# MapBiomas Atmosphere

# Algorithm Theoretical Basis Document (ATBD)

Version 1

**General coordinator**
Luciana Varanda Rizzo

**Team**
Arthur Thomas Luz
Felipe Santos da Silva
Luiz Augusto Toledo Machado
Márcio José Teixeira
Marco Aurélio de Menezes Franco
Paulo Eduardo Artaxo Netto
Victor Hugo Barreiro

*In addition to the institutions supporting MapBiomas, https://brasil.mapbiomas.org/quem-somos, this initiative also received support from FAPESP (2022/07974-0, 2020/15230-5).*

# 1. Introduction

## 1.1. Overview of the MapBiomas Atmosphere

This document describes the theoretical basis and methods applied to produce a harmonized dataset of atmospheric variables covering the Brazilian territory from 1985 to 2024, integrating data from different sources, as part of Collection 10 of the MapBiomas Project. Global datasets on meteorological variables were validated for Brazil and resampled to a common spatial resolution of 0.1 degrees and monthly temporal resolution. Table 1 lists the meteorological variables included in the MapBiomas Atmosphere dataset.

The methodological workflow consisted of: (i) resampling data from different sources to a common spatial grid; (ii) calculation of derived atmospheric variables; (iii) data validation against independent observations at surface. Subsequently, the harmonized dataset was exported to the MapBiomas workspace for integration with land use and land cover maps. Details about each step are provided below. All steps were implemented on the Google Earth Engine platform using JavaScript and Python. The codes are openly available in the MapBiomas GitHub repository.

Table 1: List of atmospheric variables included in the MapBiomas Atmosphere dataset.

| Variable | Units | Data source | Temporal coverage |
|---|---|---|---|
| Mean air temperature | ºC | ERA5 Land | 1985-2024 |
| Maximum air temperature | ºC | ERA5 Land | 1985-2024 |
| Minimum air temperature | ºC | ERA5 Land | 1985-2024 |
| Temperature anomaly | ºC | ERA5 Land | 1985-2024 |
| Land surface temperature | ºC | ERA5 Land | 1985-2024 |
| Mean temperature trend | ºC/decade | ERA5 Land | — |
| Precipitation | mm | GPCC+GPM | 1985-2024 |
| Precipitation anomaly | mm | GPCC+GPM | 1985-2024 |
| Days without rain | days | GPCC+GPM | 1985-2024 |
| Days of persistent rain | days | GPCC+GPM | 1985-2024 |
| Vapor pressure deficit | kPa | ERA5 Land | 1985-2024 |
| Water availability | mm | GPCC+GPM+ERA5 Land | 1985-2024 |
| Inhalable particulate matter (PM10) | μg/m$^3$ | CAMS | 2003-2024 |
| Fine particulate matter (PM2.5) | μg/m$^3$ | CAMS | 2003-2024 |

# 2. Methodological description

## 2.1. Data source

The MapBiomas Atmosphere dataset is a compilation of global datasets on meteorological variables, including satellite observations and atmospheric reanalysis models. Extensive validation against surface observations was performed to assess the accuracy of the final data product. This resulting database provides high quality atmospheric raster data over Brazil, covering the period from 1985 to 2024. The dataset offers a consistent monthly resolution and is complete, without data spatial or temporal gaps. Biases and limitations are discussed in section 2.6.

Table 2 summarizes the main characteristics of the data sources and lists the primary meteorological variables, from which derived metrics were calculated. Temperature and evaporation data was acquired from the ERA5-Land monthly averaged dataset (Muñoz-Sabater et al., 2019, 2021). ERA5-Land is a state-of-the-art global reanalysis dataset, providing gridded data on meteorological and surface variables derived from a combination of Earth system models and assimilation of in situ and remote sensing observations. ERA5 data was acquired and processed in the Google Earth Engine environment, using the image collection "ECMWF/ERA5_LAND/MONTHLY_AGGR". From the primary variables listed in Table 2, four additional variables were calculated, as described in Section 2.3: Temperature anomaly, Vapor Pressure Deficit, Water Availability, and Temperature trend.

