NCAR/TN-526+STR

NCAR TECHNICAL NOTE

THE SUN4CAST® SOLAR POWER FORECASTING SYSTEM: THE RESULTS OF THE PUBLIC-PRIVATE-ACADEMIC PARTNERSHIP TO ADVANCE SOLAR POWER FORECASTING

Sue Ellen Haupt, Branko Kosovic, Tara Jensen, Jared Lee, Pedro Jimenez, Jeffrey Lazo, James Cowie, Tyler McCandless, Julia Pearson, Gerry Weiner, Stefano Alessandrini, and Luca Delle Monache

Research Applications Laboratory / NCAR, Boulder, Colorado

Dantong Yu, Zhenzhou Peng, Dong Huang, John Heiser, Shinjae Yoo, and Paul Kalb

Brookhaven National Laboratory, Upton, New York

Steven Miller and Matthew Rogers

Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado

Laura Hinkleman

Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, Washington

Research Applications Laboratory

Weather Systems and Assessment Program

NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

P. O. Box 3000

BOULDER, COLORADO 80307-3000

ISSN Print Edition 2153-2397

ISSN Electronic Edition 2153-2400

Acknowledgements

This work was primarily funded by the US Department of Energy under SunShot Cooperative Agreement DE-EE0006016. The National Center for Atmospheric Research is sponsored by the National Science Foundation and portions of this work were also funded directly by NSF.

The authors are indebted to the full Sun4Cast team, which includes too many members to include all. But we particularly wish to thank

From **NCAR**: Sheldon Drobot, Seth Linden, Gary Cunning, Jimy Dudhia, Joshua Hacker, Thomas Brummet, Bill Petzke, Barbara Brown, Tressa Fowler, Amanda Anderson, Julie Prestopnik, Lisa Goodrich, William Mahoney, Thomas Aligne, Greg Thompson, Susan Dettling, Trude Eidhammer, David John Gagne, Louisa Nance, Paul Prestopnik, Karen Griggs, Stephen Dowdy, Susan Dettling, and Anna Hoffee.

From **NREL**: Manijit Sangupta, Yu Xie, Bri Matthias-Hodge

From University of Jaen: Jose Ruis-Arrias

From **NOAA**: Melinda Marquis, Stan Benjamin, Kathy Lantz, Joseph Olson, Joseph Michalsky, Christine Molling

From **Solar Consulting Services**: Chris Gueymard

From **Penn State**: George Young, Aijun Deng, David Stauffer, Jeffrey Brownson

From **University of Buffalo**: Tarunraj Singh

From **US Army Research Laboratory**: Gail Vaucher, Gary McWilliams

From **Xcel Energy**: Keith Parks, Drake Bartlett

From **Global Weather Corporation**: Kristen Larson, William Meyers, William Gail, Brett Basarab, Nancy Rehak

From **MDA Federal**: Stephen Jascourt, Daniel Kirk-Davidoff

From **Schneider Electric**: Don Leick

From **SCE**: Jack Peterson

From **SMUD**: Thomas Vargas, Obadiah Bartholomy, Megan Simone

From **CAISO**: James Blatchford

From **NYPA**: Li Kou, John Markowitz, Houtan Moaveni, Samuel Morris, Guy Sliker, Michael Swider, Edward Davis

From **AER**: Ron Isaacs, Thomas Nehrkor, Thomas Connor

From **HECO**: Dora Nakfuji

From **University of Hawaii**: Duane Stevens, Lipyeow Lim

Amendment 1: This version, edited January 20, 2017, is an amended version that reflects the new Trademark for the system, the Sun4Cast® Solar Power Forecasting System. NCAR routinely trademarks its products to maintain integrity in the use of the name. Note that the OpenSource components of the Sun4Cast System can still be used widely. NCAR requests attribution and use of the trademark on the name.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

The National Center for Atmospheric Research (NCAR) led a partnership to advance the state-of-the-science of solar power forecasting by designing, developing, building, deploying, testing, and assessing the Sun4Cast® Solar Power Forecasting System. The project included cutting-edge research, testing in several geographically and climatologically diverse high penetration solar utilities and ISOs, and wide dissemination of the research results to raise the bar on solar power forecasting technology. The partners included three other national laboratories, six universities, and industry partners. This public-private-academic team worked in concert to advance solar power forecasting by performing use-inspired and cutting-edge research to advance both the necessary forecasting technologies and the metrics for evaluating them. The project has culminated in a year-long, full-scale demonstration of providing irradiance and power forecasts to utilities and ISOs to use in their operations.

The project focused on identifying elements of a value chain, beginning with the weather that causes a deviation from clear sky irradiance and progressing through monitoring of observations, modeling, forecasting, dissemination and communication of the forecasts, interpretation of the forecasts, and through decision-making, which produces outcomes that have an economic value. The system was evaluated using metrics developed specifically for this project, which provided rich information on model and system performance.

Research was conducted on the very short range (0-6 hours) Nowcasting system as well as on the longer term (6-72 hour) forecasting system, which were then blended, converted to power, analog ensemble applied to for the Sun4Cast® Solar Power Forecasting System. The shortest range forecasts are based on observations in the field. TSICast operates on the shortest time scale, with a latency of only a few minutes and forecasts that currently extend to approximately15 min. This project facilitated research in improving hardware and software so that the new high definition cameras deployed at multiple nearby locations allow discernment of the clouds at varying levels and advection according to the winds observed at those levels. Improvements over “smart persistence” are about 29% for even these very short forecasts. StatCast uses pyranometer data measured at the site as well as concurrent meteorological observations and forecasts. StatCast is based on regime-dependent artificial intelligence forecasting techniques and has been shown to improve on “smart persistence” forecasts by 15-50%. A second category of short-range forecasting systems employs satellite imagery and uses that information to discern clouds and their motion, allowing these systems to project the clouds, and the resulting blockage of irradiance in time. CIRACast was already one of the more advanced cloud motion systems, which is the reason that team was brought to this project. During the project timeframe, the CIRA team advanced cloud shadowing, parallax removal, and implementation of better advecting winds at different altitudes. CiraCast shows generally a 25-40% improvement over Smart Persistence between sunrise and approximately 1600 UTC. A second satellite-based system, MADCast, assimilates data from multiple satellite imagers and profilers to assimilate a three-dimensional picture of the cloud into the dynamic core of WRF. This allows advection of the clouds via the WRF dynamics directly. During 2015, MADCast provided at least 70% improvement over Smart Persistence, with most of that skill being derived during partly cloudy conditions. After WRF-Solar™ showed initial success, it was also deployed in nowcasting mode with coarser runs extending to 6 hours made hourly. It provided improvements on the order of 50-60% over Smart Persistence for forecasts

extending to 1600 UTC. The advantages of WRF-Solar-Nowcasting and MADCast were then blended to develop the new MAD-WRF model that incorporates the most important features of each of those models, both assimilating satellite cloud fields and using WRF-Solar™ physics to develop and dissipate clouds. MAE improvements for MAD-WRF forecasts from 3-6 hours are improved over WRF-Solar-Now by 20%. While all the Nowcasting system components provide improvement over Smart Persistence individually, the largest benefit is derived when they are smartly blended together by the Nowcasting Integrator to produce an integrated forecast.

The development of WRF-Solar™ under this project has provided the first numerical weather prediction (NWP) model specifically designed to meet the needs of irradiance forecasting. The first augmentation improved the solar tracking algorithm to account for deviations associated with the eccentricity of the Earth's orbit and the obliquity of the Earth. Second, WRF-Solar™ added the direct normal irradiance (DNI) and diffuse (DIF) components from the radiation parameterization to the model output. Third, efficient parameterizations were implemented to either interpolate the irradiance in between calls to the expensive radiative transfer parameterization, or to use a fast radiative transfer code that avoids computing three-dimensional heating rates but provides the surface irradiance. Fourth, a new parameterization was developed to improve the representation of absorption and scattering of radiation by aerosols (aerosol direct effect). A fifth advance is that the aerosols now interact with the cloud microphysics, altering the cloud evolution and radiative properties, an effect that has been traditionally only implemented in atmospheric computationally costly chemistry models. A sixth development accounts for the feedbacks that sub-grid scale clouds produce in shortwave irradiance as implemented in a shallow cumulus parameterization Finally, WRF-Solar™ also allows assimilation of infrared irradiances from satellites to determine the three dimensional cloud field, allowing for an improved initialization of the cloud field that increases the performance of short-range forecasts. We found that WRF-Solar™ can improve clear sky irradiance prediction by 15-80% over a standard version of WRF, depending on location and cloud conditions. In a formal comparison to the NAM baseline, WRF-Solar™ showed improvements in the Day-Ahead forecast of 22-42%.

The Sun4Cast® system requires substantial software engineering to blend all of the new model components as well as existing publicly available NWP model runs. To do this we use an expert system for the Nowcasting blender and the Dynamic Integrated foreCast (DICast®) system for the NWP models. These two systems are then blended, using an empirical power conversion method to convert the irradiance predictions to power, and then applying an analog ensemble (AnEn) approach to further tune the forecast as well as to estimate its uncertainty. The AnEn module decreased Root Mean Square Error (RMSE) by 17% over the blended Sun4Cast® power forecasts and provided skill in the probabilistic forecast with a Brier Skill Score of 0.55. In addition, we developed a Gridded Atmospheric Forecast System (GRAFS) in parallel, leveraging cost share funds.

An economic evaluation based on Production Cost Modeling in the Public Service Company of Colorado showed that the observed 50% improvement in forecast accuracy will save their customers $819,200 with the projected MW deployment for 2024. Using econometrics, NCAR has scaled this savings to a national level and shown that an annual expected savings for this 50% forecast error reduction ranges from $11M in 2015 to $43M expected in 2040 with increased solar deployment. This amounts to $455M in potential discounted savings over the 26-year period of analysis.

1 INTRODUCTION

1.1 CONTEXT

In 2013, the National Center for Atmospheric Research (NCAR) was awarded a collaborative agreement with the Department of Energy (DOE) SunShot Initiative to advance the state-of-the-science of solar power forecasting using a public-private-academic partnership. The project included cutting edge research, testing in several geographically and climatologically diverse high penetration solar utilities and ISOs, and wide dissemination of the research results to raise the bar on solar power forecasting technology. The partners included three other national laboratories (Brookhaven National Laboratory [BNL], the National Renewable Energy Laboratory [NREL], and the National Oceanographic and Atmospheric Administration [NOAA] – each funded separately); six universities (Penn State University [PSU], Colorado State University [CSU], University of Washington [UW], University of Hawaii [UH], University of Buffalo [UB] – funded by New York Independent System Operators [NYISO], and Stony Brook University [SBU] – subcontractor to BNL); industry partners, including a private contractor (Solar Consulting Services [SCS]); four forecast providers (Atmospheric and Environmental Research [AER], Global Weather Corporation [GWC], MDA Information Systems [MDA], and Schneider Electric/Telvent DTN); six utilities (Xcel Energy Services, Sacramento Municipal Utility District [SMUD], Long Island Power Authority [LIPA], New York Power Authority [NYPA], Hawaiian Electric Company [HECO], and Southern California Edison [SCE]); and four balancing authorities (NYISO, California ISO [CAISO], HECO, and Public Service Company of Colorado [PSCo]). We have formally included several affiliate partners, such as the U.S. Army Research Laboratory (ARL), Commonwealth Scientific and Industrial Research Organization (CSIRO) of Australia, and University of Jaén, Spain (Jose Ruiz-Arias). Various other organizations participated less formally or beta tested the software that was developed. This public-private-academic team advanced solar power forecasting through innovative research, which improved and developed both the necessary forecasting technologies and the metrics for evaluating them. The team conducted a year-long, full-scale demonstration of integrating enhanced technologies into utility operations and evaluating its usefulness.

The goals of this project were to:

- Build a solar power forecasting system to advance the state-of-the-science through cutting-edge research;
- Test the system with appropriate metrics in several geographically diverse, high-penetration solar utilities and independent system operators (ISOs); and
- Disseminate the research results widely to raise the bar on solar power forecasting technology.

The purpose of this report is to document the accomplishment of these goals.

Advanced technologies were incorporated into a prototype solar forecasting system that was tested in collaboration with solar plant developers, utilities, and ISOs in geographically diverse areas, including Long Island, Colorado, and coastal California. Each component of the system was verified and validated using cutting-edge evaluation techniques and metrics developed specifically for the project through in-depth interactions with stakeholders. The economic value of providing the forecasts was assessed. The system will be deployed within operational environments of plant

operators, utilities, and ISO partners, with engagement of commercial forecast providers who will tailor the methods to the needs of the deployment. The results are being widely disseminated.

The prototype system, which provides solar forecasts across a wide range of temporal and spatial scales, integrates or includes:

- Solar radiation and cloud measurements
- Satellite observations
- Observations from whole sky cameras
- Local meteorological observations
- Publicly available numerical weather prediction modeling results, including the HRRR model
- Customized version of WRF-Solar™, tuned for cloud prediction and assimilating specialized data
- Fast radiative transfer model
- Statistical blending of forecast technologies, tuned to each of a variety of prediction times ranging from less than 15 min to at least 48 h, and even extend to 168 hours
- Irradiance-to-power conversion models
- Uncertainty quantification for probabilistic forecasts
- Assessment metrics

Table 1-1 details the objectives, resources, and activities of each portion of the project and relates them to the project outcomes and impacts. The research and operational application of this effort made seminal advances in the state of cloud forecasting, and thus, solar irradiance forecasting, as well as integrating it into the grid and assessing the resulting value. Thus, the work outlined herein has improved the ability to integrate solar energy and advance higher penetration of renewable energy.

The system was tested in collaboration with geographically diverse solar plant developers, utilities, and ISOs, with the engagement of commercial forecast providers that will tailor the techniques and methods to the needs of the deployment. Demonstrations included both photovoltaics (PV) and Concentrated Solar Power (CSP) systems.

Finally, a hallmark of this project was wide dissemination of all research and development results through presentations at meetings, convening special sessions at targeted conferences, peer-reviewed publications, technical reports, and workshops for end-users and commercial providers.

Table 1-1. A logic model of this project demonstrating activity relevance.

Objectives	Resources	Activities	Outcomes	Impacts
Quantify value of solar forecasting	- Statistics - Economics - Solar variability info - Grid integration info	Research - Develop metrics - Develop baselines - Develop target values - Economic assessment	Demonstrate benefits and value of improved solar forecasting	- Improve decision-making based on solar forecasts. - Advance solar energy penetration due to added value of forecasts.
Improve short-term solar irradiance forecasts	- Whole sky imagery - Satellite imagery - Real-time radiation data - Power data - Advection algorithms - Multi-sensor Advective Diffusive (MADCast) system	Research - Image processing - Advection algorithms - Nowcasting - Regime identification	Better short-term solar irradiance forecasts	Improved ability to integrate solar energy into grid for reliability
Improve day-ahead solar irradiance forecasts	- Advanced Research WRF model - Assimilation methods - Improved cloud physics and dynamics - Field data	Research - Cloud development - Cloud evolution - Radiative transfer - Aerosol physics - Blending algorithms - Power conversion	Improved day-ahead solar irradiance forecasts	Make solar energy more economical in day-ahead trading
Determine economic benefits of improved solar irradiance forecasting	- Forecast providers - Utilities - ISOs - Instrumented sites	Demonstration - Integrate forecasts into operations - Apply metrics - Evaluate Input	Show that improved solar irradiance forecasting benefits stakeholders	Advance the penetration of solar energy through stakeholder buy-in
Disseminate knowledge to stakeholder community to advance the state of practice	- Project participants - Research results - Quantified benefits - Year's forecasting experience	Dissemination - Present results - Publish - Special sessions at conferences - Workshops - Technical reports	Transfer knowledge to broad cross-section of stakeholders	Stakeholder buy-in and optimal application of improved forecasting methodologies

1.2 State-of-the-Science

The NCAR-led team was already immersed in solar forecasting research at the time of the award, along with other research teams throughout the world. Real-time solar power forecasting is reviewed in chapters in a few recent books, including Kleissl (2013) and Troccoli, Dubus, and Haupt (2014). The issue is well motivated in works like Dubus (2014) among others. Lorenz et al. (2014) reviews the extensive work of the team at the University of Oldenburg in Germany. The Australian initiative is ongoing as motivated in Davy and Troccoli (2012). Schroedter-Homscheidt et al. (2013) point out the need for excellent aerosol prediction for solar power prediction and discuss techniques leveraging European Centre for Medium-Range Weather Forecasts (ECMWF) chemistry forecasts. The difficulties in predicting cloud cover at specific locations are well known.

Solar power prediction is accomplished by different techniques for differing time scales. Solar energy is particularly variable over space and time because of its myriad complexities caused by the dynamic evolution of clouds. Lew et al. (2012) provides evidence of the challenge of solar power integration with the results showing the variability of power output was higher with high penetrations of solar than with high penetrations of wind. The response speed (ramp rate and start time), response duration, frequency of use (continuously or only during rare events), direction of use (up or down), and type of control characterize a utility company's operating reserves (Ela et al. 2013). These operating reserves are appropriately managed with accurate solar forecasts, as Curtright and Apt (2008) have shown that the cost of energy can be strategically minimized with knowledge of the short- and long-term PV variations. The quantification of temporal solar irradiance variability caused by the dynamic evolution of clouds has been extensively studied. Hinkelman (2013) found that not only are the irradiances themselves larger in the middle of the day but also the fractional change in irradiance from one time to another is larger. She also determined that cloud optical depth and cloud height are the best predictors of irradiance variability at one-minute time resolution. Gueymard and Wilcox (2011) analyzed the regional dependence of solar power and showed greater variability tends to occur in coastal areas, particularly along the California coast and in mountainous areas because of the micro-climate effects of topography. Kuszmaul et al. (2010) analyzed 1-sec PV output data and showed that it is linearly proportional to the spatial average of irradiance. Rayl et al. (2013) performed an irradiance co-spectrum analysis and concluded that solar power site aggregation could greatly reduce power variability on short time scales depending on the distance between sites.

The non-linear variations of solar irradiance result from the complex evolution of clouds in the atmosphere; thus, many studies have tested non-linear solar irradiance prediction methods (Mellit 2008; Martin et al. 2010; Bouzerdoum et al. 2013; Fu and Cheng 2013; Marquez et al. 2013a; Inman et al. 2013; Fernandez et al. 2014; Chu et al. 2014). These studies, however, have not focused on the explicit prediction of both the temporal variability and the spatial variability of solar irradiance.

There have been multiple recent studies focused on the prediction of solar radiation or solar power with statistical learning (also known as artificial intelligence or data mining). Mellit (2008) provides a summary of techniques for forecasting solar radiation and states that 37 studies have used neural networks in the modeling and prediction of solar radiation with the second most frequent method, fuzzy logic, used five times. More recently, Martin et al. (2010) showed a final model based on Artificial Neural Networks (ANN) improves accuracy 4.84% to 25.58% over persistence for half-daily radiation forecasts. Fernandez et al. (2014) concluded that the ANN model has accurate performance for days characterized by direct irradiance (clear days) and for days characterized by diffuse irradiance (cloudy days). Chu et al. (2013) used an ANN with sky image processing to predict 1-min average direct normal irradiance (DNI) for time horizons of 5 and 10 minutes. Another short-term prediction study used a regression technique on all-sky images to predict solar radiation five minutes in advance with a mean absolute error of around 22% (Fu and Cheng 2013). Autoregressive techniques have also shown solar power prediction capability, with Bouzerdoum et al. (2013) using a hybrid seasonal autoregressive moving average and support vector model to predict hourly power output. All of these studies advanced prediction of solar irradiance, but none claim to be optimal.

Sky imaging is another approach to very-short-range solar forecasting. The team has been aware of high-quality work being done at the University of San Diego and elsewhere, and has interacted with Jan Kleissl of USCD and his collaborators on several occasions. His team's work with total sky imagers (TSIs) is reported in Urquhart et al. (2015) and Bosch and Kleissl (2013). The BNL approach is unique and different from what Kleissl and his team are doing. Huang et al. (2013) describe the initial methodology for nowcasting solar irradiance with a single TSI. There are three main steps to their algorithm: 1) TSI image preprocessing, 2) cloud motion estimation, and 3) solar radiation estimation. They showed positive results for their 1-minute and 2-minute irradiance forecasts. Peng et al. (2015) extended the BNL algorithm of Huang et al. (2013) to use three TSIs to better estimate the 3D cloud field. By comparing overlapping cloud regions in multiple TSIs (located several hundred meters to about 1-2 kilometers apart), estimates of the height of each cloud feature are derived using geometry. Knowing the height of each cloud feature greatly aids in estimating cloud motion vectors, and allows for improved tracking of clouds at different levels (i.e., cumulus vs. cirrus clouds) over the single-TSI algorithm. Peng et al. (2015) also developed a support vector regression model with a radial basis function to forecast solar irradiance based on recent TSI frames and pyranometer data. This new irradiance prediction model was shown to have improved error metrics compared to other radiation prediction models tested extendingto 3-minute forecasts. Even with multiple TSIs, however, the irradiance prediction horizon is limited to approximately 10-15 minutes when low cumulus clouds are present, as they will typically transit across the image scene in that amount of time. In situations where high cirrus clouds are present, however, irradiance forecasts could potentially be extended to 30-60 minutes, as they take longer to transit across the image scene. BNL continues these advances as shown in section 2.1.

Satellite based cloud prediction is another important method that lies between the very-short-range and the medium-range time scales. An overview of the current state-of-the-art in solar forecasting is provided in the book edited by Kleissl (2013), including details of physically based satellite methods for short-term forecasting, provided in Chapter 3 by Miller et al. The problem is rooted in our ability to utilize multi-spectral satellite imagery (preferentially from the geostationary constellation) to characterize the geometric and microphysical properties of meteorological clouds. Knowledge of cloud locations, heights, and properties can be used to estimate the down-welling solar irradiance (direct and diffuse components) at the cloud shadow locations. Provided information on cloud motion, the shadows can be propagated forward in space/time to provide a time-series of solar irradiance at a given location.

For longer time scales beyond about 3-6 hours, it is necessary to employ numerical weather prediction (NWP). The initialization of clouds in NWP models is a difficult problem that has recently received increased scrutiny. Current data assimilation methods are challenged by the high spatio-temporal variability of clouds, strong non-linearities in the radiative transfer calculation and simulated microphysical properties, and non-Gaussian error distributions. Model balance via ensembles of forecasts is affected by sampling error and systematic model errors. For these reasons, NWP forecasts are usually inferior to simpler advection methods in the first few hours of the forecast (i.e. nowcasting). But over longer timeframes, it is essential to include the cloud formation and dissipation that the NWP models provide.

Under cloudless or partly cloudy conditions, aerosols have a strong impact on surface irradiance, particularly its direct and diffuse components. Under such circumstances, the solar forecasting performance of NWP models, such as WRF or ECMWF, has been found to be highly biased

(Gerstmaier et al. 2012; Ruiz-Arias et al. 2012; Troccoli and Morcrette 2012). An appropriate way to handle aerosols is to apply aerosol transport models, such as put forth in the Monitoring Atmospheric Carbon and Climate (MACC) project (Schroedter-Homscheidt et al. (2013). The aerosol model developed at ECMWF (Morcrette et al. 2008) has a remarkable 3-hourly resolution, a relatively good spatial resolution (~120-km grid spacing), and benefits from the assimilation of Moderate Resolution Imaging Spectroradiometer (MODIS) observations. Outputs of this model are now commercially used in the prediction of solar irradiance on a global scale, with noticeable improvements in the resulting global horizontal irradiance (GHI) and DNI accuracy (Cebecauer et al. 2011a,b). Data from the MACC project must be validated over the U.S. Day-ahead high-quality aerosol forecasts and could be used to predict GHI and DNI with WRF. The NASA Goddard Earth Observing System, version 5 (GEOS-5), offers another real-time analysis and prediction of aerosols (Randles et al. 2013). Substantial improvements in day-ahead solar forecasts under cloudless or partly cloudy days, particularly over areas of large aerosol variability, can be expected from these developments.

In a recent comparison of various solar forecasting models for the U.S., Perez et al. (2011) showed that NOAA's operational models (that use WRF as the underlying numerical model) lagged other international forecasting models in terms of accuracy of their solar irradiance predictions. So far, this has been interpreted partly due to shortcomings in cloud modeling and data assimilation. It is possible that the radiative transfer algorithms in the U.S. forecast models are not optimal for this application. This hypothesis was confirmed by Ruiz-Arias et al. (2012) in the case of the WRF model. That study highlighted biases in one frequently used radiative algorithm in WRF, and a need for adding aerosol data for its improvement.

The conversion of irradiance to power depends on the particular type of hardware installed at the solar farm as well as local conditions. There are models, such as PVWatts (http://pvwatts.nrel.gov/) that can do this power conversion. In prior work with wind energy, we have found however, that empirical power conversion methods can outperform models (typically power curves for wind energy) because they take into account local effects such as terrain blocking, impact of upstream turbines, density, etc. (Parks et al. 2011). For solar energy, such effects could include dust, shadowing, etc., that cannot be captured in any general model. Thus, as part of this project, we have developed such empirical methods to use across a broad range of solar technologies and geographic locations.

Any forecast requires state-of-the-science evaluation. Traditional metrics are commonly used for evaluation, such as in Marquez and Coimbra (2011). Diagnostic methods provide information that is user-relevant, such as metrics related to ramp variability. Such metrics may include application of spatial methods such as "object-based" approaches that identify and compare characteristics of spatial objects (Davis et al. 2009). In addition, they may identify specific events in time, such as changes in magnitude of insolation at a point location. They also may identify rates of change, such as ramp rates (e.g., Mathiesen et al. 2012), and they may consider temporal and spatial attributes of the forecasts (e.g., Brown et al. 2012) that are often very relevant for decision-making based on the forecasts. Metrics related to economic and other benefits are the most complex to derive, but may be the most meaningful for end users.

In summary, although there have been some important recent accomplishments in predicting solar power, there is plenty of room for advancement. This Public-Private-Academic Partnership

worked to fill the gaps in the research and then used the research results to build a functioning, seamless Sun4Cast® forecasting system.

1.3 PROJECT PROGRESSION OVERVIEW

1.3.1 First Year Full Team Workshop – Cementing the Team

The first year workshop was held 12-13 March 2013. This was an opportunity to bring the team together at NCAR to think through how to integrate all of the research into a working Sun4Cast® system. The workshop was configured to emphasize meeting the needs of the user. After an initial introduction to the project goals, a charge from the DOE SunShot Director, and introductions from NCAR Research Applications Laboratory Director, the workshop commenced with a user panel of utility and ISO representatives to explain how they use forecasts and what they need in the forecast, as well as when it must be delivered to be most useful. We saw this session as "beginning with the end in mind" as a way to picture the project outcome. The next panel was comprised of the forecast provider partners, allowing them an opportunity to discuss how they work with the end users to meet their needs, as well as with researchers to advance their systems.

With this background from the users and providers of the forecasts, the afternoon of March 13th was devoted to developing a shared conceptual model of the value chain. The group broke into five pre-assigned teams that mixed forecast users, providers, and researchers to develop mental models of the forecast value chain, with our economist, Jeff Lazo, providing guidance to this social science process. Each group was assigned a moderator and a rapporteur (see Table 1-2) The objectives of this exercise were:

- Team building
- Building a qualitative model of the weather-solar value chain
- Facilitating discussion and enhancing understanding across all participants in the project
- Explaining how research to improve forecasts will create value

The value chain guidance provided by Dr. Lazo appears as Figure 1-1. All team members were given general guidance to spend the first 5-8 minutes "drawing" their own value chain, considering issues such as:

- What values / decisions / outcomes do you think are important to end-users / decision-makers?
- How does weather impact those?
- How does weather information relate to those?
- How would changes in / improvements in weather forecasts change those outcomes?
- Who are the decision-makers?
- What are their needs, resources, constraints?
- How do different "agents" in the value change add value to information?
- What if this project improved the relevant forecast by x%?
- What is the relevant forecast information?
- What does an x% improvement mean?

- How does an x% improvement affect outcomes for weather forecast vendors, utilities, ISOs, and regional transmission operators (RTOs)?

Then, the team members were asked to discuss these issues as a team and work toward producing a unified value chain. At the end of the session, the rapporteur and moderator met over the break and developed a 3-5-minute summary of the conceptual model and outcome of their session, including

- Areas of agreement, disagreement, …
- Gaps, misunderstandings, points of confusion, …
- Issues to discuss across the broader group

We found that this approach was quite successful for team building and enabled the group to come to a joint visualization of the project goals. Hence, the following day, we delved deeper into the elements of the forecasting systems and determined how to fit them into one cohesive whole. Figure 1-2 illustrates a more complete vision that fits the value chain tothe elements of the project. The second day of the workshop included break-out discussion groups configured to bring together specific teams on the project. The project progressed with five primary teams that discussed their research and advances at least monthly. These teams were 1) Nowcasting, 2) Numerical Weather Prediction, 3) Engineering, 4) Metrics, and 5) Management (that includes all team leads). This was an effective way to manage the flow of the project.

Table 1-2. Teams assigned for the Mental Modeling exercise at the first year workshop.

Team – Room	Moderator	Rapporteur
Yellow Team	Tressa Fowler	Kirsten Orwig
Red Team	Bill Mahoney	Melinda Marquis
Blue Team	Tara Jensen	Louisa Nance
Green Team	Jeff Lazo	Jennifer Mahoney
Weather Team	Luca Delle Monache	Tom Auligné

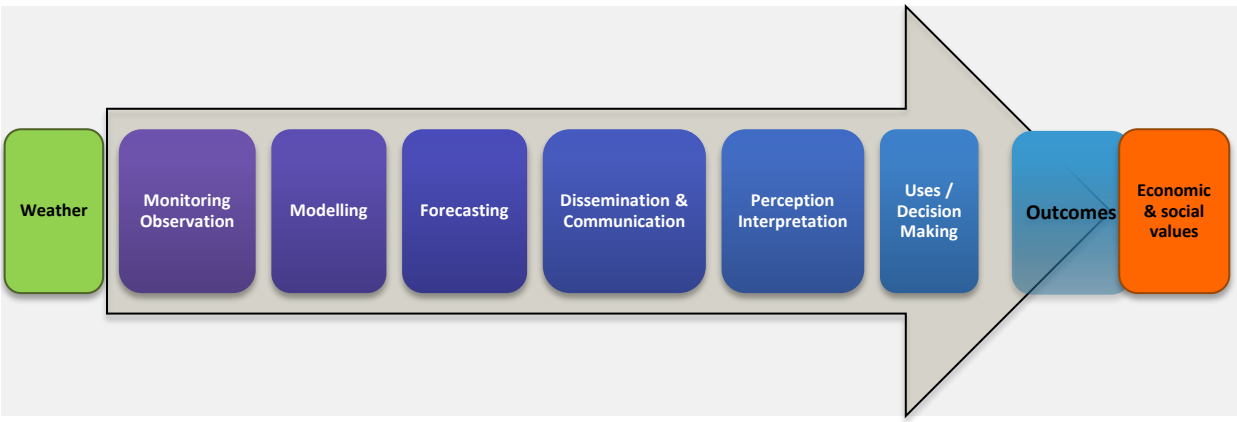


Figure 1-1. A general value chain describing the process leading tog value to the end user.

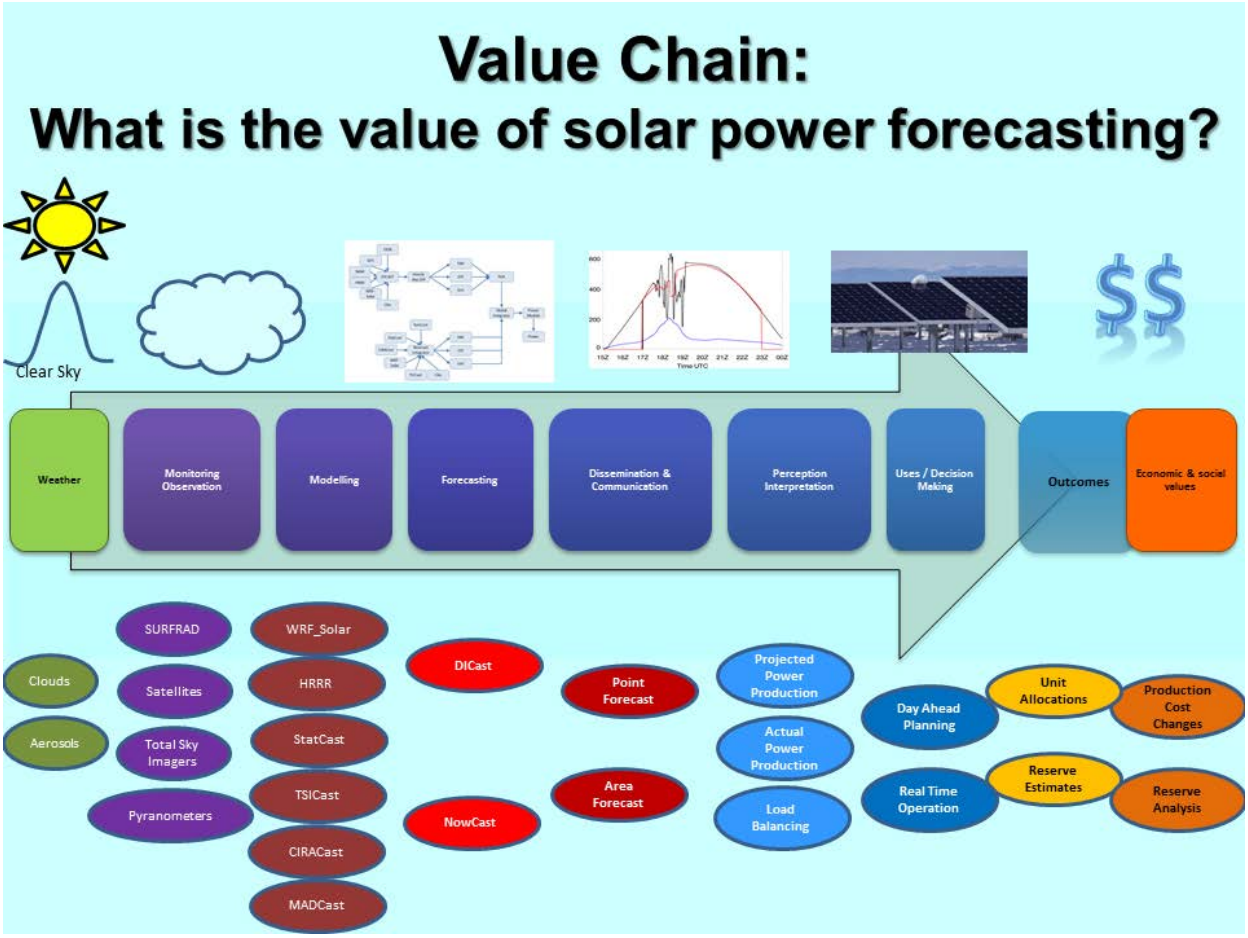


Figure 1-2. Value chain implementing a weather decision support system for solar power. At the bottom are the components of the NCAR team's system that build toward providing an economic impact of this system.

1.3.2 Metrics Development

The first year was devoted to research on individual portions of the system using the team approach discussed above. Development of metrics was accomplished jointly with a collaborative team that included DOE SunShot Leadership, the IBM Watt-Sun forecasting team, and NOAA team members. That group held several workshops that included end users associated with meetings of the Utility Variable Generation Integration Group Forecasting conferences and Annual Meetings of the American Meteorological Society. With that input and many team telecons, they designed a table of proposed metrics to exercise (Table 1-3). More details of metrics and their applications appears in section 5.2.

Table 1-3. Consensus of metrics to be exercised in evaluating solar power forecasting systems.

	Model-Reference Comparison	Utility Planning/Operations Support	
	Statistical Information	Statistical	Economic/Value
Base	BC1: Distribution of Forecast Errors BC2: Mean Absolute Error BC3: RMSE BC4: Standard Deviation / Variance BC5: Pearson's Correlation Coefficient BC6: Categorical Statistics for Event BC7: Frequency of Superior Performance	BP1: Mean Bias Error BP2: Skewness BP3: Kurtosis BP4: 99th Percentile	BV1: Operating Reserves Analysis BV2: Electricity Production Cost Analysis
Enhanced	EC1: Kolmogorov-Smirnov Test Integral EC2: OVER Metric EC3: Renyi Entropy EC4: Paired Test for Mean and Variance EC5: Performance Diagram for Categorical EC6: Taylor Diagram for Continuous Statistics	EP1: Probability Interval Forecast Evaluation EP2: Brier Score EP3: Receiver Operator Characteristic Curve and Area EP4: Reliability Diagram	EV1: Electricity Load Payments Analysis EV2: Solar Generation Curtailment EV3: Power Trading Impact P1: Load Forecast Improvement P2: Storage Optimization

In parallel, the NCAR Metrics team worked with our utility stakeholders and discussed methods to assess value provided by improved forecasting. The economic metrics developed are illustrated in Table 1-4 and discussed in more detail in section 5.3.

Table 1-4. Summary of feedback from utility partners on Economic/Reliability Metrics teleconference. Green – can be calculated and expected to be impacted by improved forecasts. Orange – difficult to calculate and/or less impacted by improved forecasts. Yellow – investigating how to calculate. White – no response.

ID	Metric	NY partners (NYISO, LIPA, NYPA)	SMUD	Xcel
BV1	Operating Reserves	See Regulation Reserves (ER2)		
ER1	Non-Spinning Reserves	Not applicable because non-spinning not centered on renewables		
ER2	Regulation Reserves			
EV1 & ER3	Production Costs			
EV2	Electricity Load Payment			
EV3 & ER3	Solar Generation Curtailment	Not applicable until more penetration		
EV4	Power Trading Impact	Not applicable until more penetration and in Day Ahead market		
ER4	NERC CPS2 Performance	Same as Regulation Reserves (ER2)		
ER5	Distribution Substation/Feeder Voltage Stability	Too Complex and forecasts not at feeder level – see P2 for suggested change		
ER6	Increased Committed Solar Power Generation	See Power Trading Impact		
P1	Load Forecasting Improvement			
P2	Storage Optimization			

1.3.3 Parallel Development of System Models

The science teams (Nowcasting and Numerical Weather Prediction) worked in parallel during thefirst year to develop their systems to the point where they could be integrated. The results of the Nowcasting research and the WRF-Solar™ day-ahead system are described in chapters 2 and 3 respectively. Figure 1-3 illustrates our view of how these Nowcasting and NWP systems work together to produce a seamless forecast across scales.

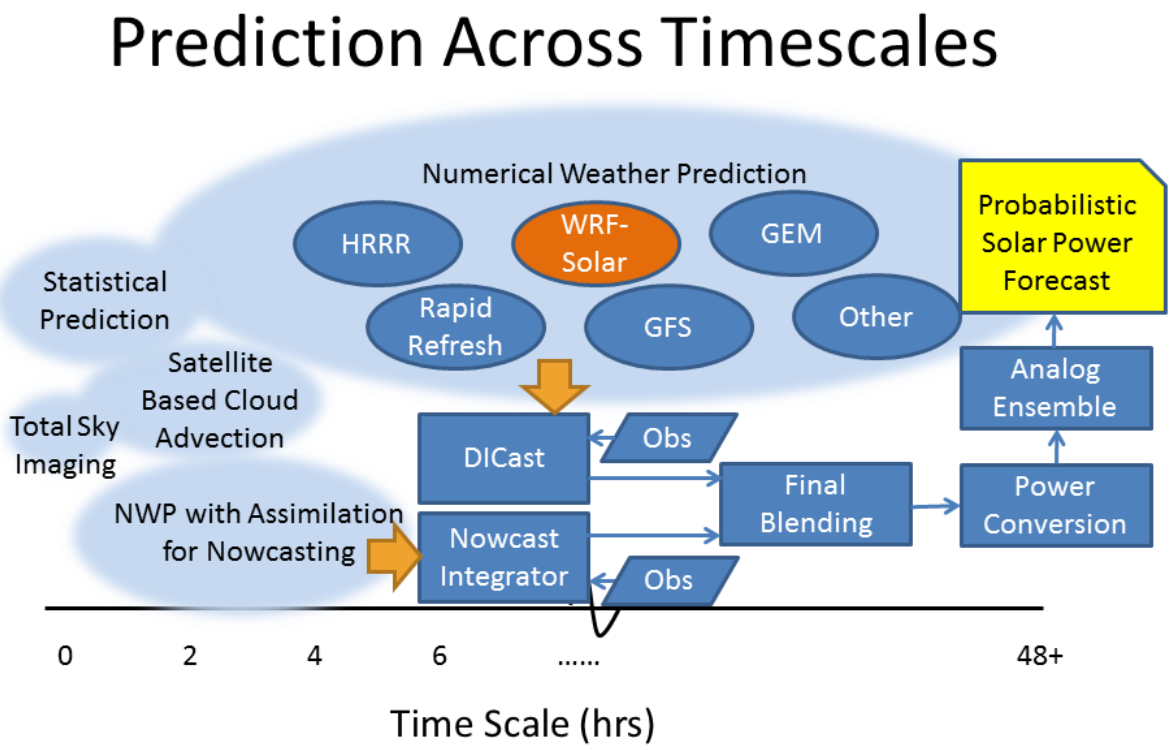


Figure 1-3. Sun4Cast® forecasting system predicts across scales.

1.3.4 System Integration

It is a large task to integrate the various models described in this document into the Sun4Cast® System. That was the job of the engineering team, comprised of professional software engineers who are skilled in bringing together multiple models and observations and configuring a system to optimize its predictive skill. More details of their systems appear in chapter 4. Part of that work includes computing the uncertainty in the forecast. This is a Big Data problem as described in more detail in Haupt and Kosovic (2015, 2016). Section 4.6 describes application of the analog ensemble (AnEn) method to this problem.

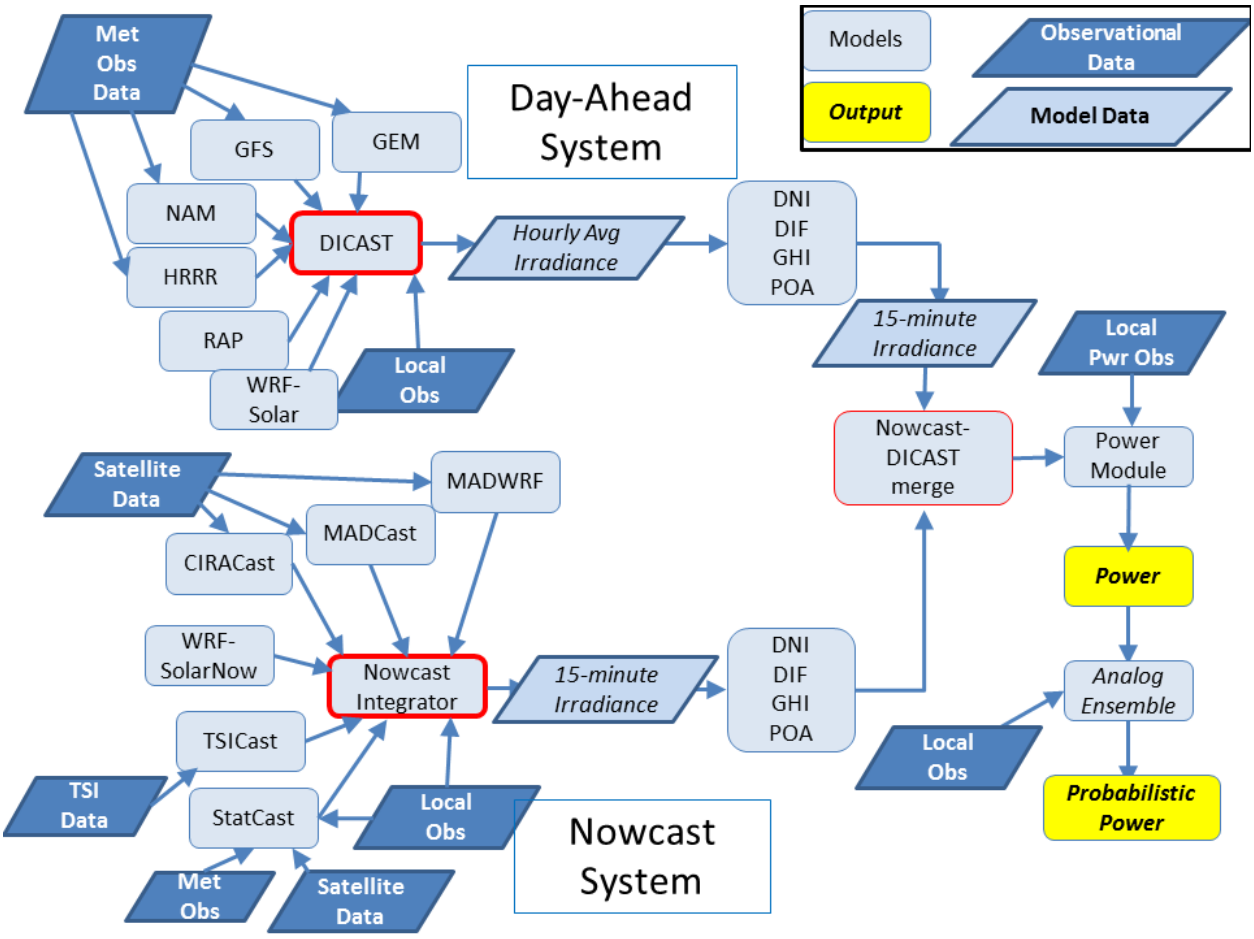


Figure 1-4. The engineered Sun4Cast® system.

1.3.5 Second Team Workshop

After each team spent about 15 months in research, model construction, testing, and tuning the models, the system components were brought together and tuned as an integrated system. A second workshop was held 26-27 August at NCAR in Boulder, CO, to discuss the advances and work yet to be accomplished (see Figure 1-5). Researchers presented details of their models and results of their initial testing. There were opportunities to discuss these advances and next steps in break-out groups, panel discussions, and in the large group. The final workshop was a panel discussion with a panel comprised of end users. Industry representatives were afforded an opportunity to provide their thoughts and feedback, formulating guidance for the team to use in the rest of the project and tuning the integrated system.

Figure 1-5. Vignettes from the second-year workshop (photos by Jared Lee).

1.3.6 Quasi-operational Demonstration

The final portion of the project was spent consistently operating the system in several key utility regions with varying climates and geography. Figure 1-6 maps these regions and the utility partners who participated in the exercise.

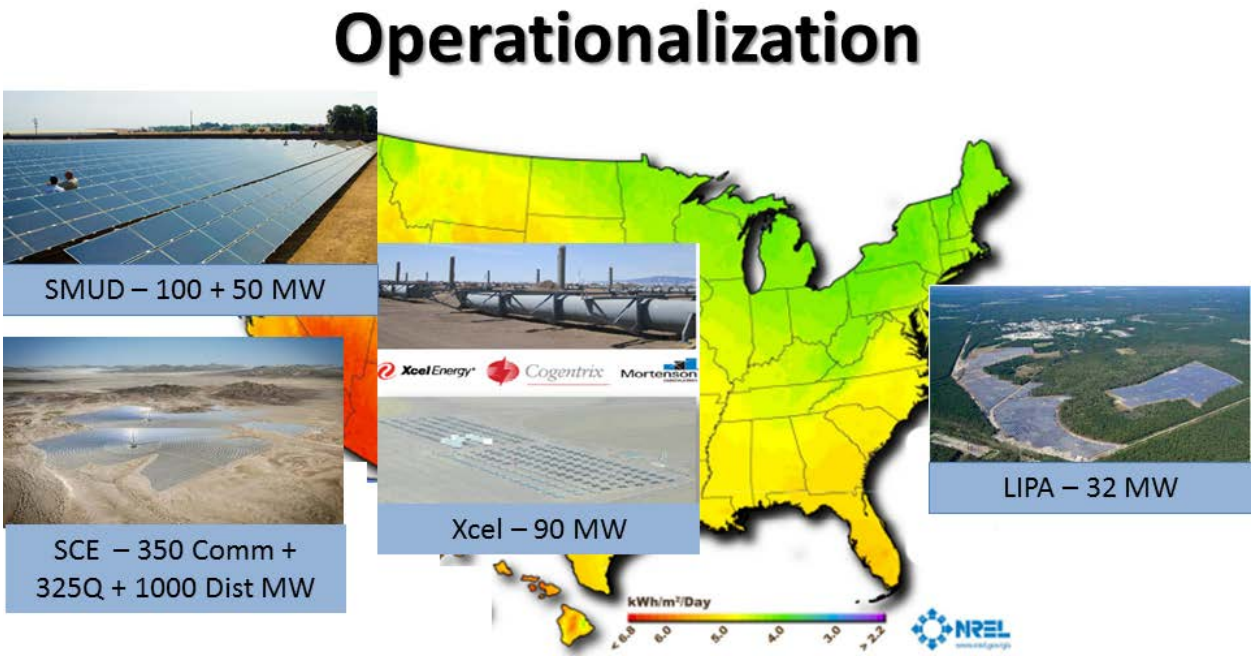


Figure 1-6. Regions of quasi-operational Sun4Cast® demonstration.

1.3.7 Formal Assessment and Documentation

Throughout the project, the metrics team performed periodic assessments. These assessments provided excellent feedback for the model research teams so that they could further improve their systems. At the end of the project, a formal project assessment wasaccomplished. The results of that assessment appear in chapter 5.

1.4 ACCOMPLISHMENTS

The team demonstrated and evaluated a working Sun4Cast® solar power prediction system that includes the multiple components described herein, including WRF-Solar™, multiple models from national centers, TSICast, CIRACast, and MADCast, MADWRF, as well as statistical models. The individual components and the overall Sun4Cast® system werevalidated using the metrics developed at the beginning of the project. The team met or exceeded target values specified in most of the milestone tables in the statement of work to DOE. Data streams from various model systems were made available to the forecasting partners, forecasts were regularly provided to the utility and ISO partners, and feedback from the partners was incorporated into the forecasting models.

The Sun4Cast® system (Figure 1-4) has two main forecast tracks: a Nowcast track that forecasts at high temporal resolution extending to 6 hours, and a DICast® track that forecasts at coarser temporal resolution out several days. Both these modules apply a consensus forecasting approach. That is, they consider multiple inputs and perform a forecast integration that takes advantage of the strengths of each input. While the consensus forecasting approach has been applied to forecasting more common weather variables (e.g., air temperature), in the past it had not previously been applied to solar irradiance forecasting in any significant way. No other public systems use a

consensus forecasting approach. In the private sector, some companies may use a consensus approach, while others rely on a single-source model; much of this is proprietary and not disseminated. Forecasts are provided every 15 minutes and extend to 72 hours and can be provided as far out as 168 hours.

The team conducted transformative research in statistical forecasting, advective/dynamic short-range forecasting, nowcasting with real-time data assimilation, satellite techniques and data assimilation for solar forecasting, numerical weather prediction with the WRF-Solar™ model (including cloud physics parameterization, convective parameterization, clear-sky aerosol estimation, and radiative transfer modeling), radiation-to-power conversion, and uncertainty quantification.

Much scientific and engineering progress was made. The rest of this report details the results. Conclusions, lessons learned, networks fostered, and recommendations of best practices in solar power forecasting appear in chapter 6.

1.5 Project Overview Publications

Book Chapters

Haupt, S.E., P.A. Jiménez, J.A. Lee, and B. Kosovic, 2016: Principles of Meteorology and Numerical Weather Prediction, in Renewable Energy Forecasting: From Models to Applications, G. Kariniotakis, Ed., Elsevier, London, UK. Submitted and in review.

Journal Papers

Haupt, S.E. and B. Kosovic, 2016: Variable Generation Power Forecasting as a Big Data Problem, submitted to IEEE Transactions on Sustainable Energy.

Conference and Workshop Presentations (presenter in Bold)

Haupt, S.E., 2016: Comparison of Solar Power Forecasting Techniques, Joint Session between 14th Conference on Artificial and Computational Intelligence and its Applications to the Environmental Sciences and Seventh Conference on Weather, Climate, Water, and the New Energy Economy, AMS Annual Meeting, New Orleans, LA, Jan. 12.

Haupt, S.E., 2016: Integrating and Operationalizing Renewable Energy Forecasts: It Takes a Community, Seventh Conference on Weather, Climate, Water, and the New Energy Economy, AMS Annual Meeting, New Orleans, LA, Jan. 11.

Haupt, S.E. and B. Kosovic, 2015: Big Data and Machine Learning for Applied Weather Forecasts: Forecasting Solar Power for Utility Operations, IEEE Symposium Series on Computational Intelligence, Capetown, South Africa, December 9. Fully reviewed paper.

Haupt, S.E., 2015: The Sun4Cast Solar Power Forecasting Decision Support System, American Solar Energy Society Conference, State College, PA, July 28.

Haupt, S.E., S. Drobot, T. Jensen, 2015: The Sun4Cast Solar Power Forecasting System. International Conference on Energy and Meteorology, Boulder, CO, June 23.

Vaucher, G., S.E. Haupt, D. Sauter, 2015: A Review of Atmospheric Forecasting Tools being Developed for Renewable Energy, 83rd Military Operations Research Society Symposium [MORSS] in Alexandria, VA, Jun 22-25.

Haupt, S.E., 2015: Renewable Energy Needs, Rapid Update Analysis/Nowcasting Workshop, NOAA ESRL, Boulder, CO, June, 4.

Haupt, S.E., 2015: NCAR's Solar Power Forecasting Research, California Utility Forecasting Meeting, Folsom, CA, April 29 (invited).

Haupt, S.E., 2015: Counting on Solar Production: Advances in Forecasting, Utility Solar Conference of the Solar Electric Power Association, San Diego, CA, April 28 (invited).

Haupt, S.E., 2015: Solar Power Forecasting: Sun4Cast and GRAFS, Utility Variable Generation Integration Group Forecasting Workshop, Lakewood, CO, Feb. 19, 2015.

Haupt, S.E., 2015: The Sun4Cast Solar Power Forecasting System, Joint Session between Sixth Conference on Weather, Climate, and the New Energy Economy and 13th Conference on Artificial Intelligence, AMS Annual Meeting, Phoenix, AZ, Jan. 7.

Haupt, S.E., B. Kosovic, and S. Drobot, 2014: Advances in Solar Power Forecasting, Fall Meeting of the American Geophysical Union, San Francisco, CA, Dec. 15.

Haupt, S.E. and S. Drobot, 2014: New Irradiance Models for Solar Energy, Solar 2014 sponsored by the American Solar Energy Society, San Francisco, CA, July 7. (full paper).

Haupt, S.E. and S. Drobot, 2014: A Public-Private-Academic Partnership to Advance Solar Power Forecasting, SunShot Summit, Anaheim, CA, May 20. Invited Poster Presentation and Review.

Haupt, S.E., 2014: Renewable Energy, UCAR Research and Partnership Meeting, Boulder, CO, April 22.

Haupt, S.E., 2014: NCAR-led SunShot Solar Forecasting Project, Utility Variable Generation Forecasting Workshop, Tuscon, AZ, Feb. 26. Invited Panel Presentation.

Haupt, S.E., 2014: Advances in Predicting Solar Power for Utilities, Fifth Conference on Weather, Climate, and the New Energy Economy, AMS Annual Meeting, Atlanta, GA, Feb. 6.

Haupt, S.E., 2014: Using Artificial Intelligence to Inform Physical/Dynamical Models, 12th Conference on Artificial and Computational Intelligence and its Applications to the Environmental Sciences, Invited Panel Presentation, Feb. 3. Invited Panel Presentation.

Haupt, S.E., 2013: A Public-Private-Academic Partnership to Advance Solar Power Forecasting, International Conference on Energy and Meteorology, Toulouse, France, June 25.

Haupt, S.E., 2013: A Public-Private-Academic Partnership to Advance Solar Power Forecasting, American Solar Energy Society Meeting, Baltimore, MD, April 18.

Haupt, S.E., 2013: A Public-Private-Academic Partnership to Advance Solar Power Forecasting, Utility Variable Generation Integration Group Workshop on Variable Generation Forecasting Application, Salt Lake City, UT, Feb. 27. (Invited panel presentation)

Haupt, S.E., 2013: A Public-Private-Academic Partnership to Advance Solar Power Forecasting, AMS Solar Metrics Workshop, AMS Annual Meeting, Austin, TX., Jan. 9.

Short Courses and Workshop Presentations

Haupt, S.E., 2015: Short-range Weather Forecasting (hours to days) for Energy Applications. International Conference on Energy and Meteorology Pre-Conference Seminar, Boulder, CO, June 22.

Haupt, S.E., 2015: Introduction to Probabilistic Forecasting, Utility Variable Generation Integration Group Tutorial on Stochastic Forecasting Methods and Applications, Lakewood, CO, Feb. 18, 2015.

Haupt, S.E., 2014: NCAR's Research including Renewable Energy, Kuwait Institute for Scientific Research presents Workshop on Solar Resource Assessment, Kuwait City, Nov. 17, 2014.

Haupt, S.E., 2013: Meteorological Forecasting I: Some Basic Considerations for Atmospheric Modeling, COST Weather Intelligence for Renewable Energy Summer School, Montegut, France, July 1, 2013.

Haupt, S.E., 2013: Meteorological Forecasting II: Predicting Atmospheric Realizations: Dealing with Uncertainty in Applied Meteorology, COST Weather Intelligence for Renewable Energy Summer School, Montegut, France, July 1, 2013.

Haupt, S.E., 2013: What is your mental model of using meteorological uncertainty information for energy? Workshop on Uncertainty in Meteorology for the Energy Sector, Preconference Seminar, International Conference on Energy and Meteorology, Toulouse, France, June 24.

Haupt, S.E., 2013: How can we better facilitate using meteorological uncertainty information for energy? Workshop on Uncertainty in Meteorology for the Energy Sector, Preconference Seminar, International Conference on Energy and Meteorology, Toulouse, France, June 24.

2 NOWCAST SYSTEMS

In this chapter we present details of each of the component models in the Sun4Cast® nowcasting system: TSICast, StatCast, CIRACast, MADCast, WRF-Solar-Nowcasting, and MAD-WRF. Additionally, in section 2.7, several case studies are analyzed, comparing predictions made by some of these systems.

2.1 TSICAST

The project team members at BNL developed the TSICast model to predict GHI in the very short-range, extending to a few minutes, relying primarily on digital photographs from total sky imagers (TSIs). TSICast is described in detail in this section. The BNL team was led by Dantong Yu, Zhenzhou Peng, Dong Huang, John Heiser, Shinjae Yoo, and Paul Kalb.

