PV Output Variability Modeling Using Satellite Imagery and Neural Networks

Matthew J. Reno[1,2] and Joshua S. Stein[2]

[1]Georgia Institute of Technology, Atlanta, GA, USA
[2]Sandia National Laboratories, Albuquerque, NM, USA

***Abstract* — Variability and ramp rates of PV systems are increasingly important to understand and model for grid stability as PV penetration levels rise. Using satellite imagery to identify cloud types and patterns can predict irradiance variability in areas lacking sensors. With satellite imagery covering the entire U.S., this allows for more accurate integration planning and power flow modeling over wide areas. Satellite imagery of southern Nevada was analyzed and methods for image stabilization, cloud detection, and textural classification of clouds were developed and tested. Artificial Neural Networks using imagery as inputs were trained on ground-based irradiance measurements and were tested and showed some promise as a means for modeling the irradiance and variability for a location at a one minute resolution without needing many ground based irradiance sensors.**

***Index Terms* — solar energy, satellites, artificial neural networks**

I. Introduction

High ramp rates, intermittency, and unpredictable fluctuations continue to be a challenge for the integration of PV at high penetration levels into the electricity grid. With more PV and bigger plants, the variability in the output will increasingly impact the stability of the grid. In the coming years with Renewable Portfolio Standards (RPS) mandates, the rate at which PV is added will only continue to increase. In order for integration planning of PV to be successful, modeling the short-term variability of the plant output for a given location and plant layout is critical, since increases in output variability may require more regulating reserves. Satellite imagery can be used to calculate irradiance and find clouds anywhere on earth. In comparison to setting up an array of irradiance sensors in a location or using a ground camera setup, this has significant advantages such as cost, time, and the ability to use historical data. National Oceanic and Atmospheric Administration (NOAA) makes satellite imagery of North America publicly available for the past thirty years. The purpose of the research is to translate satellite imagery into a model of irradiance, variability, and PV output for a fleet of PV plants at one minute resolution that can be easily implemented into a power flow model of the area.

Current techniques of estimating high frequency (<1hr) solar resource data generally rely on direct measurements of irradiance. Being able to estimate high frequency irradiance (~1 min) from satellite imagery allows the grid impacts of distributed and utility-scale solar generation to be evaluated. Utilities and energy planners need to know how solar photovoltaic plants will affect the operation of the grid in order for these plants to be built, and high frequency solar resource data is needed as input for these studies.

II. Data

The satellite imagery is from the Geostationary Operational Environmental Satellite (GOES) which is owned and operated by NOAA through their Comprehensive Large Array-Data Stewardship System (CLASS). The proposed method uses GOES West or GOES-11 which is located at longitude 135.0W at 35,790 km above the equator and has been in operation since 6/21/2006. The visible wavelength (0.55 to 0.75 Âµm) silicon detector is used for images with 1 sq-km per pixel resolution taken approximately every 15 minutes [1].

There are several groups that have developed algorithms to model average ground irradiance using satellite imagery [2, 3]. While these are often for a large geographical area (3 km grid) and one hour or 15-minute time resolution, they have been shown to be highly accurate [4], but they contain little information about the variability of a location. Therefore, this research does not focus on methods for modeling ground irradiance, but instead develops methods for using satellite imagery to characterize and model the high-resolution solar variability and ramp rates. By combining previously developed models and information with the variability modeling, a PV plant output model can be created.

The model is verified with one minute irradiance and power output data provided by Las Vegas Valley Water District (LVVWD) from six of their PV plant sites in the Las Vegas area starting in August 2006. NREL Measurement and Instrumentation Data Center (MIDC) also provides one minute irradiance data for two sites in the area starting in March 2006 at Clark Station and the University of Nevada. All of Southern Nevada has been chosen as the focus area for the year of 2008.

