IEC
Institute of Electronics
and Computer

Journal of Artificial Intelligence and Systems, 2025, 7, 76-88
https://iecscience.org/journals/AIS
ISSN Online: 2642-2859

STEPS: A Tool to Forecast Meteorological Variables on a Short Time Scale

R. Shrivastava[1], Indumathi S. Iyer[1] and Anmol Batra[1]

[1] Radiation Safety Systems Division, Bhabha Atomic Research Centre, Mumbai - 400 085, India
Email: roopa@barc.gov.in, indumati@barc.gov.in, abatra@barc.gov.in
*Corresponding Author: R. Shrivastava, Email: roopa@barc.gov.in

How to cite this paper: R. Shrivastava, Indumathi S. Iyer and Anmol Batra (2025). STEPS: A Tool to Forecast Meteorological Variables on a Short Time Scale. Journal of Artificial Intelligence and Systems, 7, 76–88. https://doi.org/10.33969/AIS.2025070105.

Received: October 27, 2025
Accepted: November 17, 2025
Published: November 21, 2025

Abstract

Accurate prediction of meteorological variables like air temperature is important in various sectors like agriculture, forecast of energy demand, climate change, transportation etc. Air temperature forecasts are also crucial in determination of heat and cold waves. Across the globe as demand for solar energy is increasing, there is a need for precise forecasts of solar radiation too. In recent times, machine learning methods are becoming popular in weather forecasting. This study describes the development of Short Term Prediction System (STEPS), a model based on Auto Regressive Integrated Moving Average (ARIMA) technique suitable for prediction of hourly values of meteorological variables like air temperature, relative humidity, solar and net radiation with a lead time of three days. The model has been validated at a single point using two years of meteorological measurements. At a lead time of one day, mean absolute error in air temperature forecast is less than 2 °C and for relative humidity is less than 11 % throughout the year and. For solar and net radiation at the same lead time, respective mean absolute errors are less than 30 W m-2 during non-monsoon season and approximately 75 W m-2 during monsoon season. Hence, model results indicate that the forecasting proficiency is comparable to traditional Numerical Weather Prediction (NWP) model at a fraction of computational cost and time.

Keywords

Short term, meteorological variables, ARIMA, machine learning, time series forecasting

1. Introduction

Rising demand for electricity, increased greenhouse gas emissions, climate change and depleting reserves of fossil fuels have led to growing interest in environmentally sustainable energy generation sources in residential, commercial and industrial sectors. Solar energy forms, one of the low-cost, infinite and freely available renewable energy resource. Forecasting solar energy on hourly time scale is challenging due to various influencing factors like weather conditions, geographical location, season, time of day, latitude, cloud cover, presence of atmospheric aerosols and dust particles. Also, solar energy is an intermittent energy source, and small variation in solar irradiance can impact the electricity production. Hence, accurate forecasts of solar radiation can improve solar power operations and substantially reduce costs. Likewise predictions of meteorological variables like air temperature, relative humidity are important in agriculture and design of heating and cooling management systems installed in buildings. Tradionally, weather forecasts have relied on Numerical Weather Prediction (NWP) Models. In recent times, due to rapid advances in data science, the application of machine learning methods in weather forecasting is increasing. In the last decade several studies have focused on the development of models based on machine learning for weather forecasting. For example, Chen et al. [1] have used Seasonal ARIMA (SARIMA) method to forecast monthly mean air temperature in Nanjing city of China during 1951-2017. Their results indicated that the forecasting accuracy for monthly mean air temperature was acceptable. Murat et al. [2] have used seasonal ARIMA technique, ARIMA with Fourier terms, and time series regression method for forecasting daily temperature at four sites in Europe. The study duration extended from 1st January 1980 to 31st December 2010. The period from 1st January 1980 to 31st December 2004 was the training data set and the remaining the test data set. The results showed that ARIMA along with Fourier terms was found to have the highest prediction accuracy among the various techniques used. Astsatryan et al. [3] have ANN to implement a weather prediction technique to improve the hourly air temperature prediction with a lead time of 24 hours at Ararat valley in Armenia. This study utilized observations from several meteorological stations as well as satellites. Experiments were conducted with several configurations of neural networks with a lead time of 24 hours and the study concluded that the best model has an accuracy of ~ 75 % for forecast of air temperature. Abhishek et al. [4] have applied ANN model with different parameters like transfer functions, hidden layers and neurons to forecast maximum, temperature for 365 days of the year with the error being in acceptable range. Baboo and Shereef [5] have used back propagation neural network to develop a forecasting system for atmospheric temperature. The training data set was for one year and then the model was tested using predictions for unseen days. Fara et al. [6] have used ARIMA and ANN based models for ten days forecast of solar radiation at a site in south Romania. Their results indicate that ARIMA model was more efficient as compared to ANN model with cloudy days being associated with larger errors as compared to clear sky days. Ettayyebi and Himdi [7] have used ARIMA and ANN with Multi Layer

