Zonal statistics datasets of climate indicators for Brazilian municipalities

Raphael Saldanha, Reza Akbarinia, Marcel Pedroso, Victor Ribeiro, Carlos
Cardoso, Eduardo Pena, Patrick Valduriez, Fabio Porto

CAMBRIDGE
UNIVERSITY PRESS

DATA PAPER

Zonal statistics datasets of climate indicators for Brazilian municipalities

Raphael Saldanha[1] , Reza Akbarinia[1], Marcel Pedroso[2], Victor Ribeiro[3], Carlos Cardoso[3],
Eduardo H. M. Pena[3], Patrick Valduriez[1,3] and Fabio Porto[3]

[1]Inria, University of Montpellier, CNRS, LIRMM, Montpellier, Occitanie, France
[2]PCDaS, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil
[3]DEXL Lab, Laboratório Nacional de Computação Científica, LNCC, Petrópolis, Rio de Janeiro, Brazil
Corresponding author: Raphael Saldanha; Email: raphael.de-freitas-saldanha@inria.fr

Abstract

Climate trends and weather indicators are used in several research fields due to their importance in statistical modeling, frequently used as covariates. Usually, climate indicators are available as grid files with different spatial and time resolutions. The availability of a time series of climate indicators compatible with administrative boundaries is scattered in Brazil, not fully available for several years, and produced with diverse methodologies. In this paper, we propose time series of climate indicators for the Brazilian municipalities produced using zonal statistics derived from the ERA5-Land reanalysis indicators. As a result, we present datasets with zonal statistics of climate indicators with daily data, covering the period from 1950 to 2022.

Impact Statement

Climate indicators are important covariates on several research fields, but datasets directly compatible with administrative boundaries in Brazil are scarce and not available for long time spans. We propose consistent dataset of climate indicators for the Brazilian municipalities derived from the ERA5-Land project, covering the period from 1950 to 2022.

1. Introduction

Climate variables are commonly used as covariates in statistical modeling on research fields like agriculture, public health, and economics, as climate trends affect several biological and physical processes (Overpeck et al., 2011).

Climate indicators can be obtained by direct measurement, taken in place by a set of different equipment like weather stations and rain gauges, or by indirect measurement, with artificial satellites and remote sensing. The usage of in situ data is usually limited in space by the uneven distribution of sensors and in time considering the starting date of data recording and possible gaps due to technical

problems and other reasons. In Brazil, weather stations are unequally scattered over the territory and limited in number due to implementation and maintenance costs and also present data with different time spans and significant gaps for some locations. Statistical interpolations in time and space from weather stations data are often used to infer the weather indicators on areas and time intervals that were not covered by sensors (de Aguiar and Lobo, 2020).

In order to improve the availability of climate indicators in longer time spans and more refined spatial scales, data assimilation techniques have been developed and applied to produce climate reanalysis, which can be defined as the "consistent reprocessing of archived weather observations using a modern forecasting system" (Dee et al., 2014). By this approach, observations from multiple sources of direct and indirect climate measurements are combined with methodological procedures for calibration, correction, and cross-validation with field data (Funk et al., 2015; Abatzoglou et al., 2018; Muñoz-Sabater et al., 2021) to produce models that are able to retrospectively infer climate variables on global scale.

Climate reanalysis products provide continuous data with regular sets of indicators for long time spans and wide spatial coverage. The exceptional continuity and integrity make reanalysis products widely applied in some research fields, as hydrological reanalysis. On this specific use case, indicators are computed to coherent register water resource indicators on long time spans and large territorial areas, as the Brazilian basin (Wongchuig et al., 2019; Zou et al., 2022).

The approaches for measuring and registering weather indicators usually result in large datasets of different formats and extensions with multiple time and spatial resolutions, presenting a multitude of indicators covering extensive portions of territory in time series formats. This imposes a challenge to some research fields like economics and public health, where statistical modeling is carried out by using lattice structures (Cressie, 1991; Morgenstern, 1995). Those models employ aggregated data representing groups or geographical regions, like the total number of disease cases in a region or the average income of a neighborhood. Administrative spatial boundaries like countries, states, counties, and municipalities are some examples of lattice aggregation used in those models.

Considering the continuous nature of weather data, one common solution to introduce weather variables into lattice models is to aggregate weather indicators to the same lattice structure as the other variables present in the model. An example would be the modeling of dengue disease spread over administrative health regions: the number of dengue disease cases is usually presented in aggregated form, by the sum of confirmed cases in a health region. Thus, spatially continuous variables used as predictors of dengue disease, such as temperature and precipitation, must be aggregated to the same region's boundaries of the dengue case. An application of this approach can be found in the study by Lowe et al. (2016).

In order to summarize gridded climate indicators into boundary lattice structures, zonal statistics can be used. Zonal statistics are descriptive statistics calculated using the set of cells that spatially intersect a given boundary (Tomlin, 1994). For each boundary in a map of regions, statistics like average, maximum value, minimum value, standard deviation, and others are computed to statistically represent the cells that intersect the boundary. This is illustrated by Figure 1.

The zonal statistical indicators may also be weighted by the proportion of the cell that overlays the geographical boundaries and by other layers, as will be further explored. This work proposes the adoption of the ERA5-Land climate reanalysis to compute zonal statistics in order to provide climate indicators for the Brazilian municipalities for a long time span.

