data

MDPI

Technical Note

ncPick: A Lightweight Toolkit for Extracting, Analyzing, and Visualizing ECMWF ERA5 NetCDF Data

Sreten Jevremović, Filip Arnaut, Aleksandra Kolarski and Vladimir A. Srećković *

Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Pregrevica 118, 11000 Belgrade, Serbia; sjevremovic@ipb.ac.rs (S.J.); filip.arnaut@ipb.ac.rs (F.A.); kolarski@ipb.ac.rs (A.K.)
* Correspondence: vlada@ipb.ac.rs

Abstract

The European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) datasets provide a rich source of climatological data. However, their Network Common Data Form (NetCDF) structure can be a barrier for researchers who are not experienced with specialized data tools or programming languages. To address this challenge, we developed ncPick, a lightweight, Windows-based application designed to make ERA5 data more accessible and easier to use. The software enables users to load NetCDF files, select points of interest manually or through shapefiles, and export the data directly to Comma-separated values (CSV) format for further processing in common tools such as Excel, R, or within ncPick itself. Additional modules allow for quick visualization, descriptive statistics, interpolation, and the generation of time-of-day heatmaps, as well as practical data handling functions such as merging and downsampling CSV files based on the time-axis. Validation tests confirmed that ncPick outputs are consistent with those from established tools (such as Panoply). The toolkit was found to be stable across different Windows systems and suitable for a range of datasets. While it has limitations with very large files and does not include automated data download for version 1 of the software, ncPick offers an accessible solution for researchers, students, and other professionals seeking a reliable and intuitive way to work with ERA5 NetCDF data.

Keywords: ncPick; ECMWF ERA5; NetCDF; software; climatological dataset; meteorological dataset; Copernicus Climate Data Store

Academic Editor: Juan-Carlos Jiménez-Muñoz

Received: 4 September 2025
Revised: 29 October 2025
Accepted: 31 October 2025
Published: 2 November 2025

Citation: Jevremović, S.; Arnaut, F.; Kolarski, A.; Srećković, V.A. *ncPick*: A Lightweight Toolkit for Extracting, Analyzing, and Visualizing ECMWF ERA5 NetCDF Data. *Data* **2025**, *10*, 178. https://doi.org/10.3390/data10110178

1. Introduction

The European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) datasets are among the most widely used sources of climate and meteorological data [1] distributed through the Copernicus Climate Data Store (CDS).

The ERA5 dataset is freely available and covers a broad range of parameters, which makes it an attractive source of data for a wide array of researchers. The ERA5 dataset has been widely applied in agriculture [2–6], forestry [7–9], traffic engineering [10–12], geosciences [13–16], and many other fields that demonstrate the need for spatially resolved climate or environmental data. However, the datasets are distributed in the NetCDF, which can be difficult to handle for users without prior experience with specialized tools or programming languages. The NetCDF is maintained and developed by the Unidata Program Center, part of the University Corporation for Atmospheric Research (UCAR), under the support of the U.S. National Science Foundation [17].

Currently, there are a number of existing applications that can process NetCDF data. Examples include Panoply, SNAP, IDV, CDAT, Met.3D, and ncBrowse [18]. These tools are highly capable and provide advanced visualization or processing options, but they are often part of larger and more complex software or tools. As a result, users who just need to extract, visualize, or convert ERA5 NetCDF files may face significant effort to master the new software. Many tools require installing multiple components, learning unfamiliar workflows, or adapting to interfaces with numerous functions beyond NetCDF processing. Many smaller tools that once supported NetCDF manipulation are now difficult to obtain, as their websites have become outdated or their download links inactive.

The core idea behind ncPick was to address these gaps by providing a lightweight, intuitive freeware Windows desktop application with a graphical user interface (GUI) dedicated specifically to ERA5 NetCDF files [19]. The main goal was to offer a simple, Windows-based application that could be understood intuitively and used with minimal technical knowledge. Instead of competing with multifunctional platforms, ncPick focuses only on the essential operations that researchers need, which are loading NetCDF data, selecting and exporting points of interest, performing descriptive statistics, visualization, interpolation, and preparing CSV files for further analysis [19]. The basic reason for choosing the CDS as the primary data source for software development lies in its extensive repository of standardized data. For most users, the CDS offers sufficient data for purposes of research, and creating software to improve accessibility to this data is considered advantageous, particularly for researchers, students, and professionals unfamiliar with the NetCDF yet requiring climatological and environmental data. The future prospects of ncPick are outlined in detail in the roadmap available on the software's website [20]. Distribution occurs through the Zenodo repository, with maintenance and updates, including new features, scheduled biannually as specified in the roadmap.

The novelty of ncPick lies in this combination of simplicity, accessibility, and availability. While many larger platforms include NetCDF support as one of many features, ncPick was designed from the beginning for this exact purpose. It is small, intuitive, portable, publicly available, and does not require advanced configuration. By lowering the technical barrier, it enables researchers from diverse disciplines to use the ERA5 datasets effectively, without investing time in mastering large software systems.

