Open Access

Earth System
Science
Data

G2DC-PL+: a gridded 2 km daily climate dataset for the union of the Polish territory and the Vistula and Odra basins

Mikołaj Piniewski[1], Mateusz Szcześniak[2], Ignacy Kardel[1], Somsubhra Chattopadhyay[1], and Tomasz Berezowski[3]

[1]Department of Hydrology, Meteorology and Water Management, Warsaw University of Life Sciences, Nowoursynowska 166, 02-787 Warsaw, Poland
[2]independent researcher
[3]Faculty of Electronics, Telecommunication and Informatics, Gdańsk University of Technology, Gabriela Narutowicza 11/12, 80-233 Gdańsk, Poland

Correspondence: Mikołaj Piniewski (mikolaj_piniewski@sggw.edu.pl)

Abstract. G2DC-PL+, a gridded 2 km daily climate dataset for the union of the Polish territory and the Vistula and Odra basins, is an update and extension of the CHASE-PL Forcing Data – Gridded Daily Precipitation and Temperature Dataset – 5 km (CPLFD-GDPT5). The latter was the first publicly available, high-resolution climate forcing dataset in Poland, used for a range of purposes including hydrological modelling and bias correction of climate projections. While the spatial coverage of the new dataset remained the same, it has undergone several major changes: (1) the time coverage was increased from 1951–2013 to 1951–2019; (2) its spatial resolution increased from 5 to 2 km; (3) the number of stations used for interpolation of temperature and precipitation approximately doubled; and (4) in addition to precipitation and temperature, the dataset consists of relative humidity and wind speed data. The main purpose for developing this product was the need for long-term areal climate data for earth-system modelling, and particularly hydrological modelling. Geostatistical methods (kriging) were used for interpolation of the studied climate variables. The kriging cross-validation revealed improved performance for precipitation compared to the original dataset expressed by the median of the root mean squared errors standardized by standard deviation of observations (0.59 vs. 0.79). Kriging errors were negatively correlated with station density only for the period 1951–1970. Values of the root mean squared error normalized to the standard deviation (RMSEsd) were equal to 0.52 and 0.4 for minimum and maximum temperature, respectively, suggesting a small to moderate improvement over the original dataset. Relative humidity and wind speed exhibited lower performance, with median RMSEsd equal to 0.82 and 0.87, respectively. The dataset is openly available from the 4TU Centre for Research Data at https://doi.org/10.4121/uuid:a3bed3b8-e22a-4b68-8d75-7b87109c9feb (Piniewski et al., 2020).

1   Introduction

Recent decades have witnessed a substantial improvement in atmospheric numerical weather prediction and climate model simulations. Precipitation and air temperature data at high spatial and temporal resolution indeed serve as major input in modelling the earth and environment. One of the major applications of such data is in distributed hydrological modelling at various spatio-temporal scales (Chattopadhyay et al., 2017). Precipitation is a critical component in rainfall–runoff models such as SWAT (Arnold et al., 1998), WetSpa (Liu and De Smedt, 2004) or TOPMODEL (Beven et al., 1995). Precipitation also plays a major role in flood–drought assessment or provision of ecosystem services (Abbaspour et al., 2015). Temperature influences evaporation, transpiration and the overall water demand. It is therefore also crucial to provide high-resolution temperature data in hydrological models. Methods of estimation of actual evapotranspiration such as Hargreaves or Penman–Monteith that require temperature data are commonly used in several models (Hargreaves, 1975; Wang and Tedesco, 2007). Climate change across the landscape has significant spatio-temporal variations which are often not uniform or consistent (Hayhoe et al., 2008). Spatial heterogeneity in the distribution of earth surface features including physical variables such as land cover, soil moisture and landscape properties such as slope and elevation interact with large-scale climate, which in turn determines microscale climate (Dobrowski et al., 2009). Herein lies the significance of fine-scale gridded analysis to study local climate and variability.

Spatial resolution of global datasets including gridded precipitation and temperature (Sheffield et al., 2006; Dee et al., 2011; Schamm et al., 2014; Weedon et al., 2014) is variable, with the highest resolution of $0.25 \times 0.25°$ translating to $28\,\text{km} \times 28\,\text{km}$ at the Equator and $28\,\text{km} \times 14\,\text{km}$ at $60°$ N. Some of the recent developments in terms of finer resolution at the global scale include Terra Climate (Abatzoglou et al., 2018), providing monthly gridded data at $4\,\text{km}$ resolution. Large-scale hydrological modelling studies often employ these global datasets (Haddeland et al., 2011; Li et al., 2013; Abbaspour et al., 2015). Previous research has shown that when the study area is smaller than one grid cell of the data, such coarser resolution is not adequate. Therefore, local meteorological gridded datasets are receiving growing attention with country-wide applications recently reported for the UK (Hollis et al., 2018), Iberia (Herrera et al., 2019), Norway (Lussana et al., 2018), Spain (Serrano-Notivolli et al., 2017) and China Peng et al. (2019). These products offer a greater advantage for hydrological modelling at the local scale compared to their global counterparts, as shown for a large dataset of French catchments and the conceptual model GR4J (in French, modèle du Génie Rural à 4 paramètres Journalier) (Raimonet et al., 2017).

Gridded datasets are constructed by interpolation from national meteorological networks. It is still unclear, however, how to choose the optimal method for spatial interpolation of these meteorological variables. The geostatistical (kriging) and inverse distance weighted (IDW) methods are generally quite popular. Some of the recent studies have used advanced interpolation techniques such as regression kriging (Brinckmann et al., 2016), iterative optimal interpolation (Lussana et al., 2018) and the area-averaged three-dimensional (AA-3D) interpolation method (Herrera et al., 2019). Kriging outperformed IDW and Thiessen polygons as evaluated by Szcześniak and Piniewski (2015) in a hydrological modelling study for several medium-sized catchments in Poland. Kriging is indeed still being used as the interpolation method for precipitation and air temperature with satisfactory results quantified by correlation coefficients or root mean squared errors (Carrera-Hernández and Gaskin, 2007; Hofstra et al., 2008; Ly et al., 2011; Herrera et al., 2012; Brinckmann et al., 2016; Herrera et al., 2019). A broad range of kriging types exists, including ordinary kriging (assumes constant mean), universal kriging (removal of trend based on spatial coordinates), kriging with external drift (mean is dependent on external variable, e.g. elevation map), co-kriging (estimates a variable based on its values and values of other variables) and others. Selection of the most appropriate kriging method is variable dependent (by studying phenomena responsible for observations of the variable, e.g. relations with elevation) and case-study dependent (by investigating whether any trend is observed at a given spatial and temporal scale, e.g. seasonal and geographical relations with climate).

Berezowski et al. (2016) described the CHASE-PL Forcing Data – Gridded Daily Precipitation and Temperature Dataset – 5 km (CPLFD-GDPT5) product, which provided input data for environmental modelling in the area defined as the union of the Polish territory and the Vistula and Odra basins (hereafter denoted as "PL+"). It was the first publicly available, high-resolution climate forcing dataset in Poland. It has been widely used for a range of purposes, including (1) bias correction of EURO-CORDEX projections (Mezghani et al., 2017); (2) forcing data for hydrological models (Piniewski et al., 2017; Terskii et al., 2019), a water quality model (Marcinkowski et al., 2017) and a simple phenology model (Marcinkowski and Piniewski, 2018); (3) a streamflow trend detection study (Somorowska, 2017); (4) a hydro-ecological classification (Piniewski, 2017); and (5) ecological modelling (O'Keeffe et al., 2018). Furthermore, the same methodological workflow as in Berezowski et al. (2016) was used to prepare hydrological model forcing in Berezowski et al. (2019). Besides a natural need to carry out a periodic update of the dataset, a number of other potential product improvements have been identified over recent years, such as increasing spatial resolution, adding new variables and increasing the number of stations used for interpolation. In this paper, we describe a new, updated product called G2DC-PL+ (a gridded 2 km daily climate dataset for the PL+ area), specifically pointing to differences between itself and its predecessor, CPLFD-GDPT5.
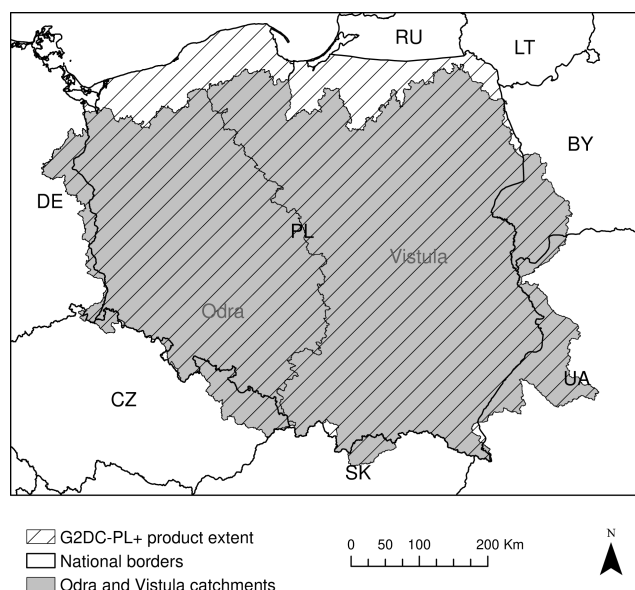
Figure 1. The spatial extent for the G2DC-PL+ dataset. Countries are labelled with black national codes. The Odra and Vistula basins are labelled in grey.
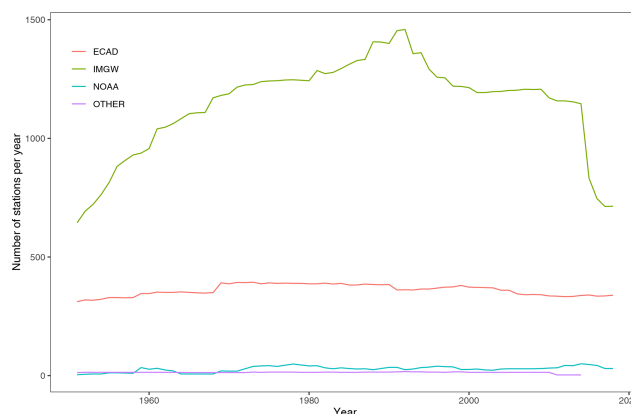

Figure 2. Number of meteorological stations for precipitation observations per year from 1951 to 2019.
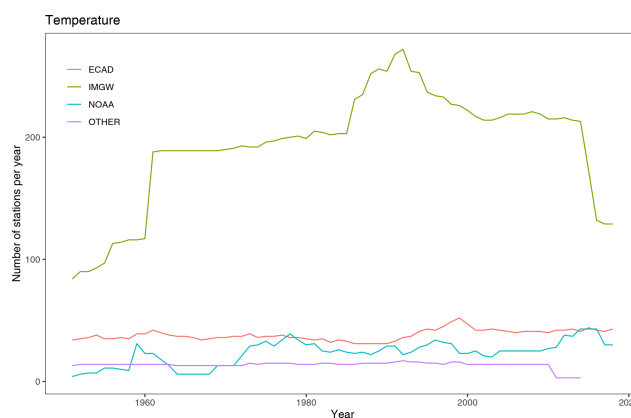
Figure 3. Number of meteorological stations for temperature observations per year from 1951 to 2019.

