Validation of Near-surface Wind Speed Monthly Averages from CHELSA Climate Data in Croatia

Dorijan Radočaj[1,], Mateo Gašparović[2], Mladen Jurišić[1]*

[1] Department of Agricultural Engineering and Renewable Energy Sources, Faculty of Agrobiotechnical Sciences Osijek, Osijek, Croatia, dradocaj@fazos.hr, mjurisic@fazos.hr

[2] Chair of Photogrammetry and Remote Sensing, Department of Cartography and Photogrammetry, University of Zagreb – Faculty of Geodesy, Zagreb, Croatia, mateo.gasparovic@geof.unizg.hr

* corresponding author

doi: 10.5281/zenodo.11510498

Abstract: This study validated monthly averages of near-surface wind speed (NSW) from the 1km Climatologies at high resolution for the earth's land surface areas (CHELSA) climate data using long-term ground truth data from 22 climate stations operated by the Croatian Meteorological and Hydrological Service (DHMZ). The CHELSA NSW monthly averages resulted in the coefficient of determination (R^2) from 0.390 to 0.691 and the normalized root mean square error in range of 0.409–0.635. A significant difference in validation results was observed between the October-April and March-September periods, with mean R^2 values of 0.655 and 0.442, respectively. There is a slight temporal discrepancy between CHELSA v2.1 (1981–2010) and the latest official DHMZ climate data (1971–2000), which may affect the accuracy of CHELSA data in Croatia. However, it could also indicate the effects of climate change on NSW, which produced higher residuals from DHMZ ground truth data in warmer May–September period and in coastal DHMZ climate stations. To complement on the current study, future research should resolve the temporal delay with the eventual publication of more recent ground truth data from DHMZ. Additionally, similar open climate datasets should be included to determine the most suitable 1 km resolution climate data source in Croatia.

Keywords: Croatian Meteorological and Hydrological Service (DHMZ); accuracy assessment; open climate data; climate change.

1 Introduction

Remote sensing technologies have become essential for producing precise and high-resolution climate datasets (Yang et al. 2013). Among them, the Climatologies at high resolution for the earth's land surface areas (CHELSA) provide rasterized climatic variables at 1 km spatial resolution, allowing for detailed analysis and modelling at regional to global scales (Karger et al. 2020). The relationship of high-resolution raster climate data based on ground truth observations determines the reliability of predictive mapping used in the creation of such datasets. Ground truth climate data, which are direct measurements of key climatic variables like temperature, precipitation, and humidity at specific locations and time intervals, are usually provided by climate and meteorological stations under governmental agencies (Colston et al. 2018). Uncertainties persist due to the intrinsic unpredictability of climatic events, sensor limits, and computational complexity, even with breakthroughs in remote sensing technology (Xu et al. 2021). Validation ensures that raster datasets accurately reflect local variations and temporal trends by closely examining the alignment between remote sensing estimates and ground-based observations across various landscapes and climatic regimes. This procedure confirms the accuracy of data obtained by remote sensing and identifies any biases or inaccuracies caused by sensor peculiarities, algorithmic errors, or environmental factors.

Enabling cross-comparison and data fusion is crucial to fully utilize a range of similar climatic datasets, such as WorldClim (Fick and Hijmans 2017). The unique advantages of each dataset can be utilized by combining them, such as the temporal continuity of ground-based observations and the geographical scope of remote sensing data. Verifying raster climate databases against ground truth observations can additionally identify discrepancies and biases in the long-term monthly climate averages, facilitating the harmonization of different datasets (Katrandzhiev et al. 2022). The integration and cross-comparison of datasets was made possible by such harmonization, leading to a more comprehensive understanding of climate dynamics and trends in previous studies. This integrated approach enhances the ability to monitor changes in the environment, assess the impacts of climate change, and develop effective plans for adaptation and mitigation in response to global warming (Abbass et al. 2022).