Perceptron (MLP) models including exogenous variables for forecasting day ahead solar radiation at Rabat city in Morocco. The study period was from 2006-2016. They concluded that MLP with exogenous variables performed better as compared to other models. Daily and monthly average global solar radiation in Seoul, South Korea was predicted using seasonal ARIMA model based on hourly solar radiation measurements during 1981-2017 [8]. Seasonal ARIMA model was able to predict accurately, solar radiation at daily and monthly average time scales. Similarly, Chodakowska et al. [9] have used ARIMA model for forecasting solar radiation at monthly and hourly time scales at Amman in Jordan and Warsaw in Poland. For each location, a suitable ARIMA model was identified, its parameters were estimated and forecasts carried out. Apart from above, numerous other studies on the same topic can be accessed in [10, 11, 12, 13, 14, 15, 16, 17].

In the present study, we describe the development and application of a model namely Short Term Prediction System (STEPS) suitable for hourly prediction of meteorological variables at a lead time of three days. STEPS is based on the well-known ARIMA algorithm. Section 2 describes the ARIMA method in brief. Section 3 describes the STEPS model. Various case studies for validation of STEPS are presented in Section 4 and finally conclusions are presented in Section 5.

2. Material and Methods

Time series is a collection of observations over regular intervals through repeated measurements. Time series analysis and forecasting has several applications in statistics, climate science, economics, market sales etc. Forecasting future values of time series is possible using simple methods like average, naïve, seasonal naïve to slightly more complex like ARIMA and ANN. Time series analysis was introduced by Box and Jenkins which include Auto Regression (AR) Model, Moving Average (MA) Model and Auto Regressive Moving Average (ARMA) [18]. In an autoregression model, we forecast the variable of interest using a linear combination of past values of the variable. Hence, an AR model of order p is defined as:

$$\boldsymbol{y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots \varphi_p y_{t-p} + \varepsilon_t} \quad (1)$$

Rather than using past values of the forecast variable in a regression, a MA model uses past forecast errors in a regression-like model. Hence, a MA model of order q is defined as:

$$\boldsymbol{y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}} \quad (2)$$

In Equation (1), using different values for parameters $\boldsymbol{\varphi_1, \dots \dots . \varphi_p}$ will result in different time series patterns. The variance of the error term $\boldsymbol{\varepsilon_t}$ will only change the scale of the series, not the patterns.

ARMA model is the combination of AR and MA models. However, ARMA requires that the input time series is stationary. Time series of meteorological observations are always non-stationary. Therefore, non-stationary time series have to be transformed into stationary time series before the application of ARMA model. This can be carried out by the taking logarithm or d^{th} order differentiation of the time series. Therefore, the addition of integration term in ARMA gives rise to ARIMA and allows for ARMA to be used with non-stationary time series. An ARIMA model is defined by the order (p, d, q) where, p is the order of the autoregressive term, d is the order of differencing and q is the order of the moving average term. A generalized ARIMA equation can be written as:

$$\left(1-\varphi_1 B-\cdots-\varphi_p B^p\right)(1-B)^d y_t=\left(1+\theta_1 B+\cdots+\theta_q B^q\right)\varepsilon_t \quad (3)$$

STEPS model carries out time series analysis using ARIMA technique in RStudio platform. At the heart of an ARIMA model is the auto.arima () function. The auto.arima () function in R uses a combination unit root tests for Stationarity and minimization of MLE (maximum likelihood estimate) to obtain the arguments of the ARIMA model. The arguments to auto.arima () provide for many variations on the algorithm. The Akaike information criterion (AIC) is a mathematical method used to assess how well a model fits the data it was generated from. In statistics, AIC is used to compare different possible models and also different variants of the same model to determine which one is the best fit for the data. The ARIMA model with the minimum value of AIC is considered to be the best and used for forecast.

