Using Multi-Pyranometer Arrays and Neural Networks to Estimate Direct Normal Irradiance

Vivek Srikrishnan\textsuperscript{a}, George S. Young\textsuperscript{a,b}, Lucas Witmer\textsuperscript{a}, Jeffrey R. S. Brownson\textsuperscript{a,∗}

\textsuperscript{a}Department of Energy and Mineral Engineering, College of Earth and Mineral Sciences, Pennsylvania State University, PA 16802, USA
\textsuperscript{b}Department of Meteorology, College of Earth and Mineral Sciences, Pennsylvania State University, PA 16802, USA

Abstract

Direct Normal Irradiance (DNI) is a critical component of solar irradiation for estimating Plane of Array (POA) irradiance on flat plate systems and for estimating photovoltaic and concentrating system power output. Current approaches to measuring or estimating DNI suffer from either high equipment costs or low precision and may require detailed environmental data. An alternative approach, using artificial neural networks to estimate DNI from the irradiance measurements of multiple pyranometers, is studied. We consider various neural network topologies and study the resulting errors. The neural network-based estimators are found to have higher accuracy than those obtained from empirical correlations of GHI measurements alone. Additionally, the use of a different GHI sensor than the one used to obtain the neural network training data does not induce significant errors. The ability of this method to be used as a quality-control instrument for pyrheliometer measurements is also discussed.

Keywords: direct normal irradiance, multi-pyranometer arrays, neural networks, machine learning

1. Introduction

Direct Normal Irradiance (DNI) is a key variable in applications where the concentration of solar energy is important, such as for solar thermal energy systems, photovoltaic systems, or concentrating systems. There are several methods for measuring or calculating DNI. The most direct (and accurate) is to measure DNI using a system such as a pyrheliometer or a rotating shadow-band pyranometer. However, these systems have a high investment cost and require regular monitoring to ensure accuracy. When these requirements are...
impractical, DNI can be estimated using Global Horizontal Irradiance (GHI) measurements as a part of one of many models. Gueymard (2010) surveyed eighteen clear-sky models. After comparing the output of the models to measured data, it was observed that models which are more physically based (and, therefore, require more atmospheric inputs) perform better and more consistently than simpler models. However, DNI model error statistics are generally optimistic, as the analysis often is restricted to the clear-sky conditions for which most of the models are developed. Additionally, it may be impractical to obtain enough atmospheric inputs for the more physical models to fully capture the anisotropy of the sky. In particular, the primary factors influencing DNI when the sun is not obscured are aerosols and precipitable water, both of which can vary significantly and are difficult to predict (Gueymard, 2010).

Another approach has been to use an array of multiple co-located pyranometers (a multi-pyranometer array, or MPA) to measure irradiance at varying angles to separate it into components. The co-located pyranometer measurements are typically used to calculate DNI via a radiation model. An artificial horizon (a barrier that prevents the pyranometers from detecting nearby ground-reflected radiation) can be added to the MPA to simplify the inverted radiation model. (Faiman et al., 1987, Munger, 1997). One potential limitation of this method is its reliance on a clear-sky model, which may reduce its effectiveness in everyday use. Additionally, even these simplified radiation models can be a computational challenge to invert, leading to the need for relatively complex postprocessing schemes.

To overcome this limitation, rather than invert a radiation model, we utilize Artificial Neural Networks (ANN), a tool from machine learning (see Subsection 1.1 for a brief overview of neural networks). We sample data from five pyranometers for input into a trained neural network, which outputs an estimated DNI measurement. Neural networks are a useful tool for this type of problem, as while it might be virtually impossible (without extensive and expensive sensor measurements) to obtain enough atmospheric data to get a full model of the anisotropic profile of the sky at any given point in time, through training on measured data a neural network can implicitly take these variables and their effects into account. Neural networks (a “black box” model) can fit almost any function to within a prescribed error bound, assuming the availability of input data that jointly captures all of the influential latent variables, and as long as there are no practical limits on the complexity of the network (Reed and Marks II, 1999). Curtiss (1993) utilized a neural network approach to DNI estimation as a “nonconventional” beam irradiance estimator, while more recently ANNs have been used to estimate GHI (Mohandes et al., 1998, Zervas et al., 2008, Paoli et al., 2010, Wang et al., 2012, al Shamisi et al., 2013), estimate DNI from sky imager data (Eissa et al., 2013, Chu et al., 2013), estimate DNI from the National Weather Service database (Marquez and Coimbra, 2011), and construct synthetic irradiance time series (Hontoria et al., 2001). Additionally, while the training phase of neural network construction can be computationally intensive, the use of a trained network is trivial for any modern computer, making the entire setup extremely portable, while also being an order of mag-
Figure 1: Overview of the sigma function implemented within a single neural network node. The inputs $x_i$ are linearly combined before the activation function is applied to the result.

ANNs are a machine learning method that generalizes logistic regression by passing $n$ input variables $\mathbf{x} = [x_1, \ldots, x_n]$ through one or more hidden layers, consisting of nodes which implement an artificial neuron model, before combining the hidden layer outputs into a final network output $y$.

Each hidden node takes inputs $\bar{x} = [\bar{x}_1, \ldots, \bar{x}_m]$ (the bar-notation is used to distinguish these hidden layer inputs from the overall network inputs). As shown in Figure 1, the neuron model implemented within each node is a combination of a sigma function, which produces a weighted sum of inputs $\sum_{i=1}^{m}(w_i x_i) + b_i$, or in vector notation, $\mathbf{w} \cdot \mathbf{x} + \mathbf{b}$, and an activation function $f$, which generates the node’s output. In the special case of linear regression, the activation function is the identity $f(z) = z$, but in a more general nonlinear network, $f$ is usually taken to be a sigmoid function, e.g. $f \equiv \tanh$, which is the activation function used in this work.