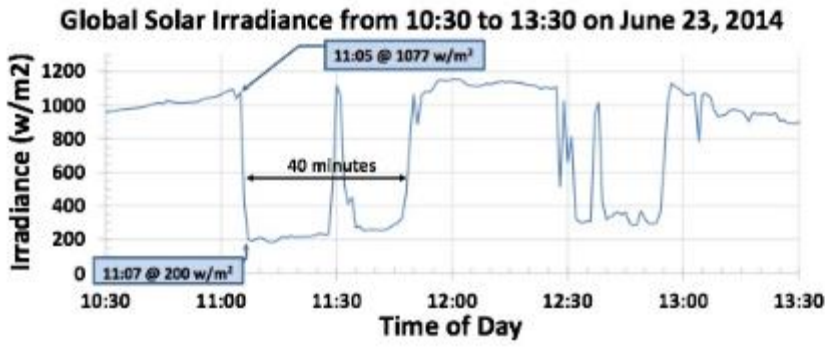
2.1.1 Motivation

Accurately predicting solar energy production, especially in short-term and mid-term horizons, becomes increasingly important for solar photovoltaic (PV) power plants. Because the sun-occlusion effect from clouds is the primary source of large fluctuations in the energy output of commercial solar farms, the ability to estimate cloud-induced variability in solar irradiance is critical to enable real-time or near real-time load shedding, balancing, and dispatching.

Although state-of-the-art techniques often rely on expensive meteorological instruments or remote sensing technology, e.g., satellite imagery to track clouds and forecast solar irradiance, some of these methods estimate clouds at a coarse granularity, such as spatial scales of 1 km or larger and temporal scales of 30 minutes or longer. However, an ever-growing number of distributed rooftop solar panels and storage solutions in smart-grids operate in a much higher spatio-temporal resolution, and so this leads to a pressing need for very short-term solar forecasts that predict minutes- or even seconds-ahead solar availability and variability. To fully address the problem, in the work detailed in this section, we integrate multiple sources of images collected from satellite imagers and ground-based cameras, and perform cloud tracking and modeling for more accurate solar energy prediction.

The intermittency and variability of irradiance compromises the reliability of solar plants and limits the adoption of solar power by utility companies. In many regions, the short-term variability within several minutes is the chief concern among utilities and balancing authorities regarding solar energy. For example, ground solar irradiance can decline by more than 80% due to cloud interference. Our data, collected from Long Island Solar Farm (LISF), New York, shows that GHI may drop over 500 W m-2 within 5 minutes, and that such ramp events occur during one out of every three days on average (Figure 2-1 and the accompanying table). Cloud-induced variability in solar radiation has become one of the biggest concerns for the operators of the power grid, particularly as the market share (penetration) of solar energy has ramped up steadily in recent years. Consequently, the ability to predict the presence of clouds and to extract their relevant features is critical for estimating the variations in solar energy, and thereby, for mitigating the effects of the fluctuations in output at utility-scale PV plants. ISOs require a cost-effective solution to manage and control the solar power intermittency. Balancing the supply-demand interaction while reducing the operation cost of electricity grids is critical to maximize the benefits of solar

energy and to ensure the quality of service for customers. As solar energy penetration continuously increases, the short-term variability of solar irradiance becomes an increasingly larger concern. Multiple decisions, such as to charge and discharge energy storage, to dispatch spinning reserves, and to merge production into smart grids in order to mitigate the instability of solar power, requires accurately estimating the ramp events of solar irradiance.



GHI (W/m^2)	2012 (days)	2013 (days)
> 100	306	311
> 200	259	271
> 300	210	224
> 400	173	185
> 500	122	137

Figure 2-1. Volatile ramp events frequently occur within one hour at BNL's 32-MW LISF.

Predicting ramps of solar irradiance is essentially equivalent to tracking and forecasting cloud coverage and sun-occlusion effects. Consequently, the capability to predict the presence of clouds and extract their relevant features is critical for estimating the variability in solar irradiance and for mitigating the effects of output fluctuations in utility-scale PV power plants. Cloud information such as cloud fraction and type can be obtained via NWP models in the hours- to days-ahead time frame. However, NWP models do not provide information on the location and movement of clouds at a hyper-local scale (order of a few km) for short-term or real-time grid operations. Cloud motion is the primary factor to be considered and estimated in our project. Cloud motion (velocity) is assimilated as a prominent input feature in most forecast models and isused to find spatio-temporal correlations between predicted clouds and solar irradiance fluctuations in the very near future. A robust methodology of tracking cloud movements in different time frames and data sources is urgently needed for modern solar forecasts.

Cloud motion tracking is a well-studied field in meteorological and remote sensing research. Cloud movement is usually represented by wind field and can be detected with atmospheric numerical models. However, a numerical model is still constrained by its resolution and is not suitable for real-time cloud tracking and minutes-ahead prediction. With the advance in imaging technologies, for example, satellite imagery or upward-facing cameras at ground level, a data-driven approach provides the necessary resolution and is much faster than NWP models. A stream of sky images provides both detailed snapshots, spectral and temporal information of clouds, and allows modern techniques of image registration and machine learning to track and predict cloud motions (Hoff and Perez 2012; Lave and Kleissl 2013). A variety of imager systems leads to requirements for cloud tracking algorithms to function by using observations from multiple sources with different temporal and spatial scales. Using multiple images offers new opportunities, for example three-dimensional tracking, accurate cloud motion detection, and minimal false estimates.

2.1.1.1 Technical Challenges

Multiple Layers of Clouds

Individual clouds must be clearly identified in images and, more importantly, located in three-dimensional space so as to provide faithful spatial correlation with ground-based measurements and actual production. However, clouds usually have complex vertical distributions and multiple layers, with the base height typically ranging from hundreds of meters to several kilometers above ground level. More importantly, each layer of clouds has distinct textural patterns, and wind velocity usually changes with height. Therefore, the sun-occlusion effect is extremely challenging to track due to the distinct layers and multiple movement patterns.

The core aspect of multi-layer cloud identification is to determine the cloud motions and height of each layer on images. When there exists only one image source, only planar 2-D information is available, and each layer must have a distinct motion pattern to differentiate layers. Any tracking algorithm that uses only single images generates inaccurate results when tracking non-rigid objects, such as clouds (Weiss 1997). Multiple image sources provide necessary depth information to determine cloud layers with a higher accuracy, by allowing the identification of clouds directly in a 3-D space. Furthermore, satellite imagery has multiple channels, each of which has different spectral sensitivity and detects cloud textural patterns under different spectra, providing the information necessary to distinguish vertical layers.

Complex Cloud Tracking

Visualization techniques can be applied to detect movements of clouds from various types of images, such as visual channels of satellite images or sky images obtained at ground level, and to generate accurate pixel-wise motion vectors to represent real spatial movements. However, given the complex physics governing cloud motion and life cycles, in addition to the texture and non-rigid shapes of clouds, tracking cloud motion for the purpose of solar prediction is a challenging problem for both computer vision and atmospheric science research.

The first issue is the variation of cloud type and color in imagery. It is extremely difficult to treat clouds as objects and detect their movements across frames based on image segmentation (Shi and Malik 2000), because observed clouds belong to different types, have different shapes, and can have obscured edges. Therefore, traditional motion tracking methods that rely upon image segmentation tend to generate inconsistent results, and need to be tuned carefully for different images and cloud conditions. In practice, tracking models are customized and vary with the different cameras in use and their resolutions. None of previous approaches have taken these types of cloud information into consideration and adjusted automatically for different cloud conditions.

Another factor that impairs tracking is cloud deformation or the arbitrary shape change with cloud. Existing cloud tracking models tend to over-simplify the problem and assume constancy in cloud shapes and velocity over a given time window. These methods work well for scattered clouds that do not undergo significant shape changes or vertical motion. However, more complex cases, such as cloud deformation and multiple cloud layers with different motions, greatly deteriorate detection accuracy.

Solar Irradiance Modeling

Once we track clouds and predict their future locations, subsequently we need to extract additional significant features from cloud images, along with solar irradiance measured by ground-based pyranometers, and estimate the potential solar irradiance value at a future time. The most significant challenge remains how to associate spectral and textural features of clouds to a particular solar irradiance value. Clouds tend to have volatile shapes and various luminance patterns over a sequence of images. Moreover, different types of clouds have distinct absorption and reflection characteristics in the solar spectrum, and directly impact the irradiance at ground level. Consequently, a rigorous method is needed to determine relevant features of clouds in images, especially those in the area interfering with sunlight. Because of instability from noisy pixels and luminance variation on images, any method needs to consider global features as well. We need to develop a new solar irradiance model that integrates both image features for prediction and actual ground-based irradiance measurements for regression, cross-validation, and model training.

2.1.2 Literature Review

Many motion-tracking techniques have been proposed to detect the motion of objects in terms of the pixel-wise movement across different images. However, only a few of them can be used to track cloud motion because of its non-rigid shape and formation. In the past, cloud motion vectors were usually obtained from satellite images (Leese et al. 1970; Cote and Tatnall 1995; Evans 2006; Corpetti et al. 2008). With the advent of inexpensive digital cameras and the emerging need for solar forecasts with the fine granularity that is beyond the spatio-temporal resolution of satellite images, recent research focused on using these ground-based cameras to track the very short-term motions of clouds (Chow et al. 2011; Wood-Bradley et al. 2012; Huang et al. 2013). These methods fall into three main categories based on the scale and tracking criterion of the motions. Here we discuss representative works in each category in terms of their approaches and the adopted optimization methods.

2.1.2.1 Optical Flow Based Motion Tracking Methods

In the field of computer vision, motion tracking is usually resolved by estimating the optical flow (OF), i.e., the pixel-wise distribution of prominent velocities of brightness/texture patterns on an image. In general, an OF method can obtain dense motion vectors at the granularity of a pixel, and was proven to be quite effective in detecting cloud motions in satellite images (Corpetti et al. 2008). Horn and Schunck (1981) (hereafter HS81) and Lucas and Kanade (1981) (hereafter LK81) proposed the original approaches of optical flow. The HS81model formulates optical flow as an optimization problem assuming that motions have global smoothness. The advantage of this model is that it can propagate information over a large distance within the image and fill texture-less regions, e.g. thin or stratus clouds, with the motions extracted from the boundaries. Consequently, it is widely incorporated into other variational models and customized for estimating cloud motions (Héas et al. 2007). Another state-of-the-art methodology, the LK81 model, constrains motions and preserves local smoothness, thereby satisfying the gradient constraint among neighboring pixels or in a pre-defined region. In particular for sky imagery, the LK81 method allows us to identify the dominant motion vector within cloud/sky regions, and therefore, is less sensitive to image noise. Wood-Bradley et al. (2012) adopted the LK81 methodology to estimate cloud motion on images captured by a laptop camera. Instead of using the brightness

intensity or the grayscale image, they converted the original image to the scale of blue-red difference to emphasize its prominent edges and corners. Wood-Bradley et al. (2012) subsequently extracted the prominent features at the points of interest with a large gradient of intensity for calculating optical flow, and then manually removed the noisy features to refine the tracking result.

Many variants of the classic models of LK81 and HS81 were developed to improve the quality of optical flow. It is beyond the scope of this technical note to review the entire literature, and to search for appropriate techniques for the purpose of cloud tracking. Hence, we focus on several typical models that can effectively address the problem of cloud tracking. Chow et al. (2015) utilized a well-known *global* smoothing method in optical flow to detect cloud motions on the basis of a new sky imager system (USI; Urquhart et al. 2014). They also adopted the robust estimation framework that originally appeared in Black and Anandan (1996) (hereafter BA96). Compared with their previous work on estimating the block-wise motion (Chow et al. 2011), the robust optical flow approach, BA96 is claimed to achieve more than a 19% improvement in cloud forecasting. Most optical flow models, including BA96, adopted a coarse-to-fine warping scheme for constructing a multi-scale image pyramid and iteratively looping through the layers of images for stable outputs of optical flow. However, in practice, the coarse-to-fine heuristic quite often does not perform well, particularly for the large movement of objects between two frames. The resulting optical flow fails to faithfully represent the real motion (Brox and Malik 2011). As fast-moving clouds are common in our observations, we therefore looked into the large-displacement optical flow (LDOF) (Brox and Malik 2011) and compared it with other classic models.

2.1.2.2 Block-wise Motion Tracking Methods

A block-matching method tries to search for the best-matching blocks from two consecutive images that maximize the pre-defined criterion of similarity. Typically, the criterion of similarity in block-matching techniques is based on cross-correlation or least-square errors, and helps find the disparity vector between the same reference block across two images within a constrained window. Different from the OF methodology that assumes constancy in brightness at the pixel level, block-matching techniques assume that the reference block retains textural constancy over time.

Block-matching methods long have been in existence for cloud tracking in satellite imagery (Leese et al. 1970; Evans 2006; Rutledge et al. 1991). Most used cross-correlation to find the maximum correspondence of regions/blocks between two consecutive image frames (Hamill and Nehrkorn 1993). Chow el al. (2011) adopted cross-correlation as the matching criterion to detect sparse cloud motion vectors (CMVs) in the images of a TSI 440A, dividing the original TSI image into the blocks of identical size. As a TSI generates an image from the reflection of a dome-shaped mirror, the regular blocks in a raw image are distorted and not uniform in real-space. Therefore, the image distortion compromises the accuracy of the detected motion vectors, especially around the boundary of an image. To resolve this issue, Huang et al. (2011, 2013) proposed pre-processing TSI images and transforming the original distorted view to a planar view. Thereafter, they located the best-matching blocks based on the normalized cross-correlation (NCC) value, and utilized a refining threshold to remove the low accuracy matches, i.e., the low NCC value. To mitigate the deformation in the cloud and predict the variation in block-wise motion, Huang et al. (2013) investigated a multi-frame motion vector tracking and back-tracking,

and extrapolated the future movement with the detected motion trend. However, these approaches are not computationally efficient and cannot take advantage of the cloud's information. To address these issues, we proposed a new method, termed "cloud-block matching" (CBM), to dynamically determine cloud blocks in sky images (Peng et al. 2015), estimate the motion vectors only in the blocks of cloud pixels, and thereby assure its efficiency and efficacy via an intelligent clustering.

2.1.2.3 Miscellaneous Other Approaches

In contrast to those two types of tracking, other tracking approaches may use image registration techniques to attain disparity vectors, namely object motion, between two consecutive frames. For example, phase correlation is a fast noise-resilient approach that estimates the translational offset between two similar frames or sub-frames in the frequency domain (Stone et al. 2001). The phase-correlation method, as shown in Huang et al. (2011), cannot discriminate multiple movements within the same sub-frame region. More importantly, compared with the block-matching technique, the phase-correlation technique incurs a higher error rate and is less accurate when no obvious cloud texture is available in the designated regions.

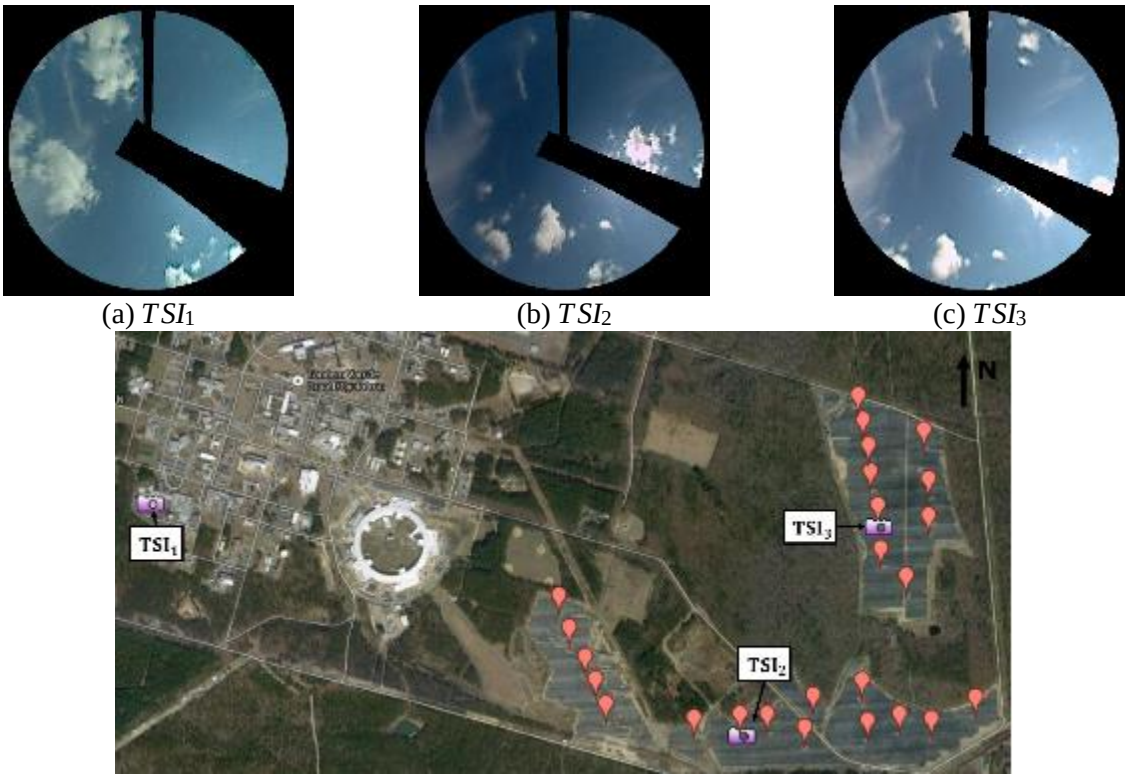
To better address the dynamics of cloud motion, and track the deformation over time, Bernecker et al. (2012, 2014) proposed using a non-rigid image registration. They used the well-studied diffusion model that is developed by Thirion et al. (1998), based on the optical flow method to detect a combined motion vector field with both global translational motions and local variations obtained from the diffusion model. Many new deformation models, such as flow and curvature models (Sotiras et al. 2013), follow the Thirion method and are widely used in medical image registration.

In addition, the particle image velocimetry (PIV) methodology is used to estimate cloud motions as a velocity field. Chu et al. (2013) and Marquez and Coimbra (2013) adopted the MPIV software (Mori and Chang 2003) to detect the block-wise cloud velocity field in TSI images. Here, MPIV partitions an image into reference blocks and searches for the best matching one based on the correspondence criterion of the minimum quadratic difference or the cross-correlation coefficient within a nearby window. Afterward, MPIV applies the post-process steps of filtering and interpolation to smooth the output velocity vectors. However, MPIV does not perform well when an image has discontinuities, such as the artifacts of the shadow-band and the camera's supporting arm on the TSI. To improve the accuracy and aggregate the sparse motion vectors of MPIV, Chu et al. (2013) proposed using a k-means clustering to extract two majority motion clusters to differentiate stationary clouds and/or clear sky from fast-moving clouds.

Huang et al. (2012) introduced a hybrid method that incorporates the stable local descriptor or local feature in estimating cloud motions. On the top of the block-wise motion, Huang et al. used the partial intensity invariant feature descriptor (PIIFD) (Chen et al. 2010) to adjust the motion vectors so to enhance the robustness to geometric and photometric variations. Similar to LK81 motion tracking, Huang's method can determine sparse motion vectors at those points with a large gradient of intensity and correct erroneous block-wise motions.

2.1.3 State-of-the-art Approaches and How the System Improved upon Them

2.1.3.1 Instruments



(a) TSI_1 (b) TSI_2 (c) TSI_3

(d) Google Maps View

Figure 2-2. (a), (b), and (c) are the pre-processed views of the three TSIs. (d) is the Google Maps view of the three TSIs from left to right (camera icon labels) and the 25 solar radiation sensors (red). The distance between TSI1 and TSI2 is 2477 meters, and that between TSI2 and TSI3 is 956 meters.

Most of the material in this subsection also appears in Peng et al. (2015).

The data used here were obtained from the Long Island Solar Farm (LISF), a 32-MW solar PV power plant built by a collaboration between BP Solar, the Long Island Power Authority (LIPA), and the Department of Energy.

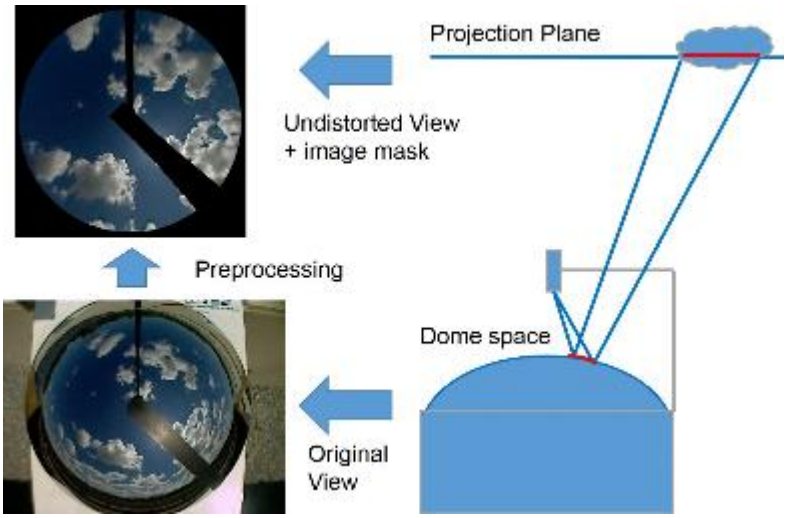


Figure 2-3. Procedure for pre-processing the TSI images. The original image is undistorted from the original dome space to the planar space via coordinate transformation. The output image is cropped based on a pre-defined field of view range and masked to remove irrelevant areas, such as the supporting arm and the shadow band.

The LISF, located at the Brookhaven National Laboratory, is currently the largest solar PV power plant in the eastern U.S. The cloud tracking system (Figure 2-2) consists of a network of three total sky imagers (TSI_1, TSI_2, and TSI_3). Their positions are triangulated to ensure good coverage of the sky above the solar farm. TSI_1 collects sky images of a region near the solar farm. TSI_2 is deployed in the middle of the farm to ensure sufficient overlap of its views with those of the other two TSIs. TSI_3 is located in the northern area of the solar farm. We installed all three TSIs at the same altitude and oriented them to ensure that their camera-supporting arms point north. Consequently, the supporting arm in each TSI image overlaps with the vertical (y) axis, and is aligned toward the north.

All TSIs uploaded real-time photo streams with a raw resolution of 640 x 480 pixels, and at a streaming rate of 10 seconds per frame to a centralized database server. We then applied pre-processing techniques, as illustrated in Figure 2-3, to un-distort the raw images and remove unrelated artifacts, i.e., the shadow band and supporting arm. The primary advantage of pre-processing the images, as described in Huang et al. (2011), is that the planar view obtained from the coordinate transformation of the raw images is more effective for estimating cloud motions. In this study, we projected the original coordinates to a flat plane with a given resolution (500 x 500 pixels) and cropped the planar image using a pre-defined view angle range (zenith range $0^\circ - 60^\circ$) to ensure a field of view (FOV) of approximately 120°. In addition, because the camera-supporting arm may not be exactly aligned with the vertical direction in a TSI image, we adjusted the orientation in the preprocessing step by rotating the images as necessary. To simplify the coordinate transformations between different TSIs, the conditions and specifications of all TSI devices were unified. Because each TSI has an identical FOV and resolution, all the pre-processed images reside on the same projection plane, wherein pixels are distributed evenly. Finally, we applied a pre-calculated mask to remove artifacts such as the supporting arm and shadow band. The pre-processed image was then generated, with all irrelevant pixels being set to black (Figure 2-2a-c).

BNL deployed 25 pyranometers in the LISF (Figure 2-2d) to measure the surface solar irradiance. These sensors measure the GHI in real-time. The measurements, which are recorded every 10 seconds, are synchronized with the TSI observations. The variations in zenith angle and the diurnal and seasonal patterns are also recorded in the raw GHI measurements, and therefore bias our subsequent irradiance forecast models. To mitigate this bias, we normalized each radiation value to a clear-sky index k^t during model training and testing. Letting GHI^t be the raw GHI measured at time *t* and GHI^t_{clear} be the corresponding clear-sky estimate, the clear-sky index k^t is calculated as follows:

$$k^t = \frac{GHI^t}{GHI^t_{clear}} \quad (2.1)$$

where k^t nominally ranges from 0 to 1. However, its maximum value can be greater than one due to the cloud enhancement caused by diffuse sunlight. The estimated clear-sky value GHI^t_{clear} at time *t* is obtained from the regression curve that best fits the distribution of historical observations (Peng et al. 2013).

2.1.3.2 State-of-the-art Cloud Tracking Algorithms and Their Limitations

Among the techniques of tracking the motions of clouds, block matching and optical flow (OF) are applied widely to various types of imagery, including ground-based cameras and satellites.

Block-matching techniques take a collection of pixels (i.e., a block) as a tracking unit, have the ability to utilize information from multiple areas, and therefore are sufficiently robust to both image noise and brightness variations within images. If the underlying motions consist of only translational velocity, and do not involve shearing and stretching, a block-matching approach can faithfully represent the true movements of clouds (Huang et al. 2011). However, the majority of block-matching approaches employ pre-defined blocks with a fixed size and position, i.e., a mesh or grid in an image (Hamill and Nehrkorn 1993; Chow et al. 2011). Consequently, this type of block-matching approach is sensitive to the block's segmentation, and an incorrect segmentation in an image can compromise its accuracy. Because non-rigid clouds have a variety of shapes and positions that invalidate any pre-defined segmentation, the performance of these block-based methods is inconsistent over the streams of images from the camera. Even with the assignment of dynamic blocks based on the cloud mask in a sky image (Peng et al. 2014), the performance of the block-matching algorithm still deteriorates when multiple pieces of cloud reside within the same block. Many recent methods adopt a hierarchical block structure to track block-wise motions, and apply post-process filtering and interpolation to the tracking outputs (Huang et al. 2011; Mori and Chang 2003). Nevertheless, with these approaches, only a limited improvement is attained, and it still relies on predefined block segmentation and the tedious tuning of block size and position.

On the other hand, OF techniques address the motion tracking problem at a lower level than does block-matching. It enables extraction of the pixel-wise motion through variational methods that first define an energy-like objective function, assume the constancy of brightness cross images, and obtain the solution via minimizing the objective function. Compared to the block-matching approach, the OF model is flexible and can accurately represent complex three-dimensional

motions, such as rotation and scaling, at the pixel level (Héas et al. 2007; Héas and Mémin 2008). However, it is sensitive to image noise and the variation in brightness, both of which are quite common in sky images. Another issue is that for texture-less regions or fast-moving, the accuracy of OF may drop significantly due to the lack of information (Brox and Malik 2011). Even with a smoothness term (Golemati et al. 2012) or a regional mask (Wood-Bradley et al. 2012) that incorporates the surrounding information, current OF approaches still neglect the important features, such as the cloud distribution and the multiple cloud layers, each of which has its own motion.

2.1.3.3 Improvements upon the State-of-the-art

To address the aforementioned issues of optical flow and block-matching, we proposed a new hybrid approach that integrates the block-matching (BM) method and the variational OF model, and uses the former method to guide/refine the latter one. This new model encompasses three main steps: (1) extracting a cloud mask, and generating cloud blocks via bottom-up merging and detecting block-wise motions; (2) identifying dominant motion patterns from detected motion vectors; and (3) estimating optical flow using our new formulation and refining based on multiple motion filters. This design recognizes that the vectors detected by the BM and the OF models are actually inter-dependent. For optimal results, they should be integrated into the same framework to ensure mutual enhancement.

All previously existing tracking algorithms use a single camera and are severely limited by the resulting two-dimensional view. A single camera cannot give height information nor recover the details of multiple layers of cloud that are critical to predicting irradiance at a solar farm. Given the low resolution of many sky cameras and severe barrel distortion at image borders, the effective field of view is limited (often less than 120º, which is too small to track fast-moving clouds. In this project, we created a new tracking algorithm to track clouds in a three-dimensional space. Figure 2-4 shows an overview of three-TSI tracking. A piece of cloud at a certain level can be projected onto/visualized as three different cloud blocks on the projection planes of the three TSIs. Because of the geometric difference among the three TSIs on the ground, the pixel coordinates of each cloud block are distinct. Therefore, these cloud blocks can be combined to recover their 3-D coordinates. Theoretically, a pair of TSIs should suffice to recover the cloud base height (CBH) using a stereography approach (Kassianov et al. 2005; Allmen and Kegelmeyer Jr. 1996; Nguyen and Kleissl 2014). In this system, we utilize the redundant information from the three-TSI network to increase the robustness of cloud tracking.

Furthermore, we increased the range of the state-of-the-art ground-based prediction system from the level of one minute to 10-15 minutes, depending on the speed at which the clouds are moving. This work includes three components: 1) a robust algorithm to detect the identical piece of cloud across different views of the TSI cameras, and recover its height information, and to track clouds across different time frames of the same camera and calculate its motion; 2) an intelligent cluster algorithm to identify different cloud layers and regimes that are based on information streams from multiple TSIs; and 3) a stitching algorithm to expand the field of view by concatenating multiple images from different cameras, and to handle the challenges of stitching images based on ambiguously defined features, such as cloud heights, borders, shapes, and textures.

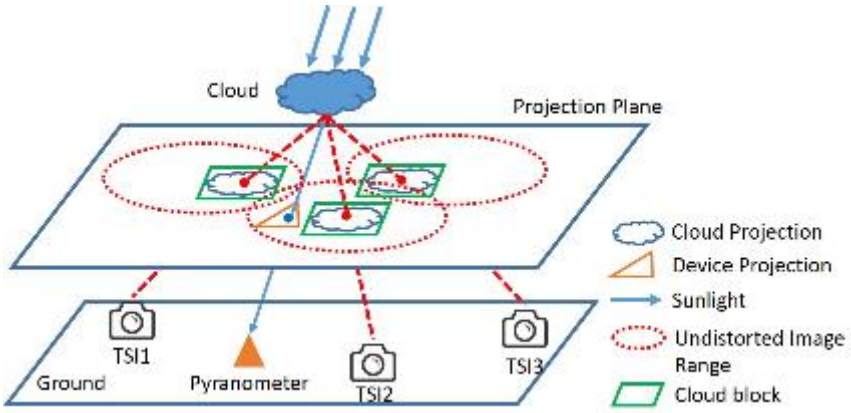


Figure 2-4. Overview of three-TSI tracking.

2.1.4 Advances to Very Short Range Forecasting

BNL designed a robust framework to identify the primary cloud layers, to extract the effective image features that are relevant to forecast the surface irradiance, and to apply a support vector machine (SVM) to predict the final solar irradiance value. Our prediction pipeline contains four technical advances: machine learning to detect clouds, a new hybrid model to track clouds, multi-TSI integration for tracking in a 3-dimensional space and expanding the field of view of cameras for a longer prediction window, and finally, the solar irradiance forecast. Much of this work is reported in more detail in Peng et al. (2015).

2.1.4.1 Machine Learning to Detect Clouds

Cloud Pixel Identification

The information in the following two subsections largely comes from section 3 of Peng et al. (2015).

To extract cloud information from images, cloud pixels need to be identified and differentiated from non-cloud pixels. The basic concept of cloud detection is to apply a classification methodology to identify cloud pixels in sky imagery and to separate the clouds from the sky at the pixel level. Beginning with the mask of classified cloud pixels, we aggregate the pixels into appropriate regions/blocks to represent pieces of cloud for subsequent tracking.

However, many challenges are encountered when identifying clouds at the pixel level. With their distinct properties, such as different optical depths and brightness variations in sky imagery, clouds have various distributions in the red-green-blue (RGB) channels and a wide range of brightness levels in different sky scenes or image frames. In particular, certain optically thin clouds may have similar textures and colors to those of the background sky pixels, whereas some clouds may appear both dark and white within the same sky image. Therefore, it is difficult to identify a fixed threshold to separate clouds from sky. For instance, cirrus clouds are often barely distinguishable from the clear sky background in TSI images.

Moreover, under different weather conditions and at different solar angles, the clouds presented on a bitmap image may appear to have different brightness levels and a large range of intensities. Therefore, we require a robust, sophisticated methodology to capture the pixel-wise differences or the regional textural differences between cloudy and clear regions. From the perspective of image processing, it may appear that we could first sharpen edges and boundaries using various derivative image filters and then apply image segmentation methods to separate clouds from clear pixels based on their boundary pixels, thereby identifying cloud segments. However, these methods often fail to segment clouds from the background due to the poorly defined edges of clouds in sky images. With the resolution and image quality limitations, a single pixel in a sky image may contain both cloud and clear sky. Consequently, a section of visually rigid cloud can appear non-rigid and blurry in sky imagery.

Furthermore, sky imagers face certain instrument-specific challenges in their practical deployment in the field and in testing. Because the sky camera adjusts its lens aperture and shutter speed in response to the amount of incident illumination, the output images can potentially suffer from variations in exposure, and may appear either brighter or darker than the ground-truth image that accurately represents the real lighting conditions.

In this work, we detail how to use a binary classifier to identify cloud versus clear-sky pixels in sky imagery. The training datasets for the classifier are generated by manually labeling cloud/sky pixels in TSI images. This process requires a considerable amount of human effort and, more importantly, may introduce uncertainties and errors into the training datasets that could significantly impact the accuracy of traditional classifiers. To overcome this challenge, we propose an outlier-aware classifier to train manually labeled pixels in sky images. Moreover, because of the many instrument-specific abnormal cases that can be generated from individual TSIs, such as over-exposed or under-exposed images, we design a classifier-based pipeline to utilize all three TSIs for multi-source image correction to enhance the overall accuracy of cloud detection. We also present a technique for aggregating cloud pixels into cloud blocks (Figure 2-5).

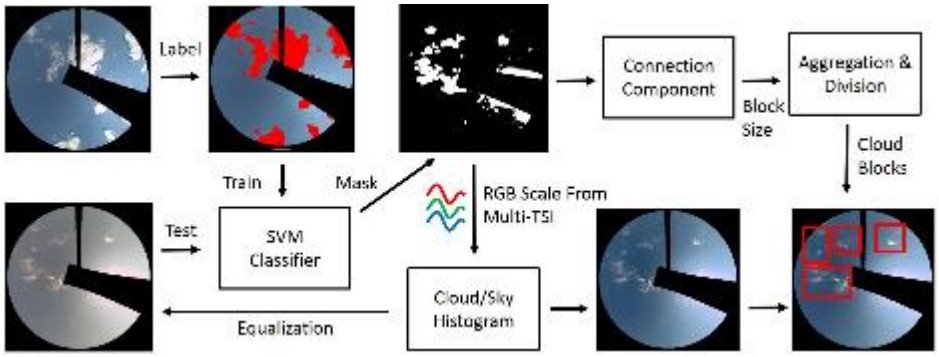


Figure 2-5. Pipeline for cloud detection using an SVM classifier and multi-source correction.

Supervised Cloud Classifier

To minimize the influence of possible outliers in the training dataset, BNL chose a support vector machine (SVM) (Lindeberg 1993) as a pixel-wise classifier and trained it using two-layer cross-validation (Kohavi 1995) to reduce over-fitting. An SVM constructs a max-margin hyper-plane to

reduce the effect caused by outliers and offers the advantage of being able to handle known outlier patterns. In particular, uncertainties and errors introduced during the manual annotation of training images are taken into account during SVM optimization. Figure 2-5 shows an example of the training and testing process for a cloud classifier. To better describe the characteristics of cloud pixels, six features are extracted from the sky images and normalized for use as the SVM training dataset: red, green, blue (RGB), and red-blue ratio (RBR) (Huang et al. 2013), all four of which are spectral characteristics extracted from the RGB color space, and two features based on the neighboring pixels, which are used to mitigate the impact of variations in illumination. In detail, we convert the sky images from the original RGB color space to the relative luminance space based on the algorithm presented in Stokes et al. (1996), and then apply the Laplacian of Gaussian (LoG) spatial filter (Boser et al. 1992) to compute the second derivative in luminance space within a fixed-size window (7x7 is used). The LoG value reflects the rapid changes in the illumination channel in a certain region and is useful for detecting sharp edges. The final feature used for cloud detection is the standard deviation of the luminance channel within a small region around each pixel (7x7 is used), representing the pixel's average difference from its neighbors. Our implementation uses the SVM package *libsvm* (Chang and Lin 2011) with a linear kernel.

Furthermore, BNL explored the possibility of multi-source abnormality correction using synchronized images from all three TSIs. Because the three TSIs are located reasonably close to each other, we can reasonably assume that their color representations have statistically similar range with regard to the RGB channels. In other words, cloudy/clear pixels in these three digitized color channels should have similar histograms across the different TSIs. Therefore, if one TSI experiences an exposure issue or abnormal brightness in the RGB color space, we can correct it by equalizing the histograms of its RGB channels to those of the two normal TSIs. We first generate the cloud masks for all TSIs using the SVM classifier and compute the histograms of the RGB channels for cloud and sky pixels separately. By calculating the Euclidean distances between the histogram vectors of the three TSIs, we can identify a device as abnormal if its image histogram is significantly different from those of the other two devices. We then apply the histogram equalization add-on to the output image from the abnormal TSI to adjust its RGB scale for cloud and sky pixels. The corrected result can be used for the next round of cloud mask generation. In practice, we iterate this procedure three times to extract the cloud mask and equalize the RGB histograms of the TSI images (Figure 2-5).

To evaluate the performance of the proposed supervised classifier and the multi-source correction algorithm, we selected various test cases from daily observations corresponding to different atmospheric conditions and cloud types and then compared the results with our manually annotated images. In this report, we adopt two evaluation metrics to measure the error in cloud classification:

$$AC_{cld} = \frac{N_{cld,cld}}{N_{cld,cld}+N_{sky,cld}}, AC_{sky} = \frac{N_{sky,sky}}{N_{sky,sky}+N_{cld,sky}} \quad (2.2)$$

where AC_{cld} and AC_{sky} are the accuracies of cloud and sky pixel classification, respectively. $N_{cld,cld}$ and $N_{sky,sky}$ denote the pixel counts of correct cloud and sky classifications, respectively, whereas $N_{sky,cld}$ and $N_{cld,sky}$ indicate the total numbers of sky and cloud pixels, respectively, that are falsely recognized by our detector. In Figure 2-6, we show the distinctive patterns of examples of scattered clouds, cloudy conditions, overcast conditions, and two multi-layer cases. Compared with the

manual classification masks, our pipeline based on the SVM classifier can accurately detect clouds (with an accuracy of more than 83.2%), except in the case of multi-layered clouds near the sun's position and very thin clouds. The image area near the sun ("sunspot") has a higher brightness and is difficult to characterize based exclusively on static textural information. Therefore, the classifier often falsely labels clear-sky pixels as clouds in that region.

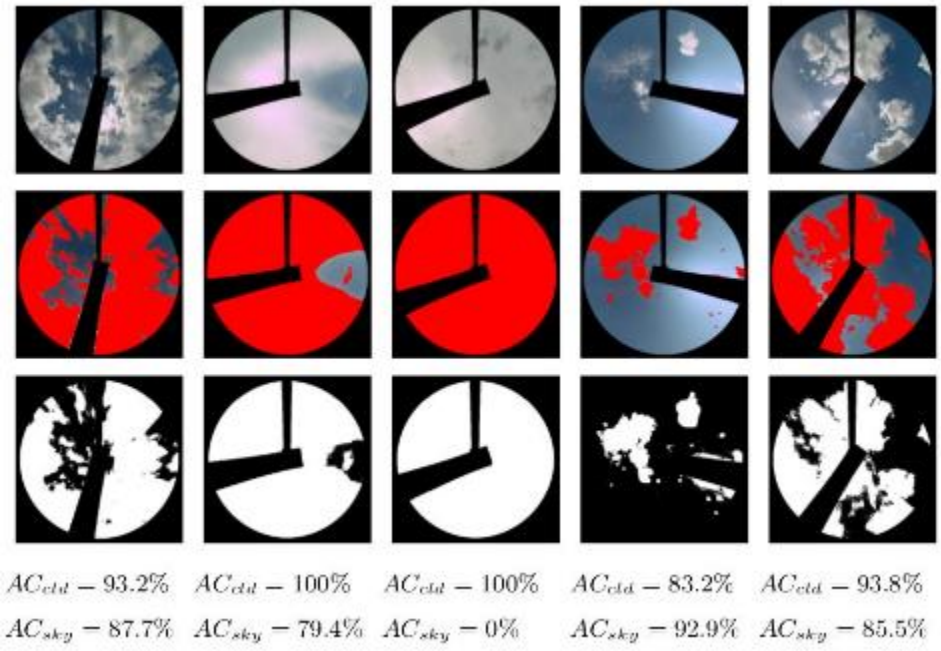


Figure 2-6. Cloud detection results (row 3) compared to manual annotation (row 2) under different weather/cloud conditions. Left to Right in row 1: Scattered cloud, mostly cloudy, overcast, multi-layer, and multi-layer with thin cloud. AC_{cld} and AC_{sky} represent the accuracy of the classification results for cloud pixels and sky pixels, respectively.

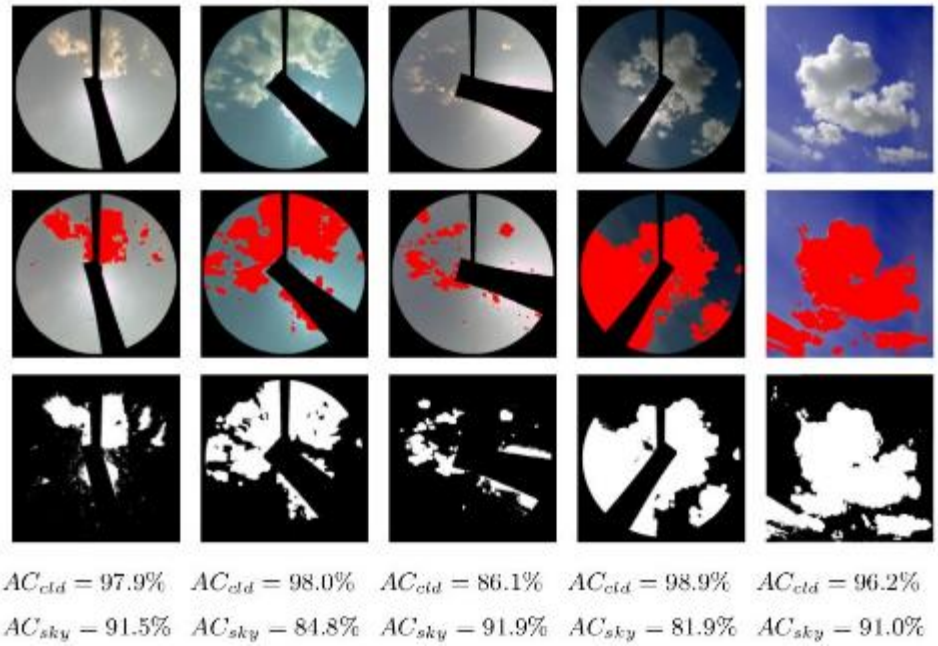


Figure 2-7. Results of cloud detection (row 3) compared to manual annotation (row 2) in the presence of device-specific bias or luminance variations. Left to Right in row 1: red-dominant, green-dominant, over-exposed, under-exposed, and a different image source. AC_{cld} and AC_{sky} represent the accuracy of the classification results for cloud pixels and sky pixels, respectively.

To validate the cloud detection performance of our method in the case of device errors or variations in exposure, we selected four abnormal images of types that are commonly observed in the field: red-color dominant, green-color dominant, over-exposed, and under-exposed (the first four cases in Figure 2-7). The accuracy of cloud pixel classification in these cases is 86.1% or higher. We also applied our classifier to another type of sky imagery configured with a different field of view and color scales (the last case in Figure 2-7), which confirmed that our classification algorithm is practical and effective. The classification accuracies for cloud and sky pixels in this case are 96.6% and 89.7%, respectively. The overall cloud detection performance in all selected cases is evaluated in the confusion matrix presented in Table 2-1. We observed our pipeline accurately (96.6%) recognizes cloud pixels.

Table 2-1. Overall confusion matrix for the cloud detection pipeline applied to the images shown in Figure 2-6 and Figure 2-7. Left to right in the Table: AC_{cld} and $(1 - AC_{cld})$ in the first row, and $(1 - AC_{sky})$ and AC_{sky} in the second row. All values are in [%].

	Cloud Detection Pipeline	
Manual	Cloud	Sky
Cloud	96.6%	3.4%
Sky	10.3%	89.7%

Cloud Block Aggregation

The cloud pixel distribution detected by supervised classifier is often utilized to aggregate the pixels into regions/blocks for motion vector discovery (Peng et al. 2014). The core idea is to divide the cloud mask into individual regions. If the texture inside is not appropriate for the

similarity criterion defined in motion tracking, then the region is often divided or merged with its neighbors to include more information (Peng et al. 2015). This process significantly improves the effectiveness of the best-match finding in block-wise motion detection, and also reduces the complexity of computation by removing non-cloud pixels during calculation. Once we successfully partition the cloud pixels into cloud blocks, we can track individual cloud blocks and predict their future locations.

2.1.4.2 New Hybrid Model to Track Clouds

Cloud Block Generation and Matching

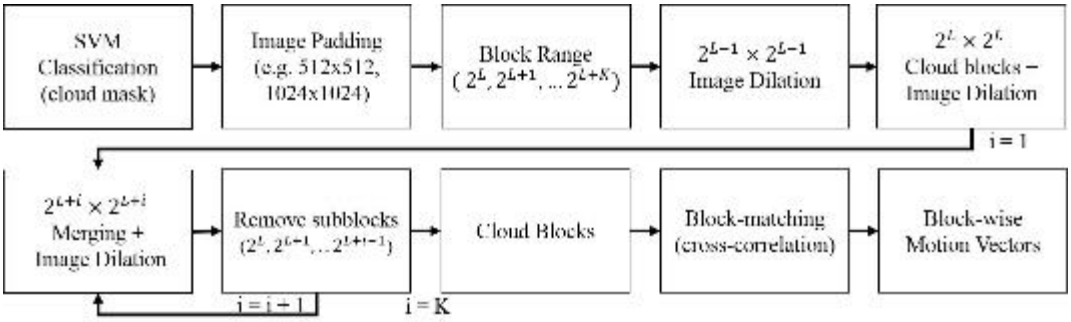
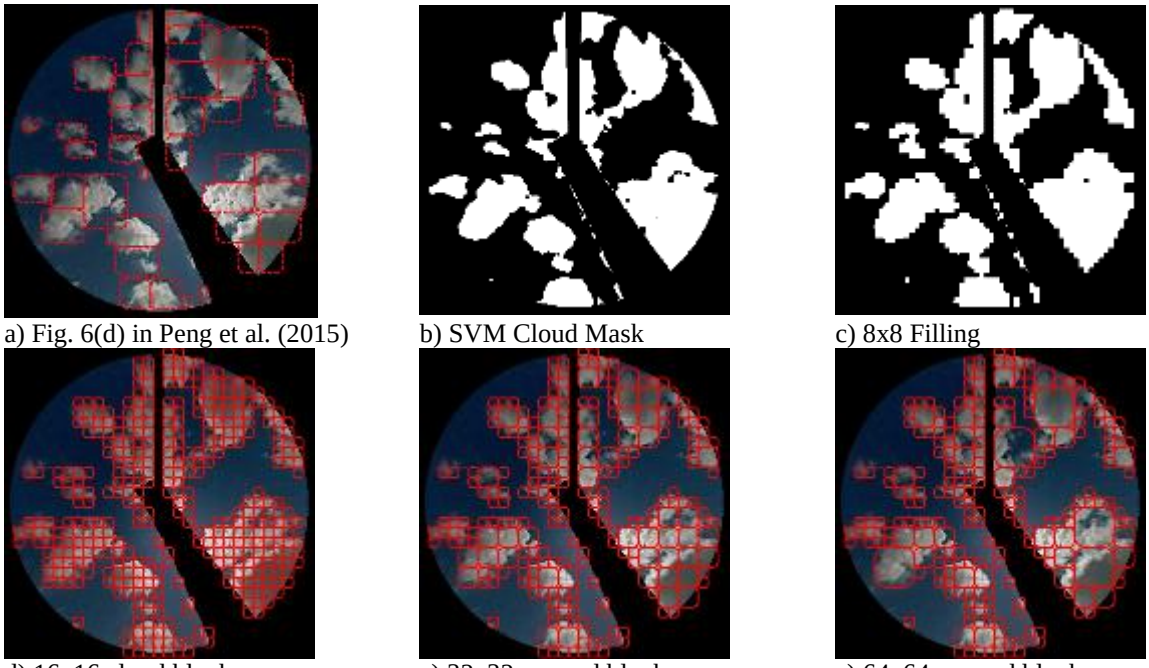


Figure 2-8. Pipeline of QCBM.

Because the block's size and position play a vital role in tracking clouds, we devised an effective algorithm for block generation and matching, quadtree-based cloud block-matching (QCBM), to take into consideration the cloud distribution in sky images. Here, a quadtree recursively decomposes images into four equal-sized square sub-images until the criterion of homogeneity is met, or the minimum block size is reached, and consequently represents an image in a top-down hierarchy with different resolutions (Shusterman and Feder 1994). Such a representation can efficiently divide an image into 2-D homogeneous regions/blocks (of similar color, texture, or structure), and ease the subsequent tracking process.

However, regions of low homogeneity, such as the edges of clouds, require excessive decomposition to meet the stopping criterion. As a result, a top-down decomposition usually generates cloud blocks that are too small to preserve meaningful textural information for motion tracking. To overcome this, we introduced a bottom-up scheme to the quadtree construction that starts from all nodes at a pre-defined level in the quadtree, identifies all cloud blocks containing cloud pixels more than the threshold CF (the fraction of cloud pixels in a block based on the cloud mask obtained in prior steps), selectively fills all pixels as clouds in these nodes, and iteratively merges homogeneous blocks.



a) Fig. 6(d) in Peng et al. (2015) b) SVM Cloud Mask c) 8x8 Filling

d) 16x16 cloud blocks e) 32x32 merged blocks e) 64x64 merged blocks

Figure 2-9. Quadtree block generation using the example shown in Figure 6 in Peng et al. (2015) with L = 4, K = 2, CF = 0.2 and T = 3.

Figure 2-8 shows the steps of the QCBM generation and matching. For simplicity in implementation, we pad the sky image to be square with the size of the power of 2. To ensure the quality of block-matching, we limit the dimension of the generated cloud blocks to be in a range of $(2^L \times 2^L, 2^{L+1} \times 2^{L+1}, \ldots, 2^{L+K} \times 2^{L+K})$, where 2^L is the minimum size allowed in generating cloud blocks. Under this new approach, there are only K+1 different sizes of cloud blocks. To ensure the quality of segmentation near the cloud's boundary, we first start at one layer lower than the finest level, (2^{L-1}) to fill the cloud mask if the cloud fraction inside is beyond CF. During the iterative merging step from the level of 2^L in our bottom-up scheme, we check the four children/sub-blocks of each node: if no less than $T=4$ children are cloud blocks, all four child nodes are eliminated, and the parent node becomes a new (bigger) cloud block. This new cloud block will then be filled with cloud pixels and undergoes subsequent merging until the level 2^{L+K} is reached. Lastly, the quadtree contains different sizes of cloud blocks for matching, as shown in Figure 2-9f. We then find the best match for each cloud block to find its block-wise movement to the next frame.

To obtain the optimal matching quality based on the correlation criterion, we choose three layers starting from the minimum block size 16×16 for sky imagery (L = 4, K = 2), and then we set CF = 0.2, T = 3 to fill and merge the lower level blocks. Figure 2-9 presents the result of the original block generation by the QCBM, based on the example in Peng et al. (2015). Figure 2-9c shows the cloud blocks with the starting size of 8×8 filled with the original cloud mask in Figure 2-9b, whereas Figure 2-9d – Figure 2-9f present the iterative merging results of cloud blocks with a size from 16×16 to 64×64. We observe that compared with the blocks generated by CBM (Figure 2-9a), the blocks by the bottom-up scheme in QCBM (Figure 2-9f) have cleaner segmentation around the boundary of clouds and do not contain overlapping regions.

More importantly, this QCBM is computationally more efficient. We implemented QCBM, BM, and CBM in MATLAB® and ran them on a laptop. Figure 2-10 shows the average run time of three different block-matching algorithms under different image resolutions and cloud fractions. As the QCBM does not involve the "divide and conquer" scheme in CBM, and eliminates most clear-sky pixels during the block generation in contrast to BM, its execution speed is significantly faster than the others. Especially when an image is large with high resolution, QCBM only incurs a marginal increase in the execution time, and always completes the matching process within 10 seconds. Though the QCBM segments an image into more cloud regions when the fraction of cloud in the image increases, the computational speed for its bottom-up scheme does not increase significantly and still is faster compared with CBM.

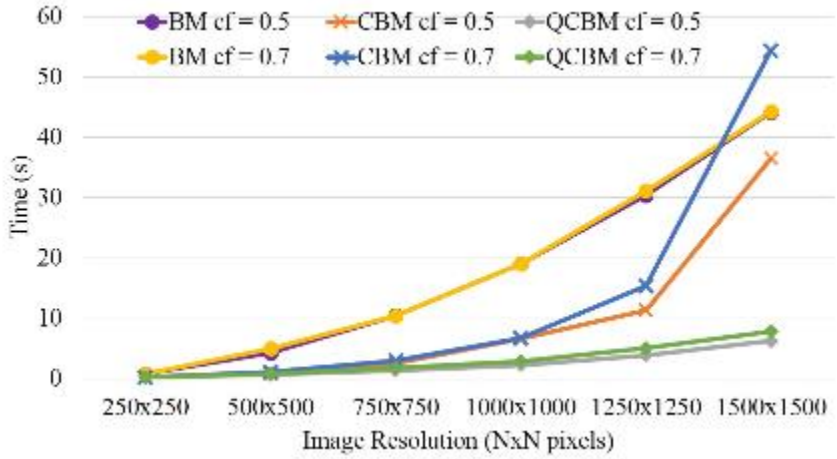


Figure 2-10. Execution time of block-matching algorithms for different image resolutions. CF indicates the cloud fraction of a sky image.

Dominant Cloud Motion Patterns

To suppress image noise and neglect falsely estimated motion vectors, we identify the dominant patterns of cloud motions and use them to refine the entire motion field. First, appropriate image pre-processing is necessary to un-distort the TSI images to ensure that the majority of cloud motions obtained from the undistorted images are simple and translational. Subsequently, we can apply a straightforward clustering (e.g. k-means) to effectively group them, to find dominant patterns, and to remove abnormal ones (Chen et al. 2013; Peng et al. 2015).

However, the clustering approach does not take into consideration the weight of each motion vector, i.e., the size of the actual cloud block, and therefore might miss some dominant vectors. In this report, we utilize the histogram statistics presented in He and Sun (2012) to extract the N most frequent cloud motions in a sky image, and use them to correct and refine the results of estimated motions. To accurately determine the significant motion modes for cloud pixels and exclude the small-scale motions caused by image noise or sky pixels, we only consider those obvious motion vectors that have adequate velocity between two consecutive frames ($||u||^2 + ||v||^2 > 1$). Thereby, stationary sky pixels or slow-moving clouds are ignored in calculating dominant motions.

In our implementation, we first produce a dense motion field based on the motion vectors detected in previous block matching. Each cloud pixel acquires the associated motion vector from the cloud block on which it resides. Thereafter, we create a 10x10 2-D histogram for the motion vectors of all cloud pixels. This design considers the size of a cloud block, and a large block has many pixels, each of which contributes to the count, namely the weight of the shared block-wise motion vectors to the histogram. Then we pick the N most-frequent motion modes, and group the pixel-wise motion vectors into N groups. For each motion group, we select the median motion vector as the dominant motion pattern, and thereby generate a collection of dominant motion vectors. Finally, we refine all non-zero motion vectors by resetting their values to the closest dominant motion vectors. Here we use $(\tilde{u}, \tilde{v})$ as the reference vector to guide the refinement of optical flow in the next section.

Context-Aware Variational Model and Its Refinement

The original OF models lack contextual information and are agnostic to existing domain knowledge. Consequently, these OF models often assign erroneous flow vectors to cloud pixels and are sensitive to image noise. To cope with those issues and utilize the information of motion layers and cloud distribution, we build a new OF model. First, we revise the energy-like objective to assimilate the aforementioned dominant motion patterns, and to use them to create a context for calculating the optical flow at each cloud pixel. However, a motion field obtained only from the straightforward process of energy minimization still suffers those issues. Previously, the general practice was to apply post-processing techniques, such as median filtering (Wedel et al. 2009) or signal-noise ratio threshold (Mori and Chang 2003) to mitigate the impact of image noise and remove the outliers. Therefore, we also follow the practice and apply three filters thereafter to further assimilate cloud information, utilize dominant motion patterns, and remove noise in the motion field. Given the facts that one iteration is not sufficient to correct the motion field and that the reference vector still needs to be refined to absorb new information to closely represent the current motion field, we design an iterative algorithm to generate and optimize the dense flow field and to update the subsequent reference motion vectors. This design recognizes that the vectors detected by the BM and the OF models are actually inter-dependent. For the optimal results, they should be integrated into the same framework to ensure mutual enhancement.

A sky filter processes the clear sky pixels that are identified by velocity threshold. We assume that the clear sky regions in images do not move between consecutive frames, and set the motions of all sky pixels acquiring small-scale motions in the previous step to be zero in both the x and y directions. It eliminates the small motion vectors caused by image noise and avoids accumulating errors in the iterative minimization step. It is noteworthy that instead of considering all clear sky pixels, we filter out only the ones with a small movement based on the threshold criterion T_{sky} because cloud pixels in thin clouds or the images containing a high level of image noise have a high probability of being falsely categorized into the class of clear sky. Consequently, relying on cloud mask only potentially neglects these special cases, and accidentally removes the prominent motions of these mis-classified pixels.

A dominant motion filter is designed to identify the outliers with reference to the dominant motion patterns. If a motion vector at position (x, y) has significant deviations from all dominant motion patterns (Ω), then we identify it as an outlier, i.e., if a motion vector meets the following condition, we will remove it.

$$\{(u_{x,y}, v_{x,y}) | (u_{x,y} - u_i)^2 + (v_{x,y} - v_i)^2 > \varepsilon^2, \forall (u_i, v_i) \in \Omega\} \quad (2.3)$$

This filter significantly helps to refine the motion field for the next round of minimization and to update the reference motion field.

