

Downscaling of Temperature Extremes for Switzerland Based on NCEP Reanalysis Data

STARDEX Deliverable D12 – Contribution from UNIBE

Dimitrios Gyalistras¹, 16. January 2004

1. INTRODUCTION

The present document describes the development and testing of a statistical procedure to downscale seasonal temperature statistics for Switzerland. This work has been undertaken in the context of the STARDEX project and follows its specifications for the rigorous comparison of downscaling procedures.

The downscaling procedure developed here uses seasonally averaged atmospheric fields as inputs. I chose to use such a coarse temporal resolution, firstly, in order to investigate the basic question in as far variations of the Swiss temperature field can be explained from average large-scale conditions. Secondly, in order to keep the downscaling models as simple and parsimonious as possible. And finally, in order to limit the input data needed to construct regional climate change scenarios from climate model runs.

2. DATA & METHODS

The statistical downscaling procedure was based on Canonical Correlation Analysis (CCA) in the space spanned by the first few Empirical Orthogonal Functions (EOFs) of the predictor (independent variables) and predictand (dependent variables) fields (VON STORCH & ZWIERS, 1999). In the present study I used varying numbers of predictor fields, according to the method proposed by GYALISTRAS *et al.* (1994).

The predictors were given by large-scale fields for sea-level pressure (slp) and 4 further atmospheric variables (see Table 1) at the 1000, 850, 700, 500 and 300 hPa levels. All fields had a 2.5° x 2.5° longitude/latitude resolution and were derived from daily NCEP/NCAR reanalysis data sets which were downloaded from the website of the Climatic Research Unit, Norwich. The fields were evaluated for 6 different sectors of varying size, which were all centered over the region of the European Alps (Table 2).

As predictands I considered 6 seasonal statistics from 21 Swiss locations (Table 3). The data were taken from the FIC data set. The statistics were derived from daily minimum and maximum temperatures (Table 4) using the STARDEX Diagnostic Extremes Indices Software.

¹ Climatology and Meteorology, Institute of Geography, University of Berne, Hallerstr. 12, CH-3012 Berne, Switzerland

Table 1: Predictor variables used to fit the downscaling models.

Predictor Variable	Identifier	Unit
Pressure at sea-level	slp	hPa
Geopotential height	gph	m
Temperature	temp	K
Specific humidity	shum	g/kg
Relative humidity	rhum	%

Table 2: Used large-scale sectors for the predictor variables.

Sector Identifier	Longitudes (°)		Latitudes (°N)		#GP
	Min.	Max.	Min.	Max.	
D1	-50	70	15	75	1225
D2	-40	60	20	70	861
D3	-30	50	25	65	561
D4	-20	40	30	60	325
D5	-10	30	35	55	153
D6	0	20	40	50	45

#GP: number of gridpoints on the used 2.5° x 2.5° longitude-latitude grid.

Table 3: Swiss locations considered.

Name	Longitude (°E)	Latitude (°N)	Elevation (m.a.s.l.)
Altdorf	8.63	46.86	451
Arosa	9.68	46.78	1840
Bad Ragaz	9.50	47.01	496
Basel-Binningen	7.58	47.55	316
Bern-Liebefeld	7.41	46.93	570
Chateau-d'Oex	7.15	46.48	985
Chur-Ems	9.53	46.86	555
Davos	9.85	46.81	1590
Geneve-Cointrin	6.13	46.25	420
Glarus	9.06	47.03	470
Locarno-Monti	8.78	46.16	379
Lugano	8.96	46.00	273
Luzern	8.30	47.03	456
Meiringen	8.16	46.73	595
Montana	7.48	46.31	1495
Montreux-Clarens	6.90	46.45	405
Neuchatel	6.95	47.00	487
Saentis	9.35	47.25	2490
Schaffhausen	8.61	47.68	437
St. Gallen	9.40	47.43	779
Zuerich	8.56	47.38	556

Table 4: Predictand variables used to fit the downscaling models.

Predictand Variable	Identifier	Unit
Seasonal mean of daily maximum temperature	txav	°C
90th percentile of daily maximum temperature in a given season	tmax90p	°C
Heat wave duration (Tmax > daily long-term mean for at least 6 consecutive days)	144HWDI	# days
Seasonal mean of daily minimum temperature	tnav	°C
10th percentile of daily daily minimum temperature in a given season	tmin10p	°C
Number of frost days (Tmin < 0 °C)	125Fd	# days

I conducted 4 types of downscaling experiments which are summarized in Table 5. In all experiments the predictors used to fit a particular downscaling model were given by one or several large-scale fields for a given season and the predictands by one of the 6 temperature statistics (Table 4) at all 21 Swiss locations (Table 3) for that same season.

