Effect of Cloudiness on Solar Radiation Forecasting

Gabriel López[1], Sergio M. Sarmiento-Rosales[2], Christian A. Gueymard[3], Aitor Marzo[4], Joaquín Alonso-Montesinos[5], Jesús Polo[6], Nuria Martín-Chivelet[6], Pablo Ferrada[4], Javier Barbero[5], Francisco J. Batlles[5] and Nieves Vela[6]

[1] Dpto. Ingeniería Eléctrica y Térmica, de Diseño y Proyectos, Universidad de Huelva, Huelva (Spain)

[2] Universidad Autónoma de Zacatecas, Zacatecas (México)

[3] Solar Consulting Services, Colebrook (U.S.A.)

[4] Centro de Desarrollo Energético de Antofagasta, Universidad de Antofagasta, Antofagasta (Chile)

[5] Dpto. Química y Física, Universidad de Almería, Almería (Spain)

[6] Photovoltaic Solar Energy Unit (Energy Department, CIEMAT), Madrid (Spain)

Abstract

Solar radiation forecasting has become a critical information technology to facilitate the integration of PV and thermal solar power plants into the electricity grid of any country. Artificial neural network (ANN) modeling of time series is known as a useful and effective forecasting tool to achieve this task, due to its ability to find non-linear relationships hidden inside historical data. Unfortunately, fast cloudiness transients add a stochastic signal to the solar radiation time series, thus diminishing the effectiveness of this methodology. In this work, ANNs are trained to provide 1-day-ahead forecasts of global solar radiation under various cloud regimes. Nine years of data measured under diverse climates at eight stations from the U.S. SURFRAD network are used. Training periods of less than two years are found too short and result in larger errors. Using a training period of eight years, the forecast accuracy is found to depend on cloud fraction (and thus location), with RMS errors ranging from 10% up to 45%.

Keywords: Forecasting, solar radiation, time series, artificial neural networks, PV performance.

1. Introduction

The increase in the use of clean energy sources grows exponentially every year. Solar energy has become a decisive part of the renewable energy market. Solar radiation can be collected and converted into electricity through the use of two mature technologies: solar-thermal power generation and photovoltaic (PV) power generation. Nevertheless, integration of these technologies into electrical networks is a new challenge, due to the variability of the incident irradiance. In order to facilitate solar-thermal and PV penetration, solar energy forecasting is thus required. In particular, forecasting of the incident global horizontal irradiance (GHI) is the first and most essential step in most PV power prediction systems. Among the different methodologies that exist to achieve this task, the use of time series forecasting models based on historical data of solar radiation has been widely employed for several decades (Diagne et al., 2013). Time series forecasting models rely on measurements over a given period of time, where each data point, $x(t)$, corresponds to a specific time t. These models then predict future outputs according to previous events. Compared with other forecasting techniques, time series forecasting is flexible and requires fewer data inputs. It is usually considered an appropriate methodology for forecasting over the short term, at least.

Time series consisting of daily GHI values, $H(t)$, are commonly used to provide one-day-ahead forecasts because of their value toward solving various decision-making problems involved in the electricity market and power system operation (Inman et al. 2013; Alsharif et al. 2019). Unfortunately, time series forecasting models

doi:10.18086/swc.2019.43.05 Available at http://proceedings.ises.org

are site-dependent. This is a consequence of the stochastic variability of solar radiation coming from the different local atmospheric and weather patterns, which are fairly specific to each location. In this respect, clouds are normally the main source of solar irradiation variability. Due to the difficulty of simulating such stochastic variability (López et al., 2008), the performance of time series forecasting is expected to be more or less dependent on cloud regime, in turn resulting into possible geographical limitations.

In this work, the effect of different annual cloud regimes on the forecasting accuracy of the daily $H(t)$ is analyzed with the goal of providing a general assessment of the suitability of using time series for solar radiation forecasting. To this end, GHI forecasting is undertaken by means of artificial neural networks (ANN) because of their known effectiveness, often performing better than other conventional algorithms based on linear models (Paoli et al., 2010).

2. Experimental data

2.1 Radiometric stations

Historical data of daily GHI, ground albedo, air temperature, relative humidity, and pressure are obtained from 1-min observations at eight U.S. stations of the NOAA Surface Radiation Budget Network (SURFRAD; www.esrl.noaa.gov/gmd/grad/surfrad/). These stations are: Alamosa (SLV), Bondville (BON), Desert Rock (DRA), Fort Peck (FPK), Goodwin Creek (GWN), Pennsylvania State Univ. (PSU), Sioux Falls (SFX), and Table Mountain (TBL). The measurement period used here is 2009–2017 for all stations, except SLV (2014–2016). Figure 1 shows the geographical distribution of the stations. They operate in climatologically and topographically diverse regions, ranging in altitude from 98 m to 2317 m above mean sea level. In addition to the diversity of cloud regimes, these sites present different annual cycles of atmospheric turbidity, precipitable water, and ground albedo (López and Batlles, 2004; López et al., 2007). Several quality tests are applied to guarantee the reliability of the irradiance data (Gueymard and Ruiz-Arias, 2016). Additional information about the SURFRAD network has been reported by Augustine et al. (2000).

