



## Data Article

# Daily dataset of precipitation and temperature in the Department of Cauca, Colombia



Kevin Blanco<sup>a,\*</sup>, Sandra R. Villamizar<sup>a</sup>, Alvaro Avila-Diaz<sup>b,c</sup>, Catalina Marceló-Díaz<sup>d</sup>, Erika Santamaría<sup>d</sup>, María Camila Lesmes<sup>c,d</sup>

<sup>a</sup> Universidad Industrial de Santander, Carrera 27 Calle 9, Bucaramanga, Postal Code 680002, Colombia

<sup>b</sup> Research Group "Interactions Climate-Ecosystems (ICE)", Earth System Science Program, Faculty of Natural Sciences, Universidad del Rosario, Carrera 24 #63C-69, Bogotá, Postal Code 111221, Colombia

<sup>c</sup> Universidad de Ciencias Aplicadas y Ambientales, Calle 222 #55-37, Bogotá, Postal Code 111166, Colombia

<sup>d</sup> Grupo de Entomología, Instituto Nacional de Salud, Avenida Calle 26 #51-20, Bogotá, Postal Code 111321, Colombia

## ARTICLE INFO

## Article history:

Received 13 July 2023

Revised 28 August 2023

Accepted 29 August 2023

Available online 3 September 2023

Dataset link: [Daily Database of Precipitation and Temperature in the Department of Cauca, Colombia \(Original data\)](#)

## Keywords:

Spatial downscaling

ERA5-Land

CHIRPS

MSWX

Kriging

## ABSTRACT

This study used the geostatistical Kriging methodology to reduce the spatial scale of a host of daily meteorological variables in the Department of Cauca (Colombia), namely, total precipitation and maximum, minimum, and average temperature. The objective was to supply a high-resolution database from 01/01/2015 to 31/12/2021 in order to support the climate component in a project led by the National Institute of Health (INS) named "Spatial Stratification of dengue based on the identification of risk factors: a pilot study in the Department of Cauca". The scaling process was applied to available databases from satellite information and reanalysis sources, specifically, CHIRPS (Climate Hazards Group InfraRed Precipitation with Station Data), ERA5-Land (European Centre for Medium-Range Weather Forecasts), and MSWX (Multi-Source Weather). The 0.1° resolution offered by both the MSWX and ERA5-Land databases and the 0.05° resolution found in CHIRPS, was successfully reduced to a scale of 0.01° across all variables. Statistical metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Person Correlation Coefficient ( $r$ ), and Mean Bias Error (MBE) were used to select the database that best estimated each variable. As a result, it was determined that the scaled ERA5-Land database yielded

\* Corresponding author.

E-mail address: [kevin2128275@correo.uis.edu.co](mailto:kevin2128275@correo.uis.edu.co) (K. Blanco).

the best performance for precipitation and minimum daily temperature. On the other hand, the scaled MSWX database showed the best behavior for the other two variables of maximum temperature and daily average temperature. Additionally, using the scaled meteorological databases improved the performance of the regression models implemented by the INS for constructing a dengue early warning system.

© 2023 The Author(s). Published by Elsevier Inc.  
This is an open access article under the CC BY license  
(<http://creativecommons.org/licenses/by/4.0/>)

## Specifications Table

Subject	Environmental Science
Specific subject area	Application of Kriging methodology for the scale reduction of climatological data
Type of data	Downscaled datasets of precipitation and temperature for the department of Cauca (Colombia) at 0.01° resolution.
How the data were acquired	The scaled data was obtained by applying the geostatistical Kriging method to the National Oceanic and Atmospheric Administration's (NOAA) CHIRPS database (precipitation), the ERA5-Land database (precipitation and temperature) from the Climate Data Store of the European Union's Earth Observation Programme, "Copernicus", and finally, MSWX from the GloH2O web portal (precipitation and temperature).
Data format	Analyzed, nc (NetCDF) files
Description of data collection	The downscaled datasets were generated using the "KrigR" R-package, which uses the "automap" R package to carry out the Kriging downscaling process. KrigR was adapted in its three-step process to work for the three climate datasets by defining the target variable (precipitation or temperature), geographic area (-78° to -75.6° latitude and 0.9° to 3.4° longitude), time period (2015-2021) and temporal resolution (daily). A DEM in raw-data resolution and target resolution (0.01°) was used as covariate for the process.
Data source location	<ul style="list-style-type: none"> <li>· Region: Cauca Department</li> <li>· Country: Colombia</li> <li>· Latitude and longitude for collected samples/data: -78° to -75.6° and 0.9° to 3.4°.</li> <li>· The raw climatic datasets are accessible on the following websites:               <ul style="list-style-type: none"> <li>• CHIRPS: <a href="https://coastwatch.pfeg.noaa.gov/erddap/griddap/chirps20GlobalDailyP05.html">https://coastwatch.pfeg.noaa.gov/erddap/griddap/chirps20GlobalDailyP05.html</a></li> <li>• ERA5-Land: <a href="https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview">https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview</a></li> <li>• MSWX: <a href="https://www.gloh2o.org/mswx/">https://www.gloh2o.org/mswx/</a></li> <li>• IDEAM: <a href="http://dhime.ideam.gov.co/atencionciudadano/">http://dhime.ideam.gov.co/atencionciudadano/</a></li> </ul> </li> </ul>
Data accessibility	Repository name: HydroShare Direct URL to data: <a href="https://www.hydroshare.org/resource/3dbf521c03134a19a138d192bbe676d0/">https://www.hydroshare.org/resource/3dbf521c03134a19a138d192bbe676d0/</a> DOI: <a href="https://doi.org/10.4211/hs.3dbf521c03134a19a138d192bbe676d0">10.4211/hs.3dbf521c03134a19a138d192bbe676d0</a>

## 1. Value of the Data

- The scaled database allowed the creation of high-resolution (0.01°) meteorological maps, which permit the description of the environmental conditions that promote the reproduction and survival of dengue vectors such as the *Aedes aegypti*, through ecological niche modeling.
- This information benefits institutions responsible for the surveillance and control of dengue, to avoid its propagation in the Department of Cauca.
- The daily downscaled data may be used in any study requiring high-resolution climatic information (precipitation and maximum, minimum, or average temperature).