Table 2: Main characteristics of the data sources for the primary meteorological variables included in the MapBiomas Atmosphere dataset.

| Data source | Original spatial resolution | Original temporal resolution | Time period | Variables | Reference |
|---|---|---|---|---|---|
| ERA5-Land | 0.1° | 1 month | 1985-2024 | temperature_2m<br>temperature_2m_min<br>temperature_2m_max<br>dewpoint_temperature_2m<br>skin_temperature<br>total_evaporation_sum | doi:10.24381/cds.68d2bb30 |
| GPM V07 | 0.1° | 30 minutes | 1998-2024 | precipitation | doi:10.5067/GPM/IMERG/3B-HH/07 |
| GPCC V2022 | 0.25° | 1 month | 1985-1997 | precipitation | doi:10.5676/DWD_GPCC/FD_M_V2022_025 |
| GPCC V2022 | 1° | 1 day | 1985-1997 | precipitation | doi: 10.5676/DWD_GPCC/FD_D_V2022_100 |
| CAMS EAC4 | 0.75° | 3 hours | 2003-2024 | Particulate matter d<10 µm (PM10)<br>Particulate matter d<2.5 µm | doi:10.24381/d58bbf47 |

| | | | | | |
|---|---|---|---|---|---|
| | | | | (PM2.5) | |

Precipitation data was acquired from two complementary datasets, to cover the entire period 1985-2024. For 1998-2024, we used the Global Precipitation Measurement (GPM) project, which provides high-quality gridded data on precipitation. GPM-IMERG (Integrated Multi-satellite Retrievals for GPM) data is derived from observations from a constellation of satellite passive microwave sensors, combined with microwave precipitation-calibrated infrared fields, and adjusted with surface precipitation data (Huffman et al., 2019). The IMERG final product (product code GPM_3IMERGHH; GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V07) provides global precipitation data at 30-minutes intervals from 1998 to 2024 . GPM data was acquired and processed in the Google Earth Engine environment, using the image collection "NASA/GPM_L3/IMERG_V07".

For the period between 1985 and 1997, when GPM data was not available, we used data from the Global Precipitation Climatology Centre (GPCC). GPCC retrieves precipitation data by combining ground-based rain gauge stations and satellite data, which is followed by quality control, harmonization, and spatial interpolation procedures. Two GPCC data products were used: Full Data Monthly Product Version 2022 (Schneider et al., 2022) and Full Data Daily Version 2022 (Zieze et al., 2022). The monthly product was used to represent total monthly precipitation and anomalies,while the GPCC daily data, provided in a coarser spatial resolution, was only used to calculate the secondary variables “Days without rain” and “Days of persistent rain” for the period 1985-1997. GPCC data was acquired from the German Weather Service (DWD) website as NetCDF files, and processed locally using scripts in Python. The combination of precipitation data from GPCC and GPM covers the entire period of interest (1985-2024).

Finally, air quality data was acquired from the global reanalysis of atmospheric composition (EAC4) produced by the Copernicus Atmosphere Monitoring Service (CAMS) (Inness et al., 2019). CAMS-EAC4 provides global gridded data on air pollutant concentrations, estimated from a combination of atmospheric transport models and assimilation of remote sensing observations. The MapBiomas Atmosphere dataset uses CAMS data on particulate matter concentrations near the surface, in two particle size ranges: inhalable particulate matter (PM10, particles with diameter below 10 μm) and fine particulate matter (PM2.5, particles with diameter below 2.5 μm). CAMS data was acquired from Copernicus Atmosphere Data Store as NetCDF files, and processed locally using scripts in Python.

## 2.2. Spatial and temporal resampling

The combination of distinct data sources results in differences in spatial and temporal resolutions among the input data products. To produce a harmonized dataset, all data were resampled to a common spatial grid and monthly temporal resolution. The gridded datasets were reprojected to the WGS 84 geographic coordinate system (EPSG:4326) using a 2D affine transformation. The data were linearly interpolated to a spatial resolution of 0.1 degrees per pixel (approximately 10 km), with the top-left corner of the image aligned to the geographic coordinates of -180.05° longitude and 90.05° latitude. Data from GPCC and

CAMS-EAC4 were linearly interpolated using the function DataArray.interp, from the xarray Python library. Data from ERA5 Land and GPM were resampled in the Google Earth Engine platform.

