39 Statistical Analysis of Climate Series

Helmut Pruscha
Mathematical Institute, University of Munich, Munich, Germany

W. Freeden, M.Z. Nashed, T. Sonar (Eds.), *Handbook of Geomathematics*, DOI 10.1007/978-3-642-01546-5_39,

Abstract The topic of this contribution is the statistical analysis of climate time series. The data sets consist of monthly temperature means and monthly precipitation amounts gained at three German weather stations. Emphasis lies on the methods of time series analysis, comprising plotting, modeling, and predicting climate values in the near future.

1 Introduction

Within the context of the general climate discussion the evaluation of climate time series gains growing importance. Here we use the monthly data of temperature (mean) and precipitation (amount) from three German weather stations, raised over many years. We analyze the series by applying statistical methods and describe the possible relevance of the results.

First the climate series (annual and seasonal data) will be considered in their own right by employing descriptive methods. Long-term trends — especially the opposed trends of temperature in the nineteenth and twentieth centuries — are statistically tested. The auto-correlations that are the correlations of (yearly, seasonally, monthly, daily) data, following each other in time, are established. These correlation coefficients are calculated before and after a trend or a seasonal component is removed. In this context, we also try to formulate well-known folk (or country) sayings about weather in a statistical language and to check their legitimacy. The notion of auto-correlation leads us to the problem, how to model the evolution of the underlying data process. To this end, we use ARMA-type time series models, applied to the differenced series, with a subsequent residual analysis to assess their adequacy. For the latter task, GARCH-type models can be employed. Based on the modeling and prediction of annual data, we proceed to do the same with monthly data.

2 Climate Series

Basic informations on three German weather stations and on the climate series, analyzed in the rest of the chapter, are presented.

2.1 Weather Stations

Our data sets stem from the following three weather stations; further information can be found in ➲ *Table 1* and in the Appendix.

Hohenpeißenberg. The mountain Hoher Peißenberg (989 m) is situated between Weilheim and Schongau (Bav.) and lies in the lee-area of the Alps. It is the place of weather recording since 1781.
Source: www.dwd.de/ (*Klima+Umwelt, Klimadaten*).
Further (Grebe 1957), (Attmannspacher 1981).

Karlsruhe. The town lies in west Germany in the upper Rhine lowlands. Weather recording started in 1799, but stopped at the end of 2008.
Source: www.klimadiagramme.de

Potsdam. Since 1893 we have weather records from this east German town near Berlin.
Source: http://saekular.pik-potsdam.de

Table 1
Survey on the three weather stations

Name	Height	Geogr. longitude	Geogr. latitude	Start of temp.series	Start of precip.series
Hohenpeiß.	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

2.2 Temperature Series

We have drawn two time series plots for each station: the annual means (upper plot) and the winter means (lower plot). The meteorological winter covers the December (of the last year) and January, February (of the actual year). Winter data are often considered as an indicator of general climate change. One finds the plots for

Hohenpeißenberg (1781–2008) in *Fig. 1*

Karlsruhe (1799–2008) and Potsdam (1893–2008) under the author's homepage.

Table 2 offers the outcomes of descriptive statistical measures that are mean value (m), standard deviation (s), auto-correlation of first order ($r(1)$). The latter describes the correlation of two variables, immediately following each other in time.

Discussion of the row *Year*: The annual mean values stand in a distinct order: Karlsruhe > Potsdam > Hohenpeißenberg. However, their oscillations s are nearly of equal size ($\approx$ 0.8), and so are even the auto-correlations $r(1)$. That is, the correlation between the averages of two consecutive years amounts to 0.29...0.36. We will see below, how much is due to the long-term trend of the series.

Discussion of the rows *Winter ... Autumn*: The winter data have the largest oscillations s and small auto-correlations $r(1)$. Even smaller are the $r(1)$ values of the autumn data (signalizing practically uncorrelation). The time series plot of the winter series (lower plot of *Fig. 1*) reflects the s and $r(1)$ values of the table. In comparison with the upper plot of the annual means it shows a high fluctuation, with no distinct trend, coming nearer to the plot of a pure random series.

2.3 Precipitation Series

Again, we have drawn two time series plots for each station: the yearly sums (upper plot) and the winter sums (lower plot). One finds the plots for

Hohenpeißenberg (1879–2008) in *Fig. 2*

Karlsruhe (1876–2008) and Potsdam (1893–2008) under the author's homepage.

Table 3 offers the outcomes of descriptive statistical measures that are total precipitation amount (h) in (mm) height, standard deviation (s), auto-correlation of first order ($r(1)$).

The *annual* precipitation amount at the mountain Hohenpeißenberg is twice the amount in Potsdam. The oscillation values s stand in the same order as the amounts h. That is different to the temperature results, where all three s values were nearly the same. Note that the precipitation scale has a genuine zero point, but the temperature scale has none (which is relevant for us).

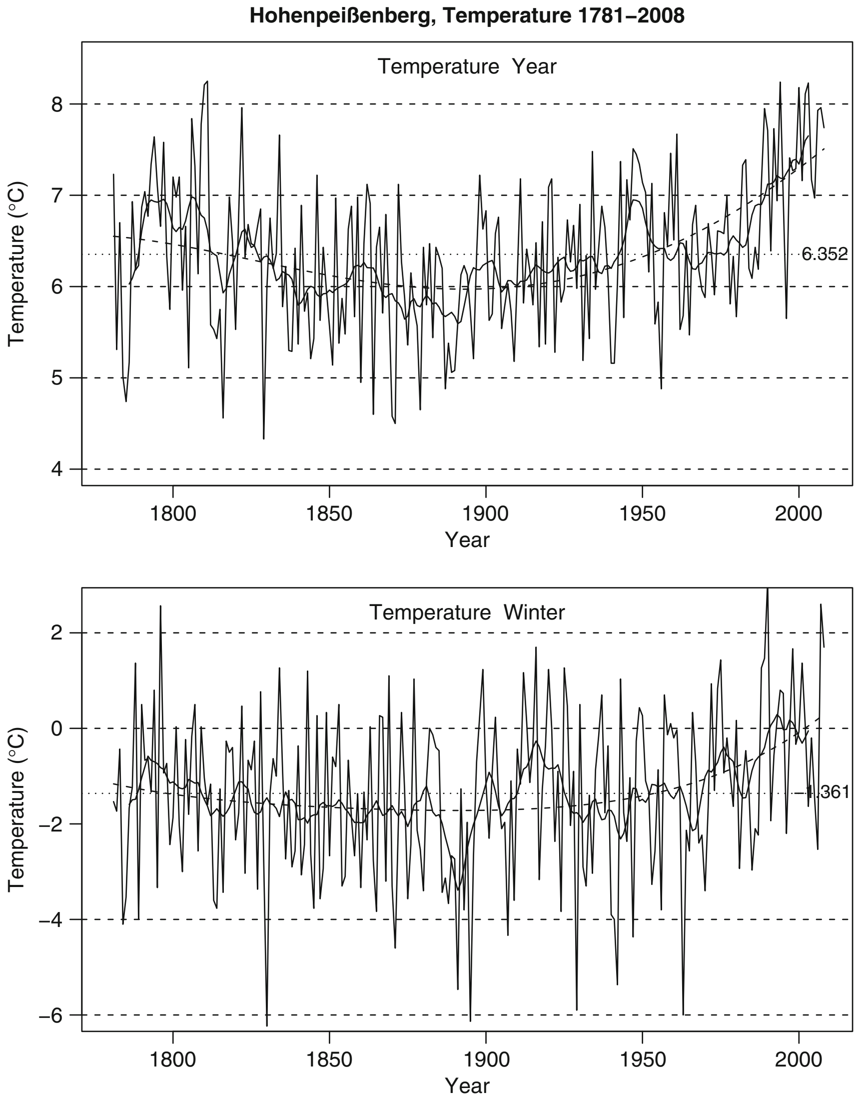

Fig. 1

Annual temperature means (*top*) and winter temperature means (*bottom*) in (°C), Hohenpeißenberg, 1781–2008; with a fitted polynomial of fourth order (dashed line), with centered moving (10-years) averages (inner solid line), and with the total average over all 228 years (horizontal dots)

The *winter* precipitation has — compared with the other three seasons — the smallest total h and the smallest oscillation s (winter temperature had the largest s)

While the precipitation series of winter and year in Karlsruhe and Potsdam — with their small $r(1)$ coefficients — resemble series of uncorrelated variables (also called *pure random series*), the series at Hohenpeißenberg, however, do not (see also sec. Section 4).

Table 2

Descriptive measures of the seasonal and annual temperature data in (°C) for the three stations

	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
Spring	5.56	1.32	0.16	10.18	1.08	0.24	8.45	1.15	0.19
Summer	14.25	1.08	0.20	18.72	1.07	0.25	17.40	1.01	0.15
Autumn	6.96	1.30	0.01	10.19	1.03	0.05	8.91	1.08	0.08
Year	6.35	0.84	0.29	10.22	0.80	0.33	8.78	0.81	0.36

3 Temperature Trends

In this section we study the long-term trend of temperature over the last two centuries.

3.1 Comparison of the Last Two Centuries

While temperature decreases in the nineteenth century, it increases in the twentieth century, see *Fig. 3*.

