2013

High Penetration PV Initiative:

PV Monitoring and Production Forecasting

Consultant Report

James Bing

NEO Virtus Engineering, Inc.

SMUD®

GAF 0448-14

NEO Virtus Engineering, INC.
410 Great Road, B-6, Littleton, MA 01460 • 978.952.2444V • 978.952.6846F • www.neovirtus.com

High Penetration PV Initiative:

PV Monitoring and Production Forecasting

James Bing

NEO Virtus Engineering, Inc.

Project Task Lead

DISCLAIMER:

This project is receiving funding under the California Solar Initiative (CSI) Research, Development, Demonstration and Deployment (RD&D) Program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the CPUC, Itron, Inc. or the CSI RD&D Program.

Task 5: Overview

The solar monitoring work performed by NEO Virtus Engineering under this grant compliments that of the other team members and shares a vision of the future of utility solar forecasting as a fundamental component of grid management for enhanced stability and optimized economic performance. In that shared vision distributed solar monitoring and forecasting of PV production will move out of the realm of research and into the utility control room as a standard component of EMS/SCADA (HECO visualization systems) and providing monitoring of energy flow at the feeder level (BEW DNV) and inform utilities and ISOs with actionable data for dispatch, demand response, energy storage, inverter VAR control or curtailment (SMUD). To achieve the optimal solution for the continuum of spatial and temporal forecast resolutions which will be needed to integrate high penetrations of PV from transmission to the far end of distribution, data will likely come from an ensemble of solar forecasting technologies. NEO's solar monitoring holds the promise that both satellite or sky imager forecasting technology will benefit from feedback from distributed ground-based irradiance measurements on a 1-minute or higher frequency. All California utilities and CAISO can exploit the technologies demonstrated in this study as the industry moves toward a “real time” irradiance monitoring and PV production forecasting network as a part of smart grid applications.

The investor owned utilities of California can leverage much of the work presented in this report immediately by applying it to their own needs for management of increasing levels of PV on their distribution networks. Lessons learned in this research which are immediately applicable include the use of a utility's full asset base --utility poles, substations, maintenance and office buildings, communications networks and automated meter infrastructure (AMI)-- as potential platforms for irradiance measurement and conduits for irradiance data. NEO Virtus Engineering's demonstrated capability for providing day ahead forecasts for SMUD's 100MWs of feed in tariff (FIT) sites using open source meteorological data from the National Digital Forecast Database (NDFD) or other sources is a process which can be duplicated by any of California's utilities. IOU's can use forecasts in the day ahead time frame to inform unit commitment and as an input to day ahead energy trading decisions. As the IOU's deploy more smart grid technologies the day ahead and near term forecasting methods described in this report can provide inputs to their control algorithms. For those utilities which are already invested in PV forecast technologies such as satellite and sky camera services this report illustrates the way in which "real time" ground truth irradiance measurements across their service territories can serve as references or corrections to the data streams coming from those systems. For those IOU's which have not yet invested in solar power forecasting services or technologies, duplicating part of the work performed in this study IOU's can leverage their own solar resource measurement capabilities to validate the claims of solar forecast venders and help select products and services suitable for their operations.

Objectives

Task 5: PV Production Forecasting

This final task included the following objectives:

a. Forecast regional irradiance for a day-ahead window 20 to 36 hours in the future
b. Model PV production for regional PV installations based on the day-ahead forecast.
c. Validate the day-ahead forecast irradiance & PV production data with measured data.
d. Forecast regional irradiance for a short-term window 1 to 3 hours in the future
e. Model PV production for regional PV installations based on the short-term forecast.
f. Validate the 1 to 3 hours forecast irradiance & PV production data with measured data.
g. Merge SMUD feeder load and PV forecast data.

Approach

Task 5: PV Production Forecasting

The original scope of work for this subtask was intended as a roadmap for obtaining the short term (0-3 hour) and long term (day ahead) solar forecasting and validation goals of this project. That roadmap laid out one possible path to achieving these goals. Most, but not all of, this path was followed as developments and discoveries presented themselves through the course of the project.

5.1 PV Production Forecasting Monitoring & Data Retrieval

NEO Virtus Engineering, Inc. (NEO), of Littleton, MA, was responsible for this subtask, including the design, fabrication, deployment[1], and maintenance of an irradiance monitoring network. The network was designed to collect one year of global horizontal irradiance (GHI) and ambient temperature in SMUD's service territory. NEO was further tasked with creating and maintaining a database of the collected data and making it accessible to SMUD.

5.1.1 Design, Manufacture, Deploy and Maintain 5km Irradiance Monitoring Network

Creating the Monitoring Network

This effort consisted of a range of technical and logistical subtasks required to design, manufacture and deploy a monitoring network of five primary and sixty-six lower cost secondary irradiance monitoring stations. The network was deployed on a five kilometer grid using autonomous PV/battery powered data loggers installed in substations and on SMUD utility

[1] NEO Virtus Engineering and SMUD collaborated on the identification and mapping of the installation sites. NEO coordinated the installations with SMUD line crews performing the actual installations on utility poles. SMUD's data technician and NEO personnel performed the installations of the five primary stations.

poles. The five primary stations used Rotating Shadowband Radiometers (RSRs) to collect and store irradiance measurements at the substations, while independent data loggers combined with global horizontal pyranometers made up the 66 secondary monitoring stations.

Figure 1: Primary station (Rotating Shadowband Radiometer) located on roof of California State University Sacramento (CSUS)

The primary RSR stations required almost no modification because the Rotating Shadowband Radiometer is an existing commercial product which NEO Virtus Engineering manufacturers for its client, Irradiance, Inc. The secondary stations, however, were designed "from scratch" for this project due to the unique requirement for mounting on utility poles. The exterior enclosure design includes a PV module and bracket, an adjustable pyranometer arm and support bracket for easy pole mount. The internal enclosure design houses a battery, data logger, cellular modem and associated electronics. Both the primary and secondary stations use Licor 200SZ silicon pyranometers for irradiance measurements. The secondary station enclosure and the internal mounting armature are entirely non-metallic allowing the modem antenna to be mounted internally.

Figure 2: Fabrication of the secondary stations at NEO Virtus Engineering, Inc.

Sheet metal and non-metallic parts were machined and the secondary units were assembled, programmed and tested at NEO Virtus Engineering's Littleton, Massachusetts facility (see figure 2). The calibration of each pyranometer was checked by comparing its output with that of the others in sets of 25 under identical irradiance conditions. Bulk rate cellular data plans were negotiated based upon data transmission bit calculations. Two prototype units were deployed, one pole mount and one on SMUD's headquarters roof, to assure cellular connectivity and to collect sample data sets. Each of the secondary station production units was bench tested and then field tested on NEO's roof in its final assembled form before shipping. Units were shipped in batches of 16 to SMUD's Sacramento warehouse where, once received they became property of SMUD.

SMUD's GIS department overlaid a map of available assets (locations for mounting primary and secondary monitoring stations such as utility poles, buildings, and secure land) on to the National Digital Forecast Database (NDFD) forecast 5km grid. The approximate "centroid" of each NDFD grid cell was targeted. SMUD staff surveyed the candidate sites and coordinated with SMUD's line crews for approval of final locations. Installation of secondary monitoring stations started in early April and was completed in late May. After each installation, prior to the line crew leaving the site, the foreman called in the completed installation to NEO for confirmation of cellular connectivity.

Figure 3: SMUD lineman installing one of the secondary stations in Sacramento, CA.

All 66 of the secondary stations were mounted on utility poles. Four of the primary stations were installed on tripods in SMUD substations and one was installed on the roof of California State University Sacramento less than a mile from SMUD's headquarters. Due to the height of the substation walls, some of which were cement block, two of the RSR tripods had to have extensions to raise the measurement device above shading obstructions. Even after elevating the RSRs some experienced early morning and late afternoon shading, especially in the winter months. Figure 4 shows the locations of the sensors in the network in SMUD's service territory.

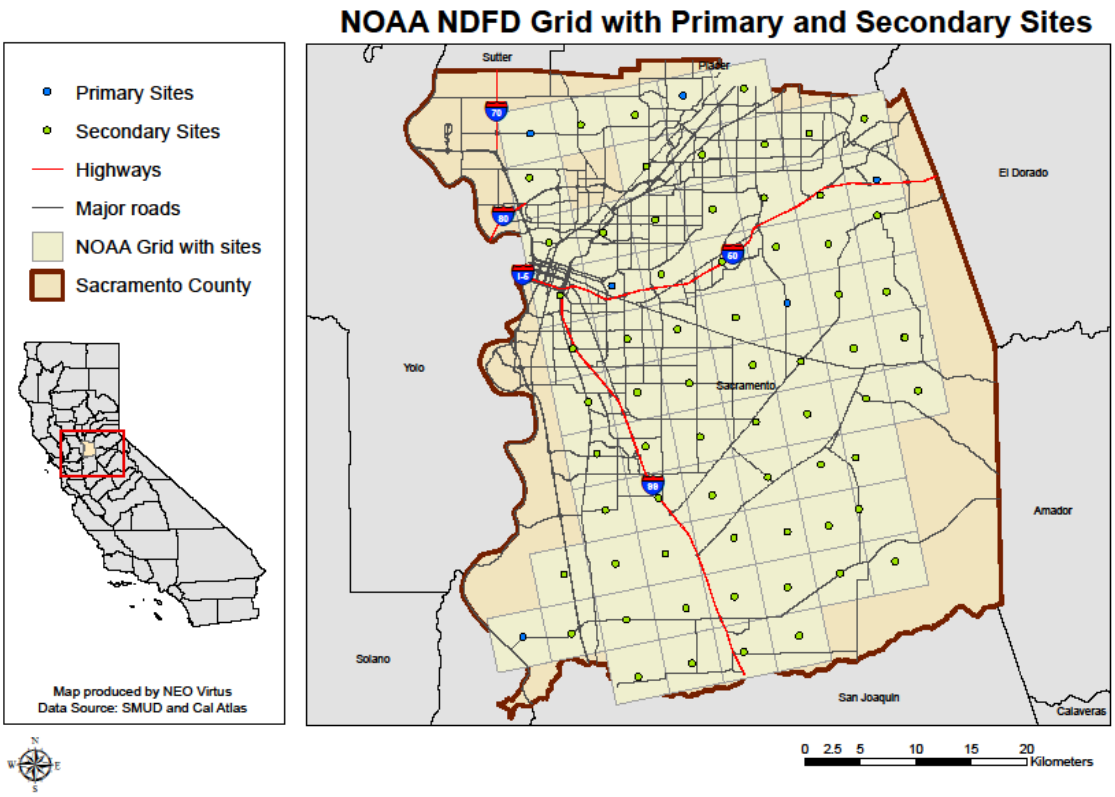


Figure 4: SMUD-NEO Irradiance Network

Data Collection

Data (irradiance and ambient temperature) is collected by connecting to each station's cellular modem. The cellular modems are programmed to turn on for one hour during the night to conserve battery power. During this nightly hourly window, Campbell Scientific Loggernet Admin software is set up to call each station, attempt connection, and if cellular connection can be made, collect data. The collection window was set up for the middle of the night, because cellular traffic is lower during this period, so the chance of connection will be greater. Should the connection fail, a second window during the day was also set up to attempt a second automatic collection. Manual collections can also be attempted during this period. Upon successful data collection, the raw data files are saved on a dedicated RAID server and are automatically pushed into a MySQL database using Loggernet Database software. NEO also maintains a redundant computer running Loggernet which connects and collects data as a backup.

Maintenance of Network

In general, the network has required very little maintenance. The RSR units collected and reported daily nearly without any gaps. The secondary stations, for the most part, also collected and reported automatically as programmed. There were, however, occasional glitches

or equipment failures that required action. Also, though generally very strong, cellular coverage for some sites was sometimes intermittent. At the end of the extended 24 month monitoring period the data set was 98.5% complete and the mean time between failure (MTBF) was 10.1 years.

To keep the network performing well NEO staff kept close watch on the data collection process. Each weekday the previous night's collection was checked for units which failed to respond. Those that did not were called and a second attempt was made to collect data during the day. Units which did not collect were logged in a tracking spreadsheet and checked the following day. The secondary station data loggers had approximately a month of data storage capacity before they began to lose data. If a unit failed to collect long enough SMUD was informed and the unit was replaced. NEO developed tools which allowed us to view multiple sensor signals simultaneously and provide a quick visual inspection, as seen in figure 5.

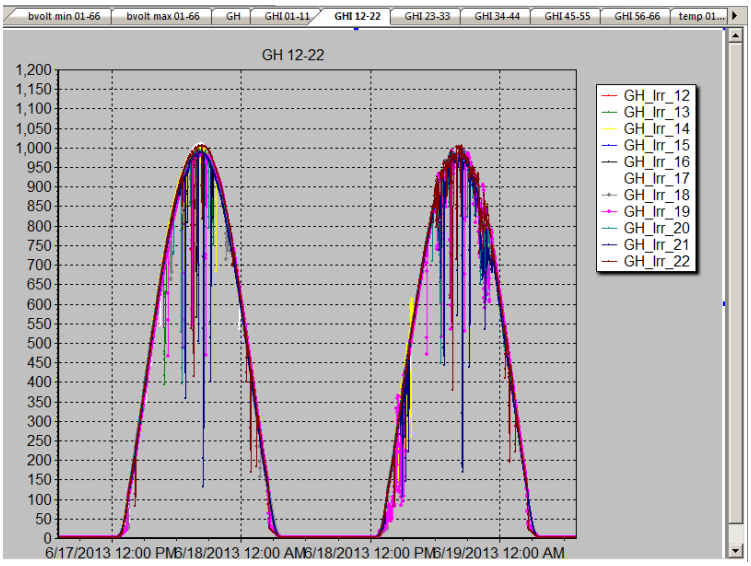


Figure 5: Loggernet RTMC data monitoring window; secondary sensors 12 to 22.

In addition to monitoring the irradiance signals NEO kept a close watch on the battery state of charge by monitoring each system's reported daily battery voltage minima and maxima. Using the Campbell Scientific data logger's RTMC application we created data windows, as shown in figure 6, which allowed our staff to view the status of all 66 secondary station batteries at a glance. This allows us to spot batteries which were beginning to discharge to unsafe levels and modify their cellular calling windows to reduce daily load.