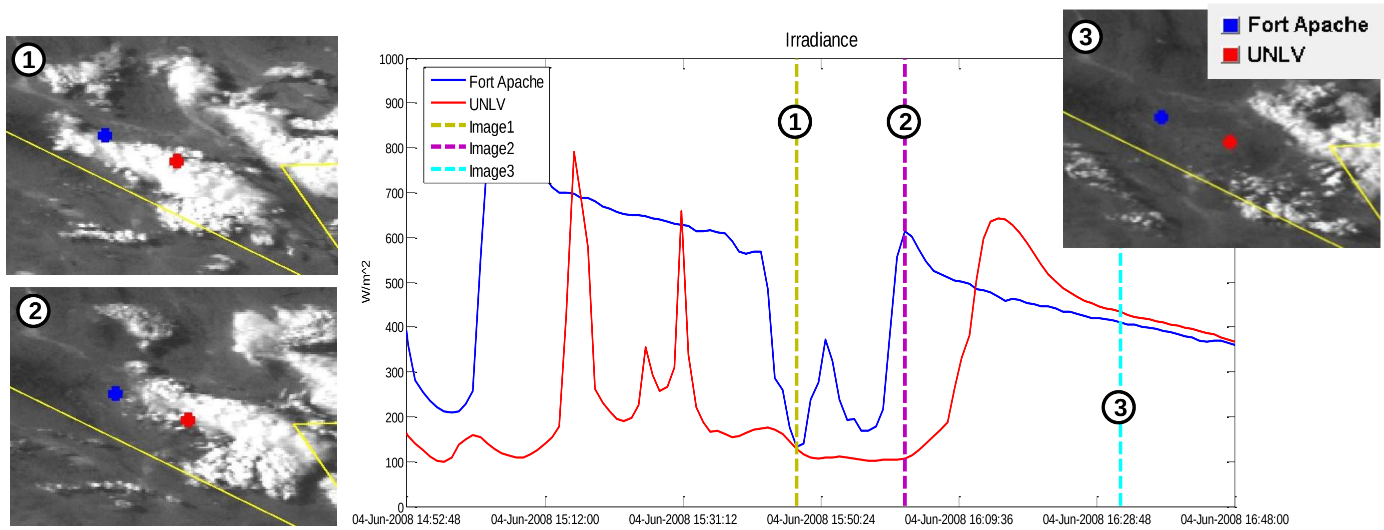


Fig. 1. Three images from GOES 11 of Las Vegas region for 6/4/2008 around 4PM (PST) with corresponding measured irradiance at two ground locations.

III. Model Overview

To develop the model for variability, the GOES satellite images is compared to the irradiance measurements from the ground locations. In Fig. 1, an example GOES images around 4:00PM (PST) on June 4, 2008, shows the correspondence with the ground irradiance measurements at two ground locations, Fort Apache and UNLV.

The model for the system is shown in Fig. 2. The irradiance is modeled at one minute resolution between two historical satellite images 15 minutes apart. First, the images go through image processing such as geographical subsetting, image stabilization, and cloud detection. The processed images have the background image of the ground on a clear day subtracted out to leave only the clouds in the image. The two images are translated into clearness indexes through the trained artificial neural network (ANN) model. The ANN was trained using images and known historical 1-minute clearness indexes from measured irradiance data. Finally, the clearness index is transformed back to irradiance measurements using the clear sky model.

The image processing and background subtraction are described in more detail in the next section, and the main solar variability modeling is accomplished using an ANN to learn the correlation between identified clouds and the high-resolution solar variability for the time period between the images. This is a type of artificial learning to automatically categorize and cluster cloud types and the matching types of variability. Training was done with multiple ground locations with thousands of satellite images throughout the year.

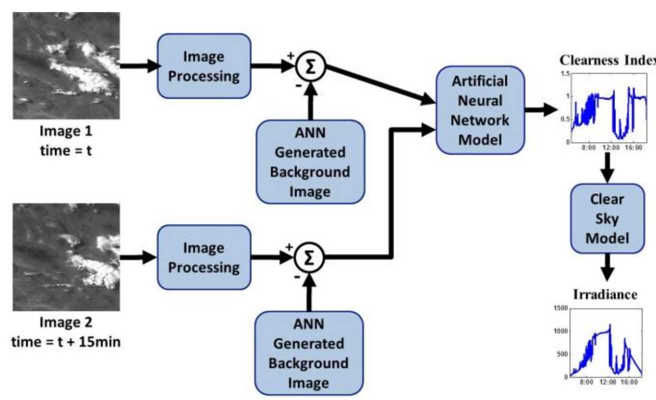


Fig. 2. Model overview using two images 15 minutes apart to generate the irradiance profile for each minute between images.