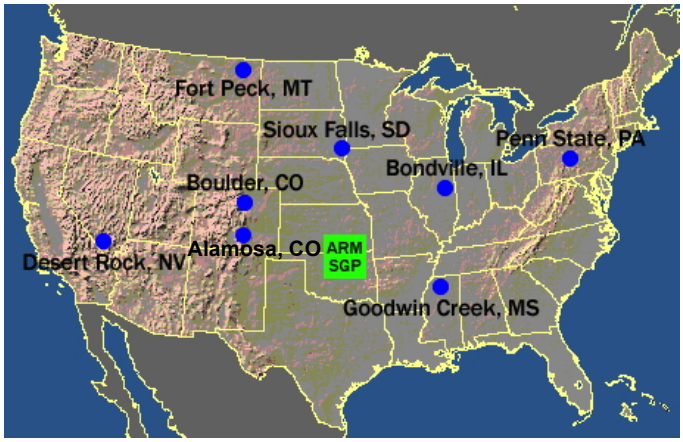

Fig. 1: Geographical distribution of the test stations

2.2 Sky condition classification

To study the impact of cloudiness on the time series forecasting accuracy, each day is classified into three categories of sky conditions: clear, partly cloudy, and cloudy. In order to keep $H(t)$ as the only measured variable, and to avoid the need for additional variables, the sky condition classification is exclusively based on the daily clear-sky index, $K_c(t)$, calculated as the ratio between $H(t)$ and the ideal daily clear-sky GHI. A number of clear-sky radiation models can provide the instantaneous GHI, as reviewed in, e.g. (Gueymard, 2012; Sun et al., 2019). For the present application, the simpler daily clear-sky model of López et al. (2007) is used to directly determine the daily clear-sky GHI, $H_{cs}(t)$, and then the daily-mean K_c through:

$$K_c = H / H_{cs.} \qquad \text{(eq. 1)}$$

The inputs to the clear-sky model are as follows: day of year, site altitude z (in meters), and daily-mean values of precipitable water (PW), w, Ångström's turbidity coefficient, β, and ground albedo, ρ. The latter three variables are derived from NASA's MERRA-2 reanalysis model, which is recommended for this kind of application (Gueymard, 2019). PW and ρ are readily available from MERRA-2, but β needs to be calculated from the aerosol optical depth at 550 nm and the Ångström exponent (both provided by MERRA-2) using Ångström's law.

The main equation of the parameterized daily clear-sky model is:

$$H_{cs} = 0.98 \exp(0.07z / 8345.3) \exp[f(w,\beta)] H_{CD}^{g(w,\beta)} h(\rho,\beta) \quad \text{(eq. 2)}$$

where the functions f, g and h are expressed as

$$f(w, \beta) = -0.249\, w^{0.31225} + 2.81375\beta^2 - 2.5948\beta \quad \text{(eq. 3)}$$

$$g(w, \beta) = 1.00324 + 0.03483 w^{0.28073} - 0.97226\beta^2 + 0.64794\beta \quad \text{(eq. 4)}$$

$$h(\rho, \beta) = 0.98613 + 0.0705\rho - 0.15225\beta + 0.77513\rho\beta \quad \text{(eq. 5)}$$

and H_{CD}—the daily GHI corresponding to an ideal ~~clear~~ clean-dry (CD) atmosphere with neither aerosol nor water vapor—is obtained from López et al. (2007) as

$$H_{CD} = \frac{0.0864}{\pi} I_{SC} E_0 (1.019 - 5.5\ 10^{-4}\varphi)[0.965\, cos\, \delta\, cos\, \varphi\, (sin\, \omega_{sr} - \omega_{sr}\, cos\, \omega_{sr}) - 0.0485\omega_{sr}] \quad \text{(eq. 6)}$$

where I_{SC} is the solar constant, E_0 the eccentricity of the Earth's orbit, φ is the latitude in degrees, δ the declination, ω_{sr} the hour angle at sunrise in radians, and H_{CD} is expressed in MJ m^{-2}.

Figure 2 shows both the measured and the clear-sky GHI time series for the years 2009–2011 at the Bondville station (BON). The modelled values agree remarkably well with the measured GHI under clear conditions, taking into account the slight imprecision due to uncertainties in the w and β inputs or in the model itself. The clear-sky index time series is also displayed in Fig. 2. Values slightly higher than one can be found because the input values used by the model, mainly w and β, can differ somewhat from the actual ones. (The MERRA-2 grid cell size is ≈55 km, which can induce local variance.) Nevertheless, K_c values close to unity represent clear-sky conditions and do not display any seasonal dependency.

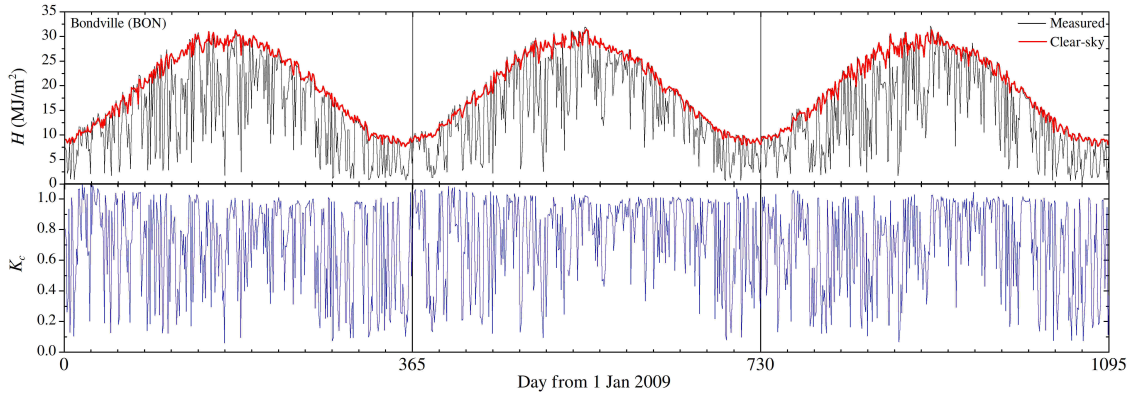

Fig. 2: Time series of both measured daily global irradiation at Bondville and modeled daily clear-sky global irradiation (top), and the corresponding daily clear-sky index (bottom).

