

## Plan Overview

---

*A Data Management Plan created using DMPTool*

**Title:** Climate effects in fauna and flora in the Amazon Basin central region near Manaus

**Creator:** andre marques

**Affiliation:** Universidade de São Paulo (www5.usp.br)

**Principal Investigator:** Shell Oil Company

**Funder:** São Paulo Research Foundation (fapesp.br)

**Template:** Digital Curation Centre (português)

**Project abstract:**

The project plans to build up correlations between bioclimatic data and flora and fauna data, using Machine Learning and Artificial Intelligence tools. Furthermore, the correlations may help other research projects related to the influence of the Amazon Rain Forest in the weather and climate changes detected recently.

**Start date:** 01-05-2022

**End date:** 01-05-2025

**Last modified:** 02-16-2022

**Copyright information:**

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

---

## **Climate effects in fauna and flora in the Amazon Basin central region near Manaus**

The data to be collected consider concentrations of pollutants, within time and spatial coordinates, related to the fauna and flora specimen. For instance, concentration of Ozone, SO<sub>2</sub>, CO, acetonitrile, particulates with diameters down to 10 nanometers, NO<sub>x</sub>, among others. The data of fauna and flora will concentrate on species to be assigned by specialists and technical advisors.

The data will be collected as source datasets, from research institutions or from specific state/private organizations related to the theme.

The metadata associated to the data and project will focus information from technical advisors, data sources (open/private) and auxiliary data references about bioclimate subjects.

The project will observe all the legal procedures and social recommendations regarding all the social, ethic, ethnical issues, commonly seen in this kind of research project. The data will come mostly from open sources.

All copyright steps of this research project will follow the instructions from Universidade de São Paulo (USP).

The data generated by this research project will be stored in open access repositories, such as the ARM-DOE/USA and others linked to the Universidade de São Paulo (USP) and to the associated research centers and universities.

The data generated by this research project will be open access, following the repositories guidelines too.

The input datasets, generated by interpolation techniques with Machine Learning and Artificial Intelligence, and the outputs generated by the correlations identified between bioclimate, fauna and flora data of the Amazon Basin Central Region.

The long term data storage plan considers the access to open access repositories and their guidelines, with version control following the Github concepts.

The data of this research project will be shared throughout the open access repositories or by online request.

No restrictions have been identified so far. Review of this statement may apply according the evolution of the research project.

The main researcher (André Marques) and his thesis advisor (Dr. Pedro Correa).

Funds provided by FAPESP related to this research project.

---

## Planned Research Outputs

### Dataset - "Bioclimatic dataset related to the Amazon Basin Central Region near the city of Manaus"

*The current research deals with environmental data related to the region around the city of Manaus-Brazil, located in the central area of the Amazon Basin.*

*The research problem encompasses the answers to the questions: 'What correlations can arise from the data about bioclimatic and fauna in the Amazon Basin central region?' 'How useful is the data from research stations related to that area, covering the fauna and bioclimatic boundary conditions?'*

*In the answer to the questions, an initial dataset was generated to assess the relations between climate and fauna/flora of that region.*

*Thus, the main data sources were from NASA Satellites, Amazon Tall Tower Observatory (ATTO) experiment and the GOAmazon project, from 2014 to 2015, along with other local data, such as the position of thermal electric power stations.*

*It has been worth noting the data available was not coherent among the sources, due to different measuring time intervals, heights and geographical locations.*

*Equally important, for future bioclimatic research, a grid space of 5km interval and a 30 minute time period was considered within the dataset, covering a surface*

*area within the boundaries of Latitude 2.00 S to 4.25 S, and Longitude 59.00 to 61.25 W. The data from NASA Satellites, ATTO and GOAmazon Projects do not have this kind of detailed grid and time mesh. To carry on, the machine learning (ML) Random Forest (RF) algorithm was selected and used, among other ML algorithms, to fill in the data into this new grid. The features considered were: data from NASA Satellites and from the two projects above, covering the pollutants: ozone, NOx, particulates with diameter down to 10 nanometers, isoprene, CO and acetonitrile. Moreover, distances and bearings to Manaus and eleven thermal power stations nearby, night and day time periods, dry and wet seasons, were also considered to compose the features.*

*The interpolation produced by the Random Forest algorithm aimed at the pollutants: acetonitrile, ozone, NOx, Particulate N, isoprene and CO, which are related to bioclimatic research commonly. In total, a dataset was created with 51,148,800 rows, with 44 columns, 38 for the features and 6 for the targets.*

---

### Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Bioclimatic dataset related to the Amazon Basin Ce ...	Dataset	Unspecified	Open	None specified		None specified	None specified	No	No