IV. Image Processing

The downloaded satellite image includes the entire Study Area of southern Nevada (latitude 34N to 38 N, longitude -118W to -112W), but the model focuses on a subset of the image around a site. Using NOAA's image coordinates, the satellite image is cropped ±25 pixels around each site. This removes any clouds that are too far from the site to affect the irradiance during the 15 minute period and allows the model to be more specific to the given location.

A. Image Stabilization

While NOAA geographically locates each pixel and calculates its closest latitude and longitude, this geo-referencing algorithm is not perfect due to slight vibrations and variations in camera, mirror, and scan parameters. The result is that a visibly identifiable geographic location will not have a consistent latitude and longitude through time as calculated by NOAA. If uncorrected, the variability in a location's latitude and longitude would result in movement of the apparent position of objects in the geographically segmented image. This image jitter is typically very small and less than ±3 pixels (3km) in the x and/or y directions. Location variability includes both major movement of the image (>1 pixel) as well as sub-pixel image jitter. An image stabilization routine was developed to correct for the jitter by comparing the image to a reference image created for each time. The reference images is constructed using the closest image in time with a clear sky and scaling it for the correct intensity for that date and time. The detection of clear sky images is described in the next section.

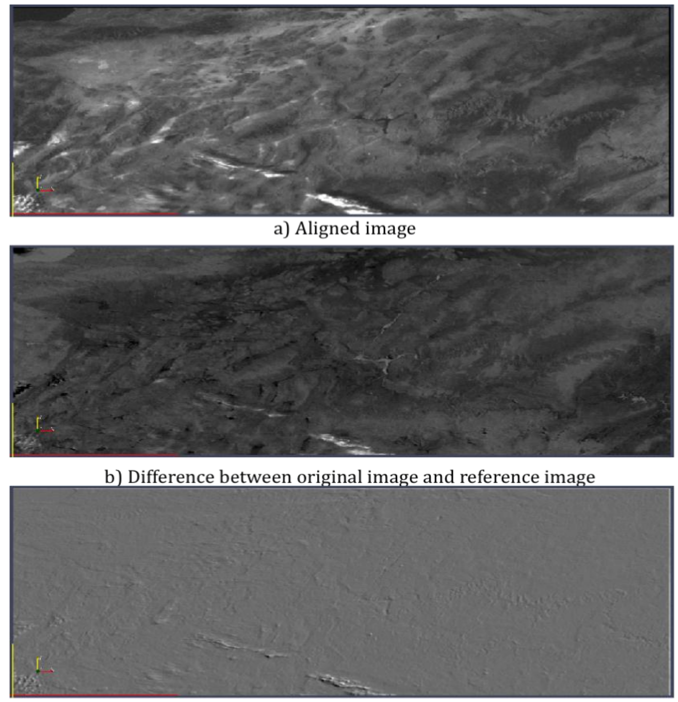
Images are aligned to the present reference image by first expanding each image's resolution four times by linearly interpolating in both x and y dimensions. Correlation coefficients are calculated between the two images for each position with the expanded resolution image shifted by +/- 16 pixels in the x and/or y directions. The location with the highest normalized cross-correlation is chosen as the aligned image. Finally, the shifted image is transformed to the original resolution through bilinear averaging and antialiasing.

With interpolation to higher resolution and image stabilization, both pixel and sub-pixel jitter were corrected to the reference image location. Fig. 3 demonstrates the necessity of performing image stabilization. This process was done using High Performance Computing parallel processing algorithms to simultaneously stabilize a large timeseries of images.

B. Cloud Detection

Identification of visible clouds in the images is not a straightforward process. Challenges include (1) variation in average image intensity and image contrast with time of day and time of year due to the variable solar intensity and angle of the sun on the land surface and (2) variability of the brightness of different ground features, such as dry lake beds and snow, which can appear very similar to clouds.

