Report Summarizing Development and Testing of Solar Forecasting for Hawai'i

Prepared for the

U.S. Department of Energy

Office of Electricity Delivery and Energy Reliability

Under Cooperative Agreement No. DE-FC26-06NT42847

Hawai'i Distributed Energy Resource Technologies for Energy Security

Subtask 11.1 Deliverable 5

Photovoltaic Systems

By the

Hawai'i Natural Energy Institute
School of Ocean and Earth Science and Technology
University of Hawai'i

September 2012

Table of Contents

List of Figures

Solar Forecasting in the Hawaiian Islands

Using Ground-based and Satellite Imagery

D. Matthews

September 14, 2012

Introduction

The integration of photovoltaic (PV) systems into the electric grid poses problems due to the natural variability of the resource, insolation. Clouds are the primary factor in modulating insolation, because they directly attenuate solar beam irradiance. We present the ongoing development of a forecasting system that predicts cloud locations and solar irradiance at minute-to-hour time scales, employing ground-based sky imagery and Geostationary Operational Environmental Satellites (GOES) radiometry.

The sky imager based system gives local (~15 km), high-resolution solar forecasts from 1 to 30 minutes ahead. The satellite based system forecasts solar conditions for the entire island chain from 15 minutes to 6 hours ahead. Both systems involve the following steps: identifying cloud-containing pixels in imagery; calculating cloud motion vectors; transforming cloud locations and motion vectors into three-dimensional space; horizontally advecting 3-D locations; and calculating cloud shadows. From cloud shadow predictions, we forecast binary clearness indices. The clearness index is the ratio of the insolation to the top of the atmosphere irradiance. In future work, observations from both systems will be combined with a numerical weather prediction system using four-dimensional variational data assimilation techniques to produce forecasts of surface irradiance from minute-to-day time scales.

Along with the forecasting system, statistical functions are being developed in MATLAB to characterize ground measurements of insolation from pyranometer instruments deployed at various PV test bed sites in Hawai'i. The information provided by these functions will be used to determine the amplitude, frequency, persistence, and time scales of fluctuations in insolation measurements, which will then be used to calibrate and

validate the forecasting system.

In the following sections, we discuss the development and application of the satellite based system in detail, while stepping through an example forecast. The algorithms presented are also designed to utilize sky imagery, but require a different cloud detection method that is also described. We then present the calibration and validation tools that are being developed and their application to pyranometer data from a PV test bed site on the island of Hawaiʻi.

Development of a Solar Forecasting System

Operational solar forecasting using GOES radiometry is currently running without human intervention on sunbeam.soest.Hawaii.edu, a Linux server specifically designed for this task. Forecasts are generated every hour, provided that satellite data is available. The following section details the satellite solar forecasting procedure and methodologies that have been developed, while walking through an example forecast for April 15, 2012 at 3:15 PM local time.

Real-time GOES imagery, collected as 10-bit GVAR counts, is provided by the Marshall Space Flight Center Earth Science Office. This raw data is converted into radiance units and the Hawaiʻi region is extracted from the full disk image. Data is converted and saved in Network Common Data Form (NetCDF) format. NetCDF allows for easy access and sharing of array oriented scientific data and is widely used in the meteorological and oceanographic modeling communities.
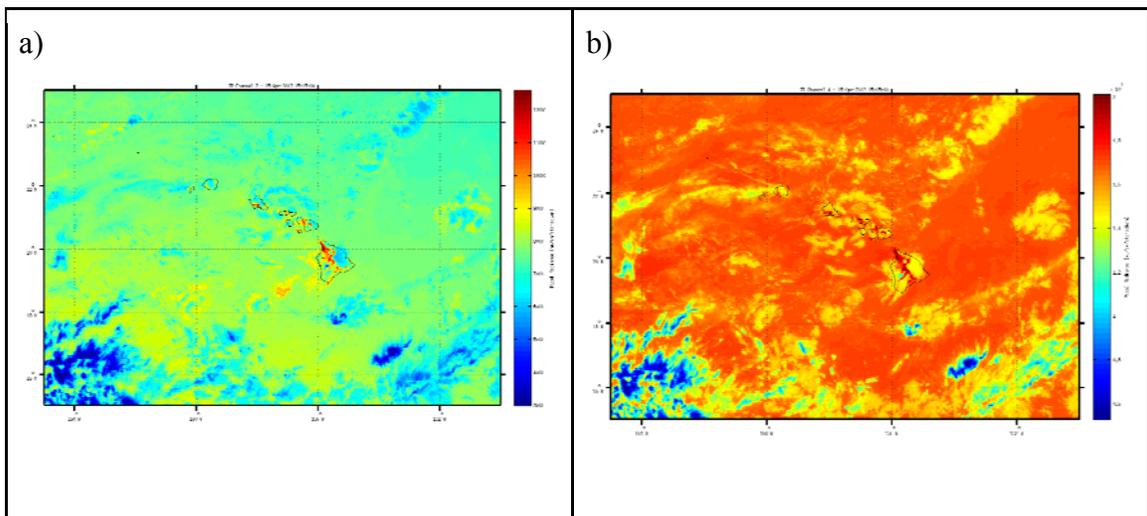
Generating Cloud Maps from Satellite Imagery

Cloud maps are generated using the bispectral composite threshold (BCT) technique [Jedlovec et al., 2008]. This method uses 4 km resolution, 3.9 and 11 μm channel, infra-red (IR) imagery to produce a binary cloud mask at single pixel resolution. The method relies on the basic principle that difference between the emissivity of the thermal

2

wavelengths and the short-wave IR wavelengths varies depending on whether emission is from that of a cloud compared to the land or ocean. This response difference can be detected in channel brightness temperature difference images (TDI), calculated from the longwave image minus the shortwave.