Figure 2 is a graphical representation of a one-hidden layer network: the network inputs are fed into each hidden neuron, the outputs of which are linearly combined in the output node to produce the overall network output. If the network has multiple hidden layers, the outputs from the nodes in the first layer become the inputs for each node in the second layer, the outputs of which become the inputs in the third layer, and so forth until the outputs in the final hidden layer are input into the output node and combined into the network output.
Figure 2: Overview of a one-hidden layer artificial neural network. Inputs are passed into hidden layer nodes, which process them for the output layer, which produces the overall network output.

In principle, a sufficiently complicated (i.e., with a large number of nodes or hidden layers) neural network can fit any convex, bounded function for either regression or classification to within a prescribed error (Funahashi, 1989, Hornik et al., 1989). The problem is to find the weight and bias vectors $w$ and $b$ for each node. This is usually done using a process called supervised learning, in which the weights and biases are initialized and a training data set is used to ascertain error which is then backpropagated through the network, updating the weights and biases using this error until some stopping point is reached.

Ideally, the training process halts when the error metric, usually the sum of squares, $E = \sum_{i=1}^{n} (t_i - y_i)^2$, where $t_i$ is the targeted output and $y_i$ is the network output, is minimized on some separate validation data set, though the stopping point can also be when a local minimum is reached or after a specified number of training iterations. This technique of stopping training prior to the network’s outputs perfectly matching the given data is called early stopping. It attempts to stop the network from overfitting the training data at the expense of generalization. This is often achieved through cross-validation, where a portion of the training data is withheld and used to test the accuracy of the network. Another approach is called regularization, which adds a penalty to the error, $\tilde{E} = E + \nu \Omega$, where $\Omega$ is a function that increases along with the magnitude of the weights. This penalizes large weights, removing a source of possible overfitting. We used regularization in this study to ensure that the training algorithm was exposed to the entirety of the data. We trained our networks for a maximum of 1,000 epochs — training was halted earlier if the error gradient had become very small, which indicates a network with sufficiently small error that further convergence will be very slow.

For more information on neural networks, the reader can consult any number of sources (e.g. Reed and Marks II, 1999, Bishop, 1995). One downside to this method is that while neural networks are an extremely effective computational tool, their weights aren’t informative; one cannot look at a neural network model and immediately learn something about the underlying dynamics of the
modeled system. The resulting lack of transparency seems to be a worthwhile tradeoff in situations where a more transparent model is insufficiently accurate or impractical for everyday use.

2. Methods

2.1. Error Metrics

To evaluate the accuracy of each model, we use the following statistics: mean absolute error (MAE), and root mean square error (RMSE), mean bias error (MBE). MAE, given by

\[ \text{MAE} = \frac{\sum_{i=1}^{n} |\text{model}_i - \text{measured}_i|}{n} \]

gives an estimation of the total error, removing the possibility of cancellations. RMSE, or

\[ \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\text{model}_i - \text{measured}_i)^2}{n}}, \]

emphasizes the magnitude of errors, and looking at the difference between RMSE and MAE can approximate the impact of the most extreme errors on the overall accuracy of the estimator. All of these measures can be converted into percentages by dividing each summand by the measured value; however, as many of our data points have a very small measured DNI, we instead divide the final error calculation by the mean of the measured DNI values. MBE, given by

\[ \text{MBE} = \frac{\sum_{i=1}^{n} \text{model}_i - \text{measured}_i}{n} \]

is useful when we are concerned with the aggregation of errors over time (as positive and negative errors may cancel).

2.2. Metrics on Frequency Distributions

The error statistics described in Section 2.1 measure the accuracy and precision of an estimator, but they do not address whether the overall predicted distribution will have the same statistics as the measured distribution. This is an important consideration when designing a solar energy conversion system installation (from site selection to constructing bankable data) and systems-level decision-making (Hoyer-Klick et al., 2010). When comparing distributions, metrics proposed by Espinar et al. (2009) and endorsed by IEA-SHP Task 36 (Hoyer-Klick et al., 2010) are useful. These are based on the Kolmogorov-Smirnov (K-S) tests for equality of distributions. While the K-S tests (one- and two-sample) are sensitive to differences in the shapes of the CDFs (and hence the underlying probability density functions), they measure only the largest deviation between
the two. The extended Kolmogorov-Smirnov Integral test extends this by dividing
the range of \( x \) into \( m \) intervals (Espinar et al., 2009). The distance between
the two CDFs is then defined on each interval \( I_n \) by

\[
D_n = \max_{x \in I_n} |F(x) - S(x)|.
\]

Finally, the Kolmogorov-Smirnov Integral (KSI) distance is obtained as the
integral

\[
D = \int_{x_{\text{min}}}^{x_{\text{max}}} D_n \, dx.
\]

In (Espinar et al., 2009), \( m \) is taken to be 100. We view this setup as a
Riemann sum and take the limit as \( m \to \infty \), yielding what we will refer to in
this work as the KSI distance

\[
\text{KSI} = \int_{x_{\text{min}}}^{x_{\text{max}}} |F(x) - S(x)| \, dx,
\]

yielding a measure of the total deviation between the two CDFs rather than an
approximation.

The second distance, called the OVER metric, is calculated by only consid-
ering deviations that are outside the K-S 99% confidence threshold. We once
again view the Espinar et al. (2009) formulation as a Riemann sum and let
\( m \to \infty \), obtaining

\[
\text{OVER} = \int_{x_{\text{min}}}^{x_{\text{max}}} \max\{0, |F(x) - S(x)| - V_c\} \, dx,
\]

where

\[
V_c = 1.63\sqrt{2/n}
\]

is the critical 99% \( p \)-value for the two-sample K-S test with equal sample pop-
ulations (Smirnov, 1939, Zwillinger and Kokoska, 1999).