The aim of this study was to evaluate the accuracy of near-surface wind speed (NSW) CHELSA monthly averages according to ground truth Croatian Meteorological and Hydrological Service (DHMZ) climate stations in Croatia. As Croatian area consists of three distinct biogeoregions (Continental, Alpine and Mediterranean), the results of this study are expected to provide observations on the effects of environmental conditions to climate data accuracy, providing guidance for future studies.

2 Materials and methods

The input data for the validation of near-surface wind speed monthly averages consisted of two main data sources: 1) ground truth monthly averages from DHMZ climate stations from 1971–2000 period, and 2) CHELSA v2.1 monthly averages according to 1981–2010 period (Karger et al. 2020). Both data sources represent the most recent official datasets from their respective sources, which ultimately produced a slight temporal discrepancy due to the lack of availability of more recent DHMZ data.

The geospatial analysis and validation were performed using R v4.0.3 with "terra" package. CHELSA rasters were reprojected to the Croatian Terrestrial Reference System (HTRS96/TM) using the bilinear interpolation.

The ground truth data consisted of only 22 DHMZ climate stations due to restrictions on their availability imposed by responsible DHMZ personnel for the purpose of this study. Their geospatial distribution across the study area, covering the entire land area of Croatia, is displayed in Figure 1.

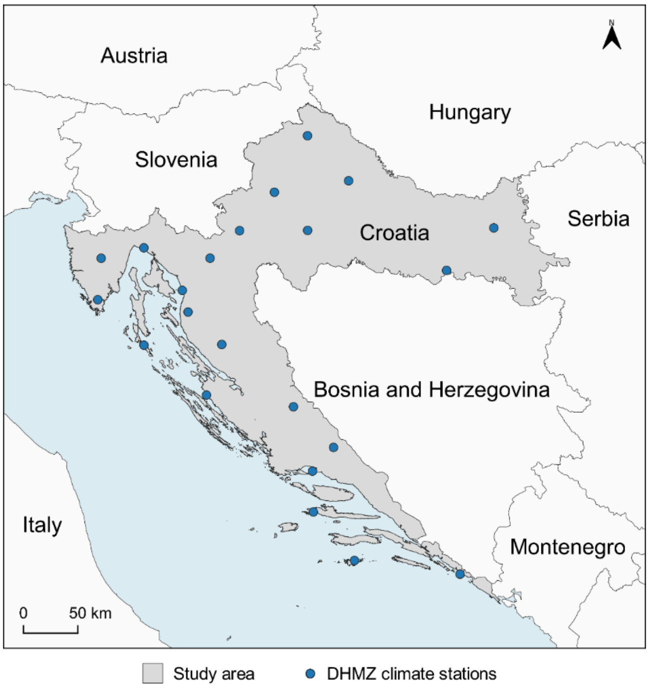

Figure 1. The geospatial distribution of ground truth climate DHMZ stations.

The CHELSA v2.1 near-surface wind speed monthly average rasters were acquired using CHELSA file browser from the Cyberduck client. The rasters were clipped to the study area and reprojected to HTRS96/TM in 1 km spatial resolution. The average monthly near-surface wind speed rasters in Croatia are displayed in Figure 2. The validation of CHELSA near-surface wind speed monthly averages according to the ground truth DHMZ climate station data was performed according to three statistical metrics: coefficient of determination (R^2), root mean square error (RMSE) and normalized RMSE (NRMSE). These metrics provide complementary information on relative (R^2) and absolute (RMSE and NRMSE) accuracy of evaluated CHELSA rasters. Moreover, R^2 and NRMSE have normalized value ranges from 0 to 1, enabling accuracy comparison between months regardless of the residuals and average near-surface wind speed values. The higher R^2 and lower RMSE and NRMSE indicated higher accuracy of CHELSA near-surface wind speed monthly averages according to the ground truth DHMZ climate stations.