Based on a visual inspection of the daily evolution of the measured GHI and the corresponding K_c values, the three types of sky conditions are then defined as:

Cloudy day: $K_c(t) < 0.56$

Partly cloudy day: $0.56 \le K_c(t) < 0.93$

Clear day: $0.93 \le K_c(t)$.

Table 1 provides details about the number of days for each sky condition at each station. Figure 3 shows the daily evolution of 1-min values of GHI, direct normal irradiance, and diffuse irradiance for seven successive days at the Bondville station. The two first days (with K_c = 1 and 0.94, respectively) are considered clear days, despite the presence of clouds in the afternoon during the second day. For generalization purposes, cases corresponding to K_c values in the range 0.93–0.96 are labeled "mostly clear days". The fourth and seventh days (with K_c equal to 0.33 and 0.10, respectively) are totally cloudy. The fifth day, with K_c = 0.59, is almost cloudy. In general, it is found that cloudy days can be described with $K_c < 0.56$. Days with K_c values ranging between 0.56 and 0.93 are affected by broken clouds and classified as partly-cloudy conditions.

Tab. 1: Number of clear, partly cloudy, and cloudy days at eight U.S. SURFRAD stations.

Station	Lat. (°)	Long. (°)	Elev. (m)	Clear days	Partly cloudy days	Cloudy days
Bondville (BON)	40.052	–88.373	230	1075	1284	911
Desert Rock (DRA)	36.624	–116.019	1007	2309	764	182
Fort Peck (FPK)	48.308	–105.102	634	1104	1549	601
Goodwin Creek (GWN)	34.255	–89.873	98	1181	1220	806
Penn. State Univ. (PSU)	40.720	–77.931	376	691	1390	1175
Sioux Falls (SXF)	43.734	–96.623	473	1173	1322	768
Alamosa (SLV)	37.697	–105.923	2317	376	277	47
Boulder (TBL)	40.125	–105.237	1689	1242	1539	490
ALL STATIONS	—	—	—	9151	9345	4980

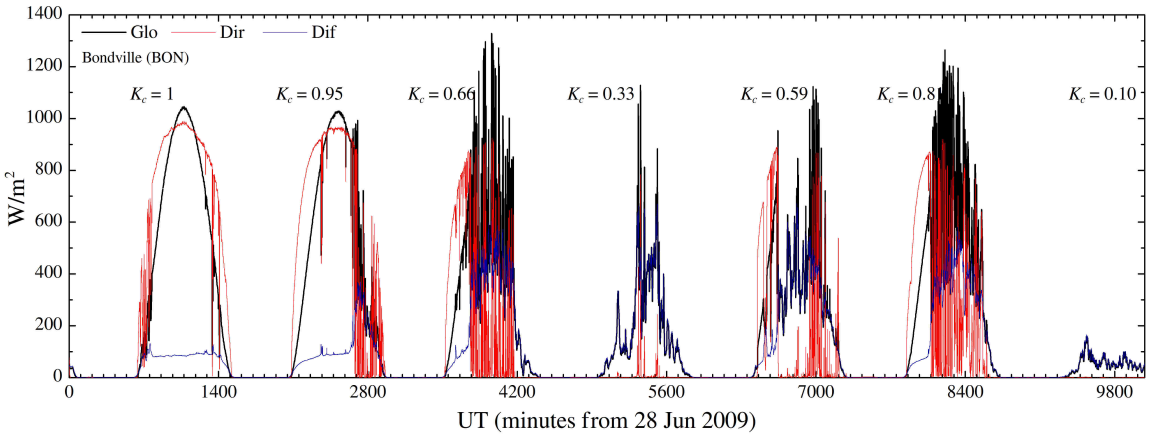

Fig. 3: Daily evolution of 1-min measured values of global horizontal, direct normal, and diffuse irradiances during seven days at the Bondville's station. The corresponding daily clearness index is also added for each day. All three categories of sky condition (clear, partly cloudy, and cloudy) are observed during this sequence.

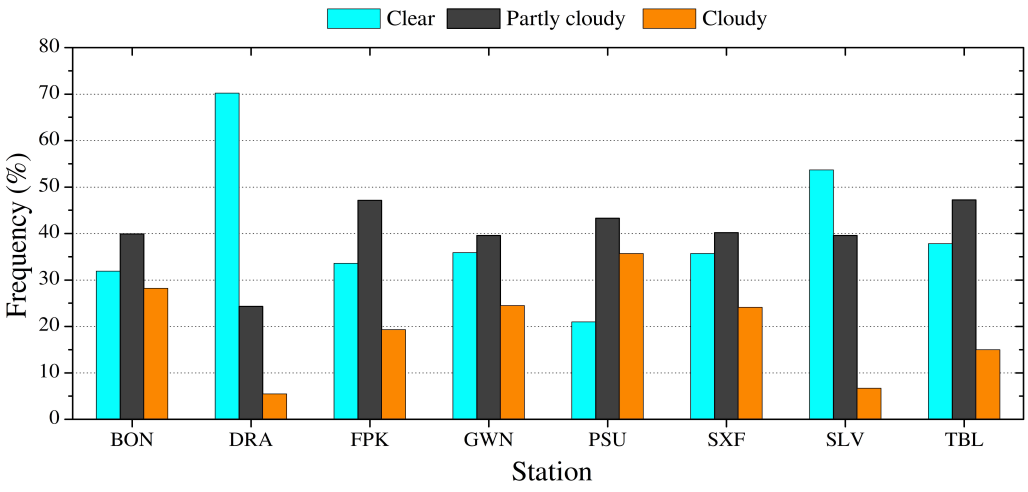

Fig. 4: Frequency of days corresponding to the three categories of sky conditions for the eight SURFRAD stations.