The BCT method uses a four-step algorithm to identify cloud-containing pixels that consist of two spectral and two spatial tests. The first step is an adjacent pixel test, conducted on the TDIs. The edges of clouds are established thorough the identification of fluctuations in variance along the scan line. If the variance between adjacent pixels is above a predetermined threshold, a cloud edge is flagged. The second step is a one dimensional spatial variability test that attempts to fill in the cloud edges established in step 1. In the TDIs there is a difference in the spatial variability of the land and ocean compared to a cloud. Using that difference, step 2 flags cloud-containing pixels by comparing values between adjacent pixels in the TDI to a predetermined threshold.

The third and fourth steps in the BCT method use statistical information generated from the previous 20 days of IR imagery to identify cloud-containing pixels. Images are regularly available every 15 minutes. By creating statistical composite images of the previous 20 days at each of these times we are attempting to capture cloud free land and ocean conditions. The third step uses positive and negative composite images, generated by taking the smallest positive and negative value for each pixel from the previous 20 days of TDIs per image time. When evaluating the current TDI, a pixel is flagged as cloud-containing if the value is larger than the positive composite value or smaller than the negative composite value (composite values are adjusted by a predetermined parameter to maximize the performance of the method). The fourth step uses a warm, cloud-free, thermal composite image, generated from the second warmest pixel of the 11 μm channel images from the previous 20 days of TDIs per image time. The second warmest pixel is used to reduce the noise in the composite image. A pixel is flagged as a cloud if its IR value is less than the composite value (adjusted by a tuning value) for that time period.

3

Figure 1a and 1b show the 3.9 and 11 μm channel images for the example forecast. Both GOES IR images give a qualitative representation of the clouds over Hawaiʻi, which reside primarily in the waters south of and directly over the island chain. In the corresponding TDI (Figure 2a), clouds are qualitatively more distinct, seen as large negative values, as this is a daytime image. The quantitative cloud mask shown in Figure 2b is consistent with that qualitative view. To generate this cloud mask, the parameters suggested by Jedlovec et al., 2008 were adjusted to account for the local tropical conditions. Further tuning and validation will be done using in situ data from various sites around Hawaiʻi.

Figure 1: Example 3.9 (a) and 11 (b) μm images used for cloud detection on April 15, 2012 at 3:15 PM.
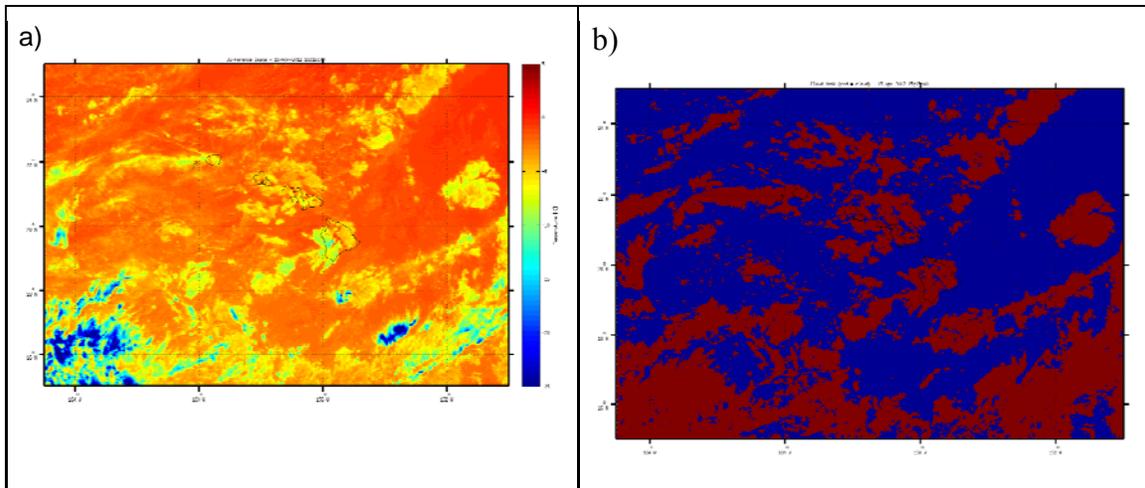
Figure 2: Brightness temperature difference image (a) and BCT cloud mask (b); red areas are clouds and blue areas are clear skies.