3. Description of STEPS

STEPS is a model developed using ARIMA technique and used to forecast meteorological variables like air temperature, relative humidity, solar and net radiation for the next three days based on last seven days of respective input data. Both, input and output are on hourly average basis. Even though several studies have focused on prediction on daily and/or monthly average time scale, the present study uses hourly average time scale. If input data are available at higher temporal resolution, then it should be possible to predict at shorter time scales too. Since this model is for short term prediction, the length of forecast is restricted to three days only. On a short time scale in the atmosphere, weather conditions are known to evolve on the basis of past few days. Hence, it is not mandatory to use only last one week of data as input and small deviations from seven days (~ five to fifteen days) should suffice. For the present study, air temperature measurements are carried out using a thermograph, relative humidity is measured using a hair hygrograph (both installed within a Stevenson Screen at 1.2 m from ground surface), solar radiation is measured using a pyranometer and net radiation using a net radiometer. The present study utilizes meteorological measurements at a single point.

For the purpose of forecast, each variable is considered as a univariate time series. As already mentioned, STEPS is developed in RStudio platform. Forecast for next three days of each variable requires few minutes of computational time. Following the above procedure, forecasts of aforementioned variables were carried out at a single point for the years 2020 and 2021. For each variable, forecasts corresponding to lead times of one, two and three days are generated. For some applications, forecast of air temperature and/or relative humidity as a function of height may be required. In such cases too, STEPS can be used. However, the present study is limited to the forecasts from STEPS at a single point and height. The next section evaluates the results of these forecasts.

4. Results and Discussion

STEPS is used to forecast meteorological variables like air temperature, relative humidity, solar and net radiation, for a single point at lead times of one, two and three days. This exercise is carried out for two years namely 2020 and 2021. Few time series plots representing forecasts from different seasons are presented in Figures 1 to 5. Time series of observations (green curve) and STEPS forecasts (blue curve) for air temperature (upper left pane), relative humidity (upper right pane), solar radiation (lower left pane) and net radiation (lower right pane) during 19th January to 21st January 2020 are presented in Figure 1. The figure shows that hourly temperature values are well predicted by STEPS even at a lead time of three days although some differences in maxima/minima are seen with respect to observations. Model performance indices for few meteorological variables have been defined in Carbonell et al. [19]. These values are used to assess the performance of STEPS. For this case study, it is seen that across the length of the forecast the MAE in air temperature is 1.0 °C which is an indication of good model performance. Likewise, prediction of relative humidity for the same case study is shown in the upper right pane. Here too, it is noted that STEPS is able to simulate the diurnal variation of relative humidity, although on the third day some differences from observations are seen. For the entire length of the forecast, MAE in relative humidity is 9 %. Comparison of solar and net radiation forecasts are shown in the lower left and right figures respectively. Especially for net radiation forecasts, peak values are found to differ slightly with respect to observations, but still the results are indicative of a good model performance.

Figure 2 is same as Figure 1 but during 13th March to 15th March 2021. Referring to temperature and forecasts for this case study, it is seen that peak values of respective variables are found to differ with respect to observations; although diurnal trend is well simulated by the model. MAE in air temperature is 1.82 °C and relative humidity is 5 %. Here it should be noted that even a difference of 10 - 20 % in forecast of relative humidity with respect to observations leads to minor differences in absolute moisture content in the atmosphere. The lower panes in Figure 2 are for comparison of solar (left pane) and net radiation (right pane). Peak values as well as diurnal variation are well simulated by the model. MAE is ~ 9 W m-2 for both solar and net radiation forecast. The previous figures were used to describe model performance for selected cases in winter and pre monsoon seasons. Figures 3 and 4 refer to case studies from monsoon season. Values of solar and net radiation obtained on the Earth’s surface are dependent on cloud amount and type. The present version of STEPS does not include forecast of cloud cover. The south west monsoon season from June to September is characterized by cloudy skies. Hence during this season, solar and net radiation forecasts can have significant variation with respect to observations as evident from Figures 3 and 4. In this season too, forecasts of air temperature and relative humidity are found to be reasonable.

Figure 5 is used to describe the performance of STEPS for a case from post monsoon season i.e.18th October, 2021 to 20th October, 2021. Referring to air temperature forecast, it is again seen that the magnitude and diurnal variation of air temperature is well forecasted by STEPS. For the length of the simulation, MAE is 0.66 °C, which is again indicative of a good model performance. Likewise, the forecast of relative humidity is also found to be reasonably good. On the third day of the forecast, there is a difference of 20 % in the observed and forecasted values of relative humidity. As pointed out before, even a difference of 20 % is less likely to have an impact on absolute moisture content in the atmosphere. The next two figures are used to describe the performance of STEPS in forecast of solar and net radiation. As compared to winter season, errors are on higher side due to exclusion of cloud cover in model formulation.