In terms of temporal resolution, all datasets were aggregated to monthly resolution, ensuring consistency across variables and allowing intra-annual analysis of temporal trends and fluctuations within the studied variables. This resampling process facilitates a more detailed understanding of seasonal changes and dynamic interactions that may not be evident in an annual dataset. GPM precipitation data, originally available every 30 minutes, was accumulated over the months. CAMS air quality data, originally available every 3 hours, was monthly aggregated by calculating mean concentrations. In addition to the monthly resolution data, MapBiomas Atmosphere also provides an annual dataset, obtained by aggregating the monthly data: annual means for temperature-related data and air quality variables, and annual sums for precipitation-related variables.

## 2.3. Calculation of derived metrics

### 2.3.1. Temperature and Precipitation anomalies

Climate and weather anomalies are derived from long-term time series of meteorological variables, indicating significant deviations from a region's typical (long-term average) conditions. Anomalies are calculated by comparing current conditions to a baseline or reference period. A positive anomaly indicates conditions are warmer or wetter than normal, while a negative anomaly indicates they are cooler or drier than normal.

In the MapBiomas Atmosphere dataset, temperature and precipitation anomalies were calculated using the period 1985-2024 as reference. For each grid point $(x,\ y)$ over Brazil, mean annual cycles (seasonal patterns) were calculated using monthly temperature and precipitation data, obtaining reference values for each month $(M)$ in the baseline period. For air temperature, the anomaly time series ($A_T$) for each grid point ($x$,$y$) was calculated by subtracting the baseline value ($B_T$) from the conditions at time $t$ in given month $M$:

$$A_T(x,\ y,\ t) = T(x,\ y,\ t) - B_T(x,\ y,\ M) \quad (1),$$

where:
T(x,y,t) is the monthly mean air temperature at grid point $(x,y)$ and time $t$, and
$B_T(x,y,M)$ is the baseline (climatological) value for month $M$, given by:

$$B_T(x,\ y,\ M) = \sum_{M=1}^{12} \sum_{Y=1985}^{2024} \frac{T(x, y, M, Y)}{2024-1985}.$$

with T(x,y,M,Y) representing the monthly mean temperature for month $M$ and year $Y$.
The same procedure was used to obtain the precipitation anomaly time series ($A_P$) for each grid point.

### 2.3.2. Days without rain

The number of days without rain can be used as a proxy for dry conditions and can be related to drought risk. For each gridpoint (x, y) and month, the number of days without rain was calculated based on daily precipitation data from the GPCC (1985-1997) and GPM

(1998-2024) datasets. The number of days with zero precipitation in 24 hours was summed up in each grid point, generating a map for each month. The number of days without rain was also aggregated by year, by summing up the corresponding monthly counts.

### 2.4.3. Days of persistent rain

Persistent rainfall with high cumulative volumes over consecutive days can increase the risk of floods and landslides. Therefore, metrics for persistent rain can help assess this risk and its potential impacts. A common metric for persistent rain is the consecutive number of days exceeding a specific rainfall threshold. In the MapBiomas Atmosphere project, an event of persistent rain was defined as at least 3 consecutive days with a cumulative rainfall of 60 mm or more. This metric was adapted from the definitions of the Expert Team on Climate Change Detection and Indices (ETCCDI) (Karl et al., 1999) to the Brazilian context. For each gridpoint (x, y) and month, the number of days of persistent rain was calculated based on daily precipitation data from the GPCC (1985-1997) and GPM (1998-2024) datasets. For each gridpoint, the daily precipitation data was processed by a rolling sum using a 3-day length window. Whenever the result of this operation exceeded 60 mm, the corresponding days were labeled as an event day. The number of days of persistent rain (event day) was then summed up for each month and grid point, generating monthly maps. The number of days of persistent rain was also aggregated by year, by summing up the corresponding monthly counts.

### 2.4.4. Vapor Pressure Deficit (VPD)

The Vapor Pressure Deficit (VPD) is a function of the temperature and amount of water vapor in the air. High VPD values indicate low humidity air conditions. This meteorological variable can be used in agriculture, environmental management to control conditions affecting plant growth and health and fire risk assessment. VPD is calculated as the difference between the current vapor pressure ($p_v$) and the saturation vapor pressure ($p_{sat}$):

$$VPD = p_{sat} - p_v \qquad (2)$$

Both $psat$ and $pv$ were estimated using the August-Roche-Magnus equation, which is a good approximation for the saturation vapor pressure under typical atmospheric conditions:

$$p_v = 0.61094 * exp(\frac{17.625T}{243.04 + T}) \qquad (2),$$

where T is the air temperature in °C. Then, the current vapor pressure was calculated using the same equation, replacing the air temperature (T) by the dew point temperature ($T_{dew}$). Monthly VDP was then computed for each grid point and month using temperature data from ERA5-Land, following Equation 2.

