~~PL1GD-T - gridded dataset of the mean, minimum and maximum daily air temperature at the level of 2 m for the area of Poland at a resolution of 1 km × 1 km~~

PL1GD-T: A high-resolution gridded daily air temperature dataset for Poland

Adam Jaczewski[1], Michał Marosz[1], Mirosław Miętus[1,2]

[1]Institute of Meteorology and Water Management – National Research Institute, Podleśna 61, 01-673 Warsaw, Poland

[2]University of Gdańsk, Department of Physical Oceanography and Climate Research, Bażyńskiego 4, 80-309 Gdańsk, Poland

Correspondence to: Adam Jaczewski (Adam.Jaczewski@imgw.pl)

Abstract. This paper presents a high-resolution gridded dataset of daily minimum (TN), mean (TG) and maximum (TX) near-surface air temperatures over Poland, covering the period from 1951 to 2020, with a spatial resolution of 1 km². ~~PL1GD-T dataset was developed using radial basis functions (RBFs) applied to quality-controlled observations from ground weather stations from~~The PL1GD-T dataset was developed using radial basis functions (RBFs), which were applied to quality-controlled observations from 347 ground weather stations at the Institute of Meteorology and Water Management – National Research Institute. TG is calculated consistently by means of the same formula throughout the entire period, and daily TN, TG, and TX fields are generated based on available daily records. Cross-validation methods evaluated the gridding procedure on a monthly basis. The linear RBF was selected~~employed~~ by hold-out cross-validation (HO-CV) as the most suitable for the gridding procedure among other RBFs. The leave-one-out cross-validation (LOO-CV) was performed to ensure the ability to reproduce the original characteristics variability. The values of the scores averaged over all stations for individual months are in the range 0.3-0.2, 0.3-0.2, 0.1-0.2 (K) for the bias, in the range 1.~~,~~23-1.~~,~~46, 0.69-0.92, 0.84-0.99 (K) for the root-mean-squared difference (RMSD) and in range 0.91-0.97, 0.98-0.99, 0.98-0.99 for the correlation for TN, TG and TX, respectively. The R~~S~~MSD is clearly altitude dependent, increasing from ~~low-land~~lowland to mountainous regions. The dataset's scope and resolution allowed robust estimation of local climate variability characteristics and observed trends. The availability of high-resolution datasets in both spatial and temporal contexts is essential for climate change impact analysis on a smaller scale. This new dataset provides a quality-validated, high-resolution, and open-access dataset that could be utilised by society, administrative bodies, or research institutions for ~~further analysis~~climate-related applications. The dataset is publicly available from the repository of the Institute of Meteorology and Water Management – National Research Institute under https://doi.org/10.26491/imgw_repo/PL1GD-T (Jaczewski et al., 2024).

— sformatowano: Angielski (Zjednoczone Królestwo)

1 Introduction

Contemporary climate change is one of the most significant issues facing civilisation. It affects societies in many ways, from direct physical impacts (droughts, floods, extreme events like tornadoes, heatwaves, and wildfires) to influencing the economies. We have no other way than to adapt, but our measures strongly depend on the quality and quantity of climate information we can acquire. Applications to research, natural resource management, and infrastructure planning typically require adequate spatial and temporal resolution (often 30 years of climate normality) and a record of past climate change (historical time series), you need long-term climate baseline data. The applicability of such data lies in the availability of measurement data from ~~European~~ National Meteorological and Hydrological Services (NMHS) networks and in the ability to assess climate change locally, which is more important in the context of adaptation. ~~The spatial resolution of most NMHS station networks is limited, and in the case of Poland, it is c.a. 50 km^2 (depending on the relief). This is a severe impediment in the adaptation or mitigation analyses, preventing the proper assessment of climate variability. The research aimed to develop a gridded database of daily mean, maximum and minimum temperatures (hereafter denoted as TG, TX and TN, respectively) with a spatial resolution of 1 km^2 (Jaczewski et al., 2024). The gridding procedures involved the application of RBFs (Radial Basis Functions) with several covariates, and the procedure's results were the subject of thorough comparison with in-situ measurements to ensure the ability to reproduce the original characteristics variability. The research outcome in a netCDF format database covers the 1951-2020 period, thus comprising 70 years of systematic measurements in the Polish NMHS network. Such database scope and the daily temporal and 1 km^2 spatial resolution allowed robust estimation of local climate variability characteristics and observed trends analysis, which society, administrative bodies, or research institutions can further utilise.~~ Climate change impact analysis requires the availability of high-resolution data sets in both spatial and temporal contexts. This is especially crucial if the study is to be performed on a smaller scale (administrative, physical-geographic). This assumption is rarely met with the stations' density operating in NMHSs. Thus, there is a clear need to prepare a quality ~~validated dataset with open access in~~, validated dataset with open access at high temporal and spatial scales.

During the XXI century, there were several attempts to provide such data. The E-OBS gridded dataset is one of the most used European daily databases (https://www.ecad.eu/download/ensembles/download.php). However, the spatial resolution is relatively coarse (0.1° × 0.1° or 0.25° × 0.25°; c. a. 11 km^2 or 25 km^2, respectively), which in the latter case may be insufficient for analyses of climatological conditions in small spatial scales. The present version is ~~28~~30.0e and was updated in ~~October~~March 202~~5~~3. However, a substantial advantage of this dataset is that it is derived from in-situ measurements. E-OBS is a daily gridded land-only observational dataset over Europe. The blended time series from the station network of the European Climate Assessment & Dataset (ECA&D) project forms the basis for the E-OBS gridded dataset. All station data are sourced directly from the NMHSs or other data-holding institutions. For many countries, the number of stations is the complete national network. Therefore, it is much more dense than the station network routinely shared among NMHSs (which is the basis of other gridded datasets). The density of stations gradually increases through collaborations with NMHSs within European research contracts. The position of E-OBS is unique in Europe because of the relatively high spatial horizontal grid

spacing, the daily resolution of the dataset, the provision of multiple variables and the length of the dataset. ~~Finally, the station data on which E-OBS is based are available through the ECA&D webpages (where the data owner has permitted to do so). In these respects, it contrasts with other datasets.~~ Aside from E-OBS, other European databases were produced ~~in the~~at a daily resolution. The set of gridded datasets covering Europe is provided in Table 1~~Table 1~~. Usually, the data available comprises mean and extreme daily ~~temperature~~temperatures.

Table 1. Selected global and European air temperature gridded datasets with a high spatial and daily temporal resolution.

Citation/dataset	Region	Variable	Interpolation method	Period	Spatial resolution
E-OBS v. 27.0e (Cornes et al., 2018)	Europe	TG, TN, TX	Monthly background fields from thin-plate splines and daily fields from the reduced-rank thin-plate spline	1950-present	0.1° × 0.1°
(Krähenmann et al., 2011)	Europe (WMO region 6)	TN, TX	Monthly background fields from block kriging and interpolation of daily anomalies from block kriging	2005-2008	5 km × 5 km
(Ntegeka et al., 2014)	Europe	TG, TN, TX	Inverse distance weighting	1990-2011	5 km × 5 km

Variables: TN – daily minimum temperature, TG – mean daily temperature, TX – maximum daily temperature

As seen in

Table 2

~~Table 2, a vast range of methodologies is~~ shows an extensive range of methods used in the calculations. Starting from the nearest neighbour (NN) approach through linear interpolation aided with IDW, distance weighting interpolation, thin plate splines, kriging, external drift kriging, inverse distance weighting, optimal interpolation, or combined methods. In most cases, the resolution rarely reaches 1 km^2 and is usually restricted to 5 km^2 or larger. In Central Europe, there are few daily databases of meteorological variables with a resolution higher than 5 km^2 and spanning at least 30 years and only Swiss datasets are routinely updated (Spatial Climate Analyses - MeteoSwiss, 2025). ~~Such is available only for Germany, Austria, and Switzerland, but in all those cases, the temporal scope is relatively short, only for Switzerland (1961-2010) and Austria, which covers 60 years (1961-2013).~~ This shows a need to produce gridded datasets of essential thermal characteristics on a daily scale covering the most extended possible period and utilising all possible, verified data quality station data. ~~In the case of our research, we were able to extend the time scope back to 1951 and forward to 2020, which resulted in 70 years of daily data.~~ Thus, for the reasons stemming from the needs of possible end users and the scientific community, it is crucial to make an effort to provide such data for Poland.

Table 2. This table lists national gridded datasets containing temperature variables with at least daily temporal resolution.

Dataset (citation)	Region	Variables	Interpolation method	Period	Temporal resolution	Spatial resolution	Validation method	Validation parameters
HYRAS	Germany and its	TN, TG, TX	interpolation method integrating	1951-2015	daily	5 km^2	LOO-CV, Fivefold	ME, MAE

(Razafimaharo et al., 2020)	catchment areas		non-linear elevation temperature profiles and IDW using multi-dimensional distances				cross-validation, Comparison with other datasets	
TRY (Krähenmann et al., 2018)	Germany	T	According to (Frei, 2014; Hiebl and Frei, 2016)	1995-2012	hourly	1 km^2	LOO-CV	RMSE, q-q
(Frei, 2014)	Switzerland	TG	A spatial scale-separation approach	1961-2010	daily	1 km^2	LOO-CV	MAE, RMSE
SPARTACUS (Hiebl and Frei, 2016)	Austria	TN, TX	A spatial scale-separation approach	1961-2013	daily	1 km^2	LOO-CV	ME, MAE, RMSE
FORBIO (Delvaux et al., 2015)	Belgium	TN, TX	EDK	1980-2013	daily	4 km^2	LOO-CV	RMSE
G2DC-PL+ (Piniewski et al., 2021)	Poland and its catchment areas	TN, TX	EDK	1951-2013	daily	5 km^2	LOO-CV	nRMSE
EMO-5 (Thiemig et al., 2021)	Europe	T, TN, TG, TX	IDW, modified SPHEREMAP, OK	1990-2019	daily 6-hourly	5 km^2	Comparison against a high-resolution regional grid	ME, MAE, MSE
STEAD (Serrano-Notivoli et al., 2019)	Spain	TN, TX		1901-2014	daily	5 km^2	LOO-CV	r, MAE, ME, RM, RSD
seNorge_2018 (Lussana et al., 2019)	Norway	TN, TG, TX	A spatial scale-separation approach	1957-2017	daily	1 km^2	LOO-CV	MAE, RMSE
SLOCLIM (Škrk et al., 2021)	Slovenia	TN, TX	GLMMs, GLMs	1950-2018	daily	1 km^2	LOO-CV	r, MAE, ME, RM, RSD
seNorge2 (Lussana et al., 2018)	Norway	T, TN	a modification of classical OI		daily, hourly			

Variables: TN – daily minimum temperature**,** TG – mean daily temperature**,** TX – maximum daily temperature, T – sub-daily temperature

Interpolation methods: OK – ordinary kriging**,** OI – optimal interpolation, EDK - External Drift Kriging, IDW – Inverse Distance Weighting, GLMM - generalised linear mixed models, GLM - generalised linear models

Validation methods: LOO-CV – the leave-one-out cross-validation

Validation metrics: r – correlation, (n)RMSE – (SD normalised) the root-mean-square error**,** RMSD – the root mean square difference, MAE – the mean absolute error, ME – the mean error (bias), RM – the ratio of means, RSD – the ratio of standard deviations, SD – standard deviation

The spatial resolution of most NMHS station networks is limited, and in the case of Poland, it is c.a. 50 km^2 (depending on the relief). This is a severe impediment in the adaptation or mitigation analyses, preventing the proper assessment of climate variability. The increasing availability and application of high-resolution Regional Climate Models (RCMs), including

convection-permitting models (CPMs), necessitate corresponding high-resolution observational datasets for robust evaluation and allows for more direct comparisons and validation. The research aimed to develop a gridded database of daily mean, maximum and minimum temperatures (hereafter denoted as TG, TX and TN, respectively) with a spatial resolution of 1 km^2 (Jaczewski et al., 2024). The gridding procedures involved the application of RBFs (Radial Basis Functions) with several covariates, and the procedure's results were the subject of thorough comparison with in-situ measurements to ensure the ability to reproduce the original climate characteristics variability. The research outcome in a netCDF format database covers the 1951-2020 period, thus comprising 70 years of systematic measurements in the Polish NMHS network. The study's main purpose is to provide a high and spatial resolution climate dataset for Poland, initially consisting of temperature variables. The provided methodology is the basis for introducing subsequent variables available from the network to the following versions of the dataset and will be improved by relying on the gained experience. ~~Such database scope and the daily temporal and 1 km^2 spatial resolution allowed robust estimation of local climate variability characteristics and observed trends analysis, which society, administrative bodies, or research institutions can further utilise.~~