1.1. Related works

Similar procedures to summarize climate indicators for Brazilian municipalities have been done in other projects. The SISAM project (INPE, 2019) created a dataset with weather and air pollution indicators for Brazilian municipalities, with daily data from January 1, 2000, to December 31, 2019. Different data sources with different spatial resolutions for specific periods were used in this project. The ERA-Interim is used as a source but is currently supplanted by the newer ERA5-Land reanalysis (Muñoz-Sabater et al., 2021), limiting its utilization. The adopted methodology to create climate indicators for municipalities consisted of observing the data exclusively on the cell where the municipality city hall geographical
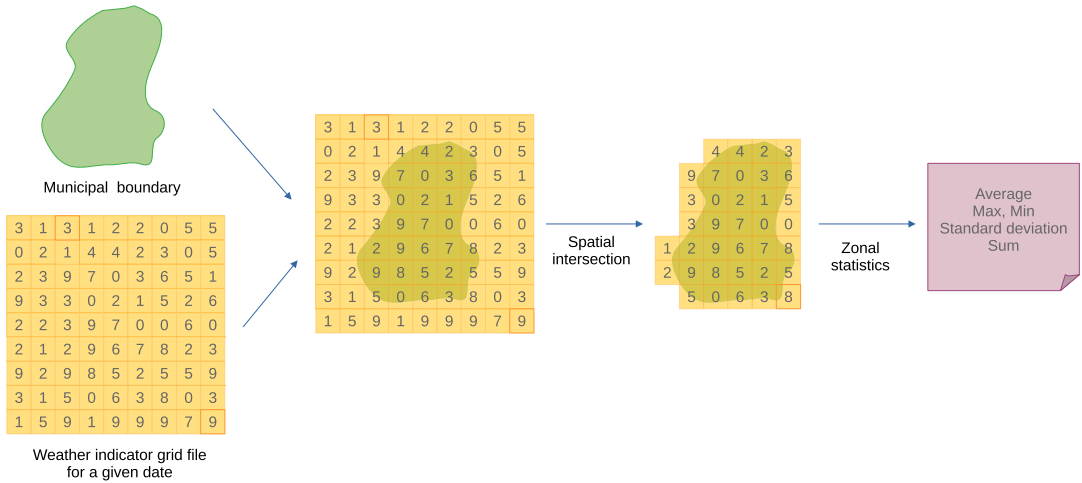
Figure 1. *Zonal statistics operation.*

coordinates are spatially located. This punctual observation, however, does not generalize to the municipality's territorial extension.

Abdalla et al. (2022) present datasets for Brazilian municipalities with different temporal resolutions and temporal coverages, including temperature and precipitation aggregated by the year, from 1981 to 2020. The adoption of yearly weather aggregation limits its usage.

It is also important to point that similar procedures are performed punctually as a methodological step by some researches to compute climate indicators on limited time spans to specific Brazilian municipalities or other boundaries in different countries (Lowe et al., 2011; Schneider et al., 2021; Abdalla et al., 2022). Usually, the adopted methodology is short described in the published papers and not directly reproducible. To the best of the authors' knowledge, there are no similarly published studies focusing on the usage of zonal statistics to compute climate indicators with climate reanalysis products. Therefore, there is still a demand for research to produce climate indicators ready-to-use for lattice modeling, aggregated at the municipal level in Brazil and other countries at daily resolution, covering a long time span.

This paper aims to propose a methodology to compute climate indicators aggregated by delimited spatial boundaries, such as administrative limits, with a Brazilian use case adopting the ERA5-Land reanalysis (Muñoz-Sabater et al., 2021) as a data source in order to produce daily zonal statistics of climate indicators for Brazilian municipalities from 1950 to 2022.

The proposed methodology may be later applied to other spatial limits in Brazil (other administrative boundaries and hydrological basins, for example) and other countries. The availability of ready-to-use climate indicator datasets has the potential to ease the work of research groups of several areas, not requiring specific knowledge of Geographic Information Systems, climate data processing, and other fields, enabling them to focus on modeling and its main research objectives.

2. Methods

2.1. Data sources

The exceptional continuity and integrity over long time spans present on climate reanalysis products make them good candidates for the creation of zonal statistics indicators, offering stable indicator definitions, methodology, and availability for several years and for large territorial areas such as the Brazilian territory.

The European ReAnalysis product, created by the European Centre for Medium-Range Weather Forecasts (ECMWF) with the Copernicus Climate Change Service (C3S), is on its fifth generation (ERA5, Hersbach et al., 2020). The ERA5 offers detailed records of the global atmosphere, land surface, and sea since 1950, replacing the European Centre for Medium-range Weather Forecasts ReAnalysis

Table 1. *Climate indicators, time and space aggregation functions*

ERA5-Land indicator	Daily time aggregating function	Spatial zonal statistic function
2 m temperature	Mean, max, min	Max, mean, min, stdev, count
2 m dewpoint temperature	Mean	Max, mean, min, stdev, count
10 m u-component of wind	Mean	Max, mean, min, stdev, count
10 m v-component of wind	Mean	Max, mean, min, stdev, count
Surface pressure	Mean	Max, mean, min, stdev, count
Total precipitation	Sum	Max, mean, min, stdev, sum, count

Abbreviations: Max, maximum; min, minimum; stdev, standard deviation; count, cells count.

Interim (ERA-Interim). The ERA5-Land reanalysis (Muñoz-Sabater et al., 2021) is derived from the ERA5, with coverage of only land areas and with a spatial resolution of 9 km and hourly outputs.

Since its release, the accuracy of ERA5-Land indicators has been evaluated, including research to assess some indicators' validity in Brazil and other countries and regions (Liu et al., 2020; Braga et al., 2021; de Araújo et al., 2022; Longo-Minnolo et al., 2022; Matsunaga et al., 2023). In general, the studies conclude that the ERA5-Land indicators are valid and fulfill their main purposes.

The following indicators from ERA5-Land were selected for this paper: 2 m dewpoint temperature, 2 m temperature, 10 m -*u*component of wind, 10m -*v*component of wind, surface pressure, and total precipitation.

2.2. Daily aggregation

Since the ERA5-Land provides hourly outputs and monthly aggregations, an initial task to aggregate the indicators at daily interval was performed.

For this, an R script was created using the `KrigR` package (Kusch and Davy, 2022). The script downloads the set of indicators from 1950 to 2022 from a geographical bounding box covering the Latin America region (coordinates $-118.47, -34.1, -56.65, 33.28$) and aggregates the data from hourly to daily intervals.

On this procedure, the hourly values of each cell within a day are summarized by a statistical function. For example, to compute the maximum temperature indicator, the maximum observed value of each cell within a day is selected. Table 1 presents the daily time aggregation functions used for each ERA5-Land original indicator. This step produces NetCDF files as a result, where each file covers a year's month and presents data layers for each day of the respective month.

The R script is available at GitHub[1] and the resulting NetCDF files are available at Zenodo[2].

2.3. Zonal statistics computation

As can be observed in Figure 1, the computation of zonal statistics considers the values from all cells that intersect the geographical boundary. For example, the mean zonal statistic can be expressed as:

$$\overline{X} = \frac{\sum_{1}^{n} x_i}{n} \tag{2.1}$$

Where x is the value of the climate indicator on the cell i and n is the total number of cells that intersect the spatial boundary. On this specification, the values of all cells that intersect the spatial boundaries are taken into account, disregarding the proportion of the cell that intersects the spatial boundary.