Apart from individual case studies described above, to have an idea of model performance for the entire year, forecasts from STEPS have been carried out for 2020 and 2021 at lead times of one, two and three days. These results are presented in Tables 1 and 2. Forecasts for multiple years give an idea of year to year variations, if any, in model performance. As in case of NWP models, even in machine learning kind of models, errors grow in magnitude with time and hence, it is seen that errors in forecast of all four variables in all seasons increase as the length of the forecast increases. Model predictions at shorter lead times are generally more accurate whereas the longer lead times provide more period for planning and decision making. Hence, depending on application and desired accuracy, the forecasting horizon can be suitably chosen. Referring to average errors in temperature forecasts, it is seen that at a lead time of one day, MAE in all seasons in both years is less than 2 °C which is an indication of good model performance [19, 20]. The monsoon season during June – September is associated with lowest errors at all lead times owing to very little diurnal variation of air temperature in this season. At lead time of three days, the MAE is higher by a factor of two as compared to the benchmark for a good model performance. Next, referring to forecasts of relative humidity, average MAE varies between ~ 4 % to 22 % across all lead times and all seasons. Again like temperature, monsoon season is characterized by least errors in forecast of relative humidity owing to less diurnal variation. At lead time of one day, maximum MAE in relative humidity is ~ 11 %.

Owing to the growing interest in solar energy as a source of renewable energy, it is of interest to examine seasonal variation of errors in forecast of solar radiation at different lead times. For reasons already mentioned before, errors occurring in monsoon season are discussed separately. Excluding this season, errors in solar radiation forecast are seen to vary between ~ 16 W m^{-2} to ~ 40 W m^{-2}. In the monsoon season, errors are seen to vary between ~ 73 W m^{-2} at lead time of one day to ~ 100 W m^{-2} at lead time of three days. Chico et al. [21] have reported MAE in solar radiation forecast at a lead time of one day to vary between 40 W m^{-2} to 140 W m^{-2} for two sites in Italy. Errors in monsoon season can be reduced by including forecasts of cloud cover. This will be taken up in future versions of the model.

Net radiation at the Earth's surface is the vector sum of incoming and outgoing short and longwave radiation components. Forecasts of net radiation are important in studies on energy balance, climate monitoring, land atmosphere interactions and agriculture meteorology [22]. Tables 1 and 2 point to the fact that forecasts of net radiation indicate a trend similar to those of solar radiation. In non-monsoon season, errors in net radiation forecast are seen to vary between ~ 15 W m^{-2} to ~ 40 W m^{-2}. In the monsoon season, errors are seen to vary between ~ 65 W m-2 at lead time of one day to ~ 85 W m^{-2} at lead time of three days. In the day time, the incoming shortwave radiation i.e. solar radiation is the major contributor to net radiation and hence improvements in solar radiation forecasts will have a positive impact on net radiation forecast as well.