For all experiments the numbers of EOFs that were retained to perform the CCA were given by as many EOFs were needed to explain 80% and 95% of the total variance of all predictors and predictands, respectively. Estimation of the predictands from independent data (in the context of model validation) was done using all canonical modes which showed a squared canonical correlation coefficient $\geq 10\%$.

Table 5: Overview of downscaling experiments.

Experiment Type	Predictors	Years Used for Model Fitting	Trend Removal	Validation Method / Validation Period
I	One predictor field (one variable at a single pressure level and for a single large-scale sector)	1958-2000	yes	C / 1958-2000
II	Selected combination of several predictor fields	1958-2000	yes	C / 1958-2000
III	Same as in II	1958-1978 \cup 1994-2000	no	C / 1958-1978 \cup 1994-2000
IV	Same as in II	1958-1978 \cup 1994-2000	no	P / 1979-1994

C: leave-one-out cross-validation; P: prediction using independent data.

Prior to fitting of the CCA models the data were prepared as follows: First, the annual cycles were removed from all predictor and predictant time series by subtracting from each seasonal value the long-term mean of the respective season. The years used to determine the long-term mean were given by the model fitting period (Table 5). Second, the predictor and predictand data sets were detrended by first computing for each season and timeseries the respective linear 43-year trend and then subtracting it from the timeseries (experiments of Type I and II only). Finally, in order to account for the decreasing grid-cell size with increasing latitude, the anomalies for the predictor variables were weighted with the square root of their latitude cosine. The predictand time series were always scaled to unit variance, such that all 21 regional time series entered the subsequent analysis with the same weight.

Each fitted single model was tested either by means of leave-one-out cross-validation (Experiment Types I-III), or by applying the model to independent predictor data from the period 1979-1994 (Experiment Type IV). The original and the statistically downscaled time series were compared using 5 different statistics (Table 6). Some critical values for the used correlation statistics are shown in Table 7.

The experiments of Type I were used to assess the prediction potential of individual atmospheric fields. They involved the fitting and cross-validation of $((1 \text{ predictor variable} \times 1 \text{ level}) + (4 \text{ predictor variables} \times 5 \text{ levels})) \times 6 \text{ sectors} \times 6 \text{ predictand variables} \times 4 \text{ seasons} = 3024$ individual downscaling models.

Table 6: Statistics used to evaluate the performance of the downscaling models.

Statistic	Identifier	Definition
Bias (Mean error)	BIAS	(to be inserted)
Root mean square error with bias removed	RMSEBR	(to be inserted)
Squared simple correlation coefficient	R2	(to be inserted)
Spearman rank correlation coefficient	SRC	(to be inserted)
Reduction of error	RE	(to be inserted)

Table 7: Critical values of correlation statistics used to validate the downscaling models.

Statistic	Sample Size	Critical Values		
		$\alpha(2)=0.1$	$\alpha(2)=0.05$	$\alpha(2)=0.01$
R2	43 (Expers. I and II)	6%	9%	14%
	28 (Exper. III)	9%	13%	21%
	15 (Exper. IV)	17%	23%	37%
SRC	43 (Expers. I and II)	0.25	0.30	0.39
	28 (Exper. III)	0.32	0.38	0.48
	15 (Exper. IV)	0.45	0.52	0.60

$\alpha(2)$: significance level for two-tailed testing of the null hypothesis “Statistic = 0”.

The experiments of Type II explored the simultaneous use of several predictor fields. The large number of fields precluded a systematic investigation of all possible combinations of predictors. Therefore I attempted to determine optimal combinations of predictor fields based on the results from the Type I experiments as follows:

First, for each predictor variable, level, predictand variable and season I determined the sector (Table 2) which yielded the highest cross-validated R2. If this R2 was $\geq 10\%$ (cf. Table 6), the given field was retained as a candidate predictor field, otherwise it was not considered any further. This yielded N_{c1} candidate fields. In a second step, these fields were sorted by their R2 in descending order, and a series of N_{c1} downscaling models were fitted, where the n-th model used the $n \leq N_{c1}$ first candidate fields from the sorted list. In a third step, a second set of N_{c2} candidate fields was compiled by determining all fields that gave an increase of the cross-validated R2 when they had been added as predictors in the previous step. Finally for these new candidate fields again a series of $n \leq N_{c2}$ downscaling models with increasing numbers of predictor fields were fitted, and the best model in terms of R2 was selected as the “best performing” model.