A weighted median filter is widely adopted to smooth and de-noise the motion field in the post-processing step (Sun et al. 2014). Our model employs this filter to update the reference (auxiliary) motion field $(\hat{u}, \hat{v})$. The weighted median filter can be calculated as follows:

$$\min_{\hat{u}_{x,y}, \hat{v}_{x,y}} \int_{(x',y') \in \Gamma(x,y)} w_{x,y}^{x',y'} \left(\left| \hat{u}_{x,y} - u_{x',y'} \right| + \left| \hat{v}_{x,y} - v_{x',y'} \right| \right) \quad (2.4)$$

where $\Gamma(x,y)$ is the set of (x,y)'s neighborhood pixels (x,y) within a pre-defined window (e.g., 5x5) centered at (x,y), and $w_{x,y}^{x'y'}$ is the weight of the affinity (similarity) between two pixels. In our implementation, we approximate its calculation with the color difference in an image:

$$w_{x,y}^{x',y'} = e^{-\frac{\|F(x,y) - F(x',y')\|^2}{2\sigma^2}} \quad (2.5)$$

F represents the color image with the RGB channels. To compute w efficiently between each pair of pixels, we adopt the joint-histogram weighted median filter (JointWMF; Zhang et al. 2014) to reduce the execution time of updating $(\hat{u}, \hat{v})$. Compared with the weighted median calculation in Sun et al. (2014), the JointWMF significantly speeds up the running time by more than a factor of 10 while still preserving the quality of estimation (Zhang et al. 2014). Afterward, we pass the updated $(\hat{u}, \hat{v})$ for the next round of the iterative minimization of $E(u,v)$. We iterate the process three times to produce the final output of the dense optical flow field.

2.1.4.3 Multi-TSI Integration for Tracking Clouds in Three-Dimensional space and Expanding Field of View

Much of the material in this subsection comes from section 4 of Peng et al. (2015). The reader is referred to that publication for equation development.

Assumptions of Consistency of Projected Images Across TSIs

The vertical height and horizontal expansion of clouds vary considerably among different types of clouds. Therefore, the projected size of a cloud observed by ground-based cameras depends strongly on the cloud type, the location of each camera, the field of view, and the solar zenith angle. With no loss of generality, we assume that the dimensions of the projected views of the cloud base in images from multiple TSIs are identical to ensure the suitability and mathematical correctness of our calculations. In practice, our cloud tracking system does not enforce these strong assumptions; however, it is sufficiently robust to tolerate the differences between projected images while still being able to identify the same object in different images based on the values of the temporal and spatial correlations.

We further simplify the cloud tracking task by adopting two assumptions: a) clouds exhibit only planar movement, without any vertical motion, and b) the velocity remains constant within our forecasting time window. Moreover, all undistorted images reside on the same projection plane, with the identical fields of view and spatial resolutions.

Therefore, to ensure the uniformity of the TSI images and the consistency of the cloud movements without any loss of generality, we make several additional assumptions about our three-TSI system. First, a cloud block and its counterparts in other TSI images have the same dimensions, without scaling or shearing. Second, the mapping from the displacement vector to the cloud base height (CBH) is independent of time because the mapping functions f and g are fixed and determined only by the locations of the TSIs. Third, a cloud block maintains a consistent size over a short time. Because a piece of cloud is assumed to exhibit only planar movement, all its cloud blocks residing on the projection plane must be of the same size at different timestamps. Finally, the fourth assumption is that a cloud block registered by one TSI and its counterparts at the other TSIs have the same motion vector.

Based on these assumptions, we designed a novel tracking system utilizing all three TSI devices for a series of consecutive timestamps. In total, nine images are used to extract two types of shift vectors at the pixel level: 1) the displacement vectors between TSI views (CBH) and 2) the motion vectors between consecutive frames from one TSI (planar motion). An intuitive example is illustrated in Figure 2-11. After identifying a cloud block segment in an image from TSI_i at time t, our goal is to find its "best" matches in the other eight images and use these matches to explicitly calculate its motion and displacement vectors. The motion vectors of the nine cloud blocks should be identical because they represent the same cloud with steady movements, so we simplify the tracking problem by setting the cloud block velocity $\mathbf{v}_t = \mathbf{v}_{t+1}$ during block matching. Moreover, we assume that the cloud base height remains consistent within a short tracking window. As a result, the estimated CBH h derived from the "best" matches must satisfy $h_{t+1} = h_t$, which is also essentially equivalent to saying that the pixel-wise displacement vector $\mathbf{d}_{i,j,t+1} = \mathbf{d}_{i,j,t}$. To identify the best estimates of $\mathbf{v}_t$ and $\mathbf{d}_{i,j,t}$, the next step is to define a criterion to match multiple cloud blocks in different images.

The Common Similarity Function for Tracking Clouds among Multiple Cameras and Sequence of Image Frames

For this system we have designed an algorithm to simultaneously incorporate both spatial and temporal correlations to improve the accuracy of the block-matching methodology. We proposed formulating the tracking problem as the maximization of a similarity function in which these two types of correlation are summed. In detail, this function consists of two components: 1) the sum of the similarity among different time frames for each TSI, and 2) the sum of similarity between any pair of TSIs at a given timestamp. Given a cloud block centered at $\mathbf{c}$ on TSI_1 at time t, the similarity function, φ, can be calculated by summing all similarity values, and the motion vector and one displacement vector are denoted by $\mathbf{v}^t$ and $\mathbf{d}^{1,2,t}$ respectively. Note that we include only the displacement vector from TSI_1 to TSI_2 in the similarity calculation. Each pixel value lies in the range $(0, 255)$. The normalized cross-correlation (NCC) between two image blocks of the same size and dimensions, is widely used in motion estimation because of its simplicity and robustness to noise and changes in intensity. Given the definition of the similarity function φ, our goal is to search for a combination of $(\mathbf{d}^{1,2,t}, \mathbf{v}^t)$ that optimizes its value. The most straightforward method is

searching all combinations and finding the ($\mathbf{d_m}^{1,2,t}$, $\mathbf{v_m}^{t}$) that maximizes the φ score (Eq. 10 in Peng et al. 2015).

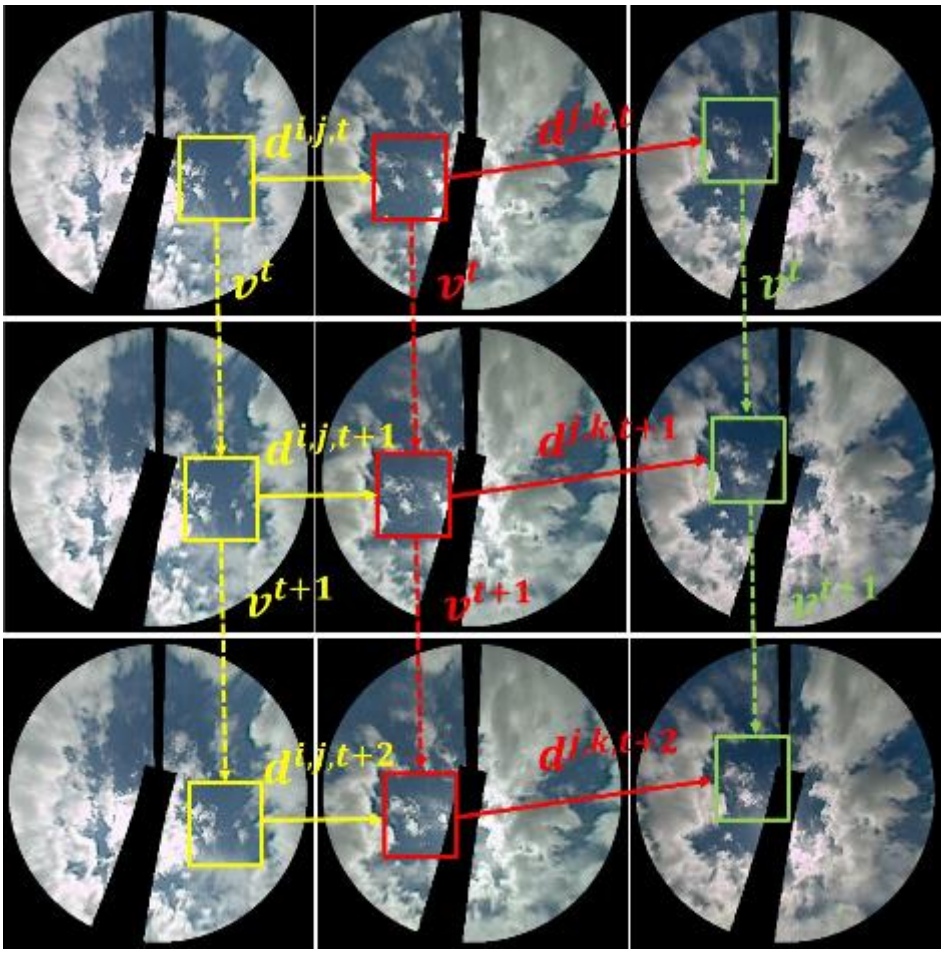


Figure 2-11. Matching cloud blocks in nine images. The cloud block of interest is indicated in yellow, red, and green boxes on the TSI_1, TSI_2, TSI_3 images, respectively. The movement of the cloud block that is detected between two consecutive frames is indicated by a dotted arrow and labeled as v. The displacement vector between a pair of TSIs at the same timestamp is represented by a solid arrow and labeled as d.

However, the maximum of φ is only the correct result if 1) the cloud block is located in the middle of the FOV and visible in all nine images, 2) the image quality is sufficient to accurately display the cloud texture within the block, and 3) the cloud inside the block and its counterparts in the images from the other TSIs are of similar shape and size, even from different angles of view. In general, several difficulties prevent us from attaining the optimal solution to maximize φ as described below. In our previous work (Huang et al. 2013), the block tracking near the image boundary was found to be inaccurate because of the loss of information. Only the blocks that are at least a certain distance from the boundary can possibly be matched using the *NCC* value. Moreover, because cloud blocks have various sizes and textures, the tracking performance achieved by applying *NCC* may vary. In practice, even for cloud blocks that are distributed on the same layer and exhibit identical movement, their optimal solutions to Eq. 10 in Peng et al. (2015) may be similar, but not identical. Another difficulty encountered in cloud-block tracking using TSI images is that some information is missing due to artifacts from

the TSI. Blank pixels (marked as black in the TSI images) significantly impact block-matching because the maximization of *NCC* always favors blocks with lower loss of information. Furthermore, noisy pixels or variations in image brightness may cause instabilities in calculating the *NCC* score and thereby influence the cloud-tracking performance. Consequently, the maximum value of φ does not always guarantee the best matches among nine images.

As discussed in previously, static features are not effective for differentiating cloud pixels near the sunspot. We propose incorporating the dynamic information from cloud tracking to enhance cloud detection in the sunspot region. To mitigate the influence of false-positive pixels around the sunspot, we apply a pre-defined sunspot mask (window) and acquire the motion vector $\mathbf{v}^t_m$ by solving Eq. 10 in Peng et al. (2015) for each cloud block within this range. If the magnitude of $\mathbf{v}^t_m$ for a cloud block is close to 0, then in the subsequent cloud-layer determination step, we convert the relevant cloud pixels within this block into clear sky pixels.

Multi-layer Detection and View Stitching

To locate the best match and find the optimal solution of the similarity function φ, we introduce clustering and multi-layer aggregation during cloud tracking to utilize all cloud blocks in the visible range. Instead of considering only the maximum of φ, we propose tracking all possible $(\mathbf{d}^{1,2,t}, \mathbf{v}^t)$ combinations as potential solutions. In practice, we store a combination for a cloud block only if each *NCC* score in the φ calculation is above a certain threshold (here, we use 0.5). Therefore, a single block may have multiple motion and displacement vectors. Our goal is to obtain **L**, a collection of multiple potential $(\mathbf{d}^{1,2,t}, \mathbf{v}^t)$ combinations detected for all cloud blocks:

$$\mathbf{L} = \{(\mathbf{d}^{1,2,t}, \mathbf{v}^t) | NCC^* \geq 0.5, \mathbf{c} \in C\} \tag{2.6}$$

As Note that NCC^* refers to any *NCC* calculation listed in the φ equation. *C* denotes the set of all cloud blocks found in the cloud detection pipeline.

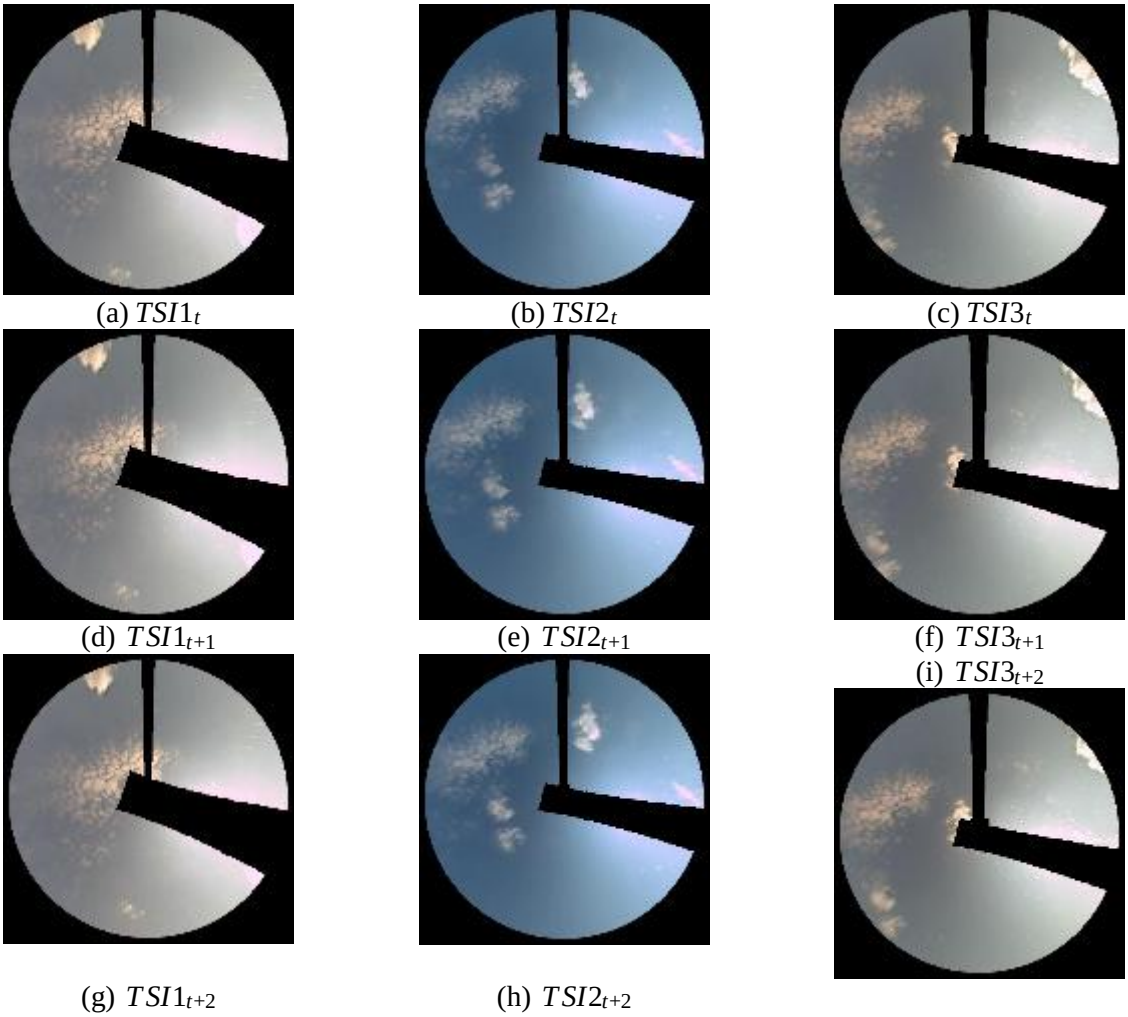
Based on **L**, we apply the k-means clustering (Singh and Glennen 2005) technique to generate layers of clouds. In our system, two layers at most are considered and aggregated. Hence, binary clustering is used to split **L** into two categories. If we use the clustering result to represent cloud layers, then the centroid of a cluster, $(\mathbf{d}^{1,2,t}, \mathbf{v}^t)$, represents the primary height and motion of this layer. Thereafter, we can group all cloud blocks into two layers/clusters based on the Euclidean distance between the reference solution $(\mathbf{d}^{1,2,t}, \mathbf{v}^t)$ and the centroids. The entire cloud block set *C* is then divided into two layers, *C*1 and *C*2, each of which contains multiple cloud blocks that should have similar motions and heights. To generate more accurate and robust information concerning the cloud layers, we assume that cloud blocks on the same layer possess only one major planar motion vector and one CBH. Similar to the form of the centroid in clustering, we define the wind field (WF), $(\mathbf{d}^{1,2,t}, \mathbf{v}^t)$, as the unique combination of the displacement and motion vectors corresponding to a given cloud layer. We calculate this combination of vectors by maximizing the summation of φ over all cloud blocks on the same layer.

If two wind fields extracted from two clusters are sufficiently similar to each other, then we treat them as a single cloud layer in which **c** can be drawn from the entire cloud block set *C*. When

the Euclidean distance between two motion vectors is less than two pixels, or their height difference is less than 500 meters, we consider these two wind fields to belong to a single layer. To further improve tracking performance and reduce computational complexity, all generated wind fields are stored as historical layers. When searching for a reasonable solution set, this historical reference is used to refine the possible range of motion vectors and height levels. This approach accelerates the search procedure and increases the system's robustness in coping with noise.

An example with two (layers of) wind fields detected is shown in Figure 2-12 and Figure 2-13. Nine images in three consecutive frames from timestamp t to $t+2$ are used to extract potential solutions. In Figure 2-13, the displacement vectors of $TSI_2 \rightarrow TSI_3$ and $TSI_2 \rightarrow TSI_1$ are shown in blue and green respectively, whereas the cloud motions are shown in red. Figure 2-13a-c confirms that to maximize φ for each individual block does not always guarantee finding the best matches. We observe that cloud blocks near the shadow band and supporting arm (black area) exhibit obvious errors due to the information loss caused by blank pixels. However, after clustering and aggregating, the cloud blocks can be categorized into two layers, $WF_1 = (1201m, 10px, -6px)$ and $WF_2 = (4184m, 7px, -2px)$. As shown in Figure 2-14d-f, these two layers are indicated by red and green boxes, respectively, and offer more stable estimations of displacement and motion than do the individual cloud blocks.

We also present another example of a single WF detected from the nine input images shown in Figure 2-11. In Figure 2-14a-c, several combinations of height and motion that achieve the maximum value of ϕ are deemed outliers and marked with green rectangles. Running the clustering algorithm identified two wind fields: $WF_1 = (4357m, 8px, 8px)$, and $WF_2 = (2098m, 8px, 9px)$. However, these two layers are very similar to each other given their similar motion vectors. Thus, we aggregated all blocks into one layer, as indicated by the red boxes in Figure 2-14d-f. The output WF_{single} is then represented as a single tuple $(2130m, 8px, 9px)$ which is less sensitive to the boundary or noisy cases.



(a) $TSI1_t$ (b) $TSI2_t$ (c) $TSI3_t$

(d) $TSI1_{t+1}$ (e) $TSI2_{t+1}$ (f) $TSI3_{t+1}$

(g) $TSI1_{t+2}$ (h) $TSI2_{t+2}$ (i) $TSI3_{t+2}$

Figure 2-12. Example of cloud tracking on nine images.

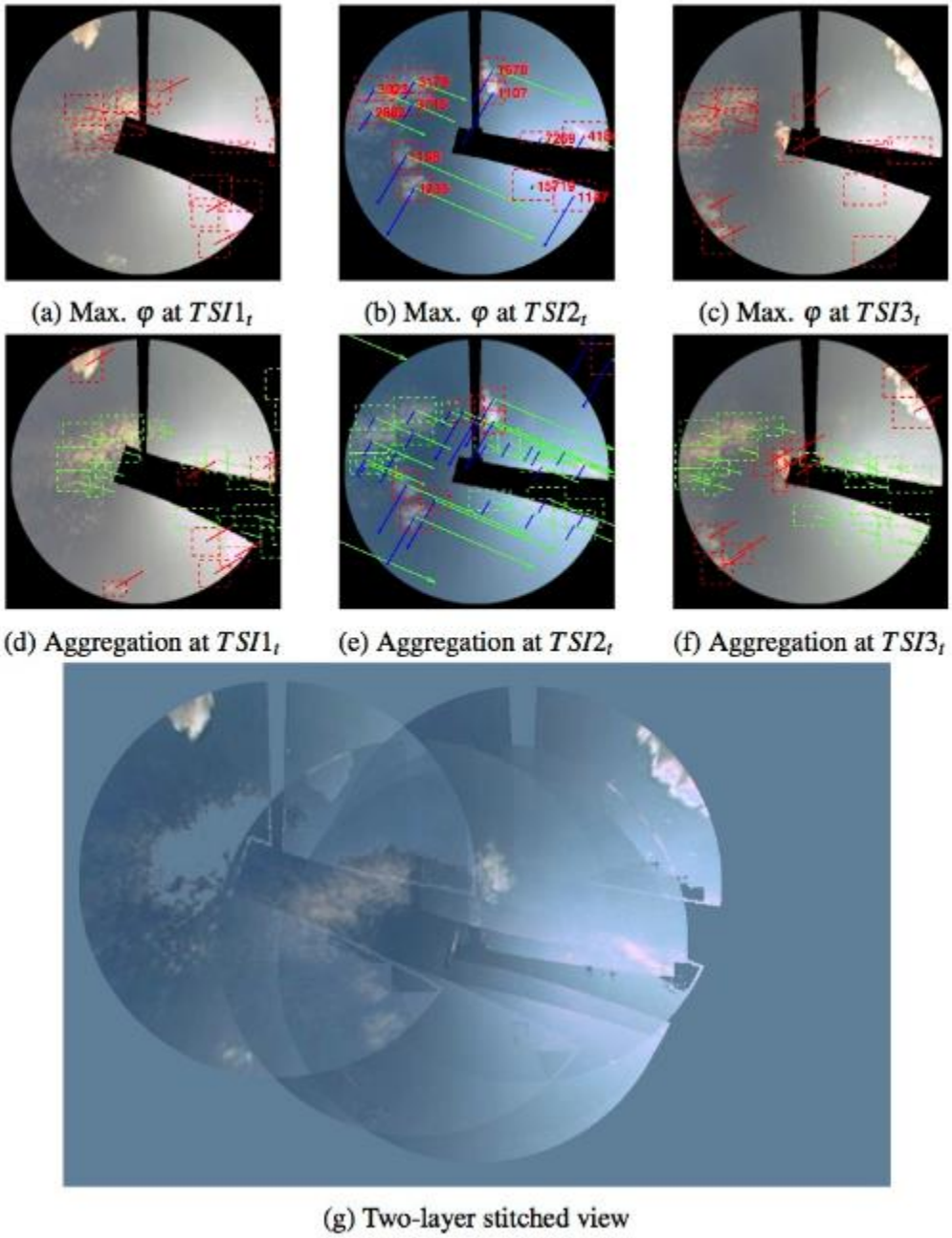


Figure 2-13. Determination of two cloud layers corresponding to the images shown in Figure 2-12. (a), (b), and (c) show the matching results for ($\mathbf{d}_m^{1,2,t}$,$\mathbf{v}_m^t$). The estimated heights are labeled in (b). (d), (e), and (f) are marked with blocks indicating the two wind fields (red and green). The arrows in the images from TSI_1 and TSI_3 represent motions, whereas the arrows in the images corresponding to TSI_2 represent the $TSI_2 \rightarrow TSI_1$ (green) and $TSI_2 \rightarrow TSI_3$ (blue) displacement vectors.

After determining the cloud layers, multiple TSI views can be stitched together block by block. We select TSI_2 as the origin/center of this combined view because it is located in the middle. The pixels of a cloud block from TSI_1 and TSI_3 can then be mapped to TSI_2's coordinate system through pixel-wise shifts of the forms $TSI_1 \rightarrow TSI_2$ and $TSI_3 \rightarrow TSI_2$ which are identical to the

extracted displacement vectors $\mathbf{d}^{\mathbf{1,2,t}}$ and $\mathbf{d}^{\mathbf{1,3,t}}$. Thus, given the estimated heights of the cloud layers, all cloud blocks from TSI_1 and TSI_3 are placed and stitched into the aggregated view from all three TSIs. An example of such a stitched view wherein all blocks belong to a single layer is shown in Figure 2-14g. For a case with more than one layer, the stitched view is generated by stitching layers one by one, i.e., from a higher altitude layer to a lower one, because low-altitude clouds can overlay those at higher altitudes. Consequently, the output view should be similar to Figure 2-16g. The black areas where gaps still exist are marked as blank, and we fill them in with the default sky color for visualization. The default value is calculated by averaging all clear-sky pixels across multiple TSI views.

2.1.4.4 Short-term Solar Forecasts

Much of the material in this section is drawn from section 5 of Peng et al. (2015).

In this section, we describe the methods for extracting image features and the different irradiance models used to forecast solar irradiance. First, given the locations of the 25 pyranometers, we need to identify the relevant pixels on the TSI images that are correlated with the surface irradiance fluctuations at these locations. Since the clouds located between the sun and the solar panels are the primary cause of ramps in irradiance, the basic concept of irradiance forecasting is to predict whether clouds will block the sunlight at a specified time. As shown in Figure 2-4, the device's projection on the sky image corresponds to the intersection of the projected plane (i.e., the image plane) with a line drawn from the sun to the pyranometer. If a cloud is located in the path of the sunlight traveling to the pyranometer, then its projected position on the image plane must overlap with the device's projection. Furthermore, we define the corresponding pixel at the projected position in the stitched view of the three TSIs as a "sun-blocking" pixel. Because this pixel is correlated with the amount of direct sunlight cast on a pyranometer, our goal is then to extract the image features of this pixel, which will subsequently serve as the input to the irradiance model. As shown in Figure 2-4, the position of the sun-blocking pixel depends on the angle of the Sun, the geographical location of the pyranometer, and the height of the cloud that is blocking the direct sunlight. Using these inputs, we can easily calculate the sun-blocking pixel for each of the 25 pyranometers, as shown in Figure 2-15a-b.

To predict which pixels will become sun-blocking pixels in the future, we apply a backtracking method to the current stitched view. In this processing step, the backtracking is guided by the motion vectors detected at previous timestamps: given that the cloud motion in a particular layer is $\mathbf{v}^{\mathbf{t}}$ where t is the current timestamp, the pixel $\hat{\mathbf{c}}^t$ that will potentially become a sun-blocking pixel $\mathbf{c}^{t+N}$ in the Nth future time frame (at time $t + N$) is then calculated by moving $\mathbf{c}^{t+N}$ in the direction opposite to the motion vector $\hat{\mathbf{c}}^t \leftarrow \mathbf{c}^{t+N} - \mathbf{v}^{\mathbf{t}} \times N$. Here, we assume that the motion vector remains unchanged between t and $t + N$ and that the pixel at $\hat{\mathbf{c}}^t$ will become $\mathbf{c}^{t+N}$ and block the sunlight after N time frames, provided that it contains a cloud at that time. If multiple layers of clouds move to this pixel location, then the lower-altitude cloud pixel is preferred because it will block the higher-altitude one.

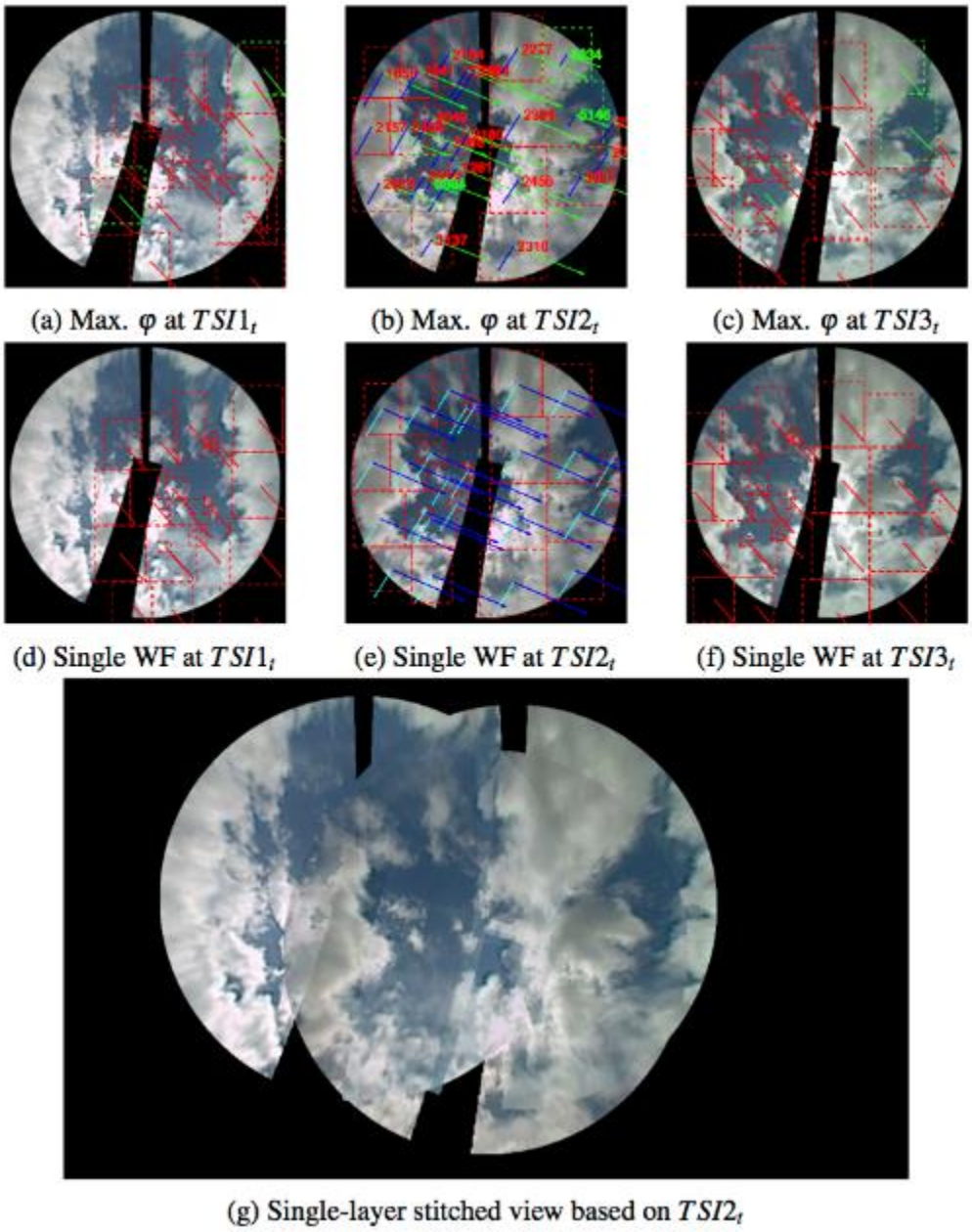
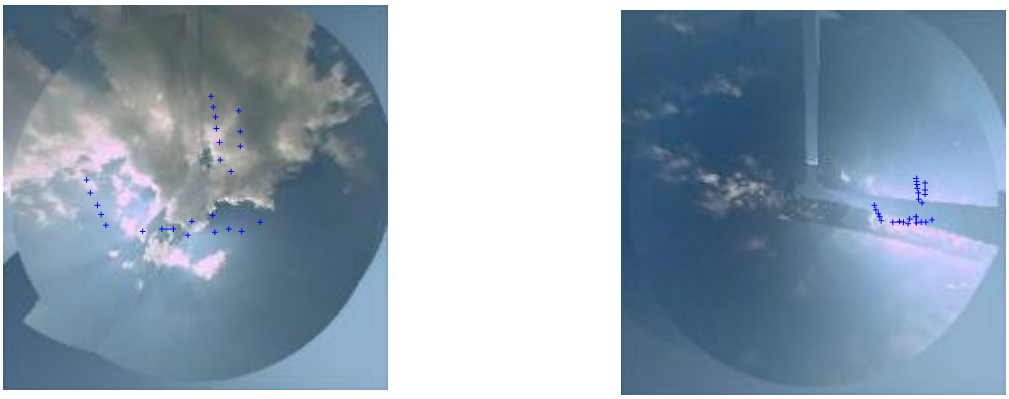


Figure 2-14. The single layer detected from the nine images shown in Figure 2-11. (a), (b),and (c) show the tracking results for ($\mathbf{d}_m^{1,2,t}$,$\mathbf{v}_m^t$). The regions marked with green boxes in (b) display obvious bias due to the boundary effect. (d), (e), and (f) show the single-layer field, (h^t, v_x^t, v_x^t) = (2130 m, 8 px, 9 px) marked in red.



(a) h=1000 at 2:00PM (b) h=3000 at 9:00AM

Figure 2-15. Examples of the locations of the sun-blocking pixels (blue crosses) for the 25 pyranometers in the stitched multi-TSI view.

As expected, the effectiveness identification of the sun-blocking pixel requires an accurate estimation of the motion vectors of each layer. Any error in this information will further accumulate in subsequent steps. To mitigate this effect, we have devised two strategies to be implemented in the feature extraction step. First, instead of focusing on a single sun-blocking pixel, we considered the 7x7 sun-blocking window surrounding this pixel. This enables us to reduce the risk of false prediction, and more importantly, to include more features from the neighboring pixels. Second, we use 23 significant features to describe the spectral variation inside the window at the current time t and the predicted time $t+N$. Among these features, nine are the average, minimum, and maximum values of the RGB channels of the window surrounding $\mathbf{c}^t$, which describe the spectral properties of the current observation. Similarly, we choose nine additional features from the sub-blocking window centered at $\hat{\mathbf{c}}^t$, which represent the estimated properties. The remaining five features, are the RBR of $\mathbf{c}^t$ (RBR^t) and $\hat{\mathbf{c}}^t$ ($R\hat{B}R^t$), the cloud fraction at time t, and the ground-truth irradiance values at the current time t (k^t) and one-minute earlier, at $t-6$ (k^{t-6}). The RBR at both timestamps are used because this quantity shows a noticeable spectral difference between cloud and sky (Morris 2005). The cloud fraction represents the overall cloud conditions. The two remaining features, the ground-truth irradiance values at t and $t-6$, enhance the performance for a short forecasting window because they incorporate persistent observations. Here, the forecasting problem is formulated as $\hat{k}^{t+N} = f(\mathbf{x}^t)$, where $\hat{k}^{t+N}$ represents the predicted irradiance at $t+N$ and $\mathbf{x}^t$ is the vector of the 23 extracted features.

To examine the predictive capabilities of the selected features, we explore four different irradiance models that use a subset or all of these features to generate a regular linear regression or more complicated non-linear relationships $f(\mathbf{x}^t)$: 1) linear RBR delta (linear_δ), 2) an ordinary linear regression model (linear_{all}),, 3) support vector regression (SVR) (Drucker et al. 1997) based on a linear kernel (SVR_{linear}), and 4) SVR with a non-linear kernel (SVR_{rfb}). For comparison, we use the persistence model as the baseline model, i.e., the radiation shift, R_{Shift}, which directly uses the current observation as the forecasting result.

In addition, we introduce one regression model that depends only on multiple irradiance values and is used to further validate image features for prediction capability in our comparison studies.

First, we simply extend the persistence model by incorporating a linear regularization term based on the information from the sun-blocking pixel. Our preliminary study of motion vectors (Huang et al. 2013) indicates that the RBR of a sun-blocking pixel is a useful indicator of the cloud transmittance at this pixel. The linear RBR delta model, denoted by $linear_\delta$, is a weighted difference between the two.

However, the ordinary model is sensitive to noise or outliers and suffers from the overfitting problem. To overcome these shortcomings, we apply the SVR (Smola and Schölkopf 2004), an extension of the SVM approach, for regression.

2.1.5 Initial evaluations

Much of the material from this section comes from section 6 of Peng et al. (2015).

2.1.5.1 Experimental Dataset

We chose the period from 13 May to 3 June 2013, encompassing various weather conditions, to evaluate the performance of 1-min to 15-min-ahead 3-D cloud tracking and irradiance forecasting. To guarantee consistent cloud visibility and zenith angle, we filtered out records whose timestamps were not between 9:00 and 16:00 Eastern Standard Time (EST). Our experimental dataset contains 9963x3 images from three TSIs that collect data simultaneously and the synchronized 9963x25 GHI records from 25 ground-based pyranometers. Furthermore, we divided the full dataset into four categories based on weather and cloud conditions: single-layer clouds (D_s), multi-layer clouds (D_m), overcast or extremely cloudy (D_o), and a mixture thereof (D_{mix}). D_s corresponds to a typical low-altitude cloud that is commonly observed on the east coast of the United States. This type of low cloud typically appears in a single layer and remains in the field of view of a TSI for several minutes at most. Given its quick movement in and out of the FOV of a TSI, D_s is an ideal dataset for demonstrating the capability of the new tracking system to capture rapid cloud motion and evolution. D_m is a collection of cases wherein multiple layers of clouds were observed within the FOV of all TSIs. Hence, the tracking and forecasting performance with respect to D_m is expected to reflect the ability of the system to categorize and track multiple wind fields. D_o corresponds to cloudy and overcast cases. Based on our observations, the textural patterns of the images acquired in extremely cloudy conditions are not obvious and are hence difficult to utilize for block-matching and the extraction of image features. Therefore, we designed D_o to demonstrate the sensitivity of this new tracking system when applied to TSI images with less obvious patterns. D_{mix} contains the remaining cases, including sunny, partially cloudy, and thin-layer conditions. The cloud tracking and irradiance forecasting performance with respect to D_{mix} reflects the system's average performance for a mixture of cloud conditions. Table 2-2 describes the details of all subsets.

2.1.5.2 Evaluation Metrics and Validation Method

As discussed above, the prediction of the sun-blocking pixels for the 25 pyranometers is performed based on cloud movements in a particular cloud layer. When a cloud moves too rapidly, or the forecasting interval is too long, one or more positions may be outside of the FOV of our TSI. In this case, we cannot extract any features for modeling and predicting irradiance. To quantify the tracking capability for sun-blocking pixels within our stitched view,

we define a metric named the successful tracking index (*STI*), which is the percentage of instances in the designated dataset for which all 25 sun-blocking pixels can be included in the FOV of the stitched image for a particular forecasting horizon.

We also use the mean absolute error (MAE) and root-mean-square-error (RMSE) as evaluation metrics. In our experiment, the MAE measures the average accuracy of cloud tracking, whereas the RMSE assigns greater penalties to large errors, such as false estimation of a cloud's presence. To avoid bias and control the over-fitting problem, we introduce the cross-validation technique (Kohavi 1995) into the modeling and evaluate the forecasting performance across all 25 stations. In the cross-validation, the original dataset is evenly divided into several independent subsets, and the average performance of the predictive model measured across these subsets. In this experiment, five-fold cross-validation ($N_{cv} = 5$) is applied, the five folds (four for training and one for testing) randomly generated. Additionally, since 25 simultaneous, location-dependent ground measurements are available, we can verify a forecast by comparing it with the observations from a different measurement station, *s*, at time *t*. Hence, the final error metrics can be formulated as the average performance of all five-fold tests across all 25 stations. The MAE and RMSE scores are calculated based on the normalized GHI values.

To evaluate the effectiveness of the image features extracted from our multi-layer cloud detection and tracking system, we created a new reference model, SVR_k, for comparison. This model also uses SVR with only radiation observations as inputs and excludes all image-based features, in contrast to our proposed forecast model SVR_{rbf}. In greater detail, SVR_{rbf} and SVR_k share the same radial basis function as their SVR kernel and use identical parameter settings for both training and testing. The difference is that SVR_k uses only the l (here, we choose $l = 6$) most recent radiation values, i.e., $k^{t-5}, k^{t-4} \ldots k^t$, as inputs. We apply both SVR_{rbf} and SVR_k to generate real-time irradiance predictions to validate the effectiveness of the image features in producing short-term solar forecasts.

2.1.5.3 Model Performance

Figure 2-17 shows the 1-15 minute *STI* scores of the entire dataset (bars) and the four subsets (lines). We observe that the *STI* scores for D_s and D_m decrease dramatically beginning with the nine-minute forecast, whereas the performance for the other two subsets remain relatively stable between one and fifteen minutes. Consequently, the success ratio for the entire dataset *overall* also decreases as the time horizon increases. The *STI* decreases for longer forecast horizons (longer than ten minutes) because of fast-moving clouds. According to our observations from the TSI images, on the east coast of United States, these clouds are mostly distributed in a single layer with a cloud base height below 3000 meters. They often have a high velocity, as detected at the pixel level, and exhibit rapid formation/dissipation within a 10-minute window. Thus, given the limited visible range and TSI resolution, 3-D cloud tracking can capture the majority of low-layer clouds for only up to nine minutes. Beginning at the 10-minute horizon, low clouds are highly likely to move out of the field of our stitched view. This cloud property, combined with the physical limitations of the tracking TSIs is consistent with the significant decrease in the *STI* of D_s at the nine-minute horizon, reaching almost 0 at the ten-minute horizon. Similarly, the *STI* of D_m decreases after nine minutes. However, since more than one layer is present in D_m, the tracking results benefit from partial estimations from the higher layers which tend to be more stable and have slower pixel-

wise motion vectors. For D_o and D_{mix}, most cases correspond to high-layer clouds or mixed conditions and are therefore suitable for ten-minute forecasts and beyond.

We evaluated the irradiance forecasting performance over the entire dataset based on the metrics of the MAE and RMSE scores, and the results are presented in Table 2-3 and Table 2-4. We excluded the out-of-FOV data points and only trained and tested the irradiance models based on the remaining available data subset. During model training, we discovered that when STI is low, the models tend to overfit the data. One reason for this behavior is that an excess of out-of-FOV records leads to a lack of observations in the training folds. For instance, if we train the model using D_s for forecast horizons longer than ten minutes, we barely have enough training records to generate the forecasting model. Another reason is that we may introduce bias into the forecasting models. Since rapidly changing cases, such as those with low-altitude clouds, are excluded for long forecasting horizons, the forecasting models will place more weight on the "easy" cases, such as those corresponding to sunny and overcast conditions. Hence, if too many records in an experimental dataset are out of the FOV, or the STI value is below a certain threshold (here, 60%), we mark the result with an asterisk to indicate a partial forecast and a potential over-fitting problem. When all records are out of the FOV (STI = 0), we denote this scenario with '-', indicating that no forecasting result is available (see Table 2-4).

Table 2-2. Descriptions of four subsets with various cloud and weather conditions. image#: number of TSI images, k#: number of GHI measurements, cf : estimated range of cloud fractions in the sky images, cfσ : mean and standard deviation of the cloud fraction, k: range of the clear-sky index. kσ : the mean and standard deviation of k, exp: observed (ab)normal condition of the TSI images, CBHest : cloud height range, vest : cloud motion in image, WF#: number of cloud layers, zenith: solar zenith range, condition: cloud conditions.

Subset	D_s	D_m	D_o	D_{mix}
$image_\#$	2517x3	2520x3	2406x3	2520x3
$k_\#$	2517x25	2520x25	2406x25	2520x25
cf	[0.07,0.94]	[0.03,0.94]	[0.94,0.95]	[0,0.95]
cf_σ	0.71 ±0.25	0.70 ±0.27	0.94 ±0.01	0.50 ±0.35
k	[0.11,1]	[0.06,1]	[0.17,1]	[0.11,1]
k_σ	0.46. ±0.22	0.57 ±0.27	0.63 ±0.21	0.81 ±0.23
exp	normal	underexposed	normal	green-dominant
$WF_\#$	1	2	2	2
$CBH_{est}\,(m)$	[1590,2960]	[1890, 4420]	[6020,15730]	[440 12330]
$v_{est}\,(px/min)$	[36,60]	[6,36]	[0,54]	[0,60]
zenith	$[40^\circ, 57^\circ]$	$[42^\circ, 58^\circ]$	$[41^\circ, 56^\circ]$	$[41^\circ, 57^\circ]$
condition	low,scattered	multi-layer	overcast	Mixture

Table 2-3. MAE and RMSE metrics for 1-minute and 5-minute irradiance forecasts. The subscript of each score indicates the subset type used to train and test the model. The subscript "avg" indicates the average performance on the entire dataset.

Dataset	R_{shift}		$linear_\delta$		$linear_{all}$		SVR_{linear}		SVR_{RBF}	
	1 m	5 m	1 m	5 m	1 m	5 m	1 m	5 m	1 m	5 m
MAE_s MAE_m	0.16	0.20	0.14	0.20	0.11	0.16	0.10	0.16	0.09	0.14
MAE_o	0.16	0.19	0.14	0.19	0.11	0.15	0.10	0.14	0.08	0.13

MAE_{mix}	0.04	0.08	0.04	0.08	0.03	0.07	0.03	0.07	0.03	0.06
MAE_{avg}	0.11	0.16	0.12	0.17	0.09	0.13	0.07	0.12	0.06	0.09
	0.12	0.17	0.12	0.17	0.10	0.16	0.09	0.15	0.07	0.12
$RMSE_S$	0.23	0.27	0.20	0.26	0.15	0.20	0.15	0.20	0.14	0.18
$RMSE_m$	0.26	0.29	0.21	0.28	0.15	0.20	0.15	0.21	0.14	0.20
$RMSE_O$	0.06	0.11	0.06	0.11	0.05	0.09	0.05	0.09	0.04	0.08
$RMSE_{mix}$	0.23	0.29	0.22	0.26	0.15	0.20	0.17	0.21	0.13	0.18
$RMSE_{avg}$	0.23	0.27	0.21	0.26	0.15	0.21	0.17	0.22	0.13	0.19

Table 2-4. MAE and RMSE metrics for 10-minute and 15-minute irradiance forecasts. '-' indicates that no forecast output is available due to cloud tracking failure. '' indicates an incomplete dataset that has a low STI value or an average performance influenced by incomplete/empty subsets.*

Dataset	R_{shift}		$linear_{\delta}$		$linear_{all}$		SVR_{linear}		SVR_{RBF}	
	10 m	15 m	10 m	15 m	10 m	15 m	10 m	15 m	10 m	15 m
MAE_S MAE_m	0.27*	-	0.25*	-	0.24*	-	0.16*	-	0.13*	-
MAE_O	0.20	0.21*	0.20	0.21*	0.16	0.18*	0.15	0.17*	0.13	0.15*
MAE_{mix}	0.09	0.11	0.09	0.11	0.08	0.10	0.08	0.09	0.07	0.08
MAE_{avg}	0.18	0.19	0.20	0.20	0.14	0.16	0.12	0.13	0.09	0.11
	0.17*	0.17*	0.18*	0.17*	0.17*	0.16*	0.15*	0.15*	0.12*	0.11*
$RMSE_S$	0.33*	-	0.30*	-	0.36*	-	0.21*	-	0.18*	-
$RMSE_m$	0.30	0.30*	0.29	0.30*	0.21	0.23*	0.22	0.24*	0.21	0.23*
$RMSE_O$	0.12	0.15	0.12	0.15	0.11	0.12	0.11	0.13	0.10	0.11
$RMSE_{mix}$	0.31	0.32	0.30	0.31	0.21	0.22	0.23	0.24	0.20	0.21
$RMSE_{avg}$	0.29*	0.29*	0.28*	0.28*	0.22*	0.22*	0.23*	0.23*	0.21*	0.20*

In Figure 2-16a-b we evaluate the effectiveness of the four forecasting models by comparing them with the persistent model for one- to fifteen-minute forecasts. To ensure that the sun-blocking pixels are contained within the stitched view for the majority of the training and test cases over the full forecast horizons, we use the experimental dataset D_{mix} that has the most stable STI values (Figure 2-17). Figure 2-16a-b show that $linear_{\delta}$ consistently acquires less large forecast errors in comparison with the persistent model (as measured by RMSE), but exhibits the worst performance in terms of average accuracy of irradiance forecast (as measured by MAE), confirming that tracking a single sun-blocking pixel leads to a high risk of deviating from its real position and thereby falsely predicting the presence of clouds. By virtue of our proposed feature extraction in the sun-blocking window, this problem is mitigated in the other three models that consistently provide better results in both metrics. In comparison to $linear_{all}$, which was trained using least-square errors, the linear SVR approach, SVR_{linear}, tends to be more robust in terms of average errors (MAE); however, it is very sensitive to large errors (RMSE). The proposed non-linear model, SVR_{rbf}, effectively addresses this problem by mapping non-linear relations to linear ones in a higher dimensional space. As is evident from these two figures, SVR_{rbf} effectively reduces the occurrence of large errors and outperforms the other four models in terms of the RMSE and MAE metrics.

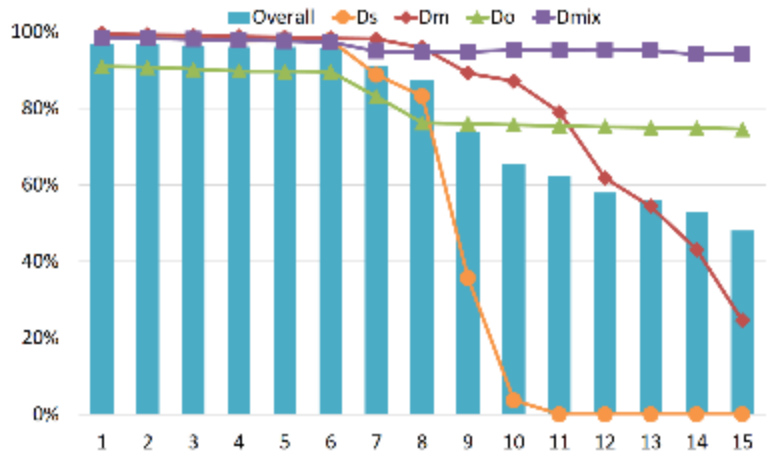
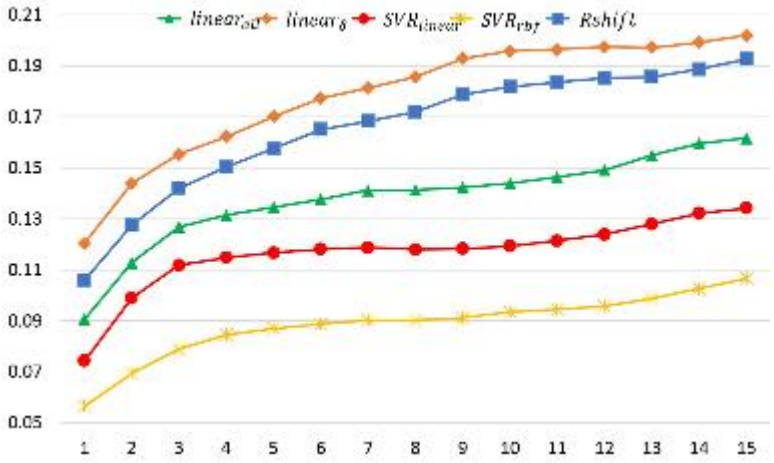
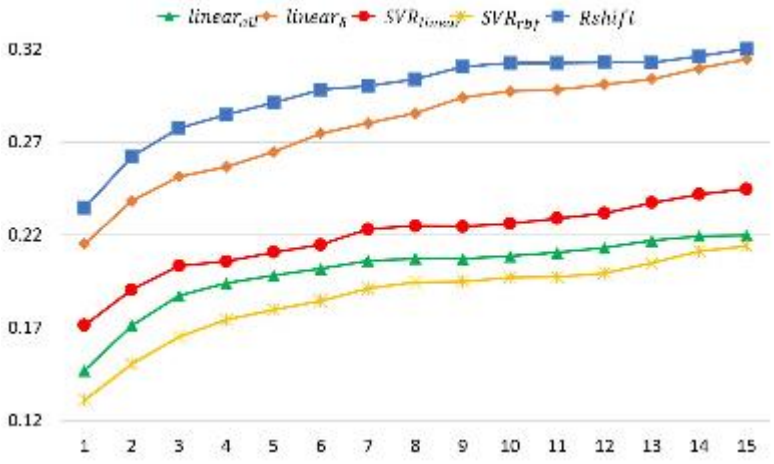


Figure 2-16. Successful Tracking Index (STI) values for the datasets in the 1-15 minute forecasting range. Overall represents the results for the entire dataset, which contains all four independent subsets.



a) Average MAE scores for five folds



b) Average RMSE scores for five folds

Figure 2-17. MAE and RMSE scores for irradiance predictions on the data subset D_{mix} over a time range from one to fifteen minutes.

We also analyzed the performance of SVR_{rbf} on the entire dataset and compared it with R_{Shift} across all 25 pyranometer measurements. Figure 2-18 shows the percentage of reduction in MAE achieved using SVR_{rbf}. The blue shaded regions represent the upper and lower improvement bounds for all 25 pyranometers, whereas the mean improvements are plotted as blue dots with standard deviation bars. We observe that beyond a nine-minute horizon, the uncertainty in performance improvement, as indicated by the upper and lower bounds, increases with the increase in forecasting horizon. This is expected because many data points are out-of-FOV for the D_s and D_m subsets, which affects the reliability beyond the nine-minute horizon. Moreover, several successful cases that remain stable even beyond the nine-minute forecast are simpler ones, such as sunny and overcast conditions. Therefore, the persistence model R_{Shift} can take advantage of these cases to minimize error, and is thus difficult to outperform. In the same plot, we observe that both the difference between the upper and lower bounds and the standard deviation increase as the time horizon increases. We observe that despite the expected uncertainties arising for a long-term forecast, the SVR_{rbf} model is nevertheless significantly superior to the persistent model, achieving at least a 26% improvement.

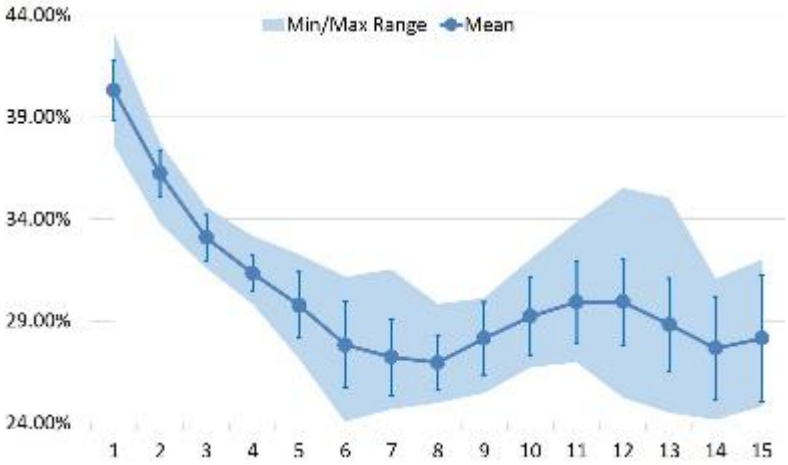


Figure 2-18. Improvements in the MAE ratio achieved by the non-linear SVR_{rbf} model in comparison with the persistent model on all available data. The Min/Max bounds represent the range of the percentage improvement values for all 25 stations. The average performance is denoted by the plotted line, which includes standard deviation bars on either side.

To apply this study to real-time forecasting, we averaged the SVR_{rbf} parameters over the entire dataset and constructed the radiation-only model SVR_k to generate 5-, 10-, and 15-minute forecasts for 14 May 2013. Figure 2-19 compares the forecasting results of SVR_{rbf} and SVR_k with the real measurements from the deployed pyranometers. In the individual panels, the normalized GHI predictions are converted back to real values using Equation 2.1.1. To be consistent with the irradiance plots in Figure 2-19, the RMSEs in the caption are calculated directly from ground-truth measurements (real GHI values) without normalization. The gray/dark areas in these figures represent gaps with no prediction, which include the cases corresponding to the low zenith angles during the early morning and late afternoon as well as the periods in which sun-blocking pixels were out of the FOV. The results show that the five-minute forecasts generated using SVR_{rbf} achieve good accuracy and capture most radiation ramps. Meanwhile, the number of detected ramps decreases for ten-minute and fifteen-minute forecasts because of the instability of the motion

vectors and the occurrence of clouds that may reside outside the field of the stitched view. Moreover, because SVR_k only relies on radiation features, we observe that it exhibits a behavior similar to that of the persistence model, often failing to detect radiation fluctuations and generating false alarms based on previous irradiance trends. For a longer forecasting horizon, such as 10 or 15 minutes, the forecasting accuracy of SVR_k decreases rapidly, and the model cannot faithfully predict irradiance ramp events. Compared with SVR_k, SVR_{rbf} incorporates multiple features derived from predicted cloud movements and sky images and consequently, introduces fewer forecasting errors and captures more ramp events.

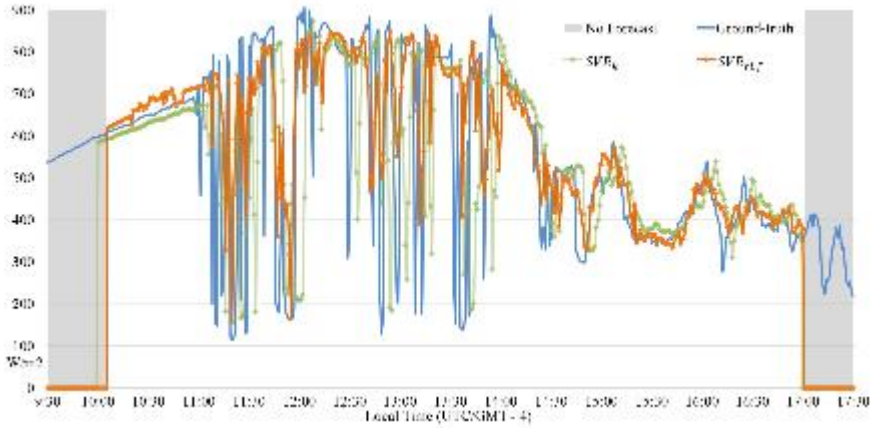
2.1.6 Summary of TSICast Contribution

BNL conducted research in short-term solar prediction and implemented a prototype system to forecast solar irradiance in real-time. In particular, we seamlessly incorporated new imaging technology, proposed and implemented cost-effective methods for image-based solar forecasts. To achieve the best performance, we studied image-based cloud tracking and feature extraction. This work integrated multiple image datasets into a very short-term solar forecast system, TSICast.