We report the following results:

1. The regression coefficients (slopes) $b = b_{\text{Temp}|\text{Year}}$, of the two — separately fitted — straight lines $\hat{y}_t = a + b \cdot t$, are tested against the hypothesis of a zero slope. The level 0.01 bound for the test statistic $T = |r|/\sqrt{1-r^2}$ is $t_{98,0.995}/\sqrt{98} = 0.265$. Herein, the correlation coefficient $r = b \cdot (s_{\text{Year}}/s_{\text{Temp}})$ is the dimension-free version of b. As *Table 4* informs us, the negative trend in the nineteenth century and the positive trend in the twentieth century are statistically well confirmed (at Hohenpeißenberg and in Karlsruhe). The test assumes uncorrelated residuals $e_t = y_t - \hat{y}_t$. This can be substantiated using the auto-correlation function of the e_t (not shown, but see Sec. Section 5 for similar analyses).
2. The total means m_1 and m_2 of the two centuries do not differ very much from each other and from the total mean m of the whole series, see *Table 4*. The average m_3 over the last 20 years is significantly larger than m, m_1 and m_2 [0.01 level]; that is immediately confirmed by a two sample test, even after a correction, discussed in the following. The warming in the last two decades is well established by our data.

When applying tests and confidence intervals to time series data, the effect of auto-correlation should be taken into account. To compensate, the sample size n is to be reduced to an *effective* sample size n_{eff}. As an example we treat the confidence interval for the true mean value μ of a climate variable, let's say the long-term temperature mean. On the basis of an observed mean value $\bar{y}$, a standard deviation s and an auto-correlation function $r(h)$, it is

$$\bar{y} - u_0 \cdot \frac{s}{\sqrt{n_{\text{eff}}}} \le \mu \le \bar{y} + u_0 \cdot \frac{s}{\sqrt{n_{\text{eff}}}}, \qquad u_0 = u_{1-\alpha/2},$$

with u_γ being the γ-quantile of the $N(0,1)$-distribution (a large n is assumed), and with Brockwell and Davis (2006) and von Storch and Zwiers (1999)

$$n_{\text{eff}} = \frac{n}{1 + 2 \cdot \sum_{k=1}^{n-1} (1 - (k/n)) \cdot r(k)}.$$

For an AR(1)-process with an auto-correlation $r = r(1)$ of first order we have $n_{\text{eff}} = n \cdot (1-r)/(1+r)$.

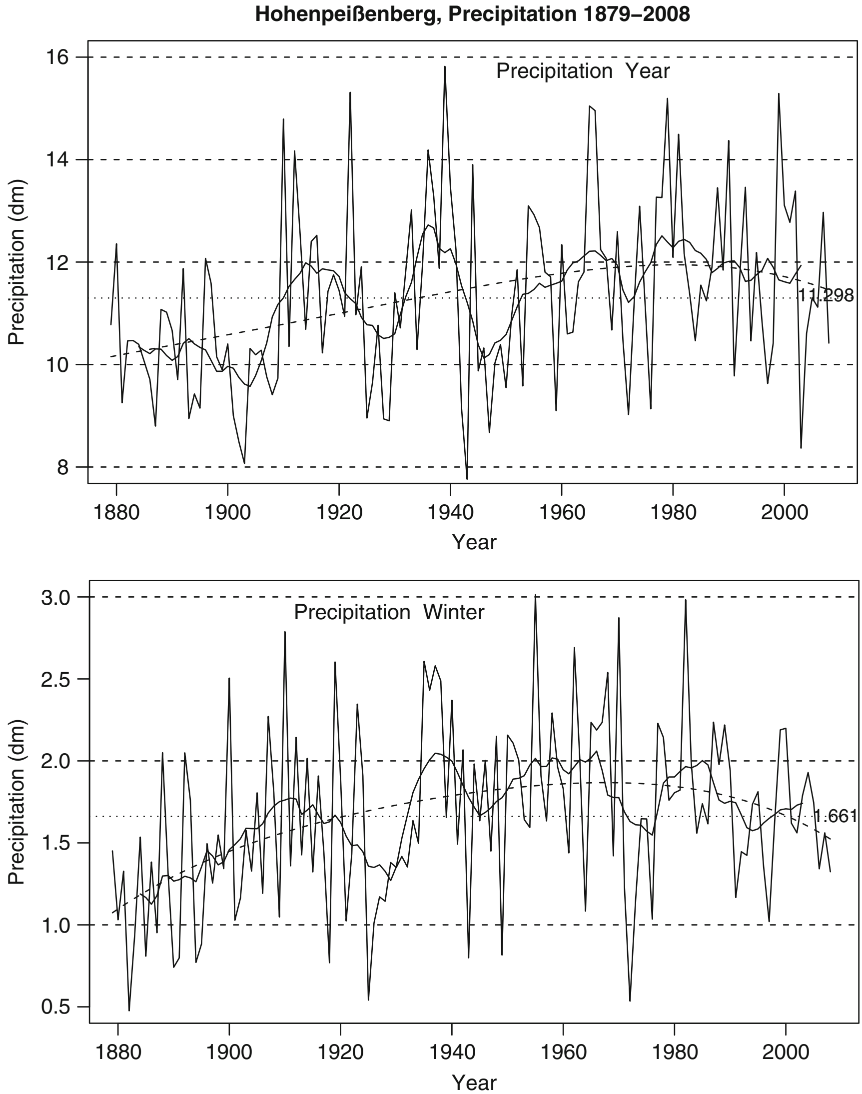

Fig. 2

Annual precipitation amounts (*top*) and winter precipitation amounts (*bottom*) in (dm), Hohenpeißenberg, 1879–2008; with a fitted polynomial of fourth order (dashed line), with centered moving (10-years) averages (inner solid line) and with the total average over all 130 years (horizontal dots)

Hohenpeißenberg: $n = 228$, $r = 0.289$, $\bar{y} = 6.352$, $s = 0.844$ lead to $n_{\text{eff}} = 125.76$ and thus to a 99% confidence interval $[6.158, 6.546]$.

Karlsruhe: $n = 210$, $r = 0.332$, $\bar{y} = 10.216$, $s = 0.802$ lead to $n_{\text{eff}} = 105.32$ and so to a 99 % confidence interval $[10.015, 10.417]$.

Table 3
Descriptive measures of the seasonal and annual precipitation amount in (mm) for the three stations

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

In both cases, at least 18 of the last 20 yearly temperature means lie above the upper 99% confidence limit, reinforcing the result 2. above.

The *winter* temperatures show the same pattern, but in a weakened form: The fall and the rise of the straight lines are no longer significant (see result 1.), 14 of the last 20 winter temperature means lie above the upper 99 % limit.

3.2 Historical Temperature Variation

Statistical results are formal statements; they alone do not allow substantial statements on Earth warming. Especially, a prolongation of the upward lines of ➋ *Fig. 3* would be dubious. An inspection of temperature variability of the last thousand years reveals that a trend (on a shorter time scale) could be turn out as part of the normal variation of climate system (Storch and Zwiers 1999; Schönwiese 1979).

4 Correlation: From Yearly to Daily Data

Scattergrams and correlation coefficients are defined for a bivariate sample $(x_1, y_1), \ldots, (x_n, y_n)$, where two variables, x and y, are measured n-times at comparable objects.

4.1 Auto-Correlation Coefficient

How strong is an observation at time point t (named x) correlated with the observation at the succeeding time point $t + 1$ (named y)? That is, we are dealing now with the case, that x and y are the same variables (e.g., temperature Tp) but observed at different time points,

$$x = Tp(t), \qquad y = Tp(t+1) .$$

The scattergram of ➋ *Fig. 5* (left) presents the 12*228 monthly temperature means at Hohenpeißenberg. The corresponding correlation coefficient is $r = r(1) = 0.79$. The large value is due to the seasonal effects, i.e., to the course of the monthly temperatures over the year. It contains, so to say, much redundant information.

