ELSEVIER

Agricultural and Forest Meteorology 96 (1999) 131–144

AGRICULTURAL AND FOREST METEOROLOGY

www.elsevier.com/locate/agrformet

Forest climatology: estimation of missing values for Bavaria, Germany

Youlong Xia*, Peter Fabian, Andreas Stohl, Martin Winterhalter

Lehrstuhl für Bioklimatologie und Immissionsforschung, Ludwig-Maximilians Universität München, Am Hochanger 13, 85354 Freising, Germany

Received 25 September 1998; received in revised form 11 March 1999; accepted 23 March 1999

Abstract

Estimation of missing values in climatological time series is an important task. In order to find an appropriate method, we examined six methods for estimating missing climatological data (daily maximum temperature, minimum temperature, air temperature, water vapour pressure, wind speed and precipitation) for different time scales at six German weather stations and three Bavarian forest climate stations. The multiple regression analysis (using the five closest weather stations) with least absolute deviations criteria (REG) predominantly gave the best estimation for daily, weekly, biweekly, and monthly maximum temperature, minimum temperature, mean temperature, water vapour pressure, wind speed, under different topographical conditions (valley, alpine foothills and mountain sites). The six methods gave similar estimates for the averaged precipitation amount. The mean absolute errors (MAE) of estimating climatological data using different techniques are of similar magnitude at the weather stations, but they are significantly different at the forest climate stations. For the same climatological variable (i.e., air temperature) for different time scales, mean absolute errors of estimated data are larger for shorter time scales (e.g., a day) than for longer ones (e.g., a month). For the different climatological variables, the most accurately estimated variables are maximum temperatures, mean temperatures and water vapour pressure, followed by minimum temperature and wind speed. The poorest results were obtained for precipitation data.

Keywords: Weather stations; Forest climate stations; Mean absolute errors (MAE); Missing data

1. Introduction

Since most weather service stations are located in suburbs and valleys, not close to forest areas, climatological data obtained from such stations cannot represent the climate of a forest region. Therefore, in order to monitor and study forest ecosystems, the Bavarian State Institute of Forestry has set up 22 twin forest climate stations in typical forest areas throughout Bavaria since 1991, monitoring climatological data both inside and outside the particular stands. These are installed in the middle of great forest regions, in which meteorological measurements can well represent the climate of these forest areas. Forest climate data are very useful not only for simulation of tree growth (Kimmins et al., 1990), study of forest water balance, damage of spring and autumn frost on trees (Cannell, 1984; Cannell et al., 1985), phenology (Kramer et al., 1996) but also for forest entomology (Russo et al., 1993). Because some forest climate stations are located in the mountains or in remote

* Corresponding author. Fax: +49-8161714753
E-mail address: xia@arbwiss.arwi.forst.uni.muenchen.de (Y. Xia)

0168-1923/99/$ – see front matter
PII: S0168-1923(99)00056-8

forest regions, observed climatological data are sometimes missing at these stations for maintenance records. Missing forest climatological data are serious hindrance to the use of climate-dependent models and forest ecosystem studies. Hence estimation of the missing forest climate data becomes more and more important, and the establishment of complete data bases at forest climate stations is an urgent task.

Estimation of missing climatological data is an important task for meteorologists, hydrologists and environment protection workers all over the world. It is particularly important in mountain and forest regions where meteorological stations are scarce, and the observed climatological data are strongly influenced by topography and the forest microclimate. The techniques of estimating missing climatological data can be grouped under empirical methods, statistical methods and function fitting (Thiebaux and Pedder, 1987). In empirical approaches, array values are computed from a distance-weighted sum of the data and the weighting function is usually predetermined. They include the simple arithmetic averaging, inverse distance interpolation (Cressman, 1959; Shepard, 1968; Barnes, 1973; Willmott et al., 1985; Willmott et al., 1991; Rudolf et al., 1992; Hubbard, 1994; Sokol and Stekl, 1994; Willmott et al., 1994; Palomino and Martin, 1995; Willmott and Matsuura, 1995; Willmott and Robeson, 1995), and ratio and difference technique (Tabony, 1983; Wallis et al., 1991). In statistical techniques, array values are also computed from a weighted sum of input data, except that weights are based on statistics of spatial covariance of the data. They include multiple regression analysis (Kemp et al., 1983; Tabony, 1983; Kim et al., 1984; Wigley et al., 1990; Ward and Folland, 1991; Young, 1992; Young and Gall, 1992; Degaetano et al., 1995; Eischeid et al., 1995), multiple discriminant analysis (Young, 1992), principal component analysis and cluster analysis (Klink and Willmott, 1989; Huth and Nemesova, 1995), kriging technique (Hevesi et al., 1992a, b; Ishida and Kawashima, 1993; Saborowski and Stock, 1994) and optimal interpolation (Bussieres and Hogg, 1989). In function fitting, data are fitted as a function (e.g., thin-plate spline). Now thin-plate splines are widely used to interpolate the climatological data (Hutchinson, 1995; Luo et al., 1998). No matter which kind of methods is used, it is dependent on the design of the measurement network and the spatial coherence of the parameter being interpolated (Rudolf et al., 1992; Young, 1992; Bardossy and Muster, 1993). Our weather network is located in the south of Bavaria which features a strongly variable topography. Due to the particular location and based on previous experience (BayForKlim, 1996), some complicated statistical methods (i.e., kriging, optimal interpolation, principal component and cluster analysis) and some empirical techniques (Barnes, Cressman) were not used. Eischeid et al. (1995) discussed the estimation accuracy of six methods which include not only empirical methods but also statistical methods for monthly mean temperature and monthly precipitation. They indicated that the multiple regression analysis is best. Kemp et al. (1983) and Degaetano et al. (1995), estimating daily minimum and maximum temperature, came to the same conclusion. From the literature review and our previous experience, we chose six techniques: simple arithmetic averaging, inverse distance method (Hubbard, 1994), normal ratio method (Young, 1992), single best estimator (Eischeid et al., 1995), multiple regression analysis (Degaetano et al., 1995), the traditional method of the UK meteorological office (constant ratio or constant difference (Tabony, 1983) and closest station method (Wallis et al., 1991).

This paper discusses the methods of estimating missing climatological data at three forest climate stations and six weather stations, and compares the accuracies of these methods at two kinds of climate stations for the different time scales.