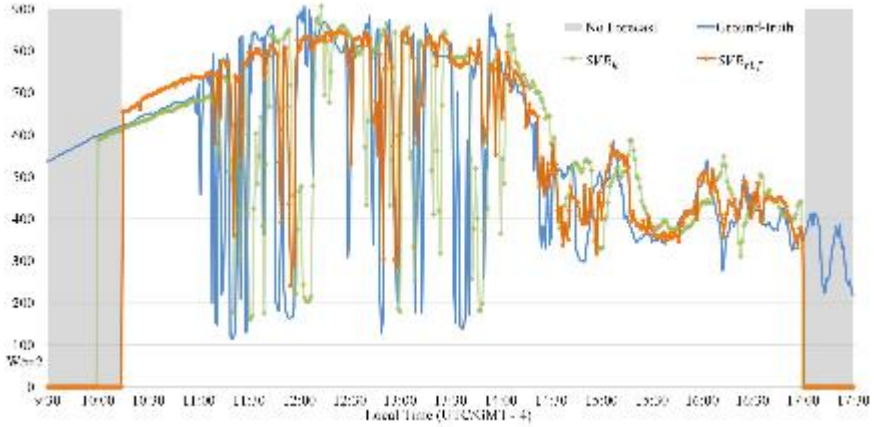
With the development of various imaging systems and growing interests in image-based solar forecasts, there is an urgent need to provide reliable approaches to extract cloud information and to build irradiance models from sky images. BNL extracted cloud spatial-temporal information and investigated a series of image analysis modules for a complete forecast system, including image preprocessing, motion tracking, multi-camera integration, and multi-layer determination in various types of sky images. To improve the robustness of image-based prediction and extend the forecasting range, our work integrates heterogeneous image datasets and ensembles a series of cloud tracking and prediction algorithms, each of which has its own strength and weakness in different cloud conditions and prediction periods. BNL's most significant contributions include:

- BNL developed a cloud detection pipeline that utilizes a supervised classifier and abnormal image correction based on histogram equalization. This research significantly improved the accuracy of extracting cloud mask and attained good performance in various cloud types, weather conditions, and lighting patterns.
- On the basis of previous work of tracking clouds in sky imagery, BNL designed a hybrid model of cloud motion tracking that combines block-matching and optical flow. The new model is able to determine local deformations of clouds, to extract cloud layers with dominant motion patterns, and to remove noise from the resulting motion field with customized motion filters.
- BNL designed a comprehensive framework of cloud image simulations and generated synthetic image sequences using motion models and Gaussian noise. The simulated image is used for the evaluation of motion tracking models under different simulations, e.g., with cloud deformation and corrupted images.
- BNL devised a short-term solar forecast system utilizing ground-based sky cameras. The system adopts multi-angle observations to undertake the task of cloud tracking based on spatial and temporal correlation and provides a pipeline to detect multi-layer motions via clustering. The robust feature extraction and irradiance models are then vetted for real production forecasts. Compared with single-camera models, the proposed

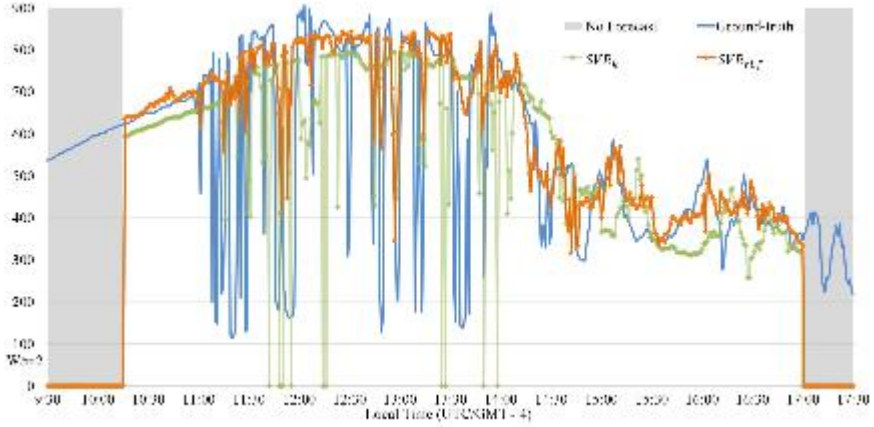
system significantly enlarges the field of view, enables 3-D cloud tracking, and obtains more accurate forecasts.



a) 5-minute forecasts with $RMSE_{rbf} = 145.3$ and $RMSE_k = 213.5$.



b) 10-minute forecasts with $RMSE_{rbf} = 171.3$ and $RMSE_k = 223.6$.



c) 15-minute forecasts with $RMSE_{rbf} = 177.5$ and $RMSE_k = 241.0$.

Figure 2-19. Real forecasts based on our new prediction system using SVR_{rbf} and SVR_k. Gray/dark areas with a flat "0" or no forecast value represent data points that are out-of-FOV or correspond to a low zenith angle. $RMSE_{rbf}$ and $RMSE_k$ are the root-mean-square square errors of SVR_{rbf} and SVR_k compared with the real GHI values.

Publications From the TSICast Work

Zhenzhou Peng, Dantong Yu, Dong Huang, John Heiser, Shinjae Yoo, Paul Kalb, "3D cloud detection and tracking system for solar forecast using multiple sky imagers", *Solar Energy*, vol.118, pp. 496-512, Aug. 2015.

Zhenzhou Peng, Dantong Yu, Dong Huang, John Heiser, Paul Kalb, "A Hybrid Approach to Estimate the Complex Motions of Clouds in Sky Images", *Solar Energy, under review*.

Hao Huang, Jin Xu, Zhenzhou Peng, Shinjae Yoo, Dantong Yu, Dong Huang, Hong Qin, "Cloud motion estimation for short term solar irradiation prediction", in *Smart Grid Communications (SmartGridComm), 2013 IEEE International Conference on*, IEEE, 2013, pp. 696-701.

Zhenzhou Peng, Shinjae Yoo, Dantong Yu, Dong Huang, "Solar irradiance forecast system based on geostationary satellite", in *Smart Grid Communications (SmartGridComm), 2013 IEEE International Conference on*, IEEE, 2013, pp. 708-713.

Zhenzhou Peng, Shinjae Yoo, Dantong Yu, Dong Huang, Paul Kalb, John Heiser, "3D cloud detection and tracking for solar forecast using multiple sky imagers", in *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, ACM, pp. 512-517.

Jin Xu, Shinjae Yoo, Dantong Yu, Hao Huang, Dong Huang, John Heiser, and Paul Kalb, "A Stochastic Framework for Solar Irradiance Forecasting Using Condition Random Field", the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2015.

Jin Xu, Shinjae Yoo, Dantong Yu, Dong Huang, John Heiser, and Paul Kalb, "Solar Irradiance Forecasting using Multi-layer Cloud Tracking and Numerical Weather Prediction", the 29th Symposium On Applied Computing (SAC'15), Salamanca, Spain.

Hao Huang, Shinjae Yoo, Dantong Yu, Dong Huang, Hong Qin, "Correlation and Local Feature Based Cloud Motion Estimation", KDD Mutimedia Data Mining (MDM) workshop - MDMKDD 2012.

Hao Huang, Shinjae Yoo, Dantong Yu, Dong Huang, and Hong Qin, "Cloud Motion Detection for Short Term Solar Power Prediction". ICML 2011 Workshop on Machine Learning for Global Challenges.

Students Funded

Zhenzhou Peng 2011-2016 **Dissertation:** Multi-source Image Integration Towards Solar Forecast

Hao Huang 2009-2014 (Co-Advise with Professor Hong Qin), **Dissertation:** A Scalable Physics-based Data Modeling Framework to Unsupervised High-Dimensional Data Mining , **First Employment:** Research Scientist at General Electric Global Research

Jin Xu 2012-2015

2.2 StatCast and Statistical Prediction

2.2.1 Motivation for StatCast

Utility companies and system operators need accurate deterministic short-range solar irradiance forecasts, including estimates of the variability, so that they can adequately balance the rapid changes as well as supply and demand peaks in the electrical grid. Artificial intelligence forecasting techniques can be used to obtain accurate short-range solar irradiance forecasts from 15- to 180-min lead times and can also be used to predict spatial and temporal solar irradiance variability over the same short-range forecast lead-times. This project has explored various ways to construct and test such models for real-time operation, focusing on the SMUD regions, but deploying it throughout the project's partner sites.

2.2.2 Statistical Characterization of Solar Irradiance Variability

It is important first to characterize the variability, which was accomplished by Laura Hinkelman of the University of Washington in collaboration with her students and Manajit Sangupta of NREL.

2.2.2.1 Relationship of Resource Variability to Cloud Type

Text below is selected from Dr. Hinkelman's report to NREL, "Relating Solar Resource Variability to Cloud Type."

The results of this work will enable downscaling of satellite-based solar resource information under various cloud conditions and create data sets for transmission studies that contain the correct variability characteristics. Relationships between the temporal variability of solar irradiance measured at several stations in different areas of the United States and cloud type/properties were derived. Statistics describing the frequency of occurrence of the different cloud types at each location and relationships among the various cloud properties were also computed.

Solar irradiances averaged over 1 minute were obtained for the year 2009 from the seven NOAA SURFRAD stations. Geostationary Operational Environmental Satellite (GOES)-based cloud data at 4 km spatial resolution and half-hourly temporal resolution over the same time period was received from NREL. This data was provided for areas around the SURFRAD stations and the DOE Atmospheric Radiation Measurement (ARM) Program Southern Great Plains (SGP) site. Information about all of the stations used in this analysis is listed in Table 2-5.

Table 2-5. Measurement station geographic information.

Station ID	Location	Latitude (°)	Longitude (°)	Elevation (m)
BON	Bondville, Illinois	40.052	-88.373	213
DRA	Desert Rock, Nevada	36.624	-116.019	1007
FPK	Fort Peck, Montana	48.308	-105.102	634
GCR	Goodwin Creek, Mississippi	34.255	-89.873	98
PSU	Rock Springs, Pennsylvania	40.720	-77.931	376
SGP	Lamont, Oklahoma	36.605	-97.485	314

Station ID	Location	Latitude (°)	Longitude (°)	Elevation (m)
SXF	Sioux Falls, South Dakota	43.734	-96.623	473
TBL	Boulder, Colorado	40.125	-105.237	1689

The satellite data set provided a cloud type as well as several other cloud and radiative properties for each 4-km diameter footprint. These cloud types and the codes used to represent them are listed in Table 2-6. Note that although 13 cloud types are designated, several did not occur in the vicinities of the SURFRAD stations for the entire year of 2009: mixed ice and water, overlapping, dust, and smoke (types 5, 8, 10, 11, and 12). In addition, none of the data was assigned to the "unknown" category (type 10). This left seven types for the analysis: clear, probably clear, fog, water, supercooled water, opaque ice, cirrus, and overshooting (types 0-4, 6, 7, and 9).

Table 2-6. GOES data set cloud types.

Category	Description
0	Clear
1	Probably clear
2	Fog
3	Water
4	Supercooled water
5	Mixed ice and water
6	Opaque ice
7	Cirrus
8	Overlapping
9	Overshooting
10	Unknown
11	Dust
12	Smoke

We first analyzed the occurrence of the seven cloud types in the vicinity of the SURFRAD and SGP stations. In order to better characterize these locations, the nine closest footprints were examined rather than just the closest one. This meant that an area of approximately 12 km x 12 km around each station was observed. This compensated for variability around the stations or slight geolocation errors in the footprint positions.

The frequency of occurrence of each cloud type determined using data from all nine footprints around each station is depicted in Figure 2-20. The typical pattern is for clear skies to dominate, followed by supercooled and cirrus clouds. Liquid water and opaque ice clouds are also common, followed by footprints determined to likely be clear. Very few cases of fog (low water clouds) and nearly no overshooting clouds were detected at any of the sites. The cloud type distribution at Desert Rock, Nevada, stands out from the rest, with a very large number of clear footprints and a correspondingly lower number of each cloudy sky type. In particular, there are many fewer supercooled clouds in that location. This difference is due to the location of Desert Rock in an arid region, where the air is dry and temperatures high, so that few clouds form.

We next examined the spatial variability of the clouds within 6 km of the measurement stations. At each satellite imaging time, the most common cloud type occurring across the nine footprints was determined and the number of footprints with this cloud type was recorded. If fewer than five footprints had matching cloud types, this time was eliminated from consideration, because no single cloud type was dominant.

The results of this analysis are shown in Figure 2-21. In this case, data for all eight stations has been combined. Each bar represents the number of times a given cloud type is dominant, i.e., at least five of the footprints have this classification. The colored segments of the bars indicate the proportion of these times that the cloud type occurs in 5, 6, 7, 8, or 9 of the footprints, respectively. Bear in mind that the number of total occurrences of each cloud type varies greatly, so that 50% of one type does not represent the same number of total occurrences as 50% of another, as shown in Figure 2-20.

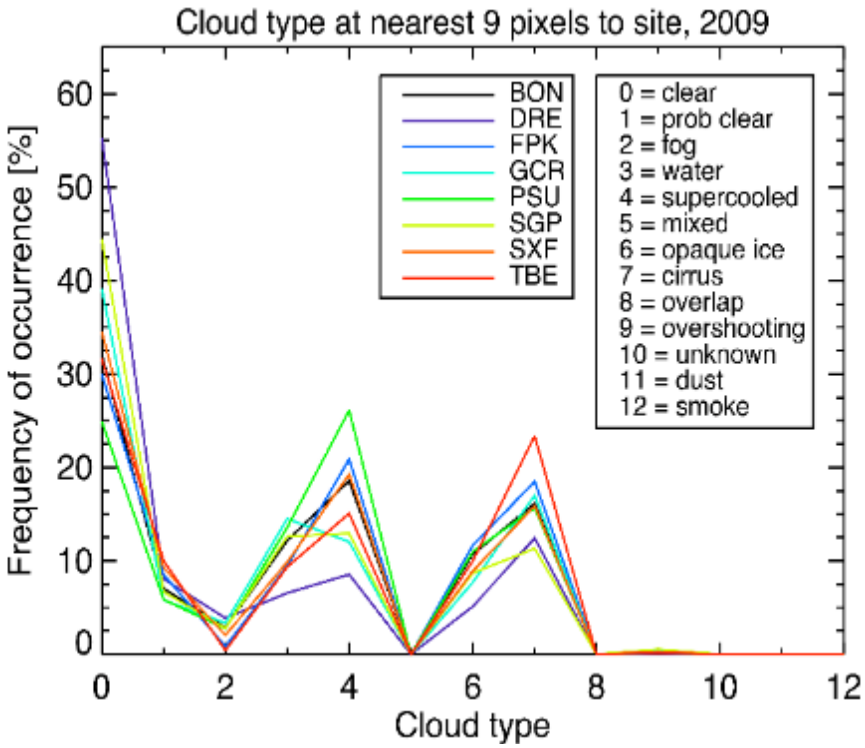


Figure 2-20. Cloud types for all SURFRAD sites and the SGP station combining data from the nine GOES-E footprints surrounding each station.

The results indicate that certain cloud types tend to occur over large areas while others are mixed with other, likely similar, types. For example, clear skies or overshooting clouds (i.e., thunderstorms) mainly occur uniformly over large areas, while fog and the "probably cloudy" class are most likely to be mixed with other clouds. This is likely due to the definitions of these classes.

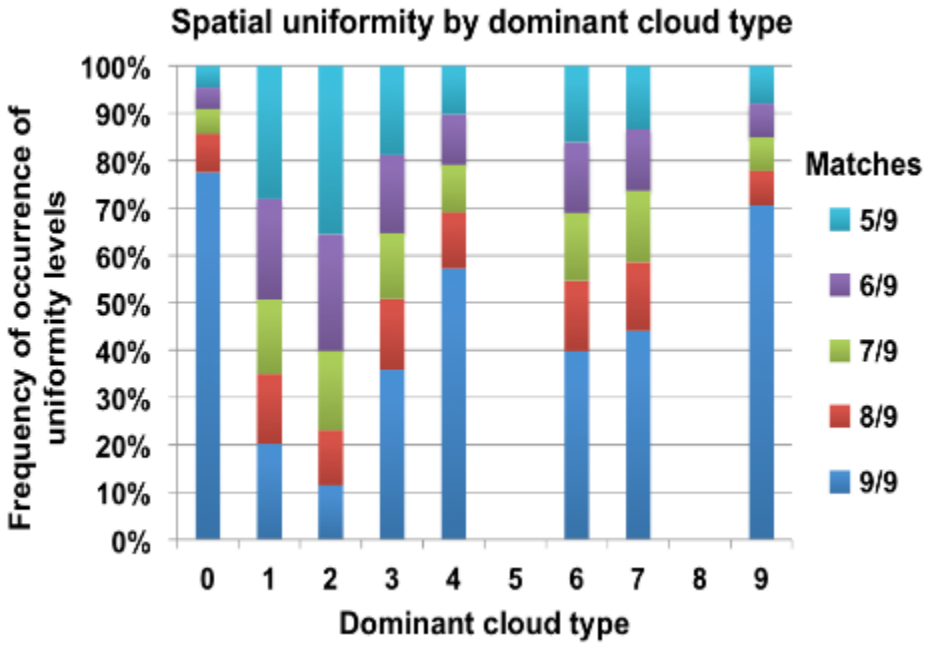


Figure 2-21. Spatial uniformity of cloud types around a ground station.

Here "fog" means low or thin liquid clouds. Thus spatial variations in liquid cloud properties will lead to different cloud classifications at neighboring footprints. Likewise, the classification of "probably clear" indicates that the retrieval is uncertain; some of these footprints are likely clear while others may contain thin clouds. The third most homogeneous cloud type is supercooled liquid cloud, while the remaining three types – liquid, opaque ice, and cirrus (types 3, 7, and 8) are completely homogeneous only about 40% of the time.

Next, the temporal variability of the incoming solar irradiance at the SURFRAD stations was analyzed as a function of cloud type. The 1-minute averaged irradiances for solar zenith angles greater than 85° were converted to transmittance (clear-sky index) using top of atmosphere incoming irradiances from an ephemeris code. The differences between consecutive samples were then computed. The fifteen transmittance differences occurring immediately before and after each satellite sampling time were associated with the corresponding cloud data, which was then sorted by cloud type. Statistics of the absolute values of the transmittance differences were then computed over the entire population of each cloud type at each SURFRAD site. These statistics included the full cumulative distribution functions as well as the magnitude of the differences at the 95th percentile of the distributions. The 95th percentile values are plotted in Figure 2-22. A consistent pattern emerged from this analysis, in which the magnitude at the 95th percentile was lowest for opaque ice clouds and highest for fog (thin water clouds), with cirrus, supercooled, and water clouds ranged from lower to higher values in between. A notable exception is the 95% point for fog at Desert Rock, NV. This may be due to actual differences between fog banks at this and other locations or difficulty identifying thin clouds over the bright desert surface.

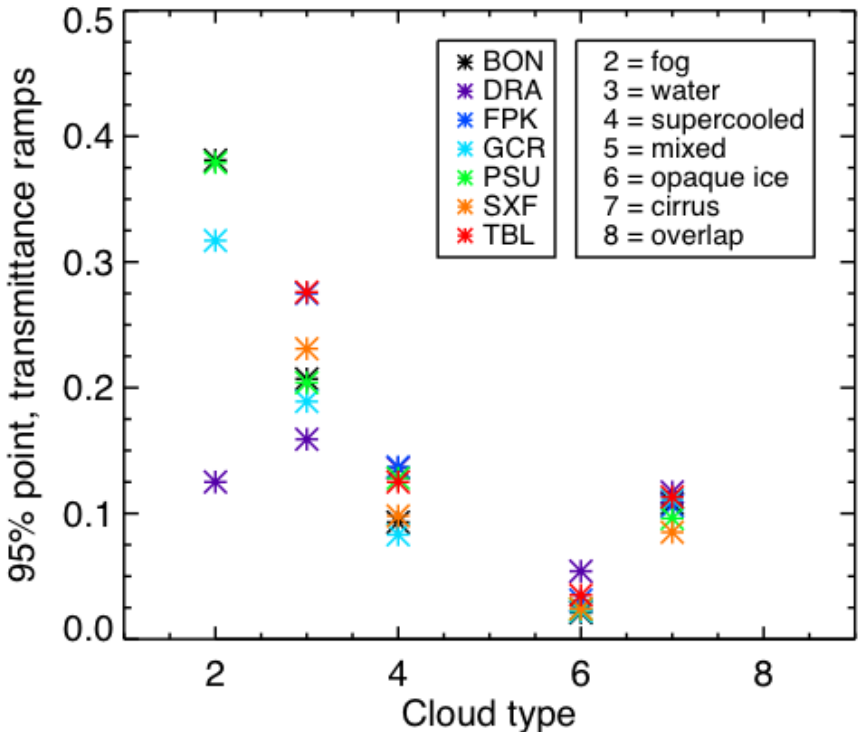


Figure 2-22. 95th percentile points in 1-minute average transmittance difference distributions as a function of location and cloud type.

Cloud type determination algorithms depend on the available measurements. For example, the instruments on the GOES-East and GOES-West satellites utilize different sensor wavelength bands. This means that relationships determined using GOES-East data, as in this study, may not be applicable to data from other satellites. We therefore continued this study with a focus on irradiance variability as a function of cloud properties. While cloud properties retrieved using data from different instruments may also differ, these values can at least be assessed using independent measurements and these differences, due to instrumentation differences, can be quantified. In addition, this approach could link to properties predictable by a numerical weather forecast model, thus providing a way to forecast rapid solar irradiance variability for the solar power industry. A finding that 95th percentile points were closely related to mean cloud transmittance suggested that this would be a fruitful avenue.

The cloud properties available from the satellite data sets included optical depth, liquid water path, cloud top height, and cloud fraction. We made no a priori assumption about how many parameters would be needed to specify the variability profile. Instead, we first compared the behavior of the different cloud properties. Since there is a clear relationship between optical depth and liquid water path given by, $\tau = 3/2\ LWP/(\rho\ r_e)$, where τ is the cloud optical depth, LWP is the liquid water path, ρ is the density of water, and r_e is the effective radius of the cloud droplet distribution, liquid water path was immediately eliminated from consideration.

Plots of the three remaining variables against each other are shown in Figure 2-23 for two of the SURFRAD stations, one in a humid mid-latitude site (Bondville, IL), the other at an arid site (Desert Rock, NV). From these plots and correlation coefficients, we see that cloud height and

fraction have a clear correspondence, but optical depth is relatively independent of both cloud fraction and height.

To be useful predictors of solar irradiance variability, the cloud properties not only need to be independent of each other but also correlated with the variations. In Figure 2-24 we plot the 95th percentile points for the populations of 1-minute averaged total hemispherical irradiances falling in a given cloud height, fraction, or optical depth bin. As seen earlier, cloud optical depth, which directly determines transmittance, has a clear, though not one-to-one, relationship with variability. Surprisingly, cloud fraction does not. We would expect broken clouds to lead to higher variability, but cloud fraction depends only on the total cloud amount, not brokenness. Another measure of brokenness should probably be sought. The surprising result of this analysis is that cloud top height exhibits a nearly linear relationship to the 95th percentile of an irradiance distribution function, indicating that it would be a good criterion from which to estimate irradiance variability.

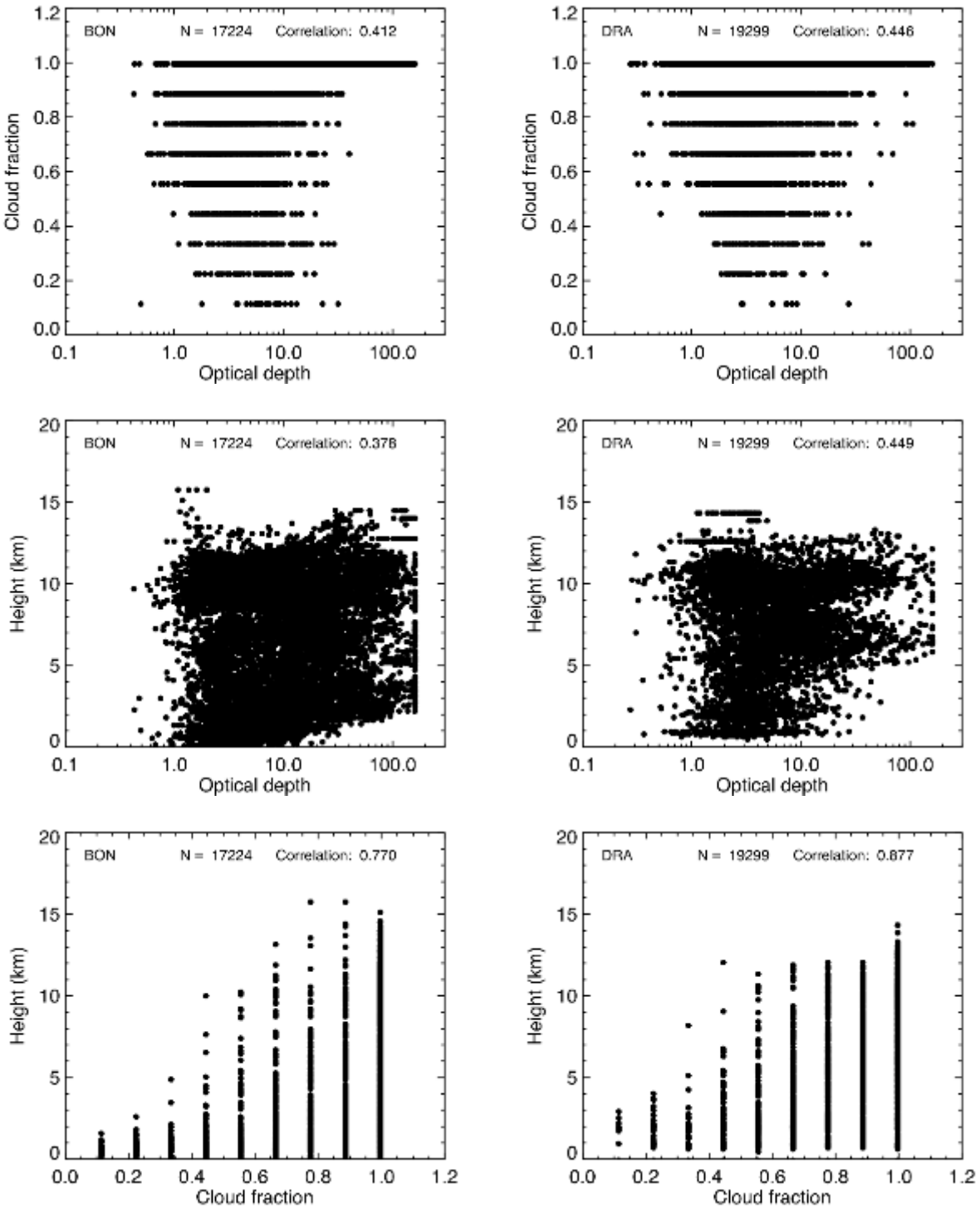


Figure 2-23. Relationships among cloud fraction, optical depth, and height at Bondville, IL (left), and Desert Rock, NV (right).

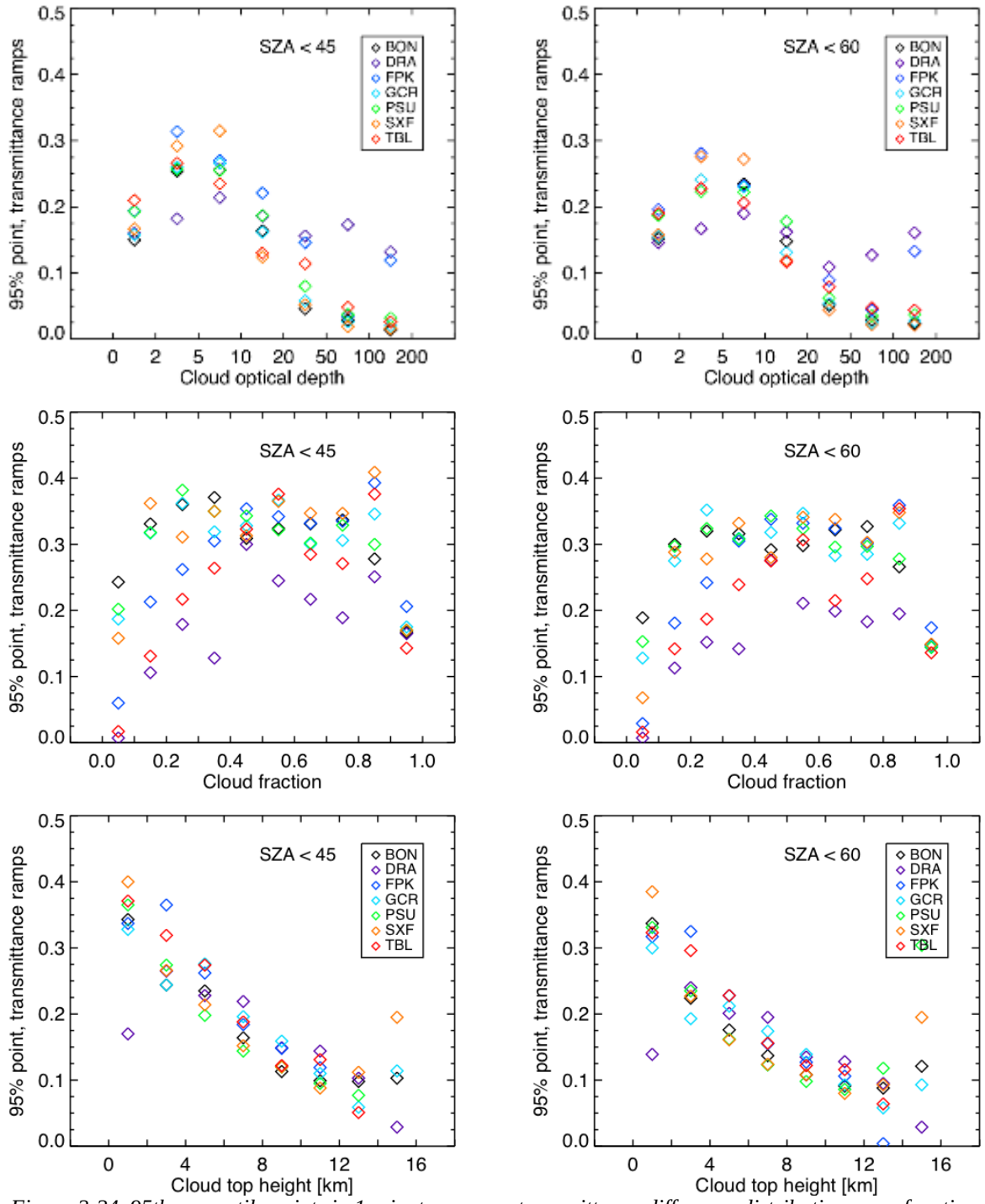


Figure 2-24. 95th percentile points in 1-minute average transmittance difference distributions as a function of location and cloud optical depth (top), fraction (middle), and height (bottom). Plots in the left (right) column include only times of day with solar zenith angles less than 45° (60°).

2.2.2.2 Solar Resource and Variability in the Northwest United States

The University of Washington additionally studied the variability of the solar resource in the Pacific Northwest. The following text is derived from the thesis of Nevin Schaeffer, a Whitman College student who worked with Laura Hinkelman at the University of Washington during the summer of 2015.

Studies have shown that with more accurate forecasting of the solar energy variability (Hoff et al. 2010) and geographically distributed PV array sites (Madrigal et al. 2013; Mills et al. 2010; Hoff et al. 2010), the costs of grid integration can be greatly minimized, especially in conjunction with systematic changes to grid operation practices, such as sub-hourly scheduling, using updated solar forecasts and diversity in the rest of the power generation sources tied to the same grid (Lew et al. 2010). Motivated by the need to understand the effect of clouds on solar power generation, to improve solar forecasting and facilitate the integration of solar energy into the national power grid, this study attempts to relate the weather and climate of five sites throughout the Northwestern United States (Table 2-7 and Figure 2-25) to site specific five minute global horizontal irradiance data and calculated clear-sky index. The five sites were chosen to be a representative spread of the region's climate zones.

Table 2-7. Distances between sites in miles, ranging from 156 to 520 miles of separation between site pairs.

From To Distance [miles]
Burns, OR Dillon, MT 336
Burns, OR Eugene, OR 204
Burns, OR Hermiston, OR 156
Burns, OR Twin Falls, ID 243
Dillon, MT Eugene, OR 520
Dillon, MT Hermiston, OR 325
Dillon, MT Twin Falls, ID 205
Eugene, OR Hermiston, OR 223
Eugene, OR Twin Falls, ID 446
Hermiston, OR Twin Falls, ID 330

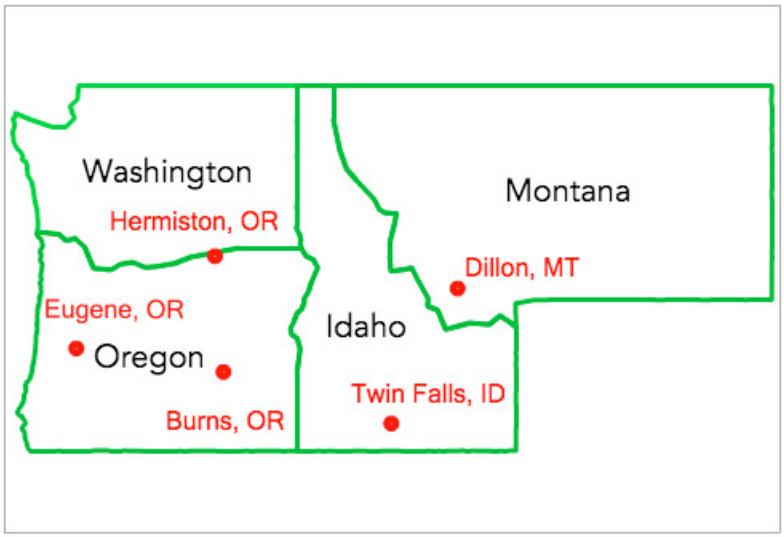


Figure 2-25. Map of the Northwestern United States showing the site locations for Burns, OR, Dillon, MT, Eugene, OR, Hermiston, OR, and Twin Falls, ID.

Five-minute resolution GHI time series were obtained from the University of Oregon Solar Radiation Monitoring Laboratory (UOSRML) (http://solardat.uoregon.edu/index.html) for five sites spread throughout the Northwest United States (see map in Figure 2-25 and distances between sites in Table 2-7). The sites were chosen based on their geographic position, grade of instrumentation, and completeness of data from the years 2004 to 2013, to ensure a variety of climate zones and reliable data. Deemed 'First-Class' stations by the UOSRML, the Burns, OR, Eugene, OR, Hermiston, OR, and Dillon, MT, sites are equipped with Eppley PSP pyranometers, Eppley NIP pyrheliometers, and Campbell Scientific CR-10 data loggers. To increase coverage of the region, the Twin Falls, ID, AgriMet site, equipped with a LI-COR pyranometer and Sutron data logger, was also analyzed. LI-COR pyranometers use a photodiode (solar cell) to measure the solar irradiance as a function of the color of the sky (blue if clear, gray if cloudy). Eppley PSP pyranometers use a multi-junction wire-wound Eppley thermopile that measures the solar irradiance as a function of radiant heat. On a day-to-day basis, the total solar irradiance measurements of a LI-COR and an Eppley PSP pyranometer are fairly similar, though measurements can differ based on the cloud cover on an hourly or shorter timescale.

As the focus of this study is the climatological effects on solar irradiance, the seasonal analysis in this study uses meteorological seasons instead of the typical astronomical seasons. Hence, winter spans December, January, and February, spring spans March, April, and May, summer spans June, July, and August, and fall spans September, October, and November.

To calculate the variability at each site, the values of the original time series were averaged every *n* rows, with *n* determined by the desired time interval, and the absolute value of the difference between adjacent values was calculated to represent the short-term variability. Time intervals of interest for this study include 5, 15, 30 and 60 minutes (with respective *n* values of 1, 3, 6, and 12). Variability statistics analyzed in this study include the mean and 95th percentile values. The 95th percentile values isolate the statistic from the extreme values and represent the value at which 95% of the data is less than or equal to that value. Two-way t-tests were computed in R using the *t.test()* function.

Looking at the short-term variability of the GHI time series at the 5-minute timescale, Figure 2-26 presents a comparison between sites and month of the year. The solar irradiance difference represents the magnitude that the PV array output ramped, either up or down, in a five-minute period as a monthly average. Greater magnitudes of the differences correspond to greater difficulty for the power grid operators to adjust to the short-term variability. One importance of Figure 2-26 lies in noting the sharp peak in solar irradiance difference at all five sites in the late spring (April (4) and May (5)). Similarly, in Figure 2-27, the estimated broken cloud cover percentages spike in the late spring as well. While this correlation cannot be quantified in this study due to the unavailability of high resolution cloud data, the visual mirroring of broken cloud cover and magnitude of solar irradiance differences in Figure 2-26 and Figure 2-27 persists as an interesting observation worthy of further research. Antón et al. (2011) similarly found the greatest variability during the month of April and concluded that short-term fluctuations increase as a function of cloud cover, except for completely overcast conditions where the short-term variability actually decreases slightly. Completely overcast conditions are much more uniform than partly cloudy conditions, and therefore it makes sense that short-term variability would decrease at such times

when the cloud cover over a solar array might be moving, but is not changing significantly in terms of light penetration.

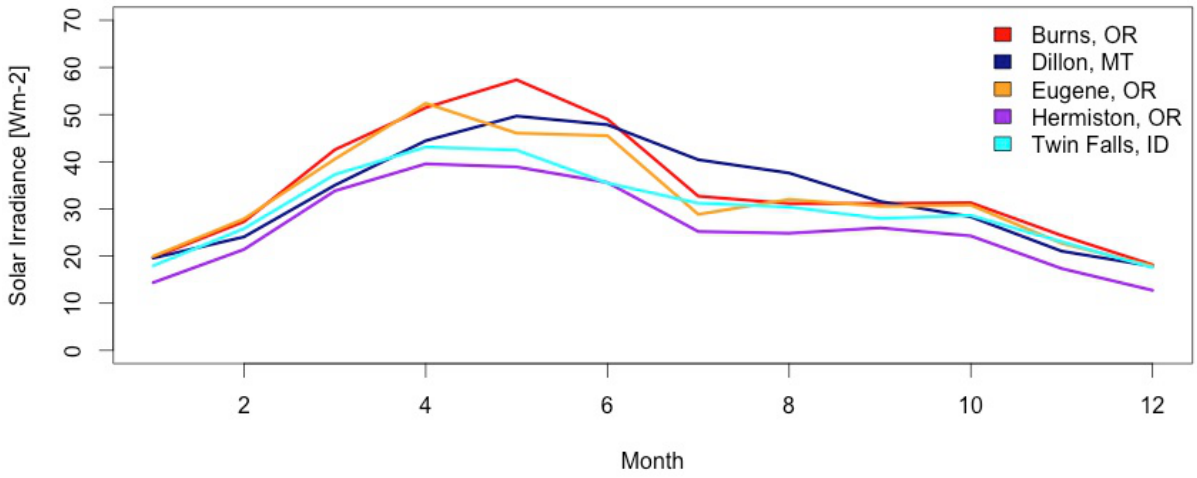


Figure 2-26. Monthly averaged five-minute variability by site (2004-2013 average).

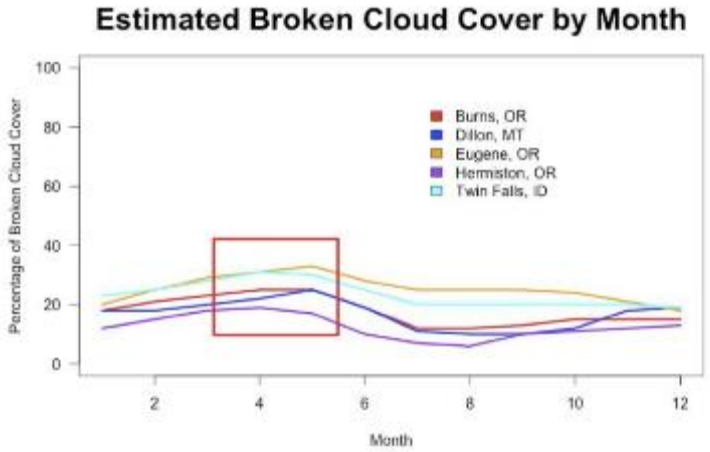


Figure 2-27. Estimated fraction of broken cloud cover (mostly cloudy, partly cloudy and mostly clear skies) within all sky conditions, by month for each site. Adapted from cloud cover plots for 30 years of averaged data (weatherspark.com). Note the peaking of broken cloud cover in the spring months for all sites (red box).

2.2.3 Introduction to StatCast

The specific forecast lead time, the type, quality, and flow of observations, and also the choice of forecasting method all impact short-range solar irradiance forecasting accuracy. There are multiple methods for short-term solar irradiance forecasting, including ground-based sky image advection techniques, satellite-based cloud advection models, and fast-running NWP models. Each of these methods, however, has strengths and limitations depending on the data sources and forecast lead time. In this study we utilize the strengths of multiple data sources by blending surface weather observations, irradiance observations, and satellite data via statistical learning algorithms into a product called StatCast. We focus on the forecast lead times of 15 minutes to three hours where

both the deterministic irradiance prediction and the forecasted variability of the irradiance are essential for utility companies and systems operators. We also show that a model tree statistical technique can predict both the temporal and the spatial irradiance variability more accurately than assuming climatology.

To achieve short-term solar irradiance predictions, we use surface weather observations and solar irradiance observations as inputs and predictors for a regime-dependent forecasting system. We also test the benefit of including GOES-East satellite data as an input and predictor. Several cloud regime-dependent short-range solar irradiance forecasting systems are also tested to make 15-minute average clearness index predictions for 15-, 60-, 120-, and 180-minute forecast lead-times.

Several versions of StatCast have been developed as part of the DOE SunShot project and of the Ph.D. dissertation of Tyler McCandless, who defended his dissertation entitled, "Artificial Intelligence Techniques for Short-Range Solar Irradiance Prediction," at The Pennsylvania State University in August 2015. StatCast actually forecasts clearness index, Kt, which is the ratio of the amount of irradiance (in this case, GHI) that reaches the earth's surface divided by the irradiance that impinges on the top of the atmosphere.

$$Kt = \frac{GHI\ Observed\ at\ the\ Surface}{GHI\ at\ the\ Top\ of\ the\ Atmosphere} \qquad (2.\ 7)$$

All of the versions of StatCast use a background solar irradiance model to compute the irradiance at the top of the atmosphere. They then predict the value of Kt and apply that to predict the irradiance at the ground. The four current incarnations of StatCast include:

1. StatCast-Persistence –The model assumes that Kt persists from the previous time step. This results in a model that recognizes the changes in solar angle, but assumes persistence of atmospheric constituents and clouds. This is sometimes called "smart persistence" and is used in much of the rest of the assessment as a baseline.
2. StatCast-Cubist – This version of StatCast uses the Cubist model regression tree to train on historical data, then predicts in real-time.
3. Regime-Dependent StatCast – RD-StatCast uses a k-means clustering method to separate instances into cloud regimes, then applies an artificial neural network (ANN) to each regime separately.
4. Regime-Dependent StatCast incorporating Satellite Data – Based on the third version, this most advanced version of StatCast also includes data from the GOES-East satellite to determine cloud state.

The inclusion of regime dependence was inspired by initial experiments with data that divided a time series into clear, cloudy, and partly cloudy days. The average behavior of those types of days is distinctly different, as shown in Figure 2-28. This observation led us to test developing specifically regime-dependent models, which has proven successful.

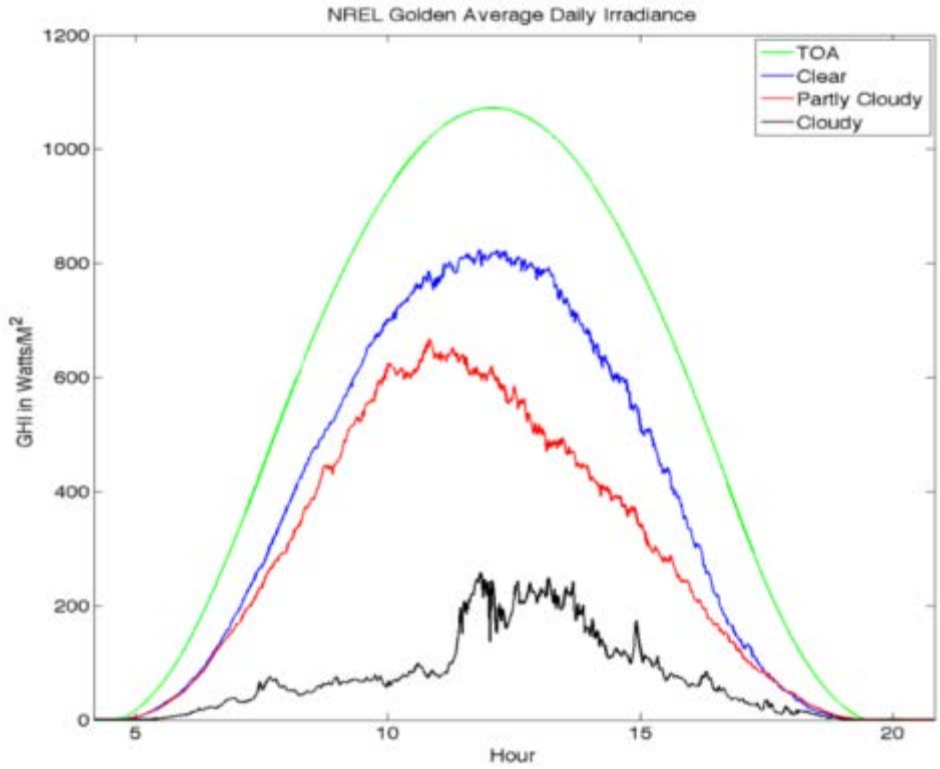


Figure 2-28. GHI as measured at the NREL Golden site for 1 year, with days separated into clear (blue), partly cloudy (red), and cloudy (black). The green line is the computed GHI at the top of the atmosphere.

2.2.4 State-of-the-Science of Short-Range Statistical Models

The optimal method for solar irradiance prediction depends on several factors, including the forecast lead time, with statistical techniques and cloud advection techniques most effective for short-range irradiance forecasting. Short-range forecasting is defined here as solar irradiance predictions from 15 to 180 minutes. Predicting solar power through statistical techniques has gained the attention of researchers in recent years. Sharma et al. (2011) found that a support vector machine approach to post-processing NWP model forecasts produced lower GHI forecast error compared to linear regression post-processing techniques. For intra-day irradiance forecasts, a combination of methods works best, including empirical models, satellite-based techniques, statistical methods, and NWP models (Lorenz et al. 2012; Kleissl 2013; Bouzerdoum et al. 2013; Voyant et al. 2013, 2014), with the combination producing the lowest forecast error depending on the specific lead time and available predictors. Hassanzadeh et al. (2010) and Yang et al. (2012) found that autoregressive integrated moving average (ARIMA) models produced lower solar irradiance and solar power errors compared to other time series short-range prediction techniques, while Morf et al. (2014) used a Markov process to predict sunshine and cloud cover. Mellit (2008) reported that artificial neural networks have been used in modeling and predicting solar radiation more than any other non-linear technique. More recently, several studies determined that models based on ANNs improve solar irradiance or solar power forecast accuracy compared to various baseline techniques (Martin et al. 2010; Hall et al. 2011; Marquez and Coimbra 2011; Wang et al. 2012; Chu et al. 2013; Cornaro et al. 2013). Several studies have examined the performance of these statistical forecast models in various weather conditions. Pedro and Coimbra (2012) found the accuracy of an ANN optimized with a genetic algorithm had a strong seasonal dependence.

Marquez et al. (2013b) correlated total sky images, infrared data, and solar radiation observations at the surface to use as input to an ANN and found the variability of solar radiation to be strongly dependent on the amount of cloud cover. Each day was classified as sunny, partly sunny, or cloudy, and an ANN was used to forecast the daily profile of the power produced by a PV plant (Mellit et al. 2014). Fernandez et al. (2014) concluded that the ANN model has lower errors for days characterized by direct irradiance (clear days) and for days characterized by diffuse irradiance (cloudy days) than for days characterized by a mix of direct and diffuse irradiance (partly cloudy days).

Statistical methods are well suited to combining multiple predictors in such blended forecast systems. Statistical models of appropriate complexity for the GHI forecast problem maximize the predictive value from the available predictors (e.g., satellite and ground-based observations). Any regression method can be applied to GHI forecasting, but the ANN is one of the most powerful, general, and therefore most widely used (e.g., Mellit 2008; Martin et al. 2010; Pedro and Coimbra 2012; Notton et al. 2012; Bhardwaj et al. 2013; Bouzerdoum et al. 2013; Diagne et al. 2013; Fu and Cheng 2013; Marquez et al. 2013b; Inman et al. 2013; Chu et al. 2013; Fernandez et al. 2014; Almonacid et al. 2014; Quesada-Ruiz et al. 2014). The relevant predictors for estimating direct normal irradiance (DNI) with a Bayesian ANN method were found to be the clearness index and the relative air mass in Lopez et al. (2005). Pedro and Coimbra (2012) found that an ANN time series model out-performed persistence, ARIMA, and k-Nearest Neighbors (kNN) models for 1-2 h solar power predictions. Marquez et al. (2013b) used processed satellite images as input into ANNs to predict GHI from 30 minutes to 120 minutes and found between 5% and 25% reduction in root mean square error (RMSE) compared to that of persistence. A challenge with ANNs, however, is the large number of tunable parameters, which is on the order of the number of predictors multiplied by number of neurons. This requires a large quantity of training data to prevent over-fitting and the consequent loss of skill on independent data (i.e., operational use). Another concern with using ANNs in operational forecasting is the lack of physical interpretability that could directly provide the user with forecast variability information.

At forecast lead-times of 15 minutes to three hours, satellite-based cloud advection techniques have traditionally been used. These techniques use cloud-motion vectors (CMVs) that are computed from consecutive satellite images and then used to advect the satellite observed clouds into the future. The use of CMVs for solar irradiance and solar power prediction was proposed by Beyer et al. (1996) with Hammer et al. (1999) and Lorenz et al. (2004) developing more advanced advection schemes. A forecasting method that uses a phase correlation between consecutive Meteosat-9 images has been used to predict 30-min cloud index values out to four hours lead time and on average showed 21% improvement in RMSE compared to cloud index persistence (Cros et al. 2014). Bilionis et al. (2014) extend the cloud advection technique to a probabilistic prediction by using principal component analysis (PCA) prior to applying a Gaussian process model. To address the errors due to assuming steady clouds during advection, Miller et al. (2014) group cloud pixels into cohesive cloud structures and then employ an appropriate steering flow that uses cloud group properties to forecast their downstream development and sheering characteristics. Their intermediate position in the lead-time spectrum makes satellite-based techniques prime candidates for blending with other forecast techniques.

We also note that separating conditions by weather or cloud regime shows promise for improving a forecast. The irradiance variability was shown to differ among satellite data derived cloud types

in Hinkelman (2014). Regime-based prediction has been used in several different solar irradiance and solar power applications. Tapakis and Charalambides (2013) provide a review of various methodologies for both supervised and unsupervised cloud classification. The unsupervised techniques classify based on the pixels of an image. The supervised techniques, which are divided into simple, statistical and artificial subgroups, classify based on available training datasets and arithmetic complexity of the technique. A one-step stochastic prediction process of cloud cover or clearness index with transition matrices dependent on the relative sunshine amount is presented in McCandless et al. (2014) and Morf (2014). Zagouras et al. (2013) used a k-means clustering algorithm with a stable initialization method to identify regimes based on step-changes of the average daily clear sky index in the region near San Diego, California. A simple approach based on the daily total solar irradiance identified clear, partly cloudy, and cloudy regimes with separate ANN models developed on each regime in Mellit et al. (2014) and showed that, particularly for the cloudy days, the ANN model trained on only those days improved on the ANN model trained on all days. McCandless et al. (2016a) used a k-means algorithm on surface weather and irradiance observations to identify regimes before applying an ANN. The separation into cloud regimes allows an artificial intelligence (AI) model to identify repeatable patterns in surface solar irradiance; however, there is a lack of research into 1) what are the most important inputs for cloud regime classification and 2) what are the most important predictors for an AI method to most efficiently make accurate short-range predictions of solar irradiance.

Finally, variability, both in space and time, is important to quantify for full application of short-range solar power forecasts. The quantification of temporal solar irradiance variability caused by the advection and dynamic evolution of clouds has been recently studied. Mills et al. (2009) showed a passing cloud at a point produces solar insolation variation exceeding 60% of peak insolation in a matter of seconds. As mentioned above, Hinkelman (2007) found that not only are the irradiances themselves larger in the middle of the day but also the fractional change in irradiance from one time to another is larger. Hinkelman (2013) also determined that cloud optical depth and cloud height are the best predictors of irradiance variability at one-minute time resolution. Kuszmaul et al. (2010) analyzed 1-sec PV output data and showed that it is linearly proportional to the spatial average of irradiance. Reikard (2009) examined data at resolutions of 5, 15, 30, and 60 minutes and found that the "data exhibits nonlinear variability, due to variations in weather and cloud cover." These studies have examined the variability of measured solar irradiance due to changes in cloud cover.

In addition to temporal variability, several studies have examined the spatial variability of solar irradiance. Zagouras et al. (2014) used cluster analysis for the determination of coherent zones of GHI for utility scale territory in California, and used step-changes of the daily average clear sky index at each location to characterize the fluctuation of GHI. Gueymard and Wilcox (2011) analyzed solar power's regional dependence and showed greater variability tends to occur in coastal areas, particularly the California coast, and mountainous areas because of the micro-climate effects of topography. Rayl et al. (2013) performed an irradiance co-spectrum analysis and concluded that solar power site aggregation could greatly reduce power variability on short time scales depending on the distance between sites.

The goal of our study is to use observed meteorological data together with a network of irradiance observing sites to better predict solar irradiance on the 15-min to 3-h timeframe as well as its temporal and spatial variability, both measured in terms of a standard deviation from the mean

value of the GHI. The focus is on short-term predictions, which as Nguyen and Kleissl (2014) state, "intra-hour solar forecasting for power production and ramp events has become an important need in the solar industry as the inevitable variability of solar power will have a greater impact on energy resource management as solar penetration increases."

2.2.5 Research Approaches to StatCast

Here we describe in detail the formulation of each of the versions of StatCast. In the following section, we detail the data used in the training and follow with specific results. The primary site of application for StatCast is the SMUD irradiance measurement sites.

2.2.5.1 StatCast – Persistence

We use clearness index persistence as our baseline prediction technique for comparison. Clearness index persistence is commonly referred to as "smart persistence." It inherently corrects for changes in solar elevation with time and can be easily converted back to GHI for operations if the clearness index forecast is multiplied by the top-of-atmosphere GHI.

This baseline technique uses the last available 15-min average observation of the clearness index as the prediction for subsequent times. For locations with either generally clear conditions or steady cloud cover, this technique is difficult to beat. In contrast, when the sky condition is characterized by mixed or variable clouds, the clearness index persistence technique performs poorly.

2.2.5.2 StatCast – Cubist

The artificial intelligence technique used for this version of StatCast is the model tree, or Cubist model, which is Quinlan's (1992) M5 model tree formatted as a set of rules (Kuhn et al. 2012). The model tree uses a "separate-and-conquer" algorithm to search for a rule that explains part of the training instances, separates these instances, and continues this process until no instances remain (Quinlan 1993). The algorithm reformulates the tree into a set of rules and places a multivariate linear model at each leaf in order to predict our continuous predictands of solar irradiance variability. See Quinlan (1987a, 1987b, and 1992) for a detailed explanation of this process. The final prediction is a weighted average of the multivariate linear regression equations at each node in the tree down to the final leaf (Kuhn et al. 2012). This weighted averaging is accomplished by a smoothing process that adjusts the predicted value from the leaf up to the root via,

$$PV(S) = \frac{n\ x\ PV(Si) + k\ x\ M(S)}{n+k} \quad , \qquad (2.8)$$

where *n* is the number of instances, *i*, in the node *S, PV(Si)* is the predicted value at node *S* and instance *i, Si, k* is a smoothing factor set equal to 15, and *M(S)* is the model prediction at the leaf of the sub-tree. This smoothing is done to capture the skill in the predicted values at nodes along the tree down to the final leaf. Figure 2-29 displays an example description of the model tree with the red branch highlighting the sub-tree used in this example prediction. Thus, this model tree is a set of rules that are paths from the top to bottom of the tree with each node's multivariate linear model output used in the final prediction.

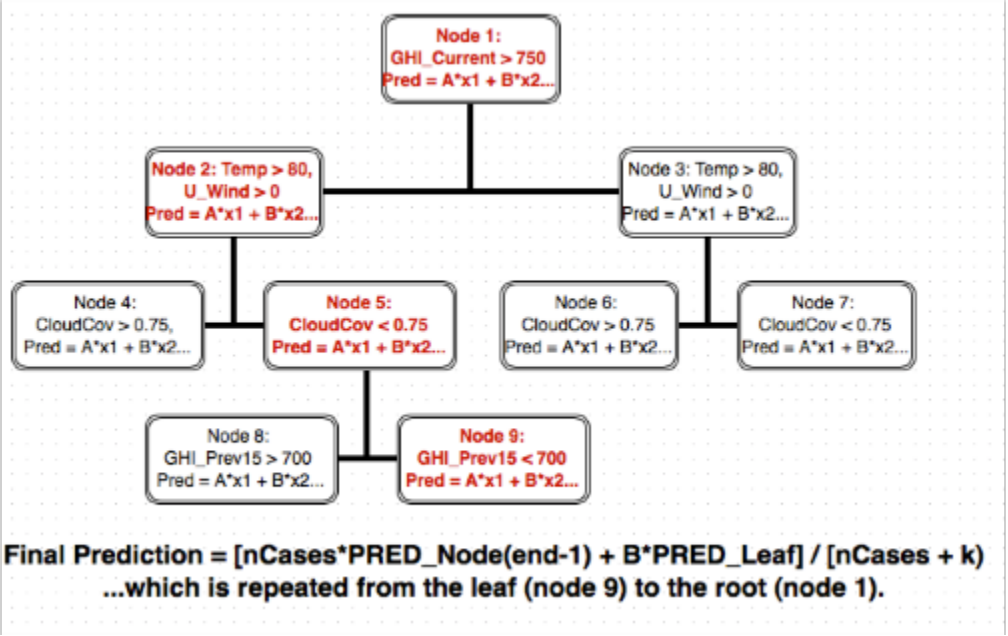


Figure 2-29. Flowchart and description of the model tree. For this arbitrary instance, the sub-tree in red is used to make the final prediction via the equation at the bottom.