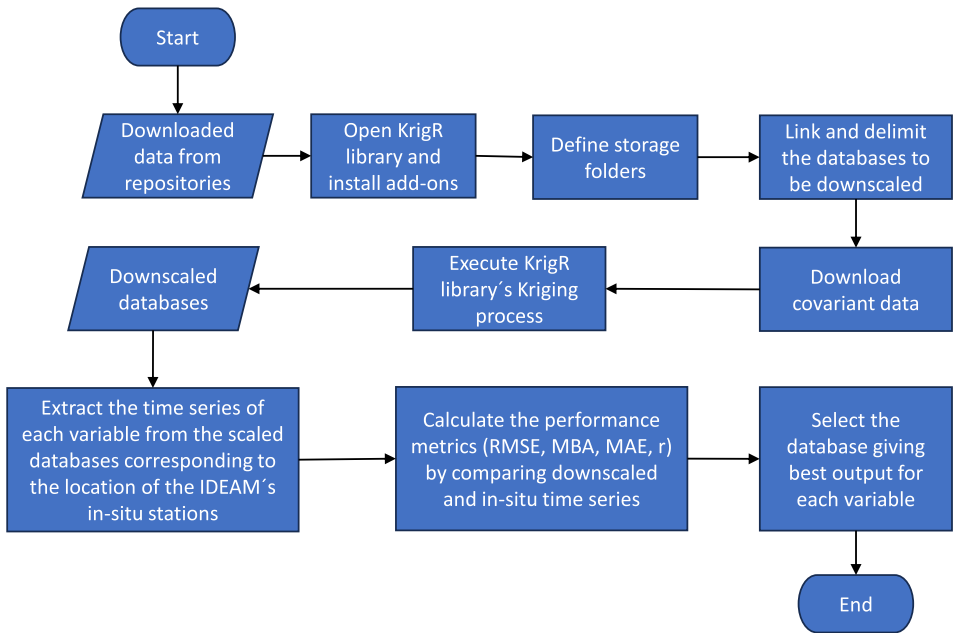


Fig. 1. Scale reduction flowchart.

## 2. Objective

A high-resolution database ( $0.01^\circ$ ) of daily climatic information (precipitation maximum temperature, minimum temperature, and average temperature) was created through the geostatistical Kriging method to support the National Institute of Health (INS) in the creation of an early warning system to monitor the risk of dengue in the municipalities of Patía, Piamonte, and Miranda, located in the department of Cauca, Colombia.

## 3. Data Description

The Department of Cauca's scaled databases were generated by applying the geostatistical Kriging methodology to the CHIRPS, ERA5-Land, and MSWX databases. Fig. 1 shows a flow chart diagram that describes the method utilized in the scaling and evaluation processes and the selection of the high-resolution database that yielded the best performance in estimating each variable.

Table 1 shows the performance metrics used for the evaluation of the proficiency of the high-resolution databases in the estimation of the variables studied, in comparison to the in-situ data available from the network stations of the Colombian Institute of Hydrology, Meteorology and Environmental Studies (IDEAM).

Where  $Q_i$  represents the observed values for each meteorological variable (total precipitation, maximum, minimum, or daily average temperature) at one station;  $P_i$  represents the values for each meteorological variable from the original or scaled databases (CHIRPS, ERA5-Land or MSWX); and  $N$  is the amount of data evaluated.

Figs 3-5 show examples of scaled data for maximum, minimum, and daily average temperature from the ERA5-Land and MSWX databases compared to their original resolutions. Fig. 6 shows an example of the scaled data for daily precipitation from the ERA5-Land, MSWX,

**Table 1**  
Estimated statistical metrics.

Name	Equation	Units	Perfect Score
Root Mean Square Error (RMSE)	$RMSE = \sqrt{\left(\frac{1}{N}\right) \sum_{i=1}^N  P_i - Q_i ^2}$	Precipitation (mm) Temperature (°C)	0.0
Mean Absolute Error (MAE)	$MAE = \frac{\sum_{i=1}^N  P_i - Q_i }{N}$	Precipitation (mm) Temperature (°C)	0.0
Pearson Correlation Coefficient (r)	$r = \frac{(N \sum_{i=1}^N (P_i \cdot Q_i) - (\sum_{i=1}^N P_i)(\sum_{i=1}^N Q_i))}{\sqrt{(N \sum_{i=1}^N P_i^2 - (\sum_{i=1}^N P_i)^2)(N \sum_{i=1}^N Q_i^2 - (\sum_{i=1}^N Q_i)^2)}}$	-	1.0
Mean Bias Error (MBE)	$MBE = \frac{1}{N} \sum_{i=1}^N  P_i - Q_i $	Precipitation (mm) Temperature (°C)	0.0

and CHIRPS databases. Lastly, Figs 7-10 present the results obtained for the estimated performance metrics.