For the eight SURFRAD stations, Fig. 4 shows the frequency of days in each of the three categories defined above to characterize the cloud regime. DRA and SLV have the higher fraction of clear days in the period considered, reaching about 70% and 55% of the total number of days, respectively. These two stations are also rarely experiencing cloudy days, with a frequency of only ≈5%. In sharp contrast, PSU is the station with the lowest number of clear days (≈21%) and with the highest occurrence of cloudy days, accounting for almost 35% of the total number of days. The remaining stations are somewhere in between these extremes.

Finally, for the purpose of developing and testing the forecasting models, each dataset is divided into two subsets: (i) a training dataset; and (ii) a test dataset. For all stations, except SLV, the years 2009–2016 are used for training, and the year 2017 is used for testing. For SLV, whose measurements lasted only from 2014 to mid-2016, the training dataset uses the period 2014–2015 and the test dataset uses the first half of 2016.

3. Artificial neural network description

The ANN used in this work is a multilayer feedforward network trained by means of the Levenberg-Marquardt algorithm. The inputs to the ANN consist of the values $H(t - i)$, i.e., from the i previous days. The assumption is that GHI information lagging i days has still forecasting skill for day t. The ANN output is the 1-day-ahead value $H(t + 1)$. Figure 5 shows the architecture of the multilayer perceptron network used here.

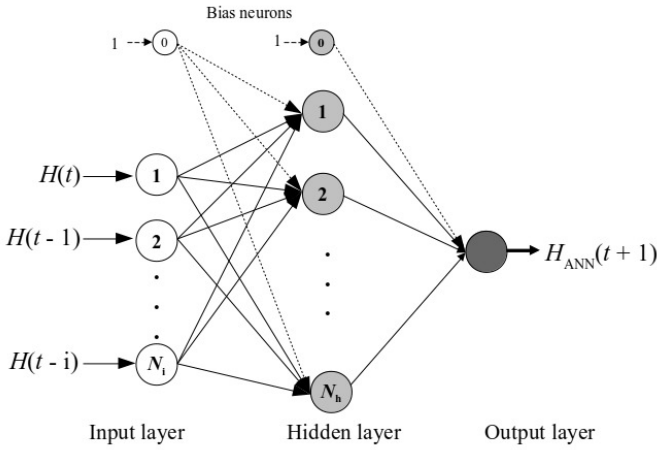

Fig. 5: ANN model architecture. The ANN output is the 1-day-ahead forecasted GHI, $H_{ANN}(t + 1)$.

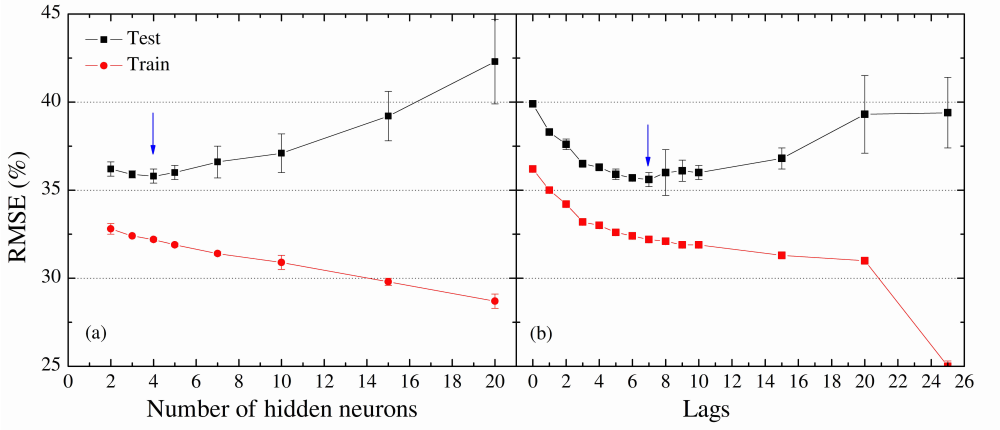

Fig. 6: ANN performance, in terms of percent RMSE of estimated versus measured $H(t + 1)$ values for the GWN station, as a function of: (a) the number of hidden neurons and using an 11-day lag in the input vector; and (b) increasing lags with 4 hidden neurons. The blue arrow shows the optimum case to be selected. Error bars denote the standard deviation of the 10-run mean.

Several trials need to be undertaken first to determine the optimum number of hidden neurons N_h and lagging days i. For that, the GWN station is selected because it presents a similar number of days in each of the three

sky categories (Fig. 4). Repeating these trials at the seven other sites did not change the results (in terms of N_h and lags) described below. The maximum lag is initially set to i = 10, based on a preliminary analysis. Several networks with different numbers of hidden neurons N_h are trained and the best performing network among them is selected. The ANN forecasting performance is evaluated in terms of three common statistical indices: coefficient of determination R^2, root-mean-square error RMSE, and mean bias error MBE. The latter two indicators are expressed in percent of the mean GHI for each station, and are additionally calculated for each class of sky condition separately. (Note that the mean GHI is specific to each data class.) To overcome potential differences in model performance that could be caused by the random initialization of the weights and the oscillating effects of local minima in the performance surface, ten training runs are undertaken for each ANN architecture, and the overall performance is reported as the mean RMSE and MBE of these ten runs. Standard deviations are also calculated to analyze the stability of the ANN predictions. For each N_h value, the ANNs predict unbiased results (MBE ≈0%). The RMSE for the training data decreases as N_h increases, as Fig. 6a shows. However, the opposite tendency is observed for the testing data, which is a consequence of overfitting—a known issue in ANN modeling. Based on the RMSE for the testing data, the best number of hidden neurons is then N_h = 4. Once the number of hidden neurons is fixed, the next step consists in optimizing the number of lag days. A procedure similar to the previous one is undertaken, but now varying the number of lags i. Figure 6b shows that the RMSE decreases when the lag increases from 0 to 7, for both the training and testing datasets. Beyond that point, the RMSE of the training data continues to decrease but the test values are then poorly forecasted. The explanation is that an increase in lag time leads to RMSE reduction as a consequence of overfitting of the training data, however associated with a dramatic deterioration in forecasting skill with the test data.

