

PAPER • OPEN ACCESS

Selection of Optimal Gridded Dataset for Application in Polish Sudetes Mountains

To cite this article: Monika Chuchro and Malgorzata Danek 2019 *IOP Conf. Ser.: Earth Environ. Sci.*
221 012120

View the [article online](#) for updates and enhancements.

You may also like

- [Calibration of accelerometer and magnetometer sensors of MWD systems in directional drilling application](#)

A Malekizadeh, H Afarideh, M
Mohamadian et al.

- [Magnetism in Isolated and Binary White Dwarfs](#)

D. T. Wickramasinghe and Lilia Ferrario

- [The development of and experiments on electromagnetic measurement while a drilling system is used for deep exploration](#)

Chunhua Lu, Guosheng Jiang, Ziqi Wang
et al.



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Selection of Optimal Gridded Dataset for Application in Polish Sudetes Mountains

Monika Chuchro¹, Malgorzata Danek²

¹ AGH University of Science and Technology, Department of Geoinformatics and Applied Computer Sciences, Mickiewicza 30 Av., 30-059 Krakow, Poland

² AGH University of Science and Technology, Department of Environmental Analysis, Mapping and Economic Geology, Mickiewicza 30 Av., 30-059 Krakow, Poland

chuchro@agh.edu.pl

Abstract. High quality measured weather data (MWD) are limited or not available for many areas. Also their time coverage can be relatively short. That is why the use of gridded climatic data (GCD) in environmental studies is very popular. However, GCD are valuable source of information, their accuracy can be sometimes insufficient for particular study. That is why GCD applicability should be checked before the study run. The objective of this study was to check the applicability of gridded data for dendroclimatological studies of larch from Sudetes Mountains. To do so we compared GCD with available MWD from several weather stations located in the mentioned area. Because many gridded time-series datasets are available, we also wanted to check which dataset is the best to use for the mentioned area. In the analysis high-resolution gridded data on monthly mean temperature and total precipitation, which cover the common period 1901-2013, in $0.5^\circ \times 0.5^\circ$ network, created by:

- Center for Climatic Research Department of Geography University of Delaware Newark (UD; model V4.01 for precipitation and temperature data),
- Global Precipitation Climatology Centre Deutscher Wetterdienst (GPCC; model V7 for precipitation data),
- Climate Research Unit, University of East Anglia (CRU; model CRU TS v.4.01 for precipitation and temperature data)

were used. The available MWD data from several weather stations started in 1951, 1956 and 1957. The common period for the analysis covered years 1951('56 or '57) – 2013.

For a given precipitation and temperature data the agreement and biases between GCDs and MWDs were assessed with the absolute mean error (ME) and root mean square error (RMSPE), L1- norm and Pearson correlation coefficient. Finally, linear regression analysis was performed to detect biases in the relationship between GCD and MWD, and the coefficient of determination (R^2) was also calculated. GCD for precipitation show high similarity to MWD. Mean Pearson correlation coefficient values equal to 0.87 for GPCC, 0.82 for CRU and 0.8 for UD GCDs. For temperature the received values of Pearson correlation were relatively high and very similar, e.g. 0.97 for CRU and 0.96 (UD). L1-norm, ME, RMSPE and regression model confirmed small differences between analysed GCDs, but with better fitting of CRU GCD.

1. Introduction

High quality measured weather data (MWD) are limited or not available for many areas. Also, their time span can be relatively short. It can limit their applicability in environmental studies. An example



can be dendroclimatological studies, where the long data records can be crucial for the analysis of tree growth response to climate. That is why the use of gridded climatic data (GCD) is very popular. Many sources of GCD exist. Variety of variables can be found as daily, monthly and annually summaries in many versions of spatial resolution, from 10' to 5°. However, GCD are valuable source of information, their accuracy can be sometimes insufficient for particular study. That is why GCD applicability should be checked before the study run.

The studies that analyze and assess the quality of GCD data by comparison with relevant MWD from weather stations are relatively rare [1-4]. Most of them were performed for the United States [4-5], some exist for Africa and Asia [1], Australia [6] and Europe [7]. The quality of gridded data for dendroclimatological studies are also checked [8-10].