2 Key features of the update

While the spatial domain of the G2DC-PL+ dataset remained the same as in the original CPLFD-GDPT5 dataset, i.e. the PL+ area (Fig. 1), the dataset has undergone several major changes: (1) the temporal range has been extended from 1951–2013 to 1951–2019; (2) the spatial resolution has increased from 5 to 2 km; (3) the number of stations used for interpolation of temperature and precipitation approximately doubled; and (4) in addition to precipitation and temperature, the new dataset consists of relative humidity and wind speed.

2.1 Temporal range

The analysis covers the period 2014–2019; the 6-year extension (2014–2019) makes the dataset applicable for studying recent earth-system phenomena. Poland and the neighbouring countries in Central and Eastern Europe have encountered three major droughts in 2015, 2018 and 2019. Inclusion of these years in the dataset will be useful for in-depth drought assessments and will help to constrain hydrological models (Pfannerstill et al., 2014).

2.2 Interpolated variables

While the previous dataset included only two climate variables, namely temperature and precipitation, the updated version includes two new ones: relative humidity and wind speed (average daily values). We originally planned to include solar radiation as well, but due to a low number of stations available in Poland (below 30) and much shorter temporal availability of data, we concluded there would be little benefit in doing so. A viable alternative for solar radiation data is offered in the E-OBS gridded dataset distributed by ECA & D. One of the major benefits of using relative humidity, wind speed and solar radiation data in addition to air temperature data is that it allows for using the energy-based Penman–Monteith method for potential evapotranspiration (PET) calculation instead of a temperature-based method, such as Hargreaves. PET is a key input in hydrological models, and its role is of particular importance in the ever-growing climate change impact studies. Some authors advocate the use of a fully physically based formulation of PET rather than temperature-based PET methods. The uncertainty of using more simple methods in climate change impact studies is huge and similar in its magnitude to the uncertainty of general circulation models (Hosseinzadehtalaei et al., 2016) or emission scenarios (Williamson et al., 2016).
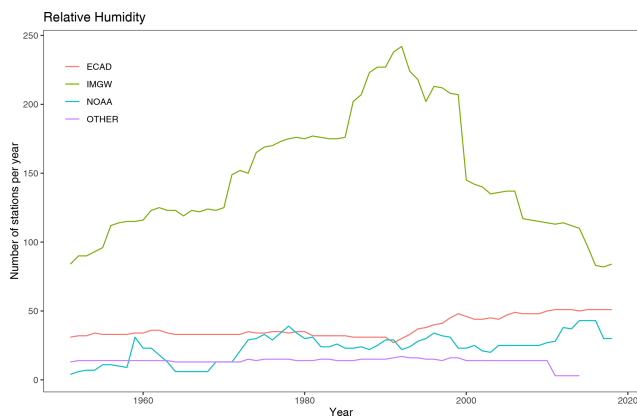
Figure 4. Number of meteorological stations for relative humidity observations per year from 1951 to 2019.
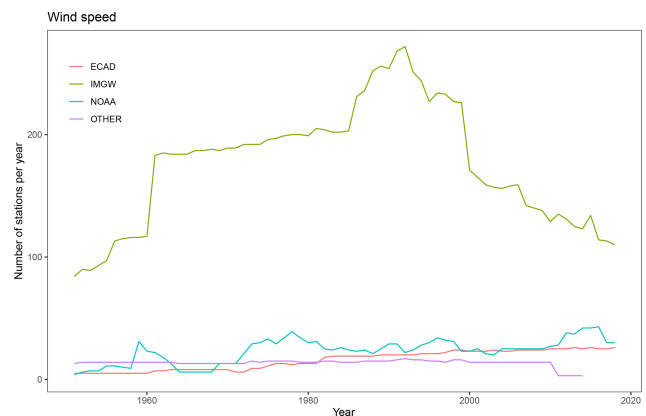
Figure 5. Number of meteorological stations for wind speed observations per year from 1951 to 2019.

2.3 Number of stations

At the time of the development of the CPLFD-GDPT5 product (2015/2016), the data sharing policy of the key institution owning climate data in Poland, the Institute of Meteorology and Water Management (IMGW-PIB), required concluding individual agreements between the data holder and the user. For this reason, only a subset of all available historical station data could be obtained at that time. Since 2018, IMGW-PIB has changed its policy and made the climate data freely available for research purposes at the https://dane.imgw.pl/ (last access: 22 March 2021) website. It has allowed us to download and use all available data for the entire period of interest, namely 1951–2019. The number of IMGW-PIB precipitation stations with available data (approaching 1500) was the highest in the early 1990s. The number is more than double the maximum number of respective stations in the former dataset. The number of IMGW-PIB stations at the beginning and end of the analysed period, however, was significantly lower (600–700) (Fig. 2). With regard to other variables, the temporal trend of data availability was the same as for precipitation, but the total numbers were significantly lower, reaching approximately 250 in early 1990s (cf. Figs. 3–5). The spatial distribution of used stations is shown in Figs. 6–9.

The major source of data from outside Poland was the European Climate Assessment and Dataset (ECA & D). Also in this case, a large, almost 10-fold, increase in data availability was observed for the area of interest. Among countries neighbouring Poland, station density was the highest in Germany. The third, least abundant data source was the National Oceanic and Atmospheric Administration National Climatic Data Center (NOAA-NCDC), but in this case the number of stations did not change much.

2.4 Spatial resolution

Considering increasing computing power and storage capacity, an increase in dataset spatial resolution is a natu-
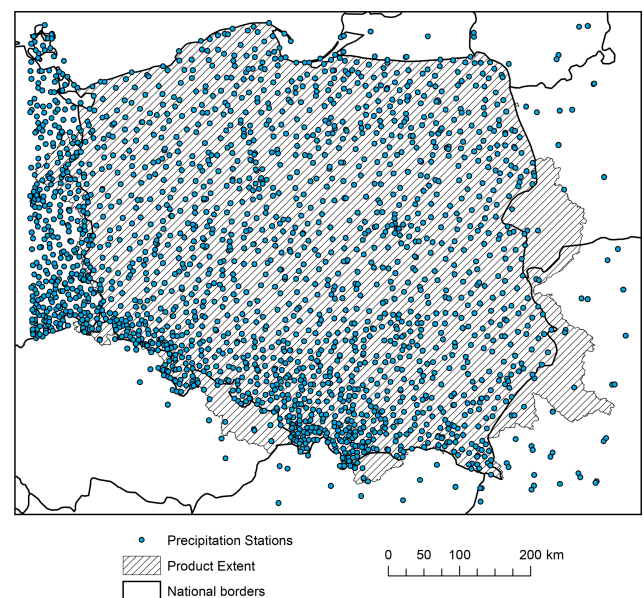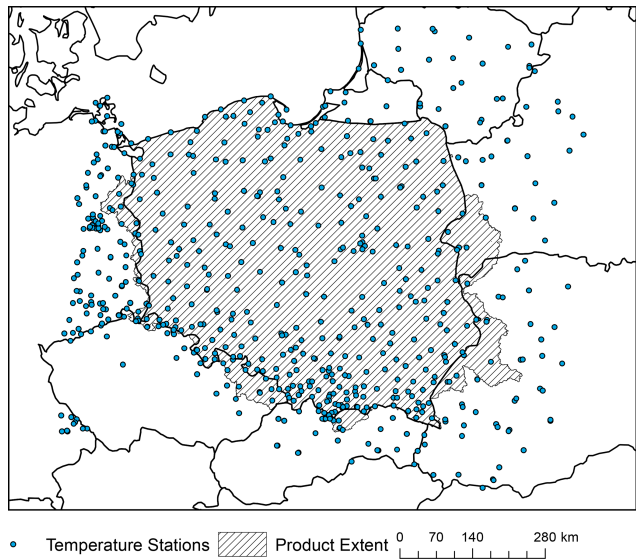

Figure 6. Spatial distribution of precipitation stations used for interpolation of the G2DC-PL+ product.

ral choice. The original 5 km resolution was not sufficiently high, in particular in mountainous areas in the south of study area. The output resolution of 2 km is of particular importance for precipitation because it is characterized by higher spatial variability. A number of gridded precipitation datasets at comparably high resolution, developed predominantly for hydrological applications, were issued recently (Duan et al., 2016; Laiti et al., 2018; Lewis et al., 2018). Another reason was that temperature, humidity and wind speed were interpolated using kriging with the external drift method, in which elevation was used as a co-variable. Elevation at 2 km is much more accurate than at 5 km resolution, especially in high-altitude areas, so this should be a clear, although indirect, benefit of this approach.

Table 1. Descriptive statistics for the kriging cross-validation results for precipitation, minimum temperature, maximum temperature, relative humidity and wind speed. "RMSEsd > 1" denotes the percentage of RMSEsd values higher than 1.