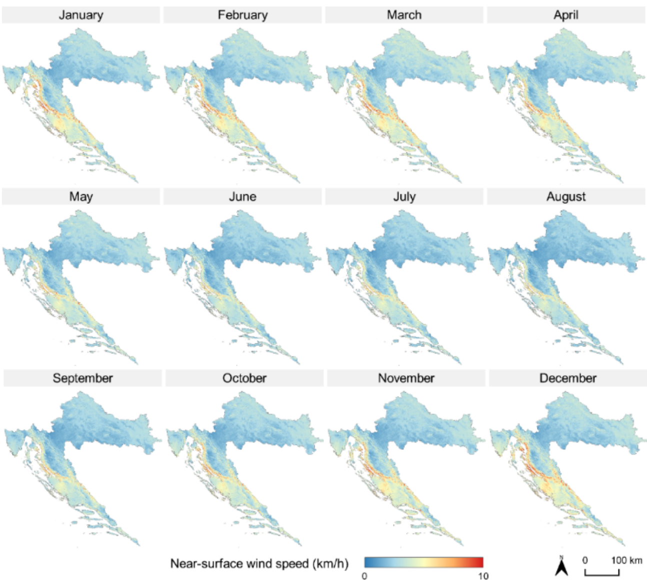

Figure 2. The display of average monthly NSW rasters in Croatia.

3 Results and Discussion

The highest accuracy of CHELSA near-surface wind speed monthly averages in terms of overall fit was observed in January, February, and December, as indicated by the high R^2 (Table 1). However, December exhibited a higher error rate, indicating a complex relationship between fit represented by R^2 and error magnitude from RMSE and NRMSE, which requires further research. The disagreement in CHELSA near-surface wind speed monthly averages, as represented by either high or low both R^2 and RMSE, could indicate a consistent bias of near-surface wind speed in CHELSA dataset, as well as the limited variability in its values (Chicco et al. 2021). Despite the presence of three distinct biogeoregions in Croatia (Continental, Alpine and Mediterranean) (Pilotto et al. 2020), the yearly averages of available DHMZ climate stations had a value range of 1.3–3.3 km/h, which may also distort validation results due to few stations with noticeably higher yearly average.

Table 1. The properties of available ground truth climate DHMZ stations used in the study.

Month	R^2	RMSE	NRMSE
January	0.689	1.16	0.578
February	0.687	1.06	0.499
March	0.620	1.08	0.488
April	0.565	0.95	0.433
May	0.451	0.87	0.438
June	0.396	0.77	0.409
July	0.390	0.80	0.437
August	0.476	0.78	0.450
September	0.496	0.97	0.541
October	0.648	1.02	0.518
November	0.687	1.15	0.559
December	0.691	1.30	0.635

A clear seasonal variability pattern was observed, with poorer performance during the summer months (June to September) in terms of both fit and error measures. This fluctuation suggests that the accuracy of CHELSA monthly average rasters may be impacted by the exclusion of important seasonal elements or other time-variant predictors from its framework. Its accuracy could be enhanced by retraining it to consider the discovered time-variant or seasonal impacts, either by incorporating new variables or by utilizing more a suitable modeling approach for different seasons of the year.

Figure 3 represents the monthly residuals magnitude of CHELSA near-surface wind speed according to the ground truth DHMZ climate stations. There is a notable distinction between residuals in northern (primarily Continental region) and southern (primarily Mediterranean region) parts of Croatia, with the largest residuals across all months concentrated in the southern coastal part of Croatia. It could also indicate the effects of climate change on more recent near-surface wind speed monthly averages, which produced higher residuals in warmer May–September period and in coastal DHMZ climate stations. There were also less available ground truth DHMZ climate stations in the eastern part of Croatia, which likely impacted CHELSA validation procedure as incomplete coverage with ground truth data produced similar results in a previous study (Radočaj et al. 2023). For smaller study areas, this issue could be resolved by establishing a local network of meteorological stations and interpolating their data, which is usually viable for valuable agricultural areas (Ramirez-Villegas and Challinor, 2012).

Therefore, future studies should be based on denser ground truth climate station data and temporal discrepancy between CHELSA or similar dataset with the ground truth data should be resolved, if possible due to DHMZ data availability. Additionally, the inclusion of other climate parameters and similar datasets in the validation approach, such as WorldClim (Fick and Hijmans, 2017), could provide a more complete assessment of the reliability of open data climate rasters in 1 km spatial resolution.

