A Contemporary Assessment of the Forecasting Accuracy of Incoming Solar Energy

Swadesh Ranjan Bhakta
Yogoda Satsanga Palpara Mahavidyalaya,WB, India
Email: ranjanswadesh123@gmail.com

Abstract: Use of results in 1705 accuracy tests reported in many geographic regions to achieve the latest in the accuracy of predicted solar resources (North America, Europe, Asia, Australia). Literature tends to avoid presenting model performance assessments on locations or on time horizons where it is known that the models are low. Usual bias and statistical indicators are useful tools, but it is necessary to take appropriate decisions by funders' managers and policy makers by using further error measures. Several types have been compared of forecast models: persistence, traditional statistics, machine learning, cloud motion monitoring, weather prediction in numbers and hybrid models (the combination of classical stats with machine learning approach and/or exogenous inputs). Forecast errors are stepped up by increased time. The performance of the model depends on the time and climatic horizon. The most effective in-hour performance in all climates is machine learning and hybrid models. Machine learning and traditional statistical models offer the best intra-hour forecasts in all climates. In tropical and winter climates, the hybrid models perform well. The hybrid models are quite good in all climates for day-to-day predictions. In general, the best performance of the hybrid models. The predictive model accuracy in the recent decade has greatly improved. Two thirds and a third, respectively, decreased the normalised MBE and RMSE.

Keywords: Solar energy, Forecasting techniques, Forecasting accuracy, Normalised statistical indicators, Support vector machine (SVM).

Introduction:

From ancient times, fossil fuels were used. The connection between the three facts makes this understandable: (1) easy-to-use fossil fuels, (2) in practice, some primary source (carbon) can be employed and (3) relatively simple technologies can be used to increase the volume density of energy (e.g., oil refining). Today, almost 80% of global energy demand is still supplied with fossil fuels. The IEA numbers indicate a gradual decrease from roughly 86 percent in 1973 to 81 percent in 2016 in global primary energy supply for fossil fuels.

Events such as the oil crisis of 1973, the catastrophic nuclear catastrophe in Chernobyl in 1986 and, in particular, knowledge of ever-increasing environmental deterioration have led to sustainable global policies on development. The World has entered a new development phase starting with the Kyoto Protocol, with new sustainable life goals supported by renewable energy sources. In this phase, worldwide populations seek varied energy-efficiencies together with the growing usage of renewable energy are a big problem [2]. The energy system mix will irrefutably play an important role in the development of a sustainable society. There are currently many concerns concerning energy sources and patterns of consumption, such as (1) increasing needs of a worldwide population who strive to improve living standards and (2) strong proof that continued use of fossil fuels in an effort to reduce greenhouse gas emissions can lead to irreversible damage to the environment. Conservation is of the highest priority with regard to these challenges. Conservation saves energy resources and decreases the environmental impact while dramatically improving conversion and end-use effectiveness. In the future, solar energy is generally recognised to make a considerable contribution to energy conservation and to meet global energy demand [1] [8].

On the other hand, electricity has acquired a privileged place in relation to our high-tech lifestyle. While some primary sources can be used in thermal or mechanical applications directly, the focus is still on electricity for research and industry. No wonder, in this setting, the weight of solar electricity was significantly increased in the energy mix, and it is anticipated that this trend would continue. In 10 years, the global total installed solar power generated 46 times, from 6.6 GWp in 2006 to 306.5 GWp by the end of 2016. With 34.5 GWp connected to the grid in 2016, China's world market was primarily dominated, which represents a 128 percent advantage over 2015 capacity. In the EU, solar power continued down in 2016 with a further 6.7 GWp, down 21 percent from the 8.6 GWp built-in 2015. Sometimes the reality exceeds the most optimistic projections, with support for beneficial measures.

An excellent example is the incredible expansion of solar power installed in Romania, with an installed capacity of fewer than two MWp at the end of 2011 and a target of 260 MWp by 2020 for government-wide photovoltaication systems [9]. Significantly, Romania's installed PV capacity exceeded 1.1 GWp in May 2014 (According to Ref., quoting the Romanian Transmission and System Operator - Transelectrica). However, following this success storey, Romania's photovoltaics sector has seen state subsidies reduce, the growth rate decreased, and the total installed photovoltaic capacity is currently in June 2018 at 1.37 GWp. It should be noted that there is adequate room to use solar power in many additional off-grid home and industrial uses, in addition to grid linked ones. Such as solar electricity, which may be used at 6 to 11 GW in extraction and transport activities and at 17 to 91 GW in refining in the worldwide petroleum sector. Another component of household solar power is solar-thermal conversion. An excellent review of solar-thermal collector and process exergy analysis is included.