However in Poland, several dozens of weather stations exist, still their coverage is limited. Also, usually the available data records are relatively short (they start in 1950's or in next decades). That is why the use of gridded data in conducted dendroclimatological studies of larch in Sudetes would be very convenient. The aim of this study was to check the applicability of gridded data for their use. Because many gridded time-series datasets are available, we also wanted to check which of them is the most suitable for the mentioned area. The choice was made on the basis of the smaller differences between GCD box and corresponding weather station data. In the analysis high-resolution gridded data on monthly mean temperature and total precipitation with the most popular 0.5° x 0.5° network were analyzed.

2. Material and Methods

The process was carried out using, on the one hand MWD, on the other hand GCD data from the three different models. A total of 14 MWD on total monthly precipitation and 5 MWD on average monthly temperature were compared with suitable GCD boxes data. Corresponding pairs of time series were compared in term of their similarity, based on statistic and data mining methods.



Figure 1. Weather stations locations (blue dots). Closed dots – temperature and precipitation records, open dots – only precipitation records

2.1. Meteorological data from weather stations.

Data on average monthly temperature and precipitation totals from fourteen weather stations located in the Polish Sudetes Mountains region were used (figure 1, table 1). Data records for twelve of them are available since 1951, for two stations MWD are several years shorter (for Szczawno Zdrój – since 1956 and Twardocice – since 1957). Only five of analysed weather stations carry out the

measurements of temperature and precipitation (figure 1 and table 1). The elevation of the stations ranges from 242.9 m a.s.l. (Twardocice) to 1602 m a.s.l. (Śnieżka), the latitudes – from 50.41 to more than 51.10 degrees north, and the longitudes – from 15.34 to 16.66 degrees east.

Table 1. Characteristics and basic information about weather stations

Localization	Measurements	Longitude E	Latitude N	Altitude [m a.s.l.]	Measurements start
Kłodzko	precip. temp.	16.66	50.43	292.8	1951
Polanica Zdrój	precip.	16.51	50.41	361.9	1951
Świeradów Zdrój	precip.	15.34	50.91	473.2	1951
Rębiszów	precip.	15.45	50.95	423.2	1951
Jelenia Góra	precip. temp.	15.72	50.90	352.8	1951
Przesieka	precip.	15.67	50.81	527.2	1951
Szklarska Poręba	precip.	15.53	50.83	629.2	1951
Śnieżka	precip. temp.	15.74	50.74	1602.0	1951
Karpacz	precip. temp.	15.76	50.78	607.9	1951
Walim	precip.	16.44	50.70	514.2	1951
Mieroszów	precip.	16.19	50.67	488.6	1951
Szczawno Zdrój	precip. temp.	16.25	50.81	422.4	1956
Gryfów Śląski	precip.	15.41	51.04	341.6	1951
Twardocice	precip.	15.78	51.10	242.9	1957

Mentioned above weather stations (table 1) are operated by the Institute of Meteorology and Water Management – National Research Institute (in Polish: Instytut Meteorologii i Gospodarki Wodnej – Państwowy Instytut Badawczy, hereafter: IMGW-PIB) [11].

2.2. Gridded data

Globally gridded precipitation and temperature data are available at many databases for example: www.esrl.noaa.gov, www.cruclimdata.uea.ac.uk, www.climdex.org. All of them enable the access to several gridded datasets, which differ in spatial and time resolution. GCD are modelled base on weather stations historical data, both daily and monthly, and interpolation algorithms. In this paper three GCD datasets, which cover the period 1901-2013, in 0.5° x 0.5° network, publicly available were analyzed:

- UD v4.01 - Center for Climatic Research Department of Geography University of Delaware Newark (UD) created grid model using a large number of weather stations from GHN2 (Global Historical Climate Network) and the archive of Legates & Willmott. In this analyse the newest model V4.01, with temporal coverage from 1901/01 to 2014/12 for both precipitation and temperature data was used [12]. Gridded data are available at esrl.noaa.gov database [13].
- CRU TS v. 4.01 - Climate Research Unit, University of East Anglia (CRU) has calculated climatic model from monthly observations for land areas. The gridded dataset includes six climatic variables (mean temperature, diurnal temperature range, precipitation totals, wet-day frequency, vapour pressure and cloud cover). Model includes diagnostics associated with each interpolated value that indicates the number of stations used in the interpolation, allowing determination of the reliability of values in an objective way. In presented study the newest

model CRU TS v.4.01 with temporal coverage from 1901/01 to 2016/12 for both precipitation and temperature data was used [14-15]. Gridded data are available at crudata.uea.ac.uk webpage [16].

