



The FORBIO Climate data set for climate analyses

C. Delvaux, M. Journée, and C. Bertrand

Royal Meteorological Institute of Belgium, Brussels, Belgium

Correspondence to: C. Delvaux (charles.delvaux@meteo.be)

Received: 13 January 2015 – Revised: 23 April 2015 – Accepted: 20 May 2015 – Published: 3 June 2015

Abstract. In the framework of the interdisciplinary FORBIO Climate research project, the Royal Meteorological Institute of Belgium is in charge of providing high resolution gridded past climate data (i.e. temperature and precipitation). This climate data set will be linked to the measurements on seedlings, saplings and mature trees to assess the effects of climate variation on tree performance. This paper explains how the gridded daily temperature (minimum and maximum) data set was generated from a consistent station network between 1980 and 2013. After station selection, data quality control procedures were developed and applied to the station records to ensure that only valid measurements will be involved in the gridding process. Thereafter, the set of unevenly distributed validated temperature data was interpolated on a $4\text{ km} \times 4\text{ km}$ regular grid over Belgium. The performance of different interpolation methods has been assessed. The method of kriging with external drift using correlation between temperature and altitude gave the most relevant results.

1 Introduction

The interdisciplinary FORBIO Climate research project wants to scrutinize the adaptive capacity of tree species and predict the future performance of tree species in Belgium under different scenarios of climate change. Towards this objective, the Royal Meteorological Institute of Belgium is in charge of providing high resolution gridded past climate data (i.e. temperature and precipitation) between 1980 and 2013. The gridded daily temperature (minimum and maximum) data set was generated from the network of climatological stations administrated by the Royal Meteorological Institute of Belgium (RMI) since the end of the 19th century. In order to provide the high resolution gridded data set from the data of the station network, data quality control procedures were developed and the performance of different interpolation methods has been assessed.

The RMI network relies on voluntary observers, professional observers on civil and military aerodromes, federal agents, regional agents and employees of private companies. RMI supplies any voluntary observers with a manual rain gauge and a meteorological shelter in about 3/5 of the stations. Air temperature is measured in a shaded enclosure (i.e., Stevenson screen) at a height of approximately 1.5 m above the ground. Maximum and minimum temperatures, for the

previous 24 h, are recorded at 08:00 LCT (local clock time). Minimum temperature is recorded against the day of the observation, and the maximum temperature against the previous day. Temperature records from both voluntary and synoptic stations were considered to generate the 34-year-long (1980–2013) high resolution daily gridded temperature data set over Belgium. Within this period, 278 time series from the RMI database of climate observations contain at least one temperature record. However, a large number of time series contain missing observations. The total number of available observations per day varies between 109 and 175 over the considered time period (see Fig. 1, left panel). The spatial distribution of the corresponding stations within the Belgian territory is provided on the right panel in Fig. 1. Stations for which less than 5 % of the temperature records are missing over the considered time period are referred to as reference stations (i.e. 65 stations, see black dots in Fig. 1). The mean temperature over the 34 years obtained using the daily temperature observations from the 65 reference stations are provided in Fig. 2 for *TN* and *TX*, respectively. Because no measuring technique is perfect and errors can run in meteorological observations for a wide variety of reasons (e.g. Aguilar et al., 2003), data were first quality controlled.

To ensure that only valid measurements will be involved in the gridding process, Sect. 2 describes the quality con-

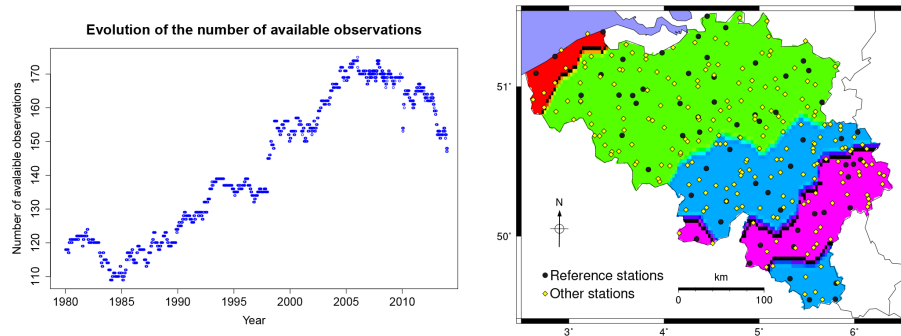


Figure 1. Evolution of the number of available observations per day for the period 1980 to 2013 (left panel) and location of the stations in operation between 1980 and 2013 with the division into 4 climate zones (right panel). Reference stations (i.e. less than 5 % of missing temperature records) are indicated by a black dot while the other stations are represented by a yellow diamond.