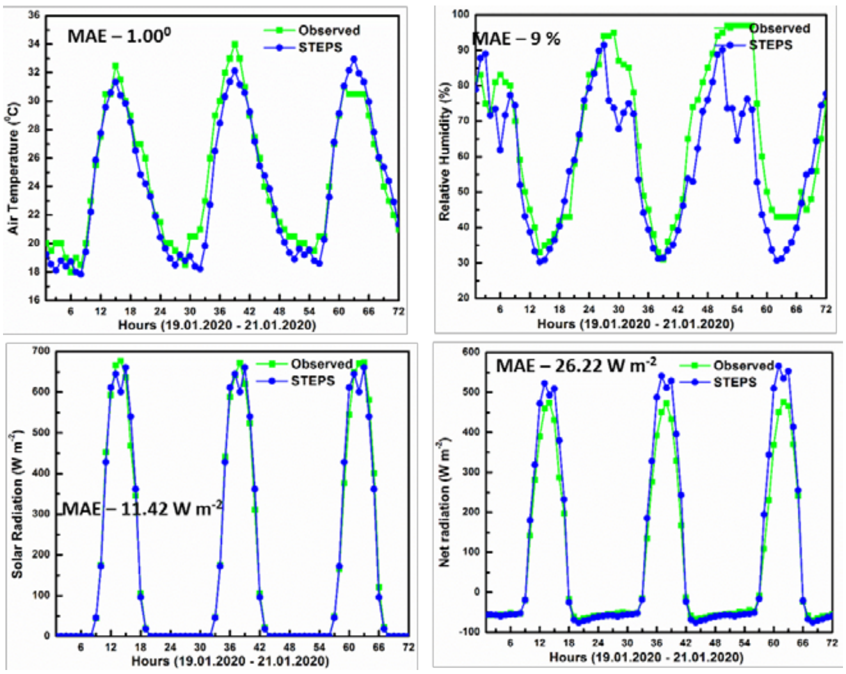

Figure 1. Time series of observations (green curve) and STEPS forecast (blue curve) for air temperature (upper left pane), relative humidity (upper right pane), solar radiation (lower left pane) and net radiation (lower right pane) during 19.01.2020 to 21.01.2020.

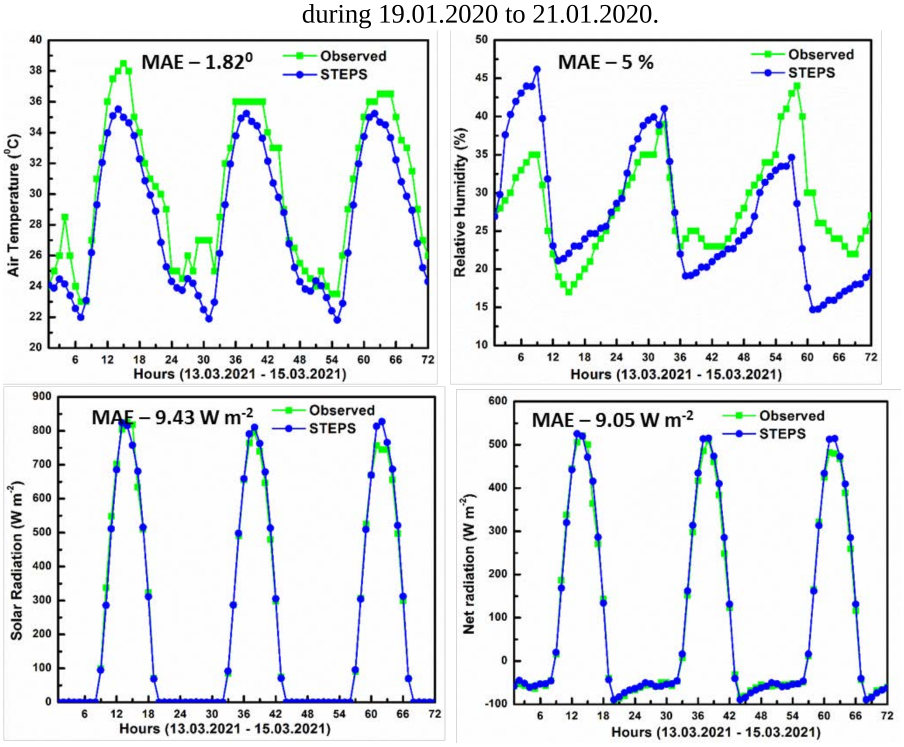

Figure 2. Same as Fig.1 during 13.03.2021 to 15.03.2021.