	Precipitation			Minimum temperature			Maximum temperature			Relative humidity			Wind speed		
	ρ	RMSEsd	RMSE	ρ	RMSEsd	RMSE	ρ	RMSEsd	RMSE	ρ	RMSEsd	RMSE	ρ	RMSEsd	RMSE
Minimum	0	0.11	0	0.02	0.16	0.53	0.05	0.15	0.6	0	0.24	0.02	0	0.48	0.66
First quartile	0.53	0.48	0.92	0.62	0.44	1.11	0.77	0.35	0.99	0.18	0.71	0.05	0.17	0.82	1.24
Median	0.66	0.59	1.64	0.72	0.52	1.33	0.84	0.40	1.11	0.36	0.82	0.06	0.24	0.87	1.56
Third quartile	0.77	0.69	2.48	0.8	0.62	1.6	0.88	0.47	1.25	0.52	0.92	0.07	0.33	0.91	1.98
Maximum	0.99	1.24	11.9	0.98	1	3.62	0.98	0.99	2.97	0.94	1.29	0.16	0.77	1.1	4.35
RMSEsd > 1	–	0.16 %	–	–	0 %	–	–	0 %	–	–	4.4 %	–	–	1.2	– %

Figure 7. Spatial distribution of temperature stations used for interpolation of the G2DC-PL+ product.
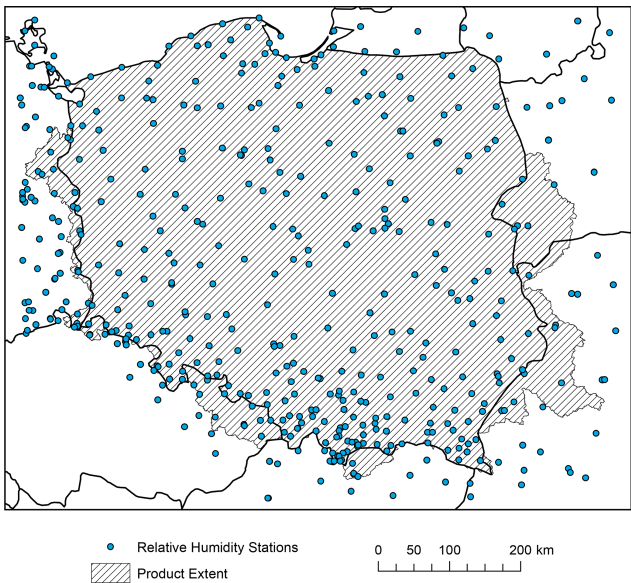

Figure 8. Spatial distribution of relative humidity stations used for interpolation of the G2DC-PL+ product.

Even though station density for variables other than precipitation was not very high (see Figs. 7–9), for practical reasons we have decided to set a uniform resolution of 2 km throughout the entire product.

2.5 Unchanged properties of the updated dataset

Other methodological features that remained unchanged between the current and previous versions are as follows.

1. The projected coordinate system for all gridded data was PUWG-92.

2. All organizations from which we have compiled the data conduct a quality control check for raw data before making them publicly available.

3. The time frequency for all variables was daily.

4. Correction for precipitation undercatch was carried out by means of the Richter method (Richter, 1995) recognized by the World Meteorological Organization. A

map showing values of coefficient b representing the effect of wind exposition of the measurement site is presented in Fig. 10. We followed the same simplified criteria for dividing stations into those with low and medium shielding as Berezowski et al. (2016). Stations located above 400 m a.s.l. and those lying within a 40 km buffer from the coast were assigned to a medium shielding category.

The values of b were set as for medium shielding for all stations apart from those in the mountains or close to the coast, where b values were set as for low shielding (Fig. 5). The rationale behind assigning different values of b for different location lies in the fact that wind speed is generally higher in mountains and at the seaside than in the lowlands.

5. The applied procedure for filling in "0" values to precipitation time series for a subset of IMGW-PIB stations was similar to that in (Berezowski et al., 2016). Due to improved metadata reporting by IMGW-PIB, however,

https://doi.org/10.5194/essd-13-1273-2021

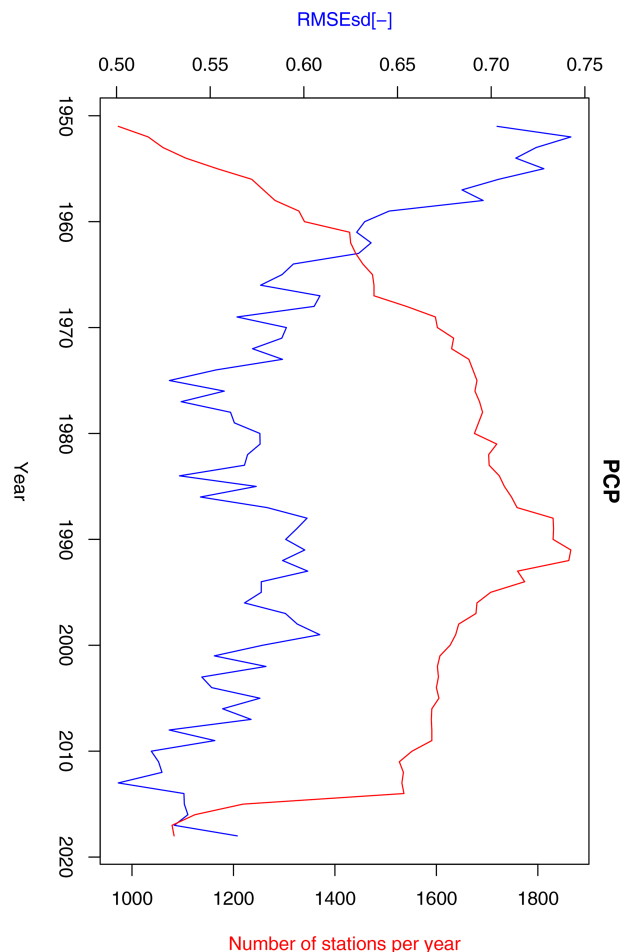Earth Syst. Sci. Data, 13, 1273–1288, 2021
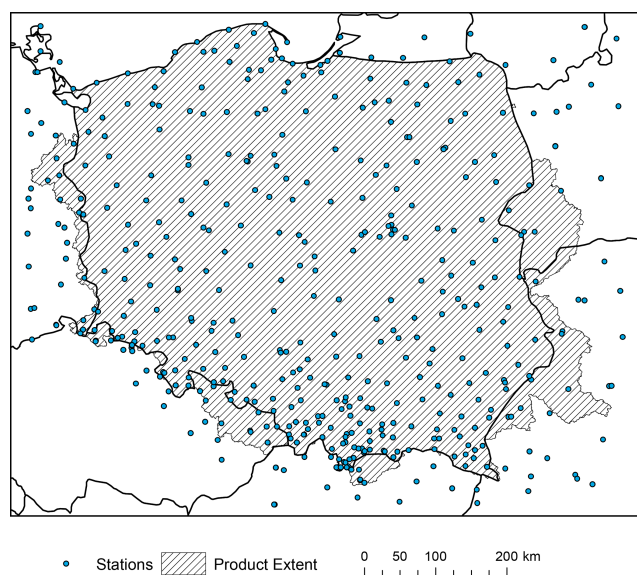
Figure 9. Spatial distribution of wind speed stations used for interpolation of the G2DC-PL+ product.
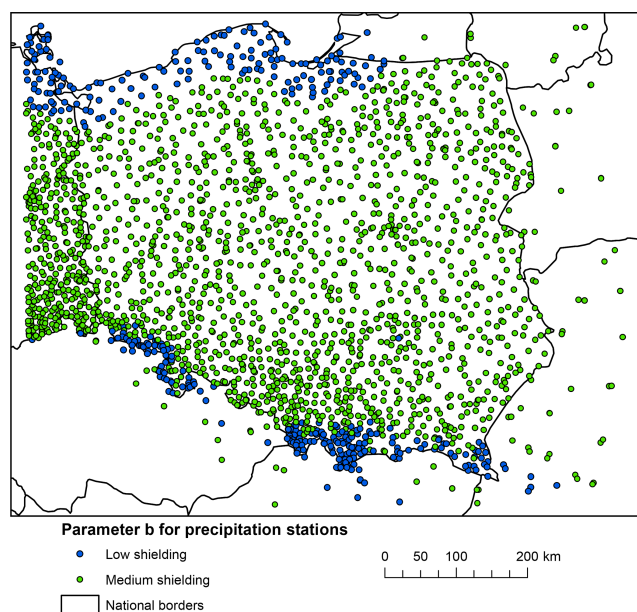
Figure 10. The Richter (1995) parameter *b* value groups for different precipitation stations.

Figure 11. Annual RMSEsd median (blue) and number of available stations per year (red) for precipitation in the period 1951–2013. The daily results, also summarized in Table 1, were used for calculating the annual medians.

it has been applied much less frequently than in the original dataset. Removal of suspicious values from the time series was also carried out in a way similar to before.

6. Minimum and maximum temperatures were interpolated with kriging with external drift and precipitation with a combination of universal and indicator kriging. The exponential variogram model was used in each case with the variogram parameters estimated automatically for each daily kriging with the weighted least squares

fit (Pebesma, 2004). The block kriging approach was applied with block size equal to the output square grid size, i.e. 2 km. The two new variables, namely relative humidity and wind speed, were interpolated using the same method as for temperature.