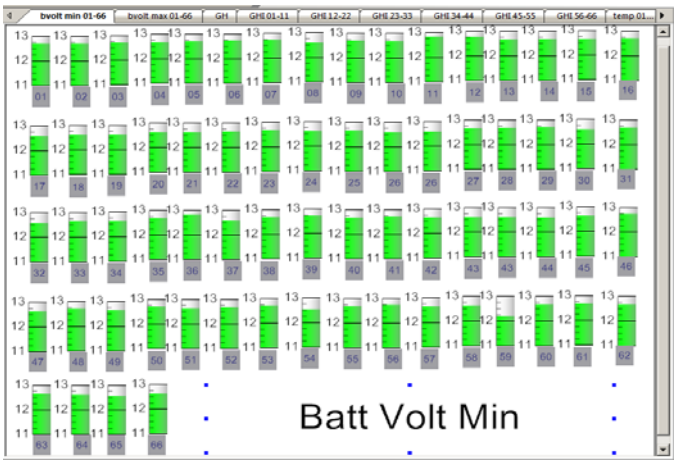


Figure 6: Battery state of charge view screen; 66 stations.

The most serious problem encountered during the first year was a software issue related to the data loggers installed in each unit (Campbell Scientific CR200). Three secondary units stopped recording data in February, 2012. Troubleshooting determined there was a failure in the flash memory of their data loggers, due to too many “writes” to a certain section of the memory. A programming fix was worked out with Campbell Scientific, and all secondary stations received a program update. NEO reprogrammed 54 secondary stations remotely and 12 units were reprogrammed/repaired/replaced in the field, completing the fix for the flash memory issue. Campbell Scientific provided warranty replacement for four failed data loggers.

Other field failures included two units with pyranometer arms that had slipped below horizontal (fixed on site), one unit with a defective solar panel junction box, and one unit with a defective data logger. Occasionally a unit would stop communicating with the loggernet program, but this was always found to be an issue with local cellular service (i.e., the modem in the unit could not be “dialed up” because of poor reception or network outage).

To ensure maximum integrity of the network, NEO Virtus built two spare secondary units in late 2011. One unit was always maintained at NEO’s offices, while the second unit was kept at SMUD’s HQ as a “hot” spare, ready to be utilized when needed. At this time, NEO also bought a customized, hard-sided shipping case (“pelican” case) to protect the units during shipment. This replacement system set up by NEO and SMUD minimized down-time for any particular network node and also maximized the chances of a unit arriving safely at its destination.

5.1.2 Data Quality Management

Early in the monitoring process we noticed anomalies in the data. During portions of the year when the solar zenith angle is relatively small the overhead wires and cross arms briefly cast shadows on the pyranometer. These shadow events present as very discrete and repeatable anomalies on clear sky days. Figure 7 has two images of global irradiance and

temperature measured by unit 64 on two clear sky days two weeks apart where characteristic shading events of the same magnitude occur at the same times of day. These shading anomalies, when viewed on clear sky days, represent a kind of signature for the monitoring unit in its unique location.

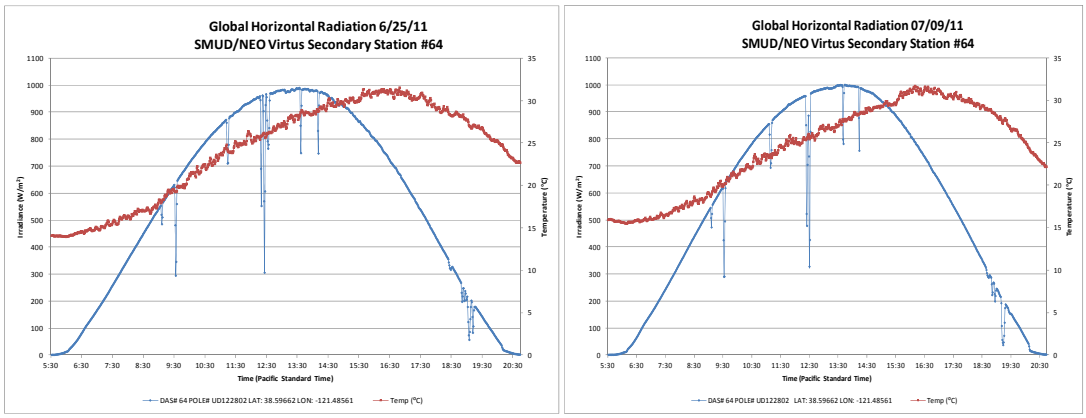


Figure 7: Irradiance signals on two clear days, two weeks apart.

Researcher at SMUD and Sandia developed a method to filter the data set for these recurring anomalies. The anomalous values were identified and the data were interpolated between the first and last "good" data point in the series. NEO used the "cleaned, interpolated data" for our own forecast validation. Both the "raw" and "cleaned, interpolated data" are available for public use on SMUD's website at ftp://ftp01.smud.org/pub/SNIData/.

During the first few months of data collection (the summer of 2011), NEO entered into discussions with SMUD regarding how we could show that the sensor network was accurately measuring irradiance. One practical limitation of the secondary sensor network is that the sensors are widely distributed and require a bucket truck to access. Thus, they would have been extremely difficult to keep clean using a typical two-week schedule. We decided that adding a small set of higher accuracy sensors to measure irradiance independently of the network would help verify the calibration of the installed pyranometers and hence the quality of overall network irradiance data.

These additional irradiance sensors were installed at four of the primary stations. Their data was incorporated into the database as part of a calibration plan for the network, and also into the distribution of daily updates of the dataset. For the four sites in question Eppley Laboratory's Precision Spectral Pyranometers (PSP), as seen in figure 8, were installed in global horizontal orientation and their signals added to the data stream. NEO provided two PSPs and SMUD provided two PSPs, all of which were calibrated at NREL using their Broadband Outdoor Radiometer Calibration (BORCAL) procedure prior to deployment.

Figure 8: One of the four RSRs with a PSP installed.

These four PSPs are distributed around SMUD's service territory, approximately at the compass points. The four RSR locations with PSPs are Anatolia, Don Julio and Twin Cities substations and on the roof of CSUS. Note that the RSR at Anatolia substation was deployed by NREL at an earlier date. NEO accesses the data from Anatolia via NREL's website. The four PSPs are maintained with an approximate two week cleaning schedule.

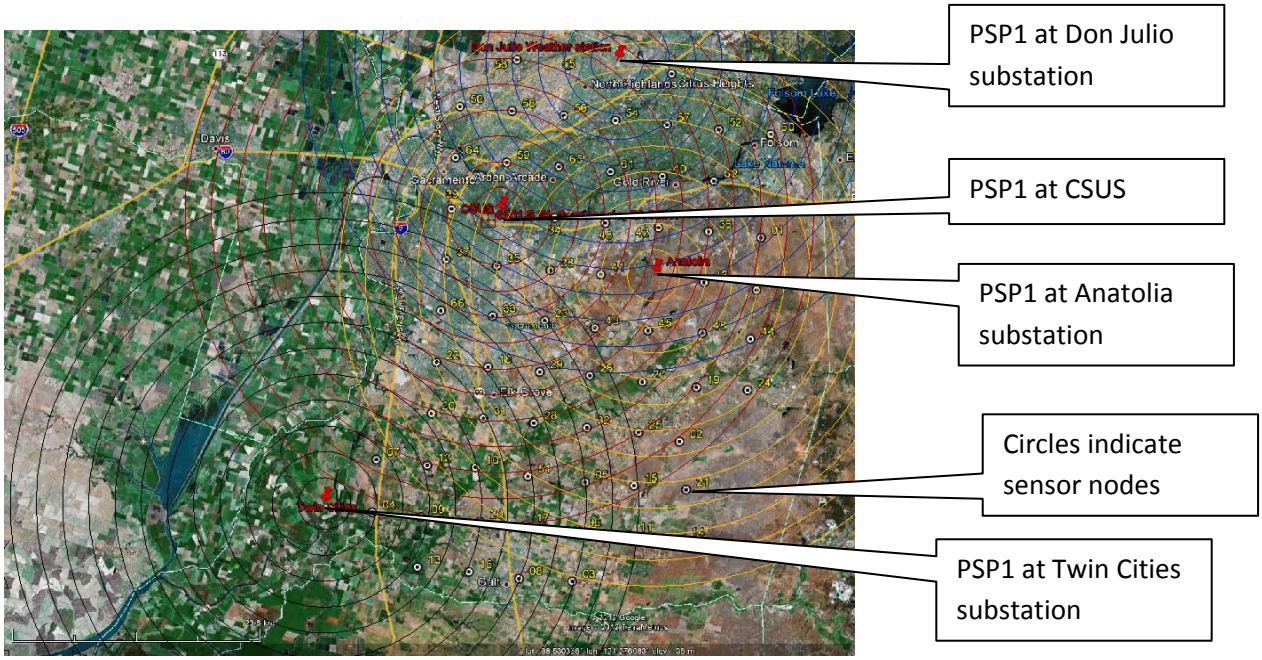


Figure 9: Locations of four Eppley PSPs.

5.1.3 Additional scope of work

During the course of this project we learned things which we had not anticipated when we crafted the original proposal. The four PSPs used to provide global calibration of the network are just one such example which, once we'd deployed the network, made sense to do and thus became an addition to our original scope.

Another development which had not been anticipated in the original proposal was the need by SMUD for enhanced power systems on three of the five primary stations. For three of the four primary units which were installed in SMUD substations NEO provided 110W PV power system (factory default is 20W) to provide additional power for a SMUD spread spectrum modem which ran 24 hours per day seven days per week and which delivered RSR data to SMUD substation SCADA. The RSR on the roof of CSUS was provided with an AC power system and a direct Ethernet connection to CSUS's LAN. This added connectivity provided a backup data path in case NEO's cellular modem failed.

Another modification which was not anticipated by the original proposal was the programming necessary for daily network data uploads on a site by site basis. NEO transmitted the first three months of collected data to SMUD in the fall of 2011, and then sent a dataset including the first six months of data in late 2011. However, handling the data in such large sets was somewhat difficult, and did not give SMUD engineers access to collected data in a timely fashion. Therefore, SMUD asked NEO to devise a method so that SMUD would be able to access the collected data in near real time. NEO found a way to transmit the data to SMUD via daily ftp upload, using Campbell Baler software to parse the large raw data files into 24 hour chunks which are automatically uploaded to the SMUD ftp site using WinSCP software.

In June of 2012, at the request of the energy trading department at SMUD, NEO began delivery of daily forecasts of feed in tariff (FIT) system power production. We provided hourly forecasts of plant production in a window 22 to 40 hours into the future, delivering the data at about 5:20am Pacific local time for Bruceville, Kost, Boessow and Point Pleasant. Due to SMUD IT security concerns the forecasts are delivered as email attachments rather than other, more direct methods, such as writing directly to their server.

Despite the present email arrangement with SMUD NEO has looked at other methods of data delivery and end user access to make these forecast services more widely available, flexible and user friendly. We evaluated the possibility of designing a scalable online solution with cloud based computing and data storage (presently all programs run on NEO server.) We created a Beta version of a graphical user interface (see Figure 10) which would provide end users with the ability to selectively access forecasts of irradiance and plant production for specific ranges of dates in the future, daily time frames and locations. The tool would provide both numeric file and graphical reporting options.

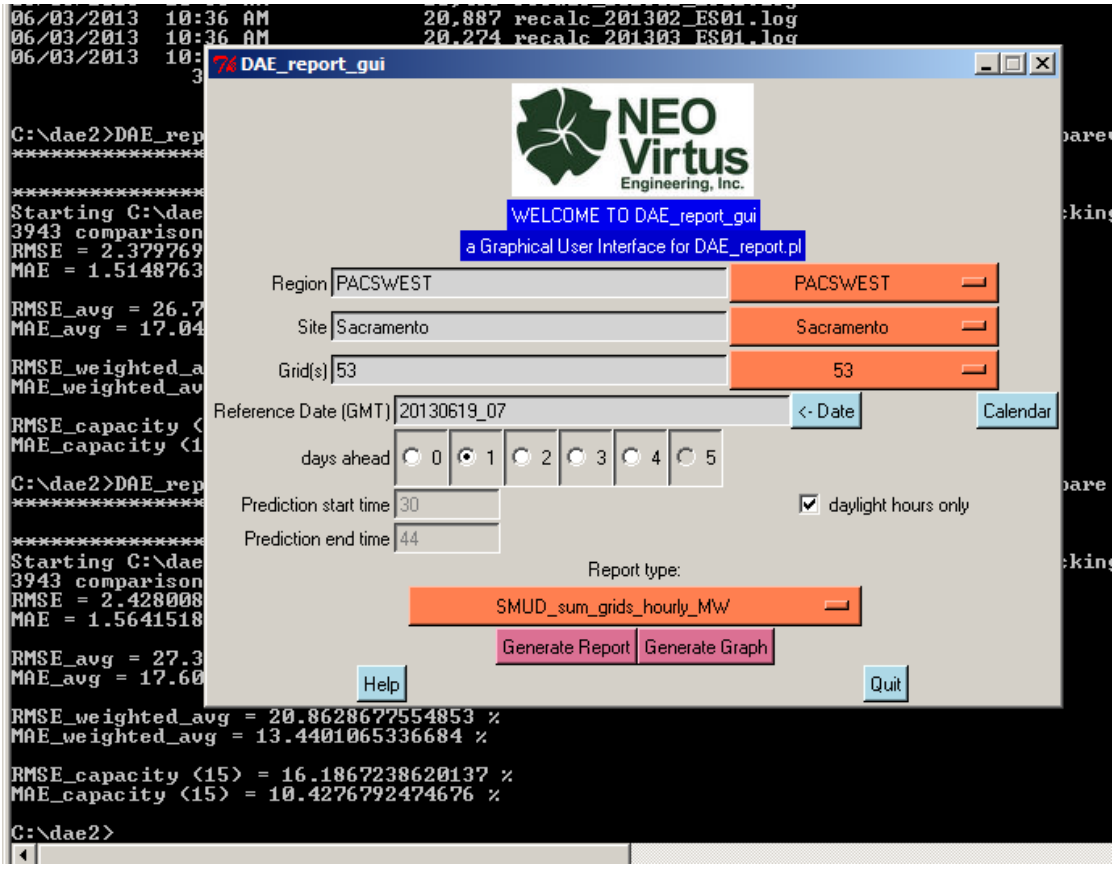


Figure 10: NEO solar forecast graphical user interface (GUI).

5.1.4 Summary

NEO designed and built the required 66 secondary stations, assembled the five primary stations (RSRs), and shipped all 71 units to Sacramento, where over the course of several weeks, they were all installed by SMUD linemen. NEO also developed an automated data retrieval system, created a MySQL database to store the collected data, and uploaded daily updates of the data to SMUD's ftp site.