4. Results

With the intention of evaluating the benefits of using the ANN methodology, the results are also compared with those obtained by means of the persistence model, used here as a naïve predictor. The persistence model supposes that global irradiation at day t + 1 is best predicted by its value at day t (Diagne et al., 2013).

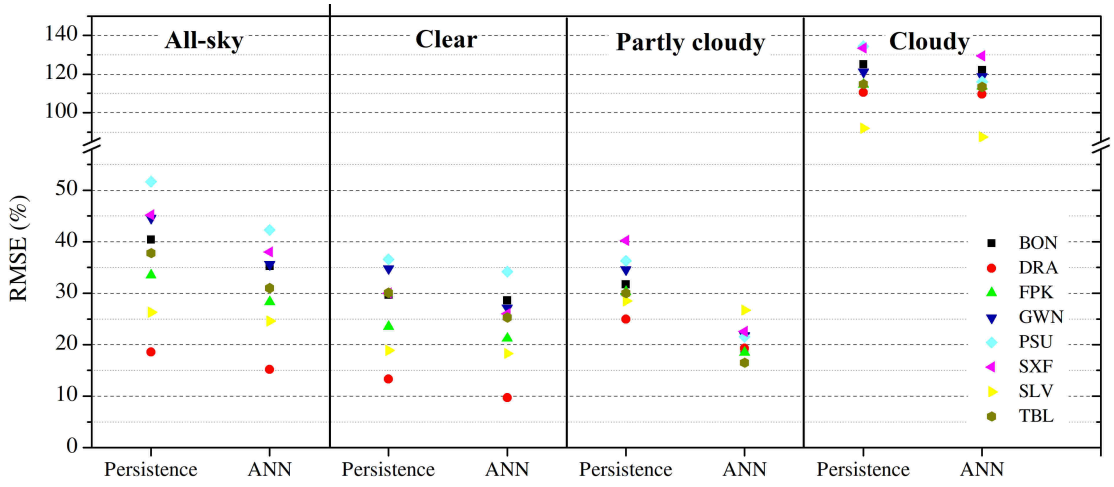

Fig. 7: RMSE for 1-day-ahead forecasting of daily GHI using the persistence model and the ANN model, for the eight SURFRAD stations and as a function of the three sky conditions defined in the text.

Figure 7 shows RMSE results for 1-day-ahead forecasting using persistence and the ANN model operated with seven successive days preceding the last day with measured data (i.e., eight successive days preceding the day being forecasted (t+1)), separately considering the three categories of sky conditions for the test data, as well as their combination (all-sky conditions). Under all-sky conditions, the RMSEs resulting from the persistence model range from 19% at DRA to 51% at PSU. As indicated in Section 2, DRA and PSU are at the two extremes regarding the number of clear and cloudy days. Still using persistence, the next second best performance with RMSE = 26% is at SLV, which has a high fraction of clear days and a low number of cloudy days, similarly to DRA. On the other hand, locations where partly-sky and overcast-sky conditions predominate, such as SXF

or GWN, forecasts using persistence become inaccurate, with RMSEs larger than 45%. As could be expected, the performance improves during clear days and degrades during cloudy days. The RMSE increases to values higher than 100% when attempting to forecast cloudy days. Nevertheless, these higher RMSE values are also a consequence of the low mean values used to calculate the percent values. In any case, this suggests that the persistence model is not appropriate for sites with a high frequency of partly cloudy and/or cloudy periods. The large disparity of results between clear and cloudy days could be expected because, in the former case, the solar irradiance follows a smooth, easily predictable deterministic progression, whereas, in the latter case, the large variability added by the presence of clouds can be described by a stochastic, hardly predictable process.

The ANN forecasting methodology improves the prediction accuracy in comparison with the baseline persistence model at each site and under any annual cloud pattern. Figure 8 shows the reduction in RMSE that results from using the ANN model rather than persistence. In general, under all-sky conditions, the ANN model decreases the forecasting error by ≈2–10% in terms of RMSE, depending on location. This result is also observed under "clear" conditions. This can be explained by how loosely a clear day is defined here: The present definition can actually accept some cloudiness (as illustrated by day 2 in Fig. 3). At most stations, a large RMSE reduction is achieved by the ANN model during partly cloudy days, decreasing the RMSE by ≈2–18% relatively to the persistence model. When considering cloudy days, a better quality of predictions is observed at all stations, and, in particular, a significant improvement occurs at PSU, with an RMSE reduction of 18%. This specific site remarkably being the cloudiest, it can be concluded that the ANN performance depends on the amount of days affected by the presence of clouds. Hence, this improvement in daily global irradiation forecasting with ANN can be tentatively explained as a sampling effect: locations with only infrequent cloudy days need longer training time series than other locations with a higher frequency of cloudy days. The satisfactory results provided by the ANN model under partly cloudy days in comparison with the persistence model are then due to the availability of long time series corresponding to that category.

In order to get a better understanding of the ANN benefits, the present performance results could profit from a comparison with those obtained using the novel forecasting skill metric proposed by Marquez and Coimbra (2013), which is well adapted to forecasting time series. This development is underway and will be reported in a subsequent and more general contribution.