### 2.4.5. Water Availability

The difference between precipitation (P) and evaporation (E) is the net flux of freshwater between the atmosphere and the Earth's surface, indicating a deficit of water when negative, and a water surplus when positive. This metric can be used as a proxy for both droughts and flooding events.

In the MapBiomas Atmosphere dataset, the water availability ($WA$) was calculated for each gridpoint (x,y) and month (M) as:

$$WA\,(x,\ y,\ M)\ =\ P(x,y,M)\ -\ |E(x,y,M)| \quad (4),$$

where P is the monthly precipitation from GPCC (1985-1997) and GPM (1998-2024) and E is the monthly evaporation data from ERA5-Land.

### 2.4.6. Mean temperature trend

Global climate change is leading to a consistent air temperature increase all over the world. However, the rate of temperature increase varies wildly in space. To assess the variability of temperature rates in Brazil, long term trends in temperature were calculated for each grid point, considering the period 1985-2024. The temperature trends were calculated for the mean, minimum and maximum air temperature, as well for the land surface temperature, based on ERA5-Land annual mean data. The nonparametric Mann–Kendall trend test (Wilks, 2011) was used to detect statistically significant monotonic trends (at a 95% confidence level). The Theil-Sen median-slope regression (Wilks, 2011) was applied to obtain the temperature rate of change (trend) in the period 1985-2024. The python library *pymannkendall* v. 1.4.3 was used for the calculation of temperature trends. Positive trend values indicate warming, while negative values indicate cooling. Trends that were not statistically significant were set to zero.

## 2.5. Validation

Annual average of maximum temperature and annual precipitation from the MapBiomas Atmosphere dataset were validated against ground-based meteorological stations from the Brazilian National Meteorological Institute (INMET). For precipitation, 2,916 stations were used for the validation, while for temperature 292 stations were selected based on data availability. The selected ground-based stations had data quality assured by Xavier et al., (2022) and only stations with a minimum daily data coverage of 75% in the period of interest (1985-2024) were retained. Co-location between model grid and station data was performed using the nearest-neighbour approach (i.e. each station was associated with the closest MapBiomas Atmosphere grid cell).

The annual average of maximum temperature from the MapBiomas Atmosphere dataset reproduced quite well the ground-based observations, with a mean Pearson correlation coefficient of 0.75. The annual average of maximum temperature differs by less than 1ºC between the model and the ground observations in 99% of the meteorological stations. Overall, the accuracy of the maximum temperature in the MapBiomas Atmosphere dataset is within the range ±1ºC.

Annual precipitation was also satisfactorily represented in the MapBiomas Atmosphere dataset, with a mean Pearson correlation coefficient of 0.72 and a relative mean bias of -10.8% (model lower than observations), which can be considered as the overall accuracy of the annual precipitation in the MapBiomas Atmosphere dataset.

## 2.6. Limitations and known biases

The MapBiomas Atmosphere dataset is based on the interpolation of global datasets on atmospheric and surface variables. These global datasets are derived from a combination of in situ and satellite observations, and Earth system model outputs. Although they are able to capture the main spatial and temporal features of atmospheric variables (e.g., Muñoz-Sabater et al., 2021; Silva et al., 2025; Rozante et al., 2018), the global datasets are subjected to biases and limitations.

The MapBiomas Atmosphere dataset satisfactorily represents atmospheric conditions and processes in spatial scales above the so-called mesoscale (>10 km). Microclimatic features, for example, temperature differences between neighborhoods within a city, are typically not represented in the MapBiomas Atmosphere dataset.