Two main methods were used for identifying clouds in the satellite imagery. The first method was thresholding the image based on simply finding pixels with intensity values above a certain threshold value. The clouds are more reflective than most geographical features, so especially certain types of clouds can be easily identified this way. Thresholding can be accomplished with a fixed threshold intensity value or with a moving threshold that depends, for example, on features within the image or the time of day. The thresholding method works better under some weather conditions than others. For example, cumulous clouds are easily distinguished from the background because they are the brightest features in the image, whereas broad thin stratus clouds are more difficult to distinguish from background. Fig. 4 shows an example of cloud detection using a threshold determined by the brightest pixel technique.



a) Aligned image

b) Difference between original image and reference image

c) Difference between aligned image and reference image

Fig. 3. Illustration of image stabilization algorithm: (a) image after jitter correction; (b) difference between original and reference image; (c) difference between aligned and reference image.

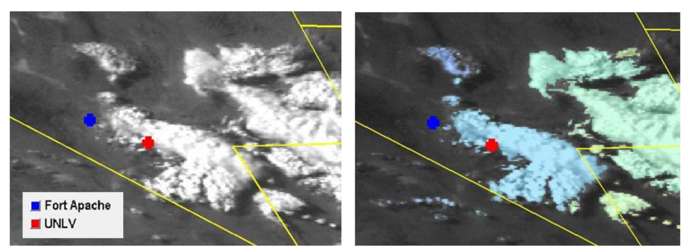


Fig. 4. Example of Cloud Detection using Thresholding: raw image (left) and detected clouds (colored features, right).

The second method used for detecting clouds was Movement Detection. In this technique, the intensity is compared between pairs of sequential images at the pixel level. After scaling each image so that there was no average change in the brightness of the image, the difference in intensity exceeding a certain threshold was assumed to be movement at that pixel. With accurate jitter correction, the only features in the image that can move are clouds; therefore pixels with movement were assumed to be clouds. One problem with this method is that it can only identify leading and trailing edges of clouds. This is because a pixel may be in the middle of the cloud in both images and the difference between intensity at these pixels may not exceed the threshold. Another problem with this method is that it depends on the accuracy of the image stabilization method, since errors in stabilization can lead to apparent movement of ground features and misidentification of these features as clouds. Fig. 5 shows an example of cloud detection using the movement detection technique, and illustrates the main problem with this method. Parts of the clouds recognized in the image are not identified as clouds by the movement detection technique. Note how the large cloud in the center of the image appears thinner in the movement detection image. Several shadows on the ground are also identified as clouds since they move between images.

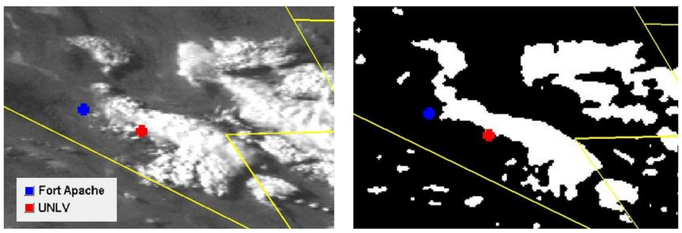


Fig. 5. Example of Cloud Detection using Movement Detection: raw image (left) and detected clouds (white features, right).

V. Background Subtraction

In order for the neural networks to learn the correlation between the clouds and the ground irradiance, the background image of the ground must be removed. This ensures that the neural network only models the impact of the clouds to the clearness index and not the geographical features. Background subtraction was accomplished by estimating what an image of the ground would look like and subtracting this image from the actual image. Areas with clouds should then show up as areas where the intensity difference is above a certain threshold. Fig. 6 shows an example of cloud detection by the method of background subtraction. Note how the background disappeared (i.e., is colored black) in the right panel of Fig. 6 and all that remains in the subtracted image are the clouds. This method allows for better detection of clouds with lower intensity because of less reflectivity. Even small changes in intensity can be detected between the image and the expected background. For example, with background subtraction a cloud pixel could be detected if it is just slightly brighter than normal, even if it is still darker than another geographical feature in the image. As a result of background subtraction, the subsequent image analysis depends only on the clouds in the image, and not on any of the background content.

