19th Brazilian Congress of Thermal Sciences and Engineering
November 6th–10th, 2022, Bento Gonçalves - RS - Brazil

ENCIT2022-0493 Development of a Solar Radiation Map Using Artificial Intelligence Techniques

Rafael Amauri Diniz Augusto
Cristiana Brasil Maia
Silvio Jamil Ferzoli Guimarães
Instituto de Ciências Exatas e Informática
radaugusto@sga.pucminas.br
cristiana@pucminas.br
sjamil@pucminas.br

Abstract. *In the last decades, one of the main goals of the brazilian Ministério de Minas e Energia has been to promote power generation by green and renewable power sources, especially solar, which is both clean and found in significant numbers in Brazil. Given Brazil is a continental country with most of its territory lying in tropical zones, investment in solar power is extremely relevant for the country, which currently generates most of its power with hydropower and fossil fuels. The usage of solar panels as well as power generation through them depends on expected values for solar irradiance in a given area. The development of a method capable of making such forecasts reliably throughout the year is of extreme relevance in the current state of world power generation. The present study uses machine learning through time series analysis and interpolation to develop a method capable of projecting the given solar irradiance in a certain area. This method uses data with direct solar irradiance values collected by the National Institute of Meteorology over 6 years in different weather stations as input to train the model. With this data, a model was developed to predict future solar irradiance in the cities in which the weather stations are. Since the model by itself only predict timely values for the cities found in the dataset, a zone of influence was set around each city, making it possible to interpolate values for areas that are not in the dataset but are found inside one or more zones of influence to generate a solar irradiance map of the Brazilian state of Minas Gerais. The end goal of this study is to develop a method capable of predicting the average daily direct solar irradiance for any month in any locality of the state of Minas Gerais.*

Keywords: *Solar irradiance, time series, machine learning, radiation map*

1. Introduction

The increase of the world population and the depletion of conventional energy resources is a problem facing the world nowadays, leading to the search for renewable energy sources. The world energy matrix is essentially fossil. Brazil has a different behavior, being third in the global ranking in terms of total renewable energy capacity at the end of 2020, with a capacity of 150 GW, behind China and the United States (REN21, 2021). Although it has the largest share of renewable energy in total final energy consumption among the G20 countries, its matrix is highly dependent on a single energy source (MINISTÉRIO DE MINAS E ENERGIA, 2021). The percentage of hydropower in the generation of domestic electricity is 62.9% (International Energy Agency, 2019). When the share of hydropower in total renewable energy capacity at the end of 2020 is disregarded, Brazil drops from third to 41st place (REN21, 2021).

Hydroelectric power plants contribute to the emission of greenhouse gases, erosion of underground carbon deposits, and destruction of green areas (BEAU LIEU et Al). In addition, hydropower plants have a set life expectancy, from fifty to one hundred years (WIELAND, 2010). In January of 2021, a report by the Institute for Water, Environment, and Health of the United Nations University showed that the majority of the 58.700 hydropower plants in the world are nearing the end of their expected life since most of them were built between 1930 and 1970 (PERERA et Al, 2021).

There is, therefore, a need to diversify the Brazilian energy matrix. Due to the increase in conventional energy prices and its environmental effects such as air pollution, global warming, ozone layer depletion, and the greenhouse effect, the use of renewable energy sources has increased considerably (Sözen, Menlik and Ünvar, 2008). Among the renewable energy sources, solar energy is the cheapest and is responsible for approximately 20% of household consumption (Bhowmik and Amin, 2017).

For the design and optimization of solar equipment, it is important to predict the solar radiation, which can be made from data measured by equipment such as pyranometers and pyrheliometers. However, equipment for measuring solar radiation is usually expensive, which makes measuring solar radiation more difficult than measuring other quantities, such as the number of hours of sunshine, temperature, humidity, and air precipitation (He et al., 2020). To try to solve the problem of insufficient solar radiation data, researchers use empirical models, machine learning models, and methods based on satellite data to estimate the global solar radiation incident on surfaces.