2. About the Software

ncPick was developed at the Laboratory for Astrophysics and Physics of the Ionosphere, Institute of Physics Belgrade, University of Belgrade, Serbia, using Python 3.12 [19]. The software runs on Windows and requires no additional setup beyond downloading and starting the executable. Its design principle is to bridge ERA5 NetCDF datasets to widely used analytical tools such as Excel, JASP, or algorithmic languages such as R and Python (Figure 1).

A significant segment of the software's development focused on establishing connections to existing tools for different functions, which are not present in other software packages that deal with NetCDF data. An illustration of this is the module designed to convert exported CSV files into a TRAINSET-compatible CSV file for the purposes of machine learning data labeling. Furthermore, what distinguishes ncPick from other software packages is the modules for merging with additional CSV data and for downsampling data. This was executed to enhance multidisciplinary research; specifically, certain researchers aim to potentially correlate their data with that acquired from CDS. The ncPick tool facilitates the seamless creation of final files for analysis, as shown in Section 3.4.

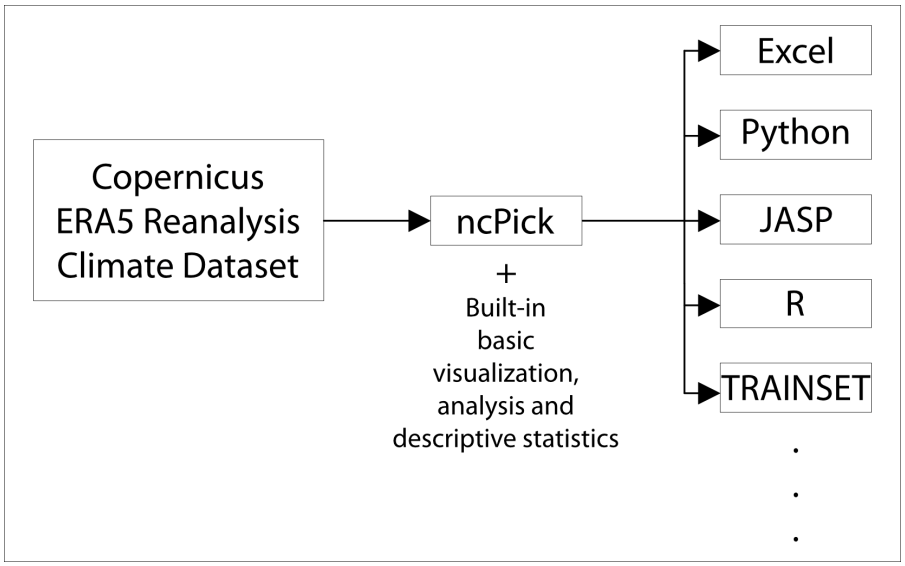

Figure 1. Location of ncPick within the data processing pipeline alongside commonly used and freely available tools.

The interface was intentionally designed to be minimal, providing only essential functions to ensure simplicity and ease of use. Users can load NetCDF files, visualize selected points, or export subsets to CSV for downstream analysis. The tool supports manual point selection, shapefile-based selection, or exporting all data points in a file [19]. Moreover, an integrated basemap offers immediate spatial context without necessitating an online connection to a web service, thereby enabling the software to function offline.

3. Workflow and Features

This section is divided into four parts, which explain the most important features, development processes, and workflow of the ncPick software, as illustrated in Figure 2.

3.1. Initialization and Basemap

At startup, ncPick loads a static basemap containing global administrative boundaries, major rivers, roads, and urban areas (Figure 3). This enables offline use and avoids dependency on online map services. The basemap was optimized to balance detail with performance, ensuring short loading times and modest memory use [19].

3.2. Data Loading and Selection

After loading the data, users can choose from several available options as needed for data preparation and analysis:

- Select single or multiple points directly on the map.
- Filter data spatially with shapefile polygons (e.g., selecting only one country or a custom-built shapefile of the given research area).
- Export all available data points from the NetCDF file in a single step.

Exports generate CSV files that can be easily integrated into standard software packages familiar to most researchers, students, and professionals.

3.3. Visualization and Statistics

For each selected point, ncPick generates a time series graph with simple zoom controls [21]. Along with the graph, the software automatically provides descriptive statistics such as mean, quartiles, standard deviation, and minimum/maximum values [22].

This makes it easy to get both a visual and numerical sense of the dataset without leaving the program.

Apart from the single-point graphs, ncPick offers tools for exploring larger patterns. The interpolation view creates a continuous surface from the NetCDF grid, which can be animated over time to reveal various spatial dynamics. The interpolation parameters provide three playback speeds: slow (1000 ms), medium (200 ms), and fast (20 ms), utilizing the linear interpolation method from SciPy (milliseconds). With sequential processing, the playback speed may fluctuate based on the displayed area.