Both the KSI and OVER metrics can be converted into relative percentages
(relative to the metric value \( a_c \) of constant deviation by the critical value \( V_c \)).
It is clear from (1) that for large sample sizes, the K-S test will reject the null
hypothesis even for very small deviations. This makes sense for large samples
when the sampled distribution is compared to a known underlying distribution.
However, the analogous result for the KSI and OVER metrics is that the relative
value is large even for very similar distributions which may not result in any
practical difference. As the data in this study is from a single site and thus
is over a similar range of measurements, we omit any relative KSI and OVER
percentages.

This is also a reason why the OVER metric may not be suitable on its own
when working with large samples. However, the two metrics are useful to view in
combination; the KSI distance measures the “total” deviation of the estimated
distribution from the measured distribution, and the difference from the OVER
distance estimates how much of this distance is statistically significant (in the sense of the null hypothesis of the K-S test).

2.3. Sensor Configuration and Data Source Considerations

One goal of this research is to establish the smallest number of pyranometers that can be used to assess DNI with these techniques to within a desired accuracy level. Witmer (2014) showed unique relationships between irradiance measurements in each cardinal direction and GHI over time, indicating the need to represent each direction in the data. As a result, the primary configuration considered for this work was a set of 5 pyranometers: one facing each cardinal direction (at a 90° angle to the horizontal) and one to capture GHI. However, a 3-pyranometer design (one south-facing pyranometer mounted at 40° and east and west-facing pyranometers mounted at 90°) was also tested. No artificial horizon is used, as the network should be robust enough to account for ground reflectance implicitly as this impact would be represented in the directional irradiance measurements.

The training data for the networks in this study was 1-minute interval data obtained from the Solar Radiation Research Laboratory (SRRL) in Golden, CO, from 1/1/2009 through 31/12/2013. The site has a latitude of 37.74° N, longitude of 105.18° W, and an elevation of 1829 m. SRRL has both thermopile (Eppley PSP — ventilated and non-ventilated) and photodiode (LICOR LI-200) pyranometers arranged in the configurations described in Section 2.3. While photodiode pyranometers such as the LI-200 are not as accurate as thermopile pyranometers due to their limited radiative response (Gueymard and Myers, 2007, Vignola et al., 2012), in principle the added errors in measurement should
not important for this application as the ANN learning process only requires sensitivity of the measurements to the same latent variables that influence DNI (this is examined in Section 3.2). As a result, the reduced cost of the LI-200 pyranometers was viewed as a major benefit to the use of these instruments and thus this configuration was treated as the primary ASE configuration.

Data points were removed from the set when they were clearly erroneous (such as a measured DNI of over 10,000 Wm$^{-2}$ or strongly negative pyranometer measurements) or when the SERI QC quality-control flagging system NREL (1993) used by SRRL suggested that the GHI measurement was unreliable (corresponding to a flag greater than or equal to 10). Otherwise all data were included.

The described methods are resilient to real-sky conditions and provide an additional feature of the ASE as a diagnostic tool. As such, SERI QC flags corresponding to the pyrheliometer measurements were intentionally ignored in the training period; they were then used for the subsequent analysis that followed to assess the ASE’s utility as a diagnostic tool (see Section 2.4). In particular, this led to the inclusion of data featuring “reliable” GHI measurements and “unreliable” DNI measurements in the training data.

The mean of the remaining DNI measurements for the all-LICOR data set is 232.76 Wm$^{-2}$ across 2,079,326 observations. For the data set with a non-ventilated Eppley PSP GHI measurement, the mean DNI is 245.09 Wm$^{-2}$ across 2,293,279 observations, and with a ventilated Eppley PSP GHI measurement, the mean DNI is 324.34 Wm$^{-2}$ over 1,537,381 observations.

One of the requirements of ANN use is a large and representative data set for training so that the training algorithm is exposed to the patterns it can be expected to encounter. To evaluate how much data is required for DNI estimation, we trained multiple networks on data sets ranging from one to four years of data before testing them on the following year and measuring the root mean squared error. This was done over all possible years from 2007 through 2013, so e.g. the 1-year error is the average error of a network trained on 2007 data and tested on 2008 data, a network trained on 2008 data and tested on 2009 data, etc. The RMSEs of these tests were 70.9 Wm$^{-2}$ for the 1-year test error, 69.4 Wm$^{-2}$ for the 2-year test error, 60.4 Wm$^{-2}$ for the 3-year test error, and 60.1 Wm$^{-2}$ for the 4-year test error.

The RMSE for one year of training is inflated by the network trained on 2008 data and tested on 2009 data, as the RMSE for that test was 103.11 Wm$^{-2}$, which was dramatically higher than any other test; excluding that gives an average RMSE of 64 Wm$^{-2}$. Given the difficulty in obtaining such large dataset for other sites, the slight reduction in average error for extended training periods can be seen as inconsequential, leading to the conclusion that a year of hourly data is sufficient for training the ASE neural network. For the rest of this study, we use 2009 data to train our networks and test on 2010–2013 data.

The variation in the data over the four testing years is illustrated with Figure 3. It can also be seen from Figure 3 that the training data followed a significantly different pattern than the testing data, with the training mean larger than any of the individual testing means in some months and smaller than any in others.
This is important for the purposes of model verification — if the testing data were similar to the training data, an overfit network would produce excellent testing statistics but would not generalize to patterns in different data. The mean of the average daily measurements will be used later as a reference for the ANN model error.