Generating Cloud Maps from Ground-based Sky Imagery

The upward looking sky imager provides high resolution visual measurements using a charge coupled device (CCD) camera and a fisheye lens. The CCD sensor has three channels at red, green, and blue frequencies. The cloud detection technique employed uses the ratio of the red channel to the blue channel to flag cloud-containing pixels. This method is based on the increased molecular scattering that occurs at shorter wavelengths. This effect causes the clear sky to be blue and return a larger relative signal measurement in that channel. Clouds, on the other hand, return a more uniform signal at visible wavelengths. Because of this, the red to blue channel ratio (RBR) is a good indicator of whether the dominant source of incoming light on a pixel is from the clear sky or a cloud. In Figure 3 we show an example sky image taken from the University of Hawai'i at Mānoa campus on November 7, 2011. The white clouds in the visual image are clearly identified in the RBR image and using a simple 0.5 threshold the clouds are qualitatively identified. In future work, a more complex methodology following the work of Chow et al. 2011 will be employed.
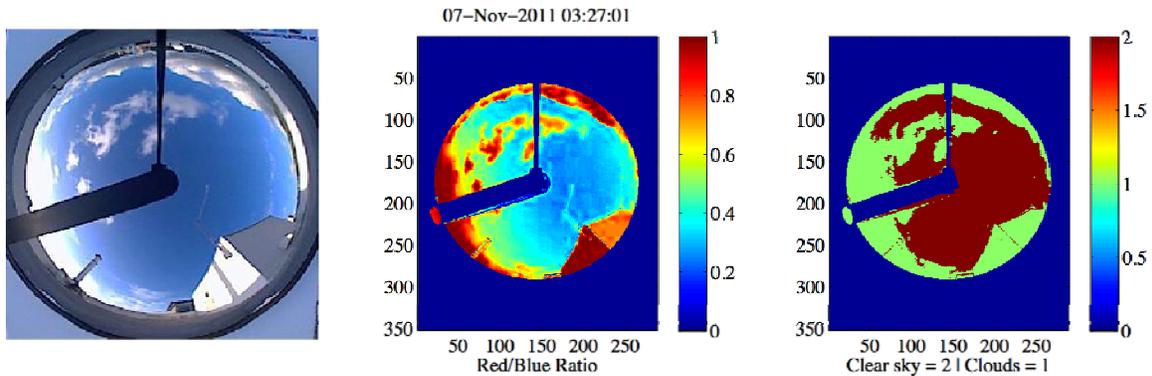
Figure 3: Processing steps of cloud detection on sky imagery. The raw sky image from November 7, 2011 shows relatively clear skies (left). The RBR is used to identify cloud containing pixels through the change in atmospheric scattering characteristics (center). Finally, the cloud map is generated from a simple RBR threshold (0.5), with clear sky shown as 2 and clouds as 1 (right).

Calculating Cloud Motion Vectors

Cloud velocities are determined by applying the maximum cross-correlation (MCC) technique to 1 km resolution, visual GOES imagery. The MCC technique, illustrated in Figure 4, is a fully automated, robust pattern tracking algorithm that calculates the displacement of small subregions from one image to another. The procedure cross correlates a template subwindow in an initial image with all possible subwindows of the same size that fall within the search window of a second image. The location of the subwindow in the second image that produces the highest cross correlation with the subwindow in the first image indicates the most likely displacement of that feature. A velocity vector is then calculated by dividing the displacement vector by the time separation between the two images. Because the method is fully automated and conditions may exist in which there are little to no features to track, filtering the raw MCC vector field is essential. We employ both a correlation cutoff, to ensure only the very similar patterns are matched, and a next-neighbor filter, which allows only a spatially smooth velocity field.

6
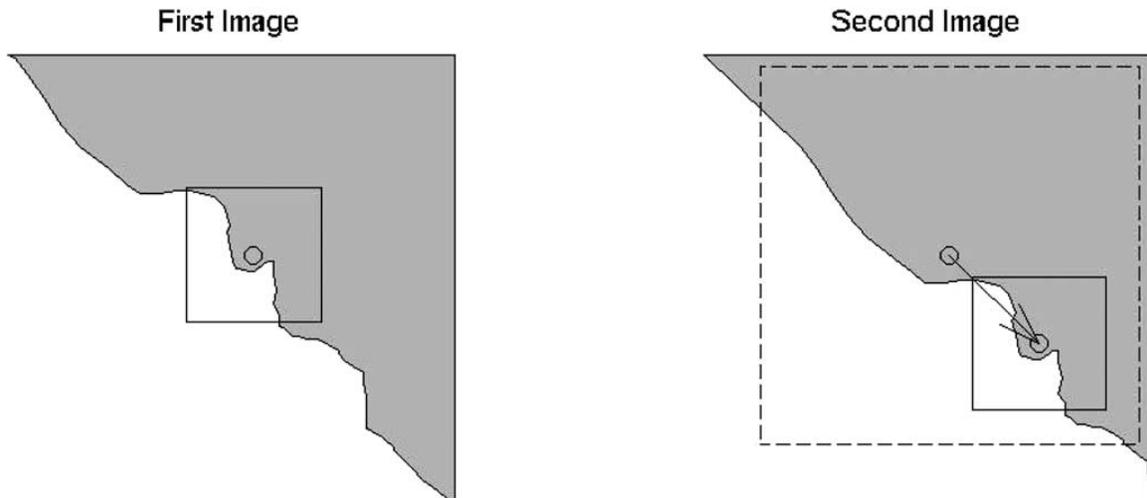
Figure 4: Illustration of the Maximum Cross Correlation method.