We use an additional model option that combines the model tree's prediction with a prediction given the training dataset's nearest neighbor to further reduce the model tree's error. The nearest neighbor option first finds the training cases that are most similar to the current instance. Then the model tree is used to make predictions for all of the nearest neighbor instances and the current instance. The value of the current instance prediction is adjusted based on the difference between the current instance prediction and the prediction for the nearest neighbor instances. A sensitivity study, shown in Table 2-8, indicates that optimal configuration of the model tree includes one nearest neighbor, which results in the lowest mean absolute error for all four forecast lead times tested.

Table 2-8. Sensitivity study for the number of nearest neighbors used in the model tree prediction. The values shown are the MAEs of the model predicting GHI temporal standard deviation (W m^{-2}). Using one nearest neighbor results in the lowest MAE for all lead times.

ALL DATA - NUMBER OF NEAREST NEIGHBORS - GHI					
Lead Time	0	1	2	5	9
15	35.04	26.86	29.04	31.98	33.45
60	49.33	35.75	38.69	43.25	45.66
120	58.91	42.31	45.44	50.45	53.50
180	59.81	44.60	47.51	52.38	52.38

2.2.5.3 Regime-Dependent StatCast

We chose to develop a cloud regime-dependent short-range solar irradiance forecast system not only to improve the deterministic forecast accuracy, but also to provide a quantification of the expected solar irradiance variability and corresponding forecast uncertainty. This section outlines

our classify-then-predict process; the details are described in the following sections. Our methodology begins by classifying the cloud regime with the k-means algorithm. We then train a separate ANN to make predictions for each individual regime as depicted in Figure 2-30. This novel work goes beyond Mellit et al. (2014), Marquez and Coimbra (2011), and others in the sophistication and automation in identifying regimes with the k-means algorithm and in the regime-dependent configuration of the ANNs that are specific to improving the final prediction algorithm. The process begins by selecting the optimal set of inputs for cloud regime classification that corresponds to the final model with the lowest forecast error. The selected set of inputs is then used by the k-means algorithm to classify and partition the datasets into an optimal number of cloud regime subsets. Finally, ANNs are constructed on each of the cloud regime datasets independently. This classify-then-predict process (with k-means then ANN) is repeated for each forecast lead-time.

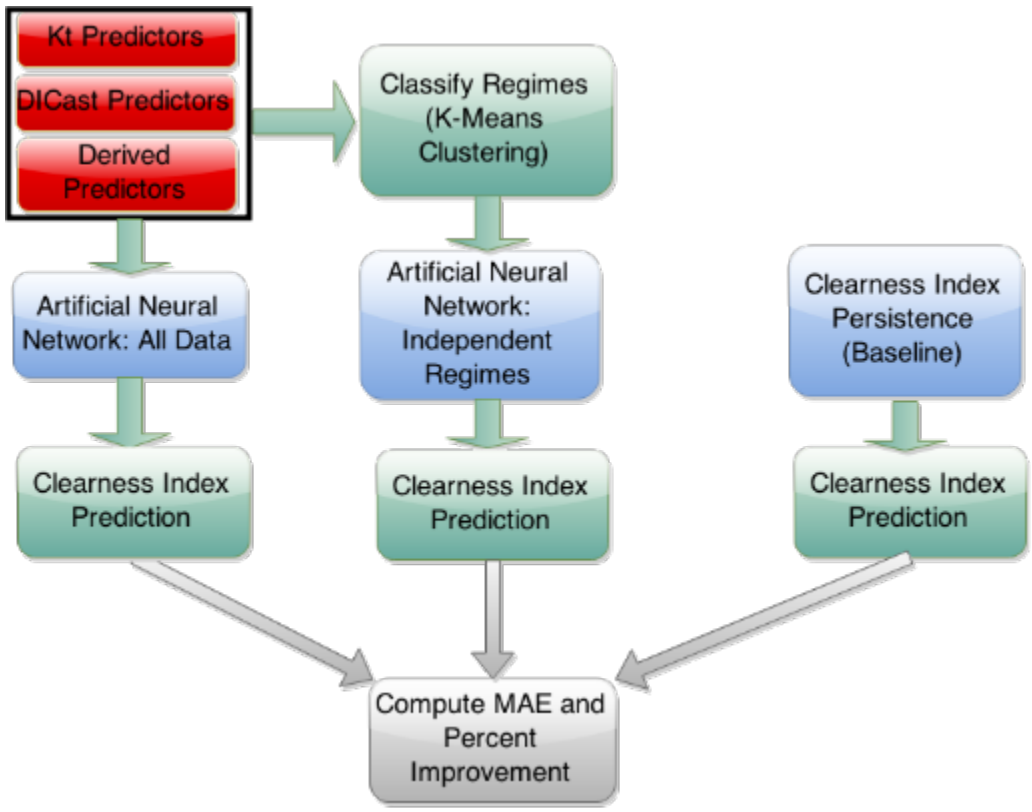


Figure 2-30. Process design: first, classify cloud regimes on the optimal set of potential inputs shown in the red rectangles outlined in the black box, and then apply ANN models to predict the clearness index on each regime independently. An ANN is also applied on all data (i.e., without regime identification), and compared to the clearness index persistence prediction.

A cloud regime-dependent AI system requires dividing the cases into distinct regimes for which the fundamental relationship between predictors and predictands is expected to differ, and therefore, to allow more accurate short-range forecasts. Thus, careful sensitivity studies determined the optimal configurations of the AI models in order to match the complexity of the relationships among the predictors in the regimes. After all data are quality controlled and additional variables are derived, the datasets are randomly split two-thirds for training and one-third for testing. All of the results shown are from the testing datasets; however, the sensitivity

tests conducted to determine the optimal configurations of the system were performed on the training datasets. The ANN and k-means sensitivity studies similarly split the training dataset into two-thirds for training and one-third for testing and the optimal configuration was determined based on this one-third independent test set. This approach avoids compromising the independence of the initial test dataset. We show results in this study for four forecast lead times: 15 min, 60 min, 120 min, and 180 min. These predictions are for the 15-minaverage clearness index ending at each lead time.

In a real-time forecasting environment, predictions are made in a three-step process. In the preprocessing step, the data are collected, quality controlled, and the additional predictors are computed. In the regime classification step, the trained k-means algorithm is applied to the current data. Then, the ANN trained for the currently classified cloud regime and forecast lead-time is used to predict the solar irradiance (clearness index). The ANNs use numerical weather prediction analysis data and irradiance observations as input to predict clearness index at multiple locations in the vicinity of Sacramento, California.

2.2.5.4 Regime-Dependent StatCast Incorporating Satellite Data

In an effort to develop a "best practices" method for regime-dependent statistical forecasting of clearness index, we tested multiple regime-dependent prediction methods for solar irradiance prediction given various inputs and predictors; therefore, we use a dataflow diagram (Figure 2-31) to describe the relationships between the various techniques. The top tier represents the data sources: irradiance observations, Meteorological Aviation Reporting (METAR) surface weather observations, derived predictors, and satellite data, which are split into two boxes for the measured and the derived variables. The GOES-East satellite derived variables are included only in the instances that are not defined as clear. The second tier illustrates this separation into the satellite-determined clear instances and satellite-determined cloudy instances. This is the first regime separation in our prediction process. The third tier of Figure 2-31 describes the prediction methods for all other instances. From left to right, the first prediction technique is the ANN applied on the clear dataset. The next prediction technique is an ANN without additional regime classification. The final three are the Regime-Dependent ANNs, which are hereafter given the name RD-ANN. The first RD-ANN method is based on regimes determined explicitly from the "cloud type" variable in the GOES-East data, which is labeled RD-ANN-GCT where GCT stands for GOES Cloud Type. The next RD-ANN technique is the k-means cloud regime classification that includes inputs from all of our data sources, which we name RD-ANN-GKtCC because it includes GOES-East data, Kt observations, and cloud cover from the METAR observations. The final prediction technique does not include the satellite measurements and is a direct comparison to previous work (McCandless et al. 2016a). This method is named RD-ANN-KtCC because it includes the Kt observations and the cloud cover. The fourth tier elements are the final predictions from all of the prediction techniques, including the baseline technique of clearness index persistence.

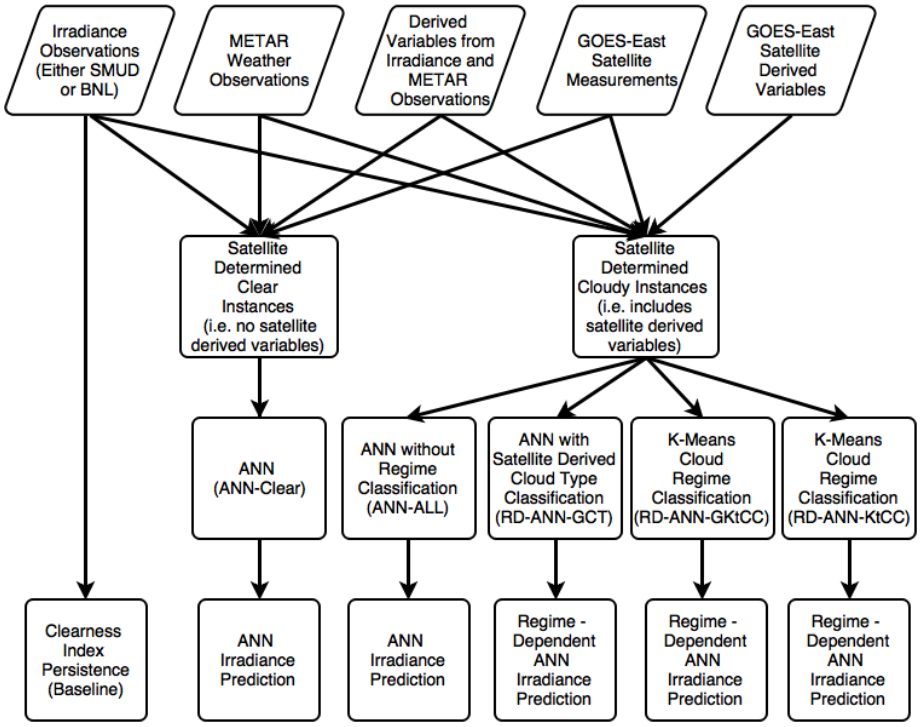


Figure 2-31. Overall process design for our regime dependent prediction technique and the comparison techniques.

2.2.6 Data for StatCast Predictions

2.2.6.1 Irradiance Data

The network of irradiance observing sites used in this study is that of the SMUD in California, which is an area of varying terrain covering approximately 900 square miles (Bartholomy et al. 2014). In order to evaluate the prediction techniques, this study follows the same procedure as the planned real-time operational implementation. We use data from eight solar power forecast sites that measure irradiance, shown in Figure 2-32 as blue triangles. The GHI observations are available for a period of 367 days from 25 January 2014 through 26 January 2015. The temporal resolution of the raw data is one minute and averages are computed over 15-minute intervals ending at :00, :15, :30, and :45 minutes for each hour. The 15-minute averaged GHI data are then converted to clearness index values. This averaging interval was selected after communication with several utility companies and agrees with the shortest time range for which a forecast is currently useful for dispatch decision-making.

Therefore, our prediction techniques ingest four predictors from the irradiance data, each converted to Kt: the average Kt from 60-45 minutes, from 45-30 minutes, from 30-15 minutes, and the past 15 minutes prior to forecast initialization time, which is the start of every hour. Hereafter, these predictors are named Kt_Prev60, Kt_Prev45, Kt_Prev30, and Kt_Prev15.

The solar irradiance data from all eight SMUD pyranometers is aggregated and all instances with missing data or nighttime observations are excluded from the final dataset. There are a total of 71,184 instances in the final dataset.

Some limited comparisons are also shown below for the Long Island Solar Farm located on Brookhaven National Laboratory's property. Details of that setup can be found in McCandless et al. (2016b).

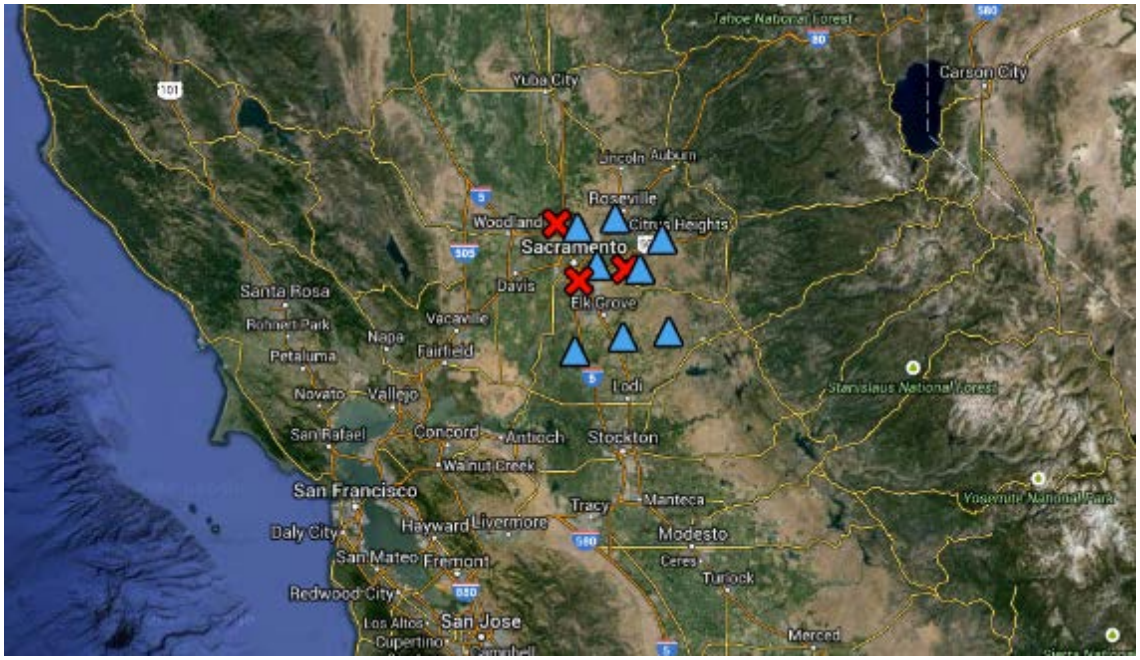


Figure 2-32. Map of the SMUD pyranometers (blue triangles) and METAR/DICast predictor sites (red Xs).

Surface weather observations are not available at the irradiance observation sites; therefore, the three nearest METAR sites are used to characterize the local weather, which represents hourly surface weather observations from stations typically located at airports across the United States. The METAR observations are quality controlled by NCAR for ingest to the Dynamic Integrated foreCast (DICast®) System (Mahoney et al. 2012). The closest METAR sites to the SMUD pyranometers are the three locations plotted as red X's in Figure 2-32. We use six weather variables: cloud cover, dewpoint temperature, categorical precipitation in the last hour (1 = precipitation occurred, 0 = precipitation did not occur), precipitation amount, temperature, and wind speed.

2.2.6.2 Satellite Data

The satellite data used as forecast predictors came from NOAA's GOES-East satellite. The GOES data were chosen for this work because they are acquired operationally every 15 minutes with a nominal nadir footprint of just 1 km in the shortwave and 4 km in the infrared channels. GOES-East was selected over GOES-West for two reasons. First, the position of GOES-East at 75°W provides views of both the California and New York forecast sites at less oblique angles than the 135°W location of GOES-West. Second, processed GOES imager data were only available from the GOES-East acquisitions at 0:15 and 0:45 after the hour and from GOES-West acquisitions at 0:00 and 0:30 after the hour. Allowing for a latency time of 15 min, the 0:45 acquisition provides the most up-to-date information for the re-initialization of our forecast system at the top of every hour.

The GOES-East data consists of both directly measured and retrieved variables provided in level-2 output from the Pathfinders Atmosphere-Extended (PATMOS-x) retrieval suite (Heidinger et al., 2013), which is run operationally by NOAA's Cooperative Institute for Meteorological Satellite Studies (CIMSS) and, for this project, by the Cooperative Institute for Research in the Atmosphere (CIRA). The directly measured variables are radiance values at wavelength bands centered on 650 nm (visible) and 3.75 μm (infrared) and brightness temperatures at 3.75 μm and 11.0 μm (water vapor window). The retrieved variables applied in this study were cloud top temperature, cloud fraction, cloud optical depth, hydrometeor effective radius, and cloud type, where the cloud types included the categories fog, liquid water clouds, supercooled water clouds, opaque ice clouds, cirrus clouds, vertically overlapping clouds, and overshooting clouds. Instantaneous solar zenith angles were also taken from the satellite data files. The data are provided as un-gridded 4-km footprints. The values supplied to the forecast system are averages over the nine footprints closest to each of the forecast locations at 0:45 after each hour.

2.2.6.3 Derived Variables

In addition to the observed irradiance and weather predictors, it is often useful to derive additional variables in order to emphasize important physical processes. Based on our previous work (McCandless et al. 2016a), we derive inputs specific to the k-means classification system as well predictors specific to the ANN prediction system. In particular, we leverage our meteorological knowledge to provide the k-means algorithm with inputs in order to identify cloud regimes and to provide the ANNs with predictors for predicting solar irradiance. Based on that previous work (McCandless et al. 2016a), variables used as inputs for the k-means algorithm include the cloud cover squared, averaged over the three nearest METAR sites, and the standard deviation of the cloud cover for the three nearest METAR sites, so as to weight higher regional cloud cover values and to quantify the regional solar irradiance variability. Another predictor, dewpoint depression, defined as the difference between the temperature and the dewpoint temperature, quantifies the nearness to saturation of the atmosphere at the surface. This derived predictor, and the cloud cover squared predictor, are averaged over the three METAR sites based on a sensitivity study that showed no improvement by including the predictor for each site independently. For the SMUD region, we derive two additional predictors by computing the spatial average and standard deviation of the clearness index at the previous 15-min interval over the remaining sites. These predictors are computed to quantify the regional distribution of cloud cover as measured by the eight solar irradiance observation sites.

2.2.7 Research Results for StatCast

We compare the various methods of prediction in a systematic manner. We first describe the specific implementation of each method compared, including applying an ANN without regime dependence.

2.2.7.1 Baseline: StatCast-Persistence

We use clearness index persistence ("smart persistence") as the baseline prediction technique for comparison. It inherently corrects for changes in solar elevation with time and can be easily converted back to GHI for operations if the clearness index forecast is multiplied by the top-of-atmosphere GHI. This baseline technique uses the last available observation of the clearness index

(i.e., 15-min average) as the prediction for subsequent times. For locations with either generally clear conditions or steady cloud cover, this technique is difficult to improve upon. In contrast, when mixed or variable clouds are observed, the clearness index persistence technique performs poorly.

2.2.7.2 Artificial Neural Network

The ANN is our choice for a nonlinear AI prediction technique because an ANN does not require *a priori* knowledge of potentially complex relationships between the predictors and the predictand. ANNs replicate how the human learning process works, and when given a sufficiently large set of training data, ANNs can model complex, nonlinear, relationships between the predictors and the predictand (Lippmann 1987). The ANN used here is a feed-forward neural network trained by a back-propagation algorithm (Reed 1998), which is commonly referred to as a multi-layer perceptron (Rosenblatt 1958). The specific neural network module used in this study is the newff model in the Neurolab python library (https://pythonhosted.org/neurolab/). The ANN used here has three layers: the input layer that consists of the predictors, the hidden layer that consists of tunable neurons, and the output layer that computes the final prediction. The actual processing is done by the neurons in the hidden layer, each of which is a linear regression that is post-processed by a nonlinear sigmoid function so that all outputs are on a common finite scale. These neuron outputs are then merged by a final linear regression neuron to yield the ANN's forecast. Each predictor of the input layer is connected to all neurons within the hidden layer, but the iterative training results in special weights for each neuron that together solve the different aspects of the problem.

Varying the number of neurons in the hidden layer changes the complexity of the model. As more neurons are added, more complex nonlinear relationships between the predictors and the predictand can be modeled. This increase in complexity, however, increases the risk of over-fitting the training data and decreasing the performance of the model on the independent data. Moreover, as the number of training epochs (iterations) is increased, an overly complex ANN may begin to tune to the random noise in the training data as well as to the real relationships. Therefore, both the number of neurons of the hidden layer and the number of training epochs determine the ANN's fit to the training and independent data. The goal of configuring the ANN is to find the best level of complexity, i.e., the number of hidden layer neurons, and the number of training epochs that model the true relationships in the training data and thus yield the lowest error on independent data. We held the learning rate (0.01) and weight decay (0.5) constant, as sensitivity studies (not shown) found these values to be best.

We have a total of 42 predictors for the SMUD sites, which includes data from SMUD irradiance observation sites, METAR weather observation sites, GOES-East satellite data, and several derived predictors. A list of all predictors for the ANN is provided in Table 2-9. For the BNL locations, the predictors, "Kt Nearby Mean" and "Kt Nearby Variability (Stdev)" are not available because, unlike SMUD, the BNL data come from a single location.

Table 2-9. List of predictors for the ANN model. The Kt Nearby Mean and Variability are marked with an asterisk because they are only available for the SMUD sites.

Solar Zenith Angle	**Satellite Measured Reflectance at 3.75um**	**Dewpoint at METAR Site 2**	**Categorical Precipitation at Site 3**	**Kt Previous 60-Min**	**Kt Temporal Variability (Stdev)**
Satellite Derived Cloud Type	Satellite Measured Temperature at 11.0um	Dewpoint at METAR Site 3	QPF at METAR Site 1	Kt Previous 45-Min	Most recent Kt Change (Kt Prev15 - Kt Prev30)
Satellite Derived Cloud Fraction	Satellite Measured Temperature at 3.75um	Cloud Cover at METAR Site 1	QPF at METAR Site 2	Kt Previous 30-Min	Kt Nearby Mean*
Satellite Derived Cloud Top Temperature	Temperature at METAR Site 1	Cloud Cover at METAR Site 2	QPF at METAR Site 3	Kt Previous 15-Min	Kt Nearby Variability (Stdev)*
Satellite Derived Cloud Optical Depth	Temperature at METAR Site 2	Cloud Cover at METAR Site 3	Wind Speed at METAR Site 1	Sine of the Julian Day	Cloud Cover Variability (Stdev)
Satellite Derived Hydrometeor Radius	Temperature at METAR Site 3	Categorical Precipitation at METAR Site 1	Wind Speed at METAR Site 2	Cosine of the Julian Day	Cloud Cover Squared
Satellite Measured Reflectance at 650nm	Dewpoint at METAR Site 1	Categorical Precipitation at METAR Site 2	Wind Speed at METAR Site 3	Dewpoint Depression (METAR Sites Average)	Kt Slope

2.2.7.3 Regime-Dependent Artificial Neural Network Applications

The goal of applying the ANN is to find the true relationship between the predictors and the predictand; therefore, we partition the dataset into cloud regime subsets in order to allow the ANN to find the simpler relationships applicable to each cloud regime rather than having to model both these relationships and regime identification with a single complex network. In order to improve the deterministic forecast, the regime identification technique must split regimes with different underlying forecast problems, each with different physical, and thus, statistical relationships, between predictors and predictand. Therefore, the regime classification method must capture differences that are directly related to short term irradiance forecasting, given the predictors available.

Two regime-identification methods, which are named after the input data, RD-ANN-KtCC and RD-ANN-GKtCC, use a k-means clustering algorithm. The k-means clustering algorithm is explained in detail in McCandless et al. (2016a). For the RD-ANN-KtCC method, the inputs to the k-means clustering algorithm are the past irradiance (converted to Kt) observations and cloud cover observations from the METAR data. This method is tested to determine the predictive skill of an RD-ANN method using only surface observations.

For the RD-ANN-GKtCC method, the inputs to the k-means clustering algorithm are the past irradiance (converted to Kt) observations, cloud cover observations from the METAR data and variables from the GOES-East data. This method is tested to determine the predictive skill of an RD-ANN method using both surface observations and satellite data. In contrast, the RD-ANN-GCT method, does not use the k-means algorithm to classify regimes, but rather uses the derived "cloud type" variable in the GOES-East data to separate regimes. This test will determine if off-the-shelf cloud typing can compete with mission-specific cloud regime typing in solar forecasting.

RD-ANN-KtCC

The first regime-dependent method tested uses the original configuration of the regime-dependent ANN of McCandless et al. (2016a), hereafter referred to as RD-ANN-KtCC. This technique does not include any GOES-East data as either inputs to the k-means regime classification or as predictors for the ANN. Sensitivity studies in McCandless et al. (2016a), showed that the best inputs to the k-means clustering algorithm are the following: Kt average in the previous 15 minutes, nearby Kt in the previous 15 minutes, standard deviation of the Kt in the previous 15 minutes among the nearby sites, the most recent change in the Kt (Kt previous 15 min – Kt previous 30 min), the slope of the Kt in the past hour, the standard deviation of the Kt over the previous hour, and standard deviation of the cloud cover. Because there are seven inputs into the k-means algorithm, there are therefore seven dimensions in the phase space of the k-means distance computation. These seven inputs provide the k-means algorithm with information that capture the meteorological state based on surface observations. Sensitivity studies indicate that the optimal number of regimes, k, was also seven. For the BNL site, only a single irradiance observation site was available; therefore, the RD-ANN-KtCC method does not include either the nearby Kt in the previous 15 minutes or the standard deviation of the Kt in the previous 15 minutes among the nearby sites.

RD-ANN-GKtCC

The RD-ANN-GKtCC method uses 16 inputs into the k-means clustering algorithm for the SMUD sites, which are shown in Table 2-10. Again, the multi-site inputs are unavailable for BNL; thus, the RD-ANN-GKtCC method does not include either the nearby Kt in the previous 15 minutes or the standard deviation of the Kt in the previous 15 minutes among the nearby sites. Because there are 16 inputs into the k-means algorithm, there are 16 dimensions in the phase space of the k-means distance computation. These 16 inputs provide the k-means algorithm with information to capture the meteorological state given both surface irradiance and weather observations, as well as satellite-based data with careful consideration given to avoiding co-linearity. The inputs include all inputs used in RD-ANN-KtCC as well as additional variables from the GOES-East observations: cloud fraction, cloud top height, cloud optical depth, hydrometeor radius, reflectance at 6.5 um (i.e. wavelength for shortwave IR), reflectance at 3.75 um (i.e. wavelength for water vapor), temperature at 6.5 um and temperature at 3.75 um.

Table 2-10. List of inputs for the k-means algorithm in the RDANN-GKtCC configuration. The Kt Nearby Mean and Variability are marked with an asterisk because they are only available for the SMUD sites.

Satellite Derived Cloud Fraction	Satellite Measured Reflectance at 650nm	Kt Previous 15-Min	Kt Nearby Variability (Stdev)*
Satellite Derived Cloud Top Temperature	Satellite Measured Brightness Temperature at 11.0um	Kt Temporal Variability (Stdev)	Cloud Cover Variability (Stdev)
Satellite Derived Cloud Optical Depth	Satellite Measured Reflectance at 3.75um	Most recent Kt Change (Kt Prev15 - Kt Prev30)	Cloud Cover Squared
Satellite Derived Hydrometeor Radius	Satellite Measured Brightness Temperature at 3.75um	Kt Nearby Mean*	Kt Slope

In order to match the level of complexity of the ANN with the number of training cases and complexity of relationships within each regime, we perform multiple sensitivity studies to determine the best number of training epochs and the best number of hidden layer neurons. We examine the mean absolute error (MAE) of the RD-ANN-GKtCC method on the sensitivity test cases for each lead-time. The MAE is calculated as

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |(obs(i) - pred(i)| \qquad (2.9)$$

where *n* is the number of instances in the testing data. We varied the number of training epochs (100, 250, 500 or 1000) and averaged the error over the regimes. The test was conducted separately for each lead time with the result for 180 minutes. The results indicate that the lowest error on the sensitivity test cases, and thus the best number of training epochs for the ANN is 500. The same result (not shown) was obtained for the other lead times.

After the sensitivity study determined the number of training epochs, the next step in configuring the RD-ANN-GKtCC model was to determine the best number of neurons and the best number of regimes for each forecast lead time and forecast location. We performed a sensitivity study with 5, 10, 15, and 20 neurons in the hidden layer and k ranging from two to nine for each forecast lead time. The best combinations (in terms of the lowest MAE on the sensitivity test datasets) are shown in Table 2-11. For the SMUD sites, the best k is two for the two shorter lead-times and three for the two longer lead-times. For the BNL location, the best k is two for all forecast lead times. The best number of neurons varies among the different locations and lead times; however, the results showed relatively minor differences between different numbers of neurons, which indicates that the increase in forecast power nearly balances the increase in over-fitting for a range of model complexities around the best configuration.

Table 2-11. Best number of regimes, K, and number of neurons in the hidden layer for all forecast lead-times at both SMUD and BNL as determined by the lowest error on the sensitivity test set.

	SMUD		**BNL**	
	K	Nodes	K	Nodes
15-Min	2	5	2	10
60-Min	2	15	2	15
120-Min	3	20	2	5
180-Min	3	15	2	10

RD-ANN-GCT

The third method of regime-dependent prediction uses the "cloud type" variable in the GOES-East data to determine regimes; therefore, this technique is named RD-ANN-GCT. Each cloud type has a separate ANN trained for that cloud type. There are seven cloud types present in the data: fog, liquid water clouds, supercooled water clouds, opaque ice clouds, cirrus clouds, overlapping clouds and overshooting clouds, in addition to the cases identified as clear due to the absence of derived satellite variables.

2.2.8 Results for StatCast

2.2.8.1 Comparison of Persistence, Cubist, and Regime Dependent Models

A first basic comparison considers clearness index persistence and two nonlinear prediction techniques: an artificial neural network and the Cubist model regression tree. In order to quantify the predictive skill in different cloud regimes before applying the k-means clustering technique, the data were divided into clear days (clearness index greater than 0.6) and cloudy days (clearness index less than 0.6). A ten-fold cross validation was performed to analyze the prediction techniques. The mean absolute error for clear days at all forecast lead times is shown in Figure 2-33 while the mean absolute error for cloudy days at all forecast lead times is shown in Figure 2-34. The results for clear days show the errors from the Cubist model are lower than the ANN or the clearness index persistence. Similarly, for the cloudy days, the Cubist model has the lowest results at all forecast lead times. The MAE for the Cubist model on the clear days ranges from 25 W m^{-2} at 15-minute lead time to 50 W m^{-2} at 180-minute lead time while the errors for the cloudy days are higher with a range from about 50 W m^{-2} at 15-minute lead time to 75 W m^{-2} at 180-minute lead time.

The baseline short-term forecasting technique is the clearness index persistence; thus, the percent improvement quantifies the improvement in forecasting error with the model tree compared to clearness index persistence. These results are plotted in Figure 2-35, with cloudy days displayed in blue and clear days displayed in red. The percent improvement for cloudy days ranged from about 10% to over 50% as lead time increases from 15 minutes to 180 minutes. This can be explained by the fact that the clearness index persistence is difficult to improve upon when the cloud cover is steady. The results for the clear days indicate the impressive predictive skill of the model tree compared to the clearness index persistence, with percent improvement between 25% and 55%.

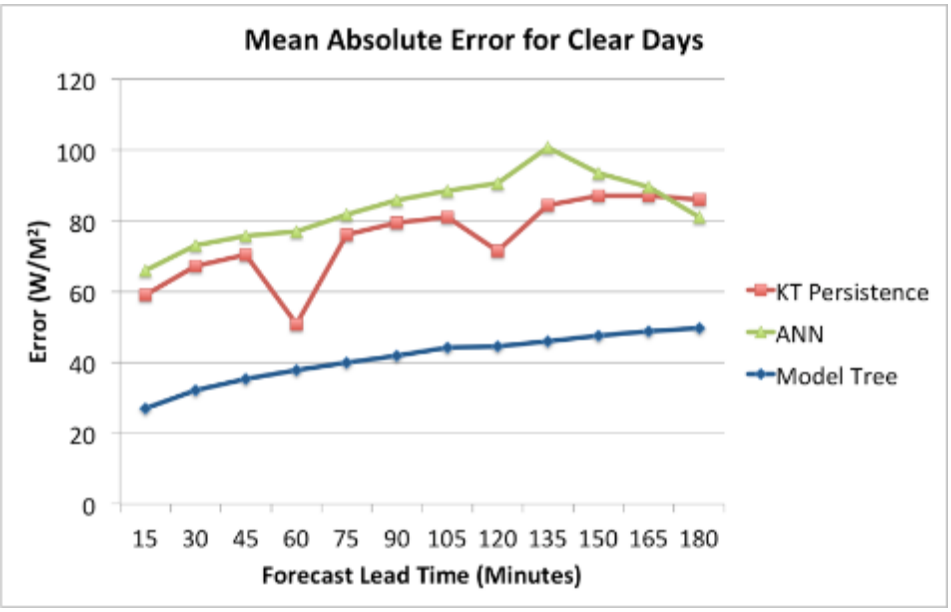


Figure 2-33. MAE for the KT persistence, ANN, and Cubist model regression tree for all forecast lead times on clear days. The Cubist regression tree has the lowest errors for all lead times.

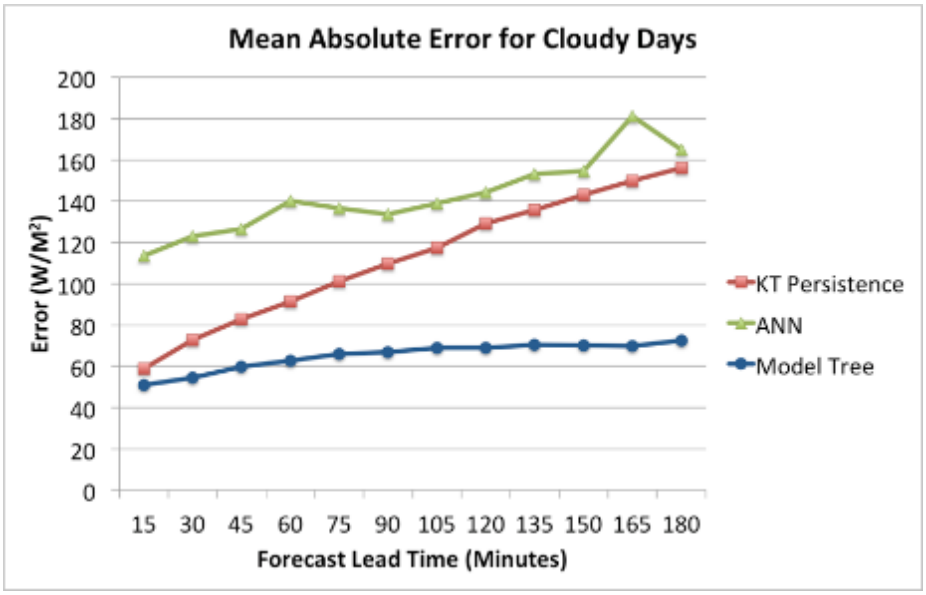


Figure 2-34. MAE for the KT persistence, ANN, and Cubist model regression tree for all forecast lead times on clear days. The Cubist model regression tree has the lowest errors for all lead times.

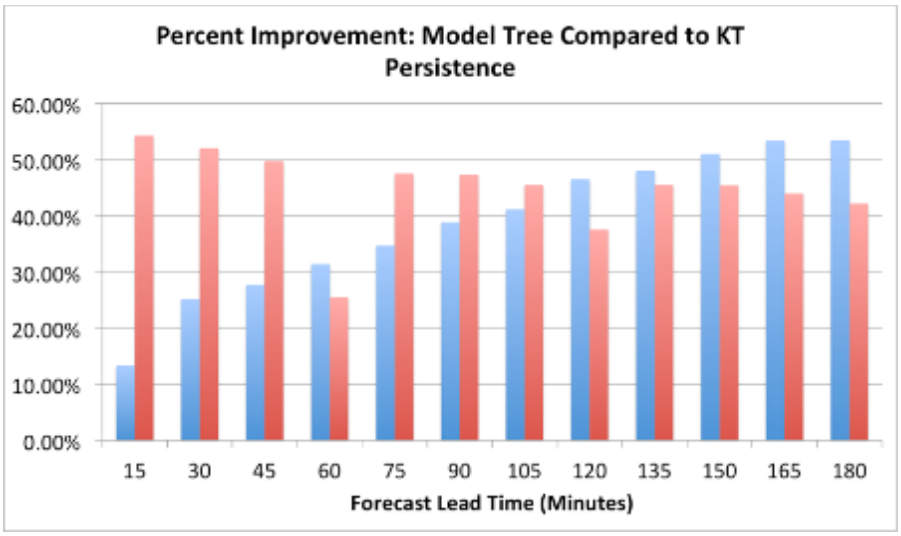


Figure 2-35. Percent improvement for the model tree's errors compared to the clearness index persistence. The results are shown for cloudy days (blue) and for clear days (red) at all forecast lead times.

2.2.8.2 Comparison of Regime Specific Methods with Baselines

We now compare the techniques that explicitly split the regimes with the baselines of StatCast-Persistence and an ANN trained without explicit regime separation. The data are initially split based on whether there are derived data in the GOES-East observations. Derived data are only available when the measured temperature and reflectance data indicate clouds are present. If an instance is identified as clear based on the GOES-East data, then an ANN trained on only those cases is used to predict the clearness index. Otherwise, the RD-ANN models and an ANN without regime identification are used to predict the clearness index.

Sacramento Sites

The results for the GOES-East defined clear cases are shown in Table 2-12 for all forecast lead times for the SMUD location. They indicate that the ANN improves upon the clearness index persistence method at the 60-, 120-, and 180-min forecast lead times. At the 15-min forecast lead-time, however, the error is nearly double that of the clearness index persistence forecast and this is likely a case of over-fitting the training data. At this forecast lead time, the magnitude of the irradiance is relatively consistent unless a cloud advects or develops over the observation site. Because these instances are rare when GOES-East data determines it to be clear, the ANN likely over-fits those uncommon cases and thus hurts the overall performance of the model.

Table 2-12. Comparison of MAE for the clearness index persistence and the ANN, CLEAR model for all forecast lead-times for the SMUD site.

	Kt Persistence	**ANN - Clear**
15-min	0.017	0.035
60-min	0.036	0.028
120-min	0.055	0.041
180-Min	0.082	0.057

Next, all of the RD-ANN methods were compared to both the ANN without regime identification (ANN-ALL) and the clearness index persistence for all the cases labeled other than clear by the GOES-East data. These MAE results are plotted in Figure 2-36 for all forecast lead times. As expected, the forecast error increases as the forecast lead time increases. The only method that generally performs worse than clearness index persistence is the RD-ANN-GCT method that uses the GOES-East derived cloud types as the regime classification method. At the 15-min lead time, the RD-ANN-KtCC, RD-ANN-GKtCC, ANN-ALL, and clearness index persistence all show similar errors. However, at the 60-minute and longer lead times, the RD-ANN-KtCC, RD-ANN-GKtCC, and ANN-ALL all show improvement over the clearness index persistence, as shown by the larger MAE of the clearness index forecasts. The method that generally performs best is RD-ANN-GKtCC method, which exploits the GOES-East data in both the k-means clustering and ANN.

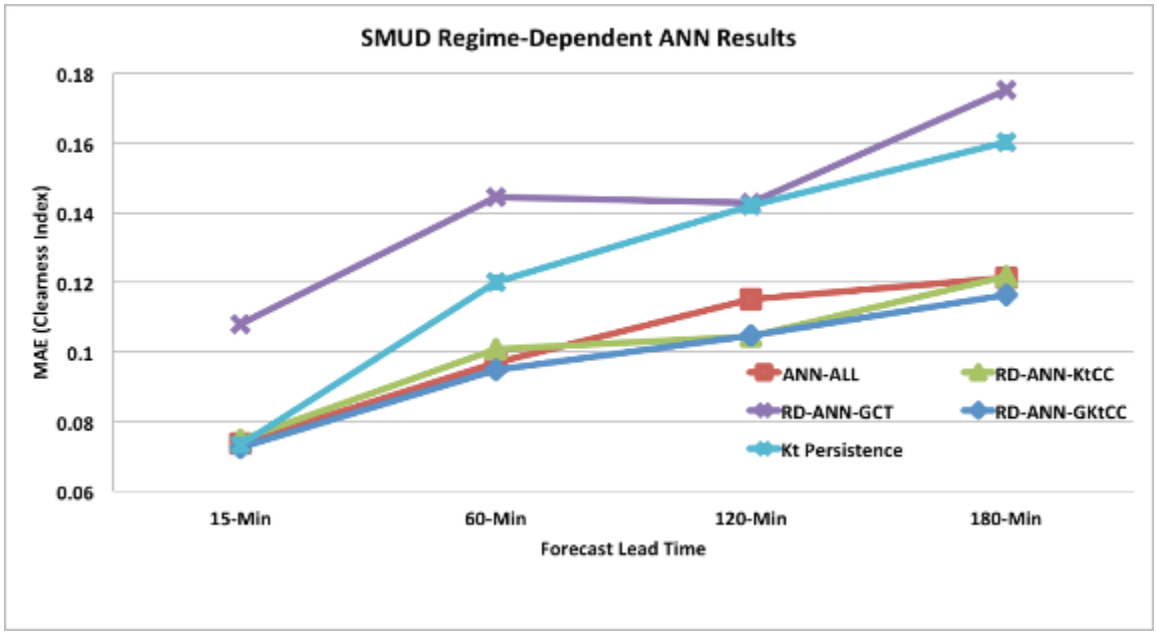


Figure 2-36. MAE as a function of lead time for all methods of the satellite determined cloudy instances for the SMUD site. The method that performs best in the majority of the forecast lead times is the RD-ANN-GKtCC method.

To quantify the forecast skill improvement with the regime-dependent methods, we compute the percent improvement over our baseline clearness index persistence technique. The percent improvement over clearness index persistence for the forecasts at the SMUD sites is shown in Figure 2-37. At the 15-min lead time, all of the methods closely mimic clearness index persistence, except for the RD-ANN-GCT method. At this lead-time only the RD-ANN-GKtCC method improves over the clearness index persistence, by 1%. In contrast, at the 60-, 120-, and 180-min lead-times, most of the RD-ANN methods show between 10% and 28% improvement over the clearness index persistence method. The RD-ANN-GOES model shows the worst performance except at the 180-minute lead-time when it begins to improve over the clearness index persistence. This poor performance is likely due to the cloud type classification because there are likely cases of misclassification by the GOES East system, and also there are cloud regimes with small data subset sizes, and thus, potentially ANN over-fitting on those regimes. At the 60-, 120-, and 180-min lead times, the RD-ANN-GKtCC method shows 21.0%, 26.4%, and 27.4% improvement over the clearness index persistence. The RD-ANN-GKtCC method is best at all lead times except at 120 minutes where the RD-ANN-KtCC produces a slightly better 26.6% improvement over clearness index persistence. These results demonstrate that the RD-ANN methods are able to improve substantially over clearness index persistence at 60-, 120- and 180-min lead times; however, the cloud regime classification makes a considerable impact on the overall performance of the models.

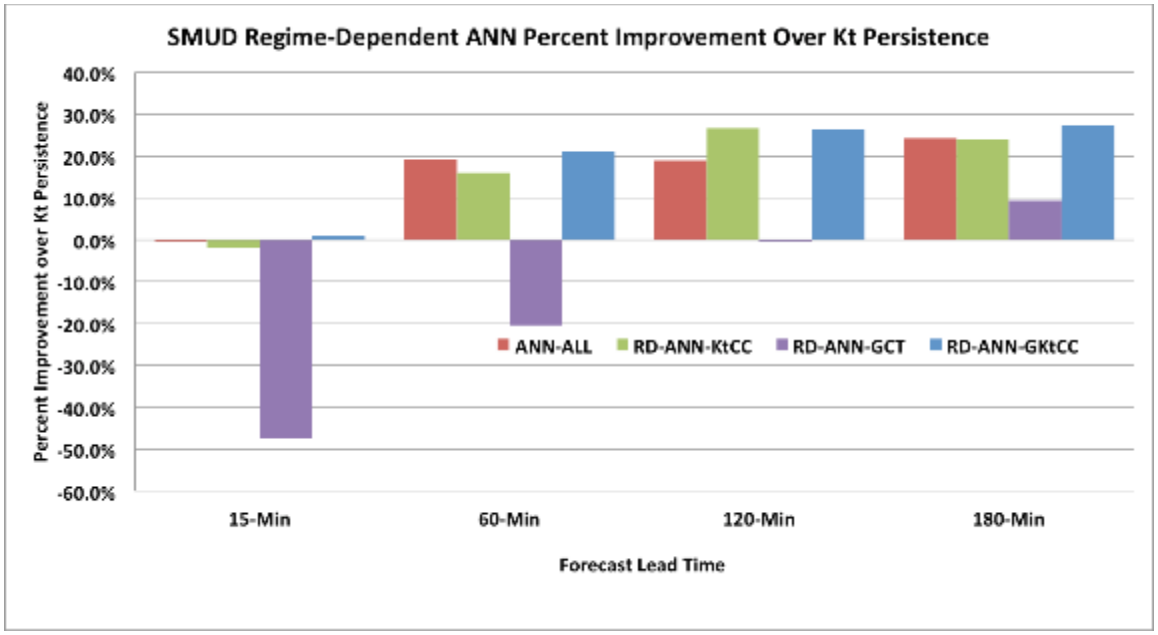


Figure 2-37. Percent improvement over the clearness index persistence forecasts for all methods on the satellite determined cloudy instances.

Long Island Solar Farm

While the SMUD dataset provides a substantial amount of data for training, sensitivity testing and independent verification, it is important to analyze how our complex regime-dependent model performs when trained with a smaller dataset. Doing so quantifies the value of obtaining larger, and thus more expensive, training datasets. In addition to redeveloping the same RD-ANN

methods using the BNL dataset, we also trained the RD-ANN-GKtCC model on the SMUD dataset and applied it to the BNL dataset (RD-ANN-SMUD) in order to determine how a general model trained at one site performs at a different site. The MAE for each method on the BNL test data is shown in Figure 2-38 for all forecast lead-times. These results indicate that the clearness index persistence method has lower error than all ANN methods for BNL. The results also indicate that, similar to the results for the SMUD sites, the RD-ANN-GCT model is the worst-performing model. At the 15-min and 60-min lead times, the best regime-dependent model is the method trained at SMUD. This highlights the importance of numerous and applicable training data, especially considering that the geostationary satellite data are distorted in different ways for locations in California versus New York, negatively impacting the forecast performance of a model trained at one location and applied to the other. The amount of data available from BNL to train the models at that site is likely too little given the number of predictors and the model complexity. With 40 predictors provided to the ANN, it may be too complex to avoid over-fitting given a training dataset of a maximum (if no regime classification is done) of 309 instances. Future work will examine how to properly down-select to the appropriate number of predictors and model complexity in order to capture the true predictive relationships among the predictors in a limited dataset.

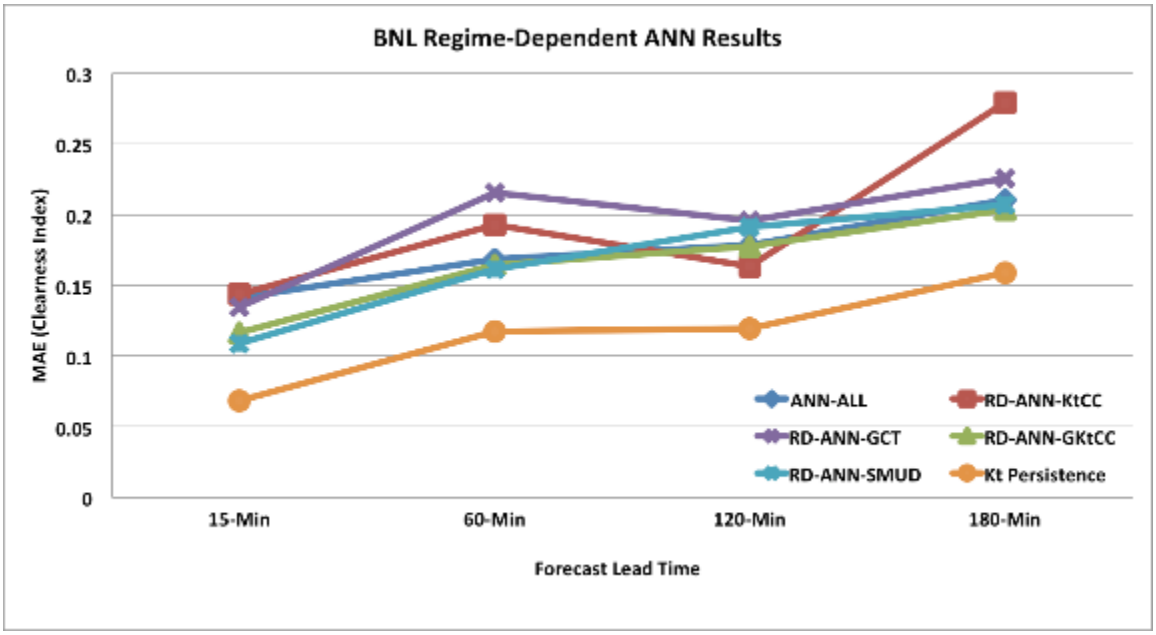


Figure 2-38. Results for all methods on the satellite determined cloudy instances for the BNL forecast site. The method that performs best in the majority of the forecast lead times is the clearness index persistence method.

Variability Prediction

Although the deterministic forecast skill is of primary interest to utility companies and systems operators, it is also valuable to predict irradiance variability (Bartholomy 2014; McCandless et al. 2015). Variability is important because the utility companies and systems operators need to allocate adequate resources to deal with variations that cannot be deterministically predicted. Here, we compute the irradiance variability as the standard deviation of the clearness index over the following three hours (i.e., the standard deviation of twelve 15-min average clearness index

values). We test the variability prediction for SMUD because the deterministic prediction results showed that the dataset has ample data for training and testing. As our baseline forecast, we compute the standard deviation of the 15-minute average clearness index values over the prior hour. Essentially, this clearness index persistence forecast predicts that variability will remain the same for the following three hours. We test this baseline technique versus an ANN trained without regime identification, and against a new version of the RD-ANN-GKtCC method that uses the same inputs and predictors as the deterministic irradiance forecast methodology, but is now trained to predict the three-hour clearness index variability. The results for the variability prediction displayed in Table 2-13 reveal that the lowest MAE comes from the RD-ANN-GKtCC prediction method. The RD-ANN-GKtCC method shows 18.6% improvement over the clearness index persistence forecast of the expected irradiance variability. The clearness index persistence, ANN-ALL and RD-ANN-GKtCC methods all show substantially lower errors than the average value of the clearness index variability, which was computed to be 0.092 for the test dataset.

Table 2-13. List of the MAEs for predicting the clearness index variability with the clearness index persistence, ANN-ALL, and RD-ANN-GKtCC methods trained to predict the variability for the SMUD sites.

	MAE	**Percent Improvement**
Kt Persistence	0.068	N/A
ANN-All	0.059	13.7%
RD-ANN-GKtCC	0.058	18.6%

2.2.9 Advances Due to this Project to Short-Range Statistical Forecasting

This work has shown that using multiple inputs into statistical techniques improves deterministic solar irradiance predictions as well as spatial and temporal solar irradiance variability predictions. This work sought to improve two major facets of short-range solar irradiance forecasting via regime-dependent statistical forecasting: deterministic irradiance forecast accuracy and irradiance variability estimates. We first classify cloud regimes with a k-means algorithm and then apply ANNs to each regime to produce a more accurate GHI forecast with variability estimates. The k-means algorithm statistically classifies the cloud regime based on surface weather and irradiance observations. This approach parallels that of Greybush et al. (2008), who classified weather regimes with principal component analysis (PCA) in order to apply regime-dependent optimal weights to ensemble temperature forecasts. After k-means clustering, ANNs are implemented for each weather regime independently with the intention of modeling the inherent predictability of each weather regime, and thus, the different causal relationships for each regime between predictors and predictand.

We utilize surface weather observations, solar irradiance observations, and GOES-East satellite data as inputs and predictors into regime-dependent techniques that first identify cloud regimes before fitting an ANN to predict clearness index. This approach allows each ANN to focus on the forecast mission for a specific cloud type. We find that a k-means cluster-based ANN method (RD-ANN GKtCC) improves upon the forecasting performance of not only the baseline clearness index persistence, but also improves upon the forecasting performance of a global ANN for lead-times of 60, 120, and 180 min. At the 15-min forecast lead time, all RD-ANN methods mimicked the

clearness index persistence, with the RD-ANN-GKtCC method managing to show a 1% gain in forecasting performance over clearness index persistence.

The RD-ANN methods not only showed improved performance for deterministic clearness index predictions, but also for predicting clearness index variability. A new version of the RD-ANN-GKtCC model trained to predict the variability of the clearness index over the next three hours showed substantial forecast error reduction compared to either using a variability persistence method or a global ANN. Thus, the RD-ANN-GKtCC model is able to improve the prediction of the deterministic irradiance and its variability for short-range lead-times, given sufficient training data.

Although the RD-ANN methods show substantial performance gain for the Sacramento, CA, (SMUD) sites that had a large training dataset, when the RD-ANN methods were trained to predict for a site on Long Island, NY, (BNL) with its small training dataset, the complex models did not perform well on the independent test dataset. In order to improve the forecasting methods at a site with a small amount of training data, the RD-ANN methods likely need be tuned with a smaller predictor set and a simpler configuration to allow the method to model the true predictive relationships among the predictors. The true predictive relationships in a small dataset are likely limited; therefore, future work can examine automatic ways of configuring RD-ANN systems depending on the amount of training data and number of available predictors. A simpler configuration with fewer predictors could potentially avoid the problem of over-fitting datasets too small (i.e. BNL) for using nonlinear models. Where insufficient data are available to apply the regime-dependent methods, the Cubist model regression tree performs quite well and is broadly applicable.

Of the three RD-ANN methods tested, the method that performed worst used a regime classification based on the cloud type derived variable in the GOES-East data. This outcome was likely due to a combination of multiple problems and so yields several ideas for future work. First, the GOES-East algorithm derives cloud types based only on the satellite measured values. Our ANN models are also provided predictors from surface weather observations and surface irradiance observations. Therefore, the RD-ANN methods that use a combination of the available data are more likely to capture clusters that represent real predictive relationships, which the ANN is able to model. The forecast error dependence on available predictors could be examined in future work by testing the forecasting skill of the RD-ANNs if the regime classification versions are the same, but the ANNs are only provided the GOES-East measured variables. Lastly, some of the cloud types are uncommon in the data, resulting in small training data subsets, and thus, giving the ANN model a higher likelihood of over-fitting the available training data.

Although the complex RD-ANN models have shown impressive forecast improvements for the SMUD sites, the clearness index persistence method still performs best when the dataset is too small to effectively train an ANN. Future work will look to quantify the amount of data required for the RD-ANN-GKtCC method to outperform a persistence-based approach.

2.2.10 Summary of Contributions to Short-Range Statistical Forecasting

This work has advanced the science of short-range statistical prediction of solar irradiance and its variability. Through understanding the variability by accomplishing statistical analysis, we

identified the efficacy of separating data into cloud regimes to build AI forecast models for each regime separately. The cubist model accomplishes this intrinsically and showed success at both improving the forecast beyond a smart persistence and producing a prediction of variability. The regime dependent versions of StatCast explicitly predict separately for each regime. We find added value by identifying the regime when sufficient data exists. With a limited amount of data, however, the value of regime identification decreases. As shown in chapter 5, StatCast is often the best performing component of the system on the short timeframes.

2.2.11 Publications from this work include:

Thesis

McCandless, T.C., 2015: Artificial Intelligence Techniques for Short-Range Solar Irradiance Prediction, Dissertation in Meteorology, The Pennsylvania State University.

Journal Papers

McCandless, T.C., G.S. Young, S.E. Haupt, and L.M Hinkelman, 2016b: Regime-Dependent Short-Range Solar Irradiance Forecasting, submitted to *Journal of Applied Meteorology and Climatology*, in revision.

McCandless, T.C., S.E. Haupt, and G.S. Young, 2016a: A Regime-Dependent Artificial Neural Network Technique for Short-Range Solar Irradiance Forecasting, *Applied Energy*, **89**, 351-359.

McCandless, T.C., S.E. Haupt, and G.S. Young, 2015: A Model Tree Approach to Forecasting Solar Irradiance Variability, *Solar Energy*, **120**, 514-524. DOI:10.1016/j.solener.2015.07.0200038-092X

Conference and Workshop Presentations (presenter in Bold)

Gagne, D.J., Haupt, S.E., Linden, S., Williams, J.K., McGovern, A., Wiener, G., Lee, J.A., and T.C. McCandless, 2015: Scaling Machine Learning Models to Produce High Resolution Gridded Solar Power Forecasts. *13th Conference on Artificial Intelligence: The Last Mile: Methods and Technologies for Delivering Custom Weather, Water, and Climate Information to Everyone in the World*, Phoenix, AZ, Amer. Meteor. Soc. T.J.1.1

Hinkelman, L. M., N. Schaeffer, and T. P. Ackerman, 2015: The character and variability of solar irradiance across the Pacific Northwest American Geophysical Union Fall Meeting, San Francisco, CA.

Hinkelman, L. M., 2014: Statistics of solar resource variability on short time scales, A Public-Private-Academic Partnership to Advance Solar Power Forecasting Project Workshop, Boulder, CO.

McCandless, T.C., **S.E. Haupt**, and G.S. Young, 2016: A Regime-Dependent Neural Network Approach to Short-Range Solar Irradiance Prediction Using Surface Observations and Satellite Data, Joint Session between 14th Conference on Artificial and Computational Intelligence and its Applications to the Environmental Sciences and Seventh Conference on Weather, Climate, Water, and the New Energy Economy, AMS Annual Meeting, New Orleans, LA, Jan. 12.