2. Data and methods

2.1. Data

Our analysis is based on two climatological data sets. The first is a climatological data set of the German Weather Service, comprising daily maximum temperature (T_{max}), minimum temperature (T_{min}), mean air temperature (T_m), relative humidity (RH), wind speed (u) and precipitation (P) at 44 weather stations, from 47.83 to 49.50°N latitude and 10.70 to 13.33°E longitude (Table 1). Due to the short observation record at forest climate stations (see below), analyses were restricted to the period 1991–1995 with complete data in the region. Despite this limited

Table 1
Description of the 44 German weather stations used in this study (six weather stations with missing data during 1991 and 1995 are written in bold. Data are missing at these stations for different periods, (at Mainburg from 1 July 1993 to 30 April 1994, at Leiblfing from 1 September 1995 to 30 September 1995, at Raisting from 1 January 1992 to 30 June 1992, at Pommelsbrunn from 1 July 1993 to 31 December 1995, at Karshuld from 1 April 1994 to 31 December 1995, and at Wasserburg from 1 January 1994 to 31 December 1995)

Station name	Station number	Latitude (°N)	Longitude (°E)	Start			End			Elevation (m)
				Day	Month	Year	Day	Month	Year	
Alderbach	4519	48.60	13.08	1	1	1991	31	12	95	325
Altomuenster	4116	48.38	11.25	1	1	1991	31	12	95	502
Amberg-Unterammersricht	4075	49.47	11.87	1	1	1991	31	12	95	386
Attenkam	4166	47.72	11.37	1	1	1991	31	12	95	655
Augsburg-Muehlhausen	4128	48.43	10.93	1	1	1991	31	12	95	461
Cham	4488	49.22	12.67	1	1	1991	31	12	95	396
Ebersberg	4171	48.08	11.95	1	1	1991	31	12	95	572
Eichstaett	4108	48.90	11.17	1	1	1991	31	12	95	397
Falkenberg	4524	48.47	12.73	1	1	1991	31	12	95	490
Gr. Arber	4489	49.12	13.13	1	1	1991	31	12	95	1437
Hoellenstein-Kraftwerk	4494	49.13	12.87	1	1	1991	31	12	95	403
Holzkirchen	4172	47.88	11.70	1	1	1991	31	12	95	685
Kaisheim-Neuhof	4107	48.77	10.78	1	1	1991	31	12	95	516
Karlshuld	4114	48.68	11.30	1	1	1991	31	03	94	374
Kaufering	4126	48.10	10.87	1	1	1991	31	12	95	585
Koesching	4109	48.83	11.48	1	1	1991	31	12	95	417
Leiblfing	4500	48.77	12.52	1	1	1991	31	12	95	365
Mainburg	4185	48.65	11.78	1	1	1991	31	12	95	431
Maisach-Gernlinden	4131	48.22	11.30	1	1	1991	31	12	95	509
Mallersdorf	4501	48.78	12.25	1	1	1991	31	12	95	410
Metten	4509	48.85	12.92	1	1	1991	31	12	95	313
Muehldorf	4528	48.28	12.50	1	1	1991	31	12	95	405
Muenchen-Nymphenburg	4121	48.17	11.50	1	1	1991	31	12	95	515
Neutraubling	4563	48.98	12.22	1	1	1991	31	12	95	332
Nuernburg-Fischbach	4078	49.42	11.20	1	1	1991	31	12	95	348
Nuernberg-Kra.	4081	49.50	11.05	1	1	1991	31	12	95	314
Oberschleissheim	4187	48.25	11.55	1	1	1991	31	12	95	484
Oberviechtach	4484	49.45	12.43	1	1	1991	31	12	95	595
Parsberg	4079	49.17	11.72	1	1	1991	31	12	95	516
Pfofeld-Langlau	4067	49.12	10.87	1	1	1991	31	12	95	435
Pommelsbrunn	4076	49.50	11.52	1	1	1991	30	6	93	368
Raisting	4164	47.92	11.10	1	1	1991	31	12	95	553
Regensburg	4499	49.05	12.10	1	1	1991	31	12	95	366
Rosenheim	4544	47.88	12.13	1	1	1991	31	12	95	444
Roth	4084	49.25	11.10	1	1	1991	31	12	95	340
Saldenburg-Entschenreuth	4512	48.78	13.32	1	1	1991	31	12	95	457
Simbach	4521	48.27	13.03	1	1	1991	31	12	95	360
Schwandorf/Opf	4483	49.33	12.12	1	1	1991	31	12	95	372
Traunstein-Axdorf	4530	47.85	12.62	1	1	1991	31	12	95	635
Trostberg	4531	48.02	12.55	1	1	1991	31	12	95	487
Wasserburg	4529	48.05	12.22	1	1	1991	31	12	93	443
Weihenstephan	4117	48.40	11.70	1	1	1991	31	12	95	470
Weissenburg	4083	49.02	10.97	1	1	1991	31	12	95	422
Zwieselberg	4493	49.00	13.22	1	1	1991	31	12	95	615

Table 2
Locations of the three Bavarian forest climatic stations used in this study

Station name	Latitude (°N)	Longitude (°E)	Start			End			Elevation (m)	Length of record (days)
			Day	Month	Year	Day	Month	Year		
Ebersberg (EBE)	48.13	11.92	1	3	1991	31	12	1995	538	1765
Mittelfels (MIT)	48.98	12.88	1	4	1991	31	12	1995	1025	1735
Riedenburg (RIE)	48.92	11.75	1	2	1991	31	12	1995	460	1795

period, data can be used to study the methods for estimating missing values (Kemp et al., 1983; Tabony, 1983). Missing climate data existed at six weather stations (in bold in Table 1) due to relocation of a station or to interruptions in observations. A total of 71214 daily climatological data were available. Water vapour pressure (e) computed from air temperature and relative humidity was considered as a basic estimated variable.

The second data set is a forest climate data set comprising of three forest climate stations. These forest climate stations are located in three typical meteorological zones and represent different terrain and elevation conditions: low elevation mountain valley (Station Riedenburg), alpine foothills (Station Ebersberg) and a high elevation ridge (Station Mitterfels). Data of this network, established and maintained by the Bavarian State Institute of Forestry, are taken with a time resolution of 15 min (Table 2). The stations are located in forest clearances (mainly forest meadows) with a diameter of at least four times the height of old trees (that is about 80–150 m). The measuring equipment is installed in the center of the clearing, in a 12 m $\times$ 12 m fenced area (for details see Preuhsler et al. (1995)). For our study daily values were computed from 15 min data according to three observation times of the weather station data (Preuhsler et al., 1995). It turned out that 15–20% of climatological data from the forest climate stations are missing in any season of a year.

2.2. Methods

The best methods of estimating missing data will, in general, depend upon the statistical properties of the data. In climatology, the two most important factors are the inter-correlations in the station network and the seasonal variations in the relations between the stations. The following six methods for estimating missing data are used in this study.

2.2.1. Simple arithmetic averging (AA)

This is the simplest method which is commonly used to fill the missing meteorological data in meteorology and climatology. The missing data are obtained by arithmetically averaging data of the five closest weather stations around a station.

2.2.2. Inverse distance interpolation (ID)

The inverse distance method is used to estimate missing data because of its simplicity (Hubbard, 1994).

$$V_0 = \frac{\sum_{i=1}^{n}(V_i/d_i)}{\sum_{i=1}^{n}(1/d_i)} \qquad (1)$$

where V_0 is the estimated value of the missing data, V_i is the value of the ith nearest weather station, and d_i is the distance between the station of missing data and the ith nearest weather station. Tronci et al. (1986) suggested that a proper choice of the influence radius is much more important than the specification of the weighting function in determining the quality of the approximation. According to an extensive literature review and the experience of Tronci et al. (1986) the influence radius was chosen as 100 km in our study. Thus, weather stations outside of 100 km will not be used.

2.2.3. Normal ratio method (NR)

The normal ratio method of spatial interpolation was first proposed by Paulhus and Kohler (1952), and the method was modified by Young (1992). The estimated data are considered as a combination of variables with different weights, i.e., $V_0 = \sum_{i=1}^{n} W_i V_i / \sum_{i=1}^{n} W_i$ where W_i is weight of the ith nearest weather station and V_i is the observational data of the ith nearest weather station. Weights for the surrounding stations used in the estimation algorithm are calculated according to:

$$W_i = \left[r_i^2 \left(\frac{n_i - 2}{1 - r_i^2} \right) \right] \quad (2)$$

where r_i is the correlation coefficient between the target station and the ith surrounding station, n_i is the number of points used to derive the correlation coefficient, and W_i is the resultant weight.

