SLOCLIM: a high-resolution daily gridded precipitation and temperature dataset for Slovenia

Nina Škrk1, Roberto Serrano-Notivoli2, Katarina Čufar1, Maks Merela1, Zalika Črepinšek3, Lučka Kajfež Bogataj3, and Martín de Luis4

1Department of Wood Science and Technology, Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia
2Department of Geography, Autonomous University of Madrid, Madrid, 28049, Spain
3Department of Agronomy, Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia
4Department of Geography and Regional Planning and Environmental Sciences Institute (IUCA), University of Zaragoza, 50009, Spain

Correspondence: Nina Škrk (nina.skrk@bf.uni-lj.si)

Received: 2 November 2020 – Discussion started: 14 January 2021
Revised: 11 June 2021 – Accepted: 22 June 2021 – Published: 29 July 2021

Abstract. We present a new publicly available daily gridded dataset of maximum and minimum temperature and precipitation data covering the whole territory of Slovenia from 1950 to 2018. It represents the great variability of climate at the crossroads between the Mediterranean, Alpine and continental climatic regimes with altitudes between 0–2864 m a.s.l. We completely reconstructed (quality control and gap filling) the data for the three variables from 174 observatories (climatological, precipitation and automatic stations) with the original records all over the country. A comprehensive quality control process based on the spatial coherence of the data was applied to the original dataset, and the missing values were estimated for each day and location independently. Using the filled data series, a grid of $1 \times 1$ km spatial resolution with 20,998 points was created by estimating daily temperatures (minimum and maximum) and precipitation, as well as their corresponding uncertainties at each grid point. In order to show the potential applications, four daily temperature indices and two on precipitation were calculated to describe the spatial distribution of (1) the absolute maximum and minimum temperature, (2) the number of frost days, (3) the number of summer days, (4) the intensity of precipitation and (5) the maximum number of consecutive dry days. The use of all the available information, the complete quality control and the high spatial resolution of the grid allowed for an accurate estimate of precipitation and temperature that represents a precise spatial and temporal distribution of daily temperatures and precipitation in Slovenia. The SLOCLIM dataset is publicly available at https://doi.org/10.5281/zenodo.4108543 and http://www.sloclim.eu (last access: 10 June 2021) and can be cited as Škrk et al. (2020).

1 Introduction

Complex territories demand high-quality climatic data to properly describe and understand those events that are usually masked by global datasets or others at coarser resolutions than those required by the high-frequency climatic phenomena, like extreme events. Slovenia is an outstanding case study of a challenging territory for climate research. The high heterogeneity of the terrestrial surfaces and the greatly varied climatic regime in a 20,271 km$^2$ country represented by lowlands and mountains within an altitudinal range of nearly 3000 m serves as example of extreme climatic diversity in a relatively small region. Continental, Alpine and sub-Mediterranean climates meet and interact in the area, including two climatic zones that are particularly affected by global warming (the Mediterranean and Alpine regions) (Giorgi,
Slovenia also faces a high frequency and magnitude of extreme events (Cegnar, 2016), which are reflected at an annual scale where, for instance, precipitation totals over 3500 mm can be locally measured in the northwest, while less than 800 mm can be found in the northeast. Moreover, temperature ranges from −35 to +40 °C (Bertalanic et al., 2006) show the very extreme character of the seasons. These extremes affect the environment and life of the population through damage to infrastructure, agricultural and afforested areas. More in-depth knowledge of climatic conditions is needed to design prevention measures in this context.

The Environmental Agency of the Republic of Slovenia (ARSO) maintains a very good archive of meteorological data (ARSO, 2020) collected through meteorological stations covering the whole territory and all the different environments. However, the density of these observations varies in space and time, as in most European countries, with a lack of meteorological stations in sparsely populated areas including mountains (especially at higher altitudes), agricultural and afforested areas, which represent most of the Slovenian territory. Furthermore, the data provided by the stations are often of varying length and partly incomplete due to various reasons, such as measurement errors, mechanical failure of devices, missing and interrupted measurements, and abandonment and relocation of meteorological stations, amongst others. In addition, considering the complex orography and proximity to the coast, the heterogeneity of terrestrial surfaces, where weather conditions change rapidly, a dense dataset with an integrated approach is needed for the whole country.

High-resolution gridded datasets allow for precise characterization of the climatic variables and are important for climatic analysis and hydrological hazard predictions, agricultural management, forestry and many other applications related to climate variables, as they help to identify extreme climatic conditions that can be analysed and used for prevention measures (Herrera et al., 2012). In this sense, gridded datasets have an advantage over station-based datasets, as they can provide climatic data for any location in the territory in the form of serially complete data series that allow long-term analysis. They allow us to examine spatial and temporal climatic variations providing an equal area grid where each grid value is under validation control (Haylock et al., 2008), and missing information from stations is reconstructed. On the other hand, they have some limitations, as they are models and not real measurements. How much they differ from real data depends on the accuracy of the method and locations/density of the meteorological stations. Furthermore, the reconstruction of long periods with missing data can lead to inhomogeneities in the temporal structure of the reconstructed series. When the temporal scale is increased from monthly to daily data, the spatial and temporal variability is also increased, and the uncertainty can quickly escalate.