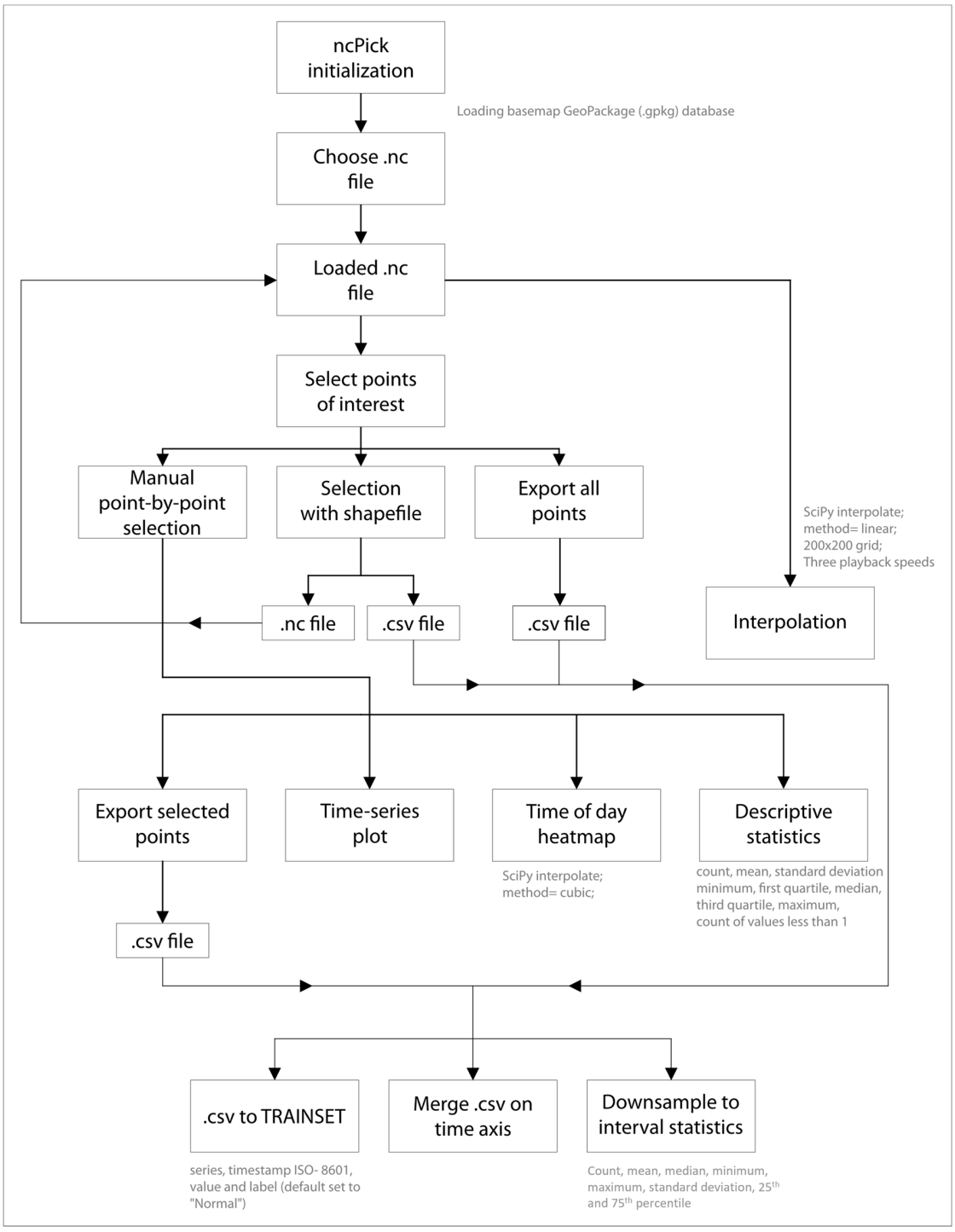

Figure 2. Structure and workflow of ncPick.