2.2.4. *Single best estimator (SIB)*

The single best estimator is simple and analogous to using the closest neighboring station as an estimate for a target station. Target station conditions are estimated using data from the neighboring station that has the highest positive correlation with the target station.

2.2.5. *Multiple regression analysis, least absolute deviations criteria (REG)*

The multiple regression analysis is a traditional interpolation approach. Kemp et al. (1983), Tabony (1983), Young (1992) and Eischeid et al. (1995) indicated many advantages of multiple regression analysis in the data interpolation and estimation of missing data. Missing data (V_0) were estimated as

$$V_0 = a_0 + \sum_{i=1}^{n} (a_i V_i) \quad (3)$$

where a_0, a_1, ..., a_n are regression coefficients. V_i is the value of ith weather station.

2.2.6. *UK traditional method (UK)*

The method traditionally used by the UK Meteorological Office to estimate missing temperature and sunshine data was based on comparisons with a single neighbouring station. For temperature, a constant difference between stations was assumed. Thus, if the January temperature at Station A was 0.1°C above that at Station B averaged over a period of overlapping records, then 0.1°C was added to the values at Station B to give the estimated values at Station A. For sunshine, a constant ratio between stations was assumed. Thus, if the July sunshine at Station A has been 1% less than at Station B during a period of overlapping records, then 1% was subtracted from the values at Station B to provide estimated values at Station A. In our study, a constant difference between stations is used for maximum temperature, minimum temperature, air temperature, water vapour pressure and wind speed. The value of Station B is an arithmetic average value resulting from the five closest weather stations.

2.2.7. *Closest station method (CSM)*

The closest station was identified, the missing data were estimated from the data of the closest station, and were adjusted by the ratio of the long-term means for that month. The method will be used for precipitation only. For the other five variables, CSM gives similar results as SIB according to our experience.

Estimate results were found to be best for (ranging from 3 to 15) maximum number of data points 5. This is in good agreement with the results of Van der Voet et al. (1994). The five closest weather stations were used for all methods. In order to compare the accuracy of the results estimated by six techniques, mean absolute errors (MAE) are used as a criterion (Nalder and Wein, 1998). Mean absolute errors provide a measure of how far the estimate can be in error, ignoring sign. In order to satisfy $T_{min} < T_m < T_{max}$ and relative humidity RH (T_m, e) ≤ 1.0, finally a quality control was used for maximum temperature, minimum temperature, mean temperature, water vapour pressure and relative humidity (Meek and Hatfield, 1994; Degaetano et al., 1995).

We used nine stations, six weather stations (written in bold in Table 1) and three forest climate stations, as reference stations to validate our method. These stations were selected, because they actually had missing data, and it was interesting to see how good a method to estimate missing data would work at these stations.

To check our method, the data for the year 1992 were removed from the dataset, and the data for the remaining years were used to establish regression relationships based on the data from the other German weather stations. These relationships were then used to estimate the data for the year 1992. Comparing these estimated data with the measured data allowed to evaluate the performance of the different interpolation methods. The year 1992 was chosen because the record was relatively complete for that year, yielding a sufficiently large database for our comparisons. If there were actual missing data for any day (week, biweek, month), that day (week, biweek, month) was not used for the comparisons.

3. Results and discussion

In order to facilitate discussion of the results, the averaged results of six weather stations (in bold of Table 1) are given. For the three forest climate stations, the results are discussed individually. To reduce random effects in the various estimation methods due to the use of a small sample, and to consider monthly (seasonal) variability as much as possible, 12 months are used for daily climatological data pooled over 4 years (Kemp et al., 1983); Four seasons (winter, spring, summer, and autumn) are used for weekly mean climatological data pooled over 4 years; Two seasons, winter half year (from October to April) and summer half year (from May to September, also called vegetation period) are used for biweekly mean climatological data pooled over 4 years. Due to the limited data series, monthly mean climatological data are used without considering seasonal variation. Since it is the purpose of our paper to discuss the overall quality of the estimated missing climatological data, 12 month averaged MAE (hereafter referred to as MAE in figures) are presented for daily climate data, four season averaged MAE are given for weekly mean climatological data, two season averaged MAE are given for biweekly mean climatological data. Finally, a seasonal variation of MAE between data observed and estimated by REG will be shown for daily maximum temperature, minimum temperature, air temperature and water vapour pressure at the weather stations and three forest climate stations.

3.1. German weather stations (GWS)

MAE between the estimated and observed climatological data at the weather stations are shown in Fig. 1. The REG and UK methods give smaller MAE than the other ones for all climatological variables except precipitation, even though the difference between the different methods is not large (Fig. 1). For precipitation, MAE are similar for all six estimation techniques. The MAE of maximum temperature, minimum temperature and mean temperature are smaller than 1°C for daily, weekly, biweekly and monthly time scales, although MAE of daily temperature are a little larger for the same method. The difference of MAE between the best method (i.e., REG) and the worst technique is 0.2–0.4°C, which depends upon the different climatological variables and time scales (Fig. 1(a–c)). The MAE of water vapour pressure are lower than 0.5 hPa for all six estimating methods. The largest difference of MAE between the six methods for the same time scale (i.e., day) is about 0.2 hPa (Fig. 1(d)). In general, the MAE of wind speed are smaller than 1.0 m/s. MAE generated by REG are about 0.5 m/s for all time scales, although monthly estimate is better (Fig. 1(e)). For precipitation, in most conditions (except daily precipitation estimated by SIB and CSM), MAE are about 0.5 mm/day (Fig. 1(f)). These results show that the climatological data at weather stations estimated by six estimating techniques are almost of the same accuracy. The REG and UK can give more accurate estimates, but they were not significantly different than the other methods. Estimates of maximum temperature, minimum temperature, air temperature, water vapour pressure, wind speed and precipitation computed by any one of six methods were relatively accurate for the four time scales.

3.2. Forest climate Station Riedenburg (RIE, Danube valley)

Forest climate Station Riedensburg is located in the Danube valley, its elevation is 460 m. Fig. 2 shows that for maximum temperature, the estimated MAE is smaller than 1.0°C for all methods except SIB. For daily maximum temperature, six techniques give similar MAE, but for the other time scales, MAE estimated by REG and UK is 0.5°C smaller than that estimated by the other methods (Fig. 2(a)). UK and REG generate the smallest MAE for minimum temperature (Fig. 2(b)). Except REG and UK, MAE produced by the other four estimating methods was 2°C for daily, weekly, biweekly and monthly mean minimum temperature. The REG and UK generate smaller errors (about 0.5°C) except for daily minimum temperature (about 1°C). For air temperature (water vapour pressure), similar results can be given except 1.5°C (0.5 hPa) MAE is estimated by AA, ID, NR and SIB (Fig. 2(c) and(d)). Mean absolute error of wind speed is below 1.5 m/s (Fig. 2(e)) except for SIB. The REG and UK give more accurate estimates