- GPCP v7 - Global Precipitation Climatology Centre Deutscher Wetterdienst (GPCP) has calculated a precipitation climatology model for the global land areas, using data from more than 67,000 rain gauge stations. Furthermore, a semi-automatic quality control processing and additional visual control was used for better quality of GCD models. In the analyses we used the newest model V7 with temporal coverage from 1901/01 to 2013/12 [17-18]. Gridded data are available in esrl.noaa.gov database [13].

2.3. Statistical analyses

The selection of the best gridded dataset for the area was made by using similarity and dissimilarity statistical assessment. Commonly used mean error (ME) and root mean square error (RMSE) were computed for the pairs of time series (MWD and GDC) for precipitation and temperature. Also dissimilarity measure (L1- norm and Pearson correlation coefficient) were calculated. Finally, linear regression analysis was performed to detect biases in the relationship between GCD and MWD, with the coefficient of determination (R^2) [8, 19-20]. The comparison of precipitation and temperature data between weather stations and gridded data from three datasets was carried out for a common period 1951-2013, separately for each time series pairs (MWD-GCD). There were two exceptions: Szczawno Zdrój and Twardocice, for which the possible periods of analysis were shorter (1956-2013 and 1957-2013 respectively).

3. Results

The results of analyses were presented in figures 2 and 3. The results obtained for precipitation show the biggest differences between three GCD datasets for two parameters: Pearson correlation coefficient and determination coefficient (R^2). Generally, GPC data showed better fit to MWD for particular weather stations than CRU and UD. The exception was Kłodzko, for which the best results were observed for UD data. Mean Pearson correlation coefficients for particular datasets were equal to 0.88 (GPCP), 0.82 (CRU) and 0.80 (UD). For GPCP data, six values of correlation coefficient exceeded 0.9 (very high positive correlation). Correlation values for CRU data were lower by value of 0.07 from GPCP's. The lowest value of correlation coefficient was observed for UD data for Jelenia Góra (0.67; figure 2a).

A high discrepancy in results was obtained for regression models. The values of a_0 and a_1 parameters differed significantly between datasets, as well as model quality of match, expressed by the determination coefficient R^2 . Again, the best match was demonstrated by the GPCP dataset (with the exception of Kłodzko). Mean R^2 equals to 0.79 for GPCP, 0.9 for CRU and 0.66 for UD dataset. Four of five the lowest values of R^2 were observed for UD (figure 2c). The low mean value of R^2 for CRU GCD was connected with poor regression model quality for Śnieżka (R^2 equal 0.54).

Dissimilarity measure L1-norm confirmed better fitting of GPCP dataset, with difference equal 4000 between analyzed gridded datasets. The lower value of L1-norm obtained for CRU data comparing to GPCP was caused by high misfit for two locations (Śnieżka, Kłodzko). Two from three very high L1-norm values were obtained for UD data, they exceed 28000 (Jelenia Góra, Przesieka), see figure 2b.

The mean error showed the cause of low fitting of UD data. ME for UD data had usually higher values than other GCDs (figure 2d), both with positive and negative signs. The biggest ME value was calculated for Kłodzko (47) for UD dataset. GPCP and CRU data had shown high correlation of ME results. For GPCP slightly lower values of ME were computed (figure 2d). The last measure – root mean square error (RMSE), confirmed problems with UD dataset. For almost all locations RMSE is lower than 5. Mean values are equal to: 0.39 (GPCP), 0.84 (CRU) and 4.45 (UD). The highest RMSE value was calculated for UD data (Mieroszów- 48). Similarly, as for the mentioned above metrics, UD

had higher values of RMSE. Elevated mean value of RMSE for CRU dataset was caused, similarly as for the rest of parameters, by one weather station – Śnieżka (figure 2e).