In Figure 5a, we show the GOES visible image from the example time period. Overlaid on the satellite image are the wind motion vectors generated by applying the MCC method to this image and the visible image 15 minutes prior. The vector field is dominated by the prevailing easterly trade winds; however the effects of the island chain can also be seen. The mountains on the island of Maui and Hawaiʻi are tall enough to penetrate the wind inversion layer, causing the wind to deflect around the islands and creating a large wake region of weakened flow.

Clearness Index Forecasting

In the next step, we generate three dimensional cloud maps and wind velocity fields using cloud height data. For the satellite-based system, cloud top height (CTH) maps are pulled from the Space Science and Engineering Center at the University of Wisconsin. These maps are generated by applying the Advanced Baseline Imager Cloud Height Algorithm (ACHA) to GOES sounder data (the CTH map for example time period is shown in Figure 5b). For the sky imager system, cloud height is provided by a ceilometer, a LIDAR system specifically designed to give the height of a cloud base.

7
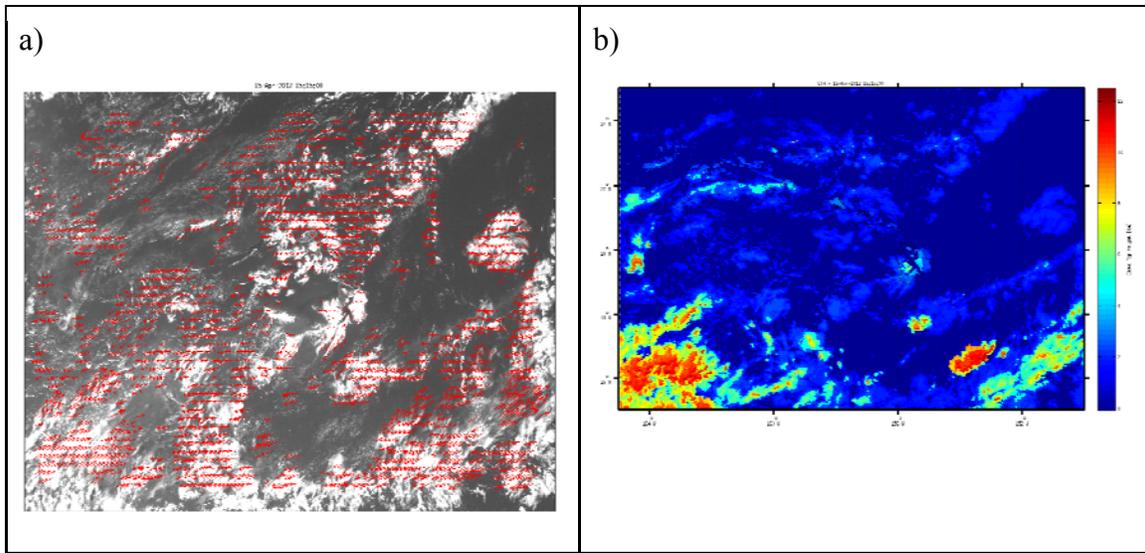
Figure 5: Example GOES visible image from April 15, 2012 at 3:15, with the wind velocity field generated from the MCC technique applied to this image and the image 15 minutes prior (a). Also, the CTH generated from GOES sounder data is shown (b).

The cloud velocity fields are then used to forecast cloud locations and heights by horizontal advection of the cloud maps. Solar zenith and azimuth positions are found using a sun position algorithm. Vectors between the Sun and ground forecasting locations (a PV test site for example) are calculated and transformed into the 3D cloud map space. Binary cloud index forecasts are generated utilizing the solar vector and the advected 3D cloud maps.

This system simply forecasts the binary clearness index and does not estimate the degree of atmospheric attenuation or diffusion of solar irradiance (by using, for example, a radiative transfer model). To test the applicability of using a binary CI to forecast solar conditions, we analyzed one year (2011) of irradiance observations collected at the Pu'u Wa'a Wa'a (PWW) PV test bed site on the island of Hawai'i. The CI estimates are calculated from 5-minute averaged, pyranometer data divided by TOA irradiance data provided by the National Renewable Energy Lab's (NREL) Measurement and Instrumentation Data Center (MIDC) solar position (SOLPOS) calculator. The CI data is analyzed using bivariate histograms, binning the data as a function of time (hour of day