2 Data and methods

2.1 Station data

Using a long series of meteorological data makes it possible to carry out analyses related to the impact of the variability of meteorological conditions over a selected area on various sectors of human activity. Ensuring data quality is crucial in this case. IMGW-PIB has taken steps to develop complete information on the daily variability of thermal and precipitation conditions in Poland since 1951. The analysis and prepared data series include information on the daily values of the following meteorological variables: average daily air temperature, maximum daily air temperature, and minimum daily air temperature. The routine measurements from SYNOP and CLIMAT stations are subject to formal checks based on WMO standards (World Meteorological Organization, 1993). On the other hand, WMO~~The official~~ have justified different methods of calculating the characteristics, particularly the average daily air temperature~~,~~. The methods varied in different periods and depended on the station's measurement program, which was different at synoptic and climatological stations and is the main source of the series' inhomogeneity. One method of calculating the average temperature was necessary to compare data and ensure consistency~~homogeneity~~ at the entire available range of stations in the whole period. ~~Due to the use of data from both synoptic and climatological stations in the set, t~~. The same method was used to calculate the ~~average~~ daily mean~~air~~ temperature, following the one used for the Polish climatological stations since 1996:

$$TG = \frac{T06_{06} + T18_{18} + TX + TN}{4} \quad (1)$$

where T06 and T18 are measurements at 06UTC and 18UTC, respectively; TX is the daily maximum temperature, and TN is the daily minimum temperature. The day for TX and TN readings begins at 18UTC of previous day. This way, consistent

information was obtained on the spatial variability of thermal conditions throughout the country in the multiannual period. ~~And even though in the case of synoptic stations, the values of the TG calculated in this way are not the official daily average (average of 8 measurements), the possibility of a uniform analysis throughout the country brings more benefits than potential losses resulting from a different calculation method.~~ In this way, the so-called uniform air temperature ~~series includes~~ data were elaborated for 347 stations. Additionally, this approach mitigated inconsistencies introduced by network automatization since 90s. Mostly, interpolated gridded datasets developed for climate applications rely on homogenised station series. The homogenisation of daily series is more complicated than on a monthly scale. The procedure must consider weather variability to incorporate stations' metadata and support higher-order adjustments. Existing homogenisation techniques regard these issues, but their application to daily data is still developing (Killick et al., 2022).

It should be noted that daily data availability varies from 1951 to 2020. A daily number of stations available for spatial interpolation is presented in Figure 2a. The number of stations increased up to the late 80s and early 90s, resulting in a maximum number of uniform data, 268 records. Since the late 90s, spatial coverage has degraded to only 119 locations. A number of stations by data availability is presented in Figure 2b. Nearly 80 stations allowed for more than 90% of records, and half of the locations overlapped for more than half of the period. Neither possible degradation in temporal consistency nor missing values filling was considered in this study.

— **sformatowano:** Kolor czcionki: Automatyczny

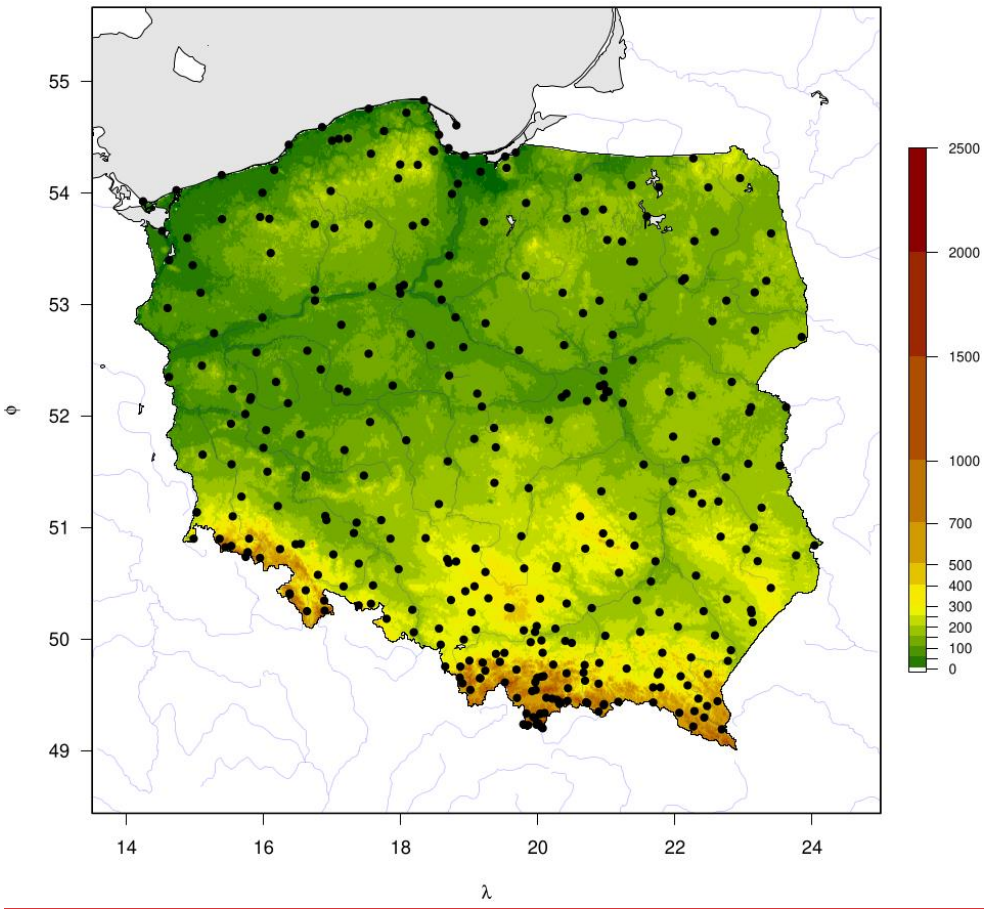
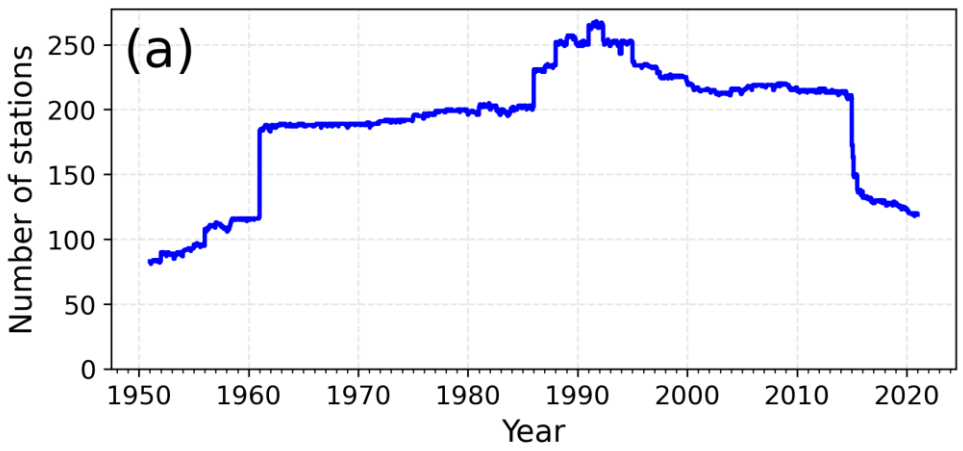


Figure 1. Orography and spatial distribution of meteorological stations in Poland with temperature time series in the 1951-2020 period.

— sformatowano: Czcionka: Nie Pogrubienie

Sformatowano: Legenda



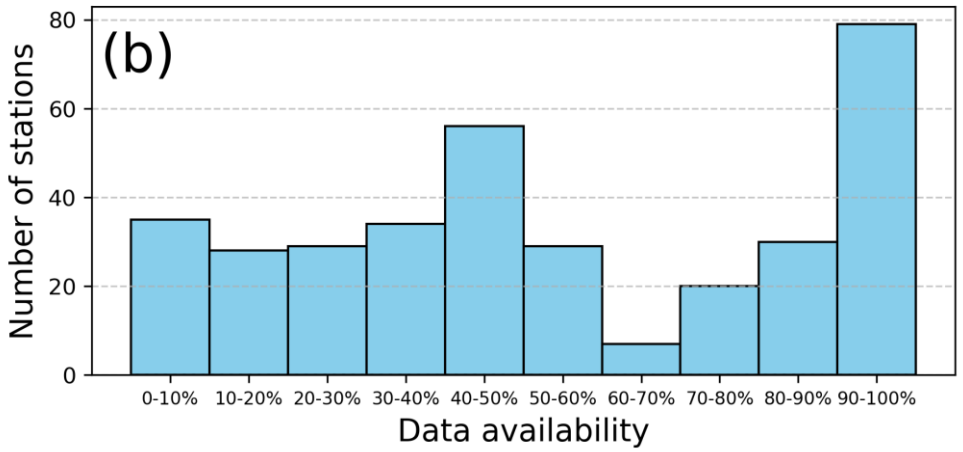


Figure 2. Number of daily records (a) and stations by data availability ranges (b).

The station density in Poland is relatively low compared to other European countries, reaching nearly one station per 1000 square kilometres in the year when the number of operating stations was the highest. Most stations are located in the lowland area, which occupies most of the country. Meteorological networks in mountainous regions across many countries are often

Sformatowano: Razem z następnym

sformatowano: Czcionka: Nie Pogrubienie

sparse, but this is not the case in Poland. The station density in the Polish mountains, covering 0.23% of the country (area with altitude exceeding 1000 m a.s.l.), is relatively high compared to the lowland and is occupied by 10 stations (Table 3), resulting in an order of magnitude higher station density in these regions. On the other hand, data availability decreases with altitude, from more than 56% at the lowlands to 42% at the highest mountains on average.

Table 3. The data records statistics by altitude ranges (m a.s.l). N: number of stations; *D*: station density; OBS: percentage of available records; *A*: percentage of country area occupied by grid points.

Altitude class	*N*	OBS (%)	*D*	*A*
0 - 300	258	56.35	9	9
300 - 500	47	53.59	2	5
500 - 1000	32	45.51	3	3
> 1000	10	42.34	1	0

2.2 Spatial interpolation

To interpolate the temperature data from the relatively ~~sparse meteorological stations across our area, we employed Radial~~ Basis Functions (RBF). While referenced in Table 1 and Table 2, the temperature datasets of high spatial and temporal resolution often use alternative interpolation methods, RBFs have been successfully applied for interpolation station observational data in meteorology (Antal et al., 2021; Fathizad et al., 2017; Piri, 2017; Ryu et al., 2024; Saaban et al., 2013; Wang et al., 2014; Yang and Xing, 2021) or hydrology (Gao et al., 2022; Wypych and Ustrnul, 2011). RBF interpolation models have also been used in agricultural applications for prediction purposes (Rocha and Dias, 2018, 2019). RBF interpolation method could be defined by Eq. (2):

$$\boldsymbol{F}(\boldsymbol{r}) = \sum_{j=1}^{m} \boldsymbol{\alpha_j} \boldsymbol{\varphi}(\boldsymbol{r_j}), \quad (2)$$

Where α_j are the weights, ~~and~~ $\varphi(r_j)$ RBF function of the distance r_j between interpolation point to m sample points and r is Euclidean vector (coordinates) of interpolation point. Selecting the appropriate RBF function determines the accuracy of the interpolation. Different RBF functions were found to be suitable for interpolation of temperature data (Piri, 2017; Ryu et al., 2024; Saaban et al., 2013). In our study, s~~S~~ix RBFs were used and given as,

linear: $\varphi(r_j) = r_j$ (3)

cubic: $\varphi(r_j) = {r_j}^3$ (4)

thin plate spline: $\varphi(r_j) = {r_j}^2 \log r_j$ (5)

gaussian: $\varphi(r_j) = e^{-\left(\frac{r_j}{\beta}\right)^2}$ (6)

multiquadric: $\varphi(r_j) = \sqrt{1 + \left(\frac{r_j}{\beta}\right)^2}$ (7)

