**RESEARCH PROTOCOL**

***Name:*** **Fernando Venâncio Mucomole**

**1.** *Title:*

**"PARAMETRIC FORECAST VARIABILITY OF SOLAR ENERGY OVER TIME APPLYING MACHINE LEARNING TECHNIQUES"**

## 2. Background and Rationale:

### 2.1. Background:

The energy produced at the output of a solar plant is significantly impacted by the various atmospheric, geographical, climatic, and spatiotemporal elements (Duffie & Beckman, n.d.; Mucomole et al., 2024a) that influence the shape of solar energy when it reaches the earth's surface (Iqbal, 1983; Mucomole et al., 2024b). Most communities would have more availability to clean energy for electrification, heating, medical care, healthcare, and other uses if they knew the true parametric variability of forecast solar energy (Prasad et al., 2019). This would also significantly boost the durability of solar systems (Mucomole et al., 2023; Page et al., 2021; Wenham et al., 2011).

### 2.2. Rationale:

The extent to which the different ecosystems of nature are impacted by climate change is of great interest. along with the pollution that human activity has brought about in the environment. A parametric model is used in this study to quantify the variability of predicted solar energy. This study will investigate the forecasting of solar energy's temporal and spatial variability, its relationship to regional climate conditions, and its possible impacts on human health and biodiversity. The findings will help refine solar energy parameters and public policy models for solar energy projects and dimensions, as well as for architectural design, climate resilience plans, fishing, aquaculture, and health advisory systems.

### 2.3. Research Questions:

1. In Response to the query, what factors go into the accurate forecast of Solar Energy Availability Using MLM?
2. How does the variability of Forecasted solar energy behave?
3. How does solar radiation influence local ecosystems, particularly plant and animal species?

### 2.4. Objectives:

#### 2.4.1. Overall objective

- Forecast parametrically Solar Energy Over Time Applying Machine Learning Techniques

### 2.4.2. Specific objective

- Systematize the methodology and the respective solar energy estimation models;
- Describe and model the parametric model of solar energy forecast;
- Apply machine learning models (MLM) to better estimate energy;
- Analyze the spatial and temporal energy copy of energy obtained;
- Interpret the results of correlational and regressional solar energy variability;
- Apply casual inference for analysis of solar energy forecast results.

## 3. Study Design and Methodology:

### 3.1. Study Design:

Focusing on the analysis of spatial variability, the research will be diverse, encompassing everything from the analysis of solar energy estimating models to the application of the parametric model that takes into account all characteristics that prevent the arrival of solar energy, but the clear sky index is crucial. by calculating the predicted solar energy clear sky index, which eliminates all solar energy variability caused by solar geometry and other variables.

### 3.2. Research Type:

The research is of a quantitative type and is based on analysis of the clear sky index of solar energy forecast parametrically using the parametric method, with a measurement interval of 1 to 10 minutes.

### 3.3. Study Population:

The analysis population in the research houses 11 measuring stations in around 11 provinces of Mozambique (study area), measured during the years 2012 to 2014 and 2019 to 2020 interpolated in the hours of solar energy incidence (6 am to 6 pm hours) , with a measurement interval of 1 to 10 minutes.

### 3.4. Sampling Method:

**3.4.1. Solar Radiation Measurement:** We will install solar radiation sensors (pyranometers) in all provinces of Mozambique, with a focus on provincial capitals, representing 11 measurement stations, within the scope of the Mozambique National Energy Fund measurement campaign.

**3.4.2. Atmospherical Data:** data on atmospheric parameters and their measurement error estimates were extracted from the AERONET database, providing measurements on: Aerosol Optical Thickness (AOT), water vapor, uniformly mixed gases, layer of ozone, among others.

**3.4.3. Climate and meteorological data:** Additional data on solar energy, and climatic parameters such as temperature, pressure, vaporization indices, water vapor, among others, were collected in all provincial capitals, from the measurement stations of the National Institute of Meteorology of Mozambique (INAM).

**Geographical data:** They were extracted from the previously described databases, based on global geographic mapping coordinates.

### 3.5. Study Setting:

- In order to investigate varying degrees of solar radiation exposure, the study will be carried out in a number of cities, including both urban and rural areas of Mozambique as a technique testing and validation site.

## 4. Data Collection and Instruments:

### 4.1. Solar Radiation Measurement:

- Use Epley pyranometers to measure global solar radiation ($W/m^2$). Sensors will be installed at various locations, with readings taken every 30 minutes throughout the study period.

### 4.2. Atmospherical and Climate Data:

- Collect data on local temperature, aerosol, water vapor, ozone, gases, humidity, and wind speed using, weather station's location, from INAM and ARONET.

## 5. Data Analysis Plan:

### 5.1. Statistical Analysis:

- ***Testing the sample collected:*** To assess ambiguities in the sample and avoid the presence of outliers, the data sample will be tested to determine its real distribution and election of the statistical treatment to be treated.

- ***Parametric Model:*** forecast parametrically the solar energy consider all Atmospherical parameters. Clear sky index determination and grouping.

- ***ANOVA:*** To compare solar radiation intensity across different days types (clear, cloudy and intermediate).

- ***Regression Analysis:*** To assess the relationship between predicted solar radiation levels in two pairs of stations using the random sample of clear–sky index.