However, the stochastic character of solar energy presents a challenge to the growth of the share of solar electricity in the energy mix. Unlike traditional fossil or nuclear power plants, the power produced by wind and photovoltaic power plants is quite variable. In the 2014 measured time series this variability is quantified. EU transmission system operators collected data. Sun thermal systems react in minutes or more to changes in solar irradiation [4], while PV reacts in milliseconds. The thermal inertia smoothes the solar thermal plants' output power. However, the development of technologies to hybridise concentrate solar-thermal concentrations and combustion is currently being investigated because of the potential for economic carbon mitigation and solid supplies. The energy production of photovoltaic plants is an important issue to lower the effective cost of integrating the photovoltaic plants into the current power system. Superior forecasts will allow grid operators to schedule traditional non-renewable capacity (e.g., gas power plants) to offset the variance in photovoltaic power supply [5].

As such, it involves two interrelated difficulties to anticipate the output power of a solar plant: (1) the provision of solar resources and (2) realistic modelling of the reaction from a photovoltaic converter. Of course, the accuracy of a PV plant's predicted energy production is dependent mostly on the accuracy of the solar resource projected [10]. This drives the study to look at solar resource prediction methods.

Objectives and motivation:

The number of publications published on solar resource forecasts has increased exponentially in the recent decade, as Fig. 1. The studies are generally site-specific, with strong findings dependent on the model nature, the time horizon of the forecast and the local climate, together with a wide range of additional data and models features. This is an important constraint, making it difficult to generalise the results. A model should be tested for all the many factors in question to measure its effect. In addition, it would be good to test the model independently in as many locations as feasible, preserving the scenario, to make clear conclusions regarding the performance of a specific forecasting model used in a given situation (time horizon, local climate, etc.). These opinions apply also to our study, which consists of a sample of statistical indicators of the accuracy of the prediction of solar irradiation [3]. The data collection must be large enough to be considered statistically important, based on statistical markers published by several research. In other words, the results of this study provide an overview of solar irradiance projection models' present performance and current trends. Fig. 1 illustrates that such an examination as we propose would never, just due to a lack of public documents, have taken place just a few (4-5) years ago. Although various review papers previously published warn that influential factors would be taken into account in comparing results from different publications, none of them has actually carried out such a study.

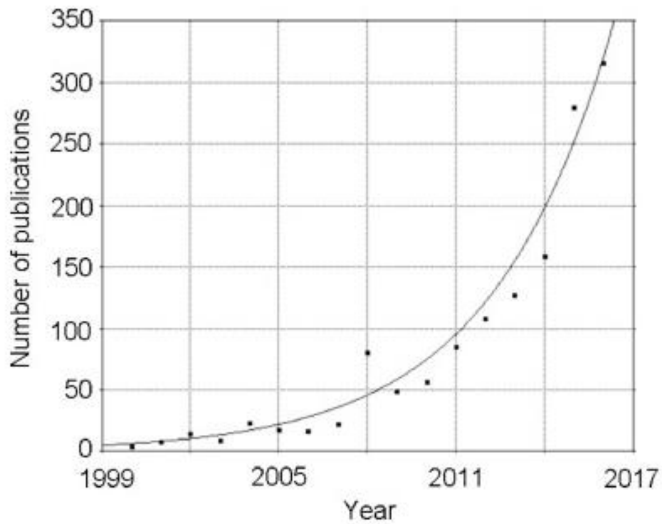

Fig. 1. Number of publications published annually on the subject of solar energy and production forecasting.

A thorough investigation of the effectiveness of solar irradiance model projections to detect current lacunae in scientific literature is presented in this paper. Many papers have been evaluated between 2006 and 2017. The data set was collected from 40 studies selected according to the following criteria: the accuracy and normalised average misconduct in the results given by the selected papers are shown by the normalised root mean quadrant error (nRMSE) (nMBE). This permits samples of various sizes to be compared, including data of various magnitudes, so assuring a consistent base for comparative analyses. The selection by this criterion results in an arbitrary sampling, and therefore this review provides a reliable image of the state-of-the-art precision in the prediction of solar resources [6]. A total of 1705 tests have been examined for nRMSE and nMBE reported in the documents selected. That is to say, a number of 1705 items, each consisting of a nRMSE and a nMBE pair, together with stamps for the model class-tested, climate zone of the place in which test data have been obtained, recording period and the projected horizon are included in the data set studied in the research. The analysis was performed from four perspectives: prediction classes (intra-hour, intra-day and day-ahead), climate (tropical, arid and temperate and snow) and trend performance with time (classic CS statistics; machine training (ML; cloud-movement tracking CM; numerical weather prediction model (NW and hybrid HY) models [17].