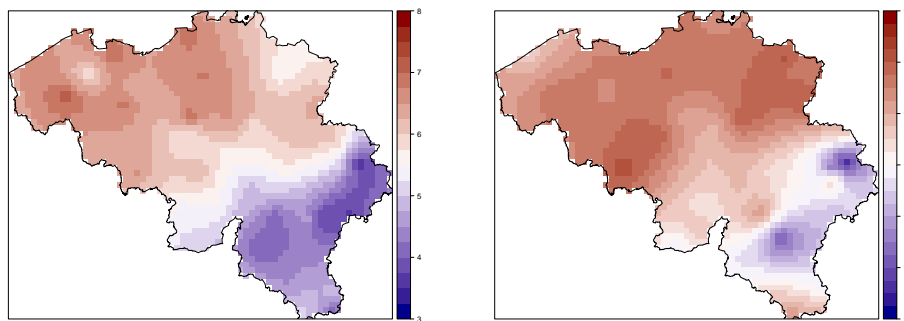


Figure 2. Averaged temperature over the 34 years [°C] for *TN* (left panel) and *TX* (right panels) using reference station records only. The interpolation was made using ordinary kriging method (see Sect. 3). The missing data are estimated by spatial interpolation.

trol (QC) procedures developed. Section 3 presents the considered interpolation methods. Results are presented in Sect. 4 and additional discussions are provided in Sect. 5. Final conclusions are given in Sect. 6.

2 Quality control procedures

2.1 Definition of quality control zones

For quality control purposes, geographical zones of similar temperature characteristics were defined based on the reference stations temperature records. These zones will be used to compare stations of the same zone to each other and to define quality control thresholds specific to each zone. The different zones were identified by *k*-means clustering approach (Hair, 2009) based on mean and variability of the reference stations *TX* and *TN* time series.

A division into 4 zones was selected (see the right panel in Fig. 1). These zones broadly correspond to the Belgian coast (red), Flanders (green), Ardenne (purple) and the rest of Wallonia (blue).

2.2 Preliminary tests (time series evaluation)

In the first instance, two preliminary tests are applied on 1-year long time series to ensure the validity of the studied station. The first test (variability test) compares the data's standard deviation σ of the given station for a time period of one year to the expected limits. The minimum (maximum) limit corresponds to the half (double) of the reference stations mean standard deviation. If σ lies outside the minimum and maximum thresholds, the entire time series is considered as not usable. The second test (lag test) verifies that no one-day time lag is affecting the data. Each 1-year long temperature record with and without lag (forwards or backwards) were compared to neighboring stations. If the data fit is better with a lag, the entire time series is considered as not usable.

2.3 Daily data quality control

All time series that succeeded the preliminary tests are further checked with the following tests for daily data. After an existence test, a first QC module checks for physical limits and flags the data violating these limits (note that values flagged as erroneous fail immediately and do not require further testing). Second, similarly to Feng et al. (2004) and Boulanger et al. (2010), automated QC procedures check the