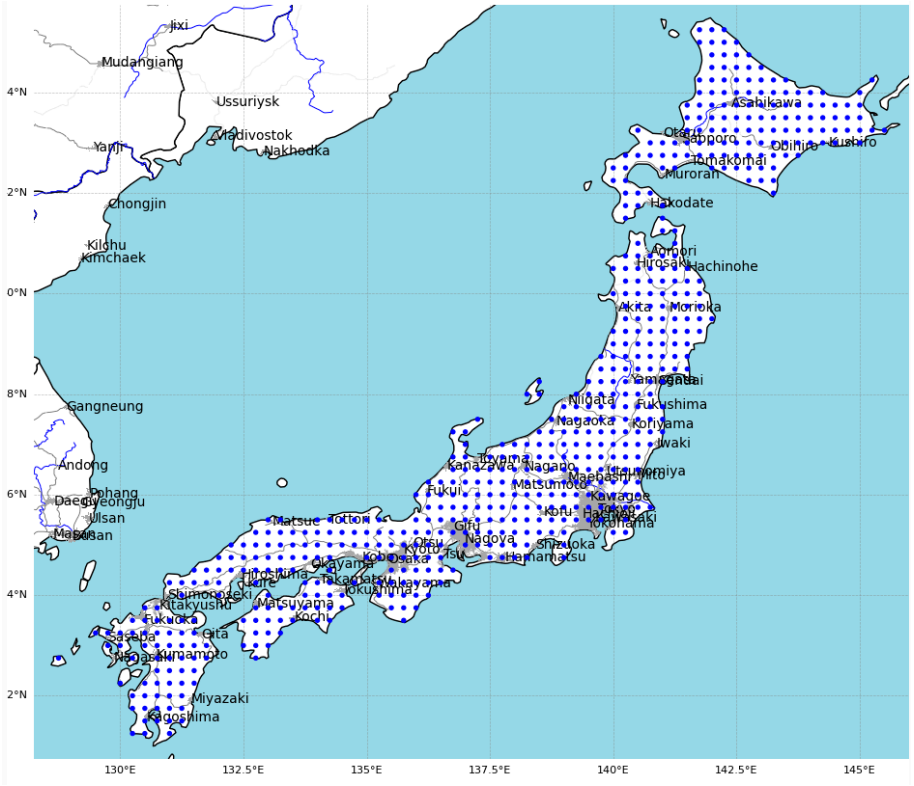

Figure 3. Basemap of Japan, with the NetCDF file obtained from ncPick after selecting the export from shapefile geometry function, with data points.

Another option is the time-of-day heatmap (Figure 4), where the day of the year is displayed along the x-axis and the time of day along the y-axis. This type of visualization is especially useful for detecting daily cycles in parameters such as temperature or humidity [23], as displayed in Figure 4 for two points located within mainland Japan, with the parameter of temperature at 2 m. Static visualizations in ncPick can be exported in a publication-ready resolution, i.e., 300 DPI.

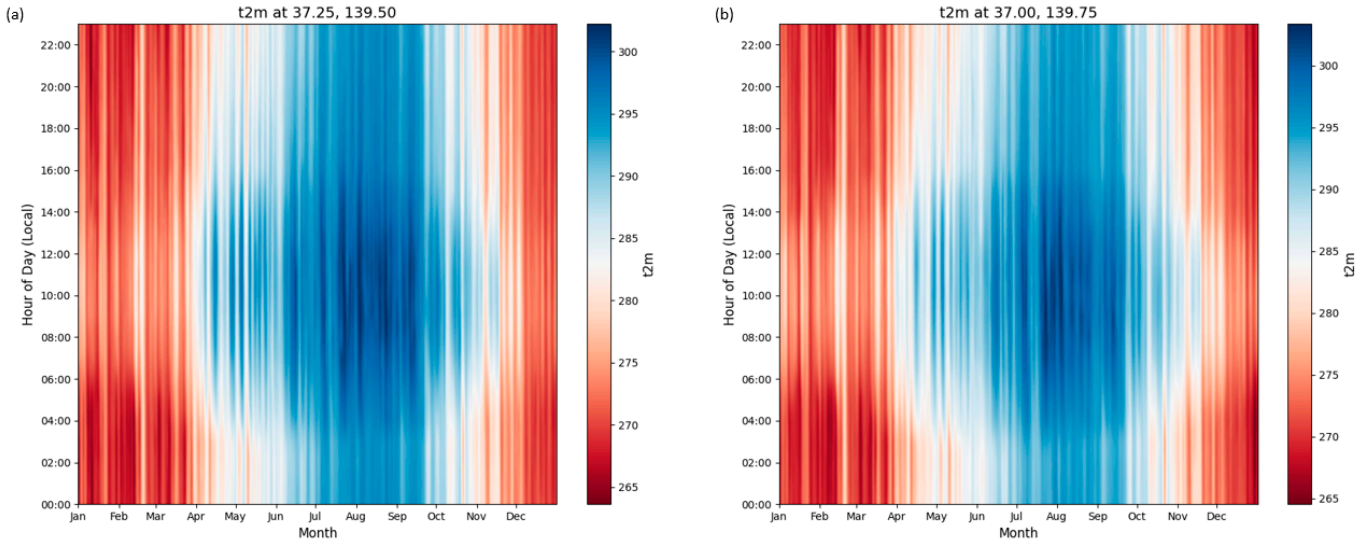

Figure 4. Temperature at 2 m time-of-day heatmap for two selected points ((**a**): 37.250° N, 139.500° E; (**b**): 37.000° N, 139.750° E) in mainland Japan, visualized in ncPick.

The visualization modules provide users with multiple ways to explore their data, whether by focusing on individual points, comparing areas, or identifying seasonal and daily patterns.