— **sformatowano:** Czcionka: Nie Pogrubienie
— **sformatowano:** Czcionka: Kursywa
— **sformatowano:** Czcionka: Nie Pogrubienie
— **sformatowano:** Czcionka: Nie Pogrubienie
— **sformatowano:** Czcionka: Kursywa
— **sformatowano:** Czcionka: Kursywa
— **sformatowano:** Czcionka: Nie Pogrubienie, Kursywa
— **sformatowano:** Czcionka: Kursywa
— **sformatowano:** Czcionka: Kursywa
Sformatowano: Z prawej
Tabela sformatowana
— **sformatowano:** Czcionka: Nie Pogrubienie
Sformatowano: Z prawej
— **sformatowano:** Czcionka: Nie Pogrubienie
Sformatowano: Z prawej
— **sformatowano:** Czcionka: Nie Pogrubienie
Sformatowano: Z prawej
— **sformatowano:** Czcionka: Nie Pogrubienie
Sformatowano: Z prawej
— **sformatowano:** Czcionka: Nie Pogrubienie
— **sformatowano:** Indeks górny
— **sformatowano:** Czcionka: Nie Pogrubienie
Sformatowano: Odstęp Po: 0 pkt

inverse multiquadric: $\varphi(r_j) = \frac{1}{\sqrt{1+\left(\frac{r_j}{\beta}\right)^2}}$ (8)

where β is shape parameter. In the three dimensional space the Euclidean distance is given by:

$$r_j = \sqrt{(x - x_j)^2 + (y - y_j)^2 + (z - z_j)^2} \quad (9)$$

Where (x, y, z) represent coordinates of interpolation point. For a given dataset of m points, the interpolated n values are obtained by following the linear system of n equations:

$$F(x_i, y_i, z_i) = \sum_{j=1}^{m} \alpha_j \varphi(r_j), i \in \{1 \dots n\} \quad (10)$$

In our case, m corresponds to the number of stations varying temporally according to daily data availability; x and y correspond to geographical coordinates, and z is elevation. Generally, the coordinates and elevation are treated as predictors, and more explanatory variables can be introduced to the interpolation procedure. It is identified that the ~~incorporating of elevation data as an auxiliary variable,~~ incorporation of elevation data as an auxiliary variable accounts for orographic effects, mitigates the smoothing effect and improves the representation of gradients in daily gridded data sets (Brinckmann et al., 2016; Delvaux et al., 2015; Frei, 2014). ~~We intentionally limited the variables to geographical coordinates, as temperature and altitude depend most on them.~~ Choosing the optimal value of the shape parameter is not a straightforward task (Fasshauer and Zhang, 2007). Often, this problem is paved over in the application studies, and it should be assumed that the value has been chosen as the mean distance between the nearest points (observational stations) as it is implemented in numerical tools as the default option. We have used this value in the interpolation procedure with gaussian, multiquadric and inverse multiquadric functions. The shape parameter ranges from 15 to 22 km for the daily maximum (268 stations) and minimum (81 stations) number of data records, respectively.

2.3 Temporal and spatial reference of the dataset

The dataset consists of daily gridded TN, TG and TX covering Polish territory at 1 km^2 spatial resolution for 1951-2020. The interpolation is in the grid projected in the PUWG 1992 (EPSG:2180) coordinate reference system. The elevation data originates from DTM for Poland (DTM 100m, 2023), corresponding to DEM. Given the data, simplify the interpolation procedure as all three coordinates (x, y and z) are in the same units (meters). Additionally, we avoided topography interpolation by defining the grid as every tenth x and y coordinate of the 100 m × 100 m DTM mesh and assigning the average elevation values over the 1 km^2 surrounding area.

— **sformatowano:** Indeks górny

3 Results

~~2.4~~3.1 Validation

The most popular technique for validating interpolation techniques is leave-one-out cross-validation (LOO-CV). It involves using all except one observation data for interpolation, and the remaining station is used for validation. The main drawback of this method is relatively high computational time, as the procedure has to be repeated for every station, which has to be removed from the whole dataset. Another cross-validation method, the hold-out technique (HO-CV), is used to avoid long calculations. This method splits the observation data into interpolation and validation sets. Daily, 95% of stations are randomly removed, and 5% are used for validation. The procedure is repeated every day. The hold-out technique was chosen to select the RBF function, and the LOO-CV technique was used to evaluate results with the selected RBF function.

To evaluate the accuracy of the interpolated data, mean error (ME), root mean square difference (RMSD), centred RMSD (cRMSD), and Pearson's correlation coefficient have been used. The ~~two latter~~latter two metrics and standard deviations have been presented through Taylor's diagrams. The tables with numeric results rounded to two decimal points were provided, including differences in the 5th and 95th percentiles. The observed data was compared to the nearest interpolation grid point, and validation was performed monthly. The following Equations define the metrics:

mean error (bias): $$\mathbf{ME} = \frac{1}{N}\sum_{n=1}^{N}(I_n - O_n) \quad (11)$$

root mean square difference: $$\mathbf{RMSD} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(I_n - O_n)^2} \quad (12)$$

Centred root-mean-square difference: $$\mathbf{cRMSD} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}[(I_n - \bar{I}) - (O_n - \bar{O})]^2} \quad (13)$$

Pearson's correlation coefficient: $$\mathbf{r} = \frac{\frac{1}{N}\sum_{n=1}^{N}(I_n - \bar{I})(O_n - \bar{O})}{\sigma_I \sigma_O} \quad (14)$$

Where I and O denote interpolated and observed values, N is the number of observations, and σ_I and σ_O are standard deviations of interpolated and observed values, respectively. $\overline{I}$ and $\overline{O}$ are the mean of interpolated and observed values, respectively.

— **sformatowano:** Czcionka: Nie Pogrubienie

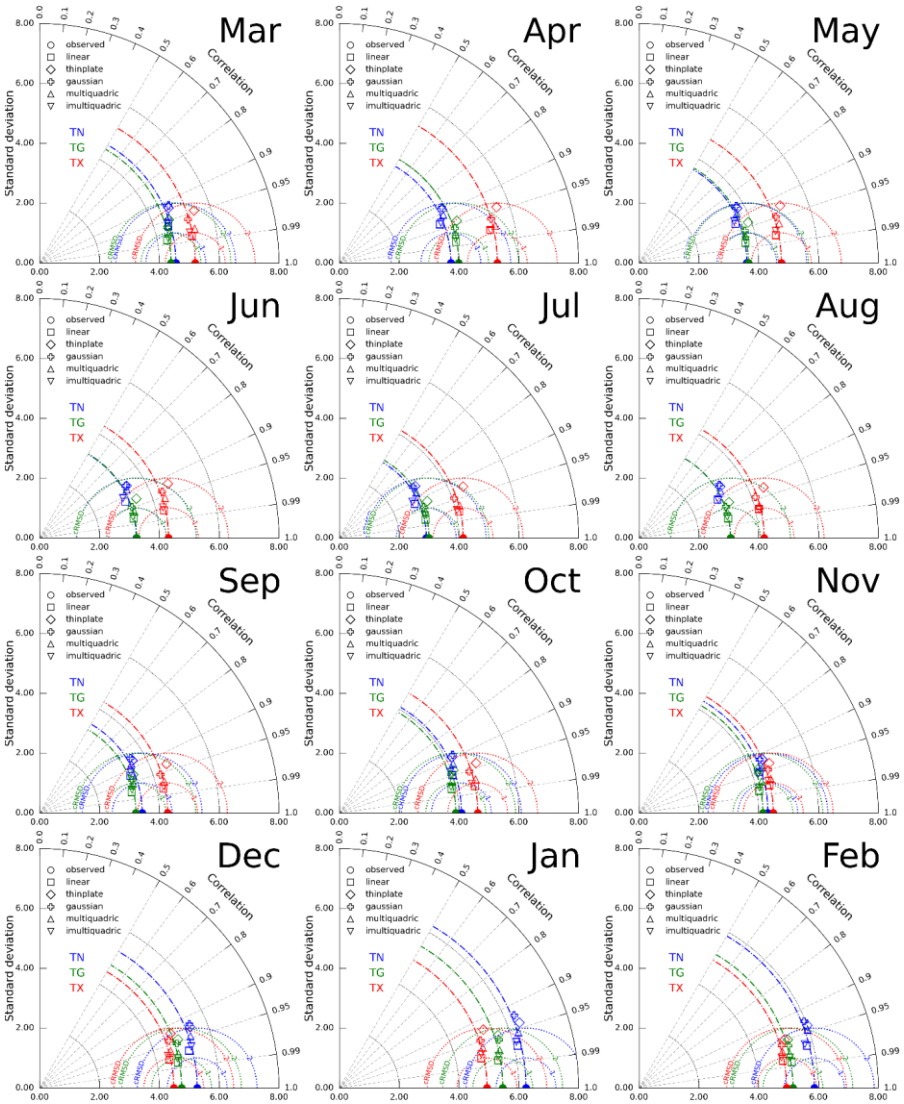
~~3.1 Results~~

~~3.1~~3.2 Hold-out cross-validation (HO-CV)

~~Using HO-CV, we will assess which RBF function is the most appropriate for interpolating temperature variables. This involves dividing the observation data into interpolation and validation sets. From the daily observational dataset, 95% of stations are randomly selected for interpolation, while 5% are used for validation. The process is repeated daily, resulting in varying interpolated and validation datasets.~~ The validation process results are presented using Taylor diagrams, which provide

a quick statistical summary of how well the interpolated~~modelled~~ and observed data align. We show mean values of correlations, centred root-mean-square differences, and standard deviations for all RBF methods every month. While a common approach is to normalise the differences and standard deviations of ~~modelled~~interpolated data by the standard deviations of observations, it may negatively impact the clarity of the figures in our case. As the variability (SDs) of the variables differ across seasons and variables, we can present the results for all methods and temperatures in the exact figure.

The linear approach is undoubtedly the most accurate for all months and variables (Figure 3~~Figure 1~~). Only for TX in April and August, values interpolated by the inverse multiquadric RBF function give similar agreement with observations. The mean cRMSD for the linear RBF function is lower than 1 for TG and TX, but for TN, it exceeds 1. This method also sustains observed variability as the standard deviations of interpolated and observed values are similar. Correlations exceed 0.95 for all variables and months except TN from April to September. These results allowed us to proceed further only with the linear method.


Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
Feb
observed
linear
thinplate
gaussian
multiquadric
imultiquadric
TN
TG
TX
Standard deviation
Correlation
CRMSD

240

Figure 3~~1~~. Taylor diagrams for HO-CV of spatial interpolation of TN, TG and TX by different RBF functions. The colours correspond to the variables in the preceding order – blue, green and red. Filled circles on the x-axis and dot-dashed minor arcs represent standard deviations of observed values; dotted semicircles correspond to cRMSD values. The azimuthal position represents the correlation coefficient. The marker's shape describes the type of RBF used for interpolation. For a detailed explanation of the diagram's principles, see Figure 2 in Taylor (2001).