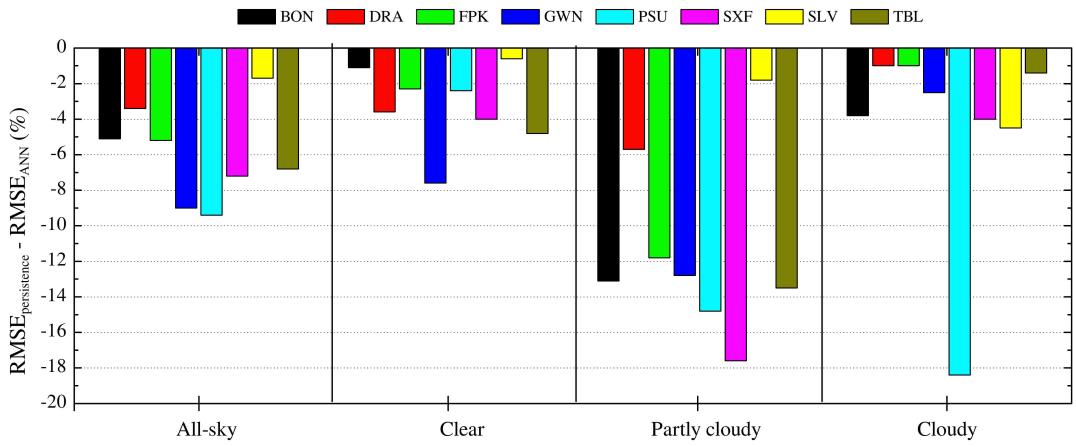

Fig. 8: Reduction in percent RMSE (compared to persistence) when using the ANN model relatively to persistence at each SURFRAD station and each type of sky condition.

The question then arises as to how many years of training are needed to improve the results using the ANN methodology. Figure 9 shows the results of ANN performance at FPK when using training time series consisting of an increasing number of years. After 4 or 5 years, the RMSE has reached a stable level, with an improvement of ≈6% under all-sky conditions, compared to persistence. A similar finding is also obtained if only clear days are rather considered, leading to a slight improvement of ≈2% compared to persistence. In contrast, forecasts of partly-cloudy days are much improved (by ≈12% compared to persistence) using a longer times series of ≈7–8 years. Conversely, the number of years used for the ANN training of cloudy conditions does not seem to have any effect at all. In any case, and to get the best performance under all possible conditions, time series of at least 7–8 years should be used.

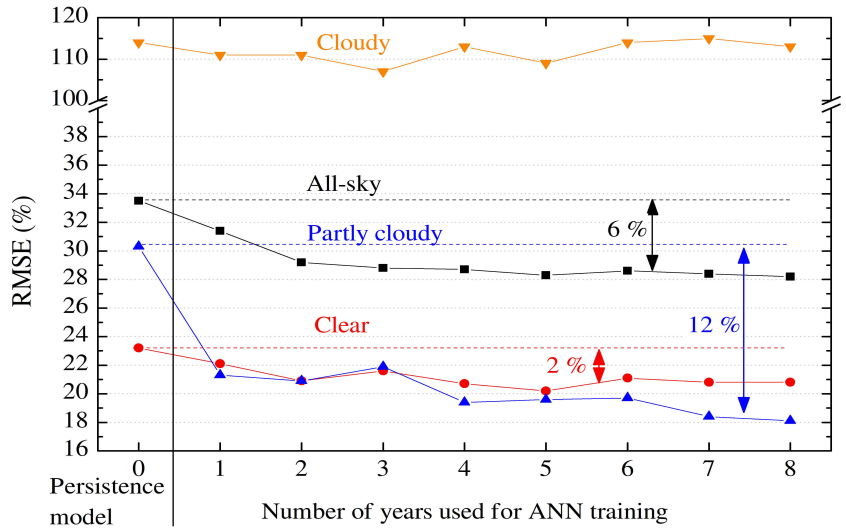

Fig. 9: RMSE when training the ANN with time series having an increasing number of years at the FPK station. The RMSE from the persistence model is also indicated (at year 0).

Figure 10 displays scatterplots comparing the forecasted and measured $H(t + 1)$ results at PSU, using either persistence or the ANN model, separately for the three classes of daily cloudiness. Good overall agreement is found when using ANN forecasting for partly cloudy days. Cloudy days, however, are affected by significant overestimation.

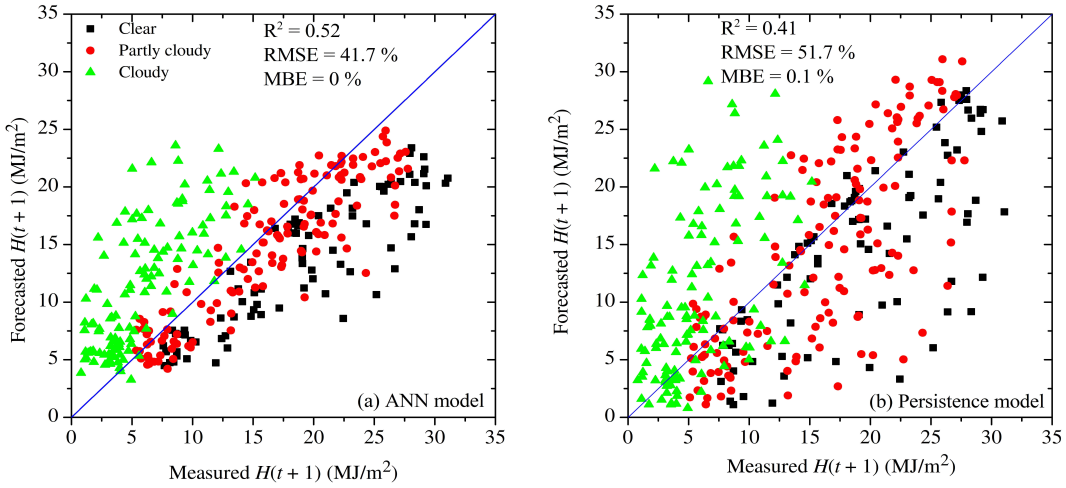

Fig. 10: Comparison at PSU between forecasted and measured next-day GHI for the 2017 test year, and using (a) the ANN model and (b) the persistence model, according to three types of daily sky conditions: clear, partly cloudy, and cloudy. Statistical indicators of model performance are also provided.

