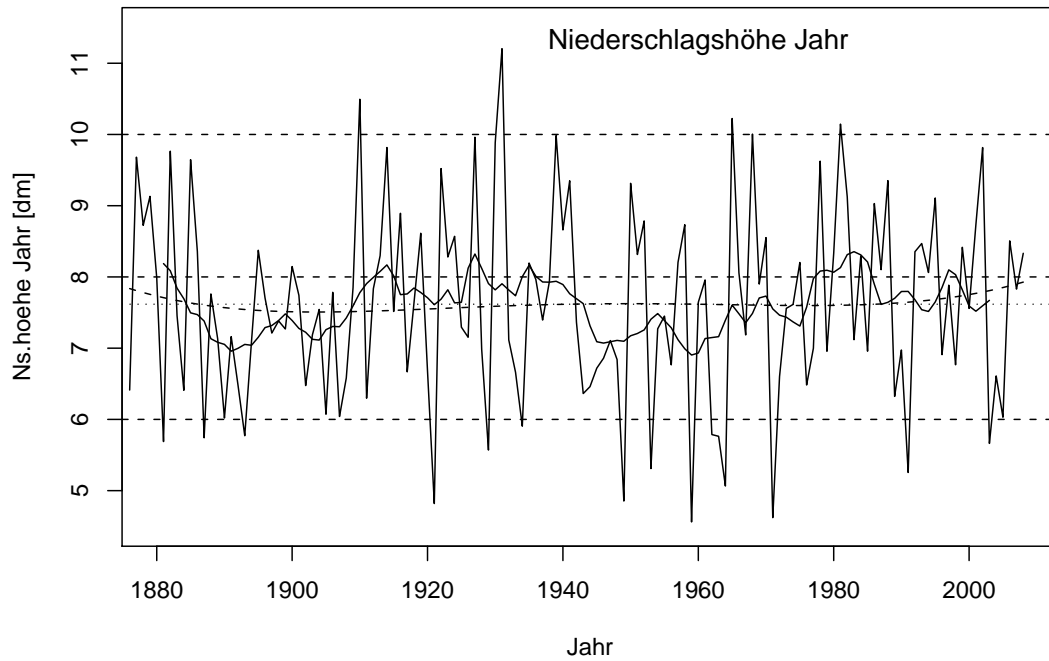


Abbildung 4: Niederschlagshöhen [dm] im Jahr und im Winter, Hohenpeissenberg, 1879–2008, jeweils mit Ausgleichspolynom vierten Grades (---), mit zentrierten gleitenden (10-Jahres) Durchschnitten (innere durchgezogene Linie) und mit dem Mittelwert über alle 130 Jahre (···)

Karlsruhe, Niederschlag 1876–2008



Karlsruhe, Niederschlag 1876–2008

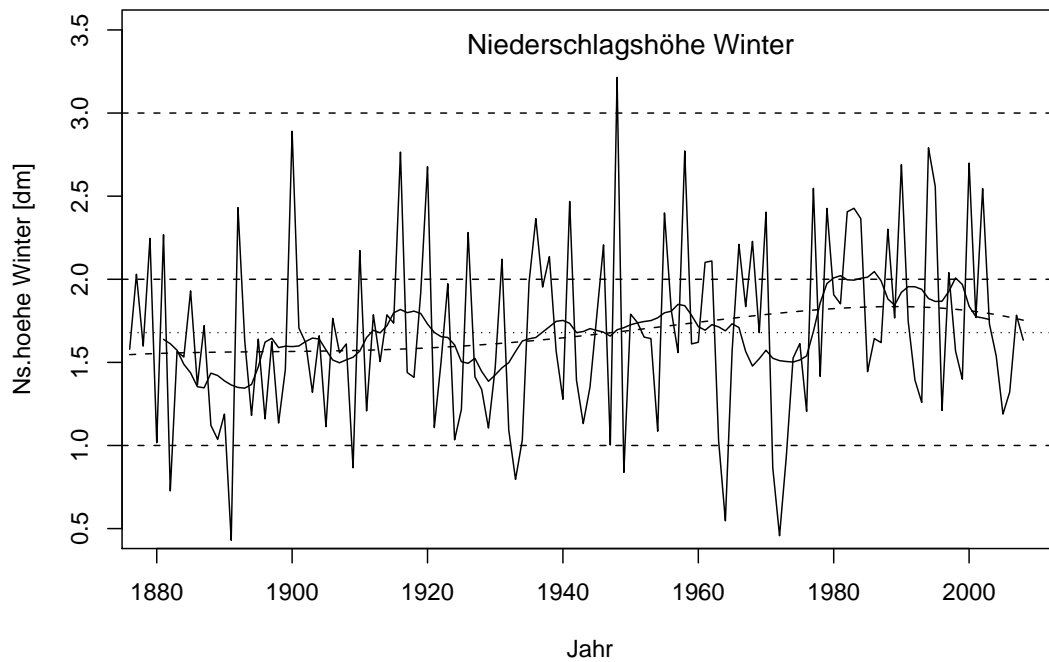
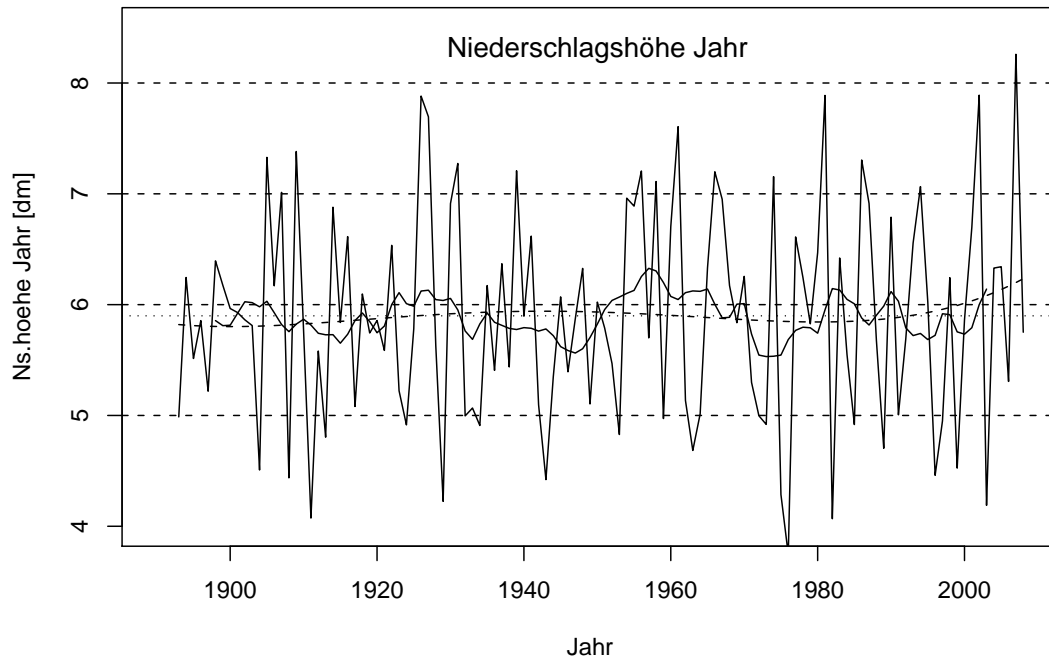


Abbildung 5: Niederschlagshöhen [dm] im Jahr und im Winter, Karlsruhe, 1876–2008, jeweils mit Ausgleichspolynom vierten Grades (---), mit zentrierten gleitenden (10-Jahres) Durchschnitten (innere durchgezogene Linie) und mit dem Mittelwert über alle 133 Jahre (···)

Potsdam, Niederschlag 1893–2008



Potsdam, Niederschlag 1893–2008

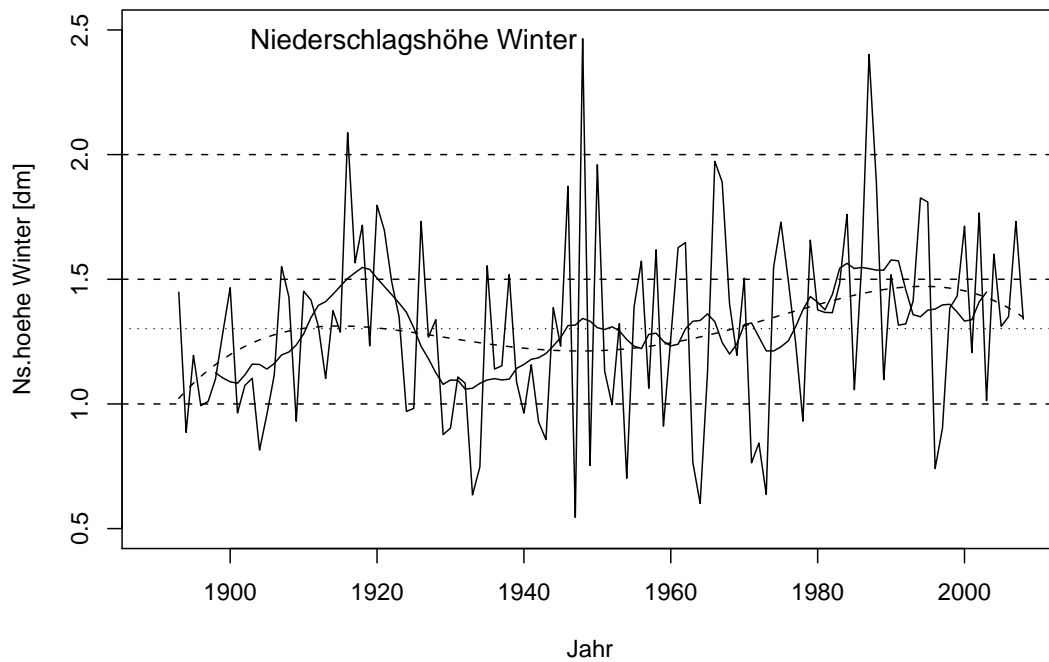


Abbildung 6: Niederschlagshöhen [dm] im Jahr und im Winter, Potsdam, 1893–2008, jeweils mit Ausgleichspolynom vierten Grades (---), mit zentrierten gleitenden (10-Jahres) Durchschnitts (innere durchgezogene Linie) und mit dem Mittelwert über alle 116 Jahre (···)

2 Temperatur-Trend und Temperatur-Anstieg

In diesem Abschnitt studieren wir den langfristigen Trend der Temperaturwerte.

2.1 Vergleich der beiden letzten Jahrhunderte

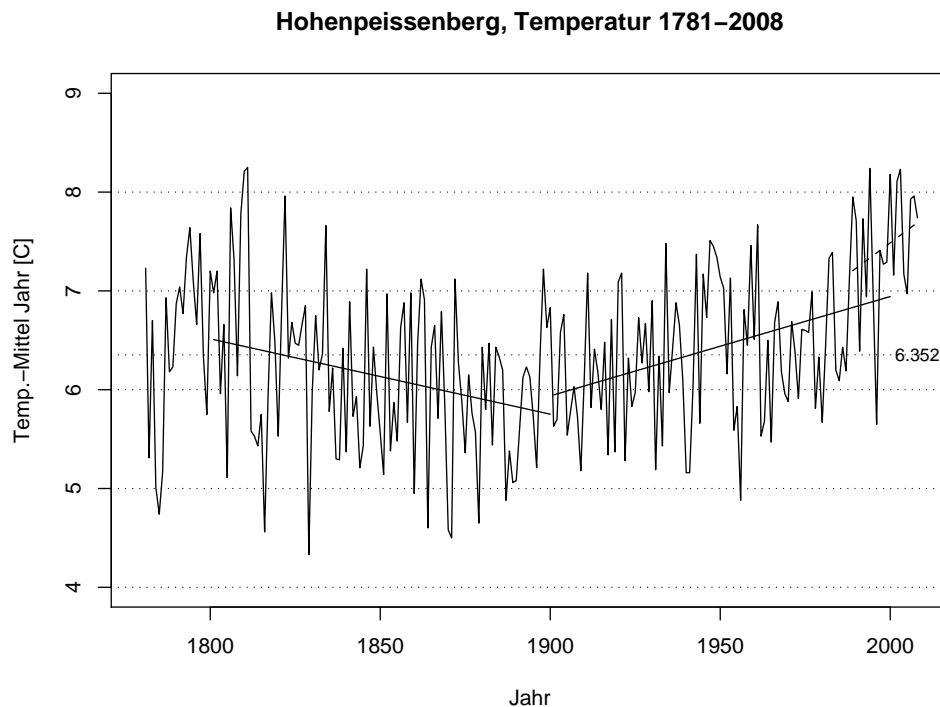


Abbildung 7: Mittlere Jahrestemperaturen [$^{\circ}\text{C}$] Hohenpeißenberg, 1781–2008, mit Ausgleichsgeraden pro Jahrhundert. Gestrichelt eingezeichnet ist die Ausgleichsgerade für die letzten 20 Jahre. Vgl. auch Schönwiese *et al*, 1993.

Während des 19ten Jahrhunderts fallen die mittleren Jahrestemperaturen ab, im 20ten Jahrhundert dagegen steigen sie wieder an, vgl. Abbildungen 7 und 8.

1. Die Regressionskoeffizienten (Steigungen) $b = b_{Temp|Jahr}$ bzw. die zugehörigen Korrelationskoeffizienten $r = b \cdot (s_{Jahr}/s_{Temp})$ der beiden –getrennt berechneten– Ausgleichsgeraden sind jeweils signifikant von Null verschieden: die $\alpha = 0.01$ -Schanke für den Betrag $|r|/\sqrt{1-r^2}$ unter der Nullhypothese „Steigung bzw. Korrelation = 0“ liegt bei $t_{98,0.995}/\sqrt{98} = 0.265$. Man vgl. dazu Tabelle 1 und Anhang A, 8.3 und 8.4. Das bedeutet, dass der negative Trend im 19ten Jhd. (in Karlsruhe) und der positive Trend im 20ten Jhd. (am Hohenpeißenberg und in Karlsruhe) statistisch signifikant sind.

2. Die Mittelwerte in den beiden Jahrhunderten fallen nicht sehr verschieden aus. Dagegen ist der Mittelwertanstieg in den letzten 20 Jahren gegenüber den vergangenen 2 Jahrhunderten signifikant, wie ein Zwei-Stichproben Test nach Anhang A, 8.2. sofort bestätigt.

Die Ergebnisse 1. und 2. sind auch noch nach einer Reduktion der Stichprobenumfänge, evtl. nötig aufgrund der Autokorrelation innerhalb der Klimareihen, gültig.

Karlsruhe, Temperatur 1799–2008

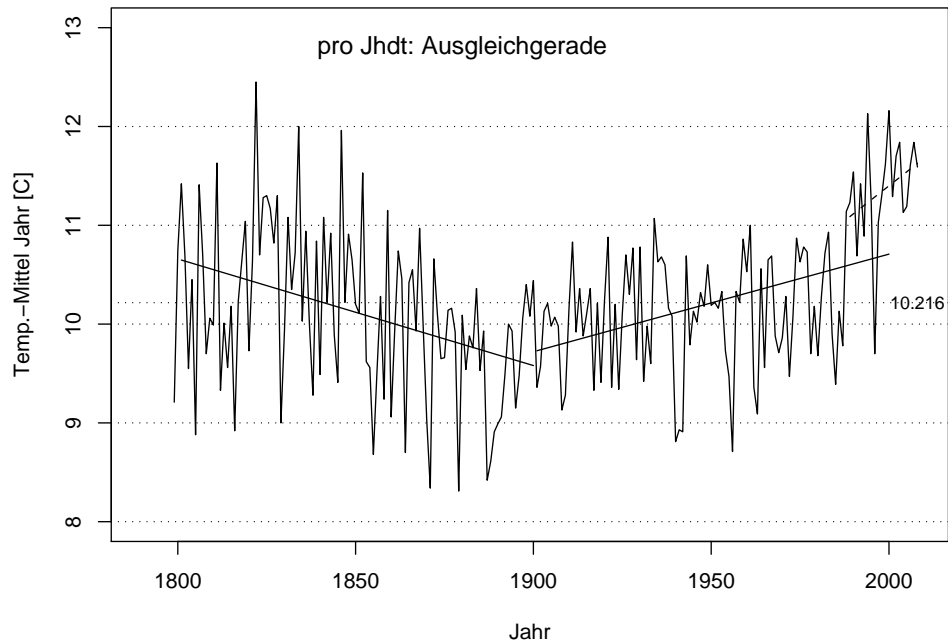


Abbildung 8: Mittlere Jahrestemperaturen [$^{\circ}\text{C}$] Karlsruhe, 1799–2008, mit Ausgleichsgeraden pro Jahrhundert. Gestrichelt eingezeichnet ist die Ausgleichsgerade für die letzten 20 Jahre.

R Plot der Temperatur-Jahresmittel, mit einer Ausgleichsgeraden pro 19tes und 20tes Jahrhundert. Der Programm-Output wird auf `C:/KLIMA/Tempout.txt` geschrieben und gespeichert.

```
attach(karlsTp)
postscript(file="C:/KLIMA/KarlsT12.ps",height=6,width=20,horizontal=F)

sink("C:/KLIMA/Tempout.txt")           #Output auf File Tempout.txt
Y<- Tjahr/100; "Temperatur 1799-2008"
cylim<- c(8.0,13.0); cabl<- c(8,9,10,11,12)
plot(Jahr,Y,type="l",lty=1,xlim=c(1790,2008),ylim=cylim,
     xlab="Jahr",ylab="Temp.-Mittel Jahr [C]",cex=1.3)
title(main="Karlsruhe, Temperatur 1799-2008",cex=1.0)
abline(h=cabl,lty=3); abline(h=mean(Y),lty=3)
text(2008,mean(Y),trunc(mean(Y)*1000)/1000,cex=0.8)

Janf<- 2;
A1<- Janf+1; B1<- Janf+100; "19.tes Jahrhundert"
Y1<- Y[A1:B1]; Jahr1<- Jahr[A1:B1]-1800      #Jahreszahlen verkleinern
tpger1<- lm(Y1~Jahr1); summary(tpger1)
lines(Jahr1+1800,predict(tpger1),lty=1)      #Ausgleichsgerade zeichnen

A2<- Janf+101; B2<- Janf+200; "20.tes Jahrhundert"
Y2<- Y[A2:B2]; Jahr2<- Jahr[A2:B2]-1900
```

```

tpger2<- lm(Y2~Jahr2); summary(tpger2)
lines(Jahr2+1900,predict(tpger2),lty=1) #Ausgleichsgerade zeichnen

detach(karlsTp)
dev.off()

```

Output Auszug aus den auf C:/KLIMA/Tempout.txt gespeicherten Ergebnissen. Die Wurzel aus R-squared = 0.137 ist der Betrag des Korrelationskoeffizienten in Tab. 1.

```

"19.tes Jahrhundert"
Call: lm(formula = Y1 ~ Jahr1)
Residuals:
    Min       1Q   Median       3Q      Max
-1.72517 -0.52365  0.05887  0.58981  2.02823
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.659115   0.159070  67.009 < 2e-16 ***
Jahr1       -0.010788   0.002735  -3.945  0.00015 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
Residual standard error: 0.7894 on 98 degrees of freedom
Multiple R-squared:  0.137, Adjusted R-squared:  0.1282
F-statistic: 15.56 on 1 and 98 DF,  p-value: 0.0001500

```

Dazu Folgendes: Bei der Anwendung von Tests und Konfidenzintervallen auf Zeitreihendaten muss auf Grund von Autokorrelationen der Stichprobenumfang (sagen wir n) zu einem *effektiven* Stichprobenumfang n_{eff} reduziert werden. Ein $(1 - \alpha) * 100$ % Konfidenzintervall z. B. für das langjährige Temperaturmittel (nennen wir es μ) lautet für große n , auf der Basis des beobachteten Mittelwertes \bar{y} und einer Autokorrelation $r \equiv r(1)$ (vgl. auch Anhang A, 8.1)

$$\bar{y} - u_0 \cdot \frac{s}{\sqrt{n_{eff}}} \leq \mu \leq \bar{y} + u_0 \cdot \frac{s}{\sqrt{n_{eff}}}, \quad u_0 = u_{1-\alpha/2}, \quad n_{eff} = n \cdot \frac{1-r}{1+r}.$$

Hohenpeißenberg: $n = 228$, $r = 0.289$, $\bar{y} = 6.352$, $s = 0.844$ führt zu $n_{eff} = 125.76$ und damit zum 99 % Konfidenzintervall [6.158, 6.546].

Karlsruhe: $n = 210$, $r = 0.332$, $\bar{y} = 10.216$, $s = 0.802$ führt zu $n_{eff} = 105.32$ und damit zum 99 % Konfidenzintervall [10.015, 10.417].

In beiden Fällen liegen mindestens 18 der letzten 20 Jahres-Mittelwerte oberhalb der oberen Konfidenzgrenze, was die Aussage in 2. bekräftigt.

Statistische Ergebnisse sind rein formale Aussagen. Nur der Klimatologe kann substanzwissenschaftliche Aussagen (wie „allgemeine Erderwärmung“ etc.) treffen. Insbesondere wäre ein „Verlängern“ der aufwärts weisenden Geradenstücke unseriös. Vor zu schnellen Verallgemeinerungen warnt auch der Blick auf die Temperaturschwankungen der letzten Jahrtausende.

Tabelle 1: Statistische Kenngrößen für die Temperaturen [$^{\circ}C$] der beiden Jahrhunderte und der letzten 20 Jahre

Zeitraum	Hohenpeißenberg				Karlsruhe			
	Mittel wert	Stand abw.	Regr.k. *100	Korrel.	Mittel wert	Stand abw.	Regr.k. *100	Korrel.
19tes Jhdt	6.129	0.843	-0.763	-0.262	10.114	0.845	-1.079	-0.370
20tes Jhdt	6.445	0.747	1.006	0.390	10.219	0.689	0.988	0.416
1989–2008	7.448	0.670	2.561	0.226	11.357	0.547	2.836	0.307

2.2 Historische Temperaturschwankungen

Der Verlauf der (geschätzten) mittleren Jahrestemperatur über die Jahrtausende mahnt zur Vorsicht in Hinblick auf generalisierende Behauptungen. Die Abbildung 9 zeigt die Temperaturschwankungen der letzten acht Jahrtausende nach Schönwiese (1979); vgl. auch Glaser (2001). Die Schätzungen erfolgten nach der Sauerstoff-Isotopenmethode aus Grönländischen Eisbohrkernen.

- KE: Kleine Eiszeit (1500 – 1700),
- MO: Mittelalterliches Optimum (800 – 1000),
- PV: Pessimum der Völkerwanderungszeit (450 – 800),
- OR: Optimum der Römerzeit (200 v. – 200 n.),
- P: Hauptpessimum (1200 v. – 600 v.).

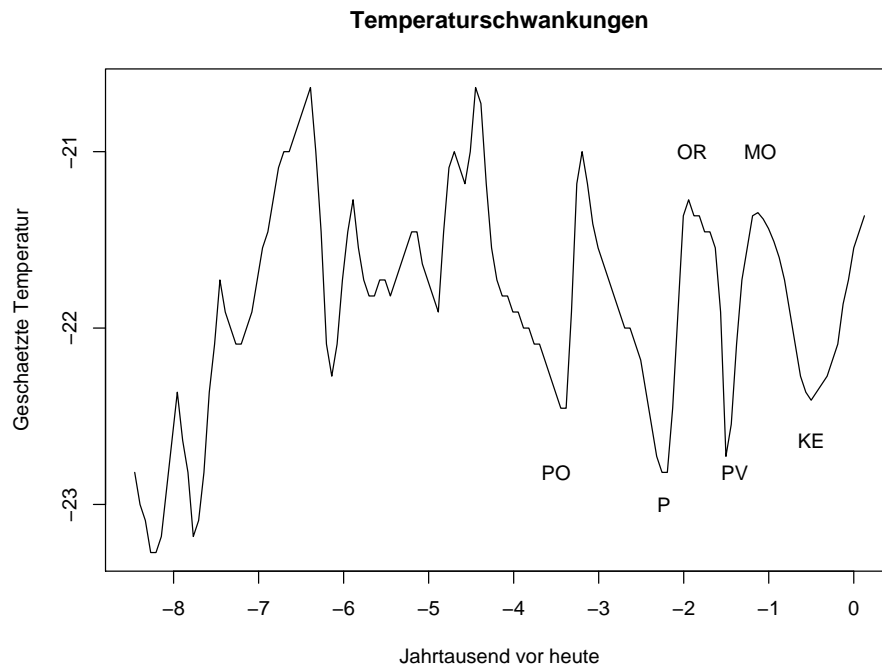


Abbildung 9: Temperaturschwankungen der letzten achttausend Jahre, qualitativer Verlauf; nach Schönwiese, 1979, Abb. 29

3 Monatswerte

Der Verlauf der Temperatur bzw. des Niederschlags über die 12 Monate des Jahres wird durch die Angabe der –über die Beobachtungsperiode gemittelten– 12 Monatswerte angegeben, vgl. die *Jahresgänge* der Abb 10. Diese sind bei den drei Temperaturverläufen ziemlich formgleich, mit einer etwas abgesenkten und gestauchten Form im Fall Hohenpeißenberg. Bei den Niederschlägen fallen die nassen Monate Juni und Juli am Hohenpeißenberg und die trockenen Monate Februar, März und Oktober in Potsdam ins Auge.

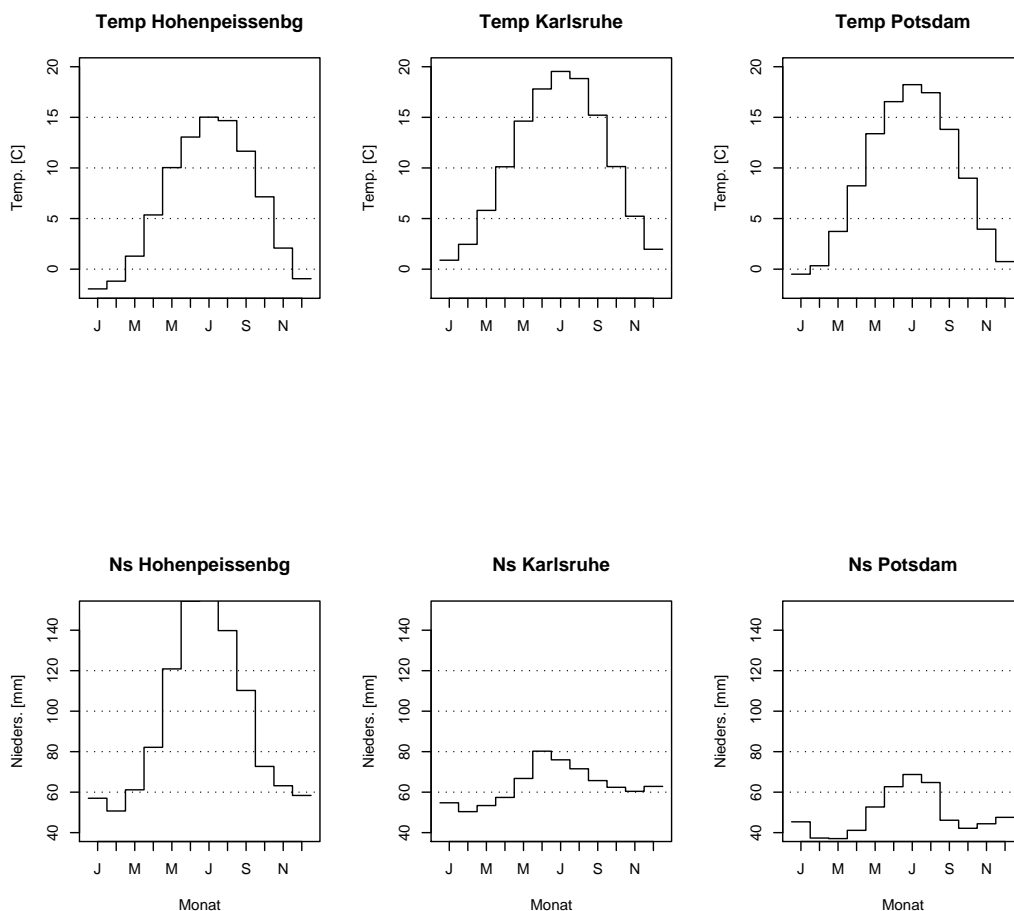


Abbildung 10: Die mittleren Monatstemperaturen (oben) und mittleren monatlichen Niederschlagshöhen in mm (unten), für die drei Stationen aufgetragen. Die abgeschnittenen Werte lauten: Hohenpeißenberg Niederschlag Juni 154.3, Juli 159.4

Nach Malberg (1997) entspricht der Jahresgang der Niederschlagssummen an den Stationen Hohenpeißenberg und Potsdam mehr einem kontinentalen und weniger einem ozeanischen Klimatyp.

R Sechs Stufen-Histogramme der Monatswerte für Temperatur und Niederschlagshöhe an drei Stationen. Innerhalb der Funktion `monaTN` wird die Funktion `plotTN` aufgerufen.

```
postscript(file="C:/KLIMA/MonatTN.ps",height=8,width=20,horizontal=F)
```

```

par(mfrow=c(2,3),pty="s")                                #2x3 quadratische Bilder

plotTN<- function(mo,ttext,cylim,tylab,cabl){
plmo<- 1:13; mon13<- mo[12]; plmo<-c(mo,mon13)
x<- seq(0.5,12.5,by=1)
plot(x,plmo,type="s",xlim=c(0.5,12.5),ylim=cylim,
     xaxt="n",xlab="",ylab=tylab)
axis(side=1,at=c(1,2,3,4,5,6,7,8,9,10,11,12),
     labels=c("J","F","M","A","M","J","J","A","S","O","N","D"))
title(main=ttext,cex=1.1)
abline(h=cabl,lty=3)
}

monaTN<- function(mon12,ttext,cylim,tylab,cabl){
mon12.mat<- as.matrix(mon12)                            #als Matrix anlegen
mon12.me<- colMeans(mon12)                             #Mittlere Monatswerte
plotTN(mon12.me,ttext,cylim,tylab,cabl)
}

cylim<- c(-2,20); tylab<- "Temperatur [C]"
cabl<- c(0,5,10,15,20)
mon12<- data.frame(hohenTp[,3:14])/10                 #jan-dez auswaehlen
monaTN(mon12,"Temp Hohenpeissenbg",cylim,tylab,cabl)
mon12<- data.frame(karlsTp[,3:14])/10
monaTN(mon12,"Temp Karlsruhe",cylim,tylab,cabl)
mon12<- data.frame(potsdTp[,3:14])/10
monaTN(mon12,"Temp Potsdam",cylim,tylab,cabl)

cylim<- c(40,150); tylab<- "Niederschlag [mm]"
cabl<- c(60,80,100,120,140)
mon12<- data.frame(hohenNs[,3:14])/10
monaTN(mon12,"Ns Hohenpeissenbg",cylim,tylab,cabl)
mon12<- data.frame(karlsNs[,3:14])/10
monaTN(mon12,"Ns Karlsruhe",cylim,tylab,cabl)
mon12<- data.frame(potsdNs[,3:14])/10
monaTN(mon12,"Ns Potsdam",cylim,tylab,cabl)

dev.off()

```

Um den Temperaturmittelwert eines bestimmten Monats zu beurteilen, vergleichen wir ihn mit demjenigen Wert, der durch die Trend- und Saison-Komponente bestimmt wird.