In recent years, a large number of different techniques have been applied to estimate solar radiation. Many of these techniques are based on machine learning algorithms, which have a great ability to provide robust results for solar radiation using different input values. Data such as latitude, longitude, temperature, wind speed and direction, and day length have been used in the literature and shown to be relevant for the production of reliable results. These and other variables can be used in different machine learning algorithms, such as different types of neural networks, regression vector support algorithms, Gaussian processes, and hybrid approaches, mixing these algorithms (Cornejo-Bueno et al., 2019). In another approach, satellite meteorological data can be used to estimate incident solar radiation.

A literature review was presented by Voyant et al (Voyant et al., 2017). The authors present an overview of the different methods presented in the literature, as well as a comparison of the methods in terms of precision and application difficulty. Manju and Sandeep (Manju and Sandeep, 2019) present a comparison between the results obtained by satellite data and experimental data for solar radiation in 12 cities in India. The authors established and evaluated 8 empirical models with correlation coefficients with the number of hours of sunlight. The results were compared with experimental data and showed a good correlation.

He et al. (He et al., 2020) explored the combination of meteorological input data and machine learning methods to estimate incident solar radiation under different climatic conditions and to improve forecast accuracy. The data were evaluated in 80 cities in China, in four different climatic regions. The optimal combinations were different in the four climatic regions evaluated. The meteorological factors that had the greatest impact on the prediction of solar radiation were the number of hours of sunlight, extraterrestrial radiation, and air temperature.

Pang, Niu, and O'Neill (Pang, Niu and O'Neill, 2020) developed a recurrent neural network model to investigate the performance of deep learning algorithms for solar radiation prediction. Weather data from a weather station in Alabama was used for the network training process. Guijo-Rubio et al (Guijo-Rubio et al., 2020) evaluated the performance of different neural network models for the estimation of solar radiation in the city of Toledo, Spain. The work was approached exclusively with measurements based on satellite images, which avoids the need to use data from local stations. The results showed excellent performance for the tested networks.

In this paper it is intended to develop a method to predict the solar irradiance in Minas Gerais, Brazil, using experimental data from the literature. A time-series analysis machine learning model was developed. A series of 6 years of solar radiation data from weather stations in selected cities of Minas Gerais and the state's vicinity available from INMET (National Institute of Meteorology) was used. A mathematical interpolation was developed to obtain a solar radiation map for the entire state of Minas Gerais, not only the cities with weather stations.

2. Materials and methods

The Brazilian state of Minas Gerais has a surface area of 586.528 km2, spanning from the Brazilian heartlands in the south to the Brazilian plateau in the north. This naturally causes Minas Gerais to be characterized by a variety of different biomes, from the semi-desert Sertao in the north to a particular type of savanna known as Cerrado in the central region and Atlantic Forest in the south and east.

The Brazilian Institute of Meteorology provides solar radiation data from several weather stations distributed along the country. These data are available from January 2015 to December 2020. In this paper, we used data from the weather stations located in Minas Gerais and some weather stations in the neighboring states, as indicated in Figure 1.

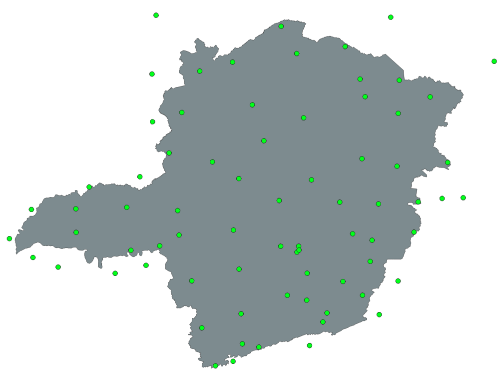

Figure 1. Location of the weather stations in the state and vicinity of Minas Gerais.

The first step was to preprocess the data from INMET to remove outliers, rows with missing entries, and incoherent values. The ending result was a collection of data from 74 weather stations, each one with six years of solar irradiance data in their respective locations. As an example, Table 1 outlines the average values collected by the first fifteen weather stations in of March 2018, for one specific day.

Table 1. Values for the average daily direct solar irradiance observed in each weather station.