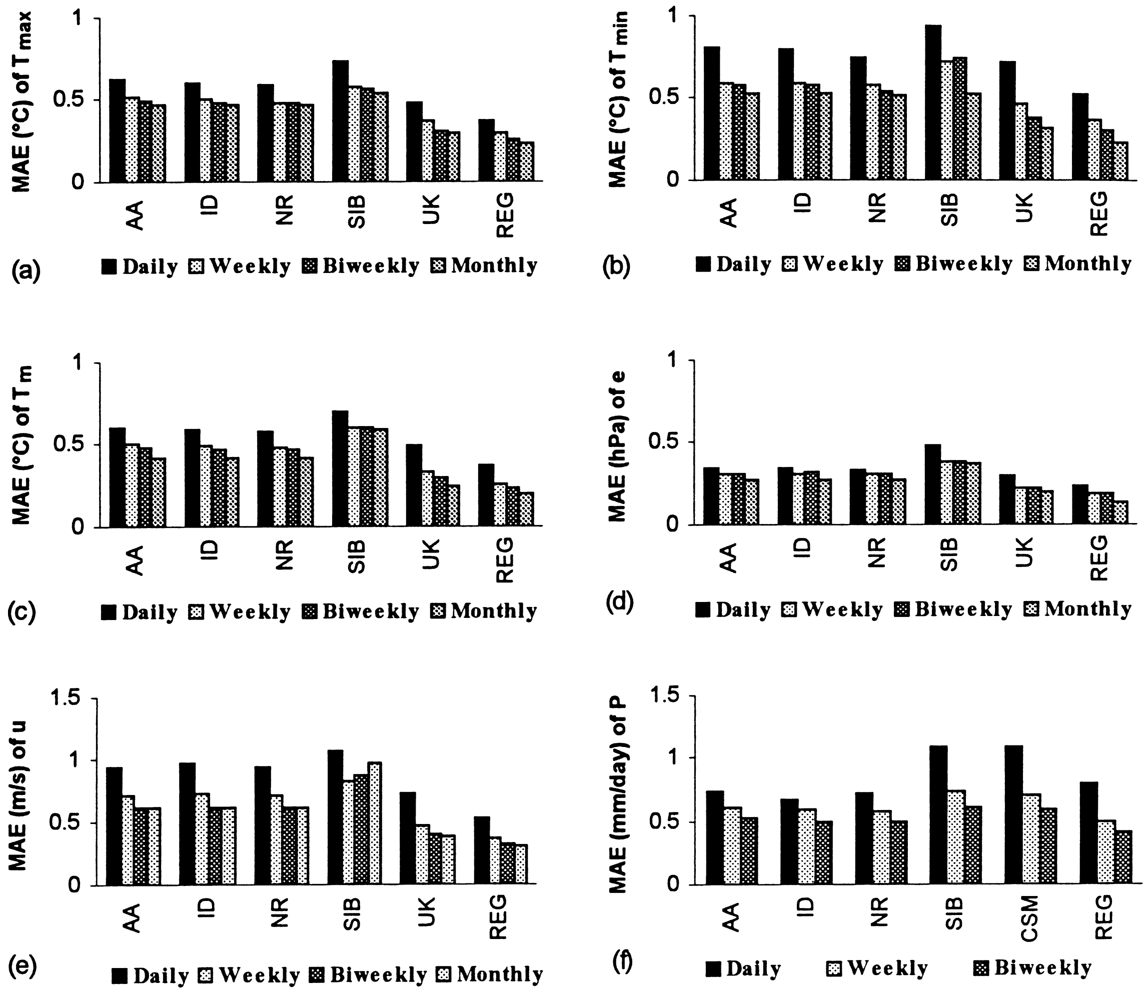

Fig. 1. The average mean absolute error (MAE) between estimated and observed climatological data at six German weather stations for (a) maximum temperature T_{max}, (b) minimum temperature T_{min}, (c) mean air temperature T_m, (d) water vapour pressure e, (e) wind speed u and (f) precipitation P (AA: simple arithmetic averaging, ID: inverse distance interpolation, NR: normal ratio method, SIB: single best estimator, UK: UK traditional method, REG: multiple regression analysis, CSM: closest station method)

(MAE < 0.3 m/s). For precipitation, again all six methods give similar results (MAE < 1.0 mm/day for daily precipitation, MAE < 0.5 mm/day for the other time scales) (Fig. 2(f)). The analysis shows that the selection of methods estimating missing climatological data is very important at Riedenburg for all climatological variables except precipitation. For minimum temperature, the MAE generated by the best estimating method (REG or UK) is 1.5°C smaller than that generated by the other methods. The difference is 0.5°C smaller for maximum temperature, about 1°C smaller for air temperature, about 0.3 hPa smaller for water vapour pressure and 1.3 m/s smaller for wind speed. Comparison with the results of the weather stations shows that the selection of the estimation methods is much more important for a forest site. As the forest climate Station Riedenburg is located in the middle of an extended forest area, forest certainly influences the observed climatological data. The AA, ID, NR and SIB did not consider the systematic forest influence and produced the larger estimated errors. Because UK used a constant difference between forest climate stations and averaged values of five closest weather stations and