7. A leave-one-out cross-validation was performed daily for all stations; i.e. each station was removed from the sample one at a time, and the remaining stations were used to predict the value of the missing station. There is one small deviation from the previous version of cross-validation affecting only the precipitation variable, based on the study of Berndt and Haberlandt (2018). Because a 6.25-fold increase in spatial resolution and more than 2-fold increase in the number of stations caused a significant increase in cross-validation calculation time, we have decided to apply cross-validation only for days with precipitation above

Earth Syst. Sci. Data, 13, 1273–1288, 2021

https://doi.org/10.5194/essd-13-1273-2021
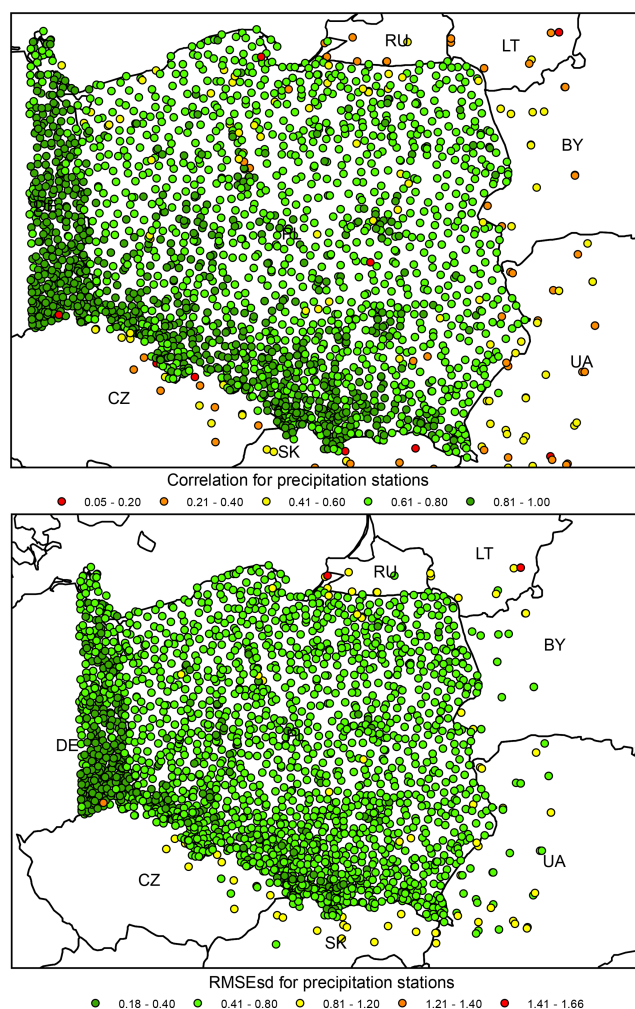
Figure 12. Precipitation ρ (top) and RMSEsd (bottom) values calculated for stations in the period 1951–2019. National borders (black lines) are labelled with country codes.
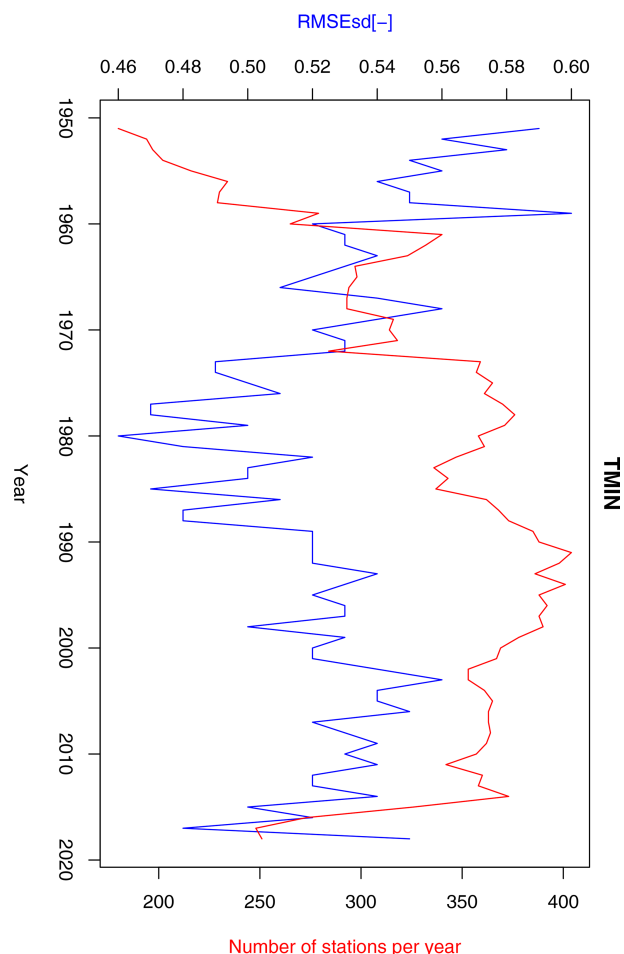

Figure 13. Annual RMSEsd median (blue) and number of available stations per year (red) for minimum temperature in the period 1951–2019. The daily results, also summarized in Table 1, were used for calculating the annual medians.

a 1 mm threshold, which allowed us to speed up the calculation process in a satisfactory way.

8. The interpolation errors were quantified using (1) Pearson's correlation coefficient (ρ) and (2) root mean squared error normalized to the standard deviation of the observed data:

$$\text{RMSEsd} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2}}{\sigma_Y},$$

where Y and $\hat{Y}$ are respectively the observed and interpolated values of a given variable, N is the number of observations (number of stations in the spatial approach or number of days in the temporal approach) and σ_Y is the standard deviation of observations.

The cross-validation was conducted on both temporal and spatial scales. On the temporal scale the errors were calcu-

lated for each day from all stations having data on this day. For all variables the standard deviation used in calculation of RMSEsd performance metrics was calculated for each Julian day separately. We will hereafter refer to the temporal scale indices as ρ_t and RMSEsd$_t$. On the spatial scale the errors were calculated for each station from all of a station's available daily values. We will hereafter refer to the spatial scale indices as ρ_s and RMSEsd$_s$.

The reader is referred to the study of Berezowski et al. (2016) for additional information related to the above-mentioned aspects.
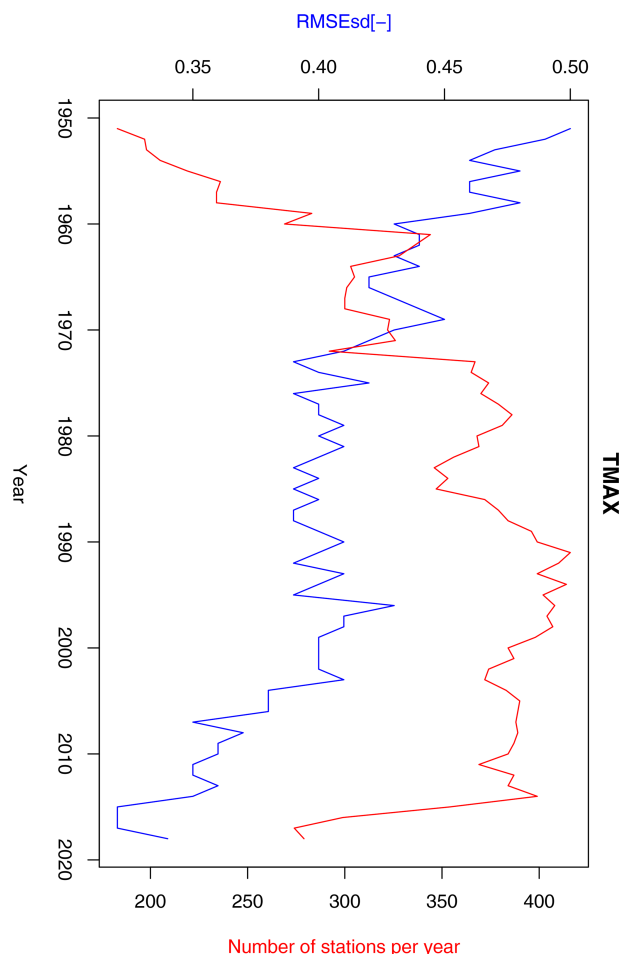
Figure 14. Annual RMSEsd median (blue) and number of available stations per year (red) for maximum temperature in the period 1951–2019. The daily results, also summarized in Table 1, were used for calculating the annual medians.
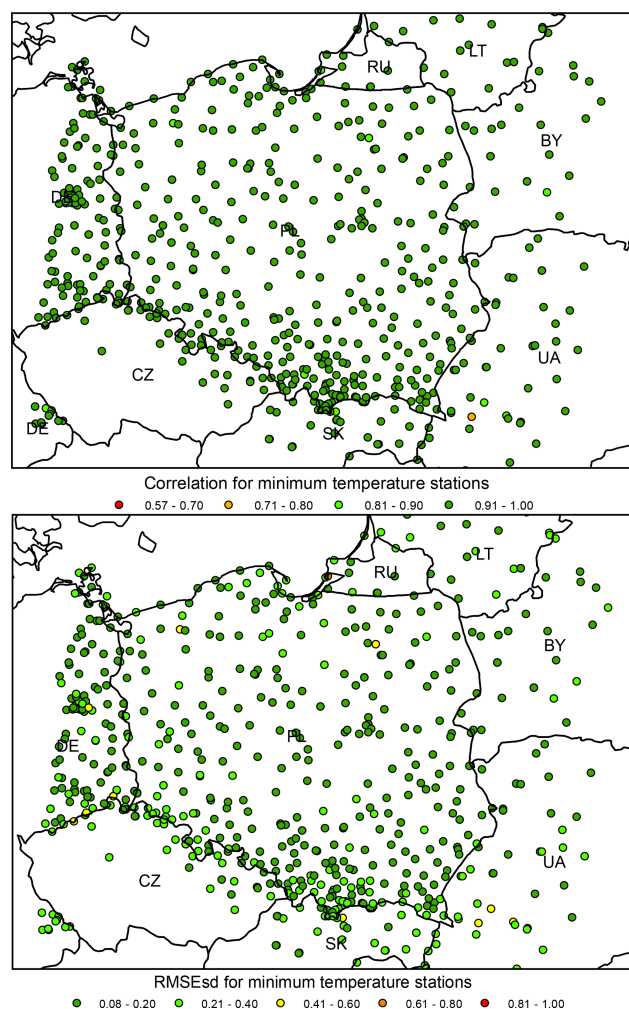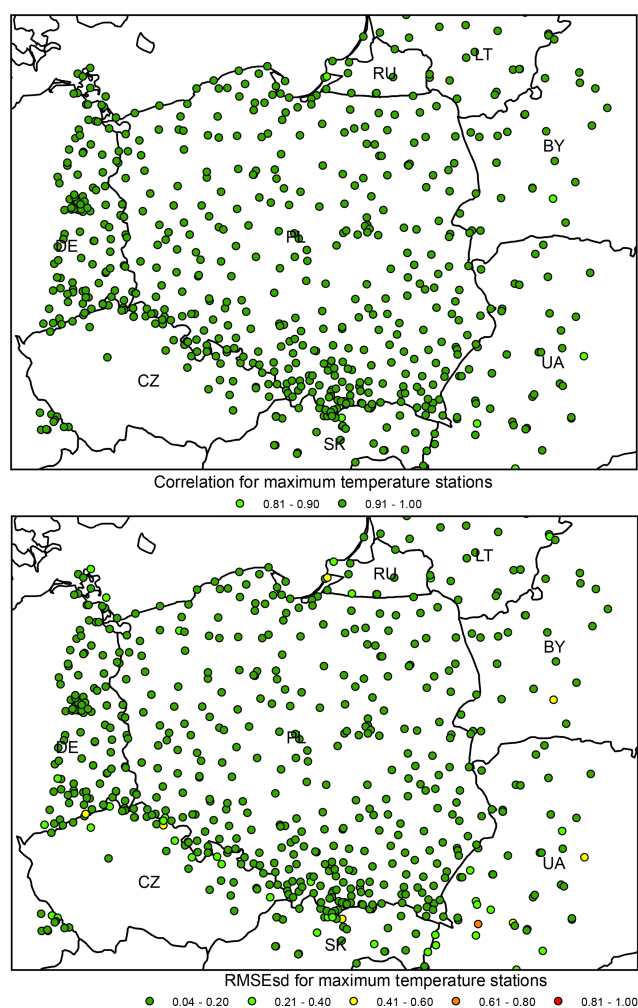

Figure 15. Minimum temperature ρ (top) and RMSEsd (bottom) values calculated for stations in the period 1951–2019. National borders (black lines) are labelled with country codes.

