Effectiveness of Cloud Cover on Solar Radiation Prediction Using ANN Algorithm

Hany Gaballa¹, Soolyeon Cho²
¹Military Technical College, Cairo, Egypt
²College of Design, North Carolina State University, Raleigh, NC, USA

Abstract
Solar radiation data is highly desirable in various areas, such as agriculture, PV industries, and building performance analysis. There are no commercial tools available for real-time solar radiation prediction using only readily available data such as temperature and humidity. The purpose of this paper is to test if solar radiation can be predicted within acceptable uncertainties using only temperature and humidity, based on ASHRAE Guideline 14-2014. It is also discussed how much “cloud cover” affects solar radiation prediction when using ANN algorithms. Three climate locations are tested in Arizona, Washington, and North Carolina. Studies show that the prediction of solar radiation using ANNs depends primarily on the correlation between cloud cover and global solar radiation, rather than on the cloud cover ratio.

Key Innovations
- This paper proved that using the ANN algorithm in predicting solar radiation could be possible using only temperature and relative humidity without including cloud cover.
- Cloud cover has a minimal effect on solar radiation prediction when using ANN algorithms.

Practical Implications
This paper provides information to those interested in real-time solar radiation data to consider using ANN algorithms in predicting solar radiation without including cloud cover as an input parameter.

Introduction
Solar energy is considered the ultimate source of energy on our planet. It is essential to store solar radiation data in weather files for building energy performance simulations. Typical Meteorological Year (TMY) files contain solar radiation data, but these are historical data, not real-time. Real-time simulation today is becoming more demanding in real-time performance analytics. Simulation software requires direct and diffuse solar radiation to perform building simulations. Suppose we have global solar radiation data. In this case, we can easily compute direct and diffuse solar radiation using the Perez model via the EnergyPlus Weather Converter Program (EERE, 2018) or via TRNSYS software (Gaballa et al., 2020). Splitting from global calculated or predicted values into direct normal and diffuse horizontal components is discussed in more detail by Perez et al., (1990), Perez et al., (1992), and Perez et al., (1993).

Researchers in previous papers have various disciplines dealing with solar radiation prediction at time intervals (hourly, daily, or monthly). Solar radiation measurements require many resources, which is why there are very few solar measurement stations around the world. Some other researchers made studies to evaluate the weakness and strength points of several sun-sky simulation models (Kider et al., 2014), and (Bruneton, 2017).

It is logical to think about finding more efficient and effective ways to predict solar radiation. Although many papers have presented different approaches, the performance ranking of these methods is complex due to the diversity of prediction methods, input data types, and time steps. Different methods were used to predict solar radiation, where Artificial Neural Networks (ANNs) are considered one of the most accurate methods (Kashyap et al., 2015). This paper focuses on the prediction of global solar radiation per hour required for real-time building simulation. According to Alam et al. (2009), ANN models predict the hourly and daily solar radiation with higher accuracy than empirical models.

Developed by Gaballa & Cho (2019) used in this paper, the ANN model aims to use parameters that can only be easily used as inputs, temperature, and relative humidity, to predict global solar radiation. Cloud covers significantly affect solar radiation prediction when used through statistical methods (Sarkar, 2016). However, cloud cover is not readily available, so this paper aims to see the effectiveness of cloud cover on solar radiation prediction when using ANN methods in different climatic regions with different conditions.

Three cases with different climatic conditions were selected for the case study, including Phoenix, AZ, Seattle, WA, and Raleigh, NC, which are located in climate zone 2B (hot and dry), 4C (mixed-marine), and 4A (mixed-humid). These three cases also have different numbers of cloudy and clear days.

Case Studies
The three cases selected are different in the weather conditions and are also different in the number of cloudy and sunny days. To make it clear, Figures 1, 2, and 3 show only the average temperature, average relative humidity, and average hourly global solar radiation for the three cases. The average temperature and global solar radiation comparison show a uniform distribution for the three
cases, as shown in Figures 1 and 3. In July, the average temperature reaches the highest in Phoenix and Raleigh, while Seattle reaches the highest temperature in August. Phoenix, Raleigh, and Seattle's highest average temperature is 96.0 F, 78.8 F, and 68.1 F, respectively. Phoenix's lowest average temperature is 53.1 F in December, while Raleigh and Seattle are 39.5 F and 41.9 F in January. From Figure 3, the highest average hourly global solar radiation is 346.6 Wh/m², 266.4 Wh/m², and 274.5 Wh/m² for Phoenix, Raleigh, and Seattle, while the lowest is 127.9 Wh/m², 93.0 Wh/m², and 32.8 Wh/m².

On the other hand, Figure 2 shows the average relative humidity comparison, which has a non-uniform distribution over the year. For example, February shows the highest average humidity percentage for Phoenix, but shows the lowest rate for Raleigh. Phoenix has the lowest average humidity percentage distribution all over the year, the highest percentage is 51.6% in February, and the lowest is 19.5% in June. In Raleigh, the highest average humidity is 80% in October, while the lowest in February is 58%. Seattle shows the highest average humidity percentage of 81.9% in January, while the lowest is 60.5% in August.