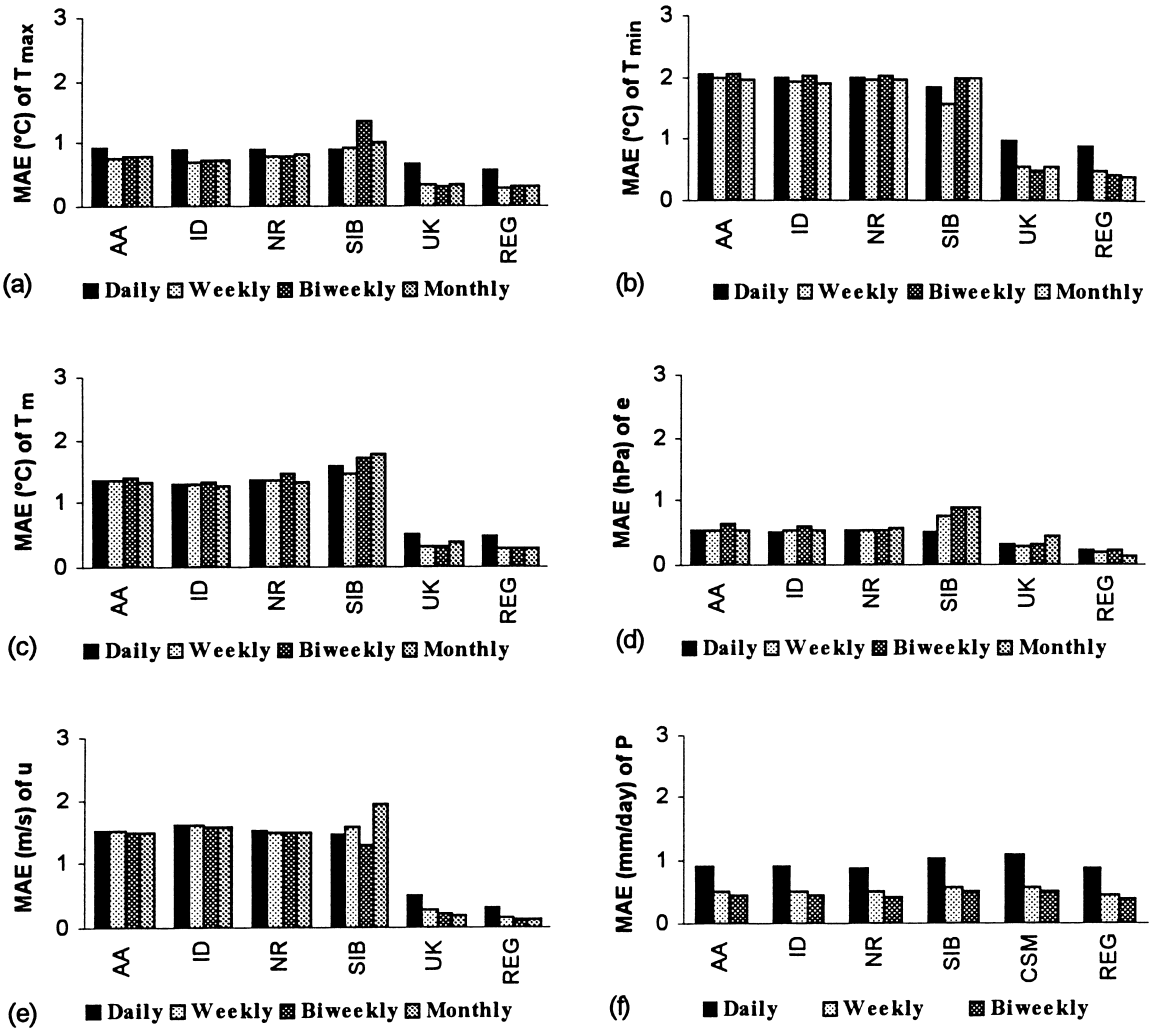

Fig. 2. The mean absolute error (MAE) between estimated and observed climatological data at forest climate Station Riedenburg for (a) maximum temperature T_{max}, (b) minimum temperature T_{min}, (c) mean air temperature T_m, (d) water vapour pressure e, (e) wind speed u and (f) precipitation P (AA: simple arithmetic averaging, ID: inverse distance interpolation, NR: normal ratio method, SIB: single best estimator, UK: UK traditional method, REG: multiple regression analysis, CSM: closest station method).

partly considered the forest influence, its results are better. The REG includes the forest influence indirectly and gives the best estimated results.

3.3. Forest climate Station Ebersberg (EBE, foothills of the Alps)

Forest climate Station Ebersberg is located in the foothills of the Alps close to Munich, its elevation is 538 m. Mean absolute errors between observed and estimated climatological data are shown in Fig. 3. The MAE of maximum temperature are smaller than 1°C, and all six methods give similar estimates (Fig. 3(a)). Evidently REG and UK give smaller MAE for minimum temperature than the other methods. Mean absolute errors (about 0.5°C) calculated by REG and UK are 1.0°C smaller than those (about 1.5°C) calculated by the other methods (Fig. 3(b)). For air temperature (water vapour pressure), similar results are obtained (Fig. 3(c) and (d)), except for the differ-

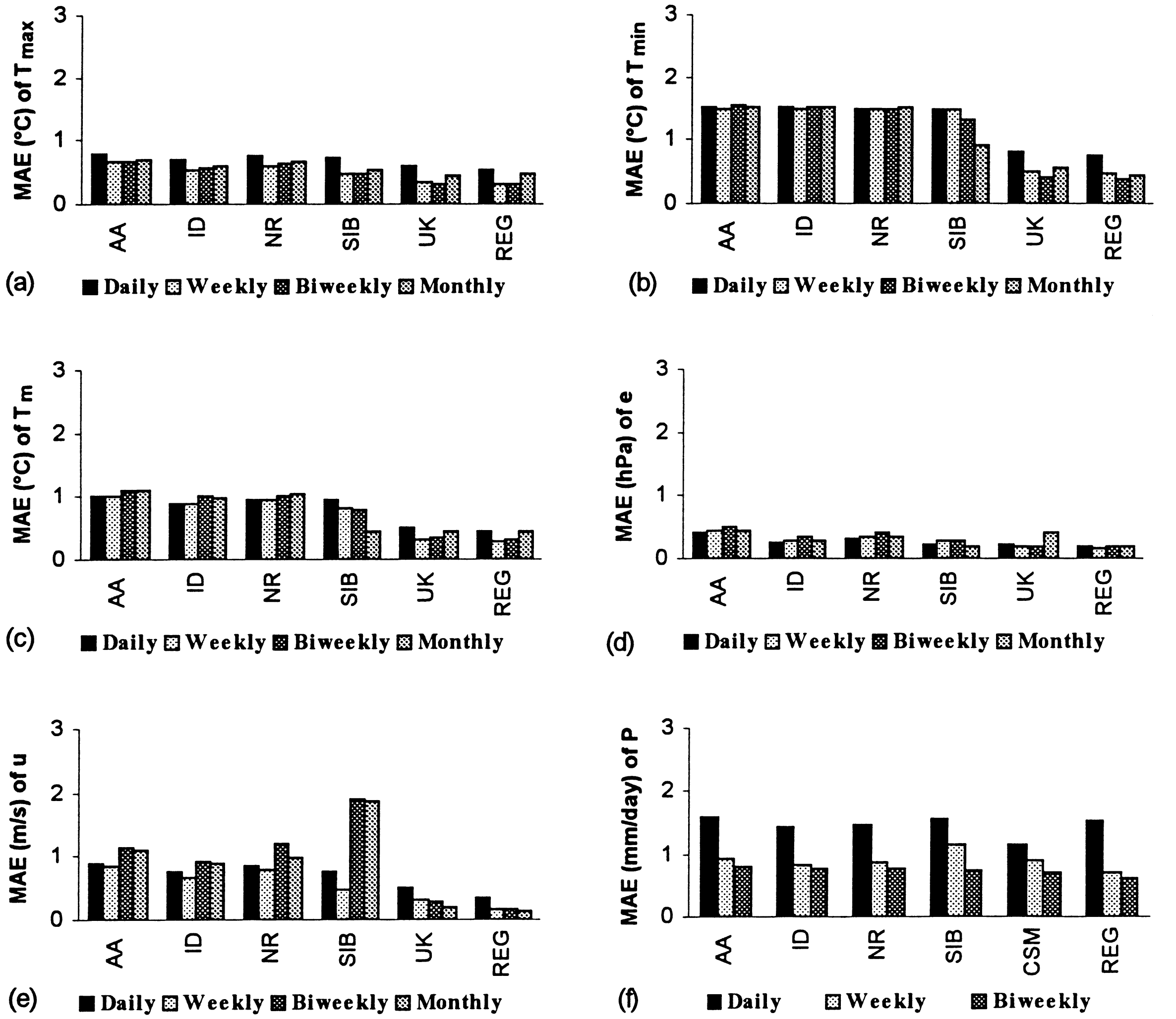

Fig. 3. The mean absolute error (MAE) between estimated and observed climatological data at forest climate station Ebersberg for (a) maximum temperature T_{max}, (b) minimum temperature T_{min}, (c) mean air temperature T_m, (d) water vapour pressure e, (e) wind speed u and (f) precipitation P (AA: simple arithmetic averaging, ID: inverse distance interpolation, NR: normal ratio method, SIB: single best estimator, UK: UK traditional method, REG: multiple regression analysis, CSM: closest station method).