Station ID	Latitude	Longitude	Average Monthly Direct Solar Irradiance ($\frac{kWh}{\frac{m^2}{day}}$)
A011	-18.96	-50.63	5.54
A024	-14.13	-47.52	5.00
A034	-18.15	-47.92	5.45
A035	-18.40	-49.19	6.08
A036	-16.78	-47.61	5.44
A045	-15.59	-47.62	5.15
A433	-14.18	-41.67	5.33
A437	-15.28	-39.09	4.87
A502	-21.22	-43.76	5.03
A505	-19.60	-46.94	5.76
A507	-16.68	-43.84	5.67
A508	-18.91	-48.25	5.46
A509	-16.16	-40.68	6.08
A510	-22.86	-46.04	5.16
A511	-20.76	-42.86	4.63

2.1 Mathematical model

Time series is a collection of data points indexed along a time axis, presenting characteristics such as trends and seasonality, according to the definition of (Box & Jenkins, 1970). The data of the present study was evaluated as a time series with seasonality and no trend. The seasonality is attributed to the characteristics of solar radiation. Although random variations are expected due to the atmospheric conditions, the translation movement of the earth around causes a predictable behavior, with lower values in autumn and winter and higher values in summer and spring.

The prediction of future values in a time series with seasonality characteristics demands an exponential smoothing technique capable of attributing exponentially decreasing weights to past data across time. Triple Exponential Smoothing (Winters, 1960) applies exponential smoothing three times to the series, which is denoted by:

$$F(t) = L(t) + T(t) + S(t) + R \tag{1}$$

where in a given timestamp t, F(t) is the value of solar irradiance for t,, L(t) is the base level for t, T(t) represents the trend of the series at time t, S(t) is the seasonality for the data and R represents random noise present in the data that cannot be explained. It is relevant to recall that the series used in this study presents no trend, which simplifies the expression to:

$$F(t) = L(t) + S(t) + R \tag{2}$$

The base level for a time series is represented by L(t) and is determined by the value for the series if it were a straight line with no slope. This component is given by:

$$L(t) = \alpha * (\frac{Yt}{S(t)}) + (1 - \alpha) * (L(t-1) + T(t-1)) \tag{3}$$

Since there is no trend in the present series, the equation can be simplified to:

$$L(t) = \alpha * (\frac{Yt}{S(t)}) + (1 - \alpha) * L(t-1) \tag{4}$$

where α represents how representative of the series past entries are. When values for α increase, the model gives increasing relevance to more recent entries, to the detriment of the relevance of older entries.

The seasonality of a time series is represented by S(t), and represents a seasonal pattern in the changes in data values that are repeated regularly over the same time period, i.e., every twelve months. This component is denoted by:

$$S(x) = \gamma(\frac{Yt}{Lt}) + (1 - \gamma)S(t - m) \tag{5}$$

where the value for γ represents how representative this component's past values are, with higher values resulting in more representativity for the more recent data.

2.2 Data interpolation

Equations 1 to 5 allow the prediction of the solar radiation data for the locations with experimental data available. In this paper, we used these data to generate a map for the entire state of Minas Gerais, by interpolating existing data.

Data interpolation refers to a mathematical method that uses known values to estimate an unknown value based on a set of rules. The interpolation data method used was the Inverse Distance Weighting. It is an interpolated multivariable method that uses known values to estimate an unknown value based on a given weight that makes the unknown value closer to its nearest neighbor (SHEPARD, 1968). This calculation aims to yield an accurate representation of the values for unknown data points, ie, cities without a weather station.

Figure 2 shows the general workflow of the methodology used, from raw data to radiation maps. The training phase utilizes the values 0.2 for α and 0.1 for γ to slightly increase the weight of more recent values in the series and minimize the risk of overfitting.

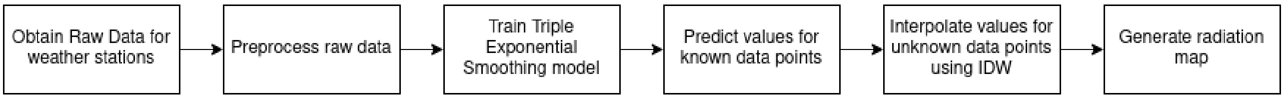

Figure 2. Simplified flowchart.