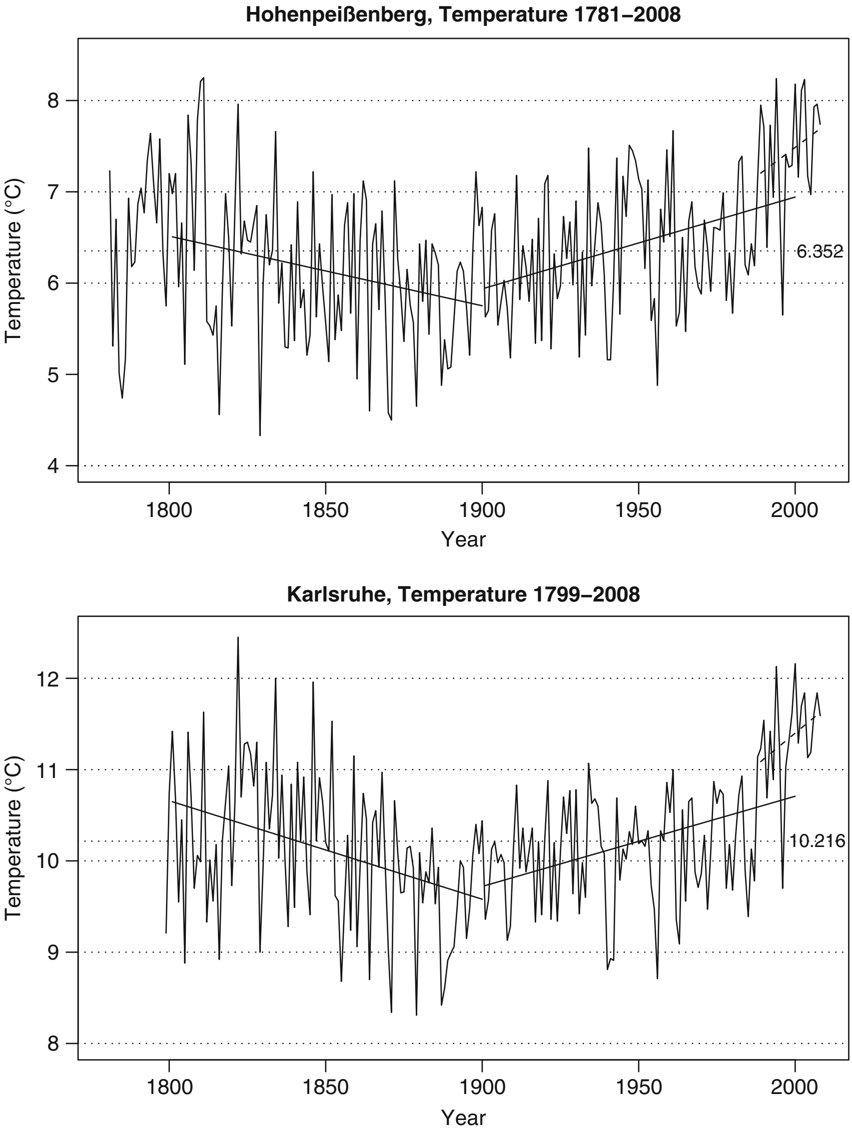

Fig. 3
Annual temperature means (°C) Hohenpeißenberg, 1781–2008 (*top*), Karlsruhe, 1799–2008 (*bottom*); with fitted straight line for each century, compare also Schönwiese et al. (1993). The fitted line for the last 20 years is also shown (dashed line)

In order to adjust, we first calculate the seasonal effects by the total averages for each month,

$$m_{\text{jan}}, \ldots, m_{\text{dec}}, \quad \text{together forming the seasonal component.}$$

Figure 4 gives the seasonal component for the three stations in form of histograms. Then we build seasonally adjusted data by subtracting from each monthly mean the corresponding

Table 4

Statistical measures for the temperature (°C) of the last two centuries and of the last 20 years

Period	Hohenpeißenberg				
	Mean value	Stand dev.	Regress. b*100	Correl. r	Test T
19th cent.	6.129	0.843	–0.763	–0.262	0.271
20th cent.	6.445	0.747	1.006	0.390	0.423
1989–2008	7.448	0.670	2.561	0.226	
Period	Karlsruhe				
	Mean value	Stand dev.	Regress. b*100	Correl. r	Test T
19th cent.	10.114	0.845	–1.079	–0.370	0.398
20th cent.	10.219	0.689	0.988	0.416	0.457
1989–2008	11.357	0.547	2.836	0.307	

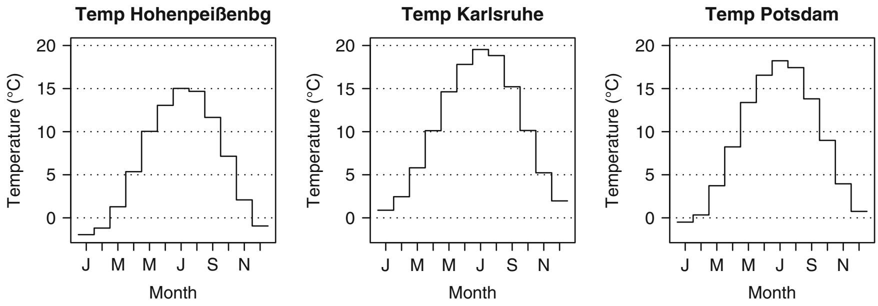

Fig. 4

Monthly temperatures; total averages at the three climate stations

seasonal effect. To the scattergram of Fig. 5 (right) belongs the correlation coefficient $r = 0.15$, which is much smaller than the $r = 0.79$ from above for the non-adjusted data.

Tables 5 and 6 bring auto-correlations $r(1) = r(Y_t, Y_{t+1})$ of climate variables for two successive time points. We deal with the variables

Y = yearly, quarterly, monthly, daily temperature and precipitation.

The $r(1)$ coefficients for *day* were gained from the 365 consecutive daily temperature and precipitation records of the year 1966 at the weather station Garching near Munich (Fiedler 1971).

Besides the auto-correlation $r(1)$ of the non-adjusted variables (put in parenthesis) we present the $r(1)$ coefficient for the adjusted variables without parenthesis. Herein adjustment refers to the removal of the trend component (here a polynomial of order 4) in the case of *year*, *quarter* and *days*. In the latter case the polynomial was drawn over the 365 days of the year (removal of the saisonal component in the case of *month*).

Note that the non-adjusted temperature variables do not have negative auto-correlations (persistence), but some precipitation variables do (switch over).

In the following, the outcomes for the adjusted series that are the figures of Tables 5 and 6 not in parenthesis, are only discussed.

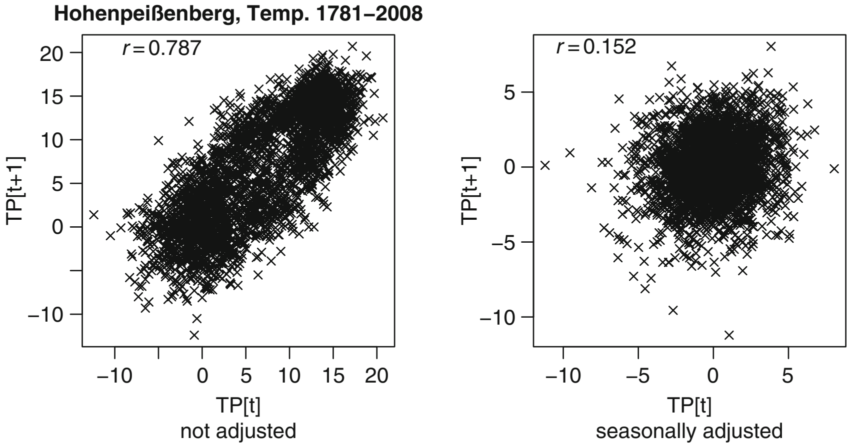

Fig. 5

Monthly temperature means TP = Y. Scattergrams $Y(t+1)$ over $Y(t)$ with $n = 12 * 228 - 1$ points, *left:* Original (not adjusted) variables, with correlation $r = 0.79$; *right:* Seasonally adjusted variables (i.e., removal of monthly total averages), with correlation $r = 0.15$

Table 5

Auto-correlation $r(1) = r(Y_t, Y_{t+1})$ for climate variables (Hohenpeißenberg; last row: Garching), without (in parenthesis) and with adjustment

	Temperature			Precipitation		
Succession	n	$r(Y_t, Y_{t+1})$		n	$r(Y_t, Y_{t+1})$	
Year → succeed.Year	227	(0.289)	0.118	129	(0.273)	0.184
Winter → succ. Wi	227	(0.077)	0.011	129	(0.144)	−0.006
Summer → succ. Su	227	(0.198)	0.104	129	(−0.116)	−0.168
Winter → succ. Su	227	(0.162)	0.101	129	(0.219)	0.170
Summer → succ. Wi	227	(0.061)	−0.016	129	(−0.025)	−0.098
Month → succ. Mo	2735	(0.787)	0.152	1559	(0.377)	0.012
Day → succeed. Day	365	(0.952)	0.833	365	(0.135)	0.117

Table 6

Auto-correlation $r(1) = r(Y_t, Y_{t+1})$ for climate variables (Karlsruhe), without (in parenthesis) and with adjustment

	Temperature			Precipitation		
Succession	n	$r(Y_t, Y_{t+1})$		n	$r(Y_t, Y_{t+1})$	
Year → succeed.Year	209	(0.332)	0.110	132	(0.009)	0.005
Winter → succ. Wi	209	(0.113)	0.060	132	(−0.041)	−0.082
Summer → succ. Su	209	(0.250)	0.064	132	(−0.201)	−0.230
Winter → succ. Su	209	(0.175)	0.121	132	(0.104)	0.127
Summer → succ. Wi	209	(0.119)	0.052	132	(−0.084)	−0.067
Month →succeed.Mo	2519	(0.811)	0.197	1595	(0.071)	0.029

Temperature: As to be expected, the auto-correlation of the *daily* data set is large. Smaller are those in the case of *month*, *season*, *year*. The seasonal auto-correlations are (with one exception) smaller than the *yearly*; especially the *winter* → *succeeding winter* correlation is small.

Precipitation: Only at the mountain Hohenpeißenberg the auto-correlation of *yearly* data differs distinctly from zero. Here the precipitation series has more inner structure than the series of Karlsruhe or Potsdam; see also the conclusions (Sec. Section 7). Completely different to the temperature situation, the correlations of the *daily* precipitation data are — perhaps against expectations — small and of the *monthly* data are nearly negligible.

What is the meaning of a particular $r(1)$ value when we are at time t and the immediately succeeding observation (at time $t+1$) is to be predicted.

4.2 Prediction of Above-Average Values

Assume that we have calculated a certain auto-correlation $r(1) = r(Y_t, Y_{t+1})$. Assume further that we have just observed an above-average (or an extreme) value of Y_t. What is the probability $\mathbb{P}$ that the next observation Y_{t+1} will be above-average (or extreme), too.

To tackle this problem, let X and Y denote two random variables, with the coefficient $\rho = \rho_{X,Y}$ of the true correlation between them. We ask for the probability that an observation X, being greater than a certain threshold value Q^x, is followed by an observation Y, exceeding a Q^y. If the X-value exceeds Q^x, then ➋ *Table 7* gives (broken up according to the coefficient ρ) the probabilities $\mathbb{P}$ for the event that the Y-value exceeds Q^y. As threshold values we choose *quantiles* Q_γ (also called $\gamma \cdot 100\%$ percentiles), for $\gamma = 0.5, 0.75, 0.90$. These threshold values could also be called: average value (more precisely an 50% value), upper 25% value, upper 10% value, respectively.