This network has been up and running since June 2011. Though originally tasked with collecting a 12 month dataset, by virtue of the contract extension, NEO has collected a total of 24 months as of June of 2013. We also collected irradiance data from four PSPs and used this data to help verify the calibration of the network sensors.

SMUD plans to maintain the monitoring network for an additional two years with NEO Virtus Engineering providing the sensor maintenance and data collection. The resulting data set continues to be of value for validation of a range of forecasting technologies from numerical weather prediction models, to satellite imagery, to sky camera approaches. SMUD has used parts of the network for validation of battery dispatch in the community of Rancho Cordova. Further SMUD is using the data set in an evaluation of trials of a number of different solar forecast products. <<<Elaine, do you want to review this. Finally as a sub-recipient of the

National Center for Atmospheric Research's grant under the DOE FOA-0000649, "Improving the Accuracy of Solar Forecasting," SMUD is providing ground truth irradiance data from the network for improving accuracy of solar forecasting in the short-term (0-6 hours) and day-ahead timeframes.

Beyond that time frame the future use of the secondary sensors has many possibilities. Of the five analog inputs on the CR200 data logger only two are presently in use meaning that the units could be repurposed for a wide range of applications in solar resource assessment or other fields.

Sub-task 5.2: Develop 20-36 Hour Feed-Forward PV Forecasting Model

The objectives of this subtask were to forecast regional irradiance and PV production of SMUD's feed in tariff (FIT) PV power plants 20 – 36 hours in the future, i.e., one day ahead. Such forecasts could provide information for making generation resource scheduling and purchase and sales decisions in the next-day energy market. The production data from the FIT sites and the installed sensor network (described in 5.1 above) provided measured power and irradiance data needed to evaluate the model's performance. This task was broken down into several steps, including development of the irradiance transmittance model, an irradiance surface model, a photovoltaic generation model, and the integration of all of these into a 20-36 hour ahead PV production forecast model to calculate FIT plant power output. All of these steps were preceded by the creation of relational database developed to store the fetched NDFD meteorological forecast data. That same relational database became the home for the FIT site parameters, the inverter parameters and the Sandia PV module database used by the forecast model.

5.2.1 REST2 Transmittance Model

The goal was to use the National Weather Service's National Digital Forecast Database (NDFD) meteorological forecasts to predict the solar irradiance incident on various PV systems from 20 to 36 hours after forecast issuance. NEO contracted with Solar Consulting Services (SCS) for this subtask. SCS' work was accomplished by Dr. Chris Gueymard, a well-known expert in this field. For this task, he used his high-performance REST2 clear-sky radiation model (Gueymard, 2008) and made various modifications, additions and improvements, as described below. REST2 calculates clear-sky radiation from site coordinates and atmospheric data.

A study was made to obtain the mean daily values of the required aerosol parameters. These were derived from satellite data, after bias correction using available ground observations in the general area. At this point, there are no reliable forecasts of hourly aerosol optical depth variables at a sufficiently high spatial resolution, but this should not be a big issue over Sacramento.

Another study was conducted to evaluate the water vapor concentration of the atmosphere (in terms of precipitable water, PW). Three local GPS stations of the NOAA network were used: Lincoln, McClellan AFB, and Dixon. A statistical relationship was found between the measured PW and the measured vapor pressure, itself derived from temperature and relative humidity through well-established physical equations. The latter two quantities are provided by NDFD, and can therefore be used to derive the necessary PW forecasts.

Reasonable monthly values of the other necessary inputs to REST2 (ozone, nitrogen dioxide and surface albedo) were determined. A validation exercise to evaluate the performance of the hourly irradiance predictions was conducted at Anatolia and showed good results.

The main study was devoted to relating the NDFD cloud fraction forecasts (at unfortunately a low temporal resolution—only a 6-hour refresh rate) to the cloud attenuation evaluated at the Anatolia RSR station by dividing the observed GHI and DNI by the calculated clear-sky values. Cloud transmittance functions were obtained for various classes of cloud fraction and solar position, with or without precipitation. It was observed that the forecasts of fog or marine layer were not correct, unfortunately (out of phase or simply missing). The evaluation of DNI under cloudy conditions was improved through an optimal combination of (i) the prediction derived by modifying the ideal clear-sky value by a cloud transmittance function; and (ii) an independent prediction based on the popular Erbs correlation that provides the diffuse/global ratio through statistical means (Erbs et al., 1982). Compared to the good results of the clear-sky predictions, the cloudy-sky situations showed large random errors, which can be explained by the insufficient refresh rate (6 hours) of the cloud forecasts, the inaccuracy of the forecasted cloud fraction under many circumstances (e.g., fog and marine layer), and the lack of additional information about layer-by-layer cloud type or cloud optical depth.

A similar study was conducted to derive the necessary cloud cover correction factor, using a similar approach, to provide day-ahead GHI, DNI and GTI forecasts close to the Honolulu area. Measured irradiance data from NREL's RSR station at Honolulu's airport were used to calibrate the model. Even more difficulty with the NDFD cloud forecasts was found there, compared to Sacramento. It is probable that the coastal location of the station, and that the typical local cloud regime (scattered cumulus clouds) adds to the difficulty.

5.2.2 PV Surface Irradiance Model

An algorithm (Gueymard, 1987) was added to the one developed in 5.2.1 above so that the global tilted irradiance (GTI)—also referred to as plane-of-array irradiance or PV surface irradiance—could be evaluated (modeled) each hour. This algorithm is normally applied to the most frequent case of fixed-tilt PV systems. A modification was introduced to make it also accept the case of single-axis tracking systems, with or without backtracking.

5.2.3 Photovoltaic Generation Model Integration

NEO's approach to calculating PV plant output power was accomplished by creating three main program components:

1. The main program.
2. An executable version of the REST2 transmittance model code.
3. The relational database with all relevant system data.

Main Program

The main program, DAE_analyze.pl, performs all of the calling functions and runs the Sandia Photovoltaic Array Model. DAE_analyze also calls the compiled executable version of REST2 and hands it the forecast meteorological data and PV plant data it needs to return a value of global tilt irradiance (GTI). In addition DAE_analyze provides all of the routines which fetch, parse and store to the database the necessary forecast meteorological data from the NDFD. Each of these items was developed separately and then merged. The Sandia Photovoltaic Array Model is the core computation for this program.

REST2 Executable

The REST2 transmittance model and the PV surface irradiance models described in 5.2.1 and 5.2.2 above were written in Fortran by Chris Gueymard and then compiled to an executable form by Bruce McArdle, the author of the main program. When run, the executable version of REST2 calculates the direct normal, diffuse horizontal, and global horizontal components of irradiance at the specified latitude and longitude based upon the forecast meteorological parameters it has been given. Those general forecast values of irradiance are then translated by REST2 into values of global tilt irradiance. Global tilt irradiance is the irradiance incident upon a specific array module surface, whether it is a fixed tilt or tracking surface. The program also calculates the effect of near field shading and array self shading. Originally the solar geometry functions were planned to be a separate routine, however Dr. Gueymard chose to include them in REST2.

Relational Database

We chose to use the open source MySQL relational database for this project based on its capabilities and cost. The MySQL database serves as the repository for the Sandia Module database, NEO's inverter database and the database of SMUD FIT system characteristics. The database was first used to store the meteorological parameters used by the transmittance model, REST2, retrieved from the National Digital Forecast Database. Principle amongst these parameters are skycover, temperature and probability of precipitation. The Sandia Module database is a listing of some of the thousands of commercially available PV modules which have been tested and characterized by Sandia. NEO used similar modules in the database to model modules installed at the various FIT sites. NEO's inverter database contains a list of inverters used at the SMUD FIT sites with DC/AC conversion efficiencies expressed as a

function of percentage of full load. The MySQL database also contains a table listing site data for all of the SMUD FIT PV plants. The parameters include site latitude, longitude, elevation, module and inverter make and model, array tilt, azimuth and tracking/no tracking status.

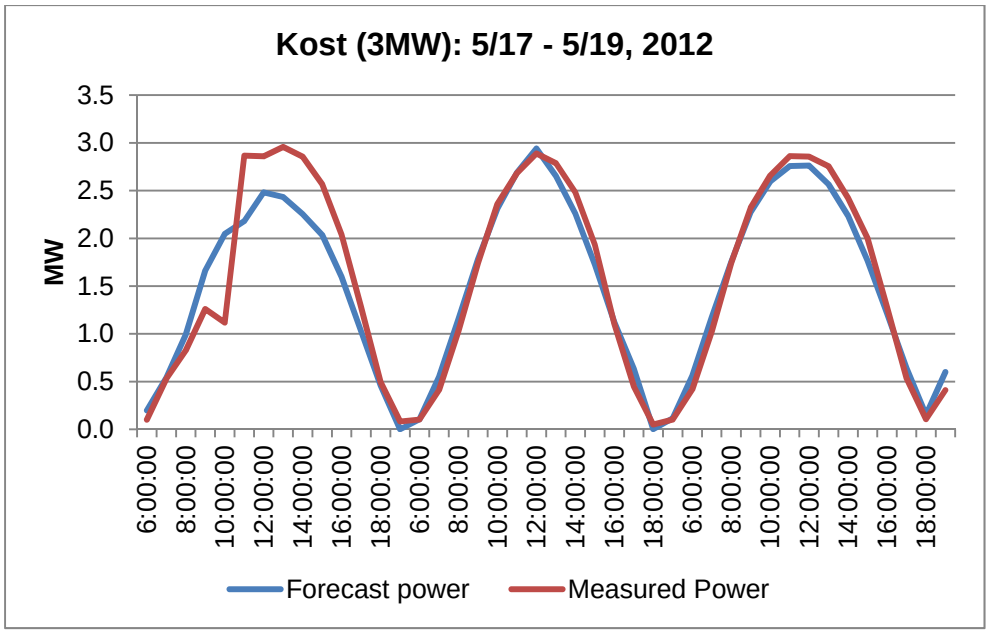
5.2.4 20 – 36 hour Feed-Forward PV Forecasting Model Integration

Each of the sub-models and the MySQL database functions described above were developed and tested independently and then, in some cases, tested together in a "piece-wise" fashion. To better understand the Sandia PV Array Model NEO consultant Robert Williams coded it in a Microsoft Excel spreadsheet and ran it a step at a time. However before the entire process could be run end to end using its separate components the main routine, DAE_analyze, was completed and the process naturally transitioned to full automation. The full process starts with DAE_fetch retrieving the forecast met data from NDFD, decoding those data from their gridded binary format to a csv format, filtering the data for the geographic subset of sites included in the sensor network, and then pushing the forecast data into MySQL. Next the program calls the REST2 executable, queries the database for forecast met data, site data, module and inverter data. Once the REST2 transmittance model has delivered a value of GTI the program calculates the instantaneous power, based upon a one hour time increment, then pushes this to the database.

One of the many challenges associated with this project was the proliferation of reference times. The end user, the SMUD Energy Trading department, operates in local Pacific time. This of course means that they use Day Light Saving time when it applies. All of the NDFD references were to Greenwich Mean Time. REST2 uses local (Pacific) Standard Time throughout the year. The calculation done in NEO's Massachusetts office occurred at Eastern time, with Day Light Saving time and a three hour difference from the end user in California.

The original intent had been to do all of the day ahead, 20-36 hour forecasting, in a post processing "hind cast" mode. In this mode historical NDFD forecasts would be retrieved from the National Climactic Data Center (NCDC) and those data would then be used to generate "historical" forecasts which would then be tested against historical FIT PV plant production data and against the historical measured data global horizontal irradiance date from the SMUD-NEO sensor network.

However as the end of 2011 approached and several of SMUD's FIT sites neared completion NEO was asked to provide a Beta version of the real time forecast. We were told that there was a need in SMUD's energy trading department for a forecast of the next day's contribution by the FIT sites. On June 5, 2012 NEO began sending forecast of the production for Bruceville, Kost, Boessow and Point Pleasant, a total of 30MW capacity, to SMUD. SMUD's energy trading schedule required that they receive the forecast at 5:00am Pacific time. To accommodate this schedule we adjusted our forecast window to 22 to 40 hours into the future. The final forecasts of this project, which were done for the purposes of model performance validation, were done as hindcast calculations using the 22 to 40 hour forecast window and the latest improvements to the component models. The time frame for the validation set of forecasts was from May 1, 2012 to April 30, 2013. Figure 11 illustrates three days comparing forecast and measured production at two of the FIT sites.



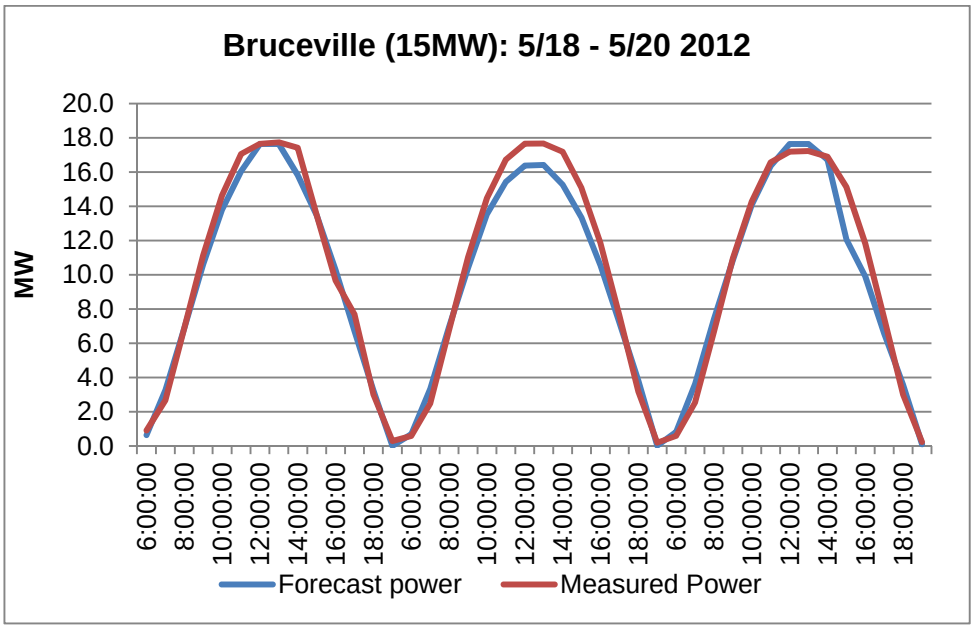


Figure 11: Sample from two sites, comparing forecast and measured production on mostly clear days. Note that hours of darkness have been eliminated, since predicting "night" output is unnecessary.