Variability of the solar resource:

Fig. 2 demonstrates the change at different time scales of global sun brightness. The global solar irradiance variability is due to two main causes:

At different timescales, solar irradiance constantly fluctuates throughout the year, seasonal and day (note the beautiful bell shape of day 7 in Fig. 2a) and declines to zero at night. The solar irradiance variability is a deterministic component, which is precisely accounted for by the sky-light models.

Clouds pass over the sky. The sun radiance often swings erratically (Fig. 2b, c). The variability generated by wandering nuclei is the major concern at a smaller temporal level. Fast variation in sun irradiance, known as the solar ramp, constitutes an extraordinary challenge to the grid operation. The cloud shadow movement across a tiny photovoltaic capacity can cause the output power to plummet to nearly 0% and to return to full capacity for a few seconds. In fact, at any point due to passing cloud, the variation in sun irradiation might surpass 60% in seconds of the maximum solar radiance. It depends on numerous parameters, namely the size of the PV plant and the cloud speed, to take a passing cloud to shade the whole plant. For instance, a ramp measured 80 per cent using a pyranometer within 60 seconds, with a power output measured simultaneously, is connected to a ramp of 50% at 13.2 MWp PV facility in Nevada.

The geographic distribution of photovoltaic plants connected to the same grid was demonstrated to be smoothing for the summary of variations.

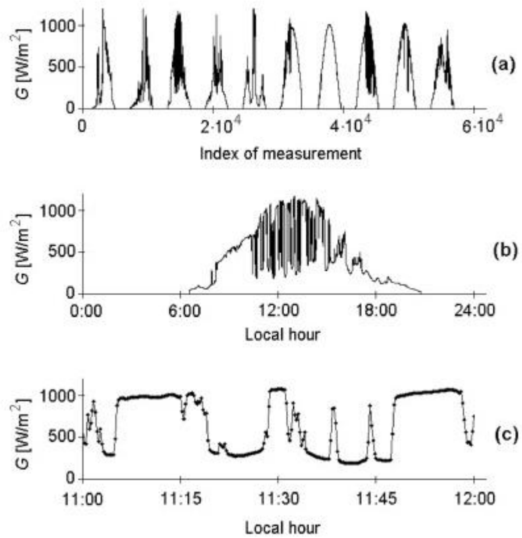

Fig. 2. Global variation in solar irradiation on a varying time scale: (a) in June 2016, first 10 days, (b) June 3, 2016, and (c) from 11 a.m. to 12:00 local time, June 3, 2016.

In the subject of Solar Energy Research, quantifying solar irradiance fluctuation is a current topic [16]. Different proposed quantifiers capture the many characteristics of variability in the solar series, and there is presently no universal consensus about which specific quantifier is best suited to classify various time intervals based on the solar radiation system variability.

In conclusion, solar irradiance can fluctuate at different time scales with large amplitude [11]. Therefore, the development of tools for both sunlight measurement and prediction is an appropriate field of research for the safe operation of a solar-powered electricity system. The following section presents the major classes of the current solar model radiation forecasts.

Classes of forecasting models:

In this section, many models are briefly examined from a proper perspective for the prediction of solar radiation. The models are divided into six types, according to their nature: classical statistical models (CSs), MLs, cloud movement monitoring models based on both ground and satellite (CMs), NWs, and Hybrid Models (HY). The sixth class includes different forms of persistence models (PS), which often provide benchmarking predictions to measure the performance of other models. A taxonomy tree diagram is displayed for each class. It should be noted that such maps of taxonomy are not unique (different authors may classify different models differently).

Performance analysis of the forecasting models:

This section assesses statistically the data series nMBE and nRMSE, acquired from the paper of this study, in terms of box plot(s) and standard deviation (quantifying the quantification of variance in the data):

$$\sigma = \sqrt{\frac{1}{M} \sum_{i-1}^{M} (m_i - \mu)^2}$$

The analysis is carried out from four angles: forecasting models classes, time horizons, climate and performance trends over time.