McCandless, T.C., Haupt, S.E., Young, G.S., 2015: A Bayesian Approach to Statistical Short-Term Solar Irradiance Forecasting. International Conference on Energy and Meteorology, Boulder, CO, June 23.

McCandless, T.C., S.E. Haupt, G.S. Young, and A.J. Annunzio, 2015: A Regime-Dependent Bayesian Approach to Short-Term Solar Irradiance Forecasts, Joint Session between Sixth Conference on

Weather, Climate, and the New Energy Economy and 13th Conference on Artificial Intelligence, AMS Annual Meeting, Phoenix, AZ, Jan. 7.

McCandless, T.C., S. E. Haupt and G. S. Young, 2014: Short Term Solar Radiation Forecasts Using Weather Regime Dependent Artificial Intelligence Techniques, Joint Session between the 12th Conference on Artificial and Computational Intelligence and its Applications to the Environmental Sciences and the Fifth Conference on Weather, Climate, and the New Energy Economy, AMS Annual Meeting, Atlanta, GA, Feb. 5.

McCandless, T. C., 2014: Sun4Cast Solar Irradiance Prediction and Statistical Short-term Solar Irradiance Prediction. Seminar at Aerospace Corporation, Pasadena, CA, October 25. **Pearson, J.M.,** Haupt, S.E., Jensen, T.L., Burghardt, C., McCandless, T.C., Brummet, T., and S. Dettling, 2015: Predicting Distributed Solar Power Production for Utilities. Sixth Conference on Weather, Climate, and the New Energy Economy: Short-Range Forecasting Modeling for Solar Electric Generation, Phoenix, AZ, Amer. Meteor. Soc. 4.2.

Schaeffer, N., L. M. Hinkelman, and T. P. Ackerman, 2016: Relating solar irradiance variations and weather across the Pacific Northwest, AMS 7th Conference on Weather, Climate, and the New Energy Economy, January 2016, New Orleans, LA.

Sengupta, M. and L. M. Hinkelman, 2014: Temporal variability of surface solar irradiance as a function of satellite-retrieved cloud properties, American Geophysical Union Fall Meeting San Francisco, CA.

2.3 CIRACAST

Team members at CIRA, led by Matthew Rogers, Steven Miller, and John Haynes, built and improved the CIRACast system, which is described below.

2.3.1 Motivation

Real-time forecasts of solar ramp events are a critical component of assuring high predictability of power output from PV and CSP plants. As part of the current SunShot initiative, the CIRA team has developed an operational, satellite-derived forecast for predicting global horizontal irradiance (GHI) on a timescale of 0-3 hours. The CIRACast method utilizes real-time retrieved cloud properties from the Pathfinder Atmospheres Extended (PATMOS-x) retrieval suite from GOES observations, projected forward in time using steering winds derived from numerical weather prediction (NWP) output, combined with a radiative transfer code to compute surface GHI, accounting for parallax, cloud shadow information (computed from sun angle and cloud-top height information) and changes in steering wind magnitude and direction.

Satellite-derived advection forecasts provide one method of addressing potential shortcomings in NWP-generated insolation forecasts, as issues with initialization of the cloud field within the model, along with model spin-up contribute both to lag time and potentially large errors in forecasting cloud shadows. As one component of the blended forecast model, CIRACast utilizes direct observation of the initial cloud field to develop a low-lag, accurate forecast with an intended peak of utility at the 0 to 1-h timeframe, reaching to the 0 to 3-h timeframe for ideal cases.

2.3.2 Literature Review

The utility of satellite-derived advection schemes has been demonstrated as filling a role between short-term (0-30 min) sky-imaging methods (Peng et al. 2015; Huang et al. 2013; Chow et al. 2011) and NWP-based methods (James et al. 2015.) Validation of satellite-derived advection forecasts (Stein et al. 2010; Miller et al. 2012; Perez et al. 2010) showcase the utility of advection-derived methods in the 0-1-hour timeframe. Variability in forecasted irradiance fields using satellite-derived methods can depend on many factors, including advective and cloud identification issues (Rogers et al. 2015), biasing in radiative transfer codes used (Stackhouse et al. 2008), and cloud evolution and weather-related issues (Stein et al. 2010.)

The CIRACast algorithm, described by Miller et al. (2012) and Rogers et al. (2015), seeks to maximize forecast accuracy by utilizing a 'best of both worlds' approach, combining the accuracy of cloud field identification from geostationary satellite observations with the generally well-forecasted wind guidance from NWP models. Results from the continued development of the CIRACast algorithm under this project will be described briefly in this report and in a forthcoming publication (Miller et al. 2016).

2.3.3 Impact of CIRACast on the State-of-the-art

An overview of the current state-of-the-art for advection-derived solar forecasting is provided by Miller et al. (2013), including details of physically-based satellite methods for short-term forecasting, provided in chapter 3 of that document by Miller et al. (2013). For the purposes of this project, the forecast window targeted by the exclusively satellite-based method falls naturally between the surface-based all-sky-camera techniques (which see the clouds in the field of regard, and can provide useful predictions of shadow crossings with the 0-30-minute timeframe) and the NWP methods that capture the dynamics responsible for cloud formation but face challenges in capturing detailed cloud distributions.

Our satellite-based / model-fusion approach to addressing this challenging problem comprises a novel attempt to leverage the best of both worlds. Satellite-based cloud location and properties are coupled with the model's inherent ability to define the dynamic, three-dimensional flow patterns in which these clouds propagate in a way that is far superior to imposing a fixed- direction vector wind. The forecasts operate within a temporal window (up to ~2 hr) where the benefits of deterministic cloud placement and realistic model advection are perceived to overcome the limiting assumptions of non-evolving cloud properties (including microphysics, growth/dissipation). We utilize high spatial/temporal resolution satellite retrievals of cloud mask, type, and optical properties (from the GOES satellites) to improve the model analysis.

The satellite-based technique utilized in this program has been some years in the making. We leverage a capability to process GOES cloud property retrievals based on NOAA/NESDIS algorithms and locally-produced ancillary datasets in real time. We apply these retrievals to the very short-term solar forecasting problem. Novel elements of this technique that represent a move toward a more physical treatment of the satellite observations and which, therefore, advance the state-of-the-art for satellite-derived advection forecasts, are:

- Deriving cloud properties and surface irradiance from NOAA's operational codes

- Accounting for satellite viewing parallax via GOES cloud height retrievals, which can result in significant cloud pixel displacements (particularly for high/deep clouds, which can lead to 10s of km, or multi-pixel, shifts)
- Accounting for the solar geometry (shadow-casting, particularly for high clouds and lower solar elevation angles, again 10s of km shifts)
- Taking advantage of differential steering effects of a divergent wind field on cloud trajectories, using GOES cloud height information, which allows for capture of cyclonic/anticyclonic flow in a way that is not possible from cloud motion vectors or single vector wind methods.

2.3.4 Advances Made

The components of the satellite-derived forecast algorithm were largely completed prior to the start date of the current project; advances made on the CIRACast algorithm under this project are the result of continued refinement and evaluation of forecast accuracy. During the project, nearly three complete years of forecasts over the three zones have been compiled, with opportunities to validate the forecast against surface observations of GHI from partner sites, including the SMUD, Southern California Edison, and Xcel Energy, as well as from the NOAA SURFRAD network.

Evaluation of forecast accuracy against SURFRAD observations for GOES-West-derived forecasts identified key sources of error leading to improvements in the CIRACast model; specifically, identification of improper advection of stationary cloud features (e.g., offshore marine stratocumulus, mountain wave clouds, etc.) led to minor improvements of the group-advection algorithm and initial development of a blended cloud-motion-vector and NWP advection product. During the course of the project, forecast accuracy, in terms of relative MAE, improved to meet target goals in many of the forecast areas, and improved by a large margin over all forecast areas.

Additionally, elements of the CIRACast algorithm were adapted to other components of the blended forecast algorithm. These components, which serve as deliverables for the project, include the adaptation of the shadow-casting and parallax correction components of the satellite algorithm, which, respectively, compute and correct for the location of cloud shadows derived from retrieved cloud-top properties and solar geometry, and the satellite geometry of observation. These two components were provided as source code to the WRF-Solar™ team and integrated successfully into that component's forecast ability. Additionally, CIRACast was adapted to work within the larger DICast® forecasting system, making the algorithm more readily ingestible by various forecast sources. Graphing and imagery code to aid in the analysis of satellite-derived forecasts were also developed during the course of the project.

As detailed previously, it was discovered during the project that the largest source of error in the CIRACast product was the incorrect advection of stationary cloud features in the advection scheme. For locations in coastal California, for example, the ubiquitous marine stratocumulus found offshore, which typically recedes seaward during the late morning, was often 'picked up' by onshore winds as diagnosed by the NWP guidance, and brought onshore, leading to significant over-forecasting of cloud ramp events for these coastal locations. Modifications to the cloud grouping algorithm that selected the model level used for advective winds based on cloud-top

properties were made, allowing for lower-level winds to provide guidance for cloud groups, ameliorated the situation somewhat, improving forecast accuracy for these cloud systems.

To further improve accuracy, work was begun on a blended forecast wind guidance, utilizing cloud-motion vector (CMV) winds in concert with NWP guidance to identify stationary cloud features. As CMV wind products are designed to identify significant wind features, it was found late in the project that stationary cloud features were often filtered out from the CMV analysis, complicating the ability to identify and utilize CMV-derived wind fields using off-the-shelf CMV analyses. Continued work on identification and quantification of near-stationary wind vectors using satellite observations, to be used in concert with NWP guidance to improve cloud advection, will continue outside of the scope of this project.

2.3.5 Initial Evaluations of CIRACast

As detailed previously, full forecast runs over the forecast domains extend from January 2014 to the present, with case runs extending back to March 2013. During that time, validation of the forecast algorithm was performed using primarily SURFRAD observations; extensive evaluation using the Desert Rock, NV, and Table Mountain, CO, SURFRAD sites were used to guide development of the algorithm.

Figure 2-39 and Figure 2-40 show scatter plots of forecasted vs. observed GHI for these SURFRAD locations, for 0-1 hour forecasts made during the forecast period of January-December 2014. As can be seen, general agreement between CIRACast and surface observations exist, with greater accuracy seen in the Desert Rock, NV, location. Errors in the Table Mountain, CO, location are somewhat higher due to the increased amount of cloud evolution seen over the Rocky Mountains in the 0-1 hour forecast period and the presence of stationary wave clouds often incorrectly advected over the Table Mountain site.

As computed, relative MAE values for the sites were 9.6% for the Desert Rock, NV, site and 21.8% for the Table Mountain, CO, site. Goals set by the project were relative MAE values under 10% for ‘simple’ sites, where significant cloud evolution and other confounding factors were minimal (such as the Desert Rock location), and under 20% for ‘complex’ sites (such as Table Mountain location.) CIRACast managed to meet the project goal for ‘simple’ sites, and was able to greatly improve forecast accuracy to come within approximately 2% of meeting target goals; continued improvement of the algorithm in identifying stationary cloud features and making appropriate modifications to the advective winds used in the algorithm would be expected to bring this error down to well below the error target.

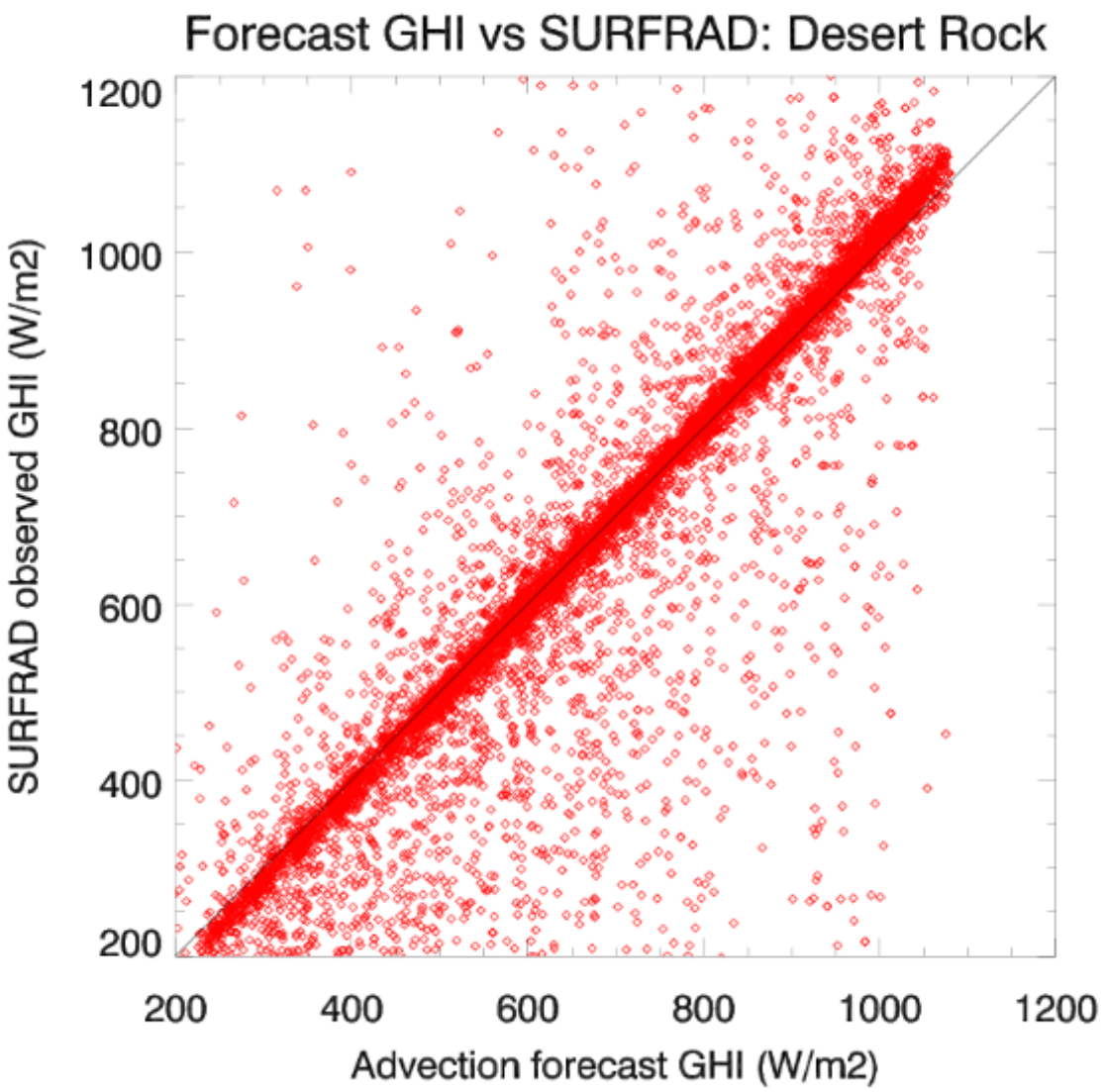


Figure 2-39. Scatter plot of forecasted GHI vs. observed GHI for the Desert Rock, NV, SURFRAD site. Forecasts are for the period between January and December 2014. Relative MAE for the period is 9.6%.

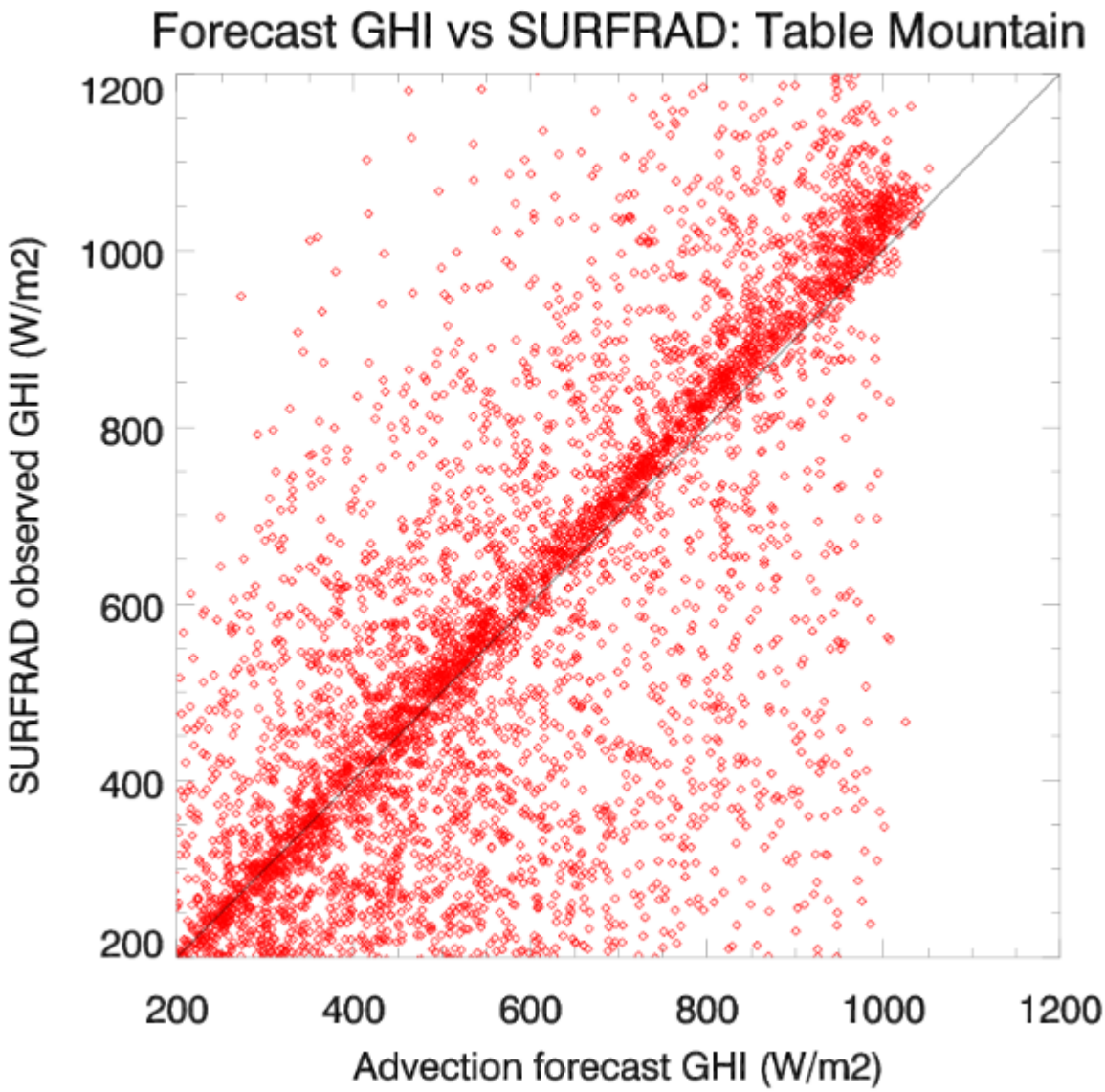


Figure 2-40. As with Figure 2-39, but for Table Mountain, CO. Relative MAE for the period is 21.8%.

2.3.6 Summary of Contribution

During the course of the project, CIRACast offered a stable, operational short-term forecast capability for the blended forecast product. As one of the first components of the project to become operational, CIRACast provided initial evaluation of the capabilities of the blended forecasts product and allowed for further development of the project goals, with regard to power generation forecasts and evaluation of project metrics. As the project evolved, continued modification and evaluation of CIRACast resulted in a state-of-the-art satellite-derived forecast mechanism, leading to several key presentations and publications.

Continued development of satellite-derived forecast mechanisms is expected to provide value to several fields, including municipal utility power generation and mobile and battlefield power generation systems, as well as continued refinement of satellite retrievals of GHI, to include enhanced methods to retrieve and account for the contribution of dust and aerosol impacts on GHI retrievals. Improvements in assessing advective guidance winds will further improve the accuracy of satellite-derived advection methods; using these improvements in the framework of an under-development probabilistic forecast method will further the efficacy of low-lag operational satellite-derived forecast products for power generation.

List of publications and presentations

Miller, S.D., A.K. Heidinger, and M. Sengupta, 2013: Physically-Based Satellite Methods. Chapter 3, *Solar Energy Forecasting and Resource Assessment,* J. Kleissl, Ed. ISBN 9780123971777

Miller, S.D., M.A. Rogers, A.K. Heidinger, I. Laszlo, and M. Sengupta, 2012: Cloud Advection Schemes for Short-Term Satellite-Based Insolation Forecasts. p 1963-1967, *World Renewable Energy Forum 2012,* C. Fellows, Ed., American Solar Energy Society, Denver, CO.

Rogers, M.A., S.D. Miller, J.M. Haynes, A. Heidinger, S.E. Haupt, and M. Sengupta, 2015: Improvements in Satellite-Derived Short Term Insolation Forecasting: Statistical Comparisons, Challenges for Advection-Based Forecasts, and New Techniques. Presentation 6.4, *Sixth Conference on Weather, Climate, and the New Energy Economy,* AMS 2015 Annual Meeting, Phoenix, AZ.

Rogers, M.A., S.D. Miller, J.M. Haynes, A. Heidinger, S. Benjamin, M. Sengupta, S.E. Haupt, and T. Auligne, 2013: Results from a Satellite-Derived Short-Term Insolation Forecast Technique: Comparison Against Surface Observations, NWP Predictions, and Challenges. Presentation A14E-2, *2013 AGU Fall Meeting,* San Francisco, CA.

Miller, S.D., M.A. Rogers, J.M. Haynes, and M. Sengupta, 2016: A Satellite-Initialized Model-Advected Scheme for Short-Term Solar Energy Forecasting. In preparation for *Solar Forecasting* 2016.

2.4 MADCAST

Pedro Jiménez of NCAR led the work of bringing in MADCast to the Sun4Cast® system.

Multisensor Advection-Diffusion NowCast (MADCast) is a new model designed for the analysis and short-term forecasting of clouds (Auligné 2014a,b; Descombes et al. 2014). The following description is inspired by Descombes et al. (2014), where the interested reader is referred to for further technical details of MADCast.

The cloud analysis is based on retrievals of multiple infrared (IR) sensors using the multivariate minimum residual (MMR) scheme (Auligné 2014a,b). MMR has been implemented in the Gridpoint Statistical Interpolation system (GSI; Kleist et al., 2009). GSI provides three-dimensional cloud fields that are subsequently advected and diffused by a modified version of the Weather Research and Forecasting model (WRF; Skamarock et al. 2008). Finally, the predicted cloud field is used to diagnose the surface irradiances completing the short-term forecast.

The following sections describe the methodology involving a MADCast forecast (section 2.4.1), the configuration of the model for the quasi-operational forecast (section 3), and the advantages of MADCast with respect to a purely numerical weather prediction (NWP) forecast (section 4).

2.4.1 Methodology

A MADCast forecast involves three steps to 1) retrieve the cloud field, 2) predict the evolution of the clouds, and 3) estimate the surface irradiance.

2.4.1.1 Cloud Fraction Retrieval

The fundamental piece of the cloud retrieval process is the MMR scheme (Auligné 2014a,b). The MMR scheme has been implemented in the GSI data assimilation system. GSI produces the cloud analysis consisting in a three-dimensional cloud fraction. This cloud fraction retrieval is performed following these steps:

- Calculate IR radiance using WRF and the Community Radiative Transfer Model (CRTM; Han et al. 2006) under the clear sky hypothesis.
- Compute the departures between the satellite radiances and WRF for multiple channels sensitive to different altitudes in the atmosphere.
- Apply the MMR scheme to the departures to solve a variational problem (similar to a 1D-Var approach) in order to retrieve a cloud fraction profile from the satellite fields of view.
- Interpolate the cloud fraction profiles from the satellite fields-of-view to the model grid points. Specific procedures are used to optimally combine the information of sounders and imagers in order to exploit their different horizontal and vertical resolutions.

The cloud retrieval has been implemented successfully for a number of IR instruments on board of polar-orbiting and geostationary platforms. These instruments include AIRS, IASI, CrIS, MODIS, GOES-Imager, GOES-Sounder, FY-2D VISSR, Himawari-7 MTSAT-2, METEOSAT-10 SEVIRI. The AIRS, IASI and CrIS radiances are available from the National Centers for Environmental Prediction. The GOES and the MODIS radiances are available from the National Aeronautics and Space Administration. Additional radiances files are available at Air Force 557th Weather Wing (formerly known as the Air Force Weather Agency).

A validation of the cloud retrieval results has been conducted with synthetic and real data, inter-comparison between instruments, and independent observations such as CloudSat (Xu et al. 2014).

2.4.1.2 Cloud Fraction Forecast

The forecasting component of MADCast is incorporated in the WRF model. WRF provides the dynamical framework to transport and diffuse the clouds over time. The three-dimensional cloud fraction is treated as a tracer, and WRF is run without activating the majority of the physical packages. The forecast is therefore faster than a full physics run and provides acceptable results in situations without significant changes in the cloud thermo-dynamical structure.

The outputs from the WRF forecasts are 1) the three-dimensional cloud fraction and 2) clear sky surface irradiance.

2.4.1.3 Irradiance Forecast

The last step in the MADCast forecast is to combine the clear sky irradiances and the three-dimensional cloud fraction from WRF into surface irradiances valid for all sky conditions. This requires some assumptions about how the clouds attenuate the shortwave radiation. In principle, clouds at different heights may have different hydrometeors and can lead to different cloud extinction properties. This makes estimation of the cloud extinction properties to be a challenge.

As a starting point, we have assumed that the clouds are homogeneous and thus have the same extinction properties. This assumption simplifies the calculation of the surface irradiance. The irradiance is calculated by vertically integrating the cloud fraction into a two-dimensional cloud fraction, and then reducing the clear sky irradiance by the fraction of the cloud field. This method has shown acceptable performance in the irradiance forecast.

2.4.2 Configuration of MADCast Quasi-Operational Forecast

MADCast has been run quasi-operationally since September 2014 until March 2016. The model runs every hour and provides a six-hour forecast length over CONUS at a 9-km horizontal grid spacing. The system assimilates IR irradiances from a geostationary satellite (GOES-13/GOES-East) and two instruments (AIRS and IASI) on board circumpolar satellites. The latency is 23 minutes for synoptic and semi-synoptic hours and 12 minutes for the rest of the hours.

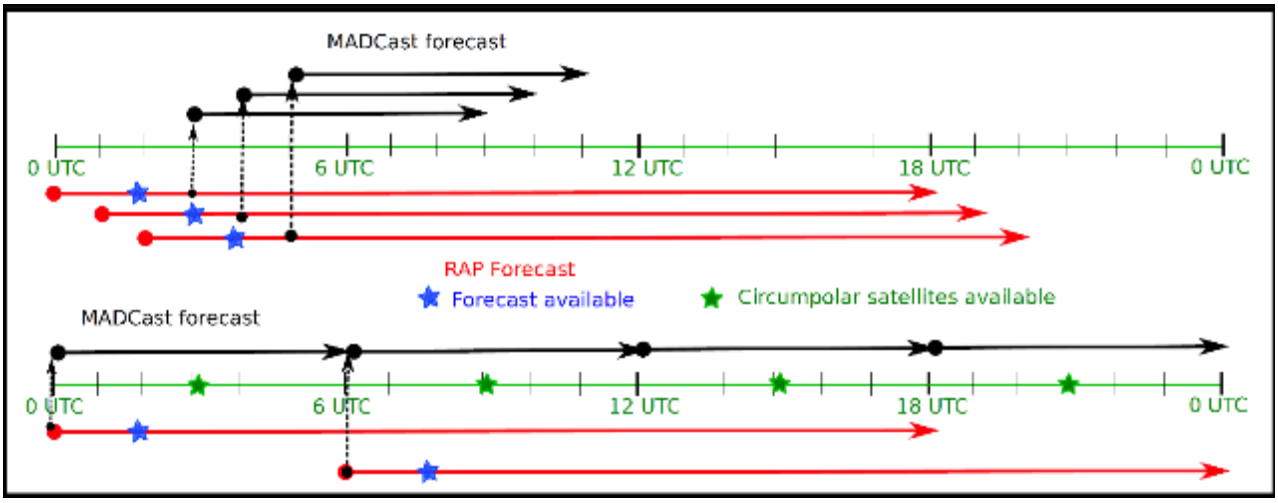


Figure 2-41. Strategy used to provide the operational MADCast forecast. Two MADCast cycles are performed. The upper part corresponds with the MADCast operational forecast whereas the lower part with the near real-time forecast.

The strategy used to provide forecasts is shown in Figure 2-41. Two different cycles of MADCast are run operationally in order to reduce the latency of the forecast. Both cycles use the Rapid Refresh (RAP) forecast run by NOAA to generate the initial and boundary conditions that drive the MADCast forecast. The "near real-time" cycle (lower part in Figure 2-41) assimilates IR irradiances from GOES as well as the radiances from the circumpolar satellites. There is a delay of 3 hours in the circumpolar radiances, so this cycle only provides a valid forecast for the last 3 hours of the simulation. The model is run four times a day on the synoptic hours doing cycling on the cloud fraction. The "operational cycle" (upper part on Figure 2-41) is responsible for the

forecast that goes into the nowcasting integrator. The cloud fraction of a given forecast is initialized using information from the "near real-time" cycle if there is a valid forecast for the initial conditions. If not, the cloud fraction from a previous run from the "operational cycle" is used. In addition, GOES radiances for the initial conditions are assimilated. GOES data are available with a delay of about 14 minutes for the synoptic and semi-synoptic hours and about 3 minutes for the rest of the hours. MADCast runs in about 9 minutes, which leads to the 23 or 12 minute latencies of the forecast. Hence, splitting the forecast in two cycles allows us to assimilate the irradiances from the circumpolar satellites without an adverse effect on the latency of the forecast.

2.4.3 MADCast Performance

MADCast added value with respect to a purely NWP forecast for some forecast horizons. Theoretically, MADCast should provide superior performance over NWP models like WRF-Solar™ at the beginning of the forecast due to the assimilation process. After some time, the WRF-Solar™ forecast should be superior given the better physics of the model.

Data from the quasi-operational forecast was used to investigate the added value of MADCast with respect to the WRF-Solar™ Nowcasting system (section 2.6). The RMSE as a function of the lead time calculated with the four months of quasi-operational runs from December 2014 to March 2015 for the GHI from MADCast and WRF-Solar™ Nowcasting are shown in Figure 2-42. Results summarize statistics over fourteen sites over the contiguous U.S. (7 SURFRAD sites and 7 Integrated Surface Irradiance Study [ISIS] sites). MADCast RMSE shows a steady increase as a function of the lead time. On the contrary, the RMSE of WRF-Solar™ shows larger values at the beginning of the forecast with near steady values after 1 hour 30 minutes. At 1 h30 min both models show similar RMSE. This indicates that, on average, MADCast forecast is superior to WRF-Solar™ Nowcasting during the first hour and a half.

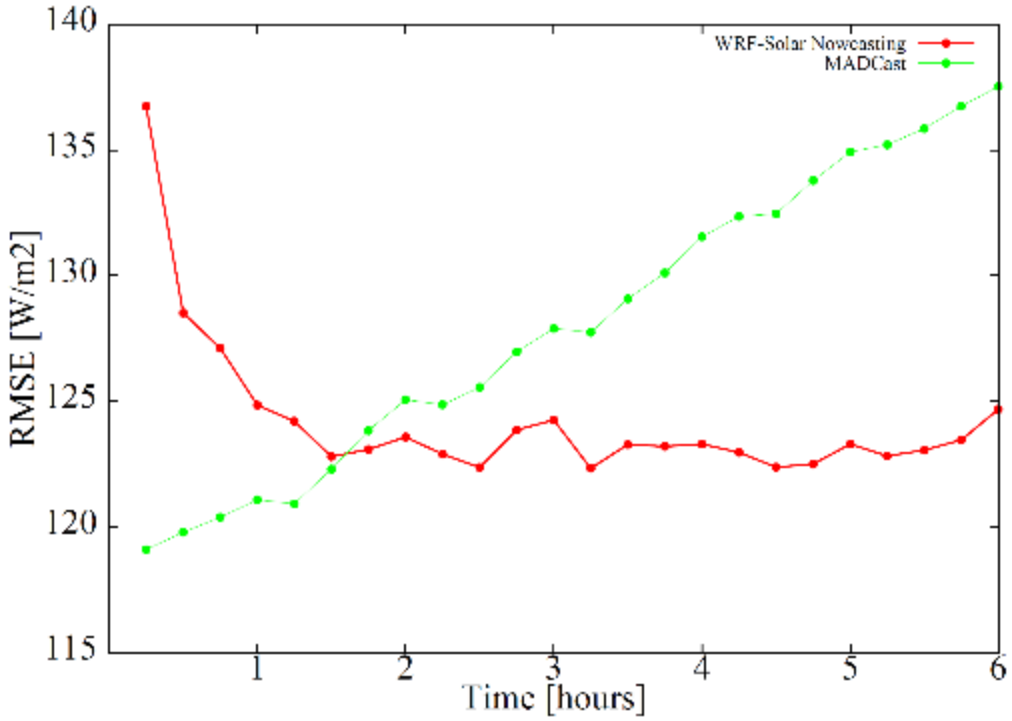


Figure 2-42. MADCast (green) and WRF-Solar™ Nowcasting (red) RMSE of the GHI as a function of the lead time.

The added value of MADCast in the short-term forecast (Figure 2-42) indicates the potential of the model for nowcasting applications. A comprehensive comparison of MADCast performance against observations, and the rest of the nowcasting components, is provided in chapter 5.

Publications regarding MADCast

Journal Papers

Auligné, T., 2014a: Multivariate Minimum Residual Method for Cloud Retrieval. Part I: Theoretical Aspects and Simulated Observations Experiments. *Mon. Wea. Rev.*, **142**, 4383 – 4398.

Auligné, T., 2014b: Multivariate Minimum Residual Method for Cloud Retrieval. Part II: Real Observations Experiments. *Mon. Wea. Rev.*, **142**, 4399 - 4415.

Descombes, G., T. Auligné, H.-C. Lin, D. Xu, C. Schwartz and F. Vandenberghe, 2014: *Multi-sensor Advection Diffusion nowCast (MADCast) for cloud analysis and short-term prediction*. NCAR Tech. Rep. TN-509STR. 21 pp.

Xu, D., Auligné T., Huang X.-Y., 2015: A Validation of the Multivariate and Minimum Residual Method for Cloud Retrieval Using Radiance from Multiple Satellites. *Advances in Atmospheric Sciences*, **32**, 349-362. doi: 10.1007/s00376-014-3258-5

2.5 WRF-Solar-Nowcasting

This section provides a description of the WRF-Solar-Nowcasting system, which was primarily developed by Pedro Jiménez of NCAR. The model is based on a specific configuration of the WRF-Solar™ model (Chapter 3 of this document; Jiménez et al. 2016a) designed to minimize

latency and maximize the speed of the forecast in order to meet nowcasting needs (section 2.5.2). The potential benefit of augmenting the WRF-Solar-Nowcasting system by including the effects of clouds not explicitly modeled is also discussed (section 2.5.3).

Much of the material in this section serves as the basis of Jiménez et al. (2016b).

2.5.1 Quasi-operational Forecast

The WRF-Solar-Nowcasting quasi-operational forecast started in September 2014. The model provides a 6-hour forecast over CONUS with a 9-km horizontal grid spacing. A schematic of the forecast strategy is shown in Figure 2-43. The model is run every hour using forecast data from the NOAA Rapid Refresh (RAP) model (Benjamin et al. 2016) to create the initial and boundary conditions. The model is initialized with the 3-h forecast from RAP, which allows us to have the forecast available around 40 minutes before the initial time of the simulation (equivalent to 40-min negative latency). The output is recorded every 15 min and it is an input to the nowcasting integrator.

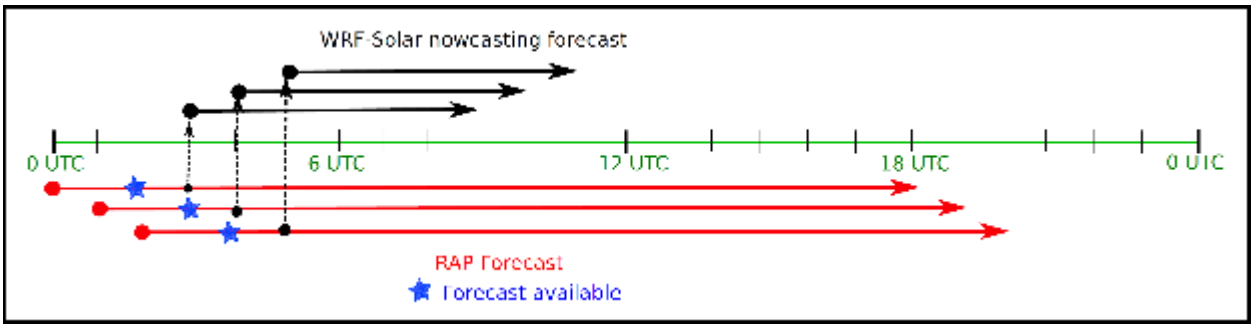


Figure 2-43. WRF-Solar™ Nowcasting quasi-operational forecast.

WRF-Solar-Nowcasting uses the solar augmentations that were available at the beginning of the quasi-operational forecasts. These include an improved solar tracking algorithm, output direct and diffuse components, and time series every model time step at sites with surface irradiance available for evaluation of the model performance. The aerosols are represented using a monthly climatology and thus avoiding the increase in the computational cost associated with the advection of aerosols. An important WRF-Solar™ augmentation for nowcasting needs, the feedback of unresolved clouds, was not available at the beginning of the quasi-operational period and thus was not activated.

The good performance of WRF-Solar-Nowcasting was already discussed in section 2.4.4, wherein it was shown that WRF-Solar-Nowcasting was competitive with the improved cloud initialization of MADCast after 1.5 hours of simulation.

A comprehensive description of the system performance is described in chapter 5.

In this section we focus on the radiative effects of unresolved clouds. Including the effects of the unresolved clouds is desirable at a 9-km horizontal grid spacing. Indeed, the WRF-Solar™ shallow cumulus (Deng et al. 2014) parameterization was specifically designed for these horizontal grid

spacings [O ~10 km]. In order to understand the potential benefit of activating the parameterization in this nowcasting framework, a sensitivity study was performed

2.5.2 Effects of the Unresolved Clouds

Three numerical experiments were performed with the WRF-Solar-Nowcasting model. The first experiment (NO-FEEDBACK) does not take into account the effects of the unresolved clouds on the atmospheric radiative transfer. The second experiment (DEEP-FEEDBACK) only activates the radiative feedbacks of deep convective cumulus clouds as an intermediate step in representing all scales of unresolved clouds. This is accomplished via the cumulus feedbacks to radiation implemented in the Grell-Freitas scheme (Grell and Freitas 2014). Finally, the third experiment (ALL-FEEDBACK) activates the radiative feedbacks of both the deep and shallow cumulus clouds using the parameterization implemented in WRF-Solar™ (Deng et al. 2014; see section 3.1).

Each experiment consists of 52 WRF-Solar-Nowcasting simulations spanning the year of 2014 to represent the different synoptic regimes over the CONUS. The first simulation is launched on 1 January 2014 followed by one simulation every week. This configuration reduces the temporal correlations and, at the same time, alleviates the computational cost of obtaining simulations for a complete year. The simulations are initialized at 15 UTC using analysis data from the RAP model with the lateral boundary conditions updated every 3 hours. Similar to the quasi-operational runs, WRF-Solar-Nowcasting is run for 6 hours, recording the output every 15 minutes.

The model uncertainty is considered by using the stochastic kinetic energy backscatter scheme (SKEBS; Shutts 2005; Berner et al. 2009). (SKEBS has recently been re-named the stochastic multi-scale perturbation scheme [STOMP].) There are a total of 10 ensemble members for each simulation. Thus there are 520 WRF-Solar™ Nowcasting runs (52 cases x 10 ensemble members) for each one of the three numerical experiments. The stochastic perturbations introduced in each ensemble member are configured to maximize the spread over the 6-h run.

The predictability of GHI is assessed against observations from the SURFRAD network (Augustine et al. 2000, 2005) and the ISIS network (Hicks et al. 1996). Each network has seven stations available over CONUS. These are among the highest quality GHI observations over CONUS. A quality control procedure to ensure consistency between the solar zenith angle and the recorded values was applied to further increase the quality of the data (Roesch et al. 2011). Additionally, the 1-min records from the SURFRAD network were averaged to 3 minutes to match the records provided by the ISIS network. Data were matched to WRF-Solar's 15-min output. A 5% error is randomly introduced to the simulated GHI to account for the observational error (Augustine et al.2000).

2.5.3 Systematic Errors

The benefits of accounting for the effects of unresolved clouds is first indicated by evaluating the GHI bias as a function of the lead time (Figure 2-44). The GHI value at the grid point nearest to each of the 14 observational sites was used in the comparison. We used bootstrap re-sampling to compute error bars in order to assess if the differences are statistically significant. The NO-FEEDBACK experiment showed a positive bias during all lead times with a mean bias of 49 W m^{-2}. Incorporating the feedbacks from cumulus clouds (DEEP-FEEDBACK experiment) reduced

the bias at all lead times (mean bias of 39 W m^{-2}, a 20% improvement over NO-FEEDBACK). When the effects of both deep and shallow cumulus were considered in the ALL-FEEDBACK experiment, the bias was considerably reduced with a statistically significant improvement for most of the lead times (mean bias of 22 W m^{-2}, a 55% improvement over NO-FEEDBACK).

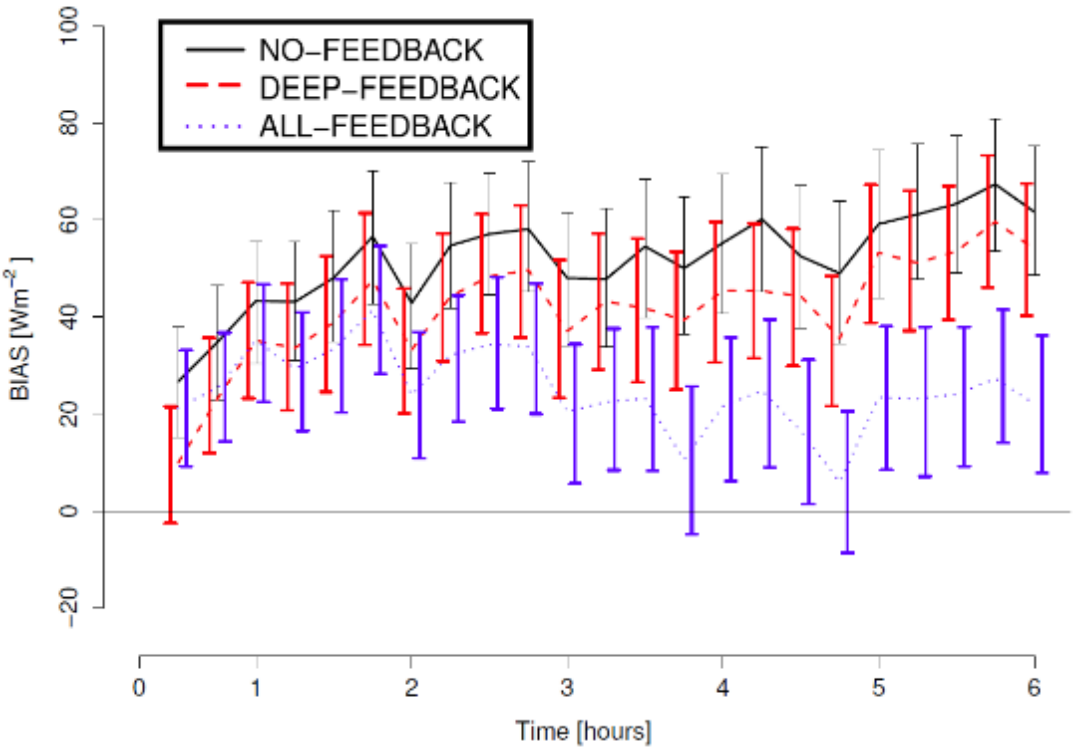


Figure 2-44. GHI bias as a function of the lead time for the three experiments (see legend). The bias is calculated using the complete set of SURFRAD and ISIS observations and the entire set of simulations. The error bars are calculated using a bootstrap re-sampling to test statistical significance.

Inspecting the bias at individual locations reinforces the conclusion that it is necessary to parameterize the cumulus feedbacks from both the shallow and deep convection schemes (Figure 2-45a). Experiments NO-FEEDBACK and DEEP-FEEDBACK displayed a positive bias at all the sites. This systematic error was corrected in the ALL-FEEDBACK experiment that largely suppressed the bias at most of the sites, and even displayed negative biases at one site.

The seasonal evolution of the bias further supports the benefits of parameterizing the radiative effects of the unresolved clouds (Figure 2-45b). Spring and summer (autumn and winter) months showed the largest (smallest) positive bias in NO-FEEDBACK. Activating the effects of the unresolved cumulus in experiment DEEP-FEEDBACK revealed a positive feedback during spring and summer. The spring and summer behavior was further improved when the effects of the shallow cumulus were activated in experiment ALL-FEEDBACK. ALL-FEEDBACK also had the largest GHI bias reduction during autumn. Winter shows little to no impacts associated with radiative effects of unresolved clouds.

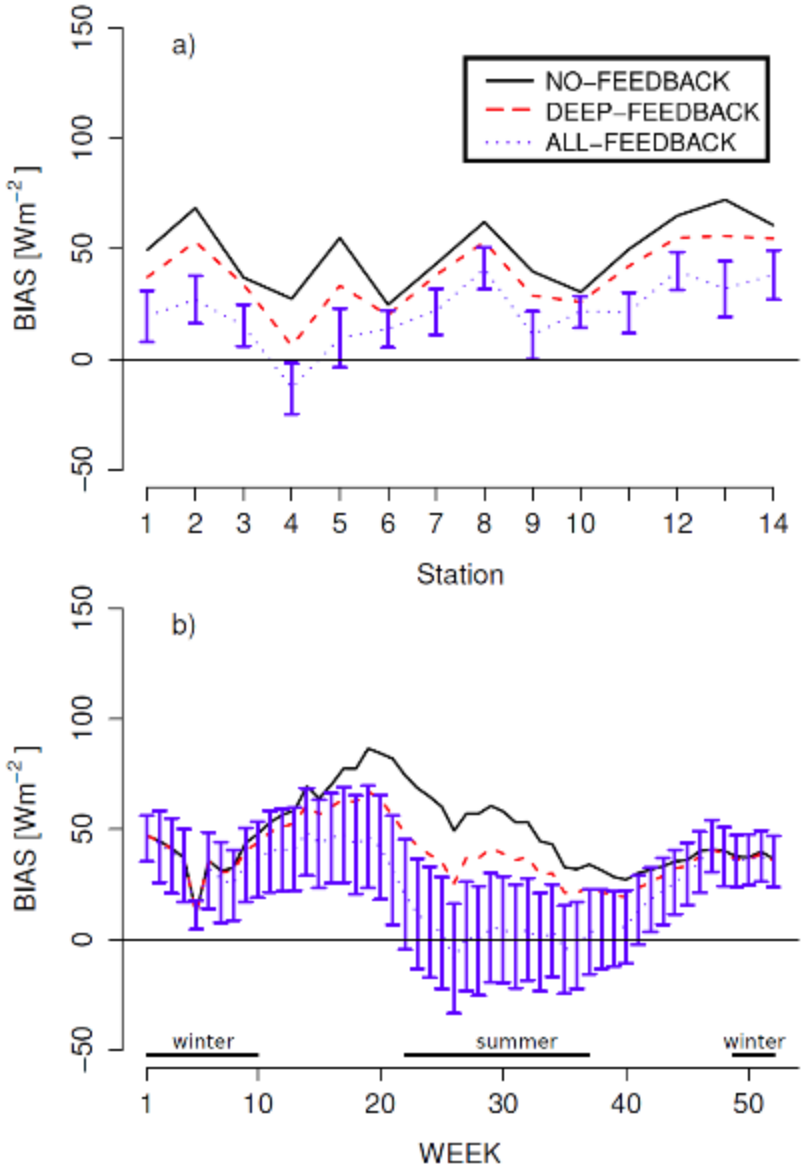


Figure 2-45. (a) Bias at each SURFRAD and ISIS observational site, and (b) seasonal evolution of the bias calculated with a running mean of the bias from each WRF-Solar™ Nowcasting simulations.

2.5.4 Error Characterization and Prediction

The RMSE and spread versus lead time are plotted in a dispersion diagram in Figure 2-46a. A perfect, statistically consistent ensemble would have identical RMSE and ensemble spread (e.g., Raftery et al. 2005; Wilks 2006). The three experiments all had similar RMSE (solid lines) at the different lead times. Experiment ALL-FEEDBACK produced slightly lower RMSE values than the other experiments but this reduction was not statistically significant. In addition, although experiments DEEP-FEEDBACK and ALLFEEDBACK produced an increase in the ensemble spread (dashed lines), no experiment displayed matching RMSE and spread.

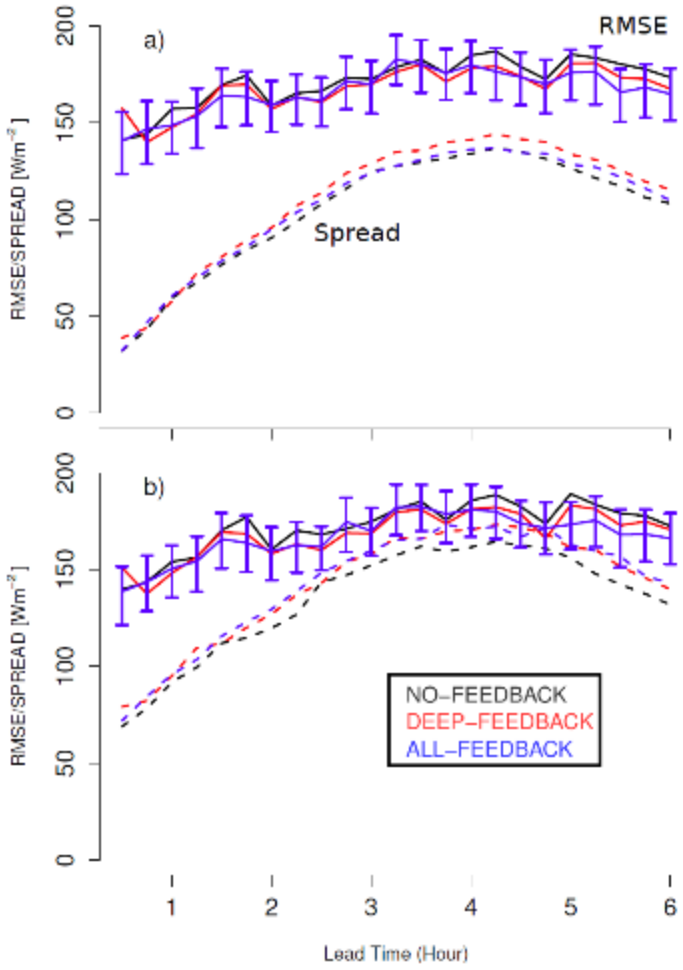


Figure 2-46. Dispersion diagrams using a) the nearest grid point to SURFRAD and ISIS sites and b) including the effects of the model's effective resolution.

The rank histograms (Anderson 1996) together with the missing rate error (MRE; Eckel and Mass 2005) provide further evidence of the under-dispersive nature of the ensembles (Figure 2-47a,b). A larger positive (negative) MRE reveals a more under-dispersive (over-dispersive) ensemble. Experiment NO-FEEDBACK exhibited a relatively flat histogram with a tendency for all ensemble members to overestimate GHI (MRE = 16.90%, Figure 2-47a). Activating the effects of the unresolved clouds in experiment ALL-FEEDBACK reduced the tendency to overestimate GHI (Figure 2-47b). However, all members underestimated GHI more frequently (MRE = 17.20 %). Experiment DEEP-FEEDBACK showed similar results (not shown).

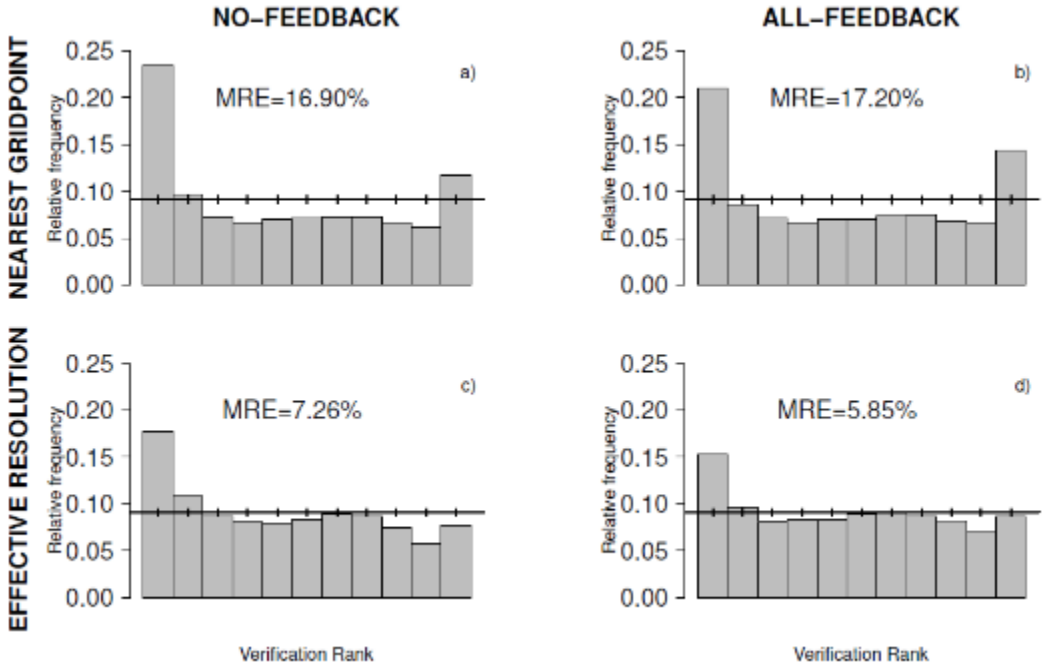


Figure 2-47. Rank histograms for GHI at the nearest grid point to SURFRAD and ISIS sites from the a) NO_FEEDBACK and b) ALL-FEEDBACK experiments. The rank histograms for GHI as a result of including the model's effective resolution for the c) NO-FEEDBACK and d) ALL-FEEDBACK experiments.

The under-dispersion can be largely ameliorated by considering the effects of the effective resolution (i.e. number of grid increments affected by numerical dissipation). The 7Δx effective resolution of WRF suggests that any phenomena below this scale are not properly resolved by the model. This implies that one must assess the resolved and unresolved clouds within 7 x 7 grid points of the nearest grid point to a given observational site. The simplest way to account for the impact of the effective resolution is to randomly perturb the GHI values at the nearest grid point using information from the grid points within the effective resolution. Here, we added a random value extracted from a normal distribution with zero mean and a standard deviation corresponding of the 49 GHI values of the 7 x 7 grid points surrounding the observational site. In order to avoid unrealistic values, the resulting GHI was forced to be greater than or equal to zero and no more than 10% higher than the clear sky GHI value.

The positive dispersive impacts of accounting for the effective resolution are illustrated by the rank histograms shown in Figure 2-47c and d. The frequency that all members of the ensemble overestimated GHI (Figure 2-47a and c) was decreased from experiment NO-FEEDBACK, although it was still under-dispersive (MRE = 7.26%; Figure 2-47c). Hence, the effects of effective resolution on the resolved clouds only partially improved the under-dispersion. The combination of both, accounting for the effective resolution and including the effects of the unresolved deep and shallow cumulus clouds, produced an ensemble with better statistical consistency (MRE = 5.85%, a 20% improvement; Figure 2-47d).

Including the impacts of the effective resolution also produces positive effects in the dispersion diagram (Figure 2-46b). The spread and the RMSE are now in better agreement, particularly for

the ALL-FEEDBACK experiment after lead times beyond 2 hours. The match of the RMSE and spread lines indicates that accounting for both the effects of unresolved clouds and the effective resolution of the model provides a meaningful quantification of the simulated uncertainty. We underline how difficult is to reach a good statistical consistency at these short lead times without any post-processing (Buizza et al. 2003).

2.5.5 Conclusions

The role of the unresolved clouds on the short-range GHI predictability has been systematically analyzed using observations and ensemble simulations spanning a one-year period. Including the radiative effects of unresolved cumulus, both deep and shallow, is necessary to significantly reduce a systematic positive bias in GHI. The bias improvements are systematic across all sites analyzed. Summer months reveal nearly unbiased GHI simulations whereas winter months show only small impacts to the radiative effects of unresolved clouds, which suggests that other cloud processes may not be properly represented in the model. Although slight improvements were produced in the RMSE, the error reduction is not statistically significant. Accounting for the unresolved clouds therefore produces a better cloudiness field, but the timing of the modeled clouds does not necessarily match with observations. This finding stresses the difficulties of simulating small atmospheric scales of motion, which are inherently uncertain, and necessitates the ensemble approach of this assessment to quantify uncertainties in the estimations.

A meaningful quantification of the prediction uncertainties (consistency between the RMSE and the spread of the ensemble) can be demonstrated with proper representation of the model's effective resolution. Thus one should consider the model's effective resolution, as well as the inherent uncertainty in modeling the specific location of convective clouds, when assessing GHI predictions that include the radiative effects of unresolved clouds.

2.6 MAD-WRF

The development of MAD-WRF was led by Pedro Jiménez of NCAR.

The two nowcasting systems, MADCast (section 2.4) and WRF-Solar-Nowcasting (section 2.5), have been coupled to form MAD-WRF in order to exploit desirable and compatible characteristics of each system. MADCast assimilates infrared irradiances from different satellite platforms to infer the presence of clouds. The assimilated cloud fraction is subsequently advected and diffused, with simplified model physics, in order to provide the cloud forecasts. Once the forecast is completed, assumptions are made on how the clouds attenuate the clear sky radiation in order to obtain the irradiance forecasts. WRF-Solar™ does not assimilate clouds, but has better physical packages to represent microphysics and the interaction of clouds with radiation. The coupled system therefore takes advantage of the data assimilation from MADCast and the better physics of WRF-Solar™.

There are three fundamental aspects that need to be considered to couple the models:

- First, the definition of the cloud fraction from MADCast is different than the WRF-Solar™ definition, which is the conventional definition for any NWP model. In MADCast, the cloud fraction of each grid volume represents the contribution to the 2-D cloud fraction (vertically integrated) that do not overlap with other grid volumes in the same vertical

column. On the contrary, the cloud fraction in NWP models is defined as the fraction of a given grid volume that has clouds. In order to use the assimilated cloud fraction in the coupled system (MAD-WRF), the MADCast cloud fraction needs to be converted to the NWP definition.