Dieser Vergleich soll durch die Abb. 11, welche die 36 mittleren Monatstemperaturen Y_t dreier aufeinanderfolgender Jahre präsentiert, illustriert werden. Die Trendkomponente \hat{m}_t ist mittels gleitender Durchschnitte über 13 Monate ermittelt, also jeweils unter Verwendung von sechs vorangehenden und sechs nachfolgenden Monaten. Die Saisonkomponente \hat{s}_t besteht aus den –auf Mittelwert Null zentrierten– Monatswerten

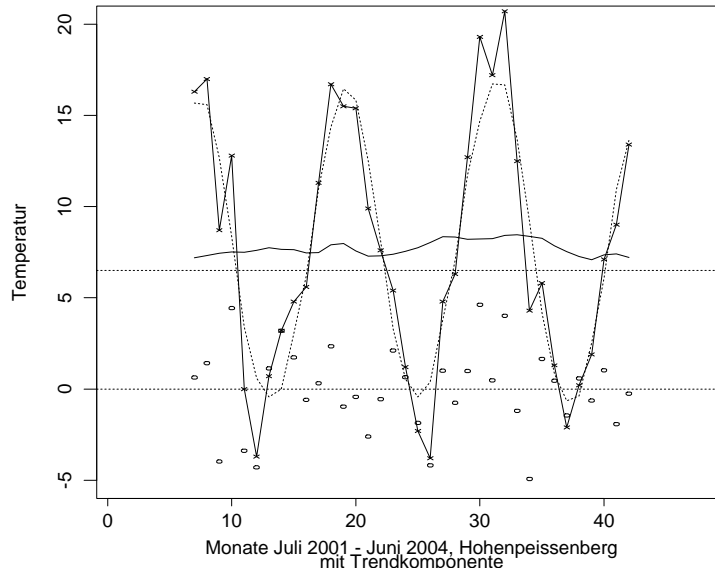


Abbildung 11: Die mittleren Monatstemperaturen Y_t (\times) über drei Jahre, Hohenpeissenberg, Juli 2001–Juni 2004. Zusätzlich mit Trendkomponente (—) und Trend+Saisonkomponente (\cdots), sowie den Residuen davon (o).

aus dem Diagramm der Abb. 10 links oben. Die Trend+Saisonkomponente ist dann

$$\hat{Y}_t = \hat{m}_t + \hat{s}_t \quad [Prädiktion].$$

Die *Residuen* $Y_t - \hat{Y}_t$ zeigen an, in welchen Monaten die Trend- und Saison-bereinigten Temperaturen zu hoch (dann mit positivem Residuum) oder zu niedrig (dann mit negativem Residuum) waren.

Der „Rekordsommer“ 2003 sticht mit überdurchschnittlich hohen Temperaturwerten im Juni und August ins Auge. Entsprechend sind die Residuenwerte deutlich positiv. Überdurchschnittlich kalt waren September, November und Dezember 2001, sowie ganz besonders der Oktober 2003 (Letzterer mit extrem negativem Residuum).

4 Korrelationen: Jahr, Jahreszeit, Monat, Tag

Wir präsentieren zunächst ein Beispiel für das Scattergramm und den zugehörigen Korrelationskoeffizienten. Diese sind für eine bivariate Stichprobe definiert, d. h. für zwei Größen x und y , die n mal (am gleichen Objekt) gemessen wurden; vgl. Anhang A, 8.3.

Bezeichne x die winterliche Niederschlagshöhe und y das winterliche Temperaturmittel. Welcher Zusammenhang besteht zwischen diesen Größen. Das Scattergramm der Abb. 12 (links) zeigt, dass sie am Hohenpeißenberg nahezu unkorreliert sind, während das Scattergramm (rechts) einen deutlichen positiven (linearen) Zusammenhang in Karlsruhe zu Erkennen gibt. Dabei ist zu beachten, dass Winterniederschläge am Hohenpeißenberg zu einem bedeutenden Teil aus (zur Messung dann getauten) Schnee bestehen, in Karlsruhe dagegen vorwiegend aus Regen. Zur Erinnerung: Der meteorologische Winter besteht aus dem Dez. des Vorjahres und aus Jan., Feb.

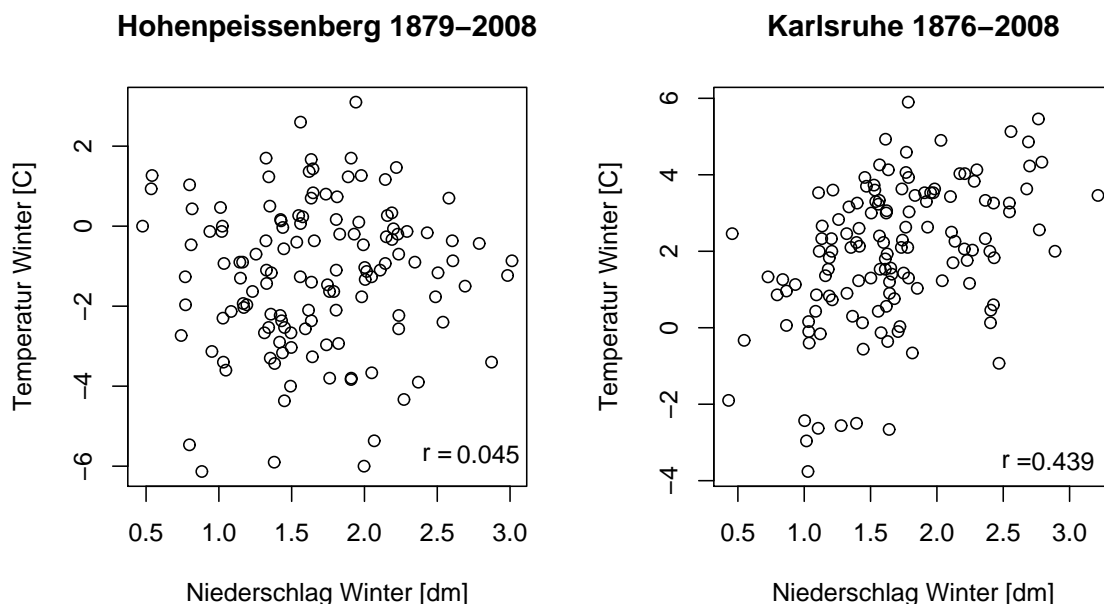


Abbildung 12: Scattergramm mit eingetragenen Korrelationskoeffizienten für winterliche Niederschlagshöhe und winterliches Temperaturmittel, für zwei Stationen.

4.1 Auto-Korrelationskoeffizient

Wie stark ist ein Messwert zum Zeitpunkt t (genannt: x) korreliert mit dem Messwert zum nachfolgenden Zeitpunkt $t + 1$ (genannt: y)?

Wir betrachten jetzt also den Fall, dass x und y die gleiche Messgröße darstellen (Temperatur bzw. Niederschlag), aber zu verschiedenen Zeitpunkten gemessen.

$$x = \text{Messwert}(t), \quad y = \text{Messwert}(t+1).$$

Das Scattergramm der Abb. 13 zeigt 12*228 mittlere Monatstemperaturen, einmal ohne (links) und einmal mit (rechts) Bereinigung von der Saisonkomponente (vgl. Abschnitt 3). Die deutlich ausgerichtete Punktwolke der linken Abbildung, mit dem hohen zugehörigen Korrelationskoeffizienten $r \equiv r(1) = 0.79$, erklärt sich zum Großteil durch den –nur allmählich sich verändernden– Jahresgang der Monatstemperaturen, enthält also, Letzteren vorausgesetzt, viel redundante Information.

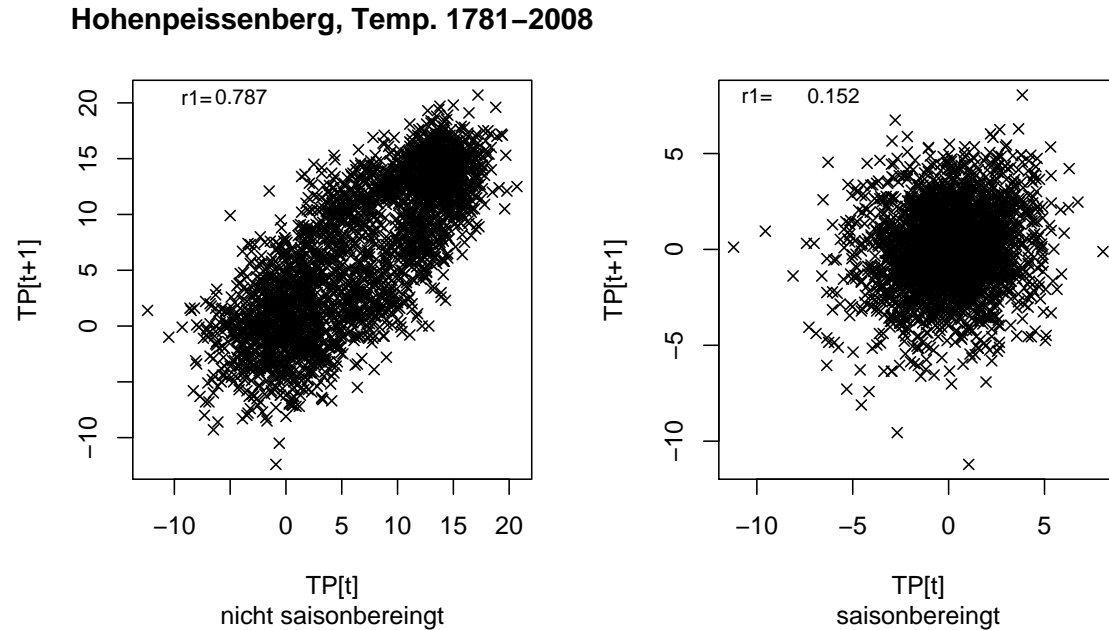


Abbildung 13: Mittlere Monatstemperaturen $TP = Y$. Scattergramms Y_{t+1} über Y_t mit $n=12*228$ Punkte, *links*: nicht bereinigt ($r=0.79$); *rechts*: saisonbereinigt ($r = 0.15$).

Die Tabellen 2 und 3 bringen Auto-Korrelationen

$$r(1) = r(Y_t, Y_{t+1})$$

von Klimawerten für aufeinanderfolgende Messzeitpunkte, jeweils ohne Bereinigung (in Klammern) und mit Bereinigung. Dabei handelt es sich um eine Bereinigung

- von der Trendkomponente (= Polynom vom Grade 4) im Fall von *Jahr*, *Quartal* und *Tag*, bzw.
- von der Saisonkomponente (= Monatswerte) im Fall von *Monat*.

R Korrelationen der Quartals-Temperaturen; nach Entfernung eines Polynomtrends der Ordnung 4.

```
attach(hohenTp)
```

```
n<- length(Jahr)
TWi<- dzvj+jan+feb; TFr<- mar+apr+mai
TSoc<- jun+jul+aug; The<- sep+okt+nov
```

```
"----Residuen vom Polynom(4)-Trend----"
```

```
Ja<- Jahr-1800; Ja2<- Ja*Ja; Ja3<- Ja2*Ja; Ja4<- Ja2*Ja2
```

Tabelle 2: Auto-Korrelationen $r(1)$ für Klimawerte (Hohenpeißenberg; letzte Zeile: Garching)

Prädiktion	Temperatur		Niederschlag	
	n	$r(Y_t, Y_{t+1})$	n	$r(Y_t, Y_{t+1})$
Jahr → nachf. Jahr	227	(0.289) 0.118	129	(0.273) 0.184
Winter → nachf. Winter	227	(0.077) 0.011	129	(0.144) -0.006
Sommer → nachf. Sommer	227	(0.198) 0.104	129	(-0.116) -0.168
Winter → nachf. Sommer	227	(0.162) 0.101	129	(0.219) 0.170
Sommer → nachf. Winter	227	(0.061) -0.016	129	(-0.025) -0.098
Monat → nachf. Monat	2735	(0.787) 0.152	1559	(0.377) 0.012
Tag → nachf. Tag	365	(0.952) 0.833	365	(0.135) 0.117

Tabelle 3: Auto-Korrelationen $r(1)$ für Klimawerte (Karlsruhe)

Prädiktion	Temperatur		Niederschlag	
	n	$r(Y_t, Y_{t+1})$	n	$r(Y_t, Y_{t+1})$
Jahr → nachf. Jahr	209	(0.332) 0.110	132	(0.009) 0.005
Winter → nachf. Winter	209	(0.113) 0.060	132	(-0.041) -0.082
Sommer → nachf. Sommer	209	(0.250) 0.064	132	(-0.201) -0.230
Winter → nachf. Sommer	209	(0.175) 0.121	132	(0.104) 0.127
Sommer → nachf. Winter	209	(0.119) 0.052	132	(-0.084) -0.067
Monat → nachf. Monat	2519	(0.811) 0.197	1595	(0.071) 0.029

```

TWres<- TWi-predict(lm(TWi~Ja+Ja2+Ja3+Ja4))
TFres<- TFr-predict(lm(TFr~Ja+Ja2+Ja3+Ja4))
TSres<- TSo-predict(lm(TSo~Ja+Ja2+Ja3+Ja4))
THres<- THe-predict(lm(THe~Ja+Ja2+Ja3+Ja4))
"1 / 2 entspricht vorangehendes / nachfolgendes Jahr"
TWres1<- TWres[1:(n-1)]; TWres2<- TWres[2:n]
TFres1<- TFres[1:(n-1)]; TFres2<- TFres[2:n]
TSres1<- TSres[1:(n-1)]; TSres2<- TSres[2:n]
THres1<- THres[1:(n-1)]; THres2<- THres[2:n]
Quar<- cbind(TWres1,TWres2,
             TFres1,TFres2,TSres1,TSres2,THres1,THres2)
cor(Quar)           #Kreuztabelle der Korrelationen
detach(hohenTp)

```

Output Kreuztabelle der Korrelationen. Bsp: $\text{cor}(\text{TWres1}, \text{TSres1})$ bezieht sich auf Winter und den direkt nachfolgenden Sommer, $\text{cor}(\text{TWres1}, \text{TSres2})$ auf Winter und den Sommer im darauffolgenden Jahr.

```

      TWres1 TWres2 TFres1 TFres2 TSres1 TSres2 THres1 THres2
TWres1  1.000  0.011  0.121  0.074  0.101  0.027 -0.072  0.032
TWres2  0.011  1.000  0.082  0.122 -0.016  0.100  0.086 -0.072
TFres1  0.121  0.082  1.000  0.073  0.163  0.128  0.162  0.104
TFres2  0.074  0.122  0.073  1.000  0.130  0.160  0.026  0.164
TSres1  0.101 -0.016  0.163  0.130  1.000  0.104  0.215 -0.044
TSres2  0.027  0.100  0.128  0.160  0.104  1.000  0.187  0.216

```

Tabelle 4: Bedingte Wahrscheinlichkeiten für das Übertreffen von Schwellenwerten

Bedingte Wahrscheinl.	Korrelation $\rho = \rho_{x,y}$							
	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70
$\mathbb{P}(Y > Q_{0.50}^y X > Q_{0.50}^x)$	0.50	0.53	0.56	0.60	0.63	0.67	0.70	0.75
$\mathbb{P}(Y > Q_{0.75}^y X > Q_{0.75}^x)$	0.25	0.29	0.34	0.40	0.45	0.51	0.57	0.64
$\mathbb{P}(Y > Q_{0.90}^y X > Q_{0.90}^x)$	0.10	0.14	0.17	0.24	0.29	0.39	0.45	0.53

THres1 -0.072 0.086 0.162 0.026 0.215 0.187 1.000 -0.097
 THres2 0.032 -0.072 0.104 0.164 -0.044 0.216 -0.097 1.000

Die $r(1)$ -Werte für „Tag“ wurden aus den 365 aufeinanderfolgenden Tageswerten des Jahres 1966 an der Station Garching b. München berechnet, vgl. Fiedler (1971) und Anhang B.

Nur die Trend- (Saison-) bereinigten Reihen –d. h. nur die nicht eingeklammerten Koeffizienten der Tabelle– werden in der folgenden Diskussion betrachtet.

Temperatur: Die Auto-Korrelation bei den Tageswerten ist –erwartungsgemäß– sehr hoch, weniger hoch bei den Monatswerten, bei den Jahreszeit-Werten und bei den Jahreswerten. Die Jahreszeitwerte sind (bis auf *Winter* → *nachf. Sommer* in Karlsruhe) kleiner als die Werte *Jahr* → *nachf. Jahr*.

Niederschlag: Nur am Hohenpeißenberg, und zwar bei den Jahreswerten und im Fall *Winter* → *nachf. Sommer* (pos.) und *Sommer* → *nachf. Sommer* (neg.) sind die Auto-Korrelationen deutlich von Null verschieden. Bei den Monatswerten kann sie vernachlässigt werden, und bei den Tageswerten ist sie auffallend gering. Die letzten beiden Aussagen sind ganz verschieden zur Situation bei der Temperatur.

Was bedeutet ein bestimmter $r(1)$ -Wert in Hinblick auf die Prädiktion?

4.2 Korrelation und die Prädiktion von Extremen

Was bedeutet ein bestimmter Autokorrelationswert $r(1)$ in Hinblick auf die Frage, mit welcher Wahrscheinlichkeit der –auf einen „extremen“ Klimawert *folgende* Wert– ebenfalls „extrem“ sein wird. Das Wort *extrem* ist hier im schwachen Sinn von *überdurchschnittlich* gemeint (Genauerer folgt sofort).

Im Folgenden wird nach der Wahrscheinlichkeit gefragt, dass auf einen X-Wert, der größer als ein Schwellenwert Q^x ist, ein Y-Wert folgt, der größer als ein Q^y ist. Da der vorangehende Wert als gegeben betrachtet wird, spricht man von bedingten Wahrscheinlichkeiten. Falls also der X-Wert größer als Q^x ist, dann ergeben sich (aufgefächert nach dem Koeffizienten $\rho = \rho_{x,y}$ der Korrelation zwischen X und Y) die Treffer-Wahrscheinlichkeiten \mathbb{P} der Tabelle 4 dafür, dass auch der Y-Wert größer als Q^y ist. Der Anhang A (8.5 mit der Abb. 31) dient zur Erklärung und Veranschaulichung.

Als Schwellenwerte haben wir *Quantile* Q_γ (auch $\gamma \cdot 100$ % Perzentile genannt) gewählt, und zwar für $\gamma = 0.5, 0.75, 0.90$. Diese Schwellen könnte man –der Reihe nach– auch Durchschnittswert, oberer 25 %-Wert, oberer 10 %-Wert, nennen.

Beispiele: • Ist X größer als der X-Durchschnitt $Q_{0,50}^x$, also ein *über*-durchschnittlicher Wert (genauer: ein „oberer 50-Prozent-Wert“), dann ist die Trefferquote dafür, dass auch Y größer als der Y-Durchschnitt ist, gleich

50 % bei $\rho = 0$; 60 % bei $\rho = 0.30$; 70 % bei $\rho = 0.60$.

• Ist X größer als der Wert $Q_{0,90}^x$, also ein „oberer 10-Prozent-Wert“, dann beträgt die Trefferquote dafür, dass auch Y ein solcher ist, gleich

10 % bei $\rho = 0$; 24 % bei $\rho = 0.30$; 45 % bei $\rho = 0.60$.

4.3 Anwendung auf Klimadaten

Nur die Trend- (Saison-) bereinigten Reihen werden betrachtet, d. h. nur die nicht eingeklammerten Werte der Tabellen 2 und 3 werden diskutiert.

Die meisten Autokorrelationen $r(1)$ liegen zwischen 0.0 und 0.2. Die Trefferrate bei der Vorhersage von überdurchschnittlichen Werten (bei Vorliegen eines solchen) liegt nach Aussage von Tabelle 4 zwischen 50 % und 56 % (50 % sind es auch beim Raten mittels „Münzwurf“).

Allein bei den Tagestemperaturen ($r(1) > 0.70$) liegt diese Trefferrate über 75 %. Liegt ein oberer „10 % Tag“ vor, so kann das immerhin noch mit einer Trefferrate von über 53 % für den nachfolgenden Tag ausgesagt werden (gegenüber 10 % beim „Raten“).

Bei den Tages-Niederschlägen ist der Koeffizient $r(1)$ nicht sehr aussagekräftig. Denn fast die Hälfte der Tage ist niederschlagsfrei. Bei Dichotomisierung (Niederschlag nein oder ja, d. h. NSchl=0 oder >0) erhalten wir für die 365 Tage in Garching 1966 die folgende 2×2 -Häufigkeitstafel

heutiger Tag	nachfolgender Tag		Σ
	NSchl=0	NSchl>0	
NSchl=0	99 0.604	65 0.396	164 1.0
NSchl>0	65 0.325	135 0.675	200 1.0
Σ	164	200	364

Bleibt ein Tag niederschlagsfrei, so können wir mit einer Trefferquote von $\frac{99}{164} \cdot 100 \approx 60\%$ dasselbe vom nachfolgenden Tag prophezeien.

Diese 60 % sind zu vergleichen mit 50 % beim „Münzwurf“, bzw. mit $164/364 \cdot 100 \approx 45\%$ bei Berücksichtigung der Häufigkeiten von trockenen/nassen Tagen.

Bei der Evaluierung von „Wetterregeln“ (bez. Temperatur, Niederschlag) kommen Zahlenwerke vom Typ der Tab. 2 oder der obigen 2×2 -Tafel zu Bedeutung.

4.4 Wetterregeln

Wetter- oder auch *Bauernregeln* beziehen sich nach Malberg (2003) meistens auf

- ein bestimmtes Gebiet (wahrscheinlich nicht gerade der Hohe Peißenberg),
- auf eine bestimmte Epoche (hier wird dagegen über Jahrhunderte gemittelt),
- auf Ernteertrag (während uns heute meistens das Urlaubswetter interessiert).

Die früheren bäuerlichen oder klösterlichen Beobachter (ohne geeignete Mess-, Auf-

Tabelle 5: Trefferquoten zu den Wetterregeln 1–6

Bsp	$X \rightarrow Y$		$r(X, Y)$	\mathbb{P}	% $[Y > \bar{y} X > \bar{x}]^1$			
			Hohen	(Tab 4)	Berlin	Karls	Hohen	
1	Tp Dez	Tp Jan	0.13	.54	[> >]	70 %	58 %	57 %
1	Tp Dez	Tp Feb	0.11	.53	[> >]	60 %	62 %	56 %
2	Tp Sep	Tp Okt	0.14	.54	[> >]	62 %	55 %	56 %
				.54	[< <]	62 %	54 %	51 %
3	Tp Nov	Tp Mai	0.05	.51	[> >]	50 %	55 %	52 %
	Ns Nov	Ns Mai	-0.02	.50	[> >]	50 %	42 %	42 %
4	Tp Aug	Tp Feb	0.08	.53	[> >]	73 %	56 %	53 %
	Ns Aug	Ns Feb	0.04	.51	[> >]	50 %	50 %	47 %
5	Tp Som	Tp Win	0.06	.51	[> >]	—	54 %	54 %
				.49	[< >]	—	46 %	46 %
	Ns Som	Ns Win	-0.02	.50	[> >]	—	41 %	42 %
				.50	[< >]	—	59 %	58 %
6	Tp Win	Tp Som	0.16	.55	[> >]	—	51 %	55 %
				.45	[> <]	—	45 %	38 %
	Ns Win	Ns Som	0.22	.57	[> >]	—	50 %	62 %
				.43	[> <]	—	43 %	42 %

¹ \diamond steht für ein > oder ein < Zeichen.

zeichnungs-, Auswertungs-Vorrichtungen) sind gewissermaßen Pioniere der Wetterprädiktion und sind entsprechend zu würdigen. Bequem dagegen machen es sich die Nachbeter in der heutigen Zeit, die sich mit der Aura der Tradition umgeben, statt Datenbanken und Auswertungs-Software zu bemühen.

Erhaltungs-(Beharrungs-)Regeln

Bsp 1: Ist Dezember lind \rightarrow der ganze Winter ein Kind

Bsp 2: Kühler September \rightarrow kalter Oktober

Sechs-Monats-Regeln

Bsp 3: Der Mai kommt gezogen \leftarrow wie der November verflogen

Bsp 4: Wie der August war \rightarrow wird der künftige Februar

Jahres-Ausgleichs-Regeln

Bsp 5: Wenn der Sommer warm ist \rightarrow so der Winter kalt

Bsp 6: Wenn der Winter kalt ist \rightarrow so der Sommer warm

Aussage der Klimadaten.

Zu den Regeln der obigen Beispiele 1 – 6 sind in den Spalten der Tabelle 5 angegeben:

- Umschreiben der Regel, mit Tp für Temperatur und Ns für Niederschlag.
- Korrelationskoeffizient r aus den Hohenpeißenberg-Daten.
- Die (zum Wert der Korrelation r gehörende) bedingte Wahrscheinlichkeit $\mathbb{P}(Y > Q_{0.5}^y | X > Q_{0.5}^x)$ aus der Tabelle 4.
- Prozentualer Anteil % $[Y > \bar{y} | X > \bar{x}]$ der Fälle, in denen auf einen überdurchschnittlichen X-Wert auch ein überdurchschnittlicher Y-Wert folgt, und zwar für Berlin-

Dahlem, Karlsruhe, Hohenpeißenberg.

Im Beispiel 2 zielt die Regel auf den Prozentsatz $\% [Y < \bar{y} | X < \bar{x}]$ ab, im Beispiel 5 auf $\% [Y < \bar{y} | X > \bar{x}]$, im Beispiel 6 auf $\% [Y > \bar{y} | X < \bar{x}]$. Diese Prozentzahlen sind ebenfalls –zusätzlich zu den Prozentzahlen $\% [Y > \bar{y} | X > \bar{x}]$ – angegeben. Ferner fügen wir bei den Beispielen 5 und 6 entsprechende Ergebnisse für Ns an.

Die Berliner Werte sind den Dahlemer Aufzeichnungen 1908–1987 entnommen, siehe Mahlberg (2003). Die Hohenpeißenberger und Karlsruher Werte entstammen den hier vorgestellten und benutzten Klimareihen.

Die aus den Hohenpeißenberger (und ähnlich aus den Karlsruher) Reihen ermittelten prozentualen Trefferquoten sind sehr mäßig und können die Regeln nicht bestätigen. Allenfalls Erhaltungsregeln finden eine (schwache) Bestätigung. In einigen Fällen wird eher eine andere Version der Regel (in Bsp 2) oder sogar das Gegenteil der Regel (in Bspen 5,6) nahegelegt. Mit einer Ausnahme (zwei Ausnahmen) bringen die Dahlem-Reihen höhere Trefferquoten hervor als die vom Hohenpeißenberg (aus Karlsruhe). Das kann daran liegen, dass die Ersteren kürzer sind und vielleicht auch klimatisch näher an den Entstehungsorten der Regeln liegen.

R Berechnung von Trefferquoten zu den Wetterregeln 1–2.

```
attach(hohenTp)
#Anzahl Ueberdurchschnittliche / Unterdurchschnittliche Faelle
BedHfkt<- function(x,y){
x0<- x[x > mean(x)]
xy0<- y[x > mean(x) & y > mean(y)]
x1<- x[x < mean(x)]
xy1<- y[x < mean(x) & y < mean(y)]
c("x> "=length(x0), "x>&y> "=length(xy0), "y>|x> "= length(xy0)/length(x0),
  "x< "=length(x1), "x<&y< "=length(xy1), "y<|x< "=length(xy1)/length(x1))
}
x<- dzvj/10; y <- jan/10; "x=Dez -> y=Jan"
c("mean(x)"=mean(x), "mean(y)"=mean(y), "cor(x,y)"=cor(x,y))
BedHfkt(x,y)
x<- sep/10; y <- okt/10; "x=Sep -> y=Okt"
c("mean(x)"=mean(x), "mean(y)"=mean(y), "cor(x,y)"=cor(x,y))
BedHfkt(x,y)
detach(hohenTp)
```

Output $y > |x >$ bzw. $y < |x <$ gibt die relative Anzahl überdurchschnittl. / unterdurchschnittl. Fälle an, die auf überdurchschnittl. / unterdurchschnittl. Fälle folgen.

```
"x=Dez -> y=Jan"
mean(x) mean(y) cor(x,y)
-0.9487 -1.9491 0.1322
  x>   x>&y>   y>|x>       x<   x<&y<   y<|x<
110.0000 63.0000 0.5727 118.0000 58.0000 0.4915
"x=Sep -> y=Okt"
mean(x) mean(y) cor(x,y)
11.6526 7.1469 0.1388
  x>   x>&y>   y>|x>       x<   x<&y<   y<|x<
119.0000 67.0000 0.5630 109.0000 56.0000 0.5138
```

Tabelle 6: Differenzen X der aufeinanderfolgenden mittl. Jahrestemp. [$^{\circ}\text{C}$]

Station	N	Mittelw.	Stand.Abw.	$r(1)$	$r(2)$	$r(3)$
Hohenp.	228	0.002	1.002	-0.469	0.019	-0.076
Karlsru.	210	0.011	0.921	-0.489	0.057	-0.052
Potsd.	116	0.018	0.924	-0.420	0.074	-0.197

5 Modell und Prädiktion: Jahreswerte

Im Folgenden sollen statistische Modelle besprochen werden, die (i) Gesetzmäßigkeiten in der Evolution der Klimazeitreihen aufdecken können und möglicherweise (ii) bei der Prädiktion zukünftiger Klimawerte von Nutzen sind. Im Zentrum stehen dabei Zeitreihenmodelle vom Typ ARMA, vgl. Anhang A, 8.8.

5.1 Differenzbildung, Prädiktionen

Es bezeichne Y die Zeitreihe der –über N Jahre sich erstreckenden– Jahresmittelwerte; gegeben also

$$Y(t), \quad t = 1, \dots, N.$$

Im Zusammenhang mit Zeitreihenmodellen vom Typ ARMA (und dem Problem der Prädiktion) wird die Entfernung des Trends gerne durch die Differenzbildung aufeinander folgender Werte vorgenommen. Von der Zeitreihe Y geht man so zur Zeitreihe X über, mit

$$X(t) = Y(t) - Y(t - 1), \quad t = 2, \dots, N, \quad [X(1) = 0]. \quad (1)$$

Die Wirkung dieser Differenzbildung wird durch einen Vergleich des oberen mit dem unteren Plot der Abbildung 14 deutlich. Tabelle 6 offenbart, dass die jährlichen Änderungen X im Mittel ≈ 0 sind und eine mittlere Abweichung (vom Mittel 0) der Größe $\approx 1[^{\circ}\text{C}]$ haben. Diese Differenzzeitreihen X weisen eine Autokorrelation $r(1)$ (also der Ordnung 1) von ca. $-0.4 \dots -0.5$ auf: Einem Anstieg der Temperatur folgt tendentiell sofort wieder ein Absinken.