### 4. Experimental Design, Materials and Methods

#### 4.1. Description of the study area

In southwest Colombia, the Cauca department covers a surface area of 29,308 km<sup>2</sup> (Fig. 2). Five large watersheds can be found within its territory: High Cauca, Pacific, High Magdalena,

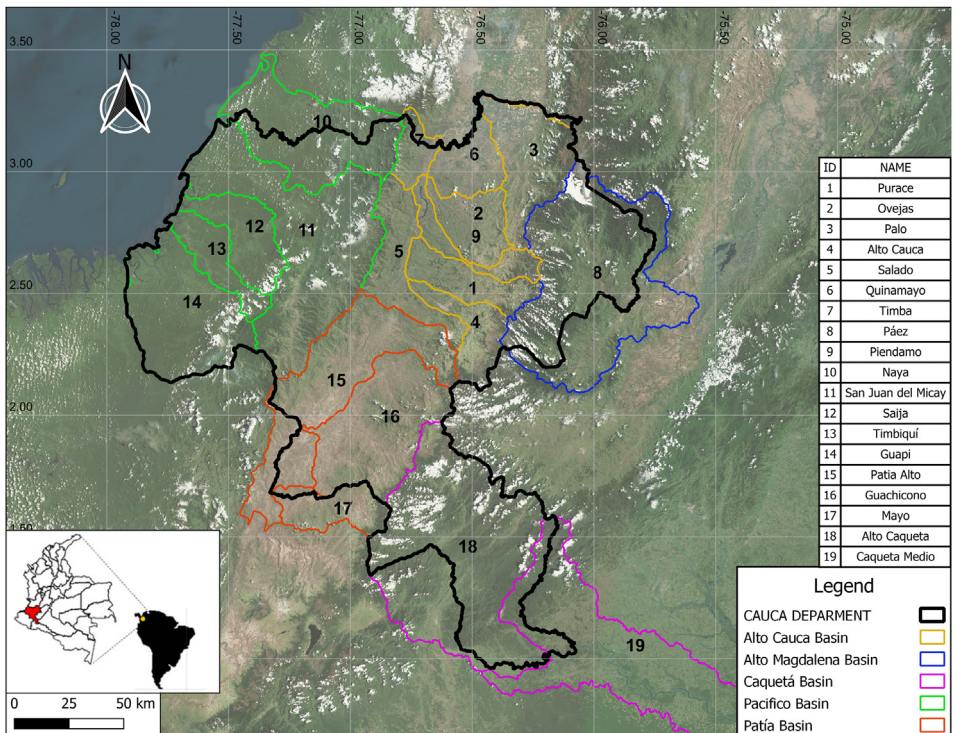


Fig. 2. Study area location.

Patía, and Caquetá. With an average altitude of 1693 m above sea level, the department presents a topographic profile that varies from 5750 m above sea level to sea level, which allows it to encompass all thermal floors. One can find a warm and humid climate in the Pacific coast to the west, while in the Patía watershed, the weather is warm and semiarid. In the central region, temperate humid and semi-humid climates predominate. The Popayán plateau has cold climates, and the *Nevado del Huila* sees perpetual snow. This diversity of thermal floors results from the altitudinal gradient that spans from the coastline to the most elevated mountainous zones [1].

#### 4.2. Raw databases

The National Oceanic and Atmospheric Administration (NOAA) web portal was used to download the CHIRPS data [2]. This platform offers precipitation data which is derived from infrared observations of cold cloud durations; the pixel size of this dataset is 0.05°. The portal permits to define the time window, the temporal resolution, and the geographical area of interest, which, in our case, were the coordinates that encompass the department of Cauca (0.9° to 3.4° latitude and -78° to -75.6° longitude).

To access the data from the ERA5-Land database, a user account was created on the Climate Data Store website, which is part of the European Union's Earth Observation Program, ("Copernicus") [3]. ERA5-Land employs a reanalysis process integrating data from climatological models with global observations. The pixel size of this dataset is 0.1°. We downloaded precipitation data and maximum, minimum, and average temperature data using the "download\_ERA" function from the "KrigR" library in R, through the RStudio interface. This function allows to configure the time window, the temporal resolution, and the geographical area of interest within the script.

