Modeling Solar Irradiance and Solar PV Power Output to Create a Resource Assessment using Linear Multiple Multivariate Regression

C. T. M. Clack, A. Alexander and A. E. MacDonald
christopher.clack@noaa.gov
Motivation and Purpose

• The create total, direct (normal and horizontal), and diffuse irradiance estimates on RUC model grid. From the irradiance estimates produce solar PV power estimates.

• Leverage satellite, model hydrometeors, and high quality surface measurements to train the technique. Compute top of atmosphere irradiance to bound the regression from above.

• Apply the technique over the CONUS domain to create an hourly data set of irradiance resource assessment (2006-2008). Working to extend this to 2014.

• Validate the methodology with observations.
Calculated Zenith Angle

\[
\cos(sza) = \sin(lat) \cdot \sin(dec) \\
\quad + \cos(lat) \cdot \cos(dec) \cdot \cos(ha)
\]

\[
dec = \varepsilon \cdot \sin[\delta + \frac{\pi}{180} \cdot (279.93 + 1.915 \cdot \sin(\delta) \\
- 0.0795 \cdot \cos(\delta) + 0.02 \cdot \sin(2\delta) - 0.00162 \cdot \cos(2\delta))]
\]

\[
ha = \pi \cdot \left(1 - \frac{hr}{12}\right) - \text{lon}
\]

Calculated top of atmosphere Irradiance

\[
DNI_0 = I_0 \cdot \left(\frac{R_{av}}{R}\right)^2
\]

\[
\left(\frac{R_{av}}{R}\right)^2 \approx 1.000110 + 0.034221 \cdot \cos(\delta) \\
+ 0.001280 \cdot \sin(\delta) + 0.000719 \cdot \cos(2\delta) \\
+ 0.000077 \cdot \sin(2\delta)
\]
Linear Multivariate Multiple Regression

\[ Y_{n \times p} = Z_{n \times (r+1)} \beta_{(r+1) \times p} + \epsilon_{n \times p}, \]
\[ E(\epsilon_{(i)}) = 0, \quad \text{Cov}(\epsilon_{(i)}, \epsilon_{(k)}) = \sigma_{i,k}I, \quad i, k = 1, 2, \ldots, p. \]

- We have \( p(=3) \) irradiance fields to calculate and \( n(=55258) \) observations of each field. The observations are taken from 10 sites (6 SURFRAD and 4 ISIS).
- The regressors (\( \beta \)) are the satellite data (3 infrared channels, a visible channel, and a water vapor channel), the RUC Assimilation Model values for water within the column (snow, ice, etc...), the temperature from the model, the calculated top of atmosphere irradiance, and the zenith angle.
- The measurements are taken from 2006 – 2008, and averaged over the top of the hour (for 12 minutes) and matched up with the model data.
- The data is quality controlled, and all night-time measurements were removed.
Linear Multivariate Multiple Regression

- Method relies on high quality ground measurements to train the regression procedure. We also use University of Oregon solar measurement sites for verification.

\[
\hat{\beta}_{(i)} = (Z'Z)^{-1}Z'Y_{(i)}
\]

Can use numerous mathematical techniques to compute the coefficients. We do not go into that here... (I used SVD).
Regression Statistics

- The regression had differing success with total, direct, and diffuse.
- We trained on 10 individual sites and computed the accuracy of the regression at those sites.
- We use one SURFRAD and one ISIS site for initial verification.