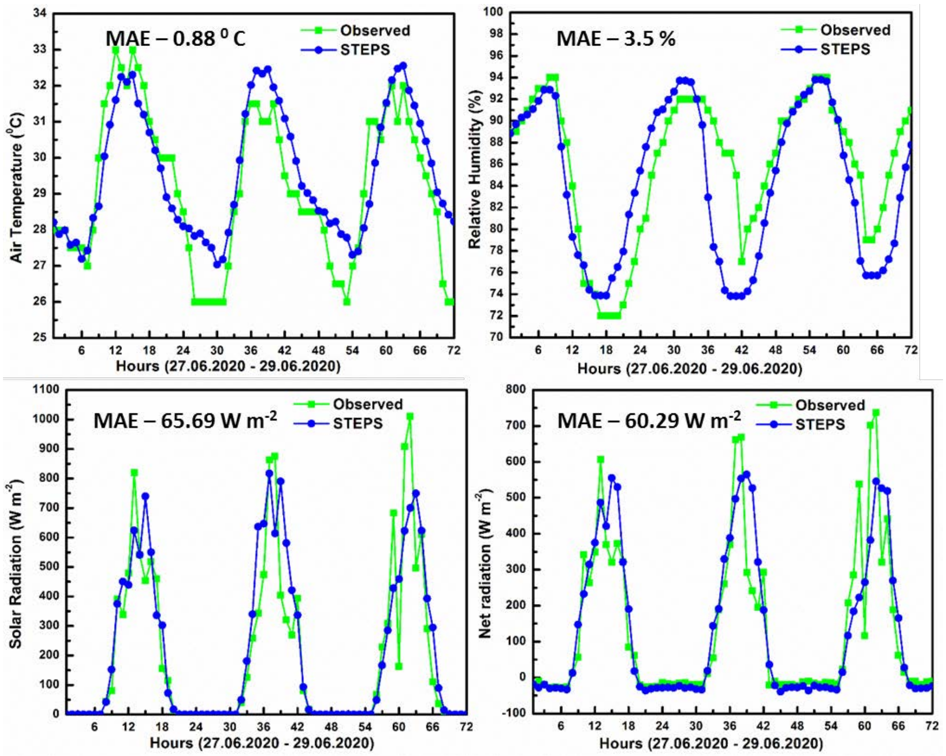

Figure 3. Same as Fig.1 during 27.06.2020 to 29.06.2020.

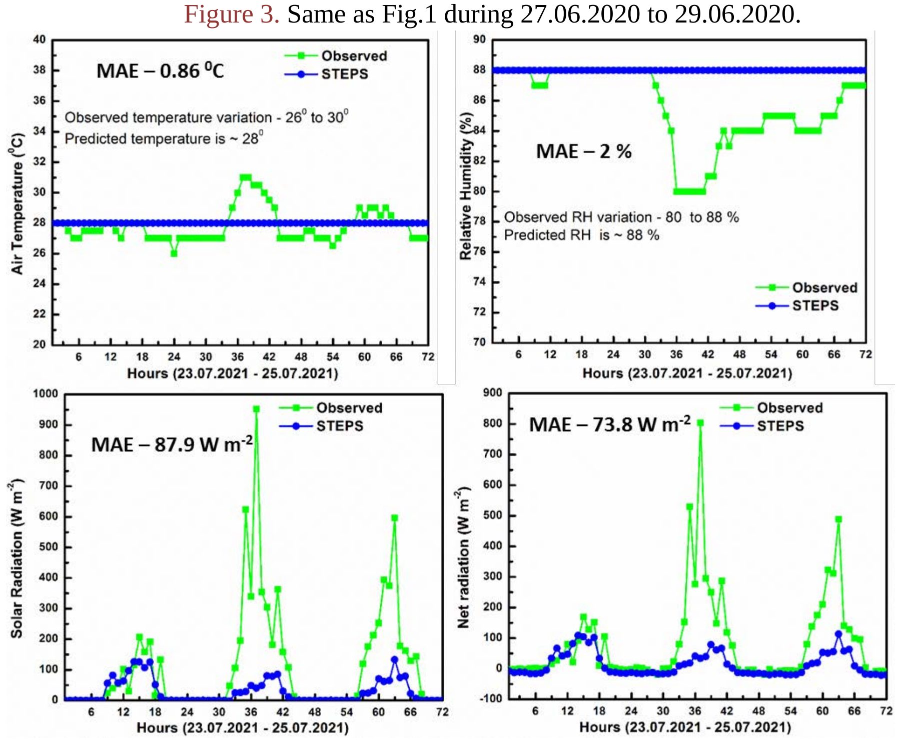

Figure 4. Same as Fig.1 during 23.07.2021 to 25.07.2021.