2.4. Diagnostics

One potential application of the ASE is as a relatively low-cost redundant system to help check for errors in other systems that directly measure DNI or rely on DNI measurements. As the data from SRRL is flagged for reliability using the SERI QC system, it is possible to determine whether large deviations between pyrheliometer measurements and ASE estimates tend to indicate unreliability of the DNI measurement. This could allow for a quick method for quality control of pyrheliometer or rotating shadowband radiometer data using an ASE as a redundant instrument, which might be appealing in applications where routine maintenance of those instruments' moving parts is not practical but their lower levels of uncertainty is preferable.

This will be examined in Section 3.4, once the best ASE design and network are identified, by assessing what percentage of each SERI QC flag category is “captured” by the set of points where the ASE estimate has a large deviation from the pyrheliometer measurement; the ideal outcome would be if none of the data identified as reliable would be represented in such a set but it contained all of the unreliable data.

3. Results

To determine the appropriate number of hidden nodes and layers, ANNs with different topologies were trained and tested. To minimize the risk of network convergence to a local minimum (as backpropagation convergence can be highly sensitive to the initial seed weights), ten networks were trained for each topology and their outputs averaged. A neural network with two hidden layers can add and subtract (with weights) convex regions of the input space, whereas a single layer can do so only for hyperplanes (Reed and Marks II, 1999). The inclusion of a network with a second hidden layer allows us to test if the underlying geometry of the problem lends itself to being solved with a multiply-layer network topology (as opposed to using a large number of neurons in the first layer).

The networks tested had the following topologies, with each topology used to train a network with both observed and simulated data:

1. NN1: one hidden layer, 1 node;
2. NN2: one hidden layer, 5 nodes;
3. NN3: one hidden layer, 10 nodes;
4. NN4: two hidden layers, 5 nodes and 3 nodes, respectively.

If the final network output (post-averaging) was negative for a particular set of inputs, the output was set to zero.
Table 1: Error statistics for tested ANNs with the 5-pyranometer implementation for 1-minute interval data over the quality-controlled data set. Networks are sorted by MAE and RMSE. Absolute error values are in Wm\(^{-2}\). Percentages are relative to the mean measured DNI.

(a) Error statistics across the entire data set. Percentages are relative to the mean measured DNI of 235.1 Wm\(^{-2}\).

(b) Error statistics when low-sun conditions are excluded. Percentages are relative to the mean measured DNI of 508.9 Wm\(^{-2}\).

<table>
<thead>
<tr>
<th>ANN Model</th>
<th>MAE</th>
<th>RMSE</th>
<th>MBE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN3</td>
<td>16.1</td>
<td>32.0</td>
<td>0.3</td>
</tr>
<tr>
<td>NN4</td>
<td>16.4</td>
<td>32.4</td>
<td>-0.5</td>
</tr>
<tr>
<td>NN2</td>
<td>17.7</td>
<td>34.5</td>
<td>-0.6</td>
</tr>
<tr>
<td>NN1</td>
<td>47.7</td>
<td>95.1</td>
<td>-2.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANN Model</th>
<th>MAE</th>
<th>RMSE</th>
<th>MBE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN3</td>
<td>31.8</td>
<td>46.6</td>
<td>-0.3</td>
</tr>
<tr>
<td>NN4</td>
<td>31.8</td>
<td>46.9</td>
<td>-2.1</td>
</tr>
<tr>
<td>NN2</td>
<td>34.4</td>
<td>50.0</td>
<td>-2.4</td>
</tr>
<tr>
<td>NN1</td>
<td>95.6</td>
<td>138</td>
<td>-5.0</td>
</tr>
</tbody>
</table>

For the non-diagnostic sections of this chapter, we worked only with quality-controlled data. This amounted to disregarding data where the DNI measurement had a SERI QC flag greater than 9. This data was included in the analysis for Section 3.4.

3.1. 5-Pyranometer Networks

All four ANN topologies were tested on ASE configurations with an LI-200 GHI sensor (see Section 2.3). Tables 1a and 1b summarize the error statistics computed as in Section 2.1 for each network model. In general, for ANNs it is expected that overall error decreases as network complexity increases. NN1, with only a single hidden node, is too simple a model to separate the direct component from the irradiance measurements with any reliability, which is reflected in its MBE and RMSE measurements. Despite NN1’s simplicity, its MBE is only slightly higher than that of the other models in two of the four testing years. This illustrates the weakness of using MBE as the sole metric of error, as a constant value (if fit well to the mean of the data) would have an MBE close to zero, and any overall bias can be removed by postprocessing. As a result, MAE and RMSE are more illustrative of actual error in operation, and with respect to those metrics NN1 is significantly worse than the other models.

The MBE and RMSE results in Table 1b can be contrasted with those found in (Gueymard, 2010, Table 5). A wide variety of empirical models were tested under similar conditions at SRRL from 2006–2008, separating global irradiance into direct and diffuse components using the diffuse fraction (though some of the tested models include other covariates). Under all weather conditions, the RMSEs of the tested models ranged from 20.2% to 29.2%. When restricted to high albedo conditions (albedo > 0.5), the RMSEs ranged from 27.5% to 47.3%.
Table 2: Error statistics on quality-controlled data under no-snow (albedo satisfies $0 < ! < 0.5$) and snowy conditions (albedo satisfies $0.5 < ! < 1$) when $Z \leq 85^\circ$. Networks are sorted by MAE and RMSE. Absolute errors are in Wm$^{-2}$. Percentages are relative to the mean measured DNI of 522.32 Wm$^{-2}$ for no-snow conditions and 342.2 Wm$^{-2}$ for snowy conditions.