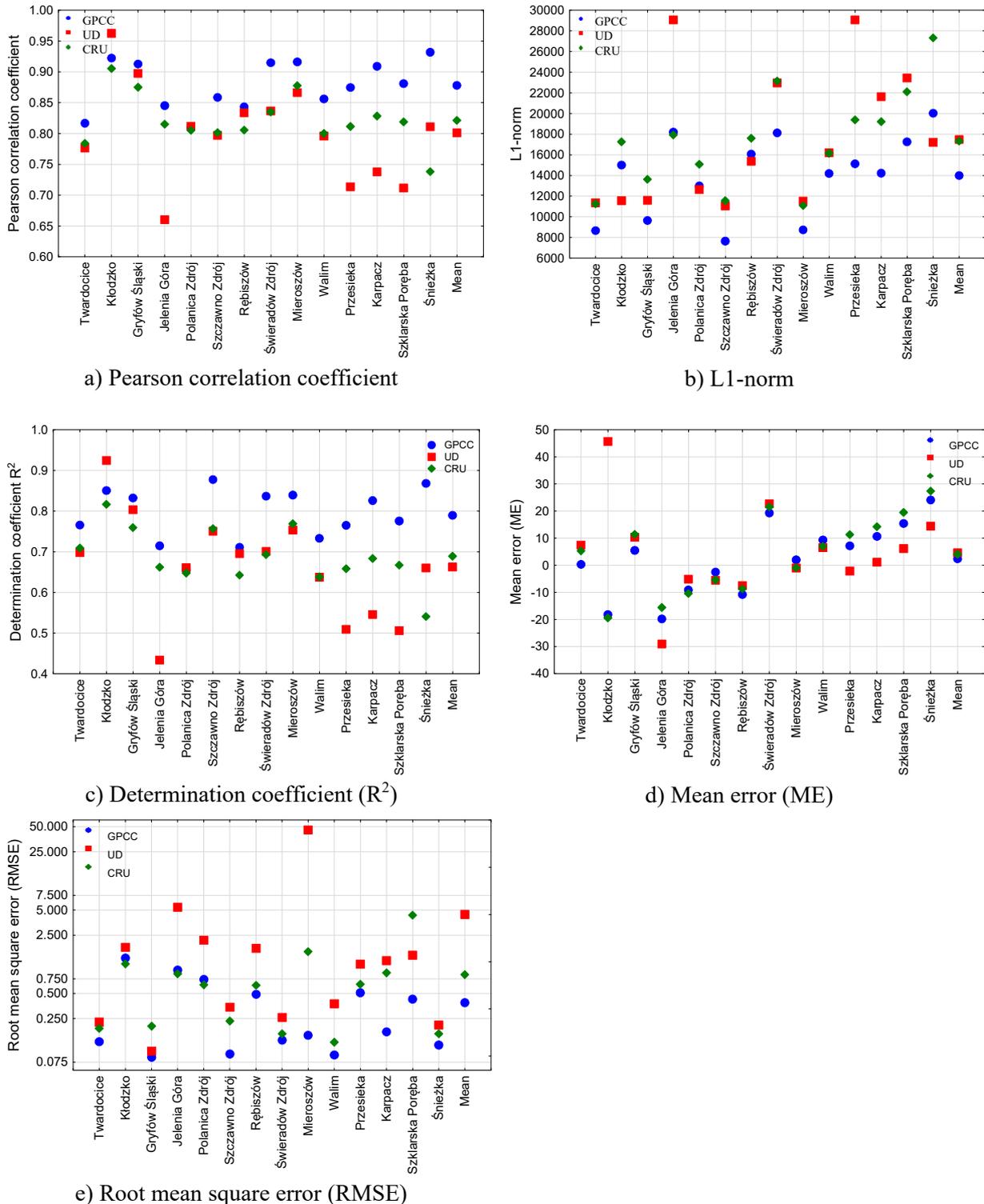
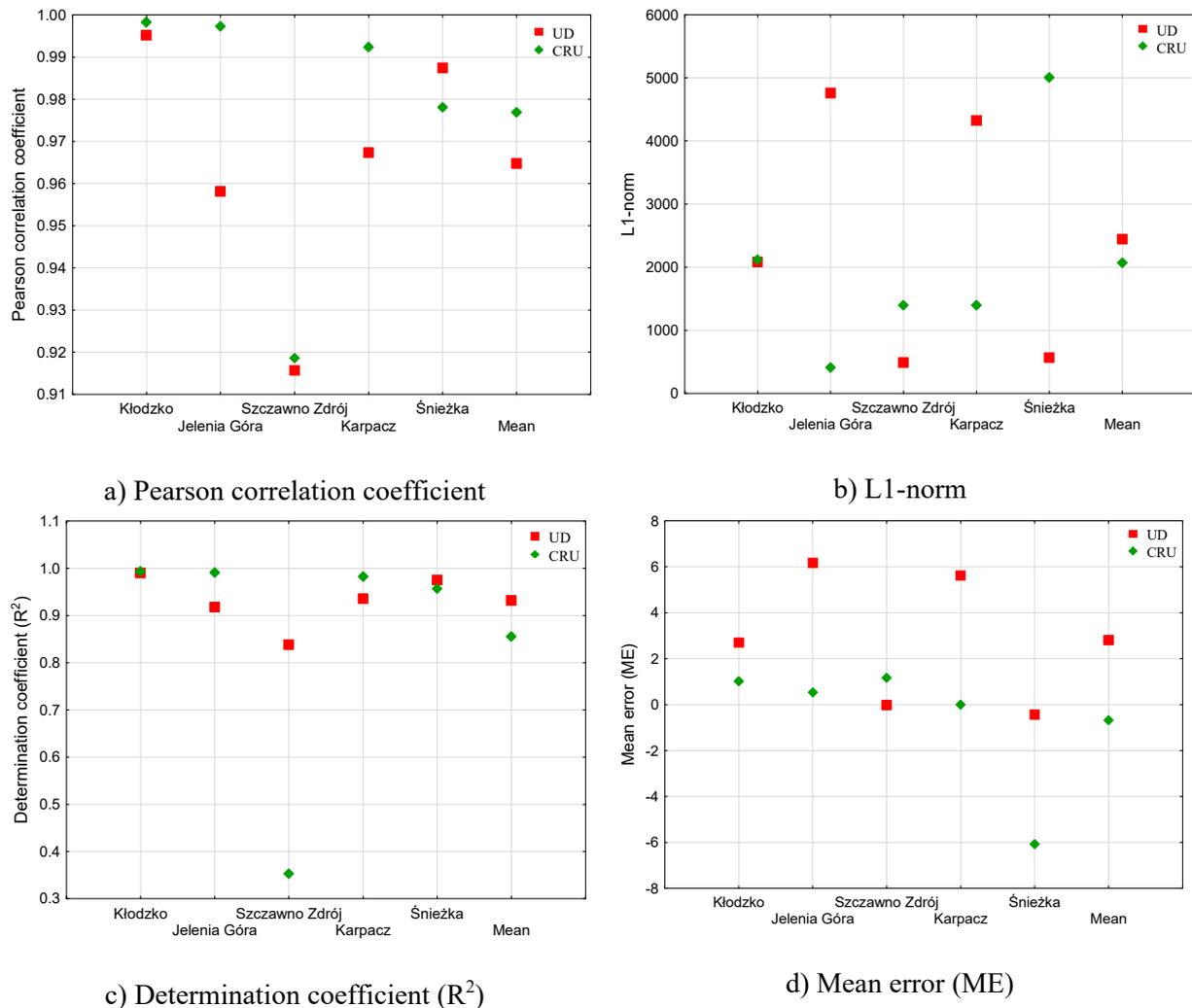
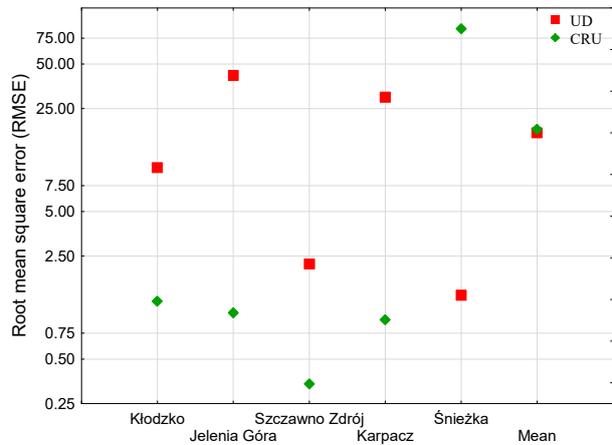