Wir erklären jetzt die Differenzzeitreihe $X(t)$ für ausreichend „trendfrei“ und passen ihr ein ARMA(p,q)-Modell an, welches also der Gleichung

$$X(t) = \alpha_p X(t - p) + \dots + \alpha_2 X(t - 2) + \alpha_1 X(t - 1) + \beta_q e(t - q) + \dots + \beta_1 e(t - 1) + e(t) \quad (2)$$

gehört. Zu jedem Zeitpunkt t lässt sich dann –wie in A 8.8 erläutert– auf der Basis der vorangegangenen Werte $X(t - 1), X(t - 2), \dots$ ein Prädiktionswert $\hat{X}(t)$ für den nächstfolgenden Wert $X(t)$ berechnen (ARMA-Prädiktion genannt). Die Güte der Prädiktion und damit des zugrunde gelegten ARMA(p,q)-Modells bemessen wir aus der Summe der mittleren Fehlerquadrate, genauer aus dem Werten

$$RootMSQ = \sqrt{\frac{1}{N} \sum_{t=1}^N (X(t) - \hat{X}(t))^2}. \quad (3)$$

Tabelle 7: ARIMA-Methode für die mittleren Jahrestemperaturen

Sta.	Ordng p, q	Koeffizienten		Root	ARIMA-Prädiktion	
		α_i	β_j	MSQ	2006–2008	2009
Ho	2,2	-0.639,0.105	-0.177,-0.667	0.768	7.25, 7.40, 7.49	7.47
Ka	2,2	-0.266,0.070	-0.567,-0.281	0.707	11.25,11.33,11.42	11.44
Po	3,1	0.155,0.044,-0.238	-0.915	0.726	9.44, 9.69, 9.82	9.70

Von der Differenzenreihe aus gelangt man durch Summation (Integrieren) zu der (trend-behafteten) Ausgangsreihe Y zurück. Die Prädiktion $\hat{Y}(t)$ für $Y(t)$ berechnet sich rekursiv zu

$$\hat{Y}(t) = Y(t-1) + \hat{X}(t), \quad t = 2, \dots, N; \quad \hat{Y}(1) = Y(1).$$

Wegen $X(t) - \hat{X}(t) = Y(t) - \hat{Y}(t)$ besitzt die Prädiktion $\hat{Y}(t)$ für $Y(t)$ die gleiche Güte $RootMSQ$ wie $\hat{X}(t)$ für $X(t)$, nämlich, vgl. Gleichung (3),

$$RootMSQ = \sqrt{\frac{1}{N} \sum_{t=1}^N (Y(t) - \hat{Y}(t))^2}. \quad (4)$$

Insgesamt sprechen wir hier von der ARIMA-Methode, die \hat{Y} -Werte werden i. F. ARIMA-Prädiktionen genannt.

5.2 Mittlere Jahrestemperaturen

$Y(t)$ bezeichnet jetzt die mittlere Temperatur im Jahr t und $X(t)$ nach Gleichung (1) die Reihe der jährlichen Änderungen. Letzterer soll ein ARMA(p, q)-Modell angepasst werden.

Es wurden (möglichst kleine) Ordnungszahlen (p, q) gewählt, so dass eine weitere Erhöhung dieser Zahlen keine wesentliche Verringerung der Güte-Maßzahl $RootMSQ$ erwirkt. Für Hohenpeißenberg und Karlsruhe erhält man $p = 2, q = 2$ und damit das ARMA(2,2)-Modell

$$X(t) = \alpha_2 X(t-2) + \alpha_1 X(t-1) + \beta_2 e(t-2) + \beta_1 e(t-1) + e(t) \quad (5)$$

(für Potsdam ergibt sich $p = 3, q = 1$). Die Tabelle 7 zeigt die errechneten Koeffizienten α_i und β_j . I. d. R. sind mindestens ein α und ein β signifikant von Null verschieden. Ferner gibt die Tabelle die Prognosen für die drei Jahre 2006–2008 und für das anstehende Jahr 2009 wieder, jeweils auf der Basis der vorangehenden Jahre. Ein Vergleich mit den Daten im Anhang B bringt zu Tage, dass die ARIMA-Methode die tatsächlich beobachteten Werte für die Jahre 2006–2007 leicht unterschätzt. Die Prognosen haben insgesamt einen ausgleichenden Charakter, vgl. die Abbildungen 15 und 16. Diese beschränken sich –einer deutlicheren Darstellung wegen– auf die Wiedergabe der letzten 50 Jahre (zur Berechnung der Koeffizienten α und β wurde aber die Gesamtreihe verwendet).

5.2.1 Zum Vergleich: Gleitende Durchschnitte

Alternativ kann die Prognosemethode der gleitenden Durchschnitte gewählt werden. Als Prädiktion $\hat{Y}(t)$ (für $Y(t)$ zum Zeitpunkt t) wird der Durchschnittswert der *vorangegangenen* Werte $Y(t-1), Y(t-2), \dots, Y(t-k)$ genommen. Die Tiefen-Zahl k gibt also

Tabelle 8: Tiefe k des gleitenden Durchschnitts und resultierender Gütewert $RootMSQ$

Tiefe k	$RootMSQ$		
	Hohenp.	Karlsru.	Potsdam
5	0.819	0.747	0.836
8	0.802	0.717	0.796
10	0.790	0.707	0.780
12	0.784	0.715	0.791
14	0.768	0.708	0.788
16	0.763	0.699	0.782
18	0.765	0.698	0.782
20	0.774	0.713	0.796
ARIMA(s. Tab. 7)	0.768	0.707	0.726

die Anzahl der verwendeten Jahre an. Wieder nach Gleichung (4) berechnen wir die Güte dieser Prädiktionsmethode. Tabelle 8 zeigt, dass erst für eine Tiefe k größer 10 die $RootMSQ$ -Werte der ARIMA-Methode erreicht werden (die mit $p+q = 4$ Kenngrößen auskommt –hier allerdings jeweils aus der Gesamtdatei ermittelt). Im Fall Potsdam werden die ARIMA-Werte gar nicht erreicht. Der obere Plot der Abb. 14 zeigt die Prädiktionen nach dem gleitenden Durchschnitt für den Wert $k = 10$.

R ARMA-Modell für die Differenzenreihe, Prädiktion für Differenzenreihe und für integrierte Reihe (=ARIMA-Prädiktion), Residuenanalyse.

```
library(TSA) #s. Cran-Software-Packages
#-----Daten bereiten, Differenzenreihe bilden-----
attach(potsdTp)

mon12<- data.frame(potsdTp[,3:14])/10; Yja<- rowMeans(mon12)
#Jahresmittel-Werte, genauer als die Alternative Yja<- Tjahr/100
N<- length(Jahr)
c("Anz Jahre"=N,"MeanYja"=mean(Yja),"StdevYja"=sqrt(var(Yja)))
Dj<- Yja; Dj[1]<- 0; Dj[2:N]<- Yja[2:N]-Yja[1:(N-1)] #Diff.reihe

c("Anz Jahre"=N,"MeanDYja"=mean(Dj),"StdevDYja"=sqrt(var(Dj)))
dacf<- acf(Dj,lag.max=8,type="corr",plot=F) #kein Plot
"Autokorrelation Differenzenreihe"; dacf$acf

ma<- 3; mb<- 1; mc<- pmax(ma,mb) #ARMA-Ordnungen

# ----- ARMA(p,q)-Modell fuer Differenzenreihe -----
darma<- arma(Dj,order=c(ma,mb)); summary(darma)
ds2<- darma$css; dcoef<- darma$coef; theta<- dcoef[ma+mb+1]
dres<- darma$residuals; dfit<- darma$fitted.values
a<- rep(0,times=12); b<- rep(0,times=12)
for (m in 1:ma){a[m]<- dcoef[m]} #alpha-Koeff.
for (m in 1:mb){b[m]<- dcoef[ma+m]} #beta-Koeff.
c("ArOrder"=ma,"MAOrder"=mb,"SSQ"=ds2,"intercept"=theta)
```

```
# ----- ARMA(p,q)-Praediktion der Differenzenreihe und Integration
DarmaNY<- theta
for (m in 1:mc){ DarmaNY<- DarmaNY + a[m]*Dj[N+1-m] + b[m]*dres[N+1-m] }
c("Forecast DNewYear"=DarmaNY)
YIarma<- Yja; YIarma[2:N]<- Yja[1:(N-1)]+dfit[2:N] # Integr. Reihe
"ARIMA-Praediktion Letzte Dekade"; YIarma[(N-9):N]
c("Forecast NewYear"=Yja[N],DarmaNY,DarmaNY+Yja[N]) # Integr. Reihe

#----- Residuenanalyse, Mean Outputs -----
Drer<- dres[(mc+1):N]; rsq<- mean(Drer*Drer)
c("MeanDpred"=mean(dfit[(mc+1):N]), "StdDpred"=sqrt(var(dfit[(mc+1):N])),
  "MeanDres"=mean(Drer), "StdDres"=sqrt(var(Drer)))
c("Differenz MSQ"=rsq, "RootMSQ"=sqrt(rsq))
racf<- acf(Drer,lag.max=8,type="corr",plot=F)
"Autokorrelation ARIMA-Residuen = ARMA-Resid der Differenzen"; racf$acf

YIres<- Yja - YIarma #Test: YIres = dres s.o.
YIrer<- YIres[(mc+1):N]; risq<- mean(YIrer*YIrer)
c("Integr MSQ"=risq, "RootMSQ"=sqrt(risq))

detach(potsdTp)
```

Output ARMA-Kenngrößen für Differenzenreihe, Praediktion für Differenzenreihe und für Integrierte Reihe, Residuen mit RootMSQ

```
Anz Jahre MeanDYja StdevDYja
116.00000 0.01796 0.92412
[1] "Autokorrelation Differenzenreihe"
[1,] -4.202e-01
[2,] 7.444e-02
[3,] -1.972e-01
... ..
```

Model: ARMA(3,1) Coefficient(s):

	Estimate	Std. Error	t value	Pr(> t)
ar1	0.1553	0.0944	1.65	0.100 .
ar2	0.0443	0.0954	0.46	0.643
ar3	-0.2383	0.0942	-2.53	0.011 *
ma1	-0.9149	0.0476	-19.23	<2e-16 ***
intercept	0.0156	0.0068	2.29	0.022 *

Fit: sigma² estim. as 0.531, Cond. Sum-of-Squares = 59.54, AIC = 265.8

ArOrder	MAOrder	SSQ	intercept
3.00000	1.00000	59.54390	0.01560

```
Forecast DNewYear
-0.5138
[1] "ARIMA-Praediktion Letzte Dekade"
[1] 9.668 9.388 9.542 9.188 9.221 9.577 9.408 9.439 9.691 9.824
```

Tabelle 9: Autokorrelationsfunktion $r_e(h)$ bis *time lag* $h = 8$ für die ARIMA-Residuen, mittlere Jahrestemperatur

	$r_e(1)$	$r_e(2)$	$r_e(3)$	$r_e(4)$	$r_e(5)$	$r_e(6)$	$r_e(7)$	$r_e(8)$	b_1	b_8
Ho	0.014	-0.003	-0.012	0.035	-0.059	-0.038	-0.003	-0.056	0.130	0.181
Ka	0.005	0.016	-0.022	-0.071	-0.016	-0.009	0.045	-0.033	0.135	0.189
Po	0.007	0.009	0.004	-0.053	-0.008	-0.102	0.075	0.009	0.182	0.254

Forecast NewYear

10.2167 + -0.5138 = 9.7029

MeanDpred StdDpred MeanDres StdDres

0.002125 0.603713 0.018893 0.728891

Differenz MSQ RootMSQ

0.5269 0.7259

[1] "Autokorrelation ARIMA-Residuen = ARMA-Resid der Differenzen"

[1,] 0.006674

[2,] 0.009542

[3,] 0.003875

[4,] -0.052651

[5,] -0.007806

[6,] -0.102221

[7,] 0.074662

[8,] 0.009001

Integr MSQ RootMSQ

0.5269 0.7259

#identisch zu oben!

5.2.2 ARIMA-Residuen, GARCH-Modell

Bezeichnet $\hat{Y}(t), t = 1, \dots, N$ die ARIMA-Prädiktionen für $Y(t), t = 1, \dots, N$, so bilden

$$e(t) = Y(t) - \hat{Y}(t), \quad t = 1, \dots, N,$$

die Residuen von der Prädiktion; vgl. die oberen Plots in Abbildungen 15 und 16. Residuen wurden bereits oben in Gleichung (4) verwendet; wie schon festgestellt, gilt auch $e(t) = X(t) - \hat{X}(t)$. Welche Struktur hat diese Residuen-Zeitreihe $e(t), t = 1, \dots, N$. Die Autokorrelationsfunktion $r_e(h), h = 1, \dots, 8$, besitzt nur Werte nahe 0, vgl. Tab. 9. Die simultane Signifikanzschranke zum Niveau 0.05 für das Maximum von $|r_e(h)|, h = 1, \dots, 8$, beträgt (in Hinblick auf die Nullhypothese einer unkorrelierten Reihe)

$$b_8 = u_{1-0.025/8}/\sqrt{N},$$

vgl. Anhang A, 8.6. Aber noch nicht einmal die Schranke b_1 für ein individuelles $|r_e(k)|$, das ist $u_{0.975}/\sqrt{N}$, wird überschritten: Die Reihe $e(t)$ lässt sich also als eine unkorrelierte Reihe auffassen.

Wir gehen der Frage nach, ob die (wahre) Varianz der ARIMA-Residuen $e(t)$ über die Zeit konstant ist –oder ob sich Perioden echt stärkerer und Perioden echt schwächerer Oszillation ablösen. Pro fünf Jahre wird (gleitend) die empirische Varianz $\hat{\sigma}^2(t)$

berechnet. Ihre Wurzel ist in den Abbildungen 17 und 18, in denen auch die Residuen $e(t)$ selber geplottet sind, eingezeichnet. Es ergibt sich für $\hat{\sigma}(t)$ eine oszillierende Linie um den Wert 0.76 (Hohenpeißenberg) bzw. 0.72 (Potsdam) herum, die aber noch keine gesicherte Aussage zulässt.

GARCH-Modellierung. Bezeichnen wir mit $\sigma^2(t) = \text{Var}(e(t))$ die wahre Varianz des Residuums $e(t)$, so soll die Struktur der Zeitreihe $\sigma^2(t)$, $t = 1, 2, \dots, N$, untersucht werden. Mit Hilfe des GARCH-Modells können Zeitreihen mit (möglicherweise) variierenden Varianzen analysiert werden, vgl. Anhang A, 8.8. Dazu wird eine ARMA-ähnliche Gleichung für $\sigma^2(t)$ aufgestellt, Ordnungen (p, q) bestimmt (hier $p = 3, q = 1$), und $p + q + 1$ Koeffizienten α, β geschätzt. Diese Koeffizienten sind für unsere drei Stationen nicht signifikant von der Null verschieden, so dass sich wie eben die Residuen $e(t)$ als eine unkorrelierte Reihe auffassen lassen. Mittels der errechneten Koeffizienten α, β wird eine Prädiktion für $\sigma^2(t)$ durchgeführt (GARCH-Prädiktion genannt, siehe Abbildungen 17 und 18). Diese Prädiktion reproduziert i. W. die Horizontale 0.76 für Hohenpeißenberg (bzw. 0.72 für Potsdam) und demonstriert, dass wir $e(t)$ letztlich als unkorrelierte Reihe mit *konstanter* Varianz $\sigma^2(t) = \sigma^2$ (d. h. als *white noise* Prozess) betrachten können. Die Differenzenreihe $X(t)$ kann demnach durch ein ARMA-Modell, das ja *white noise* Residuen verlangt, gut angepasst werden.

R GARCH-Modell für die ARIMA-Residuenreihe, GARCH-Prädiktion für $\sigma(t)$, einmal per R-Funktion `garch`, einmal per Eigenfunktion `garchpr`. Ferner empirische Schätzung von $\sigma(t)$ durch Bildung gleitender Blöcke und Streuungsberechnung.

```
library(TSA)                                #s. Cran-Software-Packages
postscript(file="C:/KLIMA/GARCHmod.ps",height=6,width=16,horizontal=F)

garchpr<- function(y,n,a,ma,b,mb,sde){
ys<- rep(sde,times=n)                        #GARCH Praediktion
mc<- pmax(ma,mb)                             #per Eigenrechnung
for (t in (mc+1):n){
suma<- a[1]; sumb<- 0
for (m in 1:mc){
suma<- suma + a[m+1]*y[t-m]^2
sumb<- sumb + b[m]*ys[t-m]^2}
ys[t]<- sqrt(suma+sumb)}
invisible(ys)                                #Return ohne Ausgabe
}

#-----Daten-----
Htpres<- read.table("C:/KLIMA/Hot33Res.txt",header=T)
sink("C:/KLIMA/GARCHout.txt")

attach(Htpres)
"Jahrestemp. Hohenp. 1781-2008, Y Residuen vom ARIMA(3,3)-Trend"
N<- length(Jahr)
c("MeanY"=mean(Y), "StdevY"=sqrt(var(Y)), "Anz Jahre"=N)

#-----GARCH-order-----
ma<- 3; mb<- 1 ; mc<- pmax(ma,mb)+1          #ma,mb <= 6
c("ma"=ma, "mb"=mb, "mc"=mc)
```

```

ord<- c(mb,ma) #vertauschte Eingabe
#-----R function garch, Schaetzen, Testen -----
zgarch<- garch(Y,order=ord,maxiter=40)
summary(zgarch) #u.a. Diagnostic tests
a<- rep(0,times=mc); b<- rep(0,times=mc)
for (m in 1:(ma+1)) {a[m]<- zgarch$coef[m]} #a[m] = alpha(m-1)
if (mb > 0) {for (m in 1:mb) {b[m]<- zgarch$coef[m+ma+1]} }
nll<- zgarch$n.likeli
c("alpha0"=a[1],"alpha1"=a[2],"alpha2"=a[3],
  "alpha3"=a[4],"alpha4"=a[5],"negloglikelihood"=nll)
c("beta1"=b[1],"beta2"=b[2],"beta3"=b[3],"beta4"=b[4])
asu<- 0; for(i in 1:ma){asu<- asu + a[i+1]}
if (mb > 0){ for(i in 1:mb){asu<- asu + b[i]}}
sde<- a[1]/(1-asu); c("stdev(Y)"=sde) #plug-in Schaetzung

#-----GARCH-Praediktion-----
"Spre GARCH-Praediktion fuer sigma(t)"
Spre<- zgarch$fitted.values
"GARCH(p,q)-Praediktion letzte 10 Werte"; Spre[(N-9):N]
"Shat emp. Schaetzer fuer sigma(t)"
ka<- 5; kc<- pmax(ka,mc); Shat<- rep(1,times=N)
for(t in ka:N) {Shat[t]<- sqrt(var(Y[(t-ka+1):t]))}
"Laenge der Gltten Bloecke zur Berechnung der Streuung"; ka
"Emp. Schaetzer letzte 10 Werte"; Shat[(N-9):N]
c("MeanSpre"=mean(Spre[kc:N]),"StdevSpre"=sqrt(var(Spre[mc:N])),
  "MeanShat"=mean(Shat[kc:N]))

"Eigenrechnung per Funktion garchpr"
spre<- garchpr(Y,N,a,ma,b,mb,sde)
"spre GARCH(p,q)-Praediktion letzte 10 Werte"; spre[(N-9):N]
c("Meanspre"=mean(spre[mc:N]),"Stdevspre"=sqrt(var(spre[mc:N])))

#-----Plot-----
cylim<-c(-1.5,1.5); tehoe<- -1.5
ytext<- "Y Input-Reihe, sigma-Praed_GARCH, sigma-hat"
plot(Jahr,Y,type="l",lty=1,xlim=c(1780,2008),
      ylim=cylim, xlab="Jahr",ylab=ytext,cex=1.3)
title(main="Mittl. Jahrestemp. Hohenp. 1781-2008, ARIMA-Residuen")
title(sub= "Input-Reihe Y_____,sigma-Praed(GARCH) _____,sigma-hat_ _ _")
lines(Jahr[mc:N],Spre[mc:N],type="l",lty=1)
lines(Jahr[mc:N],Shat[mc:N],type="l",lty=2)
abline(h=c(-1,0,1),lty=3)
text(1996,tehoe,"p,q=",cex=0.8)
text(2003,tehoe,ma,cex=0.8); text(2007,tehoe,mb,cex=0.8)
detach(Htpres)
dev.off()

```

Output GARCH-Kenngrößen für die ARIMA-Residuenreihe.

[1] "Jahrestemp. Hohenp. 1781-2008, Y Residuen vom ARIMA(3,3)-Trend"

Tabelle 10: Differenzen der aufeinanderfolgenden jährlichen Niederschlagshöhen [dm]

Station	N	Mittelw.	Stand.Abw.	$r(1)$	$r(2)$	$r(3)$
Hohenp.	130	-0.003	2.085	-0.459	0.003	0.026
Karlsru.	133	0.014	1.900	-0.429	-0.114	0.025
Potsd.	116	0.007	1.417	-0.461	-0.064	-0.040

```

      MeanY      StdevY      Anz Jahre
2.0961e-03  7.6397e-01  228
ma mb mc
 3  1  4
Model:  GARCH(1,3)
Coefficient(s):
      Estimate  Std. Error  t value  Pr(>|t|)
a0 4.916e-01   1.091e+01   0.045    0.964
a1 2.741e-02   1.131e-01   0.242    0.809
a2 7.141e-02   6.202e-01   0.115    0.908
a3 2.124e-12   1.563e+00  1.36e-12  1.000
b1 5.942e-02   2.085e+01   0.003    0.998

negloglikelihood = 5.2559e+01      stdev(Y) = 0.58401

[1] "Spre GARCH-Praediktion fuer sigma(t)"
[1] "GARCH(p,q)-Praediktion letzte 10 Werte"
0.7609 0.7291 0.7552 0.7975 0.7476 0.7903 0.7736 0.7317 0.7348 0.7515

[1] "Shat emp. Schaetzer fuer sigma(t)"
[1] "Laenge der GltDen Bloecke zur Berechnung der Streuung"      5
[1] "Emp. Schaetzer letzte 10 Werte"
0.8260 1.0022 0.5390 0.5791 0.5792 0.7198 0.6701 0.6494 0.5703 0.4467
  MeanSpre      StdevSpre      MeanShat
0.76383      0.03482      0.72556

[1] "Eigenrechnung per Funktion garchpr"      #Identischer Output!
[1] "spre GARCH(p,q)-Praediktion letzte 10 Werte"
0.7609 0.7291 0.7552 0.7975 0.7476 0.7903 0.7736 0.7317 0.7348 0.7515
  Meanspre      Stdevspre
0.76362      0.03487

```

5.3 Jährliche Niederschlagsmengen

$Y(t)$ bezeichnet jetzt die jährliche Niederschlagshöhe. Von dieser Zeitreihe Y aus geht man wieder durch Differenzenbildung zur Zeitreihe X über, wobei also gilt

$$X(t) = Y(t) - Y(t - 1), \quad t = 2, \dots, N.$$

Tabelle 10 zeigt, dass die jährlichen Änderungen X im Mittel ≈ 0 sind, eine mittlere Abweichung (vom Mittel 0) der Größe $\approx 1.5 \dots 2.0$ [dm] haben und eine Autokorrelation

Tabelle 11: ARIMA-Methode für die jährlichen Niederschlagshöhen [dm]

Station	Ordng	Koeffizienten		Root	ARIMA-Praediktion		
	p, q	α_i	β_1	MSQ	2006–2008		2009
Hohenp.	3,1	0.106,0.085,0.010	–0.935	1.673	11.81,11.84,12.12		11.91
Karlsru.	3,1	0.046,–0.163,–0.062,	–0.943	1.378	7.82, 8.00, 7.59		7.61
Potsd.	3,1	–0.128,–0.102,–0.059	–0.989	0.972	6.10, 6.10, 5.86		5.94

Tabelle 12: Autokorrelationsfunktion $r_e(h)$ bis *time lag* $h = 8$ für die ARIMA-Residuen, jährliche Niederschlagshöhen. H = Hohenpeißenberg, K = Karlsruhe, P = Potsdam.

	$r_e(1)$	$r_e(2)$	$r_e(3)$	$r_e(4)$	$r_e(5)$	$r_e(6)$	$r_e(7)$	$r_e(8)$	b_1	b_8
H	0.083	0.043	0.048	–0.044	–0.018	–0.054	–0.134	–0.005	0.172	0.240
K	–0.036	0.040	0.051	0.023	–0.051	–0.003	–0.063	0.054	0.170	0.237
P	0.047	–0.057	–0.051	0.021	–0.096	–0.020	0.101	–0.099	0.182	0.254

$r(1)$ (der Ordnung 1) von ca. $-0.4 \dots -0.5$ aufweisen: Einem Anstieg der Niederschläge folgt tendentiell sofort wieder ein Absinken.

Auf die Differenzenreihe X passen wir ein ARMA(p,q)-Modell an. Es wurden (möglichst kleine) Ordnungszahlen (p, q) gewählt, so dass eine weitere Erhöhung dieser Zahlen keine wesentliche Verringerung der Maßzahl $RootMSQ$ erwirkt. Man erhält $p = 3, q = 1$ und damit das ARMA(3,1)-Modell

$$X(t) = \alpha_3 X(t-3) + \alpha_2 X(t-2) + \alpha_1 X(t-1) + \beta_1 e(t-1) + e(t). \quad (6)$$

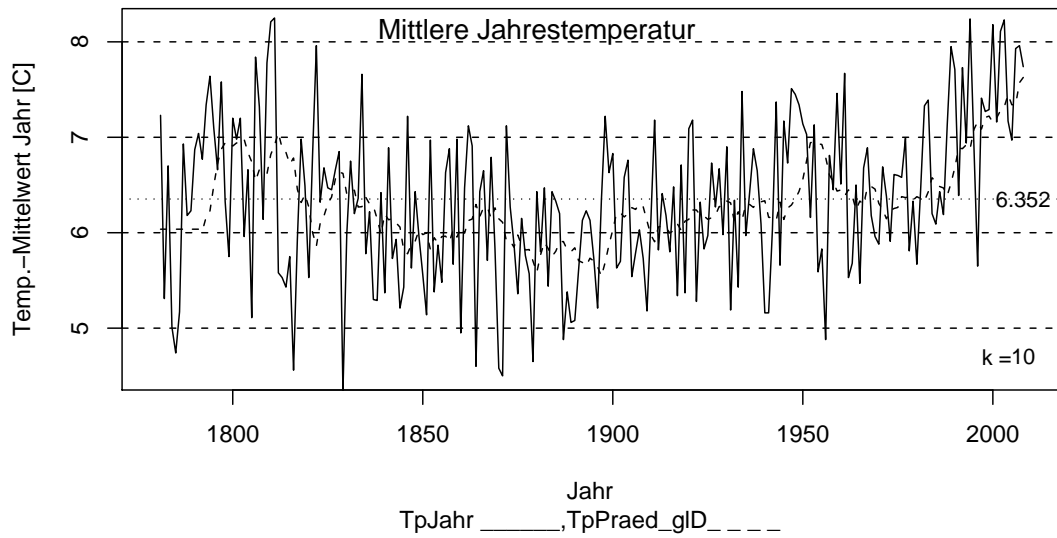
Die Tabelle 11 zeigt die errechneten Koeffizienten α_i und β_1 ; der Koeffizient β_1 ist für alle drei Stationen signifikant von Null verschieden. Ferner gibt sie die Prognosen für die drei Jahre 2006–2008 und für das anstehende Jahr 2009 wieder, jeweils auf der Basis der vorangehenden Jahre. Die Prognosen haben ausgleichenden Charakter und liegen teils oberhalb, teils unterhalb der tatsächlich beobachteten Werte, vgl. Anhang B bzw. die Abbildungen 19 und 20. Diese beschränken sich –einer deutlicheren Darstellung wegen– auf die Wiedergabe der letzten 50 Jahre (zur Berechnung der Koeffizienten α und β wurde aber die Gesamtreihe verwendet).

Bezeichnet $\hat{Y}(t), t = 1, \dots, N$ die ARIMA-Prädiktionen für $Y(t), t = 1, \dots, N$, so bilden wir wieder mit

$$e(t) = Y(t) - \hat{Y}(t), \quad t = 1, \dots, N,$$

die Residuen von der Prädiktion; vgl. die oberen Plots in Abbildungen 19 und 20. Die Autokorrelationsfunktionen $r_e(h), h = 1, \dots, 8$, sind in Tab. 12 wiedergegeben. Die Schranke b_1 für ein individuelles $|r_e(k)|$, das ist $u_{0.975}/\sqrt{N}$, wird nicht überschritten, und so erst recht nicht die simultane Signifikanzschranke b_8 zum Niveau 0.05. Die Residuenreihe $e(t)$ lässt sich also als eine unkorrelierte Reihe auffassen, was das angewandte ARIMA-Modell bestätigt (auf eine GARCH-Modellierung der Residuen wie in 5.2 wird hier verzichtet).

Hohenpeissenberg, Temperatur 1781–2008



Hohenpeissenberg, Temperatur 1781–2008

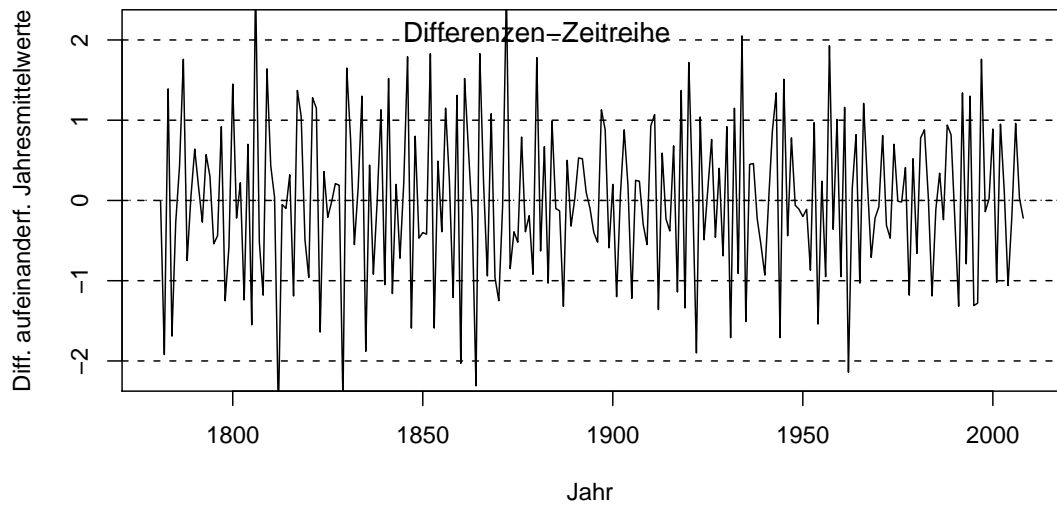


Abbildung 14: Hohenpeissenberg, 1781–2008. Oben: Zeitreihe der mittleren Jahrestemperaturen [$^{\circ}\text{C}$], zusammen mit der Prädiktion gemäß gleitender Durchschnitte (gestrichelt). Unten: Zeitreihe aufeinanderfolgender Differenzen.