- Second, MADCast provides no information about the cloud water and ice mixing ratios that need to be specified in order to compute the shortwave radiative transfer.
- The third aspect is to consolidate the two sets of cloud fractions and cloud mixing ratios from MADCast and WRF-Solar-Nowcasting in order to provide one single set to the shortwave radiation package. This consolidation should take into account the better performance of MADCast during the beginning of the forecasts. Having consistent cloud fractions and mixing ratios and a strategy to consolidate the two sets are the fundamental components of the coupled MAD-WRF nowcasting system.

The cloud fraction from MADCast is converted to the NWP definition, assuming that the clouds are homogeneous. The first step is to calculate the 2-D cloud fraction from MADCast. This is straightforward given the MADCast definition. One just needs to add the cloud fractions of a given vertical column and repeat the process for each column. Assuming that the clouds are homogeneous implies that the 2-D cloud fraction is the cloud fraction of each grid volume that has clouds. In order to suppress the tendency of MADCast to produce clouds that reach the ground, a threshold has been imposed to classify the grid volumes into clear and cloudy. Figure 2-48 shows a conceptual diagram to illustrate this process.

The cloud and ice mixing ratios were specified with the assumption that the clouds are warm and homogeneous. Hence, we neglect the contribution of ice in the shortwave radiative transfer and impose the same cloud water mixing ratio for all the grid volumes identified as cloudy.

The two sets of cloud fractions and cloud mixing ratios were merged based on our experience with the quasi-operational runs performed with MADCast and WRF-Solar-Nowcasting. The analysis of four months of forecasts, from December 2014 to March 2015, indicates that MADCast is, on the mean, superior to WRF-Solar-Nowcasting during the first 1.5 hours of the simulation (see section 2.4). After that point WRF-Solar-Nowcasting is usually superior to MADCast. Based on these empirical results, the cloud fraction and mixing ratios from MADCast are imposed during 0-1 h; the two sets are averaged from 1-1.5 h, and the set from WRF-Solar-Nowcasting is imposed after 1.5 h.

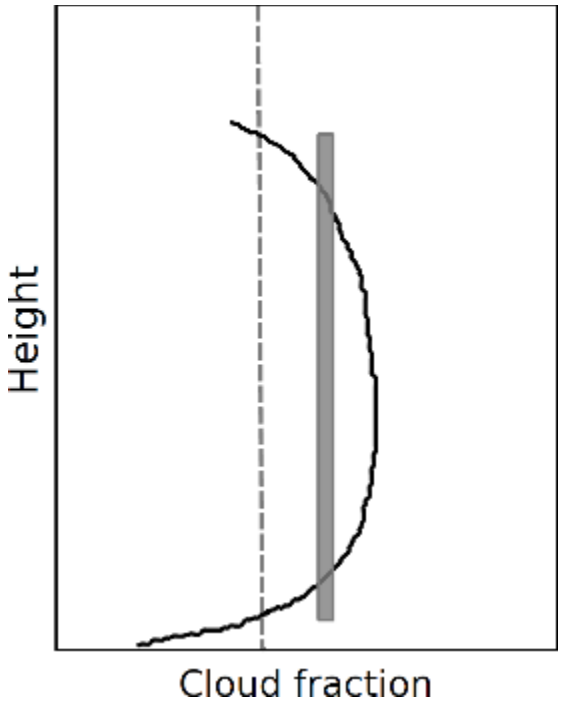


Figure 2-48. Conceptual diagram showing the cloud fraction from MADCast (black line) and its conversion to the NWP cloud fraction (gray rectangle). The threshold in the cloud fraction is also shown (gray dashed line).

The consolidation of the two sets of cloud fractions and mixing ratios will be revisited in the future to incorporate a more flexible framework. Comparisons of the quasi-operational forecasts from MADCast and WRF-Solar-Nowcasting will be used to redefine the merging strategy as a function of the month of the year and even as a function of the location. This is expected to provide better performance than the current static merging. However, the static merging is a first solid step forward that facilitates us to understand the behavior of MAD-WRF.

In order to compare the performance of the coupled system with respect to its both components, MADCast and WRF-Solar-Nowcasting, we are running the coupled system quasi-operationally twice a day starting in August 2015. The forecasts are launched at 1400 and 1600 UTC and run for 6 hours, recording the output every 15 minutes like the other nowcasting components. Results from this experiment are provided in chapter 5.

2.7 INTER-COMPARISON CASE STUDIES

Much of the material in this section serves as the basis for Lee et al. (2016).

2.7.1 Motivation

To better understand the performance of the various Sun4Cast® components in specific situations, a series of inter-comparison case studies was undertaken by Jared Lee of NCAR. Four case days were chosen for the region near Sacramento, California. These four case days represent canonical cloud cover regimes of the region, and thus are a good test forecast systems that predict GHI. Fifteen-minute average GHI predictions were compared against observations from seven pyranometers owned and operated by the SMUD. A map indicating the locations of these

pyranometers and nearby METAR sites appears as Fig. 2-49. Results indicate that each forecast system has its own strengths and weaknesses in the various regimes, times of day, and forecast lead times.

The specific components of Sun4Cast® that were used in the case studies were StatCast-Cubist, CIRACast, MADCast, and four configurations of WRF-Solar™ Nowcasting.

The StatCast-Cubist model was trained with ten-fold cross-validation on data for all daylight hours (i.e. both METARs and GHI observations > 25 W m^{-2} at forecast initialization time) for the period from 25 January–28 May 2014, excluding the four case study days. Twelve individual StatCast-Cubist models were trained independently to predict each of the twelve forecast lead times from 15 minutes and extending to 180 min. This allowed the StatCast-Cubist model to correct for any inherent systematic biases for those lead times. StatCast-Cubist forecasts were initialized every 15 minutes.

CIRACast forecasts were initialized every 30 minutes (:00 and :30), and produced forecasts for lead times from 15 minutes and extendingto 165 min. Forecasts initialized in the morning before 1500 UTC were not included in this analysis because of the algorithm's difficulty in detecting cloud objects in the visible satellite imagery with such a low sun angle.

The MADCast simulations are 6 hours long with the boundary conditions updated every 3 hours. Three simulations are performed for each case study with the model initialized at 1200, 1500, and 1800 UTC. The model output is recorded every 15 min. The aerosol direct effect is represented using a 4º-latitude x 5º-longitude monthly climatology from Tegen et al. (1997). While MADCast is designed to assimilate irradiances from different satellite instruments, the forecasts for these case studies only assimilated GOES-East irradiances to analyze the potential of imagers on board of geostationary satellites for solar irradiance forecasting.

For these case studies we ran four versions of WRF-Solar™ Nowcasting. The first experiment, "Baseline," only activated the solar augmentations introduced in support of solar energy applications, which include an improved solar tracking algorithm and updates in the modeled surface irradiances every time step of the model. The second experiment, "Aero," activated the aerosol direct effect using a high-resolution aerosol monthly climatology over CONUS (0.05º x 0.05º) generated from aerosol climatological data provided by Solar Consulting Services (SCS). The third WRF-Solar™ experiment, "ShCu," activated the Deng et al. (2014) mass-flux parameterization to represent radiative effects of unresolved deep and shallow convective clouds. While it is technically a mass-flux parameterization, it can also be considered a shallow cumulus scheme. The last WRF-Solar™ experiment, "Aero+ShCu," used both the higher-resolution climatological aerosol data and the Deng mass-flux scheme. As with the MADCast runs, the four WRF-Solar™ experiments were initialized at 1200, 1500, and 1800 UTC on each case day, with each forecast extending to a lead time of 6 hours.

2.7.2 Case Descriptions

With the aid of archived METAR sky cover observations from the Sacramento Executive Airport (KSAC) and GOES-15 (GOES-West) visible (0.65 μm) imagery, four days in April 2014 were identified that exhibited typical cloud cover regimes over the Sacramento area:

- Case 1 (17 Apr 2014): This day had mostly clear skies around Sacramento, with only a few small, isolated clouds that briefly moved over individual SMUD pyranometers during the middle of the day. This is essentially a control case. Extensive cloud cover in northern and central/southern California made it a challenging clear day to model, however.
- Case 2 (13 Apr 2014): This day had a low deck of marine stratocumulus in the Sacramento area in the morning. By about 1900 UTC (1200 PDT), the stratocumulus had burned off, leaving clear skies over each SMUD pyranometer. Predictions of the development, advection, and burn-off of marine stratocumulus clouds are challenging, and are of acute interest to utility operators on the U.S. West Coast.
- Case 3 (22 Apr 2014): This day had mostly clear skies in the early morning, before a field of small cumulus clouds formed over the Sacramento area around 1800 UTC (1100 PDT). The mix of clouds and sun remained through sunset. This case is a challenging one for any model, as the field of fair-weather cumulus clouds was essentially stochastic.
- Case 4 (25 Apr 2014): This day had mostly overcast, rainy conditions from sunrise to sunset, and 0.42" of rain fell at KSAC. From the GOES-15 imagery it was apparent that occasionally thinner clouds moved over some of the SMUD pyranometers throughout the day. This case was expected to provide another tough test for the models.

Each of these four case days presents a unique challenge for models, as will be discussed in the next section.

2.7.3 Results

2.7.3.1 Case 1

Mostly clear skies predominated over the Sacramento region on 17 April 2014, though there was considerable cloudiness to the northwest and to the south of Sacramento throughout the day, as seen in the GOES visible image shown in Fig. 2-50. As the day progressed, clouds advanced toward Sacramento from both directions. The SMUD pyranometers in general also exhibited a clear sky profile for GHI (Fig. 2-51), with little variability across the network. Network-averaged GHI observations and forecasts for the various models are shown in Fig. 2-52 and Fig. 2-53, and the network-averaged GHI mean error and mean absolute errors as a function of lead time are found in Fig. 2-54.

Note that in Fig. 2-52a,b, only the lead times extendingto 60 minutes for StatCast and CIRACast are presented in order to reduce clutter. The remaining panels (Fig. 2-52c,d and Fig. 2-53) show the three forecast initialization times (1200 UTC, 1500 UTC, and 1800 UTC) for MADCast and WRF-Solar™. Error bars denoting ±1 standard deviation across the SMUD network are always displayed for the observations, but only for select forecast time series in order to reduce clutter. The WRF-Solar™ Aero+ShCu forecasts are shown in panel d for both Fig. 2-52 and Fig. 2-53 to facilitate easier comparison among the various forecasts. Additionally, the thin lines in Fig. 2-54a,b denote the ±1 standard deviation ranges in the errors for each time series. The same plotting conventions and ranges are followed for all the cases.

With the exception of MADCast (Fig. 2-52c), all of the forecasting systems performed well for this day. MADCast incorrectly predicted that some of the clouds in the region would advect over

the SMUD sites, thus attenuating the modeled GHI substantially and leading to substantial under-forecasts. This highlights a limitation of satellite-based techniques, though CIRACast (Fig. 2-52b), the other satellite-based technique here, did not suffer the same problems as MADCast. On the other hand, CIRACast displayed a small consistent over-forecasting bias. This positive GHI bias most likely stems from insufficient modeled aerosol loading in CIRACast for this case day. This conclusion is further reinforced by the WRF-Solar™ experiments with the default aerosol modeling (Baseline and ShCu, Fig. 2-53a,c) having an even larger, consistent positive GHI bias than CIRACast. The WRF-Solar™ experiments with a representation of the aerosol direct effect (Aero and Aero+ShCu, Fig. 2-53b,d) noticeably reduced this positive GHI bias. This is strong evidence that improved handling of aerosols is necessary for numerical models to have better clear sky GHI predictions, consistent with Jiménez et al. (2016a). The best forecast model on this day was StatCast (Fig. 2-52a), which had a mean error of nearly 0 W m^{-2} for nearly all lead times (Fig. 2-54). As StatCast was trained on real GHI observations and weather conditions around the same time, it already accounts well for the aerosol loading in this region at this time. This result highlights one of the strengths of statistical forecasting, that no direct knowledge of the local aerosol loading is needed, provided it is not radically different from other days in the training dataset, and that the observed irradiance on that day forms the basis for the prediction.

2.7.3.2 Case 2

An area of marine stratocumulus clouds was widespread across coastal regions of California, and had infiltrated through the San Francisco Bay along the Sacramento River to the Central Valley around the SMUD region early in the morning of 13 April 2014, as seen in Fig. 2-55. The stratocumulus gradually burned off by about 1900 UTC, though not uniformly over all the sensors, and clear sky conditions prevailed the rest of the day (Fig. 2-56). Network-average observations and forecasts are shown in Fig. 2-57 and Fig. 2-58, and GHI forecast errors as a function of lead time are displayed in Fig. 2-59.

For the first 75-90 minutes of forecast lead time, StatCast unquestionably provided the best forecasts, as it had the lowest errors (Fig. 2-57a andFig. 2-59). For these short lead times, not only did StatCast perform well during the cloudy morning, but also during the clear afternoon, as in Case 1. Errors increased steadily with lead time, however, indicating that a longer training dataset may be helpful to capture additional marine layer events similar to this one.

CIRACast forecasts for the early morning failed to capture the marine layer clouds, but forecasts initialized later in the morning improved in predicting the cloud-attenuated GHI for 15-45 minute lead times (Fig. 2-57b). This short-term improvement for mid-morning was likely due to improved cloud object identification on the visible satellite data owing to a higher sun angle. However, even when CIRACast identified the presence of clouds in the SMUD area at initialization, CIRACast appeared to advect these stationary clouds out of the region after about an hour, thereby returning GHI forecasts to near clear sky values. Correctly simulating stationary clouds remains a difficult problem for satellite-based forecasts. For the clear skies later in the day, CIRACast suffered from the positive GHI biases that also occurred during Case 1, but otherwise performed well.

MADCast, in contrast to CIRACast, did not forecast any clouds over the SMUD region, and so overforecasted GHI in the morning (Fig. 2-57c). In the clear sky conditions, however, MADCast had a negative GHI bias, unlike any of the other models. This can likely be attributed to MADCast

diffusing and advecting marine stratocumulus from the northern California coast to the SMUD region, which provided a thin cloud cover in the model over the SMUD sensors, instead of dissipating the largely stationary clouds as was observed around midday. This finding illustrates that satellite-based methods like MADCast simply do not capture cloud evolution after forecast initialization, and sometimes have difficulty handling terrain-locked or other semi-stationary cloud features. And yet, despite these flaws in the forecast, MADCast had the lowest errors of any model for lead times of 120 minutes and beyond.

For the WRF-Solar™ experiments, none of the 1200 UTC runs simulated any clouds, while the 1500 UTC runs all nearly perfectly handled the marine layer for most of the morning, except for burning off the stratocumulus about 1-2 hours too quickly (Fig. 2-58). The 1800 UTC runs burned off the clouds at nearly the correct time. These findings illustrate the tremendous potential value that NWP can have in difficult forecast situations, even for nowcasting time scales, though additional work is needed to identify why the 1200 UTC runs failed to capture the marine layer clouds. For the clear skies later in the day, WRF-Solar™ forecasts exhibited the same positive GHI biases that were endemic to the Case 1 forecasts, with reduced biases in the Aero and Aero+ShCu experiments (Fig. 2-58b,d) than in the Baseline and ShCu experiments (Fig. 2-58a,c).

2.7.3.3 Case 3

Because a robust and long-lasting field of fair-weather cumulus clouds formed in place over the SMUD region by 1800 UTC on 22 April 2014 (Fig. 2-60), following clear skies in the morning, Case 3 is a particularly interesting and challenging case to model. The immense variability in observed GHI across the SMUD network throughout the day can be seen in Fig. 2-61. For a situation like this, correct predictions of the extreme variability in the forecast across the network would be critical to grid operators. GHI forecasts and observations for StatCast, CIRACast, MADCast, and WRF-Solar™ are shown in Fig. 2-62 and Fig. 2-63. Forecast errors as a function of lead time appear in Fig. 2-64.

Fig. 2-62 and Fig. 2-63 together indicate that all of the forecast methods actually performed reasonably well for this complex and difficult case, though StatCast (Fig. 2-62a) and CIRACast (Fig. 2-62b) generally had a positive GHI bias through the first 60 minutes of lead time. Later lead times generally had larger errors for this case (Fig. 2-64), which is to be expected in a case with a highly variable cloud field that forms in place overhead essentially stochastically, rather than being advected into the area. Cases like this are difficult if not impossible for satellite-based forecasting systems like CIRACast to capture well. MADCast, somewhat surprisingly, had the lowest errors overall of any of the forecast modules for Case 3 because its SMUD-average predictions remained in the middle of the variability (Fig. 2-62c). The small standard deviations on the MADCast predictions are indicative of a smooth cloud field that was consistent over the entire network at any given time. So even though the details of the cloud field were incorrect in MADCast, and even though the variability was decidedly insufficient on this day, the average GHI prediction was, on balance, good for this particular day. The WRF-Solar™ runs with the Deng ShCu scheme turned on (ShCu and Aero+ShCu, in Fig. 2-63c,d) also had generally reduced errors compared to the runs without the ShCu scheme (Baseline and Aero, in Fig. 2-63a,b), illustrating the benefits of using that scheme to better represent unresolved clouds.

2.7.3.4 Case 4

Skies were overcast throughout 25 April 2014, but not uniformly so, as can be seen in the satellite image at 2100 UTC in Fig. 2-65. The occasional thinning of the clouds over certain sensors was reflected in the occasional peaks in observed GHI in Fig. 2-66, though observed GHI was quite low most of the day. GHI forecasts from the various models are shown in Fig. 2-67 and Fig. 2-68, with the forecast errors as a function of lead time displayed in Fig. 2-69.

StatCast predictions were generally good in a lead time range of 15-45 minutes (Fig. 2-67a), but the errors quickly grew beyond that lead time (Fig. 2-69). The short positive trends in GHI as clouds occasionally thinned appear to have led to StatCast often predicting those thinning trends to continue, leading to large errors at longer lead times. Overcast days are far less common in Sacramento than clear days, but a longer training dataset with additional overcast days included, or potentially a regime-dependent version of StatCast, as described in section 2.2 (e.g., McCandless et al. 2016a), would likely have led to better StatCast forecasts.

CIRACast predictions at lead times less than 60 minutes were also reasonable, but with a marked positive GHI bias indicating that the clouds were not optically thick enough in the model, and occasional near-clear sky forecasts for short periods (Fig. 2-67b). The satellite loop for that day periodically revealed pockets of clear sky or thin clouds to the west of Sacramento that eventually were filled in by other, thicker clouds before reaching the region. It is these clear pockets that CIRACast detected and predicted would advect over the SMUD region. As with Case 3, satellite-based forecast methods have extreme difficulty performing well when the cloud field rapidly develops or dissipates after initialization.

MADCast did predict cloud cover, but, like CIRACast, the cloud cover did not attenuate enough irradiance (Fig. 2-67c). Also, as in Case 3, the variability in the MADCast forecasts was generally far smaller than the observed variability, and far smaller than the CIRACast variability. Even so, it was one of the better forecasts on this day, from an average error standpoint, though the errors grew steadily with time (Fig. 2-69).

The WRF-Solar™ forecasts without the Deng scheme (Baseline and Aero, in Fig. 2-68a,b) poorly simulated the clouds through most of the day. Positive GHI biases were often large, due to insufficient cloud coverage and thickness, and the average errors were frequently the largest of any model. On the other hand, in the experiments with the Deng shallow cumulus scheme (ShCu and Aero+ShCu, in Fig. 2-68c,d), the GHI biases were markedly reduced, and even nearly eliminated during the middle of the day. These experiments generally produced the lowest errors of any model for Case 4. Even more so than Case 3, these results for Case 4 highlight the vital importance of simulating unresolved clouds for obtaining good GHI forecasts.

2.7.4 Summary

In this inter-comparison study the GHI forecasts from several forecast models — StatCast-Cubist, CIRACast, MADCast, and four versions of WRF-Solar™ — were compared over four case days with canonical sky cover regimes for the region of Sacramento, California. The various forecasts were compared against observations from seven pyranometers operated by the Sacramento Municipal Utility District (SMUD). While this study did not exhaustively study all types of sky

cover conditions, the four case days included a clear sky day, a day with morning marine layer stratocumulus followed by clear sky, a day with a field of fair-weather cumulus forming in place over the region, and an overcast day.

Statistical forecasting, as accomplished with StatCast-Cubist, was the best forecast under clear skies, because its training dataset and the observations on that day had already accounted for attenuation from typical aerosol loading in the area. Furthermore, there was an abundance of clear days in the training dataset, which benefitted StatCast's performance for Case 1. For all case days, StatCast often had some of the lowest errors in the first 45-60 min. When clouds were present, however, GHI forecast errors for longer lead times increased, unsurprisingly, especially in cases when trends in GHI reversed themselves due to rapidly changing cloud cover.

Satellite-based forecasting methods, as with CIRACast and MADCast, were also generally good at short lead times, but struggled on days when clouds were rapidly forming, growing, and decaying after forecast initialization. These types of situations are a significant limitation for satellite-based forecast models. MADCast generally predicts a smoother cloud field than does CIRACast, which at times resulted in better mean GHI predictions across the network for MADCast, but at the cost of grossly underestimating the GHI variability. Furthermore, the treatment of aerosols and the corresponding GHI attenuation is an important factor in making accurate predictions of GHI: on the clear afternoon of Case 2, when both CIRACast and MADCast also predicted clear skies, CIRACast had a positive GHI bias, while MADCast had a negative GHI bias because it advected some thin marine stratocumulus clouds from the coast, even though the mostly stationary observed clouds instead dissipated. In many locations, model details of how to handle nearly stationary or terrain-locked clouds are a vital consideration.

Numerical weather prediction with WRF-Solar™ was shown to provide comparatively accurate GHI predictions for all the cases studied here, especially when using a high-resolution aerosol dataset to represent the aerosol direct effect and when using the Deng mass-flux scheme to represent radiative effects of unresolved shallow cumulus clouds. The improved aerosol treatment made a noticeable difference in clear sky conditions, while the shallow cumulus scheme led to substantially reduced GHI errors when cloud cover was extensive.

Each of the four case days that we examined revealed various strengths and weaknesses of the component Sun4Cast® nowcasting systems. At various times of day, lead times, and sky cover regimes, different components do better or worse. This finding reinforces the need to include all these diverse components in the Sun4Cast® system, and also indicates the need for further research to develop a dynamic weighting algorithm to make the most advantage of each system.

List of Publications and Presentations from this work

Lee, J. A., S. E. Haupt, P. A. Jiménez, M. A. Rogers, S. D. Miller, and T. C. McCandless, 2016: Solar energy nowcasting case studies near Sacramento. *Wea. Forecast.*, submitted.

Lee, J. A., S. E. Haupt, P. A. Jiménez, T. C. McCandless, M. A. Rogers, S. D. Miller, and X. Zhong, 2016: Nowcasting case studies with Sun4Cast. *7th Conf. on Weather, Climate, Water, and the New Energy Economy at the 96th AMS Annual Meeting*, New Orleans, LA, 11 Jan 2016. Abstract available at: https://ams.confex.com/ams/96Annual/webprogram/Paper288480.html.

Lee, J. A., S. E. Haupt, P. A. Jiménez, T. C. McCandless, M. A. Rogers, and S. D. Miller, 2015: Solar energy nowcasting case studies near Sacramento. *AMS 27th Conf. on Weather Analysis and Forecasting/23rd Conf. on Numerical Weather Prediction*, Chicago, IL, 3 Jul 2015. Abstract available at: https://ams.confex.com/ams/27WAF23NWP/webprogram/Paper273565.html.

Lee, J. A., S. E. Haupt, P. A. Jiménez, T. C. McCandless, M. A. Rogers, and S. D. Miller, 2015: Comparison of solar energy nowcasting techniques. *6th Conf. on Weather, Climate, and the New Energy Economy at the 95th AMS Annual Meeting*, Phoenix, AZ, 6 Jan 2015. Abstract available at: https://ams.confex.com/ams/95Annual/webprogram/Paper264080.html.

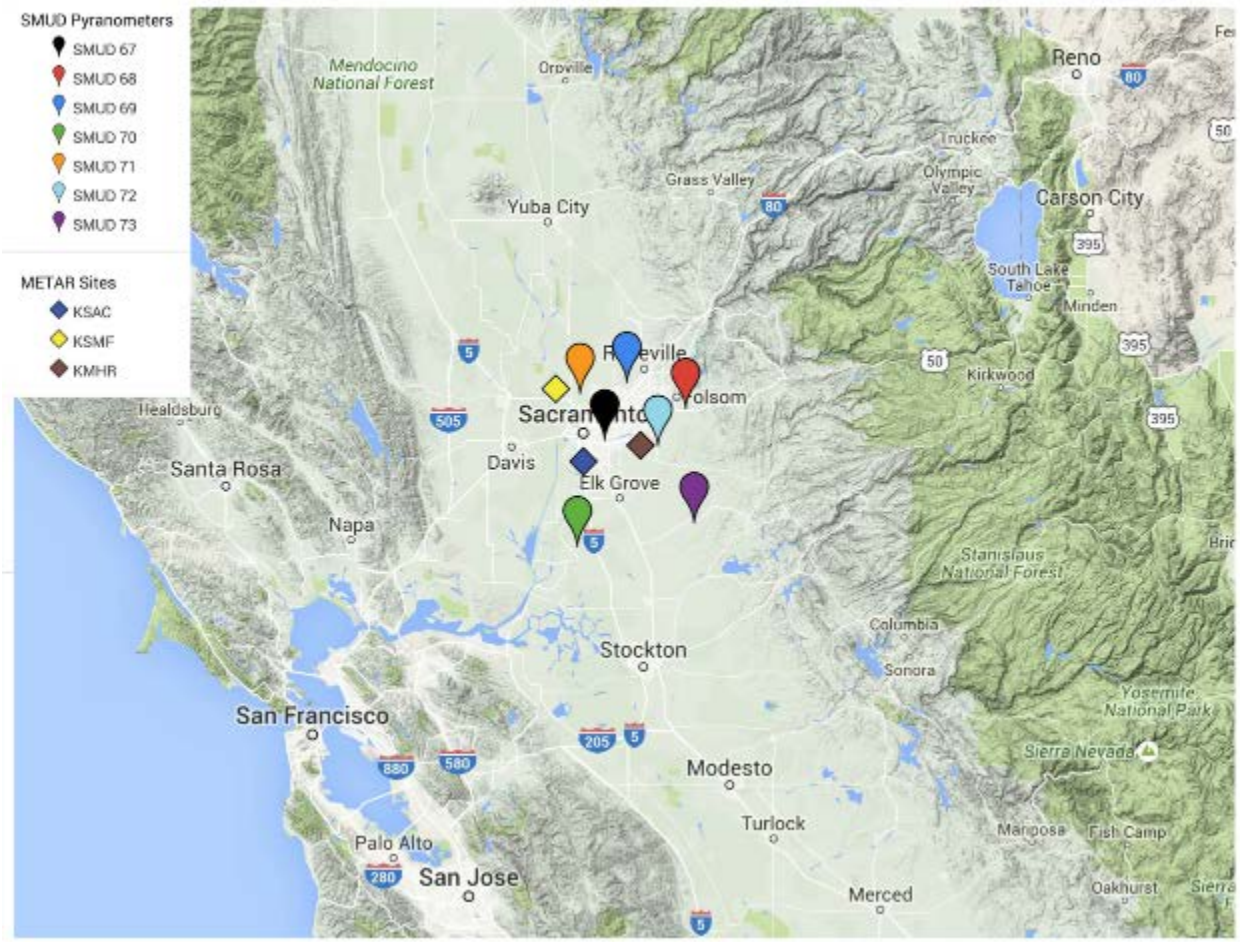


Fig. 2-49. Map of the locations of the SMUD pyranometers (pins) and nearby METAR sites (diamonds). See the legend for site names and corresponding symbol colors.

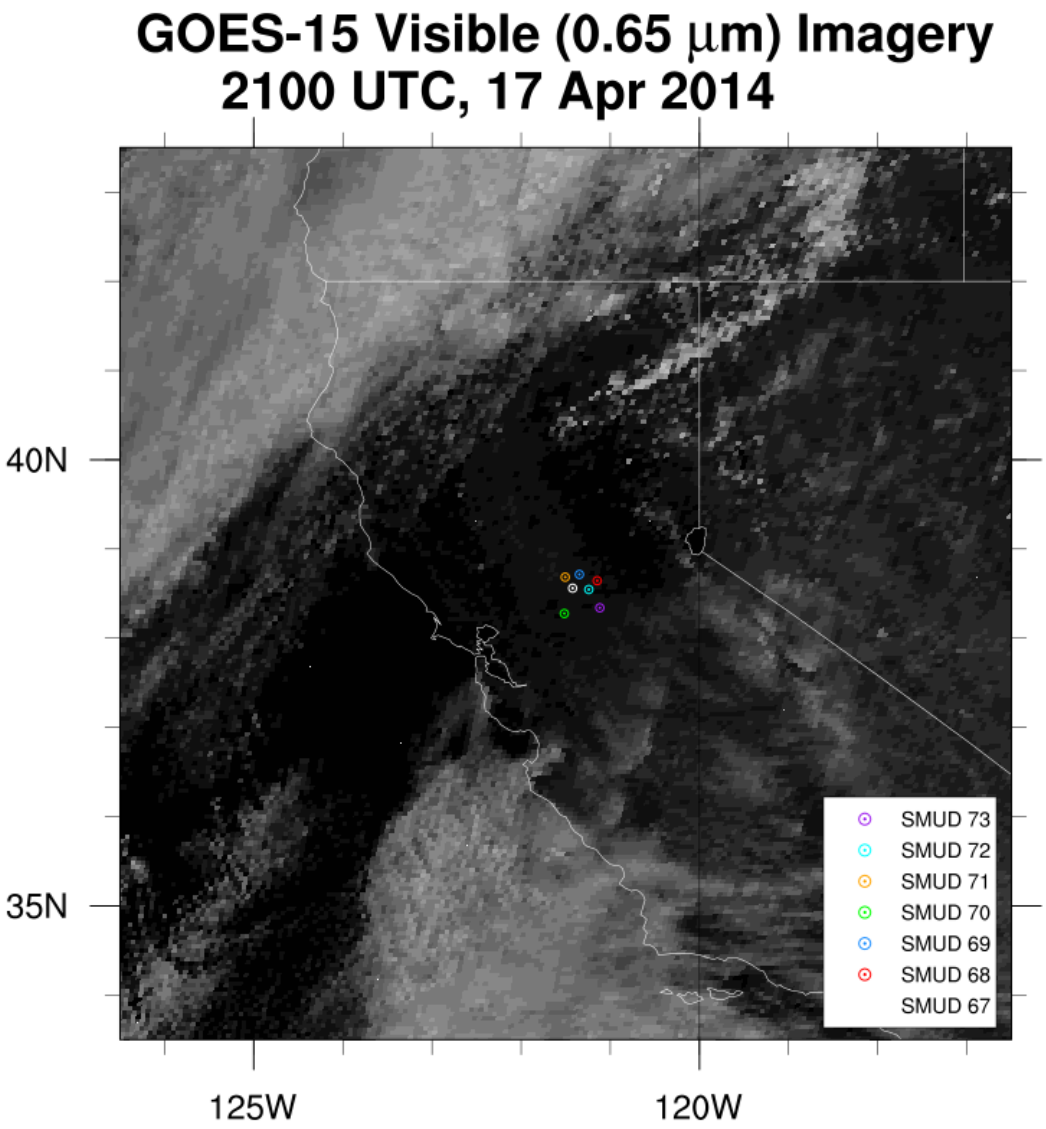


Fig. 2-50. Visible satellite image at 2100 UTC during Case 1. The locations of SMUD sensors 67-73 are indicated by the colors in the legend.

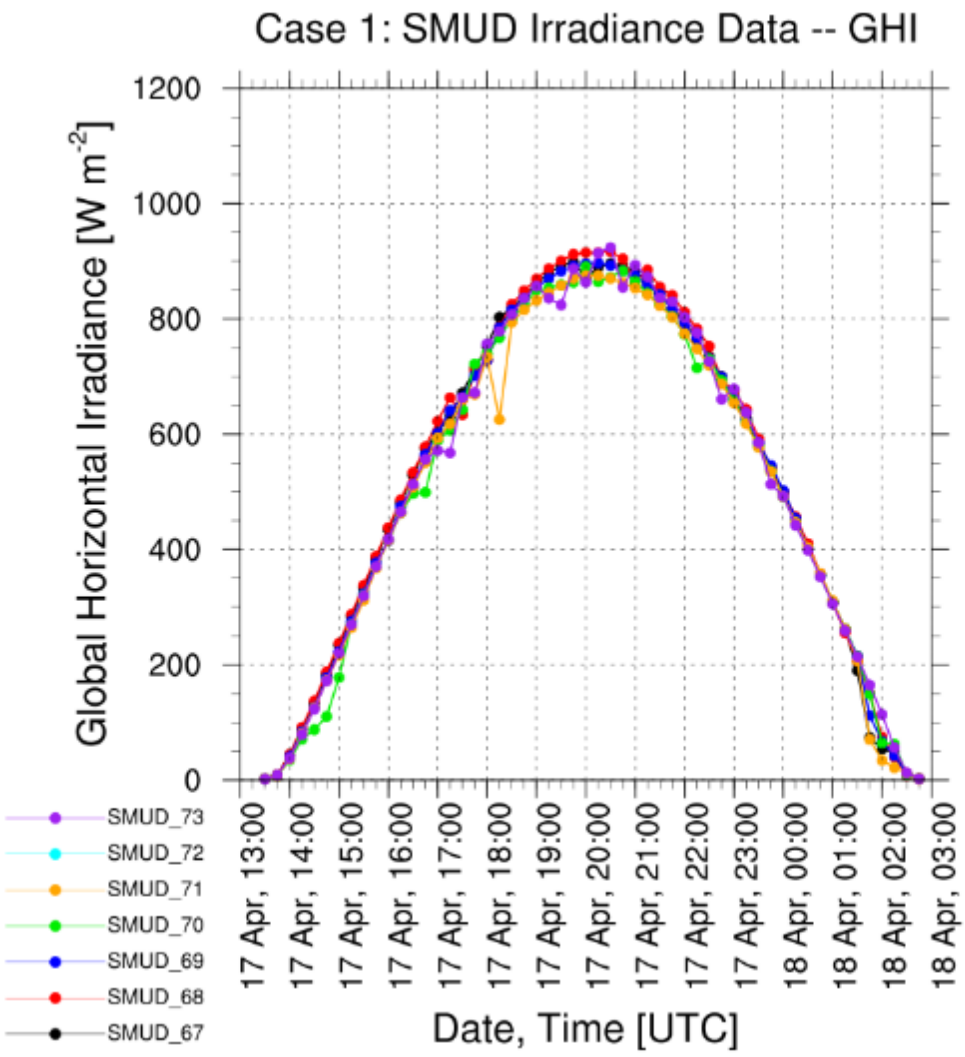


Fig. 2-51. Time series of 15-minute average GHI recorded by SMUD sensors 67-73 for Case 1. The color for each sensor is indicated in the legend.

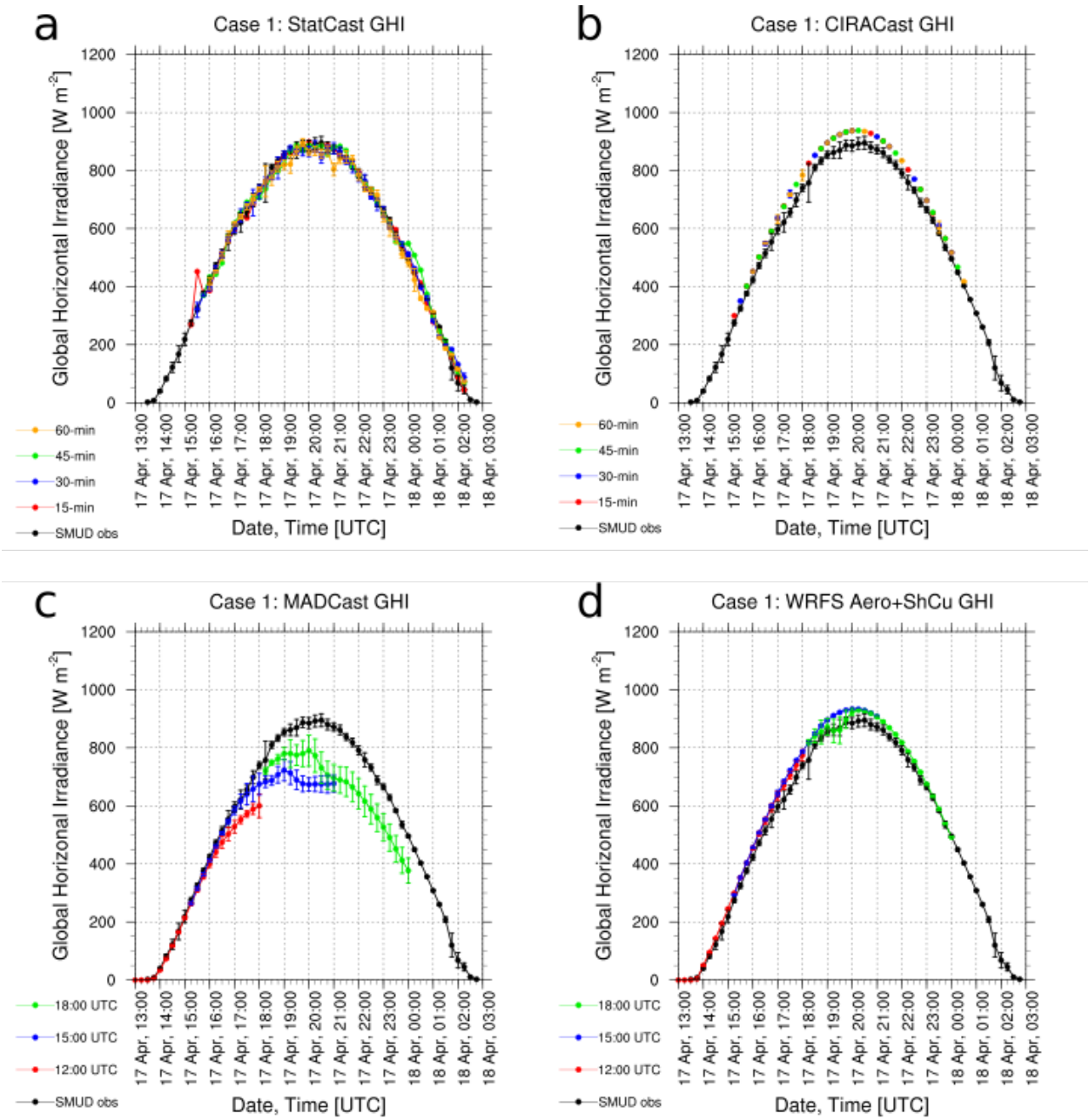


Fig. 2-52. 15-minute GHI predictions and observations for Case 1, averaged over the SMUD 67-73 sensors, for a) StatCast, b) CIRACast, c) MADCast, and d) WRF-Solar™ Aero+ShCu. The error bars for select time series denote ±1 standard deviation across the SMUD sensors.

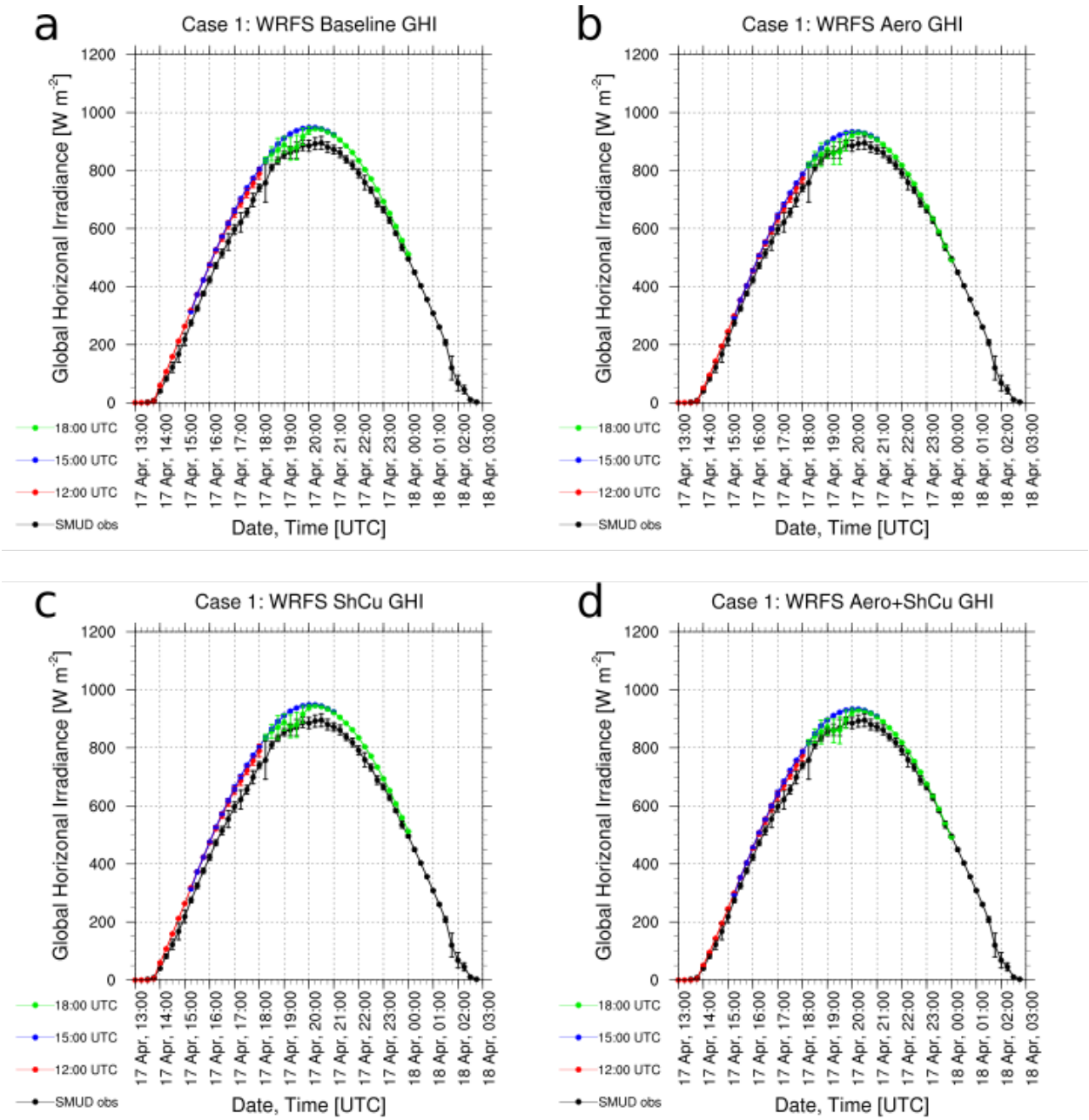


Fig. 2-53. 15-minute GHI predictions and observations for Case 1, averaged over the SMUD 67-73 sensors, for a) WRF-Solar™ Baseline, b) WRF-Solar™ Aero, c) WRF-Solar™ ShCu, and d) WRF-Solar™ Aero+ShCu. The error bars denote ±1 standard deviation across the SMUD sensors.

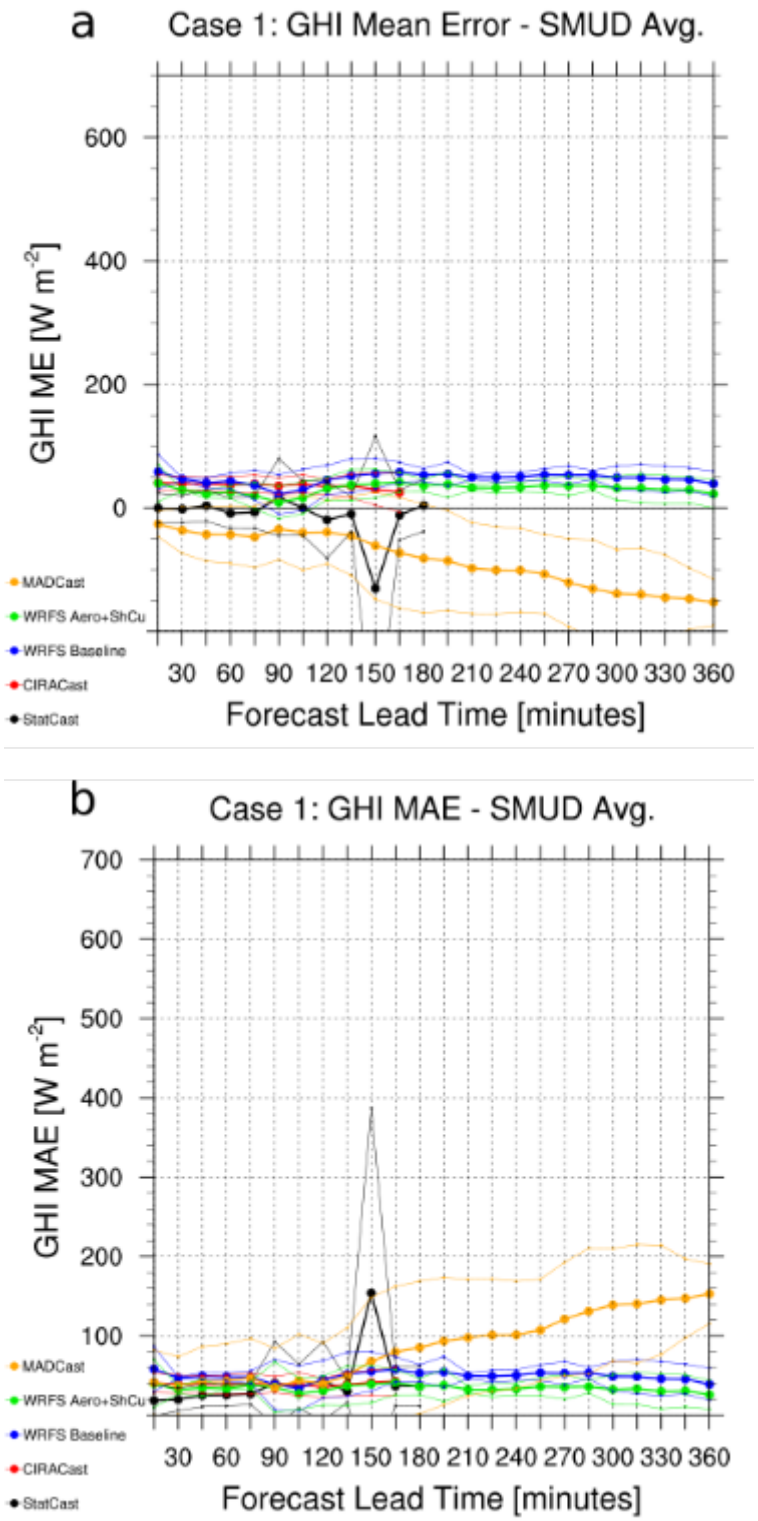


Fig. 2-54. a) Mean error and b) mean absolute error for the nowcasting components for Case 1, as a function of forecast lead time. Thick lines denote average values, while thin lines denote the ±1 standard deviation range. Colors are described in the legends.

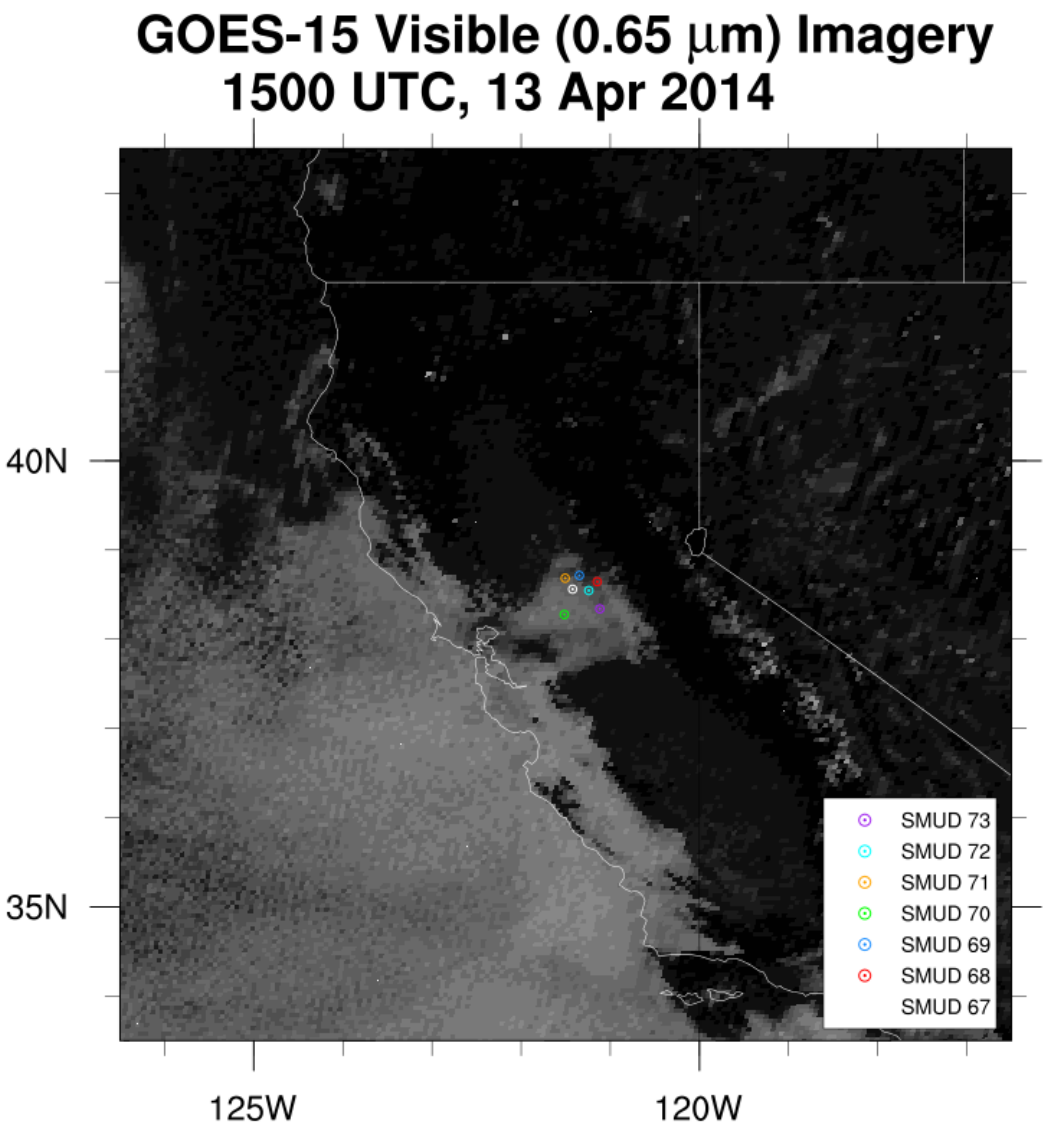


Fig. 2-55. Visible satellite image at 1500 UTC during Case 2. The locations of SMUD sensors 67-73 are indicated by the colors in the legend.

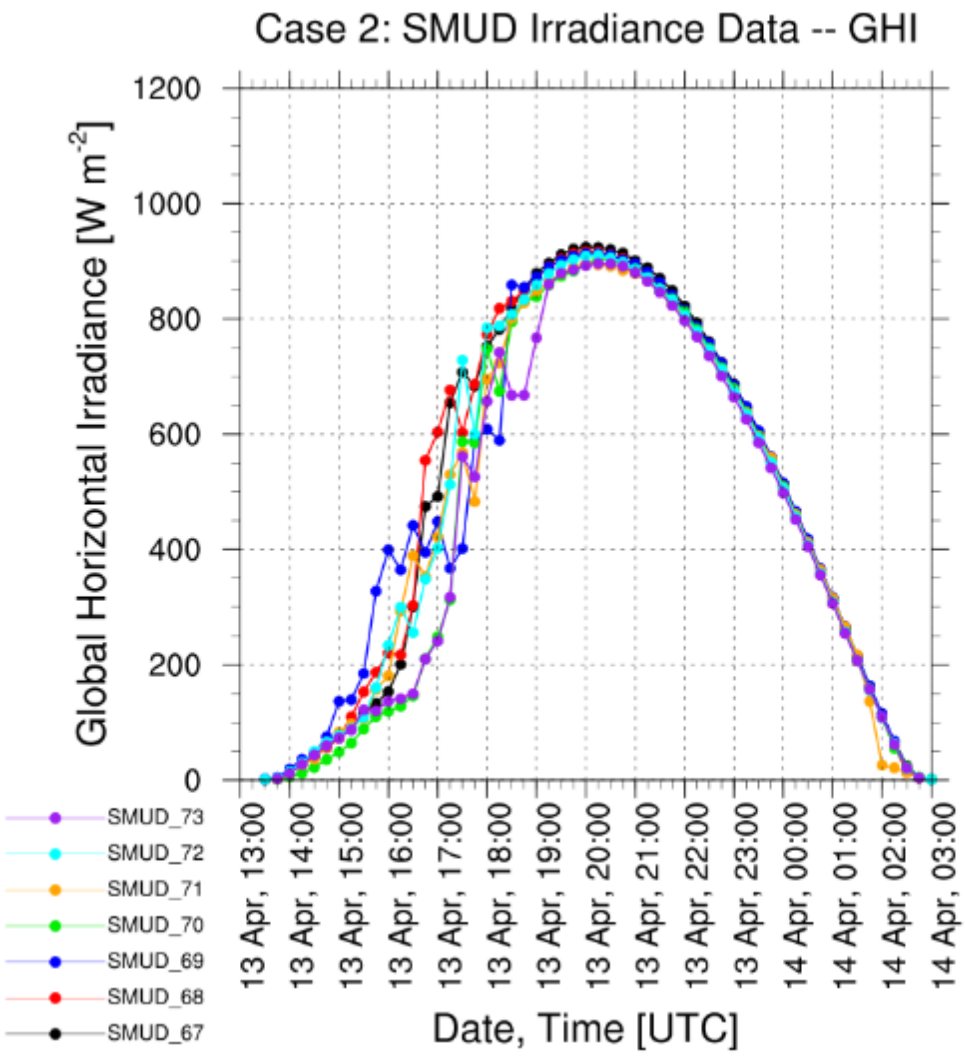


Fig. 2-56. Time series of 15-minute average GHI recorded by SMUD sensors 67-73 for Case 2. The color for each sensor is indicated in the legend.

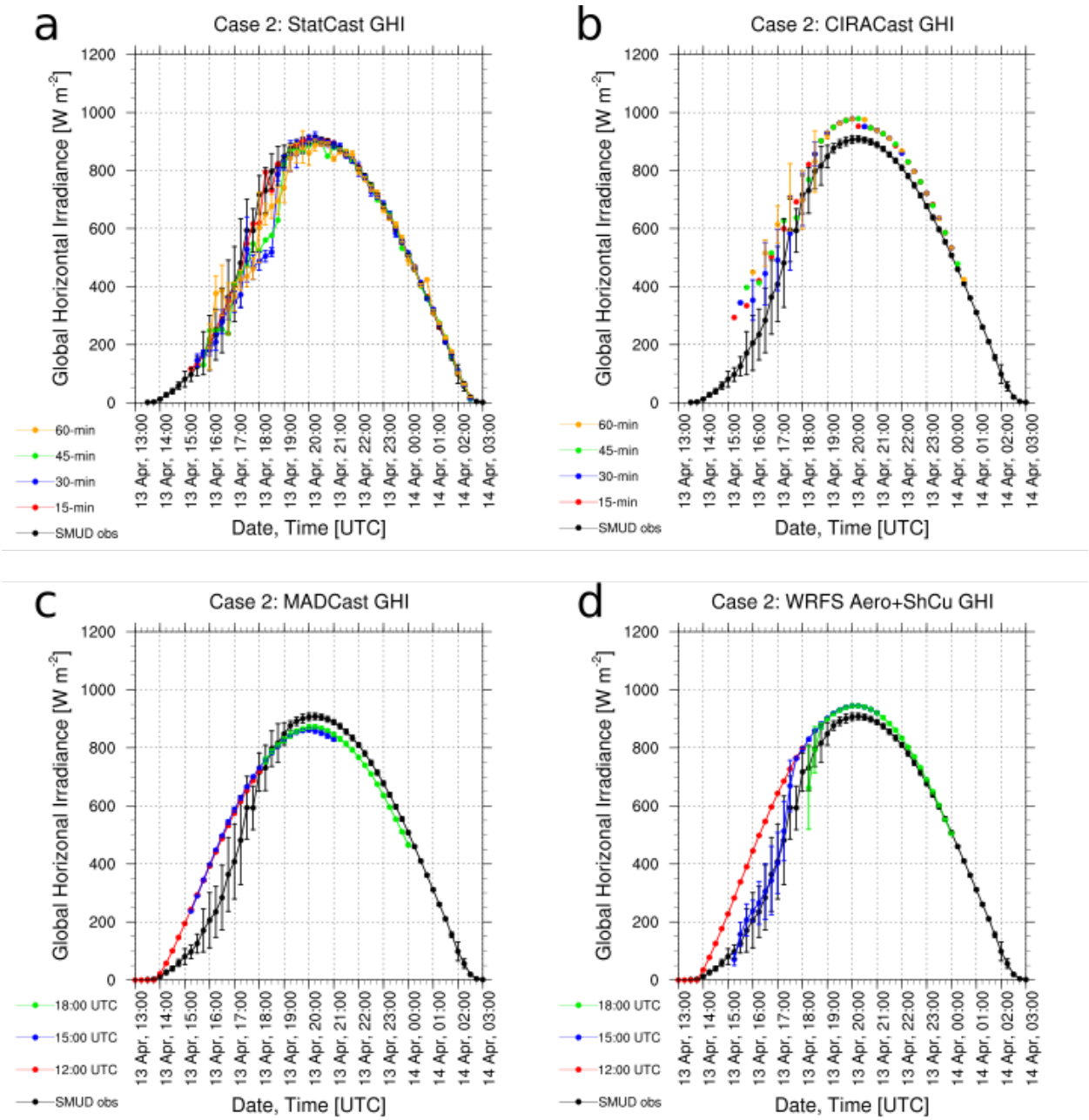


Fig. 2-57. 15-minute GHI predictions and observations for Case 2, averaged over the SMUD 67-73 sensors, for a) StatCast, b) CIRACast, c) MADCast, and d) WRF-Solar™ Aero+ShCu. The error bars for select time series denote ±1 standard deviation across the SMUD sensors.

