

Statistical downscaling in central Europe: evaluation of methods and potential predictors

Radan Huth*

Institute of Atmospheric Physics, Bořní II 1401, 141 31 Praha 4, Czech Republic

ABSTRACT: Several statistical downscaling methods and large-scale predictors are evaluated to ascertain their potential to determine daily mean temperatures at 39 stations in central Europe. The methods include canonical correlation analysis, singular value decomposition, and 3 multiple regression models. The potential large-scale predictors are 500 hPa heights, sea level pressure, 850 hPa temperature and 1000–500 hPa thickness. The performance of the methods is evaluated using cross-validation and root-mean-squared error as a measure of accuracy. The stepwise screening of gridpoint data is found to be the statistical model that performed the best. Among the predictors, temperature variables yield more accurate results than circulation variables. The best predictor is the combination of 500 hPa heights and 850 hPa temperature. Geographical variations of the specification skill, mainly the differences between the elevated and lowland stations, are also discussed.

KEY WORDS: Statistical downscaling · Daily temperature · Canonical correlation analysis · Multiple regression · Cross-validation

1. INTRODUCTION

There are several methodologies to bridge the gap that exists between what the general circulation models (GCMs) are able to simulate with enough accuracy and what is needed in climate impact studies (von Storch 1995, Kattenberg et al. 1996, Wilby & Wigley 1997). Statistical downscaling methods (or, in a somewhat narrower sense, regression methods) are perhaps the ones most widely used. Their essence is to seek statistical relationships between the variables simulated well by GCMs, which are treated as predictors, and those required by impact researchers, treated as predictands.

GCMs do a reasonable job in simulating large-scale upper-air features but fail to reproduce surface variables on regional and local scales (Grotch & MacCracken 1991, Gates et al. 1996), which are essential in assessments of climate change impacts. Therefore, large-scale circulation (geopotential height or sea level pressure) patterns have been used as the only

predictor in many downscaling studies for various variables, including temperature (Hewitson & Crane 1992, Hewitson 1994, Schubert & Henderson-Sellers 1997), precipitation (von Storch et al. 1993, Noguer 1994, Corte-Real et al. 1995, Saunders & Byrne 1996) and sea level (Heyen et al. 1996). However, the assumption that changes of surface climate elements due to the enhanced greenhouse effect are attributable to changes in circulation only does not appear to be realistic because observed changes in circulation are not a dominant agent in long-term surface climate trends (Yarnal 1985). The application of downscaling based on circulation only to a GCM run for the $2 \times \text{CO}_2$ climate may lead to an entirely unrealistic result of virtually no temperature change, although a direct GCM output indicates warming by several degrees (Schubert 1998). Several recent studies avoid that questionable assumption and include among predictors other variables such as upper air temperature (or equivalent thickness) (Kaas & Frich 1995, Cavazos 1997) and moisture variables (Crane & Hewitson 1998, Wilby et al. 1998).

Several statistical methods are applicable to the description of relationships between large-scale

*E-mail: huth@ufa.cas.cz

upper-air fields and local climate elements. Multiple linear regression (MLR), either based directly on grid-point data or on principal components (PCs) of predictor fields, and canonical correlation analysis (CCA) have been used most widely. Recently, non-linear methods have emerged, including multi-variate splines (Corte-Real et al. 1995) and neural networks (Cavazos 1997, Crane & Hewitson 1998, Wilby et al. 1998).

The aim of this study is to evaluate and compare the performance of several linear methods of downscaling the daily mean temperature for a variety of predictors, including both upper-air circulation and temperature variables. The evaluation is carried out on a network of 39 stations, covering 6 countries in central and western Europe, for the observations under present climate conditions. The study is confined to winter (December to February), when circulation is best pronounced in the northern mid-latitudes. However, circulation's links with surface temperature have been shown to be of similar strength both in winter and summer at several stations in the Czech Republic (Huth 1997).

2. DATASETS USED

The study examines 8 winter seasons (December to February) from 1982–1983 to 1989–1990. The source of large-scale upper-air fields was the NCEP reanalyses. The potential predictors include 500 hPa heights (Z500), sea level pressure (SLP), 850 hPa temperature (T850) and thickness (relative topography) of the 1000–500 hPa layer (RT), all defined on a $5^\circ \times 10^\circ$ grid extending from 45° W to 45° E and from 35° to 70° N, which covers most of Europe and adjacent parts of the Atlantic Ocean.

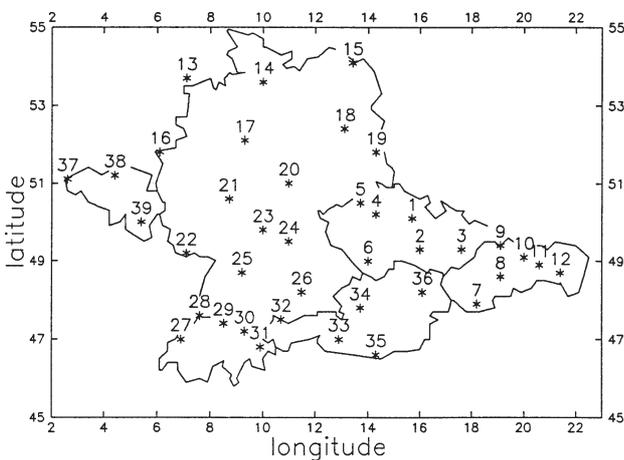


Fig. 1. Geographic locations of the stations used in the study. For their names see Table 1

The station network with daily mean temperature data is composed of 39 stations in Austria, Belgium, the Czech Republic, Germany, Slovakia and Switzerland, situated in different orographic conditions, ranging from the sea coast to mountain valleys and summit stations. For the locations of the stations see Fig. 1; a list of the station names together with their altitudes is provided in Table 1.

3. METHODOLOGY

3.1. Downscaling methods. Three linear methods of statistical downscaling are examined: (1) canonical correlation analysis (CCA) prefiltered by principal component analysis (PCA); (2) singular value decomposi-

Table 1. Stations used in the study. Altitudes are in metres above the mean sea level. Station numbers are used to show the stations' positions in Fig. 1

