Project No. 037005

CECILIA

Central and Eastern Europe Climate Change Impact and Vulnerability Assessment

Specific targeted research project
1.1.6.3.I.3.2: Climate change impacts in central-eastern Europe

D 3.1: Observed data for SDS models building, output localization, upscaling, and model validation: station data and reanalysis data for SDS predictors

Due date of deliverable: 1st June 2007
Actual submission date: 29th February 2008

Start date of project: 1st June 2006 Duration: 36 months

Lead contractor for this deliverable: Czech Hydrometeorological Institute (CHMI)

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Table of content:

1 CENTRAL EUROPEAN DATASET, DATA PREPARATION 3

1.4 DATA QUALITY CONTROL 5

2 CALCULATION OF STATION TECHNICAL SERIES AND GRIDDED DATASET.... 7

3 FINDING THE BEST SETTINGS FOR STATION TECHNICAL SERIES AND GRIDDED DATASETS CALCULATION.... 9

3.1 MAXIMUM TEMPERATURE 11
3.2 MINIMUM TEMPERATURE.... 14
3.3 RELATIVE HUMIDITY 17
3.4 PRECIPITATION 21
3.5 SUNSHINE DURATION.... 24

4 SUMMARY.... 28

1 Central European dataset, data preparation

All the tasks in WP3 need observed data for the development of the methods and their calibration. Therefore, the original idea was to create a common dataset for the development and calibration of the methods. After discussions at the Bucharest (2006) and Semmerning (2007) meetings, it was decided that the common dataset will cover the area along the boundaries of the Czech Republic, Austria, Slovakia, and Hungary. The main intention was that the majority of the impact target areas in the central European region be covered by this dataset. Another leading thought in this decision was the idea that it would be easier to get the meteorological data from the meteorological services for only relatively small parts of the countries than for the large parts or even whole countries. The area covered in individual countries by the dataset can be seen in Fig.1-2. It includes:

- in the Czech Republic: southern and south-eastern part, consisting of regions České Budějovice, Highlands (Vysočina), South Moravian, Zlín, and small southern parts of Central Bohemian;
- in Austria: federal states Lower Austria, Upper Austria, Vienna and Burgenland;
- in Slovakia: western part, consisting of regions Bratislava, Trnava, Nitra, Trenčín, and Banská Bystrica;
- in Hungary: regions Györ-Moson-Sopron and Komárom-Esztergom.

The central European area covers the following impact target areas: agriculture – Lower Austria (AT), southern Moravia (CZ), Danube lowlands (SK); forestry –southern central Slovakia (SK); hydrology – Dyje and upper Vltava catchments (CZ), Hron catchment (SK).

It was agreed that the dataset will be composed of daily data. Variables available in dataset are given in Table 1. Potential evapotranspiration was not calculated in the end since several ways for its calculation exist and it can be calculated from the available elements by individual users.

Table 1. Meteorological elements available in the common dataset.

Abbreviation	Description	Unit
TMI	Maximum temperature	°C
TMA	Minimum temperature	°C
H	Relative humidity	%
SRA	Precipitation	mm
SSV	Sunshine duration	h

The following comments on the variables selected and not selected should be made:

1. Daily mean temperature was not included because of regional differences in its calculation and change in the practice in its calculation in Austria in early 1970's, which could induce an inhomogeneity in the time series and inconsistency along the state boundaries.
2. Relative humidity was selected, and not another measure of atmospheric moisture that is not affected by daily temperature cycle, such as specific humidity, because some of the impact models require just relative humidity as their input.

3. Wind speed and direction will not be subject to gridding and creating the technical series because of the necessity of working separately with the two wind components, which would cause considerable complications, making the resultant technical series doubtful and unreliable.
4. Potential evaporation can be calculated from other variables (air temperature, relative humidity, …) available as the technical series (except wind speed).
5. Solar radiation can be easily approximated from the sunshine duration data.

The dataset covers the period 1961-2000. Even incomplete time series were allowed to enter the database. The data were prepared and provided by the following partners: CHMI for the Czech Republic (about 90 stations), FRI for Slovakia (40 stations), BOKU for Austria (30 stations), and OMSZ for Hungary (10 stations). To facilitate the procedure of getting data, an official letter signed by the project coordinator was sent to the director of the Hungarian Meteorological Service.

The data policy of some of the involved meteorological services does not allow a distribution of raw station data. Therefore, it was decided to create technical series from the station data available, which will further be distributed among the project participants. The technical series of two kinds are being constructed: (i) gridded datasets covering the area where station data are available; this will be a primary dataset; (ii) station technical series, which will have the advantage over the raw data in their better homogeneity and completeness.

In the CECILIA central European domain, about 150 climatological stations are available – see Fig. 1, in comparison with 832 grid points adopted from RCM ALADIN-CLIMATE/CZ – see Fig. 2. Number of stations available for individual countries and meteorological elements is given in Table 2.

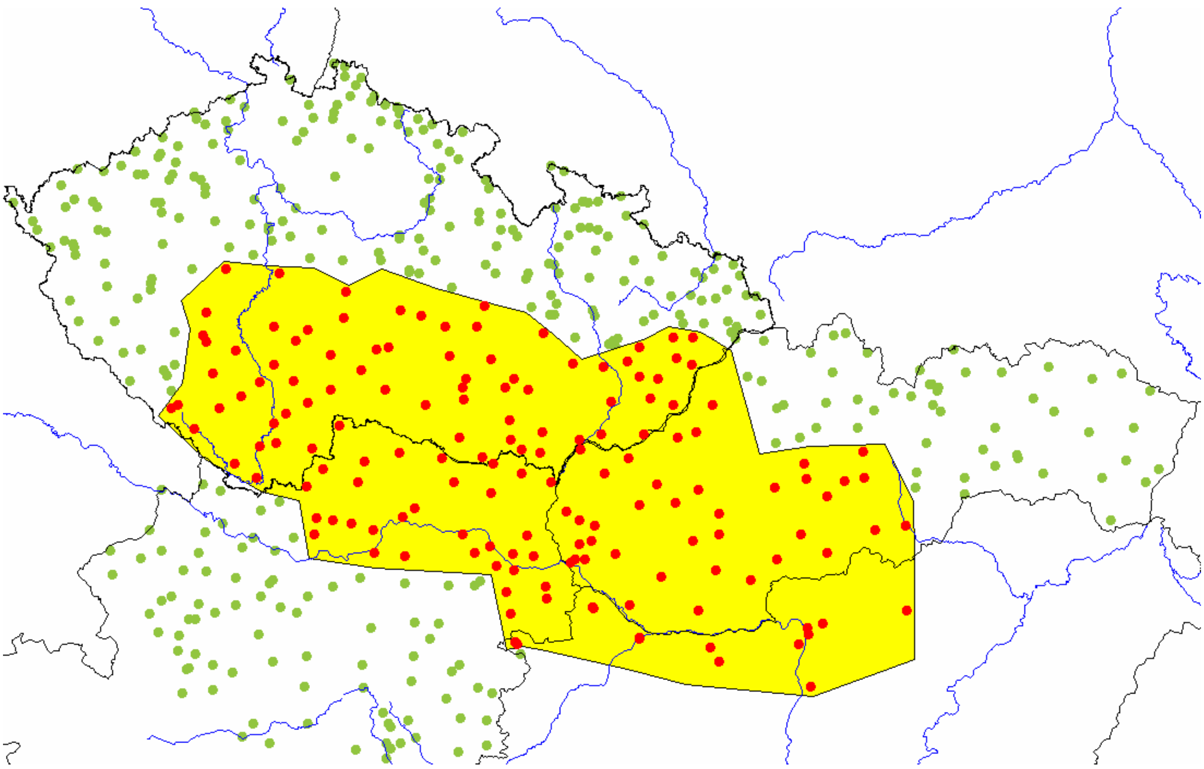


Fig. 1. CECILIA central European domain (yellow area) with available climatological stations.

Table 2. Number of stations available per individual country (AT – Austria, CZ – Czech Republic, SK – Slovakia, HU - Hungary) and meteorological element (see Table 1 for explanatory notes).

Country	Element				
	TMA	TMI	SRA	SSV	H
AT	33	33	35	11	30
CZ	90	90	90	68	91
SK	39	39	39	39	40
HU	11	11	11	6	11
total	173	173	175	124	172

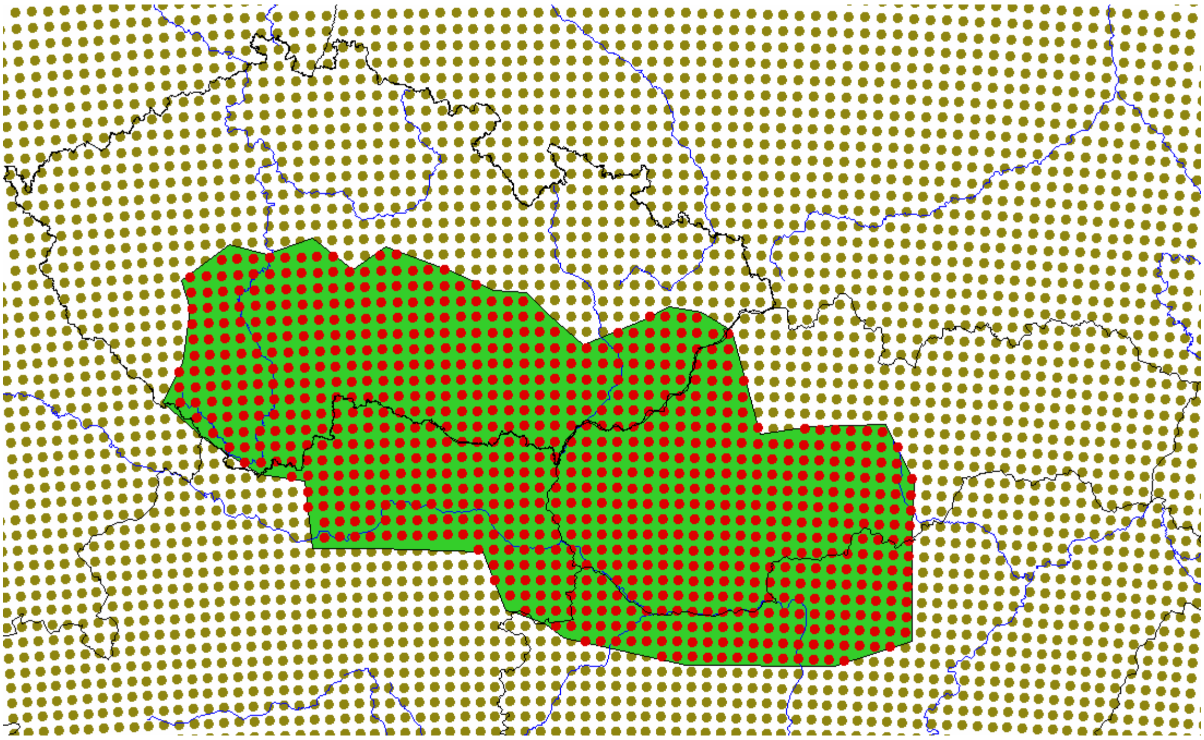


Fig. 2. Grid points of ALADIN-CLIMATE/CZ available in the CECILIA central European domain (green area).

1.4 Data quality control

Before station technical series and gridded dataset calculation, raw station data were subject of thorough quality control using AnClim and ProClimDB software of Petr Stepanek, CHMI (Štěpánek, 2007, more details can be found in documentation of the software, www.climahom.eu). Tools available in the software were designed in such a way that they can be applied for automated finding of errors in datasets. The outliers were found by combination of several methods: percentage of neighbour stations which are significantly (p=0.05) different from the base station (found from standardized differences between neighbours and base station, limit value: more than 75%); difference of base station value and median calculated from values of neighbours standardized to base station altitude (using linear regression) divided by standard deviation of base station, expressed as CDF of normal distribution (limit value: more than 0.95), coefficient (multiple) of distance of the base station value above (below) q75 (q25) quartile calculated from the standardized (to base station altitude) values of neighbour stations (higher the value, the more similar neighbour values are compared to base station value, limit value: coefficient more than 5); difference of expected value (details of its calculation are given in the chapter 2 dealing with technical series creation) and median calculated from original values of neighbour stations divided by standard deviation of base station values (expressed as CDF of normal distribution, the value should be low otherwise it indicates that expected value calculation is probably wrong, limit value: less than 0.75). The calculation was carried out for each meteorological element and individual day separately.