5.2.7 Summary

Under subtask 5.2 of this project NEO Virtus Engineering and its consultants successfully implemented the irradiance and PV power production forecast methodology described in the project scope of work. Our efforts resulted in PV production forecasts in the day-ahead time frame which have been delivered to SMUD on a daily basis since the beginning of June, 2012, and which were then validated in section 5.4 of this effort.

Sub-task 5.3: Develop 1-3 Hour Feed-Forward PV Forecasting Model

The objective of this subtask was to forecast regional irradiance and PV production of SMUD's feed in tariff (FIT) PV power plants 1-3 hours out, i.e., in the short term, using the global horizontal data from the sensing network. The time frame was initially described as a "1-3 hour" forecast, in order to describe the time period in the first hour ahead and the first three hours ahead. Because this was the source of some confusion we changed the description of the 5.3 subtask to a "0-3 hour" forecast (henceforth referred to here as "0-3 hours".)

5.3.1 Develop 0–3 Hour Feed-Forward PV Forecasting

Ground Level Irradiance Velocity Vector Recognition

The approach to short term sensor based irradiance forecasting which we proposed in our initial scope of work was one in which we would use all of the measured irradiance data from the monitoring network at the same time. The basic concept for forecasting irradiance based upon ground sensor measurements was one of pattern recognition similar to that described in Bosch et al under the concept name, Most Correlated Pair or "MCP" but with the benefit of a much larger array of sensors. The process entails synthesizing an image of irradiance levels at an instant in time comprised of global horizontal irradiance (GHI) levels measured by our network of sensors. This irradiance image would be like an image created by a black and white digital camera using a charge coupled device (CCD) where each value of irradiance reported from the sensor network is similar to a light level in a pixel of a CCD element. As illustrated in figure 12, the first image, I, at time t1 (It1) in a daily cycle is taken shortly after sunrise because this process is only run during the daylight hours. At some time after It1, for example one minute, which is the equivalent of one record, the program would synthesize a second image, It2, from global horizontal irradiance levels measured by our network. After synthesizing It2, the program would then compare It1 to It2. Using an approach similar to the "Most Correlated Pair" method, we would derive a velocity vector that represents the displacement in time and distance between It1 and It2. Using the synthesized irradiance image It2, in combination with the image velocity vector, it would then forecast a future irradiance image by estimating the location of irradiance image It2, at time t3. This forecast irradiance image, FIt3, is the "result" which we are seeking. This forecast image of irradiance will be transformed into a forecast of ac power generated by solar PV systems in the area of the sensor network.

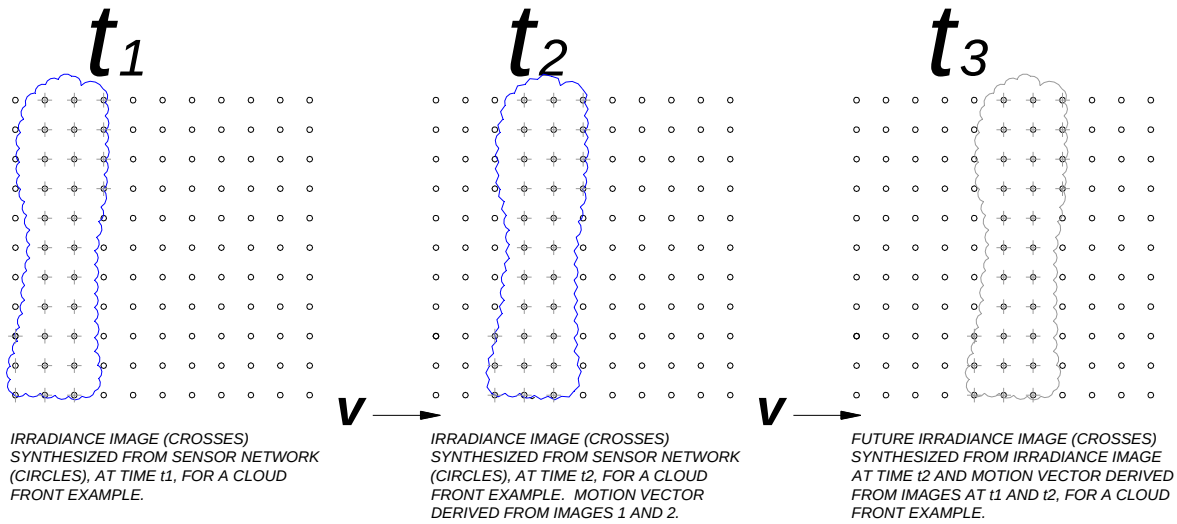


Figure 12: Example of initial methodology for 0 – 3 hour irradiance forecast: Ground Level Irradiance Velocity Vector Recognition

Unfortunately, for the majority of the network wide GHI data we examined there was no discernible, coherent pattern in the changing images of ground level irradiance on days with intermittent conditions. The work we did based on this approach indicated that forecasts will probably result in very high error in many instances. We concluded that the reasons for this were that the network is too small geographically, and the spacing between sensors too large, in relation to cloud speed and cloud size, to allow patterns in ground level irradiance to provide information about the sensor area farther than about an hour into the future. Put another way, if a cloud pattern is traveling from west to east and it takes a single cloud feature an hour to travel from the western edge to the eastern edge, then the furthest into the future a model which relies solely on sensor measurements can forecast a ground level irradiance state is about an hour (and that would be for a location on the eastern side of the sensor network.) In addition, there will be no sensor data history for that day on which to build the forecast model output for the early morning hours. Finally, and possibly most significantly, we believe that our network is spatially too course, i.e. 5km between sensors is too great a distance, to implement our originally intended method. In our attempts to implement this process we were able to find only one day --June 22, 2012-- where visual pattern recognition was possible.

Neural-Network Forecasting Approach

After concluding that our initial approach to short term forecasting based upon ground level irradiance sensing was unworkable with our sensor network, our consultants at Wichita State University (WSU), led by Dr. Yanwu Ding, recast the problem as a neural-network approach. More systematically, WSU applied a multi-layered neural-network model, as shown in Figure 13, which has three hidden layers, input, and output layers, to generalize data from a sensor array so that the irradiance at any point can be predicted by the model automatically through the network training process. The input of the network can be historical GHI data or exogenous data such as temperature, humidity, DNI, and DHI, depending on the locations of

the sensors in the grid. Since the exogenous data is only available at the primary sensors, to predict the GHI at the secondary sensors that are far away (beyond 10 or 15km) from any of the primary sensors, the inputs are historical GHI data: from the sensor to be predicted and from nearby secondary sensors. On the other hand, to predict the GHI at the sensors that are close (within 15km) to a primary sensor, the inputs include the following historical data: GHI from the sensor to be predicted, GHI from nearby secondary sensors, and exogenous data from the nearby primary sensor. The studies by WSU suggest that the exogenous data can help improve the performance of prediction.

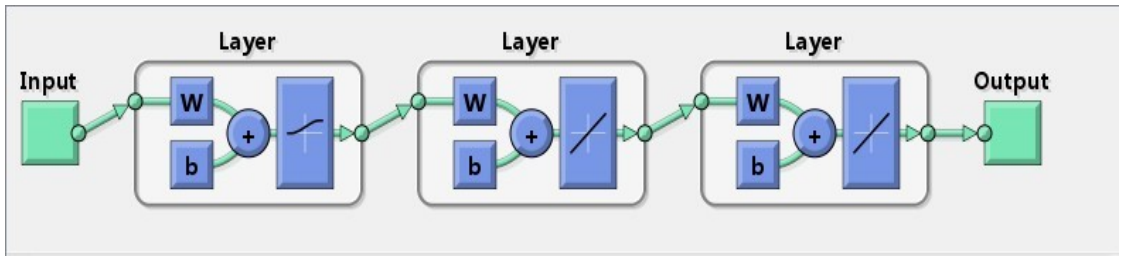


Figure 13: Diagram of neural-network applied in the prediction

Figure 14 plots the 3-hours ahead GHI predictions at Sensor 10 on 9/19/12, which is a clear day with little cloud. The forecast is performed on an interval of 15 minute increment. The inputs are the 3-hours historical data: GHI from the sensor 10 (the sensor to be predicted), GHI from sensors 9, 10, and 12, and exogenous data from primary sensor 4. Specifically, prediction of GHI at time 6:14am uses historical data at time 3:14am; prediction of GHI at 6:29am uses data at time 3:29am, etc. Figure 14 indicates that results from the neural-network prediction model show a good agreement to the measured GHI values for a clear day.

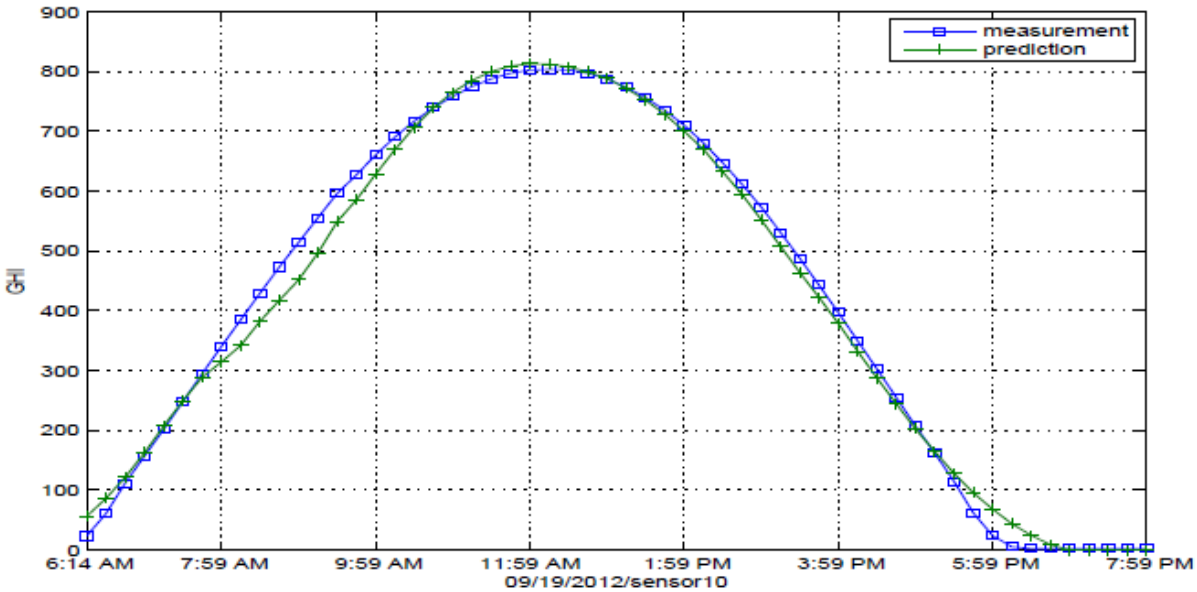


Figure 14: Output from neural net 0 – 3 hour GHI forecast compared with measured GHI on a clear day.

Figure 15 plots the 3-hours ahead GHI predictions at Sensor 10 on 9/21/12, which is a day with intermittent clouds. As can be seen from the plot the neural network prediction does less well under cloudy conditions.

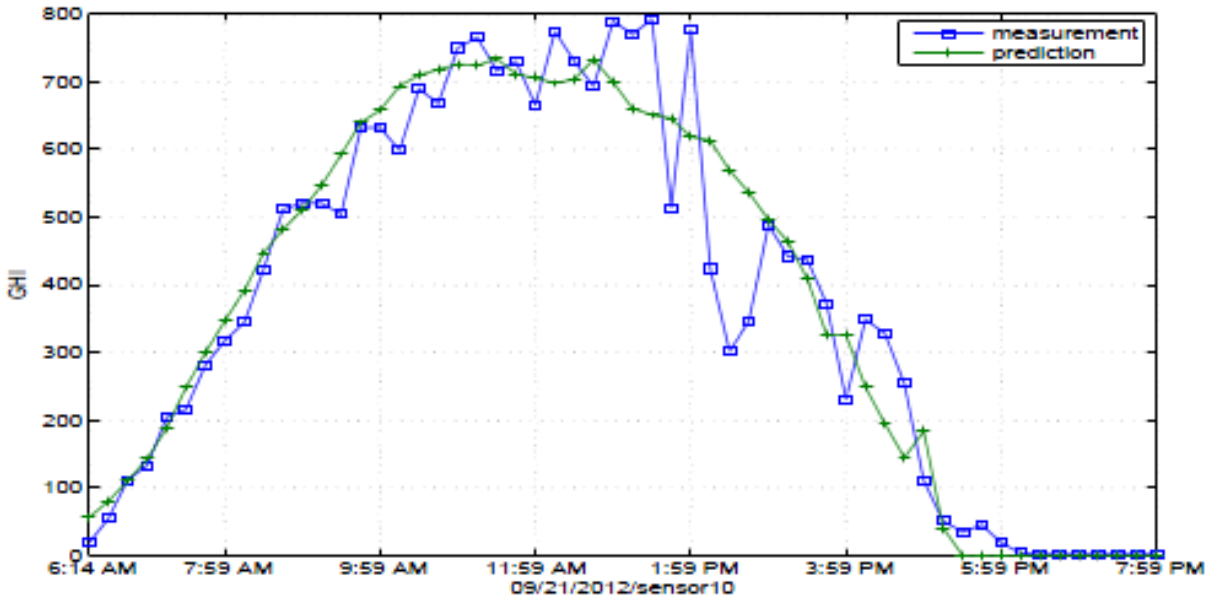


Figure 15: Output from neural net 0 – 3 hour GHI forecast compared with measured GHI on a cloudy day.

5.3.2 0-3 Hour Feed-Forward PV Production Forecasting Model Integration

Once the short term forecasts of GHI were complete they were converted, using a modification of the model described in 5.2.2 above created by SCS, to GTI, without preliminary knowledge of DNI. These values of GTI were in turn used by DAE_analyze to create forecast of PV plant power as in 5.2.3.

5.3.3 Summary

For our 0-3 hour ahead forecasting subtask, after several attempts, we concluded that our initial approach using ground level irradiance velocity vector recognition was nearly entirely unworkable with the 5km sensor grid for which we had irradiance data. The neural network approach shows some promising results.