This method allows the ANN to learn the connection between cloud images and irradiance variability without the ground data included in the image. An additional ANN can be used to generate the background image that varies with the seasonal and daily changes. This ANN is automatically trained by detecting and using only images of the location without clouds throughout the year. It can then generate what an image would look like for any date for that location without clouds. The synthetic background images are verified to match the min, max, and mean intensity for each time and day of the year.

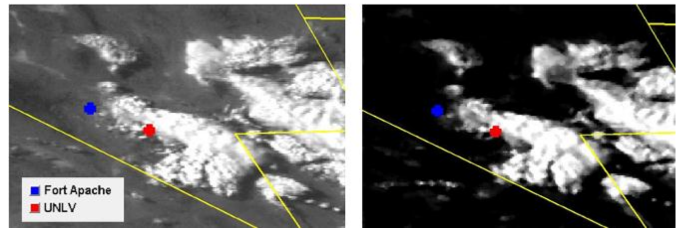


Fig. 6. Example satellite image with background subtraction for southern Nevada. Approximate state boundaries in yellow.

A. Determining Background Portion of the Image

Background Subtraction requires determination of a background image without the presence of clouds. First, Movement Detection was used to select a subset from available images that contain no clouds. If any movement was detected the image was flagged as having clouds, and every day that had images with clouds was classified as a cloudy day. This high sensitivity in detecting clouds guarantees that only days that were completely clear are used to generate the background images.

For each of the images with clear skies, image statistics (mean, minimum and maximum) of the pixel intensity are computed. These statistics vary in a smooth manner during daylight hours and in a more complex but non-random manner annually. Pixel intensity in an image of the background varies due to diurnal and seasonal changes in solar illumination. Because the background image varies by season and time of day, and clear sky images are not available at all times, a neural network that is trained on the clear images throughout the year was used to generate images for all other times during the year.

B. Neural Network Learning of Clear Images

Pixel intensities do not necessarily vary algebraically between clear sky images because of changes in earth's albedo, the occurrence of snowfall, and atmosphere properties. In order to generate suitable background images of the ground for all times of interest, a neural network was developed and trained to produce reference background images. The feed-forward backpropagation ANN was set up with two hidden layers of 300 neurons with a log-sigmoid transfer function. The BFGS quasi-Newton backpropagation algorithm in MATLAB was used to train the ANN with the detected clear day images. The ANN was trained to take the date and times as inputs and produce minimum, average, and maximum pixel values for any time when clear sky images were not available. This setup can be seen in Fig. 7.

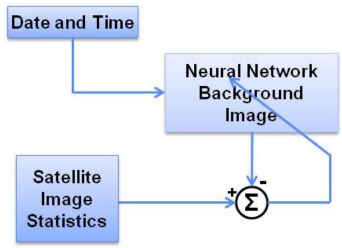


Fig. 7. Training ANN to generate clear background images

C. Neural Network Generation of Clear Images

Once the ANN has been trained, for any input time *t*, it produces minimum ($I_{t,min}$), average ($I_{t,mean}$), and maximum ($I_{t,max}$) pixel values. Clear images for each day of the year were generated by scaling the pixels in the baseline image I_B to the neural network generated statistics for every time *t*. The closest clear sky image, as determined by the cloud detection, was selected as the base reference image I_C. The baseline image I_C was normalized to I_B using (1) to create an image that is easily scaled for any given time I_t. This resulting image I_t represents what the image would look like for a clear, cloudless sky at that time.

$$[I_B] = \frac{[I_C] - \mathrm{mean}([I_C])}{\max([I_C]) - \min([I_C])} \tag{1}$$

$$[I_t] = [I_B](I_{t,max} - I_{t,min}) + I_{t,mean} \tag{2}$$

Fig. 8 compares pixel intensities for synthetic images to those from images during clear sky days throughout 2008. The comparison shows that the neural network and scaling produced images for which the average pixel intensity follows the annual pattern.