Fig. 3 shows example of found suspicious values. Such values were found in all available raw datasets (Austria, Czech Republic, Slovakia and Hungary, numbers are given in Table 2b) and were sent to individual data providers. Unfortunately even if only the most evident errors have been sent (tens of values per each available country dataset), it is unrealistic to use corrected / verified values back in this deliverable due to time demand for such a task. The combination of methods and used limits were set in the way that found outliers can be regarded as evident errors (except for summer precipitation, the temporal and spatial variability of this element have not allowed us to make reliable conclusions like in case of other elements), that is why such values were rejected from further processing (they were replaced with code for missing value).

	Station				Suspected val	Expected value		Neighbouring stations				
Original nai	ID	YEAR	MONTH	DAY	ST_BASE	EXPECT_VAL	REMARK	ST_1	ST_2	ST_3	ST_4	ST_5
TMIN	10000				492.0		Altitude	648.0	480.0	695.0	810.0	842.0
TMIN	9900						st_1, di	22.0				
TMIN	13301						st_2, di		43.1			
TMIN	9811						st_3, di			50.1		
TMIN	15900						st_4, di				56.9	
TMIN	16000						st_5, di					62.7
TMIN	**10000**	**1961**	**3**	**18**	**8.0**	**-1.8**		**-2.9**	**-1.7**	**-1.5**	**-1.8**	**-2.0**
TMIN	Absolute numbers	[illegible]	[illegible]	**22**	**10.0**	**2.9**		**1.1**	**3.2**	**3.8**	**3.1**	**4.0**
TMIN	**10000**	**1962**	**4**	**23**	**13.0**	**0.9**		**0.1**	**1.3**	**1.8**	**0.6**	**2.8**
TMIN	**10000**	**1962**	**5**	**22**	**7.0**	**1.1**		**1.3**	**0.8**	**2.9**	**0.7**	**1.4**
TMIN	**10000**	**1962**	**7**	**21**	**13.0**	**8.4**		**7.4**	**8.6**	**9.1**	**8.5**	**9.0**
TMIN	**10000**	**1963**	**5**	**30**	**10.6**	**3.3**		**3.1**	**3.3**	**4.1**	**2.7**	**3.2**
TMIN	**10000**	**1964**	**1**	**5**	**-10.0**	**-18.5**		**-19.7**	**-18.4**	**-16.5**	**-16.4**	**-17.0**
TMIN	**10000**	**1968**	**4**	**15**	**5.0**	**-0.6**		**-1.3**	**-0.5**	**0.6**	**-1.4**	**-1.4**
TMIN	**10000**	**1975**	**4**	**6**	**9.4**	**4.0**			**4.2**	**2.1**	**2.1**	**2.2**
TMIN	**10000**	**1976**	**2**	**8**	**-1.2**	**-8.9**			**-9.0**	**-7.9**	**-6.9**	**-8.3**
TMIN	10000				492.0		Altitude	648.0	790.0	480.0	670.0	695.0
TMIN	9900						st_1, di	22.0				
TMIN	10200						st_2, di		37.2			
TMIN	13301						st_3, di			43.1		
TMIN	16101						st_4, di				44.5	
TMIN	9811						st_5, di					50.1
TMIN	**10000**	**1975**	**5**	**11**	**12.3**	**5.6**			**3.6**	**5.7**	**2.4**	**4.1**
TMIN	**10000**	**1975**	**7**	**29**	**13.3**	**9.1**			**7.8**	**9.2**	**8.2**	**9.2**
TMIN	**10000**	**1978**	**7**	**16**	**13.1**	**8.4**		**8.3**	**8.2**	**8.6**	**7.3**	**9.6**
TMIN	**10000**	**1980**	**7**	**30**	**11.2**	**16.2**		**15.0**	**13.9**	**16.5**	**14.0**	**15.9**
TMIN	**10000**	**1986**	**8**	**19**	**10.0**	**15.3**		**14.1**	**12.2**	**15.0**	**14.0**	**13.8**
TMIN	**10000**	**1992**	**7**	**27**	**10.0**	**15.2**		**14.3**	**13.9**	**15.4**	**14.4**	**15.0**
TMIN	**10000**	**1993**	**3**	**1**	**2.8**	**-3.0**		**-4.8**	**-5.5**	**-2.9**	**-4.7**	**-4.7**

Fig. 3. Example of found suspicious values in raw dataset (yellow column) compared to values of five neighbour stations (five rightmost columns).

Table 2b. Number of suspicious values (evident errors) per individual country (AT – Austria, CZ – Czech Republic, SK – Slovakia, HU - Hungary) and meteorological element (see Table 1 for explanatory notes).

Absolute numbers						Relatively per number of stations					
Country	Element					Country	Element				
	TMA	TMI	SRA	SSV	H		TMA	TMI	SRA	SSV	H
AT	28	74	195	309	118	AT	0.85	2.24	5.57	28.09	3.93
CZ	36	157	489	910	498	CZ	0.40	1.74	5.43	13.38	5.47
SK	8	37	72	975	346	SK	0.21	0.95	1.85	25.00	8.65
HU	1	10	33	374	201	HU	0.09	0.91	3.00	62.33	18.27
total	73	278	789	2568	1163	total	0.42	1.61	4.51	20.71	6.76

The prepared datasets were used further for station technical series and gridded dataset calculation. Originally it was planned to perform also homogenization of raw station data, after finding outliers, unfortunately such a task is beyond current time possibilities of CHMI. Nonetheless, the way of calculation of station technical series and gridded dataset makes it possible to presume that possible inhomogeneities in input station data are suppressed in final series so that the series can be regarded as homogeneous. The process of the calculation is discussed in further chapters.

2 Calculation of station technical series and gridded dataset

Several methods can be applied to calculate values of a given meteorological element at a certain geographical position (e.g. at a grid point). Inverse distance weighting belongs among the simpler ones but still it gives good results even compared to modern geostatistical methods such as krigging, co-krigging, universal krigging (Kliegrova et al., 2007). The weights can be applied as inverse distances or correlations (Isaaks and Srivastava 1989), possibly powered, to control lower or higher spatial correlations of a given meteorological element. Applying geostatistical methods on time series is not an easy task (mainly due the computational demands) but some attempts that combine time and spatial analysis already exist (e.g. Szentimrey 2002; Květoň and Tolasz 2003) and such methods have started to be used more widely in recent time.

For purposes of this deliverable, daily series of several meteorological elements for hundreds of locations (grid points) had to be calculated. Utilizing GIS environment for such a task would be advantageous in a sense of choosing from variety of interpolation methods. Nonetheless current GIS environments (e.g. ArcMap, ESRI ArcView, ArcGIS) are not designed for comfortable retrieving information for time series (calculation for each time step). That is why it was needed to create own tool which would be automated enough. For the computation, software ProClimDB (Štěpánek, 2007) was extended to include required functionality. The created software is freely available.

After quality control (see previous chapter), a particular grid point (station location) technical series of daily values were calculated from up to 6 neighbouring (nearest) stations within distance of 300 km, with allowed maximum difference in altitude of 500 m. Before applying inverse distance weighting, the neighbour stations data were standardized with respect to the base grid point (station location) altitude. The standardization was carried out by means of linear regression, dependence of values of a particular meteorological element on altitude, for each day individually, regionally. Each standardized value was checked if it does not differ too much from an original value (if CDF does not exceed 0.99, in such a case linear regression was regarded not to be a good model for the particular case, and an original – i.e. not standardized - value was used for further calculation). In case of precipitation, neighbours with original values equal to zero were not standardized. For the weighted average (using inverse distances as weights), power of weights like 1 (all meteorological elements except precipitation) or 3 (precipitation) was applied. In case of temperatures, standardized neighbour values outside 0.2 and 0.8 percentiles were not considered in final value calculation (i.e. trimmed mean was applied). Example of settings (in ProClimDB) of parameters for technical series calculation is given in Fig. 4.

Originally, the “raw” station data (only with suspicious values excluded), i.e. with gaps and many of them not measuring in whole period 1961-2000, were used both for stations technical series and grid points calculation. Even if the same statistical properties of original measured data were preserved in calculated technical series (calculated for each day separately), some of the time series showed inhomogeneities, which could flow either from inhomogeneous original station data or from the way of calculation: if some stations measured only for a short time, then the selection of neighbours during the whole period 1961-2000 varied. To avoid inhomogeneities following from changing list of used neighbours, we proceeded during technical series calculation in this way: first missing values were filled in station data (calculated from neighbouring stations standardized to candidate station average and standard deviation, for each month individually, daily step). Second, for station series with filled gaps, station technical series were calculated, applying standardization of neighbours to base station altitude average and standard deviation (estimated using linear regression for the neighbouring region, for each month individually), thus all stations were prolonged to have values in the whole period 1961-2000. Third, only these equally long station technical series were used for grid points calculation.

Altitudes applied for grid points calculation were real altitudes, read from 1 km resolution model of terrain. However for purposes of RCM outputs validation, it would be better to gain the altitudes from smoothed terrain (e.g. 20x20 km, 10x10 km smoothing) to characterize vicinity of a grid point as is the case of models. The same is valid for power of weights (inverse distances). Applying power like 0.5 (square root) better characterizes wider vicinity of a grid point. The meaning of the calculation for this deliverable was however creation of technical series at a location of a station or a grid point and to preserve statistical characteristic of the particular point. Thus the calculated series provide point-specific data rather than area-aggregated data. The other reason is that area aggregation is various for different models. Using these technical series for dynamical downscaling models validation purposes has to be performed with caution.

Settings of parameters differ for individual meteorological elements. Next chapter describes finding of the best solution for each meteorological element on example of selected stations in the area of the Czech Republic.