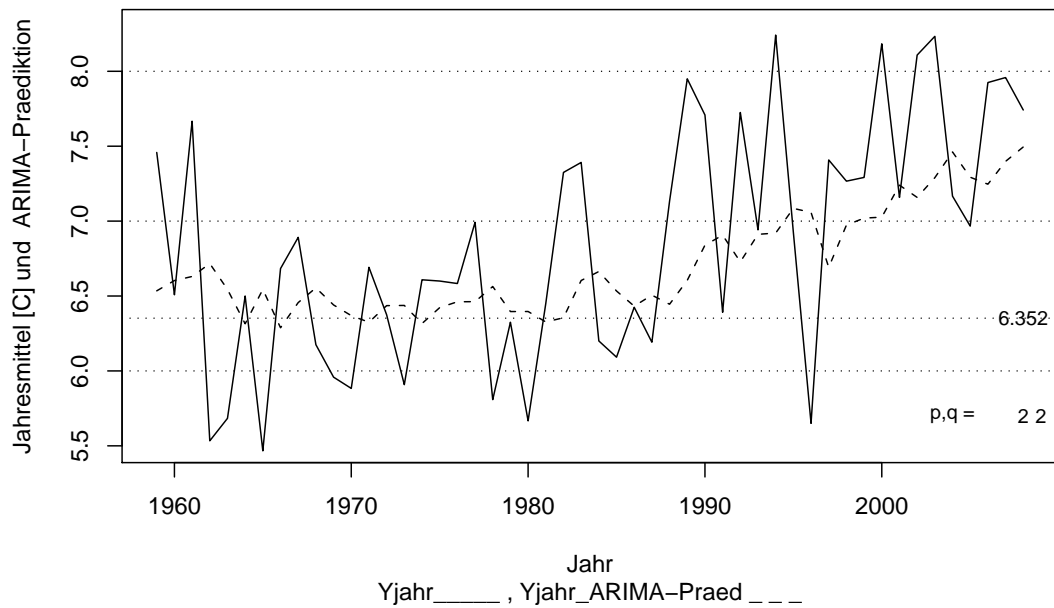
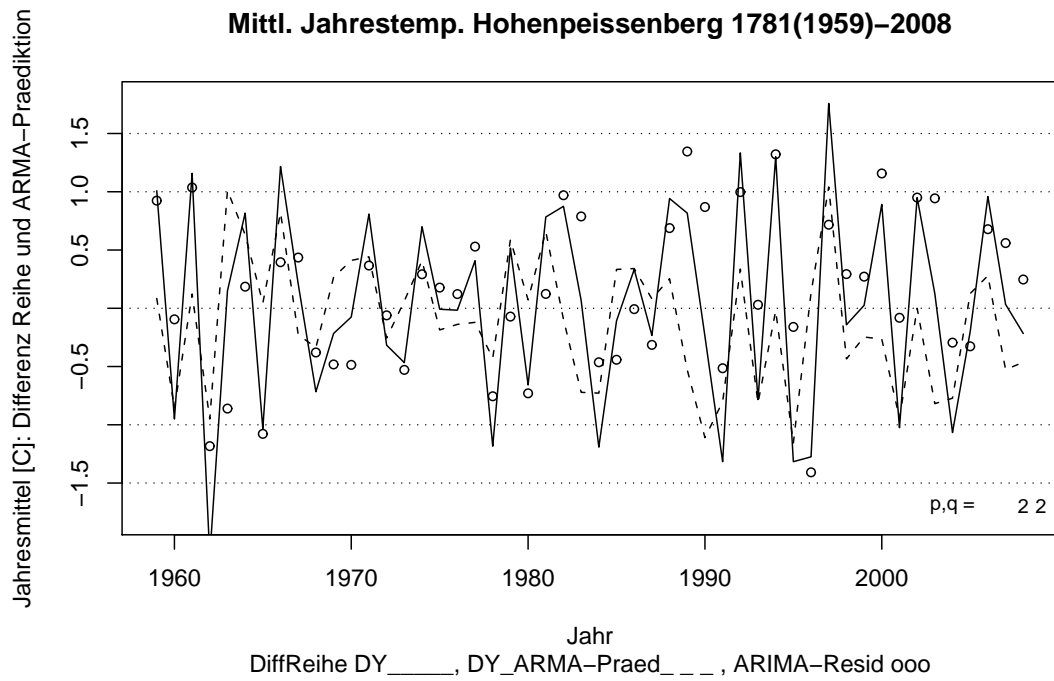
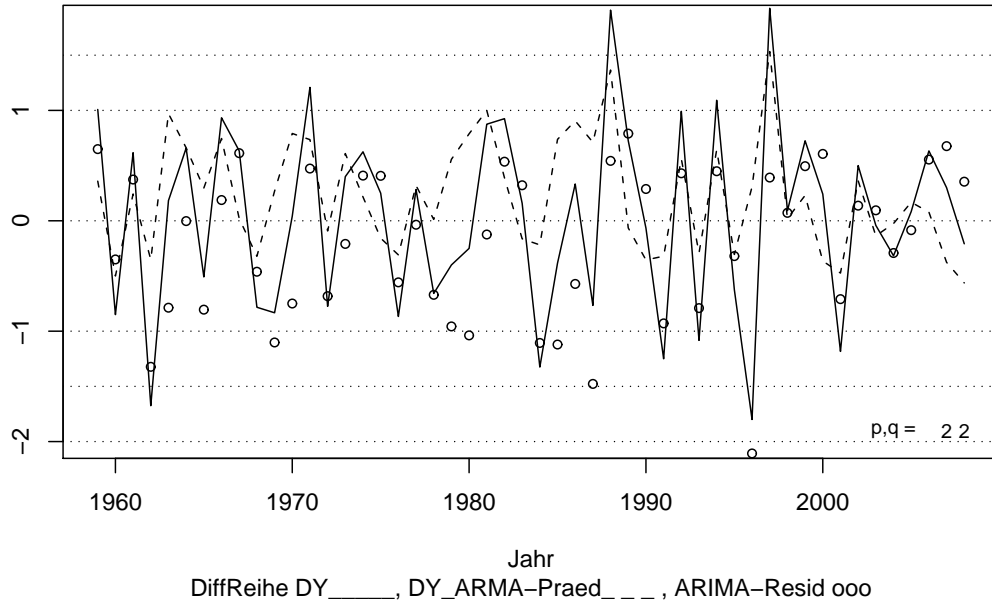


Abbildung 15: Hohenpeißenberg, 1879–2008. Oben: Differenzenzeitreihe, zusammen mit der ARMA-Prädiktion (gestrichelt) und mit den Residuenwerten (als Kreise). Unten: Zeitreihe Mittlere Jahrestemperaturen [$^{\circ}C$], zusammen mit der ARIMA-Prädiktion (gestrichelt). Gezeigt werden nur die letzten 50 Jahre der Zeitreihen.

Jahresmittel [C]: Differenz Reihe und ARMA-Praediktion

Mittl. Jahrestemp. Potsdam 1893(1959)–2008



Jahresmittel [C] und ARIMA-Praediktion

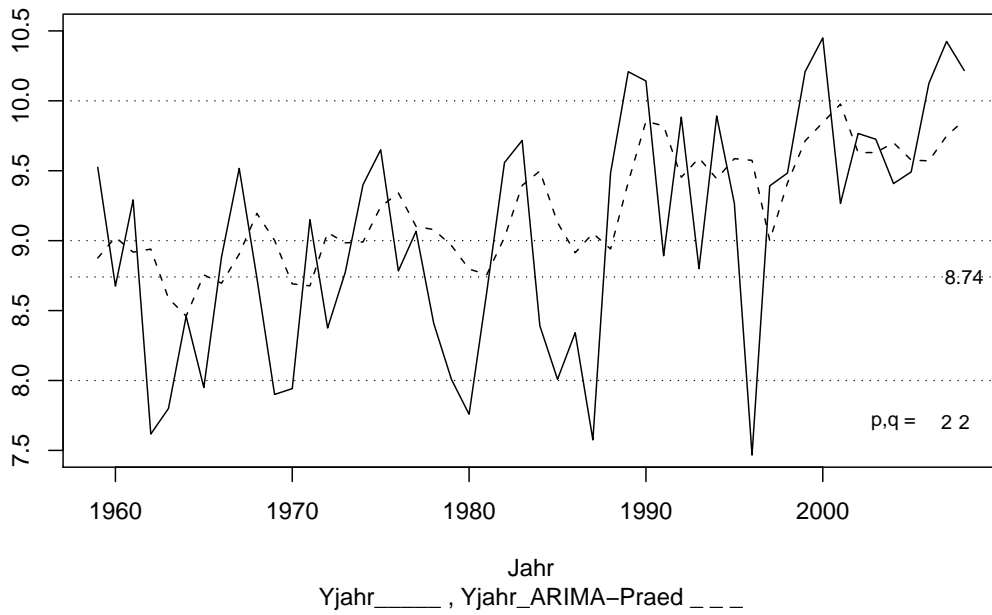


Abbildung 16: Potsdam, 1893–2008. Gleiche Legende wie zu Abb. 15.

Mittl. Jahrestemp. Hohenpeissenb. 1781–2008, ARIMA–Residuen

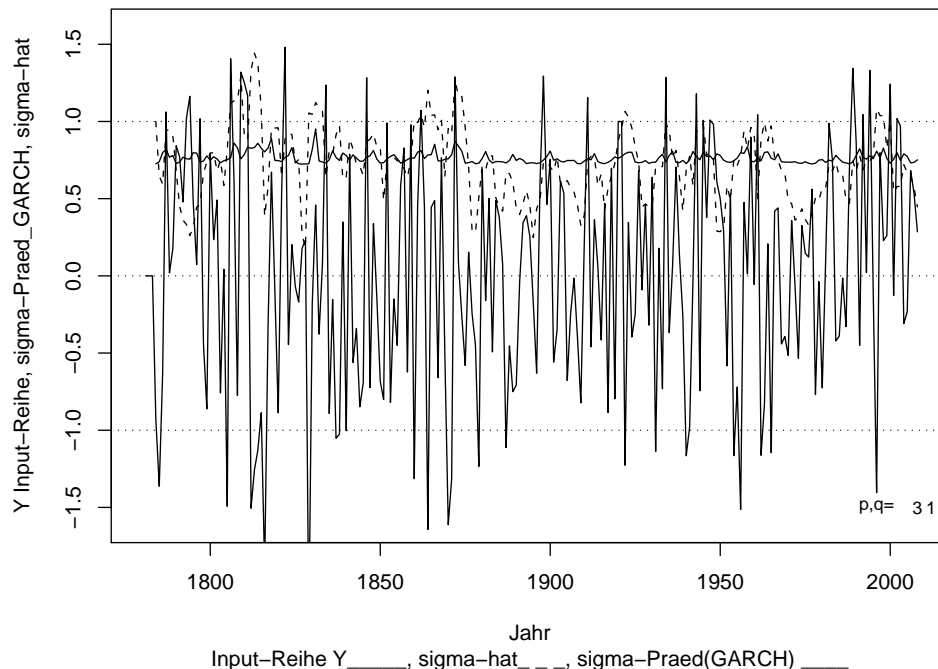


Abbildung 17: Hohenpeissenberg, 1781–2008. ARIMA-Residuen Zeitreihe, Gleitende 5-Jahresstreuung $\hat{\sigma}$ (gestrichelt), GARCH-Prädiktion für σ (durchgezogen).

Mittl. Jahrestemp. Potsdam 1893–2008, ARIMA–Residuen

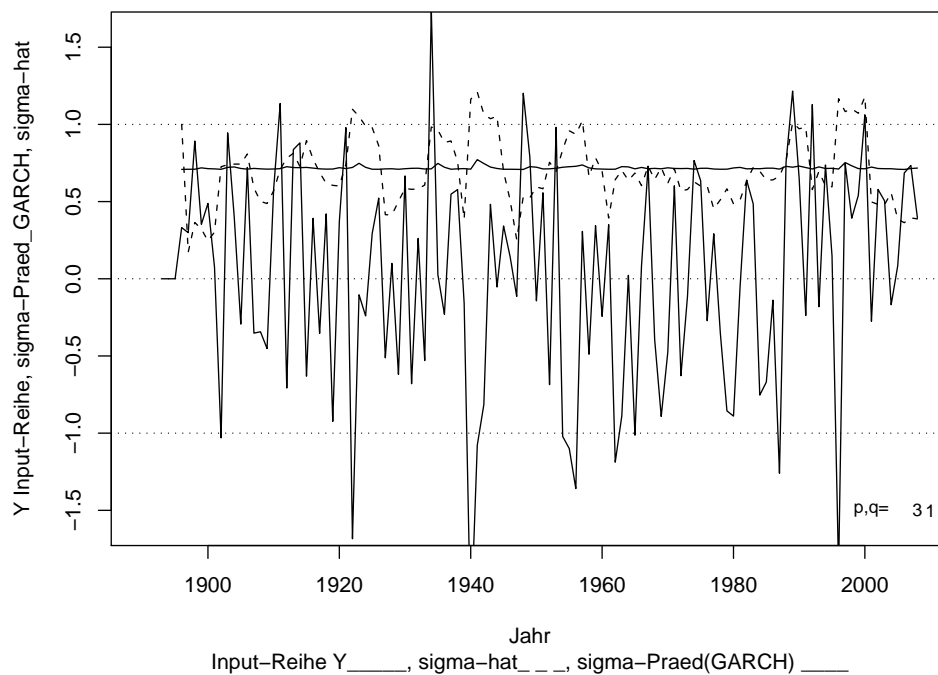
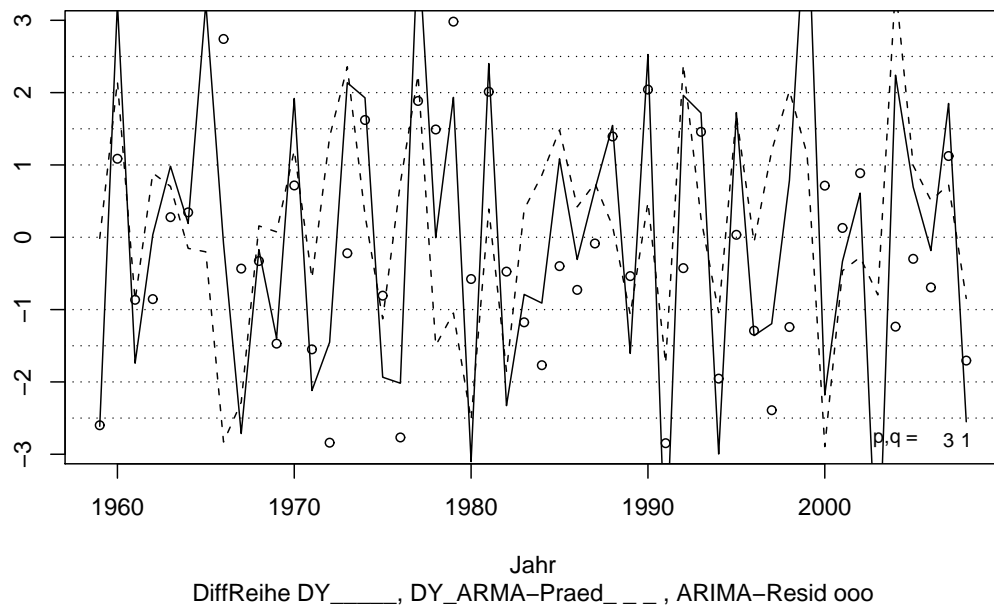


Abbildung 18: Potsdam, 1893–2008. Gleiche Legende wie zu Abb. 17.

Jahressummen [dm]: Differenz Reihe und ARMA-Prädiktion

Jaehrl. Niederschlag Hohenpeissenberg 1879(1959)–2008



Jahressummen [dm] und ARIMA-Prädiktion

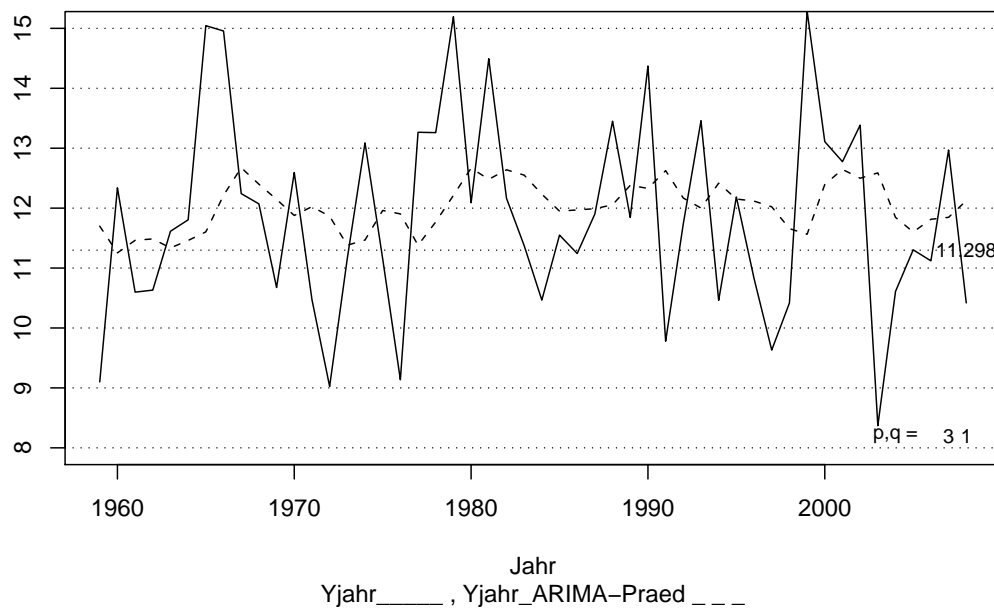
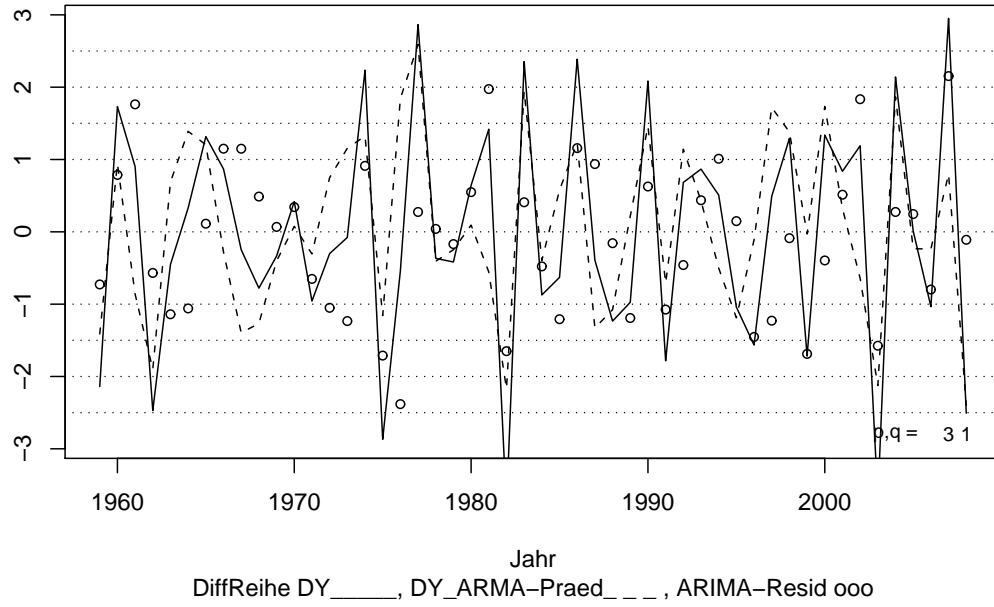


Abbildung 19: Hohenpeissenberg, 1879–2008. Oben: Differenzenzeitreihe, zusammen mit der ARMA-Prädiktion (gestrichelt) und mit den Residuenwerten (als Kreise). Unten: Zeitreihe der Niederschlag-Jahressummen [dm], zusammen mit der ARIMA-Prädiktion (gestrichelt). Gezeigt werden nur die letzten 50 Jahre der Zeitreihen.

Jahressummen [dm]: Differenz Reihe und ARMA-Prädiktion

Jaehrl. Niederschlag Potsdam 1893(1959)–2008



Jahressummen [dm] und ARIMA-Prädiktion

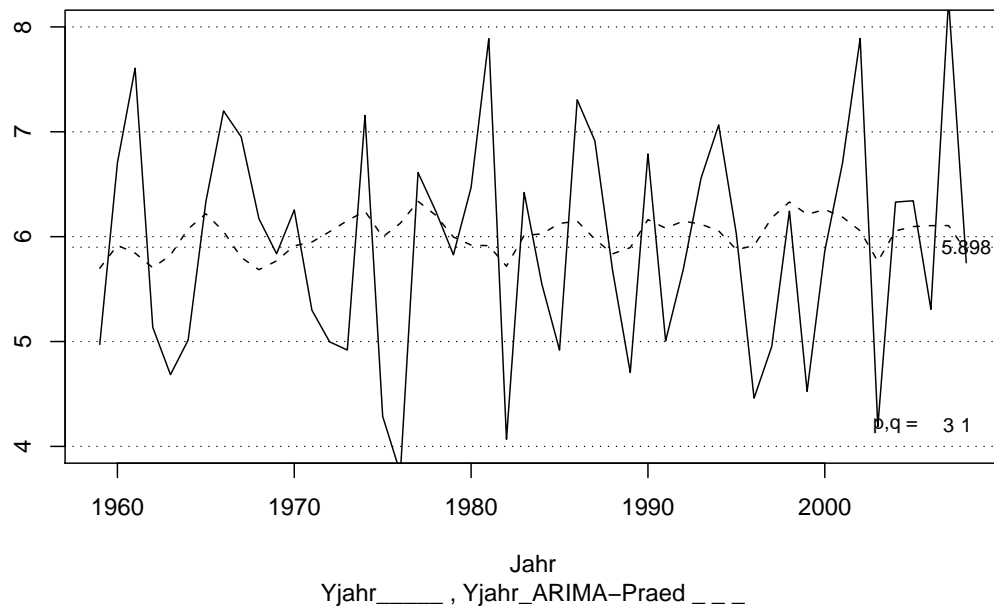


Abbildung 20: Potsdam, 1893–2008. Gleiche Legende wie zu Abb. 19.

6 Modell und Prädiktion: Monatliche Werte

Zur Untersuchung von monatlichen Werten (nur die monatlichen Temperatur-Mittelwerte werden betrachtet) geben wir –ähnlich wie im Abschnitt 3– eine (Jahres)Trend- und eine (Monats)Saison-Komponente an. Anders als dort gewinnen wir die Jahreskomponente als Prädiktion mittels eines ARIMA-Modells und die Monatskomponente mittels eines ARMA-Modells. Deshalb nennen wir das folgende Verfahren eine doppelte ARMA-Methode (kurz: ARIMA+ARMA).

6.1 Doppelte ARMA-Methode

Für die Temperatur-Monatsmittelwerte $Y(t)$ machen wir wie in Abschn. 3 den Ansatz *Trendkomponente + Saisonkomponente*,

$$Y(t) = m(t) + X(t), \quad t = 1, 2, \dots, \quad (7)$$

wobei t die aufeinanderfolgenden Monate zählt, $m(t)$ den langfristigen (Jahres)Trend bezeichnet und die (Monats)Saison-Komponente hier mit $X(t)$ notiert wird. Zur Vereinfachung wollen wir i. F. auch die Schätzungen von $m(t)$ und $X(t)$ wieder $m(t)$ und $X(t)$ bezeichnen. Die Trendschätzung $m(t)$ wird aus 5.2 genommen, als ARIMA-Prädiktion der mittleren Jahrestemperaturen (mit $p = q = 2$ für Hohenpeißenberg und Karlsruhe, und mit $p = 3, q = 1$ für Potsdam); $m(t)$ ist für die Monate t desgleichen Kalenderjahres konstant. Die trendbereinigte Reihe

$$X(t) = Y(t) - m(t), \quad t = 1, 2, \dots,$$

wird in den oberen Plots der Abbildungen 21 und 22 dargestellt. Der Reihe $X(t)$ wird ein ARMA(p, q)-Modell angepasst, mit $p = 3, q = 2$ (das erwies sich als ausreichend). Tab. 13 gibt die geschätzten Koeffizienten α_i und β_j wieder, die (fast) alle signifikant von Null verschieden sind. Mittels ARMA-Prädiktion $\hat{X}(t)$ für $X(t)$, die ebenfalls in den oberen Plots dargestellt ist, wird die trendbehaftete Reihe (genauer: die Prädiktion für die trendbehaftete Reihe) zurückgewonnen, nämlich durch

$$\hat{Y}(t) = m(t) + \hat{X}(t), \quad t = 1, 2, \dots \quad (8)$$

Vgl. dazu die unteren Plots der Abbildungen 21 und 22, in denen $\hat{Y}(t)$ zusammen mit den tatsächlichen monatlichen Werten $Y(t)$ dargestellt wird. Die Güte RootMSQ nach Gleichung (4) ist ebenfalls in Tabelle 13 wiedergegeben. Mit insgesamt nur 4+5 Parametern verläuft diese ARIMA+ARMA-Prädiktion $\hat{Y}(t)$ eng am tatsächlichen Verlauf $Y(t)$, kann aber besonders warmen Sommern und besonders kalten Wintern nicht folgen. Man vergleiche den „Rekordsommer“ 2003 (in Abb. 21 unten um den Monat Nr. 55 herum) und den mit $-2.7, -1.3$ bzw. -2.1 °C (in Hohenp., Karlsru. bzw. Potsd.) relativ kalten Jan. 2009. Vgl. dazu die Prädiktion in der letzten Spalte der Tab. 13.

R `Ytr` ist der Jahrestrend, der gemäß 5.2 berechnet wurde und von File `PoT31Pre.txt` eingelesen wird. Trendbereinigung, ARMA-Modell für die trendbereinigte Reihe, Prädiktion für die trendbereinigte und trendbehaftete Reihe, Residuenanalyse.

`library(TSA)`

`#s. Cran-Software-Packages`

Tabelle 13: ARIMA+ARMA-Methode für die Temperatur-Monatsmittel

Station	Ordng	Koeffizienten		Root	Prädiktion	Jan
	p, q	α_i	β_j	MSQ	Okt-Dez 2008	2009
Hohenp.	3,2	1.758,-1.045,0.026	-1.725,0.992	2.149	8.25, 4.01, 0.63	-0.94
Karlsru.	3,2	1.691,-0.928,-0.041	-1.691,0.953	1.933	11.31, 6.67, 3.38	2.30
Potsd.	3,2	1.810,-1.136,0.078	-1.682,0.951	1.953	9.65, 5.08, 1.80	0.42

```
#-----Daten bereiten, Trendbereinigung-----
attach(potsdTp)

"Mittl. Monatstemp. Potsdam 1893-2008"
mon12<- data.frame(potsdTp[,3:14])/10          #jan-dez auswaehlen
NJahr<- length(Tjahr); N<- NJahr*12
c("Anz Jahre"=NJahr,"Anz Monate"=N)
detach(potsdTp)

#Einlesen des Jahrestrends Ytr, dann verzwoelffachen auf Ytre
potsdPr<- read.table("C:/KLIMA/PoT31Pre.txt",header=T)
attach(potsdPr)
"Trend letzte Dekade"; Ytr[(NJahr-9):NJahr]
Ytre<- 1:N
for(m in (1:N)){j<- trunc((m-1)/12)+1; Ytre[m]<- Ytr[j]}
YtreN1<- 9.7029                                #aus 5.2

Ymat<- as.matrix(mon12)                        #als NJahr x 12 Matrix anlegen
Yobs<- 1:N                                     #Yobs Vektor der dim N
for(j in (1:NJahr)){Yobs[((j-1)*12+1):(j*12)]<- Ymat[j,(1:12)]}

Yde<- Yobs - Ytre                              #detrended
"trendbehaftete und trendbereinigte Werte, letzte 12 Monate"
Yobs[(N-11):N]; Yde[(N-11):N]
ma<- 3; mb<- 2 ; mc<- pmax(ma,mb)              #ARMA order

#-----ARMA-Modell fuer trend-bereinigte Monatswerte"
#-----a) Koeffizienten schaezen, Praediktionsanalyse-----
Ydarma<- arma(Yde,order=c(ma,mb))
summary(Ydarma); Ydcoef<- Ydarma$coef; Yds2<- Ydarma$css
Ydfit<- Ydarma$fitted.values; Ydres<- Ydarma$residuals

a<- rep(0,times=12); b<- rep(0,times=12)
for (m in 1:ma){a[m]<- Ydcoef[m]}
if (mb > 0) for (m in 1:mb){b[m]<- Ydcoef[ma+m]}
asu<- 0; for (j in 1:mc){asu<- asu+a[j]}
theta1<- (1-asu)*mean(Yde); theta2<- Ydcoef[ma+mb+1]
c("ArOrder"=ma,"MAOrder"=mb,"SSQ"=Yds2)
c("alpha1"=a[1],"alpha2"=a[2],"alpha3"=a[3],
  "alpha4"=a[4],"alpha5"=a[5],"alpha6"=a[6])
c("theta1"=theta1,"theta2"=theta2)
```

```

c("beta1"=b[1], "beta2"=b[2], "beta3"=b[3],
  "beta4"=b[4], "beta5"=b[5], "beta6"=b[6])

"ARMA-Praediktion, trendbereinigt, Letzte 12 Monate"; Ydfit[(N-11):N]
YdarmaNJ<- theta2
for (m in 1:mc){YdarmaNJ<- YdarmaNJ+a[m]*Yde[N+1-m]+b[m]*Ydres[N+1-m]}
c("Forecast Jan_NewYear trendbereinigt"=YdarmaNJ)
Yarma<- Ytre; Yarma[mc:N]<- Ytre[mc:N]+Ydfit[mc:N] #Trendbehaftete Reihe
Yres<- Yobs - Yarma
"ARMA-Praediktion und Residuals, trendbehaft, Letzte 12 Monate"
Yarma[(N-11):N]; Yres[(N-11):N]
"Forecast Jan_NewYear"
c(YtreN1, YdarmaNJ, YdarmaNJ+YtreN1)

#-----b) Residuenanalyse, Mean Stdev Outputs-----
mc1<- mc+1; Yrer<- Yres[(mc1):N]; rsq<- mean(Yrer*Yrer)
c("MeanYpred"=mean(Yarma[(mc1):N]), "StdYpred"=sqrt(var(Yarma[(mc1):N])),
  "MeanYres"=mean(Yrer), "StdYres"=sqrt(var(Yrer)))
c("RSQ"=rsq, "RootRSQ"=sqrt(rsq))
racf<- acf(Yrer, lag.max=12, type="corr", plot=F) #kein Plot
"Autokorrelation der Residuen"; racf$acf

```

```
detach(potsdPr)
```

Output ARMA-Kenngrößen für trendbereinigte Reihe, Praediktion für trend-bereinigte und trend-behaftete Reihe, Residuen mit Berechnung von RootMSQ.

```

[1] "Mittl. Monatstemp. Potsdam 1893-2008"
  Anz Jahre Anz Monate
      116      1392
[1] "Trend letzte Dekade"
  9.6683 9.3879 9.5418 9.1881 9.2206 9.5768 9.4079 9.4394 9.6909 9.8238
[1] "trendbehaftete und trendbereinigte Werte, letzte 12 Monate"
[1]  3.8  4.6  4.5  8.6 15.8 18.6 19.1 18.2 13.2  9.3  5.4  1.5
[1] -6.023 -5.223 -5.323 -1.223  5.976  8.776  9.276  8.376  3.376
     -0.523 -4.423 -8.323

```

```

Model: ARMA(3,2)
Coefficient(s):
      Estimate Std. Error t value Pr(>|t|)
ar1      1.8105   0.0234   77.28 < 2e-16 ***
ar2     -1.1359   0.0405  -27.99 < 2e-16 ***
ar3      0.0784   0.0234    3.35 0.0008 ***
ma1     -1.6821   0.0113 -148.54 < 2e-16 ***
ma2      0.9513   0.0113   83.50 < 2e-16 ***
intercept 0.0004   0.0140    0.03 0.9765

```

```

Fit: sigma^2 estim. as 3.82, Cond. Sum-of-Squares = 5299.5, AIC = 5827
ArOrder  MAOrder  SSQ
      3.0      2.0  5299.5

```

```

alpha1    alpha2    alpha3    alpha4    alpha5    alpha6
1.81052  -1.13599  0.07849  0.00000  0.00000  0.00000
      theta1          theta2
      0.005279      0.000414
beta1     beta2     beta3     beta4     beta5     beta6
-1.68210  0.95136  0.00000  0.00000  0.00000  0.00000
[1] "ARMA-Praediktion, trendbereinigt, Letzte 12 Monate"
[1] -9.3895 -7.7728 -4.3038 -0.0363  4.4494  8.0950  9.3116  8.0021
      4.6540 -0.1687 -4.7442 -8.0257
Forecast Jan_NewYear trendbereinigt
      -9.2795
[1] "ARMA-Praediktion und Residuals, trendbehaft, Letzte 12 Monate"
      0.4343  2.0510  5.5199  9.7874 14.2733 17.9188 19.1354 17.8260 14.4779
      9.6551  5.0796  1.7981
      3.3657  2.5490 -1.0199 -1.1874 1.5267 0.6811 -0.0354 0.3740 -1.2778
      -0.3550  0.3204 -0.2980
[1] "Forecast Jan_NewYear"
      9.7029 - 9.2795 = 0.4234
MeanYpred  StdYpred  MeanYres  StdYres
      8.7353   6.6947   0.0253   1.9538
      RSQ    RootRSQ
      3.8153   1.9533

```

6.2 Zum Vergleich: Gleitende Durchschnitte und Lag-12 Differenzen

Auf der Grundlage von Modell (7) kann alternativ die Prognosemethode der gleitenden Durchschnitte gewählt werden, angewandt sowohl für den Jahrestrend $m(t)$ als auch für die trendbereinigte Reihe $X(t)$. Als Trendschätzung $m(t)$ wird der Durchschnittswert der *vorangegangenen* Werte $Y(t-1), Y(t-2), \dots, Y(t-k*12)$ genommen. Die Zahl k gibt also die Anzahl der verwendeten Jahre an. Als Prädiktion $\hat{X}(t)$ für die (Monats)Saison-Komponente $X(t)$ wird der Durchschnittswert der *vorangegangenen* Werte

$$Y(t-1) - m(t-1), Y(t-2) - m(t-2), \dots, Y(t-k) - m(t-k)$$

genommen. Die Zahl k gibt hier die Anzahl der verwendeten Monate an. Wieder nach Gleichungen (8) und (4) berechnen wir die Güte dieser Prädiktionemethode. Tabelle 14 zeigt, dass erst ab einer Tiefe von $k \geq 20$ die *RootMSQ*-Werte der ARIMA+ARMA-Methode erreicht werden (die mit $4 + 5 = 9$ Kenngrößen auskommt –hier allerdings jeweils aus der Gesamtdatei ermittelt).