Sub-task 5.4: Validate Irradiance/PV Production/Load Forecast Models

This final task assigned to NEO culminates our work on this project. Here, we compare our forecasts of irradiance with actual irradiance measurements, collected by the network of sensors installed in subtask 5.1. We also compare our forecasts for power production of SMUD's FIT PV power plants with their actual production, as recorded by SMUD's meters. Load forecasting was not accomplished as SMUD was not able to provide feeder-level load data. Acceptable margins of error for irradiance and power forecasts are still TBD (the DOE is funding research work on this topic—see DE-FOA-0000649: IMPROVING THE ACCURACY OF SOLAR FORECASTING). To validate our forecasts, therefore, NEO turned to error analysis, using metrics recommended by other experts in the solar forecasting field. The resulting values of those error metrics indicate that our day-ahead forecasts, especially for irradiance, were more accurate than our 0 – 3 hour forecasts.

5.4.1 Error analysis of 20-36 hour regional irradiance forecast model

In their paper, "Reporting of Relative Irradiance Prediction Dispersion Error", Hoff et al. determined that the best metrics to use when analyzing the error of an irradiance forecast were Root Mean Square Error (RMSE) over Rated Capacity and Mean Absolute Error (MAE) over average. In the case of irradiance, "capacity" is given as 1000 W/m^2, while average would be the average irradiance over the time interval (typically an hour). The formulas for calculating these metrics are:

$$\frac{RMSE}{Capacity} = \frac{\sqrt{\frac{\sum_1^N \left(I_t^{test} - I_t^{ref}\right)^2}{N}}}{1000}$$

where I_t^{test} is the forecast irradiance at time t, I_t^{ref} is the measured irradiance at time t, and N is the number of observations.

$$\frac{MAE}{Average} = \left(\frac{1}{\sum_{t=1}^N I_t^{ref}}\right) \sum_{t=1}^{N} \left| I_t^{test} - I_t^{ref} \right|$$

NEO calculated the above metrics using measured and forecast data from 6/1/2012 through 12/12/12. This was the time period for which we had "clean, interpolated" measured irradiance data. "Raw" measured data contained artifacts caused by short-term shadows cast by overhead lines and crosstrees. SMUD and Sandia developed a program to find these shadows, and replace the measured irradiance with an interpolated value.

NEO used the cleaned, interpolated measured irradiance values for each of the 66 secondary stations, averaging the 1-minute data from 30 minutes prior to the hour to 30 minutes after the hour, to create hourly measured Global Horizontal Irradiance (GHI) values in W/m^2. Also, we eliminated hours in the early morning and late afternoon when measured GHI was less than 5 W/m^2, as irradiance measurements at those low sun angles suffer from greater inaccuracy because of the "cosine" effect. We then calculated the various error metrics using these hourly values of measured GHI and the hourly GHI forecasts produced by our 20 -36 hour model developed in task 5.2. Table 1 provides the results for each sensor plus the average metrics for all the sensors.

Table 1: 20 – 36 hour irradiance forecast error metrics

Sensor #	MAE (W/m^2)	MAE/AVERAGE	RMSE (W/m^2)	RMSE/1000	MAE/1000
1	63.63	13.1%	123.62	12.4%	6.4%
2	65.46	14.3%	108.83	10.9%	6.5%
3	56.28	11.9%	105.38	10.5%	5.6%
4	47.12	9.0%	94.10	9.4%	4.7%
5	57.14	12.0%	106.65	10.7%	5.7%
6	55.94	11.8%	106.70	10.7%	5.6%
7	67.69	14.7%	117.64	11.8%	6.8%
8	55.38	11.6%	105.24	10.5%	5.5%
9	78.54	15.1%	114.66	11.5%	7.9%
10	55.99	11.9%	105.99	10.6%	5.6%
11	58.54	12.2%	110.90	11.1%	5.9%
12	56.47	11.7%	106.69	10.7%	5.6%
13	66.21	14.6%	122.44	12.2%	6.6%
14	55.95	11.6%	104.47	10.4%	5.6%
15	59.11	12.6%	110.94	11.1%	5.9%
16	57.78	12.2%	106.80	10.7%	5.8%
17	85.06	16.6%	136.07	13.6%	8.5%
19	62.36	13.3%	111.53	11.2%	6.2%
20	57.58	12.2%	109.26	10.9%	5.8%
21	58.64	12.3%	106.91	10.7%	5.9%
22	59.59	12.7%	111.24	11.1%	6.0%
23	60.63	12.8%	114.05	11.4%	6.1%
24	64.12	13.5%	115.52	11.6%	6.4%
25	60.64	12.7%	109.11	10.9%	6.1%
26	61.70	13.2%	110.97	11.1%	6.2%
27	57.77	12.1%	107.87	10.8%	5.8%
28	98.70	23.4%	164.26	16.4%	9.9%
29	74.20	15.9%	128.89	12.9%	7.4%
30	55.72	11.9%	102.38	10.2%	5.6%

31	57.63	12.1%	106.52	10.7%	5.8%
32	64.97	12.9%	112.46	11.2%	6.5%
33	59.17	12.7%	110.28	11.0%	5.9%
34	71.22	16.1%	117.36	11.7%	7.1%
35	59.24	12.5%	109.80	11.0%	5.9%
36	69.65	13.9%	118.99	11.9%	7.0%
37	62.34	13.5%	110.79	11.1%	6.2%
38	56.86	12.0%	106.61	10.7%	5.7%
39	58.64	12.3%	106.96	10.7%	5.9%
40	56.79	11.8%	106.34	10.6%	5.7%
41	57.23	12.0%	106.80	10.7%	5.7%
42	58.38	11.9%	108.38	10.8%	5.8%
43	56.02	11.7%	104.29	10.4%	5.6%
44	61.09	12.8%	113.71	11.4%	6.1%
45	60.17	12.8%	109.84	11.0%	6.0%
46	58.99	12.6%	105.54	10.6%	5.9%
47	62.27	13.1%	117.22	11.7%	6.2%
48	63.70	13.8%	112.30	11.2%	6.4%
49	54.36	11.4%	102.09	10.2%	5.4%
50	56.09	12.0%	105.75	10.6%	5.6%
51	56.91	12.0%	109.34	10.9%	5.7%
52	62.10	13.2%	111.08	11.1%	6.2%
53	54.63	11.5%	103.93	10.4%	5.5%
54	59.04	12.6%	109.39	10.9%	5.9%
55	58.54	12.4%	109.57	11.0%	5.9%
56	56.51	12.0%	109.28	10.9%	5.7%
57	59.76	12.6%	109.26	10.9%	6.0%
58	58.25	12.3%	110.16	11.0%	5.8%
59	54.48	11.5%	104.83	10.5%	5.4%
60	62.68	13.4%	110.75	11.1%	6.3%
61	66.32	14.3%	120.16	12.0%	6.6%
62	59.06	12.5%	109.42	10.9%	5.9%
63	64.99	14.4%	110.08	11.0%	6.5%
64	57.77	12.5%	104.02	10.4%	5.8%
65	72.32	14.4%	116.90	11.7%	7.2%
66	60.09	12.7%	111.11	11.1%	6.0%
	MAE (W/m^2)	**MAE/AVERAGE**	**RMSE (W/m^2)**	**RMSE/1000**	**MAE/1000**
Average:	**61.26**	**12.9%**	**111.24**	**11.1%**	**6.1%**

To provide a context for these error metrics we compared the performance of our day ahead forecast with a persistence model of day ahead forecasting[2]. We used a persistence model because it was a means by which we could get a second forecast for the same sites, with the same forecast horizon and with the same averaging interval. The persistence model we used was based upon the measured data from the sensor network and was expressed in kWh/m^2/day. To do this we used sensors which we knew had complete uninterrupted data for the same six month period as we had used for the previous analysis[3] in section 5.4.1. For this persistence model total measured daily kWh/m^2 were used as a predictor for total daily kWh/m^2 on the subsequent day. For each sensor we summed the daytime hour values in each 24 hour period over the course of the selected 6 month sample period. We then divided the sum of one minute measured values by 60. The resulting value is the number of Wh/m^2 for the day in question. We divided this value by 1000 to get kWhr/m^2 or "sun hours." The integral of all of the sun hours in a 24 hour period is referred to as equivalent sun hours per day or "ESH/day." Using this method we calculated a day ahead forecast of ESH for the latitude and longitude for each sensor based on the previous day's ESH for that sensor. From this persistence calculation the error metrics (RMSE, MAE, RMSE/average, MAE/average, RMSE/capacity, MAE/capacity) were calculated by comparing the persistence model prediction of ESH to the measured ESH at each sensor site. For a value for capacity we took the maximum global horizontal ESH for Sacramento of 7.9 kWhr/m^2, as reported in the NREL Red Book.

In a similar manner we calculated the ESH for each day of NEO's day ahead forecast which was created using NDFD data. From NEO's day ahead model the same error metrics were calculated by comparing the NEO model prediction of ESH to the measured ESH at each sensor site. This is was necessary to achieve the same forecast horizon and time interval as the persistence model.

The table below is a comparison of forecast error for all 66 measurement sites for a six month period for NEO's NDFD method and the persistence method described above. The NEO RMSE absolute error values are lower by about 0.2 kWhr/m^2/day and RMSE relative error values are lower by between 3.5 and 2.6% for NEO's day ahead forecast method than persistence in all cases. However the MAE, both absolute and relative, are almost identical for NEO and the persistence model. This equivalence in apparent errors may be due to the fact that the Sacramento area has a very high percentage of very clear days, resulting in more instances where the daily ESH from one day are a good predictor of the subsequent day. In addition the process of aggregating the hourly irradiance data in NEO's day ahead forecasts into a single daily summation of ESH "throws away" much of the value of the hourly forecast for

[2] Our "day ahead" forecast predicts hourly irradiance in a window 20 to 36 hours in the future. Our prediction is made and our forecast is delivered at approximately 05:15 local time in California, the day before the forecast interval. The time window for the forecast is a period from slightly before sunrise to slightly after sunset on the target day. The choice of when the forecast is delivered, and thus how far in advance the forecast is calculated, is based upon SMUD's energy trading department schedule. SMUD's energy trading department plans their purchases and sales in their day ahead market between 5:00 and 8:00 am each morning for the next day.

[3] This is a six month period for which we had sensor data which had been filtered and interpolated using the process developed by Sandia and SMUD.

the NEO model and creates a lower resolution output. This can be seen where the RMSE metrics captures more of the impact of outliers in the dataset than the MAE metric.

Table 2: NEO NDFD day ahead forecasts compared to a persistence model.

Day Ahead Forecast Models: NEO NDFD vs. Persistence			
66 Sensors, Six Months			
		NEO NDFD	Persistence
RMSE	(ESH)	0.7249	0.9268
MAE	(ESH)	0.4758	0.4761
RMSE/avg	(%)	12.41	15.87
MAE/avg	(%)	8.14	8.15
RMSE/capacity (7.9)	(%)	9.18	11.73
MAE/capacity (7.9)	(%)	6.02	6.03

5.4.2 Error analysis of 20 – 36 hour forecasts for PV power

The same metrics could also be applied to forecasting the power of a PV plant. In this case, capacity would be the rated capacity of the plant (MWdc) and average would be the average measured power (MWdc) over the time interval. Note that since we are using one hour as the interval, if the average power during that a particular hour for a plant was 10 MW, then one could also say that it produced 10 MWh of energy during that hour, since energy = power * time. For our work, we use “power” and “production” interchangeably.

The following tables present the metrics for the 22 – 40[4] hour (day-ahead) power/production forecast for the eight FIT PV sites studied in this project. The model was run using “hindcasts” from 5/1/2012 through 5/1/2013 for all the sites except for McKenzie. This last site did not come on line until November 2012; hence our forecasts and related error analysis for McKenzie include only 11/1/2012 through 4/30/2013. Output from the models was then compared with measured production (power) from the sites, provided to NEO by SMUD.

Our metric algorithm eliminated instances when PV plant output and forecast output both equaled zero (i.e., night time hours), when measured output < 1% of rated capacity (thus eliminating early morning and late afternoon low irradiance hours), and when measured output was negative (usually occurred during nighttime but not always). Once all these conditions were filtered, we then lined up the remaining hours of output with the same hours of forecast, and performed our metric calculations. The tables below provide the monthly values for the error metrics for each FIT site; the last table provides annual averages for all the sites.

[4] Mid-way through the project, SMUD asked us to adjust the forecast window from 20-36 hours ahead to 22 – 40 hours ahead. Therefore, when we ran our production models, we produced hourly forecasts for the latter time period.

Table 3

Bruceville (18MW Fixed Tilt Array)						
Monthly Hourly Averages	**RMSE (MW)**	**MAE (MW)**	**RMSE/Capacity**	**MAE/Capacity**	**MAE/Average**	**Ave Measured Pwr**
May-12	1.90	1.32	10.5%	7.4%	12.8%	10.33
Jun-12	1.80	0.91	10.0%	5.1%	9.2%	9.89
Jul-12	1.29	0.83	7.2%	4.6%	8.3%	9.93
Aug-12	1.33	0.81	7.4%	4.5%	7.9%	10.25
Sep-12	1.52	0.89	8.4%	5.0%	9.3%	9.61
Oct-12	2.32	1.33	12.9%	7.4%	16.1%	8.24
Nov-12	3.83	3.10	21.3%	17.2%	44.7%	6.95
Dec-12	4.19	3.27	23.3%	18.2%	65.9%	4.96
Jan-13	3.34	2.75	18.5%	15.3%	39.2%	7.01
Feb-13	3.35	2.38	18.6%	13.2%	27.4%	8.67
Mar-13	3.55	2.53	19.7%	14.1%	28.0%	9.03
Apr-13	1.80	1.07	10.0%	5.9%	10.8%	9.91
	RMSE (MW)	**MAE (MW)**	**RMSE/Capacity**	**MAE/Capacity**	**MAE/Average**	**Ave Measured Pwr**
Annual Averages	2.52	1.77	14.0%	9.8%	23.3%	8.73

Table 4

Eschinger (15MW Tracking Array)						
Monthly Hourly Averages	**RMSE (MW)**	**MAE (MW)**	**RMSE/Capacity**	**MAE/Capacity**	**MAE/Average**	**Ave Measured Pwr**
May-12	2.72	1.79	18.1%	11.9%	16.3%	10.96
Jun-12	2.63	1.51	17.5%	10.1%	14.9%	10.16
Jul-12	2.24	1.34	14.9%	8.9%	12.1%	11.09
Aug-12	2.48	1.93	16.5%	12.9%	18.2%	10.63
Sep-12	2.89	2.18	19.2%	14.5%	23.0%	9.49
Oct-12	2.82	2.02	18.8%	13.5%	26.0%	7.75
Nov-12	3.65	2.74	24.3%	18.3%	43.1%	6.36
Dec-12	3.98	3.05	26.5%	20.3%	71.4%	4.27
Jan-13	2.76	1.98	18.4%	13.2%	29.5%	6.71
Feb-13	2.90	2.11	19.3%	14.1%	25.1%	8.43
Mar-13	3.63	2.64	24.2%	17.6%	30.0%	8.80
Apr-13	2.97	2.29	19.8%	15.3%	23.2%	9.91
	RMSE (MW)	**MAE (MW)**	**RMSE/Capacity**	**MAE/Capacity**	**MAE/Average**	**Ave Measured Pwr**
Annual Averages	2.97	2.13	19.8%	14.2%	27.7%	8.71

