**Forecasting Solar radiation using hybrid ARIMA-ANN model**

**Abstract:** Solar radiation prediction has a great importance in electricity generation from solar energy and helps to size photovoltaic power systems. Therefore, the Global Horizontal Irradiance (GHI) was predicted at 1-hour duration in this paper.

Direct Normal Irradiance (DNI), Direct Horizontal Irradiance (DHI), Temperature, Pressure, Relative Humidity, Dew Point, Wind Direction and Wind Speed parameters were used as atmospheric input variables for time series model by AutoRegressive Integrated Moving Average(ARIMA) model. ARIMA is one of the popular linear models in time series forecasting.

As time series also have non linearity hence, a hybrid model is made afterwards using residuals of ARIMA as input for Artificial Neural Networks. Statistical error measures such as the mean error (ME), the mean square error (MSE) and the mean absolute error (MAE) were calculated to compare the two methods.

The results showed that the Hybrid models predict the solar radiations with a higher accuracy than the ARIMA model in the four examined sites.

Keywords: SARIMAX; Box–Jenkins methodology; Artificial neural networks; Time series forecasting; Combined forecast, hybrid SARIMAX-ANN

Table of contents

Title Page No.

Acknowledgement 2

Abstract 3

List of tables/figures 5

Abbreviations and Nomenclature 6

Chapter 1: Introduction 7

Chapter 2: Software Analysis 10

2.1: Software

2.2: Hardware

2.3: Functional Requirements

2.4: Non-Functional Requirements

Chapter 3: Detailed Design 11

4.1: SARIMAX

4.2: Artificial Neural Network

4.3 Hybrid model

Chapter 5: Observation 12

Chapter 6: Implementation 13

Chapter 7: Testing Reports 14

Chapter 8: Conclusion 17

Chapter 9: The proposed work plan for the remaining period 18

Chapter 10: References 19

Chapter 11: PowerPoint Presentation 20

List Of Figures/Tables

Figure Title Page

6.1 Implementation Of Model 13

7.1 Actual v/s Predicted emotion 14

7.2 Confusion Matrix of Speech Based Model 15

7.3 Confusion Matrix of Facial Based Model 15

7.4 Accuracy of Facial Images Based Model 16

9.1 Gantt Chart Of Implemented Work 18

**ABBREVIATIONS AND NOMENCLATURE**

SARIMAX: Seasonal Auto Regressive Integrated Moving Average with Exogenous variable

ANN: Artificial Neural Networks

RMSE: Root Mean Squared Error

MAAPE: Mean Absolute Arc-tangential Error

**Chapter 1: Introduction**

Solar radiation prediction is an important problem with direct applications in renewable energy. Solar is one of the most important green sources of energy, that is currently under expansion in many countries of the world, especially in those with more solar potential, such as Rajasthan. An accurate estimation of the energy production in solar energy systems involves the accurate prediction of solar radiation, depending on different atmospheric variables. We have used humidity Direct Normal Irradiance (DNI), Direct Horizontal Irradiance (DHI), Temperature, Pressure, Relative Humidity, Dew Point, Wind Direction and Wind Speed as inputs ([Khatib](https://www.sciencedirect.com/science/article/abs/pii/S1364032112000767) et al., 2012; [Inman](https://www.sciencedirect.com/science/article/pii/S0360128513000294) et al., 2013; [Sozen](https://www.sciencedirect.com/science/article/pii/S0196890404000172) et al., 2004; [Voyant](https://www.sciencedirect.com/science/article/abs/pii/S0360544210005955) et al., 2011).

In recent years, several works have been developed to try to predict solar radiation using machine learning techniques and environmental parameters. They used different input geographical and atmospheric parameters like latitude, longitude, temperature, wind speed and direction, daily global irradiation, sunshine duration or precipitation ([Mellit and Kalogirou](https://www.sciencedirect.com/science/article/pii/S0360128508000026), 2008; [Mubiru](https://www.sciencedirect.com/science/article/abs/pii/S0960148108000074), 2008). According to [Bilgili and Ozoren](https://link.springer.com/article/10.1007%2Fs00703-011-0137-9) (2011), sunshine duration, air temperature and relative humidity are the most widely used meteorological parameters to predict daily solar radiation and its components