<table>
<thead>
<tr>
<th>Irradiance</th>
<th>Mean (W/m²)</th>
<th>MBE (%)</th>
<th>$R^2$ (%)</th>
<th>RMSE (%)</th>
<th>CV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GHI</td>
<td>442.00</td>
<td>-2.82</td>
<td>94.17</td>
<td>20.67</td>
<td>20.48</td>
</tr>
<tr>
<td>B</td>
<td>442.00</td>
<td>-3.33</td>
<td>92.96</td>
<td>22.63</td>
<td>22.39</td>
</tr>
<tr>
<td>C</td>
<td>442.00</td>
<td>-4.26</td>
<td>91.08</td>
<td>25.60</td>
<td>25.25</td>
</tr>
<tr>
<td>DNI</td>
<td>512.37</td>
<td>-12.41</td>
<td>77.75</td>
<td>41.82</td>
<td>39.94</td>
</tr>
<tr>
<td>B</td>
<td>512.37</td>
<td>-15.33</td>
<td>71.80</td>
<td>47.92</td>
<td>45.40</td>
</tr>
<tr>
<td>C</td>
<td>512.37</td>
<td>-22.16</td>
<td>54.29</td>
<td>57.46</td>
<td>53.01</td>
</tr>
<tr>
<td>DHI</td>
<td>148.66</td>
<td>-4.19</td>
<td>82.87</td>
<td>42.42</td>
<td>42.21</td>
</tr>
<tr>
<td>B</td>
<td>148.66</td>
<td>-4.63</td>
<td>80.83</td>
<td>44.56</td>
<td>44.32</td>
</tr>
<tr>
<td>C</td>
<td>148.66</td>
<td>-6.90</td>
<td>69.20</td>
<td>55.40</td>
<td>54.97</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Irradiance</th>
<th>Mean (W/m²)</th>
<th>MBE (%)</th>
<th>$R^2$ (%)</th>
<th>RMSE (%)</th>
<th>CV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GHI</td>
<td>458.13</td>
<td>2.41</td>
<td>89.37</td>
<td>19.57</td>
<td>19.42</td>
</tr>
<tr>
<td>B</td>
<td>458.13</td>
<td>2.67</td>
<td>88.16</td>
<td>20.67</td>
<td>20.50</td>
</tr>
<tr>
<td>C</td>
<td>458.13</td>
<td>1.08</td>
<td>83.91</td>
<td>24.03</td>
<td>24.01</td>
</tr>
<tr>
<td>DNI</td>
<td>468.03</td>
<td>2.35</td>
<td>65.91</td>
<td>39.51</td>
<td>39.44</td>
</tr>
<tr>
<td>B</td>
<td>468.03</td>
<td>0.21</td>
<td>58.98</td>
<td>43.27</td>
<td>43.27</td>
</tr>
<tr>
<td>C</td>
<td>468.03</td>
<td>-9.80</td>
<td>41.86</td>
<td>52.93</td>
<td>52.01</td>
</tr>
<tr>
<td>DHI</td>
<td>164.60</td>
<td>-9.26</td>
<td>66.26</td>
<td>40.33</td>
<td>39.25</td>
</tr>
<tr>
<td>B</td>
<td>164.60</td>
<td>-10.32</td>
<td>63.43</td>
<td>42.08</td>
<td>40.80</td>
</tr>
<tr>
<td>C</td>
<td>164.60</td>
<td>-10.60</td>
<td>48.26</td>
<td>49.92</td>
<td>48.78</td>
</tr>
</tbody>
</table>

Composite metrics for training sites

Composite metrics for initial verification sites
Regression Statistics

Mean Bias Error at Independent Validation Sites

Root-Mean-Squared Error at Independent Validation Sites

Figure 6: Time series of measured (dashed red) and estimated (solid blue) DNI for Hermiston, OR. The top panel is for the 31 days from January 2 to June 1, 2006, and the bottom panel is for the 31 days following June 1, 2006. The panels show high correlation between the estimated and the measured.
Regression Statistics

Figure 3: The difference between the estimated irradiance and the measurement versus the measured irradiance. The top panel is for GHI, the middle panel is for DNI, and the bottom panel is for DHI. The black is for regression scheme A, red is for scheme B, and blue is for C (similar to all other figures). The light green line designates the zero-line.

Caused by interference of the beam by clouds, aerosols and atmospheric disturbances in neighboring grid cells (nearby locations) that are not in the regression. The effect is over a large range of zenith angle values due to (a smaller effect of) high level clouds and then as the sun progresses through the sky the DNI is blocked by lower, and usually thicker, atmosphere in surrounding areas.
Secondly, we calculate the power falling onto the panel from the irradiance fields. Fourthly, the current and voltage are combined to calculated the panel power. There are equations that compute the derating due to the panel structure and material. The output of the panel is restricted to 115% of the nameplate capacity. After the algorithm has finished, the power for the panel is calculated. The equations from the irradiance fields are used to calculate the power falling onto the panel within the algorithm, which are based on NREL SAM, E. Boyson (2004) and NREL SAM are empirically derived. Finally, the current and voltage are combined to derived the power for the panel. There are equations that compute the derating due to the panel structure and material. The output of the panel is restricted to 115% of the nameplate capacity. After the algorithm has finished, the power for the panel is calculated.