8

and month to look for diurnal and seasonal signals) and CI level, with results shown in Figure 6a and 6b.   The histograms show a clear bimodal distribution in CI estimates, with a majority of the data falling in bins either below 0.2 or between 0.6 and 0.8. This pattern shows very little seasonal variation and suggests that a binary CI would be useful in forecasting insolation levels. Also shown in the histogram, is the formation of afternoon clouds, with CI values primarily below 0.2 late in the day.
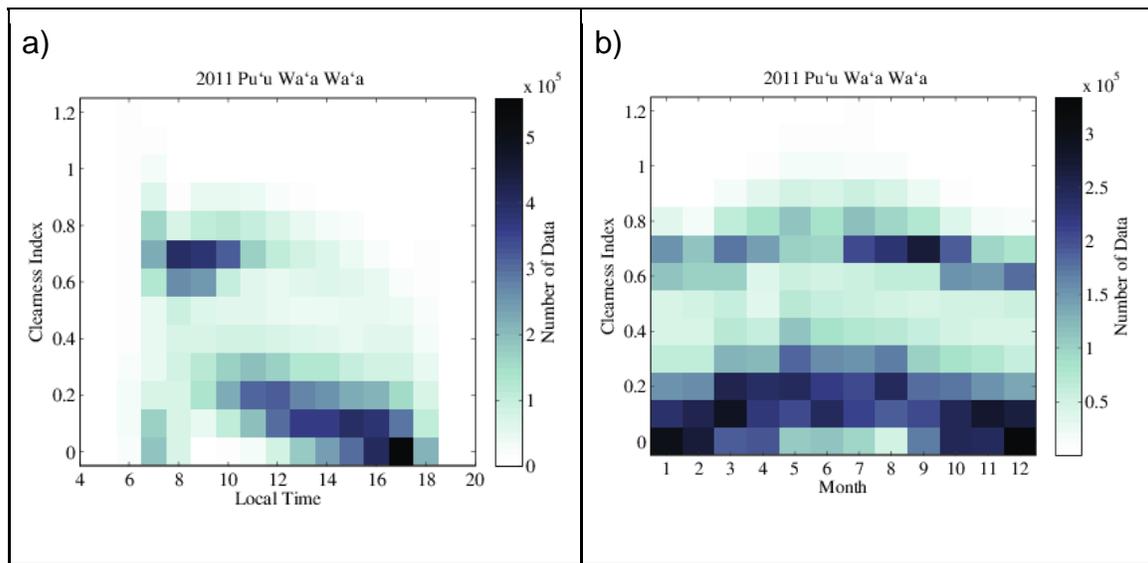
Figure 6: Bivariate histograms of 5-minute CI data from the PWW PV test bed site on the island of Hawaiʻi. Data are binned according to CI level and local time (a) or month (b).

In Figure 7, we show the CI forecast generated from the example satellite imagery presented above at the La Ola, Lānaʻi solar farm. CI estimates are calculated from pyranometer observations, gathered by the Solar Resource & Meteorological Assessment Project (SOLRMAP), divided by TOA irradiance from SOLPOS. From the satellite information at 15:15 local time, the system predicts the relatively cloudy conditions (CI < 20%) found at the Lānaʻi PV site for the next hour, along with the transition to relatively clear conditions (CI ~= 80%) at 6:15 (within 15 minutes). In future work forecasts such as this will be compared with pyranometer-derived CI estimates to directly validate and calibrate the forecasting system.
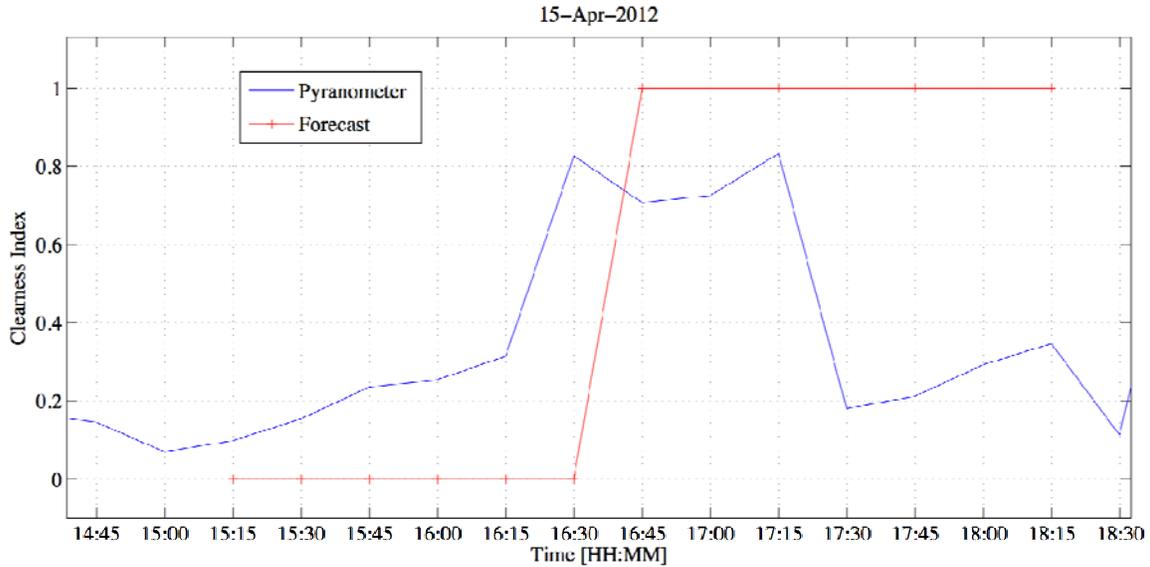
Figure 7: Binary CI forecast generated from the satellite-based solar forecasting system on April 15, 2012 for the La Ola, Lāna'i solar farm (red line). Also shown are the CI estimates, calculated from pyranometer observations taken at the solar farm (blue line).