An initial category to distinguish is time-series models, which can be further divided into Box and Jenkins techniques, smoothing techniques, Kalman-ﬁltering theory, and spectral analysis. Early applications of Box and Jenkins techniques in the ﬁeld of traffic forecasting were implemented by [Ahmed and Cook](https://trid.trb.org/view/148123) and [Nihan and Holmesland](https://sci-hub.se/https:/link.springer.com/article/10.1007/BF00167127). More advanced techniques have been applied recently, including autoregressive integrated moving average (ARIMA) models with intervention x-variables (ARIMAX) by [Williams, B. M., and L. A. Hoel](https://sci-hub.se/https:/ascelibrary.org/doi/abs/10.1061/(ASCE)0733-947X(2003)129:6(664)), seasonal ARIMA models (SARIMA) by [Williams, B. M](https://sci-hub.se/https:/ascelibrary.org/doi/abs/10.1061/(ASCE)0733-947X(2003)129:6(664)), seasonal ARIMA with intervention x-variables(SARIMAX) by [SI Vagropoulos, GI Chouliaras](https://sci-hub.se/https:/ieeexplore.ieee.org/abstract/document/7514029/).

Neural network models are the second category of techniques that can be identiﬁed. [Y Jiang](https://sci-hub.se/https:/www.sciencedirect.com/science/article/pii/S0301421508003133) was among the first who applied ANN on solar radiation prediction. [Amit Kumar Yadav , Hasmat Malik](https://sci-hub.se/https:/www.sciencedirect.com/science/article/pii/S1364032113008228) wrote about the most relevant input parameters for artificial neural network based solar radiation prediction models.

A suitable combination of linear and nonlinear models provides a more accurate prediction modelthan an individual linear or nonlinear modelfor forecasting time series data originating from various applications. The linear autoregressive integrated moving average (ARIMA) and nonlinear artificial neural network (ANN) models are explored in this paper to devise a new hybrid ARIMA–ANN model for the prediction of time series data. The results obtained from all of these data sets show that for both one-step-ahead and multistep-ahead forecasts, the proposed hybrid model has higher prediction accuracy.

A hybrid ARIMA–ANN model was proposed by [Zhang](https://www.sciencedirect.com/science/article/abs/pii/S0925231201007020?via%3Dihub) [14], which was shown to give more accurate predictions than the individual models. On Wolf’s sunspot data, Canadian lynx data, and exchange rate time series data, this hybrid model was shown to outperform individual ARIMA and ANN models in the case of one-step-ahead prediction. Another hybrid ARIMA–ANN method was proposed by [Khashei and Bijari](https://www.sciencedirect.com/science/article/abs/pii/S1568494610002759?via%3Dihub) [15], which was shown to give better performance for one-step-ahead forecasting than the method proposed by Zhang [14]. The hybrid method proposed by Zhang was also used for [electricity price forecasting](http://dx.doi.org/10.1049/iet-gtd.2012.0263) in [16] and for [water quality time series prediction](https://www.sciencedirect.com/science/article/abs/pii/S0952197609001390?via%3Dihub) in [17].

For error detection,we have used MAAPE and RMSE values. The mean absolute percentage error (MAPE) is one of the most widely used measures of forecast accuracy, due to its advantages of scale-independency and interpretability. However, MAPE has the significant disadvantage that it produces infinite or undefined values for zero or close-to-zero actual values. In order to address this issue in MAPE, we have used a new measure of forecast accuracy called the mean arctangent absolute percentage error (MAAPE) by [Sungil Kima and Heeyoung Kim](https://sci-hub.se/https:/www.sciencedirect.com/science/article/pii/S0169207016000121). MAAPE has been developed through looking at MAPE from a different angle. In essence, MAAPE is a slope as an angle, while MAPE is a slope as a ratio, considering a triangle with adjacent and opposite sides that are equal to an actual value and the difference between the actual and forecast values, respectively. MAAPE inherently preserves the philosophy of MAPE, overcoming the problem of division by zero by using bounded influences for outliers in a fundamental manner through considering the ratio as an angle instead of a slope.

Chapter 2: Requirement Analysis

3.1 Software Requirements

3.1.1 Library used:

• Numpy

• Pandas

• OpenCV

• Keras

• Sklearn

3.1.2 Other Requirements:

• Anaconda Platform (spyder, Jupyter Notebook)

• Python 3 or higher

3.2 Hardware Requirements

• Microsoft Windows 10

• Processor: Intel ® Core (TM) i5 -6200U CPU @2.30GHz 2.40GHz

• Ram : 4 GB and above

• Disk Space : 1 TB

3.3 Functional Requirements

• Appropriate data set to work on.

• An eligible software to implement our ideas.

• Suitable libraries for using algorithm in the source code.

3.4 Non-functional Requirements

• Validation: Applying benchmark functions for checking the accuracy and performance of algorithms.