Development of a data set for distributed hydrological simulation of Paraiba do Sul Basin

Lina M. Osório Olivos¹, Arisvaldo Vieira Mêlo Júnior¹, Silvana Susko Marcellini², Dário Hachisu Hossoda², Rogério Moreira Chagas², Gabriel A. dos Santos Soares², Flávia Ribeiro dos Santos Jacobsem², André Schardong²

¹Hydraulic and Environmental Engineering Section
Polytechnic School – São Paulo University (USP) – São Paulo, SP – Brazil

²Decision Support System Laboratory (LabSid)
Polytechnic School – São Paulo University (USP) – São Paulo, SP – Brazil

{lmosorioo, arisvaldo}@usp.br, {silvana, dario, rogerio, gabriel, flavia}@labsid.eng.br, andreschardong@gmail.com

Abstract. This work describes a dataset created for the purpose of simulating the Paraiba do Sul basin with hydrological distributed models. The dataset contains gridded numerical and descriptive information related to different characteristics of the basin. Distributed hydrological models are useful tools for the management of water resources, as they allow estimating the hydrological regime in a basin according to its physical, rainfall and land use change characteristics, seeking to accurately quantify the availability of water allowing planning future use according to scenarios of land use change or climate change.

Keywords: Gridded Data, Distributed models, Spatio-Temporal series, Hydrology

1. Introduction

Basic data related to climate, topography, land cover and soil characteristics are necessary for the hydrological simulation of basins [Xavier, et.al. 2016]. Quantifying water balance in tropical areas requires knowledge of the influence of heterogeneity in soil-vegetation-atmospheric relationships [Mêlo Junior, et al. 2022]. Distributed models achieve to represent the spatial-temporal variation in the variables that force the hydrological cycle, however the availability of these type of data is a challenge in many countries [Kepeng and Juncang 2021].

Located in the South-East, the Paraiba do Sul (Figure 1) is one of the most significant river basins of Brazil in terms of urban water supply. This interstate basin forms part of the states of Minas Gerais, Rio de Janeiro and São Paulo, and presents particular socio-economics characteristics and a significant increasing in water demand in the last years [OCDE, 2011]. A complete set of base maps have been compiled at 250 m spatial resolution, the gridded data are available at PCRaster format [Karssenberg, et al. 2010] and can be used in models like RUBEM [Mêlo Júnior, et al. 2022], SPHY [Terink, et al. 2015], PCR-GLOBWB [Sutanudjaja, et al. 2018], LISFLOOD [Van der Knijff and Younis 2010], WetSpass [Batelaan and De Smedt 2001] and WFLOW
[Köhler, et al. 2006]. This document gives a brief summary of the source of the geo-spatial data and the applied methodology.

Figure 1. Basin geographic location.

2. Data General Information
- Data Format: GeoTiff, PCRaster raster files and temporal series (*.txt);
- Projection: EPSG:4326 – WGS 84;
- Extent (dd):
  - $x_{\text{min}}$: -46.755;
  - $x_{\text{max}}$: -40.7457;
  - $y_{\text{min}}$: -23.9934;
  - $y_{\text{max}}$: -20.3418.
- Spatial Resolution: 250m;
- Temporal Resolution: monthly;
- Time period: Jan/2000 – Dec/2021;
- Available data:
1. Monthly meteorological raster series of: rainfall [mm/month], potential evapotranspiration [mm/month], evaporation coefficient - kp [-];

2. Monthly filled images of NDVI [-];

3. Maps of gridded of soil characteristics and tables- wilting point [θ (cm³/cm³)], field capacity [θ (cm³/cm³)], hydraulic conductivity [mm/month], depth root zone [cm], saturated capacity water content [θ (cm³/cm³)], soil bulk density [g/cm³];

4. Annual land use raster maps and land use parameters tables;

5. Digital elevation model.

3. Meteorological Data

The meteorological data necessary to run the model is summarized in:

- Rainfall;
- Potential Evapotranspiration;
- Class A Pan Coefficient (Kp).

Rainfall was obtained from stations distributed in each basin area, 389 stations were used for the spatialized product (Figure 1), regional vector method was used for filling gaps after a data evaluation of the raw data. Point measures were converted to raster data by kriging interpolation and variogram parameters were optimized using the SciPy-library algorithm in a Python script. For potential evapotranspiration, point values (at station locations) were calculated using the Penman-Montheit method, data from NASA POWER Project were used to fill gaps and kriging interpolation were used to generate the raster series. The Class A pan coefficient was calculated using the wind speed, and relative humidity.

4. Digital Elevation Model

The raster map with elevation has as data resource the global elevation data at 1 arc second spacing NASADEM [NASA 2020]. NASADEM data products were derived from original telemetry data from the Shuttle Radar Topography Mission (SRTM). These NASA project applied multiples methodologies for improve the accuracy of the data, using images from other resources to filling gaps and sinks.