---

[1] https://github.com/rfsaldanha/era5daily.
[2] https://zenodo.org/communities/brclim.

Administrative boundaries may follow irregular patterns, leading to the inclusion of several cells on the zonal statistical computation that only slightly intersects the geographical boundary.

As suggested by Baston (2023), it is possible to weight the computation of the zonal statistics by considering the overlapping proportion of the cells over the spatial boundaries:

$$\overline{X}_c = \frac{\sum_1^n x_i c_i}{\sum_1^n c_i} \tag{2.2}$$

Where c is the proportional area of the cell i covered by the spatial boundary. This speciation results in a more realistic summarization of the climate indicator over a spatial boundary and was adopted in this work.

To perform the computation of the zonal statistics on the daily aggregated indicators, an R package called `zonalclim` was developed. The package contains a function to create a list of tasks to be computed, splitting the layers available at the indicator files and the geographical boundaries into a number of chunks. This strategy was adopted to avoid overloading the random access memory by loading a potentially large number of climate data files and geographic boundaries at once. This list of tasks feeds a function that wraps the `exactextractr` package (Baston, 2023) and performs the zonal statistics computation, writing its results in an SQLite database.

The Brazilian municipality boundaries from 2010 were used as a reference, provided by the "Instituto Brasileiro de Geografia e Estatística" (IBGE), the Brazilian official authority of administrative boundaries. The simplified boundaries version presented by the *geobr* (Pereira and Goncalves, 2022) package was used, which preserves the original topology with a tolerance of 100 m. The specific year of 2010 was adopted due to its compatibility with the population census conducted in the same year.

Table 1 summarizes the daily time aggregation functions and spatial zonal statistical functions applied to the climate indicators.

Thus, for each climate indicator, a number of spatial zonal statistical functions are applied to the daily time aggregation results. There are a total of eight daily climate indicators, leading to a total of 41 indicators computed for each day between 1950 and 2022.

The resulting daily climate indicators are converted to databases in the DuckDB (Mühleisen and Raasveldt, 2023) format for faster columnar-vectorized queries and exported as parquet files (Richardson et al., 2023) for dissemination of results.

3. Results

The available ERA5-Land data from 1950 to 2022 have been processed, and the computed zonal statistics are available to access at the parquet files listed in Table 2. These files follow a common schema, as

***Table 2.** Parquet files, indicators, and units*

File	Indicator	Unit
`prec.parquet`	Total precipitation	M
`tmax.parquet`	2 m maximum temperature	K
`tmin.parquet`	2 m minimum temperature	K
`tmean.parquet`	2 m mean temperature	K
`tdew.parquet`	2 m dewpoint temperature	K
`surfp.parquet`	Surface pressure	Pa
`vwind.parquet`	10 m v-component of wind	ms^{-1}
`uwind.parquet`	10 m u-component of wind	ms^{-1}

Table 3. *Common parquet file schema*

Field	Description	Format
code_muni	Municipality code	Numeric
date	Date	Date
name	Indicator acronym and zonal statistic	Text
value	Value	Numeric

presented in Table 3. As ERA5-Land data are updated monthly with a delay of three months, we plan to continuously update the zonal statistics dataset with new data at yearly intervals or more often if feasible.

Together, those files present 6,085,749,761 records of daily climate indicators for Brazilian munici- palities. The files are available for download at Zenodo: https://zenodo.org/doi/10.5281/ zenodo.10036211.

Figure 2 presents a sample from the dataset, with the zonal statistic of average temperatures from the 27 Brazilian capitals. It is possible to observe at the figure different climate regimes, where municipalities like Belém, São Luiz, and Fortaleza present almost flat and stable temperatures over the months of all years. Differently, municipalities like Campo Grande, Curitiba, and Porto Alegre show lower temperat- ures over winter months (from June to September) on all years.

Figures 3 and 4 are maps of the maximum and minimum zonal statistics of temperature (k) for the Brazilian municipalities at January 1 on the years of 1950, 1960, 1970, 1980, 1990, 2000, 2010, and 2020. It is possible to observe at those maps how the maximum and minimum zonal statistics are spatially clustered over some groups of municipalities at this specific date, highlighting north and northeast parts of the country and also the municipalities at the south region of the country, localized in a mountain range.

Figure 5 presents the sum and standard deviation zonal statistics of precipitation (m) at the Rio de Janeiro municipalities on January 1, 2010. The standard deviation zonal statistic represents the spatial variability of the precipitation, where higher standard deviation values implies on precipitation concen- trated on few grid cells, and the sum zonal statistic is the sum of the cell's values. Considering those statistics together, it is possible to observe that the total precipitation was higher on two municipalities (*Rio Claro* on the center-west and *Cachoeiras de Macacu* on the east) but more concentrated in a different municipality (*Angra dos Reis*) on that specific date. The higher standard deviation of precipitation occurrence is directly related to a massive landslide occurred on that same day, which resulted in 50 deaths (de Mendonca and da Silva, 2020; Memória Globo, 2021).

Some of the zonal statistics—for example, mean, minimum, and maximum—may be re-aggregated in time by the user in order to create time series with monthly, quarterly, and other intervals.

Figure 6 presents the Köppen Climate Classification for the Brazilian municipalities, based on the proposed dataset, using data from 1990 to 2020. The daily zonal statistics of precipitation, maximum temperature, minimum temperature, and average temperature were aggregated to monthly periodicity using averages values, and the municipalities were classified with the ClimClass package (Eccel et al., 2016). The obtained classes are well aligned with results from a study that adopted weather station data for the 1961–2015 period (Dubreuil et al., 2017).