Stn no.	Stn name	Country	Altitude
1	Hradec Králové	Czech Republic	278
2	Velké Meziříčí	Czech Republic	452
3	Holešov	Czech Republic	224
4	Milešovka	Czech Republic	833
5	Teplice	Czech Republic	225
6	Husinec	Czech Republic	536
7	Hurbanovo	Slovakia	115
8	Sliač	Slovakia	313
9	Oravská Lesná	Slovakia	780
10	Žtrbské Pleso	Slovakia	1360
11	Poprad	Slovakia	695
12	Košice	Slovakia	230
13	Norderney	Germany	11
14	Hamburg	Germany	13
15	Greifswald	Germany	2
16	Kleve	Germany	46
17	Hamel	Germany	66
18	Potsdam	Germany	81
19	Cottbus	Germany	69
20	Erfurt	Germany	316
21	Gießen	Germany	186
22	Saarbrücken	Germany	319
23	Würzburg	Germany	268
24	Nürnberg	Germany	314
25	Stuttgart	Germany	373
26	München	Germany	515
27	Neuchâtel	Switzerland	487
28	Basel	Switzerland	317
29	Zürich	Switzerland	569
30	Säntis	Switzerland	2498
31	Davos	Switzerland	1590
32	Reutte	Austria	870
33	Sonnblick	Austria	3105
34	Feurkogel	Austria	1618
35	Klagenfurt	Austria	447
36	Wien	Austria	202
37	Koksijde	Belgium	5
38	Deurne	Belgium	10
39	Saint Hubert	Belgium	556

tion (SVD); and (3) multiple linear regression (MLR). The CCA and SVD methods are discussed in detail by Bretherton et al. (1992). To the author's knowledge, SVD has not been applied in the context of downscaling research before, but it has been used as an alternative to CCA in various climatological studies (e.g. Lau & Nath 1994, Ward & Navarra 1997). Basically, both CCA and SVD seek pairs of patterns, one from predictors and one from predictands, that share maximum correlation (in the case of CCA) or covariance (SVD). The second, third, etc. pairs (modes) maximize correlation/covariance not captured by the preceding pairs. Prefiltering and orthogonalization of predictors prior to CCA, which is performed using PCA, have been recommended e.g. by Barnett & Preisendorfer (1987) and Barnston & Ropelewski (1992). Three different MLR models are considered: (1) stepwise screening of principal components (PCs) (hereafter referred to as 'stepwise regression'), (2) MLR on PCs without screening, that is, all PCs being forced to enter the model ('full regression'), and (3) stepwise screening of gridded values ('pointwise regression'). Unlike CCA and SVD, MLR calculates the predictand at target stations independently of each other. In stepwise screening, each potential predictor (PC or gridpoint value) is evaluated for its individual significance level before including it in the regression equation, and, after each addition, each variable within the equation is evaluated for its significance as part of the model. A variable is included (retained) in the equation if the corresponding significance level exceeds 90% (95%).

For each method, several plausible variants of the number of PCs and the number of modes have been examined.

3.2. Analysis procedure. A preliminary analysis showed that the specification of daily mean surface air temperature from large-scale upper-air fields is more accurate for predictors and predictands expressed in terms of normalized anomalies than for raw values and non-normalized anomalies (not shown here). Therefore, both predictors and predictands are treated in the form of normalized anomalies, i.e. long-term winter means have been subtracted from the actual station/gridpoint values and then divided by standard deviations.

Since the downscaling methods allow only a part of the total variance of daily temperature time series to be explained, the downscaled temperatures (i.e. anomalies resulting from the downscaling multiplied by the observed standard deviation and increased by the observed mean) have smaller variance than the observed series. To retain the original variability, the downscaled time series are inflated, i.e. their standard deviation is increased by dividing the temperature anomaly on each day by the correlation between orig-

inal and downscaled series (Karl et al. 1990). This has been the most common way of handling the unexplained variance so far. However, the variance inflation implicitly assumes that all local variability is related to large-scale variability, which is not valid (von Storch 1999). An alternative to inflation is to add noise to the downscaled series to represent the processes unresolved by the large-scale predictor. In this study I use inflation; as a result, the downscaled temperatures reproduce the mean and variance of the observed series. A comparison of the downscaling schemes based on inflation with noise addition is a subject of further research.

The downscaling methods are evaluated within the cross-validation framework, which allows an unbiased estimate of potential 'predictability'. Cross-validation consists of omitting one case in turn, building the whole statistical model on the remaining dataset, and applying the statistical model to the omitted case (Michaelsen 1987). Since time series of daily temperatures exhibit considerable autocorrelation, the above procedure would lead to an optimistic bias of downscaling skill if only 1 day were omitted each time. I have used a more appropriate approach: 1 season at a time is removed, and the statistical model built on the remaining 7 seasons is tested on the omitted period. All the statistical models are thus built 8 times.

The accuracy of the downscaled values is quantified in terms of the root-mean-squared error (RMSE) of the downscaled values relative to the observed. The results based on RMSE may, however, be biased because the variance inflation leads to an overestimation of the RMSE (von Storch 1999). The results were qualitatively the same, nevertheless, i.e. the rating of the methods did not change, when the correlation coefficient was used as a measure of accuracy (not shown). Since inflation has no effect on the correlation coefficient, we may conclude that the inflation does not impact adversely the results of this study in terms of RMSE. To obtain 1 number for each method to characterize its performance, the RMSE values are averaged over the stations.

There are other possible criteria according to which the degree of reproduction of original time series may be evaluated: for example, their time structure (persistence) and the spatial structure (correlations, modes of variability) are also of importance in many applications. The way the unexplained part of variance is treated is likely to affect the time and space structure of downscaled series quite considerably. The analysis here, however, concentrates on the accuracy of specification; investigations on the time and space structure of downscaled temperatures, with emphasis on the effect of the treatment of the unexplained variance, are underway.

Table 2. Percentage of variance explained by the 15 leading PCs (principal components) of 500 hPa heights in the whole dataset (All) and in datasets used in cross-validation by omitting the first to eighth year (1 to 8)

PC	All	1	2	3	4	5	6	7	8
1	21.80	21.30	22.31	21.40	21.84	21.98	22.09	22.30	21.96
2	16.21	16.38	15.87	16.03	16.24	15.99	16.73	16.66	16.15
3	12.67	12.59	12.64	13.11	12.49	12.85	12.64	12.09	12.99
4	10.64	10.80	10.40	10.80	10.79	10.87	10.86	10.47	9.86
5	7.61	7.29	7.59	7.51	7.63	7.58	7.45	7.77	7.60
6	5.35	5.67	5.33	5.11	5.17	5.23	5.25	5.48	5.65
7	4.36	4.23	4.31	4.66	4.47	4.27	4.36	4.25	4.22
8	3.16	3.08	3.18	3.24	3.19	3.14	3.14	3.15	3.21
9	2.69	2.79	2.69	2.73	2.67	2.63	2.60	2.68	2.81
10	2.42	2.56	2.43	2.40	2.33	2.41	2.49	2.21	2.40
11	2.20	2.22	2.28	2.06	2.26	2.20	2.00	2.17	2.27
12	1.51	1.49	1.53	1.52	1.55	1.55	1.49	1.51	1.45
13	1.33	1.35	1.34	1.34	1.36	1.27	1.25	1.38	1.33
14	0.92	0.96	0.94	0.95	0.91	0.93	0.86	0.90	0.96
15	0.85	0.93	0.87	0.86	0.86	0.87	0.78	0.84	0.86

4. EVALUATION OF METHODS

The performance of the downscaling methods is evaluated for 500 hPa heights used as a predictor. Three of the 5 methods (CCA and stepwise and full regression) require PCs of predictors as input; CCA also needs PCs of predictands. The percentage of variance explained by the leading PCs is displayed in Tables 2 & 3 for the Z500 field and daily mean temperatures, respectively. Values are shown both for the whole dataset and for the 8 subsets used in cross-validation. Plausible numbers of PCs to retain and enter the statistical models can be derived from the separation of eigenvalues; according to O'Lenic & Livezey's (1988) recommendation, the PCs should be cut just behind the section with a relatively small slope on an eigenvalue versus PC-number diagram. Tables 2 & 3 confirm that the suitable numbers of PCs to retain are the same in all datasets: 4, 7 and 11 for the Z500 field and 4, 7 and 9 for temperature.