Examples: Assume that X turns out to exceed the X-average $Q^x_{0.50}$ (X-value being above-averaged). Then the probability that Y is above-averaged, too, equals

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

If X exceeds $Q^x_{0.90}$ (X being an upper 10% value) the probability that Y is an upper 10% value, too, equals

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

In the sequel X and Y will denote climate variables, where Y follows X in time.

4.2.1 Application to Climate Data

Once again, only the outcome for the adjusted series that are the figures in ➋ *Tables 5* and ➋ *6* not in parenthesis are discussed.

The absolute value $|r(1)|$ of most auto-correlations fall into the interval from 0.0 to 0.2. The ratio of hits — when observing an above-average climate value and predicting the same for the next observation — lies between 50% and 56% (according to ➋ *Table 7*). This is to compare with the 50% when pure guessing via "coin tossing" is applied. These modest chances of a successful prediction will find their empirical counterparts in ➋ *Table 8*.

The daily temperatures, with $r(1) > 0.70$, have a ratio above 75% for the prediction *above-average* → *above-average*. If we have an upper 10% day, then we can predict the same for the next day with success probability above 53% (to compare with 10% when merely guessing).

Table 7

Conditional probabilities for exceeding threshold values Q_γ

Conditional probability	Correlation $\rho = \rho_{X,Y}$							
	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70
$\mathbf{P}(Y > Q^y_{0.50} \mid X > Q^x_{0.50})$	0.50	0.53	0.56	0.60	0.63	0.67	0.70	0.75
$\mathbf{P}(Y > Q^y_{0.75} \mid X > Q^x_{0.75})$	0.25	0.29	0.34	0.40	0.45	0.51	0.57	0.64
$\mathbf{P}(Y > Q^y_{0.90} \mid X > Q^x_{0.90})$	0.10	0.14	0.17	0.24	0.29	0.39	0.45	0.53

Each entry is calculated by means of 40,000 simulations of a pair (X, Y) of two-dimensionally Gaussian random variables.

Table 8

Hit ratio of the rules 1–6. Explanations in the text

Ex	$X \to Y$		$r(X,Y)$	$\mathbb{P}$	% $[Y\diamond\ \bar{y}\ \mid X\diamond\ \bar{x}]$*			
			Hohen			Berlin	Karls	Hohen
1	Tp Dec	Tp Jan	0.13	.54	$[>\mid>]$	70 %	58 %	57 %
1	Tp Dec	Tp Feb	0.11	.53	$[>\mid>]$	60 %	62 %	56 %
2	Tp Sep	Tp Oct	0.14	.54	$[>\mid>]$	62 %	55 %	56 %
				.54	$[<\mid<]$	62 %	54 %	51 %
3	Tp Nov	Tp May	0.05	.51	$[>\mid>]$	50 %	55 %	52 %
	Pr Nov	Pr May	–0.02	.50	$[>\mid>]$	50 %	42 %	42 %
4	Tp Aug	Tp Feb	0.08	.53	$[>\mid>]$	73 %	56 %	53 %
	Pr Aug	Pr Feb	0.04	.51	$[>\mid>]$	50 %	50 %	47 %
5	Tp Sum	Tp Win	0.06	.51	$[>\mid>]$	–	54 %	54 %
				.49	$[<\mid>]$	–	46 %	46 %
	Pr Sum	Pr Win	–0.02	.50	$[>\mid>]$	–	41 %	42 %
				.50	$[<\mid>]$	–	59 %	58 %
6	Tp Win	Tp Sum	0.16	.55	$[>\mid>]$	–	51 %	55 %
				.55	$[<\mid<]$	–	55 %	62 %
				.45	$[>\mid<]$	–	45 %	38 %
	Pr Win	Pr Sum	0.22	.57	$[>\mid>]$	–	50 %	62 %
				.43	$[>\mid<]$	–	43 %	42 %

* $\diamond$ stands for a > or a < sign

4.2.2 Folk Sayings

Folk (or country) sayings about weather relate to

a narrow region (probably not covered here)
a particular time epoch (here centuries are involved)

and to the crop (Malberg 2003). The former weather observers (from the country or from monasteries) without modern measuring, recording, and evaluation equipments were pioneers of weather forecasting.

The following sayings are selected from Malberg (2003) and from popular sources. We kept the German language, but we have transformed them in *Table 8*.

Persistence rules

Ex. 1: Ist Dezember lind → der ganze Winter ein Kind
Ex. 2: Kühler September → kalter Oktober

Six-months rules
Ex. 3: Der Mai kommt gezogen ← wie der November verflogen
Ex. 4: Wie der August war → wird der künftige Februar
Yearly-balance rules
Ex. 5: Wenn der Sommer warm ist → so der Winter kalt
Ex. 6: Wenn der Winter kalt ist → so der Sommer warm

The columns of *Table 8* present: Transcription of the weather rules 1–6, with Tp standing for temperature and Pr for precipitation, correlation coefficient r from Hohenpeißenberg data, conditional probability $\mathbb{P}\left(Y > Q^y_{0.5} | X > Q^x_{0.5}\right)$, belonging to the r-value according to *Table 7*,

Percentage $\%\left[Y > \bar{y} | X > \bar{x}\right]$ of cases, in which an above-average X-value is followed by an above-average Y-value. This is given for Berlin–Dahlem 1908–1987 (Malberg 2003), Karlsruhe, Hohenpeißenberg.

Rule 2 aims at the percentage $\%\left[Y < \bar{y} | X < \bar{x}\right]$, rule 5 at $\%\left[Y < \bar{y} | X > \bar{x}\right]$, rule 6 at $\%\left[Y > \bar{y} | X < \bar{x}\right]$. These percentages are presented, too, in addition to the percentage $\%\left[Y > \bar{y} | X > \bar{x}\right]$.

The hit ratios, gained from the Hohenpeißenberg and from the Karlsruhe data, are rather poor and cannot confirm the rules. At most the persistence rules find a weak confirmation. In some cases another version of the rule (Ex. 2) or even the opposite rule (Ex. 5, Ex. 6) are proposed by our data. In connection with summer/winter prognoses the figures of the tables favor precipitation rules more than temperature rules.

With one (two) exceptions the Berlin–Dahlem series brings higher hit ratios than the series from Hohenpeißenberg (Karlsruhe). The reason could be, that the Dahlem series is shorter and is perhaps (climatically) nearer to the place of origin of the rules.

Note that the theoretical $\mathbb{P}$ values from *Table 7* are consistent with the empirical percentages in *Table 8* (both evaluated for Hohenpeißenberg). There are two or three exceptions, which relate to the precipitation data, showing once again their somewhat more irregular character.

5 Model and Prediction: Yearly Data

In the following we discuss statistical models, which can reveal (i) the mechanism of how a climate series evolves, and can support (ii) the prediction of climate values in the near future. Time series models of the ARMA-type will stand in the center of our analysis.

5.1 Differences, Prediction, Summation

Let Y be the time series of N yearly climate records; i.e., we have the data $Y(t)$, $t = 1, \ldots, N$. In connection with time series modeling and prediction the trend of the series is removed preferably by forming differences of consecutive time series values. From the series Y we thus arrive at the time series X, with

$$X(t) = Y(t) - Y(t-1), \quad t = 2, \ldots, N, \quad [X(1) = 0]. \tag{1}$$

Table 9 shows that the yearly changes X of temperature have mean ≈ 0 and an average deviation (from the mean 0) of $\approx 1\,(^\circ\text{C})$, at all three stations. The first-order auto-correlations

Table 9
Differences X of temperature means [°C] in consecutive years

Station	N	Mean	Stand.dev.	$r(1)$	$r(2)$	$r(3)$
Hohenp.	228	0.002	1.002	−0.469	0.019	−0.076
Karlsr.	210	0.011	0.921	−0.489	0.057	−0.052
Potsd.	116	0.018	0.924	−0.420	0.074	−0.197

$r(1)$ of the differences X lie in the range $-0.4\ldots-0.5$. After an increase of temperature follows — as a tendency — an immediate decrease in the next year, and vice versa.