3.4. CSV Processing Tools

To simplify and streamline data preparation, ncPick also includes:

- Convert CSV files to TRAINSET-compatible format. TRAINSET is a web-based, client-side application used for labeling datasets in machine learning [24], and ncPick provides a direct conversion module for this purpose. Consequently, it was considered advantageous to create a module capable of simply converting users' CSV files into TRAINSET-compatible CSV files, which can subsequently be manually labeled in TRAINSET.
- Merge CSV files by aligning time axes. This function allows users to quickly and efficiently combine two CSV files based on the shared intersecting time axis.
- Aggregate data into interval-based summaries (daily, hourly, etc.) using descriptive measures. This function allows users to downsample the time resolution of their data to a selected interval, specifically daily (24 h) or hourly intervals, such as 3, 6, or 12 h, or any other value.

4. Testing and Quality Control

Testing was carried out to ensure that all functions of ncPick operated correctly and produced the expected results. In addition, exports from ncPick were cross-validated with those from an established software package in the domain of NetCDF files obtained from the CDS. This process involved checking the accuracy of outputs, verifying that all features worked as intended, and ensuring overall software stability. Testing was carried out by comparing outputs with those from Panoply and independent scripts [25]. The tests confirmed:

- Exported point values were within 0.5% of Panoply outputs for equivalent parameters (differences attributable only to rounding).
- Interpolations and visualizations matched visual outputs of other software (Figure 5).
- CSV conversion and downsampling produced correct results when verified manually.

Stability tests on Windows 10 and 11 showed consistent operation across different systems, with all modules executing without errors or crashes during repeated runs.

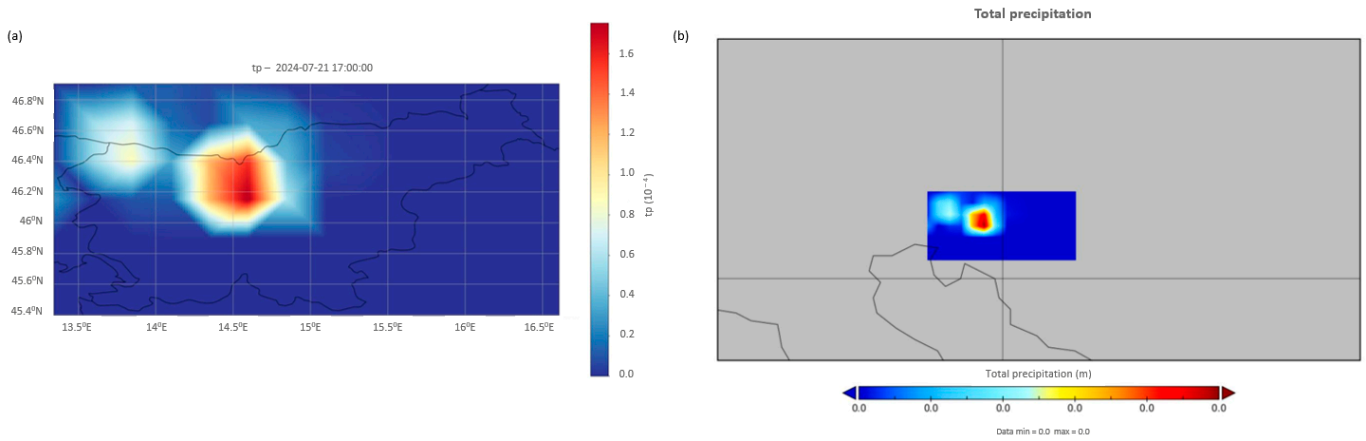

Figure 5. Interpolation validation results: ncPick (**a**) and Panoply (**b**), for precipitation data for Slovenia.

Stress Test

To evaluate ncPick under demanding conditions, a large ERA5 NetCDF file was tested, covering a 50° × 50° region over North America (almost all of Canada, the continental United States, and the northern part of Mexico) with hourly 2 m temperature data for 2024 [19] (Figure 6). This dataset contained around 40,000 grid points and nearly 350 million values, far exceeding the scale of most everyday use cases. The software loaded the file successfully, and basic operations such as point selection, time series visualization, and descriptive statistics functioned as intended. Exporting individual points was also completed without issues, while applying the shapefile filter to extract data for the state of Texas produced a valid 85 MB CSV, though with longer processing time and some temporary unresponsiveness.