<table>
<thead>
<tr>
<th>ANN Model</th>
<th>No-Snow</th>
<th></th>
<th></th>
<th>Snow</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAE</td>
<td>RMSE</td>
<td>MBE</td>
<td>MAE</td>
<td>RMSE</td>
<td>MBE</td>
</tr>
<tr>
<td>NN3</td>
<td>6.1%</td>
<td>8.9%</td>
<td>0.0%</td>
<td>8.4%</td>
<td>14.4%</td>
<td>-2.5%</td>
</tr>
<tr>
<td>NN4</td>
<td>6.1%</td>
<td>8.9%</td>
<td>-0.4%</td>
<td>9.3%</td>
<td>15.4%</td>
<td>-1.8%</td>
</tr>
<tr>
<td>NN2</td>
<td>6.6%</td>
<td>9.5%</td>
<td>-0.4%</td>
<td>9.8%</td>
<td>16.1%</td>
<td>-3.2%</td>
</tr>
<tr>
<td>NN1</td>
<td>18.4%</td>
<td>26.4%</td>
<td>-1.9%</td>
<td>27.0%</td>
<td>41.3%</td>
<td>16.7%</td>
</tr>
</tbody>
</table>

Figure 4: Frequency distributions for training and testing measured DNI and 5-pyranometer ANN model (LI-200 GHI) outputs when low-sun conditions are excluded.

which may be compared to the results in Table 2. In comparison to this study, only NN1 yields errors in this range (albeit at the high end), suggesting that NN1 has component-separating skill similar to that of the studied empirical models.

Interestingly, despite the two-hidden layer network topology, NN4 fails to outperform NN3 in any setting. This could be a result of insufficient training time (around half of the networks with the NN4 topology did not complete training in the specified 2000 epochs), though it suggests that at a single site, a single hidden layer network has sufficient skill at irradiance component separation. Even if further gains were achieved by extended training, it is unlikely that an information criterion aimed at minimizing both error and network complexity (such as the Akaike information criterion) would endorse NN4.

The kernel density functions for the outputs of each model, as well as the
<table>
<thead>
<tr>
<th>ANN Model</th>
<th>Metric</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN3</td>
<td>KSI</td>
<td>9.6</td>
<td>3.0</td>
<td>7.3</td>
<td>3.5</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>OVER</td>
<td>6.5</td>
<td>1.1</td>
<td>3.3</td>
<td>1.5</td>
<td>1.7</td>
</tr>
<tr>
<td>NN4</td>
<td>KSI</td>
<td>10.4</td>
<td>3.5</td>
<td>8.3</td>
<td>6.2</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>OVER</td>
<td>7.1</td>
<td>0.9</td>
<td>4.2</td>
<td>3.4</td>
<td>3.0</td>
</tr>
<tr>
<td>NN2</td>
<td>KSI</td>
<td>10.9</td>
<td>3.4</td>
<td>8.1</td>
<td>7.6</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>OVER</td>
<td>7.3</td>
<td>0.8</td>
<td>3.9</td>
<td>4.5</td>
<td>2.9</td>
</tr>
<tr>
<td>NN1</td>
<td>KSI</td>
<td>32.4</td>
<td>34.5</td>
<td>32.6</td>
<td>35.9</td>
<td>32.6</td>
</tr>
<tr>
<td></td>
<td>OVER</td>
<td>27.3</td>
<td>29.0</td>
<td>27.4</td>
<td>30.4</td>
<td>29.9</td>
</tr>
</tbody>
</table>

Table 3: KSI and OVER metrics for 5-pyranometer ASE models (with LI-200 GHI) when low-sun conditions are excluded, both overall and by year. Metrics have units Wm$^{-2}$. Networks are sorted by performance across both metrics.

<table>
<thead>
<tr>
<th>ASE Model</th>
<th>MAE</th>
<th>RMSE</th>
<th>MBE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN3</td>
<td>4.0%</td>
<td>5.3%</td>
<td>-0.1%</td>
</tr>
<tr>
<td>NN4</td>
<td>4.1%</td>
<td>5.4%</td>
<td>-0.4%</td>
</tr>
<tr>
<td>NN2</td>
<td>4.4%</td>
<td>5.8%</td>
<td>-0.5%</td>
</tr>
<tr>
<td>NN1</td>
<td>11.9%</td>
<td>15.3%</td>
<td>-1.0%</td>
</tr>
</tbody>
</table>

Table 4: Average error of each 5-input (LI-200 GHI) model’s predicted daily irradiance over the quality-controlled data set when low-sun conditions are excluded. Networks are sorted by MAE and RMSE.

measured data, are shown in Figure 4. The impact of the choice of training year can be directly observed, as it serves a role similar to a Bayesian prior distribution for the more capable models (NN2, NN3, and NN4). All four estimators do not fully capture the probability mass at the high measured mode around 975 Wm$^{-2}$. This is consistent with Figure 3, which shows that the year used for training the ANNs had lower average daily DNI during high-irradiance months. NN4 appears to be more capable of deviation from the training pattern, though as Table 3 indicates, the NN4 estimated distribution deviates from the measured distribution further than the NN3 estimated distribution. This could indicate that a two-layer network would be appropriate if a trained ANN were intended for use at sites other than the one at which the training data was collected.

Each model’s prediction for total daily irradiance was calculated for each day in the test set (low-sun conditions were excluded as these are less important for solar energy conversion). Table 4 summarizes the average percentage error of these predictions. While the mean bias of the models remains the same as in the 1-minute data (see Table 1b), many of the residuals, particularly the large ones that are responsible for the difference between the first-order error MAE and the second-order error RMSE, average away over the course of a day. The daily uncertainties (interpreting standard error as a crude estimate of instrument uncertainty) are 5.3% for NN3, 5.4% for NN4, 5.7% for NN2, and 15.3% for
Another consideration is the accuracy of the models when their performance is broken out by month rather than across the entire year. This can give us more information about how the models respond to different climate regimes. Figure 5 shows the error from the average daily measured DNI per month across the four year testing period for the entire dataset.