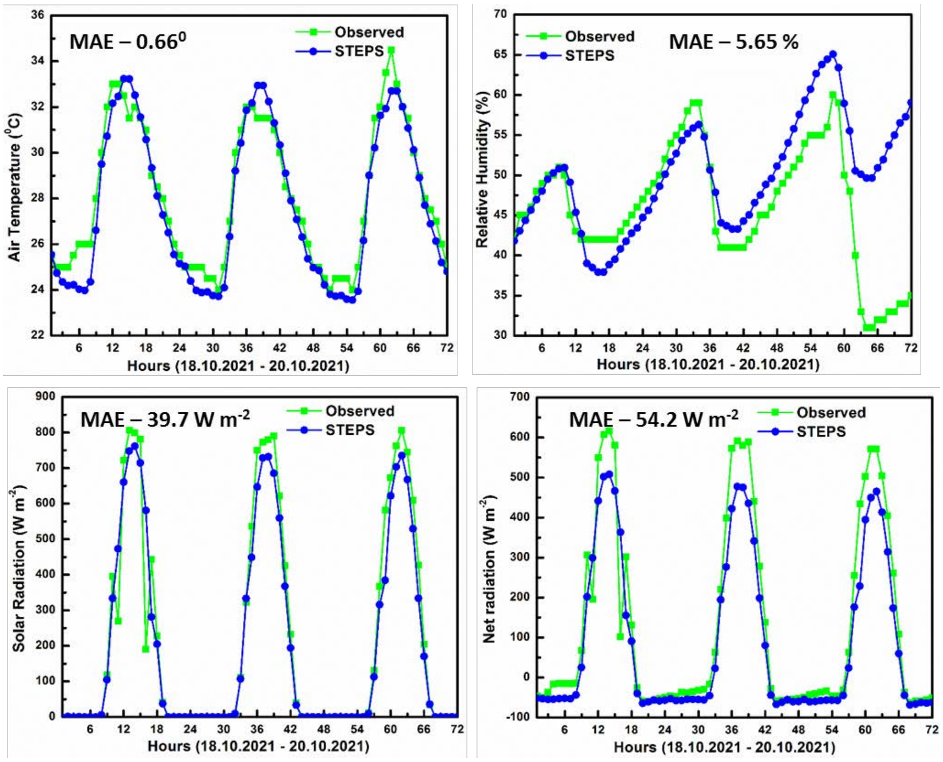

Figure 5. Same as Fig.1 during 18.10.2021 to 21.10.2021.

Table 1. Seasonal evaluation of forecasts from STEPS for 2020

Season	Average MAE (Lead time = 1 day)	Average MAE (Lead time = 2 days)	Average MAE (Lead time = 3 days)
	Air Temperature (^{0} C)		
Winter	1.81	3.00	3.94
Pre-monsoon	1.85	3.45	3.85
Monsoon	1.05	1.52	1.81
Post-Monsoon	1.20	1.72	2.25
	Relative Humidity (%)		
Winter	9.84	14.67	22.51
Pre-monsoon	10.97	13.61	16.53
Monsoon	4.33	5.72	6.73
Post-Monsoon	6.00	8.89	11.42
	Solar Radiation (W m^{-2})		
Winter	16.66	20.29	21.69
Pre-monsoon	23.96	26.19	28.35
Monsoon	73.68	87.53	100.69
Post-Monsoon	29.11	29.21	32.55
	Net Radiation (W m^{-2})		
Winter	26.13	32.79	37.78
Pre-monsoon	23.51	27.28	29.96
Monsoon	68.05	75.18	84.91
Post-Monsoon	28.26	29.48	32.60

Acknowledgements

The authors express their gratitude to Shri Probal Chaudhury, Associate Director, Health, Safety and Environment Group, BARC, and Dr. R. B. Oza, Head, Radiation Safety Systems Division, BARC, for constant encouragement and fruitful discussions during the period of study. The authors also thank Shri B. B. Ghorpade of Radiation Safety Systems Division for assistance in measurements of meteorological variables.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] P. Chen, A. Niu, D. Liu, W. Jiang and B. Ma, "Time Series Forecasting of Temperatures using SARIMA: An Example from Nanjing," in IOP Conf. Series: Materials Science and Engineering, 2018.

[2] M. Murat, I. Malinowska, M. Gos and J. Krzyszczak, "Forecasting daily meteorological time series using ARIMA and regression models," International Agrophysics, vol. 32, pp. 253-264, 2018.

[3] H. Astsatryan, H. Grigoryan, A. Poghosyan, R. Abrahamyan, S. Asmaryan, V. Muradyan, G. Tepanosyan, Y. Guigoz and G. Giuliani, "Air temperature forecasting using artificial neural network for Ararat valley," Earth Science Informatics, vol. 14, p. 711–722, 2021.

[4] K. Abhishek, M. Singh, S. Ghosh and A. Anand, "Weather forecasting model using Artificial Neural Network," Procedia Technology, vol. 4, pp. 311-318, 2012.

[5] S. Baboo and I. Shereef, "An Efficient Weather Forecasting System using Artificial Neural Network," International Journal of Environmental Science and Development, vol. 1, no. 4, pp. 321-326, 2010.