3. Results and discussion

Data from INMET are available for years 2015 to 2020. When predicting a given year N, the training data consisted of years 2015 to N-1. Therefore, for each location, a total of 48 predictions were performed (one prediction per month for four years). To test the predicted values, the Mean Absolute Error (MAE) metric was used to test a given time frame and the Absolute Error metric was used to test individual predictions. These are the most common metrics for testing variance and error in regression models.

Table 2 shows the monthly average direct solar irradiance obtained by the daily data for station A507, corresponding to the city of Uberlândia. This sample is a representation of one of the 74 weather stations used to train the machine learning model for the predictions.

Table 2. Preprocessed data for station A507 in 2015 and 2016.

Date	Average Monthly Direct Solar Irradiance $(\frac{kWh}{\frac{m^2}{day}})$
01/2015	6.49
02/2015	5.85
03/2015	4.42
04/2015	4.69
05/2015	4.30
06/2015	4.37
07/2015	4.44
08/2015	5.59
09/2015	5.33
10/2015	6.25
11/2015	5.89
12/2015	5.48
01/2016	4.70
02/2016	5.44
03/2016	5.43
04/2016	5.46
05/2016	4.50
06/2016	4.53
07/2016	4.94
08/2016	5.64
09/2016	5.88
10/2016	6.21
11/2016	5.26
12/2016	5.91

Table 3 shows the predictions of direct solar radiation for 2017 for the same location, using the years 2015 and 2016 as reference. For comparison, Table 3 also shows the real values and the difference between the values. The results are also presented in Figure 3. It can be seen that the predictions were accurate for 2017, with a maximum error of 0.4 $\frac{kWh}{\frac{m^2}{day}}$

Table 3. Direct Solar Irradiance for weather station A507 in 2017, using data from 2015 and 2016 as training.

Date	Average Monthly Direct Solar Irradiance ($\frac{kWh}{\frac{m^2}{day}}$)	Predicted Average Monthly Direct Solar Irradiance ($\frac{kWh}{\frac{m^2}{day}}$)	Absolute Error
01/2017	5.33	5.62	0.29
02/2017	5.98	5.71	0.27
03/2017	5.34	5.00	0.34
04/2017	4.88	5.13	0.25
05/2017	4.60	4.44	0.16
06/2017	4.59	4.49	0.10
07/2017	4.85	4.73	0.12
08/2017	5.21	5.64	0.43
09/2017	6.15	5.64	0.51
10/2017	6.04	6.25	0.21
11/2017	5.43	5.59	0.16
12/2017	5.15	5.58	0.53