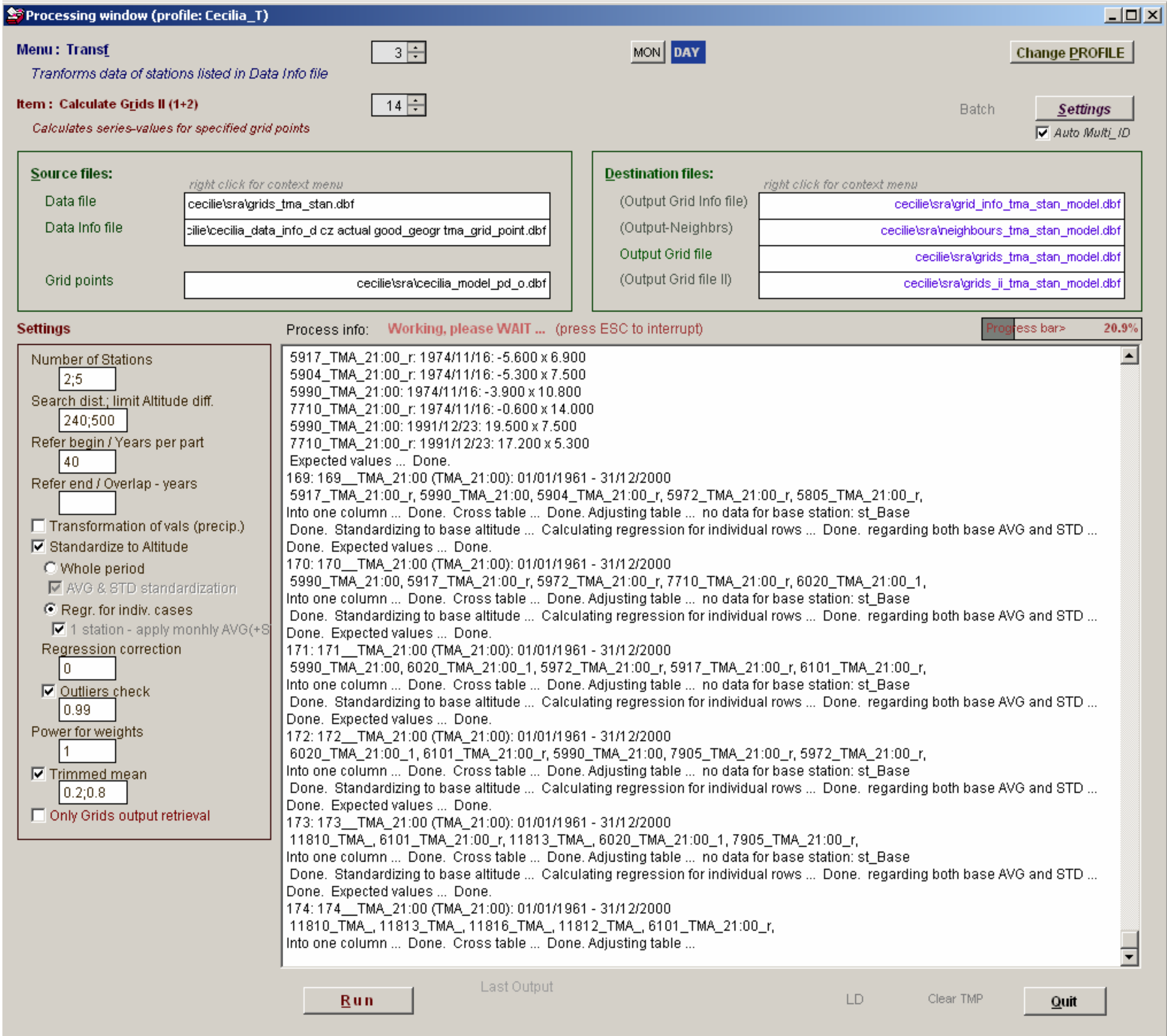


Fig. 4. Example of settings parameters for technical series calculation in ProClimDB (maximum temperature).

3 Finding the best settings for station technical series and gridded datasets calculation

Parameter settings for station technical series and gridded datasets calculation differ for various meteorological elements. The "ideal" settings of parameters of ProClimDB software was searched using selection of four base stations in the area of the Czech Republic. The stations were selected in the way they represent different climatological conditions, so both lowland and highland stations were selected and stations were selected both on eastern and western part of the area to capture differences between maritime and continental weather regime which is already manifested in the area of the Czech Republic. The four selected base stations with their neighbour station are shown on Fig. 5, information about the base stations is given in Table 3. The parameters settings were tuned by comparing original and calculated values (in previous text – quality control - marked as "expected") using various verification criterions.

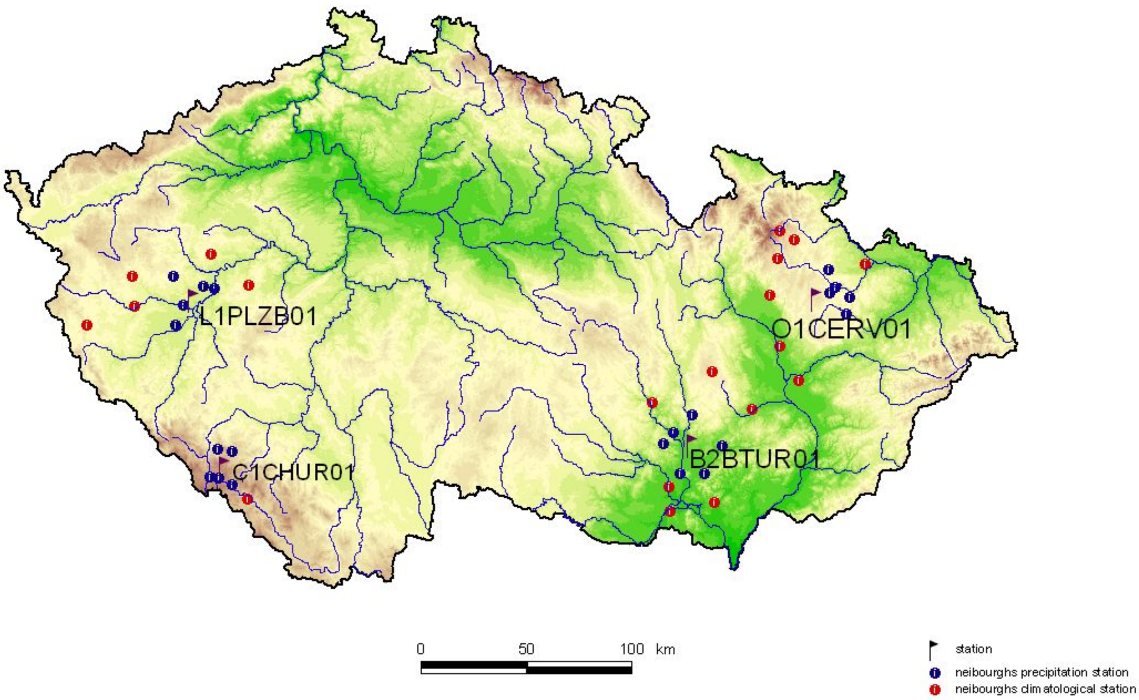


Fig. 5. Four base stations (marked with flag) and their neighbours (distinguished for precipitation and climatological stations) used for verification of calculated technical series.

Table 3. Base stations used for verification of calculated technical series.

NAME	ID	LATITUDE	LONGITUDE	ALTITUDE
Brno-Tuřany	B2BTUR01	49.1597	16.6956	241.00
Plzeň-Bolevec	L1PLZB01	49.7892	13.3867	328.00
Červená	O1CERV01	49.7772	17.5419	750.00
Churáňov	C1CHUR01	49.0683	13.6131	1118.00

Altogether, 11 various parameters of ProClimDB were tested to find an "ideal" settings, individually for all the required meteorological elements: maximum and minimum temperature, relative humidity, precipitation and sunshine duration. Daily values of the meteorological elements in the period 1991-2007 were used. The changed (controlled) parameters were: transformation of input values (log, square root, …); standardization of neighbour stations values to base station monthly averages (and/or standard deviations), standardization of neighbour stations values to base station altitude (in this case you can further control: calculating regression for the whole period – monthly, or for each time step – day - individually, set behaviour in case of presence of the only station, correction coefficient for regression to control dependence on altitude); checking if standardized values did not become outliers; power of weights for a new (Expected) value calculation; applying trimmed mean during a new value calculation (and setting the limits in such a case).

It was more difficult to find a solution for precipitation and relative humidity than for the other meteorological elements. Unfortunately it seems impossible to get 100% of realistic values during the calculation (e.g. non-negative relative humidity, precipitation, number of unrealistic values is given in the last chapter). The unrealistic values are caused mainly by:

poor quality of input (raw station) data, insufficient lengths of neighbour stations (the same time gaps in several neighbour stations diminishes number of values used for regression), and greater difference among altitudes of stations used in regression model. These factors can be controlled to some extend. The input data were controlled for quality before calculation (see previous chapter). Stations allowed for the calculation can be filtered to have certain lengths (the same like base stations), and without longer time gaps. The third factor – difference in altitudes – is not easy to solve since we selected the nearest neighbours for the calculation, which is e.g. in case of precipitation the only solution (selection of nearest and best correlated stations is the same, in case of temperatures we could select neighbour stations also according to correlations but this was not performed for shortage of time and since some other problems can arise from such a selection). The main problems were encountered during finding solution for mountainous station (Churáňov), since its altitude is higher than those of its neighbours and thus we have to extrapolate values instead of interpolate them.

Further sections show settings using ProClimDB software, version 7.989 (Štěpánek, 2007), switched to daily version, using menu Transf, item Calculate Grids II. Comparison of calculated and original values is shown with more details on example of station Brno-Tuřany.

3.1 Maximum temperature

For the station technical series and gridded dataset maximum temperature calculation, parameters were set in this way (see Fig. 6):

- standardization of neighbour stations values to base station altitude (checked option *Standardize to Altitude*)
- station technical series calculation: regression (dependence on altitude) calculated for the whole period, monthly (both *AVG and STD standardization*), while for gridded dataset: regression calculated for each day individually (option *Regr. for invid. cases*)
- checking if standardized values do not differ too much from original values, for CDF greater than 0.99, in such a case original values are used for further calculation (option *Outliers Check*), settings 0,95 or 0,90 lead to much worse results
- Weights - inverse distances - were powered by 1 (option *Power for weights*)
- Applying trimmed mean for a new value calculation with quantile limits 0.2 and 0.8 (option *Trimmed mean*)

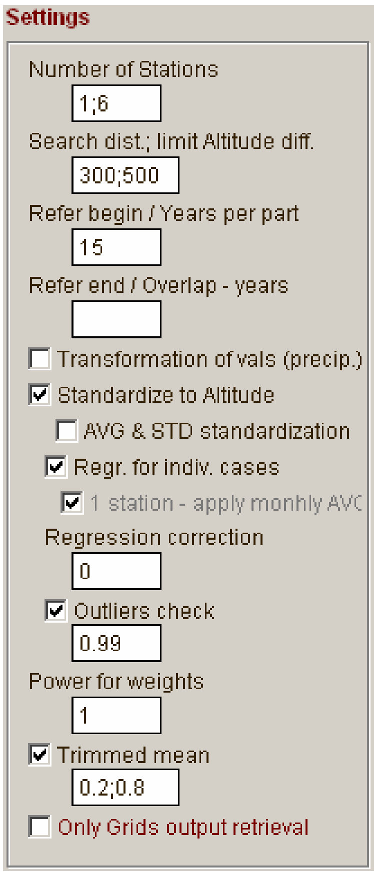


Fig. 6. Settings for gridded dataset calculation for maximum temperature (ProClimDB software, menu Transf, item Calculate Grids II).

From Fig. 7 it can be seen that stations with lower altitudes show unclear annual cycle of RMSE (root mean square error applied on calculated and original – measured – values). The value varies up to 1. On the contrary mountainous station Churáňov reaches very high values of RMSE during winter months, the discrepancies can be given by occurrence of inversions in mountains in winter when the lowland stations used for the calculation have different weather conditions.

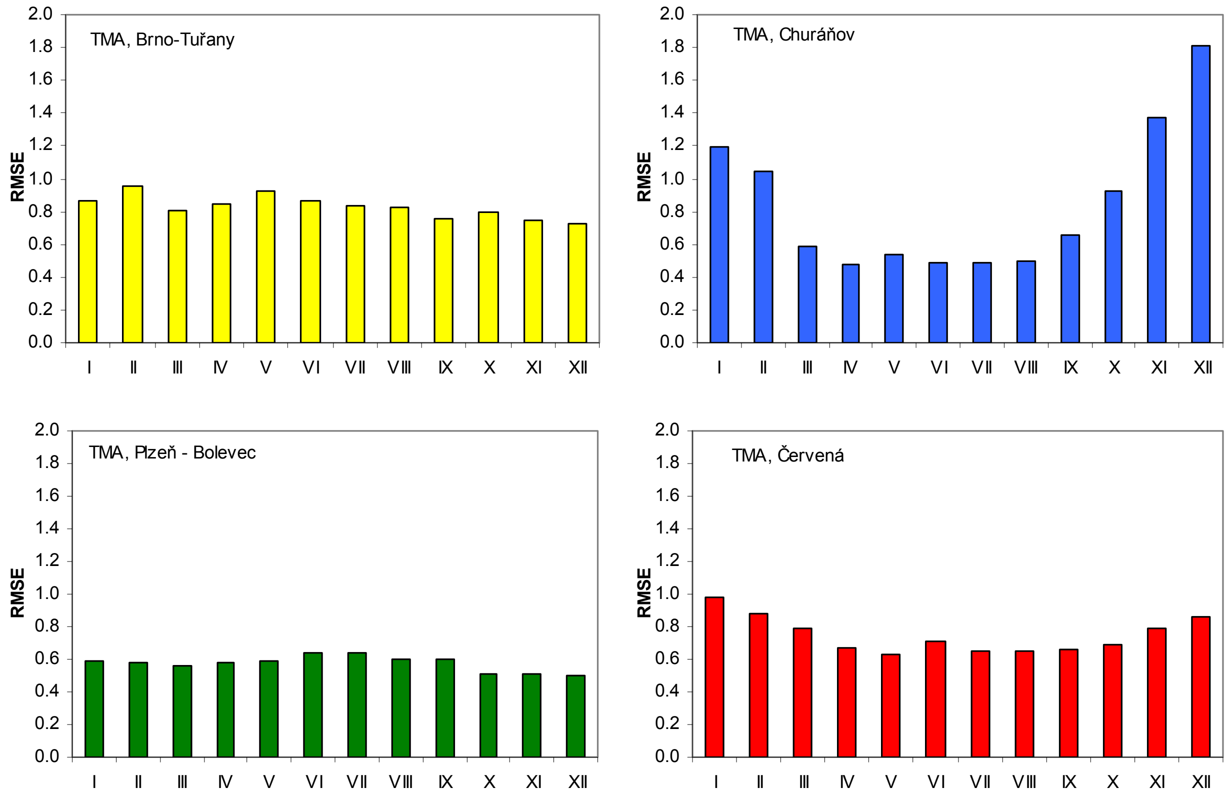


Fig. 7. RMSE for four base (tested) stations and maximum temperature.