Eine weitere –zu 6.1 alternative– Methode ähnelt der ARIMA-Methode aus 5.2; doch anstatt Differenzen $Y(t-1) - Y(t)$ aufeinanderfolgender Werte (lag-1 Differenzen) bilden wir mit

$$X(t) = Y(t) - Y(t-12), \quad t = 13, 14, \dots,$$

Differenzen der um 12 Zeiteinheiten (also um 12 Monate) auseinanderliegender Werte (lag-12 Differenzen). Dem Differenzenprozess $X(t)$ wird ein AR(12)-Modell angepasst (AR(12) = ARMA(12,0)), und es wird nach Gleichung (3) bzw. gleichwertig (4) die Güte der Anpassung berechnet. Wir verwenden das Kürzel ARIMA(lag12). Tabelle

Tabelle 14: Tiefe k des gleitenden Durchschnitts und resultierender Gütewert $RootMSQ$. Ferner die Gütewerte der Methoden ARIMA+ARMA und ARIMA(lag12)

Tiefe k	$RootMSQ$		
	Hohenp.	Karlsru.	Potsdam
5	2.366	2.137	2.229
10	2.248	2.012	2.079
12	2.227	1.991	2.077
15	2.193	1.971	2.065
20	2.171	1.939	2.058
ARIMA+ARMA	2.149	1.933	1.953
ARIMA(lag12)	2.544	2.301	2.318

Tabelle 15: Autokorrelationsfunktion $r_e(h)$ bis $time\ lag\ h = 8$ für die ARIMA+ARMA-Residuen, mittlere Monatstemperatur

Station	$r_e(1)$	$r_e(2)$	$r_e(3)$	$r_e(4)$	$r_e(5)$	$r_e(6)$	$r_e(7)$	$r_e(8)$	b_1	b_8
Hohen.	0.102	0.029	0.013	0.023	0.014	0.037	0.035	0.029	0.130	0.181
Karlsru.	0.204	0.051	0.003	0.017	0.050	0.075	0.051	0.022	0.135	0.189
Potsd.	0.152	0.063	0.012	0.006	0.003	0.001	0.039	0.041	0.182	0.254

14 zeigt, dass diese Methode der Methode ARIMA+ARMA und auch den gleitenden Durchschnitten unterlegen ist.

6.3 Residuenanalyse

Bezeichnet $\hat{Y}(t), t = 1, \dots, N$ die ARIMA+ARMA-Prädiktionen für $Y(t), t = 1, \dots, N$, so bilden

$$e(t) = Y(t) - \hat{Y}(t), \quad t = 1, \dots, N,$$

die Residuen von der Prädiktion; vgl. die oberen Plots in Abbildungen 21 und 22. Welche Struktur hat diese Residuen-Zeitreihe $e(t), t = 1, \dots, N$? Die Autokorrelationsfunktion $r_e(h), h = 2, \dots, 8$, besitzt nur Werte nahe 0, vgl. Tab. 15. Allein die Autokorrelation erster Ordnung (d. h. die Korrelation zwischen zwei unmittelbar aufeinanderfolgenden Monaten) weist relativ große Koeffizienten auf. Die Signifikanzschranke b_1 (ja sogar die simultane Schranke b_8) zum Niveau 0.05 wird vom Wert $r_e(1)$ im Fall Karlsruhe überschritten. In diesem Fall lässt also die Prädiktions-Methode ARIMA+ARMA (mit den gewählten Ordnungszahlen p,q) für die Residuen eine zu große Autokorrelation erster Ordnung zurück.

Zur Prädiktion der monatlichen Temperaturwerte sollten auch noch weitere Methoden zur Trend- und Saisonbestimmung kombiniert und getestet werden.

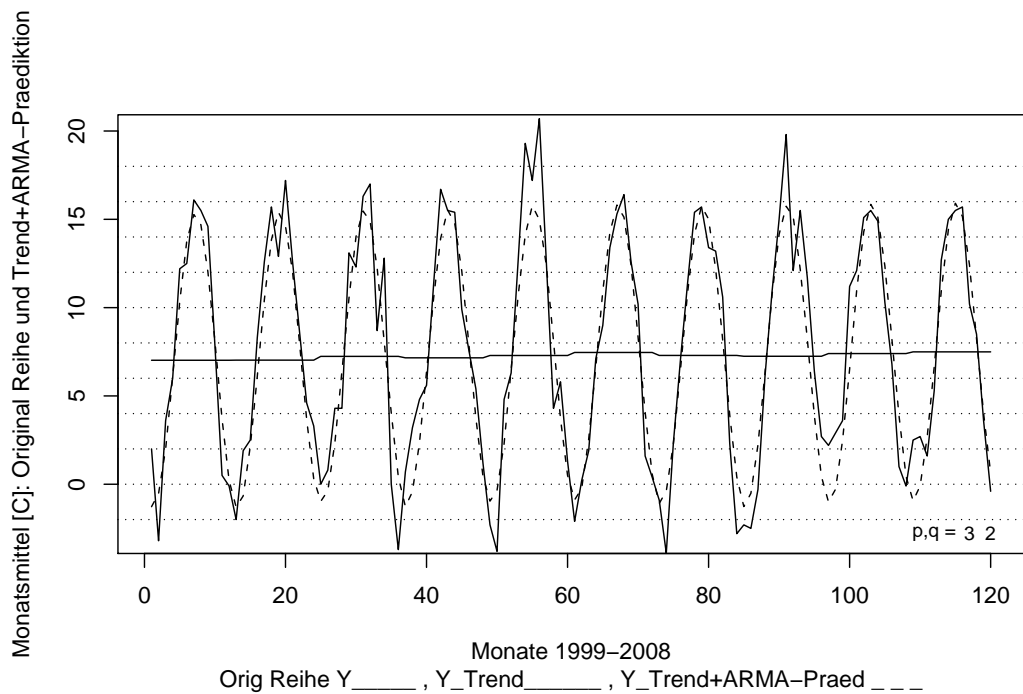
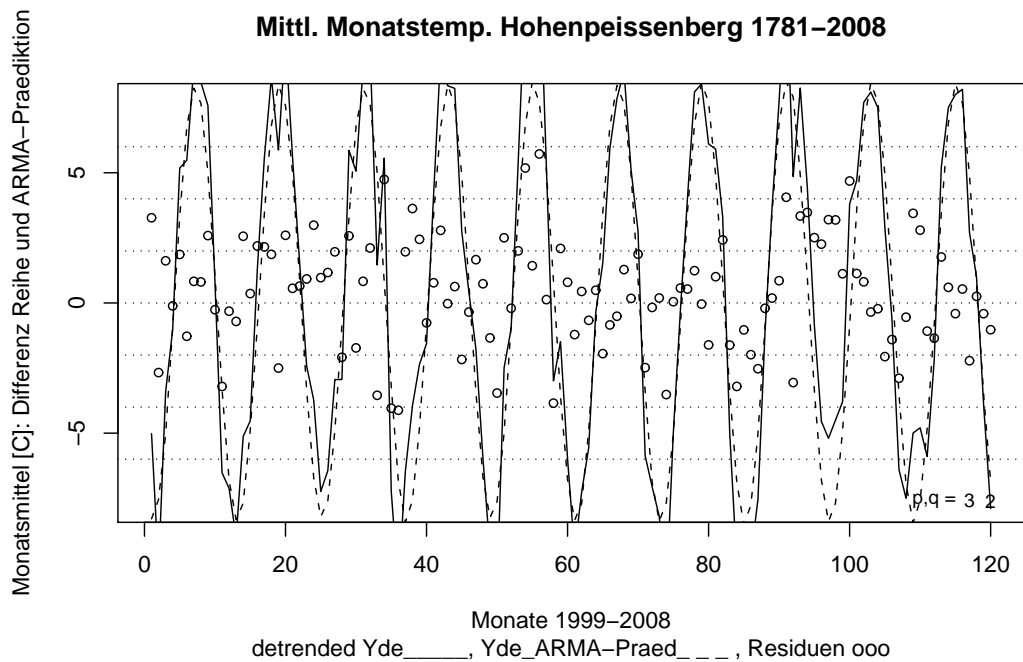


Abbildung 21: Hohenpeissenberg, Temperatur Monatsmittel 1781–2008. Oben: Trendbereinigte Zeitreihe, zusammen mit der ARMA-Prädiktion (gestrichelt) und mit den Residuenwerten (als Kreise). Unten: Monatsmittelwerte [°C], zusammen mit der Trend+ARMA-Prädiktion (gestrichelt). Gezeigt werden die letzten 10 Jahre der Zeitreihen.

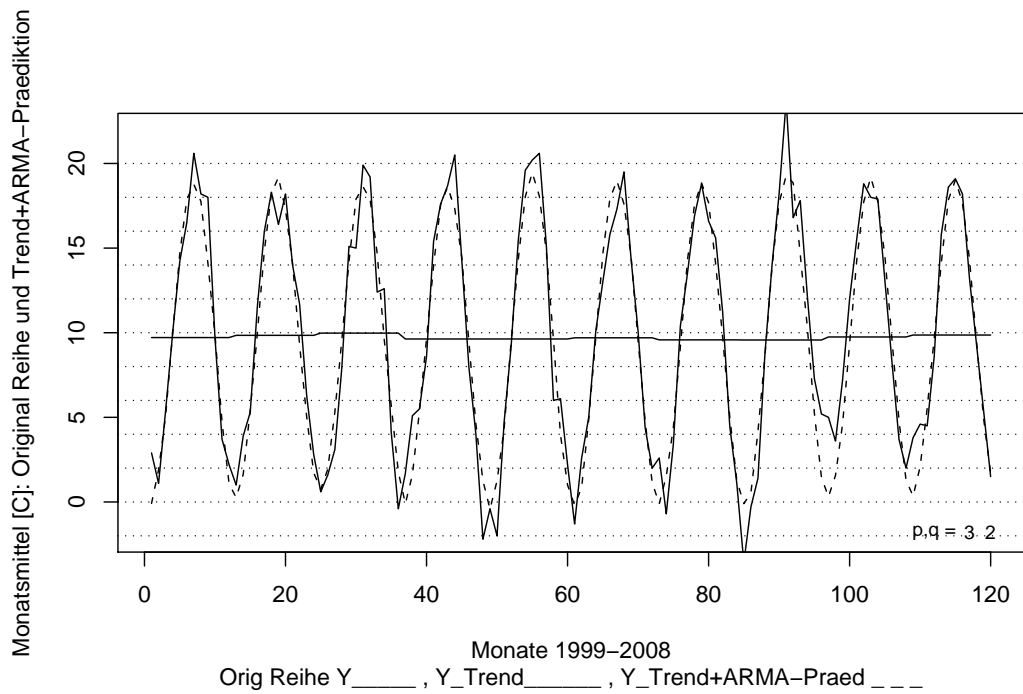
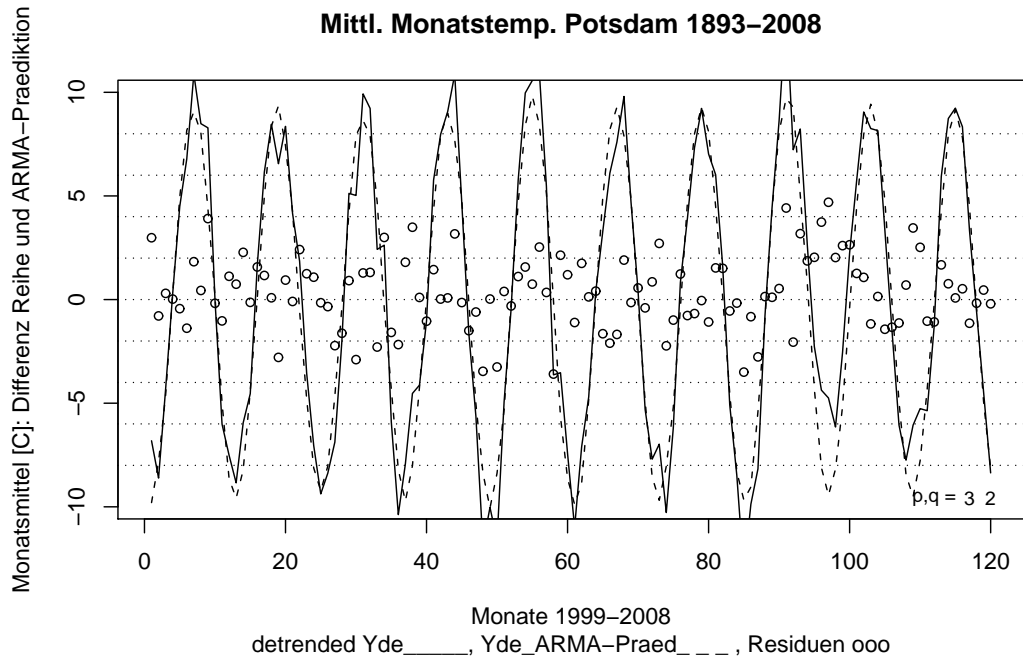


Abbildung 22: Potsdam, Temperatur Monatsmittel 1893–2008. Gleiche Legende wie zu Abb. 21

7 Periodogramm- und Spektraldichte-Analyse

Die nun folgende *Frequenzanalyse* der Zeitreihe geht von der Vorstellung aus, dass die „Oszillation“ der beobachteten Reihe durch Überlagerung von harmonischen (sin, cos) Schwingungen entsteht. Details zu diesem Abschnitt finden sich in Anhang A, 8.7.

7.1 Periodogramm, auch geglättet

Zur Beschreibung einer Schwingung verwenden wir –untereinander gleichbedeutend–

- die Schwingungszahl k , $k = 1, \dots, n/2$; d. i. die Anzahl der Schwingungen im Zeitbereich $[1, n]$,
- die Periodenlänge $T = T_k = n/k$,
- die Kreisfrequenz $\omega = \omega_k = 2\pi/n \cdot k$.

Der Periodogrammwert $I(\omega_k)$, der mittels der Fouriekoeffizienten der Zeitreihe $Y(t)$, $t = 1, \dots, n$, berechnet wird, gibt an, wie stark eine „Schwingung“ mit Schwingungszahl k an der Oszillation der Zeitreihe beteiligt ist. Der Verlauf des Periodogramms ist i. A. sehr „gezackt“. Durch Glättung des Periodogramms entsteht die *Spektraldichte*-Schätzung $\hat{f}(\omega_k)$ der Zeitreihe.

Eine einfache Möglichkeit der Glättung besteht mit Hilfe des sog. diskreten *Daniel-fensters*, und zwar durch die Bildung gleitender Durchschnitte über eine Anzahl $2M+1$ von Periodogrammwerten $I(\omega_l)$: jeweils M Werte links und M Werte rechts von ω_k . Diese Glättung führt zu einer Version der Spektraldichte-Schätzung $\hat{f}(\omega_k)$ (und nur diese Version wird hier betrachtet).

Es sei noch erwähnt, dass wir die Ordinaten der folgenden Abbildungen in dem Sinne *normiert* haben, dass $I(\omega)$ –und damit auch $\hat{f}(\omega)$ – jeweils durch s^2 , der empirischen Varianz der Zeitreihe, geteilt wurde.

Die in den folgenden Abbildungen eingetragenen Schranken b_l bzw. B_l , das sind

$$b_l = -\frac{1}{\pi} \cdot \ln\left(\frac{\alpha}{l}\right), \quad B_l = \frac{1}{\pi\nu} \cdot \chi_{\nu, 1-\alpha/l}^2, \quad \nu = 4M, \quad (9)$$

beziehen sich auf das normierte Periodogramm bzw die normierte Spektraldichte einer reinen Zufallsreihe (mit gleicher Varianz wie die beobachtete Reihe $Y(t)$). Dabei bedeutet $\chi_{\nu, \gamma}^2$ in (9) das γ -Quantil der χ^2 -Verteilung mit ν Freiheitsgraden. Beim Periodogramm z. B. ist die Bedeutung dieser Schranken die folgende: Die Wahrscheinlichkeit, dass das Maximum von l Werten (an l vorher fixierten Stellen ω_k bzw. T_k) die Schranke b_l überschreitet, beträgt im Fall einer Zufallsreihe $\alpha = 0.05$. Die *Bonferroni*-Korrektur in (9), das ist α/l anstatt α , bezieht sich dabei auf die Tatsache, dass man sich nicht für den Wert des Periodogramms bzw. der Spektraldichte an einer einzigen (vorher festgelegten) Stelle ω bzw. T interessiert, sondern an mehreren (nämlich l) Stellen. Die $l = 1$ -Schranke jedenfalls ist zu niedrig: Man beachte, dass im Mittel 5 % der Periodogramm-Werte einer Zufallsreihe über der Schranke b_1 liegen. Wir werden (etwas willkürlich) bei Werten über der $l = 4$ -Schranke von schwacher Signifikanz sprechen, bei Werten über der $l = 12$ -Schranke von Signifikanz.

Berücksichtigt man noch –unter der Voraussetzung eines AR(1)-Prozesses– die Autokorrelation $r = r(1)$ der Zeitreihe, so ist das Periodogramm $I(\omega)$ bzw. die

Spektraldichte-Schätzung $\hat{f}(\omega)$ noch durch die Größe

$$\lambda(\omega) = \frac{1 - r^2}{1 - 2r \cos \omega + r^2} \quad (10)$$

zu teilen, damit die Schranken b_l bzw. B_l weiterhin gültig bleiben. In der Tat stellt $\frac{s^2}{\pi} \cdot \lambda(\omega)$ eine Schätzung der Spektraldichte $f(\omega)$ eines AR(1)-Prozesses dar. Diese Korrektur wurde hier nicht durchgeführt; die $r = r(1)$ -Werte sind nach Ausweis von Tabellen 2 und 3 für die bereinigten Reihen (die wir benutzen werden, siehe gleich) sehr klein, die Korrekturgröße $\lambda(\omega)$ in (10) also ≈ 1 .

Stationarität. Die hier vorgestellte Frequenzanalyse verlangt die *Stationarität* der Zeitreihe. Wir gehen auf diese Forderung insofern ein, dass wir nicht die beobachteten Werte selber, sondern die *Residuen* vom polynomialen Trend in die Analyse eingehen lassen (Polynome vom Grade 4 wurden verwendet). Ohne diese *Trendbereinigung* gehen langfristige Schwankungen ($T \geq 10$ oder 20 Jahre etwa) ganz dominant in die Frequenzanalyse ein, mit Trendbereinigung machen diese sich nur noch abgeschwächt bemerkbar.

7.2 Temperatur: Jahres- und Winterwerte

Die Periodogramme der Abbildungen 23, 24, 25 weisen Gipfel (*peaks*) an diversen Periodenlängen T auf.

Für die Jahresreihe Hohenpeißenberg werden (schwach signifikant) hohe Werte bei langjährigen Perioden (insbes. über 20 Jahre) ausgewiesen. Ob dies echte Zyklen oder Reste eines nicht bereinigten Trends sind, bleibt weiteren Analysen vorbehalten. Vergleiche Schönwiese *et al*, 1993, Abb. 1 für eine Spektraldichte-Schätzung der unbereinigten Reihe, mit dominant hohen Werten für $T \gg 20$ Jahre.

Im Folgenden wollen wir nur noch die *Winterreihen* von Hohenpeißenberg und Potsdam betrachten.

Während die Winterreihe Hohenpeißenberg keine auffälligen Perioden aufweist (Abb. 24), zeigt das Periodogramm der Winterreihe Potsdam (Abb. 25) bei $T = 7.8$ Jahre einen signifikanten Peak, der auch noch in der geglätteten Version (unterer Plot) zum Vorschein kommt.

In den Periodogrammen der Temperaturreihen gibt es einen „Nebengipfel“ bei $T = 2.2$ Jahre. Eine mögliche meteorologische Begründung ist die quasi-periodische 2.2-jährige (d. i. 26-monatige) Schwingung der Windrichtung (engl. QBO = quasi-biannual oscillation) zwischen Ost- und Westwind in der tropischen Stratosphäre (vgl. Malberg, 1997).

7.3 Niederschlag: Winterwerte und Summarium

Das Periodogramm der Abbildung 26 für Hohenpeißenberg zeigt einen scharfen, signifikanten Gipfel bei $T = 4.0$ Jahre. Die geglättete Version bestätigt den Periodenbereich um 4 Jahre zumindest als bedeutsam. Möglicherweise ist diese Periode –als ungefähre Verdopplung der eben genannten Zahl 2.2– auch ein Ausfluss des QBO-Phänomens. Die Abb. 27 der Winter-Niederschlagsreihe Potsdam weist einen solchen Gipfel nicht

auf. Der große Wert am rechten Rand des Periodogramms (entspricht $T = 2$) könnte aber vielleicht auch auf diesem Phänomen beruhen.

Insgesamt sind nur die Perioden $T = 7.8$ Jahre in der Winter-Temperaturreihe Potsdam und $T = 4.0$ Jahre in der Winter-Niederschlagsreihe Hohenpeißenberg statistisch wirklich gesichert. Die (hier nicht gezeigten) Winterreihen Karlsruhe weisen eine solche signifikante Periode nicht auf. Allein das Periodogramm der Karlsruher Winter-Temperaturreihe hat bei $T = 8.4$ einen schwach signifikanten Gipfel.

Summa summarum lassen sich keine verlässlichen allgemeineren Aussagen über Periodizitäten der Temperaturen oder Niederschläge in Jahres- oder Winterreihen machen.

R Berechnung des Periodogramms der Zeitreihe $Y[1], \dots, Y[n]$ durch Bestimmung der Fourierkoeffizienten $a[k]$ und $b[k]$, $k = 1, \dots, n/2$. Plot des (standardisierten) Periodogramms mit simultanen Schranken.

```
attach(hohenNs)
postscript(file="C:/KLIMA/HNpgram.ps",height=6,width=15,horizontal=F)

plotP<-function(k,z,nh,tylab,ybo,yli,bc,bct,lte,xte){
plot(k,z,type="l",lty=1,xlim=c(0,nh),ylim=c(ybo,yli),
      ylab=tylab,xlab="")
segments(1,bc,nh-4,bc,lty=2)           #simultane Schranken eintragen
text(nh,bc,bct,cex=0.7)
mtext(lte,side=1,line=2,at=xte[1:7])   #Text am unteren Rand
mtext("k",side=1,line=1,at=xte[8])
mtext("T",side=1,line=2,at=xte[8])
}

#-----Auszuwertenden Vektor Y bereitstellen-----
Y1<- (dzvj+jan+feb)/1000                #Niederschlagssummen Winter
c("N1"=length(Y1), "Mean1"= mean(Y1), "SD1"= sqrt(var(Y1)))
Ja<- Jahr-1900; Ja2<- Ja*Ja; Ja3<- Ja2*Ja; Ja4<- Ja2*Ja2
Y<- Y1 - predict(lm(Y1~Ja+Ja2+Ja3+Ja4))  #Polynomtrend-Bereinigung
n<- length(Y); nh<- round(n/2); SD<- sqrt(var(Y))
c("N"=n, "Nhalbe"=nh,"Mean"= mean(Y), "SD"= SD)

#Periodogramm berechnen mittels Fourierkoeffizienten a,b
seq<- 1:nh; per<- n/seq; nu<- seq/n; pg<- 1:nh
for (k in 1:nh)
{a<- 0; b<- 0; omk<- 2*pi*k/n
for (i in 1:n)
{a<- a+ Y[i]*cos(omk*i)
b<- b+ Y[i]*sin(omk*i)
pg[k]<- (a*a+b*b)/(n*pi)}}

#-----Plotten des Periodogramms-----
Umr<- SD^2; Pgr<- pg/Umr                #Standardisieren
tylab<- "Periodogramm R(k)^2*n/(4 pi s^2)"
```

```
#Simultane Schranken  $b_l = -\ln(\alpha/l)/\pi$  berechnen,  $l=1,4,12$ 
bc<- c(-log(0.05/1)/pi,-log(0.05/4)/pi,-log(0.05/12)/pi)
bct<- c("b_1","b_4","b_12")
lte<- c("26","13","6.5","4.3","3.3","2.6","2.2")
xte<- c(5,10,20,30,40,50,60,63)
ybo<- 0.0;    yli<- 2.0

plotP(seq,Pgr,nh,tylab,ybo,yli,bc,bct,lte,xte)
title(main="Hohenpeissenberg, Niederschl. Winter (Resid.) 1879-2008")

detach(hohenNs)
dev.off()
```

Hohenpeißenberg, Temp. Jahr (Resid.) 1781–2008

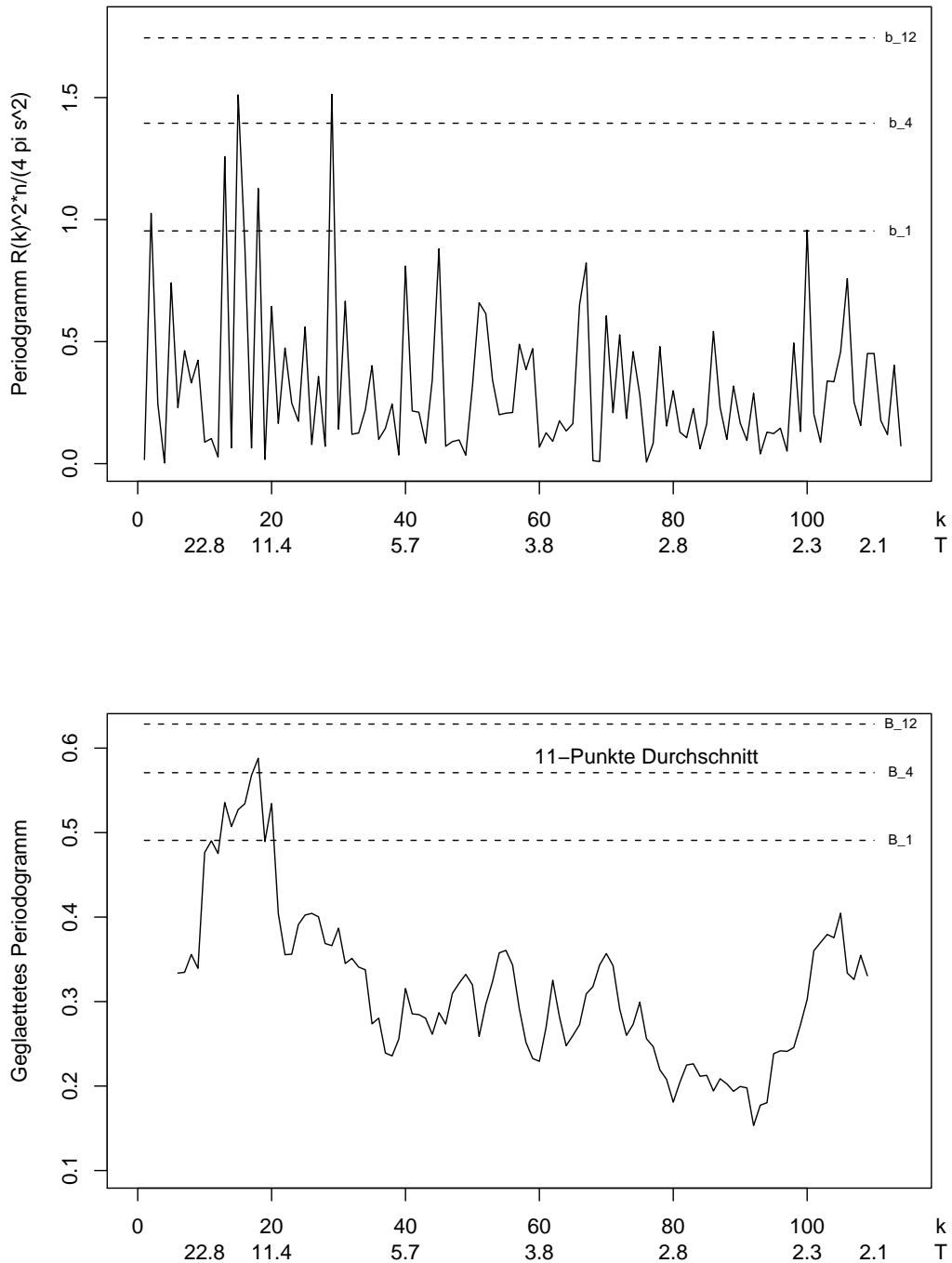


Abbildung 23: Hohenpeißenberg. Temperatur-Mittelwerte Jahr. Zeitreihe der Residuen vom polynomialen Trend. In das normierte Periodogramm (oben) bzw. das geglättete norm. Periodogramm (unten) eingetragen sind Schranken b_1, b_4, b_{12} bzw. B_1, B_4, B_{12} . Das Periodogramm weist Maxima auf bei Periodenlängen von $T = 17, 14, 10.7, 7.5, 3.5, 2.2, 2.1$ [Jahre].