We consider now the differenced time series $X(t)$ as sufficiently "trendfree" and try to fit an ARMA(p,q)-model. Such a model obeys the equation

$$\begin{aligned} X(t) =& \alpha_p X(t-p) + \cdots + \alpha_2 X(t-2) + \alpha_1 X(t-1) \\ &+ \beta_q e(t-q) + \cdots + \beta_2 e(t-2) + \beta_1 e(t-1) + e(t)\,, \end{aligned} \tag{2}$$

with error (residual) variables $e(t)$. For each time point t we can calculate a prognosis $\hat{X}(t)$ for the next observation $X(t)$, called ARMA-prediction. This is done on the basis of the preceding observations $X(t-1), X(t-2), \ldots$ in the following way. Eq. (2) is converted into

$$\begin{aligned} e(t) = X(t) - &\big(\alpha_p X(t-p) + \cdots + \alpha_1 X(t-1)\big) \\ &- \big(\beta_q\, e(t-q) + \cdots + \beta_1\, e(t-1)\big) \end{aligned} \tag{3}$$

for $t = 1, \ldots, n$. Here the first q error variables e and the first p observation variables X must be predefined. Then the further error variables can be recursively gained from equ. (3). The prognosis $\hat{X}(t)$ for the next observation $X(t)$ uses eq. (2), setting $e(t)$ zero, while the other variables $e(t-1), e(t-2), \ldots$ are recursively gained as described under (3). We have then the ARMA-prediction

$$\begin{aligned} \hat{X}(t) =& \alpha_p X(t-p) + \cdots + \alpha_2 X(t-2) + \alpha_1 X(t-1) \\ &+ \beta_q\, e(t-q) + \cdots + \beta_2\, e(t-2) + \beta_1\, e(t-1). \end{aligned} \tag{4}$$

The goodness of the prediction and hence the goodness-of-fit of the ARMA-model is assessed by the mean sum of squared errors, more precisely, by

$$\text{RootMSQ} = \sqrt{(1/N)\sum_{t=1}^{N}\big(X(t) - \hat{X}(t)\big)^2}\,. \tag{5}$$

From the differenced series X we get back by summation (also called integration) the original series Y. The prediction $\hat{Y}(t)$ for $Y(t)$ is gained by

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

Note that the calculation of $\hat{Y}(t)$ uses information up to time $t-1$ only. Due to $X(t) - \hat{X}(t) = Y(t) - \hat{Y}(t)$, the prediction $\hat{Y}(t)$ for $Y(t)$ is as good as the prediction $\hat{X}(t)$ for $X(t)$, namely by (5)

$$\text{RootMSQ} = \sqrt{(1/N)\sum_{t=1}^{N}\big(Y(t) - \hat{Y}(t)\big)^2}\,. \tag{6}$$

This procedure is called the ARIMA-method, the variables $\hat{Y}$ are referred to as ARIMA-predictions for $Y(t)$.

Table 10
ARIMA-method for the annual temperature means; coefficients, goodness-of-fit, prediction. H = Hohenpeißenberg, K = Karlsruhe, P = Potsdam

	Order	ARMA-coefficients		Root	ARIMA-prediction	
	p, q	α_i	β_j	MSQ	2006–2008	2009
H	2,2	−0.639,0.105	−0.177,−0.667	0.768	7.25, 7.40, 7.49	7.47
K	2,2	−0.266,0.070	−0.567,−0.281	0.707	11.25,11.33,11.42	11.44
P	3,1	0.15,0.04,−0.24	−0.915	0.726	9.44, 9.69, 9.82	9.70

5.2 Yearly Temperature Means

$Y(t)$ denotes now the temperature mean of year t and $X(t)$ – according to equ. (1) – the differenced series, i. e., the series of the yearly changes. It is X to which an ARMA(p,q)-model is fitted.

We choose order numbers (p, q) as small as possible, such that an increase of these numbers brings no essential improvement of the goodness measure RootMSQ. For the Hohenpeißenberg and Karlsruhe data we get $p = q = 2$ and therefore the ARMA(2,2)-model

$$X(t) = \alpha_2 X(t-2) + \alpha_1 X(t-1) + \beta_2 e(t-2) + \beta_1 e(t-1) + e(t) \tag{7}$$

(for Potsdam we obtain $p = 3, q = 1$). Table 10 shows the estimated coefficients α_i and β_j. As a rule, at least one α and one β is significantly different from zero. Further, the table offers the forecasts for the 3 years 2006–2008 as well as for the year 2009, each time on the basis of the preceding years. The actual observations 2006–2008 are slightly underestimated; compare the data excerpt in the Appendix and Fig. 6 (lower plot). This plot also shows the smoothing character of the predictions. For a clearer presentation we confine ourselves to the reproduction of the last 50 years (but for calculating the coefficients α, β the whole series was used).

5.2.1 Comparison with Moving Averages

Alternatively, prediction according to the method of left-sided moving averages can be chosen. As prediction $\hat{Y}(t)$ – for $Y(t)$ at time point t – the average of the preceding observations $Y(t-1), Y(t-2), \dots, Y(t-k)$ is taken. The depth number k denotes the number of years involved in the average. Once again by equ. (6) we calculate the goodness of this prediction method. Table 11 demonstrates that for a depth k smaller than 11 the RootMSQ-values of the ARIMA-method are not improved. Note that the latter method only needed $p + q = 4$ coefficients (but see also the remark in the conclusions).

5.2.2 ARIMA-Residuals

Having calculated the ARIMA-predictions $\hat{Y}(t)$ for $Y(t), t = 1, \dots, N$, we then build residuals

$$e(t) = Y(t) - \hat{Y}(t), \quad t = 1, \dots, N, \tag{8}$$

from these predictions; see Fig. 6 (top). Note that we already used residuals in equ. (6); as stated above we also have $e(t) = X(t) - \hat{X}(t)$. We ask now for the structure of the residual

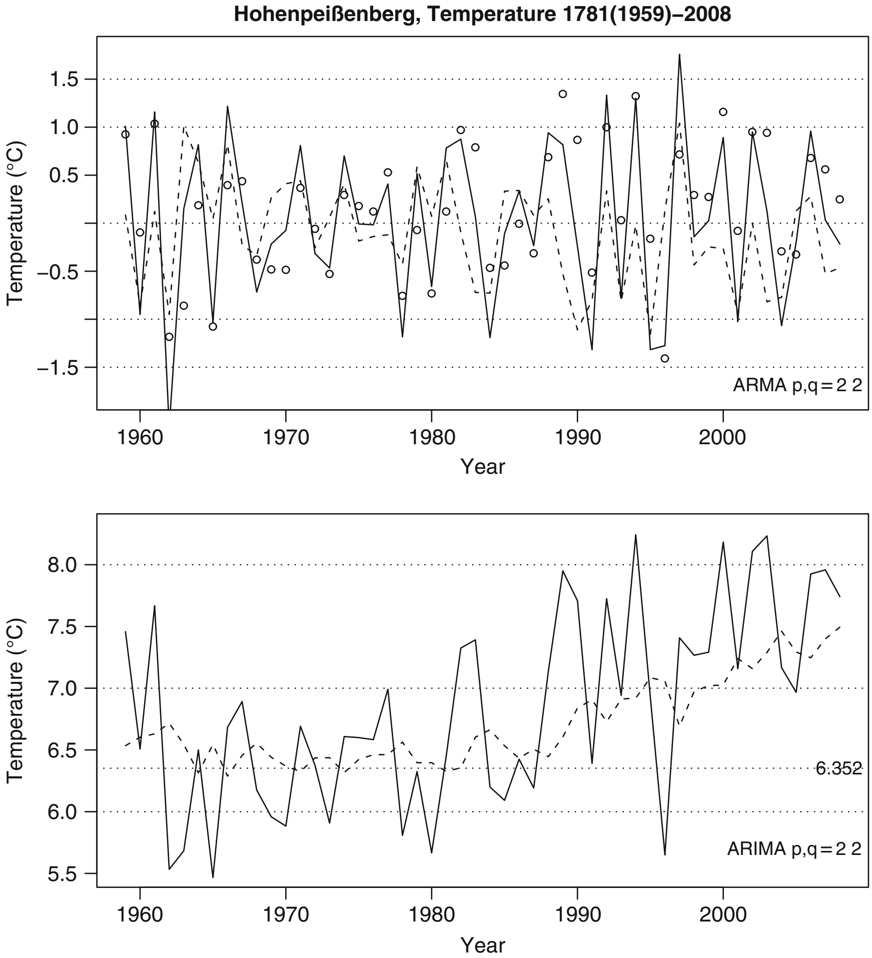

Fig. 6

Hohenpeißenberg, annual temperature means, 1781–2008. *Top:* Differenced time series, with ARMA-predictions (dashed line) and with residual values (as circles o). *Bottom:* Time series of annual temperature means [°C], together with the ARIMA-prediction (dashed line). The last 50 years are shown

time series $e(t), t = 1, \ldots, N$. All values of the auto-correlation function $r_e(h), h = 1, \ldots, 8$, are close to zero, cf. Table 12. The bound for the maximum of $|r_e(h)|, h = 1, \ldots, 8$ (i. e. the simultaneous bound with respect to the hypothesis of a pure random series) is

$$b_8 = u_{1-0.025/8}/\sqrt{N} \quad \text{(significance level 0.05)},$$

and is not exceeded, not even the bound $b_1 = u_{0.975}/\sqrt{N}$ for an individual $|r_e(h)|$. We can assume, that the series $e(t)$ consists of uncorrelated variables.

Table 11

Depth *k* of the left-sided moving averages and resulting goodness-of-fit RootMSQ. The latter is given for the ARIMA-method, too

Depth	RootMSQ		
k	Hohenp.	Karlsr.	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. Table 10)	0.768	0.707	0.726

Table 12

Auto-correlation function $r_e(h)$ up to time lag $h = 8$ of the ARIMA-residuals; annual temperature means. 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	0.014	−0.003	−0.012	0.03	−0.06	−0.04	−0.00	−0.06	0.130	0.181
K	0.005	0.016	−0.022	−0.07	−0.02	−0.01	0.04	−0.03	0.135	0.189
P	0.007	0.009	0.004	−0.05	−0.01	−0.10	0.07	0.01	0.182	0.254

Next we ask, whether the (true) variances of the ARIMA-residuals $e(t)$ are constant over time — or whether periods of (truly) stronger and periods of (truly) weaker oscillation alternate. To this end, we calculate — moving in 5-years time blocks $[t-4, t]$ – the empirical variances $\hat{\sigma}^2(t)$. The roots $\hat{\sigma}(t)$, plotted in ➋ *Fig. 7*, form an oscillating line around the value 0.77 (Hohenpeißenberg), but a definite answer to the above question cannot be given.