4. Discussion

The proposed methodology resulted in a dataset that will be useful for further researches that require climate indicators at the municipal scale of analysis. In comparison with published datasets of similar objective, our proposed dataset excels previous research by providing a longer time coverage of more than six decades at a finer daily time resolution. Also, our results present a rich set of six climate indicators that were spatially aggregated by zonal statistics with different strategies, which allows a more comprehensive representation of the spatial variability of the climate indicators over the municipalities' territory.
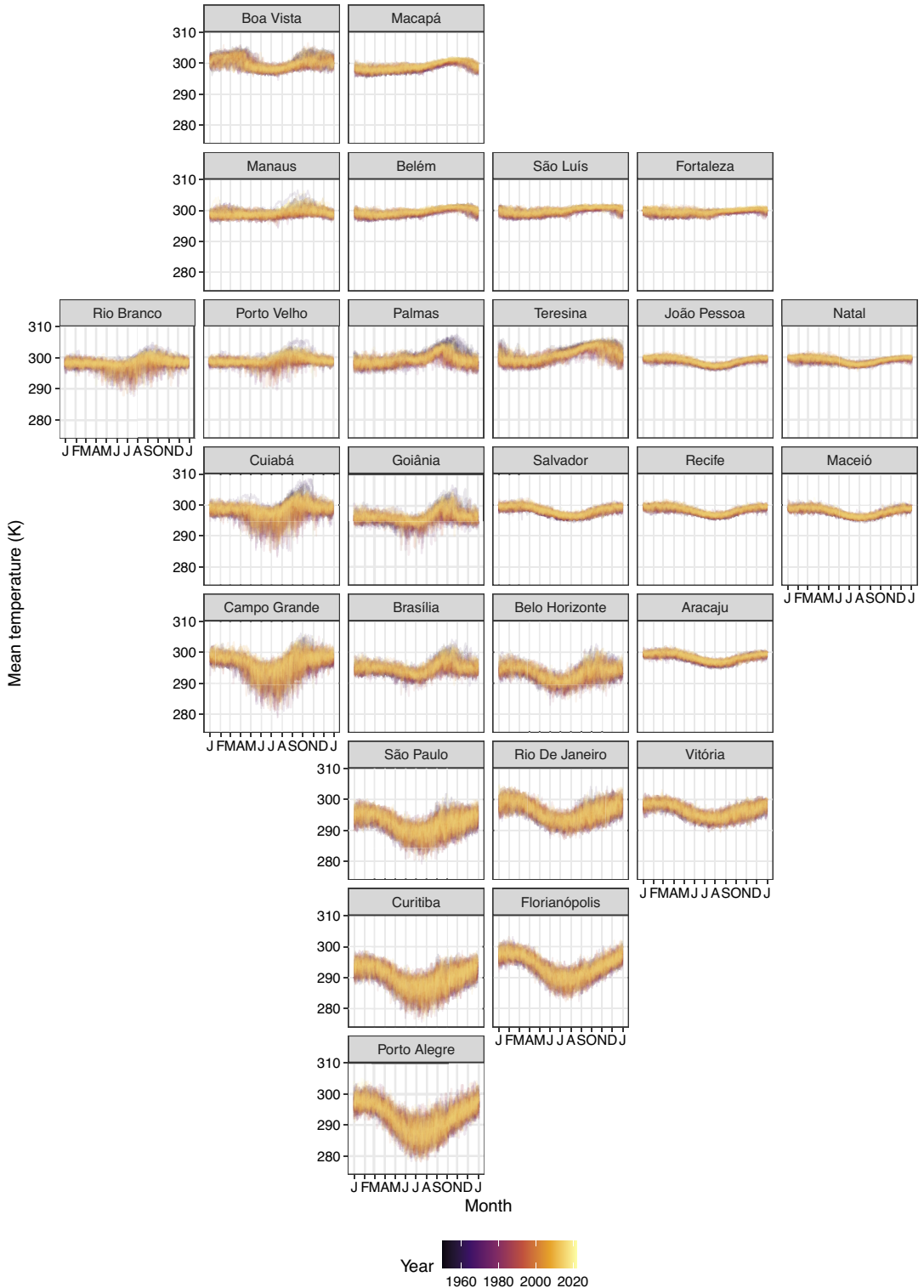
Figure 2. Average daily temperatures (*K*) from the 27 Brazilian capitals, from 1950 to 2022.
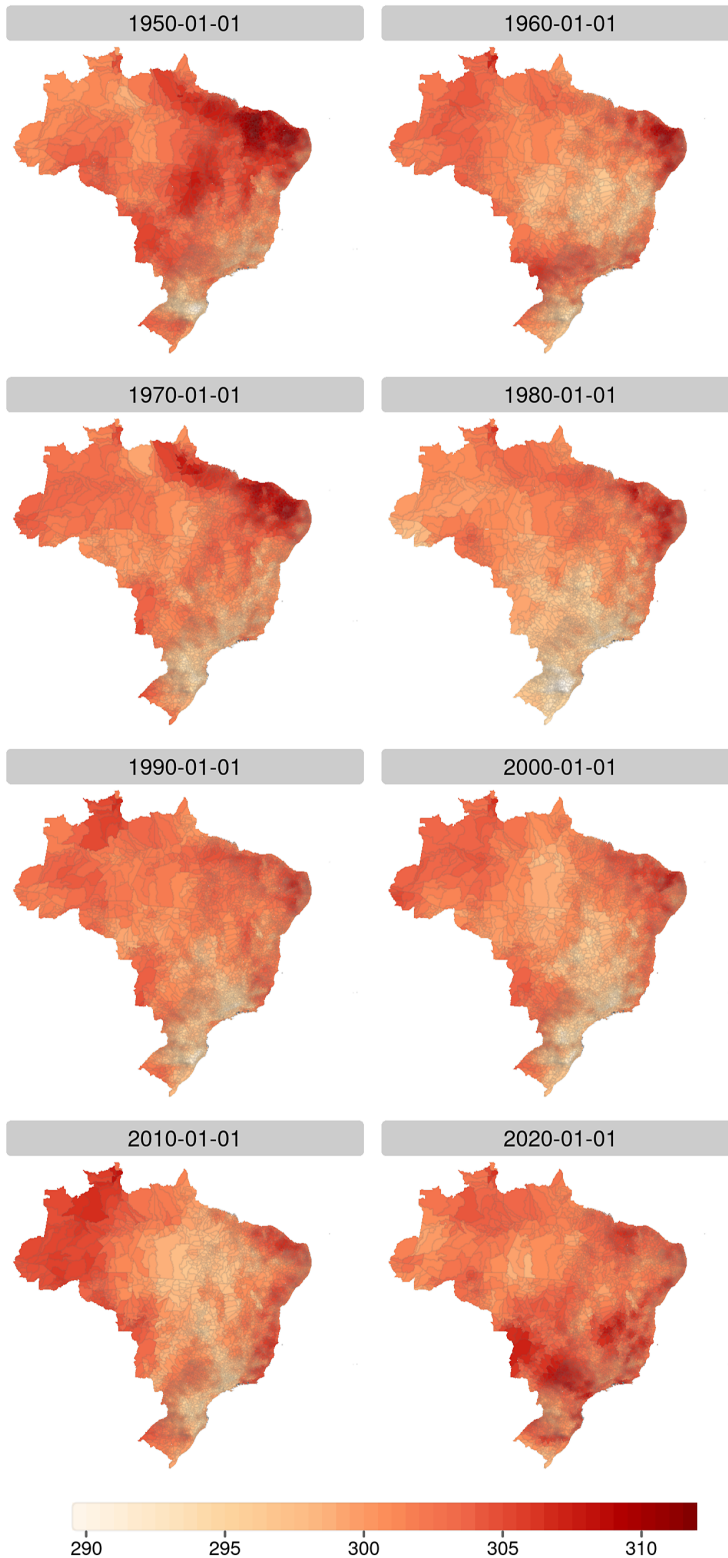
Figure 3. Maximum zonal temperatures (*K*) on Brazilian municipalities, for selected years.
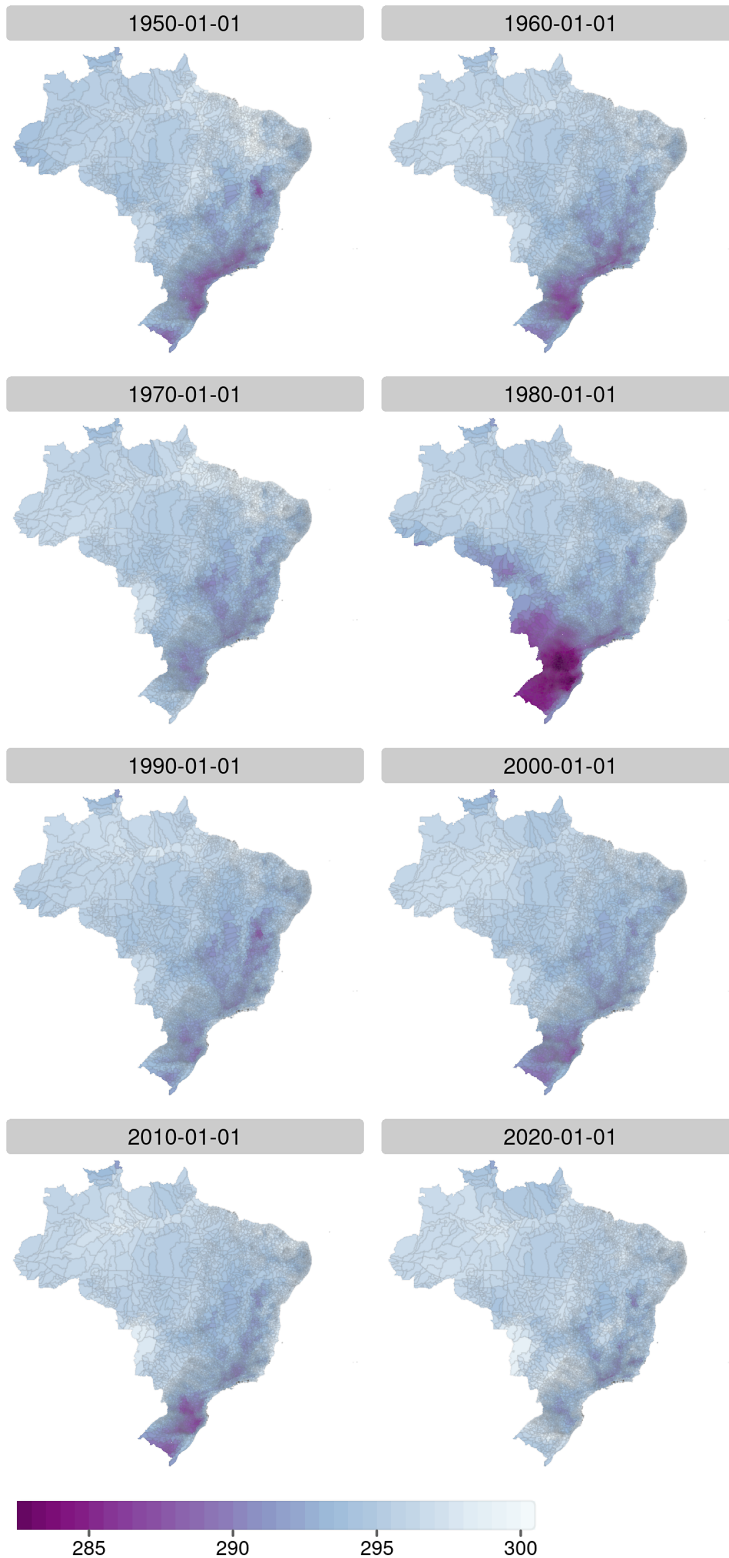
Figure 4. Minimum zonal temperatures (*K*) on Brazilian municipalities, for selected years.
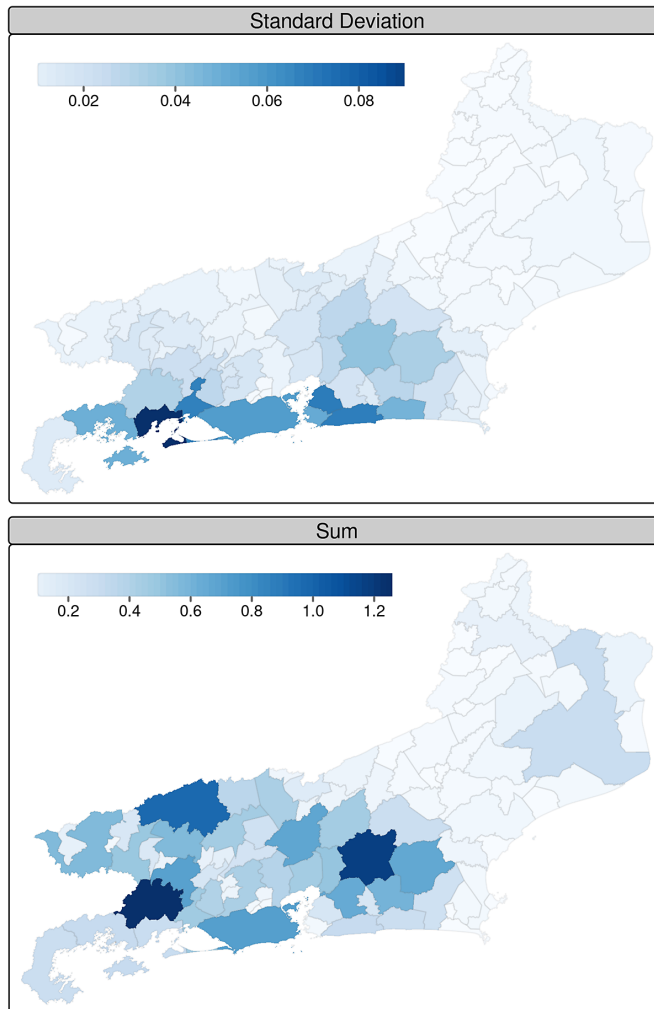
Figure 5. Standard deviation and sum zonal statistics of precipitation (*m*) at the Rio de Janeiro municipalities on January 1, 2010.