Calibration and Validation

In order to utilize the pyranometer measurements for calibration, noise levels in the observations need to be characterized. This is done following the methodology of Zanife et al. 2003, in which the power spectral density (PSD) of a time series is computed and variance is deduced from the high-frequency part of the spectrum. We show an example of this method applied to 1 second data from the 2011 PWW pyranometer time series (Figure 8a shows a 1-day example of this data set). Prior to using this method the frequency of the noise floor must be estimated. The spectra of several test cases taken from the PWW time series showed the noise floor in the data ranged from 0.1 to 0.5 Hz. Noise levels are then determined from the summation of the PSD over the noise floor frequency range (converted to the observation unit by taking the square root of the PSD and dividing by the sampling period). Figure 8b shows the noise level estimates for 20-day sections of the 2011 PWW time series. Noise levels were found to be consistent throughout the year, ranging from 2 to 18 W/m^2. This type of calculation may also be

used to flag instrument failures in an operational sense by looking for increasing noise level trends or noise spikes in the data.
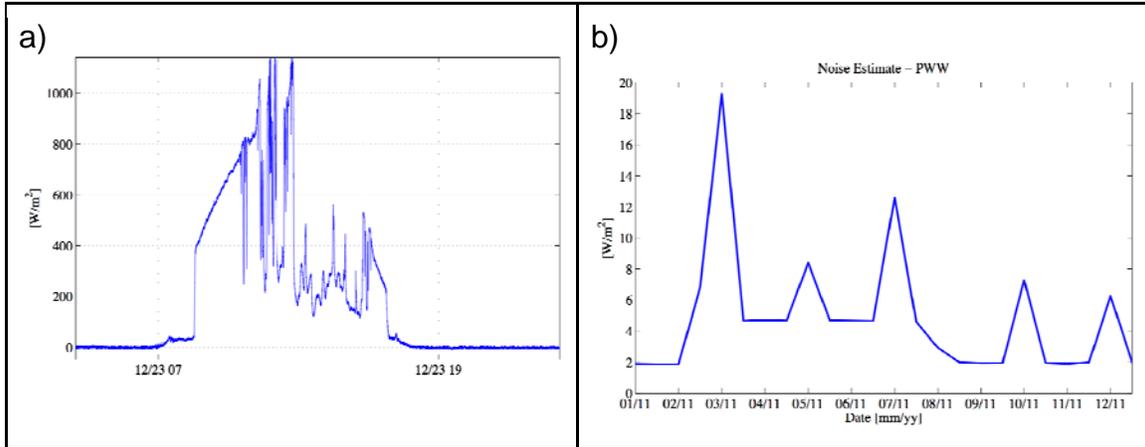
Figure 8: Example pyranometer data from the 2011 PWW time series used to test the calibration and validation tools being developed (a) . Noise estimates using PSDs of 20-day sections of the PWW time series (b).

CI time series, described in the previous section, lack the diurnal and seasonal signals of the raw insolation measurements and by definition are stochastic, stationary time series. This allows for time scale estimates of irradiance observations that are not biased by the diurnal cycle. In the following, we describe the semivariogram technique [Matthews et al. 2011] that is used to estimate dominant temporal or spatial scales and variability. We also show an application of the method to CI estimates from the 2011 PWW pyranometer time series, in order to find the dominant time scale of atmospheric attenuation and variability imposed by the atmosphere on the irradiance signal.

The semivariogram method describes the covariance of data as a function of the data distance by summing residual squared values according to a time or space lag. It is defined as the expected squared difference of the data values separated by the lag distance, h [Banerjee et al., 2004]:

$$\gamma(h) = \frac{1}{2}$$

where z(x) represents the data at a given location, x, and γ(h) is defined as the semivariogram function. The empirical semivariogram is computed for all available lag distances within the data by binning and averaging all squared data differences, for the specified h. Bin sizes should be chosen to be as small as possible, while maintaining significant and consistent sample numbers per bin.

A statistical model is then fit to the empirical semivariograms for a mathematical representation of the variance [Journel and Huijbregts, 1978] and to provide a tool for describing how a measurement varies as it is perturbed from its location. In the following example we use a stationary Gaussian model [Kitanidis, 1997] to fit to the empirical semivariogram. The Gaussian model is given by:

$$\gamma(h) = C_o + (\sigma^2 - C_o)\left(1 - \exp\left(-h^2/L^2\right)\right)$$

where the model parameters, Co and L, are solved for using the method of least squares. The "range," α, is defined to be the lag at which the variance has reached an upper limit. Because the Gaussian function decays asymptotically, the range is estimated from the model parameter L [Kitanidis, 1997]:

$$\alpha \approx 7L/4.$$

The "sill," defines the upper limit of the variance, with the "nugget," Co, defining unresolved variance (measurement error) given by the semivariagram at lag zero. The difference between the sill and nugget is attributed to the geophysical variability captured by the observations.