With regard to Slovenia, and to the best of our knowledge, until recently the only gridded climate dataset was the one created by Dolinar (2016), and this contains temperature and precipitation data with a 1 × 1 km spatial resolution from 1961 onwards. Although this dataset is valuable to describe the mean climate and regimes, its monthly temporal scale is insufficient to address the frequency, duration and intensity of the extremes, which clearly define the variable climate of Slovenia. Based on the experience of previous studies creating daily datasets at the regional level, both for precipitation (Chaney et al., 2014; Hernández et al., 2016; Serrano-Notivoli et al., 2017b; Yatagai et al., 2012) and temperature (Lussana et al., 2018; Serrano-Notivoli et al., 2019), a new high-resolution climate dataset for the Slovenian territory is suitable considering the high-density network of stations that is available. Serrano-Notivoli et al. (2017a, b, 2019) developed a method to reconstruct and create daily gridded datasets using all the available information in three stages: (1) quality control of raw data, (2) estimate of missing values and (3) prediction of new values over a gridded dataset. This methodology has been proven to work well in dense networks such as the one available for mainland Spain and its island territories (Serrano-Notivoli et al., 2017a, b, 2019) but also for specific mountain areas with a large diversity of environments (Cuadrat et al., 2020; Martínez del Castillo et al., 2019).

We present a new high-resolution daily gridded precipitation and temperature dataset for Slovenia (SLOCLIM) with a spatial resolution of 1 × 1 km for the period 1950 to 2018. We used all the available climatic information in the country gathered from stations in all possible environments to create a grid based on thoroughly reconstructed data series in three steps: (a) a comprehensive quality control of the original data, (b) the estimate of new values corresponding to the missing data and (c) the creation of the gridded dataset using the reconstructed series.

2 Input data

We used the meteorological records from 174 stations of the Slovenian Meteorological Agency (ARSO) regularly distributed over the territory of Slovenia (Fig. 1) to create the gridded datasets. Although some data series begin before 1950, we decided to limit the research to the years 1950 to 2018, when the station network remained stable over time and space.

All the stations recorded daily precipitation, while 167 recorded both precipitation and temperature (ARSO, 2020). The altitudinal pattern showed a lower availability of data at the highest elevations, with less than five stations over 1000 m a.s.l. (Fig. 2b) and corresponding with less populated and afforested areas. The average amount of annual data from stations also varied over time (Fig. 2a), with a rapid increase in the first 20 years, reaching the highest availability...
Figure 1. Location of Slovenia in Europe (left) and spatial distribution (right) of the 174 meteorological stations with temperature and precipitation information (yellow dots) and only precipitation (blue dots) (Esri, 2020; GISCO, 2020; GURS, 2020).

in 1970 (107) and decreasing thereafter. The early measurements contained more precipitation than temperature data.

3 Methodology

A complete reconstruction of the original daily data series was performed following the three-step methodology developed in Serrano-Notivoli et al. (2017a, b, 2019), which is based on an (1) exhaustive quality control (QC) of the raw series, (2) the estimate of new values for missing records using all the available information and local regression models, and (3) the estimates of new values in a regular grid using the reconstructed series.

The QC detected and removed the most obvious erroneous data in the original dataset based on different criteria created to keep the internal and the spatial coherence of the data. For precipitation five criteria were used: (1) suspect data – the observed value was over zero and all its 10 nearest observations were zero; (2) suspect zero – the observed value was zero and all its 10 nearest observations were over zero; (3) suspect outlier – the magnitude of the observed value was 10 times higher or lower than that one predicted by its 10 nearest observations; (4) suspect dry day – the observed value was zero, wet probability was over 99 %, and predicted magnitude was over 5 mm; and (5) suspect wet day – the observed value was over 5 mm, dry probability was over 99 %, and predicted magnitude was under 0.1 mm. In order to determine a day as wet or dry, for each day and location independently, the 10 nearest observations were codified as 0 or 1 depending on the absence or not of precipitation on that day, and then a binomial generalized linear model (GLM) was fitted to these values using their longitude, latitude and altitude as covariates. The model was then applied to obtain a prediction at the target location using its corresponding covariates values.

If the predicted value (predicted probability), ranging from 0 to 1, was higher than 0.5, it was considered a wet day.

For temperature we also used five criteria: (1) internal coherence ($T_{\text{MAX}} > T_{\text{MIN}}$), (2) removal of months containing less than 3 d of data, (3) removal of those days out of range considering maximum temperature ($T_{\text{MAX}} \geq 50 ^\circ \text{C}$ or $T_{\text{MAX}} \leq -30 ^\circ \text{C}$) and minimum temperature ($T_{\text{MIN}} \geq 40 ^\circ \text{C}$ or $T_{\text{MIN}} \leq -35 ^\circ \text{C}$), (4) removal of all days in a month with a standard deviation of temperature equal to zero (suspect repeated values in the series), and (5) removal of all days in a month if the sum of the differences between $T_{\text{MAX}}$ and $T_{\text{MIN}}$ was equal to zero (suspect duplicated values in $T_{\text{MAX}}$ and $T_{\text{MIN}}$). These criteria helped us to evaluate, amongst others, situations of internal incoherence like abnormal values, significant deviations from other values or duplicated values. In such cases all suspects were removed. In order to perform the QC process by applying the above-referenced criteria, we used generalized linear mixed models (GLMMs) and generalized linear models (GLMs) to calculate the reference values (RVs) for each day and location of the original dataset. The information from the 10 nearest stations (15 in temperature) varied according to the data availability. In order to estimate precipitation and temperature variables, the observations acted as dependent variables, and the associated geographical information of each station participating in the model (latitude, longitude, altitude and distance from the coast) was the independent variable. The calculated RVs were then compared with their corresponding original data to assess their quality and remove suspect values following the QC criteria.