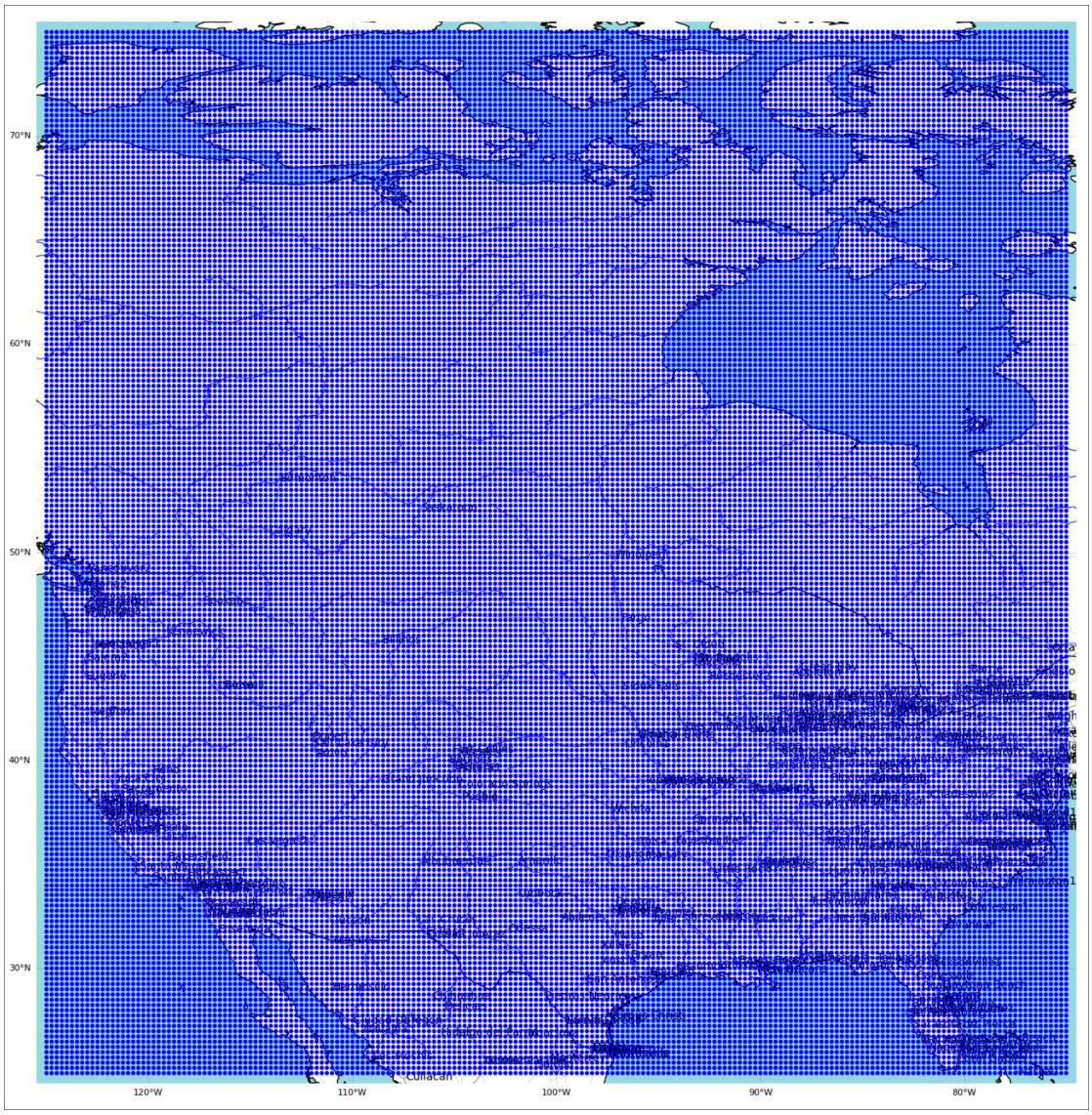

Figure 6. The 50° × 50° area over North America used for the stress test.

The test showed that performance limits are mainly tied to the spatial extent of the dataset rather than its temporal length [19]. While ncPick remained stable, some modules, particularly interpolation, became slower or temporarily unresponsive at this scale. Memory consumption stayed moderate, but heavy operations were CPU- and I/O-bound. For practical use, very large domains should be divided into smaller regions to maintain

responsiveness. For typical research areas, such as smaller countries in Europe or other regions of moderate size, ncPick runs smoothly and without noticeable performance issues.

5. Future Development

Several potential features are being considered. These include automation capabilities for retrieving data directly from the CDS, support for additional file formats, e.g., hierarchical NetCDF data files, and the development of new modules such as an anomaly detection module or spatial analysis tools [19]. These features are at the planning stage and will be implemented according to the roadmap on the ncPick website [20], with updates released biannually as described in the user manual.

6. Conclusions

This technical note presents ncPick, a newly developed lightweight Windows-based application designed to make Copernicus ERA5 NetCDF datasets more accessible. The tool lowers the entry barrier for researchers, students, and professionals who need fast and reliable access to climatological data without relying on large and complex platforms. By focusing on essential functions, such as loading and exporting NetCDF data, visualization, descriptive statistics, interpolation, and CSV processing, ncPick fills a clear gap between multifunctional software packages and outdated or unavailable similar tools.