5.2.3 GARCH-Modeling

Denoting by $\sigma^2(t) = \mathrm{Var}(e(t))$ the true variance of the error variable $e(t)$, we are going to investigate the structure of the time series $\sigma^2(t)$, $t = 1, 2, \ldots, N$. By means of GARCH-models we can analyze time series with (possibly) varying variances. For this reason, an ARMA(p, q)-type equation for $\sigma^2(t)$ is established, namely

$$\begin{aligned}\sigma^2(t) &= \alpha_p Z^2(t-p) + \cdots + \alpha_2 Z^2(t-2) + \alpha_1 Z^2(t-1) + \alpha_0 \\ &\quad + \beta_q \sigma^2(t-q) + \cdots + \beta_1 \sigma^2(t-1), \quad t = 1, 2, \ldots, \end{aligned} \tag{9}$$

(α's, β's nonnegative). A zero-mean process of uncorrelated variables $Z(t)$ is called a GARCH(p, q)-process ($p, q \geq 0$), if the (conditional) variance of $Z(t)$, given the information up to time $t-1$, equals $\sigma^2(t)$, where $\sigma^2(t)$ fulfills equ. (9) (Kreiß and Neuhaus 2006), (Cryer and Chan 2008).

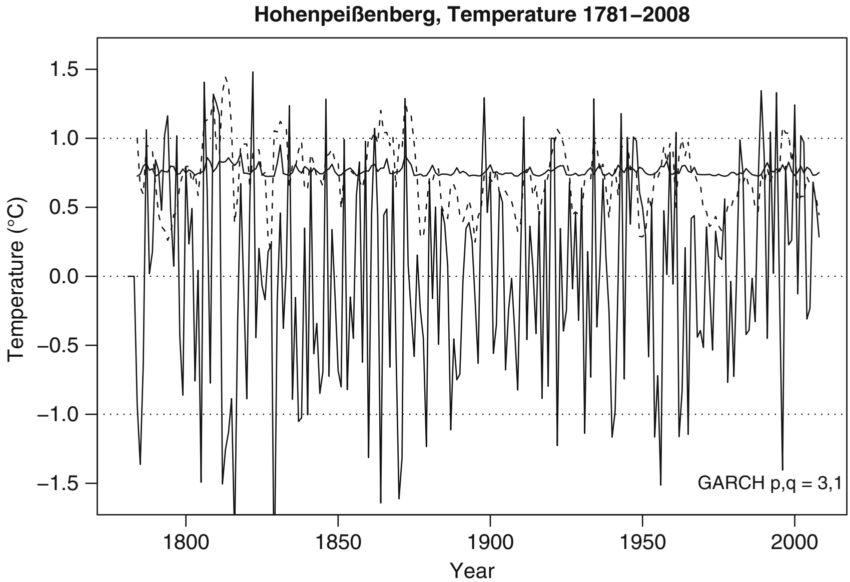

Fig. 7
Hohenpeißenberg, annual temperature means, 1781–2008. Time series of ARIMA-residuals, standard deviation $\hat{\sigma}$ of left-sided moving (5-years) blocks (dashed line), GARCH-prediction for σ (solid line)

Order numbers (p, q) are to be determined (here $p = 3, q = 1$) and $p + q + 1$ coefficients α, β must be estimated. Then we build predictions $\hat{\sigma}^2(t)$ for the series $\sigma^2(t)$ in this way: Let the time point t be fixed. Having observed the preceding $Z(t-1), Z(t-2), \ldots$ (and having already computed $\hat{\sigma}^2(t-1), \hat{\sigma}^2(t-2), \ldots$), then we put $\hat{\sigma}^2(t)$ according to Eq. (9), but with $\sigma^2(t-s)$ replaced by $\hat{\sigma}^2(t-s)$. Here the first q $\hat{\sigma}^2$-values must be predefined, for instance by the empirical variance of the time series Z. We are speaking of the GARCH-prediction for the variance $\sigma^2(t)$.

Now we apply this method to our data and put $Z(t) = e(t)$, the ARIMA-residuals from eq. (8). For the Hohenpeißenberg series these GARCH-predictions reproduce in essence the horizontal line 0.77, see ➋ *Fig. 7*. This means that we can consider $e(t)$ as a series of uncorrelated variables with *constant* variance $\sigma^2(t) = \sigma^2$, i.e., as a white noise process. From there we can state that the differenced sequence $X(t)$ can sufficiently be fitted well by an ARMA-model, since the latter demands a white noise error process.

5.3 Yearly Precipitation Amounts

$Y(t)$ denotes now the precipitation amount in the year t. From Y we pass to the series X by building differences, where $X(t) = Y(t) - Y(t-1)$, $t = 2, \ldots, N$, $X(1) = 0$.

➋ *Table 13* shows that the yearly changes X equal $\approx$ 0 in the mean and have an average deviation (from the mean 0) of $\approx$ 1.5...2.0 [dm]. The auto-correlations $r(1)$ lie in the range

Table 13

Differences X of precipitation amounts [dm] in consecutive years

Station	N	Mean	Stand.Dev.	$r(1)$	$r(2)$	$r(3)$
Hohenp.	130	–0.003	2.085	–0.460	0.003	0.026
Karlsr.	133	0.014	1.900	–0.429	–0.114	0.025
Potsd.	116	0.007	1.417	–0.461	–0.064	–0.040

Table 14

ARIMA-method for the annual precipitation amounts in [dm]. H Hohenpeißenberg, K Karlsruhe, P Potsdam

	Order	Coefficients		Root	ARIMA-prediction	
	p, q	α_i	β_1	MSQ	2006–2008	2009
H	3,1	0.106,0.085,0.010	–0.935	1.673	11.81,11.84,12.12	11.91
K	3,1	0.046,–0.163,–0.062,	–0.943	1.378	7.82, 8.00, 7.59	7.61
P	3,1	–0.128,–0.102,–0.059	–0.989	0.972	6.10, 6.10, 5.86	5.94

–0.4 ... –0.5. An increase of precipitation is immediately followed by a decrease, as a tendency, and vice versa.

We fit an ARMA(p, q)-model to the differenced series X. As order numbers we get $p = 3$, $q = 1$ and therefore the ARMA(3,1)-model

$$X(t) = \alpha_3 X(t-3) + \alpha_2 X(t-2) + \alpha_1 X(t-1) + \beta_1 e(t-1) + e(t)\,. \tag{10}$$

Table 14 presents the estimated coefficients α_i and β_1; the coefficient β_1 is significantly different from zero for all three stations. Further, prognoses for the three years 2006–2008 as well as for the year 2009 were made, each time on the basis of the preceding years.

The predictions lie partly above, partly below the actually observed values, demonstrating their smoothing character, see Fig. 8 (lower plot). The residuals $e(t)$ from the predictions are shown in the upper plot of Fig. 8. The auto-correlations $r_e(h)$, $h = 1, \ldots, 8$, of the residuals were calculated (not reproduced in a table). The bound b_1 for an individual $|r_e(h)|$ is not exceeded and thus — all the more — not the simultaneous bound b_8 (significance level 0.05). The residual series $e(t)$ can be comprehended as a pure random series, confirming the applied ARIMA-model. We abstain here from a GARCH application to the residual series.

6 Model and Prediction: Monthly Data

For the investigation of monthly climate data we confine ourselves to the monthly *temperature* means. We first estimate a trend by the ARIMA-method of Sec. Section 5 (as well as by alternative methods) and model the detrended series as an ARMA-process.

6.1 Trend+ARMA Method

In order to model the monthly temperature means $Y(t)$, we start with

$$Y(t) = m(t) + X(t)\,, \quad t = 1, 2, \ldots, \tag{11}$$

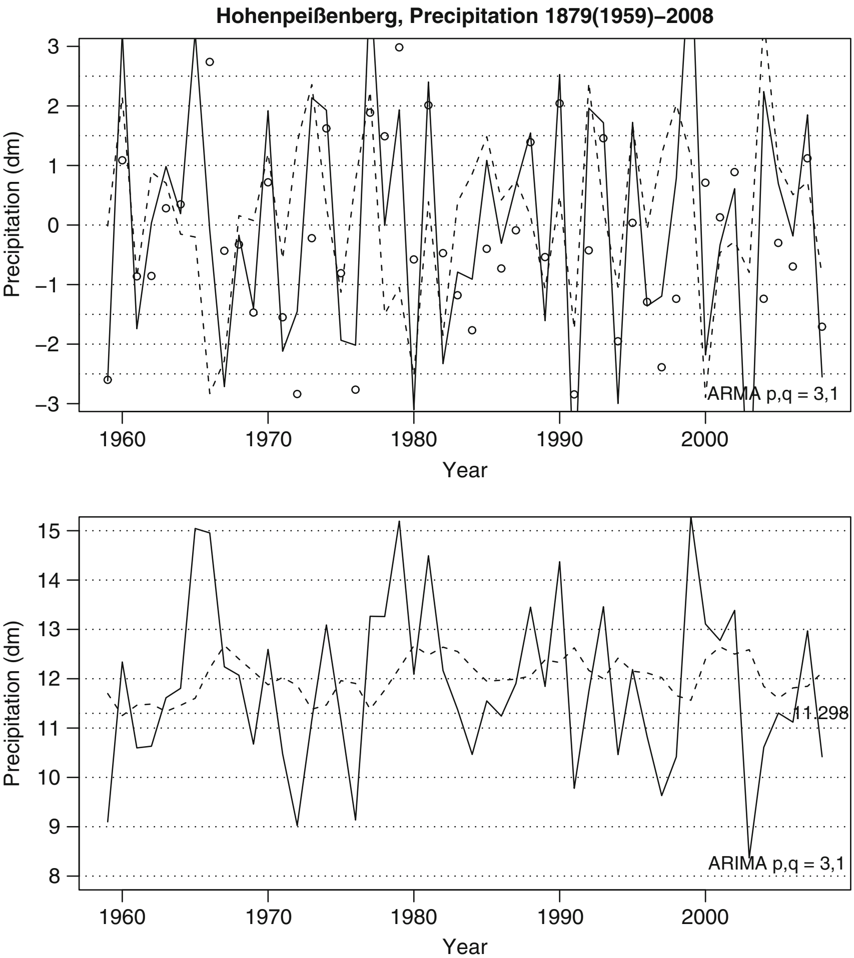

Fig. 8

Hohenpeißenberg, annual precipitation amounts 1879–2008. *Top:* Differenced time series, together with the ARMA-prediction (dashed line) and with the residual values (as circles o). *Bottom:* Time series of annual precipitation amounts in (dm), together with the ARIMA-prediction (dashed line). The last 50 years are shown