In the MapBiomas Atmosphere dataset, temperature-related variables are accurate within ±1ºC, while annual precipitation is accurate within 10.8% (refer to Section 2.5). Temperature may be overestimated over land water bodies, like rivers and lakes, due to limitations in the ECMWF-ERA5 H-TESSEL land surface scheme, which is primarily designed for land tiles (soil + vegetation) (Muñoz-Sabater et al., 2021). In the H-TESSEL model, a “river grid cell” may be classified as land (often “bare soil” or sparse vegetation) rather than water. Precipitation-related variables before the year of 1998, retrieved from GPCC, may present higher biases, especially in regions of complex topography. The water deficit (Section 2.4.5) can be biased due to errors in total evaporation. In Brazil, monthly errors on evaporation from ERA5 Land are typically overestimated by tenths of mm/month, but it can reach 40 mm/month in humid regions of the Amazon (Ruhoff et al., 2022), possibly leading to overestimated water deficit values.

Data on particulate matter (PM) concentrations were retrieved from atmospheric composition reanalysis model outputs. As such, PM concentrations are prone to biases, mostly due to model assumptions on the magnitude and spatial distribution of emission sources in Brazil. An evaluation study for the Brazilian city of Sao Paulo reported 50% and 16% overestimations for PM2.5 and PM10 respectively (Paiva et al., 2025).

# 4. References

Huffman, G.J., E.F. Stocker, D.T. Bolvin, E.J. Nelkin, Jackson Tan (2019), GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: September 24, 2025, doi:10.5067/GPM/IMERG/3B-HH/07

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, Atmos. Chem. Phys., 19, 3515–3556, https://doi.org/10.5194/acp-19-3515-2019, 2019

Karl, T.R., N. Nicholls, and A. Ghazi, 1999: CLIVAR/GCOS/WMO workshop on indices and indicators for climate extremes: Workshop summary. Climatic Change, 42, 3-7.

MapBiomas – Collection 10 of the annual series of Maps of Land Cover and Use of Brazil, accessed on September 24, 2025, through the link: https://brasil.mapbiomas.org/en/colecoes-mapbiomas

Muñoz-Sabater, J., (2019): ERA5-Land monthly averaged data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (date of access: September 24, 2025), doi:10.24381/cds.68d2bb30

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, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Paiva MS, Franco MA, Rizzo LV. (2025) Evaluation of near-surface atmospheric composition reanalysis data in the metropolis of São Paulo, Brazil. Journal of Southern Hemisphere Earth Systems Science 75, ES24041. https://doi.org/10.1071/ES24041

Rozante, J.R.; Vila, D.A.; Barboza Chiquetto, J.; Fernandes, A.D.A.; Souza Alvim, D. Evaluation of TRMM/GPM Blended Daily Products over Brazil. Remote Sens. 2018, 10, 882. https://doi.org/10.3390/rs10060882

Ruhoff, A.; de Andrade, B.C.; Laipelt, L.; Fleischmann, A.S.; Siqueira, V.A.; Moreira, A.A.; Barbedo, R.; Cyganski, G.L.; Fernandez, G.M.R.; Brêda, J.P.L.F.; et al. Global Evapotranspiration Datasets Assessment Using Water Balance in South America. Remote Sens. 2022, 14, 2526. https://doi.org/10.3390/rs14112526

Schneider, Udo; Hänsel, Stephanie; Finger, Peter; Rustemeier, Elke; Ziese, Markus (2022): GPCC Full Data Monthly Product Version 2022 at 0.25°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historical Data. DOI: 10.5676/DWD_GPCC/FD_M_V2022_025

Silva, A.L., Sheridan, S., Ibebuchi, C. et al. Comparing observed and reanalysis data for trends in extreme temperature events in Brazil (1996–2022). Theor Appl Climatol 156, 546 (2025). https://doi.org/10.1007/s00704-025-05777-5

Wilks, Daniel S. Statistical methods in the atmospheric sciences. Vol. 100. Academic press, 2011.

Xavier, A. C., Scanlon, B. R., King, C. W., & Alves, A. I. (2022). New improved Brazilian daily weather gridded data (1961–2020). International Journal of Climatology, 42(16), 8390–8404. https://doi.org/10.1002/joc.7731

Ziese, Markus; Rauthe-Schöch, Armin; Becker, Andreas; Finger, Peter; Rustemeier, Elke; Hänsel, Stephanie; Schneider, Udo (2022): GPCC Full Data Daily Version 2022 at 1.0°: Daily Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data. DOI: 10.5676/DWD_GPCC/FD_D_V2022_100