Analysis of forecasting models:

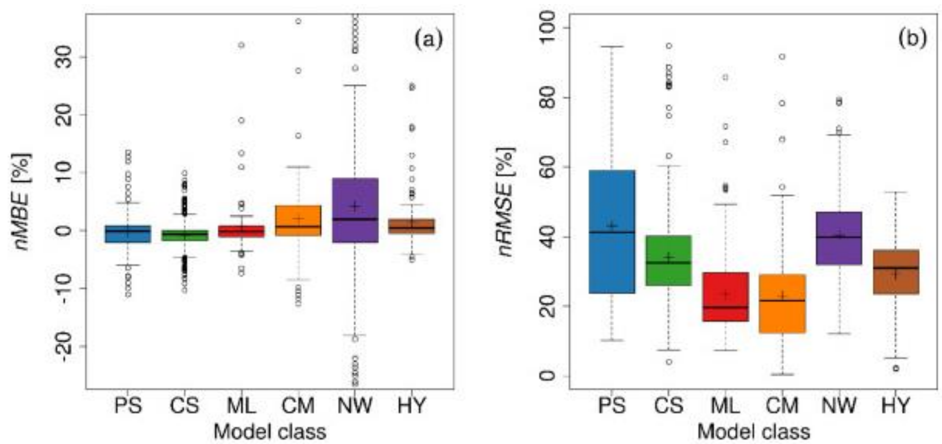

Fig 3. Forecasting accuracy (a) *nMBE* and (b) *nRMSE* for each class of models.

In the box plot, the variation in nMBE and nRMSE is shown in Fig. 3. In the six kinds of prediction models, the distribution is qualitatively substantially different:

PS: These class models are the simplest sort of forecasts, often used as benchmarks and predicted poor performance. This applies to the predictive mistakes that are measured as nRMSE, for which the highest values in every class (mean nRMSE = 43.1 percent) are presented, and the greatest error volatility is found (p nRMSE = 21.9). In terms of medium bias error, the average error is actually pretty minimal (mean nMBE = −0.14 per cent), much better than CM and NW models than CS, ML and HY types. There is also strong volatility of the mean bias errors (p nMBE = 3,23), albeit the greater the volatility is constantly below 15 per cent. This demonstrates that persistence forecasts often have low, but can cause very significant, peculiar errors. Given that persistence models anticipate the last data, we may assume that when the radiative regime is unstable, i.e. rapidly moving broken clouds in the skies, the big inaccuracies occur.

CS: models in this class have the least average errors of mean bias, and the volatility of the errors (µnMBE=1.89) is likewise the smallest. The overall performance of these CS models with PS models is qualitatively similar, with very little systemic but significant nRMSE biases. However, the CS models are largely higher on all accounts than nRMSE (µ nRMSE = 12.92) because of their volatility.

ML: This class models have relatively low average biases. Nevertheless, the bias' volatility is somewhat high (Ţ nMBE = 8.52). This is the result of enormous positive outlines, some of which are greater than 15%, while all outlines are less than 10%. These major inaccuracies occur, as we will see later, when applying ML models over broad horizons. The error is quite minor in relation to nRMSE, with an average error of nRMSE = 23.3%. The nRMSEs are equivalent to the CS class volatility.

CM: Models in this class have significantly greater bias than PS, with a noticeable overcasting tendency. At (Ţ nMBE = 5.46), the volatility of the bias errors is also extremely wide. On the other hand, the NRMSE performance is the best in every class with a mean error (mean nRMSE = 22.8%) and error volatility just a little higher than the CS model.

NW: The models in this class appear to have the worst performance at first appearance. They are highly prejudicial and volatile (Ţ nMBE = 8.88). There is an important tendency for overestimation, although outliers for both over- and underestimation are higher than 20 percent. The situation is highly fascinating with regard to nRMSE. The average error is big, at nRMSE=40.4%, but the volatility is comparable with the CS, ML, and CM models and substantially lower than the PS class models. As we can see later, the big mistakes are partly explained because the NW models are typically utilised on broad horizons (mainly DA). The huge nRMSE with small volatility suggests that these models routinely, not stochastically, produce big mistakes, which leads to the optimism that model precision can be increased after processing of output.