After QC, the second stage consisted of calculating new RVs based on the cleaned dataset that were used to replace the missing values. A leave-one-out cross-validation (LOOCV) was implemented, where RVs were calculated for all
days and locations including those for which an observation exists but without using that observation, for comparison and validation purposes. This whole process produced a serially complete dataset.

In the third stage, new values at each grid location of the $1 \times 1$ km gridded dataset were estimated based on the reconstructed point-based dataset, including a measure of uncertainty for each location and day, as calculated from the standard error of the models. The advantage of this methodology is that all of the available data can be used as there are no restrictions imposed due to the length or structural characteristics of the series.

Based on the gridded dataset created from the original data and with the described method, we computed four temperature and two precipitation indices to show the potential applications of the grid: (1) the mean annual maximum value of daily maximum temperature, (2) the mean annual minimum value of daily minimum temperature, (3) the average annual count of days when daily minimum temperature was below 0°C (frost days), (4) the average annual count of days when daily maximum temperature was over 25°C (summer days), (5) maximum number of consecutive dry days per year with daily precipitation of less than 1 mm and (6) daily precipitation intensity (annual precipitation/number of wet days). All the indices were presented with their corresponding uncertainty estimates. The indices were calculated through a Monte Carlo approach by computing 1000 random realizations for each grid cell. These realizations were based on a normal distribution where the mean was the value of the estimate and the standard deviation was the value of the uncertainty. The indices were computed for all realizations, and the final maps represent the median and the percentage of variation based on the standard deviation of all the realizations.

The maps throughout this article were created using ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and used herein under license.

Copyright ©Esri. All rights reserved. For more information about Esri® software, please visit http://www.esri.com (last access: 10 June 2021).

4 Results

4.1 Quality control

The original missing daily data, considering the whole period (1950–2018) for all stations, were 60.2% for precipitation and 59.7% for maximum and minimum temperature. The quality control was applied to the remaining 39.8% and 40.3% of the data, respectively.

4.1.1 Precipitation

The number of removed data varied over the years, increasing from 1950 onwards and reaching a peak in 1967, and then decreasing until the end of the period (Fig. 3a). From the removed data, the Suspect data and Suspect zeros criteria, which usually represent estimates of low precipitation values based on nearest observations, were the most frequent. While these criteria are commonly frequent in this kind of QC (Serrano-Notivoli et al., 2017b, 2019), only 1.26% of the original raw observations were removed.

4.1.2 Temperature

The quality control process removed 1.28% of the original $T_{\text{MAX}}$ data and 0.86% of the original $T_{\text{MIN}}$ data (Fig. 3b). According to the monthly criteria, 1090 (0.06%) pieces of data for the $T_{\text{MAX}}$ and 2139 (0.13%) for the $T_{\text{MIN}}$ were removed, and according to the daily criteria, 20 491 (0.21%) values were removed for the $T_{\text{MAX}}$ and 12 283 (0.73%) values for the $T_{\text{MIN}}$. 

Figure 2. Original data: (a) mean annual number of observations of precipitation and temperature, and (b) number of stations by elevation ranges.
Figure 3. Total number of removed precipitation (a) and temperature (b) observations from the original dataset by the different criteria per year. The criteria are explained in more detail in Serrano-Notivoli et al. (2017b, 2019).

4.2 Measurements – estimation comparison

4.2.1 Precipitation

The number of observed zero precipitation days (dry days) in Slovenia was 1 006 891, and the number of estimated ones (only for corresponding days with measurements) was 1 011 337, with the difference being +0.44 % in favour of estimated ones. Terming the wet days as positive (observed $P > 0$) and the dry days as negative (observed $P = 0$), the true negative rate (RV = 0 and $P = 0$) was over 92 %, and the true positive rate (RV > 0 and $P > 0$) was over 85 % (Table 1). To a large extent the false negatives (RV = 0 and $P > 0$) and false positives (RV > 0 and $P = 0$) were due to the prediction of precipitation for days with low amounts of this. In events with very low amounts of precipitation, the estimate of the probability of occurrence was likely to be dry, despite the fact that the station could register a minimum quantity of rain (usually under 1 or 2 mm). This causes a small difference in amounts and is becoming more distinct in the dry/wet accuracy assessment.

The highest Pearson correlations (> 0.94) were found in the comparison by days between the original dataset and corresponding estimates rather than in the comparison by stations (Fig. 4). This indicates the complexity of the geographical variability (elevation, orientation, landscape configuration, etc.) between stations. The highest correlation between the measurements and estimates by stations was found for medians without zeros (only considering days with $P > 0$ in measurements and estimations). The comparison showed the lowest correlation in the 95th percentiles of wet days when comparing by both stations and days. Also, a slight underestimation of high precipitation amounts was noted, probably due to a smoothness of the estimates in situations with very local high precipitation events.