### 3.2. Impact of Changing GHI Sensor

In light of the results of Section 3.1, an ensemble of networks with the NN3 topology were trained on data sets using measurements made with ventilated and non-ventilated Eppley PSP GHI measurements. Due to quality-controlling the data sets prior to training on the basis of the SERI QC flags assigned to the GHI measurement, the results are not comparable in relation to any common baseline (when restricted to zenith angle $Z \leq 85$ and quality-controlled, the LI-200 GHI test data set consists of 750,042 points, the non-ventilated PSP data set consists of 790,433 points, and the ventilated PSP data set consists of 782,727 points). Instead, we compare the results to the remaining DNI measurements corresponding to the quality-controlled GHI measurements. These error statistics are shown in Table 5.

All three implementations offer similar levels of performance, both absolutely and relative to the measured DNI. Table 3 shows the frequency distribution metrics in each testing year and across the entire data set. There is relatively little consistency from year to year in the KSI and OVER metrics, though in aggregate the LI-200 and ventilated PSP networks perform similarly, with the
Table 5: Overall error statistics for 5-pyranometer NN3 implementations with varying GHI sensors on quality-controlled data when low-sun conditions are excluded. Networks are sorted by MAE and RMSE. Absolute error values are in Wm\(^{-2}\). Percentages are relative to mean measured DNI values of 508.9 Wm\(^{-2}\) for the LI-200 GHI data set, 553.6 for the non-ventilated PSP GHI data set, and 492.8 for the ventilated PSP GHI data set.

<table>
<thead>
<tr>
<th>GHI Sensor</th>
<th>MAE</th>
<th>RMSE</th>
<th>MBE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LI-200</td>
<td>31.8</td>
<td>46.6</td>
<td>-0.3</td>
</tr>
<tr>
<td>PSP (ventilated)</td>
<td>32.2</td>
<td>47.6</td>
<td>1.8</td>
</tr>
<tr>
<td>PSP (non-ventilated)</td>
<td>34.7</td>
<td>49.9</td>
<td>-2.6</td>
</tr>
</tbody>
</table>

Table 6: KSI and OVER metrics for 5-pyranometer ASE models (with LI-200 GHI) when low-sun conditions are excluded, both overall and by year. Metrics have units Wm\(^{-2}\). Networks are sorted by performance across both metrics.

<table>
<thead>
<tr>
<th>GHI Sensor</th>
<th>Metric</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>LI-200</td>
<td>KSI</td>
<td>9.6</td>
<td>3.0</td>
<td>7.3</td>
<td>3.5</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>OVER</td>
<td>6.5</td>
<td>1.1</td>
<td>3.3</td>
<td>1.5</td>
<td>1.7</td>
</tr>
<tr>
<td>PSP (ventilated)</td>
<td>KSI</td>
<td>9.8</td>
<td>6.6</td>
<td>10.1</td>
<td>4.6</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>OVER</td>
<td>6.6</td>
<td>3.4</td>
<td>6.0</td>
<td>2.0</td>
<td>1.5</td>
</tr>
<tr>
<td>PSP (non-ventilated)</td>
<td>KSI</td>
<td>11.0</td>
<td>5.8</td>
<td>5.3</td>
<td>6.0</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>OVER</td>
<td>8.3</td>
<td>3.3</td>
<td>1.9</td>
<td>3.3</td>
<td>3.9</td>
</tr>
</tbody>
</table>

estimated distribution from the non-ventilated PSP network deviating further from the measured DNI distribution. Table 7 shows that this trend continues when looking at the average daily predicted irradiance; all of the models perform similarly.

Table 8 shows the impact of using a different type of GHI sensor for operational use than that used for collecting the training data. For Table 8a, the networks were trained on data collected with a thermopile-type pyranometer, which are preferred to photodiode-based pyranometers as they have a wider spectral responsivity. Both ventilated and non-ventilated measurements were tested (the ventilator increases power requirements and the footprint of the sensor but helps to keep the dome of the sensor free of dust and snow, decreasing maintenance needs). Table 8 shows the impact of using a photodiode-type pyra-

Table 7: Average error of each 5-input NN3 implementation’s predicted daily irradiance over the quality-controlled data set when low-sun conditions are excluded. Networks are sorted by MAE and RMSE.

<table>
<thead>
<tr>
<th>GHI Sensor</th>
<th>MAE</th>
<th>RMSE</th>
<th>MBE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LI-200</td>
<td>4.0%</td>
<td>5.3%</td>
<td>-0.1%</td>
</tr>
<tr>
<td>PSP (non-ventilated)</td>
<td>4.1%</td>
<td>5.4%</td>
<td>-0.5%</td>
</tr>
<tr>
<td>PSP (ventilated)</td>
<td>4.1%</td>
<td>5.5%</td>
<td>0.4%</td>
</tr>
<tr>
<td>Training GHI Sensor</td>
<td>MAE</td>
<td>RMSE</td>
<td>MBE</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>PSP (ventilated)</td>
<td>0.6</td>
<td>0.8</td>
<td>1.9</td>
</tr>
<tr>
<td>PSP (non-ventilated)</td>
<td>1.5</td>
<td>1.3</td>
<td>-1.7</td>
</tr>
</tbody>
</table>

(a) Change in error statistics (from Table 1b) on data collected using an LI-200 GHI sensor when using NN3 implementations trained on ventilated and non-ventilated PSP GHI sensor data.

<table>
<thead>
<tr>
<th>Data GHI Sensor</th>
<th>MAE</th>
<th>RMSE</th>
<th>MBE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSP (ventilated)</td>
<td>0.1</td>
<td>0.1</td>
<td>-2.0</td>
</tr>
<tr>
<td>PSP (non-ventilated)</td>
<td>-1.0</td>
<td>-0.5</td>
<td>1.7</td>
</tr>
</tbody>
</table>

(b) Change in error statistics (from Table 5) on data collected using a ventilated and non-ventilated PSP GHI sensor when using an NN3 implementation trained on LI-200 GHI sensor data.