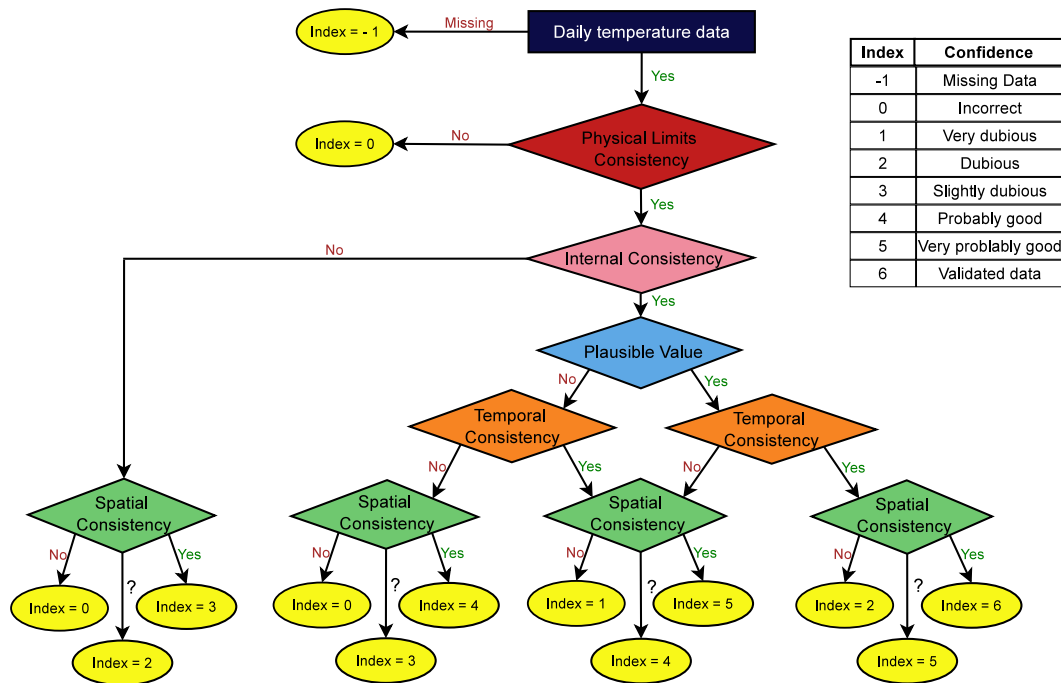


Figure 3. Automated QC applied to temperature records and associated confidence index.

temperature values for more subtle errors. Implementation of the automated QC is diagrammed in Fig. 3. Data are checked for internal consistency, plausible value test (using adapted limits to reflect climatic conditions more precisely than in the first range test), temporal consistency and spatial consistency. At the end of the process, a confidence index ranging from -1 to 6 (i.e. missing data to validated data) is attributed to each datum.

2.3.1 Physical limit consistency test

The aim of the physical limit consistency test is to ensure that temperature records stays within acceptable range limits using extreme climatological boundaries. These boundaries are based on the extreme values of the reference stations for each season and climate zone. To account for extreme values within the different zones, the boundaries are adjusted with a tolerance of 3 °C. Lower and upper limits for the four seasons are given for the Flanders and Ardenne zones in Table 1 for illustration.

2.3.2 Internal consistency test

The internal consistency test ensures that for a given day *TN* is not higher than *TX*. Such an incoherence is technically possible because *TN* is recorded one day later than *TX* but is attached to the same day as *TX*.

Table 1. Extreme climatological boundaries for the Flanders/Ardenne zones per seasons used in the physical limit consistency test.

	Lower limit <i>TN</i> [°C]	Upper limit <i>TN</i> [°C]	Lower limit <i>TX</i> [°C]	Upper limit <i>TX</i> [°C]
winter	-26.5/-27.6	16.5/15.8	-15.2/-17.6	26.8/25.7
spring	-12.8/-15.4	25.7/24	-2.8/-5.9	39.4/36
summer	-1/-5.4	26/25.3	6.8/6	41.2/40
autumn	-18.2/-23.2	21/18.2	-13.9/-14.2	31/29.9

2.3.3 Plausible value test

For the plausible value test, daily *TN* and *TX* values are compared to daily lower and upper bounds. For each of the 4 climate zones, lower and upper bounds for a given temperature data series (i.e. *TN* and *TX*) are constructed by retrieving the highest and lowest daily values on each calendar day *d* of the year from 34 years of data. Assuming that the annual temperature variations follow a sinusoidal wave, a fit of the annual variation of these extreme temperatures is defined using wave function $T_{L/U}(d)$. An observation succeeds the test if it stays within the lower and upper bounds of the corresponding day (Sciuto et al., 2013).

2.3.4 Temporal consistency test

The temporal consistency test analyzes the rate of daily change in order to detect possible anomalies. A spike or step test ($\Delta\max$) checks for a plausible rate of change from a pre-

Table 2. Percentage of the different indexes obtained after the QC procedures. Note that to maintain a good comparison between both types of stations, the “missing data” are not included in the index percentage of index 0 to 6.