Figure 3: Average hourly global solar radiation comparison
Figure 4 shows a comparison between the numbers of cloudy and sunny days. Phoenix has the highest portion of sunny days, it reaches 247 days, while Seattle has the lowest number, only 64 sunny days. Raleigh came in the middle; it has 125 sunny days. About the cloudy days, Seattle has the highest overcast days, 138 days, while Phoenix has only 17 overcast days, and again Raleigh has a moderate number, 86 days.

For more clarity, Figure 5 shows a graph for the number of sunny, partly cloudy, and overcast days distributed all over the year. We can see clearly from the three graphs that Seattle has the highest cloud coverage, Phoenix has the lowest, while Raleigh came in the middle.
Data Acquisition and ANN Architecture

In this paper, the ANN algorithm is used to predict the global solar radiation every hour. One of the most effective steps in dealing with the ANN techniques is the data required for training and testing the algorithm. In this paper, the developed ANN model by Gaballa & Cho (2019) is used twice for the three case studies to predict the global solar radiation. The training and testing data are picked from the Typical Meteorological Year (TMY) weather data files for the three cases to perform the prediction.

First, the training data tells the algorithm about the correlation between the inputs and the output. It gives a weight factor for each input parameter correlated to the output. Each weight reflects the input parameter's strength affecting the output result, and it changes in each step to find the optimum weight value through the back-propagation process. After this step, the activation function takes place, which is the sigmoid function in this paper. Between the input and the output layer, a hidden layer consists of some of the hidden neurons where the calculations are made, and the number of hidden neurons is considered a variable. The process passing through the input, hidden, and output layer is called a feed-forward neural network. The number of epochs is regarded as a second variable; the number of iterations to train the algorithm through the feed-forward and back-propagation process. The third variable is the learning rate, which settles the weight changes in each epoch. In this algorithm, an optimization method is used to find the optimum value of the three variables.

The testing data feed the algorithm with the input parameters to predict the output. Generally, in this paper, the data used are temperature, relative humidity, solar zenith angle, and cloud cover from the TMY.

In the first scenario, temperature, relative humidity, solar zenith angle, month, day, and hour are used as input parameters, as shown in Figure 6. This figure shows the original ANN architecture used by Gaballa & Cho (2019). In the second scenario, cloud cover is added to the input parameters, which means seven layers are used as input parameters. The output layer remains the same, which is the hourly global solar radiation.

The error difference between the predicted and measured solar radiation is calculated in each simulation step to ensure that the data predicted meets ASHRAE Guideline 14-2014 limits.

Statistical Indices

To consider a model calibrated, there are limits of the CV(RMSE), and NMBE should be respected. Depending on the calibration time interval, if it is monthly or hourly, the limit is subjected to slight differences, as shown in Table 1, which shows ASHRAE Guideline 14-2014 limits.

Table 1: Limits of statistical indices required for calibration

<table>
<thead>
<tr>
<th>Statistical Indices</th>
<th>Monthly Calibration</th>
<th>Hourly Calibration</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV(RMSE) [%]</td>
<td>±5</td>
<td>±10</td>
</tr>
<tr>
<td>NMBE [%]</td>
<td>±5</td>
<td>±10</td>
</tr>
</tbody>
</table>

If a model is calibrated within these limits, it means that the simulated model is sufficiently close to the actual measurements with acceptable predictive capabilities. In this research, all data are predicted hourly; that is why the used statistical indices limits are 30% and 10% for CV(RMSE) and NMBE, respectively.

Simulation and Results

To ensure that the ANN algorithm is valid in different climate conditions, three different cases are used. Phoenix, Raleigh, and Seattle were tested for the prediction accuracy of global solar radiation using the ANN algorithm developed by Gaballa & Cho (2019). Two scenarios have been tested as mentioned before; the first scenario is using temperature (T), relative humidity (RH), and solar zenith angle (SZA); the second scenario is using the same parameters mentioned besides cloud cover (CC). In the following few graphs, Figure 7, 8, and 9, a comparison is presented between measured and predicted solar radiation before and after including cloud cover as an input parameter in the ANN algorithm. Python 3.7 software is used to perform the simulation process.
CV(RMSE), NMBE, and R-squared values were calculated to see the difference in prediction before and after including the cloud cover. Table 2 shows these percentages for the three cases in both scenarios.

<table>
<thead>
<tr>
<th>Table 2: Ann input parameters and their effect on the output results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phoenix, AZ</td>
</tr>
<tr>
<td>T, RH, SZA</td>
</tr>
<tr>
<td>CC</td>
</tr>
<tr>
<td>CV(RMSE)</td>
</tr>
<tr>
<td>NMBE</td>
</tr>
<tr>
<td>R-squared</td>
</tr>
</tbody>
</table>

Both scenarios show a prediction accuracy within ASHRAE Guideline 14-2014 limits. However, Phoenix showed the best accuracy with or without including cloud cover than Raleigh and Seattle.