[6] L. Fara, A. Diaconu, D. Craciunescu and S. Fara, "Forecasting of Energy Production for Photovoltaic Systems Based on ARIMA and ANN Advanced Models," International Journal of Photoenergy, 2021.

[7] H. Ettayyebi and K. Himdi, "Artificial Neural Network for Forecasting One Day Ahead of Global Solar Irradiance," in The Second International Conference on Smart Applications and Data Analysis for Smart Cities, 2018.

[8] M. Alsharif, M. K. Younes and J. Kim, "Time Series ARIMA Model for Prediction of Daily and Monthly Average Global Solar Radiation: The Case Study of Seoul, South Korea," Symmetry, vol. 11, pp. 240-256, 2019.

[9] E. Chodakowska, J. Nazarko, L. Nazarko, H. Rabayah, R. Abendeh and R. Alawneh, "ARIMA Models in Solar Radiation Forecasting in Different Geographic Locations," Energies, vol. 16, pp. 5029-5042, 2023.

[10] N. Junhuathon and K. Chayakulkheeree, "Comparative Study of Short-Term Photovoltaic Power Generation Forecasting Methods.," in Proceedings of the 2021 International Conference on Power, Energy and Innovations (ICPEI), Nakhon Ratchasima, 2021.

[11] Y. Jiang, L. Zheng and X. Ding, "Ultra-Short-Term Prediction of Photovoltaic Output Based on an LSTM-ARMA Combined Model Driven by EEMD," J. Renew. Sustain. Energy 2021, vol. 13, pp. 046103-046116, 2021.

[12] S. Mughal, Y. Sood and R. Jarial, "Design and Optimization of Photovoltaic System with a Week Ahead Power Forecast Using Autoregressive Artificial Neural Networks," Mater. Today Proc., vol. 52, p. 834–841, 2022.

[13] J. Rogier and N. Mohamudally, "Forecasting Photovoltaic Power Generation via an IoT Network Using Nonlinear Autoregressive Neural Network," Procedia Comput. Sci., vol. 151, p. 643–650, 2019.

[14] S. Boubaker, S. Kamel, L. Kolsi and O. Kahouli, "Forecasting of One-Day-Ahead Global Horizontal Irradiation Using Block-Oriented Models Combined with a Swarm Intelligence Approach," Nat. Resour. Res., vol. 30, pp. 1-26, 2021.

[15] R. Basmadjian, A. Shaafieyoun and S. Julka, "Day-Ahead Forecasting of the Percentage of Renewables Based on Time-Series Statistical Methods," Energies, vol. 14, p. 7443, 2021.

[16] G. Reikard, "Predicting Solar Radiation at High Resolutions: A Comparison of Time Series Forecasts," Sol. Energy, vol. 83, pp. 342-249, 2009.

[17] M. Alomar, F. Khaleel, M. Aljumaily, A. Masood, S. Razali, M. AlSaadi, N. Ansari and M. Hameed, "Data-driven models for atmospheric air temperature forecasting at a continental climate region," PLOS One, vol. 17, no. 11, pp. 1-31, 2022.

[18] G. Box, G. Jenkins, G. Reinsel and G. Ljung, Time Series Analysis Forecasting and Control (Fifth Edition), New Jersey: John Wiley and Sons, 2016.

[19] L. Carbonell, G. Mastrapa, Y. Rodriguez, L. Escudero, M. Gacita, A. Morlot, I. Montejo, E. Ruiz and S. Rivas, "Assessment of the Weather Research and Forecasting model implementation in Cuba addressed to diagnostic air quality," Atmos Poll Res, vol. 4, pp. 64-4, 2013.

[20] R. Borge, V. Alexandrov, J. Vas, J. Lumbreras and E. Rodri´guez, "A comprehensive sensitivity analysis of the WRF model for air quality applications over the Iberian Peninsula," Atmospheric Environment, vol. 42, pp. 8560-8574, 2008.

[21] G. Chicco, V. Cocina, P. Leo, F. Spertino and A. Pavan, "Error Assessment of Solar Irradiance Forecasts and AC Power from Energy Conversion Model in Grid-Connected Photovoltaic Systems," Energies, vol. 9, no. 8, pp. 1-27, 2016.
[22] O. Ojo, B. Adeyemi and D. Oluleye, "Artificial neural network models for prediction of net radiation over a tropical region," Neural Computing and Applications, vol. 33, no. 12, pp. 6865-6877, 2020.