ence of MAE estimated by the best and worst estimating methods is 0.5°C (0.2 hPa). The REG and UK gave the smallest MAE (0.2 m/s), and the estimated errors of the other methods are within 1.5 m/s (Fig. 3(e)). For precipitation, all six methods show similar errors (approximately 1.5 mm/day for daily precipitation, and 1.0 mm/day for the other time scales) (Fig. 3(f)). The REG and UK can give the lowest mean errors for maximum temperature (about 0.5°C), minimum temperature (0.5°C), mean temperature (0.5°C), water vapour pressure (0.2 hPa) and wind speed (0.2–0.4 m/s). The analysis shows that the selection of methods for estimating missing climatological data is very important for mean minimum temperature, mean air temperature, water vapour pressure and wind speed for each of the time scales. For maximum temperature and precipitation, however, the choice of methods is not important at this site. For all six estimation methods, MAE do not vary much with different time scales (from day to month) for maximum temperature, minimum temperature, air temperature, water vapour pressure and wind speed. Mean absolute errors were reduced by 30% from daily precipitation to weekly precipitation. The MAE estimated by REG for daily maximum and minimum temperature in our study were smaller than those from Kemp et al. (1983) and Degaetano et al. (1995), because our weather network is dense (average station separation (square root of the ratio: area/number of stations) is of the order 25 km).

3.4. Forest climate Station Mitterfels (MIT, mountain)

Mitterfels is a mountain forest climate station at an elevation of 1025 m. Clearly its climate will be influenced by elevation and the forest land cover. The average elevation of the weather stations close to forest climate Station Mitterfels is about 400 m. For maximum temperature, minimum temperature and mean air temperature, if a temperature lapse rate of 0.6°C/100 m is applied, a difference of 4.1°C between German weather stations and Mitterfels results. As expected, MAE produced by four methods (AA, ID, NR, SIB) were large and ranged from 2.0 to 3.0°C for maximum temperature (Fig. 4(a)). The REG and UK showed their advantage and produced the smaller MAE (0.5–1.0°C) for maximum temperature. For minimum temperature and mean air temperature, the method producing the lowest errors is REG (0.4–1.0°C), the next is NR (about 1.0°C), the other four methods produced larger MAE ranging from 1.0 to 2.0°C (Fig. 4(b) and (c)). Mean absolute errors of REG and UK are much smaller than those of the other methods for water vapour pressure. The difference between MAE estimated by the REG and ID is about 2.0 hPa (Fig. 4(d)). The REG and UK produced the lowest MAE for wind speed (0.5 m/s), and AA and ID also give smaller MAE (1.0 m/s). NR and SIB gave the largest MAE (above 2.5 m/s) for wind speed (Fig. 4(e)). For precipitation, MAE ranged from 1.0 to 3.0 mm/day. The results showed that all of six estimating methods cannot produce accurate estimates for precipitation at Mitterfels (Fig. 4(f)). Although the climatological data at Mitterfels were influenced by topography and forest, REG can give rather accurate estimates of four time-scale averaged maximum temperature, minimum temperature, air temperature, water vapour pressure and wind speed.

To compare the results from Sections 3.1, 3.2, 3.3, 3.4, we found that the selection of the method for estimating missing climatological data is important for the three typical forest climate stations. The REG can give the most accurate estimate at three forest climate stations for daily, weekly, biweekly, monthly climatological data, except for precipitation. UK is another useful method to estimate missing climatological data at forest climate stations. For German weather stations, the choice of methods is less important, although REG can give more accurate estimates.

4. Seasonal analysis of MAE

The seasonal variation of MAE calculated by REG at German weather stations and forest climate stations is shown in Tables 3–6. The seasonal variation of MAE at the weather stations is smaller than that at forest climate stations for daily T_{max}, T_{min}, T_m and e. The REG produced MAE in daily maximum temperature ranging from 0.45°C in February to 0.31°C in July, at German weather station, but MAE ranged from

Table 3
Mean absolute error of maximum temperature T_{max} at German weather stations and forest climate stations (°C)

Month	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
GWS[a]	0.43	0.45	0.38	0.36	0.34	0.33	0.31	0.32	0.34	0.36	0.36	0.39
EBE[b]	0.83	0.55	0.51	0.42	0.61	0.61	0.41	0.38	0.44	0.46	0.64	0.51
MIT[c]	1.17	0.83	0.85	0.61	0.60	0.57	0.54	0.48	0.61	0.84	1.18	1.13
RIE[d]	0.49	0.70	0.47	0.68	0.59	0.76	0.58	0.58	0.51	0.53	0.38	0.50

[a] GWS: German weather stations.
[b] EBE: Ebersberg.
[c] MIT: Mitterfels.
[d] RIE: Riedenburg.