Hohenpeissenberg, Temp. Winter (Resid.) 1781–2008

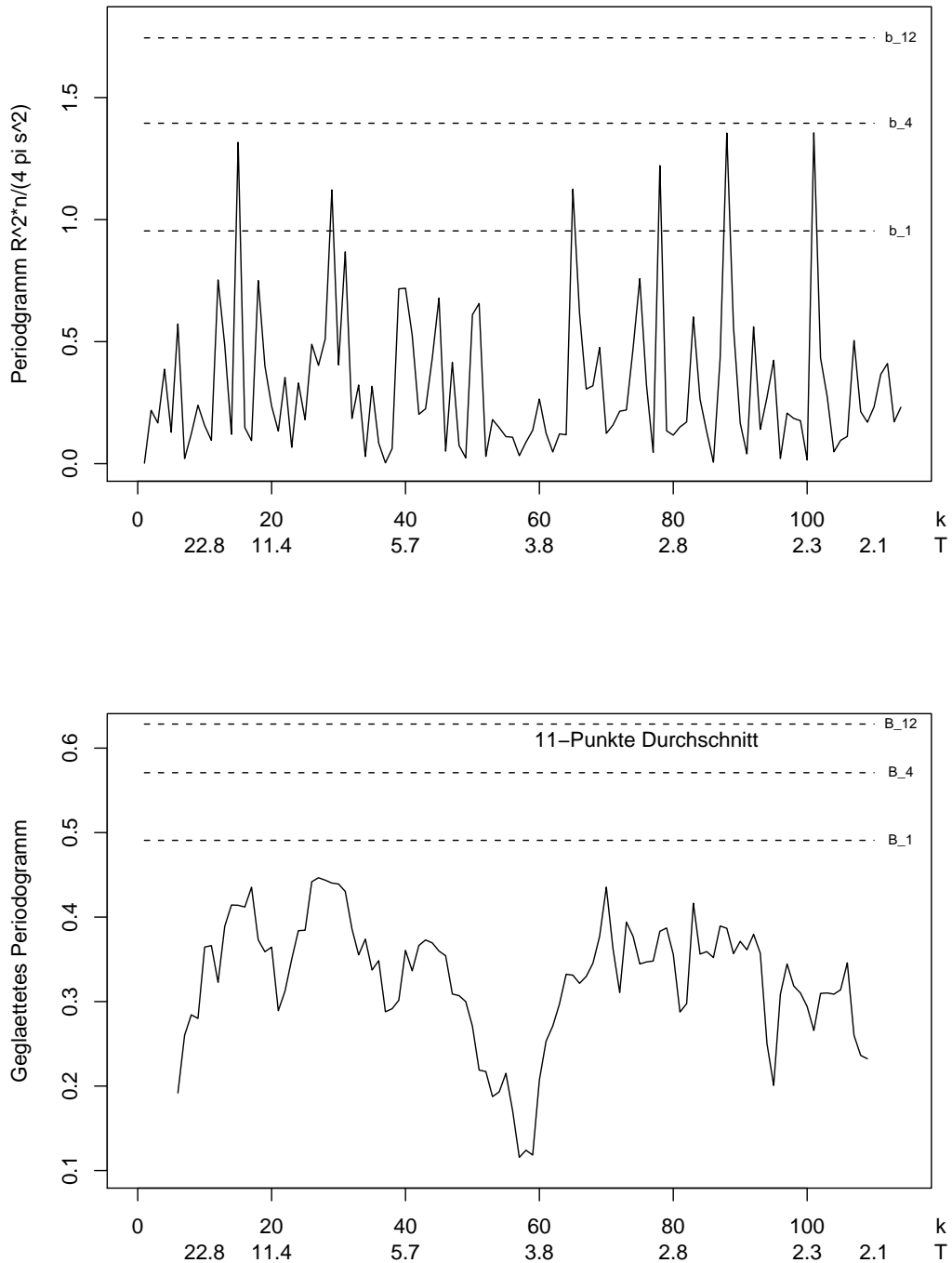


Abbildung 24: Hohenpeissenberg. Temperatur-Mittelwerte Winter. Zeitreihe der Residuen vom polynomialen Trend. In das normierte Periodogramm (oben) bzw. das geglättete norm. Periodogramm (unten) eingetragen sind Schranken b_1 , b_4 , b_{12} bzw. B_1 , B_4 , B_{12} . Das Periodogramm weist Maxima auf bei Periodenlängen von $T = 7.8$, 5.8 , 3.5 , 2.6 , 2.2 [Jahre].

Potsdam, Temp. Winter (Resid.) 1893–2008

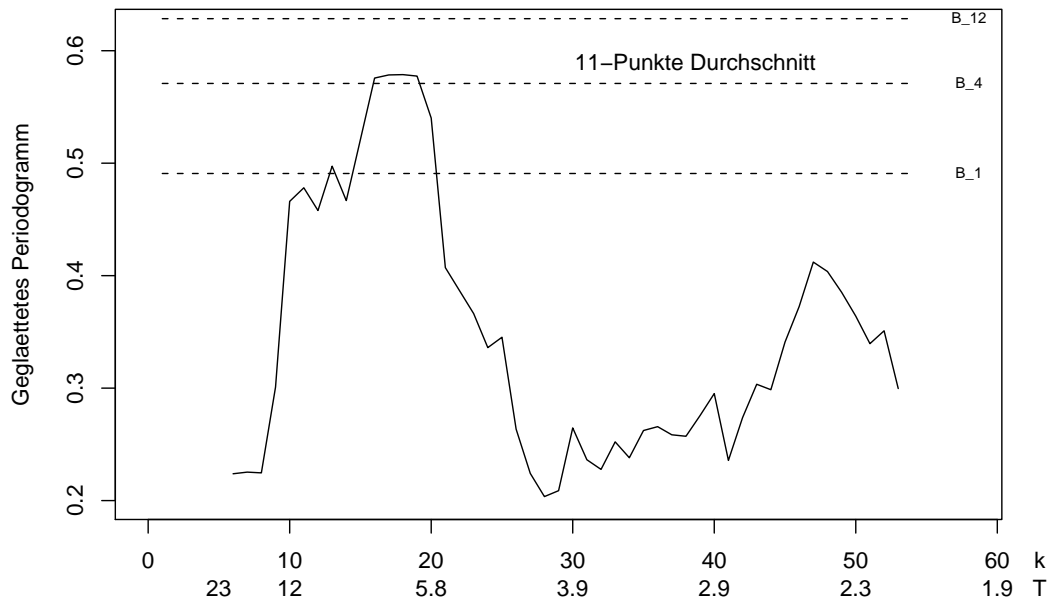
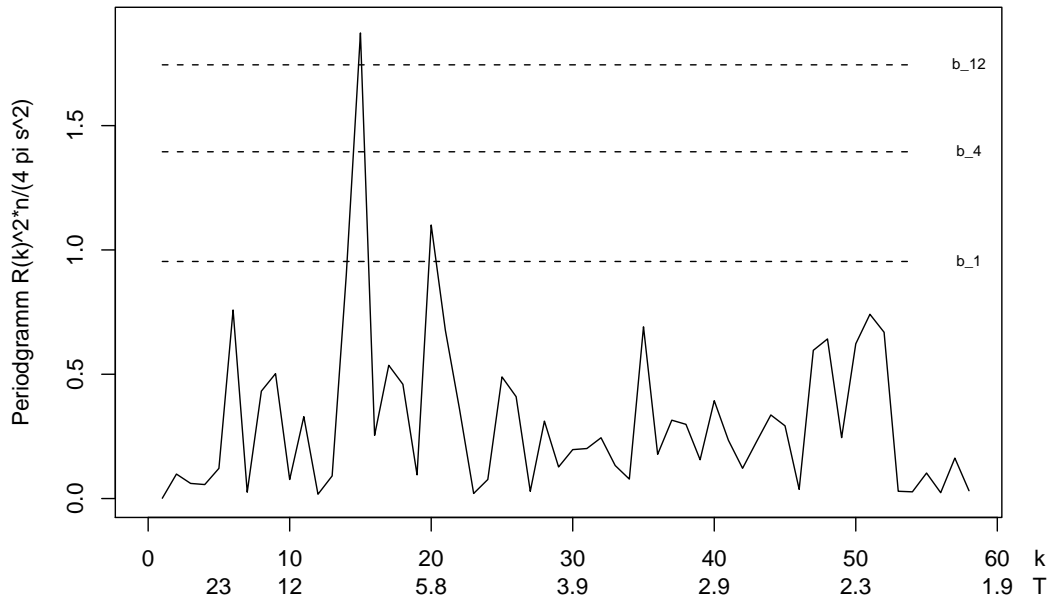


Abbildung 25: Potsdam. Temperatur-Mittelwerte Winter. Zeitreihe der Residuen vom polynomialen Trend. In das normierte Periodogramm (oben) bzw. das geglättete norm. Periodogramm (unten) eingetragen sind Schranken b_1 , b_4 , b_{12} bzw. B_1 , B_4 , B_{12} . Das Periodogramm weist Maxima auf bei Periodenlängen von $T = 7.8, 5.8, 3.5, 2.6, 2.2$ [Jahre].

Hohenpeissenberg, Niederschl. Winter (Resid.) 1879–2008

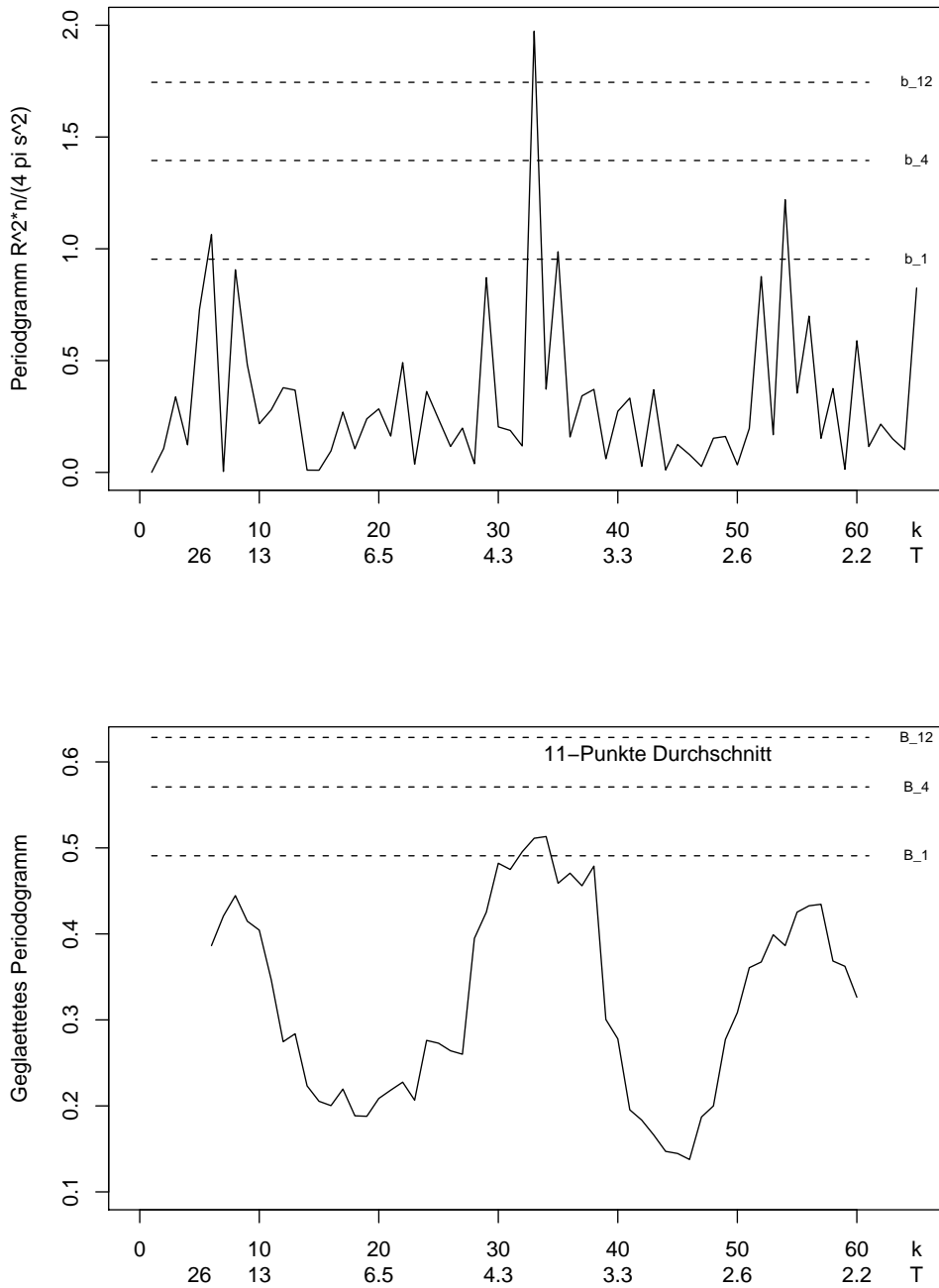


Abbildung 26: Hohenpeissenberg. Niederschlagshöhe Winter. Zeitreihe der Residuen vom polynomialen Trend. In das normierte Periodogramm (oben) bzw. das geglättete norm. Periodogramm (unten) eingetragen sind Schranken b_1 , b_4 , b_{12} bzw. B_1 , B_4 , B_{12} . Das Periodogramm weist ein Maximum auf bei einer Periodenlänge von $T = 4.0$ [Jahre].

Potsdam, Nieders. Winter (Resid.) 1893–2008

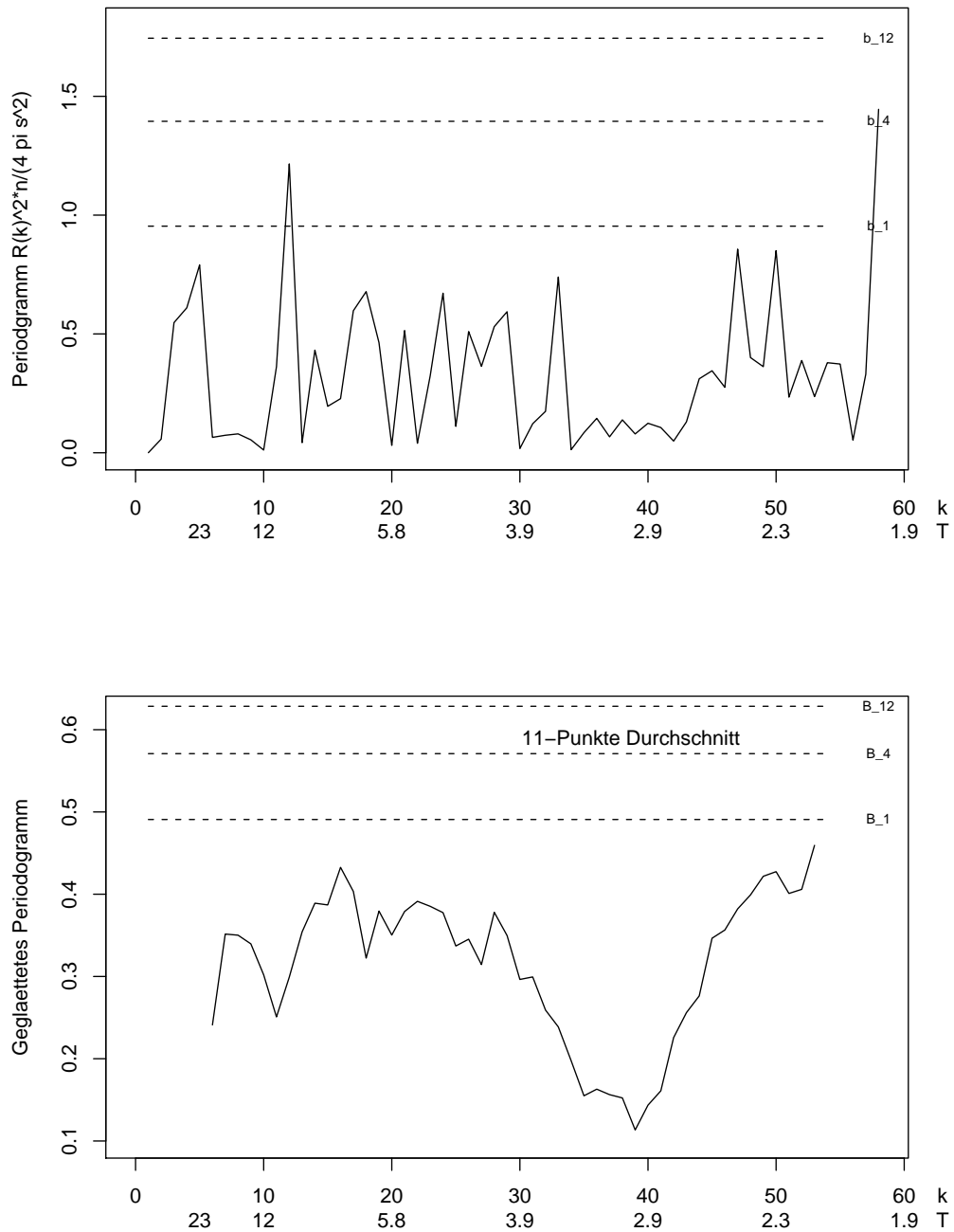


Abbildung 27: Potsdam. Niederschlagshöhe Winter. Zeitreihe der Residuen vom polynomialen Trend. In das normierte Periodogramm (oben) bzw. das geglättete norm. Periodogramm (unten) eingetragen sind Schranken b_1 , b_4 , b_{12} bzw. B_1 , B_4 , B_{12} . Das Periodogramm weist ein Maximum auf bei einer Periodenlänge von $T = 9.7$ [Jahre] sowie einen großen Wert am rechten Rand.

8 Anhang A: Statistische Methoden

Wir benutzen die folgenden Wahrscheinlichkeits-theoretischen Notationen.

Mit μ und σ^2 bezeichnet man i. d. R. Erwartungswert $\mathbb{E}(X)$ und Varianz $\text{Var}(X)$ einer Zufallsvariablen X , mit $\rho_{x,y}$ die Korrelation zweier Zufallsvariablen X und Y .

$N(\mu, \sigma^2)$, χ_p^2 , t_p und $F_{m,p}$ bezeichnen wie üblich die Normalverteilung mit Parametern μ und σ^2 , die χ^2 -Verteilung mit p Freiheitsgraden, die t -Verteilung mit p Freiheitsgraden und die F -Verteilung mit m und p Freiheitsgraden; ferner sind u_γ , $\chi_{p,\gamma}^2$, $t_{p,\gamma}$ und $F_{m,p,\gamma}$ die entsprechenden γ -Quantile, $0 < \gamma < 1$.

Für viele Verfahren, wie z. B. (13) oder (14) unten, wird vorausgesetzt, dass die Beobachtungswerte normalverteilt sind. Für große Stichproben-Umfänge n wird aber oft darauf verzichtet, diese Voraussetzung nachzuprüfen.

8.1 Univariate Stichproben

Pro Fall (Versuchseinheit) wird eine Variable x gemessen (ein Merkmal x erhoben). Die n -malige Wiederholung ergibt eine Stichprobe (x_1, x_2, \dots, x_n) vom Umfang n . Kenngrößen der *Lage* geben die zentrale Position der Stichprobenwerte auf der Messskala an, Kenngrößen der *Streuung* beschreiben die Schwankungsbreite dieser Werte.

Lageparameter

Das gebräuchlichste Maß der zentralen Lage ist der *Mittelwert* \bar{x} der Stichprobe, auch arithmetisches Mittel genannt, also

$$\bar{x} = \frac{x_1 + \dots + x_n}{n} \equiv \frac{1}{n} \sum_{i=1}^n x_i. \quad (11)$$

Streuungsparameter

Stichproben werden nicht nur durch ihre zentrale Lage auf der Messskala, sondern auch durch ihre Streubreite auf derselben charakterisiert. Wir beginnen mit der empirischen *Varianz*

$$s^2 = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1} \equiv \frac{1}{n - 1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (12)$$

bei der manchmal auch der Faktor $1/n$ statt $1/(n - 1)$ benutzt wird. Man rechnet nach der sog. *Verschiebungsformel*

$$s^2 = \frac{1}{n - 1} \left(\sum_{i=1}^n x_i^2 - n \cdot (\bar{x})^2 \right).$$

Aus s^2 wird – und das ist das gebräuchlichste Streuungsmaß – die (empirische) *Standardabweichung*

$$s = \sqrt{s^2}$$

abgeleitet. Bei einer Umskalierung der Werte x_i in $b \cdot x_i + a$ ($b > 0$) verändert sich der Wert s der Standardabweichung zu $b \cdot s$.

Konfidenzintervall und Test

Die Stichprobe $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ sei die Realisation von n unabhängigen, $N(\mu, \sigma^2)$ -verteilten Zufallsvariablen X_1, \dots, X_n . Wir schätzen die Parameter μ und σ durch den

Mittelwert \bar{x} bzw. die empirische Standardabweichung $s = \sqrt{s^2}$. Wegen $\text{Var}(\bar{x}) = \sigma^2/n$ lautet der *standard error* des Mittelwertes: $se(\bar{x}) = \sqrt{s^2/n} = s/\sqrt{n}$.

Konfidenzintervall. Ein Konfidenzintervall für μ zum Niveau $1 - \alpha$ wird durch

$$\bar{x} - t_{n-1, 1-\alpha/2} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{n-1, 1-\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad (13)$$

gebildet. Dabei bezeichnet, wie oben eingeführt, $t_{n-1, 1-\alpha/2}$ das $(1 - \alpha/2)$ -Quantil der t -Verteilung mit $n - 1$ F.G. Für große Stichproben-Umfänge n kann das Quantil $t_{n, \gamma}$ durch das Quantil u_γ der $N(0, 1)$ -Verteilung angenähert werden.

Signifikanztest. Es soll eine der folgenden (i) einseitigen bzw. (ii) zweiseitigen Hypothesen geprüft werden (μ_0 fest gewählt)

Zwei einseitige Versionen

Eine zweiseitige Version

$$(i) \quad H_0 : \begin{cases} \mu \geq \mu_0 \\ \mu \leq \mu_0 \end{cases} \quad \text{vs} \quad H_1 : \begin{cases} \mu < \mu_0 \\ \mu > \mu_0 \end{cases} \quad (ii) \quad H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0.$$

Wir bilden die Teststatistik

$$T = \sqrt{n} \frac{\bar{x} - \mu_0}{s}$$

und verwerfen H_0 zugunsten von H_1 , falls

$$(i) \quad \begin{cases} T < -t_{n-1, 1-\alpha} = t_{n-1, \alpha} \\ T > t_{n-1, 1-\alpha} \end{cases} \quad (ii) \quad |T| > t_{n-1, 1-\alpha/2} \quad [\text{t-Test}]. \quad (14)$$

Zwischen dem zweiseitigen Test (ii) aus (14) und dem Konfidenzintervall (13) besteht der folgende Zusammenhang: Die Zahl μ_0 liegt genau dann im Konfidenzintervall, wenn $|T| \leq t_{n-1, 1-\alpha/2}$, d. h. wenn H_0 nicht verworfen wird.

8.2 Univariate 2-Stichproben Situation

Deskription

Ein und dieselbe Variable (Messgröße) wird in zwei verschiedenen Situationen, z. B. in zwei verschiedenen Gruppen, unabhängig voneinander gemessen, und zwar in der ersten Gruppe n_1 mal, in der zweiten n_2 mal. Bezeichnen wir die Werte in der ersten Gruppe mit x , diejenigen in der zweiten Gruppe mit y , so haben wir das Datenschema

Stichprobe	Umfang	Stichprobenwerte	Mittelwert	Standardabw.	Erwartungswert	Varianz
1	n_1	x_1, \dots, x_{n_1}	\bar{x}	s_1	μ_1	σ^2
2	n_2	y_1, \dots, y_{n_2}	\bar{y}	s_2	μ_2	σ^2

Wir setzen voraus, dass die Werte Realisierungen von unabhängigen Zufallsvariablen sind, mit eventuell verschiedenen Erwartungswerten μ_1, μ_2 (aber mit gleichen Varianzen σ^2 bzw. gleichen Standardabweichungen σ) in den beiden Gruppen. Aus den beiden empirischen Varianzen s_1^2 und s_2^2 bilden wir die sog. *pooled* Varianz als Schätzer für σ^2 , d. i.

$$s^2 = \frac{1}{n_1 + n_2 - 2} \left((n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2 \right) \quad [\text{daraus } s = \sqrt{s^2}].$$

Es interessiert hier die Differenz $\mu_1 - \mu_2$ der Erwartungswerte, die durch $\bar{x} - \bar{y}$ geschätzt wird. Wegen $\text{Var}(\bar{x} - \bar{y}) = \sigma^2 (1/n_1 + 1/n_2)$ ergibt sich der standard error für $\bar{x} - \bar{y}$ zu

$$se(\bar{x} - \bar{y}) = s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = s \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}. \quad (15)$$

Konfidenzintervall und Test

Wir setzen die Normalverteilung der Beobachtungen voraus.

Ein Konfidenzintervall für $\mu_1 - \mu_2$ zum Niveau $1 - \alpha$ lautet mit der Abkürzung $t_0 = t_{n_1+n_2-2, 1-\alpha/2}$ und mit $se(\bar{x} - \bar{y})$ gemäß (15)

$$(\bar{x} - \bar{y}) - t_0 \cdot se(\bar{x} - \bar{y}) \leq \mu_1 - \mu_2 \leq (\bar{x} - \bar{y}) + t_0 \cdot se(\bar{x} - \bar{y}).$$

Signifikanztest. Die Hypothese $H_0 : \mu_1 = \mu_2$ gleicher Erwartungswerte, und damit völlig gleicher Verteilungen in den beiden Gruppen, wird mit Hilfe der Teststatistik

$$T = \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \cdot \frac{\bar{x} - \bar{y}}{s}$$

geprüft: Man verwirft H_0 zugunsten von $H_1 : \mu_1 \neq \mu_2$ auf dem Signifikanzniveau α , falls

$$|T| > t_{n_1+n_2-2, 1-\alpha/2} \quad [2\text{-Stichproben t-Test}].$$

8.3 Bivariate Stichprobe: Scattergramm und Korrelationskoeffizient

Deskription

Pro Fall (Versuchseinheit) werden jetzt zwei Variablen x und y gemessen (zwei Merkmale x und y erhoben). Die n Wertepaare

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

werden wie in Abb. 28 als Punkte in ein x-y-Diagramm eingetragen; die resultierende Punktwolke nennt man das Scattergramm (Streuungsdiagramm) der bivariaten Stichprobe.

Ein Maß für die Stärke der linearen Ausrichtung der Punktwolke stellt der (gewöhnliche oder Pearson-) *Korrelationskoeffizient* $r_{x,y}$ dar,

$$r_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad -1 \leq r_{x,y} \leq 1,$$

wobei $\bar{x} = \sum_{i=1}^n x_i/n$ und $\bar{y} = \sum_{i=1}^n y_i/n$ die Mittelwerte der x - bzw. y -Werte bezeichnen. Mit der (empirischen) *Kovarianz*

$$s_{x,y} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}) \quad [s_{x,x} = s_x^2, s_{y,y} = s_y^2]$$

lässt sich

$$r_{x,y} = \frac{s_{x,y}}{s_x \cdot s_y}$$

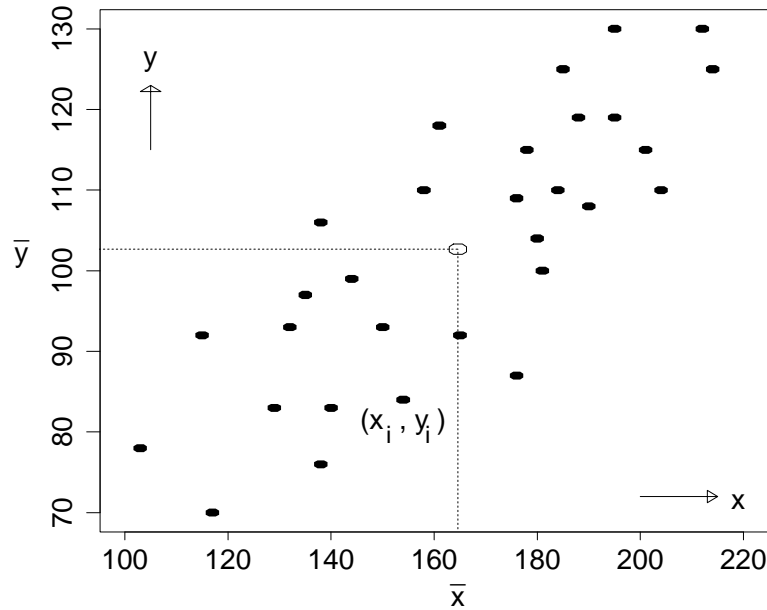


Abbildung 28: Scattergramm mit $n = 30$ Punkten, dem Korrelationskoeffizienten $r_{x,y} = 0.8$ und mit dem eingezeichneten Schwerpunkt (\bar{x}, \bar{y}) der Punktwolke.

schreiben. Ellipsenförmige (x, y) -Punktwolken und zugehörige Werte der Korrelation $r_{x,y}$ sind in der Abb. 29 illustriert.

Signifikanztest

Die Hypothese *unkorrelierter* Variablen x und y , das heißt $H_0: \rho_{x,y} = 0$, wird mit der Teststatistik

$$t = \sqrt{n-2} \frac{r_{x,y}}{\sqrt{1-r_{x,y}^2}} \quad (16)$$

geprüft. H_0 wird verworfen, falls

$$|t| > t_{n-2, 1-\alpha/2} \quad [\text{t-Test auf } \rho_{x,y} = 0].$$

Es reicht aus, dass $|r_{x,y}| > t_{n-2, 1-\alpha/2} / \sqrt{n-2}$ gilt. Für große n gibt es die Faustregel: Verwirf H_0 , falls $|r_{x,y}| > u_{1-\alpha/2} / \sqrt{n}$.

8.4 Bivariate Stichprobe: Einfache Lineare Regression

Deskription

Während die Rollen der Variablen x und y beim Korrelationskoeffizienten als gleichwertig anzunehmen sind (z. B. gilt ja $r_{x,y} = r_{y,x}$), ist dies in der sog. Regressionsanalyse nicht mehr der Fall.

Im Folgenden seien

- x eine Kontroll- (Einfluss-, Regressor-) Variable,
- Y eine Kriteriums-Variable.

Es werden n Werte x_1, \dots, x_n der x -Variablen voreingestellt und die zugehörigen Y -Werte gemessen, i. F. mit y_1, \dots, y_n bezeichnet. Durch das Scattergramm der bivariaten

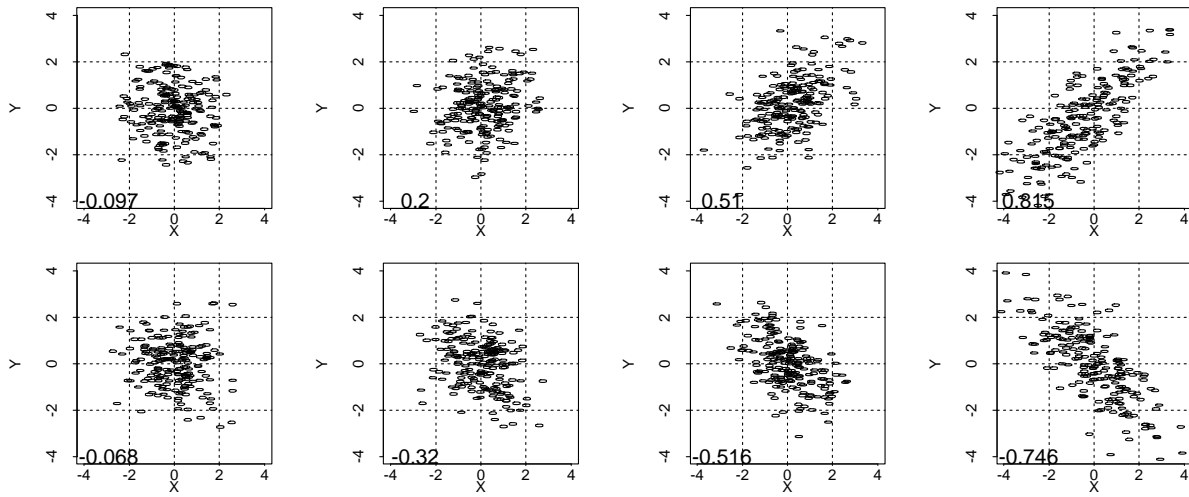


Abbildung 29: Scattergramms mit $n=200$ Punkten und mit variierendem $r_{x,y}$, *oben*: wachsendes (positives) $r_{x,y}$; *unten* fallendes negatives $r_{x,y}$

Stichprobe

$$(x_1, y_1), \dots, (x_n, y_n)$$

vom Umfang n wird eine Ausgleichsgerade (*Regressionsgerade*)

$$y = a + b \cdot x$$

gelegt, vgl. Abb. 30. Dabei wird der Ordinatenabschnitt a und die Steigung b nach der *Methode der kleinsten Quadrate* bestimmt, nämlich so, dass die Summe der vertikalen Abweichungsquadrate

$$SQ = \sum_{i=1}^n (y_i - (a + b \cdot x_i))^2$$

minimal wird. Das führt zu

$$\begin{aligned}
 b &= \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} && \text{[Regressionskoeffizient]} \\
 &= \frac{\sum_{i=1}^n x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^n x_i^2 - n \cdot (\bar{x})^2} && \text{[Rechenformel],} \\
 a &= \bar{y} - b \cdot \bar{x} && \text{[Ordinatenabschnitt].}
 \end{aligned} \tag{17}$$

Mit den Bezeichnungen $r_{x,y}$, $s_{x,y}$, $s_x^2 = s_{x,x}$ von 8.3 können wir auch

$$b = \frac{s_{x,y}}{s_x^2} = r_{x,y} \cdot \frac{s_y}{s_x}$$

schreiben. Man nennt die Werte der Regressionsgeraden an den Stellen x_i , das sind die Werte

$$\hat{y}_i = a + b \cdot x_i, \quad i = 1, \dots, n,$$

Schätzwerte oder *Prädiktionswerte*; die Differenzen zum Beobachtungswert, das sind

$$\hat{e}_i = y_i - \hat{y}_i = y_i - (a + b \cdot x_i), \quad i = 1, \dots, n,$$

die *Residuenwerte*.