Average difference between calculated and original values of Plzeň-Bolevec is -0.0°C (see table 4 and table 5). In all the months the bias is not higher than 0.1 °C, the highest being in April. Correlation coefficient is very high, 0,998 (see Fig. 8). Root mean square error is similar for individual months, around 0.55. From the histogram of differences between calculated and original values of Plzeň-Bolevec (Fig. 9) it can be seen that 50% of values lies in interval (0,1 - 0,5> or (- 0,1 - -0,5>.

Calculated maximum temperature for Brno – Tuřany is on the contrary biased, most of the values are overestimated (e.g. 26% of values lies in interval (0,1;0,5>), the bias for a whole year is 0.4 °C. On the other hand, average daily temperature is without bias. From metadata of the station it follows that there were problems in the past with sensor for temperature measurements. The problems were encountered also in case of Churáňov mountainous station.

Table 4. Basic statistical characteristics for original and calculated values for Plzeň-Bolevec in the period 1991-2007.

	Plzeň-Bolevec	Calculated value
Mean	13.1	13.1
Median	12.9	12.9
Minimum	-14.4	-14.2
Maximum	35.1	35.0
Lower quartile	5.5	5.4
Upper quartile	20.5	20.5
10% quantile	1.4	1.4
90% quantile	25.7	25.7
Standard deviation	9.3	9.3
Skewness	0.1	0.1
Kurtosis	-0.9	-0.9
Correlation coeff.		0.998

Table. 5. Validation criterion applied on daily values of Plzeň-Bolevec, for individual months, 1991-2007.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Min. difference (Bias)	-2.46	-2.42	-1.44	-1.93	-1.48	-1.74	-2.29	-2.46	-2.05	-1.95	-1.97	-1.70
Max. difference (Bias)	3.08	3.48	1.72	1.77	2.32	2.12	1.94	2.30	1.99	2.24	2.18	2.10
Bias (mean)	0.05	0.06	0.03	-0.09	-0.06	0.06	-0.02	-0.04	0.07	0.02	0.07	0.03
MAE	0.43	0.41	0.43	0.38	0.38	0.44	0.41	0.47	0.41	0.42	0.44	0.41
RMSE	0.59	0.57	0.56	0.49	0.51	0.58	0.54	0.62	0.55	0.55	0.58	0.52
NRMSE	0.02	0.03	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.03	0.02	0.02
Mean original value	1.19	3.91	8.79	12.65	18.73	22.31	23.66	24.10	19.37	12.35	6.45	3.16
Mean calculated value	1.24	3.97	8.82	12.56	18.67	22.36	23.63	24.07	19.44	12.37	6.52	3.19

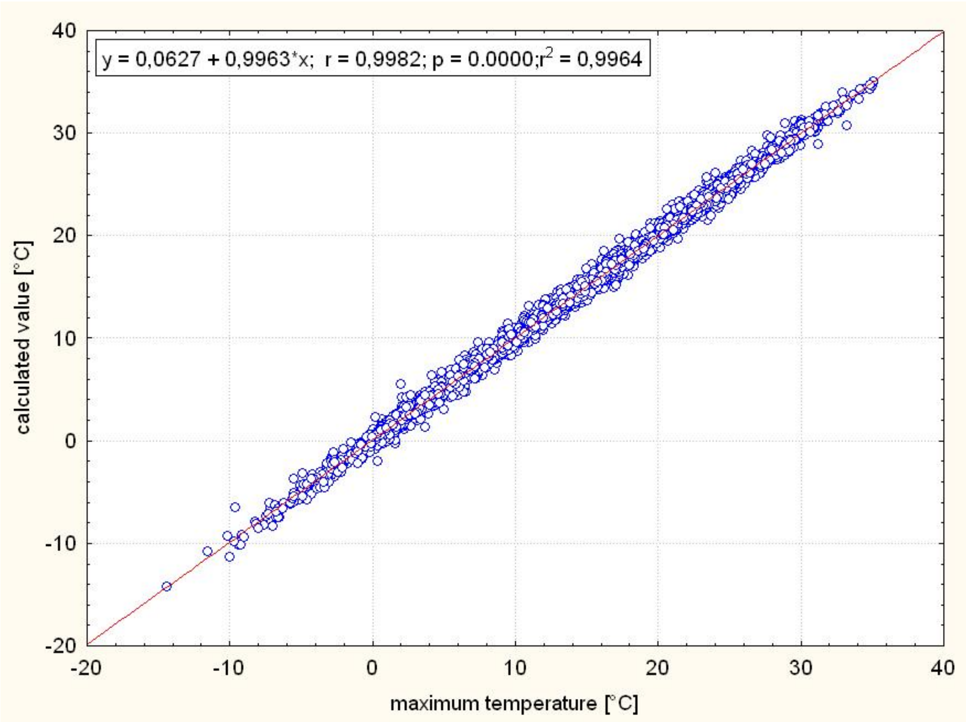


Fig. 8. Scatter plot for calculated and original values of Plzeň-Bolevec.

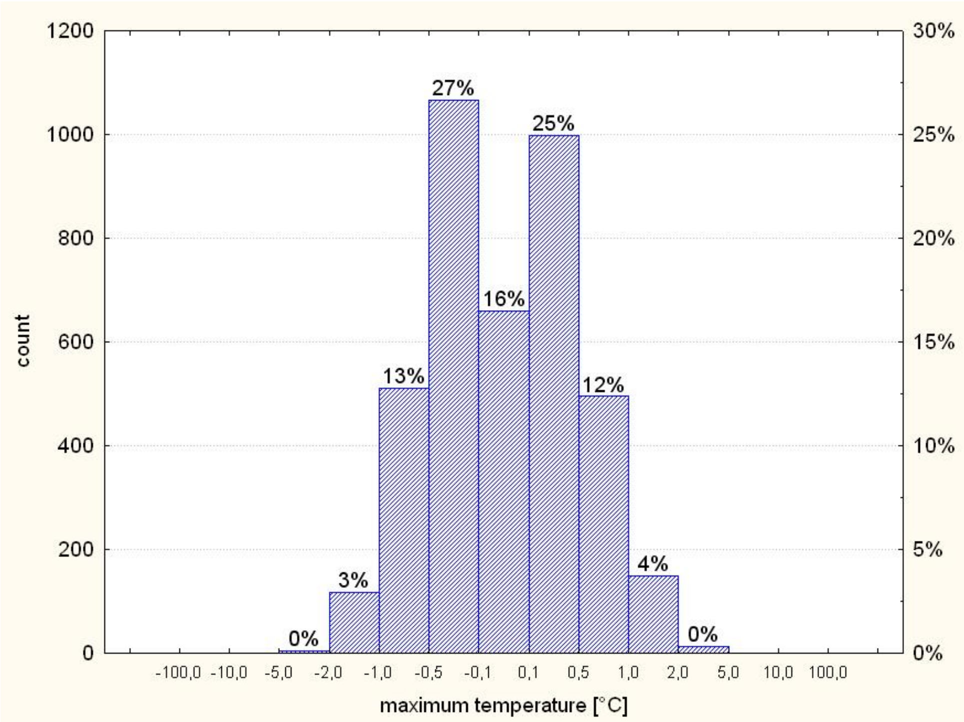


Fig. 9. Histogram for differences of calculated and original values of Plzeň-Bolevec.

3.2 Minimum temperature

The setting for minimum temperature is the same like in case of maximum temperature.

In case of minimum temperatures differences between lowland and other stations are not so great like in case of maximum temperature (see Fig. 10). Plzeň-Bolevec and Červená stations show surprisingly annual cycle of RMSE with higher values in spring and summer months.

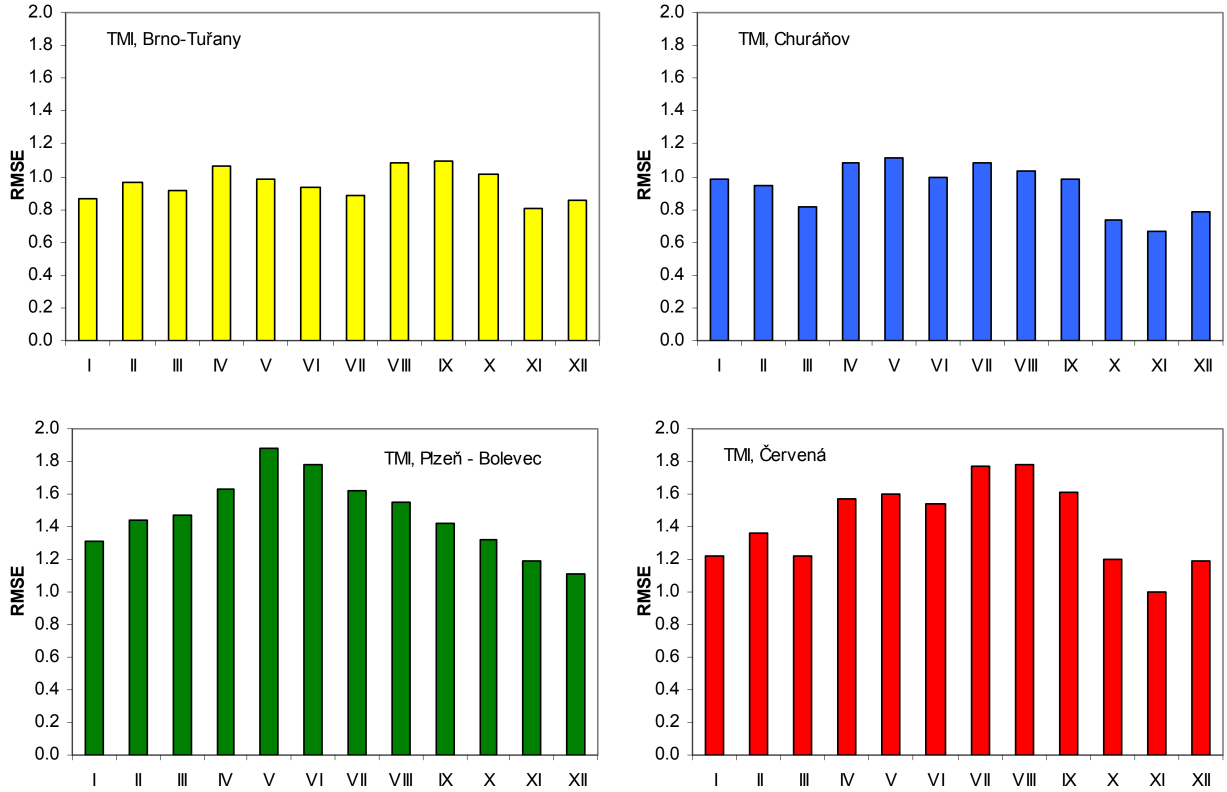


Fig. 10. RMSE for four base (tested) stations and minimum temperature.