The resulting “best” models, one per predictand variable and season, were then used to conduct the Type III and IV experiments. These experiments served to cross-check the cross-validation results from experiment Types I and II by using different years and methods for the fitting and testing of the downscaling models (see Table 5).

3. RESULTS

The R² results from the Type I experiments for the regional variables related to the daily maximum temperatures are summarized in Figure 1. R²-values above 0.3 (highlighted in orange or red) were obtained for txav in all seasons and for tmax90p in winter and summer. In all other cases the downscaling models performed rather poorly. The best predictor field varied with predictand and season. Generally, the lower-tropospheric temperature fields for the sectors D4-D6 yielded the best results.

The cross-validation results for the regional statistics related to the daily minimum temperatures are summarized in Figure 2. Here, R²-values above 0.3 were obtained in all seasons for tnav, and for tmin10p and 125Fd in winter and spring. The best predictor fields were again in most cases the lower-tropospheric temperature fields for the sectors D4-D6, but the dynamical (slp and gph) and moisture fields (rhum, shum) also showed in several cases reasonable skill.

Table 8 summarizes the results of the Type II experiments. It can be seen that in 50% of the cases the use of several predictor fields (#F > 1 in Table 8) lead to improved downscaling models as opposed to the use of individual fields only. Generally, in cases where the prediction skill was low for the individual fields, such as for the predictands 144 HWDI and 125Fd, no combination of fields could be found that improved the results.

The most frequently used predictor fields were the temperature at the 1000 hPa level (occurrence in 14 downscaling models), followed by the 850 and 700 hPa temperatures (7 occurrences each), the specific humidity at 1000 hPa (6 occurrences) and sea-level pressure (4 occurrences). The most frequently used sector was sector D6 (33 occurrences), followed by sectors D5 (14 occurrences), D2 (6 occurrences) and D3 (4 occurrences).

The validation statistics from all 4 types of downscaling experiments are juxtaposed in Table 9. The BIAS was found to be generally small. The experiments of Type IV tended to give somewhat higher values than the other experiments.

The mean RMSEBR from all experiments and seasons for txav was 0.8 °C, and for tnav it was 0.6 °C. Larger mean RMSEBR values were obtained for the percentile statistics tmax90p (1.3 °C) and tmin10p (1.4 °C). For 144HWDI and 125Fd the mean RMSEBR was 4.4 d and 4.7 d, respectively.

The validation statistics R² and SRC showed generally high values for txav and tnav: the mean SRC from all experiments and seasons was in both cases 0.78. Less good results were obtained for 125Fd (average SRC: 0.43), tmin10p (0.39), and tmax90p (0.37). The tests suggested no skill for 144HWDI (-0.18).

According to the last statistic shown in Table 9, N(RE>0.5), the downscaling models were able to reconstruct txav and tnav on average over all experiments and seasons with RE > 0.5 at 12.4 and 13.3 of the 21 stations, respectively. Good results (here defined as RE > 0.5 at more

than 10 stations) were also obtained for $t_{\max 90p}$, $t_{\min 10p}$ and 125Fd in winter. Note, the generally good results obtained for 125Fd in summer are not very representative, since in the warm season for most locations and years 125Fd was zero.

Table 9 also gives an indication of the improvement obtained thanks to the use of several predictor fields (cf. Table 8). The average increases in R^2 that were obtained for the Type II multi-predictor models as compared to the respective Type I models was 0.05 for $t_{\max 90p}$ and t_{nav} , 0.04 for t_{xav} and $t_{\min 10p}$, and 0.03 for 125Fd.

As can be seen from Table 9, the best results were generally obtained for winter. Interestingly, the largest individual increases in R^2 due to the use of several predictor fields (Type II vs. Type I experiments) were also found for the cold season. The found gain in wintertime R^2 was +0.07 for t_{nav} , +0.06 for $t_{\max 90p}$ and $t_{\min 10p}$, and +0.05 for t_{xav} .