— **sformatowano:** Czcionka: Nie Pogrubienie

~~3.2~~3.3 Leave-one-out cross-validation (LOO-CV)

LOO-CV has emerged as a widely used verification methodology for spatial interpolation techniques in meteorology and climatology. In executing this method, the given station is removed from the daily observation dataset, and the remaining stations are used to construct the interpolated field. The value at the grid nearest to the removed station is compared to the observed value. The procedure is repeated for all stations and days in the analysed period, and the mean values of the comparison metrics are presented in Taylor diagrams (Figure 4~~Figure 2~~). The charts depict the level of agreement between the observed and interpolated values. The results show that the best agreement is achieved for TG in winter and spring, with the lowest cRMSD and highest correlation values. Although the agreement for TX is slightly worse than the mean, it is still relatively favourable overall, with the best agreement seen in autumn. However, the interpolation methodology performs worse in capturing the nuances of TN, with the best agreement observed in winter. It is important to note that the variability of all temperatures is preserved in the standard deviation, with similar values between observed and interpolated data. ~~The precise values for these and additional metrics are in the following tables.~~

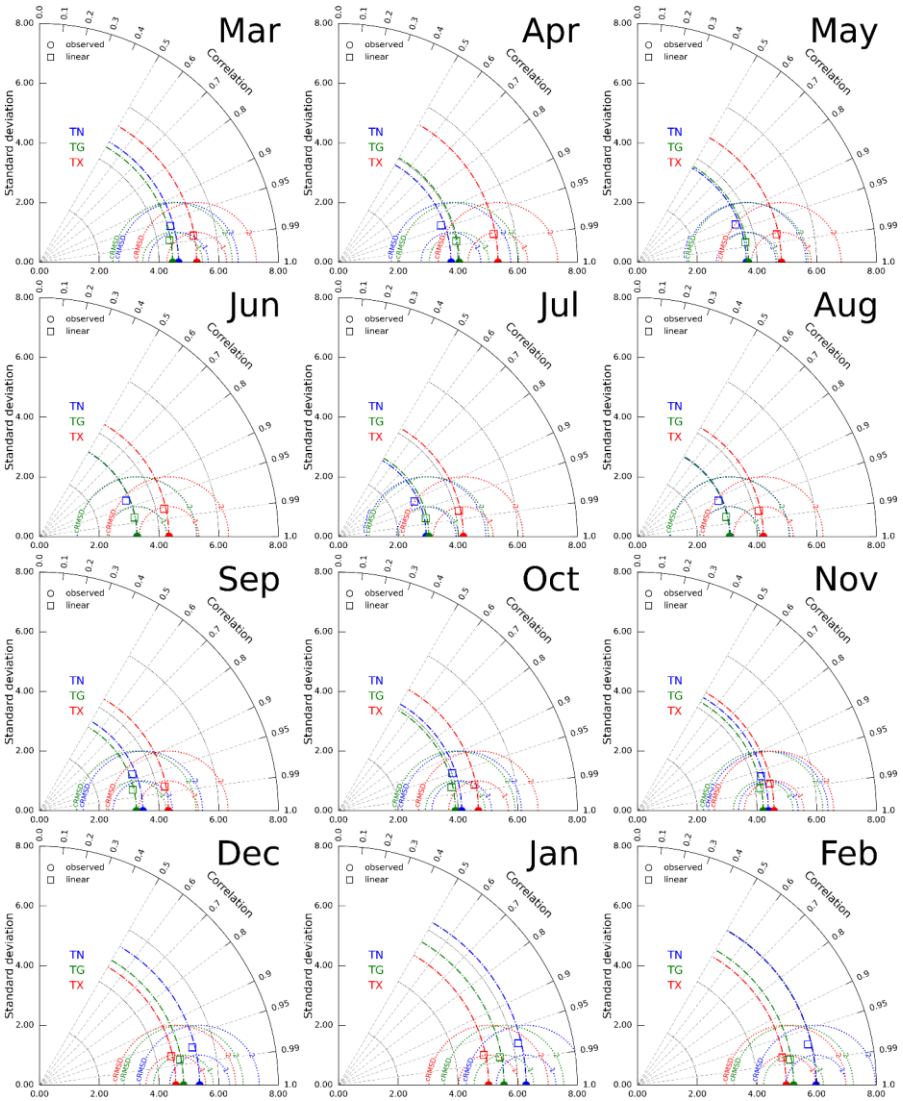


Figure 42. Taylor diagrams for LOO-CV of spatial interpolation of TN, TG and TX by linear RBF function. See Figure 3 for more details.

— sformatowano: Czcionka: Nie Pogrubienie

Kod pola został zmieniony

Table 4~~Table 3~~ provides the set of quality metrics for RBF interpolation. Mean error (ME, a.k.a BIAS) is minimal and, for all variables, does not exceed 0.03 with minuscule interannual variability. TN is characterised by the highest RMSD, reaching, on average, 1.29. The RMSD annual value for TG equals 0.75, whereas for the TX, it is slightly higher (0.92). The highest average annual value is recorded for TN (1.29). For all analysed variables, Pearson's correlation coefficient is very high. An annual course ranges from 0.94 (TN) to 0.98 (TG, TX). Only in the case of TN, there is a slight seasonal variability, with the lowest values (0.91) in July and the highest (0.97) in winter (Dec-Jan). RM (the ratio of means) is close to 1 with a range of 0.99 – 1.01 for all months. RSD (the ratio of standard deviation) is very high, nearing 1 for all variables, pointing to slight variability underestimation. Slight ~~with minimal~~ seasonal variability is observed for TN – lower values are seen in the summer. ~~Such a situation suggests a slight variance flattening in the outcome RBF data and further assures the high quality of the RBF interpolation results.~~

The difference in the values of ~~quantiles (5% and 95%) suggests slight underestimation in the case of Q95 and~~ the 5th and 95th percentiles suggest a slight underestimation in the case of Q95 and an overestimation for Q05. This pattern repeats for all variables (TN, TG, and TX), with the highest differences in average annual values recorded for TN.

Table 4~~3~~. Overall monthly validation measures (OBS vs GRIDDED).

VAR		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
TN	ME	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02
	RMSD	1.46	1.45	1.31	1.40	1.44	1.38	1.37	1.41	1.39	1.38	1.23	1.32
	cRMSD	1.43	1.39	1.25	1.28	1.32	1.26	1.24	1.25	1.27	1.30	1.18	1.28
	Pearson	0.97	0.97	0.96	0.94	0.93	0.92	0.91	0.92	0.93	0.94	0.96	0.97
	RM	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	RSD	0.99	0.98	0.98	0.97	0.96	0.96	0.95	0.96	0.96	0.97	0.98	0.98
	Q95D	-0.12	-0.06	-0.18	-0.17	-0.15	-0.14	-0.17	-0.22	-0.23	-0.13	-0.18	-0.13
	Q05D	0.27	0.27	0.25	0.25	0.31	0.28	0.33	0.32	0.35	0.23	0.14	0.28
TG	ME	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	RMSD	0.92	0.87	0.77	0.78	0.74	0.70	0.69	0.74	0.78	0.84	0.78	0.85
	cRMSD	0.94	0.85	0.74	0.72	0.68	0.64	0.62	0.67	0.71	0.80	0.76	0.86
	Pearson	0.99	0.99	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
	RM	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	RSD	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99
	Q95D	-0.10	-0.02	-0.04	-0.02	-0.04	-0.07	-0.07	-0.10	-0.02	-0.02	-0.11	-0.03
	Q05D	0.20	0.14	0.14	0.10	0.06	0.04	0.10	0.12	0.06	0.16	0.08	0.07
TX	ME	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.02	0.02	0.02
	RMSD	0.99	0.94	0.91	0.95	0.98	0.96	0.92	0.90	0.84	0.89	0.91	0.95
	cRMSD	1.02	0.93	0.90	0.96	0.95	0.94	0.87	0.87	0.82	0.89	0.92	0.97
	Pearson	0.98	0.98	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
	RM	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	RSD	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	Q95D	-0.07	-0.06	-0.09	-0.05	-0.06	-0.11	-0.05	-0.08	-0.11	-0.14	-0.07	-0.11
	Q05D	0.16	0.15	0.12	0.13	0.22	0.13	0.11	0.11	0.08	0.16	0.08	0.13

ME – mean error (K), RMSD – the root mean square difference (K), cRMSD – – centred root mean square difference (K), Pearson – Pearson's correlation coefficient, RM – the ratio of means, RSD – the ratio of standard deviations, Q95D and Q05D – the difference in 95th and 5th between gridded and observed values (K).

— **sformatowano:** Indeks górny

— **sformatowano:** Indeks górny

Verification of RBF interpolation quality based on RM~~SD~~ for distinct altitude (m a. s. l.) classes confirms overall quality characteristics. A table is intentionally not shown as the values for particular stations are in the range 0.992 – 1.008, resulting in 1.00 after rounding to two decimal places for all months and altitude ranges.~~, but there is an apparent increase in RMSD values with the rise in altitude for TN. Up to 300 m, the RMSD values are relatively uniform throughout the year and range from 1.08 in November to 1.35 in May. On average, the annual RMSD value equals 1.26. In the 300-500 m altitude class, there is an increase in RMSD value to 1.58 (annual) with the range from 1.45 (June) to 1.83 (January). The next class (500-1000 m) shows yet another increase to 2.20 (annual average), with the range between 2.00 (June) to 2.65 (January). In the highest class above 1000 m, the RMSD value increases again, reaching an average of 2.59 with a range between 2.24 (June, July) and 3.16 (January). What is apparent is the increase in seasonal RMSD variability with altitude, with relatively small interannual differences at a lower altitude to the greatest in the case of the highest altitude class.~~

~~For daily mean air temperature (TG), the overall pattern of RMSD increase with altitude is confirmed. However, what is apparent is that this RMSD increase is much smaller than in the case of extreme temperatures. The average annual values range from 0.71 (0-300m) to 1.97 (> 1000 m). The seasonal variability in the lowest altitude class is minimal and ranges from 0.65 (July) to 0.77 (January). With the altitude increase, the seasonal variability increases with higher RMSD values in colder months and lower during the warm part of the year. At 300-500 m, the annual average RMSD increases slightly to 0.91, but what appears is more apparent seasonal variability with values ranging from 0.72 (June) to 1.18 (January). For the following two altitude classes, the average annual RMSD exceeds 1.00. In the 500-1000 altitude class, the annual average RMSD equals 1.33; stations above 1000 m reach 1.97. The increased seasonal variability with the RMSD range for both classes at 0.96 is also typical for higher altitude types.~~

~~TX follows the pattern mentioned above with some distinctive differences. The annual average values range from 0.88 (0-300 m) to 2.14 (> 1000 m). What strikes relatively small RMSD values exist for two lower altitude classes closer to the TG RMSD values than TN. Also, the seasonal variability is less pronounced than for TN. The highest interannual variability of TX RMSD is recorded for the 500-1000 m altitude class, ranging from 1.00 (July) to 1.82 January).~~

~~**Table 5.** Overall monthly RMSD (OBS vs. GRIDDED) values for station altitude classes.~~

	altitude class[m]	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
TN	0 - 300	1.28	1.30	1.20	1.29	1.35	1.32	1.29	1.32	1.31	1.25	1.08	1.14
	300 - 500	1.83	1.72	1.51	1.56	1.56	1.45	1.47	1.53	1.55	1.62	1.49	1.65
	500 - 1000	2.65	2.47	2.05	2.20	2.19	2.00	2.03	2.17	2.12	2.13	2.00	2.37
	> 1000	3.16	2.88	2.56	2.43	2.38	2.24	2.24	2.37	2.49	2.85	2.60	2.85
TG	0 - 300	0.77	0.76	0.70	0.72	0.71	0.68	0.65	0.69	0.71	0.74	0.66	0.70

	300 - 500	1.18	1.04	0.88	0.86	0.79	0.72	0.73	0.80	0.87	0.97	0.95	1.08
	500 - 1000	1.90	1.62	1.27	1.24	1.09	0.94	0.97	1.11	1.24	1.40	1.43	1.74
	> 1000	2.52	2.18	1.88	1.83	1.77	1.56	1.57	1.67	1.85	2.22	2.20	2.38
TX	0 - 300	0.85	0.84	0.87	0.99	1.03	0.99	0.92	0.88	0.79	0.80	0.79	0.81
	300 - 500	1.23	1.10	0.98	0.95	0.94	0.93	0.90	0.91	0.91	1.02	1.10	1.18
	500 - 1000	1.82	1.51	1.23	1.11	1.07	1.03	1.00	1.03	1.09	1.29	1.51	1.77
	> 1000	2.51	2.25	2.06	2.15	2.11	1.92	1.84	1.85	1.93	2.20	2.38	2.47

The ratio of standard deviation (RSD) showed slight variance underestimation (Table 5), the highest for TN at 500-1000 m in summer. The ratio for TN also experiences the highest seasonal variability, where the smallest values are seen in summer and the highest in winter. Surprisingly, the cases of agreement close to perfect are seen most often at >1000 m.