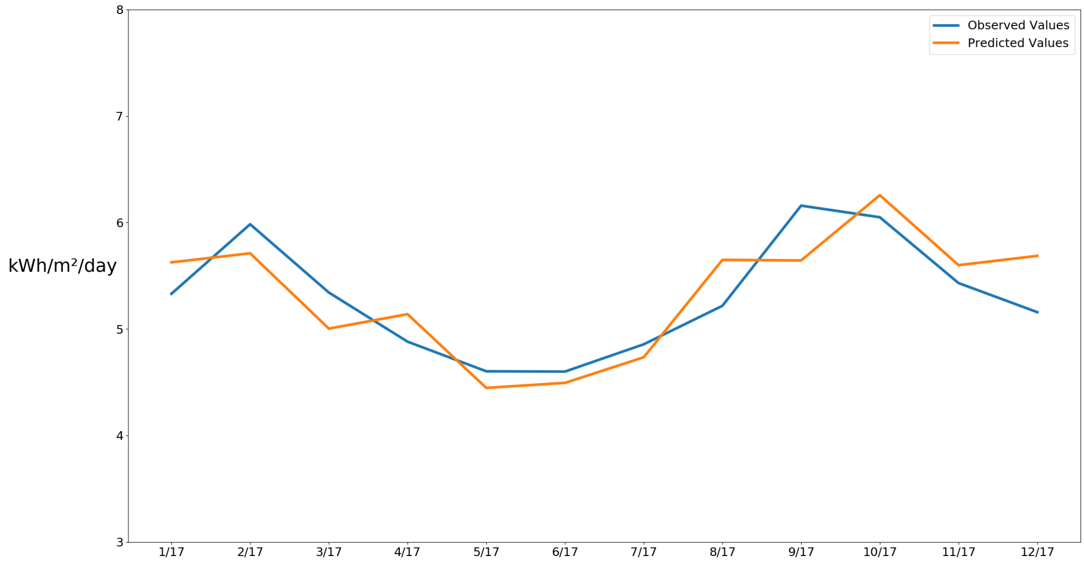

Figure 3. Predicted and Observed values in 2017.

The previous results were obtained using the years 2015 and 2016 for training. To predict solar irradiance for a given year N, the Triple Exponential Smoothing method uses data from the year 2015 to year N-1 for training, adding more data at each iteration. Figure 3 shows predicted and real values for solar radiation in 2017 using the previously described method.

The desired result is the overlapping of the lines, indicating that the machine learning model was able to accurately predict the solar irradiance for each timestep. The MAE score obtained was 0.3, indicating that an average error of 0.3 kWh/m2.day was found. This value is significant considering that most entries are between 4 and 6 $\frac{kWh}{\frac{m^2}{day}}$, but it is worth noting that the solar radiation varies from year to year, due to variations in atmospheric conditions. It is demonstrated in Table 2, which shows real values of solar irradiance in 2015 and 2016. For instance, the difference in the measured values in January 2015 and January 2016 was 1.89 $\frac{kWh}{\frac{m^2}{day}}$.

Figure 4 presents the solar irradiance predicted for the 74 locations, for March 2018. It can be noticed that it is not possible to predict the solar irradiance in a location without any weather station. Blue represents irradiance values equal to 3 $\frac{kWh}{\frac{m^2}{day}}$, while red represents values equal to 6 $\frac{kWh}{\frac{m^2}{day}}$.

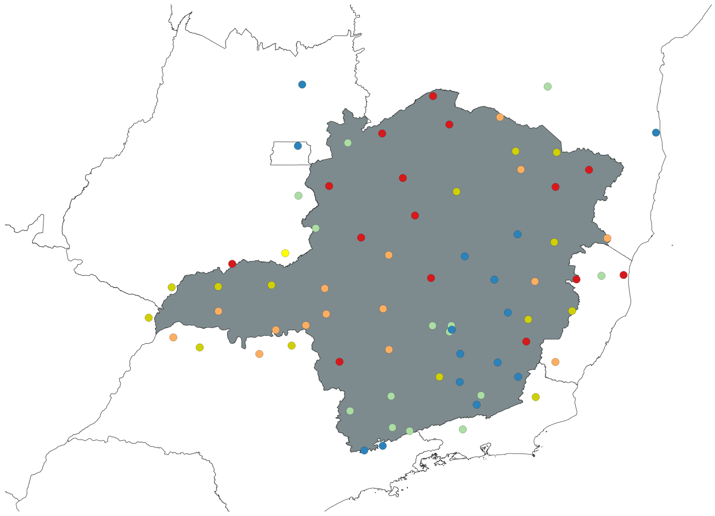

Figure 4. Predicted values before the interpolation step.

An interpolation method was used to generate a solar radiation map for the state of Minas Gerais, shown in Figures 5 and 6. For comparison, Figure 5 shows the real values and Figure 6 shows the predicted values. The MAE score was 0.29 $\frac{kWh}{\frac{m^2}{day}}$, and the differences can be seen as hot-red and light-green dots on the map. The differences between predicted and experimental data are small, and it is considered that the model was accurate.