The area-averaged standard deviation of daily mean temperature is 4.96°C. This may be considered as a lower benchmark for downscaling methods, representing the accuracy of a 'climatological downscaling', since it is identical with the RMSE if daily temperatures were replaced with the winter seasonal mean. The spatially averaged values of RMSE for several variants of regression, SVD and CCA are displayed in Fig. 2, manifesting a considerable improvement for a large majority of variants over the 'climato-

logical downscaling'. Several observations can be inferred from Fig. 2: (1) The pointwise regression performs best, and seems to represent an upper limit of performance for other methods. (2) The higher the number of PCs, representing the predictor field, the better the performance of the methods that use prefiltering by PCA (stepwise and full regression, and CCA). (3) The performance of stepwise and full regression is almost identical. (4) In CCA, the increase in the number of canonical modes leads to an improvement first, but after attaining a 'saturation level' (3 to 4 modes), adding more modes into the model deteriorates the results. (5) The performance of the SVD method worsens with the inclusion of a larger number of modes.

The results are somewhat surprising in several aspects. First of all, the PCs explaining a small portion of variance, which are frequently considered to be noise and to adversely affect the regression results, contribute non-negligibly to an improvement in the accuracy of specification by both CCA and regression methods. This effect occurs naturally if the evaluation is performed on the dependent sample; the surprising fact is that no apparent overfitting of the statistical models by weakly significant predictors is observed. The improvement due to the inclusion of additional PCs into the statistical models is not spatially homogeneous. Fig. 3 suggests that there are sites where the PCs of higher order do not bring a pronounced improvement of specification (e.g. at Norderney, Stn 13, the addition of the 5th to 11th PCs lowers the RMSE by 0.2°C only). Adding PCs 5 to 7 to the full regression

Table 3. Percentage of variance explained by the 12 leading PCs of surface daily mean temperature in the whole dataset (All) and in datasets used in cross-validation by omitting the first to eighth year (1 to 8)

PC	All	1	2	3	4	5	6	7	8
1	76.22	76.84	76.71	73.60	76.26	76.11	76.73	77.28	75.52
2	6.37	6.06	6.30	7.59	6.17	6.18	6.17	6.33	6.62
3	4.99	4.92	4.90	5.47	5.18	4.69	5.01	4.62	5.31
4	3.41	3.38	3.41	3.60	3.51	3.57	3.43	3.04	3.34
5	1.26	1.24	1.23	1.36	1.33	1.35	1.18	1.22	1.28
6	0.98	0.96	0.93	1.03	0.95	1.08	0.92	0.96	1.02
7	0.86	0.83	0.84	0.93	0.86	0.90	0.84	0.84	0.90
8	0.66	0.67	0.65	0.74	0.67	0.70	0.64	0.66	0.70
9	0.63	0.65	0.60	0.71	0.59	0.61	0.57	0.62	0.66
10	0.49	0.44	0.46	0.55	0.39	0.52	0.47	0.47	0.53
11	0.38	0.38	0.36	0.41	0.35	0.38	0.38	0.39	0.39
12	0.31	0.31	0.31	0.36	0.32	0.34	0.31	0.30	0.32

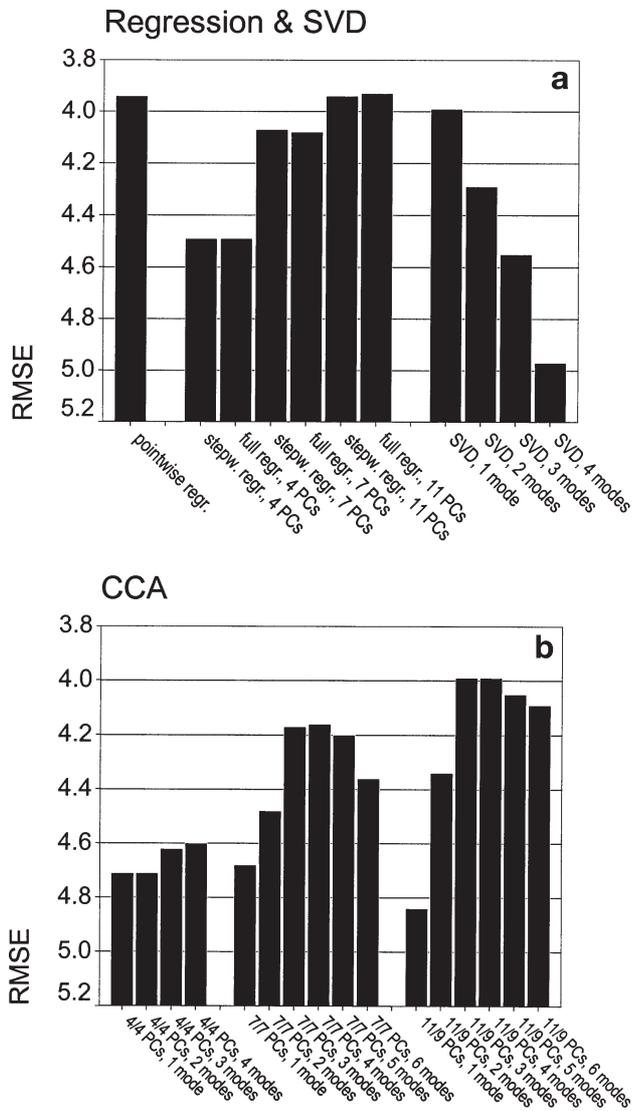


Fig. 2. Accuracy (in terms of spatially averaged RMSE; in °C) for various downscaling methods and numbers of PCs/modes: (a) regression and SVD (singular value decomposition), (b) CCA (canonical correlation analysis); all for Z500 (500 hPa height) as predictor. Note that the RMSE scale is reversed: the higher the bar, the more accurate the specification

model (Fig. 3a) brings about the largest improvement at elevated stations both in the Alps and in Slovakia (with a maximum RMSE decrease of more than 1°C at the Austrian summit stations) whereas the improvement is only marginal (decrease by RMSE of less than 0.2°C) at the maritime sites in the northwest of the region. The 8th to 11th PCs lead to an improvement of more than 0.2°C in the central part of the region (mainly in southern Germany) whereas a negligible improvement is observed in the north; at Košice (the easternmost station), even a slight deterioration appears. A qualitatively similar picture appears for the

CCA method. It is difficult to identify physical reasons for spatial differences in the RMSE improvement in Fig. 3, especially because the higher-order PCs constitute a mixture of processes with various areal extent and time behaviour. I do not make such an attempt here.