Table 8: Change in error statistics for 5-pyranometer NN3 implementations when varying the sensor used to measure GHI from the GHI sensor which collected the training data. Data is restricted to quality-controlled data and low-sun conditions are excluded. Absolute error values are in Wm\(^{-2}\). Percentages are relative to mean measured DNI values of 508.9 Wm\(^{-2}\) for the LI-200 GHI data set, 553.6 for the non-ventilated PSP GHI data set, and 492.8 for the ventilated PSP GHI data set.

3.3. 3-Pyranometer Networks

Comparing Tables 1a and 9a, it can be seen that there is a general loss in accuracy with the 3-pyranometer sensor configuration. In every measure (see also Table 10), the 3-pyranometer ASEs have less skill at separating irradiance into components than the analogous 5-pyranometer implementation, indicating that there is added value in the information provided by each of the five pyranometers (a result which parallels the results obtained by Witmer (2014)). While there is increased error, the 3-pyranometer models still perform well relative to the empirical models surveyed by Gueymard (2010).

NN4 marginally outperforms NN3 when low-sun angles are excluded; however, as NN4 is a significantly more complex model than NN3, it would not be recommended for such a small increase in performance. This is particularly true in light of the average daily irradiance prediction errors summarized in Table 11 and visualized in Figure 6. What improvement NN4 has disappears over the course of a day, and even when considered by month, the average daily predictions are effectively identical.
ANN Model | MAE | RMSE | MBE  
---|---|---|---
NN3 | 18.2 | 7.7% | 37.8 | 16.1% | -3.3 | -1.39%  
NN4 | 18.4 | 7.8% | 37.7 | 16.0% | -3.3 | -1.42%  
NN2 | 19.7 | 8.4% | 39.9 | 17.0% | -3.3 | -1.39%  
NN1 | 51.3 | 21.8% | 101 | 43.0% | -6.8 | -2.9%  

(a) Error statistics across the entire data set. Percentages are relative to the mean measured DNI of 235.1 Wm$^{-2}$.

ANN Model | MAE | RMSE | MBE  
---|---|---|---
NN4 | 36.2 | 7.1% | 54.9 | 10.8% | -8.1 | -1.6%  
NN3 | 36.3 | 7.1% | 55.2 | 10.8% | -8.0 | -1.6%  
NN2 | 38.6 | 7.6% | 58.0 | 11.4% | -7.8 | -1.5%  
NN1 | 102 | 20.1% | 145 | 28.5% | -11.7 | -2.3%  

(b) Error statistics when low-sun conditions are excluded. Percentages are relative to the mean measured DNI of 508.9 Wm$^{-2}$.

Table 9: Error statistics for tested ANNs with the 3-pyranometer implementation for 1-minute interval data over the quality-controlled data set. Networks are sorted by MAE and RMSE. Absolute error values are in Wm$^{-2}$. Percentages are relative to the mean measured DNI.

ANN Model | Metric | 2010 | 2011 | 2012 | 2013 | Overall  
---|---|---|---|---|---|---
NN3 | KSI | 9.5 | 12.4 | 11.4 | 13.5 | 10.7  
     | OVER | 6.6 | 9.2 | 7.3 | 9.9 | 8.8  
NN4 | KSI | 10.2 | 12.5 | 12.2 | 14.1 | 11.6  
     | OVER | 0 | 8.5 | 0 | 9.5 | 9.3  
NN2 | KSI | 11.0 | 12.2 | 12.6 | 14.7 | 12.1  
     | OVER | 7.3 | 9.4 | 8.0 | 11.5 | 9.6  
NN1 | KSI | 35.3 | 44.2 | 34.1 | 42.2 | 36.9  
     | OVER | 30.3 | 38.5 | 28.9 | 36.9 | 34.1  

Table 10: KSI and OVER metrics for 5-pyranometer ASE models when low-sun conditions are excluded, both overall and by year. Metrics have units Wm$^{-2}$. Networks are sorted by performance across both metrics.

ASE Model | MAE | RMSE | MBE  
---|---|---|---
NN4 | 4.9% | 7.0% | -1.6%  
NN3 | 4.9% | 7.1% | -1.6%  
NN2 | 5.4% | 7.6% | -1.6%  
NN1 | 13.6% | 17.3% | -2.4%  

Table 11: Average error of each 3-input model's predicted daily irradiance over the quality-controlled data set when low-sun conditions are excluded. Networks are sorted by MAE and RMSE.
3.4. Diagnostics

To evaluate the potential of the ASE as a diagnostic/quality control tool, following the results of Sections 3.1 and 3.3 we restrict our attention to the 5-pyranometer (with an LI-200 GHI sensor, as Section 3.2 indicates there is no practical difference between GHI sensor choices) implementation of NN3, which was the most successful ASE implementation. In this section, we no longer limit the data under consideration to that which is flagged as reliable by SERI QC, though we will exclude low-sun conditions.

First, we examine the relationship between SERI QC flags and the difference between the NN3 estimation and the measured DNI, which is shown in Figure 7. The goal is to determine a value $d_{\text{flag}}$ such that if $|\text{(NN3 estimate)} - \text{(measured DNI)}| > d_{\text{flag}}$, flagging the data for quality-control purposes will mark a large number of unreliable data and not many reliable data points. Either a deviation of 150 Wm$^{-2}$ or 200 Wm$^{-2}$ seem to be appropriate choices depending on how conservative we would like to be with the good data. While setting $d_{\text{flag}} = 100$Wm$^{-2}$ eliminates approximately 27% of unreliable data, it also eliminates 5% of data given a SERI QC flag of 4 and 4.1% of data flagged 4 and below overall.