The histogram of estimated and observed precipitation showed a good agreement (Fig. 5). However, values below

<table>
<thead>
<tr>
<th></th>
<th>Observed = 0</th>
<th>Observed &gt; 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted = 0</td>
<td>92.55</td>
<td>14.13</td>
</tr>
<tr>
<td>Predicted &gt; 0</td>
<td>7.45</td>
<td>85.87</td>
</tr>
</tbody>
</table>

Table 1. Accuracy of the wet/dry day estimates: percent of observed dry ($P$ (precipitation) = 0) and wet ($P > 0$) days, and percent of predicted dry (RV = 0) and wet (RV > 0) days on observed dry and wet days.
Figure 4. Comparison between measurements and estimates by stations (a, b, c) and by days (d, e, f). (a, d) Comparison of mean precipitation; (b, e) comparison of median precipitation in wet days; (c, f) comparison of 95th percentiles in wet days. Dashed lines represent ±1 standard deviation of the data.

Figure 5. Histograms of observed and predicted daily precipitation frequency.

1 mm were slightly underestimated in some cases. The agreement between the histograms was very high above 1 mm.

The fit of the estimated values by altitudinal ranges showed a slight overestimation in the ratio of means (RM) between 700 and 900 m a.s.l. (Table 2). Overestimation was higher above 1100 m a.s.l. with the exception of 1500–2000 m a.s.l., where the estimation was lower than the measurements. The ratio of the standard deviation (RSD) also showed differences between estimated and observed values, especially between 1100 and 1300 m a.s.l. The mean absolute error (MAE) increased with altitude, which can be attributed to the decreasing number of available stations with altitude.

The ratio of means (RM) indicated that the monthly aggregates of daily precipitation in all months were slightly underestimated. The ratio of standard deviation (RSD) was very close to 1 through all months (Table 3). In December, however, the RSD was below 0.95, indicating a bias in the variance estimate. In the other months there are no substantial biases in the variance estimation. Overall, all monthly statistics are relatively homogeneous over the year due to the evenly distributed precipitation over the year.
Table 2. The leave-one-out cross-validation (LOO-CV) statistics showing the goodness of fit between measurements and estimations of daily precipitation separated by altitudes (m a.s.l.). N: number of stations; %OBS: percentage of observed precipitation; %PRE: percentage of predicted precipitation; MAE: mean absolute error; ME: mean error; RM: ratio of means; RSD: ratio of standard deviations. The results were rounded to two decimal places.

<table>
<thead>
<tr>
<th>Altitude</th>
<th>N</th>
<th>%OBS</th>
<th>%PRE</th>
<th>MAE</th>
<th>ME</th>
<th>RM</th>
<th>RSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–300</td>
<td>48</td>
<td>33.30</td>
<td>34.20</td>
<td>1.33</td>
<td>−0.01</td>
<td>1.00</td>
<td>0.97</td>
</tr>
<tr>
<td>300–500</td>
<td>43</td>
<td>28.70</td>
<td>27.10</td>
<td>1.41</td>
<td>−0.31</td>
<td>0.91</td>
<td>0.86</td>
</tr>
<tr>
<td>500–700</td>
<td>18</td>
<td>15.50</td>
<td>14.50</td>
<td>1.61</td>
<td>−0.45</td>
<td>0.90</td>
<td>0.87</td>
</tr>
<tr>
<td>700–900</td>
<td>8</td>
<td>7.90</td>
<td>8.80</td>
<td>2.38</td>
<td>0.39</td>
<td>1.09</td>
<td>1.16</td>
</tr>
<tr>
<td>900–1100</td>
<td>8</td>
<td>8.30</td>
<td>7.30</td>
<td>2.72</td>
<td>−0.84</td>
<td>0.85</td>
<td>0.81</td>
</tr>
<tr>
<td>1100–1300</td>
<td>2</td>
<td>0.80</td>
<td>1.20</td>
<td>3.16</td>
<td>1.36</td>
<td>1.37</td>
<td>1.38</td>
</tr>
<tr>
<td>1300–1500</td>
<td>2</td>
<td>1.80</td>
<td>2.10</td>
<td>3.17</td>
<td>0.59</td>
<td>1.13</td>
<td>1.28</td>
</tr>
<tr>
<td>1500–2000</td>
<td>3</td>
<td>3.60</td>
<td>2.70</td>
<td>3.72</td>
<td>−1.78</td>
<td>0.74</td>
<td>0.67</td>
</tr>
<tr>
<td>&gt; 2000</td>
<td>1</td>
<td>2.10</td>
<td>3.70</td>
<td>6.05</td>
<td>4.11</td>
<td>1.74</td>
<td>1.95</td>
</tr>
</tbody>
</table>

Table 3. The leave-one-out cross-validation (LOO-CV) statistics for precipitation, showing the goodness of fit between measurements and estimates of monthly aggregates. MAE: mean absolute error; ME: mean error; RM: ratio of means; RSD: ratio of standard deviations. The results were rounded to two decimal places.