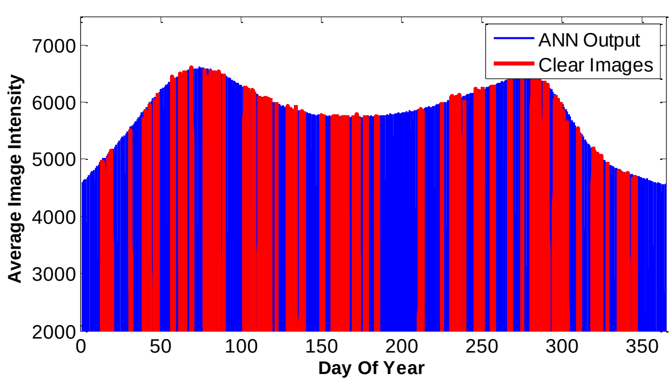


Fig. 8. Comparison of average pixel intensity of clear sky images and ANN simulated output through the year.

Moreover, for individual clear sky days, Fig. 9a shows the neural network was found to produce synthetic images which had statistics reasonably close to the statistics for the actual clear day images. Fig. 9a shows that the synthetic image retains the general structure and characteristics evident in the GOES-11 image. Fig. 9b shows that the ANN also learned the diurnal variation through the year to account for different lengths of days and solar intensity.

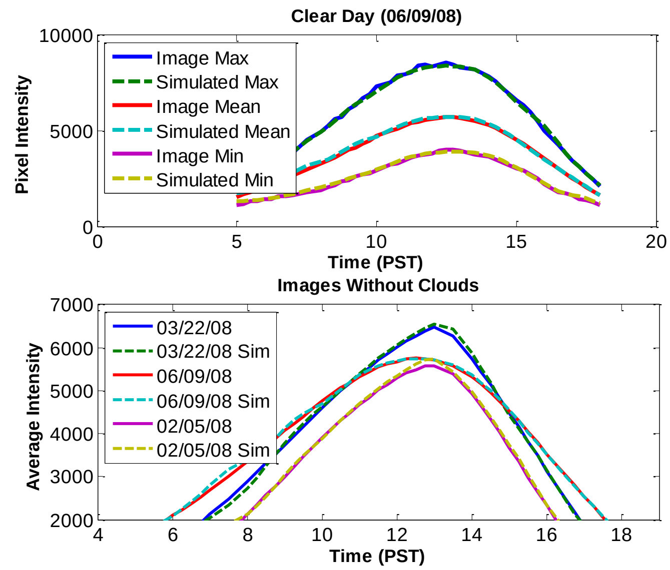


Fig. 9. Comparison of pixel intensity of clear sky images and ANN simulated output for a) Diurnal variation in image statistics and b) diurnal variation for different days of the year.

VI. Training the Neural Network Model

The ANN model was trained using measured ground irradiance between the two images 15 minutes apart. The irradiance was transformed to clearness index by dividing by the clear sky model irradiance. The feed-forward back-propagation ANN was set up with three hidden layers of 300 neurons with a log-sigmoid transfer function. The BFGS quasi-Newton backpropagation algorithm in MATLAB was used to train the ANN with the satellite images as inputs and the ground clearness index as the output as shown in Fig. 10.

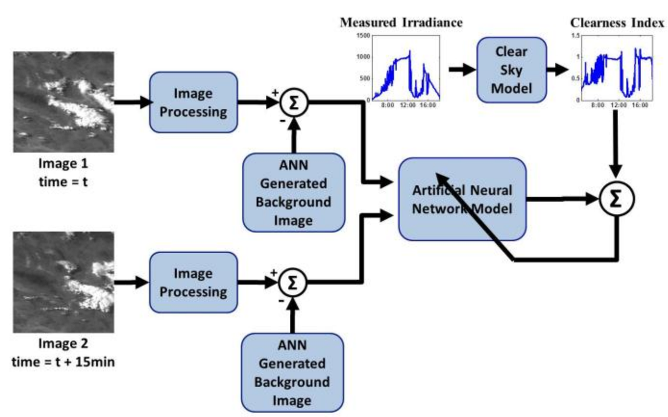


Fig. 10. Training the Neural Network model to generate clearness index from background subtracted satellite images.

One week of images and data was used the train the ANN model. An example of the model learning the training data is shown in Fig. 11 where the model learned the correlation between the images and irradiance very accurately.