Second, the inclusion in the regression model of the PCs not selected by the stepwise screening, i.e. the redundant ones not having a discernible effect on surface temperature, does not deteriorate the results. In other words, the stepwise screening of PCs seems to be a useless complication in the downscaling procedure. Again overfitting does not take place.

Third, the pointwise regression is superior to (or, if enough PCs are retained, at least as good as) any other method. The superiority of the pointwise regression to regressions based on PCs has already been reported by Klein & Walsh (1983) in their study on specification of monthly temperatures from 700 hPa heights. This superiority occurs despite a larger stability from one cross-validated subsample to another of PCs and canonical modes relative to the gridpoints selected to enter the pointwise regression. Table 4 shows the congruence coefficients for the left and right heterogeneous correlation maps between the whole sample and 8 cross-validated subsamples, for 11/9 PCs and 4 canonical modes (which is the optimum number). [The congruence coefficient is defined similarly to the correlation coefficient except that it does not remove the mean of the variables. Its use when similarity of magnitude is important in addition to pattern similarity has been recommended e.g. by Richman & Lamb (1985); according to them, congruence exceeding 0.98 indicates a 'perfect' match. The k th left (right) heterogeneous correlations ('left' and 'right' stand for the predictor and predictand, respectively) are correlation coefficients between the gridpoint values of the left (right) field with the k th right (left) expansion coefficient (i.e. the time series of 'intensity' of the k th canonical mode). The right heterogeneous correlation indicates how well the station temperatures can be specified from the k th left expansion coefficient (Bretherton et al. 1992).] The congruence coefficients in Table 4 exceed the 0.98 threshold in their majority, indicating a perfect match between the modes in the whole dataset and in the cross-validated samples. In other words, the modes are almost unaffected by the omission of 1 year from the data. The stability of the pointwise regression model is much less: various gridpoints are selected in different cross-validated samples. Fig. 4 displays which gridpoints enter the pointwise regression for Würzburg (Stn 23) in the whole dataset and in the 8 cross-validated samples. Although there is a pronounced spatial coherence in the selected gridpoints and the sign with which they contribute to

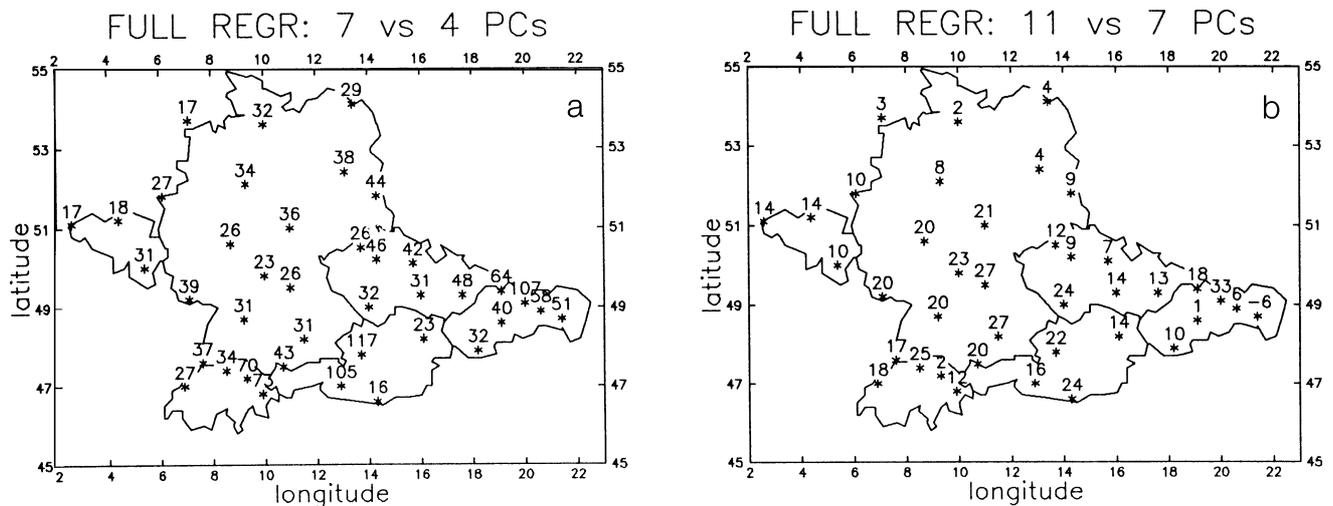


Fig. 3. Improvement of specification (decrease of RMSE; in hundredths of °C) for the full regression method and Z500 predictor; (a) 7 versus 4 PCs, (b) 11 versus 7 PCs

Table 4. Congruence coefficients between heterogeneous correlation maps based on the whole dataset and those based on the cross-validated subsamples for 4 leading canonical modes; CCA based on 11 Z500 PCs and 9 station temperature PCs. For an explanation of the congruence coefficient see Section 4

Subsample	Left map of mode no.				Right map of mode no.			
	1	2	3	4	1	2	3	4
1	99.8	99.8	98.9	99.2	100.0	99.9	99.9	98.3
2	99.5	99.6	99.7	99.0	100.0	100.0	99.9	96.1
3	99.2	99.6	98.8	97.7	99.8	100.0	99.7	97.8
4	99.7	99.1	99.8	99.0	100.0	99.8	99.8	98.5
5	99.6	99.4	-99.7	99.5	100.0	99.9	-100.0	99.3
6	99.9	99.8	99.3	99.1	100.0	100.0	99.8	98.9
7	99.6	99.6	99.4	98.3	99.9	100.0	99.8	99.4
8	99.7	99.1	98.5	98.6	100.0	99.8	99.6	99.3

the model (warmer temperatures are connected with negative 500 hPa anomalies northwest and north of the station and with positive anomalies south and east of it), their spread over the map is quite remarkable: there is only 1 gridpoint (east of Würzburg) that enters all the 9 regression models.

The left and right heterogeneous maps of the leading 4 canonical modes for 11/9 PCs of the whole dataset are shown in Fig. 5 for illustration. Note the different behaviour of the elevated stations: they are more sensitive than the surrounding lowland stations to the 500 hPa anomaly over southeastern Europe (Mode 1), in its positive polarity inducing a warm advection into the region of interest, but less sensitive to the anomaly over the North Sea (mainly the Alpine sites; Mode 3) and to a dipole with anomalies centred over Spain and Scandinavia (mainly the stations in Slovakia; Mode 2).

The overall performance of the pointwise regression and full regression on 11 PCs is the same: the area averaged RMSE equals 3.94°C. However, the performance differs regionally (Fig. 6). The regression based on PCs is much better at the Alpine stations; at most lowland stations, the pointwise regression tends to perform better by a narrow margin. The PC regression is superior at elevated stations probably because of their proximity to the free atmosphere: they are more tightly related to large-scale circulation features such as variability modes (teleconnections). On the other

hand, at lowland stations, where more local influences are in effect, the pointwise regression allows some of the local peculiarities to be explained by a single grid-point value (or a few of them) specific for the site and selected so as to fit it best.