Index	Index percentage	
	Reference stations [%]	Other stations [%]
–1	3.32	61.22
0	0.001	0.002
1	0.003	0.005
2	0.008	0.008
3	0.029	0.036
4	0.008	0.014
5	0.593	0.640
6	99.358	99.301

ceding acceptable level. The maximum probable change is based on the 99th percentile change for the 34 years of data. The test is applied by season and climate zone with a one-day time step. Similarly a persistence test (Δ_{\min}) flags the measurements that fail to change by more than a minimum amount.

2.3.5 Spatial consistency test

The spatial consistency test compares the observations at a given station with the observations of the same day made at neighboring stations. The result is suspicious (represented by a question mark in Fig. 3) if the station record differs for more than 2.5 °C from each of its 3 closest neighboring stations or by more than 4 °C of the averaged values of the 3 closest neighboring stations. The observation fails the spatial consistency tests if the two values above become respectively 5 and 7 °C (these values are based on the expertise of the well qualified RMI’s data quality agents). A nearby station is considered only if the difference in elevation with the studied station is below 150 m.

2.4 Confidence Index

More than 5 millions of temperature records (*TN* and *TX*) were faced to the data quality control protocol described above. More than 99.3 % of the analyzed records have succeeded the different tests (confidence index of 6). The proportion of data classified as at least “very probably good” (confidence index greater than 5) was of 99.95 %. Table 2 compares the results in terms of attributed confidence index obtained at the reference stations and at the other stations.

Table 3. Correspondance between Corine Land Cover database and the five land cover types used in this study.

Index	Corine index	Corine description
1	1–11	Artificial surfaces
2	12–15	Arable land
3	16–22	Permanent crops, pastures and heterogenous agricultural areas
4	23–34	Forest and semi-natural areas
5	35–44	Wetlands and water bodies

3 Interpolation methods

Long-term climate patterns observed across the globe are a result of a combination of many different processes that manifest themselves at many spatial scales. It is assumed that most of those patterns occur at scales large enough to be adequately reflected in the station data, and thus are not explicitly accounted for by the major interpolation methods. The main physiographic features affecting spatial patterns of climate are terrain and water bodies (e.g., Daly, 2006). Several additional spatial climate foreign factors are most important at scales of less than 1 km but may also have effects at larger scales. These factors include slope and aspect, riparian zones and land use/land cover (Lookingbill and Urban, 2003; McCutchan and Fox, 1986; Bolstad et al., 1998; Dong et al., 1998).

For each day of the considered 34-year-long time period (i.e. 1 January 1980 to 31 December 2013), all the validated (i.e. confidence index of 6) stations temperature records (i.e. *TN* and *TX*) were interpolated on a regular 4 km × 4 km grid over Belgium. To predict the unknown values from the unevenly distributed temperature records observed at known locations, the performance of different interpolation methods has been assessed: inverse distance weighting (IDW), ordinary kriging (OK) and kriging with external drift (KED; Wackernagel, 1995) that is able to handle densely sampled auxiliary variables highly correlated with the parameter of interest. In this study, KED was used with either the orography (KED₁) or land cover types (KED₂) as a drift, as well as simultaneously with the two drifts (KED₁₂). For the kriging methods, the parameters used to estimate the semivariogram were fixed in order to have an exact interpolation at the station location (i.e. the estimation satisfies the observation at the station location) with a fixed range of 50 km and an exponential semivariogram model.

Left panel in Fig. 4 displays the Belgian orography at the 4x4 km spatial resolution. Similarly, right panel in Fig. 4 presents the spatial distribution of the 5 land cover classes considered here. These classes are derived from the 100 m spatial resolution CORINE land cover database (Bossard and Feranec, 2000). Basically the 44 classes of the CORINE database are merged in 5 main classes as detailed in Table 3 and reprojected in a 4x4 km regular grid over Belgium (the