3   Cross-validation

3.1   Precipitation

According to daily ρ_t statistics for precipitation show that 75 % of ρ_t values are higher than 0.53 (0.47[1]) and RMSEsd$_t$ values are lower than 0.69 (0.93), respectively (Table 1). Median ρ_t is 0.66 (0.65) and median RMSEsd$_t$ is 0.59 (0.79). The fraction of RMSEsd values larger than 1 dropped from 14.2 % to 0.16 %.

We also quantified the effect of modifying the method of calculation of RMSEsd$_t$ (see Sect. 2.5). For a subset of approximately 10 % of years RMSEsd$_t$ values were calculated using both the new, faster approach, involving removal of dates with low precipitation, and the original approach. We concluded that removal of low-precipitation data led to

[1]Respective values referring to the cross-validation errors reported in Berezowski et al. (2016) will be herein shown in parentheses.

a slight improvement (average difference between RMSEsd$_t$ values equal to 0.04). The scale of this improvement, however, is much lower than the overall improvement discussed in the previous paragraph.

A strong negative correlation exists between the median of daily RMSEsd$_t$ values and the number or available precipitation stations for the first approximately 20 years of the dataset (Fig. 11). While the number of stations steadily increased over the period 1971–1990, the cross-validation error in the same period oscillated without any clear trend. Interestingly, the decline in the number of stations after 1991 was associated with the decrease in RMSEsd$_t$ value. The median of daily RMSEsd$_t$ values aggregated in years is negatively correlated (-0.72) with the number of available stations (Fig. 11), sharply decreasing as the number of station increases and reaching an equilibrium in 1980s. This suggests that the kriging errors for precipitation are dependent on the density of the observation network. We also found
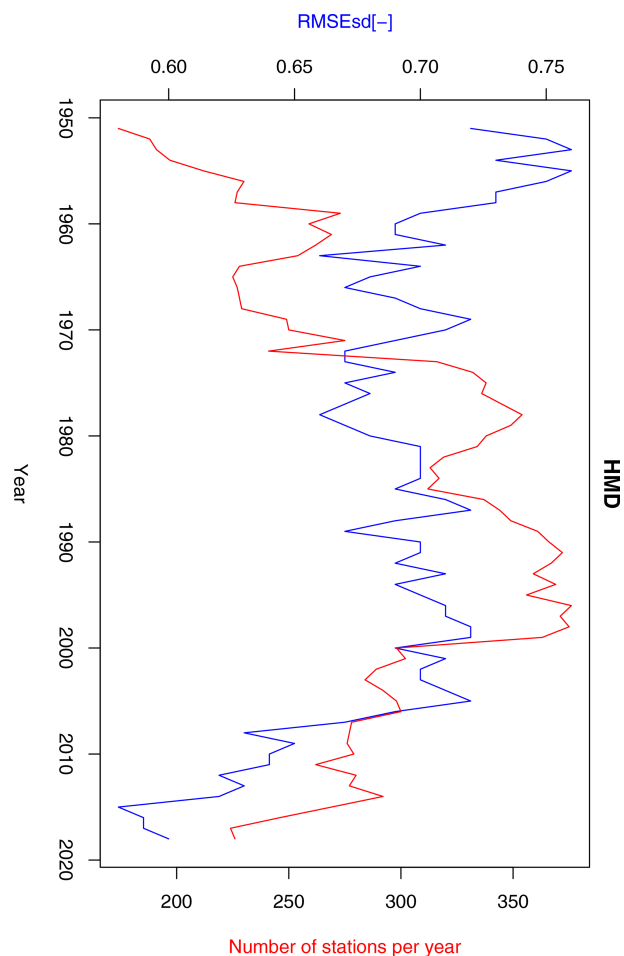
Figure 16. Maximum temperature ρ (top) and RMSEsd (bottom) values calculated for stations in the period 1951–2019. National borders (black lines) are labelled with country codes.


Figure 17. Annual RMSEsd median (blue) and number of available stations per year (red) for relative humidity in the period 1951–2019. The daily results, also summarized in Table 1, were used for calculating the annual medians.

that the interpolation results before 1960 are associated with higher uncertainty (Fig. 11). We do not have any reliable data on how measurement quality changed over time and thus are unable to confirm a potentially relevant hypothesis that it was an important factor affecting the evolution of the interpolation error.

Like in the original dataset, when considering the RMSEsd$_s$ (calculated spatially) for all stations, the results show a clear pattern of higher errors at the edge of the interpolation area, particularly in all neighbouring countries except for Germany (Fig. 12). Note that Germany features much higher density of stations than other countries neighbouring with Poland, which explains the spatial pattern.

The interpolation error for precipitation expressed in absolute, not-standardized values, i.e. RMSE, equals 1.6 mm (see Table 1 for other statistics).

3.2 Temperature

The statistics of daily RMSEsd$_t$ show the median equal to 0.52 (0.54) for minimum temperature and 0.4 (0.47) for maximum temperature (Table 1). None of the RMSEsd$_t$ values in any of the cases exceeds the value of 1 (meaning that the root mean squared errors are always below the standard deviation of observations), and all the ρ_t values are positive. The median of ρ_t is 0.72 (0.84) for minimum temperature and 0.84 (0.88) for maximum temperature. This suggests that the value of ρ_t measuring the collinearity between simulations and observations used to be higher in the former version of the dataset, particularly for minimum temperature.

Inter-annual variability of RMSEsd$_t$ exhibits quite different behaviour for minimum (Fig. 13) and maximum (Fig. 14) temperature. In the former case, a sharp decreasing trend, possibly connected to the rising number of stations, can be observed until 1980, followed by an increase until the early 2000s and another decrease afterwards. In the latter case
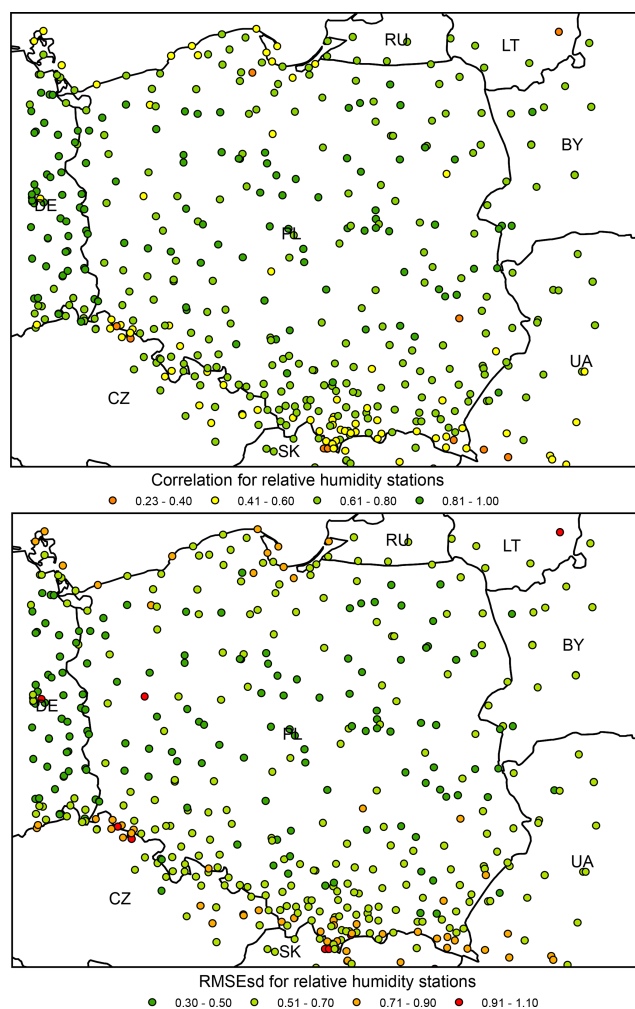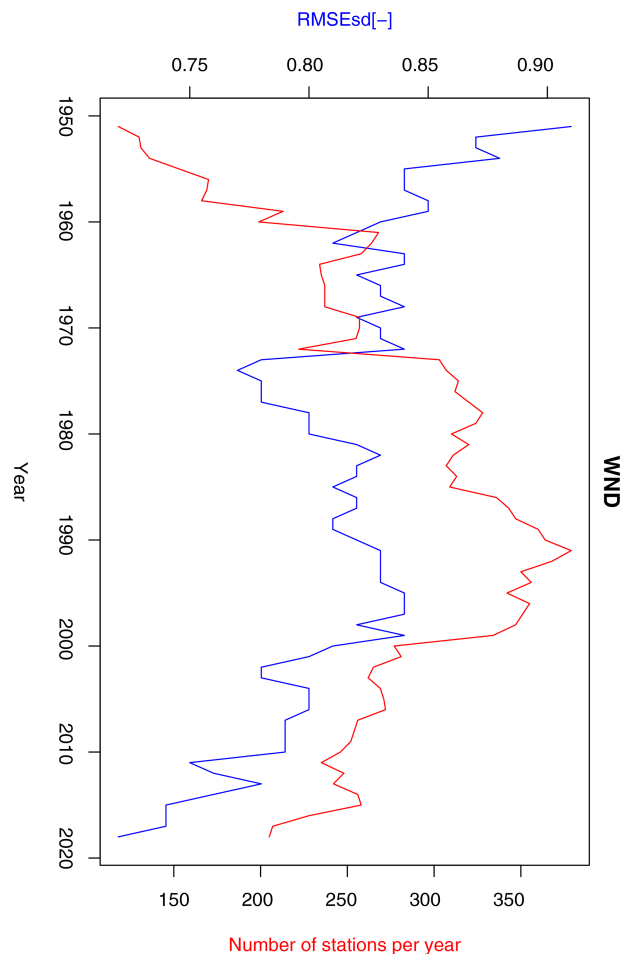
Figure 18. Relative humidity ρ (top) and RMSEsd (bottom) values calculated for stations in the period 1951–2019. National borders (black lines) are labelled with country codes.