The adopted methodology is fully reproducible, allowing the continuous update of the results as new data are made available by the ERA5-Land project and the inclusion of new climate indicators and other zonal statistics. The proposed methodology can be also easily extended to other territories of the globe to produce climate zonal statistics to other countries.

We believe that the proposed dataset will be helpful for climate-aware policymaking by presenting climate indicators spatially aggregated by municipalities, a fairly recognized and common boundary division for strategical planning in Brazil.

It is important to highlight that the interpretation and usage of zonal statistics to summarise the climate of administrative boundaries must consider the existing relationship between the boundaries area and the spatial resolution of the grid indicator. As with other statistical summary measures, the zonal statistics' capacity to summarize values is directly related to the sample size and variance.

Figure 7 presents a histogram with the weighted area cell count in each municipality, with the *x* axis on logarithmic scale. The weighted area cell count represents the sum of entire cells and fractions used to compute the zonal statistics indicators in each municipality. The ERA5-Land grid is constant over the years for all indicators.
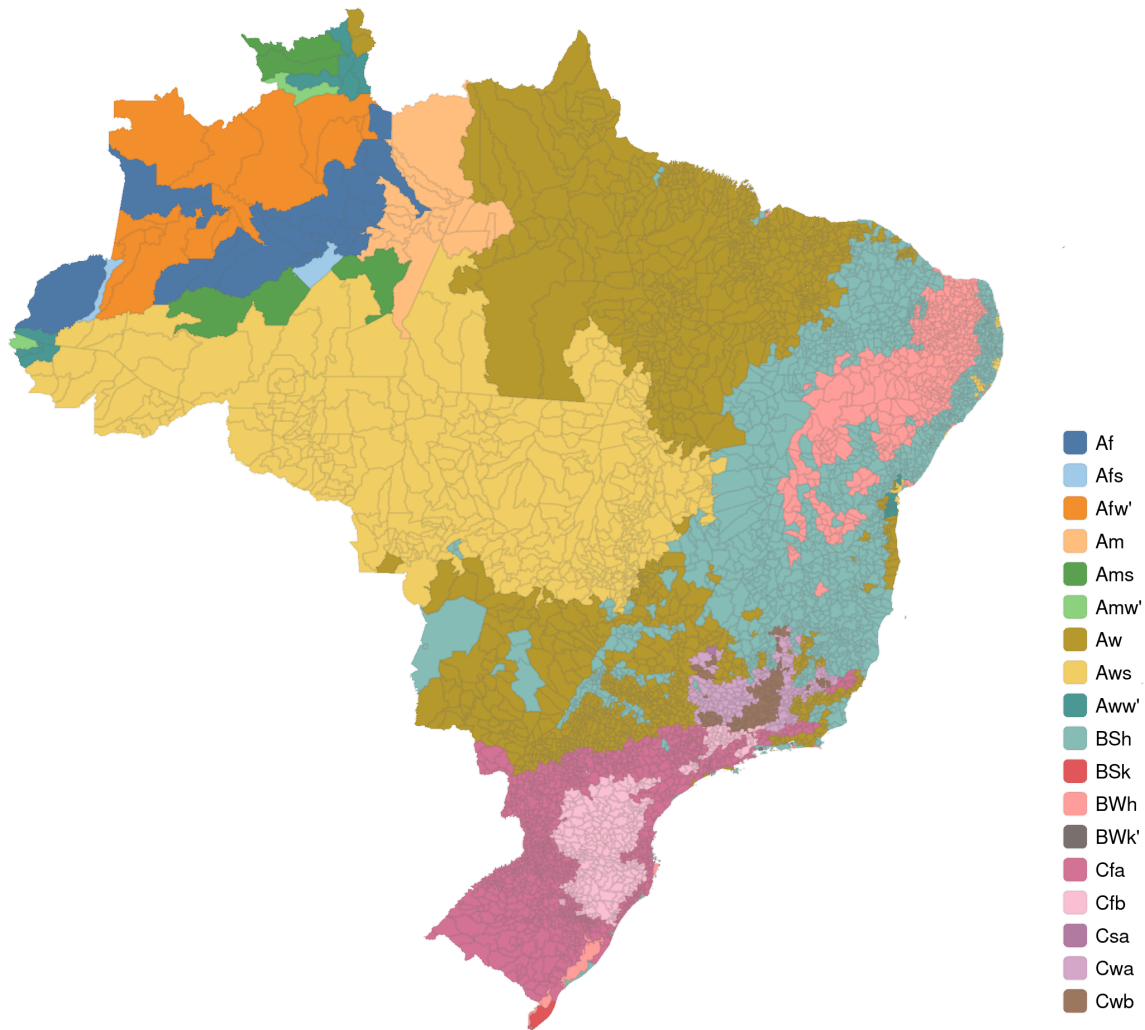
Figure 6. *Köppen-Geiger climate classification of Brazilian municipalities, 1990 to 2020.*
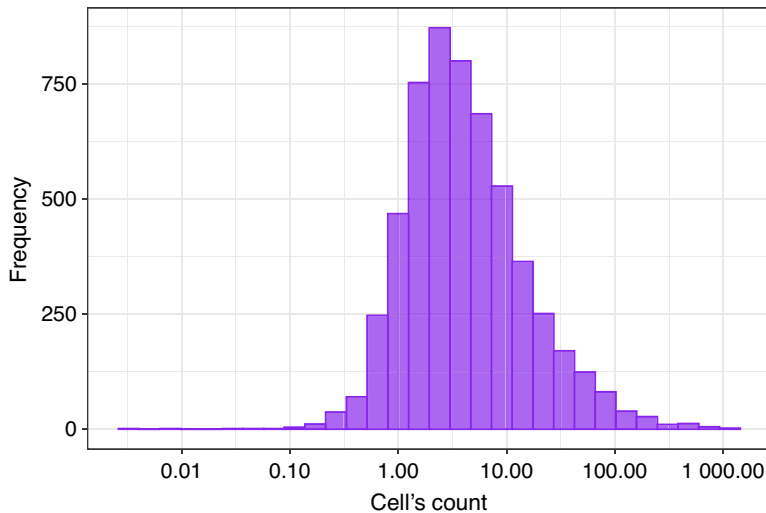
Figure 7. *Histogram of weighted number of cells on a logarithmic scale.*