As with any lightweight solution, ncPick has certain limitations. Performance may decrease when very large NetCDF files are loaded, and extensive CSV exports (more than a million rows of data) may exceed the capacity of spreadsheet applications such as Microsoft Excel. The standalone executable may occasionally be flagged by antivirus software, which is common for unsigned applications. Furthermore, ncPick was designed primarily for Copernicus ERA5 files, so compatibility with other NetCDF datasets cannot yet be guaranteed but will be tested in future iterations. The software also does not provide automated downloading of datasets, which means users must obtain the files manually.

Despite these limitations, ncPick has proven to be a practical and reliable solution, as validation tests (Section 4) confirmed its consistency with established tools such as Panoply and its stable performance on multiple Windows systems. Its main contribution lies in its simplicity and availability: researchers from various disciplines can now access ERA5 data with minimal effort and integrate it into their existing workflows. The software may be expanded (in accordance with the needs) with features such as automated data retrieval, anomaly detection, or broader format support, but its central aim will remain unchanged: to provide an intuitive, efficient, and user-friendly gateway to Copernicus ERA5 reanalysis data.

Author Contributions: Conceptualization, methodology, software, formal analysis, writing—original draft, writing—review and editing, S.J.; conceptualization, software, validation, formal analysis, writing—review and editing, visualization, F.A.; conceptualization, writing—review and editing, supervision, A.K.; V.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge funding provided by the Institute of Physics Belgrade, through a grant from the Ministry of Science, Technological Development and Innovation of the Republic of Serbia.

Data Availability Statement: The data used for testing the developed ncPick software, as well as the ncPick software and supporting documentation, examples and instructions are openly available in: https://doi.org/10.5281/zenodo.17453305 or on https://arnautf.github.io/ncpick_website/ (accessed on 29 October 2025).

Acknowledgments: The developers/authors would like to express their gratitude to Copernicus (Program of the European Union), Natural Earth and nvkelso for providing vector data used in the