HY: The models in this class combine the finest of the other classes. Therefore, both with tiny volatitudes, they have small nMBEs and good nRMSEs. In fact, these models' performance is confusing in part because HY comprises projections on all time scales ranging from IH-adapted models to DA-adapted models. In analysing HY models by class of forecast horizon, section 3.2 will show that these models perform quite well in each class.

Climate influence:

The data set in box plot, divided throughout the climate zone, is shown in Fig. 4. The models are evaluated in relation to nMBE and nRMSE's mean and standard deviation. The mean and standard deviation of all existing data in a climate zone were determined. A major positive or negative average value for nMBE indicates that measurements have been overestimated or underestimated and the standard deviation measures the amplitude of the prediction mistakes. Tropical zone forecasts have very good nMBE and nRMSE values performance. It actually has the lowest bias error (mean nMBE = −0,64%) and error volatility (Ţ A nMBE = 2,06%). In terms of nRMSE the models likewise have good performance and lowest volatility (µA nRMSE=11,96). The arid climate forecasts tend to overestimate observations with a mean nMBE rate of 2,1%, with moderate volatility but the smallest mean nRMSE of 2,1%. For temperate climates, both nMBE and nRMSE (mean nMBE = 1.7% and mean nRMSE = 26.4%) are predicted to be major mistakes.

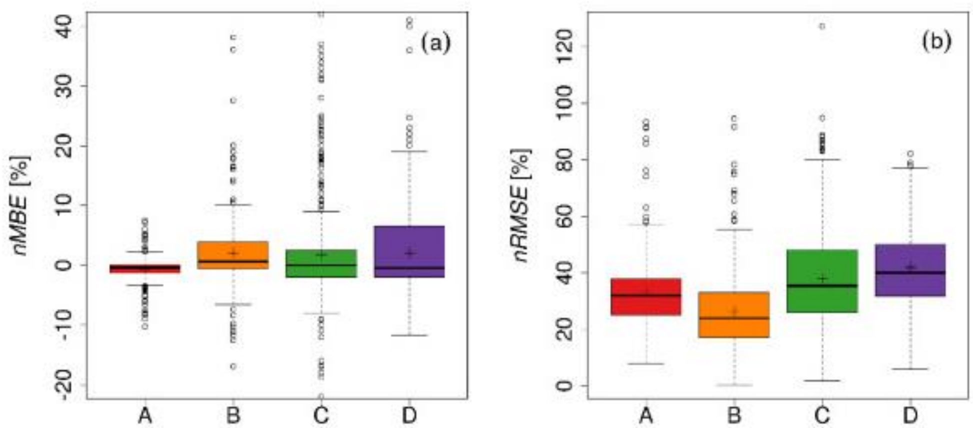

Fig 4. Box plots of a) *nMBE* and b) *nRMSE* for each climate region (A – tropical; B – arid; C – temperate and D – snow).

Also, they are very volatile (σ A nMBE = 9.55, σ A nRMSE = 17.7). The errors for snow climate zones are the largest on average but have smaller volatility than those performed in moderate climates. From the above observations, we can conclude that the available solar resource may be forecasted in a reliable way in regions with a tropical and arid climate, with accuracy greater than in moderate climates [12].

Climate / Model class	**Persistence (PS)**	**Classical statistics (CS)**	**Machine learning (ML)**	**Cloud motion (CM)**	**Numerical weather prediction (NW)**	**Hybrid models (HY)**
Tropical (A)	42	231	46	0	27	41
Arid (B)	68	34	17	134	113	9
Temperate (C	102	205	72	16	197	57
Snow (D)	54	42	2	23	149	24

Table 1. Data set for each climatic area in model classes.

Next we examine the data distribution for each climatic group in model classes and the prediction horizon classes. In any of the four climate groups, the sub-sets of data are well balanced, with the partition reported in Tables 1. Very few use CM and NW models in the tropical climate zone among the research included in this analysis. This is because the radiative regime in tropical places is typically unstable, especially given that majority of the locations in our study are island. This makes it particularly untrustworthy to forecast in general and long-term expectations. Considering that studies that employ NW models typically make daily forecasts, as we have seen from Figure 10, it is therefore natural that there is a very small number of such studies that are used in tropical areas. Table 1 confirms the unwillingness to carry out day-to-day forecasts in the tropical zones. The other remarkable feature in Table 1 is the fact that there are very few ML models used in snowy settings. we note that IH predictions are rarely conducted in snowy locations. We assume that this could be linked to more stable radiative regimes in snow conditions, enabling rather precise projections in the morning.