— **sformatowano:** Czcionka: Nie Pogrubienie

Table 5. Overall monthly RSD values for station altitude classes.

	altitude class	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
TN	0 - 300	0.99	0.99	0.98	0.97	0.97	0.96	0.96	0.96	0.97	0.98	0.98	0.99
	300 - 500	0.97	0.98	0.98	0.97	0.96	0.96	0.96	0.97	0.96	0.97	0.98	0.98
	500 - 1000	0.98	0.97	0.95	0.95	0.94	0.95	0.93	0.93	0.95	0.96	0.96	0.98
	> 1000	1.00	0.99	0.99	0.98	1.00	0.98	0.97	0.97	1.00	0.98	0.98	1.00
TG	0 - 300	0.99	0.99	0.99	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	300 - 500	0.98	0.99	0.99	1.00	0.99	0.99	1.00	1.00	0.99	0.99	0.99	0.99
	500 - 1000	0.98	0.98	0.97	0.98	0.97	0.98	0.97	0.98	0.97	0.98	0.97	0.98
	> 1000	0.98	0.97	0.98	0.99	0.99	0.98	0.98	0.97	0.98	0.96	0.97	0.98
TX	0 - 300	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.99	0.99
	300 - 500	0.98	0.99	0.99	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98
	500 - 1000	0.98	0.99	0.99	0.97	0.98	0.98	0.99	0.99	0.98	0.98	0.98	0.98
	> 1000	0.98	0.97	0.99	0.997	0.99	1.00	1.00	0.99	1.00	0.98	0.97	0.96

Figures Figure 3, Figure 4, Figure 5 present the spatial variability of RMSD for all analysed variables on a monthly scale (with seasons in rows beginning from spring). There is an apparent pattern with relatively low RMSD values in central Poland in the lowlands and substantially higher in the mountains. This pattern is quite common for other variables in most months.

Suppose one divides RMSD values into bins of equal width, equalling 1 with breaks from 0 to 5. In that case, one can assess the RBF method's overall quality and share the station's occurrence in those five bins, with the first referring to the 0-1 RMSD boundaries and the fifth with values of RMSD between 4 and 5.

TN (Figure 3) exhibits the most significant overall variability, and it is the only variable for which the fifth bin (RMSD values from 4 to 5) is filled in the annual cycle. However, the fifth bin's average annual share of stations is only 0.1% (3 cases out of 4164). Also, there is a noticeable difference between the share of the first bin, which, on average, is only 57.8% of the cases. In contrast, for TG and TX, it is over 90%, which suggests that in the case of minimum temperature, the RBF interpolations'

general quality seems lower than in the case of the other two variables. The second bin (RMSD - 1 to 2) comprises, on average, 39.7% of cases, which is substantially higher than for TG and TX. In the case of the first two bins, there is also high seasonal variability. For the first bin, the values range from 51.3% (May) to 70.9% (November). For the second bin (RMSD values from 1 to 2), the share varies from 27.1% (November) to 45.8% (May).

TG (Figure 4), an even higher share (94.8%) of the stations falls into the first RMSD bin (0-1 values). The seasonal variability is relatively low, ranging from 89.3% (January) to 98.0% (June, July). In the second bin (on average 4.4% of the stations), there is a substantial seasonal variability with RMSD values ranging from 1.4% (June) to 8.9% (January, February)

The overall share of RMSD values in the two upper bins (third and fourth) is less than 1%, with more pronounced seasonal variability in the third bin.

TX (Figure 5), on average (annual), over 90% of stations fall into the first bin (0 to 1 RMSD value). The seasonal variability is relatively low and ranges from 86.5% (January) to 94.5 (September). The second bin's (1-2 RMSD value) average equals 8.5%, with more pronounced seasonal variability ranging from 4.9% (September, October)) to 12.1% (January). Two higher classes of RMSD values (2-3 and 3-4) comprise, in total, on average, only 0.9% of stations, but in the case of the 2-3 class, one can see a high seasonal variability with RMSD values ranging from 0.3 to 1.4 (May).

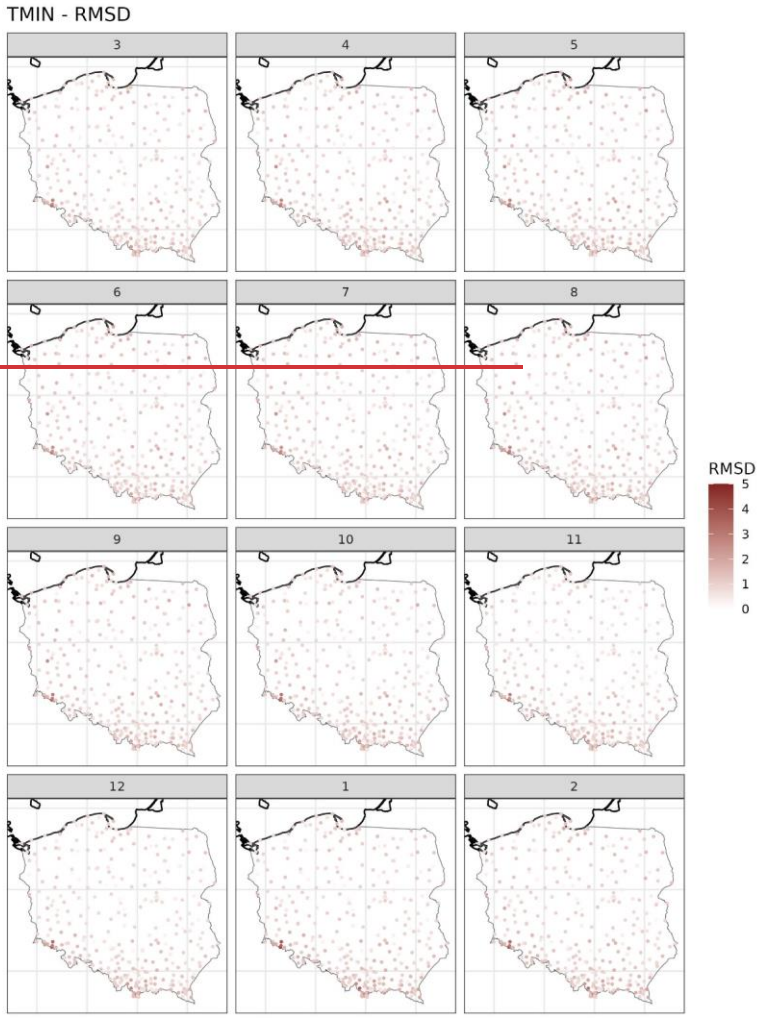
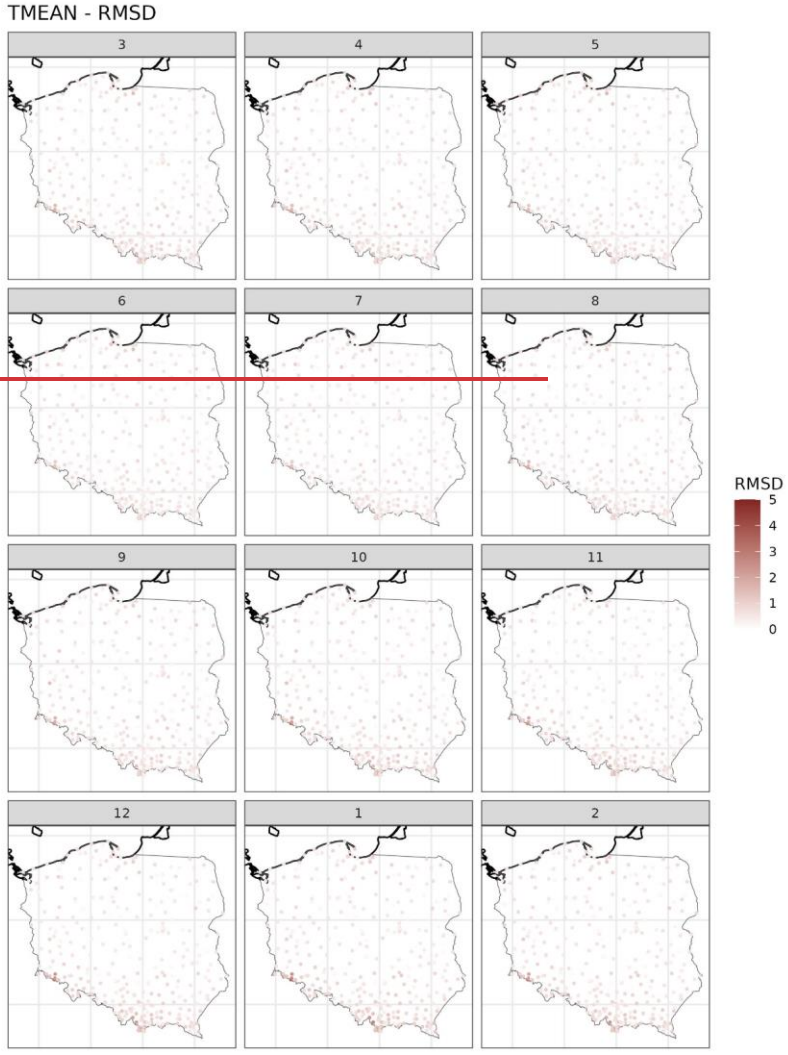


Figure 3. Monthly RMSD values (degrees Celsius) for minimum daily temperature (TN) at selected stations.



340

Figure 4. Monthly RMSD values (degrees Celsius) for mean daily temperature (TG) at selected stations.

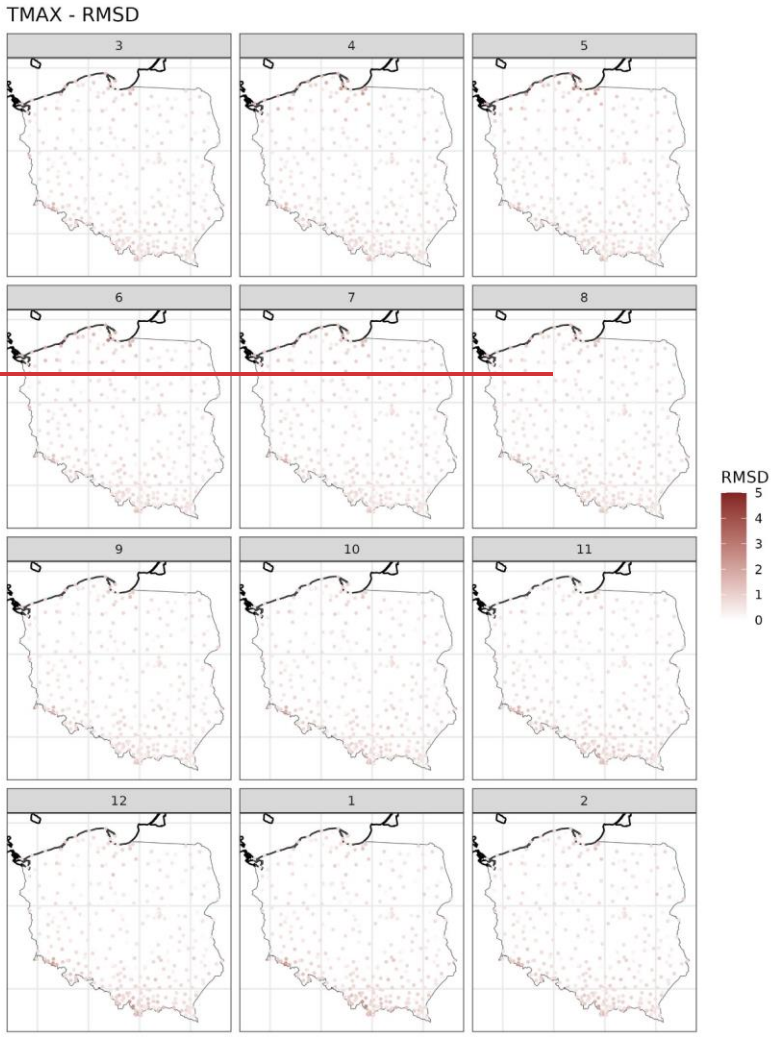


Figure 5. Monthly RMSD values (degrees Celsius) for maximum daily temperature (TX) at selected stations.

Derived with RBF data series may be a source of valuable information on recorded climate change characteristics in Poland. The daily temporal resolution of the data set allows the possibility of deriving multiple statistical characteristics with very high spatial resolution and detailed analysis of the thermal characteristics on the regional or even local scale (counties, small cities, or villages). For example, Figure 6 provides insight into the spatial variability of TX 90% quantile in Poland for two 30 years: 1951-1980 and 1991-2020 in January and July. As expected in January, there is a substantial shift in the values of TX. The overall range of recorded values changes from -1.4°C - 7.7°C (1951-1980) to 1.1°C-10.6°C (1991-2020), which indicates a 2.5°C shift in the lower boundary and 2.9°C in the upper indicating higher variability in the maximum values of TX. The overall spatial pattern of the TX 90% quantile remains relatively intact. However, the shift is visible in the whole country. In July, the change in the range of TX 90% quantile also reaches 2°C from 13.6°C-30.8°C to 15.6°C-32.9°C. This is also clearly visible for nearly the whole country but less pronounced in the mountains and the vicinity of the Baltic Sea.