txav							tmax90p							144HWDI																	
DJF		D1	D2	D3	D4	D5	D6	MAM		D1	D2	D3	D4	D5	D6	JJA		D1	D2	D3	D4	D5	D6	SON		D1	D2	D3	D4	D5	D6
slp							slp							slp							slp										
gph.1000							gph.1000							gph.1000							gph.1000										
gph.850							gph.850							gph.850							gph.850										
gph.700							gph.700							gph.700							gph.700										
gph.500							gph.500							gph.500							gph.500										
gph.300							gph.300							gph.300							gph.300										
temp.1000							temp.1000							temp.1000							temp.1000										
temp.850							temp.850							temp.850							temp.850										
temp.700							temp.700							temp.700							temp.700										
temp.500							temp.500							temp.500							temp.500										
temp.300							temp.300							temp.300							temp.300										
shum.1000							shum.1000							shum.1000							shum.1000										
shum.850							shum.850							shum.850							shum.850										
shum.700							shum.700							shum.700							shum.700										
shum.500							shum.500							shum.500							shum.500										
shum.300							shum.300							shum.300							shum.300										
rhum.1000							rhum.1000							rhum.1000							rhum.1000										
rhum.850							rhum.850							rhum.850							rhum.850										
rhum.700							rhum.700							rhum.700							rhum.700										
rhum.500							rhum.500							rhum.500							rhum.500										
rhum.300							rhum.300							rhum.300							rhum.300										

Figure 1: Cross-validation results for the prediction of seasonal statistics of daily maximum temperatures from individual atmospheric predictor fields. Shown are the average squared correlations (R^2) between the original and cross-validated time series ($n = 43$) from 21 Swiss locations (see Table 3). txav: seasonal mean of daily maximum temperatures; tmax90p: 90th percentile of daily maximum temperatures; 144HWDI: heat wave duration index; slp: sea-level pressure; gph.x: geopotential height of pressure level x; temp.x, shum.x, rhum.x: temperature, specific humidity and relative humidity at pressure level x; D1-D6: sectors used to define the extent of the predictor field (see Table 2); --: $R^2 < 0.1$; white fields: $0.1 \leq R^2 < 0.15$; yellow fields: $0.15 \leq R^2 < 0.3$; orange fields: $0.3 \leq R^2 < 0.5$; red fields: $0.5 \leq R^2 < 1$.

Table 8: Overview of the found best combinations of predictor fields.

Predictand	#F	Predictor Level Season	slp		temp				shum				rhum				
			gph		700	500	300	1000	850	700	500	300	1000	850	700	500	
txav	4	DJF					D6	D5	D5		D4						
	4	MAM					D6			D5					D2		D5
	3	JJA					D6								D6		D5
	3	SON					D6	D6	D6								
tmax90p	6	DJF		D2	D4		D6		D5	D6		D4					
	1	MAM							D6								
	4	JJA			D6		D6								D6		D5
	1	SON		D6													
144HWDI	1	DJF		D5													
	1	MAM								D6							
	1	JJA												D2			
	1	SON		D3													
tnav	6	DJF					D6	D6	D6		D5		D5	D6			
	1	MAM					D6										
	3	JJA					D6	D6									D5
	4	SON					D6	D6			D5	D5					
tmin10p	4	DJF			D3		D6	D5	D3								
	1	MAM					D6										
	6	JJA					D2	D1	D1		D6				D2	D2	
	1	SON		D6													
125Fd	3	DJF		D3							D6	D6					
	1	MAM					D6										
	1	JJA					D6										
	1	SON		D6													

#F: number of predictor fields in the found best performing downscaling model; slp: sea-level pressure; temp: temperature; shum: specific humidity; rhum: relative humidity; txav: seasonal mean of daily maximum temperatures; tmax90p: 90th percentile of daily maximum temperatures; 144HWDI: heat wave duration index; tnav: seasonal mean of daily minimum temperatures; tmin10p: 10th percentile of daily minimum temperatures; 125Fd: number of frost days; DJF: December–February; MAM: March–May; JJA: June–August; SON: September–November; D1–D6: sectors used to define the extent of the predictor field (see Table 2). Bold face denotes the best performing individual predictor field for the given predictant and season (cf. Figs. 1 and 2).

Table 9: Summary of results from the validation of the downscaling models. The shown statistics refer to 21 Swiss locations (Table 3) and 4 different types of downscaling experiments (Table 5).