The trend over time:

As illustrated in Fig. 1, the number of papers on predicting solar irradiance has increased dramatically within the previous decade. It is important assessing if the increasing interest in the subject becomes a measurable general improvement in the accuracy of predictions. To evaluate this, we examine our forecasting error data set with respect to the year each entry is published.

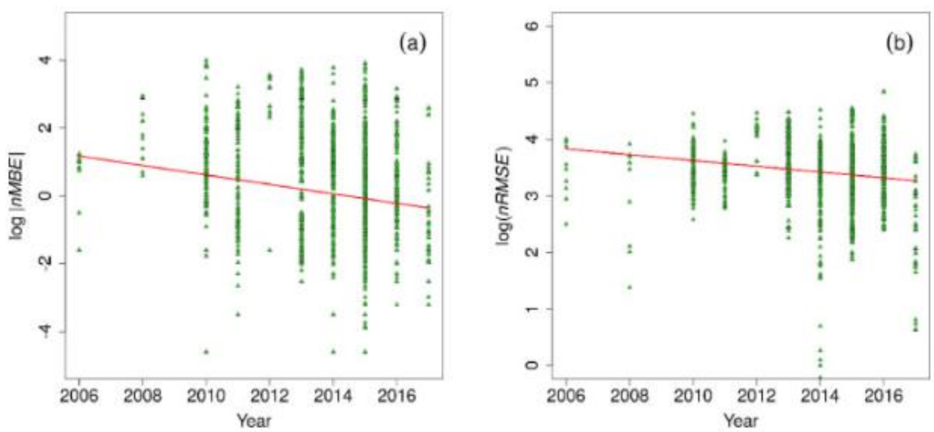

Fig 5. Performance of models developed between 2006 and 2017.

Fig. 5. shows in semi-logarithmic scale the nMBE and nRMSE for the year of study publication. Predicting mistakes can never be reduced to nil. This would only be conceivable if the physical model and the beginning circumstances were adequately high to simulate the entire atmosphere perfectly. It is, therefore reasonable for new research to gradually increase the accuracy of the forecast over time. An exponential fit that decreases is most likely to measure the annual trend of the forecast inaccuracy [13].

The linear fit α with the logarithmic amounts shown in Fig. 5 is a fluctuation in the value of an error with a factor exp (α). As a result of the publishing year, the linear fit for log |nMBE| is α = −0.12; this suggests that average error is decreased in the bias with an exp factor (-0.12) = 0.89 every year. In a five-year period, the bias error decreases to about half its value by a factor of 0.895=0.55 and decreases it in 10 years to around a third from the original amount. The slope of the linear fit of log(nRMSE) is α = −05, which suggests that the root is 0.95 each year, with a factor of 0.78 for a period of 5 years and a factor of 0.61 for 10 years (i.e. roughly two thirds from its initial value).

Finally, we would like to note that one factor in preserving the stability of the grid-supplied partially by pV plants is the development made in forecasting solar energy output [15]. In a separate context, a great deal is done to predict a power grid load accurately. The improvement of the energy weight of photovoltaics can be the outcome of both fields.

Conclusions:

A thorough investigation of solar radiation model efficiency was presented in this paper. The study has gone beyond earlier work's principal constraints, notably that the enormous number of various metrics used to evaluate the performances of model models makes comparing the different prediction models challenging. A strict framework has been established in order to compare different model classes, therefore giving a credible image of the cutting-edge accuracy of the prediction of solar resources. From four views, the results can be stratified:

(1) Prospective models classes ((persistence, classical statistics, machine learning, cloud-motion tracking, numerical weather prediction and hybrid models) [14]. The hybrid class forecast models (combining classical data with machine learning methods and/or exogenous inputs) show the best performance not expected at first glance.

(2) Horizon of time (intra-hour, intra-day and day ahead). With an increase in the period of the forecast, the forecast error assessed by nRMSE increases. On average, in-hour forecasts reach lower nRMSE values than intra day projections, while day-to-day predictions are less trustworthy.

(3) Influence on climate (tropical, arid, temperate and snow). In all climate areas, intra-hour forecasts are recommended for model of machine learning and hybrid classes. In spite of the traditional statistical class, which in its entirety has only limited performance, individual ARIMA models can also perform well. The models of the machine education and classical statistics classes regularly show minimal errors over all climate types for intraday forecasts. The hybrid models work well at sites in tropical and snow-climate environments. The hybrid models offer excellent performance for day-to-day forecasts and are suitable for use in all climate regions.