where t counts the successive months, $m(t)$ denotes the long-term (yearly) trend, and where $X(t)$ is the remainder series. We estimate the trend by the ARIMA-method of Sec. Section 5: The variable $m(t)$ is the ARIMA-prediction of the yearly temperature mean (with $p = q = 2$ for Hohenpeißenberg and Karlsruhe, and with $p = 3, q = 1$ for Potsdam), called trend(ARIMA), and is the same for all 12 months t of the same year. The detrended series

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

is shown in the upper plot of Fig. 9. We fit an ARMA(p, q)-model to the series $X(t)$, with $p = 3, q = 2$ (that turned out to be sufficient). In Table 15 one can find the estimated

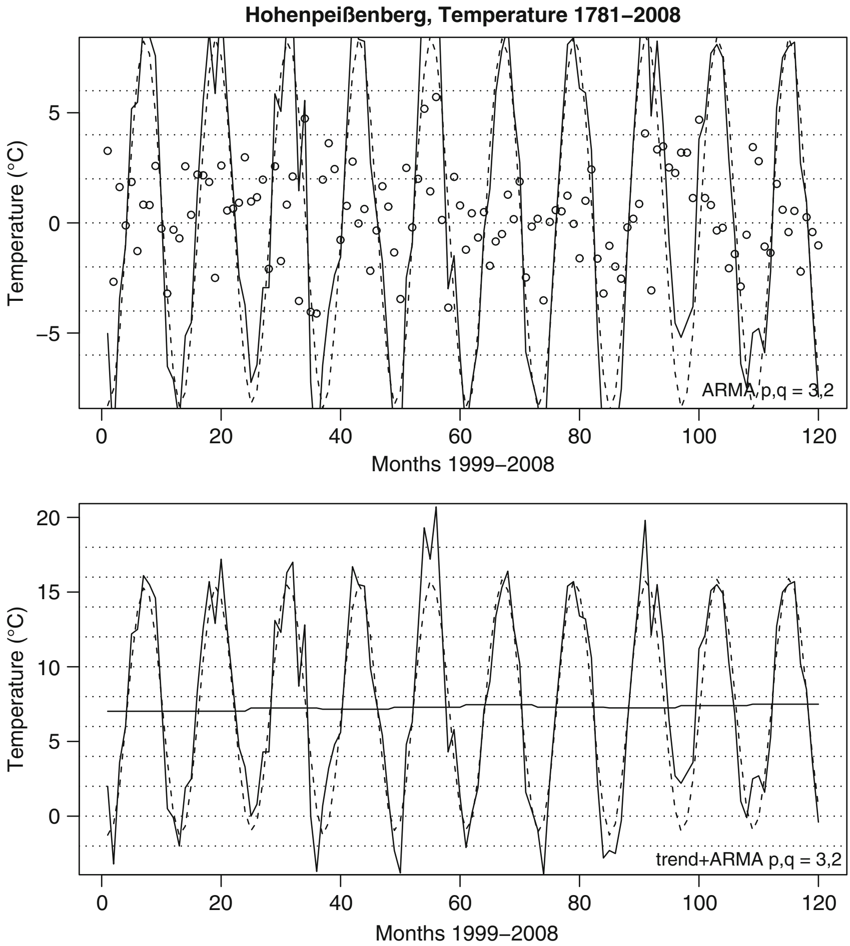

Fig. 9

Hohenpeißenberg, monthly temperature means 1781–2008. *Top:* Detrended time series, together with the ARMA-prediction (dashed line) and with the residual values (as circles o). *Bottom:* Monthly temperature means in [°C], together with the trend (inner solid line) and the trend+ARMA-prediction (dashed line). The last 10 years are shown

coefficients α_i and β_j, (nearly) all of them being significantly different from zero. The ARMA-prediction $\hat{X}(t)$ for $X(t)$ is plotted in the upper part of *Fig. 9*, too. By means of $\hat{X}(t)$ we gain back the original (trend-affected) series, more precisely: the trend(ARIMA)+ARMA-prediction $\hat{Y}(t)$ for $Y(t)$. We put

$$\hat{Y}(t) = m(t) + \hat{X}(t), \quad t = 1, 2, \ldots, \tag{12}$$

compare the lower plot of *Fig. 9*, where the $\hat{Y}(t)$ are portrayed, together with the actual observations $Y(t)$. The goodness-of-fit RootMSQ according to Eq. (6) and the predictions for Oct. 08 to Jan. 09 are presented in *Table 15*, too. With only 4 + 5 parameters these trend(ARIMA)+ARMA-predictions $\hat{Y}(t)$ run close to the actual observed values $Y(t)$. They cannot, however, follow extremely warm summers or cold winters. To give examples, we point

Table 15

Trend(ARIMA) + ARMA-method for the monthly temperature means. H = Hohenpeißenberg, K = Karlsruhe, P = Potsdam

	Order	Coefficients		Root	Prediction	Jan
	p,q	α_i	β_j	MSQ	Oct–Dec 2008	2009
H	3,2	1.758,–1.045,0.026	–1.725,0.992	2.149	8.25, 4.01, 0.63	–0.94
K	3,2	1.691,–0.928,–0.041	–1.691,0.953	1.993	11.31, 6.67, 3.38	2.30
P	3,2	1.810,–1.136,0.078	–1.682,0.951	1.953	9.65, 5.08, 1.80	0.42

Table 16

Depth *k* of the left-sided moving average and resulting goodness-of-fit RootMSQ. The latter is listed for the trend(ARIMA)+ARMA-method and for the ARIMA(lag12)-method, too

Depth	RootMSQ		
k	Hohenp.	Karlsr.	Potsdam
5	2.366	2.137	2.229
10	2.248	2.012	2.079
12	2.227	1.992	2.077
15	2.193	1.971	2.065
20	2.171	1.939	2.058
Trend(ARIMA)+ARMA	2.149	1.993	1.953
ARIMA(lag12)	2.544	2.301	2.318

to the "record summer" 2003 (in *Fig. 9* around the month no. 55) or to the relatively cold January 2009. For the latter, compare the predictions in the last column of *Table 15* with the actual observed values –2.7, –1.3, –2.1 [$^\circ C$], in Hohenpeißenberg, Karlsruhe, and Potsdam, respectively.

6.2 Comparisons with Moving Averages and with Lag-12 Differences

On the basis of approach (11) we can alternatively choose the method of left-sided moving averages, applied for estimating the (yearly) trend $m(t)$ as well as for predicting the detrended series $X(t)$. As trend estimation $m(t)$ we take the average of the preceding observations $Y(t-1), Y(t-2), \ldots, Y(t-k*12)$. The depth number k indicates the number of the employed years. As prediction $\hat{X}(t)$ for the variable $X(t)$ we take the average of the preceding detrended observations

$$Y(t-12)-m(t-12), Y(t-24)-m(t-24), \ldots, Y(t-k*12)-m(t-k*12).$$

The integer k indicates here the number of the employed months. Again according to Eq. (12) and (6) we compute the goodness of this prediction method. *Table 16* shows that for no depth smaller than $k = 21$ (Karlsruhe $k = 12$) the RootMSQ values of the trend(ARIMA) + ARMA-method are attained. Recall that in the latter method only $4 + 5 = 9$ parameters are involved.

Another alternative procedure resembles the ARIMA-method of Sec. Section 5. Instead of using differences $Y(t) - Y(t-1)$ of two consecutive variables (lag-1 differences), however, we

Table 17

Auto-correlation function $r_e(h)$ up to time lag $h = 8$ of the trend(ARIMA)+ARMA-residuals; monthly temperature means. 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	0.102	0.029	0.013	0.02	0.01	0.04	0.03	0.03	0.037	0.052
K	0.204	0.051	0.003	0.02	0.05	0.07	0.05	0.02	0.039	0.054
P	0.152	0.063	0.012	0.01	0.00	0.00	0.04	0.04	0.052	0.073

form lag-12 differences, that are differences

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

of two observations being separated by 12 months. We fit an AR(12)-model to this differenced process $X(t)$, and determine the goodness-of-fit by Eq. (5) or — equivalently — Eq. (6). We will use the short-hand notation ARIMA(lag12). *Table 16* shows, that this procedure is inferior to the method trend(ARIMA) + ARMA and to the method of moving averages as well.