4 Conclusions

This study proved an overall moderate accuracy of near-surface wind speed monthly averages from CHELSA according to available ground truth DHMZ climate stations, with high variability of accuracy metrics across the year. There was a slight disagreement in terms of accuracy from a relative (quantified by R^2) and absolute standpoint (quantified by RMSE and NRMSE), likely indicating a consistent bias of near-surface wind speed in CHELSA

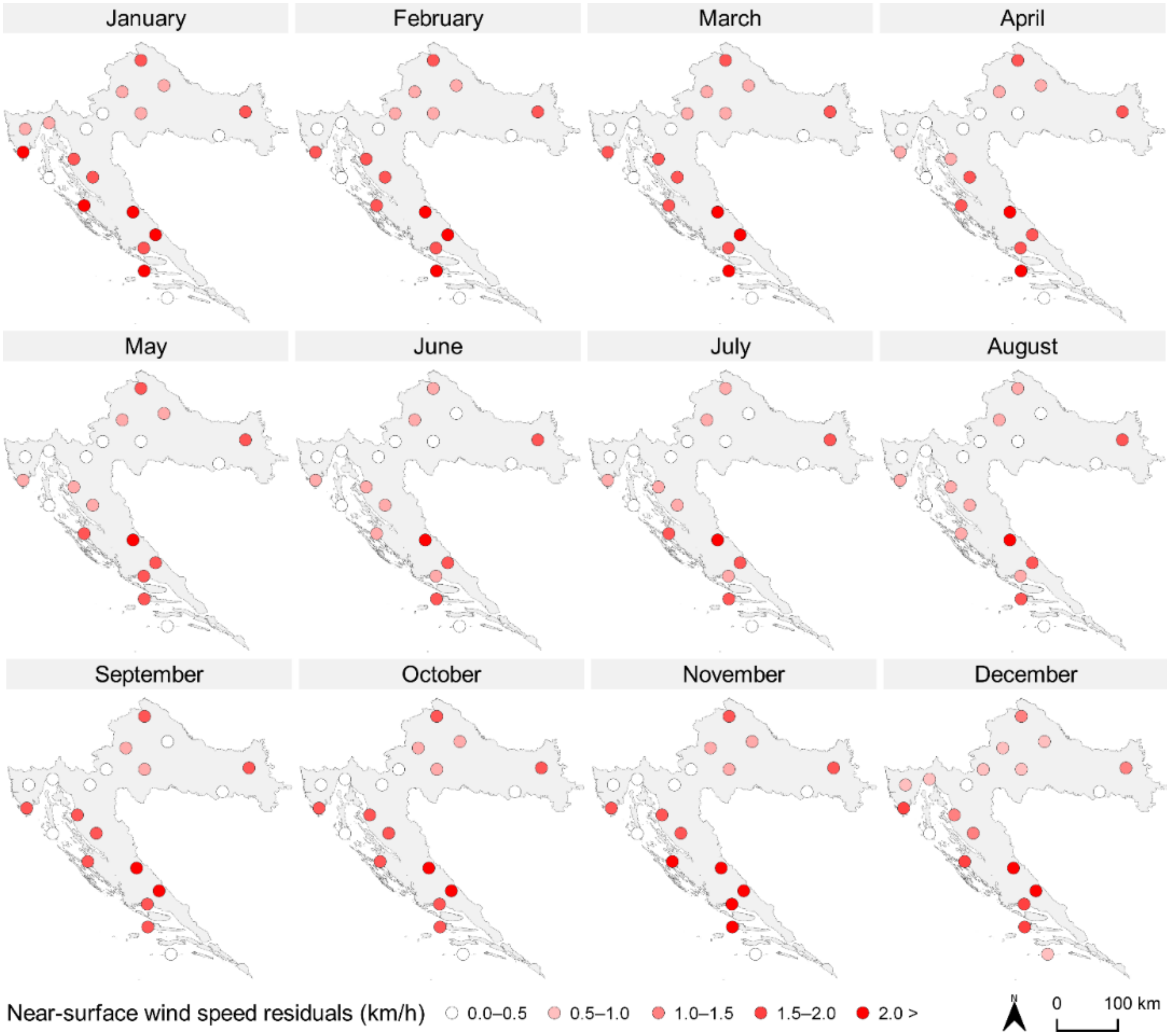

Figure 3. The display of NSW residuals of CHELSA monthly averages according to the ground truth DHMZ climate stations.

dataset. Despite the presence of three distinct biogeoregions in Croatia, the yearly averages of available DHMZ climate stations had a relatively limited variability in its values. A notable distinction between residuals in northern continental region and southern Mediterranean region of Croatia was observed, with the largest residuals concentrated in the southern coastal part of Croatia. It could also indicate the effects of climate change on more recent near-surface wind speed monthly averages, which produced higher residuals in warmer May–September period and in coastal DHMZ climate stations. To improve the study, future research should resolve the temporal discrepancy with the eventual publication of more recent ground truth data from DHMZ. Additionally, similar open climate datasets should be included to determine the most suitable 1 km resolution climate data source in Croatia.

Acknowledgments: This research was funded by the European Space Agency (ESA) for the project "Cropland suitability prediction method for land management planning based on Copernicus data for soybean – CropSuit" under contract 4000143723/24/NL/MH/yd and supported by the Croatian Science Foundation for the ALCAR project: "Assessment of the Long-term Climatic and Anthropogenic Effects on the Spatio-temporal Vegetated Land Surface Dynamics in Croatia using Earth Observation Data" (Grant No. HRZZ IP-2022-10-5711).

5 References