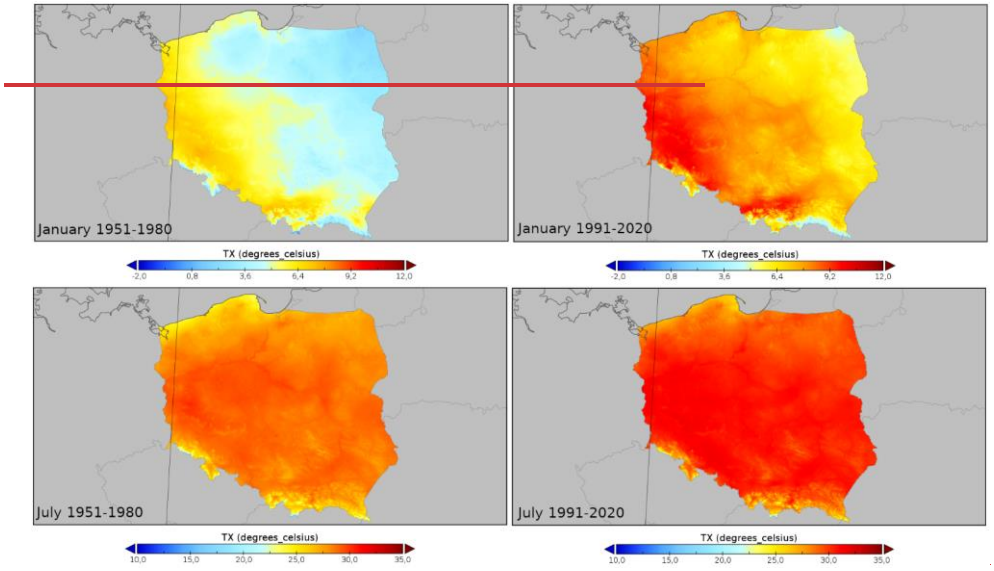


Figure 6. Spatial variability of 90% quantile of TX in Poland for selected 30-year periods (1951-1980 left panel, 1991-2020 right panel) in January (upper panel) and July (lower panel).

Figure 7 presents a similar analysis, but in this case, it is for the minimum temperature (TN) probabilistic characteristics. Again, the comparison between selected 30-year periods shows a substantial change in the derived values, concordant with the observed trends in the TN values in Central Europe (Ustrnul et al., 2021). For January, the change in the range of recorded values is 1.3C (an increase from −21.1°C to −19.8°C) in the case of minimum values and 1.2°C (an increase from −8.1°C to −6.9°C) for maximum grid values. Also, in this case, the overall observed change is characteristic for nearly the entire country, except the Baltic Sea coast, where the positive change is less pronounced. In July, the overall change in the range of values is +2.2°C (for minimum) and +0.9°C (for maximum values).

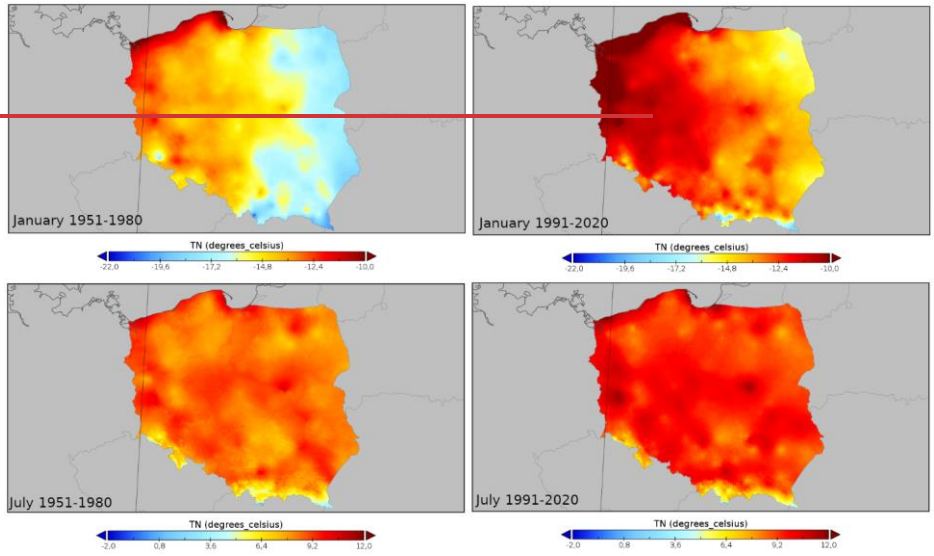


Figure 7. Spatial variability of 10% quantile of TN in Poland for selected 30-year periods (1951-1980 left panel, 1991-2020 right panel) in January (upper panel) and July (lower panel)

Figure 8 and Table 5 provide an additional example of the potential application of the data set. In this case, we present it for the whole country. Still, with the spatial resolution of ca.1km, it is equally applicable for the counties or even smaller areas desired by, for example, organisations preparing climate adaptation plans. The analysis presents the differences between the values of selected characteristics (in this case, extremes: 90% quantile of TX and 10% quantile of TN) for selected 30-year

periods: 2020-1991 and 1980-1951. The map indicates the areas where the observed climate change (expressed as the difference in temperature's extreme characteristics) was most pronounced and also allows further insight into the spatial variability of chosen metrics.

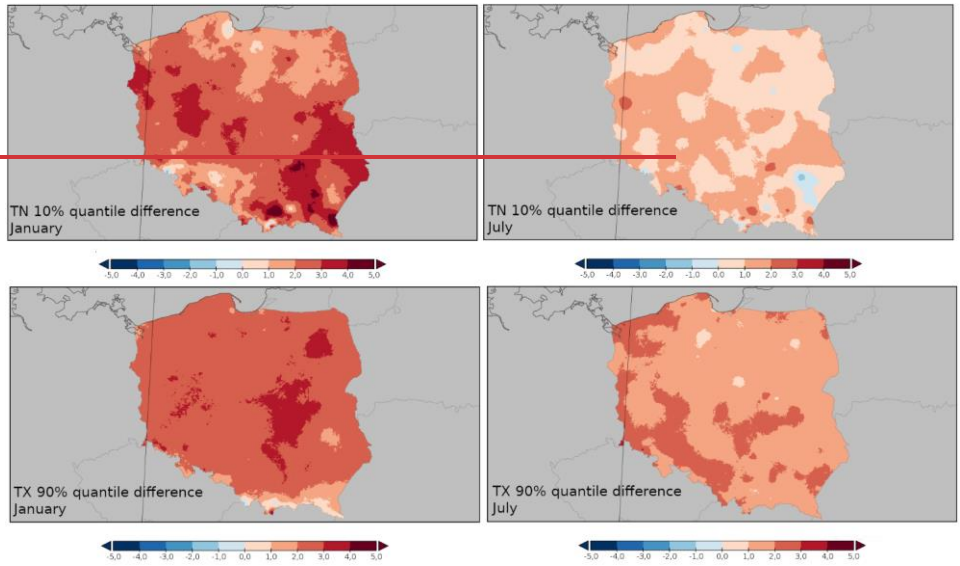


Figure 8. January (left) and July (right) differences (°C) between the values of TN 10% quantile (upper panel) and TX 90% quantile (lower panel) between periods 1991-2020 and 1980- 1951

The further analysis of such fields provides additional context. Table 5 shows the precise information on the area average change in the values of the abovementioned characteristics. It not only refers to the averages but, thanks to the calculation of the percentiles, provides information on the distribution of the shift magnitude in a given area. In this case, providing the percentiles' values cutting off the extreme 10% per cent of grid point values. For example, it shows that the area average change for TX and TN quantiles (10% and 90%, respectively) average was relatively similar in January (ca. +2.5°C) and much lower in July; however, in July, this change was more pronounced for TX (+1,75°C) than TN (+0.96°C) extremes.

In the case of the TX 90% quantile, 10% of the analysed area experienced a shift of over +3.00 °C in January. For June, this value is lower and equals +2.15°C. For the TN 10% quantile, 10% of the area had a shift more significant than +3.28°C in January, whereas in July, it was much lower (+1.44°C).

~~The lower quantile (10%) of the TX 90% and TN 10% quantiles, showing the threshold for the area experiencing less or equal values, also indicates a more pronounced shift magnitude in January than in July. It is worth noting that in the case of TX, 90% quantile in January, only 10% of the country experienced a change lower than +2.13°C (1.31°C in July). In the case of TN 10% quantile, this is only +0.46°C in July and +1.65°C in January.~~

~~**Table 5.** Essential field characteristics (mean, quantiles: 0.1, 0.5 and 0.9) of TX90% and TN10% quantile differences (°C) between 1951-1980 and 1991-2020.~~

~~Variable~~	~~Month~~	~~Mean~~	~~Q10~~	~~Q50~~	~~Q90~~
~~TX 90%~~	~~Jan~~	~~2.58~~	~~2.13~~	~~2.66~~	~~3.00~~
	~~Jul~~	~~1.75~~	~~1.31~~	~~1.76~~	~~2.15~~
~~TN 10%~~	~~Jan~~	~~2.45~~	~~1.65~~	~~2.45~~	~~3.28~~
	~~Jul~~	~~0.96~~	~~0.46~~	~~0.97~~	~~1.44~~

4 Example ~~application~~ applications of the datasets

4.1 Trend analysis

Sformatowano: Nagłówek 2

One of the most important features required from grid data, which ~~are~~ is to be used in climatological analysis and subsequently in the application of climate adaptation schemes, is the ability to represent the trends of analysed variables. Figure 5~~Figure 9~~ presents the inter-annual course of trend coefficients for TG~~AVE~~, TX and TN at the nodes nearest to selected stations. The locations represent the low-lying stations (Hel, Łódź-Lublinek, Wrocław) and stations from mountainous areas: Zakopane (~900 m a.s.l.) and Kasprowy Wierch (1984 m a.s.l.). There is striking complete coherence of trend coefficients for stations outside the mountainous areas, where the values are virtually equal when comparing observational (OBS) and interpolated (RBF) data. More significant discrepancies occur in the mountainous regions, but the differences are within the 0.05°C/10 years range. In the case of Zakopane, there is a systematic difference in the case of minimum temperature (TN) with slightly higher trend coefficient values for observational data. This situation, but to a much lower extent, is also visible in mean temperature (TG~~AVE~~). In the case of Kasprowy Wierch (a station which is a mountain observatory and located in the terrain of highly variable relief), the differences in trend coefficients present themselves in the case of all analysed variables. For TX, there is a bias with higher trend values for interpolated values, most visible in cold months (with the difference not exceeding 0.05°C/10 years, however). It is less evident for TG~~AVE~~, where greater bias is recorded for winter/cold months (October – March), with much more minor differences for the warm half of the year. Minimum temperature exhibits a reversing pattern with higher values of observed trends in the warmer part of the year and reversed bias (RBF > OBS) during cold months (October-March). It should be noted that the significance of trends based on interpolated data follows the results for observed trends. Exceptionally, non-significant observable TX trends in January and February appear significant for interpolated data

in Kasprowy Wierch. Conversely, significant observable TN trends in November and December are non-significant in Zakopane based on interpolated data. The above examples confirm the possibility of using the RBF gridded data to analyse multiannual air temperature variability in Poland, including even the areas of high relief variability.