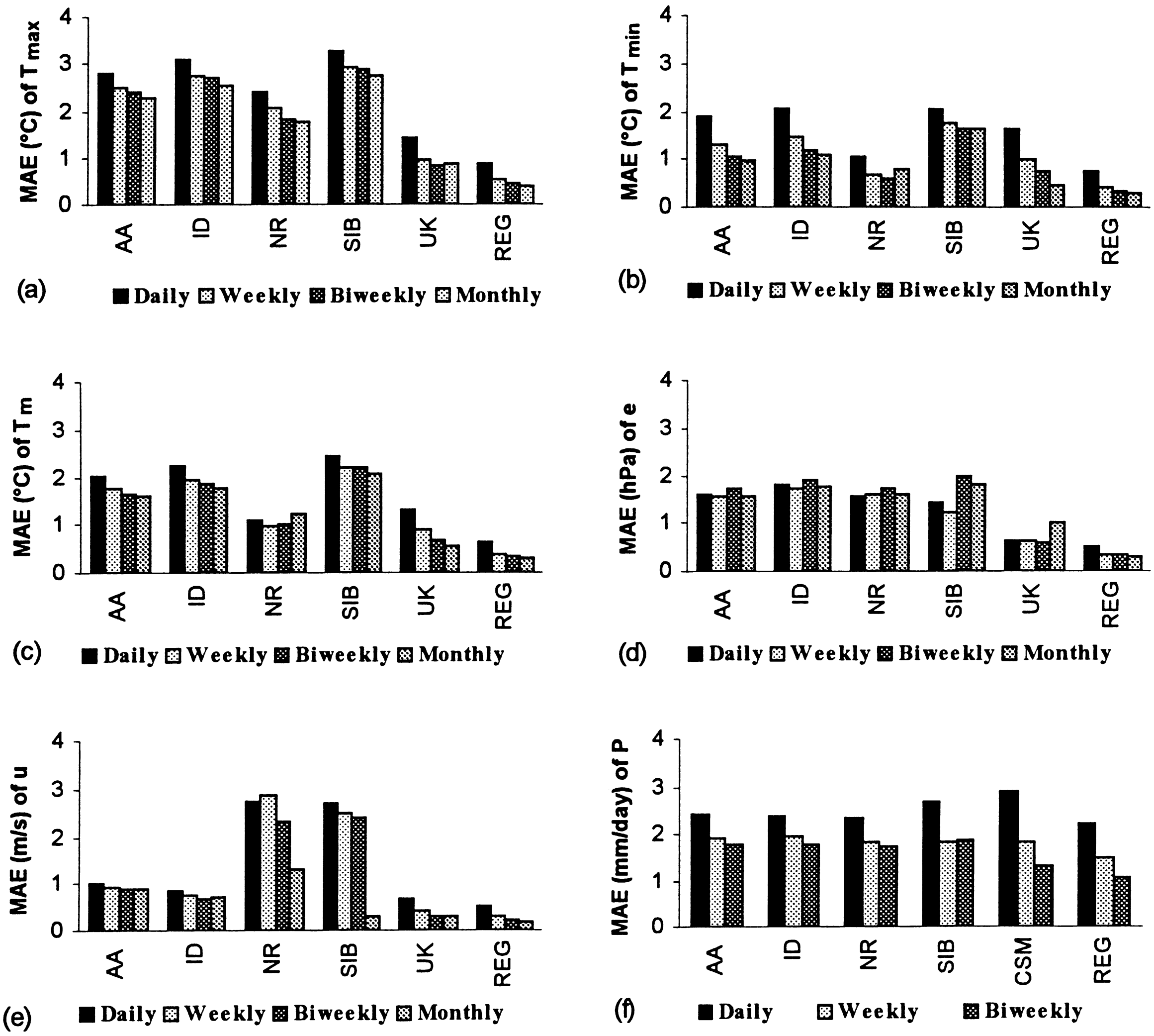

Fig. 4. The mean absolute error (MAE) between estimated and observed climatological data at forest climate station Mitterfels for (a) maximum temperature T_{max}, (b) minimum temperature T_{min}, (c) mean air temperature T_m, (d) water vapour pressure e, (e) wind speed u and (f) precipitation P (AA: simple arithmetic averaging, ID: inverse distance interpolation, NR: normal ratio method, SIB: single best estimator, UK: UK traditional method, REG: multiple regression analysis, CSM: closest station method).

Table 4
Mean absolute error of minimum temperature T_{min} at German weather stations and forest climate stations (°C)

Month	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
GWS[a]	0.55	0.61	0.49	0.56	0.60	0.52	0.50	0.53	0.53	0.53	0.53	0.46
EBE[b]	0.73	0.70	0.75	0.93	0.85	0.8	0.66	0.9	0.71	0.67	0.65	0.94
MIT[c]	1.02	0.72	0.54	0.55	0.81	0.63	0.76	0.64	0.55	0.85	0.91	1.05
RIE[d]	0.67	0.95	0.89	1.03	1.04	0.95	0.78	0.79	0.94	0.88	0.85	0.73

[a] GWS: German weather stations.
[b] EBE: Ebersberg.
[c] MIT: Mitterfels.
[d] RIE: Riedenburg.

Table 5
Mean absolute error of mean temperature T_m at German weather stations and forest climate stations (°C)

Month	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
GWS[a]	0.38	0.40	0.36	0.39	0.40	0.35	0.37	0.38	0.36	0.36	0.27	0.31
EBE[b]	0.46	0.41	0.55	0.56	0.47	0.39	0.41	0.46	0.46	0.40	0.45	0.27
MIT[c]	0.72	0.52	0.41	0.5	1.21	0.42	0.51	0.49	0.45	0.74	0.77	0.85
RIE[d]	0.44	0.43	0.54	0.54	0.57	0.49	0.55	0.53	0.56	0.45	0.32	0.36

[a] GWS: German weather stations.
[b] EBE: Ebersberg.
[c] MIT: Mitterfels.
[d] RIE: Riedenburg.

Table 6
Mean absolute error of water vapour pressure e at German weather stations and forest climate stations (hPa)

Month	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
GWS[a]	0.16	0.15	0.17	0.22	0.27	0.34	0.40	0.37	0.26	0.22	0.14	0.12
EBE[b]	0.18	0.16	0.20	0.22	0.34	0.34	0.54	0.50	0.31	0.38	0.18	0.10
MIT[c]	0.24	0.29	0.29	0.45	1.02	0.64	0.73	0.65	0.49	0.44	0.35	0.27
RIE[d]	0.23	0.23	0.22	0.30	0.43	0.54	0.82	0.78	0.39	0.26	0.24	0.18

[a] GWS: German weather stations.
[b] EBE: Ebersberg.
[c] MIT: Mitterfels.
[d] RIE: Riedenburg.

0.83°C in January to 0.38°C in August at EBE, from 0.76°C in June to 0.38°C in November at RIE, and from 1.18°C in November to 0.48°C in August (Table 3). For daily minimum temperature and mean air temperature, MAE generated by REG are larger at the weather stations and forest climate stations (Tables 4–5). Mean absolute errors of daily water vapour pressure are largest in summer and smallest in winter at both weather stations and forest climate stations (Table 6).

5. Conclusions

Of the methods evaluated, REG was consistently the most accurate at German weather stations and the Bavarian forest climate stations. The next best method is the UK traditional method. As Kemp et al. (1983); Eischied et al. (1995); Degaetano et al. (1995) suggested, REG is very useful over limited areas. Our results indicate that REG is more useful for forest climate stations since it produces the most accurate estimates at different time scales for all climatological variables except for precipitation. However, the MAE at forest climate stations are a little larger than those at German weather stations. Thus multiple regression analysis with the least absolute errors (REG) using the five closest weather stations should be used where complete data bases are required, since REG can account for the local effect (i.e., topography, forest), particularly for forest climate stations.

For German weather stations, the choice of methods to estimate missing climatological data at different time scales is less important. REG again produced the most accurate estimates, but these are not significantly different from the other methods. For Bavarian forest climate stations, which are included in the EU/ICP forests Level II Monitoring Programme (International Co-operation Programme on Assessment and Monitoring of Air Pollution Effects on Forests (ICP-Forests of UN/ECE)) and Level II monitoring network, the selection of methods for estimating missing climatological data is very important, because the difference of MAE estimated by the REG and the other estimating methods is significant. We suggest that REG with the five closest weather stations should be used to estimate the missing climatological data for different time scales at forest climate stations in Level II network, because a rather dense weather station network exists in Europe (i.e., 20–25 km distance in Germany). The filled complete data bases generated with REG as described in this paper are used to reconstruct long-term forest climate data at Bavarian forest sites in the companion paper (Xia et al., 1999).

Acknowledgements

We wish to thank two anonymous reviewers and Dr. John Stewart whose comments contributed greatly to improve the readability and quality of this paper. The work was partly funded by the Bavarian Ministry of Food, Agriculture and Forestry as Project No. M26.

References

Bardossy, A., Muster, H., 1993. Comparison of interpolation methods for daily rainfall amounts under different meteorological conditions. WMO/TD-No. 558, A/72

Barnes, S.L., 1973. Mesoscale objective map analysis using weighted time-series observations. NOAA Tech. Memo. ERL NSSL-62. US Dept. of Commerce, 60 pp.

BayForKlim, 1996. Bavarian Climate Atlas. Munich, Germany (in German).

Bussieres, N., Hogg, W., 1989. The objective analysis of daily rainfall by distance weighting schemes on a mesoscale grid. Atmosphere-Ocean 27, 521–541.

Cannell, M.G.R., 1984. Spring frost damage on young picea stitchensis: 1. Occurrence of damaging frosts in Scotland compared with Western North America. Forestry 57, 159–175.

Cannell, M.G.R., Sheppard, L.J., Smith, R.I., Murray, M.B., 1985. Autumn frost damage on young picea stichensis: 2. Shoot frost hardening, and the probability of frost damage in Scotland. Forestry 58, 145–166.