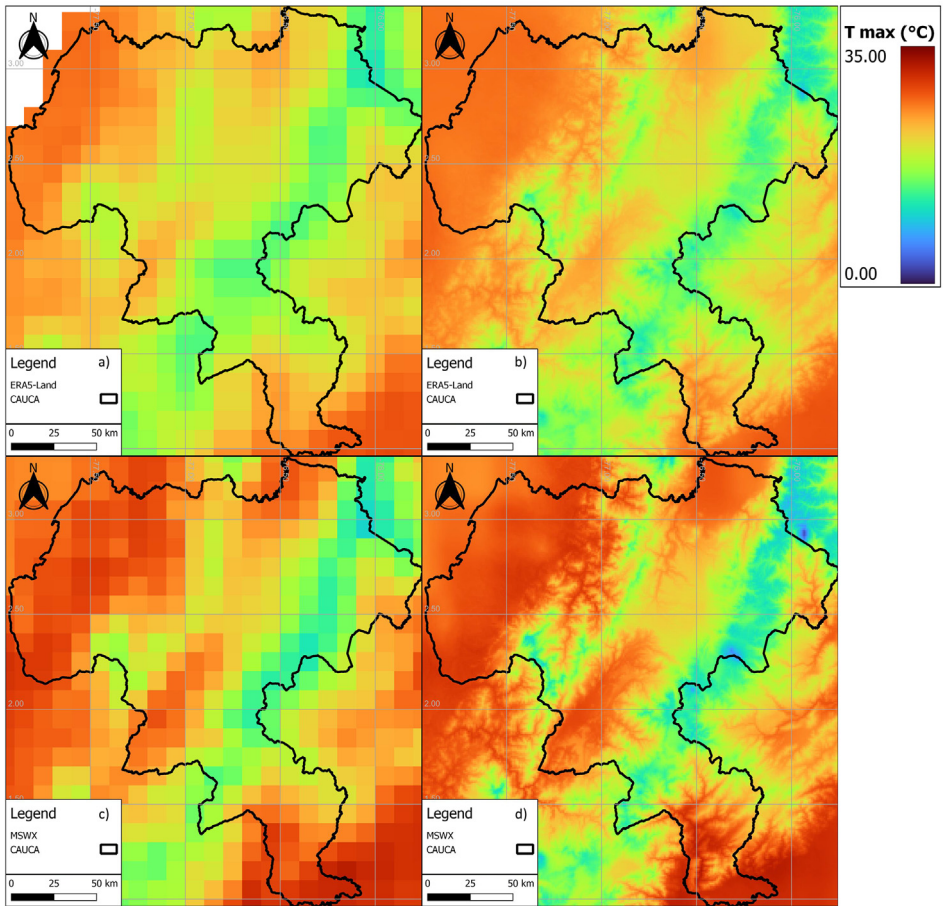
Lastly, before downloading the MSWX data (precipitation data, maximum, minimum, and average temperature data), it was necessary to complete the license request process on the GloH2O website [4], which is the institution responsible for developing MSWX. This dataset, with 0.1°-pixel size, consists of data generated from ERA5 that was corrected using high-resolution reference climatology. After the license request, an e-mail containing an access link to a Google Drive folder housed the data was received. Post-processing using CDO (Climate Data Operators) [5] allowed the extraction of the geographical area of interest (the raw files are for the entire globe) and to integration de individually downloaded daily files into the complete time window of the study.

#### 4.3. Methodology

Kriging is a geostatistical method used for estimating spatial values in a specified area. It is based on the analysis of averages, covariance functions and variograms of the data available in said area. The method is widely used because it effectively allows interpolating data to finer resolutions by incorporating covariate information that helps determine the uncertainty associated with the downscaling [6–8]. Furthermore, this method ensures a notable reproducibility level, guaranteeing result consistency and the ability to replicate findings. There are three main forms of linear Kriging: simple Kriging, which utilizes linear estimators with a known mean and a covariance function; ordinary Kriging, which is used when there is no known average and utilizes only covariance functions; and universal Kriging, which employs variograms to construct the estimator [9]. This project used the universal Kriging to carry out the estimations.

$$\sum_{\beta} \lambda_{\beta} \gamma_{\alpha\beta} + \sum_l \mu_l f_{\alpha}^l = \gamma_{\alpha 0} \quad \alpha = 1, \dots, N$$

$$\sum_{\alpha} \lambda_{\alpha} f_{\alpha}^l = f_0^l \quad l = 0, \dots, L$$



**Fig. 3.** Daily maximum temperature for 01/17/2015. (a) ERA5-Land unscaled (0.1°); (b) ERA5-Land scaled (0.01°); (c) MSWX unscaled (0.1°); (d) MSWX scaled (0.01°).

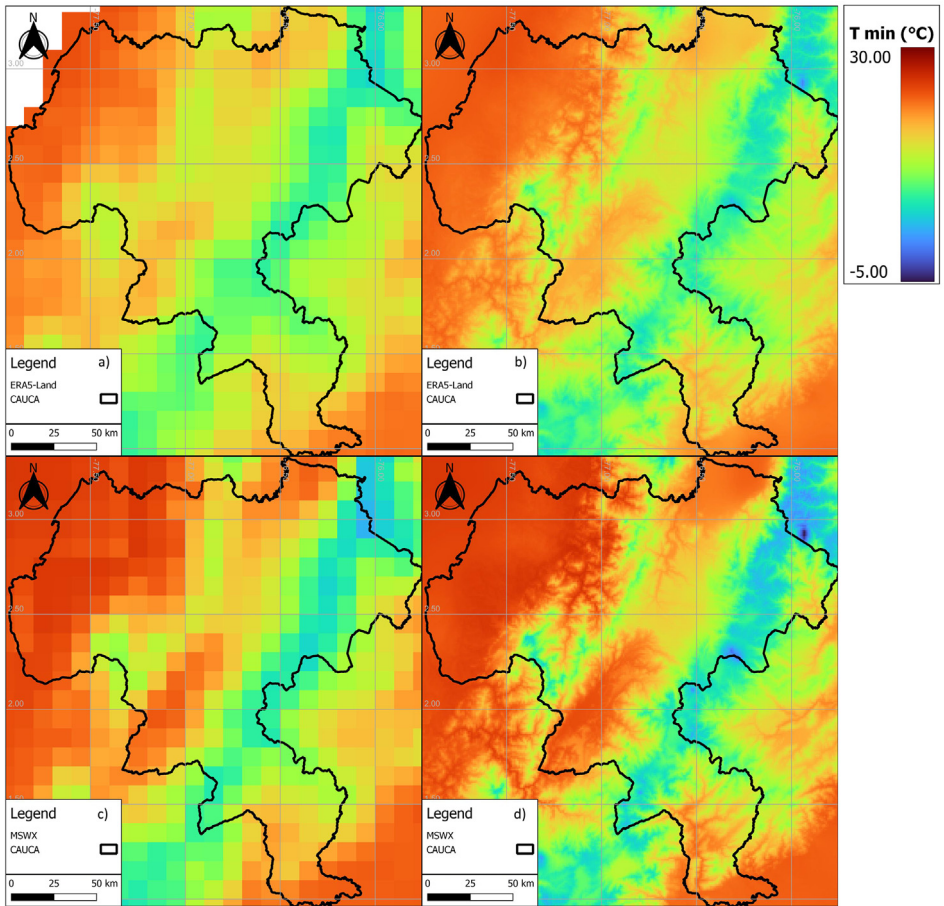
Where  $\lambda\beta$  is the weight of the dataset at point  $\beta$ ,  $\gamma\alpha\beta$  is the variogram between the datasets at points  $\alpha$  and  $\beta$ ,  $f_{\alpha}^l$  are the known base functions of the dataset at point  $\alpha$ ,  $\mu_l$  are  $L + 1$  unknown variables used utilizing Lagrange multipliers,  $\gamma\alpha 0$  is the variogram between the dataset at point  $\alpha$  and the dataset at point 0,  $\lambda\alpha$  is the weight of dataset  $\alpha$ , and  $f_0^l$  are known base functions of the dataset at point 0 [9].