Figure 11 shows the time series of 1-day-ahead GHI forecasted by the ANN model and the corresponding measured GHI at the DRA and PSU stations. A good overall agreement is observed at DRA. This is a consequence of two effects: (i) an extensive time series of clear days exists at that site, and such forecasts are typically accurate; and (ii) even though the ANN learning ability decreases due to the random nature of cloud effects, there are only a few cloudy days there, so that the overall performance is not affected. At PSU, the ANN reproduces the variability in the time series acceptably well. However, it is neither able to correctly predict the maximum GHI values (during very clear days) nor the minimum GHI values (during dense overcast days). Considering the low MBE observed at that site (Fig. 10), it is likely that this issue is site-specific and a consequence of the dataset characteristics. A possible remedy, that remains to be investigated, would be to use a longer GHI time series for training.

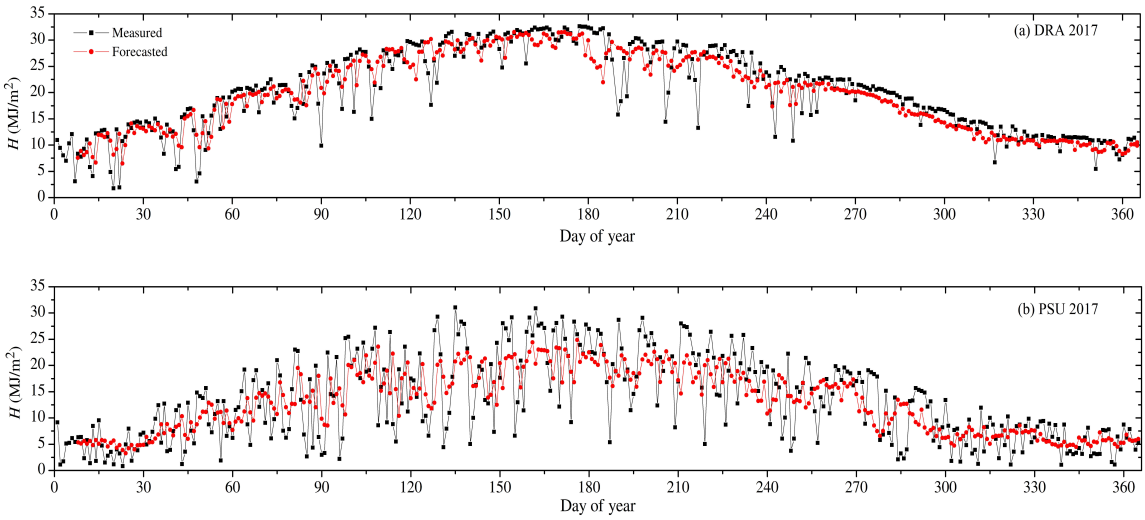

Fig. 11: Time series of 1-day-ahead GHI forecasted by the ANN model and the corresponding measured GHI at DRA (top) and PSU (bottom) using the test data (year 2017).

At SLV, the ANN forecasts perform similarly to the persistence model despite the relatively large fraction of clear days. This is apparently caused by the small amount of data that was available for training—only one year and a half. Hence, it can be concluded that an ANN model does require long training time series to improve performance relatively to persistence and to achieve reasonably accurate forecasts under all possible sky conditions.

To evaluate the effect of the amount of data in each of the three types of sky conditions on the ANN model performance, a second ANN model is developed for the GWN station, but now using the same number of clear days, partly cloudy days, and cloudy days. The total training GWN dataset of the initial model had 864 clear days, 1098 partly cloudy days, and 694 cloudy days. For the new ANN model, the number of days corresponding to clear and partly cloudy categories is reduced to 694 (the same as the number of cloudy days) by means of a random selection. The performance of this new ANN model, evaluated in terms of percent RMSE for the test data year, amounts to 29%, 23% and 102% respectively for clear, partly cloudy, and cloudy cases, comparatively to 26%, 19% and 112% for the original model. It is thus found that the new ANN model improves the forecasting skill during cloudy days but at the expense of performance degradation under clearer conditions, just because of their lower number of training days.

Finally, other ANN trials have been undertaken using additional exogenous inputs, such as daily mean values of temperature, relative humidity, atmospheric pressure, or direct irradiance during the last few days. As a result, the ANN overfitted the training datasets, consequently precluding the proper forecasting of the test data. This suggests that the exogenous variables selected in this preliminary evaluation do not add any significant information compared to historic GHI data. Further tests, now underway, will involve other exogenous variables, as well as more advanced ANN architectures, including multi-stage ANN (Kemmoku et al., 1999), hybrid models (Blaga et al., 2019), and multimodel ensembles (Zemouri et al., 2019).