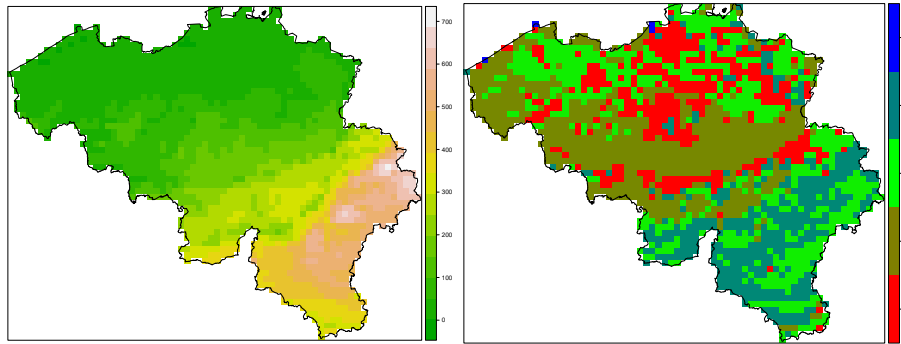


Figure 4. 4 km resolution orography of Belgium [m] (left panel) and 4 km resolution CORINE Land Cover in Belgium (right panel).

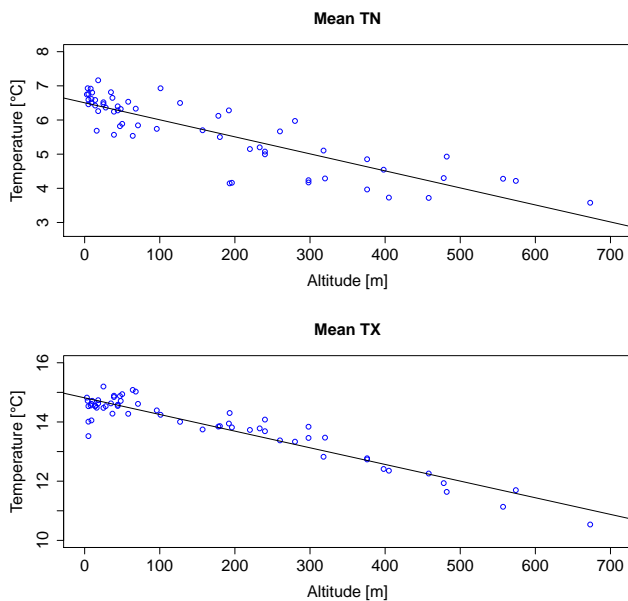


Figure 5. Correlation between station elevation and daily mean temperature (for the reference stations). The correlation coefficients are of -0.85 and -0.93 for TN and TX , respectively.

dominant land cover class within a given grid point is attributed to the entire grid point).

The various interpolation approaches are validated by leave-one-out cross-validation. For each day of the considered period, the cross-validation root mean square error ($RMSE_{cv}$) is computed from the differences between the prediction P and the actual measurements M at the n stations:

$$RMSE_{cv} = \sqrt{\frac{1}{n} \sum_{i=1}^n (P(x_i) - M(x_i))^2}$$

where $\{x_i\}_{i=1, \dots, n}$ is the location of the climate stations.

To determine the best performing method, two indices are derived from the daily $RMSE_{cv}$. First, the averaged over the 34 years (i.e. $RMSE_{cv}^{AVG}$) and second, the frequency of the method providing the best daily $RMSE_{cv}$ over the 34 years (i.e. $RMSE_{cv}^{FREQ}$).

4 Results

Table 4 summarizes the performance of the investigated interpolation methods in terms of $RMSE_{cv}^{AVG}$ and $RMSE_{cv}^{FREQ}$ index for both TN and TX , respectively. Results indicate that for TN all methods perform quite similarly: the averaged accuracy ($RMSE_{cv}^{AVG}$) ranges from 0.867 for the “worst” method (IDW) to 0.818 for the best one (KED_1). By contrast the interpolated TX fields appear more sensitive to the considered methods. As for TN , the best performing method is KED_1 and the worst one is IDW but the magnitude of the difference between the two methods in terms of $RMSE_{cv}^{AVG}$ is larger (i.e. 0.14 °C vs. 0.05 °C for TN). This difference between TN and TX can be explained by the larger correlation found over our domain between the elevation and TX than for TN (see Fig. 5).