Figure 19. Annual RMSEsd median (blue) and number of available stations per year (red) for wind speed in the period 1951–2019. The daily results, also summarized in Table 1, were used for calculating the annual medians.

there is a clear negative trend for the entire period, although in its middle (1970–2000) the values were fluctuating around the mean without any trend. This behaviour is not fully consistent with the temporal evolution of RMSEsd$_t$ for these two variables in Berezowski et al. (2016), but the previously stated conclusion that kriging errors for temperature are not dependent on the density of the observation network seems to hold true.

In the analysis of ρ_s and RMSEsd$_s$ for all stations both the minimum and maximum temperatures show rather uniformly distributed values with very few outliers, usually located in the proximity of the edge of the interpolation area (Figs. 15–16). It is noteworthy that the source of data (IMGW-PIB or international databases) does not influence the errors for temperature as much as for precipitation.

The interpolation error expressed in absolute, not-standardized values, i.e. RMSE, equals 1.33 and 1.11 °C for

minimum and maximum temperature, respectively (see Table 1 for other statistics).

3.3 Relative humidity

The median values of ρ_t and RMSEsd$_t$ for relative humidity were equal to 0.36 and 0.82, respectively (Table 1). Because relative humidity was not included in the first version of the dataset, no prior statistics are available for comparison, as occurred for precipitation and temperature. ρ_t values for relative humidity are generally considerably lower than for temperature and precipitation, pointing to low collinearity between simulations and observations. The same holds true for RMSEsd$_t$, although the fraction of data with root mean squared errors exceeding 1 standard deviation is relatively low, reaching 4.4 %.

RMSEsd$_t$ values fluctuate in a range of 0.75–0.9 for most of the analysed years. Three sub-periods can be distinguished: 1951–1990 with a low, decreasing trend; 1991–
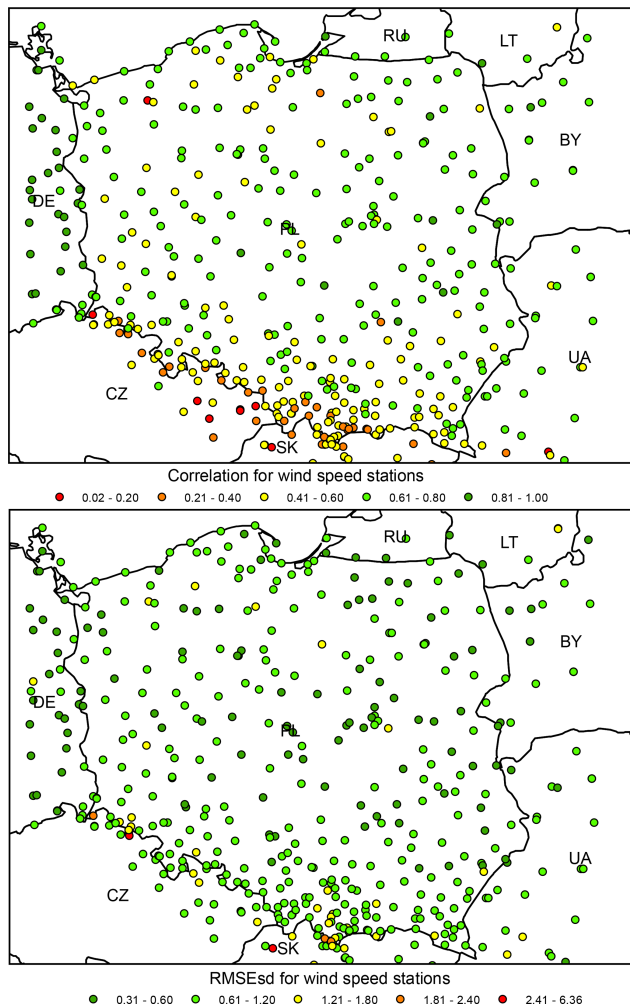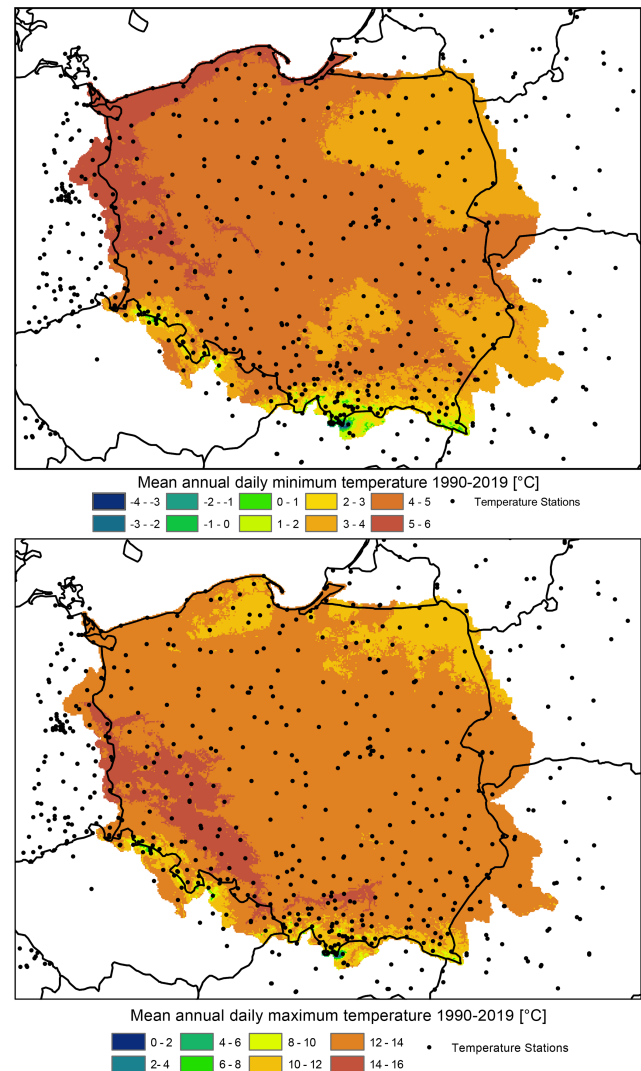
Figure 20. Wind speed ρ (top) and RMSEsd (bottom) values calculated for stations in the period 1951–2019. National borders (black lines) are labelled with country codes.

Figure 21. Mean annual daily minimum and maximum temperature in the time period 1990–2019: output from the G2DC-PL+ dataset.

2005 with an increasing trend; and 2005–2019 with a decreasing trend (Fig. 17). Trends in the first two sub-periods can be related to changes in relative humidity station density (increasing until 1990 and then decreasing).

Spatial variability in ρ_s and RMSEsd$_s$ is much higher for relative humidity than for temperature or precipitation (Fig. 18). The dataset uncertainty evidently increases with elevation, with stations located in the mountainous southern belt of the study domain showing the lowest ρ_s and the highest RMSEsd$_s$ values. The proximity of the coast is another possible cause of higher errors. For the great majority of the Polish Plain, which covers the interior part of the study domain, Pearson's correlation exceeds 0.6 and RMSEsd$_s$ is below 0.7.

The interpolation error for relative humidity expressed in absolute, not-standardized values, i.e. RMSE, equals 0.06 (see Table 1 for other statistics).

3.4 Wind speed

According to Table 1 median values of ρ_t and RMSEsd$_t$ for wind speed were found to be 0.24 and 0.87. As was the case with relative humidity, wind speed was not included in the first version of the dataset; hence, the comparison which was possible for precipitation and temperature was not feasible in this case. ρ_t values for wind speed were found to be substantially lower than for other studied variables. A similar pattern was noticed for RMSEsd$_t$, although the fraction of data with root mean squared errors exceeding 1 standard deviation is very low, reaching 1.2 %.

Wind speed exhibits lower inter-annual variability of cross-validation errors than other variables (Fig. 19). The range of variability during the period 1951–2019 is 0.84–0.92. Furthermore, no trend in the data and no correlation of RMSEsd$_t$ with station density exists.
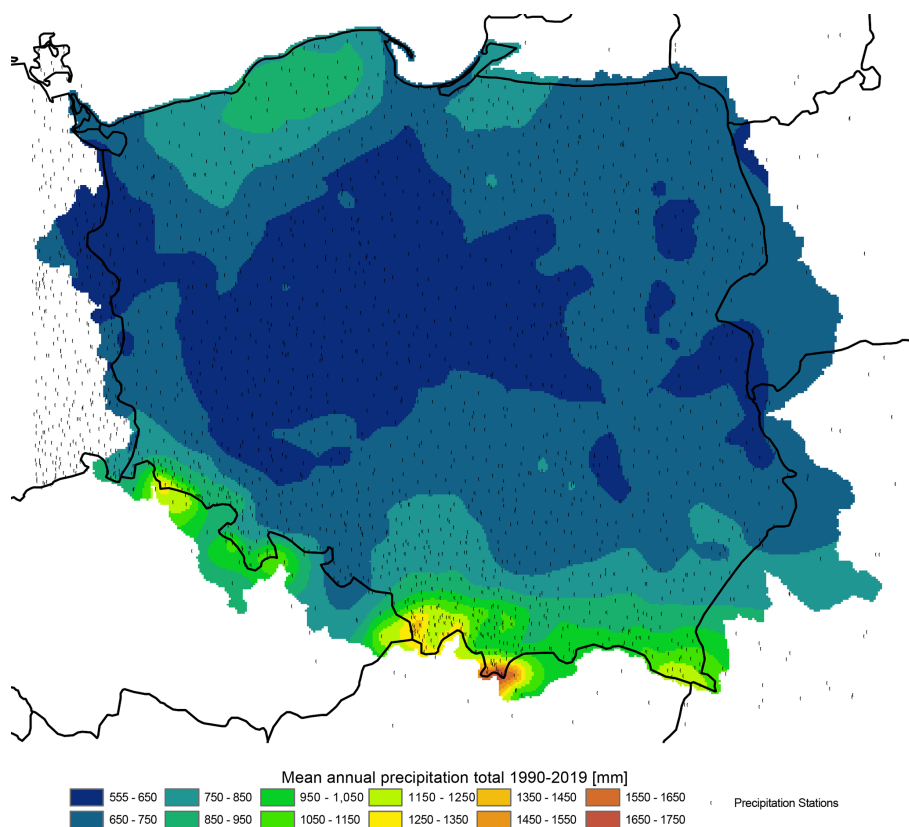
Figure 22. Mean annual precipitation in the time period 1990–2019: output from the G2DC-PL+ dataset.