Difference between calculated and original value of Brno-Tuřany is -0,4°C (bias for maximum temperature was on the contrary 0.4 °C). The highest differences are reached in summer months, better results are again reached for winter months, in December the bias is only 0.1 °C, while in August it is -0,6°C. From the histogram (Fig. 12) it can be seen that the calculation is underestimated mainly in interval -0,1 to -0,5 °C. RMSE is greater than in case of maximum temperature. Correlation coefficient is again very high (0,993), see Fig. 11.

Table 6. Basic statistical characteristics for original and calculated values for Brno-Tuřany in the period 1991-2007.

	Brno-Tuřany	Calculated value
Mean	5.2	4.8
Median	5.5	5.3
Minimum	-21.8	-23.2
Maximum	22.9	23.1
Lower quartile	-0.6	-0.7
Upper quartile	11.5	11.0
10% quantile	-4.7	-4.8
90% quantile	14.9	14.4
Standard deviation	7.7	7.5
Skewness	-0.3	-0.3
Kurtosis	-0.6	-0.4
Correlation coeff.		0.993

Table 7. Validation criterion applied on daily values of Brno-Tuřany, for individual months, 1991-2007.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Min. difference (Bias)	-3.52	-4.17	-4.57	-4.74	-3.92	-3.77	-3.91	-4.08	-4.28	-3.55	-4.39	-2.78
Max. difference (Bias)	3.41	2.53	5.78	3.52	1.72	1.76	1.91	2.51	1.94	2.09	1.96	3.24
Bias (mean)	-0.11	-0.24	-0.22	-0.45	-0.47	-0.45	-0.40	-0.60	-0.54	-0.48	-0.13	-0.01
MAE	0.62	0.72	0.64	0.77	0.74	0.71	0.66	0.85	0.81	0.75	0.61	0.63
RMSE	0.87	0.97	0.92	1.06	0.99	0.94	0.89	1.08	1.09	1.01	0.81	0.86
NRMSE	0.03	0.04	0.04	0.05	0.05	0.05	0.05	0.06	0.07	0.05	0.03	0.03
Mean original value	-3.90	-3.19	0.18	4.55	9.35	12.39	14.30	14.16	10.03	5.76	1.10	-3.17
Mean calculated value	-4.01	-3.43	-0.03	4.10	8.88	11.94	13.90	13.56	9.49	5.28	0.97	-3.18

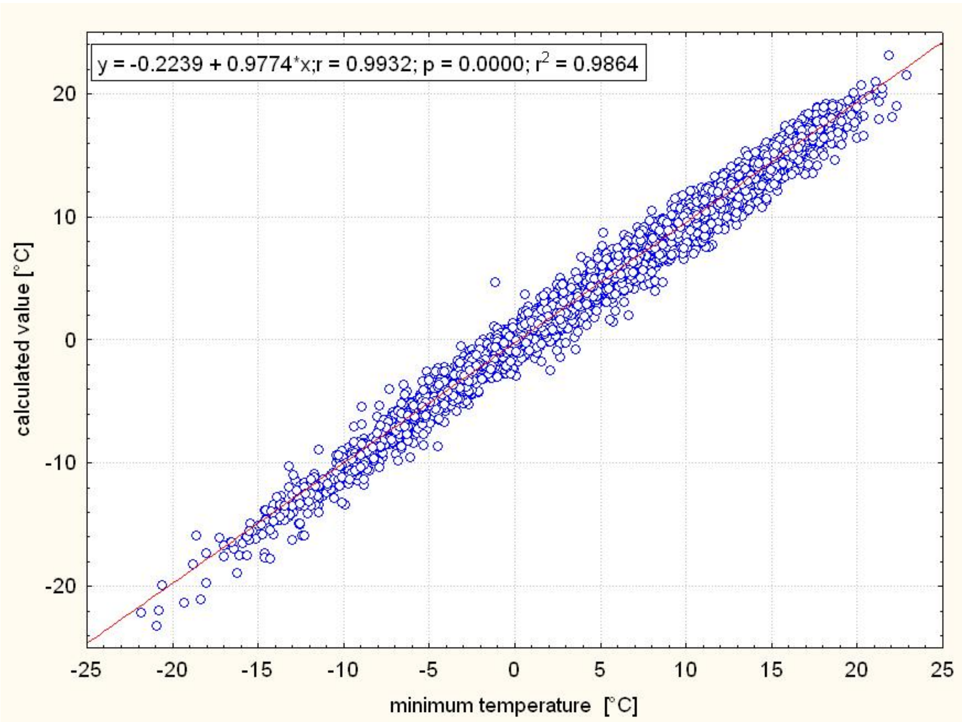


Fig. 11. Scatter plot for calculated and original values of Brno-Tuřany.

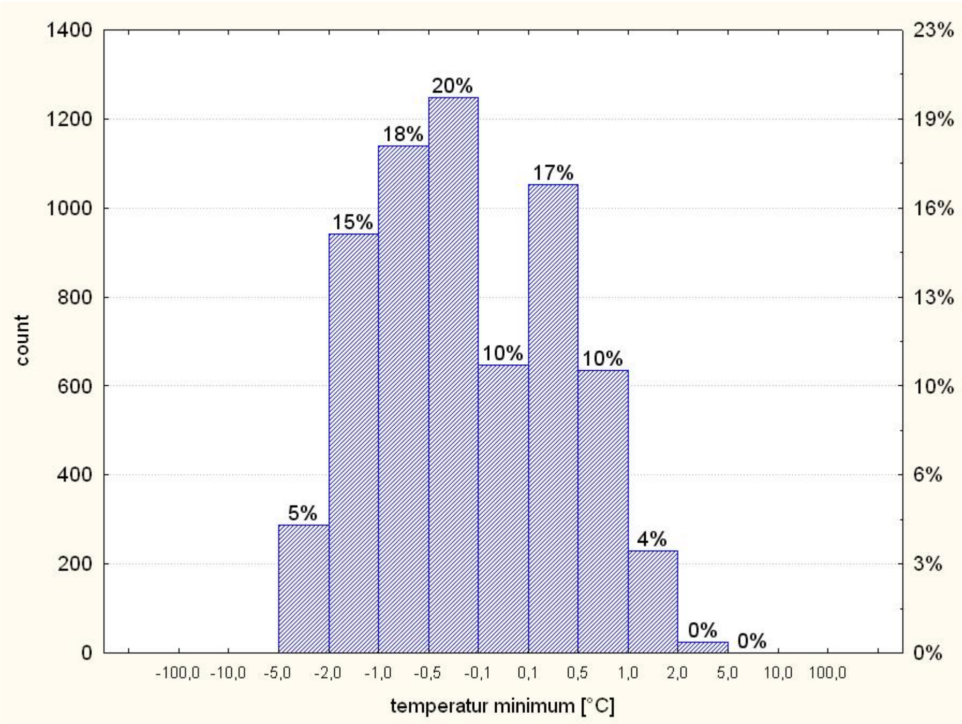
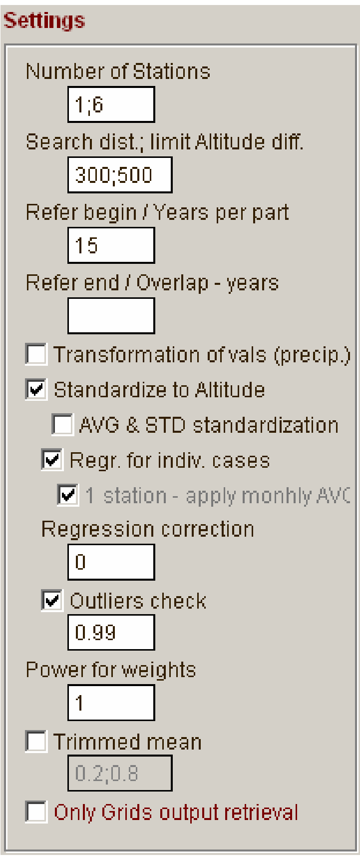


Fig. 12. Histogram for differences of calculated and original values of Brno-Tuřany.

3.3 Relative humidity



Settings of parameters for calculation of relative humidity is similar to that for air temperature, it means standardization to base station altitude, regression calculated for the whole period (monthly, both AVG and STD standardization, station technical series) or for each day individually (gridded dataset), power for weights being 1, but trimmed mean is not applied in this case (see Fig. 13).

Mountainous station Churáňov shows pure results in winter months (see RMSE, Fig. 14), similar to results of maximum temperature, the reason will be probably the same: different weather regime in winter months of lowland stations used for the calculation and possibly errors in measurements (see the end of this chapter).

Fig. 13. Settings for gridded dataset calculation for relative humidity (ProClimDB software, menu Transf, item Calculate Grids II).

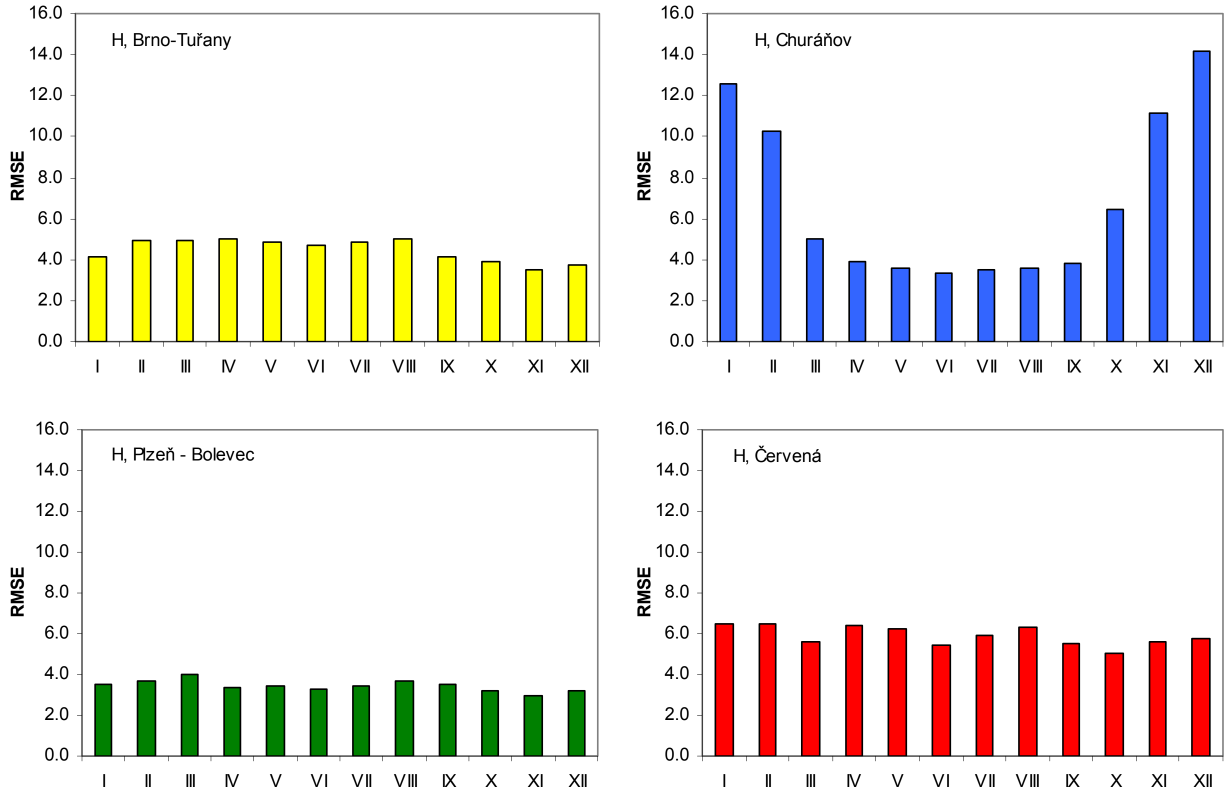


Fig. 14. RMSE for four base (tested) stations and relative humidity.

Average difference between calculated and original values for Brno-Tuřany (1991-2007) is only 0,3%. For individual months the difference varies from 0,04 % to 2,24 %. In winter months the bias is about -1%. In March and October we find the lowest values (0,04-0,12%). In the rest of a year the bias is positive, the highest difference occurring in August. Scatter plot (Fig. 15) is distinguished by wider spread of data compared to air temperature, but correlation coefficient is still very high (0,952). Calculated values can sometimes exceed slightly 100% or to get below 0%, the final output has to be checked for such cases.