Accounting for the land cover type in the interpolation scheme does not improve the results for both TN and TX . The difference in terms of $RMSE_{cv}^{AVG}$ between OK and KED_2 is clearly negligible (i.e. 0.852 °C vs. 0.854 °C for TN and 0.721 °C vs. 0.723 °C for TX). Similarly using the land cover type in addition to the orography in the kriging with external drift approach does not significantly modify the performance obtained when only considering the orography as drift (e.g. $RMSE_{cv}^{AVG}$ of 0.818 °C for KED_1 vs. 0.820 °C for KED_{12} in the case of TN and 0.601 °C for KED_1 vs. 0.605 °C for KED_{12} in the case of TX , respectively). This is not so surprising in view of the spatial resolution adopted for the gridding (e.g. $4 \text{ km} \times 4 \text{ km}$) as the land use/land cover variations are expected to be most important below 1 km. Analysis of the $RMSE_{cv}^{FREQ}$ provided in Table 4 confirm that the best performing interpolation method in terms of $RMSE_{cv}$ is the kriging using the orography as external drift (KED_1) for both TN and TX (e.g. $RMSE_{cv}^{FREQ}$ of 43 % for TN and $RMSE_{cv}^{FREQ}$ of 77.1 % for TX). What is interesting to note regarding the $RMSE_{cv}^{FREQ}$ in Table 4 is the score obtained by the kriging using the land cover type as external drift (KED_2). Indeed, for both TN and TX , the value of this index for the KED_2 is lower than for IDW (i.e. 7.9 % vs. 14.5 % for TN and 0.3 % vs. 0.6 % for TX).

Table 4. Overall performance of the different interpolation methods over the entire time period expressed in terms of $\text{RMSE}_{\text{cv}}^{\text{AVG}}$ and $\text{RMSE}_{\text{cv}}^{\text{FREQ}}$ for both TX and TN .

	$\text{RMSE}_{\text{cv}}^{\text{AVG}}$ (TN) [$^{\circ}\text{C}$]	$\text{RMSE}_{\text{cv}}^{\text{AVG}}$ (TX) [$^{\circ}\text{C}$]	$\text{RMSE}_{\text{cv}}^{\text{FREQ}}$ (TN) [%]	$\text{RMSE}_{\text{cv}}^{\text{FREQ}}$ (TX) [%]
IDW	0.867	0.738	14.5	0.6
OK	0.852	0.721	15.1	1.1
KED ₁	0.818	0.601	43	77.1
KED ₂	0.854	0.723	7.9	0.3
KED ₁₂	0.820	0.605	19.5	20.9

Finally, Fig. 6 displays the time evolution of the annual mean daily RMSE_{cv} over the considered 34-year-long time period obtained by using the OK and the KED₁ daily interpolation methods, respectively. Both methods present a clear reduction of the RMSE_{cv} throughout the time for TX with the largest reduction obtained by the KED₁ approach (see left panel in Fig. 6). Similar behavior is also observed for TN while for this parameter the dispersion is a bit more larger than for TX (see right panel in Fig. 6). The reduction of the annual mean daily RMSE_{cv} has to be put in connection with the increasing number of stations involved in the interpolation process with time (see Fig. 1, left panel).

5 Discussion

As an additional validation exercise, the newly developed FORBIO climate data set was compared to existing data sets.

Towards this objective, the RMSE results presented in Sect. 4 are compared to these obtained from the HYRAS data set which cover Germany (Frick et al., 2014). This data set has a similar resolution (5 km) and use the Optimal Interpolation method. The fivefold cross-validation of HYRAS gives an averaged RMSE of 1.39 $^{\circ}\text{C}$ versus values from 0.601 $^{\circ}\text{C}$ to 0.867 $^{\circ}\text{C}$ for the FORBIO climate data set. The better station density and the smaller elevation differences between stations can explain the better results obtained in the FORBIO data set.