A variogram can be mathematically defined as “Half of the mathematical expectancy of the square of the difference of two values of the study variable  $Z$ , corresponding to positions separated by a distance of  $h$ ” [10].

$$\gamma(h) = \frac{1}{2}E[Z(X) - Z(X + h)]^2$$

Where  $E[Z(x)]$  is the mathematical expectancy operator,  $X$  is a vector of coordinates, and  $h$  is the distance that separates the two points where variable  $Z$  is measured [10].

The implementation of the geostatistical Kriging method in the scale reduction of the study’s climatological variables of interest used the krigR function (multi-core Kriging) of the KrigR library found within R. This function requires training data which are to be scaled, along with covariable data in two resolutions: the resolution of the training data and the target resolution



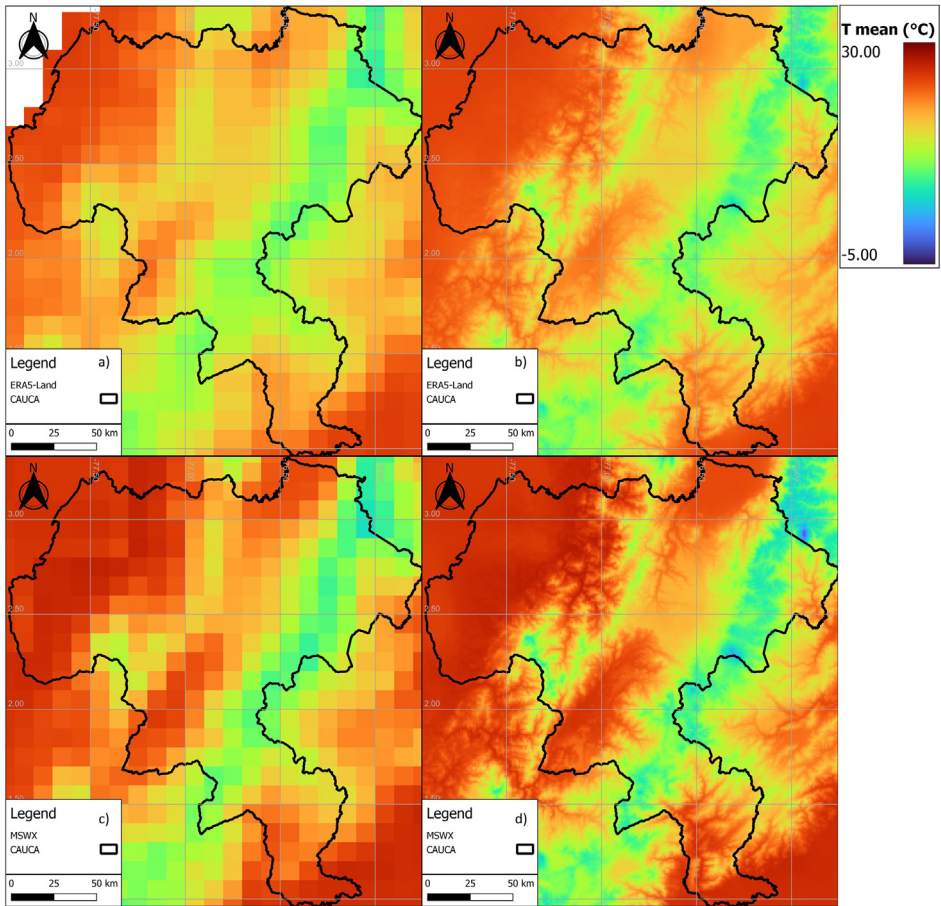
**Fig. 4.** Daily minimum temperature for 01/17/2015. (a) ERA5-Land unscaled ( $0.1^\circ$ ); (b) ERA5-Land scaled ( $0.01^\circ$ ); (c) MSWX unscaled ( $0.1^\circ$ ); (d) MSWX scaled ( $0.01^\circ$ ).

to be reached by the scaling process. From this information, the scaling procedure is carried out in two steps; first, the tool adjusts variograms to the training dataset and establishes functions of covariance with the covariable data in the resolution of the training data, generating functions that make it possible to describe the spatial correlation between the training data and the covariables [6]; second, the prediction of the values of the variable in the new location is carried out, parting from the covariable data in the target resolution [11]. In the process of scaling the databases, the digital elevation model (DEM) ( $0.01^\circ$  resolution) from the United States Geological Service (USGS) was utilized as covariant information [12].