Table 8. Basic statistical characteristics for original and calculated values for Brno-Tuřany in the period 1991-2007.

	Brno – Tuřany	Calculated value
Mean	73.9	74.2
Median	75.0	74.6
Minimum	32.0	37.6
Maximum	100.0	99.8
Lower quartile	64.0	65.0
Upper quartile	85.0	84.1
10% quantile	54.0	56.7
90% quantile	93.0	91.1
Standard deviation	14.3	12.7
Skewness	-0.2	-0.2
Kurtosis	-0.6	-0.7
Correlation coeff.		0.952

Table 9. Validation criterion applied on daily values of Brno-Tuřany, for individual months, 1991-2007.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Min. difference (Bias)	-13.43	-16.83	-12.76	-12.28	-11.80	-11.82	-10.44	-8.61	-9.69	-9.22	-10.91	-13.11
Max. difference (Bias)	12.63	12.82	21.58	15.28	15.06	12.73	15.98	21.47	15.92	12.62	9.67	10.51
Bias (mean)	-1.02	-0.70	0.04	1.22	0.90	0.65	0.85	2.24	1.57	0.12	-0.98	-1.17
MAE	3.24	4.00	3.89	4.05	3.90	3.80	3.88	3.82	3.37	3.19	2.76	3.00
RMSE	4.13	4.96	4.93	5.02	4.83	4.68	4.82	4.99	4.13	3.89	3.51	3.76
NRMSE	0.08	0.10	0.08	0.08	0.08	0.08	0.08	0.08	0.07	0.07	0.08	0.09
Mean original value	84.06	79.10	73.62	63.67	64.88	66.45	65.08	64.73	72.71	79.80	85.61	86.80
Mean calculated value	83.04	78.40	73.66	64.89	65.78	67.10	65.93	66.96	74.28	79.92	84.63	85.63

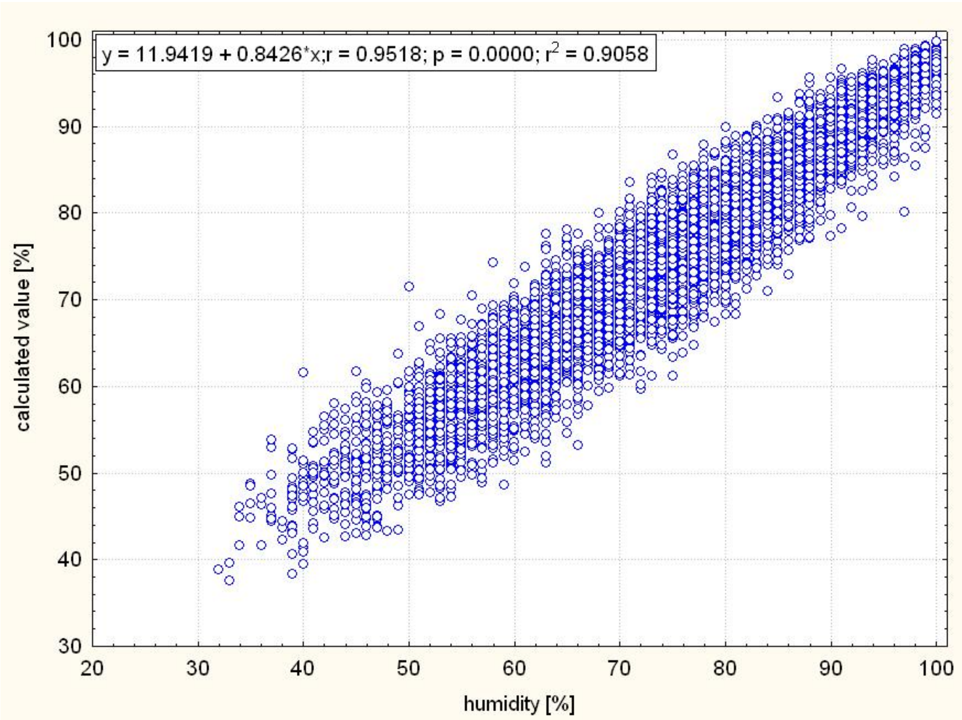


Fig. 15. Scatter plot for calculated and original values of Brno-Tuřany.

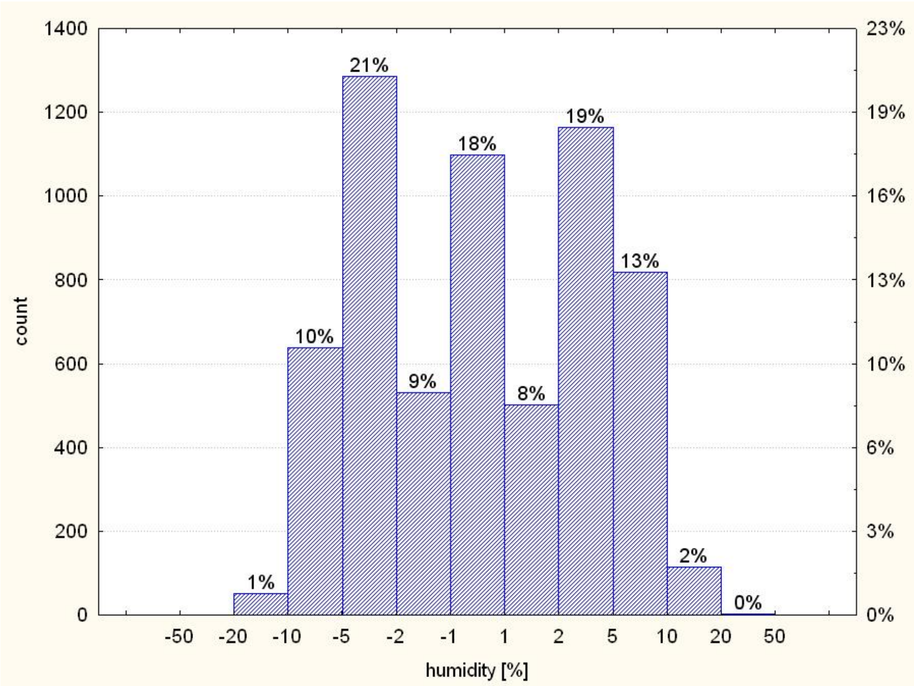


Fig. 16. Histogram for differences of calculated and original values of Brno-Tuřany.

The problems during relative humidity calculation were found mainly in case of Churáňov station (C1CHUR01). It is an example of mountainous station for which many problems during calculation occur. One of them is insufficient number of surrounding stations. Their difference in altitude can be very high (in this case, Churáňov has 1118 m, the neighbours have 803, 737, 480, 360, 328 m), moreover the neighbours are more distant then in case of lowland base stations. The other problem is different whether regime of Churáňov from the neighbour stations. The discrepancies in calculation were found mainly for winter months. The question is whether the differences can be assigned to different climate conditions or errors in measurements (for example low values around 20 % compared to high values of 90% in neighbour stations as shown in Fig. 17).

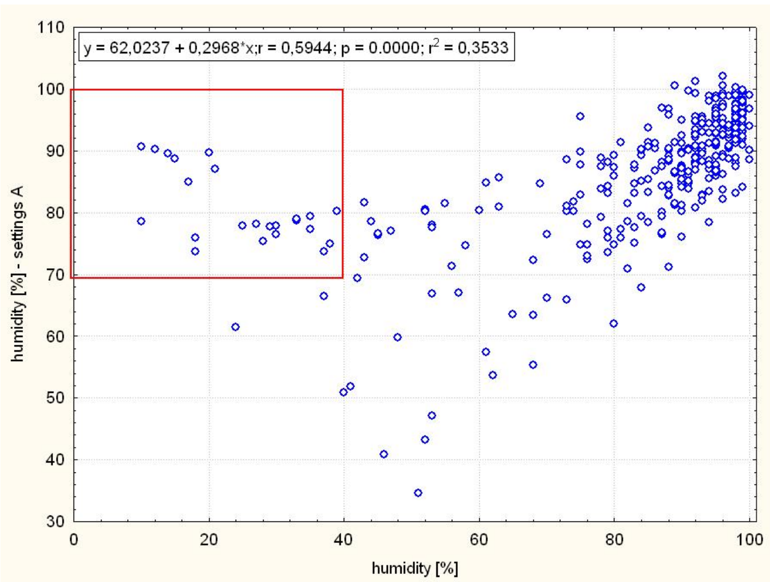


Fig. 17. Scatter plot for calculated and original values with problematic values (in red rectangle) for mountainous station Churáňov in the period 1981-1991.

3.4 Precipitation

For calculation of technical series in case of precipitation, standardization to altitude for the whole period (station technical series), or applied individually for each day (gridded dataset) was carried out again. The difference from previous settings is that power for weight is 3 to reflect lower spatial correlations of precipitation, and trimmed mean is not applied like in case of relative humidity. No transformation of input values was performed since it gave poorer results. Example of the settings is given in Fig. 18.

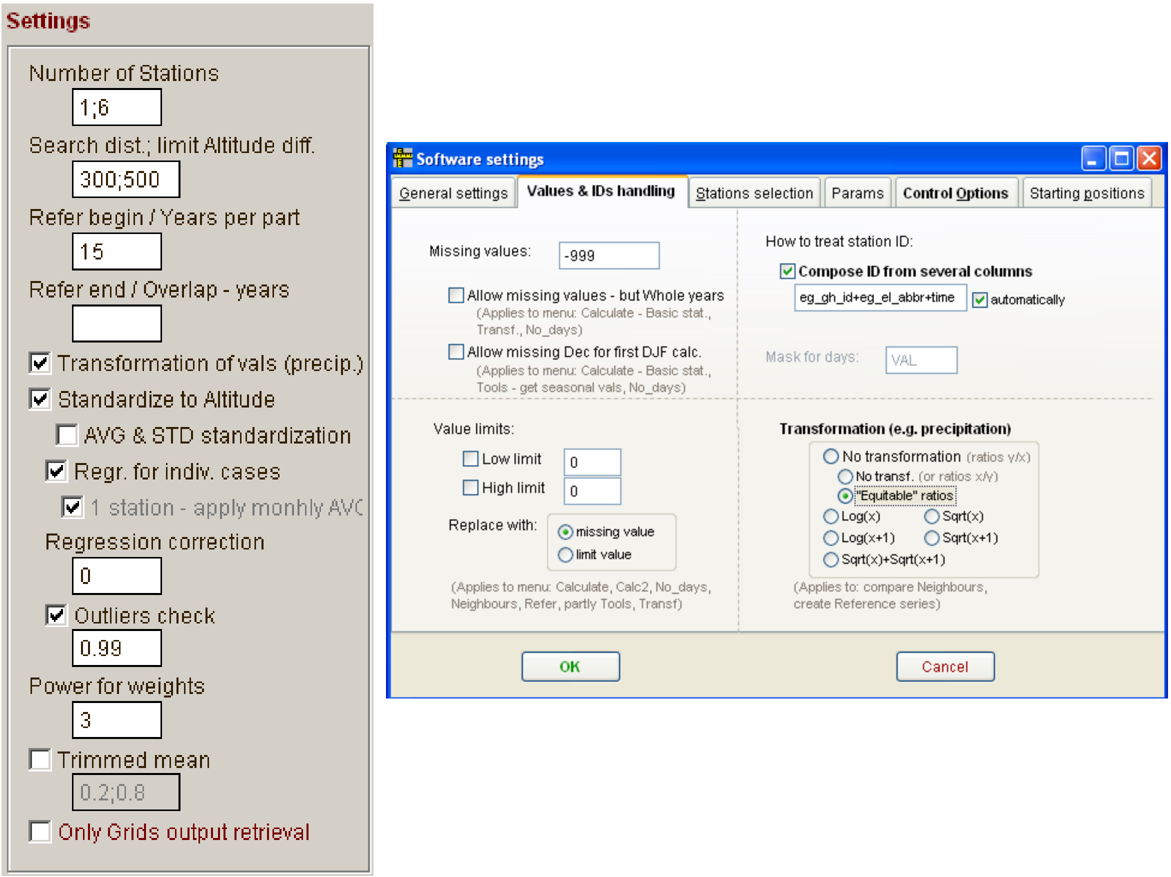


Fig. 18. Settings for gridded dataset calculation for precipitation (ProClimDB software, menu Transf, item Calculate Grids II).