Figure 2. Similarity assessment results for precipitation data. The weather stations were ordered from the lowest (left) to highest (right) locations

Similarly, assessment for temperature show big disproportion between two analysed datasets (i.g. CRU and UD). Pearson correlation coefficient was higher for CRU than for UD (mean values equal to 0.98 and 0.96 respectively). Generally, the lowest value of correlation between MWD and GCD was received for Szczawno Zdrój (0.915), the highest – for Kłodzko (0.998). The biggest distance between correlations for one weather station was observed fore Jelenia Góra (0.04), between CRU and UD data (figure 3a). Values of L1-norm showed high similarity between analyzed data. For two weather stations (Szczawno Zdrój, Śnieżka) UD had smaller values than CRU, for Jelenia Góra and Karpacz CRU data showed a better fit (lower L1-norm values). The highest L1-norm value was received for Śnieżka for CRU data (5003; see figure 3b). Regression parameters were more stable for temperature in comparison with precipitation. Parameter $a1$ varied slightly from 1, and indicated good quality of both gridded datasets. Determination coefficient (R^2) confirmed high similarity of both GCD's, mean R^2 value equal: 0.93 for CRU and 0.85 for UD. The lowest value of R^2 was observed for Szczawno Zdrój (0.35) for CRU data (figure 3c). Mean error (ME) and root mean square error (RMSE) showed the same pattern. Usually CRU data were better fitted, with lower parameters values. However, for Szczawno Zdrój and Śnieżka ME and RMSE results were better for UD than for CRU data (figure 3d and 3e).





e) Root mean square error (RMSE)

Figure 3. Similarity assessment results for temperature data. The weather stations were ordered from the lowest (left) to highest (right) locations.

4. Conclusions

The obtained results for precipitation showed weaker similarity of GCD to WMD than for temperature, which can be explained by high spatial variability of this climatic factor. The weakest fit for precipitation was observed for UD dataset. CRU and GPCC dataset better represent real data (MWD). In general, for GPCC datasets better results were obtained (especially R^2 and Pearson correlation coefficient). But it has to be admitted, that lower values of calculated parameters obtained for CRU data were mainly caused by the results obtained for one weather station, with the highest altitude (Śnieżka). With the results for Śnieżka be excluded, the obtained results for CRU and GPCC are similar.

Obtained results for temperature show high quality of GCD for two analyzed datasets, however better results were obtained for CRU data. Only for one weather station clearly better fit was found for UD data (Śnieżka). Both models had some problems with prediction of some weather station data (Szczawno Zdrój, Śnieżka). UD GCD were underestimated, with ME equal -2.8. Lower ME, but overestimated, was calculated for CRU GCD.

In addition, no relation was found between altitude and similarity measures values (figure 2 and 3). All analyzed parameters were also checked for correlation with latitude and longitude, the correlations were negligible.

To sum up, lower differences between gridded and weather station data for temperature were observed for CRU data. This is supported by higher correlation values and better results for regression models: higher values of R^2 and lower values of mean error received for this factor. However for Śnieżka location, the UD data should be preferred. The precipitation data results can be affected by high spatial variability of this factor, especially in mountain areas, which can be difficult to preserve in gridded data. GPCC and CRU GCD had enough high similarity to weather stations data, that is confirmed by Pearson correlation coefficient, R^2 , and L1- norm. Over all it can be stated that analyzed CRU and GPCC gridded data can be used in dendroclimatological studies in the area, but in some cases with caution, observing the difference in altitude.

Acknowledgment(s)

The study was supported by the National Science Centre, Poland, project no 2014/13/B/ST10/02529 and AGH University of Science and Technology as a part of statutory projects no. 11.11.140.626 and 11.11.140.613.