#### 4.4. Validation

The scaled databases were validated by using a comparison between the downscaled data and the in-situ measurements at the local IDEAM station network [13].

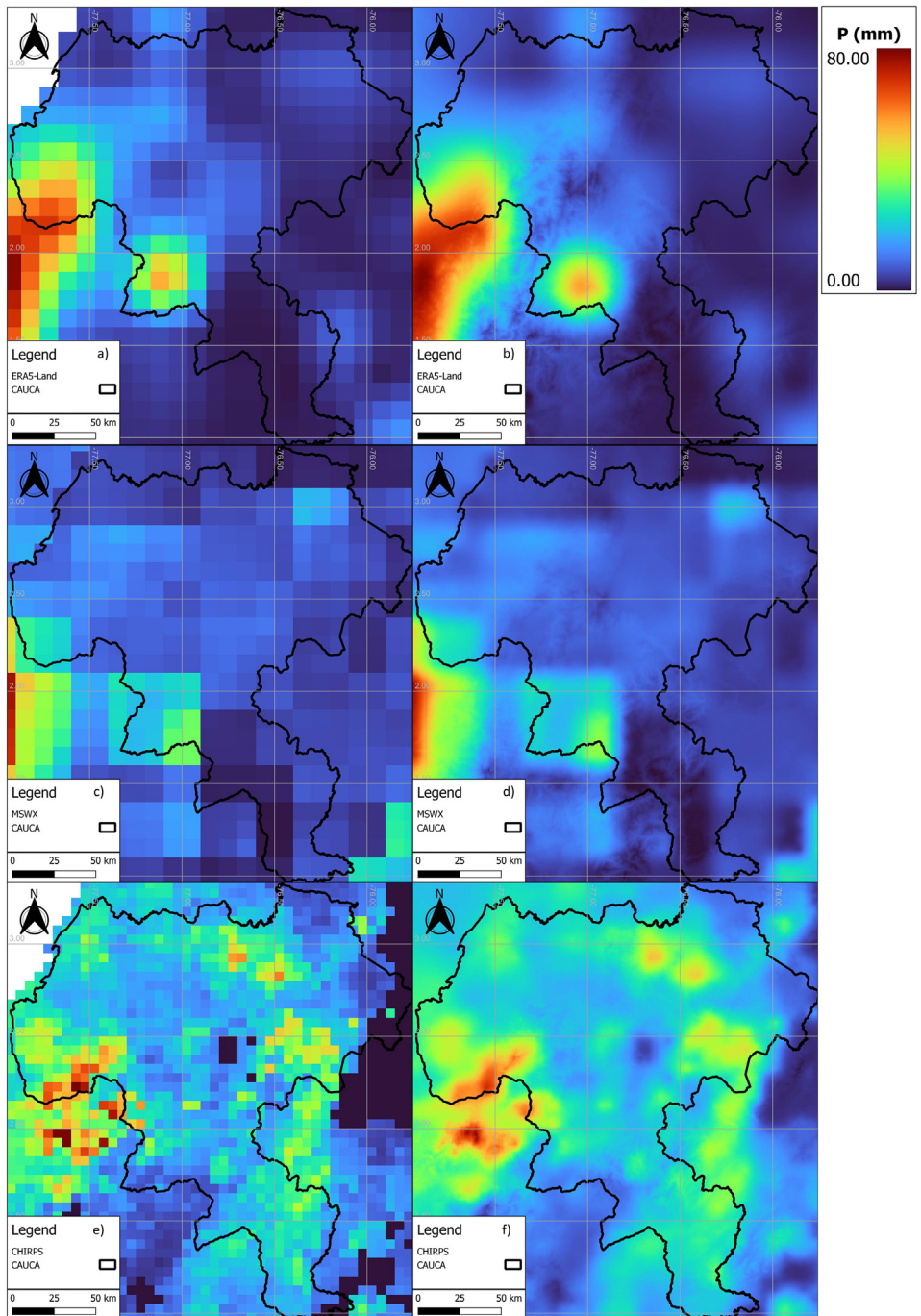
Obtaining the in-situ measurements for the comparison involved consulting and downloading the information available at the local stations in the department of Cauca for each of the



**Fig. 5.** Daily average temperature for 01/17/2015. (a) ERA5-Land unscaled ( $0.1^\circ$ ); (b) ERA5-Land scaled ( $0.01^\circ$ ); (c) MSWX unscaled ( $0.1^\circ$ ); (d) MSWX scaled ( $0.01^\circ$ ).

meteorological variables of interest in the period between 01/01/2015 and 31 /12/2021 (which corresponds to the period of the study carried out by the INS). The data initially downloaded in comma-separated values (.csv) format contained daily values for total precipitation and maximum, minimum, and average temperatures. Subsequently, the data sets were processed through Microsoft Excel to identify and eliminate erroneous data, specifically, values that had typographical errors, and to verify the temporal continuity of the time series, which allowed the identification of a total of 12 stations for the analysis. validation of precipitation data, 13 stations for the maximum temperature variable, 14 stations for minimum temperature and only one station for the case for the average temperature. It should be noted that, since the IDEAM data was used to assess performance in the original and scaled databases, the missing data-filling procedure was not performed. This was done to avoid introducing errors associated with this procedure to the estimated metrics (Table 1) for the evaluation of this study.