In Figure 9, we show the semivariograms calculated from 2011 PWW irradiance observations. The empirical semivariogram begins to flatten out with increasing lag at around 100 minutes and 0.23 CI. From the model fit to the empirical data, the range and sill values are given as 104 minutes and 0.25 CI, respectively. This information suggests that the dominant variability caused by atmospheric attenuation is over 20% of the

incoming TOA irradiance level and occurs at a time scale of 100 minutes. In future work, this method will be used to characterize other PV sites around the Hawaiian Islands, as well as seasonal and time of day variations in time scale and variability. In addition, the method will be applied to satellite and sky imager observations and forecasts to validate and calibrate the systems.
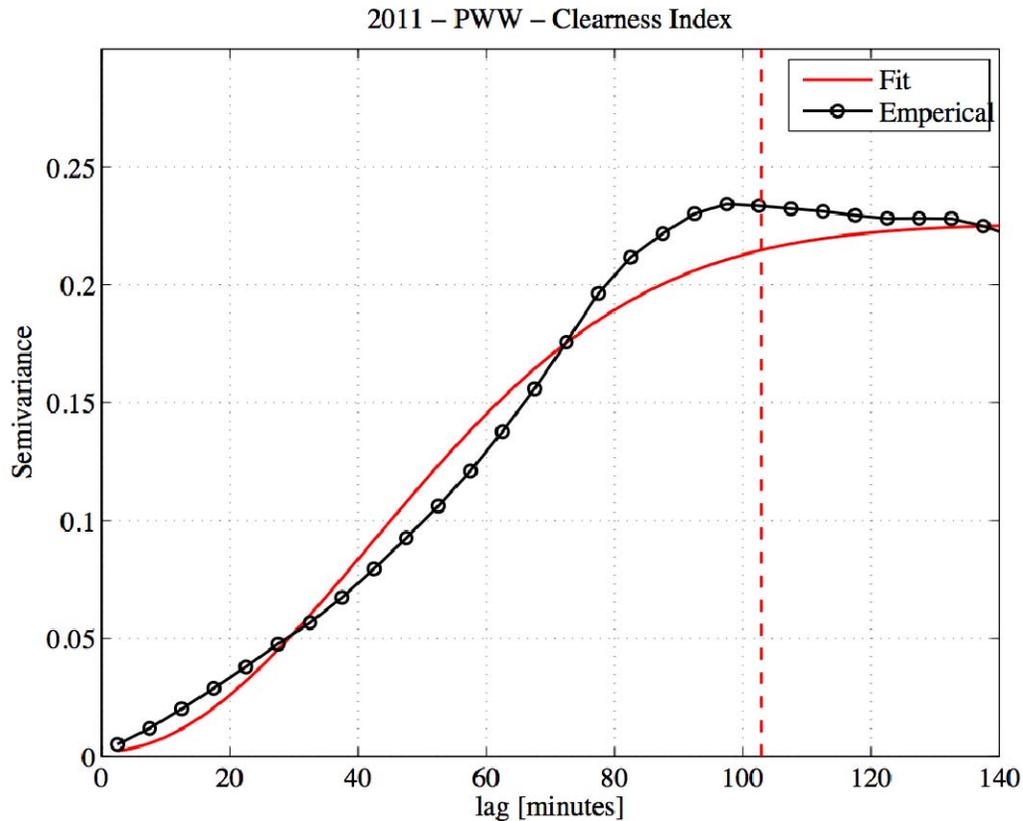
Figure 9: Semivariogram calculated from 2011 PWW pyranometer time series.

To characterize the magnitude and duration of ramp events, we apply the swinging door compression algorithm [Bristol, 1990], a data trending technique that identifies dominant points in a time series. The method obtains the optimal straight line approximation to a segment of a time series, given the initial point in that line. The estimated line is optimal in the sense that it is the longest straight line possible for the data, given a set error; and

13

the line is computed with the minimum number of computations possible.

The method, illustrated in Figure 10a, starts at an initial point, setting two pivot points above and below the initial point. The pivot points construct two closed doors, set at a distance that defines the allowable error in the straight line estimation. The doors are then swung open at the pivot points, along the time series, with the top door swinging up and the bottom swinging down. As each new point in the time series is considered the doors are opened further to include that point within the area bounded by the doors. When a data point is reached, beyond which the doors would be opened past parallel, a midline segment is generated, starting at the initial point and extending to the time of the last point considered. These two points are considered the dominate points in the time series. The last data point considered is then used as the initial point in a new data segment, continuing the process to the end of the time series.
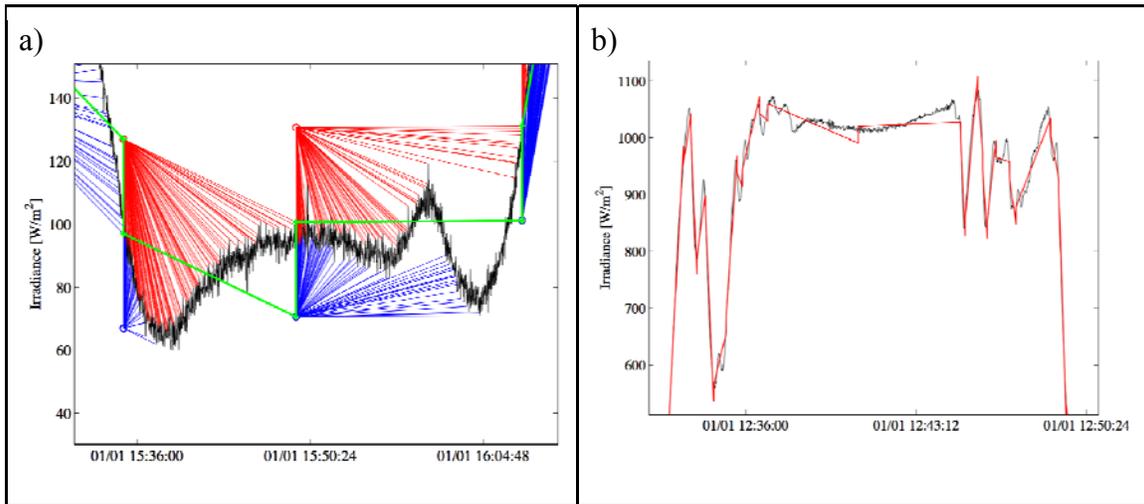
Figure 10: Illustration of the swinging door algorithm applied to 1-second pyranometer data (a), along with a small 15-minute section of the resulting dominant points (red line) and raw data (black line) (b).