<table>
<thead>
<tr>
<th>MAE</th>
<th>ME</th>
<th>RM</th>
<th>RSD</th>
<th>Pearson</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>1.93</td>
<td>−0.32</td>
<td>0.96</td>
<td>0.95</td>
<td>0.82</td>
</tr>
<tr>
<td>Feb</td>
<td>1.74</td>
<td>−0.29</td>
<td>0.96</td>
<td>0.95</td>
<td>0.86</td>
</tr>
<tr>
<td>Mar</td>
<td>1.95</td>
<td>−0.26</td>
<td>0.97</td>
<td>0.96</td>
<td>0.84</td>
</tr>
<tr>
<td>Apr</td>
<td>2.35</td>
<td>−0.35</td>
<td>0.97</td>
<td>0.99</td>
<td>0.78</td>
</tr>
<tr>
<td>May</td>
<td>2.62</td>
<td>−0.34</td>
<td>0.97</td>
<td>0.99</td>
<td>0.78</td>
</tr>
<tr>
<td>Jun</td>
<td>3.08</td>
<td>−0.58</td>
<td>0.96</td>
<td>0.97</td>
<td>0.76</td>
</tr>
<tr>
<td>Jul</td>
<td>2.95</td>
<td>−0.61</td>
<td>0.95</td>
<td>0.98</td>
<td>0.79</td>
</tr>
<tr>
<td>Aug</td>
<td>2.86</td>
<td>−0.58</td>
<td>0.96</td>
<td>0.96</td>
<td>0.82</td>
</tr>
<tr>
<td>Sep</td>
<td>2.94</td>
<td>−0.44</td>
<td>0.97</td>
<td>0.96</td>
<td>0.85</td>
</tr>
<tr>
<td>Oct</td>
<td>3.04</td>
<td>−0.38</td>
<td>0.97</td>
<td>0.97</td>
<td>0.85</td>
</tr>
<tr>
<td>Nov</td>
<td>3.21</td>
<td>−0.38</td>
<td>0.97</td>
<td>0.97</td>
<td>0.80</td>
</tr>
<tr>
<td>Dec</td>
<td>2.50</td>
<td>−0.35</td>
<td>0.97</td>
<td>0.94</td>
<td>0.82</td>
</tr>
</tbody>
</table>

4.2.2 Temperature

The results showed very high correlations between the estimated values and measurements both for TXM and TMIN when considering daily means, as the Pearson correlation coefficients were 0.99 and 0.93, respectively (Fig. 6). Similarly, good agreements were obtained when considering the extremes (5th percentile of TMIN and 95th percentile of TMAX) with Pearson correlations of 0.99 and 0.91, respectively.

The frequency of the observed temperatures and their estimates (Fig. 7) showed very good general agreement. There was a slight underestimation of TMAX (in the temperature intervals 0 to 5 and 10 to 15 °C) and TMIN (interval 5 to 10 °C). Overestimation was found for TMAX and TMIN for the interval −5 to 0 °C.

Most of the stations in Slovenia are located at lower altitudes, with 60 % of them between 100 and 500 m a.s.l., which leads to a slight underestimation of TMIN values at 900–1100 and 1500–2000 m a.s.l. due to having fewer data. Also, an overestimation of TMIN was observed at the 1300–1500 m a.s.l. range, but the similarity between the observed and estimated temperatures was generally higher for TMAX (Table 4).

The ratio of means (RM) showed that the differences in the means between observed and estimated temperatures were lower for TMAX, with all the values very close to 1 (Table 5). TMIN showed, overall, a higher underestimation, especially in March. Moreover, the ratio of standard deviation (RSD) was higher in TMAX, while TMIN showed values under 0.95 from April to September, indicating a bias in variance estimation in these months. ME values were close to 0 in all cases, and Pearson correlations were higher than 0.98 in TMAX in all months, ranging from 0.93 to 0.98 in TMIN.

4.3 Spatial distribution and uncertainty in daily precipitation

Mean precipitation on wet days (SDII) showed that the highest values occurred in the central to northwestern part of the country and also at some individual locations on the south and north, with amounts up to 27 mm (Fig. 8a). The lowest amounts were in the northeastern part of Slovenia (less than 10 mm). The average maximum length of dry spells (CDD) reached the highest values in the northeastern part of the country, with 70 d, while this number decreased in the southeast and northwest until the values were lower than 35 d (Fig. 8b). The uncertainty of SDII was very low (<10 %) all over the country (Fig. 9a), while the uncertainty of CDD averaged between 16 % and 20 % in most of it, with higher values (21 %–25 %) in the northeastern part (Fig. 9b).

The highest values of the mean annual absolute daily maximum temperature (TXa) were found in the southwest and east of Slovenia, with temperatures up to 36 °C (Fig. 10a), while values below 20 °C were more frequent at higher elevations. As expected, the mean annual absolute daily minimum temperature (TNa) showed a similar pattern to TX (Fig. 10b), with the lowest values in the north and northwestern part of Slovenia (−22 °C) and much higher values in the southwest, where the influence of the sea affects temperatures. However, in the eastern half of the territory there was a lower spatial variability compared to that seen for TXa. The mean annual number of summer days (SU) was calculated from the annual count of days when daily TXa was over 25 °C. Similar to the TXa, the highest values were found in the southwestern and also eastern part of the country (Fig. 10c). In the mountain areas, however, only a small number of summer days were found. The mean annual number of frost days (FD) was calculated from the annual count of days when the minimum daily temperature was below 0 °C. The number of frost days varied from 36 in the coastal area to 240 in the highest elevations, where orography had a significant influence (Fig. 10d).
Figure 6. Comparison between measurements and estimates, by stations, of the mean maximum temperatures (a) and their 95th percentiles (b), and comparison of the mean minimum temperatures (c) and their 5th percentiles (d). Dashed lines represent ±1 standard deviation of the data.