From the computed performance metrics (RMSE, MAE,  $r$  & MBE), it was determined that the scaled ERA5-Land databases yielded the best performance for the variables of precipitation and daily minimum temperature. On the other hand, the scaled MSWX databases obtained the best results for daily maximum temperature and daily average temperature (Fig. 10).



**Fig. 6.** Daily total precipitation for 01/17/2015. (a) ERA5-Land (m) unscaled ( $0.1^\circ$ ); (b) ERA5-Land (m) scaled ( $0.01^\circ$ ); (c) MSWX (mm) unscaled ( $0.1^\circ$ ); (d) MSWX (mm) scaled ( $0.01^\circ$ ); (e) CHIRPS (mm) unscaled ( $0.05^\circ$ ); (f) CHIRPS (mm) scaled ( $0.01^\circ$ ).

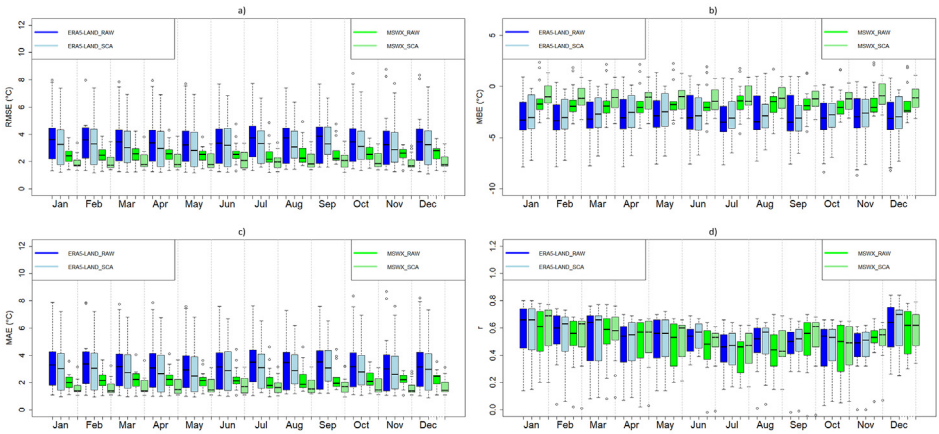


Fig. 7. Daily maximum temperature metrics for the original and scaled databases. (a) RMSE; (b) MBE; (c) MAE (d) r.

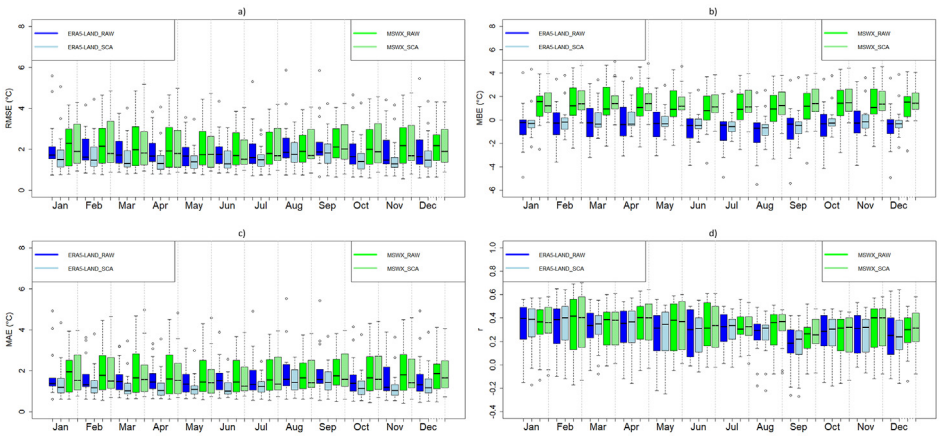


Fig. 8. Daily minimum temperature metrics for the original and scaled databases. (a) RMSE; (b) MBE; (c) MAE (d) r.

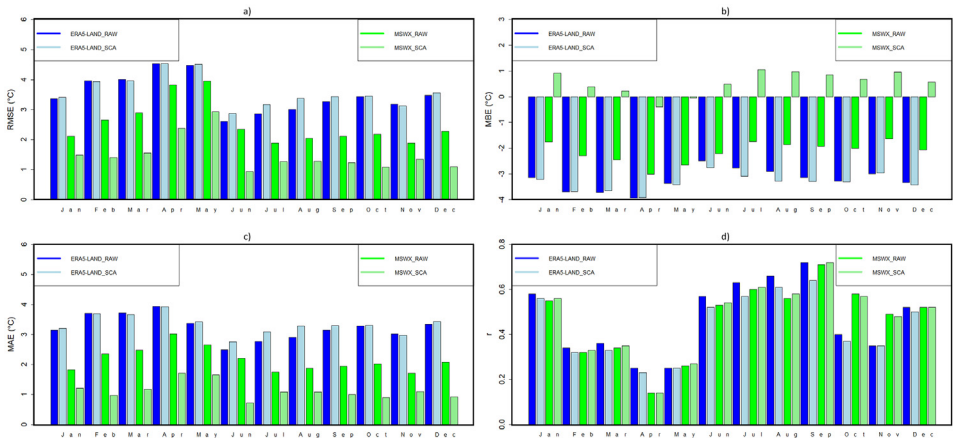
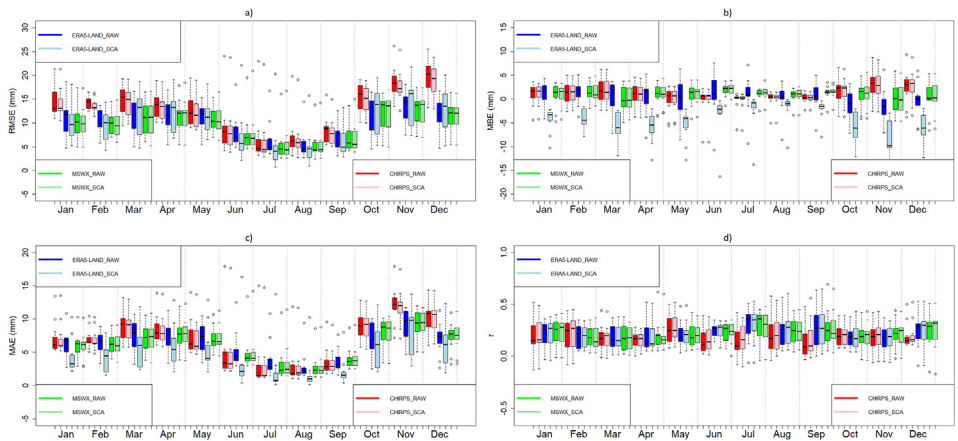