- ***Correlation:*** between predicted solar radiation levels in two pairs of stations using the random sample of clear–sky index.

### 5.2. Spatial Analysis:

- Use of clear sky index statistical sample for statistical determination of special solar energy intensity. Spatial interprovincial relationship of solar energy intensity.

### 5.3. Solar energy Outcome Analysis:

- Solar energy forecasted variability in different locations of Mozambique province and its incidence levels.

## 6. Ethical Considerations:

### 6.1. Informed Consent:

- The projects on which researchers collect data, namely: FUNAE, INAM, AERONET and the Eduardo Mondlane University, will be fully informed about the objective of the study, its role and the nature. Informed consent will be obtained prior to any data collection.

**6.2. Confidentiality:**

Not applicable.

**6.3. Ethics Approval:**

This study will be reviewed and approved by an Institutional Review Board (IRB) or Ethics Review Committee before commencement.

**6.4. Risks and Benefits:**

There are minimal since the researchers carried out a preliminary feasibility study and identified all needs, in addition to the fact that the sample will be processed without physical exposure, in addition to the fact that the study involves non-invasive methods of data collection. With the study, you can learn about the variability of solar energy resources for various applications, from improving the output of a solar plant, to architectural, construction, fishing, aquaculture, and health applications, among others.

**7. Timeline and Milestones:**

| Phase | Duration | Milestone |
|---|---|---|
| Study Design & Setup | 2 months | Finalized protocol, site selection, and equipment procurement. |
| Data Collection | 6 months | Begin data collection (solar radiation, health surveys, ecological assessments). |
| Data Analysis | 3 months | Complete statistical and spatial analysis. |
| Final Report & Dissemination | 2 months | Draft final report, submit to journals, and present at conferences. |

**8. Budget and Resources:**

**8.1. Equipment:**

- Pyranometers (11 units): $110,000
- UV meters (5 units): $5,000
- Weather stations (11 units): $7,000
- Data analysis software (e.g., SPSS, GIS tools): $4,000
- Travel and fieldwork costs: $10,000

**8.2. Personnel:**

- Principal Investigator (1): $60,000
- Research Assistants (2): $48,000
- Field Technicians (2): $22,000
- Total Budget: $125,000

**9. Dissemination Plan:**

**9.1. Publication:**

The study's findings will be submitted to peer-reviewed journals such as: energies–MDPI energies and Advancing in Energies–Elsevier.

**9.2. Conference Presentations:**

Present results at relevant conferences, including the: International Conference of material energies and environmental Engineering (ICEEE), Physics Society, SAMSA –Southern Afrina Mathematical Association.

**9.3. Public Engagement:**

Create a website to share key findings with the public, and work with local energy organizations to distribute solar energies maps materials.

**10. Registration Information:**

**10.1. Protocol Registration:**

The protocol will be registered with PROSPERO (for systematic reviews) or Open Science Framework (OSF) to ensure transparency and public access to the research methodology.

**11. Conclusion:**

This research protocol provides a comprehensive approach to studying the variability of forecasted solar energy using machine learning models, focusing on atmospheric parameters that reduce the intensity of solar energy reaching Earth's surface such as: aerosols, water vapor, ozone layer, mixed uniformly gases among others. By measuring solar radiation levels under various conditions, determining the clear sky index and statistically separating day types, the study aims to provide valuable information for energy planning, solar photovoltaic energy, public health policies and management strategies. adaptation to the current panorama of climate change.

**12. Bibliographic references**

Duffie, J. A., & Beckman, W. A. (n.d.). *Solar Engineering of Thermal Processes*.
Iqbal, M. (1983). *An introduction to solar radiation*. Academic Press.
Mucomole, F. V., Silva, C. A. S., & Magaia, L. L. (2023). Temporal Variability of Solar Energy Availability in the Conditions of the Southern Region of Mozambique. *American Journal*

*of Energy and Natural Resources*, *2*(1), Article 1. https://doi.org/10.54536/ajenr.v2i1.1311

Mucomole, F. V., Silva, C. A. S., & Magaia, L. L. (2024a). Regressive and Spatio-Temporal Accessibility of Variability in Solar Energy on a Short Scale Measurement in the Southern and Mid Region of Mozambique. *Energies*, *17*(11), 2613. https://doi.org/10.3390/en17112613

Mucomole, F. V., Silva, C. S. A., & Magaia, L. L. (2024b). Quantifying the Variability of Solar Energy Fluctuations at High–Frequencies through Short-Scale Measurements in the East–Channel of Mozambique Conditions. *American Journal of Energy and Natural Resources*, *3*(1), Article 1. https://doi.org/10.54536/ajenr.v3i1.2569

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1), 89. https://doi.org/10.1186/s13643-021-01626-4

Prasad, R., Ali, M., Kwan, P., & Khan, H. (2019). Designing a multi-stage multivariate empirical mode decomposition coupled with ant colony optimization and random forest model to forecast monthly solar radiation. *Applied Energy*, *236*, 778–792. https://doi.org/10.1016/j.apenergy.2018.12.034

Wenham, S. R., Green, M. A., Watt, M. E., Corkish, R., & Sproul, A. (Eds.). (2011). *Applied Photovoltaics* (3rd ed.). Routledge. https://doi.org/10.4324/9781849776981