Table 5

Boessow (3MW Fixed Tilt Array)						
Monthly Hourly Averages	**RMSE (MW)**	**MAE (MW)**	**RMSE/Capacity**	**MAE/Capacity**	**MAE/Average**	**Ave Measured Pwr**
May-12	0.33	0.23	10.9%	7.7%	13.9%	1.67
Jun-12	0.35	0.16	11.6%	5.4%	9.7%	1.66
Jul-12	0.22	0.12	7.4%	3.9%	7.2%	1.64
Aug-12	0.22	0.13	7.4%	4.4%	7.8%	1.71
Sep-12	0.27	0.18	9.0%	6.0%	10.7%	1.68
Oct-12	0.40	0.24	13.5%	8.1%	17.0%	1.44
Nov-12	0.68	0.53	22.6%	17.8%	45.0%	1.18
Dec-12	0.73	0.57	24.4%	19.0%	74.0%	0.77
Jan-13	0.57	0.47	19.2%	15.5%	42.8%	1.09
Feb-13	0.60	0.50	19.9%	16.7%	34.5%	1.45
Mar-13	0.60	0.43	20.0%	14.2%	28.2%	1.52
Apr-13	0.34	0.20	11.3%	6.6%	12.0%	1.64
	RMSE (MW)	**MAE (MW)**	**RMSE/Capacity**	**MAE/Capacity**	**MAE/Average**	**Ave Measured Pwr**
Annual Averages	0.44	0.31	14.8%	10.4%	25.2%	1.45

Table 6

Kammerer (15MW Tracking Array)						
Monthly Hourly Averages	**RMSE (MW)**	**MAE (MW)**	**RMSE/Capacity**	**MAE/Capacity**	**MAE/Average**	**Ave Measured Pwr**
May-12	2.93	1.98	19.5%	13.2%	18.5%	10.67
Jun-12	3.03	1.98	20.2%	13.2%	20.1%	9.88
Jul-12	2.33	1.41	15.5%	9.4%	12.8%	10.98
Aug-12	2.49	1.83	16.6%	12.2%	16.9%	10.84
Sep-12	2.92	2.17	19.5%	14.4%	22.8%	9.52
Oct-12	2.81	2.07	18.7%	13.8%	27.0%	7.65
Nov-12	3.77	2.82	25.1%	18.8%	44.3%	6.35
Dec-12	3.96	2.98	26.4%	19.9%	71.9%	4.15
Jan-13	3.49	2.54	23.3%	16.9%	36.9%	6.87
Feb-13	3.50	2.61	23.3%	17.4%	31.2%	8.38
Mar-13	3.56	2.62	23.7%	17.5%	29.6%	8.88
Apr-13	3.06	2.26	20.4%	15.0%	23.1%	9.78
	RMSE (MW)	**MAE (MW)**	**RMSE/Capacity**	**MAE/Capacity**	**MAE/Average**	**Ave Measured Pwr**
Annual Averages	3.15	2.27	21.0%	15.1%	29.6%	8.66

Table 7

Point Pleasant (1MW Fixed Tilt Array)						
Monthly Hourly Averages	**RMSE (MW)**	**MAE (MW)**	**RMSE/Capacity**	**MAE/Capacity**	**MAE/Average**	**Ave Measured Pwr**
May-12	0.11	0.08	11.0%	7.9%	15.5%	0.509
Jun-12	0.11	0.07	10.5%	7.5%	13.6%	0.551
Jul-12	0.12	0.08	11.7%	8.4%	16.6%	0.509
Aug-12	0.10	0.08	10.3%	7.5%	14.7%	0.511
Sep-12	0.11	0.09	11.0%	8.9%	16.4%	0.539
Oct-12	0.13	0.11	13.4%	11.1%	20.8%	0.535
Nov-12	0.25	0.20	24.6%	20.0%	48.5%	0.413
Dec-12	0.23	0.18	22.8%	18.1%	52.5%	0.344
Jan-13	0.19	0.16	18.5%	16.0%	36.6%	0.437
Feb-13	0.19	0.16	19.1%	15.8%	31.3%	0.505
Mar-13	0.20	0.15	20.0%	15.3%	29.6%	0.516
Apr-13	0.14	0.10	14.0%	9.6%	17.4%	0.553
	RMSE (MW)	**MAE (MW)**	**RMSE/Capacity**	**MAE/Capacity**	**MAE/Average**	**Ave Measured Pwr**
Annual Averages	0.16	0.12	16%	12%	26%	0.49

Table 8

Kost (3MW Fixed Tilt Array)						
Monthly Hourly Averages	**RMSE (MW)**	**MAE (MW)**	**RMSE/Capacity**	**MAE/Capacity**	**MAE/Average**	**Ave Measured Pwr**
May-12	0.34	0.21	11.3%	6.9%	13.6%	1.54
Jun-12	0.31	0.15	10.4%	5.0%	9.4%	1.59
Jul-12	0.23	0.13	7.8%	4.4%	8.3%	1.57
Aug-12	0.22	0.14	7.4%	4.6%	8.3%	1.66
Sep-12	0.28	0.16	9.3%	5.3%	10.5%	1.52
Oct-12	0.38	0.24	12.5%	7.9%	17.0%	1.40
Nov-12	0.68	0.53	22.6%	17.6%	43.4%	1.22
Dec-12	0.69	0.54	23.1%	18.1%	63.2%	0.86
Jan-13	0.59	0.49	19.6%	16.4%	41.5%	1.18
Feb-13	0.57	0.48	18.9%	15.9%	32.2%	1.48
Mar-13	0.59	0.42	19.8%	13.9%	28.0%	1.49
Apr-13	0.32	0.20	10.7%	6.7%	12.0%	1.66
	RMSE (MW)	**MAE (MW)**	**RMSE/Capacity**	**MAE/Capacity**	**MAE/Average**	**Ave Measured Pwr**
Annual Averages	0.43	0.31	14%	10%	24%	1.43

Table 9

Dillard (9.4MW Tracking Array)						
Monthly Hourly Averages	**RSME (MW)**	**MAE (MW)**	**RSME/Capacity**	**MAE/Capacity**	**MAE/Average**	**Ave Measured Pwr**
May-12	2.01	1.38	21.4%	14.7%	21.6%	6.42
Jun-12	1.98	1.28	21.1%	13.6%	19.8%	6.46
Jul-12	1.95	1.25	20.8%	13.3%	19.9%	6.28
Aug-12	1.97	1.44	21.0%	15.4%	22.7%	6.37
Sep-12	2.24	1.68	23.8%	17.9%	29.2%	5.74
Oct-12	1.98	1.44	21.1%	15.3%	29.9%	4.80
Nov-12	2.48	1.85	26.3%	19.6%	47.6%	3.88
Dec-12	2.71	2.06	28.8%	21.9%	82.4%	2.50
Jan-13	2.12	1.56	22.5%	16.6%	37.5%	4.17
Feb-13	2.25	1.68	24.0%	17.9%	32.3%	5.20
Mar-13	2.39	1.74	25.4%	18.5%	31.3%	5.55
Apr-13	2.18	1.58	23.2%	16.9%	26.5%	5.97
	RSME (MW)	**MAE (MW)**	**RSME/Capacity**	**MAE/Capacity**	**MAE/Average**	**Ave Measured Pwr**
Annual Averages	2.19	1.58	23.3%	16.8%	33.4%	5.28

Table 10

McKenzie (30MW Tracking Array)						
Monthly Hourly Averages	**RMSE (MW)**	**MAE (MW)**	**RMSE/Capacity**	**MAE/Capacity**	**MAE/Average**	**Ave Measured Pwr**
May-12						
Jun-12						
Jul-12						
Aug-12						
Sep-12						
Oct-12						
Nov-12	6.51	4.92	21.7%	16.4%	43.7%	11.28
Dec-12	8.02	6.24	26.7%	20.8%	73.3%	7.73
Jan-13	5.27	4.14	17.6%	13.8%	30.0%	13.79
Feb-13	5.26	4.37	17.5%	14.6%	24.8%	17.64
Mar-13	7.26	5.51	24.2%	18.4%	30.0%	18.40
Apr-13	8.87	7.42	29.6%	24.7%	44.3%	16.75
	RMSE (MW)	**MAE (MW)**	**RMSE/Capacity**	**MAE/Capacity**	**MAE/Average**	**Ave Measured Pwr**
Annual Averages	6.87	5.43	22.9%	18.1%	41.0%	14.27

Table 11

Annual Metrics						
FIT Site	**RMSE**	**MAE**	**RMSE/Capacity**	**MAE/Capacity**	**MAE/Average Measured Power**	**Ave. Measured Power**
Bruceville (18MW Fixed)	2.52	1.77	14.0%	9.8%	23.3%	8.73
Eschinger (15MW Tracking)	2.97	2.13	19.8%	14.2%	27.7%	8.71
Boessow (3MW Fixed)	0.44	0.31	14.8%	10.4%	25.2%	1.45
Dillard (9.4MW Tracking)	2.19	1.58	23.3%	16.8%	33.4%	5.28
Kammerer (15MW Tracking)	3.15	2.27	21.0%	15.1%	29.6%	8.66
Point Pleasant (1MW Fixed)	0.16	0.12	15.6%	12.2%	26.1%	0.49
Kost (3MW Fixed)	0.43	0.31	14.4%	10.2%	23.9%	1.43
McKenzie (30 MW Tracking)	6.87	5.43	22.9%	18.1%	41.0%	14.27
Aggregate*	**NA**	**NA**	**17.6%**	**12.7%**	**27.0%**	**NA**
***Does not include McKenzie.**						

5.4.3 Error analysis of 0 – 3 hour forecasts for power production

Error analysis for our 0 – 3 hour PV production forecasts was performed in a similar manner to the day-ahead PV production forecasts. Note that these forecasts produce power estimates in 15-minute increments, whereas measured power was only provided in hourly increments. To adjust for this, we assumed that the power for each hour remained the same for each 15-minute increment within that hour. This in itself may have increased the magnitude of the errors. The results of the error analyses are displayed in the table and graph below. The McKenzie FIT site is not included, since it did not come on-line until mid-November 2012.

Table 12: 0 - 3 Hour PV Power Forecast Metric Summary

FIT Site	**RMSE**	**MAE**	**RMSE/Capacity**	**MAE/Capacity**	**MAE/Average Measured Power**	**Ave. Measured Power**
Bruceville (18MW Fixed)	3.57	2.91	19.8%	16.2%	33.3%	8.93
Eschinger (15MW Tracking)	3.09	2.12	20.6%	14.2%	22.9%	9.28
Boessow (3MW Fixed)	0.60	0.49	19.9%	16.2%	32.0%	1.51
Dillard (9.4MW Tracking)	2.53	2.07	26.9%	22.1%	38.3%	5.41
Kammerer (15MW Tracking)	3.13	2.18	20.9%	14.6%	24.9%	9.15
Point Pleasant (1MW Fixed)	0.23	0.18	22.9%	17.9%	36.4%	0.49
Kost (3MW Tracking)	0.50	0.38	16.7%	12.7%	27.7%	1.37
Aggregate	NA	NA	21.1%	16.2%	30.8%	NA

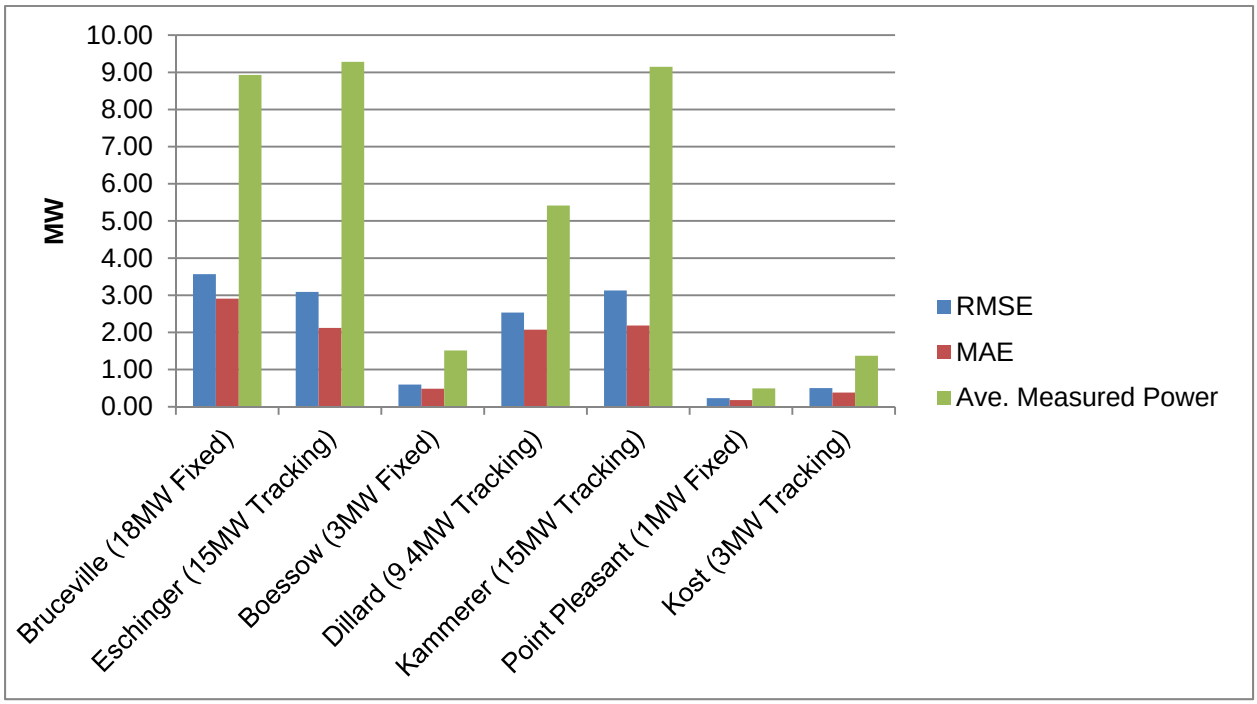


Figure 16: 0-3 Hour PV power forecast error metrics and average measured power for FIT sites

5.4.4 Comparing results of all the forecasts

A comparison of the day-ahead average metrics between irradiance and PV production indicates that the irradiance forecasts were more accurate:

Table 13

	Irradiance	PV Production
MAE/AVERAGE MEASURED	12.9%	27.0%
RMSE/CAPACITY	11.1%	17.6%
MAE/CAPACITY	6.1%	12.7%

There are a number of possible reasons. First, NEO could not know when a particular sub-array of a PV plant was offline (for maintenance, for example; much of this time was during the startup period for these systems). Our metric algorithm would not have eliminated these time periods of partial PV plant output, and hence they would have been compared to the model's full output forecast, thus introducing error. Second, NEO did not receive exact specifications for a number of the FIT sites (especially the four single-axis tracking sites), and hence we had to use best-fit specifications for modules, array layout, etc., when creating the PV plant models. For all of the single axis tracking systems we assumed that they employed a backtracking algorithm, however we had no confirmation of this and we had to assume the characteristics of the backtracking algorithm. If the actual components at a FIT site varied significantly from our best-fit components, errors would have been introduced.