The new FORBIO data set can also be compared to the existing E-OBS data set (version 10.0) (Haylock et al., 2008), which provides daily temperature data for the period 1950–2013 on a regular 0.25 $^{\circ}$ grid (approximately 25 km). 16 Belgian synoptic stations are used in the E-OBS data set. To allow a comparison at the same resolution, the FORBIO data set was first degraded at the E-OBS resolution. Figure 7 shows the differences between the two data sets using the mean maximum temperature over the 34 years. The benefit of the original FORBIO data set spatial resolution is evident for the user. At the degraded resolution, a major difference between the two data sets (> 1 $^{\circ}\text{C}$) appears in the Hainaut Province. Only one station located in the Hainaut Province is considered with the E-OBS data set (i.e. the Chièvre synoptic station, black dot in Fig. 7) and this station appears

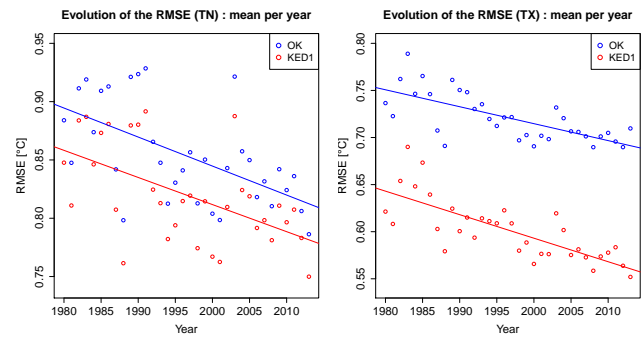


Figure 6. Evolution of the averaged annual RMSE_{cv} and regression line using ordinary kriging (OK) and kriging with external drift using orography (KED₁) for TN (left panel) and TX (right panel).

too cold compared to the neighboring climatological stations used in the FORBIO data set in addition to the Chièvre station records. This comparison is also true for minimum temperature.

6 Conclusions

To meet the needs of the FORBIO Climate research project, a 34-year-long (1980–2014) daily gridded temperature (minimum and maximum) has been produced over Belgium at a 4 km \times 4 km spatial resolution. Because no measuring technique is perfect and errors can run in meteorological observations for a wide variety of reasons, data quality control procedures were developed and applied to temperature records performed within the Belgian climatological station network operated by RMI prior to undergo the daily gridding process. More than 5 millions of daily temperature records (TN and TX) were analyzed in depth and about 0.7 % of these were discarded. The performance of 5 different interpolation methods was assessed over the Belgian domain ranging from the simple inverse-distance weighting approach to the kriging with external drift methods. Two auxiliary drifts have been considered (i.e. the orography and the land cover type) either individually or in combination. Because of the spatial resolution of 4 km \times 4 km adopted for the data set, it has been found that accounting for the land cover type in the interpolation process of temperature over Belgium is not relevant. By contrast using the orography as external drift in the kriging interpolation scheme provides the best results in term of RMSE_{cv} and $\text{RMSE}_{\text{cv}}^{\text{FREQ}}$ for both TX and TN . It is however worth pointing out that the influence of the selected methods in the accuracy of the interpolated temperature field is only well apparent for TX . For TN the performance of the different interpolation methods were found rather similar. Based on these results the daily gridded temperature (TN and TX) were carried out with the kriging with external drift method using the orography as auxiliary highly correlated parameter. The results obtained are similar to the E-OBS data set excepted in

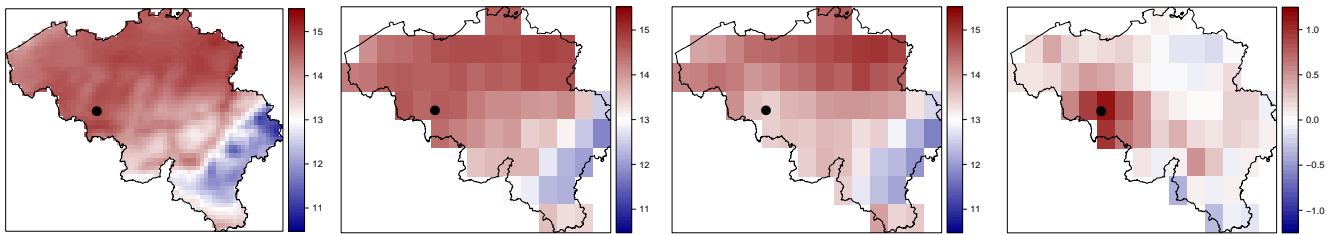


Figure 7. Averaged value for TX over the 34 years [$^{\circ}\text{C}$]: FORBIO (4 km resolution), FORBIO (25 km resolution), E-OBS (25 km resolution), differences between FORBIO and E-OBS. The black dot corresponds to the Chièvre station.

the Hainaut province where the E-OBS data set is colder. A minimum of 80 stations per day were always used in the gridding process for both TN and TX . Note that in average over the 34 years a bit more stations were involved in the spatial interpolation of TX than for TN (i.e. 137 vs. 132 stations, respectively). Finally, because the number of stations involved in the daily gridding process has increased with time (even if the total number of available stations has started to decrease in the last decade covered by the dataset), the daily RMSE_{cv} present a clear reduction as a function of time for both TX and TN . As an example, the annual mean daily RMSE_{cv} has decreased by 0.085°C from 1980 to 2013 for TX and by 0.068°C for TN .