5. Conclusions

Using eight radiometric stations providing high-quality irradiance data in the USA, this work has shown the capability of ANN models to provide 1-day-ahead forecasts of daily GHI with satisfactory results. Improved statistical results have been found in comparison with the conventional persistence model. For cloudy days, the ANN model performance is not as good (in relative terms) as that under clearer conditions because of the relative lack of data and the lower irradiance. During cloudy periods, the forecast results are improved only at those stations where such days are frequent. With the present database of eight reference stations in the USA, almost all stations have a high frequency of partly cloudy days. It is here shown that this is a condition needed in order for the ANN methodology to perform better than the simple persistence model. It is also shown that, to guarantee accurate and well-balanced ANN forecasts, it can be beneficial to rearrange the training dataset so as to

select a similar fraction of days in each of the three cloud regimes defined here. In this sense, eight years of daily data may not be enough to obtain very accurate predictions.

Finally, the addition of times series of exogenous inputs (in the form of common meteorological variables) has been found disappointing because it does not improve the current ANN model, possibly because it has a simple architecture. More efficient architectures, possibly involving other machine-learning techniques, are expected to improve this situation. Nevertheless, the ANN-based method introduced here for the forecasting of daily GHI values appears useful as an input to other existing methodologies aimed at forecasting the 1-day-ahead PV power production and/or to improve their results. Further research will evaluate the optimum number of years that is necessary to maximize the forecasting performance of GHI under various cloud climates.

6. Acknowledgments

The authors would like to thank the PVCastSOIL Project with reference (ENE2017-83790-C3-1, 2 and 3), which is funded by the Spanish Ministry of Economy and Competitiveness and co-financed by the European Regional Development Fund. At the international level, the authors were also supported by CONICYT/FONDAP/15110019 "Solar Energy Research Center" SERC-Chile, the Chilean Economic Development Agency (CORFO) with contract No 17PTECES-75830 under the framework of the project "AtaMoS TeC" and the generous financial support provided by the GORE-Antofagasta, Chile, Project FIC-R Antofagasta 2017 BIP code 30488824-0.

7. References

Alsharif, M.H., Younes, M.K., Kim, K., 2019. Time series ARIMA model for prediction of daily and monthly average global solar radiation: The case study of Seoul, South Korea. Symmetry, 11, 240. DOI: 10.3390/sym11020240.

Augustine, J.A., DeLuisi, J.J., Long, C.N., 2000. SURFRAD—a national surface radiation budget network for atmospheric research. Bull. Am. Met. Soc. 81, 2341–2357. DOI:10.1175/1520-0477(2000)081<2341: SANSRB>2.3.CO;2.

Blaga , R. , Sabadus, A., Stefu, N., Dughir, C., Paulescu, M., Badescu, V., 2019. A current perspective on the accuracy of incoming solar energy forecasting. Prog. Energy Combust. Sci., 70, 119–144.

Diagne, M., David, M., Lauret, P., Boland, J., Schmutz, N., 2013. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. Renew. Sust. Energy. Rev., 27, 65–76. DOI: 10.1016/j.rser.2013.06.042.

Gueymard, C.A., 1994. Analysis of monthly atmospheric precipitable water and turbidity in Canada and northern United States. Sol. Energy, 53, 57–71. DOI:10.1016/S0038-092X(94)90606-8.

Gueymard, C.A., 2012. Clear-sky irradiance predictions for solar resource mapping and large-scale applications: Improved validation methodology and detailed performance analysis of 18 broadband radiative models. Sol. Energy, 86, 2145-2169. DOI:10.1016/j.solener.2011.11.011.

Gueymard, C.A., 2019. Clear-sky radiation models and aerosol effects. In: Polo, J., Martin-Pomares, L., Sanfilippo, A. (Eds.), Solar Resources Mapping Fundamentals and Applications, Springer, pp. 137–182.

Gueymard, C.A. and Ruiz-Arias, J.A., 2016. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance. Sol. Energy, 128, 1–30.

Inman, R.H., Pedro, H.T.C., Coimbra, C.F.M., 2013. Solar forecasting methods for renewable energy integration. Prog. Energy Combust. Sci., 39, 535–576. DOI:10.1016/j.pecs.2013.06.002.

Kemmoku, Y., Orita, S., Nakagawa, S., Sakakibara, T., 1999. Daily insolation forecasting using a multi-stage neural network. Sol. Energy, 66, 193–199.

López, G. and Batlles, F. J., 2004. Estimate of the atmospheric turbidity from three broad-band solar radiation algorithms. A comparative study. Ann. Geophys., 22, 2657–2668. DOI:10.5194/angeo-22-2657-2004.

López, G., Batlles, F. J., Tovar-Pescador, J., 2007. A new simple parameterization of daily clear-sky global solar radiation including horizon effects. Energy Convers. Manag., 48, 226 –223. DOI:10.1016/j.enconman.2006.04.019.

López, G., Batlles, F. J., Bosch, J.L., 2008. Nonlinear analysis of daily global solar radiation time series. Proc. EuroSun 2008 Conf., ISES.

Marquez, R., Coimbra, C.F.M., 2013. Proposed metric for evaluation of solar forecasting models. J. Solar Energy Eng., 135, 011016. DOI:10.1115/1.4007496.

Paoli, C., Voyant, C., Muselli, M., Nivet, M.L., 2010. Forecasting of preprocessed daily solar radiation time series using neural networks. Sol. Energy, 84, 2146–2160. DOI:10.1016/j.solener.2010.08.011.

Sun, X., Bright, J.M., Gueymard, C.A., Acord, B., Wang, P., Engerer, N.A., 2019. Worldwide performance assessment of 75 global clear-sky irradiance models using Principal Component Analysis. Renew. Sust. Energy. Rev., 111, 550–570. DOI: 10.1016/j.rser.2019.04.006.

Zemouri, N., Bouzgou, H., Gueymard, C.A., 2019. Multimodel ensemble approach for hourly global solar irradiation forecasting. Eur. Phys. J. Plus (2019) 134: 594. DOI:10.1140/epjp/i2019-12966-5.