4. Discussion & Conclusions

We have shown that the ASE design of an MPA (using either five or three pyranometers) and an ANN shows better skill at irradiance component separation than other empirical models. (see e.g. Gueymard, 2010, Table 5)). It seems
Figure 7: Percentage of data by SERI QC flag with errors of 100 Wm$^{-2}$, 150 Wm$^{-2}$, and 200 Wm$^{-2}$ between 5-pyranometer (LI-200 GHI) NN3 estimates and measured DNI. Low-sun conditions are excluded. Percentages are shown on a logarithmic scale. The data in Section 1 is untested (flag of zero) or reliable; in Section 2, the data is less reliable but not necessarily unreliable; in Section 3, the data is unreliable (in increasing magnitude); and in Section 4, the data is physically impossible or missing.
that a single GHI measurement, even in combination with some environmental
variables, is a low-quality data source for the purposes of component separation.
This can also be seen by the results in Section 3, which show an improvement in
irradiance component separation skill with five pyranometer measurements over
three. This indicates that there is a dimensionality requirement for the input
data to be suitable to separate the beam component from the diffuse radiation
across all sky conditions. However, we would expect diminishing returns after
a certain sensor resolution is reached. It is possible that this ideal number of
sensors is greater than five, though we are not aware of available data from such
a sensor configuration.

The method for processing the pyranometer measurements is also a key part
of the ASE design. The same MPA configuration (with different computational
methods) used in (Perez et al., 1986) had reduced error (10.3%) when DNI was
greater than 50 Wm$^{-2}$, though when DNI was below 50 Wm$^{-2}$, the RMSE
percentage was 261%. That study was also done with data collected over 45
days (from May through mid-June), indicating that conditions such as albedo
may have been the same, which was not true for our data.

There is room for future work fine-tuning the ANN training procedures uti-
lized. Here, the individual networks in each ensemble were trained to a stopping
point appropriate for an individual network, and the ensemble was formed by
simple averaging (to account for varying degrees of sensitivity to the inputs in
each individual network). A more targeted approach to aggregation might yield
superior results, and a different aggregation approach might be best served with
a different training methodology. For example, over-trained networks have been
shown to perform better in large ensembles than under-trained networks (Naf-
taly et al., 1997, Granitto et al., 2005).

It is important here to make two notes related to generalizing the results in
this thesis. The first is that at least a year of data is required to thoroughly
train the ANNs studied. However, if the year chosen is not representative of
typical conditions, the ANN might not be able to accurately estimate DNI in
a more typical year. This effect seemed to be responsible for some of the error
seen in this study, particularly around the high-irradiance mode. Using the
single-site methodology employed here, this is a potentially difficult problem
in practice, as it cannot be known a priori which year would be suitable to
use for training. Comparing satellite DNI measurements during the training
year to those from previous years could help inform a decision to extend the
training period for another year (or through a period which exposes the ANN
to the appropriate irradiance, such as a more typical summer if high irradiance
periods are missing).

The second note is that, as mentioned in the preceding paragraph, this study
has focused on training and testing data from a single location. There is no
reason to suppose that these particular networks would be capable of accurate
modeling in a location with a different microclimate. Thus, it remains to be seen
if it is possible to train a single network on a year’s data from a variety of sites
that could accurately model DNI across a broad region, or if each location would
need a more customized network. In the former case, the network would have a
more complicated topology than those studied here, and it would be advisable
to utilize a self-pruning training technique to determine the appropriate network
topology for such a complex problem. This could be achieved by adapting the
backpropagation algorithm (Williams, 1995, Wan et al., 2009) or adopting a
self-pruning mechanism along the lines of Dropout (Srivastava et al., 2014).

It remains to be seen how effective an ASE could be for ground-truthing
satellite DNI measurements to help create bankable data sets or if there are
other potential applications, such as the construction of synthetic time series or
gathering data to help understand the spatiotemporal variability of sub-hourly
DNI measurements. Given the relative paucity of DNI measurements in the
United States and many developing countries and the relative inaccuracy of
GHI to DNI models, ASEs (or similar measuring devices) could form the basis
for a “solar resource exploration platform,” providing observations to reduce the
standard error of satellite-derived solar data to surface DNI models over short
time steps (Ruiz-Arias et al., 2015). Once the maximum distance over which a
trained ANN is useful is established, based on the results of Section 3.2 it could
be possible to have “master” ASEs using a thermopile pyranometer to provide
high-quality GHI measurements and several lower-cost photodiode-based ASEs
in other locations to provide DNI estimates across a region with distributed PV
assets.

At this point, the ASE methodology seems like a high-value approach to
solar resource estimation — the instrumentation is of relatively low cost, even
if a single thermopile pyranometer is used for continued GHI collection, and the
MPA can be used as a predictor of not only DNI and DHI (which would give
it the same functionality as an RSR), but also environmental variables such as
albedo or AOD if measurements are collected for training (though this would
require a complex ANN). There are no moving parts, and the ASE provides
information that can provide better estimates for tilted collectors, as it pro-
vides 90° measurements, including ground-reflected radiation, in every cardinal
direction, along with the GHI and DNI measurements.

Acknowledgements

The authors would like to thank the National Renewable Energy Laboratory
for allowing us access to the Solar Radiation Resource Laboratory data used in
this study. We would also like to acknowledge Tyler McCandless for his feedback
and suggestions.

References

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University
Press.

Chu, Y., Pedro, H.T.C., Coimbra, C.F.M., 2013. Hybrid intra-hour dni forecasts
with sky image processing enhanced by stochastic learning. Solar Energy 98,
592–603.