So, looking at the first scenario before including cloud cover, Phoenix, Raleigh, and Seattle show 14%, 19.5%, and 21.6% of the CV(RMSE), respectively. To see the reason behind these differences, a regression analysis was made between each input parameter and global solar radiation for each case, as shown in Figure 10.
Figure 10: Regression analysis between individual input parameter and global solar radiation

a) Phoenix, AZ  
b) Raleigh, NC  
c) Seattle, WA

Figure 11: Correlation between cloud cover and global solar radiation

Looking at the second scenario after adding cloud cover to the input layer, Phoenix, Raleigh and Seattle show 12.0%, 18.2%, and 17.0% of the CV(RMSE), respectively, with a difference of 2.0%, 1.3%, and 4.6% when compared to the first scenario. These results should get attention, as Seattle showed the most significant
difference after adding cloud cover; however, it has the lowest cloud coverage. To come up with the reason of these differences before and after including cloud cover, the correlation between cloud cover and global solar radiation is presented in Figure 11.

Raleigh and Phoenix show no correlation between cloud cover and global solar radiation. While in Seattle, the ANN algorithm understands that there is no solar radiation over 600 Wh/m² when it is overcast, (cloud cover is 10). So, this clarifies why Seattle has a 4.6% difference of CV(RMSE) after adding cloud cover to the input layer.

So, solar radiation prediction using ANN doesn't mainly depend on the cloud cover percentage on a specific location but on the correlation between cloud cover and global solar radiation. To prove that and make it clearer, Cairo, Egypt has a hot climate, looks like the weather in Phoenix but with different correlations between solar radiation and other parameters as shown in Figure 12.

The graph shows that the correlation between global solar radiation and solar zenith angle in Cairo, Egypt has more consistency than Phoenix, Raleigh, and Seattle. It also shows a good relationship between solar radiation and cloud cover. To see the effect of these correlations, Table 3 shows the results after running the prediction process using the ANN model before and after adding cloud cover to the inputs.

<table>
<thead>
<tr>
<th>Cairo, Egypt</th>
<th>T, RH, SZA</th>
<th>T, RH, SZA, CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV(RMSE)</td>
<td>12.0%</td>
<td>3.6%</td>
</tr>
<tr>
<td>NMBE</td>
<td>-0.4%</td>
<td>0.3%</td>
</tr>
<tr>
<td>R-squared</td>
<td>96.8%</td>
<td>99.7%</td>
</tr>
</tbody>
</table>

It was found that before including cloud cover in the input layer, the prediction showed a CV(RMSE) of 12%; however, after adding cloud cover, it showed only 3.6%. This clarifies the significant effect of cloud cover on prediction only if it has a good correlation with global solar radiation.

**Conclusion**

Four different climate zone cases were tested using the ANN prediction algorithm: Phoenix, Raleigh, Seattle, and Cairo. The prediction process was made using two scenarios, 1) using temperature, humidity, and solar zenith angle, 2) using temperature, humidity, solar zenith angle, and cloud cover. The ANN output, the hourly global solar radiation, showed promising results compared to measured data with uncertainty rates within ASHRAE Guideline 14-2014 limits.

- Cloud covers have a significant impact on solar radiation prediction when used through statistical methods, but minimize impact when used in ANN models for short time intervals. There are several points in the background as follows.
- Cloud cover has a significant impact on surface temperature. That is why the surface temperature can indirectly reflect the effect of cloud cover on solar radiation when using the ANN models.
- ANN algorithm must first normalize the input parameters between 0 and 1 before starting the prediction process. For temperature, there are hundreds of different values in infinite decimal places. However, there are only ten decimal places in the cloud cover: 0.1, 0.2, ..., and 1. The reason why cloud cover has the least impact on ANN models is that ANN is very sensitive to handling decimal places. The higher the decimal places for each input, the greater the opportunity for the ANN training process and consequently the higher the accuracy of the predicted output.
- The regression analysis showed individually the correlation between each parameter and global solar radiation. Cloud covers are not correlated with global solar radiation in most cases, but in some cases show minimal correlation.
- The prediction process using the ANN model showed some differences when adding cloud cover to the input parameters. These differences do not depend on the cloud cover in a specific location but on how much a good correlation is between cloud cover and solar radiation in this location.
- Based on the cloud cover minimal effect on the prediction of hourly solar radiation when using ANN models, cloud cover can be excluded from the input layer especially that without including cloud cover, we have a prediction accuracy within ASHRAE Guideline 14-2014 limits.
- Therefore, global solar radiation can be predicted using ANN algorithms with readily available input parameters such as temperature, humidity, and solar zenith angle, time of day, day of the month, and month of the year.

References