Table 4. The leave-one-out cross-validation (LOO-CV) statistics showing the goodness of fit between measurements and estimates of daily maximum and minimum temperature separated by altitudes (m a.s.l.). N: number of stations; OBSm: mean observed maximum and minimum temperature (maximum and minimum temperature in °C); PREDm: mean predicted temperature (°C); MAE: mean absolute error (°C); ME: mean error (°C); RM: ratio of means; RSD: ratio of standard deviations. The results were rounded to two decimal places.

<table>
<thead>
<tr>
<th>Altitude (m)</th>
<th>0–100</th>
<th>100–300</th>
<th>300–500</th>
<th>500–700</th>
<th>700–900</th>
<th>900–1100</th>
<th>1100–1300</th>
<th>1300–1500</th>
<th>1500–2000</th>
<th>&gt; 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>T_{MAX} Obsm</td>
<td>18.30</td>
<td>15.80</td>
<td>14.70</td>
<td>13.50</td>
<td>12.00</td>
<td>11.30</td>
<td>9.30</td>
<td>8.70</td>
<td>7.30</td>
<td>7.30</td>
</tr>
<tr>
<td>Predm</td>
<td>18.50</td>
<td>15.80</td>
<td>14.60</td>
<td>13.70</td>
<td>12.20</td>
<td>11.50</td>
<td>9.40</td>
<td>8.70</td>
<td>7.50</td>
<td>7.10</td>
</tr>
<tr>
<td>MAE</td>
<td>0.56</td>
<td>0.35</td>
<td>0.34</td>
<td>0.45</td>
<td>0.44</td>
<td>0.36</td>
<td>0.36</td>
<td>0.50</td>
<td>0.50</td>
<td>0.46</td>
</tr>
<tr>
<td>ME</td>
<td>0.50</td>
<td>−0.02</td>
<td>−0.01</td>
<td>0.14</td>
<td>0.04</td>
<td>0.06</td>
<td>0.13</td>
<td>0.03</td>
<td>−0.01</td>
<td>−0.20</td>
</tr>
<tr>
<td>RM</td>
<td>1.03</td>
<td>0.99</td>
<td>0.99</td>
<td>1.01</td>
<td>0.99</td>
<td>1.01</td>
<td>1.02</td>
<td>0.99</td>
<td>0.99</td>
<td>0.96</td>
</tr>
<tr>
<td>RSD</td>
<td>1.03</td>
<td>1.01</td>
<td>0.99</td>
<td>0.98</td>
<td>1.01</td>
<td>0.99</td>
<td>0.99</td>
<td>1.02</td>
<td>0.99</td>
<td>0.76</td>
</tr>
<tr>
<td>N</td>
<td>12</td>
<td>61</td>
<td>59</td>
<td>35</td>
<td>11</td>
<td>10</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

| T_{MIN} Obsm | 9.60 | 5.50 | 4.60 | 4.30 | 3.50 | 3.30 | 3.10 | 0.80 | 0.90 | 1.80 |
| Predm | 8.80 | 5.50 | 4.70 | 3.90 | 3.80 | 2.60 | 3.00 | 0.99 | 0.70 | 1.70 |
| MAE | 1.04 | 0.57 | 0.69 | 0.84 | 1.06 | 0.79 | 0.59 | 0.60 | 0.74 | 0.32 |
| ME | −0.18 | −0.02 | 0.10 | −0.30 | 0.25 | −0.43 | −0.13 | 0.27 | −0.63 | −0.13 |
| RM | 0.98 | 0.99 | 1.02 | 0.92 | 1.07 | 0.81 | 0.96 | 1.19 | 0.62 | 0.99 |
| RSD | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.97 | 0.99 | 1.01 | 1.01 | 1.01 |
| N | 12.00 | 61.00 | 59.00 | 35.00 | 11.00 | 10.00 | 3.00 | 4.00 | 4.00 | 2.00 |

The highest uncertainty values of TXx and SU were found in the northwestern part of the country (Fig. 11a, c), while the highest values of uncertainty of TNn were found in the southwestern part (Fig. 11b). Uncertainty was very low for the number of frost days (Fig. 11d).

The temporal evolution of the annual values of uncertainty (Fig. 12) showed a homogeneous behaviour in precipitation all over the period except from 2013, when uncertainty rose both in median and highest values. Maximum temperature showed also a relatively static behaviour, with higher values...
Table 5. The leave-one-out cross-validation (LOO-CV) statistics showing the goodness of fit between maximum and minimum temperature measurements and estimates of monthly aggregates. MAE: mean absolute error (°C); ME: mean error (°C); RM: ratio of means; RSD: ratio of standard deviations; Pearson: Pearson correlation coefficient. The results were rounded to two decimal places.

<table>
<thead>
<tr>
<th></th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{\text{MAX}}$</td>
<td>MAE</td>
<td>0.48</td>
<td>0.42</td>
<td>0.33</td>
<td>0.31</td>
<td>0.31</td>
<td>0.33</td>
<td>0.32</td>
<td>0.33</td>
<td>0.37</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>ME</td>
<td>0.05</td>
<td>0.07</td>
<td>0.01</td>
<td>0.00</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>RM</td>
<td>1.01</td>
<td>1.01</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>RSD</td>
<td>0.97</td>
<td>0.97</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
<td>0.97</td>
<td>0.98</td>
<td>0.97</td>
<td>0.99</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>Pearson</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.98</td>
<td>0.98</td>
<td>0.99</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
</tbody>
</table>