Abbass, K., Qasim, M. Z., Song, H., Murshed, M., Mahmood, H., Younis, I., 2022. A review of the global climate change impacts, adaptation, and sustainable mitigation measures. Environmental Science and Pollution Research, 29 (28), 42539-42559.

Chicco, D., Warrens, M. J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. Peerj computer science, 7, e623.

Colston, J. M., Ahmed, T., Mahopo, C., Kang, G., Kosek, M., de Sousa Junior, F., ... The, M. E., 2018. Evaluating meteorological data from weather stations, and from satellites and global models for a multi-site epidemiological study. Environmental research, 165, 91-109.

Fick, S. E., Hijmans, R. J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International journal of climatology, 37 (12), 4302-4315.

Karger, D. N., Schmatz, D. R., Dettling, G., Zimmermann, N. E., 2020. High-resolution monthly precipitation and temperature time series from 2006 to 2100. Scientific data, 7 (1), 248.

Katrandzhiev, K., Gocheva, K., Bratanova-Doncheva, S., 2022. Whole System Data Integration for Condition Assessments of Climate Change Impacts: An Example in High-Mountain Ecosystems in Rila (Bulgaria). Diversity, 14 (4), 240.

Pilotto, F., Kühn, I., Adrian, R., Alber, R., Alignier, A., Andrews, C., ... Haase, P., 2020. Meta-analysis of multidecadal biodiversity trends in Europe. Nature communications, 11 (1), 3486.

Radočaj, D., Jurišić, M., Rapčan, I., Domazetović, F., Milošević, R., Plaščak, I., 2023. An Independent Validation of SoilGrids Accuracy for Soil Texture Components in Croatia. Land, 12 (5), 1034.

Ramirez-Villegas, J., Challinor, A., 2012. Assessing relevant climate data for agricultural applications. Agricultural and forest meteorology, 161, 26-45.

Xu, L., Chen, N., Chen, Z., Zhang, C., Yu, H., 2021. Spatiotemporal forecasting in earth system science: Methods, uncertainties, predictability and future directions. Earth-Science Reviews, 222, 103828.

Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., Dickinson, R., 2013. The role of satellite remote sensing in climate change studies. Nature climate change, 3 (10), 875-883.