We compress the data set to 0.3% of its initial size by applying the swinging door method to the 2011 PWW time series, using an allowable error of 30 W/m2. Figure 10b shows a

14

15-minute example section of the raw and compressed data. The swinging door method captures dominant features of the raw data, allowing for large changes in irradiance that would lead to a ramp event in solar energy to be easily characterized. In Figures 11a and 11b, we show bivariate histograms of the change in magnitude and the duration between dominate points binned according to the hour of the day. The histograms show that the majority of changes in irradiance occur on time scales of minutes and are on the order of magnitude of 100 W/m². The larger changes in irradiance are most commonly found from 10:00 to 13:00 local time. Results such as these will be helpful in future tuning and validation of the satellite and ground-based forecasting system, as well as for characterizing local irradiance conditions.
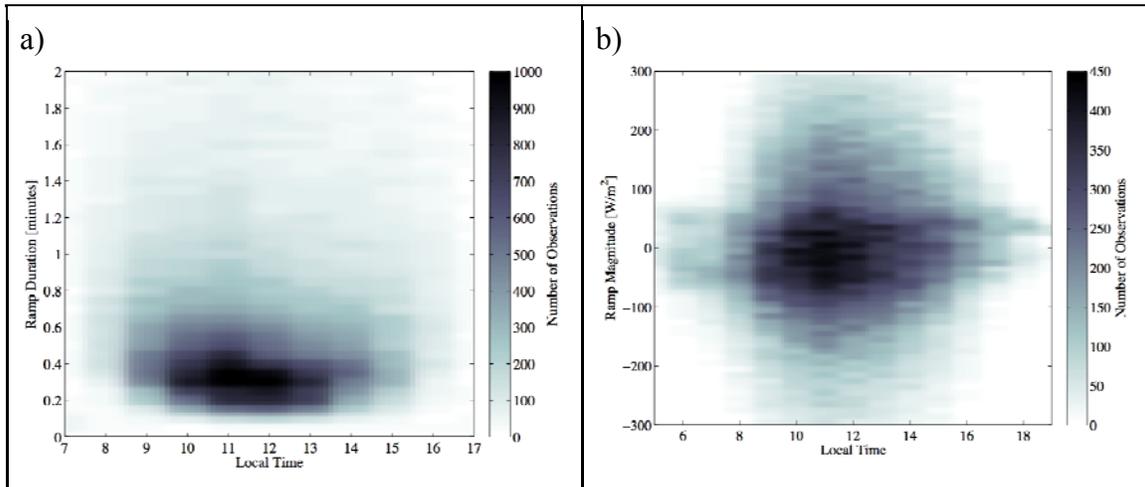
Figure 11: **Bivariate histograms of the duration (a) and change in magnitude (b) of ramp events established by the swinging door algorithm applied to the 2011 PWW time series.**

Current Status

The satellite-based solar forecasting system has been operational since July 8, 2011 and initial validation using pyranometer insolation measurements has begun. A 6-month time series of binary CI forecasts is being generated. This time-series will be used for comprehensive validation and calibration of the forecasting system using the tools described above. Future work will focus on the sky imager based solar forecasting

15

system and combining the satellite based and sky imager based forecasting methods using data assimilation techniques in the Weather and Research Forecasting (WRF) model system.

References

Bristol, E., "Swinging Door Trending: Adaptive Trend Recording" 1990, ISA National Conference Proceedings.

Chow, C., Urquart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., Washhom, B., "Intra-hour Forecasting with a Total Sky Imager at the UC San Diego Solar Energy Testbed" Solar Energy, 2011, 85, 2881-2893.

Jedlovec, G., Haines, S., LaFontaine, F., "Spatial and Temporal Varying Thresholds for Cloud Detection in GOES Imagery" IEEE Trans. on Geo. and Remote Sensing, 2008, 6, 1705-1717.

Journel, A.G., Huijbregts, C., 1978. Mining Geostatistics. Acedemic Press, Harcourt Brace Jovanovich Pubilshers, New York, New York 600 pp.

Kitanidis, P.K., 1997. Introduction to Geostatistics, Applications in Hydrogeology. Cambridge University Press, New York, New York 249 pp.\

Matthews, D., Powell, B., Milliff, R., "Dominant Spatial Variability Scales from Observations around the Hawaiian Islands", Deep Sea Res., 2011, Part I, 58(10), 979–987.

Zanifé, O. Z., Vincent, P., Amarouche, L., Dumont, D., Thibaut, P., Labroue, S., "Comparison of the Ku-band Range Noise Level and the Relative Sea-state Bias of the Jason-1, TOPEX, and Poseidon-1 Radar Altimeters", Mar. Geod., 2003, 26, 201–238.