Analogous conclusions regarding the relative performance of the methods apply if other predictors, discussed in the following section, are used.

5. EVALUATION OF PREDICTORS

The area averaged RMSE for the statistical model that performed the best, i.e. stepwise regression, is presented in Fig. 7 for 5 predictors: SLP, Z500, RT, T850, and the combination of Z500 and T850. Obviously, the large-scale free-atmosphere temperature variables (T850, RT) are much better predictors of sur-

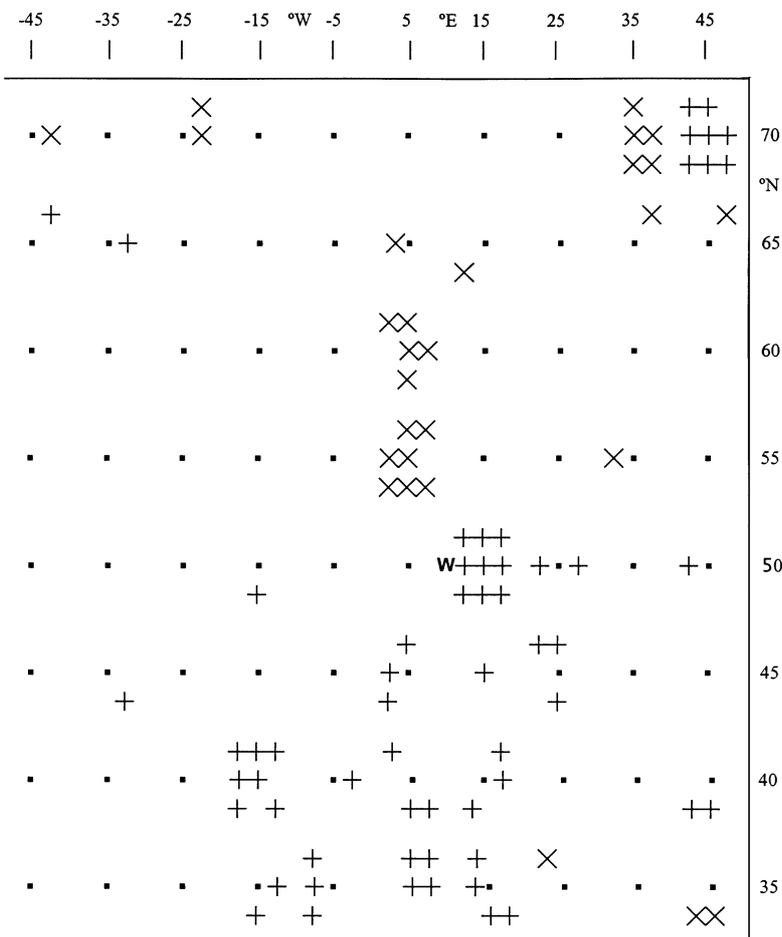


Fig. 4. A scheme indicating which gridpoints (■) enter the pointwise regression model (Z500 as predictor) for Würzburg in the whole dataset and 8 cross-validated samples. +: positive contribution in the regression equation; x: negative contribution. A symbol at the centre, replacing the square, indicates that the gridpoint was selected into the regression model for the whole dataset; symbols around the centre indicate that the gridpoint was selected for a cross-validated subsample. W: approximate position of Würzburg

face air temperature than the circulation variables (SLP, Z500). The best performance is achieved if the temperature and circulation variables are combined; of all pairs of variables, 500 hPa heights together with 850 hPa temperatures yield the lowest RMSE of 2.88°C. The corresponding correlation coefficient between the specified and observed values reaches 0.82.

In the following, the spatial distribution of RMSE is considered. All results reported in this section are obtained by the pointwise regression model. First of all, the performance of SLP and Z500 predictors is compared (Fig. 8). If averaged over all stations, mid-tropospheric circulation is a slightly better predictor than SLP. However, the superiority of the former stems mainly from its better performance at a few elevated

stations, whereas at most lowland stations, SLP appears to perform better, although by a narrow margin.

The improvement of specification using T850 as predictor relative to Z500 exhibits a different spatial structure (Fig. 9). There is no pronounced difference between the elevated and lowland stations, except for the 2 stations with highest altitude (Sonnblick, Stn 33, and Säntis, Stn 30) where both predictors perform with almost the same accuracy. The improvement (T850 being a better predictor than Z500) tends to gradually increase from west to east. That is, at maritime sites, mid-tropospheric geopotential is almost as good predictor as lower-tropospheric temperature, whereas in more continental regions temperature performs considerably better.

The regional distribution of the RMSE for the combined Z500 and T850 predictor is shown in Fig. 10a. For the sake of convenience, Fig. 10b displays the ranks of the RMSE for each station. Two effects contribute to the degree of accuracy of the downscaling. First, elevation plays a role: the specification is most successful at elevated stations, regardless of their location on summits (Säntis, Stn 30; Sonnblick, Stn 33; Feurkogel, Stn 34; and also Milešovka, Stn 4) or elevated plains (2trbské Pleso, Stn 10) or in mountain valleys (Davos, Stn 31). Second, the accuracy improves from east to northwest, being highest (apart from the mountain stations) at the maritime sites. The largest error is observed at the station of Teplice (Stn 5), located in a valley where frequent temperature inversions occur and local

weather is only weakly related to large-scale circulation patterns. The improvement in the specification based on the combined predictor over the Z500 predictor is 1.1°C on average; at individual sites it ranges from 0.5°C at Säntis and Sonnblick to 1.7°C at 4 stations in the Czech Republic and Slovakia (see Fig. 11a). The pattern resembles the difference between T850 and Z500 in that the improvement increases from west to east in general, with the exception of the 2 most elevated stations, where the improvement is smallest. The improvement over 850 hPa temperature, being 0.3°C on average, exhibits no pronounced regional dependence (Fig. 11b). Worth noting is the fact that at all stations, adding 500 hPa geopotential as a predictor to 850 hPa temperature leads to a more accurate spec-