| $T_{\text{MIN}}$ | MAE  | 0.78 | 0.74 | 0.71 | 0.67 | 0.75 | 0.66 | 0.69 | 0.67 | 0.7  | 0.68 | 0.59 | 0.64 |
|       | ME   | -0.05 | -0.07 | -0.03 | -0.02 | 0.01 | -0.04 | -0.04 | -0.07 | -0.07 | -0.07 | 0.00 | 0.07 |
|       | RM   | 0.99 | 0.97 | 0.71 | 0.97 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.89 | 0.97 |
|       | RSD  | 0.99 | 0.99 | 0.96 | 0.93 | 0.93 | 0.93 | 0.94 | 0.93 | 0.94 | 0.96 | 0.95 | 0.96 |
|       | Pearson | 0.98 | 0.98 | 0.97 | 0.95 | 0.93 | 0.94 | 0.93 | 0.93 | 0.95 | 0.96 | 0.97 | 0.96 |

Figure 7. Comparison of frequencies by categories between observed (solid) and predicted (transparent) maximum (red) and minimum (blue) temperatures.

5 Discussion

SLOCLIM is the first climatic reconstruction for Slovenia at a high spatial resolution, providing daily data of $T_{\text{MAX}}$, $T_{\text{MIN}}$ and the amount of precipitation from 1950 to 2018. This new daily gridded dataset contributes significantly to the climate description of the country, and it is expected to be of high value for multiple potential applications, especially in areas where observed climate records are not available, such as agricultural land, forests and mountains.

There were no climatic reconstructions available for Slovenia until the work of Dolinar (2016), who established a grid on a 1 km scale for the monthly temperature and precipitation from 1961 onwards, using explanatory variables such as longitude, latitude, altitude and their second degree polynomial terms to improve the reconstruction. Besides this, there have been other activities aimed at the homogenization of climatic data, such as the project Climate Change and Variability in Slovenia (Vertačnik et al., 2015), which was based on homogenized monthly data with the goal of obtaining high-quality meteorological data from 1961 onwards.

In order to fill the gap, the SLOCLIM database was created with a methodology that is suitable for a country like Slovenia with large climatic and orographic differences, as it provides a reliable independent assessment of climate parameters, considering the local character of precipitation and temperature at each particular location. The period 1950 to 2018 represents the longest possible time span for data reconstruction at the given resolution, which is mainly determined by the number, location and quality of meteorological stations. The method makes use of all available information and includes a comprehensive iterative quality control process that checks the spatial and temporal consistency of the data until no suspect values are detected.

The presented results show very high correlations between the estimated values and the measurements for both temperatures and precipitation, indicating the ability to reproduce...
realistic climatic situations. The comparison by days showed higher Pearson correlations between the original precipitation dataset and the corresponding estimates than the comparison by stations. The reason for this is the complexity of the high geographical variability between stations, which could, to a certain extent, also influence the underestimation of minimum temperatures at higher altitudes. This is shown in the LOO-CV by altitudes, where the bias in the estimates at high elevations is probably due to the lower number of available observations. Grid estimates at these altitudinal ranges (especially > 2000 m a.s.l.) may reflect the lower reliability through higher values of uncertainty. The dataset used in Dolinar (2016) showed slightly higher correlation coefficients between predicted and measured values than found in our study, but it was constructed on a monthly basis using a completely different approach. Furthermore, Dolinar (2016) obtained lower correlations between predicted and measured values for precipitation than for temperature, which is consis-
Figure 10. Mean annual (1950–2018) values for (a) $TX_x$ – absolute daily maximum temperature, (b) $TN_n$ – absolute daily minimum temperature, (c) SU – number of summer days, and (d) FD – number of frost days (Esri, 2020; GISCO, 2020).
The agreement between the original dataset and the corresponding estimates in our study was high due to the good quality of the original data, mainly because ARSO regularly undertakes a validation control of the climatological data collected by meteorological stations, resulting in consolidated data (Bertalanič et al., 2006). Although the spatial coverage with meteorological stations decreases over time, it is still relatively high compared to other countries.

In Slovenia, precipitation is highest in the northwest (in the Julian Alps) and southwest in the Dinaric Mountains and lowest in the northeast (ARSO, 2016b), as confirmed also by our reconstruction. The highest uncertainties in precipitation were found in the northeast, in the area with the lowest amount and intensity of precipitation. Therefore, the uncertainty is higher because it is more difficult to make solid pre-

Figure 11. Uncertainty of (a) TX – absolute daily maximum temperature, (b) TN – absolute daily minimum temperature, (c) SU – number of summer days and (d) FD – number of frost days (Esri, 2020; GISCO, 2020).

Figure 12. Temporal evolution of annual median values (lines) in precipitation and temperature gridded datasets. Shaded areas show the range between the 5th and 95th percentile of uncertainty values.
dictions in areas where precipitation is very low and the number of consecutive dry days is higher. The highest temperatures are reached in the southwestern part of the country and the lowest in the northwest (in the Alpine region), as shown by ARSO (2016a) and confirmed by our reconstruction. The presented indices of mean annual absolute daily maximum and minimum temperatures are in agreement with the observations of ARSO (2016a). The highest uncertainty levels for the high temperature indices (TX, SU) were found for the northwest, and those for the lowest temperatures (TN, FD) were found in the southwest, possibly due to the high orographic effect and the distance to the sea. In the southwest, lower temperatures occur less frequently than in other parts of the country, which possibly explains why the uncertainty is greater. In this regard, the uncertainty of the number of summer days (SU) was particularly high in the areas at the highest altitudes. The geographical distribution of the temperature indices is consistent with the climatic regions defined by Kozjek et al. (2017).