(4) Time-long trend in performance. In general, prediction models have been more accurate in time. Our study demonstrates that forecasting errors have fallen dramatically over the past 10 years, whereas nMBE has fallen by two-thirds on average as compared to the values at the start of the reporting period and nRMSE by one-third.

On a general level, it is possible to conclude that the literature still has substantial gaps. This is partly because of the scientific paradigm that does not encourage unfavourable results to be published. In the context of our investigation, this would include reporting model performance evaluations in situations in where the performance or timelines of these models are projected to be poor and not practically relevant. To make an investigation a major contribution to our common knowledge, authors are highly requested to present prediction performance with a broader set of statistical accuracy metrics to give a more comprehensive insight into the capabilities of the models. It is still necessary to use the classic MBE and RMSE indicators, preferably standardised to mean values. However, it is highly advised to make other statistical mistakes, such as the Kolmogorov-Smirnoff Integral or quality metrics. This would allow fundraisers and policymakers to make suitable judgments.

References:

1. Sun S, Wang S, Zhang G, Zheng J. A decomposition-clustering-ensemble learning approach for solar radiation forecasting. Solar Energy, 2018, 163: 189–199

2. Alhmoud L, Wang B. A review of the state-of-the-art in wind-energy reliability analysis. Renewable & Sustainable Energy Reviews, 2018, 81: 1643–1651

3. Antonopoulos V Z, Papamichail D M, Aschonitis V G, . Solar radiation estimation methods using ANN and empirical models. Computers and Electronics in Agriculture, 2019, 160: 160–167

4. Antonanzas J, Osorio N, Escobar R, . Review of photovoltaic power forecasting. Solar Energy, 2016, 136: 78–111

5. Gigoni L, Betti A, Crisostomi E, . Day-ahead hourly forecasting of power generation from photovoltaic plants. IEEE Transaction on Sustainable Energy, 2018, 9(2): 831–842

6. Khosravi A, Nunes R O, Assad M E H, . Comparison of artificial intelligence methods in estimation of daily global solar radiation. Journal of Cleaner Production, 2018, 194: 342–358

7. Behrang M A, Assareh E, Ghanbarzadeh A, . The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data. Solar Energy, 2010, 84(8): 1468–1480

8. Fouilloy A, Voyant C, Notton G, . Solar irradiation prediction with machine learning: forecasting models selection method depending on weather variability. Energy, 2018, 165: 620–629

9. Mazorra-Aguiar L, Díaz F. Solar radiation forecasting with statistical models. In: Wind Field and Solar Radiation Characterization and Forecasting: A Numerical Approach for Complex Terrain, Springer Verlag, 2018: 171–200

10. Liu Y, Shimada S, Yoshino J, . Ensemble forecasting of solar irradiance by applying a mesoscale meteorological model. Solar Energy, 2016, 136: 597–605

11. Yadav H K, Pal Y, Tripathi M M. A novel GA-ANFIS hybrid model for short-term solar PV power forecasting in Indian electricity market. Journal of Information and Optimization Science, 2019, 40(2): 377–395

12. Wan C, Zhao J, Song Y, . Photovoltaic and solar power forecasting for smart grid energy management. CSEE Journal of Power Energy Systems, 2015, 1(4): 38–46

13. Yule G U. On the time-correlation problem, with especial reference to the variate–difference correlation method. Journal of the Royal Statistical Society, 1921, 84(4): 497

14. Wang Y, Wang C, Shi C, . Short-term cloud coverage prediction using the ARIMA time series model. Remote Sensing Letters, 2018, 9(3): 274–283

15. Atique S, Noureen S, Roy V, . Forecasting of total daily solar energy generation using ARIMA: a case study. In: 9th Annual Computing and Communication Workshop Conference, Las Vegas, NV, USA, 2019, 114–119

16. Alsharif M H, Younes M K, Kim J. Time series ARIMA model for prediction of daily and monthly average global solar radiation: the case study of Seoul, South Korea. Symmetry, 2019, 11(2): 240

17. Doorga J, Rughooputh S, Boojhawon R. Modelling the global solar radiation climate of Mauritius using regression techniques. Renewable Energy, 2019, 131: 861–878