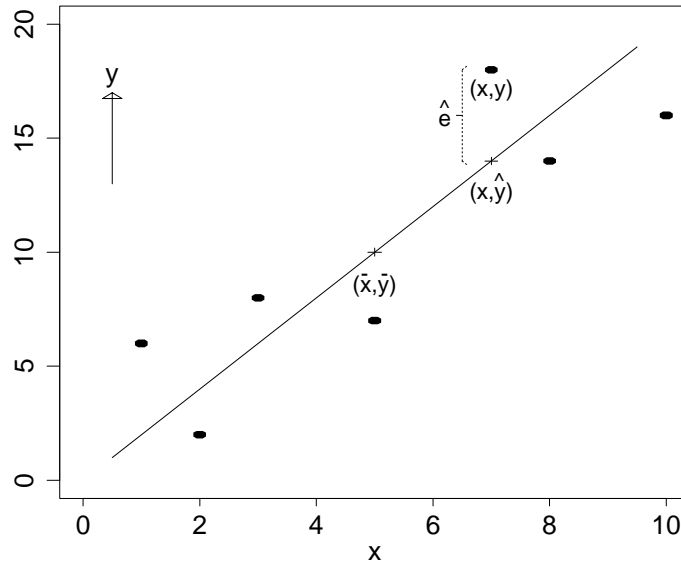


Abbildung 30: Scattergramm mit der Regressionsgeraden, auf welcher der Schwerpunkt (\bar{x}, \bar{y}) und die Punkte (x_i, \hat{y}_i) , $i = 1, \dots, n$, liegen.

Anders als der Korrelationskoeffizient ist der Regressionskoeffizient weder symmetrisch in x und y noch invariant unter linearen Skalentransformationen.

Test und Konfidenzintervalle

Wir machen die Modellannahme

$$Y_i = \alpha + \beta \cdot x_i + e_i, \quad i = 1, \dots, n,$$

wobei die Fehlervariablen e_i unabhängig sind, mit $\mathbb{E}(e_i) = 0$ und $\text{Var}(e_i) = \sigma^2$ für alle i . Schätzer für die unbekannt Parameter α und β sind $\hat{\alpha} = a$ und $\hat{\beta} = b$ gemäß Gleichungen (17). Ein Schätzer für die unbekannte Varianz σ^2 ist

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - (\hat{\alpha} + \hat{\beta} \cdot x_i))^2. \quad (18)$$

Wird zusätzlich eine Normalverteilung für die Fehlervariablen e_i angenommen, so können wir die folgenden Signifikanztests zum Niveau α (hier nicht zu verwechseln mit dem Ordinatenabschnitt α) und Konfidenzintervalle zum Niveau $\gamma = 1 - \alpha$ aufstellen:

- Testen des Regressionskoeffizienten β . Mit der Teststatistik

$$t = (\hat{\beta} - \beta_0) / se(\hat{\beta})$$

wird die Hypothese $H_0: \beta = \beta_0$ verworfen, falls $|t| > t_0$ gilt. Dabei ist $t_0 = t_{n-2, 1-\alpha/2}$ das $(1 - \alpha/2)$ -Quantil der t_{n-2} -Verteilung und

$$se(\hat{\beta}) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} = \frac{1}{\sqrt{n-1}} \frac{\hat{\sigma}}{s_x} \quad [\hat{\sigma} = \sqrt{\hat{\sigma}^2}, \text{ siehe (18)}]$$

der standard error von $\hat{\beta}$. Der Test auf $\beta = 0$ ist gleichwertig mit dem Test aus 8.3 auf $\rho_{x,y} = 0$, wenn man dort $r_{x,y} = (s_x/s_y)\hat{\beta}$ setzt und eine geeignete Verteilungsannahme

für das Paar X, Y trifft.

- Konfidenzintervall für den Parameter β . Mit dem eben eingeführten Quantil t_0 und standard error $se(\hat{\beta})$ lautet es

$$\hat{\beta} - t_0 \cdot se(\hat{\beta}) \leq \beta \leq \hat{\beta} + t_0 \cdot se(\hat{\beta}).$$

- Individuelles Konfidenzintervall für den Wert $\alpha + \beta x$ der (wahren) Regressionsgeraden an der Stelle x . Mit

$$se(\hat{\alpha} + \hat{\beta}x) = \hat{\sigma} \cdot \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$$

als standard error von $\hat{\alpha} + \hat{\beta}x$ und wieder mit $t_0 = t_{n-2, 1-\alpha/2}$ heißt es

$$(\hat{\alpha} + \hat{\beta}x) - t_0 \cdot se(\hat{\alpha} + \hat{\beta}x) \leq \alpha + \beta x \leq (\hat{\alpha} + \hat{\beta}x) + t_0 \cdot se(\hat{\alpha} + \hat{\beta}x). \quad (19)$$

8.5 Bedingte Wahrscheinlichkeiten

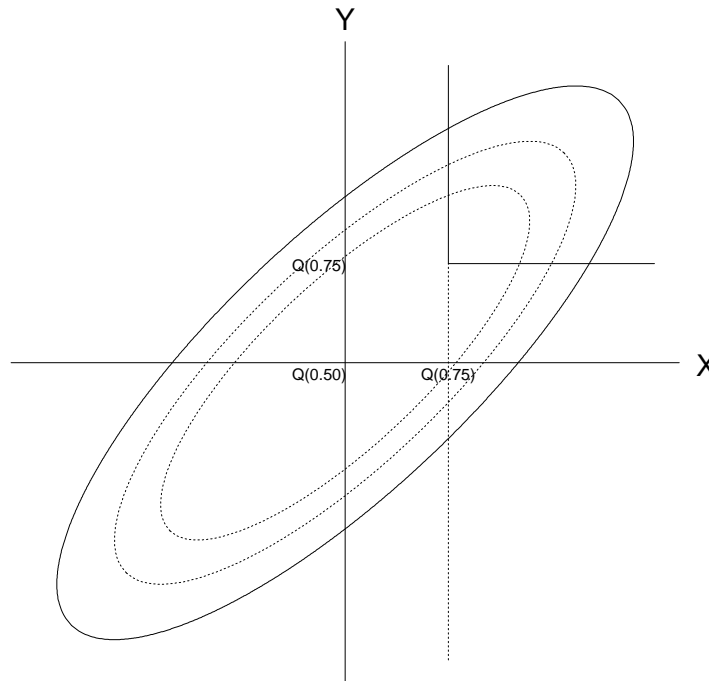


Abbildung 31: Elliptische Punktwolke für ein positives $r_{x,y}$. Obere rechte Quadranten beinhalten Messpunkte, die zu überdurchschnittlichen x- und y-Werten gehören.

Im Folgenden ist die Annahme zugrunde gelegt, dass X und Y zwei (2-dimensional) normalverteilte Zufallsgrößen sind, korreliert mit dem Koeffizienten ρ . Es wird nach der Wahrscheinlichkeit gefragt, dass auf einen X -Wert, der größer als ein bestimmter Schwellenwert Q^x ist, ein Y -Wert mit der analogen Eigenschaft folgt, nämlich größer als ein bestimmtes Q^y zu sein. Falls also der X -Wert größer als Q^x ist, dann ergeben sich (aufgefächert nach dem Korrelationskoeffizient $\rho = \rho_{x,y}$) die Treffer-Wahrscheinlichkeiten \mathbb{P} der Tabelle 4 dafür, dass auch der Y -Wert größer als Q^y ist. In der Abb.

31, die zur Veranschaulichung dient, haben wir als Schwellenwerte für X und für Y ihre 75 % Perzentile gewählt, und die Variablen normiert, so dass wir die Gleichheit $Q_{0.75}^x = Q_{0.75}^y \equiv Q_{0.75}$ haben. Die X -Werte, die in die Halbebene rechts von $Q_{0.75}$ fallen, liegen über dem 75 % Niveau, die (X, Y) -Werte, die in den rechten oberen Quadranten fallen, weisen beide diese Eigenschaft auf. Der Quotient der durch diese Flächen bestimmten Wahrscheinlichkeiten ist gerade die fragliche bedingte Wahrscheinlichkeit, die mit

$$\mathbb{P}(Y > Q_{0.75}^y | X > Q_{0.75}^x)$$

angegeben wird.

8.6 Zeitreihen: Autokorrelation

Deskription

Zeitreihen entstehen in Situationen, in denen eine Größe Y in regelmäßigen zeitlichen Abständen gemessen wird, und in denen die Möglichkeit besteht, dass der Wert von Y im Zeitpunkt t von den Messwerten zu den vorangegangenen Zeitpunkten $t - 1, t - 2, \dots$, abhängt. Wir bezeichnen die äquidistanten Zeitpunkte der insgesamt n Messungen (Beobachtungen) mit $t = 1, 2, \dots, n$. Die Zeitreihe können wir dann in der Form

$$Y_1, Y_2, \dots, Y_n$$

niederschreiben. Werden diese n Messwerte über der Zeitachse aufgetragen, d. h., werden die n Punkte (t, Y_t) , $t = 1, \dots, n$, in ein Koordinatensystem eingetragen und verbunden, so entsteht ein Zeitreihenplot.

Erwartungswert, Autokorrelation, Stationarität

Wir betrachten die Erwartungswerte $\mathbb{E}(Y_t)$ und die Kovarianzen

$$\text{Cov}(Y_t, Y_{t+h}) = \mathbb{E}((Y_t - \mathbb{E}(Y_t))(Y_{t+h} - \mathbb{E}(Y_{t+h}))).$$

Die Zeitreihe Y_t heißt *stationär*, falls die Erwartungswerte und die Kovarianzen nicht vom Zeitpunkt t abhängen; Letztere wohl aber von der Zeitdifferenz $t + h - t = h$. Wir können also unter Stationarität

$$\begin{aligned} \mathbb{E}(Y_t) &= \mu && \text{für alle } t = 1, 2, \dots, \\ \text{Cov}(Y_t, Y_{t+h}) &= \gamma(h), && \text{für alle } t = 1, 2, \dots; \quad h = 0, 1, \dots, \end{aligned}$$

schreiben. Im Spezialfall $h = 0$ haben wir

$$\gamma(0) = \mathbb{E}(Y_t - \mu)^2 \equiv \sigma^2 \quad \text{für alle } t = 1, 2, \dots \quad [\text{Varianz der } Y_t],$$

wobei stets $\sigma^2 > 0$ vorausgesetzt wird. Es gilt die Verschiebungsformel

$$\gamma(h) = \mathbb{E}(Y_t Y_{t+h}) - \mu^2.$$

Die Zeitdifferenz h wird auch *time lag* genannt; mit der symmetrischen Fortsetzung

$$\gamma(-h) = \gamma(h)$$

lassen wir den *time lag* h in ganz $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ variieren. Die Funktion $\gamma(h)$, $h \in \mathbb{Z}$, wird Kovarianzfunktion, auch *Auto-Kovarianzfunktion*, genannt.

Aus der Auto-Kovarianzfunktion $\gamma(h)$, $h \in \mathbb{Z}$, leitet man die Auto-Korrelationsfunktion $\rho(h)$, $h \in \mathbb{Z}$, ab. Die Korrelation der Variablen Y_t und Y_{t+h} , das ist

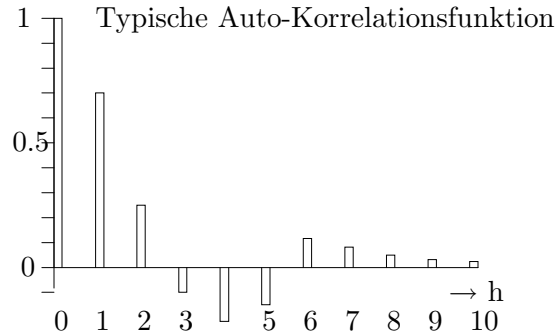
$$\rho(Y_t, Y_{t+h}) = \frac{\text{Cov}(Y_t, Y_{t+h})}{\sqrt{\text{Var}(Y_t) \cdot \text{Var}(Y_{t+h})}},$$

schreibt sich nämlich für stationäre Prozesse in der Form

$$\rho(Y_t, Y_{t+h}) = \frac{\gamma(h)}{\gamma(0)} \equiv \rho(h),$$

$h = 0, 1, \dots$, und hängt nicht mehr von t ab. Wir setzen sie wieder mit $\rho(-h) = \rho(h)$ symmetrisch fort. Für die *Autokorrelation* $\rho(h)$, $h \in \mathbb{Z}$, gilt

$$\rho(0) = 1, \quad -1 \leq \rho(h) \leq 1.$$



Eine stationäre Zeitreihe mit $\mu = 0$ und mit

- $\rho(h) = 0$ für alle $h = 1, 2, \dots$ heißt eine reine Zufallsreihe oder *white noise* Prozess,
- $\rho(h) = \alpha^h$ für alle $h = 1, 2, \dots$ heißt ein autoregressiver Prozess der Ordnung 1, kurz ein AR(1)-Prozess ($-1 < \alpha < 1$ fest vorgegeben, siehe 8.8 unten).

Schätzen und Testen

Die oben eingeführten –unbekannten– Kenngrößen einer stationären Zeitreihe müssen aus einer Beobachtung (Realisation) Y_1, Y_2, \dots, Y_n der Zeitreihe geschätzt werden.

Den Erwartungswert $\mu = \mathbb{E}(Y_t)$ (für alle t) schätzt man durch den Mittelwert $\bar{Y} = (1/n) \sum_{t=1}^n Y_t$. Einen Schätzer $c(h)$ der Autokovarianz $\gamma(h) = \mathbb{E}((Y_t - \mu)(Y_{t+h} - \mu))$ einer stationären Zeitreihe gewinnt man mit Hilfe der bivariaten Stichprobe

$$(Y_1, Y_{h+1}), \dots, (Y_{n-h}, Y_n) \tag{20}$$

vom Umfang $n - h$. Und zwar definiert man in der Zeitreihenanalyse die empirische Kovarianz von (20) durch

$$c(h) = \frac{1}{n} \sum_{t=1}^{n-h} (Y_t - \bar{Y})(Y_{t+h} - \bar{Y}), \quad h = 0, 1, \dots, \quad c(-h) = c(h).$$

Letzteres bedeutet wieder, wie schon bei $\gamma(h)$ geschehen, dass $c(h)$ für negative h -Werte symmetrisch ergänzt wird.

Der Schätzer $\hat{\sigma}^2$ für die Varianz $\sigma^2 = \text{Var}(Y_t)$, das ist

$$\hat{\sigma}^2 \equiv c(0) = \frac{1}{n} \sum_{t=1}^n (Y_t - \bar{Y})^2,$$

hat den Faktor $1/n$ und nicht – wie in der Standardstatistik üblich – den Faktor $1/(n-1)$.

Als empirische Korrelation $r(h)$ der bivariaten Stichprobe (20) verwendet man in der Zeitreihenanalyse den folgenden –aus $c(h)$ abgeleiteten– Schätzer,

$$r(h) = \frac{c(h)}{c(0)} = \frac{\sum_{t=1}^{n-h} (Y_t - \bar{Y})(Y_{t+h} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}, \quad h = 0, 1, \dots, \quad r(-h) = r(h).$$

Es ist $r(0) = 1$. Die Größen $r(h)$, über $h = 1, 2, \dots$ aufgetragen, bilden das sogenannte *Korrelogramm* der Zeitreihe. Es gibt eine Empfehlung, Analysen mittels Korrelogramme erst ab einem Stichprobenumfang $n \geq 50$ durchzuführen und $r(h)$ nur bis zu einem time lag $h \leq n/4$ zu verwenden. Die folgenden Tests sind für ein großes n konzipiert.

Test auf $\rho(1) = 0$. Die wichtige Autokorrelation zum time lag 1 wird auf den hypothetischen Wert 0 wie folgt getestet: Verwirf $H_0 : \rho(1) = 0$ zugunsten von $\rho(1) \neq 0$, falls

$$|r(1)| > u_{1-\alpha/2}/\sqrt{n} \quad [u_\gamma \text{ das } \gamma\text{-Quantil der } N(0,1)\text{-Verteilung}].$$

Test auf Zufallsreihe (oder *white noise* Prozess). Er basiert auf den k Werten $r(h)$, $h = 1, \dots, k$, des Korrelogramms. Man verwirft die Hypothese

$$H_0 : \text{ die Zeitreihe ist Realisation einer reinen Zufallsreihe}$$

zum Niveau α , falls, mit der *Bonferroni*-Schranke $b_k = u_{1-\beta/2}/\sqrt{n}$, $\beta = \alpha/k$,

mindestens einer der Werte $|r(1)|, |r(2)|, \dots, |r(k)|$ größer ist als b_k .

8.7 Zeitreihen: Frequenzanalyse

In der Frequenzanalyse wird die Oszillation der Zeitreihe in verschiedene Frequenzanteile zerlegt. Dabei lässt man sich von der Vorstellung leiten, dass die beobachtete Zeitreihe eine Überlagerung von zyklischen Komponenten mit den Kreisfrequenzen ω , welche zwischen 0 und π variieren, darstellt. Anstelle der Kreisfrequenz ω gibt man oft auch die Frequenz $\nu = \omega/(2\pi)$ an, welche dann zwischen 0 und $1/2$ liegt, oder die Periode(nlänge) $T = 1/\nu$, die größer gleich 2 ist.

Spektraldichte

Die Spektraldichte $f(\omega)$, $0 \leq \omega \leq \pi$, kurz auch *Spektrum* genannt, ist mit der Kovarianzfunktion $\gamma(h)$, $h \in \mathbb{Z}$, durch die Formeln

$$\begin{aligned} \gamma(h) &= \int_0^\pi f(\omega) \cos(h\omega) d\omega, & h = \dots, -1, 0, 1, \dots \\ f(\omega) &= \frac{1}{\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \cos(h\omega), & 0 \leq \omega \leq \pi, \end{aligned} \tag{21}$$

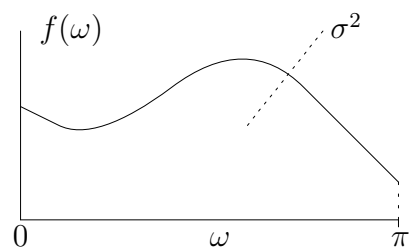
verknüpft. Die Funktionen γ und f bilden ein Paar von Fouriertransformierten. Im Spezialfall $h = 0$ drückt die Formel

$$\gamma(0) = \sigma^2 = \int_0^\pi f(\omega) d\omega \tag{22}$$

eine Zerlegung der Varianz σ^2 in die Frequenzanteile $f(\omega)$ aus. Aufgrund der Symmetrie der Cosinusfunktion, d. i. $\cos(-x) = \cos x$, lässt sich $f(\omega)$ auch

$$f(\omega) = \frac{1}{\pi} \left(\gamma(0) + 2 \sum_{h=1}^{\infty} \gamma(h) \cos(h\omega) \right)$$

schreiben. Vorausgesetzt wird hier stets, dass $\sum_{h=1}^{\infty} |\gamma(h)| < \infty$ gilt, was für die wichtigsten Beispiele von Zeitreihen der Fall ist.



- Für die reine Zufallsreihe ist die Spektraldichte konstant über dem Intervall $[0, \pi]$, nämlich

$$f(\omega) = \frac{\sigma^2}{\pi}, \quad 0 \leq \omega \leq \pi.$$

Alle Kreisfrequenzen aus dem Intervall $[0, \pi]$ liefern zur Varianz σ^2 der Zufallsreihe einen gleich großen Beitrag. Aus diesem Grund wird eine reine Zufallsreihe auch *weißes Rauschen* (*white noise*) genannt.

- Die Spektraldichte eines AR(1)-Prozesses (siehe 8.8 unten) lautet

$$f(\omega) = \frac{\sigma^2}{\pi} \frac{1 - \alpha^2}{1 - 2\alpha \cos \omega + \alpha^2}, \quad 0 \leq \omega \leq \pi.$$

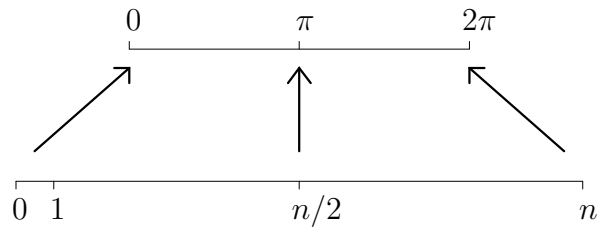
Fourierkoeffizienten

Wir verwenden das *Orthogonalsystem* der folgenden n Funktionen,

$$1, \quad \cos(\omega_j t), \quad j = 1, \dots, n/2, \quad \sin(\omega_j t), \quad j = 1, \dots, n/2 - 1, \quad (n \text{ gerade}),$$

mit den sog. Fourierfrequenzen

$$\omega_j = \frac{2\pi}{n} \cdot j,$$



vergleiche Abb. 32. Durch die Transformation $t \rightarrow (2\pi/n)t$ wird das Beobachtungsintervall $[0, n]$ auf das Intervall $[0, 2\pi]$ transformiert. Auf der Basis dieser Funktionen lautet die Fourier-Darstellung der Zeitreihe Y_1, \dots, Y_n

$$Y_t = 1 \cdot a_0 + \sum_{j=1}^{n/2} a_j \cdot \cos(\omega_j t) + \sum_{j=1}^{n/2-1} b_j \cdot \sin(\omega_j t), \quad t = 1, 2, \dots, n,$$

mit den *Fourierkoeffizienten* a_j, b_j .

Periodogramm

Wir definieren das Periodogramm $I(\omega_k)$, $k = 1, 2, \dots, n/2$, durch die Gleichung

$$I(\omega_k) = \frac{n}{4\pi} \cdot (a_k^2 + b_k^2), \quad k = 1, 2, \dots, n/2 - 1,$$

und tragen es auf über

- die Schwingungs-Anzahl $k = 1, 2, \dots, n/2$ bzw. über
- die Periode(n-Länge) $T = n, n/2, \dots, 2$ bzw. über
- die Kreisfrequenz $\omega = 2\pi/n, \dots, \pi$.

Der Periodogrammwert $I(\omega_k)$ gibt an, wie stark eine „Schwingung“ mit Schwingungsanzahl k (Periodenlänge $T = n/k$, Kreisfrequenz $\omega = 2\pi/n \cdot k$) an der Oszillation der Zeitreihe beteiligt ist.

In der Praxis trägt man das Periodogramm meistens in der normierten Form $I(\omega_k)/\hat{\sigma}^2$ auf, mit der empirischen Varianz $\hat{\sigma}^2 = c(0)$ der Zeitreihe Y_1, \dots, Y_n .

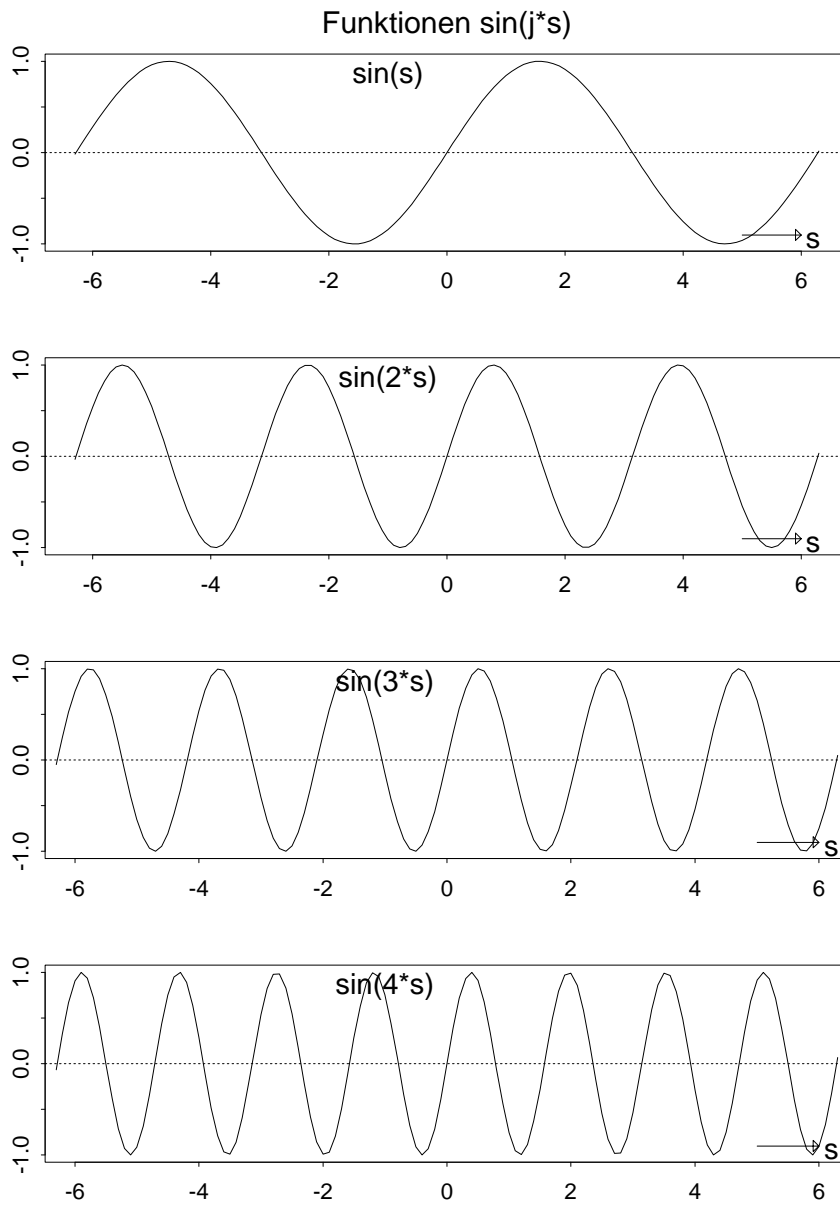
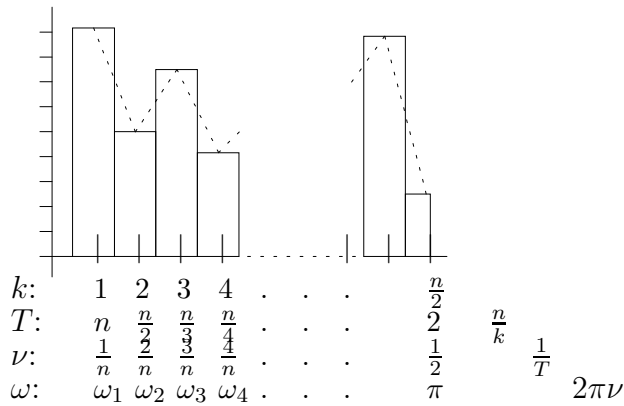


Abbildung 32: Die trigonometrischen Funktionen $\sin(j \cdot s)$, aufgetragen über s , für $j = 1, 2, 3, 4$. Die Substitution $s = \frac{2\pi}{n} \cdot t$ führt zu den Funktionen $\sin(\omega_j \cdot t)$.

Histogramm-Darstellung von $I(\omega_k)$



Das Periodogramm ist ein „Histogramm-Spektrum“ und sein Verlauf ist i. A. sehr „irregulär“. Durch Glättung des Periodogramms (dazu gibt es verschiedene Methoden) entsteht die **Spektraldichte**-Schätzung $\hat{f}(\omega)$ der Zeitreihe.

Der Quotient $I(\omega)/f(\omega)$ ist asymptotisch exponentialverteilt zum Parameter 1 (das Doppelte dieses Quotienten also χ_2^2 -verteilt). Der Quotient

$$\nu \cdot \frac{\hat{f}(\omega)}{f(\omega)}$$

ist asymptotisch χ_ν^2 -verteilt, mit F.G. ν , die noch von der Glättungsmethode abhängen.

8.8 Zeitreihen: Modelle

Wir stellen die Modellfamilie der Zeitreihen vom ARMA-Typ vor. Bei sämtlichen dieser Prozesse ist der Erwartungswert gleich Null.

Autoregressive und Moving average Prozesse

Eine Zeitreihe Y_t , $t \in \mathbb{Z}$, heißt ein *autoregressiver* Prozess der Ordnung p , kurz ein AR(p)-Prozess, falls

$$Y_t = \alpha_p Y_{t-p} + \dots + \alpha_2 Y_{t-2} + \alpha_1 Y_{t-1} + e_t, \quad t \in \mathbb{Z}, \quad (23)$$

gilt. Dabei ist e_t , $t \in \mathbb{Z}$, eine reine Zufallsreihe (*white noise* Prozess), e_t unabhängig von Y_{t-1}, Y_{t-2}, \dots . Die $\alpha_1, \alpha_2, \dots, \alpha_p$ sind unbekannte Parameter und $p \geq 0$ ist eine vorgegebene ganze Zahl. Die Parameter α_i müssen eine bestimmte Bedingung erfüllen (*Stationaritätsbedingung*), damit die Zeitreihe Y_t stationär ist, was i. F. stets stillschweigend vorausgesetzt wird. Es wird $\sigma_e^2 = \text{Var}(e_t)$ gesetzt.

Beispiel: Ein AR(1)-Prozess erfüllt die Gleichung

$$Y_t = \alpha Y_{t-1} + e_t, \quad t \in \mathbb{Z};$$

die Stationaritätsbedingung ist $-1 < \alpha < 1$. Die Autokovarianz- und Autokorrelationsfunktion lauten für den AR(1)-Prozess

$$\begin{aligned} \gamma(h) &= \sigma_e^2 \cdot \frac{\alpha^h}{1 - \alpha^2}, \quad h = 0, 1, \dots, & \text{insbesondere } \sigma^2 &\equiv \gamma(0) = \frac{\sigma_e^2}{1 - \alpha^2}, \\ \rho(h) &= \alpha^h, \quad h = 0, 1, \dots & \text{(also } \rho(h) &= \alpha^{|h|} \text{ für alle } h \in \mathbb{Z}). \end{aligned}$$

Eine Zeitreihe Y_t , $t \in \mathbb{Z}$, heißt ein *moving average* Prozess der Ordnung q , kurz ein MA(q)-Prozess, falls

$$Y_t = \beta_q e_{t-q} + \dots + \beta_2 e_{t-2} + \beta_1 e_{t-1} + e_t, \quad t \in \mathbb{Z}, \quad (24)$$

gilt. Dabei ist e_t , $t \in \mathbb{Z}$, eine reine Zufallsreihe, $\beta_1, \beta_2, \dots, \beta_q$ sind unbekannte Parameter und $q \geq 0$ ist eine vorgegebene ganze Zahl.