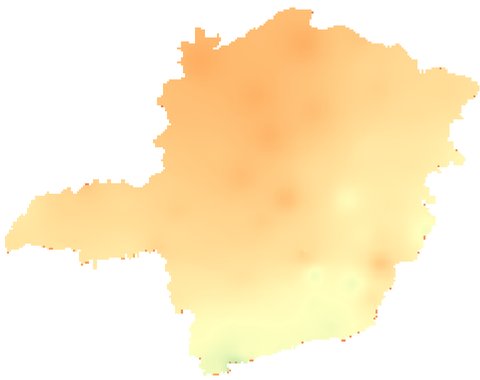

Figure 5. Real radiation map for March 2018.

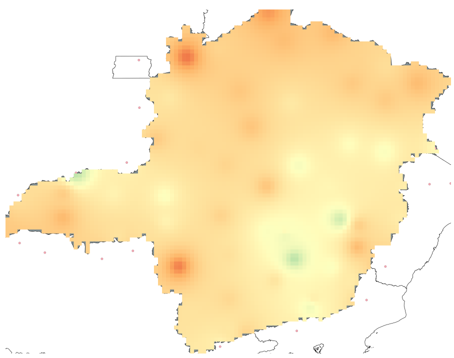

Figure 6. Predicted radiation map for March 2018.

Figure 7 shows the solar radiation maps for the entire dataset, using experimental data from 2015 to 2020. It can be seen that the higher values were obtained for October and the lower values were obtained for July, consistent with experimental data. The MAE score for this test yielded a result of 0.3 $\frac{kWh}{\frac{m^2}{day}}$.

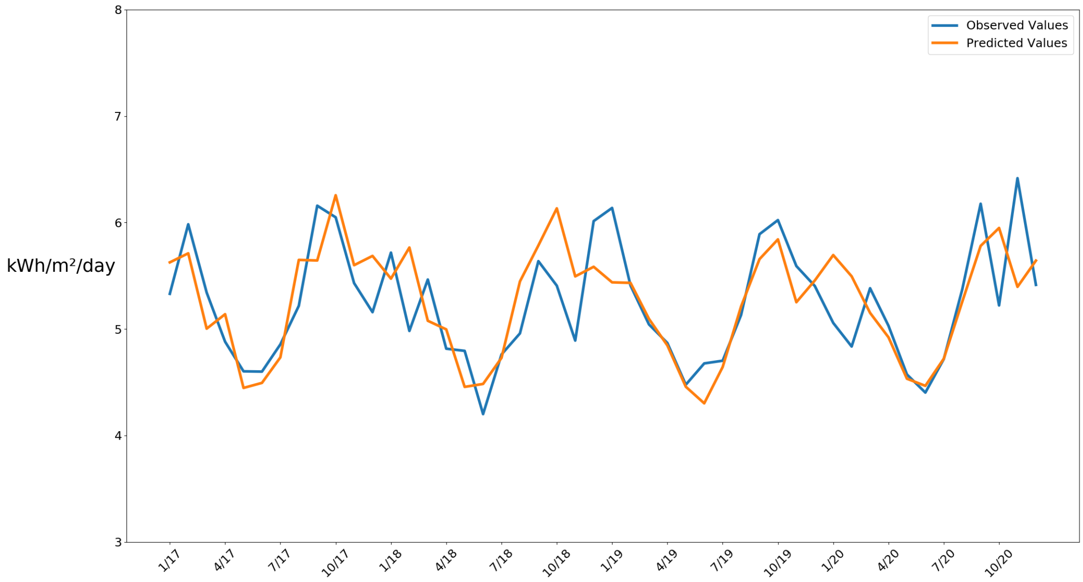

Figure 7. Predictions and Observed values spanning from January 2017 to December 2020.

4. Conclusions

In this paper machine learning techniques were used to predict solar radiation data for the state of Minas Gerais, Brazil. 2015 to 2020 experimental data from INMET were used for training and validation of the results. Data were available for discrete locations, corresponding to cities with weather stations. Interpolation methods were used on the predicted values to generate a solar radiation map for the entire state. Comparing the predicted values to experimental data, it was concluded that the proposed model was accurate. The differences obtained are lower than the differences found experimentally from year to year, considering that the solar variation depends on random atmospheric conditions. It is expected that the results help the design of solar equipment to operate in the state of Minas Gerais. Also, the model developed can be simulated for other regions.