There are 595 municipalities where the zonal statistics were computed using one cell or a fraction due to its small territorial area in comparison with the ERA5-Land cell size. The zonal statistics from 3,736 municipalities were computed with data from 1 to 10 cells, 1,138 municipalities with 10 to 100 cells, and 98 municipalities with more than 100 cells.

Municipalities with larger territorial areas will have more data points used for the zonal statistical computation. This may introduce a spatial variability in the zonal statistics, as the municipal area may cover a diversity of regions with different climate regimes. The spatial variability of the indicator may be observed by considering the standard deviation of the zonal statistic and the weighted cell count. Higher or lower values of standard deviations imply that there is, respectively, a larger or smaller variability in the values of the totality of cells used to compute the zonal statistic. Thus, the spatial variability measured by the standard deviation shows how concentrated or dispersed an indicator is in space, as illustrated by Figure 5.

5. Conclusion

In this work, we presented a consistent methodology to produce updated zonal statistics from gridded climate indicators for the Brazilian municipalities, resulting in the creation of a dataset of zonal statistics that covers a long time span at daily resolution and presents a full set of relevant climate indicators. The dataset access is facilitated by the adoption of the parquet file format for dissemination and may be updated periodically as new data are made available by the Copernicus program.

There are some possibilities that future research could address. To enhance the spatial resolution of the grid files, the application of statistical downscaling is possible by using interpolation or kriging methods (Chiles and Delfiner, 2012). For this, a list of suitable and available covariates for each climate indicator at higher spatial resolution must be compiled, and a study of the errors in the estimates is required.

Furthermore, the daily zonal statistics of climate indicators offered by this proposal may also be used to calculate weekly or monthly statistical time features of trends, persistence, and intensity. For example, indicators such as temperature variation ratio, number of consecutive days of rain, and number of days of an indicator below, above, or within some threshold may be computed by week or month interval. Also, some of the indicators proposed by the Expert Team on Climate Change Detection and Indices (ETCCDI) may be computed (World Meteorological Organization, 2009).

Punctual measures of weather indicators may also be extracted and presented as municipalities' centroids or city hall localization, as adopted by the SISAM project. Considering the use of more refined

spatial resolution grids, there is also the possibility of generating zonal statistics for different land covers and land use. The availability of zonal statistics for urban and rural areas, for example, may be useful for studies of urban vector diseases or for agriculture production.

These zonal statistics presented by this research may be used as a common source of data for different research projects, by presenting a standardized set of weather indicators for all municipalities for a long time span. The availability of direct access to weather indicators at the municipal level may speed up research in modeling lattice data, saving research time and effort.

References

Abatzoglou JT, **Dobrowski SZ**, **Parks SA and Hegewisch KC** (2018) Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data 5*, 170191. https://doi.org/10.1038/sdata.2017.191

Abdalla L, **Augusto DA**, **Chame M**, **Dufek AS**, **Oliveira L and Krempser E** (2022) Statistically enriched geospatial datasets of brazilian municipalities for data-driven modeling. *Scientific Data 9*, 489. https://doi.org/10.1038/s41597-022-01581-2

Baston D (2023) Exactextractr: Fast extraction from raster datasets using polygons. Available at https://isciences.gitlab.io/exactextractr/, https://github.com/isciences/exactextractr

Braga RAHW, **Santos EB and Barros MFD** (2021) Validação de dados de vento da reanálise ERA5-LAND Para estimativa de potencial eólico no Estado do Rio de Janeiro. *Revista Brasileira de Energia 27*(4). https://doi.org/10.47168/rbe.v27i4.592

Chiles J-P and Delfiner P (2012) *Geostatistics: Modeling Spatial Uncertainty*, Vol. *713*. New Jersey: John Wiley & Sons.

Cressie N (1991) *Statistics for Spatial Data*. New Jersey: John Wiley & Sons.

de Aguiar JT and Lobo M (2020) Reliability and discrepancies of rainfall and temperatures from remote sensing and Brazilian ground weather stations. *Remote Sensing Applications: Society and Environment 18*, 100301. https://doi.org/10.1016/j.rsase.2020.100301

de Araújo CSP, e Silva IAC., Ippolito M and de Almeida CDGC (2022) Evaluation of air temperature estimated by ERA5-land reanalysis using surface data in Pernambuco, Brazil. *Environmental Monitoring and Assessment 194*(5), 381. https://doi.org/10.1007/s10661-022-10047-2

de Mendonca MB and da Silva DR (2020) Integration of census data based vulnerability in landslide risk mapping - The case of Angra dos Reis, Rio de Janeiro, Brazil. *International Journal of Disaster Risk Reduction 50*, 101884. https://doi.org/10.1016/j.ijdrr.2020.101884

Dee DP, **Balmaseda M**, **Balsamo G**, **Engelen R**, **Simmons AJ and Thépaut J-N** (2014) Toward a consistent reanalysis of the climate system. *Bulletin of the American Meteorological Society 95*(8), 1235–1248. https://doi.org/10.1175/BAMS-D-13-00043.1

Dubreuil V, **Pechutti Fante K**, **Planchon O and Neto JLS** (2017) Les types de climats annuels au Brésil: Une application de la classification de Köppen de 1961 à 2015. *EchoGéo 41*. https://doi.org/10.4000/echogeo.15017

Eccel E, **Cordano E and Toller G** (2016) *Climclass: Climate classification according to several indices* [R package version 2.1.0]. Available at https://CRAN.R-project.org/package=ClimClass.