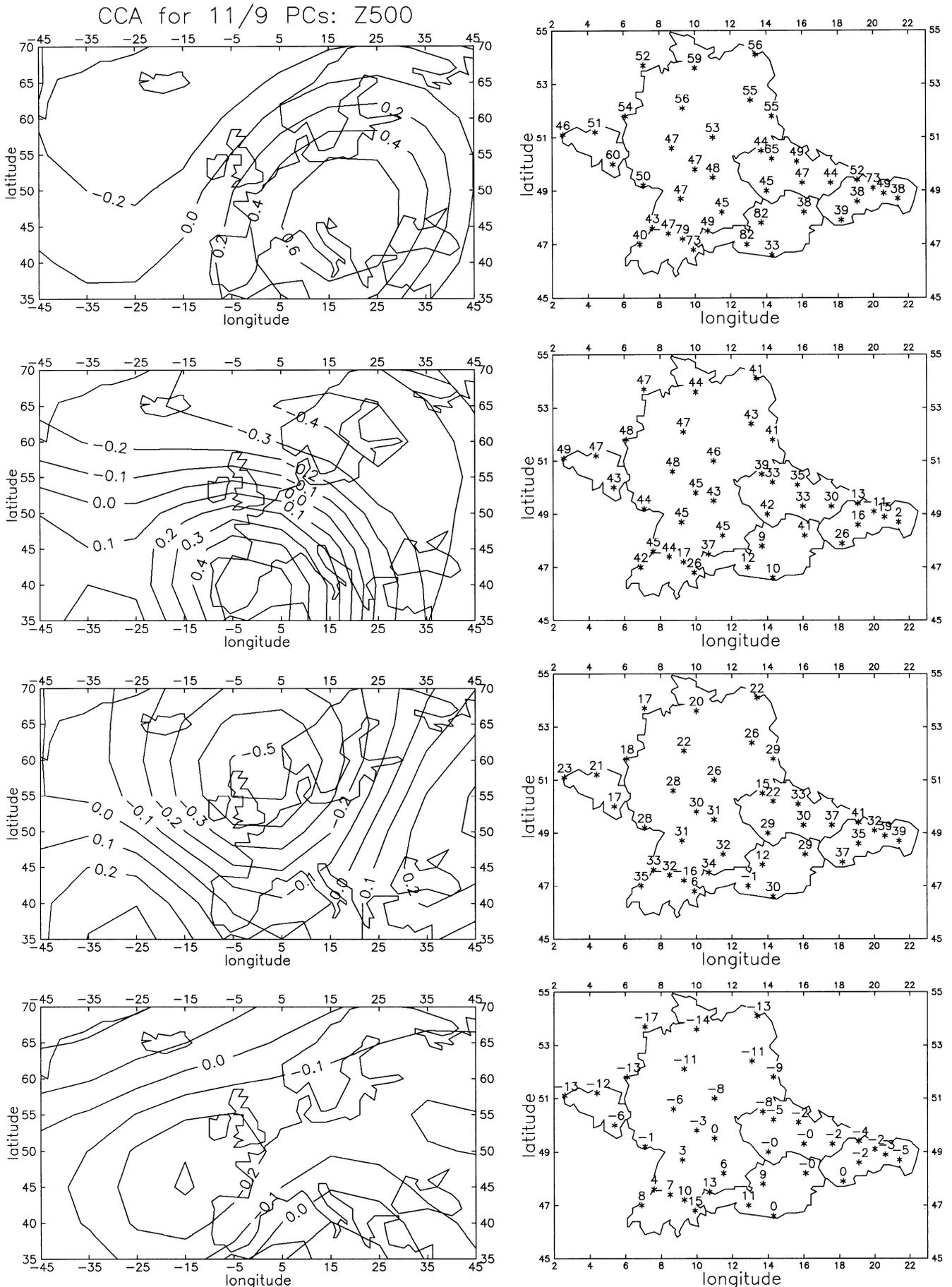


Fig. 5. Left and right heterogeneous correlation maps of the leading 4 canonical modes (Modes 1 to 4 from top to bottom) for 11/9 PCs; Z500 field as predictor; whole dataset. Values in the right maps are multiplied by 100

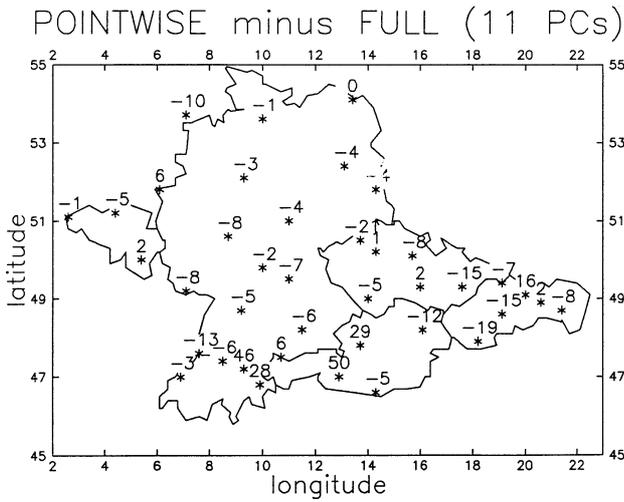


Fig. 6. Difference in RMSE (in hundredths of °C) between the pointwise and full regression of 11 PCs, Z500 as predictor. Negative values indicate that the pointwise regression is more accurate

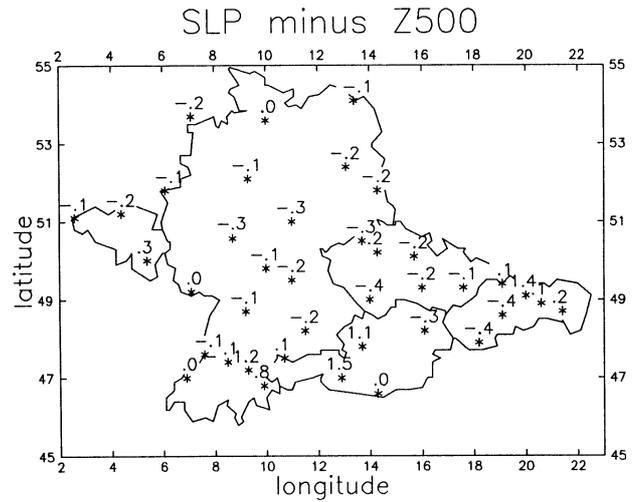


Fig. 8. Difference in RMSE (in °C) between SLP and Z500 as predictors, for pointwise regression method. Positive values indicate that Z500 performs better

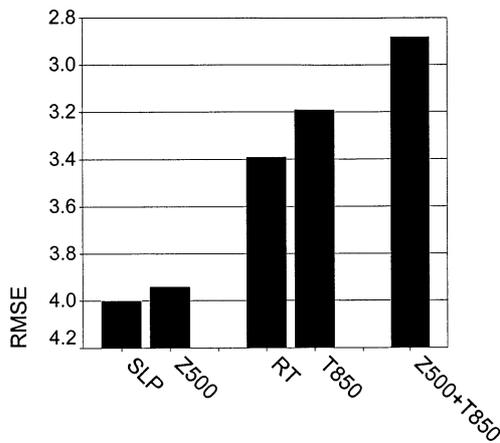


Fig. 7. Accuracy (in terms of spatially averaged RMSE; in °C) for various predictors; pointwise regression method is used. SLP: sea level pressure; Z500: 500 hPa height; RT: thickness (relative topography) of the 1000–500 hPa layer; T850: 850 hPa temperature

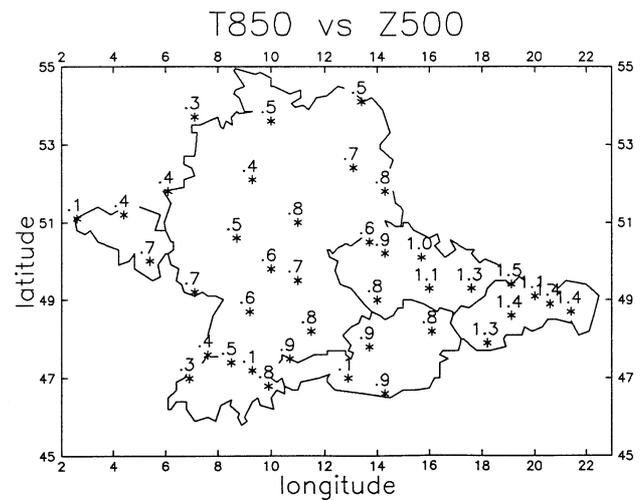


Fig. 9. Difference in RMSE (in °C) between Z500 and T850 as predictors, for pointwise regression method. Positive values indicate that T850 performs better

ification, although the improvement approaches zero (but remains positive) at the 2 southernmost stations of Davos (Stn 31) and Klagenfurt (Stn 35).