Cressman, G.P., 1959. An operational objective analysis system. Monthly Weather Rev. 87, 367–374.

Degaetano, A.T., Eggleston, K.L., Knapp, W.W., 1995. A method to estimate missing maximum and minimum temperature observations. J. Appl. Meteorol. 34, 371–380.

Eischeid, J.K., Baker, C.B., Karl, T.R., Diaz, H.F., 1995. The quality control of long-term climatological data using objective data analysis. J. Appl. Meteorol. 34, 2787–2795.

Hevesi, J.A., Istok, J.D., Flint, A.L., 1992. Precipitation estimation in mountainous terrain using multivariate geostatistics Part 1. Structural analysis. J. Appl. Meteorol. 31, 661–675.

Hevesi, J.A., Flint, A.L., Istok, J.D., 1992. Precipitation estimation in mountainous terrain using multivariate geostatistics Part 2. Isohyetal maps. J. Appl. Meteorol. 31, 677–688.

Hubbard, K.G., 1994. Spatial variability of daily weather variables in the high plains of the USA. Agric. Forest Meteorol. 68, 29–41.

Hutchinson, M.F., 1995. Stochastic space-time weather models from ground-based data. Agric. Forest Meteorol. 73, 237–264.

Huth, R., Nemesova, I., 1995. Estimation of missing daily temperature: can a weather categorization improve its accuracy. J. Appl. Meteorol. 34, 1901–1916.

Ishida, T., Kawashima, S., 1993. Use of cokriging to estimate surface air temperature from elevation. Theoretical Appl. Climatol. 47, 147–157.

Kemp, W.P., Burnell, D.G., Everson, D.O., Thomson, A.J., 1983. Estimating missing daily maximum and minimum temperatures. J. Climate Appl. Meteorol. 22, 1587–1593.

Kim, J.W., Chang, J.T., Baker, N.L., Wilks, D.S., Gates, W.L., 1984. The statistical problem of climate inversion: determination of the relationship between local and large-scale climate. Monthly Weather Rev. 112, 2069–2077.

Kimmins, J.P., Comeau, P.G., Kurz, W., 1990. Modeling the interactions between moisture and nutrients in the control of forest growth. Forest Ecol. Manage. 30, 361–379.

Klink, K., Willmott, C.J., 1989. Principal components of the surface wind field in the United States: a comparison of analysis based upon wind velocity, direction, and speed. Int. J. Climatol. 9, 293–308.

Kramer, K., Friend, A., Leinonen, I., 1996. Modeling comparison to evaluate the importance of phenology and spring damage for the effects of climate change on growth of mixed temperate-zone deciduous forests. Climate Res. 7, 31–41.

Luo, Z., Wahba, G., Johnson, D.R., 1998. Spatial-temporal analysis of temperature using smoothing spline ANOVA. J. Climate 11, 18–28.

Meek, D.W., Hatfield, J.L., 1994. Data quality checking for single station meteorological databases. Agric. Forest Meteorol. 69, 85–109.

Nalder, I.A., Wein, R.W., 1998. Spatial interpolation of climatic normals: test of a new method in the Canadian boreal forest. Agric. Forest Meteorol. 92, 211–225.

Palomino, I., Martin, F., 1995. A simple method for spatial interpolation of the wind in complex terrain. J. Appl. Meteorol. 34, 1678–1693.

Paulhus, J.L.H., Kohler, M.A., 1952. Interpolation of missing precipitation records. Monthly Weather Rev. 80, 129–133.

Preuhsler, T., Gietl, G., Grimmeisen, W., 1995. Bayerische Waldklimastationen Jahrbuch 1993. Bayerische Landesanstalt fuer Wald und Forstwirtschaft. Sachgebiet Forsthydrologie, Freising, Germany.

Rudolf, B., Hauschild, H., Reiss, M., Schneider, U., 1992. Die Berechnung der Gebietsniederschläe im 25°-Raster durch ein objectives Analyseverfahren. Meteorologische Zeitschrift 1, 32–50.

Russo, J.M., Liebhold, A.M., Kelley, J.G., 1993. Mesoscale weather data as input to a gypsy moth (Lepidoptera: Lymantriidae) phenology model. J. Econ. Entomol. 86, 838–844.

Saborowski, J., Stock, R., 1994. Regionalization of precipitation data in the Harz mountains. Allgemeine Forst und Jagdzeitung 165, 117–122.

Shepard, D., 1968. A two-dimensional interpolation function for irregularly spaced data. Proc. 1968 ACM Nat. Conf., pp.517–524.

Sokol, Z., Stekl, P., 1994. 3-D mesoscale objective analysis of selected elements from SYNOP and SYRED reports. Zeitschrift für Meteorologie 3, 242–246.

Tabony, R.C., 1983. The estimation of missing climatological data. J. Climatol. 3, 297–314.

Thiebaux, H.J., Pedder, M.A., 1987. Spatial Objective Analysis with Applications in Atmospheric Science. Academic Press, London, 455pp.

Tronci, N., Molteni, F., Bozzini, M., 1986. A comparison of local approximation methods for the analysis of meteorological data. Arch. Meteorol., Geophys. Bioclimatol. 36, 189–211.

Van der Voet, P., Van Diepen, C.A., Oude Voshaar, J., 1994. Spatial interpolation of daily meteorological data. A knowledge-based procedure for the regions of the European Communities. Wageningen, DLO Winand Staring Centre for Integrated Land, soil, and Water Research, Report 53.3, 117pp.

Wallis, J.R., LettenMayer, D.P., Wood, E.F., 1991. A daily hydroclimatological data set for the continental United States. Water Resources Res. 27, 1657–1663.

Ward, M.N., Folland, C.K., 1991. Prediction of seasonal rainfall in the north Nordeste of Brazil using eigenvectors of sea-surface temperature. Int. J. Climatol. 11, 711–743.

Wigley, T.M.L., Jones, P.D., Briffa, K.R., Smith, G., 1990. Obtaining sub-grid-scale information from coarse-resolution general circulation model output. J. Geophysical Res. 95, 1943–1953.

Willmott, C.J., Rowe, C.M., Philpot, W.D., 1985. Small-scale climate maps: a sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring. Am. Cartography 12, 5–16.

Willmott, C.J., Robeson, S.M., Feddema, J.J., 1991. Influence of spatially variable instrument network on climatic averages. Geophys. Res. Letters 18, 2249–2251.

Willmott, C.J., Robeson, S.M., Feddema, J.J., 1994. Estimating continental and terrestrial precipitation averages from raingauge network. Int. J. Climatol. 14, 403–414.

Willmott, C.J., Matsuura, K., 1995. Smart interpolation of annually averaged air temperature in the United States. J. Appl. Meteorol. 34, 2577–2586.

Willmott, C.J., Robeson, S.M., 1995. Climatologically aided interpolation (CAI) of terrestrial air temperature. Int. J. Climatol. 15, 221–229.

Xia, Y., Fabian, P., Stohl, A., Winterhalter, M., 1999. Forest climatology: reconstruction of mean climatological data for Bavaria, Germany. Agric. Forest Meteorol. (this issue).

Young, K.C., 1992. A three-way model for interpolating monthly precipitation values. Monthly Weather Rev. 120, 2561–2569.

Young, K.C., Gall, R.L., 1992. A streamflow forecast model for central Arizona. J. Appl. Meteorol. 31, 465–479.