Beispiel: Für den MA(1)-Prozess reduziert sich (24) auf

$$Y_t = \beta \cdot e_{t-1} + e_t, \quad t \in \mathbb{Z}.$$

Wir erhalten für den MA(1)-Prozess $\sigma^2 \equiv \gamma(0) = \sigma_e^2 \cdot (1 + \beta^2)$, sowie

$$\rho(1) = \frac{\beta}{1 + \beta^2}, \quad \rho(h) = 0 \quad \text{für } h \geq 2.$$

ARMA(p,q)

Kombinationen von AR(p)- und MA(q)-Termen ergeben die ARMA(p,q)-Prozesse. Ein solcher Prozess ist also durch die Gleichung

$$Y_t = \alpha_p Y_{t-p} + \dots + \alpha_2 Y_{t-2} + \alpha_1 Y_{t-1} + \beta_q e_{t-q} + \dots + \beta_2 e_{t-2} + \beta_1 e_{t-1} + e_t \quad (25)$$

für alle $t \in \mathbb{Z}$ gegeben. Dabei setzen wir voraus, dass die Koeffizienten α_i die Stationaritätsbedingungen erfüllen. Ein ARMA(p,0)- bzw. ARMA(0,q)-Prozess ist ein AR(p)- bzw. MA(q)-Prozess.

Die ersten q Autokorrelationen $\rho(1), \dots, \rho(q)$ werden hier nicht wiedergegeben; sie hängen von den α_i als auch von den β_j ab. Für $h > q$ berechnet man $\rho(h)$ rekursiv allein unter Verwendung der α_i , nämlich nach der Formel

$$\rho(h) = \alpha_p \rho(h-p) + \dots + \alpha_1 \rho(h-1), \quad h > q.$$

Beispiel: Für den ARMA(1,1)-Prozess steht nur ein AR- und ein MA-Term in der Gleichung (25), das heißt

$$Y_t = \alpha Y_{t-1} + \beta e_{t-1} + e_t, \quad t \in \mathbb{Z}, \quad |\alpha| < 1.$$

Parameterschätzung und Prädiktion

Zum Schätzen der unbekannt Parameter

$$\alpha_1, \dots, \alpha_p, \quad \beta_1, \dots, \beta_q$$

eines ARMA(p,q)-Modells gehen wir von einer Realisation Y_1, Y_2, \dots, Y_n der Zeitreihe aus. Im ersten Schritt formen wir die Gleichung (25) nach der Residuen-(Fehler-) Variablen um,

$$e_t = Y_t - (\alpha_p Y_{t-p} + \dots + \alpha_1 Y_{t-1}) - (\beta_q e_{t-q} + \dots + \beta_1 e_{t-1}) \quad (26)$$

für $t = 1, \dots, n$. Dabei müssen wir die ersten q Residuenwerte e und die ersten p Beobachtungswerte Y vorgeben (z. B. durch $e = 0$ und $Y = \bar{Y}$) und dann die weiteren Residuenwerte rekursiv aus der Gleichung (26) berechnen (Wir wollen notationsmäßig nicht zwischen wahren Residuenvariablen e_t wie in (25) und den berechneten unterscheiden). Man beachte, dass die Werte der Residuenvariablen von den Werten α_i, β_j der Parameter abhängen. Darauf bildet man die Residuenquadrat-Summe

$$S_n(\alpha, \beta) = \sum_{t=1}^n e_t^2, \quad (27)$$

und bestimmt diejenigen Werte der Parameter α_i, β_j , bei denen (27) minimal wird (MQ-Methode).

Liegen zum Zeitpunkt t die vorangehenden Beobachtungen Y_{t-1}, Y_{t-2}, \dots vor, so wird die Prognose (Prädiktion) \hat{Y}_t für die anstehende Beobachtung Y_t unter Benutzung von Gleichung (25) berechnet. Dabei wird in (25) e_t gleich Null gesetzt und die übrigen Fehler-(Residuen-) Größen e_{t-1}, e_{t-2}, \dots werden rekursiv, wie unter (26) beschrieben, ermittelt. Mit Hilfe dieser so berechneten e_{t-j} haben wir dann die *ARMA-Prädiktion*

$$\hat{Y}_t = \alpha_p Y_{t-p} + \dots + \alpha_2 Y_{t-2} + \alpha_1 Y_{t-1} + \beta_q e_{t-q} + \dots + \beta_2 e_{t-2} + \beta_1 e_{t-1}. \quad (28)$$

ARCH/GARCH-Modelle

Der Prozess Y_t heißt ein ARCH(p)-Prozess ($p \geq 0$), falls die folgenden zwei Gleichungen (29) und (30) erfüllt sind. In der ersten, d. i.

$$Y_t = \sigma_t \cdot e_t, \quad t \in \mathbb{Z}, \quad (29)$$

ist e_t eine reine Zufallsreihe unabhängiger Variablen, mit $\text{Var}(e_t) = 1$; der (nichtnegative) Prozess σ_t erfüllt die Gleichung

$$\sigma_t^2 = \alpha_p Y_{t-p}^2 + \dots + \alpha_2 Y_{t-2}^2 + \alpha_1 Y_{t-1}^2 + \alpha_0, \quad t \in \mathbb{Z}, \quad (30)$$

mit $\alpha_0 > 0$, $\alpha_1, \dots, \alpha_p \geq 0$. Der ARCH(p)-Prozess Y_t hat einen Erwartungswert gleich 0; die bedingte Varianz von Y_t ist σ_t^2 (bedingt durch die vorangegangenen Beobachtungen bis zum Zeitpunkt $t-1$), und im Fall $\sum_{i=1}^p \alpha_i < 1$ erhält man die Momente

$$\text{Var}(Y_t) = \alpha_0 / (1 - \sum_{i=1}^p \alpha_i), \quad \text{Cov}(Y_t, Y_{t+h}) = 0 \text{ für } h \neq 0. \quad (31)$$

Der ARCH(p)-Prozess selber stellt bloß eine Folge unkorrelierter Zufallsvariablen dar; es ist seine variierende (bedingte) Varianz σ_t^2 , die ihn interessant macht. Der quadrierte Prozess Y_t^2 besitzt eine AR(p)-Gestalt der Form

$$Y_t^2 = \alpha_p Y_{t-p}^2 + \dots + \alpha_2 Y_{t-2}^2 + \alpha_1 Y_{t-1}^2 + \alpha_0 + \eta_t, \quad t \in \mathbb{Z},$$

mit einer reinen Zufallsreihe η_t .

Beispiel ARCH(1): Gleichung (30) heißt $\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2$, so dass (29) als

$$Y_t = \sqrt{\alpha_0 + \alpha_1 Y_{t-1}^2} \cdot e_t$$

geschrieben werden kann. Es gilt $\text{Var}(Y_t) = \alpha_0 / (1 - \alpha_1)$ (falls $\alpha_1 < 1$).

Der Prozess Y_t heißt ein GARCH(p, q)-Prozess ($p, q \geq 0$), falls wieder $Y_t = \sigma_t \cdot e_t$ gilt (e_t wie oben bei ARCH), und der (nichtnegative) Prozess σ_t die Gleichung

$$\sigma_t^2 = \alpha_p Y_{t-p}^2 + \dots + \alpha_2 Y_{t-2}^2 + \alpha_1 Y_{t-1}^2 + \alpha_0 + \beta_q \sigma_{t-q}^2 + \dots + \beta_1 \sigma_{t-1}^2, \quad t \in \mathbb{Z}, \quad (32)$$

erfüllt, mit α 's wie oben bei ARCH und mit $\beta_1, \dots, \beta_q \geq 0$. Es gelten die Aussagen, die um die Gleichung (31) herum für ARCH gemacht wurden, hier für GARCH analog.

Der GARCH(p, q)-Prozess (32) lässt sich in einen ARMA(p, q)-Prozess für Y_t^2 umschreiben, und zwar der Form (sei $p \geq q$ angenommen)

$$Y_t^2 = \gamma_p Y_{t-p}^2 + \dots + \gamma_2 Y_{t-2}^2 + \gamma_1 Y_{t-1}^2 + \alpha_0 - (\beta_q \eta_{t-q} + \dots + \beta_1 \eta_{t-1}) + \eta_t, \quad t \in \mathbb{Z}.$$

Dabei ist η_t eine reine Zufallsreihe und $\gamma_i = \alpha_i + \beta_i$ (falls $i \leq q$, sonst $\gamma_i = \alpha_i$).

Die Prognose (Prädiktion) $\hat{\sigma}_t^2$ für die anstehende Größe σ_t^2 kann unter Benutzung von Gleichung (32) rekursiv gewonnen werden. Liegen nämlich zum Zeitpunkt t die vorangehenden Beobachtungen Y_{t-1}, Y_{t-2}, \dots vor (sowie die bereits berechneten $\hat{\sigma}_{t-1}^2, \hat{\sigma}_{t-2}^2, \dots$), so setzen wir

$$\hat{\sigma}_t^2 = \alpha_p Y_{t-p}^2 + \dots + \alpha_2 Y_{t-2}^2 + \alpha_1 Y_{t-1}^2 + \alpha_0 + \beta_q \hat{\sigma}_{t-q}^2 + \dots + \beta_1 \hat{\sigma}_{t-1}^2. \quad (33)$$

Dabei müssen die ersten q $\hat{\sigma}^2$ -Werte vorgegeben werden (z. B. durch die empirische Varianz der Zeitreihe). Wir sprechen dann von einer *GARCH-Prädiktion* für die Varianz $\sigma_t^2 = \text{Var}(Y_t)$.

9 Anhang B: Datenauszüge

In den folgenden Tabellen ist der Dezemberwert des Vorjahres (dzvj) zu Beginn jedes Jahres nochmals aufgeführt, um die drei meteorologischen Wintermonate Dez., Jan., Feb. nebeneinander zu haben. Einige Datensätze sind vollständig unter

www.math.lmu.de/~pruscha/ Stichwort: Klimareihen

zu finden

Hohenpeißenberg

Mittlere Monats-[Jahres-] Temperaturen in 1/10 °C [1/100 °C]. Einen Zeitreihenplot der Jahresmittel und der Wintermittel bringt Abb. 1.

Jahr	dzvj	jan	feb	mar	apr	mai	jun	jul	aug	sep	okt	nov	dez	Tjahr	
1781	-18	-18	-10	24	87	122	145	154	166	126	44	15	12	723	
1782	12	-10	-54	0	38	94	156	176	144	108	36	-28	-23	531	
1783	-23	7	3	-4	64	108	131	163	144	118	82	12	-24	670	
1784	-24	-53	-46	0	21	128	132	152	136	143	23	12	-47	501	
1785	-47	6	-65	-60	13	91	117	131	131	141	60	23	-19	474	
1786	-19	-1	-30	-5	71	91	139	118	123	94	35	-5	-10	517	
1787	-10	-35	8	33	38	72	141	143	161	120	93	18	39	693	
1788	39	-19	21	23	56	116	148	176	140	135	56	-6	-105	618	
1789	-105	-10	-5	-34	74	131	110	147	144	110	65	4	12	623	
1790	12	-6	9	16	38	120	144	135	157	109	85	28	-11	687	
1791	-11	9	-18	20	89	97	130	148	164	110	72	14	10	704	
1792	10	-9	-20	37	78	91	138	154	158	98	81	21	-15	677	
.....														
1991	-17	-14	-27	50	42	61	119	168	166	141	58	25	-22	639	
1992	-22	-4	9	28	53	125	136	164	191	124	51	49	1	773	
1993	1	22	-18	14	81	125	138	139	150	112	61	-9	18	694	
1994	18	5	1	61	46	106	145	189	170	117	75	62	12	824	
1995	12	-20	30	5	61	104	111	181	146	98	113	20	-18	692	
1996	-18	-16	-32	-12	61	98	141	140	139	80	73	27	-21	565	
1997	-21	-10	27	45	37	110	128	139	170	133	62	37	11	741	
1998	11	5	34	18	64	115	149	152	161	109	73	-10	2	727	
1999	2	20	-32	36	60	122	125	161	155	146	78	5	-1	729	
2000	-1	-20	19	25	83	126	157	129	172	126	86	46	33	818	
2001	33	-0	8	43	43	131	123	163	170	87	128	0	-37	716	
2002	-37	7	32	48	56	113	167	155	154	99	76	54	12	811	
2003	12	-23	-38	48	63	127	193	172	207	125	43	58	13	823	
2004	13	-21	2	19	71	90	134	153	164	126	102	16	4	717	
2005	4	-9	-39	23	70	113	154	157	134	132	106	23	-28	697	
2006	-28	-23	-25	-3	61	109	150	198	121	155	117	64	27	793	
2007	27	22	29	36	112	121	151	155	149	103	68	10	-1	796	
2008	-1	25	27	16	53	127	150	155	157	102	85	36	-4	774	
		-----					-----								
		Winter					Sommer								

Niederschlagshöhen in 1/10 mm. Einen Zeitreihenplot der Jahressummen und der Wintersummen bringt Abb. 4.

Jahr	dzvj	jan	feb	mae	apr	mai	jun	jul	aug	sep	okt	nov	dez	Njahr
1879	578	254	619	272	1071	1039	1009	1473	1457	1645	685	861	393	10778
1880	393	385	253	315	967	1203	1991	1870	1212	997	1784	473	907	12357
1881	907	188	232	448	809	1332	1404	885	1490	1173	828	263	202	9254
1882	202	186	88	400	696	952	1565	1802	1314	1275	788	896	501	10463
1883	501	310	127	421	504	1179	2096	2020	835	1152	526	614	684	10468
1884	684	649	202	412	1164	440	1846	1957	1130	534	1360	268	432	10394
1885	432	100	277	643	285	1185	1437	1644	925	1366	761	462	968	10053
1886	968	244	171	436	828	625	2214	972	2288	382	430	436	682	9708
1887	682	145	125	888	291	1712	435	1550	718	699	647	637	952	8799
1888	952	431	667	494	1438	690	1575	1288	1733	1887	607	211	55	11076
1889	55	174	1084	502	701	1151	1738	1408	1019	1610	678	696	260	11021
1890	260	344	137	327	718	698	1426	2045	2265	1251	708	639	107	10665
...														...
1991	509	482	176	519	766	1503	1984	1216	599	861	328	694	651	9779
1992	651	88	707	1228	821	166	1557	1605	1371	713	961	1993	531	11741
1993	531	618	275	493	583	1376	1587	3465	2168	1032	767	376	719	13459
1994	719	688	328	862	1104	1073	783	1419	1201	938	497	733	836	10462
1995	836	450	526	713	1035	1070	2173	1169	2133	568	385	1059	905	12186
1996	905	138	314	588	597	1353	1018	1532	1849	1018	1082	920	417	10826
1997	417	18	585	673	934	425	1724	2404	457	319	947	189	956	9631
1998	956	398	280	1012	474	569	1389	1174	666	1632	1496	923	404	10417
1999	404	598	1187	474	927	3507	1625	1560	1194	1357	423	1341	1097	15290
2000	1097	300	802	1514	549	1537	1422	1743	2006	1413	997	551	275	13109
2001	275	681	664	1162	1107	628	2183	967	1626	1621	345	1018	773	12775
2002	773	103	685	992	723	952	1541	1509	1908	2203	821	1342	606	13385
2003	606	670	513	327	265	817	800	1504	815	450	1352	485	371	8369
2004	371	1127	431	677	547	1009	1573	1712	857	944	769	509	455	10610
2005	455	551	740	431	1191	1310	693	1840	2522	593	454	432	547	11304
2006	547	400	395	1025	1601	1123	1524	293	2453	690	679	504	433	11120
2007	433	608	520	488	173	2377	1079	2117	2065	1627	421	726	769	12970
2008	769	414	141	694	1546	942	879	1700	1709	701	735	510	448	10419

Winter

Sommer

Karlsruhe

Mittlere Monats-[Jahres-] Temperaturen in 1/10 °C [1/100 °C]. Einen Zeitreihenplot der Jahresmittel und der Wintermittel bringt Abb. 2.

Jahr	dzvj	jan	feb	mae	apr	mai	jun	jul	aug	sep	okt	nov	dez	Tjahr
1799	-7	-41	36	48	90	137	172	190	195	158	105	51	-36	921
1800	-36	30	04	27	146	173	154	194	208	169	98	66	20	1074
1801	20	37	33	79	108	167	168	196	187	167	116	67	45	1142
1802	45	-36	32	65	117	148	198	185	222	160	121	47	24	1069
1803	24	-18	-12	39	118	121	172	205	199	128	93	59	42	955

1804	42	55	08	41	98	167	192	191	184	159	115	53	-09	1045
1805	-09	-21	17	46	93	134	171	182	178	160	77	16	12	888
1806	12	49	48	62	76	173	179	198	188	153	95	72	76	1141
1807	76	-02	38	20	87	170	177	227	232	138	120	67	10	1070
1808	10	15	09	11	84	185	175	222	205	151	88	49	-30	970
1809	-30	13	65	61	64	160	165	190	189	148	87	28	37	1006
1810	37	-42	-10	71	102	148	173	189	187	177	103	63	38	999
...														...
1991	18	30	-05	88	97	123	166	226	220	178	98	51	11	1069
1992	11	18	38	74	105	169	183	212	224	159	86	76	26	1142
1993	26	50	09	66	129	165	187	190	195	144	93	23	56	1089
1994	56	45	29	96	98	149	191	241	203	154	98	94	58	1213
1995	58	23	73	57	112	149	167	229	204	141	134	48	11	1123
1996	11	01	10	44	110	134	184	189	195	127	110	63	-03	970
1997	-03	-21	61	88	95	153	177	192	222	163	97	58	39	1103
1998	39	37	52	78	107	167	190	192	203	154	110	35	32	1131
1999	32	45	21	77	114	166	180	217	199	188	105	46	36	1162
2000	36	30	61	78	122	168	202	178	208	162	121	77	52	1216
2001	52	34	52	80	90	172	169	212	212	133	143	44	14	1129
2002	14	12	72	78	106	149	204	197	200	147	111	83	45	1170
2003	45	10	08	89	111	162	230	217	243	163	84	73	31	1184
2004	31	23	45	63	116	138	183	201	208	165	120	58	16	1113
2005	16	32	07	66	115	152	202	207	182	176	128	55	21	1119
2006	21	-11	17	45	108	156	196	248	173	189	138	84	50	1161
2007	50	63	64	77	148	167	195	196	193	143	104	49	22	1184
2008	22	50	52	66	98	178	195	208	197	141	109	73	24	1159

Winter							Sommer							

Niederschlagshöhen in 1/10 mm. Einen Zeitreihenplot der Jahressummen und der Wintersummen bringt Abb. 5.

Jahr	dzvj	jan	feb	mae	apr	mai	jun	jul	aug	sep	okt	nov	dez	Njahr
1876	539	239	801	920	227	245	844	357	642	893	91	614	539	6412
1877	539	548	943	1050	915	824	732	1400	481	1230	326	584	648	9681
1878	648	615	335	654	1190	820	813	629	1230	653	759	365	661	8724
1879	661	748	837	94	998	535	1046	890	1379	877	720	557	450	9131
1880	450	57	510	101	1030	172	1420	372	359	415	1636	460	1454	7986
1881	1454	315	499	743	656	360	261	484	876	591	339	278	287	5689
1882	287	200	240	274	382	543	1036	1297	691	1444	685	1967	1006	9765
1883	1006	290	270	384	328	340	351	1766	382	841	981	878	645	7456
1884	645	513	374	415	285	526	686	532	901	245	582	360	990	6409
1885	990	171	769	1170	512	853	997	852	951	1133	1024	458	756	9646
1886	756	513	95	578	254	851	1252	603	343	670	861	779	1557	8356
1887	1557	73	91	680	310	1170	611	424	515	587	497	271	513	5742
...														...
1991	924	597	228	270	230	170	550	660	100	710	330	860	550	5255
1992	550	422	422	1039	430	574	1230	616	637	333	1150	1188	313	8354
1993	313	774	172	220	410	890	820	910	330	910	1066	300	1667	8469

1994	1667	629	495	710	910	835	916	329	681	1062	330	375	790	8062
1995	790	1241	527	848	641	1020	653	859	576	995	361	819	568	9108
1996	568	35	608	223	303	1183	853	695	511	172	674	1150	502	6909
1997	502	203	1335	402	256	575	1417	1033	515	116	641	655	736	7884
1998	736	616	216	341	942	290	719	514	229	740	1325	589	246	6767
1999	246	596	557	880	485	736	749	700	370	700	487	499	1657	8416
2000	1657	270	772	617	489	893	375	1120	520	878	432	727	465	7558
2001	465	705	600	1344	814	374	803	477	315	972	523	823	981	8731
2002	981	277	1288	590	346	1422	537	1560	478	825	900	1007	586	9816
2003	586	1018	128	241	276	675	614	644	241	236	931	428	231	5663
2004	231	1013	295	229	236	558	501	611	1069	388	981	389	340	6610
2005	340	331	518	351	762	657	218	883	734	311	345	295	626	6031
2006	626	254	442	854	336	642	652	1317	1567	268	1360	356	458	8506
2007	458	447	879	838	0	1100	1100	668	968	591	76	560	602	7829
2008	602	508	524	931	933	532	715	469	926	492	1070	604	628	8332

Winter

Sommer

Potsdam

Mittlere Monats-[Jahres-] Temperaturen in 1/10 °C [1/100 °C]. Einen Zeitreihenplot der Jahresmittel und der Wintermittel bringt Abb. 3.

Jahr	dzvj	jan	feb	mar	apr	mai	jun	jul	aug	sep	okt	nov	dez	Tjahr
1893	2	-82	18	45	85	127	164	183	172	127	105	24	8	817
1894	8	-15	24	53	101	120	145	190	155	111	77	50	3	848
1895	3	-35	-53	19	90	135	166	180	178	157	75	38	-9	791
1896	-9	-7	2	58	65	115	178	179	152	129	101	9	-8	812
1897	-8	-39	-1	50	78	113	180	164	178	127	73	26	18	810
1898	18	26	17	36	74	124	159	144	182	136	77	47	39	890
1899	39	24	24	28	78	121	148	180	168	126	82	72	-37	847
1900	-37	1	5	7	66	120	164	192	172	142	91	47	27	866
1901	27	-40	-37	25	83	136	162	195	170	134	102	31	9	815
1902	9	33	-19	32	69	96	162	156	144	119	68	8	-26	706
1903	-26	6	41	62	54	136	157	171	157	136	94	42	-13	872
1904	-13	-12	9	31	88	125	157	189	171	124	83	38	27	861
...														
1991	6	18	-26	66	78	103	143	203	181	155	89	44	13	895
1992	13	11	35	50	87	150	200	200	202	137	60	47	7	989
1993	7	20	0	40	111	160	159	164	161	124	85	1	31	887
1994	31	30	-8	57	93	133	161	225	184	136	75	66	35	995
1995	35	2	45	36	88	131	154	213	196	136	120	22	-31	929
1996	-31	-43	-29	8	98	120	164	163	186	113	97	48	-29	748
1997	-29	-25	42	55	67	137	175	190	212	145	80	33	16	942
1998	16	27	54	46	101	152	174	172	172	140	83	12	5	950
1999	5	29	11	54	100	142	165	206	182	180	97	37	22	1026
2000	22	10	39	53	115	160	183	164	182	140	117	63	28	1047
2001	28	6	16	31	80	151	150	199	192	124	126	41	-4	933
2002	-4	17	51	55	86	154	176	187	205	143	80	40	-22	980

2003	-22	-4	-20	47	92	154	196	202	206	151	60	61	22	981
2004	22	-13	26	50	101	130	158	173	195	144	101	44	20	941
2005	20	26	-7	34	104	135	169	188	166	156	112	45	11	955
2006	11	-36	-3	14	88	140	182	237	168	178	122	73	52	1017
2007	52	50	36	72	120	155	188	180	179	131	83	37	20	1046
2008	20	38	46	45	86	158	186	191	182	132	93	54	15	1024

Winter

Sommer

Niederschlagshöhen in 1/10 mm. Einen Zeitreihenplot der Jahressummen und der Wintersummen bringt Abb. 6.

Jahr	dzvj	jan	feb	mae	apr	mai	jun	jul	aug	sep	okt	nov	dez	Njahr
1893	375	277	796	262	10	170	362	678	366	354	789	712	211	4987
1894	211	189	486	299	382	467	951	614	1346	361	551	198	401	6245
1895	401	560	234	399	195	257	1057	405	416	173	720	551	547	5514
1896	547	284	162	577	289	343	564	731	960	1005	427	154	360	5856
1897	360	374	276	621	275	578	140	1177	315	689	239	277	258	5219
1898	258	371	471	856	587	696	549	1284	354	157	549	45	476	6395
1899	476	619	191	190	363	1572	439	1062	98	812	185	297	347	6175
1900	347	736	383	284	565	431	978	602	320	451	415	418	382	5965
1901	382	392	190	220	428	568	291	1110	514	478	564	724	450	5929
1902	450	456	168	619	463	645	577	805	785	736	262	32	315	5863
1903	315	290	498	128	387	559	393	987	597	645	653	601	73	5811
1904	73	222	520	116	405	723	498	284	143	354	324	597	322	4508
...														...
1991	739	266	311	406	347	354	883	280	889	136	250	467	417	5006
1992	417	514	390	867	343	186	114	765	681	250	336	678	565	5689
1993	565	487	360	90	232	958	889	956	414	692	183	317	978	6556
1994	978	762	86	869	691	715	677	390	1148	616	303	309	499	7065
1995	499	656	655	531	373	435	878	344	513	662	80	574	324	6025
1996	324	5	412	186	132	895	307	669	378	372	561	407	136	4460
1997	136	76	691	218	299	568	275	991	284	261	362	180	746	4951
1998	746	498	141	522	412	253	616	613	469	733	1153	404	429	6243
1999	429	410	594	447	351	472	458	217	366	114	90	278	728	4525
2000	728	488	497	743	397	456	299	450	1052	623	239	253	366	5863
2001	366	333	507	595	357	470	556	753	396	1257	406	375	695	6700
2002	695	297	774	374	508	860	365	750	1845	406	731	756	224	7890
2003	224	708	81	244	183	328	389	449	249	320	450	349	439	4189
2004	439	627	535	222	325	386	846	983	816	234	237	804	315	6330
2005	315	609	387	220	175	855	264	1498	572	462	368	298	634	6342
2006	634	228	488	466	534	545	371	863	792	84	247	343	346	5307
2007	346	851	535	554	3	1519	1114	987	867	789	108	626	306	8259
2008	306	813	221	502	642	63	363	519	520	616	688	415	389	5751

Winter

Sommer

Garching

Mittlere Tages-Temperaturen [1/10 °C] und tägliche Niederschlagshöhen [1/10 mm]
 Januar - Dezember 1966, Garching b. München, aus Fiedler (1970).

Tag	Mon	TP	NS				
1	1	66	63	1	7	135	55
2	1	62	13	2	7	152	0
3	1	32	3	3	7	192	0
4	1	-7	16	4	7	202	0
5	1	-68	1	5	7	183	24
6	1	-87	0	6	7	161	215
7	1	-53	0	7	7	118	158
8	1	-47	0	8	7	118	152
9	1	-86	0	9	7	145	0
10	1	-77	1	10	7	171	0
11	1	-83	0	11	7	178	0
12	1	-93	12	12	7	170	12
13	1	-92	10	13	7	194	0
14	1	-92	3	14	7	167	20
15	1	-121	1	15	7	148	36
16	1	-144	0	16	7	151	12
17	1	-150	0	
18	1	-129	0	
19	1	-127	0	11	12	23	74
20	1	-115	0	12	12	24	107
21	1	-72	0	13	12	35	28
...			...	14	12	13	8
...			...	15	12	-3	1
15	6	188	0	16	12	-19	0
16	6	195	0	17	12	7	21
17	6	212	0	18	12	48	24
18	6	214	0	19	12	42	0
19	6	115	79	20	12	51	12
20	6	129	63	21	12	0	79
21	6	144	55	22	12	-9	8
22	6	165	65	23	12	15	46
23	6	178	0	24	12	48	51
24	6	161	73	25	12	2	103
25	6	132	108	26	12	-29	2
26	6	135	54	27	12	-63	0
27	6	138	62	28	12	-14	0
28	6	150	226	29	12	6	6
29	6	103	0	30	12	51	77
30	6	109	101	31	12	30	13

Danksagung

Dem Deutschen Wetterdienst in Offenbach am Main bin ich für die Überlassung ihrer historischen Daten zu Dank verpflichtet. Insbesondere habe ich Herrn Wolfgang Fricke vom Observatorium Hohenpeißenberg für die ausdauernde Beratung und die persönlichen Mitteilungen zu danken.

Literatur

1. Klimareihen

- W. Attmannspacher (Hrsg.) *200 Jahre meteorologische Beobachtungen auf dem Hohenpeißenberg 1781–1980*, Bericht Nr. 155 des DWD. Offenbach/m 1981
- F. Fiedler. *Klimawerte zur Temperatur- und Windschichtung*. Wiss. Mitt. Nr. 18 des meteor. Inst. der Univ. München 1971.
- H. Grebe. *Temperaturverhältnisse des Observatoriums Hohenpeißenberg*, Bericht Nr. 36 des DWD. Offenbach/m 1957

2. Meteorologie

- R. Glaser. *Klimageschichte Mitteleuropas*, Wissensch. Buchgem. Darmstadt 2001.
- H. Malberg. *Meteorologie und Klimatologie*, 3. Aufl. Springer 1997
- H. Malberg. *Bauernregeln*, 4. Aufl. Springer 2003.
- C.D. Schönwiese *et al.* *Klimatrend-Atlas*. Bericht Nr. 20 des Z. f. Umweltforschung Frankfurt/m 1993
- C.D. Schönwiese. *Klimaschwankungen*, Springer 1979.

3. Zeitreihenanalyse

- G.E.P. Box & G.M. Jenkins. *Time Series Analysis: Forecasting and Control*, Holden-Day 1976.
- P.J. Brockwell & R.A. Davis. *Time Series: Theory and Methods*, 2nd ed. Springer 2006.
- W.A. Fuller *Introduction to Statistical Time Series*, Wiley 1976.
- H. Pruscha. *A Note on Time Series Analysis of Yearly Temperature Data*. J. Royal Statist. Soc. A, 149, 174–185, 1986.
- R. Schlittgen & H.J. Streitberg. *Zeitreihenanalyse*, 8.Aufl. Oldenburg 1999.

4. Statistische Methoden, einschl. R oder S

- P. Dalgaard. *Introductory Statistics with R*, 2nd ed. Springer 2008.
- H. Pruscha. *Statistisches Methodenbuch: Verfahren, Fallstudien, Programmcodes*, Springer 2005.
- L. Sachs & J. Hedderich. *Angewandte Statistik. Methodensammlung mit R*, 12.Aufl. Springer 2006.
- W.N. Venables & B.D. Ripley. *Modern Applied Statistics with S*, 4th ed. Springer 2002.