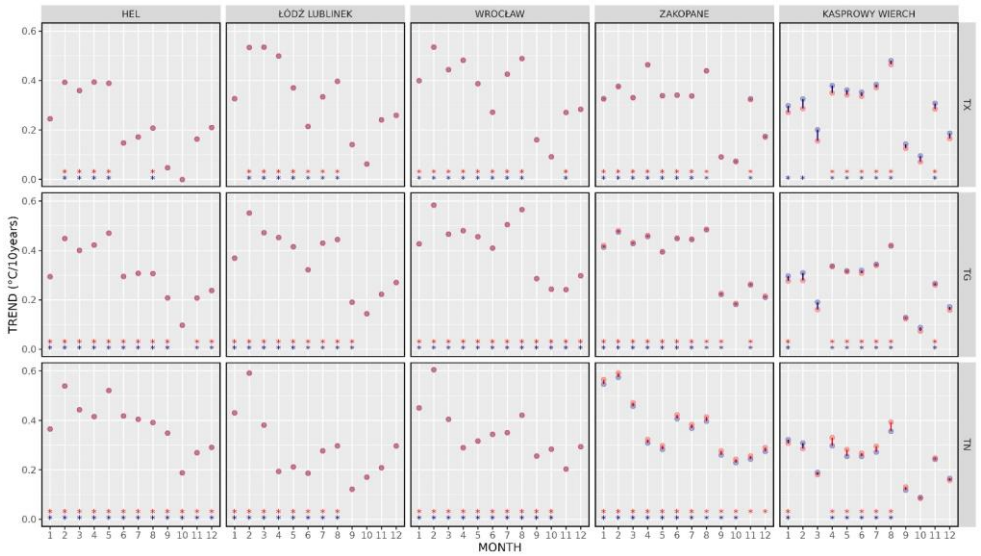


Figure 5.~~Figure 9.~~ Monthly trend coefficients (°C/10 years) at selected stations 1951-2020. Red dots – OBS, blue dots – RBF. The asterisks in the corresponding colours indicate significant trends.

4.2 Extreme analysis

The daily temporal resolution of the dataset allows the possibility of deriving multiple statistical characteristics with very high spatial resolution and detailed analysis of the thermal characteristics on the regional or even local scale (counties, small cities, or villages). For example, Figure 6 provides insight into the spatial variability of the TX 90th quantile in Poland for two 30-year periods: 1951-1980 and 1991-2020 in January and July. As expected in January, there is a substantial shift in the values of TX. The overall range of recorded values changes from –1.4°C to 7.7°C (1951-1980) and from 1.1°C to 10.6°C (1991-2020), which indicates a 2.5°C shift in the lower boundary and 2.9°C in the upper limit of the spatial distribution, indicating higher variability in the maximum values of TX. The overall spatial pattern of the TX 90th quantile remains relatively intact. However, the shift is visible in the whole country. In July, the change in the range of TX 90th quantile also reaches 2°C and is

Sformatowano: Razem z następnym

— sformatowano: Kolor czcionki: Automatyczny, Angielski (Stany Zjednoczone)

Sformatowano: Nagłówek 2

from 13.6°C to 30.8°C in case of first sub-period and from 15.6°C to 32.9°C for the second one. This is also clearly visible for nearly the whole country but less pronounced in the mountains and the vicinity of the Baltic Sea.

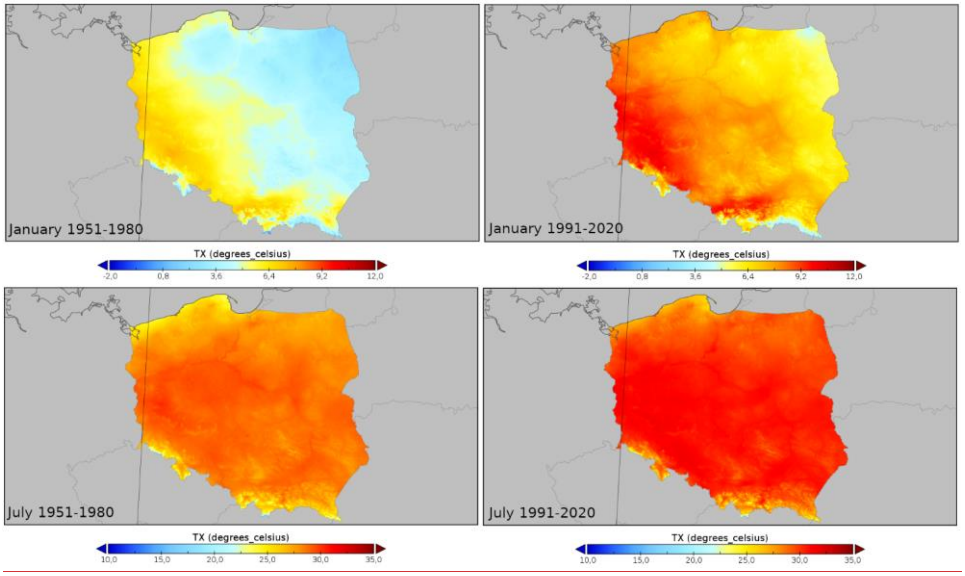


Figure 6. Spatial variability of the 90th quantile of TX in Poland for selected 30-year periods (1951-1980 left panel, 1991-2020 right panel) in January (upper panel) and July (lower panel).

Figure 7 presents a similar analysis, but in this case, it is for the minimum temperature (TN) probabilistic characteristics. Again, the comparison between selected 30-year periods shows a substantial change in the derived values, concordant with the observed trends in the TN values in Central Europe (Ustrnul et al., 2021). For January, the change in the range of recorded values is 1.3°C (an increase from –21.1°C to –19.8°C) in the case of the lower limit and 1.2°C (an increase from –8.1°C to –6.9°C) for the upper limit. Also, in this case, the overall observed change is characteristic for nearly the entire country, except the Baltic Sea coast, where the positive change is less pronounced. In July, the overall change in the range of values is +2.2°C (for minimum) and +0.9°C (for maximum values).

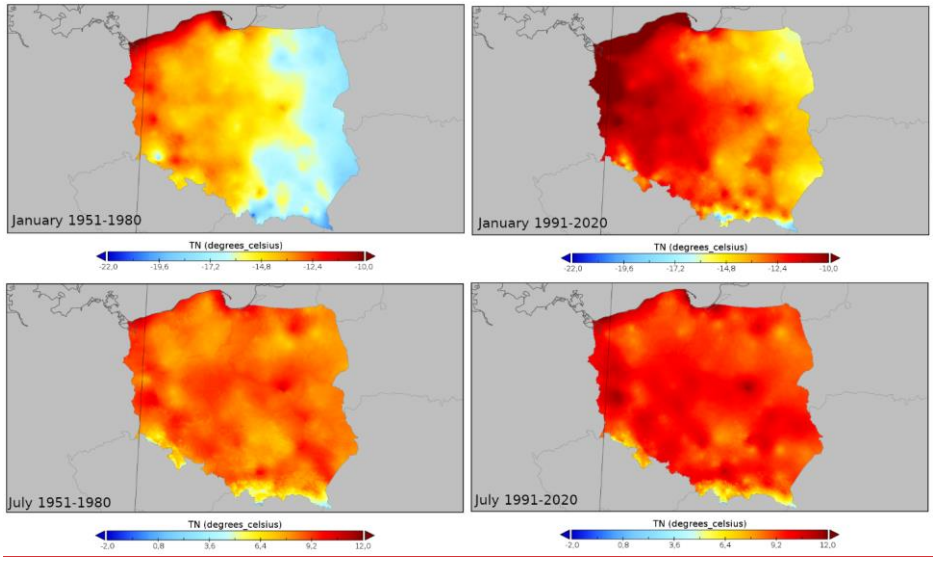


Figure 7. Spatial variability of the 10th quantile of TN in Poland for selected 30-year periods (1951-1980 left panel, 1991-2020 right panel) in January (upper panel) and July (lower panel).

Figure 8 and Table 6 provide an additional example of the potential application of the data set. In this case, we present it for the whole country. Still, with the spatial resolution of 1 km^2, it is equally applicable for the counties or even smaller areas desired by, for example, organisations preparing climate adaptation plans. The analysis presents the differences between the values of selected characteristics (in this case, extremes: 90th quantile of TX and 10th quantile of TN) for selected 30-year periods: 2020-1991 and 1980-1951. The map indicates the areas where the observed climate change (expressed as the difference in temperature's extreme characteristics) was most pronounced and also allows further insight into the spatial variability of chosen metrics.

— **sformatowano:** Indeks górny

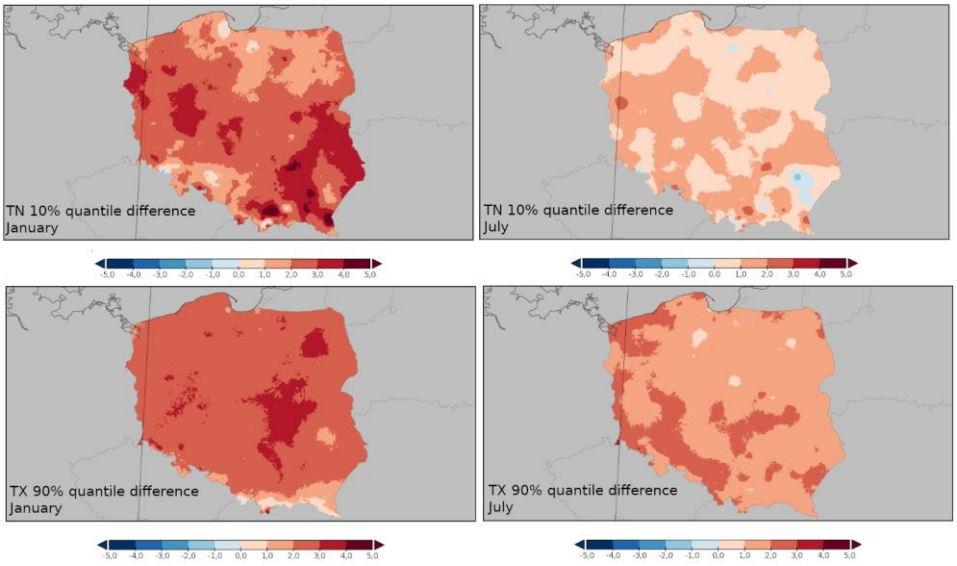


Figure 8. January (left) and July (right) differences (°C) between the values of TN 10th quantile (upper panel) and TX 90th quantile (lower panel) between periods 1991-2020 and 1980- 1951

The further analysis of such fields provides additional context. Table 6 shows the precise information on the area average change in the values of the abovementioned characteristics, extended by comparison to the E-OBS dataset. The analysis not only refers to the averages but, thanks to the calculation of the percentiles, provides information on the distribution of the shift magnitude in a given area. In this case, providing the percentiles' values cutting off the extreme 10% per cent of grid point values. In the case of PL1GD-T dataset, it shows that the area average change for TX and TN quantiles (10th and 90th, respectively) average was relatively similar in January (ca. +2.5°C) and much lower in July; however, in July, this change was more pronounced for TX (+1,75°C) than TN (+0.96°C) extremes. In the case of the TX 90th quantile, 10% of the analysed area experienced a shift of over +3.00 °C in January. For June, this value is lower and equals +2.15°C. For the TN 10th quantile, 10% of the area had a shift more significant than +3.28°C in January, whereas in July, it was much lower (+1.44°C). The lower quantile (10th) of the TX 90th and TN 10th quantiles, showing the threshold for the area experiencing less or equal values, also indicates a more pronounced shift magnitude in January than in July. It is worth noting that in the case of TX, the 90th quantile in January, only 10% of the country experienced a change lower than +2.13°C (1.31°C in July). In the case of TN 10th quantile, this is only +0.46°C in July and +1.65°C in January.

Table 6. Essential field characteristics (mean, quantiles: 0.1, 0.5 and 0.9) of TX 90th and TN quantile differences (°C) averaged over the country area between 1951-1980 and 1991-2020.