basemap, rbende for the Sun Valley theme utilized in ncPick, and opendatasoft for making the World Administrative Boundaries-Countries and Territories dataset freely available. Without such openly accessible datasets and resources, software solutions like ncPick would not be possible.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hersbach, H.; Bell, B.; Berrisford, P.; Biavati, G.; Horányi, A.; Muñoz Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Rozum, I.; et al. *ERA5 Hourly Data on Single Levels from 1940 to Present*; European Environment Agency (EEA): Copenhagen K, Denmark, 2023.
2. Garcia-Prats, A.; Carricondo-Antón, J.M.; Ippolito, M.; De Caro, D.; Jiménez-Bello, M.A.; Manzano-Juárez, J.; Pulido-Velazquez, M. High-Resolution Spatially Interpolated FAO Penman-Monteith Crop Reference Evapotranspiration Maps of Sicily Island (Italy) and Jucar River System (Spain) Using AgERA5 and ERA5-Land Reanalysis Datasets. *J. Hydrol. Reg. Stud.* **2025**, *60*, 102531. [CrossRef] [PubMed]
3. Ippolito, M.; De Caro, D.; Cannarozzo, M.; Provenzano, G.; Ciraolo, G. Evaluation of Daily Crop Reference Evapotranspiration and Sensitivity Analysis of FAO Penman-Monteith Equation Using ERA5-Land Reanalysis Database in Sicily, Italy. *Agric. Water Manag.* **2024**, *295*, 108732. [CrossRef]
4. Vanella, D.; Longo-Minnolo, G.; Belfiore, O.R.; Ramírez-Cuesta, J.M.; Pappalardo, S.; Consoli, S.; D'Urso, G.; Chirico, G.B.; Coppola, A.; Comegna, A.; et al. Comparing the Use of ERA5 Reanalysis Dataset and Ground-Based Agrometeorological Data under Different Climates and Topography in Italy. *J. Hydrol. Reg. Stud.* **2022**, *42*, 101182. [CrossRef]
5. Buontempo, C.; Hutjes, R.; Beavis, P.; Berckmans, J.; Cagnazzo, C.; Vamborg, F.; Thépaut, J.N.; Bergeron, C.; Almond, S.; Amici, A.; et al. Fostering the Development of Climate Services through Copernicus Climate Change Service (C3S) for Agriculture Applications. *Weather Clim. Extrem.* **2020**, *27*, 100226. [CrossRef]
6. Chevuru, S.; de Wit, A.; Supit, I.; Hutjes, R. Copernicus Global Crop Productivity Indicators: An Evaluation Based on Regionally Reported Yields. *Clim. Serv.* **2023**, *30*, 100374. [CrossRef]
7. Schönauer, M.; Asabere, S.B.; Sauer, D.; Drollinger, S. Topographic Indices and ERA5-Land Data to Describe Soil Moisture Variability in a Central European Beech Forest. *J. Hydrol. Reg. Stud.* **2025**, *59*, 102456. [CrossRef]
8. Chebbi, W.; Rubio, E.; Markos, N.; Yakir, D.; García-Morote, F.A.; Andrés-Abellán, M.; Arquero-Escañuela, R.; Picazo-Córdoba, M.I.; Rotenberg, E.; Radoglou, K.; et al. Refinement of Gross Dry Matter Productivity (GDMP) Product from Copernicus Land Monitoring Service (CLMS): An Ecophysiological Assessment of Mediterranean Evergreen Forests. *Remote Sens. Environ.* **2025**, *328*, 114856. [CrossRef]
9. Fraccaroli, C.; Marini Govigli, V.; Briers, S.; Peña Cerezo, N.; Paz Jiménez, J.; Romero, M.; Lindner, M.; Martínez de Arano, I. Climate Data for the European Forestry Sector: From End-User Needs to Opportunities for Climate Resilience. *Clim. Serv.* **2021**, *23*, 100247. [CrossRef]
10. Taszarek, M.; Kendzierski, S.; Pilguj, N. Hazardous Weather Affecting European Airports: Climatological Estimates of Situations with Limited Visibility, Thunderstorm, Low-Level Wind Shear and Snowfall from ERA5. *Weather Clim. Extrem.* **2020**, *28*, 100243. [CrossRef]
11. Yu, P.; Xu, R.; Zhang, Y.; Ye, T.; Yu, W.; Chen, G.; Wen, B.; Yue, X.; Li, S.; Guo, Y. Spatiotemporal Variations and Inequalities in Global Traffic-Sourced PM2.5 Exposure. *Environ. Int.* **2025**, *199*, 109492. [CrossRef] [PubMed]
12. Su, L.; Xie, Z.; Yong, M.; Wu, C.; Cheng, Z.; Purevjav, G.; Bueh, C.; Bao, Y. Influence of Cut-off Lows on Dust Transport from the Great Lakes Basin to Northern China. *Glob. Planet. Change* **2025**, *247*, 104738. [CrossRef]
13. Meadows, M.; Reinke, K.; Jones, S. Explaining Machine Learning Models Trained to Predict Copernicus DEM Errors in Different Land Cover Environments. *Artif. Intell. Geosci.* **2025**, *6*, 100141. [CrossRef]
14. Liu, L.; Yi, Y.; Jiang, H.; Ran, Y.; Chen, D. ERA5-Land Overestimates Runoff Coefficient but Underestimates Runoff Recession Rate in the Central Tibetan Permafrost Region. *J. Hydrol. Reg. Stud.* **2024**, *53*, 101792. [CrossRef]
15. Liu, Z.; Filhol, S.; Treichler, D. Retrieving Snow Depth Distribution by Downscaling ERA5 Reanalysis with ICESat-2 Laser Altimetry. *Cold Reg. Sci. Technol.* **2025**, *239*, 104580. [CrossRef]
16. Castant, J.; Vantrepotte, V.; Frouin, R.; Beaugrand, G. Comprehensive Gridded Dataset of Photosynthetically Active Radiation in the Upper Ocean from 1958 to 2022. *Remote Sens. Environ.* **2024**, *311*, 114305. [CrossRef]
17. Unidata Program Center. NetCDF Credits. Available online: https://docs.unidata.ucar.edu/netcdf-c/current/credits.html (accessed on 29 October 2025).
18. National Science Foundation. Unidata Software for Manipulating or Displaying NetCDF Data. Available online: https://www.unidata.ucar.edu/software/netcdf (accessed on 23 August 2025).
19. Arnaut, F.; Jevremović, S.; Kolarski, A. *NcPick- A Toolkit for Extracting, Analyzing, and Visualizing Copernicus ERA5 NetCDF Reanalysis Climate Datasets—User Manual & Testing Documentation*; Belgrade, Serbia; 2025. Available online: https://arnautf.github.io/ncpick_website/docs/ncPick%20v1.0-%20User%20manual.pdf (accessed on 3 September 2025).

20. Arnaut, F.; Jevremović, S.; Kolarski, A.; Srećković, V.A. NcPick. Environmental Data. Simplified. Available online: https://arnautf.github.io/ncpick_website/ (accessed on 25 October 2025).
21. Siegel, A. Chapter 14—Time Series: Understanding Changes over Time. In *Practical Business Statistics*, 6th ed.; Elsevier: Amsterdam, The Netherlands, 2012; pp. 429–464.
22. Gudivada, V.N. *Data Analytics: Fundamentals*; Elsevier Inc.: Amsterdam, The Netherlands, 2017; ISBN 9780128098516.
23. Li, L.; Zha, Y. Mapping Relative Humidity, Average and Extreme Temperature in Hot Summer over China. *Sci. Total Environ.* **2018**, *615*, 875–881. [CrossRef] [PubMed]
24. Geocene Inc. Welcome to TRAINSET. Available online: https://trainset.geocene.com/ (accessed on 29 October 2025).
25. National Aeronautics and Space Administration & Goddard Institute for Space Studies. Panoply NetCDF, HDF and GRIB Data Viewer. Available online: https://www.giss.nasa.gov/tools/panoply/ (accessed on 26 August 2025).