Annual cycle of RMSE in case of precipitation (Fig. 19) is similar for all the tested stations, lower values occurring in winter months due to prevailing circulation factors and thus with more similar conditions over wider area, compared to higher values of RMSE occurring in summer months, due to local effects coming form prevailing radiative factors.

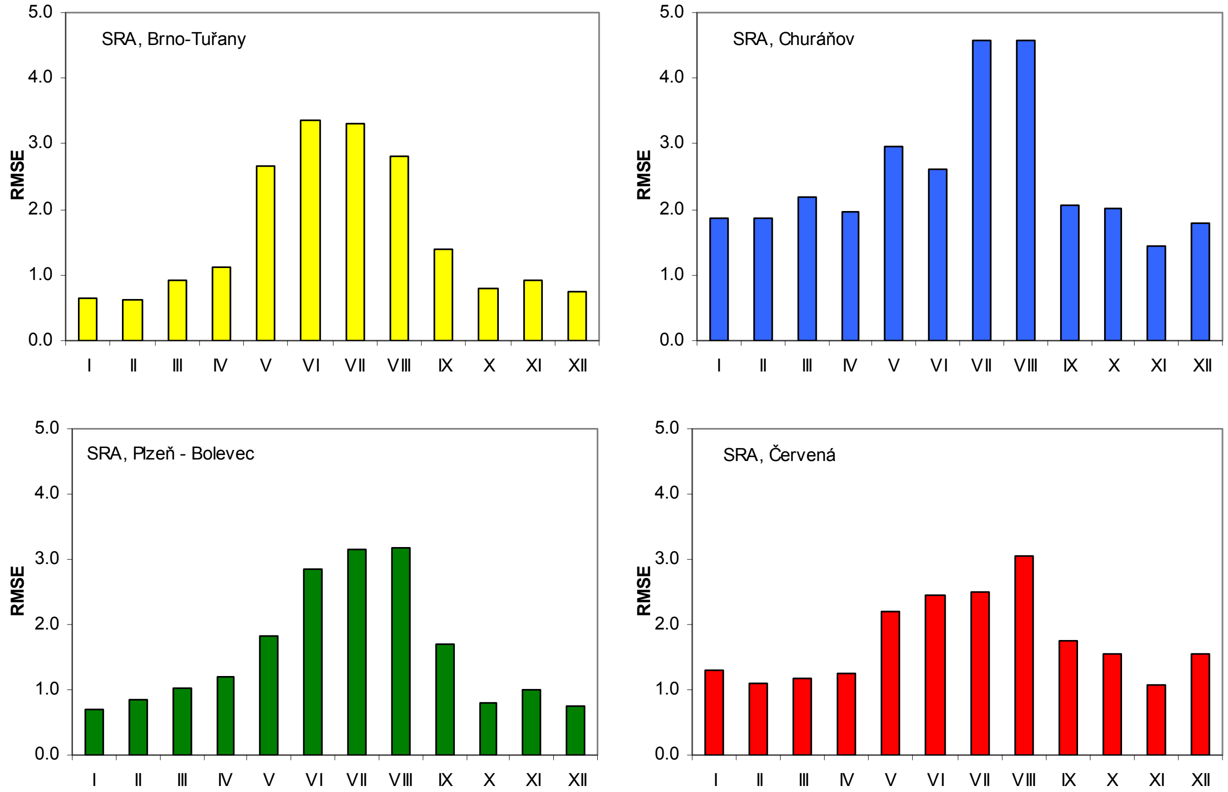


Fig. 19. RMSE for four base (tested) stations and precipitation.

Average difference for precipitation for Brno-Tuřany is 0.0 mm, in most of the months the value does not exceed 0.1 mm. The highest difference occurs for June, 0.27 mm, RMSE values are highest for summer months as well. Precipitation are influenced by local effects much more than the other meteorological elements, even in adjacent places we can get great differences (in some cases we can encounter 30-60 mm for two neighbour stations, for other two stations no precipitation at all). For this reason also correlation coefficient is lower, only 0,875. From the scatter plot (Fig. 20) we can see several outliers which influence value of the correlation coefficient. From the histogram (Fig. 21) it follows that 62% values differ only negligibly.

Table 10. Basic statistical characteristics for original and calculated values for Brno-Tuřany in the period 1991-2007.

	Brno – Tuřany	Calculated value
Mean	1.4	1.4
Median	0.0	0.0
Minimum	0.0	-0.2
Maximum	56.5	61.4
Lower quartile	0.0	0.0
Upper quartile	0.7	0.7
10% quantile	0.0	0.0
90% quantile	4.0	4.0
Standard deviation	3.9	3.8
Skewness	5.2	5.3
Kurtosis	37.4	40.0
Correlation coeff.		0.875

Table 11. Validation criterion applied on daily values of Brno-Tuřany, for individual months, 1991-2007.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Min. difference (Bias)	-4.84	-3.83	-6.48	-10.19	-26.93	-24.40	-35.58	-34.41	-15.50	-5.36	-4.63	-5.03
Max. difference (Bias)	7.48	3.87	10.41	7.32	20.74	16.00	23.24	15.17	8.09	7.47	10.24	4.90
Bias (mean)	-0.01	-0.09	0.03	0.03	-0.16	-0.27	-0.01	0.17	-0.03	0.00	0.00	-0.04
MAE	0.24	0.23	0.33	0.42	0.87	1.26	1.20	0.95	0.52	0.30	0.35	0.33
RMSE	0.64	0.61	0.92	1.13	2.67	3.36	3.32	2.82	1.40	0.79	0.91	0.75
NRMSE	0.05	0.03	0.05	0.05	0.05	0.10	0.08	0.06	0.03	0.03	0.05	0.03
Mean original value	0.69	0.74	0.99	1.01	1.76	2.37	2.23	1.84	1.65	1.12	1.18	1.03
Mean calculated value	0.67	0.65	1.02	1.04	1.61	2.09	2.22	2.01	1.62	1.12	1.18	0.99

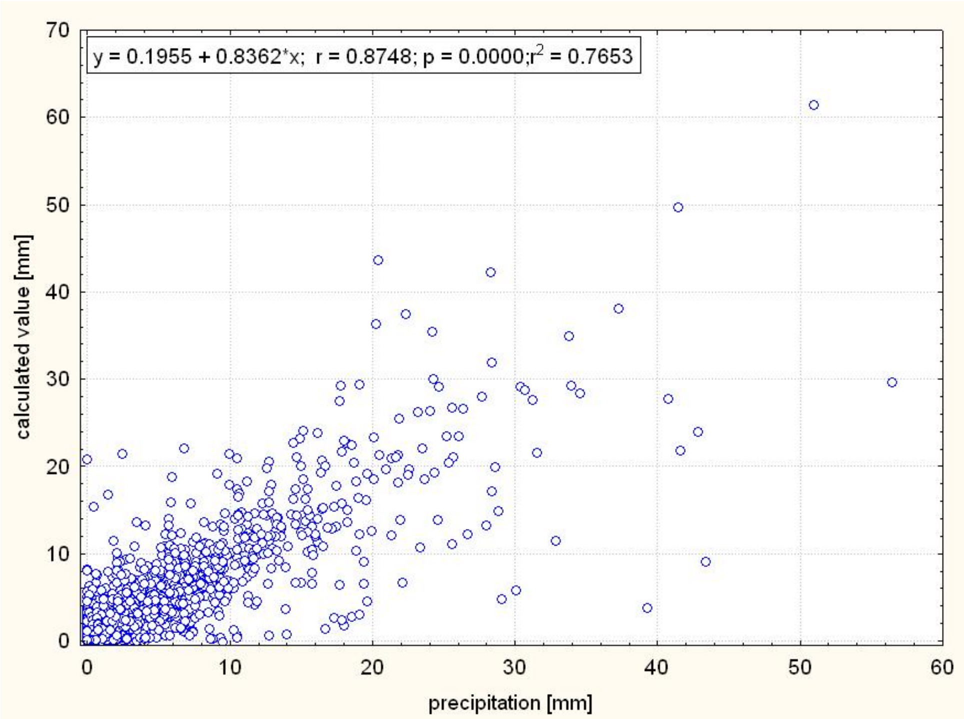


Fig. 20. Scatter plot for calculated and original values of Brno-Tuřany.

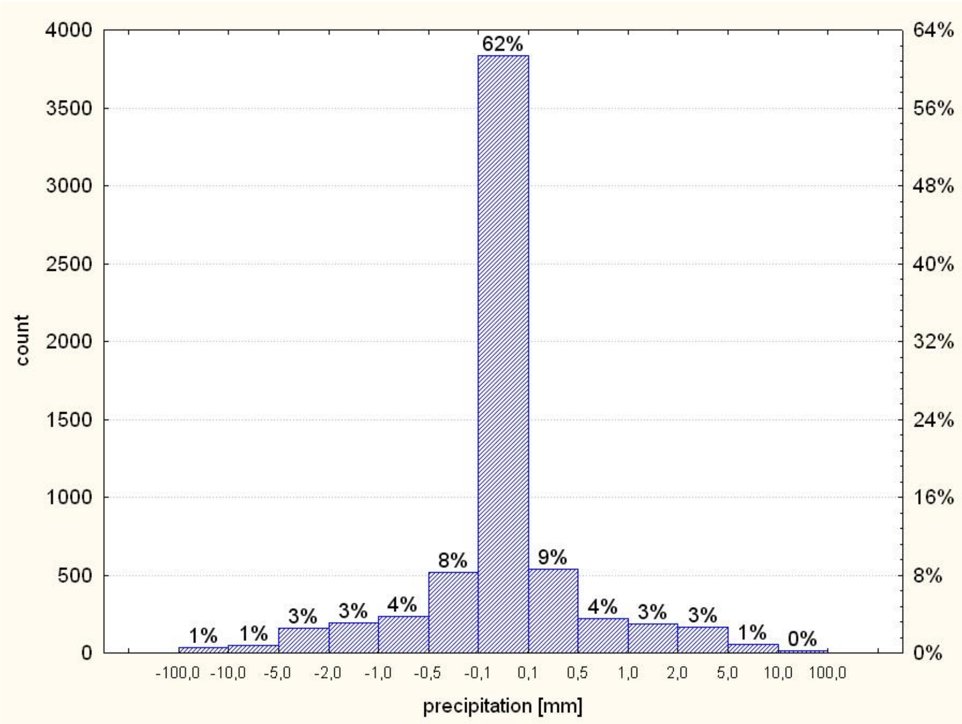
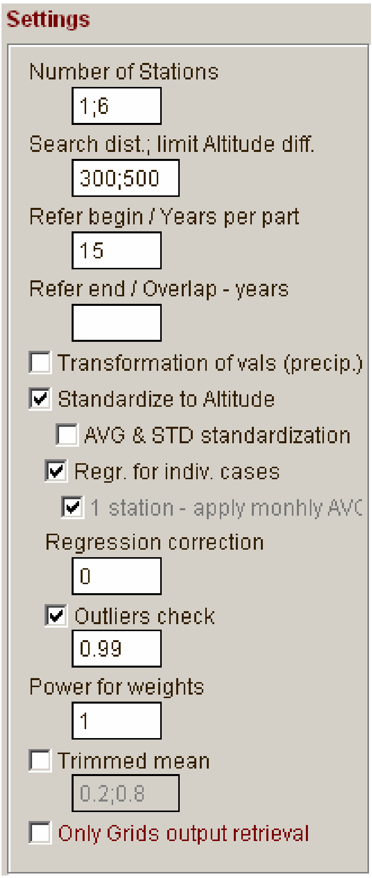


Fig. 21. Histogram for differences of calculated and original values of Brno-Tuřany.

3.5 Sunshine duration



For calculation of calculated value in case of sunshine duration, the settings is the same like for temperature, but trimmed mean is not applied (see Fig. 22).