6.3 Residual Analysis

Denoting by $M = N * 12$ the total number of months and by $\hat{Y}(t)$ the trend(ARIMA)+ARMA-prediction for $Y(t), t = 1, \ldots, M$, we obtain by

$$e(t) = Y(t) - \hat{Y}(t), \quad t = 1, \ldots, M,$$

the residuals from the prediction; compare the upper plot in *Fig. 9*. Which structure has this residual time series $e(t), t = 1, \ldots, M$? Its auto-correlation function $r_e(h), h = 2, \ldots, 8$, consists of values more or less near zero, cf. *Table 17*. It is particularly the auto-correlation $r_e(1)$ of first order (i.e., the correlation between $e(t), e(t+1)$ of two immediately succeeding months), which turns out to be relatively large. The simultaneous bound $b_8 = u_{1-0.025/8}/\sqrt{M}$ is exceeded by all three $r_e(1)$ values (at the significance level 0.05). Our chosen prediction method trend(ARIMA)+ARMA leaves behind residuals, which are correlated too strong (at least of order one), and thus do not fulfill the demand on residual variables $e(t)$. A similar statement is to be made with respect to the method of moving averages.

7 Conclusions

First we state that the separation of the trend/saison component on one side and the auto-correlation structure on the other side is crucial in our analysis. To handle the latter, ARMA-type modeling the detrended series or the differenced series (with subsequent integration: ARIMA) is performed and turns out to work quite satisfactory.

The correlation and prediction analysis reveals that *precipitation* is more irregular and closer to a random phenomenon than *temperature* is; see also (von Storch and Navarra 1993). This statement is also confirmed by *Table 18*, where the coefficients of correlation $r(Y, Y1)$ and of multiple correlation $r(Y, (Y1 \ldots Y6))$ are presented, with Y = temperature or Y = precipitation. By Y1 to Y6 we denote lagged variables, from lag = 1 year to lag = 6 years.

Table 18

Correlation analysis for annual climate data. Temperature (Temp.) and precipitation (Prec.) with lagged variables

Station	$r(Y, Y1)$		$r(Y, (Y1 \ldots Y6))$		$r(Y, (Y1 \ldots Y6, Z1 \ldots Z6))$
	Y = Temp.	Y = Prec.	Y = Temp.	Y = Prec.	Y = Prec.
Hohenp.	0.289	0.273	0.370	0.321	0.338
Karlsr.	0.332	0.009	0.455	0.136	0.230
Potsd.	0.358	−0.079	0.402	0.277	0.289

In Karlsruhe and in Potsdam, the correlations between temperature variables are distinctly larger than those between precipitation variables. If precipitation (Y = Prec.) is correlated with the set $(Y1 \ldots Y6, Z1 \ldots Z6)$, comprising the lagged precipitation variables $Y1 \ldots Y6$ *and* the lagged temperature variables $Z1 \ldots Z6$, the coefficient remains — nevertheless — far below that of temperature (Y = Temp.).

The exception is Hohenpeißenberg, as already mentioned in Sects. 2 and 4, where the level of correlation for precipitation is closer to that for temperature than it is in the other two stations.

For *predicting* a climate variable Y_t at time t, observations only up to time $t-1$ are allowed: Y_{t-1}, Y_{t-2},... Therefore polynomials, drawn over the whole time interval $t = 1, \ldots, N$, are not qualified as a (yearly) trend component in Sects. 5 and 6. Our numerical procedures, unfortunately, violate this rule in one aspect: The coefficients of the ARMA-models, the α's and β's, were estimated from the whole series. Here in future work the time-consuming amendment should be introduced and the coefficients should be calculated for each time point t anew.

For predicting *monthly* climate variables one has to tune the estimation of the trend- and of the seasonal component. Here further procedures should be tested, since the residual series in Sect. 6 are not sufficiently close to a pure random series.

Winter data alone are (only) a weak indicator for the general climate development. This can also be documented by spectral analysis methods (Pruscha 1986).

Appendix: Excerpt from Hohenpeißenberg Data

The complete data sets can be found under www.math.lmu.de/~pruscha/ keyword: Climate series.

Monthly- [yearly-] **temperature** means in 1/10°C [1/100°C]. A time series plot of the yearly and the winter means can be found in *Fig. 1* and further analysis of these data in (Pruscha 2005).

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Tyear
1781	−18	−10	24	87	122	145	154	166	126	44	15	12	723
1782	−10	−54	0	38	94	156	176	144	108	36	−28	−23	531
1783	7	3	−4	64	108	131	163	144	118	82	12	−24	670
1784	−53	−46	0	21	128	132	152	136	143	23	12	−47	501
1785	6	−65	−60	13	91	117	131	131	141	60	23	−19	474
1786	−1	−30	−5	71	91	139	118	123	94	35	−5	−10	517
1787	−35	8	33	38	72	141	143	161	120	93	18	39	693

1788	−19	21	23	56	116	148	176	140	135	56	−6	−105	618
1789	−10	−5	−34	74	131	110	147	144	110	65	4	12	623
1790	−6	9	16	38	120	144	135	157	109	85	28	−11	687
1791	9	−18	20	89	97	130	148	164	110	72	14	10	704
.....													.....
1996	−16	−32	−12	61	98	141	140	139	80	73	27	−21	565
1997	−10	27	45	37	110	128	139	170	133	62	37	11	741
1998	5	34	18	64	115	149	152	161	109	73	−10	2	727
1999	20	−32	36	60	122	125	161	155	146	78	5	−1	729
2000	−20	19	25	83	126	157	129	172	126	86	46	33	818
2001	−0	8	43	43	131	123	163	170	87	128	0	−37	716
2002	7	32	48	56	113	167	155	154	99	76	54	12	811
2003	−23	−38	48	63	127	193	172	207	125	43	58	13	823
2004	−21	2	19	71	90	134	153	164	126	102	16	4	717
2005	−9	−39	23	70	113	154	157	134	132	106	23	−28	697
2006	−23	−25	−3	61	109	150	198	121	155	117	64	27	793
2007	22	29	36	112	121	151	155	149	103	68	10	−1	796
2008	25	27	16	53	127	150	155	157	102	85	36	−4	774

Monthly and yearly **precipitation** amounts in 1/10 mm height. A time series plot of yearly and winter amounts can be found in ➋ *Fig. 2*.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Pyear
1879	254	619	272	1071	1039	1009	1473	1457	1645	685	861	393	10778
1880	385	253	315	967	1203	1991	1870	1212	997	1784	473	907	12357
1881	188	232	448	809	1332	1404	885	1490	1173	828	263	202	9254
1882	186	88	400	696	952	1565	1802	1314	1275	788	896	501	10463
1883	310	127	421	504	1179	2096	2020	835	1152	526	614	684	10468
1884	649	202	412	1164	440	1846	1957	1130	534	1360	268	432	10394
1885	100	277	643	285	1185	1437	1644	925	1366	761	462	968	10053
1886	244	171	436	828	625	2214	972	2288	382	430	436	682	9708
1887	145	125	888	291	1712	435	1550	718	699	647	637	952	8799
1888	431	667	494	1438	690	1575	1288	1733	1887	607	211	55	11076
1889	174	1084	502	701	1151	1738	1408	1019	1610	678	696	260	11021
...													...
1996	138	314	588	597	1353	1018	1532	1849	1018	1082	920	417	10826
1997	18	585	673	934	425	1724	2404	457	319	947	189	956	9631
1998	398	280	1012	474	569	1389	1174	666	1632	1496	923	404	10417
1999	598	1187	474	927	3507	1625	1560	1194	1357	423	1341	1097	15290
2000	300	802	1514	549	1537	1422	1743	2006	1413	997	551	275	13109
2001	681	664	1162	1107	628	2183	967	1626	1621	345	1018	773	12775
2002	103	685	992	723	952	1541	1509	1908	2203	821	1342	606	13385
2003	670	513	327	265	817	800	1504	815	450	1352	485	371	8369
2004	1127	431	677	547	1009	1573	1712	857	944	769	509	455	10610
2005	551	740	431	1191	1310	693	1840	2522	593	454	432	547	11304
2006	400	395	1025	1601	1123	1524	293	2453	690	679	504	433	11120
2007	608	520	488	173	2377	1079	2117	2065	1627	421	726	769	12970
2008	414	141	694	1546	942	879	1700	1709	701	735	510	448	10419

References

Attmannspacher W (1981) 200 Jahre meteorologische Beobachtungen auf dem Hohenpeißenberg 1781–1980. Bericht Nr. 155 des DWD, Offenbach/m

Brockwell PJ, Davis RA (2006) Time series: theory and methods, 2nd ed. Springer, New York

Cryer JD, Chan KS (2008) Time series analysis, 2nd ed. Springer, New York

Fiedler F (1971) Klimawerte zur Temperatur- und Windschichtung. Wissensch. Mitteilung Nr. 18 des Meteor. Instituts der Univ. München

Grebe H (1957) Temperaturverhältnisse des Observatoriums Hohenpeißenberg. Bericht Nr. 36 des DWD, Offenbach/m

Kreiß JP , Neuhaus G (2006) Einführung in die Zeitreihenanalyse. Springer, Berlin

Malberg H (2003) Bauernregeln, 4th ed. Springer, Berlin

Pruscha H (1986) A note on time series analysis of yearly temperature data. J Royal Statist Soc A 149:174–185

Pruscha H (2005) Statistisches Methodenbuch. Springer, Berlin

Schönwiese CD (1979) Klimaschwankungen. Springer, Berlin

Schönwiese CD, et al (1993) Klimatrend-Atlas. Bericht Nr. 20 des Zentrums f. Umweltforschung, Frankfurt/m

von Storch H, Navarra A (eds) (1993) Analysis of climate variability. Springer, Berlin

von Storch H, Zwiers FW (1999) Statistical analysis in climate research. Cambridge University Press, Cambridge