Funk C, **Peterson P**, **Landsfeld M**, **Pedreros D**, **Verdin J**, **Shukla S**, **Husak G**, **Rowland J**, **Harrison L**, **Hoell A and Michaelsen J** (2015) The climate hazards infrared precipitation with stations: A new environmental record for monitoring extremes. *Scientific Data 2*, 150066. https://doi.org/10.1038/sdata.2015.66

Hersbach H, **Bell B**, **Berrisford P**, **Hirahara S**, **Horányi A**, **Muñoz-Sabater J**, **Nicolas J**, **Peubey C**, **Radu R**, **Schepers D**, **Simmons A**, **Soci C**, **Abdalla S**, **Abellan X**, **Balsamo G**, **Bechtold P**, **Biavati G**, **Bidlot J**, **Bonavita M**, **De Chiara G**, **Dahlgren P**, **Dee D**, **Diamantakis M**, **Dragani R**, **Flemming J**, **Forbes R**, **Fuentes M**, **Geer A**, **Haimberger L**, **Healy S**, **Hogan RJ**, **Hólm E**, **Janisková M**, **Keeley S**, **Laloyaux P**, **Lopez P**, **Lupu C**, **Radnoti G**, **de Rosnay P**, **Rozum I**, **Vamborg F**, **Villaume S**

and Thépaut J-N (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society 146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

INPE (2019) *Sistema de informações ambientais integrado a saúde – sisam.* Available at https://queimadas.dgi.inpe.br/queimadas/sisam/v2/.

Kusch E and Davy R (2022) KrigR—A tool for downloading and statistically downscaling climate reanalysis data. *Environmental Research Letters 17*(2), 024005. https://doi.org/10.1088/1748-9326/ac48b3

Liu J, **Hagan DFT and Liu Y** (2020) Global land surface temperature change (2003–2017) and its relationship with climate drivers: AIRS, MODIS, and ERA5-land based analysis. *Remote Sensing 13*(1), 44. https://doi.org/10.3390/rs13010044

Longo-Minnolo G, **Vanella D**, **Consoli S**, **Pappalardo S and Ramírez-Cuesta JM** (2022) Assessing the use of ERA5-land reanalysis and spatial interpolation methods for retrieving precipitation estimates at basin scale. *Atmospheric Research 271*, 106131. https://doi.org/10.1016/j.atmosres.2022.106131

Lowe R, **Bailey TC**, **Stephenson DB**, **Graham RJ**, **Coelho CA**, **Sá Carvalho M and Barcellos C** (2011) Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil. *Computers and Geosciences 37*(3), 371–381. https://doi.org/10.1016/j.cageo.2010.01.008

Lowe R, **Cazelles B**, **Paul R and Rodó X** (2016) Quantifying the added value of climate information in a spatio-temporal dengue model. *Stochastic Environmental Research and Risk Assessment 30*, 2067–2078. https://doi.org/10.1007/s00477-015-1053-1

Matsunaga WK, **Sales ESG**, **Júnior GCA**, **Silva MT**, **Lacerda FF**, **de Paiva Lima E**, **dos Santos CAC and de Brito JIB** (2023) Application of ERA5-land reanalysis data in zoning of climate risk for corn in the state of Bahia—Brazil. *Theoretical and Applied Climatology 155*, 945–963. https://doi.org/10.1007/s00704-023-04670-3

Memória Globo (2021) Deslizamentos em Angra. Retrieved December 12, 2023, from https://memoriaglobo.globo.com/jornalismo/coberturas/deslizamentos-em-angra/noticia/deslizamentos-em-angra.ghtml.

Morgenstern H (1995) Ecologic studies in epidemiology: Concepts, principles, and methods. *Annual Review of Public Health 16*, 61–81. https://doi.org/10.1146/annurev.publhealth.16.1.61

Mühleisen H and Raasveldt M (2023) *Duckdb: Dbi package for the duckdb database management system* [R package version 0.7.0]. Available at https://CRAN.R-project.org/package=duckdb.

Muñoz-Sabater J, **Dutra E**, **Agustí-Panareda A**, **Albergel C**, **Arduini G**, **Balsamo G**, **Boussetta S**, **Choulga M**, **Harrigan S**, **Hersbach H**, **Martens B**, **Miralles DG**, **Piles M**, **Rodríguez-Fernández NJ**, **Zsoter E**, **Buontempo C and Thépaut JN** (2021) Era5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data 13*, 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Overpeck JT, **Meehl GA**, **Bony S and Easterling DR** (2011) Climate data challenges in the 21st century. *Science 331*, 700–702. https://doi.org/10.1126/science.1197869

Pereira RHM and Goncalves CN (2022) *Geobr: Download official spatial data sets of Brazil* [R package version 1.7.0]. Available at https://CRAN.R-project.org/package=geobr.

Richardson N, **Cook I**, **Crane N**, **Dunnington D**, **François R**, **Keane J**, **Moldovan-Grünfeld D**, **Ooms J** and **Apache Arrow** (2023) *Arrow: Integration to 'apache' 'arrow'* [R package version 11.0.0.2]. Available at https://CRAN.R-project.org/package=arrow.

Schneider R, **Sebastianelli A**, **Spiller D**, **Wheeler J**, **Carmo R**, **Nowakowski A**, **Garcia-Herranz M**, **Kim D**, **Barlevi H**, **El Raiss Cordero Z**, **Liberata Ullo S**, **Mathieu P-P and Lowe R** (2021) *Climate-based ensemble machine learning model to forecast dengue epidemics. In ICML 2021 Workshop on Tackling Climate Change with Machine Learning.* Available at https://www.climatechange.ai/papers/icml2021/10.

Tomlin C (1994) Map algebra: One perspective. *Landscape and Urban Planning 30*, 3–12. https://doi.org/10.1016/0169-2046(94)90063-9

Wongchuig SC, **De Paiva RCD**, **Siqueira V and Collischonn W** (2019) Hydrological reanalysis across the 20th century: A case study of the Amazon Basin. *Journal of Hydrology 570*, 755–773. https://doi.org/10.1016/j.jhydrol.2019.01.025

World Meteorological Organization (2009) *Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation.* Technical report No. 72. WMO, Geneve.

Zou J, **Lu N**, **Jiang H**, **Qin J**, **Yao L**, **Xin Y and Su F** (2022) Performance of air temperature from ERA5-land reanalysis in coastal urban agglomeration of Southeast China. *Science of the Total Environment 828*, 154459. https://doi.org/10.1016/j.scitotenv.2022.154459