Acknowledgements. This work was supported by the Belgian Science Policy Office (BELSPO) through the BRAIN-be program contract NR BR|132|A1|FORBIO CLIMATE “Adaptation potential of biodiverse forests in the face of climate change”. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (<http://ensembles-eu.metoffice.com>) and the data providers in the ECA & D project (<http://www.ecad.eu>). The “gstat” Package (version 1.0-22) in R (Pebesma, 2004) was used for geostatistical interpolation.

Edited by: I. Auer

Reviewed by: three anonymous referees

References

- Aguilar, E., Auer, I., Brunet, M., Peterson, T., and Wieringa, J.: Guidelines on climate metadata and homogenization, World Climate Programme Data and Monitoring WCDMP-No. 53, WMO-TD No. 1186, World Meteorological Organization, Geneva, 2003.
- Bolstad, P. V., Swift, L., Collins, F., and Régnière, J.: Measured and predicted air temperatures at basin to regional scales in the southern Appalachian mountains, *Agr. Forest Meteorol.*, 91, 161–176, 1998.
- Bossard, M. and Feranec, J.: Otahel, Jaffrain, Gabriel, CORINE Land Cover technical guide – Addendum 2000, Tech. rep., Technical Report 40, EEA, Copenhagen, <http://www.eea.eu.int> (last access: 8 January 2015), 2000.
- Boulangier, J.-P., Aizpuru, J., Leggieri, L., and Marino, M.: A procedure for automated quality control and homogenization of historical daily temperature and precipitation data (APACH): part 1: quality control and application to the Argentine weather service stations, *Climatic Change*, 98, 471–491, 2010.
- Daly, C.: Guidelines for assessing the suitability of spatial climate data sets, *Int. J. Climatol.*, 26, 707–721, 2006.
- Dong, J., Chen, J., Brosofske, K., and Naiman, R.: Modelling air temperature gradients across managed small streams in western Washington, *J. Environ. Manage.*, 53, 309–321, 1998.
- Feng, S., Hu, Q., and Qian, W.: Quality control of daily meteorological data in China, 1951–2000: a new dataset, *Int. J. Climatol.*, 24, 853–870, 2004.
- Frick, C., Steiner, H., Mazurkiewicz, A., Riediger, U., Rauthe, M., Reich, T., and Gratzki, A.: Central European high-resolution gridded daily data sets (HYRAS): Mean temperature and relative humidity, *Meteorol. Z.*, 23, 15–32, 2014.
- Hair, J. F.: *Multivariate data analysis*, Prentice Hall, Upper Saddle River, 2009.
- Haylock, M., Hofstra, N., Klein Tank, A., Klok, E., Jones, P., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *J. Geophys. Res.-Atmos.*, 113, D20119, doi:10.1029/2008JD10201, 2008.
- Lookingbill, T. R. and Urban, D. L.: Spatial estimation of air temperature differences for landscape-scale studies in montane environments, *Agr. Forest Meteorol.*, 114, 141–151, 2003.
- McCutchan, M. H. and Fox, D. G.: Effect of elevation and aspect on wind, temperature and humidity, *J. Clim. Appl. Meteorol.*, 25, 1996–2013, 1986.
- Pebesma, E. J.: Multivariable geostatistics in S: the gstat package, *Comput. Geosci.*, 30, 683–691, 2004.
- Sciuto, G., Bonaccorso, B., Cancelliere, A., and Rossi, G.: Probabilistic quality control of daily temperature data, *Int. J. Climatol.*, 33, 1211–1227, 2013.
- Wackernagel, H.: *Multivariable geostatistics: an introduction with applications*, Springer-Verlag, Berlin, 1995.