6. CONCLUSIONS

Several statistical downscaling methods and several large-scale upper air predictors were evaluated to determine their ability to specify local daily mean temperatures at 39 central European stations. The methods include canonical correlation analysis (CCA), singular value decomposition (SVD) and 3 multiple

regression models: stepwise screening of the predictor's PCs (stepwise regression), regression of the PCs without screening (full regression) and stepwise screening of the predictor's gridpoint values (pointwise regression). The performance of the methods was evaluated using cross-validation and quantified in terms of root-mean-squared error (RMSE). As large-scale predictors, 500 hPa geopotential, sea level pressure, 850 hPa temperature and 1000–500 hPa thickness are examined.

The pointwise regression proved to be the best method of those considered. The other 2 regression methods and CCA were comparable if a relatively

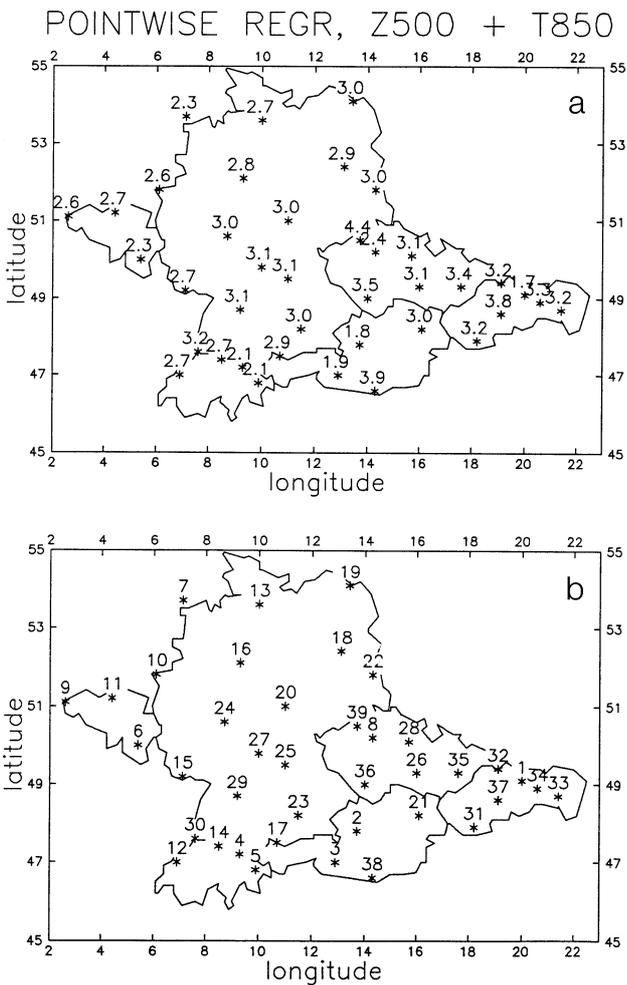


Fig. 10. (a) RMSE (in °C) for the combined T850 and Z500 predictor, pointwise regression; (b) rank of RMSE (1 denotes the best performance)

large number of the predictor's (and the predictand's in the case of CCA) PCs, including those explaining a small share of the variance, enter the statistical models. The SVD method appeared to be the worst.

Among the predictors, temperature fields result in a more accurate specification than circulation variables. The combination of 850 hPa temperature and 500 hPa heights leads to the smallest error both in a regional average and at individual stations. The accuracy of downscaling exhibits geographical variations: it is highest at elevated stations and tends to increase from the continent towards more maritime sites.

The results presented in this study were derived for winter in central and western Europe. Nevertheless, some generalization appears to be possible. I believe that the rating of the methods and predictors will apply to all regions where large-scale fields and local climate variables are related in a similar way as in the region investigated, i.e. at least in northern mid-latitudes over continents. The results may, however, be somewhat different for other seasons, especially for summer, when the effect of large-scale, low-frequency circulation features such as teleconnections on surface temperature is weaker than in winter (Huth 1997).

GCMs are likely to simulate different predictors with different accuracy as well as to be more confident in simulating large-scale features (teleconnections, modes of variability) than individual gridpoint values. Therefore, the rating of methods and predictors may be different from that presented here if the downscaling is applied to a GCM run simulating present climate conditions. Further research in this direction is needed to find out which method(s) and predictor(s) are preferable for constructing site-specific climate change scenarios using statistical downscaling of a GCM output.

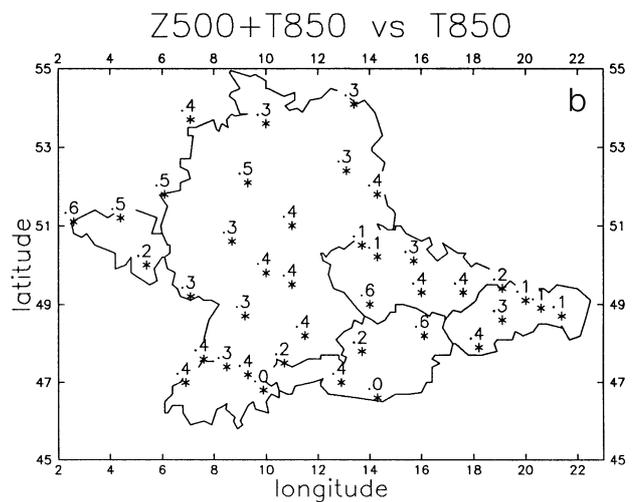
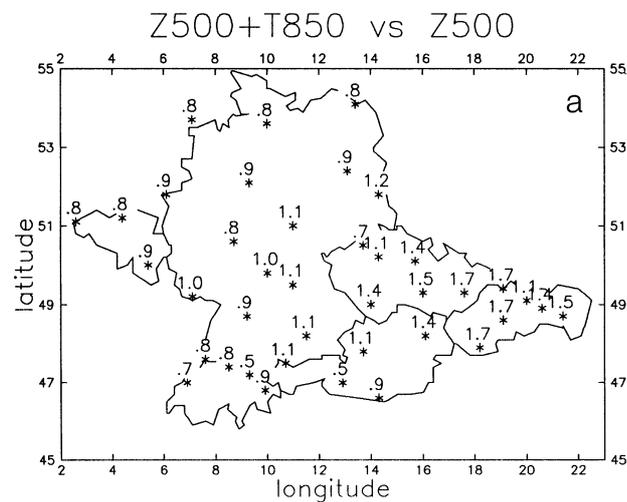