5. Acknowledgments

This study was financed in part by stipends from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The authors are also thankful to PUC Minas, FAPEMIG, and CNPq. Special thanks to Felipe Andrade Martins for the support in the early stages of the study.

6. References

BEAULIEU, Jake; SMOLENSKI, Rebecca; NIETCH, Christopher; TOWNSEND SMALL, Amy; ELOVITZ, Michael. High Methane Emissions from a Midlatitude Reservoir Draining an Agricultural Watershed. Environmental science & technology. 48. 10.1021, 2014

BHOWMIK, H. and AMIN, R. (2017) 'Efficiency improvement of flat plate solar collector using reflector', Energy Reports. doi: 10.1016/j.egyr.2017.08.002.

BOX, G; JENKINS, G. Time series analysis: forecasting and control. Holden-Day, 1 edition, 1970

CORNEJO-BUENO, L. et al. (2019) 'Machine learning regressors for solar radiation estimation from satellite data', Solar Energy, 183. doi: 10.1016/j.solener.2019.03.079.

GUIJO-RUBIO, D. et al. (2020) 'Evolutionary artificial neural networks for accurate solar radiation prediction', Energy, 210. doi: 10.1016/j.energy.2020.118374.

HE, C. et al. (2020) 'Improving solar radiation estimation in China based on regional optimal combination of meteorological factors with machine learning methods', Energy Conversion and Management, 220. doi: 10.1016/j.enconman.2020.1

INTERNATIONAL ENERGY AGENCY (2019) Energy balance, Key World Energy Statistics. doi: 10.1016/B978-0-12-409548-9.00927-1.

MINISTÉRIO DE MINAS E ENERGIA. Balanço Energético Nacional 2021, 2021. Disponível em: <https://www.epe.gov .br/pt/publicacoes-dados-abertos/ publicacoes/balanco-energetico-nacional-2021>

MANJU, S. and SANDEEP, M. (2019) 'Prediction and performance assessment of global solar radiation in Indian cities: A comparison of satellite and surface measured data', Journal of Cleaner Production, 230. doi: 10.1016/j.jclepro.2019.05.108.

PANG, Z., NIU, F. and O'NEILL, Z. (2020) 'Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons', Renewable Energy, 156. doi: 10.1016/j.renene.2020.04.042.

PERERA, D.; SMAKHTIN, V.; WILLIAMS, S.; NORTH, T.; CURRY, A. Aging Water Storage Infrastructure: An Emerging Global Risk.UNU-INWEH Report Series, Issue 11, 2021

REN21 (2021) Facts from the Renewables 2021 Global Status Report: Trends in Brazil. Available at: $https : //www.ren21.net/wp - content/uploads/2019/05/REN21_GSR2021_Factsheet_Brazil_EN.pdf$.

SHEPARD, Donald. A two-dimensional interpolation function for irregularly-spaced data. In Proceedings of the 1968 23rd ACM national conference (ACM '68). Association for Computing Machinery, New York, NY, USA, p. 517-524, 1968

SÖZEN, A., MENLIK, T. and ÜNVAR, S. (2008) 'Determination of efficiency of flat-plate solar collectors using neural network approach', Expert Systems with Applications, 35(4). doi: 10.1016/j.eswa.2007.08.080.

VOYANT, C. et al. (2017) 'Machine learning methods for solar radiation forecasting: A review', Renewable Energy. doi: 10.1016/j.renene.2016.12.095.

WEIGEND, A. S.; GERSHENFELD, N. A. Time Series Prediction: Forecasting the Future and Understanding the Past, 1994

WIELAND, Martin. Life-span of storage dams. 62. 32-35, 2010

WIENER, N. Extrapolation, Interpolation, and Smoothing of Stationary Time Series, MIT Press. 1949

WILLSON, R.; HUDSON, H. The Sun's luminosity over a complete solar cycle. Na ture 351, p. 42-44, 1991.

WINTERS, Peter R. Forecasting Sales by Exponentially Weighted Moving Averages. Manage. Sci., p. 324-342, 1960

Responsibility notice

The authors are the only responsible for the printed material included in this paper