Finally, we suspect our models—especially for the tracking systems—could have used more adjusting to better fit the measured production. As just one example, the figure below shows a series of clear-sky days at Kammerer, one of the single-axis tracking sites.

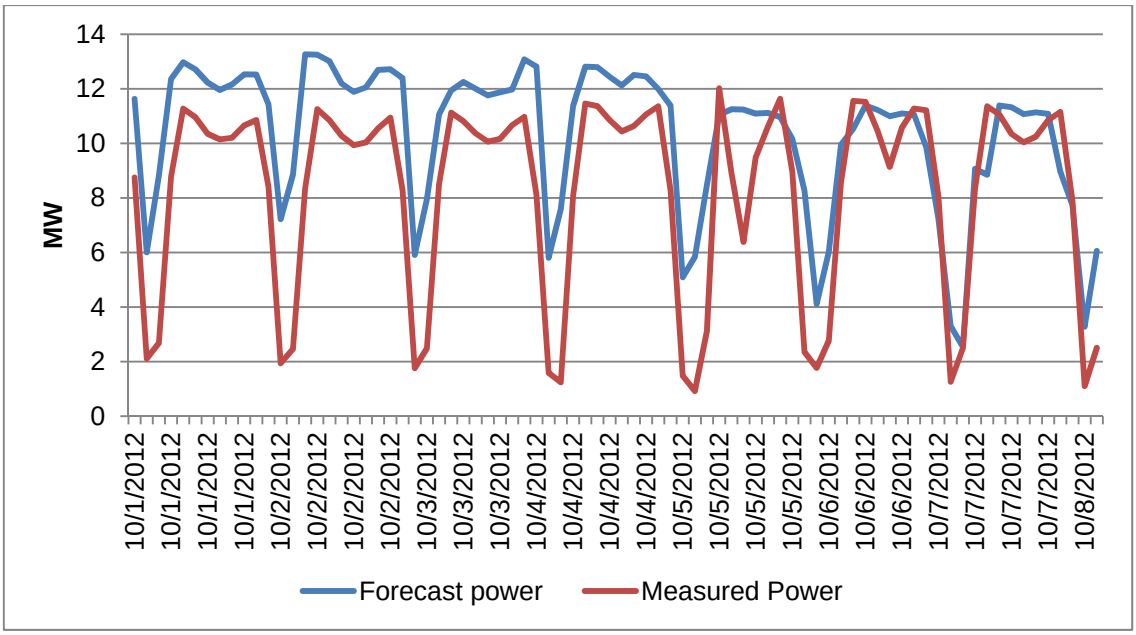


Figure 17: Kammerer forecast vs. measured power, Oct 1 - 7, 2012

One can see that for the first four days, the model overestimates output, both during the peak hours and especially during the "troughs" that occur early and late in each day. However, for the final three days, the model is much closer to the measured power. The variation in forecast peak power can be attributed to forecast irradiance, as that was also lower later in the week, as can be seen in the following graph:

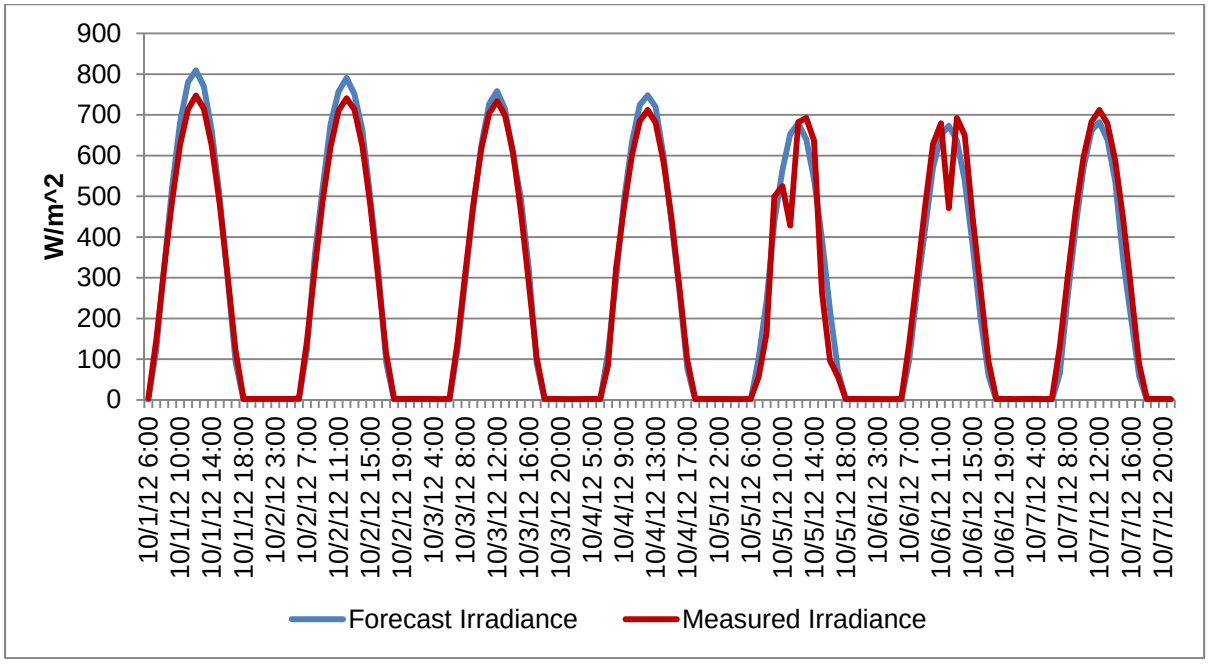


Figure 18: Measured vs. forecast irradiance for sensor #10 (closest to Kammerer), Oct 1 - 7, 2012

However, the decline in forecast peak irradiance does not explain the large deltas that occur early in the week. Also of note: the measured irradiance declines during the week, but the measured power remains essentially unchanged. Thus, we are left with a mystery as to why the forecast power follows the forecast irradiance (as one would expect, since the former is dependent upon the latter), but the measured power does not follow the measured irradiance!

This week is a good example of how a slight disagreement between forecast and measured irradiance manifests itself for only a couple of hours around local noon, whereas the same magnitude disagreement between forecast and measured power extends over the entire day. This extended period directly contributes to the relatively large error metrics for the power forecast.

Next, the following graph illustrates how the output of each FIT site varied over the year. The sharp decline in output during the winter months is attributable to both cloudier weather and to shorter daylight hours.

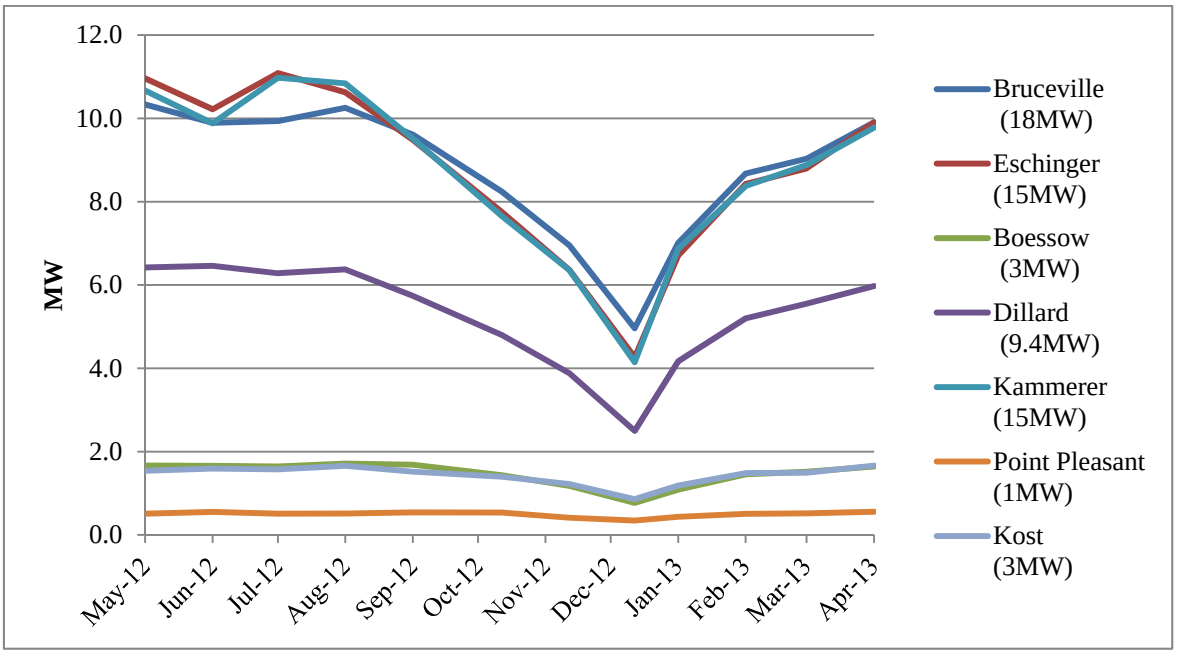


Figure 19: Average Monthly Measured Power per FIT Site

This sharp decline (approximately 50%) in average output is inversely proportional to the MAE/average power metric, as can be seen in the graph below.

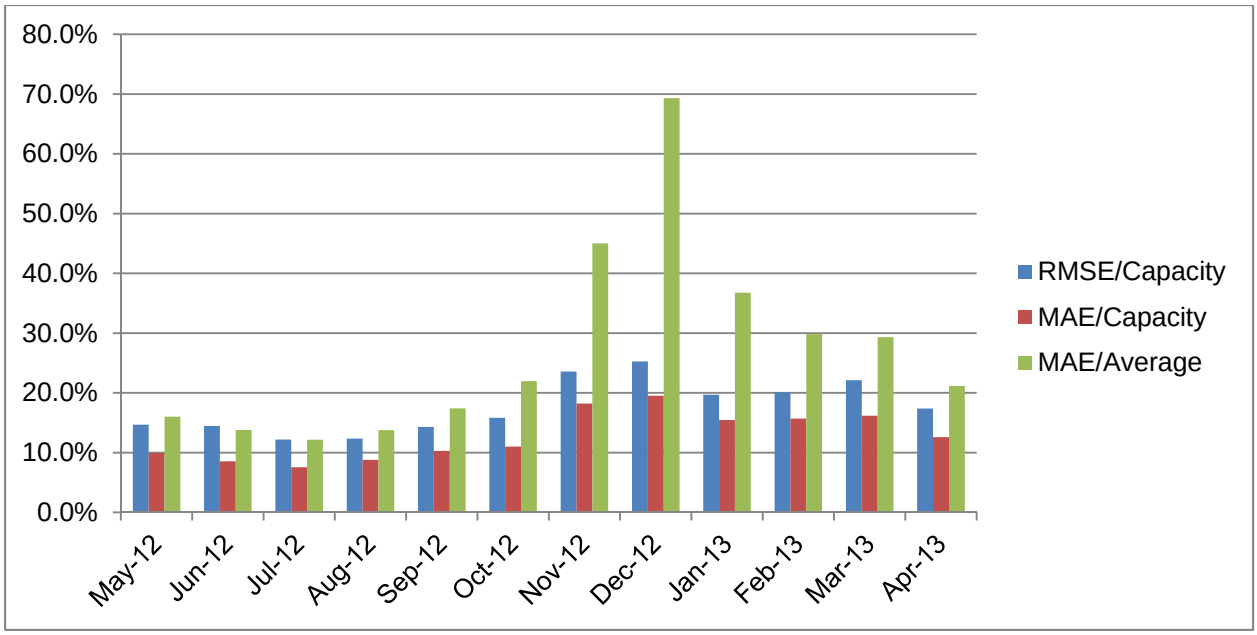


Figure 20: Metric monthly variation

At this time, we have no way of knowing exactly how much error to attribute to shorter days and how much to cloudier weather. However, one can see that the two metrics related to capacity do not vary nearly as much, giving some indication that the shorter days (and hence lower average power output) contribute more to the MAE/average error than the increase in clouds. This gives some validity to the accuracy of our forecasts vis a vis cloud cover.

We could also see from our analysis that our fixed-axis model performed better than our single-axis tracking model (see the two scatter plots below). Again, part of this difference could be caused by inaccurate system specifications for the tracking arrays, and part could be the tracking model itself.