RMSE for sunshine duration is shown in Fig. 23. Again, mountainous station Churáňov shows purer results in winter months like in case of maximum temperature and relative humidity. Remarkable are high values of station Červená over a whole year.

Fig. 22. Settings for technical series and gridded dataset calculation for sunshine duration (ProClimDB software, menu Transf, item Calculate Grids II).

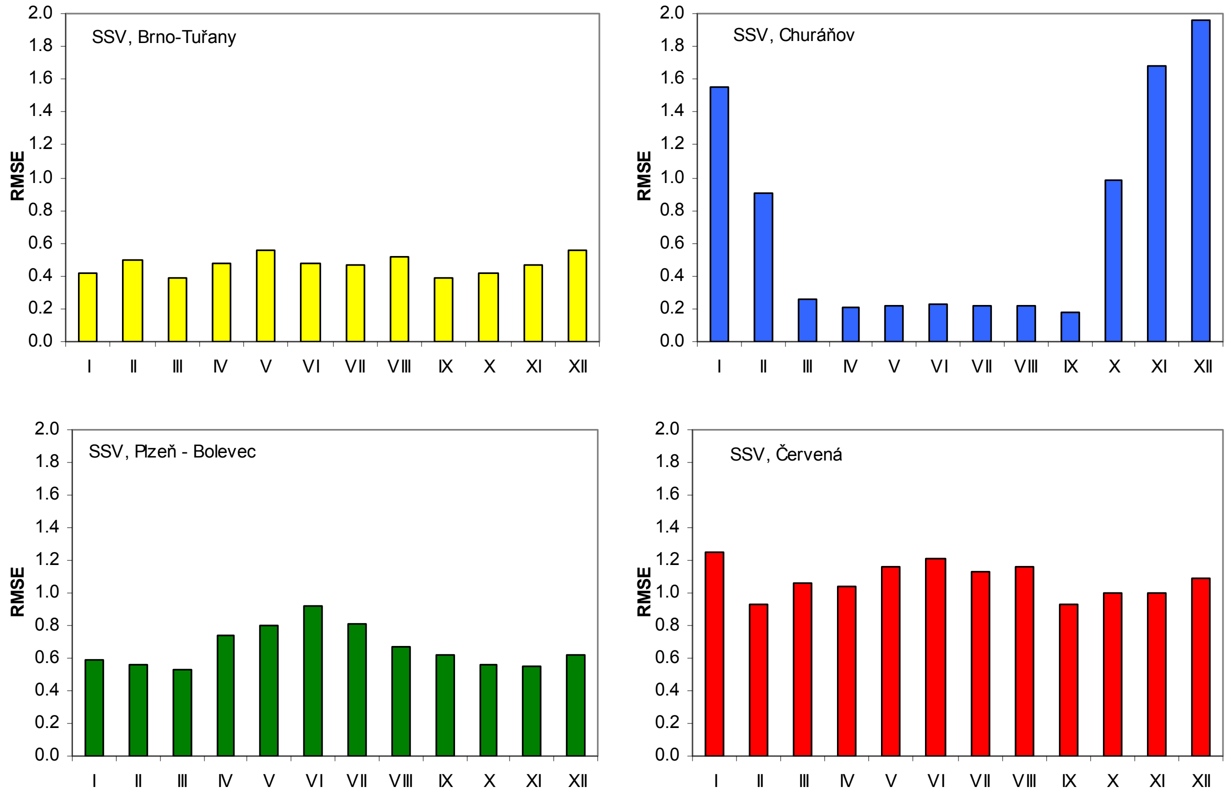


Fig. 23. RMSE for four base (tested) stations and sunshine duration.

Average difference between calculated and original value for Brno-Tuřany is only -0,1 hour. The highest differences are reached in July and August, the lowest in winter months. Correlation coefficient is again very high (0,995). From the scatter plot (Fig. 24) a few outliers are apparent, mainly in cases of low original values which is probably caused by errors in input data file. Histogram (Fig. 25) shows that 79% values is biased to ±0,1 hour. RMSE is about 0,5. Besides negative values, values exceeding physical limits (e.g. 16 hours in summer) have to be corrected in the output files.

Table 12. Basic statistical characteristics for original and calculated values for Brno-Tuřany in the period 1991-2007.

	B2BTUR01	Calculated value
Mean	5.2	5.1
Median	4.6	4.6
Minimum	0.0	0.0
Maximum	15.6	15.6
Lower quartile	0.4	0.5
Upper quartile	8.8	8.7
10% quantile	0.0	0.0
90% quantile	12.0	11.8
Standard deviation	4.5	4.5
Skewness	0.4	0.4
Kurtosis	-1.1	-1.1
Correlation coeff.		0.995

Table 13. Validation criterion applied on daily values of Brno-Tuřany, for individual months, 1991-2007.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Min. difference (Bias)	-2.49	-3.63	-2.87	-2.75	-3.94	-3.22	-2.33	-4.29	-3.58	-2.34	-1.95	-5.40
Max. difference (Bias)	1.77	3.09	2.76	1.92	2.26	2.79	2.37	4.03	2.36	2.18	7.50	6.20
Bias (mean)	-0.05	-0.04	-0.03	-0.05	-0.09	-0.08	-0.10	-0.11	-0.01	-0.03	0.03	-0.01
MAE	0.15	0.18	0.14	0.19	0.21	0.18	0.19	0.21	0.16	0.16	0.11	0.13
RMSE	0.42	0.50	0.39	0.48	0.56	0.48	0.47	0.52	0.39	0.42	0.47	0.56
NRMSE	0.05	0.05	0.03	0.04	0.04	0.03	0.03	0.04	0.03	0.04	0.05	0.06
Mean original value	1.94	3.22	4.35	6.53	8.13	8.51	8.50	8.16	5.74	3.69	1.78	1.37
Mean calculated value	1.89	3.19	4.32	6.48	8.03	8.43	8.39	8.05	5.73	3.67	1.81	1.37

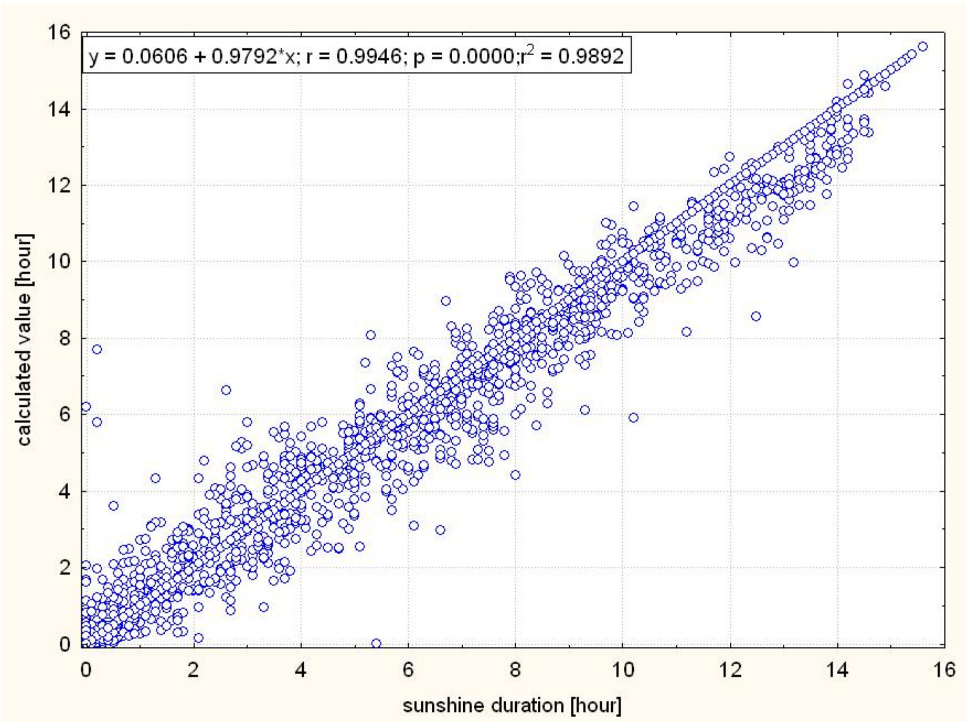


Fig. 24. Scatter plot for calculated and original values of Brno-Tuřany.

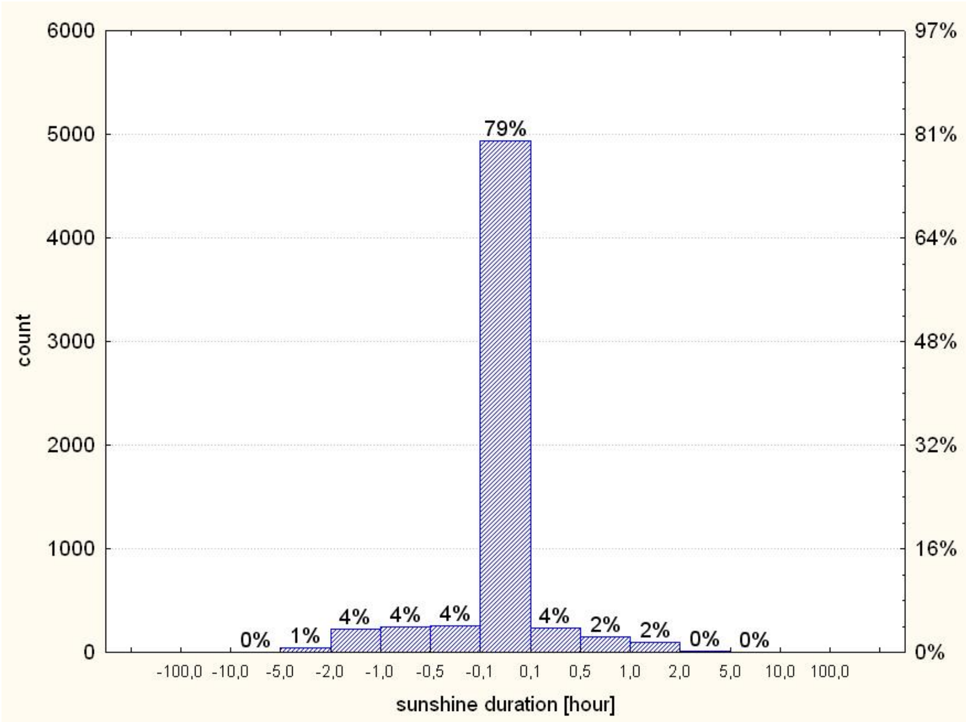


Fig. 25. Histogram for differences of calculated and original values of Brno-Tuřany.

4 Summary

As can be seen from the given validation results, the calculated station technical series and gridded datasets reflect very well behaviour of measured values of applied meteorological elements (maximum and minimum temperature, relative humidity, precipitation, sunshine duration), so it is capable of using for purposes of WP3.

The output files for CECILIA central European domain in the period 1961-2000 can be distributed in various formats and table structures according to demands of end users (thanks to tools implemented in the ProClimDB software).

Web page, from which it is possible to download data (zipped in DBF file format), is: http://www.cecilia-eu.org/restricted/participant/WP3/data/CECILIA_WP3_D3_1.html

References:

Isaaks E., Srivastava R., 1989: *An Introduction to Applied Geostatistics.* New York 1989. Oxford University Press. 561 pp.

Kliegrova, S., Dubrovsky, M., Metelka, L. (2007): Interpolation methods of weather generator parameters. Program & Abstracts (10th International Meeting on Statistical Climatology, Beijing, China, August 20-24, 2007) , p. 122-123.

Květoň, V., Tolasz, R. (2003): Spatial Analysis of Daily and Hourly Precipitation Amounts with Respect to Terrain. http://www.map.meteoswiss.ch/icam2003/468.pdf

Szentimrey, T. (2002): Statistical Problems Connected with the Spatial Interpolation of Climatic Time Series. Home page: http://www.knmi.nl/samenw/cost719/documents/Szentimrey.pdf

Štěpánek, P. (2007): ProClimDB – software for processing climatological datasets. CHMI, regional office Brno. http://www.climahom.eu/ProcData.html.