Fig. 11. Improvement of specification (difference in RMSE; in °C) by the combined T850 and Z500 predictor; over (a) Z500 and (b) T850

Acknowledgements. This study was supported by the Grant Agency of the Czech Republic under project 205/96/1670. Thanks for the provision of the data and for assistance in acquiring them go to I. Auer, M. Beniston, A. Berger, A. Kästner, M. Lapin, E. Nieplová, R. Schweitzer, C. Tricot and M. Wolek. The NCEP reanalyses were obtained from the University of Colorado at Boulder. Critical comments by 3 anonymous reviewers, Hans von Storch, Robert Wilby and Ivana Nemešová led to a considerable improvement of the manuscript. The datasets were preprocessed by N. Klimperová.

LITERATURE CITED

- Barnett TP, Preisendorfer R (1987) Origins and levels of monthly and seasonal forecast skill for United States surface air temperature determined by canonical correlation analysis. *Mon Weather Rev* 115:1825–1850
- Barnston AG, Ropelewski CF (1992) Prediction of ENSO episodes using canonical correlation analysis. *J Clim* 5: 1316–1345
- Bretherton CS, Smith C, Wallace JM (1992) An intercomparison of methods for finding coupled patterns in climate data. *J Clim* 5:541–560
- Cavazos T (1997) Downscaling large-scale circulation to local winter rainfall in north-eastern Mexico. *Int J Climatol* 17: 1069–1082
- Corte-Real J, Zhang X, Wang X (1995) Downscaling GCM information to regional scales: a non-parametric multivariate regression approach. *Clim Dyn* 11:413–424
- Crane RG, Hewitson BC (1998) Doubled CO₂ precipitation changes for the Susquehanna basin: down-scaling from the GENESIS general circulation model. *Int J Climatol* 18:65–76
- Gates WL, Henderson-Sellers A, Boer GJ, Folland CK, Kitoh A, McAvaney BJ, Semazzi F, Smith N, Weaver AJ, Zheng QC (1996) Climate models—evaluation. In: Houghton JT, Meira Filho LG, Callander BA, Harris N, Kattenberg A, Maskell K (eds) *Climate change 1995. The science of climate change*. Cambridge University Press, Cambridge, p 229–284
- Grotch SL, MacCracken MC (1991) The use of general circulation models to predict regional climatic change. *J Clim* 4: 286–303
- Hewitson B (1994) Regional climates in the GISS general circulation model: surface air temperature. *J Clim* 7:283–303
- Hewitson BC, Crane RG (1992) Regional-scale climate prediction from the GISS GCM. *Palaeogeogr Palaeoclim Palaeoecol* 97:249–267
- Heyen H, Zorita E, von Storch H (1996) Statistical downscaling of monthly mean North-Atlantic air-pressure to sea level anomalies in the Baltic Sea. *Tellus* 48A:312–323
- Huth R (1997) Potential of continental-scale circulation for the determination of local daily surface variables. *Theor Appl Climatol* 56:165–186
- Kaas E, Frich P (1995) Diurnal temperature range and cloud cover in the Nordic countries: observed trends and estimates for the future. *Atmos Res* 37:211–228
- Karl TR, Wang WC, Schlesinger ME, Knight RW, Portman D (1990) A method of relating general circulation model simulated climate to the observed local climate. Part I: Seasonal statistics. *J Clim* 3:1053–1079
- Kattenberg A, Giorgi F, Grassl H, Meehl GA, Mitchell JFB, Stouffer RJ, Tokioka T, Weaver AJ, Wigley TML (1996) Climate models—projections of future climate. In: Houghton JT, Meira Filho LG, Callander BA, Harris N, Kattenberg A, Maskell K (eds) *Climate change 1995. The science of climate change*. Cambridge University Press, Cambridge, p 285–357
- Klein WH, Walsh JE (1983) A comparison of pointwise screening and empirical orthogonal functions in specifying monthly surface temperature from 700 mb data. *Mon Weather Rev* 111:669–673
- Lau NC, Nath MJ (1994) A modeling study of the relative roles of tropical and extratropical SST anomalies in the variability of the global atmosphere—ocean system. *J Clim* 7:1184–1207
- Michaelsen J (1987) Cross-validation in statistical climate forecast models. *J Clim Appl Meteorol* 26:1589–1600
- Noguer M (1994) Using statistical techniques to deduce local climate distributions. An application for model validation. *Meteorol Appl* 1:277–287
- O’Lenic EA, Livezey RE (1988) Practical considerations in the use of rotated principal component analysis (RPCA) in diagnostic studies of upper-air height fields. *Mon Weather Rev* 116:1682–1689
- Richman MB, Lamb PJ (1985) Climatic pattern analysis of three- and seven-day summer rainfall in the central United States: some methodological considerations and a regionalization. *J Clim Appl Meteorol* 24:1325–1343
- Saunders IR, Byrne JM (1996) Generating regional precipitation from observed and GCM synoptic-scale pressure fields, southern Alberta, Canada. *Clim Res* 6:237–249
- Schubert S (1998) Downscaling local extreme temperature changes in south-eastern Australia from the CSIRO Mark2 GCM. *Int J Climatol* 18:1419–1438
- Schubert S, Henderson-Sellers A (1997) A statistical model to downscale daily temperature extremes from synoptic-scale atmospheric circulation patterns in the Australian region. *Clim Dyn* 13:223–234
- von Storch H (1995) Inconsistencies at the interface of climate impact studies and global climate research. *Meteorol Z NF* 4:72–80
- von Storch H (1999) On the use of ‘inflation’ in statistical downscaling. *J Clim* (in press)
- von Storch H, Zorita E, Cubasch U (1993) Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. *J Clim* 6: 1161–1171
- Ward MN, Navarra A (1997) Pattern analysis of SST-forced variability in ensemble GCM simulations: examples over Europe and the tropical Pacific. *J Clim* 10:2210–2220
- Wilby RL, Wigley TML (1997) Downscaling general circulation model output: a review of methods and limitations. *Progr Phys Geogr* 21:530–548
- Wilby RL, Wigley TML, Conway D, Jones PD, Hewitson BC, Main J, Wilks DS (1998) Statistical downscaling of general circulation model output: a comparison of methods. *Water Resour Res* 34:2995–3008
- Yarnal B (1985) A 500 mb synoptic climatology of Pacific north-west coast winters in relation to climatic variability, 1948–1949 to 1977–1978. *J Climatol* 5:237–252

Editorial responsibility: Hans von Storch,
Geesthacht, Germany

Submitted: October 9, 1998; *Accepted:* May 13, 1999
Proofs received from author(s): September 22, 1999