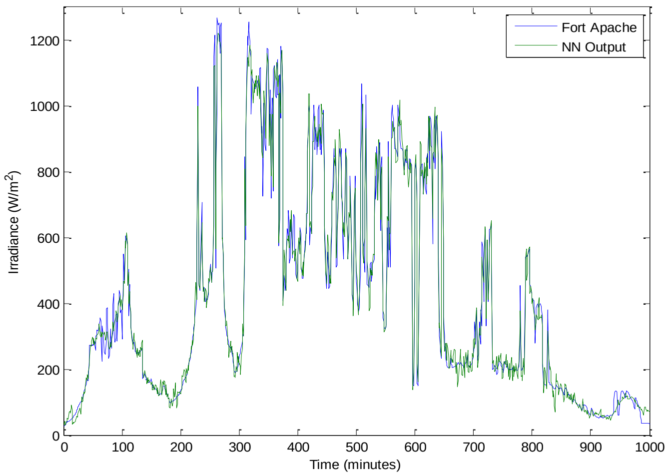


Fig. 11. Measured and simulated (NN Output) irradiance for Fort Apache at 1 minute resolution for May 25, 2008.

VII. Results

After the model has been developed using known ground irradiance values, it can be implemented anywhere with satellite images. The current hypothesis is that the trained ANN will only be able to work with similar weather patterns as in the training data, so it may only work for geographically similar locations with relatively similar weather and clouds. The simulation results will have to be verified with some of the sample sites used in the model development process and some new sites with minute irradiance data to compare the model accuracy for weather type. The correlation and residuals of the irradiance and variability can then be analyzed.

Current model results can be seen in Fig. 12 for Fort Apache for the week after the training data. The model very accurately models the large transitions of the cumulous clouds later in the day, but has more trouble with the variability produced from the high thin cirrus clouds earlier in the day.

Acknowledgement

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

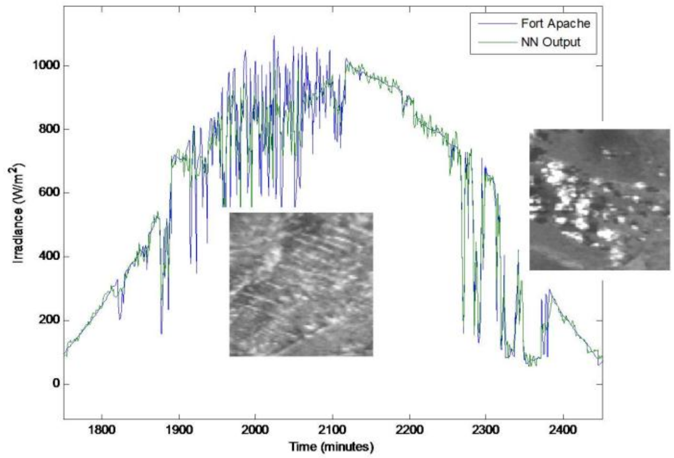


Fig. 12. Measured and simulated (NN Output) irradiance for Fort Apache at 1 minute resolution for May 27, 2008.

VIII. Conclusion

A proof of concept model was developed to predict high frequency irradiance variability in areas with no ground sensors. Artificial Neural Networks (ANN) can be used to generate clear background images to do background subtraction, cloud identification, and cloud classification in satellite imagery. The ANN model has difficulty modeling all possible images to irradiance patters, but categorizing clouds and using separate neural networks for each cloud type could improve accuracy. The overall processing is very intensive and utilizing High Performance Computing Resources is necessary. For interconnection studies modeling solar power on the electric grid, a good model for system variability is needed. This method shows the possibility of modeling high-resolution solar variability using only satellite images.

Reference

[1] NOAA. (2010, July). *NOAA's Comprehensive Large Array-data Stewardship System*. Available: http://www.class.ngdc.noaa.gov/saa/products/welcome

[2] R. Perez, P. Ineichen, K. Moore, M. Kmiecik, C. Chain, R. George, and F. Vignola, "A new operational model for satellite-derived irradiances: Description and validation," *Solar Energy*, vol. 73, pp. 307-317, 2002.

[3] 3TIER. (2011, Feb.). *3TIER: Renewable Information Services*. Available: http://www.3tier.com

[4] J. S. Stein, A. Parkins, and R. Perez, *Validation of PV performance models using satellite-based irradiance measurements : a case study*, 2010.