Predictand	Statistic Exper. Season	M(BIAS)				M(RMSEBR)				M(R2)				M(SRC)				N(RE>0.5)			
		I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
txav	DJF	-.02	-.06	-.08	.08	0.8	1.1	1.2	1.1	.69	.74	.74	.85	.81	.84	.81	.90	17	20	20	20
	MAM	.00	.00	.00	.46	0.7	0.7	0.8	0.6	.64	.69	.64	.80	.77	.80	.78	.85	14	14	12	11
	JJA	-.01	-.01	-.01	.01	0.6	0.6	0.8	0.5	.66	.67	.54	.76	.76	.77	.68	.85	13	14	7	2
	SON	.01	.00	.01	.32	0.7	0.6	0.7	0.5	.57	.60	.58	.66	.73	.75	.71	.75	9	11	13	2
tmax90p	DJF	-.03	-.05	-.04	.35	1.2	1.2	1.4	1.3	.53	.59	.60	.64	.70	.75	.74	.69	0	15	15	13
	MAM	.02	.02	.01	.30	1.4	1.4	1.4	1.4	.20	.20	.08	.39	.34	.34	.21	.56	0	0	0	0
	JJA	-.01	-.02	-.02	.26	0.9	0.9	1.1	0.8	.42	.45	.32	.60	.61	.65	.51	.74	0	1	0	2
	SON	.01	.01	.01	-.31	1.7	1.7	1.8	1.6	.21	.21	.27	.07	-.28	-.28	-.49	.09	0	0	0	0
144HWDI	DJF	.04	.04	.05	-.62	4.9	4.9	4.2	6.1	.20	.20	.04	.03	-.32	-.32	-.27	-.04	0	0	0	1
	MAM	.00	.00	.01	1.87	5.4	5.4	5.9	4.6	.15	.15	.18	.05	-.19	-.19	-.17	.05	0	0	0	0
	JJA	-.07	-.07	-.09	.98	4.4	4.4	4.3	3.0	.12	.12	.19	.09	-.22	-.22	-.31	-.06	0	0	0	4
	SON	.00	.00	.00	-.71	3.3	3.3	3.2	3.4	.22	.22	.10	.05	-.27	-.27	-.05	-.08	0	0	1	3
tnav	DJF	-.04	-.03	-.04	.04	0.8	1.0	0.9	0.9	.62	.69	.72	.75	.76	.83	.78	.90	13	19	18	19
	MAM	.01	.01	.00	.04	0.6	0.6	0.7	0.5	.67	.67	.59	.74	.82	.82	.77	.85	16	16	8	15
	JJA	.00	.00	-.01	.11	0.4	0.4	0.6	0.4	.58	.63	.61	.64	.74	.79	.73	.79	7	9	13	9
	SON	.00	.00	-.03	-.04	0.6	0.6	0.8	0.5	.57	.61	.54	.61	.72	.75	.70	.69	11	15	14	10
tmin10p	DJF	.00	-.02	.04	.30	1.6	1.5	1.7	1.2	.47	.52	.48	.68	.61	.66	.58	.83	3	6	1	12
	MAM	.01	.01	.01	-.46	1.4	1.4	1.4	1.5	.40	.40	.46	.40	.63	.63	.68	.60	0	0	2	1
	JJA	.01	-.01	.02	.04	0.9	0.9	1.0	0.9	.24	.27	.10	.17	.48	.49	.29	.39	0	0	0	0
	SON	.00	.00	.00	.32	1.6	1.6	1.6	1.7	.18	.18	.17	.11	-.24	-.24	-.34	.13	0	0	0	0
125Fd	DJF	.23	.09	.70	2.35	6.5	6.0	8.5	6.8	.49	.52	.37	.45	.66	.70	.54	.66	8	11	7	3
	MAM	-.03	-.03	-.06	.85	5.5	5.5	5.7	5.7	.38	.38	.42	.42	.61	.61	.61	.61	0	0	1	1
	JJA	.00	.00	.00	-.22	0.6	0.6	0.6	0.6	.27	.27	.20	.07	.64	.64	.71	.74	15	15	15	16
	SON	-.02	-.02	-.02	-.15	5.7	5.7	6.1	5.1	.25	.25	.22	.07	-.29	-.29	-.31	.08	0	0	0	0

M(): mean value from all 21 locations; BIAS: mean error; RMSEBR: root mean square error with bias removed; R2: squared simple correlation coefficient; SRC: Spearman rank correlation coefficient; N(RE>0.5): number of climate stations with reduction of error > 0.5; I-IV: Types of downscaling experiments; txav: seasonal mean of daily maximum temperatures; tmax90p: 90th percentile of daily maximum temperatures; 144HWDI: heat wave duration index; tnav: seasonal mean of daily minimum temperatures; tmin10p: 10th percentile of daily minimum temperatures; 125Fd: number of frost days; DJF: December–February; MAM: March–May; JJA: June–August; SON: September–November. Bold face denotes for R2 and SRC values significantly different from zero at the 99% confidence level, for N(RE>0.5) it is used to highlight numbers of stations > 10.

4. REFERENCES

- GYALISTRAS, D., VON STORCH, H., FISCHLIN, A. & BENISTON, M. (1994). Linking GCM-simulated climatic changes to ecosystem models: case studies of statistical downscaling in the Alps. *Clim. Res.* 4(3): 167-189.
- VON STORCH, H. & ZWIERS, F. W. (1999). *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge, United Kingdom, 484 pp.