Like in the case of relative humidity, the spatial variability of ρ_s and $\mathrm{RMSEsd_s}$ was much higher than for temperature or precipitation (Fig. 20). The majority of stations with low correlation (below 0.2) are located in Czechia, and the great majority of stations with correlation below 0.4 are located at the southern edge of the interpolation domain. ρ values for German stations are higher than for Polish ones. In the case of $\mathrm{RMSEsd_s}$, spatial variability is much lower, but several outliers are also located in the south.

The interpolation error for wind speed expressed in absolute, not-standardized values, i.e. RMSE, equals $1.56\,\mathrm{m\,s^{-1}}$ (see Table 1 for other statistics).

4   Consistency with climatic data

Berezowski et al. (2016) compared the consistency of the CPLFD-GDPT5 dataset with maps of climatic statistics for the period 1971–2000 provided by IMGW-PIB. The comparison revealed a high level of consistency as well as certain differences that could be attributed to different data processing methods. Here, we have updated the previous set of these comparison maps and found only a minor difference in climatic means. Therefore, the previous conclusions remain valid. One difference is that due to a larger number of stations in the northern part of Poland close to the Baltic Sea coast,

the updated dataset predicts higher long-term mean precipitation for this area than the previous one. Figures 21–23 demonstrate the spatial pattern of temperature, precipitation, relative humidity and wind speed respectively during 1990–2019 from the G2DC-PL+ dataset.

In addition, we present a supplementary analysis focused on the comparison of maps of climatic statistics for selected years. One warm and dry year (2015) and one cool and wet year (2017) were selected for comparison. The maps can be found in the Supplement (Figs. S1–S6). Although the G2DC-PL+ dataset precipitation is higher than the IMGW-PIB precipitation due to the applied correction for precipitation undercatch, spatial patterns in both dry and wet years remain very similar between both data sources (Figs. S1 and S4). The spatial agreement was also very high for minimum temperature (Figs. S2 and S5) as well as maximum temperature (Figs. S3 and S6).

5   Data availability

The G2DC-PL+ product is available in NetCDF and Geo-TIFF formats. The gridded structure of the data and the NetCDF and GeoTIFF data formats ensure easy processing in GIS and data analysis software (e.g. R for both NetCDF and GeoTIFF; list of NetCDF manipulation soft-
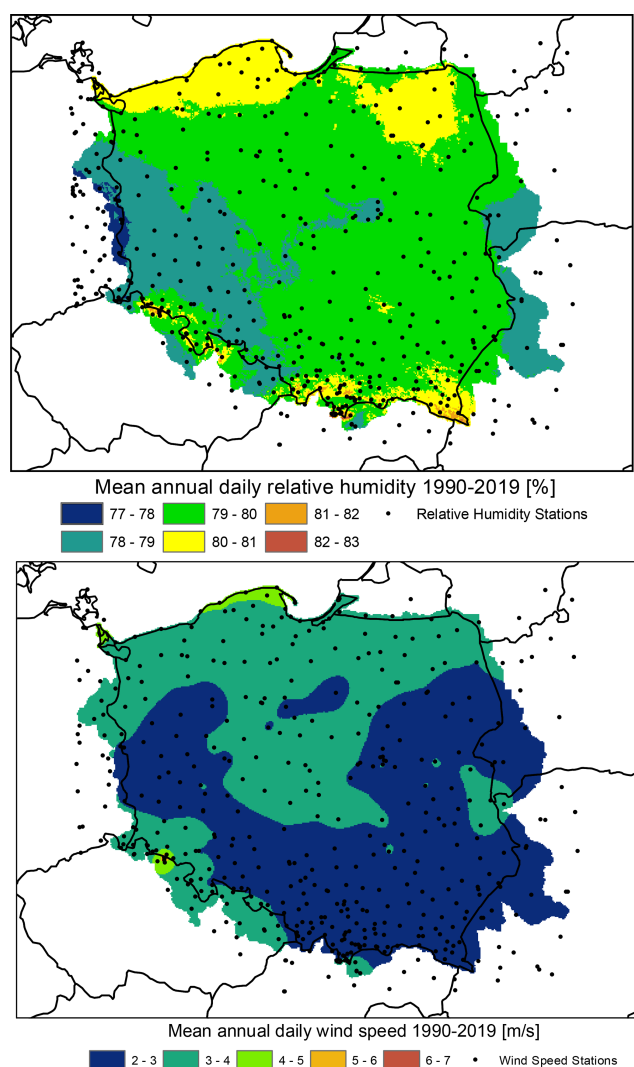
Figure 23. Mean annual relative humidity and wind speed in the time period 1990–2019: output from the G2DC-PL+ dataset.

ware: http://www.unidata.ucar.edu/software/netcdf/software. html, last access: 22 March 2021). Example R scripts allowing the reading of the data and conducting basic processing can be found in Berezowski et al. (2016).

The data are publicly available in the 4TU Centre for Research Data repository at https://doi.org/10.4121/uuid:a3bed3b8-e22a-4b68-8d75-7b87109c9feb (Piniewski et al., 2020). Files with daily data are organized in decades, e.g. 1951–1960, whereas monthly and annual data are stored in single files. The NetCDF file naming convention is *variable_time step_from year_to year*.nc (e.g. pre_d_1951_1960.nc). *Time step* can be "d" for day, "m" for month, "a" for annual aggregation period. Every NetCDF file follows the CF-1.0 convention. *Variable* can be $T_{\min}/T_{\max}$ for minimum/maximum air temperature [°C], "Pre" for precipitation [$\mathrm{kg\,m^{-2}}$], "Hmd" for relative humidity [%] or "Wnd" for wind speed [$\mathrm{m\,s^{-1}}$].

Each daily grid for each variable is also stored as a separate GeoTIFF file. The naming convention for zipped collections of GeoTIFF files is *variable_time step_from year_to year_tif*.zip, with "pre" for precipitation, "tmin" for minimum temperature, "tmax" for maximum temperature, "hmd" for relative humidity and "wnd" for wind speed, whereas time is coded as YYYYMMDD, YYYYMM or YYYY for daily, monthly and annual time step, respectively.

6 Conclusions

In the conclusions of the paper by Berezowski et al. (2016) we stated that the dataset update was planned on a 3-year basis. It took slightly more time, but the dataset has been updated with very recent data reaching 2019. The product's spatial resolution has also been increased from 5 to 2 km, which may be crucial for some high-resolution applications. Inclusion of two additional variables, namely relative humidity and wind speed, although associated with higher interpolation errors as shown in this study, is an important step towards using the energy-based Penman–Monteith method for PET estimation in hydrological or agricultural modelling. Taking advantage of a more open data sharing policy of the key climate data provider, IMGW-PIB, we have also substantially increased the number of stations available for interpolation of temperature and precipitation, which in the latter case has led to a noticeable reduction of interpolation error. This holds promise that future applications of this dataset for hydrological modelling will benefit from better input data and will therefore deliver more reliable predictions.

Note on former version

A former version of this article was published on 21 March 2016 and is available at https://doi.org/10.5194/essd-8-127-2016.

References

Abatzoglou, J., Dobrowski, S., Parks, S., and Hegewisch, K.: TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015, Sci. Data, 5, 477–507, https://doi.org/10.1038/sdata.2017.191, 2018.

Abbaspour, K., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H., and Klove, B.: A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model, J. Hydrol., 524, 733–752, 2015.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R.: Large Area Hydrologic Modeling And Assessment Part I: Model Development, J. Am. Water Resour. As., 34, 73–89, 1998.

Berezowski, T., Szcześniak, M., Kardel, I., Michałowski, R., Okruszko, T., Mezghani, A., and Piniewski, M.: CPLFD-GDPT5: High-resolution gridded daily precipitation and temperature data set for two largest Polish river basins, Earth Syst. Sci. Data, 8, 127–139, https://doi.org/10.5194/essd-8-127-2016, 2016.

Berezowski, T., Partington, D., Chormanski, J., and Batelaan, O.: Spatiotemporal Dynamics of the Active Perirheic Zone in a Natural Wetland Floodplain, Water Resour. Res., 55, 9544–9562, https://doi.org/10.1029/2019WR024777, 2019.

Berndt, C. and Haberlandt, U.: Spatial interpolation of climate variables in Northern Germany – Influence of temporal resolution and network density, Journal of Hydrology: Regional Studies, 15, 184–202, https://doi.org/10.1016/j.ejrh.2018.02.002, 2018.

Beven, K., Lamb, R., Quinn, P., Romanowicz, R., Freer, J., and Singh, V.: Topmodel, in: Computer models of watershed hydrology, Water Resources Publications, 627–668, 1995.

Brinckmann, S., Krähenmann, S., and Bissolli, P.: High-resolution daily gridded data sets of air temperature and wind speed for Europe, Earth Syst. Sci. Data, 8, 491–516, https://doi.org/10.5194/essd-8-491-2016, 2016.