5. Land Use Maps and Parameters

The database available in the Annual Mapping of Land Cover and Land Use Project in Brazil (MapBiomas) - Collection 6 were used as source of data. MapBiomas images are based on machine learning algorithms for classification of land cover, resulting in to 49 classes and subclasses, for Collection 6 products, legend is available at https://mapbiomas-br-site.s3.amazonaws.com/C%C3%B3digos_Classes_Legenda_Cole%C3%A7%C3%A3o_7_-_PT.pdf. Some parameters associated to each type of land use are available in txt tables, these parameters are related to vegetated, soil bare, open water and
impervious fraction areas and Manning roughness (see Mello Junior et al. 2022 for more information).

The characterization and analysis of the historical series of watershed coverage (LULC) and the assignment of modeling parameters involved the following steps:

1. Download the series of coverages for the period 2000 – 2021 from Mapbiomas collection 6 (2021 was provided by collection 7) through the GEE platform;
2. Preparation of files to PCRaster format;
3. Assignment of features for modeling (area fractions per pixel and Manning roughness).

6. Soil Data

The different soil types in the basin were established from a reclassification of available soil maps from the States of São Paulo, Rio de Janeiro and Minas Gerais. The differentiation of the soils considering their depth and textural class (from semi-detailed surveys) greatly influences their hydrological behavior. The reclassification also sought to simplify the number of soil types, as the great diversity of associations in the surveys made it impossible to calibrate a hydrological model.

The primary data source for obtaining the soil parameter was the HYBRAS database [Ottoni, 2018], a hydro-physical database of soils in Brazil, with data on water retention and saturated hydraulic conductivity. Other references to obtain averages for each parameter were also included. Maps and tables are available for the parameter mentioned in section 2. Some type of soils present in the area are: Argissolos (Ultisols), Cambissolos (Inceptisols), Latossolos (Oxisols), Organossolos (Histosols), Neossolos (Entisols quartzipsamment), Espodossolos (Spodosols) and Gleissolos (Entisols - Aqu-alf-and-ent-ect). Raster maps of each parameter and a table with the legend associated are available in the dataset.

7. NDVI Data

NDVI describes the normalized ratio between the near-infrared and red bands, formally described according to Equation 1:

\[
NDVI = \frac{\rho_{\text{nir}} - \rho_{\text{red}}}{\rho_{\text{nir}} + \rho_{\text{red}}}
\]

Where \(\rho_{\text{nir}}\) corresponds to the value of the near-infrared band and \(\rho_{\text{red}}\) to the value of the red band of the electromagnetic radiation spectrum. The NDVI index allows monitoring seasonal and interannual changes in vegetation development and activity, being widely used in environmental and natural science research [Jensen, 2009].

Based on the analysis of image availability, the product MODIS - MOD09A1.061 Terra Surface Reflectance 8-Day Global 500 m, was selected to obtain NDVI data. Images collected presented gaps in some areas as a filter for cloud cover and pixel quality were applied.
7.1 Gap-Filling in NDVI images - Proposed Methodology

A simplified methodology for NDVI gap-filling was developed to obtain full monthly images with consistent information linked to annual coverage maps. This association proved coherent. Even if the frequency of coverage is annual, the reading of NDVI monthly reflects the characteristics and state of maturity of the vegetation cover.

The developed methodology consists of 4 steps:

- **Step 1: NDVI image pairing with faults and associated coverage map**: In this step, the monthly NDVI image is associated with the corresponding annual coverage map;

- **Step 2: Obtaining the average NDVI values by coverage class**: From the valid or flawless pixels of the NDVI image, a reading associated with each coverage class of the annual map was taken to obtain the average NDVI values for the month of analysis;

- **Step 3 – Definition of filling criteria**: The criterion for gap-filling in faulty pixels is selected after compiling the monthly average NDVI values for each coverage class. Three criteria were listed for filling gaps, which depend on flawless pixel and coverage availability, in the month under analysis. The criteria are follows:
  - Criteria 1: filling pixels with flaws associated with the average NDVI values obtained in the same month, by coverage class;
  - Criteria 2: completion with mean NDVI values obtained in the immediately preceding and subsequent months, by coverage class;
  - Criteria 3: completion with annual average NDVI values, by coverage class.

- **Step 4 – Filling the gaps**: Finally, monthly NDVI gaps were filled in according to the criterion selected for filling in a group of pixels associated with a given coverage class.

8. Data Availability

The dataset is available at HydroShare web based hydrologic information system [https://www.hydroshare.org/resource/07d089fccc88345b4846f35fa0eced92/](https://www.hydroshare.org/resource/07d089fccc88345b4846f35fa0eced92/). Additional information about data processing (methodologies, scripts) are available in metadata files in the folders database and at [https://rubem.readthedocs.io/en/latest/preprocessing.html](https://rubem.readthedocs.io/en/latest/preprocessing.html).

References