Fig. 9. Daily average temperature metrics for the original and scaled databases. (a) RMSE; (b) MBE; (c) MAE (d) r.



**Fig. 10.** Daily total precipitation metrics for the original and scaled databases. (a) RMSE; (b) MBE; (c) MAE (d)  $r$ .

## Ethics Statements

This paper adheres to the ethical requirements set forth by the journal Data in Brief for the publication of summary data.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Daily Database of Precipitation and Temperature in the Department of Cauca, Colombia (Original data) (HYDROSHARE).

## CRediT Author Statement

**Kevin Blanco:** Conceptualization, Methodology, Software, Data curation, Writing – original draft, Visualization; **Sandra R. Villamizar:** Conceptualization, Methodology, Software, Data curation, Visualization, Writing – review & editing; **Alvaro Avila-Diaz:** Conceptualization, Methodology, Software, Data curation, Visualization, Writing – review & editing; **Catalina Marceló-Díaz:** Supervision, Writing – review & editing; **Erika Santamaría:** Supervision, Writing – review & editing; **María Camila Lesmes:** Supervision, Writing – review & editing.

## Acknowledgements

Financing: This research was funded by the National Institute of Health, Bogotá, Colombia (research project [CEMIN 13-2019](#)), the Secretaria de Salud Departamental del Cauca, the Universidad de Ciencias Aplicadas y Ambientales-UDCA and Minciencias (research project [210484467217](#)).

## References

- [1] Universidad del Valle Perfil Epidemiológico De Muertes Por Causa Externa, 2023 <https://prevencionviolencia.univalle.edu.co/observatorios/cauca/departamental/departamental.html> (Accessed 10 August 2023).
- [2] C. Funk, P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell, J. Michaelsen, The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, *Sci. Data* 2 (2015) 150066, doi:10.1038/sdata.2015.66.
- [3] Copernicus Climate Change Service ERA5-Land Hourly Data from 2001 to Present, 2019, doi:10.24381/CDS.E2161BAC.
- [4] H.E. Beck, A.I.J.M. van Dijk, P.R. Larraondo, T.R. McVicar, M. Pan, E. Dutra, D.G. Miralles, MSWX: global 3-hourly 0.1° bias-corrected meteorological data including near real-time updates and forecast ensembles, *Bull. Am. Meteorol. Soc.* 1 (2021) 1–55, doi:10.1175/BAMS-D-21-0145.1.
- [5] Max-Planck-Institut für Meteorologie Vistazo - CDO - Project Management Service, 2023 <https://code.mpimet.mpg.de/projects/cdo/> (Accessed 11 July 2023).
- [6] E. Kusch, R. Davy, KrigR—a tool for downloading and statistically downscaling climate reanalysis data, *Environ. Res. Lett.* 17 (2022) 024005, doi:10.1088/1748-9326/ac48b3.
- [7] N. Zhao, A Method for Merging Multi-Source Daily Satellite Precipitation Datasets and Gauge Observations over Poyang Lake Basin, 15, *Remote Sens. China*, 2023, doi:10.3390/rs15092407.
- [8] K. Alsafadi, S. Bi, B. Bashir, E. Sharifi, A. Alsalman, A. Kumar, S. Shahid, High-resolution precipitation modeling in complex terrains using hybrid interpolation techniques: incorporating physiographic and MODIS cloud cover influences, *Remote Sens.* 15 (2023), doi:10.3390/rs15092435.
- [9] J.-P. Chilès, P. Delfiner, *Geostatistics: Modeling Spatial Uncertainty: Second Edition*, 2012, doi:10.1002/9781118136188.
- [10] J.A.P. Oropeza, E.P. Vélez, O.P. Vélez, Ajuste de un variograma esférico de la precipitación anual de las normales climatológicas 1951-2010 - Adjustment of a spherical variogram of the annual precipitation of the Climatological Normals 1951-2010, *Tecnol. Cienc. Agua*. 9 (2018) 209–222, doi:10.24850/j-tyca-2018-06-10.
- [11] R. Davy, E. Kusch, Reconciling high resolution climate datasets using KrigR, *Environ. Res. Lett.* 16 (2021) 124040, doi:10.1088/1748-9326/ac39bf.
- [12] USGS.gov | Science for a changing world, (2022). <https://www.usgs.gov/> (Accessed 13 December 2022).
- [13] IDEAM Consulta y Descarga de Datos Hidrometeorológicos, 2022 <http://dhime.ideam.gov.co/atencionciudadano/> (Accessed 23 January 2022).