Thirdly, we calculate the power falling onto the panel from the irradiance fields. Fourthly, the current and voltage are combined to calculated the panel power. There are equations that compute the derating due to the panel structure and material. The output of the panel is restricted to 115% of the nameplate capacity. After the algorithm has finished, the power for the panel is calculated. The equations from the irradiance fields are used to calculate the power falling onto the panel within the algorithm, which are based on NREL SAM, E. Boyson (2004) and NREL SAM are empirically derived. Finally, the current and voltage are combined to derived the power for the panel. There are equations that compute the derating due to the panel structure and material. The output of the panel is restricted to 115% of the nameplate capacity. After the algorithm has finished, the power for the panel is calculated.

East has more DHI than the West. All boundaries have been removed to display the detail of the data. The Gulf Coast has the most DNI, and the black is for the DHI.

The South contiguous USA over the three year period of 2006–2008. The South has much poorer. All boundaries have been removed to display the detail of the data.

Figure 8: Mean Biased Error (MBE) and Root-Mean-Squared Error (RMSE) for the seven independent verification sites and the initial regression.

Figure 9: The average estimated GHI in kWh/day for the contiguous USA over the three year period of 2006–2008 (the range is 1.0–1.6 kWh/day). The South has much poorer. All boundaries have been removed to display the detail of the data.

Figure 10: The average estimated DNI in kWh/day for the contiguous USA over the three year period of 2006–2008 (the range is 3.0–4.0 kWh/day). The East has the greatest resource area whereas the rest of the USA is much poorer. All boundaries have been removed to display the detail of the data.

Figure 11: The average estimated DHI in kWh/day for the contiguous USA over the three year period of 2006–2008 (the range is 1.4–1.6 kWh/day). The East has more DHI than the West. All boundaries have been removed to display the detail of the data.
Regression Created Dataset

Figure 5: Time series of measured (dashed red) and estimated (solid blue) GHI for Burns, OR. The top panel is for the 31 days from January 1, 2006, and the bottom panel is for the 31 days following June 1, 2006. The panels show high correlation between the estimated and the measured.

The present regression technique is superior in terms of MBE and RMSE. For example, at the Burns, OR site, the current technique has an MBE of -1.64% for GHI, while the SUNY dataset over the same period has an MBE of -2.00%. Similar statistical differences were found with the other irradiance species and different sites. The differences are not very large, and a review of the SUNY dataset statistics can be found in, e.g., Notrott and Kleissl (2010); Djebbar et al. (2012). More comparisons need to be done at more sites to establish if indeed the current technique is consistently more accurate.

The linear multivariate multiple regression method has provided accurate estimates of the solar irradiance over the contiguous USA. The dataset is comprised of $\approx 152,000$ geographic cells that each contain $\approx 26,000$ hourly data points. In Figs 9–11, we show the three-dimensional GHI estimated error for Burns, OR.
Irradiance to Solar Photovoltaic Power

- Take the output GHI, DNI and DHI and use them as inputs to a power modeling algorithm.
- In addition take temperature (at 2 m) and wind speed (at 10 m) from the RUC to help provide an estimate of the panel temperature.
Conclusions and Future Work

• The results are promising, even though an older, lower resolution model was utilized for the regression.

• The regression technique, once trained, is very computationally inexpensive to be used in real-time to improve GHI, DNI and DHI estimates.

• The solar irradiance estimates are comparable to other products available.
  
  ✓ We will extend the dataset to 2014 at 13 km and perform the same technique on 3 km HRRR.

  ✓ Will start to incorporate the GOES East/West composite satellite data.

  ✓ Calculating the line-of-sight model data rather than vertical column.

  ✓ Utilizing NREL’s solar measurements and other sources of measurements to improve the accuracy of the regression.

  ✓ Extend the estimates to forecast hours
Questions?

christopher.clack@noaa.gov  @clacky007  www.linkedin.com/in/clacky007

Photo: Chuck Coker via Flickr