Carrera-Hernández, J. and Gaskin, S.: Spatio temporal analysis of daily precipitation and temperature in the Basin of Mexico, J. Hydrol., 336, 231–249, 2007.

Chattopadhyay, S., Edwards, D., Yao, Y., and Hamidesepehr, A.: An Assessment of Climate Change Impacts on Future Water Availability and Droughts in the Kentucky River Basin, Environ. Process., 4, 477–507, https://doi.org/10.1007/s40710-017-0259-2, 2017.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553–597, 2011.

Dobrowski, S. Z., Abatzoglou, J., Greenberg, J., and Schladow, S.: How much influence does landscape-scale physiography have on air temperature in a mountain environment?, Agr. Forest Meteorol., 149, 1751–1758, https://doi.org/10.1016/j.agrformet.2009.06.006, 2009.

Duan, Z., Liu, J., Tuo, Y., Chiogna, G., and Disse, M.: Evaluation of eight high spatial resolution gridded precipitation products in Adige Basin (Italy) at multiple temporal and spatial scales, Sci. Total Environ., 573, 1536–1553, https://doi.org/10.1016/j.scitotenv.2016.08.213, 2016.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Vos, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P.: Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results, J. Hydrometeorol., 12, 869–884, 2011.

Hargreaves, G.: Moisture availability and crop prodution, T. ASAE, 18, 980–984, 1975.

Hayhoe, K., Wake, C., Anderson, B., Zhong-Liang, X., Maurer, E., Zhu, J., Bradbury, J., DeGaetano, A., Stoner, A., and Wuebbles, D.: Regional climate change projections for the Northeast USA, Mitig. Adapt. Strat. Gl., 13, 425–436, https://doi.org/10.1007/s11027-007-9133-2, 2008.

Herrera, S., Gutiérrez, J. M., Ancell, R., Pons, M. R., Frias, M. D., and Fernandez, J.: Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over Spain (Spain02), Int. J. Climatol., 32, 74–85, 2012.

Herrera, S., Cardoso, R. M., Soares, P. M., Espírito-Santo, F., Viterbo, P., and Gutiérrez, J. M.: Iberia01: a new gridded dataset of daily precipitation and temperatures over Iberia, Earth Syst. Sci. Data, 11, 1947–1956, https://doi.org/10.5194/essd-11-1947-2019, 2019.

Hofstra, N., Haylock, M., New, M., Jones, P., and Frei, C.: Comparison of six methods for the interpolation of daily, European climate data, J. Geophys. Res.-Atmos., 113, D21110, https://doi.org/10.1029/2008JD010100, 2008.

Hollis, D., McCarthy, M., Kendon, M., Legg, T., and Simpson, I.: HadUK-Grid – A new UK dataset of gridded climate observations, Geosci. Data J., 6, 151–159, https://doi.org/10.1002/gdj3.78, 2018.

Hosseinzadehtalaei, P., Tabari, H., and Willems, P.: Quantification of uncertainty in reference evapotranspiration climate change signals in Belgium, Hydrol. Res., 48, 1391–1401, https://doi.org/10.2166/nh.2016.243, 2016.

Laiti, L., Mallucci, S., Piccolroaz, S., Bellin, A., Zardi, D., Fiori, A., Nikulin, G., and Majone, B.: Testing the Hydrological Coherence of High-Resolution Gridded Precipitation and Temperature Data Sets, Water Resour. Res., 54, 1999–2016, https://doi.org/10.1002/2017WR021633, 2018.

Lewis, E., Quinn, N., Blenkinsop, S., Fowler, H. J., Freer, J., Tanguy, M., Hitt, O., Coxon, G., Bates, P., and Woods, R.: A rule based quality control method for hourly rainfall data and a 1 km resolution gridded hourly rainfall dataset for Great Britain: CEH-GEAR1hr, J. Hydrol., 564, 930–943, https://doi.org/10.1016/j.jhydrol.2018.07.034, 2018.

Li, L., Ngongondo, C. S., Xu, C.-Y., and Gong, L.: Comparison of the global TRMM and WFD precipitation datasets in driving a large-scale hydrological model in Southern Africa, Hydrol. Res., 10, 770–788, https://doi.org/10.2166/nh.2012.175, 2013.

Liu, Y. B. and De Smedt, F.: WetSpa Extension, A GIS-based Hydrologic Model for Flood Prediction and Watershed Management, Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, 2004.

Lussana, C., Saloranta, T., Skaugen, T., Magnusson, J., Tveito, O. E., and Andersen, J.: seNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day, Earth Syst. Sci. Data, 10, 235–249, https://doi.org/10.5194/essd-10-235-2018, 2018.

Ly, S., Charles, C., and Degré, A.: Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium, Hydrol. Earth Syst. Sci., 15, 2259–2274, https://doi.org/10.5194/hess-15-2259-2011, 2011.

Marcinkowski, P. and Piniewski, M.: Effect of climate change on sowing and harvest dates of spring barley and maize in Poland, Int. Agrophys., 32, 265–271, https://doi.org/10.1515/intag-2017-0015, 2018.

Marcinkowski, P., Piniewski, M., Kardel, I., Szczesniak, M., Benestad, R., Srinivasan, R., Ignar, S., and Okruszko, T.: Effect of Climate Change on Hydrology, Sediment and Nutrient Losses in Two Lowland Catchments in Poland, Water, 9, 156, https://doi.org/10.3390/w9030156, 2017.

Mezghani, A., Dobler, A., Haugen, J. E., Benestad, R. E., Parding, K. M., Piniewski, M., Kardel, I., and Kundzewicz, Z. W.: CHASE-PL Climate Projection dataset over Poland – bias adjustment of EURO-CORDEX simulations, Earth Syst. Sci. Data, 9, 905–925, https://doi.org/10.5194/essd-9-905-2017, 2017.

O'Keeffe, J., Piniewski, M., Szczesniak, M., Oglecki, P., Parasiewicz, P., and Okruszko, T.: Index-based analysis of climate change impact on streamflow conditions important for Northern Pike, Chub and Atlantic salmon, Fisheries Manag. Ecol., 26, 474–485, https://doi.org/10.1111/fme.12316, 2018.

Pebesma, E. J.: Multivariable geostatistics in S: the gstat package, Comput. Geosci., 30, 683–691, 2004.

Peng, S., Ding, Y., Liu, W., and Li, Z.: 1 km monthly temperature and precipitation dataset for China from 1901 to 2017, Earth Syst. Sci. Data, 11, 1931–1946, https://doi.org/10.5194/essd-11-1931-2019, 2019.

Pfannerstill, M., Guse, B., and Fohrer, N.: Smart low flow signature metrics for an improved overall performance evaluation of hydrological models, J. Hydrol., 510, 447–458, https://doi.org/10.1016/j.jhydrol.2013.12.044, 2014.

Piniewski, M.: Classification of natural flow regimes in Poland, River Res. Appl., 33, 1205–1218, https://doi.org/10.1002/rra.3153, 2017.

Piniewski, M., Szczesniak, M., Kardel, I., Berezowski, T., Okruszko, T., Srinivasan, R., Schuler, V. D., and Kundzewicz, Z.: Hydrological modelling of the Vistula and Odra river basins using SWAT, Hydrolog. Sci. J., 62, 1266–1289, https://doi.org/10.1080/02626667.2017.1321842, 2017.

Piniewski, M., Szcześniak, M., Kardel, I., and Berezowski, T.: G2DC-PL+ A gridded 2 km daily climate dataset for the union of Polish territory and the Vistula and Odra basins, Dataset in 4TU Centre for Research Data, https://doi.org/10.4121/uuid:a3bed3b8-e22a-4b68-8d75-7b87109c9feb, 2020.

Raimonet, M., Oudin, L., Thieu, V., Silvestre, M., Vautard, R., Rabouille, C., and Le Moigne, P.: Evaluation of Gridded Meteorological Datasets for Hydrological Modeling, J. Hydrometeorol., 18, 3027–3041, https://doi.org/10.1175/JHM-D-17-0018.1, 2017.

Richter, D.: Ergebnisse methodischer Untersuchungen zur Korrektur des systematischen Messfehlers des Hellmannniederschlagsmessers, Vol. 194, Berichte des Deutschen Wetterdienstes, Offenbach am Main, 1995.

Schamm, K., Ziese, M., Becker, A., Finger, P., Meyer-Christoffer, A., Schneider, U., Schröder, M., and Stender, P.: Global gridded precipitation over land: a description of the new GPCC First Guess Daily product, Earth Syst. Sci. Data, 6, 49–60, https://doi.org/10.5194/essd-6-49-2014, 2014.

Serrano-Notivoli, R., Beguería, S., Saz, M. Á., Longares, L. A., and de Luis, M.: SPREAD: a high-resolution daily gridded precipitation dataset for Spain – an extreme events frequency and intensity overview, Earth Syst. Sci. Data, 9, 721–738, https://doi.org/10.5194/essd-9-721-2017, 2017.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling, J. Climate, 19, 3088–3111, 2006.

Somorowska, U.: Climate-driven changes to streamflow patterns in a groundwater-dominated catchment, Acta Geophys., 65, 789–798, https://doi.org/10.1007/s11600-017-0054-5, 2017.

Szcześniak, M. and Piniewski, M.: Improvement of Hydrological Simulations by Applying Daily Precipitation Interpolation Schemes in Meso-Scale Catchments, Water, 7, 747–779, 2015.

Terskii, P., Kuleshov, A., Chalov, S., Terskaia, A., Belyakova, P., Karthe, D., and Pluntke, T. D.: Assessment of Water Balance for Russian Subcatchment of Western Dvina River Using SWAT Model, Front. Earth Sci., 7, 241, https://doi.org/10.3389/feart.2019.00241, 2019.

Wang, J. and Tedesco, M.: Identification of atmospheric influences on the estimation of snow water equivalent from AMSR-E measurements, Remote Sens. Environ., 111, 398–408, 2007.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Water Resour. Res., 50, 7505–7514, 2014.

Williamson, T., Nystrom, E., and Milly, P.: Sensitivity of the projected hydroclimatic environment of the Delaware River basin to formulation of potential evapotranspiration, Climatic Change, 139, 215–228, https://doi.org/10.1007/s10584-016-1782-2, 2016.