Variable	Month	Dataset	Mean	Q10	Q50	Q90
TX 90th	Jan	PL1GD-T	2.58	2.13	2.66	3.00
		E-OBS	2.58	2.18	2.63	2.98
	Jul	PL1GD-T	1.75	1.31	1.76	2.15
		E-OBS	1.75	1.30	1.75	2.17
TN 10th	Jan	PL1GD-T	2.45	1.65	2.45	3.28
		E-OBS	2.46	1.64	2.46	3.31
	Jul	PL1GD-T	0.96	0.46	0.97	1.44
		E-OBS	0.97	0.51	0.97	1.41

Sformatowano: Wyjustowany

5 Summary and outlook

The paper presents a comprehensive study on developing a high-resolution gridded dataset of daily air temperatures at 2 m above the ground (minimum, mean, and maximum) over Poland from 1951 to 2020, with a spatial resolution of 1 km². The dataset was created using quality-controlled observations from ground weather stations. Hold-out cross-validation was chosen to select appropriate radial basis functions (RBFs) for interpolation. Leave-one-out cross-validation assessed product quality on a seasonal, monthly, and station basis. The findings indicate the best performance for mean temperature in winter and spring, with high correlation and low root mean square deviation. Although the performance for maximum temperature is slightly below average, it still shows relatively good agreement, especially in autumn. The method is less successful at accurately capturing minimum temperature, with the best results seen in winter. The analysis also confirms that the standard deviation of temperatures, representing variability, is consistent between observed and interpolated data. All variables have a bias that does not exceed 0.03, with minimal interannual variability. Pearson's correlation coefficient is very high, ranging from 0.94 to 0.98. The difference between the 5th and 95th percentiles suggests slight underestimation for Q95 and overestimation for Q05. There is a noticeable increase in seasonal RMSD variability with altitude, with relatively small interannual differences at a lower altitude ~~to,~~ and the greatest in the case of the highest altitude class. Mean values are preserved, and the interpolated data's standard deviation is slightly lower. Finally, we have ~~will~~ shown an example of the application of the resulting gridded product in the field of climate change.

The validation results demonstrate the dataset's accuracy and suitability for climatological applications. However, some limitations and potential areas for improvement have been identified. Notably, while computationally efficient, the RBF interpolation method may smooth out extremes and underestimate spatial variability, particularly in areas with complex terrain.

Future work could explore refining the RBF implementation to mitigate this, such as by: investigating adaptive shape parameters that vary spatially based on station density or terrain complexity, employing anisotropic RBFs to account for directional dependencies in temperature variability or using an ensemble of RBF interpolations with varying parameters to capture uncertainty in the interpolation. Furthermore, while beyond the scope of this study, future releases will include a measure of uncertainty for all estimates, enhancing the dataset's utility for users. In addition, future work may consist of homogenising the input station data. The availability of detailed station metadata (https://klimat.imgw.pl/pl/meta-dane/) will be crucial for applying appropriate homogenisation techniques. Finally, given the dataset's demonstrated value, we envision transitioning to a quasi-routine production system and incorporating other surface meteorological variables, enabling regular updates and wider accessibility for the climate research community and related applications.

This open-access dataset is crucial for climate change impact studies on a smaller scale and can serve a wide range of users, including researchers, administrative bodies, and society. One important application of such a dataset is serving as reference data for bias correction of regional dynamical downscaling results (e.g., EURO-CORDEX initiative) to elaborate effective adaptation and mitigation strategies.

6 Data availability

PL1GD-T - gridded dataset of the mean, minimum and maximum daily air temperature at the level of 2 m for the area of Poland at a resolution of 1 km × 1 km – are publicly available in the public repository of the Institute of Meteorology and Water Management – National Research Institute under the https://doi.org/10.26491/imgw_repo/PL1GD-T (Jaczewski et al., 2024). The data are archived as NetCDF files in 10-year chunks for every variable separately. The resulting datasets consist of fields with values rounded to one decimal, with data values encoded as integers with a scale factor of 0.1. It allowed the files' size to be decreased by a factor of two. The naming convention is *variable_startyear_endyear.nc* with *variable* prefixes 'tn', 'tg', and 'tx' for minimum, mean and maximum temperature, respectively. The NetCDF files comply with the CF-1.6 convention.

Author contribution. AJ designed the study with conceptual support from MMi and MMa. AJ developed the model code and prepared the dataset for the repository. AJ and MMa performed calculations for model validation. MMa prepared spatial maps. AJ and MMa prepared the manuscript. MMi contributed with critical review and commentary.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The study was financially supported by the Institute of Meteorology and Water Management – National Research Institute under the project FBW10/2023, "Implementation of advanced monitoring and analysis of contemporary climate variability in Poland in the context of observed climate change on a regional scale". Computations were supported using the computers of Centre of Informatics Tricity Academic Supercomputer & Network.

References

DTM 100m: https://dane.gov.pl/en/dataset/792/resource/30643,numeryczny-model-terenu-o-interwale-siatki-co-najmniej-100m-usuga-atom/table, last access: 24 May 2023.

Spatial Climate Analyses - MeteoSwiss: https://www.meteoswiss.admin.ch/climate/the-climate-of-switzerland/spatial-climate-analyses.html, last access: 27 March 2025.

Antal, A., Guerreiro, P. M. P., and Cheval, S.: Comparison of spatial interpolation methods for estimating the precipitation distribution in Portugal, Theor. Appl. Climatol., 145, 1193–1206, https://doi.org/10.1007/s00704-021-03675-0, 2021.

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.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., and Jones, P. D.: An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets, J. Geophys. Res. Atmospheres, 123, 9391–9409, https://doi.org/10.1029/2017JD028200, 2018.

Delvaux, C., Journée, M., and Bertrand, C.: The FORBIO Climate data set for climate analyses, Adv. Sci. Res., 12, 103–109, https://doi.org/10.5194/asr-12-103-2015, 2015.

Fasshauer, G. E. and Zhang, J. G.: On choosing "optimal" shape parameters for RBF approximation, Numer. Algorithms, 45, 345–368, https://doi.org/10.1007/s11075-007-9072-8, 2007.

Fathizad, H., Mobin, M. H., Gholamnia, A., and Sodaiezadeh, H.: Modeling and mapping of solar radiation using geostatistical analysis methods in Iran, Arab. J. Geosci., 10, 391, https://doi.org/10.1007/s12517-017-3130-x, 2017.

Frei, C.: Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances, Int. J. Climatol., 34, 1585–1605, https://doi.org/10.1002/joc.3786, 2014.

Gao, Y., Guo, J., Wang, J., and Lv, X.: Assessment of Three-Dimensional Interpolation Method in Hydrologic Analysis in the East China Sea, J. Mar. Sci. Eng., 10, 877, https://doi.org/10.3390/jmse10070877, 2022.

Hiebl, J. and Frei, C.: Daily temperature grids for Austria since 1961—concept, creation and applicability, Theor. Appl. Climatol., 124, 161–178, https://doi.org/10.1007/s00704-015-1411-4, 2016.

Jaczewski, A., Marosz, M., and Miętus, M.: PL1GD-T - gridded data of the mean, minimum and maximum daily air temperature (2 m) for the Polish area at a resolution of 1 km x 1 km and the period 1951-2020, https://doi.org/10.26491/imgw_repo/PL1GD-T, 2024.

Killick, R. E., Jolliffe, I. T., and Willett, K. M.: Benchmarking the performance of homogenization algorithms on synthetic daily temperature data, Int. J. Climatol., 42, 3968–3986, https://doi.org/10.1002/joc.7462, 2022.

Krähenmann, S., Bissolli, P., Rapp, J., and Ahrens, B.: Spatial gridding of daily maximum and minimum temperatures in Europe, Meteorol. Atmospheric Phys., 114, 151, https://doi.org/10.1007/s00703-011-0160-x, 2011.

Krähenmann, S., Walter, A., Brienen, S., Imbery, F., and Matzarakis, A.: High-resolution grids of hourly meteorological variables for Germany, Theor. Appl. Climatol., 131, 899–926, https://doi.org/10.1007/s00704-016-2003-7, 2018.

Lussana, C., Tveito, O. E., and Uboldi, F.: Three-dimensional spatial interpolation of 2 m temperature over Norway, Q. J. R. Meteorol. Soc., 144, 344–364, https://doi.org/10.1002/qj.3208, 2018.

Lussana, C., Tveito, O. E., Dobler, A., and Tunheim, K.: seNorge_2018, daily precipitation, and temperature datasets over Norway, Earth Syst. Sci. Data, 11, 1531–1551, https://doi.org/10.5194/essd-11-1531-2019, 2019.

Ntegeka, V., Salamon, P., Gomes, G., Sint, H., Lorini, V., Thielen, J., and Zambrano-Bigiarini, M.: EFAS-Meteo : a European daily high-resolution gridded meteorological data set for 1990-2011, Publications Office of the European Union, https://doi.org/doi/10.2788/51262, 2014.

Piniewski, M., Szcześniak, M., Kardel, I., Chattopadhyay, S., and Berezowski, T.: G2DC-PL+: a gridded 2 km daily climate dataset for the union of the Polish territory and the Vistula and Odra basins, Earth Syst. Sci. Data, 13, 1273–1288, https://doi.org/10.5194/essd-13-1273-2021, 2021.

Piri, I.: DETERMINATION OF THE BEST GEOSTATISTICAL METHOD FOR CLIMATIC ZONING IN IRAN, Appl. Ecol. Environ. Res., 15, 93–103, https://doi.org/10.15666/aeer/1501_093103, 2017.

Razafimaharo, C., Krähenmann, S., Höpp, S., Rauthe, M., and Deutschländer, T.: New high-resolution gridded dataset of daily mean, minimum, and maximum temperature and relative humidity for Central Europe (HYRAS), Theor. Appl. Climatol., 142, 1531–1553, https://doi.org/10.1007/s00704-020-03388-w, 2020.

Rocha, H. and Dias, J.: Honey Yield Forecast Using Radial Basis Functions, in: Machine Learning, Optimization, and Big Data, Cham, 483–495, https://doi.org/10.1007/978-3-319-72926-8_40, 2018.

Rocha, H. and Dias, J. M.: Early prediction of durum wheat yield in Spain using radial basis functions interpolation models based on agroclimatic data, Comput. Electron. Agric., 157, 427–435, https://doi.org/10.1016/j.compag.2019.01.018, 2019.

Ryu, S., Song, J. J., and Lee, G.: Interpolation of Temperature in a Mountainous Region Using Heterogeneous Observation Networks, Atmosphere, 15, 1018, https://doi.org/10.3390/atmos15081018, 2024.

Saaban, A., Hashim, N. M., and Murat, R. I. Z.: Estimating monthly temperature using point based interpolation techniques, AIP Conf. Proc., 1522, 783, https://doi.org/10.1063/1.4801206, 2013.

Serrano-Notivoli, R., Beguería, S., and de Luis, M.: STEAD: a high-resolution daily gridded temperature dataset for Spain, Earth Syst. Sci. Data, 11, 1171–1188, https://doi.org/10.5194/essd-11-1171-2019, 2019.

Škrk, N., Serrano-Notivoli, R., Čufar, K., Merela, M., Črepinšek, Z., Kajfež Bogataj, L., and de Luis, M.: SLOCLIM: a high-resolution daily gridded precipitation and temperature dataset for Slovenia, Earth Syst. Sci. Data, 13, 3577–3592, https://doi.org/10.5194/essd-13-3577-2021, 2021.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res. Atmospheres, 106, 7183–7192, https://doi.org/10.1029/2000jd900719, 2001.

Thiemig, V., Gomes, G. N., Skøien, J. O., Ziese, M., Rauthe-Schöch, A., Rustemeier, E., Rehfeldt, K., Walawender, J. P., Kolbe, C., Pichon, D., Schweim, C., and Salamon, P.: EMO-5: A high-resolution multi-variable gridded meteorological data set for Europe, Earth Syst. Sci. Data Discuss., 1–35, https://doi.org/10.5194/essd-2021-339, 2021.

Ustrnul, Z., Wypych, A., and Czekierda, D.: Air Temperature Change, in: Climate Change in Poland: Past, Present, Future, edited by: Falarz, M., Springer International Publishing, Cham, 275–330, https://doi.org/10.1007/978-3-030-70328-8_11, 2021.

Wang, S., Huang, G. H., Lin, Q. G., Li, Z., Zhang, H., and Fan, Y. R.: Comparison of interpolation methods for estimating spatial distribution of precipitation in Ontario, Canada, Int. J. Climatol., 34, 3745–3751, https://doi.org/10.1002/joc.3941, 2014.

World Meteorological Organization: Guide on the Global Data-processing System, Secretariat of the World Meteorological Organization, 214 pp., 1993.

Wypych, A. and Ustrnul, Z.: Spatial differentiation of the climatic water balance in Poland, IDOJARAS, 115, 111–120, 2011.

Yang, R. and Xing, B.: A Comparison of the Performance of Different Interpolation Methods in Replicating Rainfall Magnitudes under Different Climatic Conditions in Chongqing Province (China), Atmosphere, 12, 1318, https://doi.org/10.3390/atmos12101318, 2021.