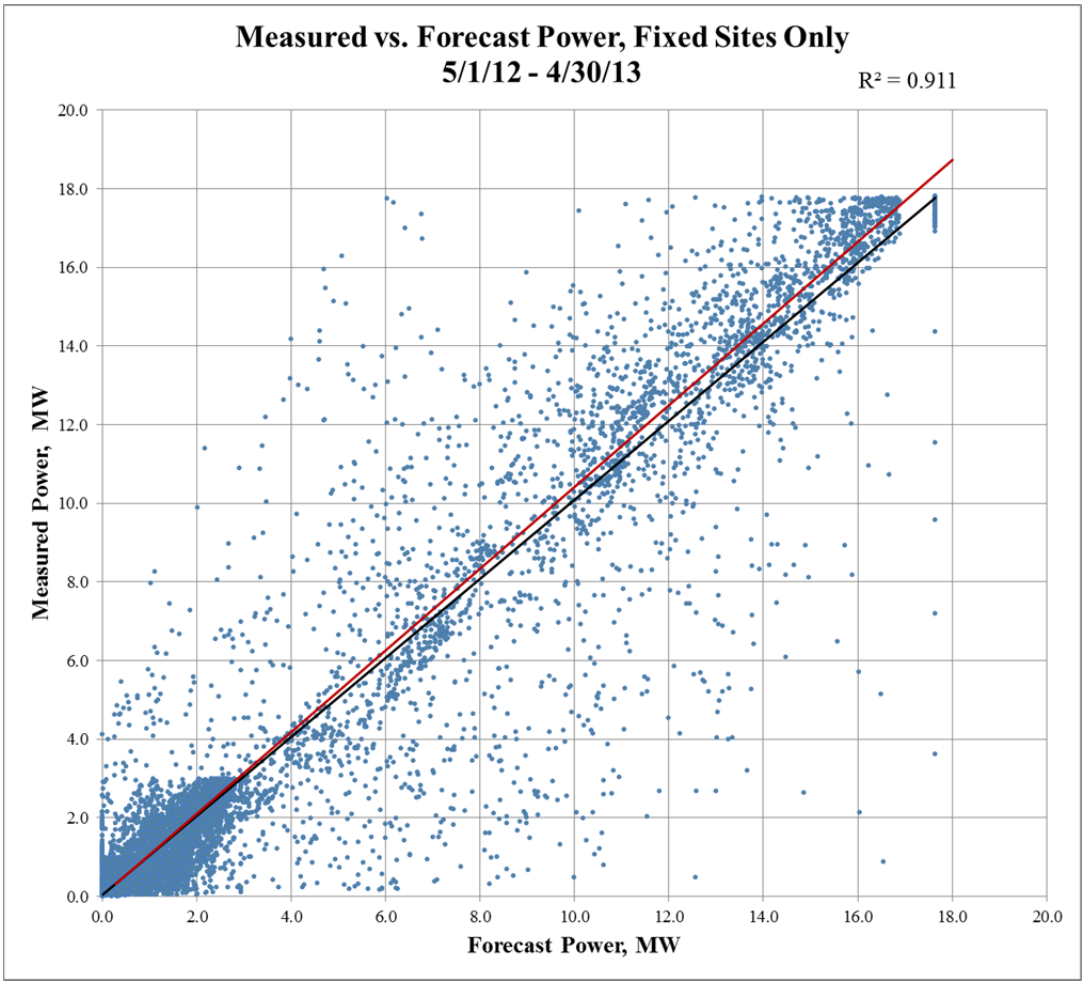


Figure 21: Scatter plot of hourly measured vs. 20 – 36 hour forecast PV power for the four fixed array FIT sites.

Note that the large cluster of points at low power levels are due to the fact that of the four fixed arrays (Bruceville, Boessow, Point Pleasant and Kost), three were relatively small (3 MW or less). Thus, the points from these relatively small systems all occur in the lower left quadrant of this plot.

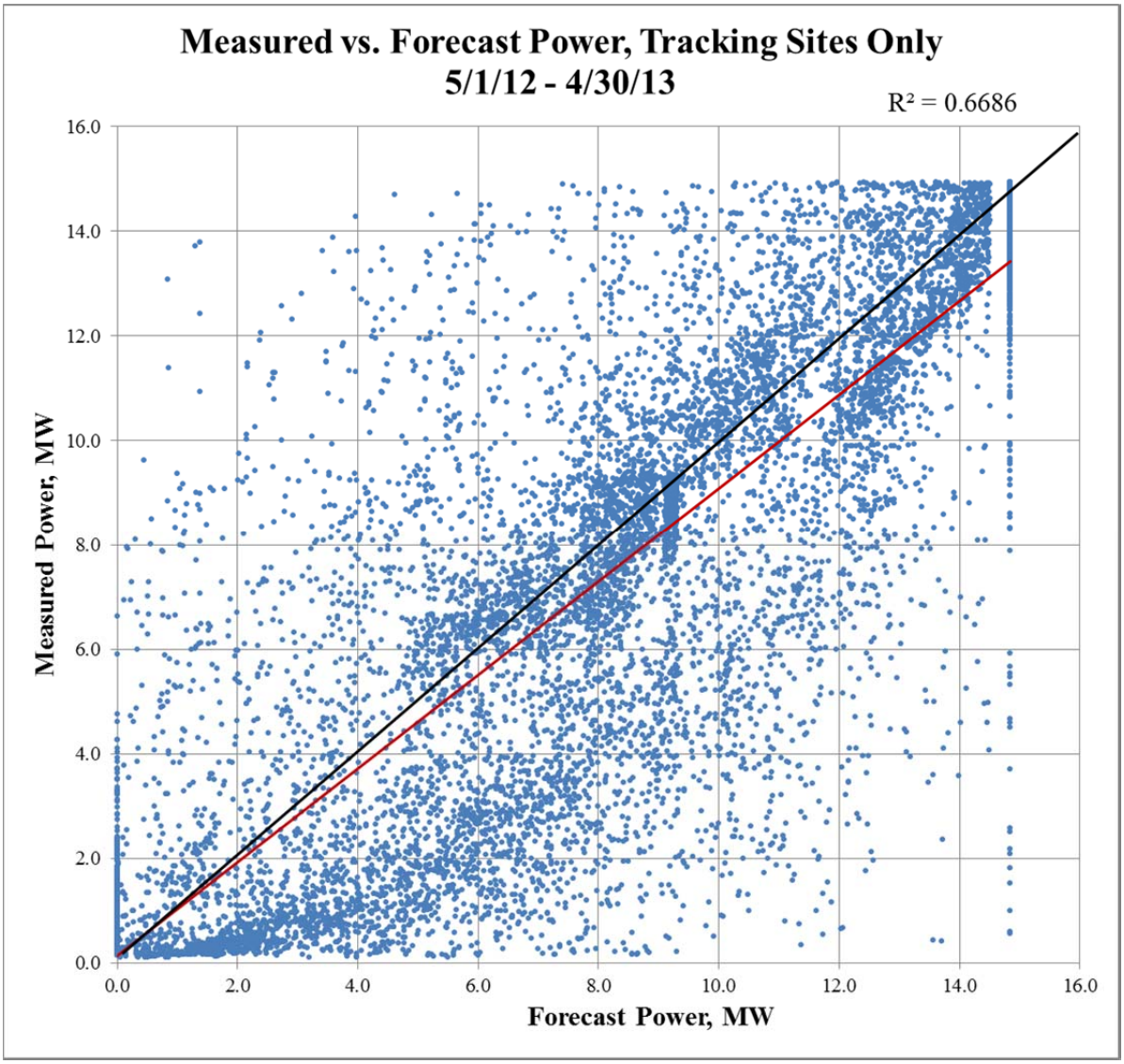


Figure 22: Scatter plot of hourly measured vs. 20 – 36 hour forecast power for the three tracking FIT sites.

This plot of the tracking systems includes Kammerer, Eschinger and Dillard. Note the relatively low R2 (coefficient of determination) value. This value indicates how closely data points fit a line, i.e., an R2 value equal to one would indicate a perfect fit and hence perfect forecast (the black line in the plot).

5.4.5 Summary

The table below summarizes the error metrics for the three forecasts. It is clear that the irradiance day-ahead forecast yields the best aggregate results, followed by our day-ahead PV production forecasts, and then the 0 – 3 hour PV production forecasts.

Table 14

	Day Ahead Irradiance	Day Ahead PV Production Forecast	0 – 3 Hour PV Production Forecast
MAE/AVERAGE MEASURED	12.9%	27.0%	30.8%
RMSE/CAPACITY	11.1%	17.6%	21.1%
MAE/CAPACITY	6.1%	12.7%	16.2%

Our 0 – 3 hour forecasts are based on a neural network model that "learns" from measured irradiance data from sensors in the grid cells adjacent to each PV FIT site. This is a completely different methodology from our day-ahead forecasts, which begin with NDFD weather forecasts. Hence, we believe that the 5 km special resolution of our sensor network was the primary cause for the limited accuracy of our short-term predictions. We conclude that this is a function of our sensor spacing relative to typical cloud size.

It should be noted that the primary purpose of the sensor grid (or the SMUD-NEO Irradiance Network, "SNI") in Sacramento is to measure irradiance and temperature data over a large area at a high temporal rate and for multiple years. For future work that specifically focuses on short-term forecasting, we would recommend a denser network of sensors in close proximity to PV plants. Such a network would have a higher likelihood of capturing shading events caused by the motion of smaller clouds, and hence would provide more accurate input into our neural network model.

Finally, the consistency of the irradiance models' and PV power models' metrics across the different sensors and FIT sites, respectively, is in itself an indication of the validity of the models. For example, the standard deviation of the MAE error for the 66 sensors was only 7.58 W/m^2, with an average MAE of 61.23 W/m^2. Likewise, the four MAE/averages for the fixed-array FIT sites all fell between 23.3% and 26.1%. The rest of the metrics exhibited similar consistency. Therefore, we conclude that our models do provide forecasts consistently, with our irradiance forecast performing the best, followed by our fixed array model, and then the single-axis tracking model.

Project Outcomes

Task 5: PV Production Forecasting

5 km grid irradiance network (NEO and SMUD)

This network was up and running by May 31, 2011. It consists of 66 secondary stations (designed and built by NEO) plus 6 primary stations (RSR2s, designed by Irradiance, Inc.; 5 of the 6 were assembled by NEO, one was already in-place). This network has provided 24 months of GHI and temperature data, and continues to function.

Forecast PV production piece-wise program (NEO and contractors)

NEO downloaded required NDFD data manually from NOAA site. NDFD program tkdegrib was run manually to "unpack" NDFD gridded binary forecast meteorological files. FIT system modules were modeled based upon similarities with modules already in the Sandia Module database and the database was updated. NEO created an Excel version of the Sandia PV Array Model for manual processing and sample runs were done.

The final automated forecasting program, DAE_analyze.pl, was coded and since April 17, 2012, NEO has been sending daily power production forecasts to SMUD.

One year measured 5km, 1 min rate, irradiance & ambient temp data (NEO and contractors)

As of May 31, 2013, data from the full primary and secondary sensor network is available on the SMUD website. Due to a one year extension, as of May 31 the data set contains a full 24 months of data. The full data record can be found on SMUD's ftp site:

ftp://ftp01.smud.org/pub/SNIData/

The file will be a zip file, and it will be a backup of NEO's MySQL database table. There is a "read me" file that includes information about the data. The data will include raw "shadow events" and interpolated values. The data will be updated monthly starting in May 2013.

Review of the data has determined that 98.5% of the dataset contains valid and accurate data.

One year day-ahead forecast PV production data by site, 1 hour intervals (NEO and contractors)

NEO created forecasts from 5/1/2012 through 4/30/13 for the following seven FIT sites:

Bruceville (18MW)
Kammerer (15MW)

Eschinger (15MW)
Dillard (9.4MW)
Kost (3MW)
Boessow (3MW)
Point Pleasant (1MW)

NEO created forecasts from only 11/16/2012 – 4/30/2013 for the eighth FIT site, McKenzie (30 MW) since this site did not come on-line until November 2012. These forecasts can be found on SMUD's ftp site, in the spreadsheet entitled:

Deliverable IV_PV Production Day Ahead Forecasts 130610.xlsx

The spreadsheet also contains the measured power for each FIT site to allow for easy comparison.

One year 3-hour forecast PV production data by site, 15 min intervals (NEO and contractors)

NEO created forecasts from 2/1/2012 through 11/30/12 for the following seven FIT sites:

Bruceville (18MW)
Kammerer (15MW)
Eschinger (15MW)
Dillard (9.4MW)
Kost (3MW)
Boessow (3MW)
Point Pleasant (1MW)

NEO created forecasts from only 11/16/2012 – 11/30/2012 for the eighth FIT site, McKenzie (30 MW) since this site did not come on-line until November 2012. These forecasts can be found on SMUD's ftp site, in the spreadsheet entitled:

Deliverable V_PV Production 0-3 Hour Forecasts 130610.xlsx

The spreadsheet also contains the measured power for each FIT site to allow for easy comparison. There are also 10 spreadsheets that contain 0-3 hour forecast and measured GHI for 72 5 km grid cells that make up the SNI network. Each spreadsheet contains one month worth of data for every grid cell.

Quantify forecast vs. measured error for PV production (NEO and contractors)

NEO created a detailed report on our error analysis of our PV production and irradiance forecasts. This report was provided to SMUD on June 10, 2013, and can be found on SMUD's ftp site in the document entitled:

Deliverable VI_Report 130610.pdf

Recommendations

(aka lessons learned)

1. Mounting secondary stations on existing utility poles had the unexpected consequence of requiring post collection data filtering. Shadows from overhead wires and crosstrees caused drops in GHI levels as the sun moved across the sky. These were very noticeable on clear days—see section 5.1.2 and figure 7 above. For future installations on utility poles we recommend further developing and automating the filtering algorithm developed by SMUD and Sandia. Where possible, we recommend placing the devices only on poles without crosstrees.
2. Future installations of a network of irradiance sensors such as the one installed under this grant should include a means for calibrating the sensors to ensure accurate measurements. NEO and SMUD installed four PSPs in our network as reference devices for use in calibrating all the other pyranometers. However, these devices were an afterthought rather than a planned component. Similarly, our "global calibration" methodology for this project was developed after the sensors were installed. Future projects of this nature should have a calibration plan included as part of the initial scope of work.
3. The NDFD data we used in our 20 - 36 hour forecast model had 3-hour forecast periods (time intervals) for most parameters, e.g., 0600 – 0900, 0900 – 1200, etc.). Since we were forecasting irradiance and power production every hour, these intervals were not optimal. In addition, the forecast period for a key parameter—sky cover—varied between geographic locations, further complicating our modeling efforts. The lesson here is that there (now) exists a better resource —the High Resolution Rapid Refresh (HRRR) forecast. This forecast has shorter forecast periods (1 hour) and smaller geographic size (3 km vs. 5 km grid) and a larger number of critical parameters.
4. Some of the RSRs used in the network were connected via short-haul modem to SMUD's SCADA network. Occasionally, their internal clocks would be reset to local time vs. GMT. This caused a misalignment of "time/date" stamps for the daily collected data. For future work, any programming and datalogger clock resets need to be limited to the network's administrator only (in this case, NEO).
5. To improve PV plant modeling, better access to plant data, especially back-tracking algorithms, should be made available to the forecasters. In addition, to improve the accuracy of forecast metrics, plant availability (i.e., down times for maintenance) should be provided to the forecasters. In this current project, NEO was not told of these down times, and hence had to make some assumptions while comparing forecast to actual plant production (if the measured output of a plant was zero two days straight, for example, we assumed it was off line, and did not use those days in our metrics).