References

[1] J. Ramirez-Villegas, and A. Challinor, "Assessing relevant climate data for agricultural

- applications”, *Agric. For. Meteorol.*, vol. 161, pp. 26-45, 2012.
- [2] J. Van Wart, P. Grassini, and K.G. Cassman, “Impact of derived global weather data on simulated crop yields”, *Global Change Biol.*, vol.19, iss.13, pp. 3822-3834, 2013.
- [3] J. Van Wart, P. Grassini, H.S. Yang, L. Claessens, A. Jarvis, and K.G. Cassman, “Creating long-term weather data from the thin air for crop simulation modelling”, *Agric. For. Meteorol.*, vol. 208, pp. 49-58, 2015.
- [4] C. Daly, G.H. Taylor, W.P. Gibson, T.W. Parzybok, G.L. Johnoson, and P.A. Pasteris, “High-quality spatial climate data sets for the United States and beyond, Transaction of the ASAE, vol. 43, iss. 6, pp. 1957-1962, 2000.
- [5] S. Mourtzinis, J.I. Rattalino Edreira, S.P. Conley, and P. Grassini, “From grid to field: Assessing quality of gridded weather data for agricultural applications”, *European Journal of Agronomy*, vol. 82(A), pp. 163-172, 2017.
- [6] D.A. Jones, W. Wang, and R. Fawcett, “High-quality spatial climate data-set for Australia”, *Australian Meteorological and Oceanographic Journal*, vol. 58, pp. 233-248, 2009.
- [7] J.L.J. Ledesma, and M.N. Futter, “Gridded climate data products are an alternative to instrumental measurements as inputs to rainfall–runoff models”, *Hydrological Processes*, vol. 31, iss. 18, pp. 3283-3293, 2017.
- [8] M. Chuchro, and M. Danek, “Applicability of gridded climate data in tree-ring studies of larch from southern part of Poland”, *SGEM 2017 conference proceedings*. vol. 17, Water resources, forest, marine and ocean ecosystems. Iss. 32, Soils, forest ecosystems, pp. 717-725, 2017.
- [9] R. Sánchez-Salguero, R.M. Navarro-Cerillo, J.J. Camarero, and A. Fernández-Cancio, “Selective drought-induced decline of pine species in southeastern Spain”, *Clim Ch.*, vol. 113, pp. 767–785, 2012.
- [10] V. Treml, T. Ponocná, and U. Büntgen, “Growth trends and temperature responses of treeline Norway spruce in the Czech-Polish Sudetes Mountains”, *Clim.Res.*, vol. 55, pp. 91–103, 2012.
- [11] IMGW-PIB webpage: <http://www.imgw.pl>
- [12] C.J. Willmott, and K. Matsuura, “Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1950-1999)”: http://climate.geog.udel.edu/~climate/html_pages/README.ghcn_ts2.html
- [13] GRID model database: <https://www.esrl.noaa.gov/>.
- [14] I. Harris, P.D. Jones, T. J. Osborn, and D.H. Lister, “Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset”. *Int. J. Climatology*, vol.34, pp. 623–642, 2014.
- [15] RELEASE NOTES FOR CRU TS v3.25 / v4.01 OBSERVATIONS: September 2017: https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_3.25/observation.v3.25.v4.01/Release_Notes_CRU_TS3.25_4.01_observations.txt
- [16] CRU TS v. 4.01 dataset- crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.01/.
- [17] U. Schneider, A. Becker, P. Finger, A. Meyer-Christoffer, B. Rudolf, and M. Ziese, “GPCC Full Data Reanalysis Version 6.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data”, Global Precipitation Climatology Centre, DOI: 10.5676/DWD_GPCC/FD_M_V6_050, 2011.
- [18] U. Schneider, A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf, “GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle”, *Theoretical and Applied Climatology*, vol.115, iss. 1-2, pp. 15-40, 2014.
- [19] M. Yetkin, and C. Inal, “ L_1 Norm Minimization in GPS Networks”, *Survey Review*, vol. 43, iss. 323, pp. 523-532, 2013.
- [20] N.R. Draper, and H. Smith, “Applied Regression Analysis,” *Wiley Series in Probability and Statistics*, Third Edition, 1998.