The SLOCLIM database opens up many possibilities for applications in areas which need data with high spatial and temporal resolution. It has the potential to address a number of relevant climate-related topics, including recognition of climatic trends and extremes as well as their mitigation and planning of prevention measures. SLOCLIM is expected to support ongoing and future research activities which are focused on basic climatic questions in Slovenia and surrounding regions, such as trend analyses and projections (Dolinar et al., 2018; de Luis et al., 2014; Milošević et al., 2017), as well as reconstruction of past climate and extreme events based on tree-ring parameters (Cook et al., 2015; Čufar et al., 2008; Hafner et al., 2014). Research where climate data are needed also includes that focused on the effects of climate change on the well-being and health of the population (Ciuhu et al., 2019; Pogačar et al., 2019, 2020), as well as agroclimatic shifts and the food supply (Ceglar et al., 2012, 2019a, b). In Slovenia, a densely afforested country, climate affects the survival and productivity of forest trees due to shifts in leaf phenology (Čufar et al., 2012, 2015) and the production of crucial tree tissues, like wood and phloem (Delpierre et al., 2019; Martínez del Castillo et al., 2018; Prislan et al., 2013). While forest trees can be damaged due to extreme events like ice storms (Decuyper et al., 2020; Klopcič et al., 2020), the effects of climate differ along various gradients (Čater and Levanič, 2019; Jevšenak et al., 2020) and at the microsite level (Diaci et al., 2020; Kermavnar et al., 2019). In most cases the use of precise daily climatic parameters can help to improve recognition of climate-growth relationships (Jevšenak, 2019; Jevšenak and Levanič, 2018).

6 Data availability

The SLOCLIM dataset is freely available in the web repository Zenodo. It can be accessed through https://doi.org/10.5281/zenodo.4108543 and cited as Škrk et al. (2020). The data are arranged in six files (daily maximum and minimum temperature, daily amount of precipitation, and their uncertainties). The data are also freely available through web page http://www.sloclim.eu (last access: 10 June 2021).

7 Conclusions

The result of the presented study is a high-resolution daily gridded precipitation and temperature dataset for Slovenia with a spatial resolution of 1 × 1 km for the period 1950 to 2018 (SLOCLIM). Advantages of SLOCLIM compared to other existing datasets in Slovenia are the use of all available data and the provision of reconstructed daily data over a dense grid. It delivers the daily amount of precipitation as well as maximum and minimum temperature. The available measurement data with regard to latitude, longitude, altitude and distance from the coast were used to calculate the dataset. All the data were quality controlled by using cross-validation. Missing values were calculated using the 10 nearest observations in precipitation and 15 in temperature. All results showed high correlation between estimated and observed values, and six climatic indices were also calculated: precipitation intensity, length of dry period, mean annual absolute maximum and minimum temperatures, mean annual number of frost days, and mean annual number of summer days. The indices correlated well with altitudes and distance from the coast.

SLOCLIM represents a novel publicly available climatic database for Slovenia with multiple potential applications for recognition and mitigation of climate-related events directly affecting the environment, well-being and health of the population, agriculture and food supply, and forestry. High-resolution gridded daily data are expected to facilitate research activities in numerous scientific disciplines studying climatic trends and projections; reconstruction of past climate and prediction of future climate, environment, human and animal populations; and agricultural and forest systems. The data can help to identify extreme climatic events and can be also used for prediction, adaptation and mitigation measures.

Author contributions. RSN, MdL, KC and NS designed the research; RSN and MdL developed the methodological approach; LKB, ZC and KC organized the original data; MdL, RSN and NS made the reconstruction of the climatic data and developed the gridded dataset. RSN and NS performed the validation processes and performed the analysis and presentation of the results. NS and RSN designed and wrote the manuscript with contribution from all co-authors. MM contributed in the creation of the web page.
Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. Climatic data were provided by the Environmental Agency of the Republic of Slovenia (ARSO) within the Ministry of the Environment and Spatial Planning. We thank Gregor Senegačnik from the Slovenia Forest Service for help with ArcMap. We thank Zorko Vičar for his great help with the original climatic data. We thank Paul Steed for language editing.

Financial support. The study was supported by the Slovenian Research Agency ARRS (programs P4-0015 and P4-0085 and young researchers’ program). Roberto Serrano-Notivoli and Martín de Luis were partially supported by the Government of Aragón through the “Program of research groups” (group H09_20R, “Clima, Agua, Cambio Global y Sistemas Naturales”).

Review statement. This paper was edited by Lukas Gudmundsson and reviewed by Miquel Tomas-Burguera and Martin Hanel.

References


Ćufar, K., De Luis, M., Prislan, P., Gricar, J., Crepinšek, Z., Merela, M., and Kajfež-Bogataj, L.: Do variations in leaf phenology affect radial growth variations in Fagus sylvatica?, Int. J. Biomete-


Esri: Imagery [basemap], World Imag., available at: https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9 (last access: 10 June 2021), 2020.


Miloševec, D. D., Savić, S. M., Stankov, U., Žiberna, I., Pantelić, M. M., Dolinaj, D., and Leščen, I.: Maximum temperatures over Slovenia and their relationship with at-