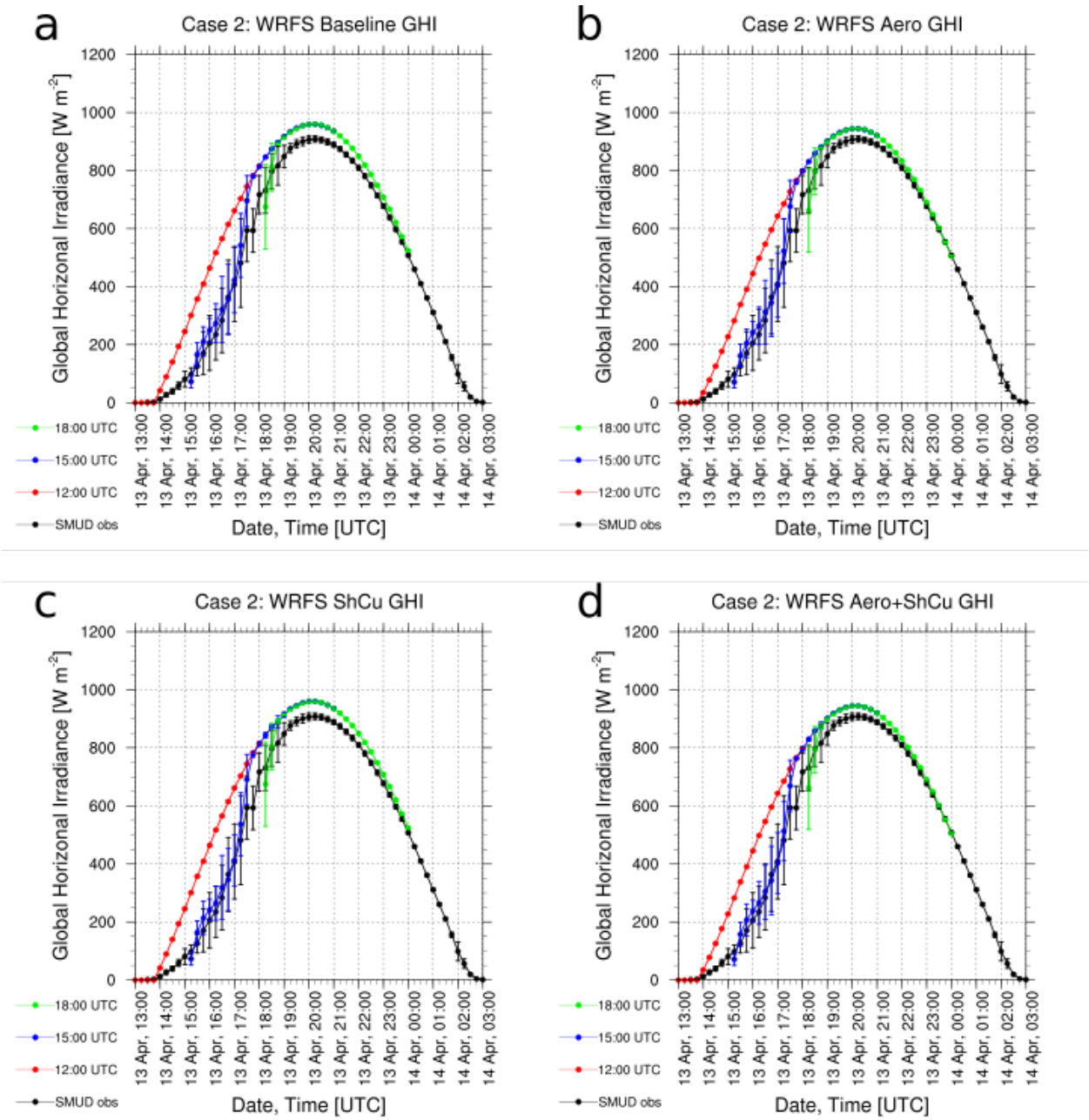


Fig. 2-58. 15-minute GHI predictions and observations for Case 2, averaged over the SMUD 67-73 sensors, for a) WRF-Solar™ Baseline, b) WRF-Solar™ Aero, c) WRF-Solar™ ShCu, and d) WRF-Solar™ Aero+ShCu. The error bars denote ±1 standard deviation across the SMUD sensors.

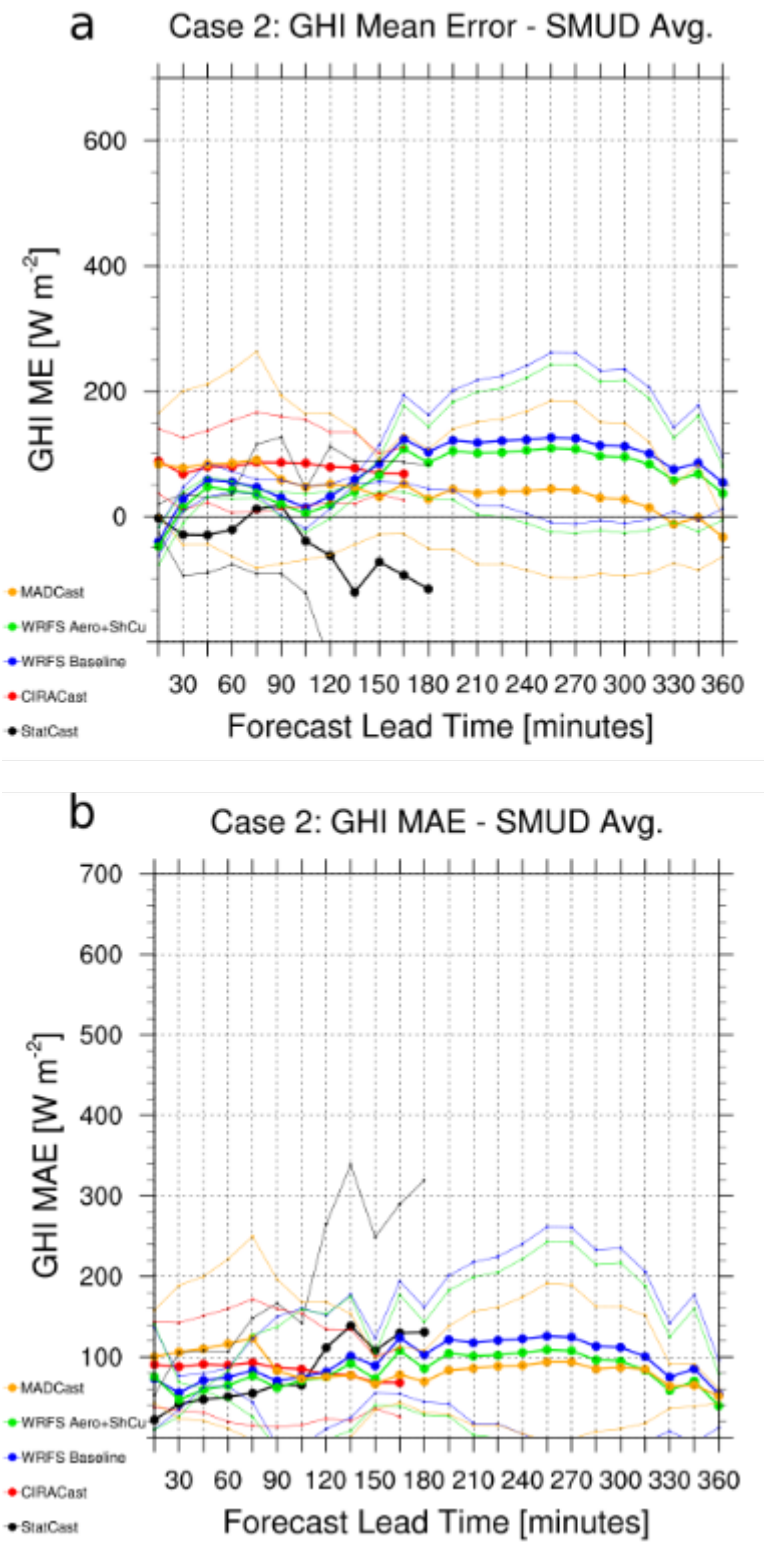


Fig. 2-59. a) Mean error and b) mean absolute error for the nowcasting components for Case 2, as a function of forecast lead time. Thick lines denote average values, while thin lines denote the ±1 standard deviation range. Colors are described in the legends.

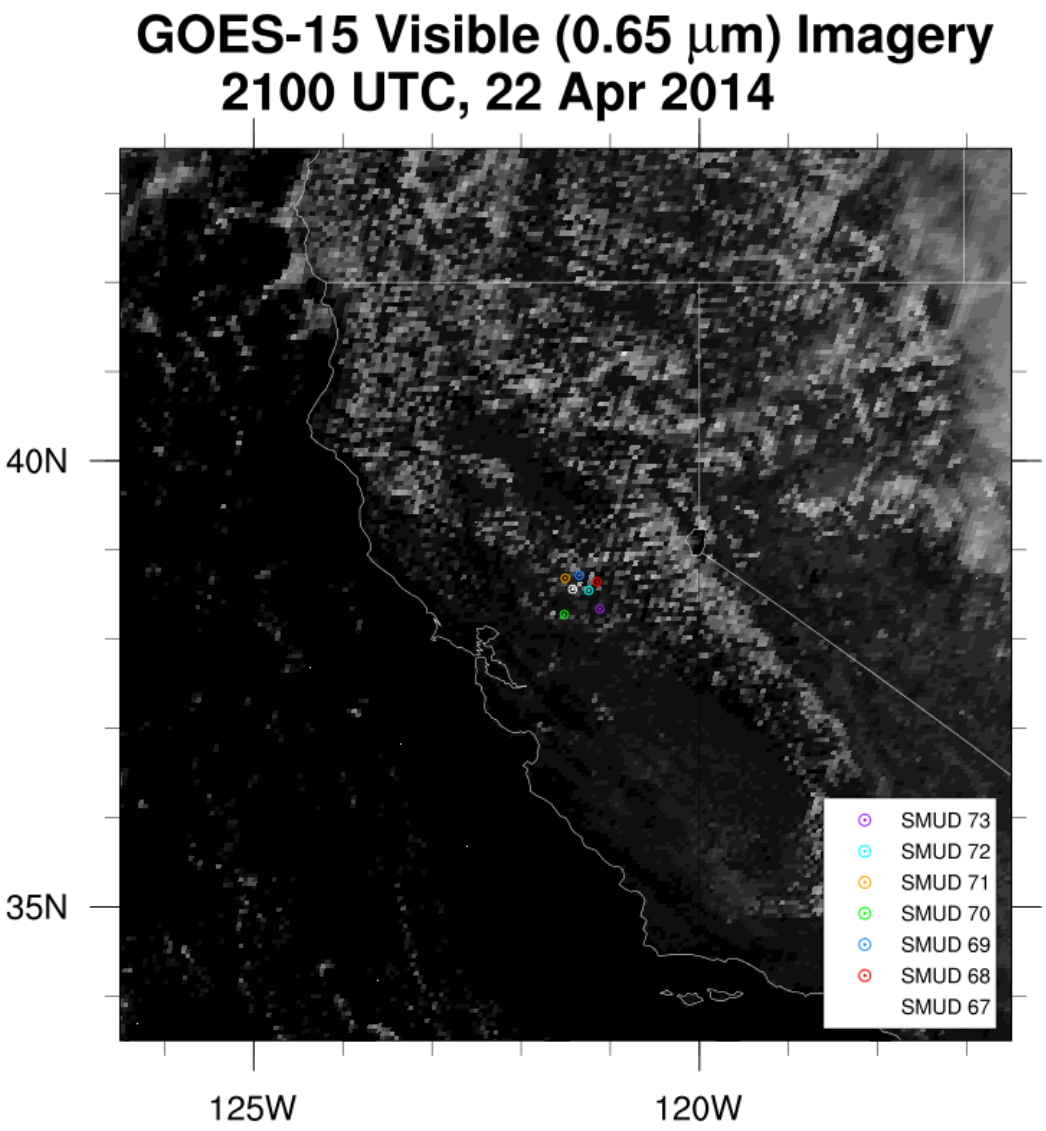


Fig. 2-60. Visible satellite image at 2100 UTC during Case 3. The locations of SMUD sensors 67-73 are indicated by the colors in the legend.

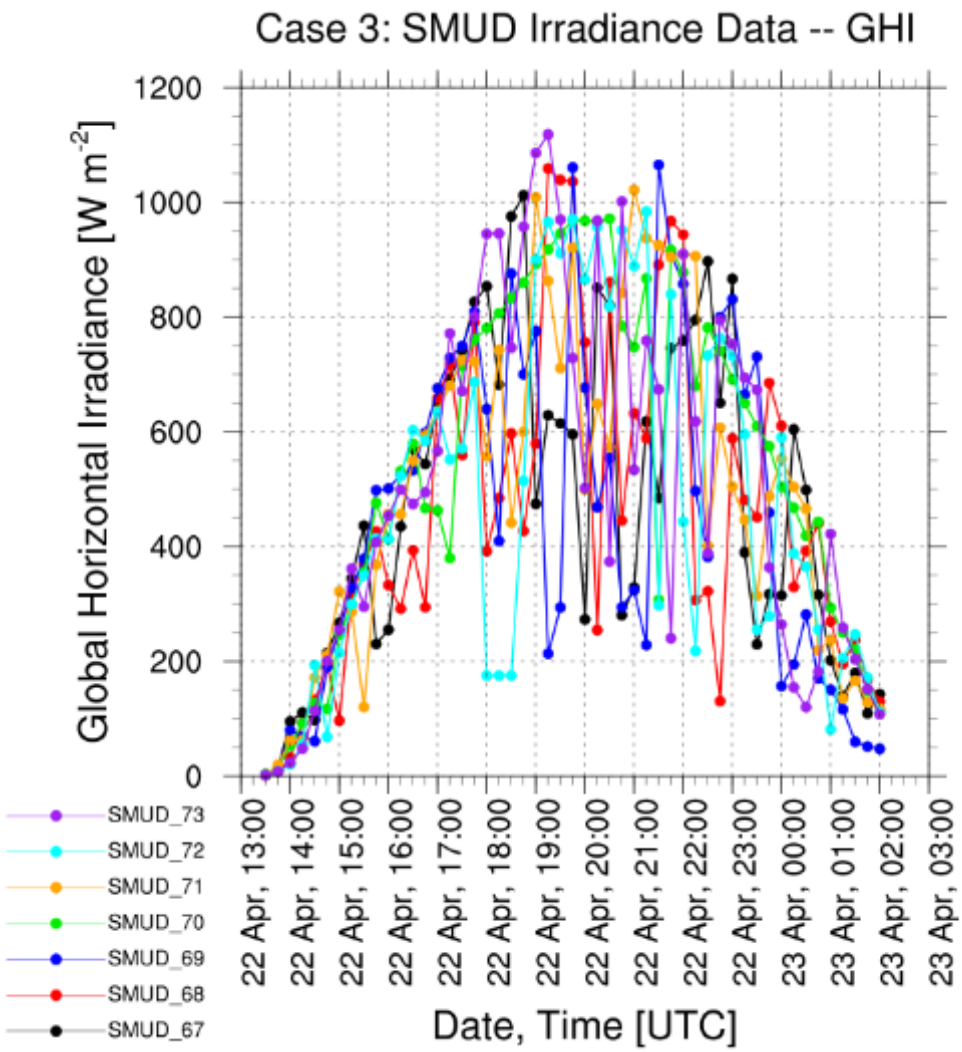


Fig. 2-61. Time series of 15-minute average GHI recorded by SMUD sensors 67-73 for Case 3. The color for each sensor is indicated in the legend.

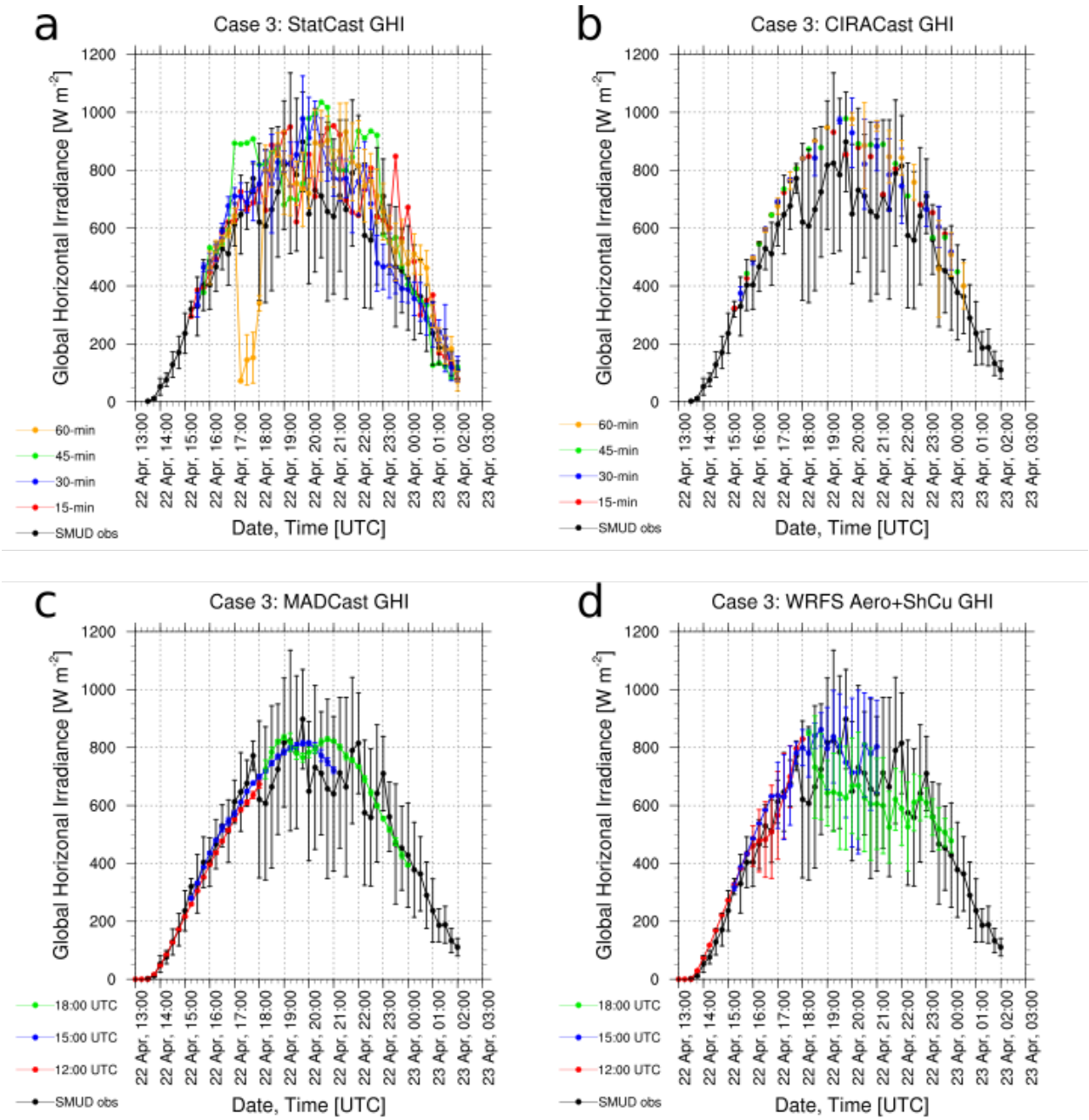


Fig. 2-62. 15-minute GHI predictions and observations for Case 3, averaged over the SMUD 67-73 sensors, for a) StatCast, b) CIRACast, c) MADCast, and d) WRF-Solar™ Aero+ShCu. The error bars for select time series denote ±1 standard deviation across the SMUD sensors.

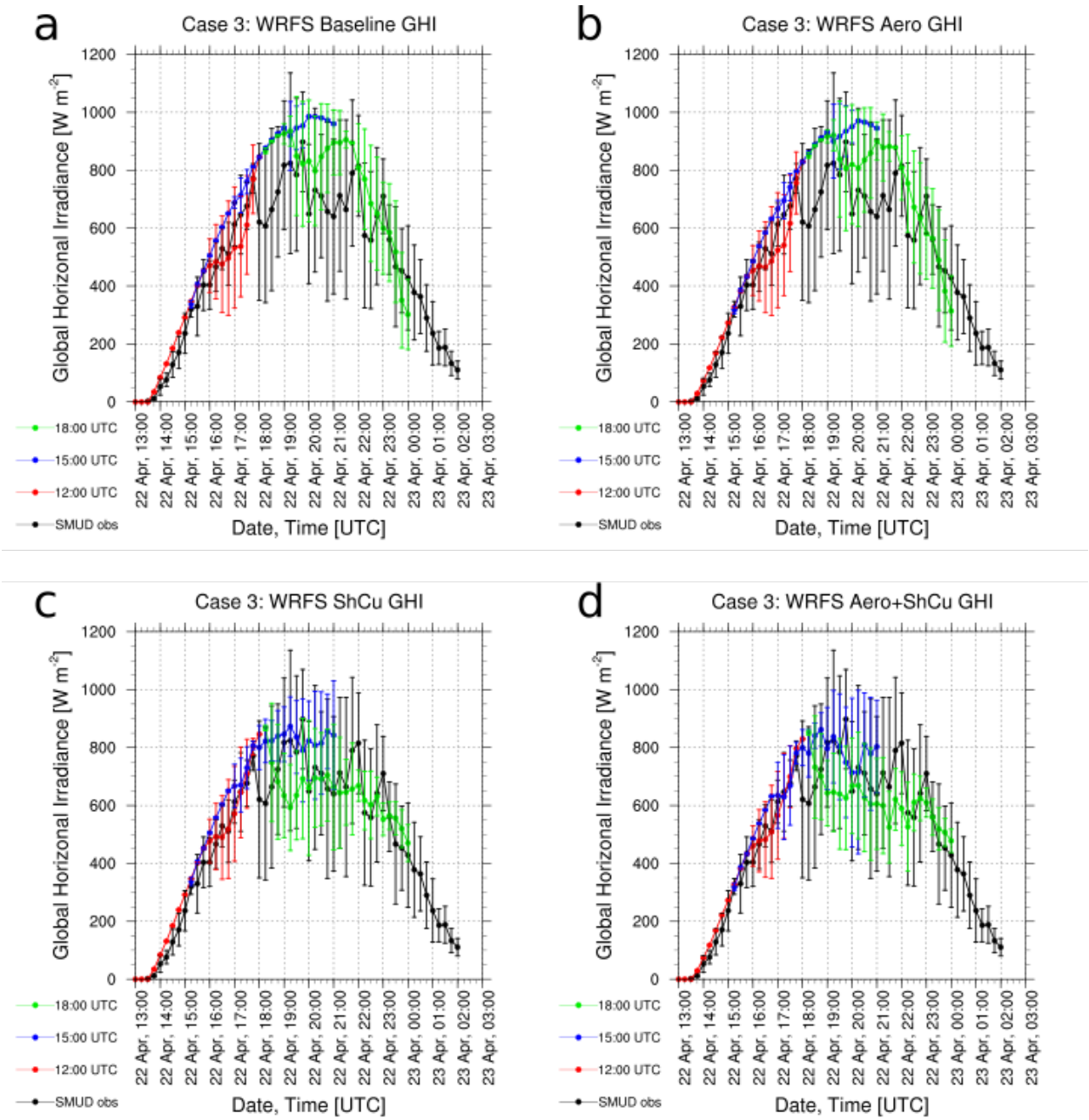


Fig. 2-63. 15-minute GHI predictions and observations for Case 3, averaged over the SMUD 67-73 sensors, for a) WRF-Solar™ Baseline, b) WRF-Solar™ Aero, c) WRF-Solar™ ShCu, and d) WRF-Solar™ Aero+ShCu. The error bars denote ±1 standard deviation across the SMUD sensors.

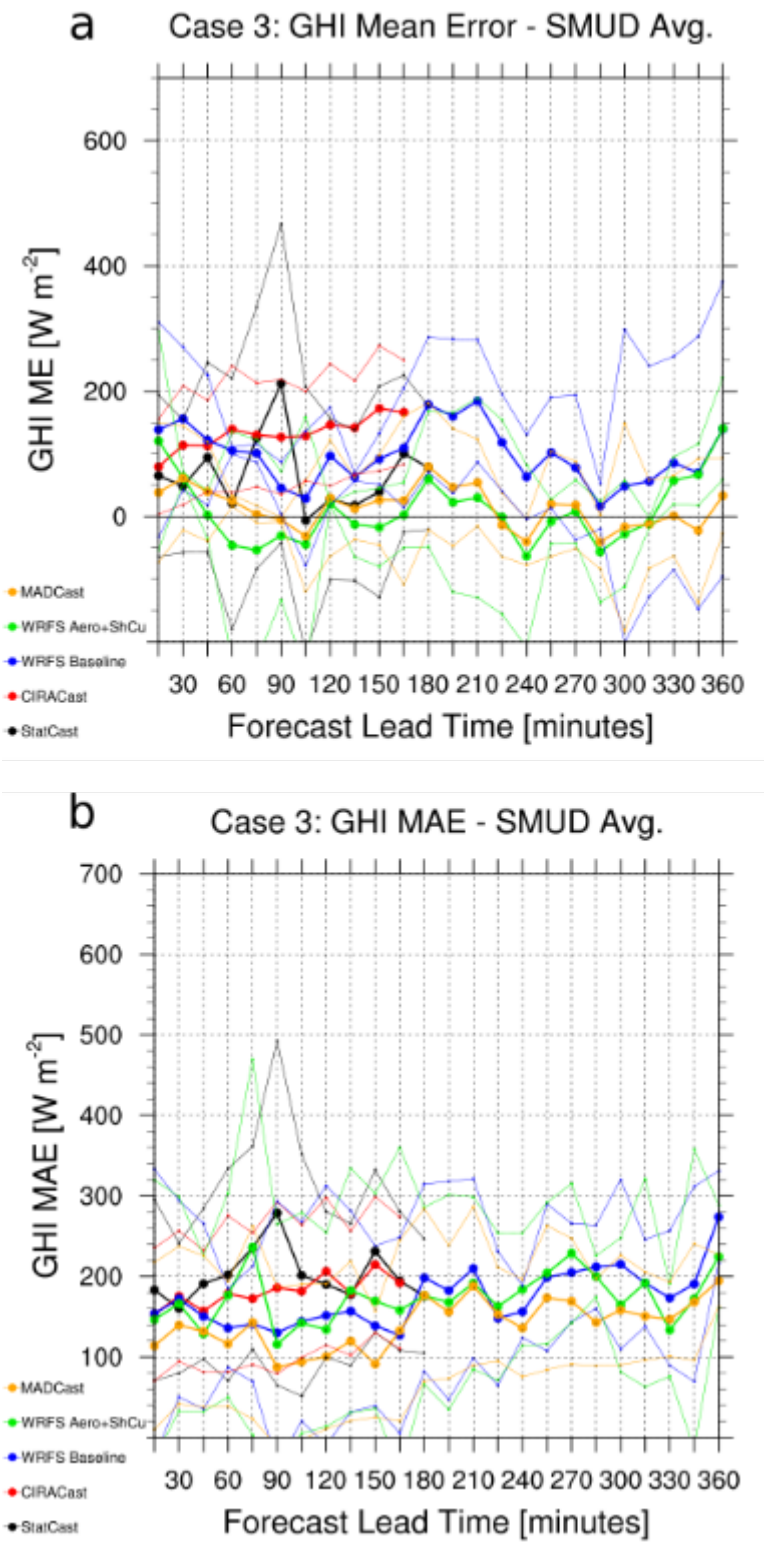


Fig. 2-64. a) Mean error and b) mean absolute error for the nowcasting components for Case 3, as a function of forecast lead time. Thick lines denote average values, while thin lines denote the ±1 standard deviation range. Colors are described in the legends.

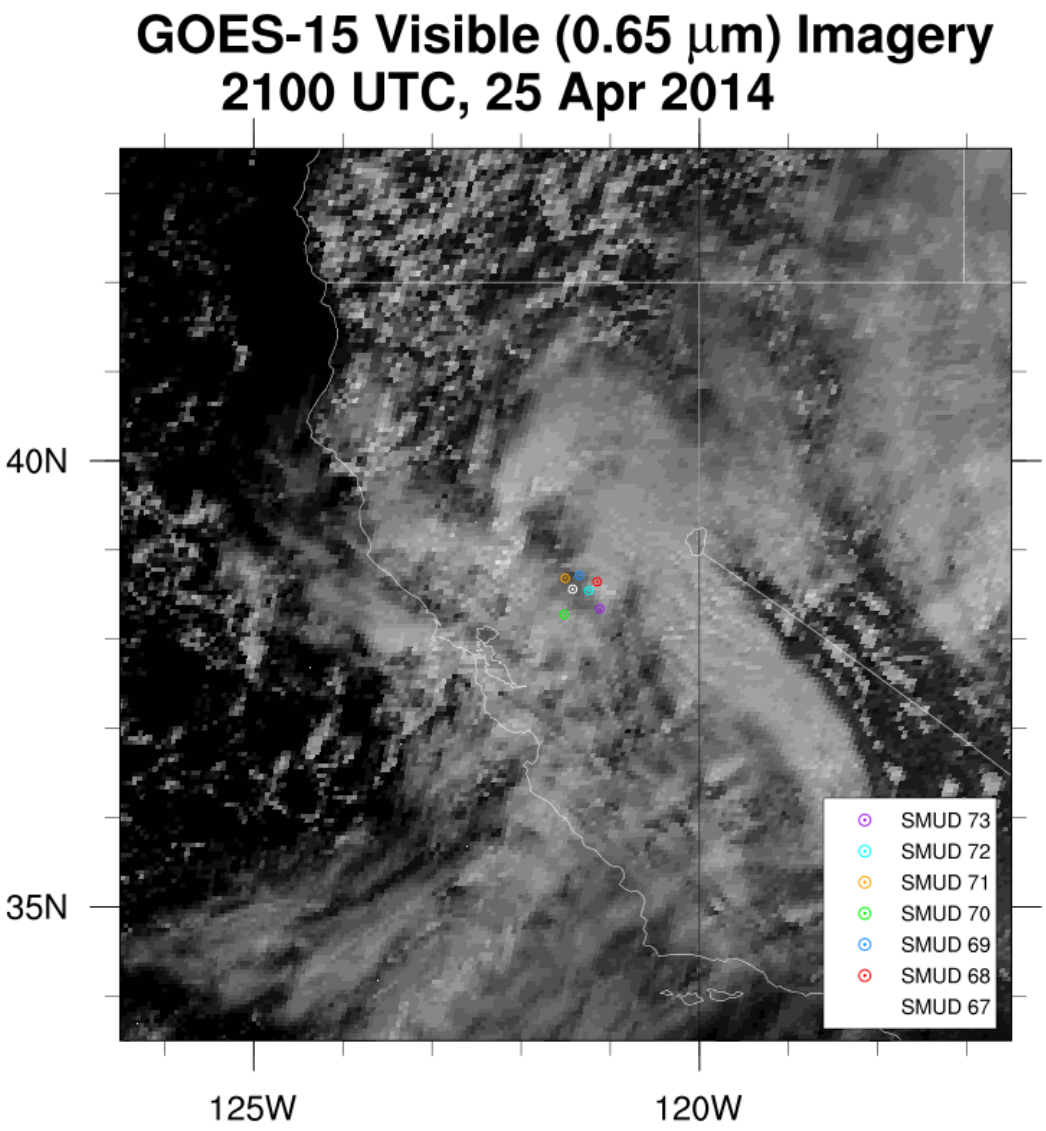


Fig. 2-65. Visible satellite image at 2100 UTC during Case 4. The locations of SMUD sensors 67-73 are indicated by the colors in the legend.

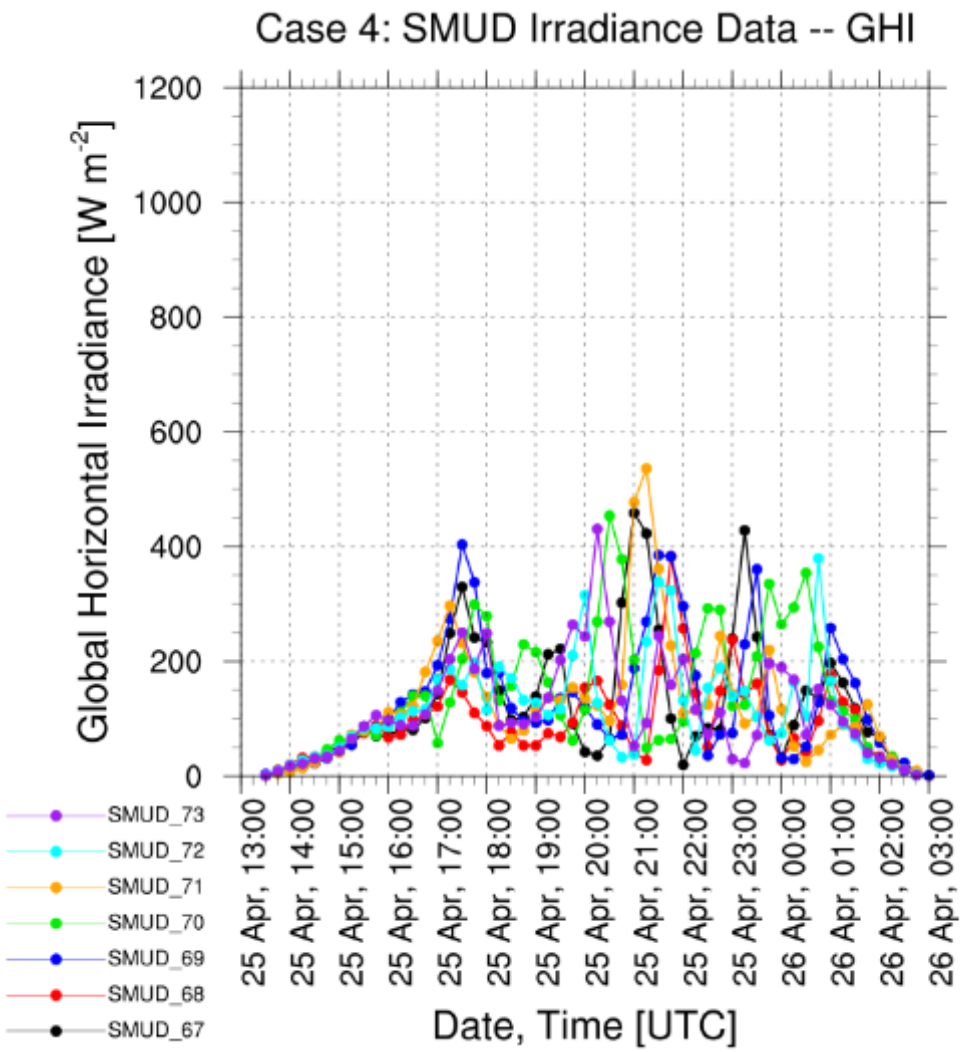


Fig. 2-66. Time series of 15-minute average GHI recorded by SMUD sensors 67-73 for Case 4. The color for each sensor is indicated in the legend.

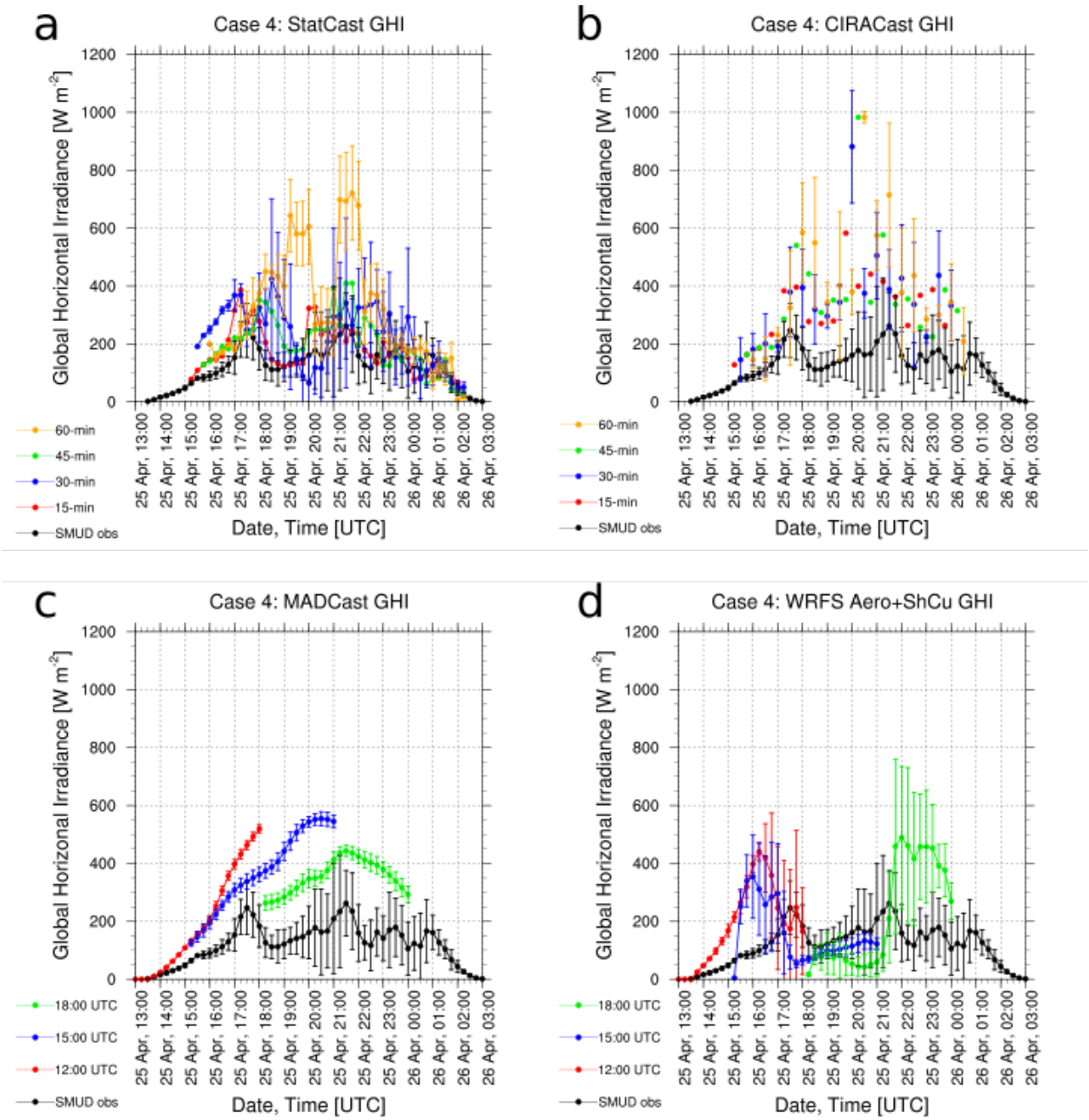


Fig. 2-67. 15-minute GHI predictions and observations for Case 4, averaged over the SMUD 67-73 sensors, for a) StatCast, b) CIRACast, c) MADCast, and d) WRF-Solar™ Aero+ShCu. The error bars for select time series denote ±1 standard deviation across the SMUD sensors.

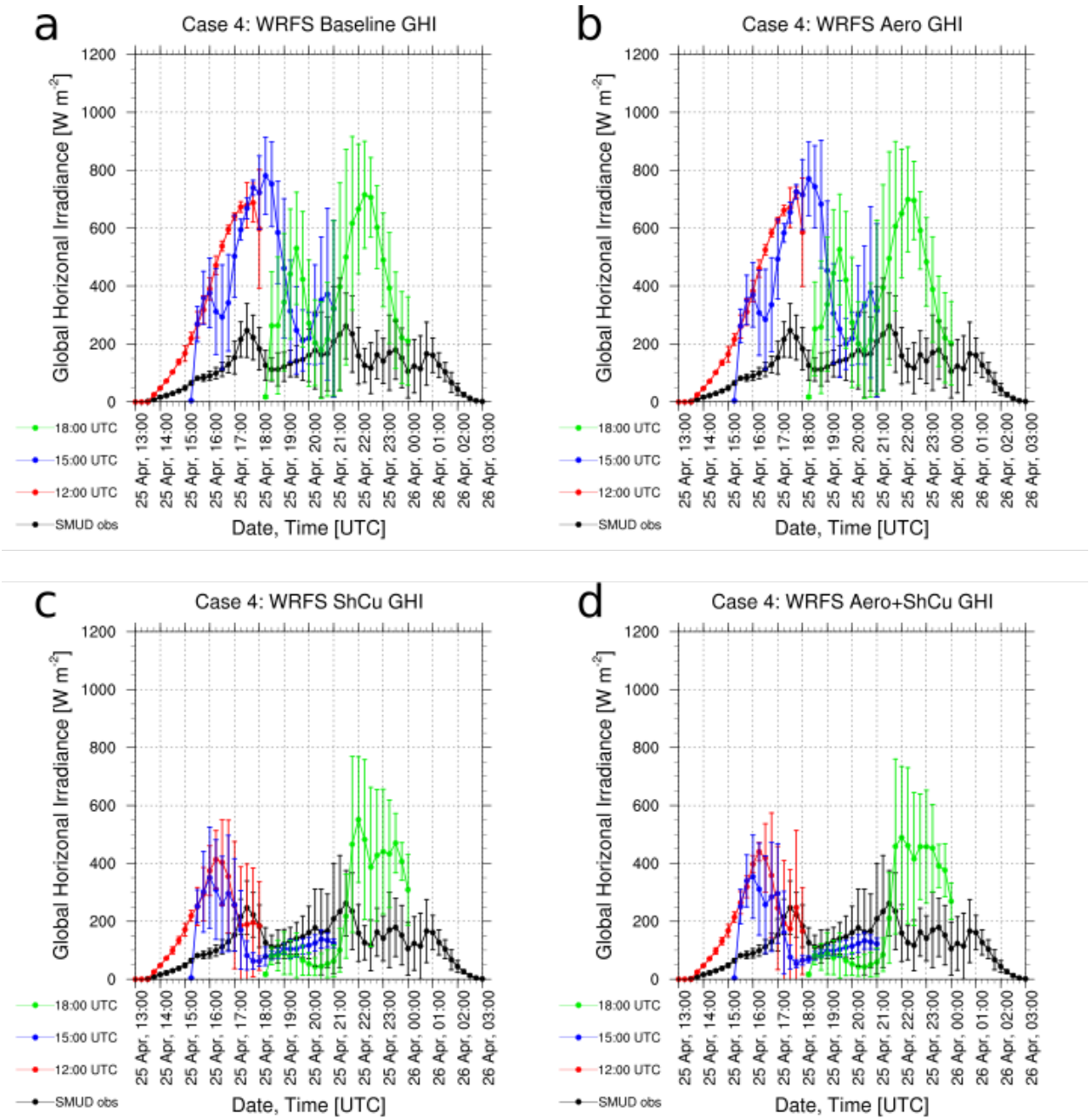


Fig. 2-68. 15-minute GHI predictions and observations for Case 4, averaged over the SMUD 67-73 sensors, for a) WRF-Solar™ Baseline, b) WRF-Solar™ Aero, c) WRF-Solar™ ShCu, and d) WRF-Solar™ Aero+ShCu. The error bars denote ±1 standard deviation across the SMUD sensors.

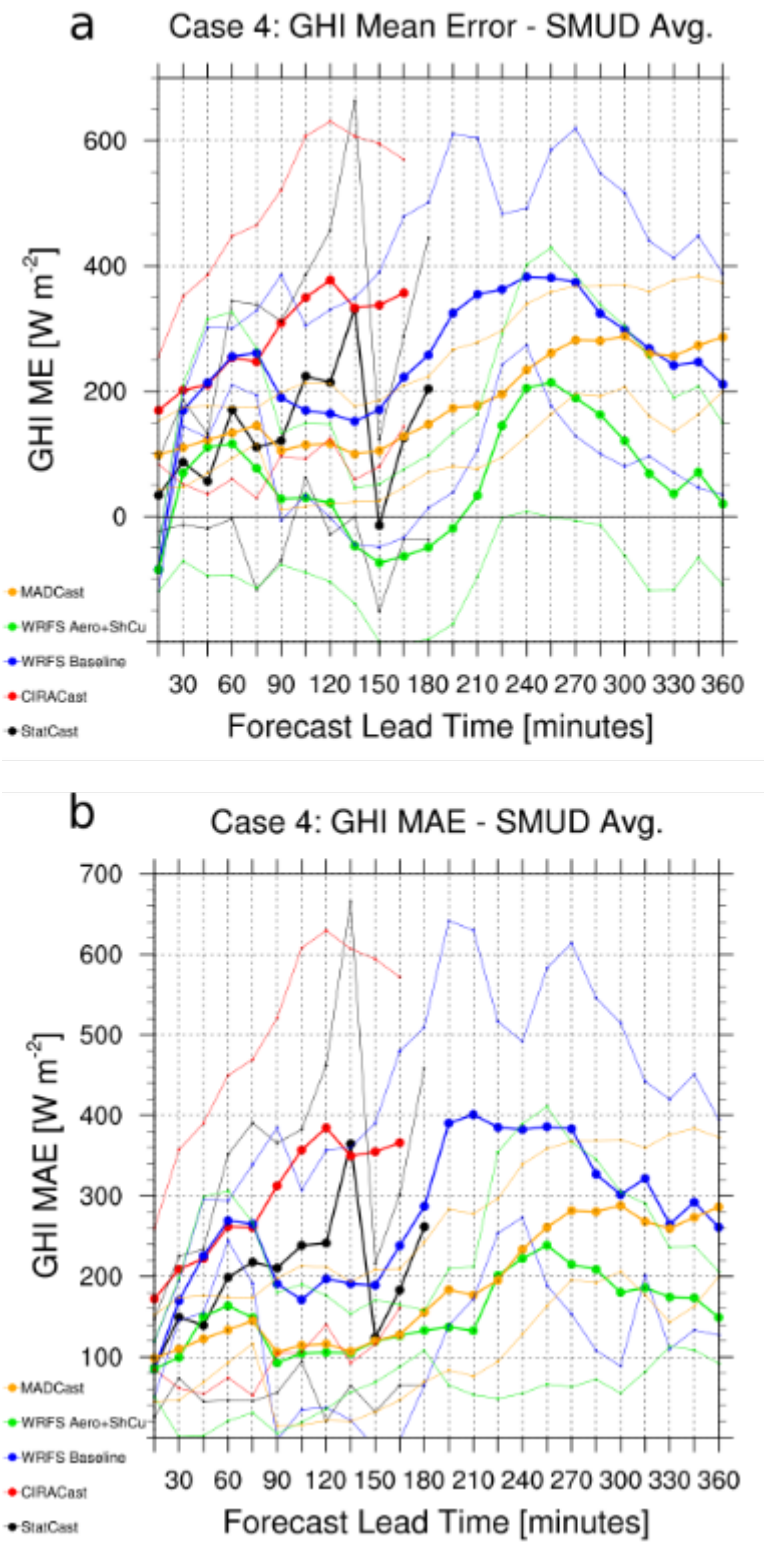


Fig. 2-69. a) Mean error and b) mean absolute error for the nowcasting components for Case 4, as a function of forecast lead time. Thick lines denote average values, while thin lines denote the ±1 standard deviation range. Colors are described in the legends.

3 WRF-SOLAR™ ENHANCEMENTS

WRF-Solar™ is a specific configuration and augmentation of the Weather Research and Forecasting (WRF) model designed for solar energy applications. Recent upgrades to the WRF model contribute to making the model appropriate for solar power forecasting, and include:

- Developments to diagnose internally relevant atmospheric parameters required by the solar industry
- Improved representation of aerosol-radiation feedback
- Incorporation of cloud-aerosol interactions
- Improved representation of cloud-radiation feedback

The WRF-Solar™ developments are presented in section 3.1. A comprehensive characterization of the model performance for forecasting during clear skies is presented in section 3.2. A first indication of the impact in the surface irradiance forecast of the cloud-aerosol interactions is also discussed in section 3.3. The configuration of the quasi-operational forecast is presented and discussed in section 3.4, and the conclusions are presented in section 3.5.

Much of the material in this chapter serves as the foundation for Jiménez et al. (2016a). The interested reader is referred there for some additional details.

3.1 THE WRF-SOLAR™ MODEL: ENHANCEMENTS TO THE STANDARD WRF MODEL

WRF-Solar™ is a specific augmentation of the Advanced Research WRF (WRF-ARW) model (Skamarock et al. 2008) designed to provide an improved NWP tool for solar energy applications. We call the WRF lacking solar augmentation the "standard WRF" unless otherwise noted. In this section we document the WRF-Solar™ additions to WRF-ARW version 3.6, which was publicly released in April 2014. These include making DNI and DIF available at a temporal frequency limited only by the model time step, and several changes to model physics that account for feedbacks between aerosols, solar irradiance, and clouds. These mechanisms are illustrated in a conceptual diagram in Figure 3-1.

Figure 3-1 also shows the different components of the irradiance. DNI, or direct normal irradiance, is the irradiance received per unit of area over a surface perpendicular to incoming rays from the sun. DIF, or diffuse irradiance, is the amount of surface irradiance that has been scattered by the atmosphere. GHI, or global horizontal irradiance, is the total amount of shortwave irradiance received by a horizontal surface at ground level, and combines DIF and the component of DNI perpendicular to the horizontal surface.

3.1.1 Physical Enhancements for Solar Energy Applications

Many solar applications require direct and diffuse solar irradiance components in addition to GHI. For example, if the aim is to calculate the shortwave irradiance that impinges onto the plane of a solar panel, the typical computational method includes (i) projecting DNI onto the direction normal to the plane of the panel; (ii) adding a fraction of DIF evaluated from a sky view factor (the fraction of the total sky hemisphere that is visible from a point in the panel); and (iii) adding a fraction of

GHI resulting from surface reflectance onto the tilted plane of the panel. An additional application is production from concentrated solar plants, for which DNI is the fuel responsible for the energy production (DIF cannot be concentrated). WRF-Solar™ addresses the need for separate direct and diffuse components by making them available in the WRF output, and also in auxiliary output files at arbitrary time intervals as short as the time step length of the model.

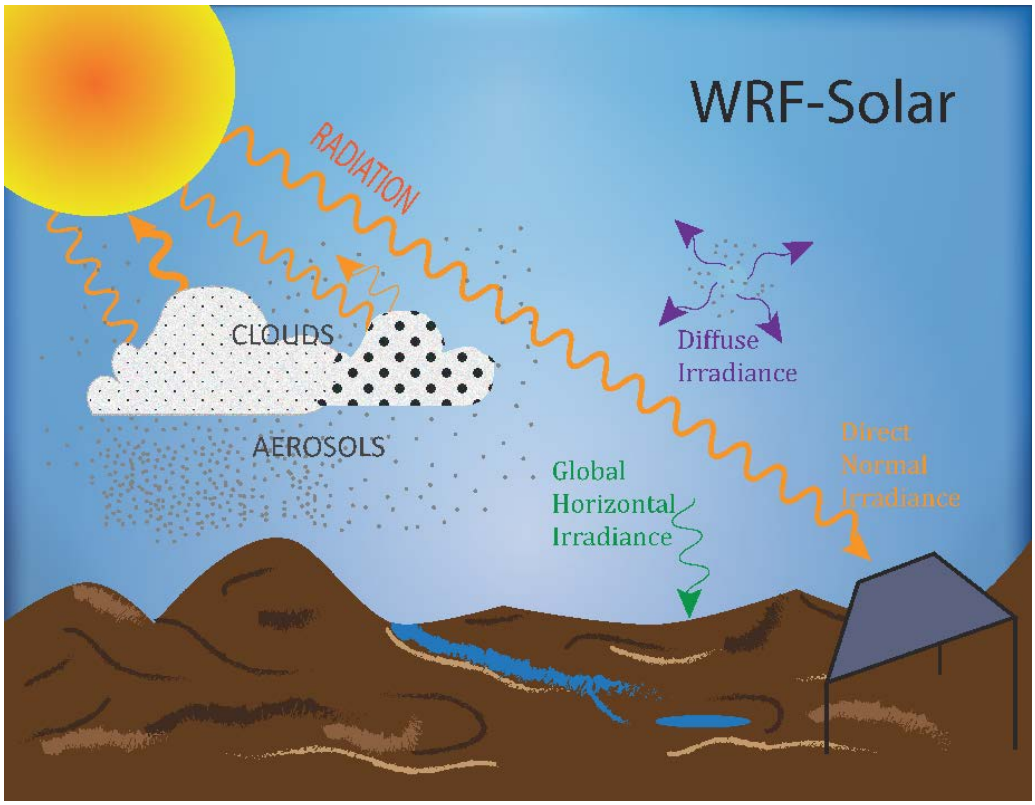


Figure 3-1. Sketch representing the physical processes that WRF-Solar™ improves. The different components of the radiation are indicated.

At least two of the existing shortwave parameterizations in WRF calculate the direct and diffuse radiative transfer equations (e.g. the Goddard scheme [Chou and Suarez 1999] and the Rapid Radiative Transfer Model for Global models scheme [RRTMG; Iacono et al. 2008]). In these cases, WRF-Solar™ adds the surface irradiance components to the model output. Other shortwave schemes do not explicitly solve for the direct and diffuse components, but provide GHI at the surface. Examples are the Dudhia scheme (Dudhia 1989), and an older Goddard scheme (Chou 1992). The advantage of these simpler parameterizations is that they are faster at solving the radiative transfer equation. For these parameterizations, the direct and diffuse components are estimated from GHI using a regression model trained on world-wide observations (Ruiz-Arias et al. 2010). The optical air mass and the clearness index modulate the regression from GHI to DIF.

Improvements to the solar position algorithm used in previous versions of WRF, particularly the equation of time (EOT; Muller 1995), are now included. Deviations associated with the eccentricity of the Earth's orbit and the obliquity of the Earth previously caused irradiance leads

and lags of up to 16 mins, depending on the day of the year. The standard WRF, before version 3.5.1, neglects the EOT correction, which is normally inconsequential when running the radiation scheme approximately every half hour. Solar energy applications demand more frequent calls to the radiation scheme, and the leads/lags then introduce non-negligible error. Simulations performed during a day close to a maximum lag reduced the RMSE in GHI by 31%.

WRF-Solar™ aims to provide irradiance components at every time step while avoiding unphysical discontinuities. Typically, the computational time of a model integration step calling the radiation parameterization is an order of magnitude longer than an integration step that skips it, and the radiative computations are done at most every ten model steps. In between calls, the standard WRF assumes that irradiance is constant, introducing temporal discontinuities (steps) in irradiance values computed at the surface. In WRF-Solar™ two different methods have been implemented. The first one is based in a computationally efficient algorithm that interpolates the irradiance between successive calls to the radiation scheme considering only the change in the actual solar position, and assuming the cloud extinction effect remains fixed to the latest computed value (a smart persistence approach). The second one was built by NREL as part of this project: Fast All-Sky Radiation Model (FARMS, Xie et al. 2016) for solar applications, which uses a simplified clear-sky radiative transfer model, REST2, and the simulated cloud transmittances and reflectances from RRTM with a sixteen-stream discrete ordinates radiative transfer (DISORT). FARMS therefore accounts for changes in hydrometeor content in between calls to the radiation parameterization.

A particularly useful augmentation is irradiance output at time intervals limited only by the model time step. High-frequency time series of surface irradiance components are helpful when modeling solar ramps (i.e. an abrupt change in the surface irradiance). Figure 3-2 illustrates a solar ramp event accurately predicted by WRF-Solar™. The first part of the day features cloudy skies that block the DNI so that GHI equals the DIF. WRF-Solar™ predicts this with only a slight overestimation. Around 1700 UTC, scattered cloud conditions appear and clear-sky conditions alternate with clouds. WRF-Solar™ then simulates the different components of the irradiance, but with less high-frequency variability because the modeled irradiance is more representative of a temporal (and spatial) average.

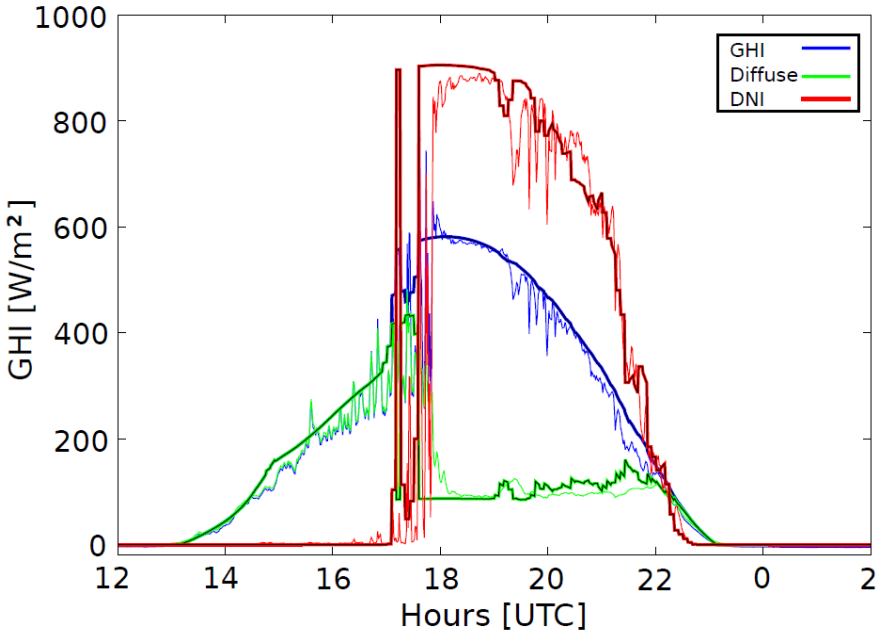


Figure 3-2. Observed and simulated surface irradiance components during a solar ramp. The simulated irradiances are highlighted with a thick solid line on top of the thin colored line that represents the observed irradiances (see legend).

3.1.2 Aerosol-Radiation Feedback: Aerosol Direct Effect

Standard WRF simulations neglect the primary effects of atmospheric aerosols. The radiative impact of aerosols on GHI is relatively small, explaining the lack of attention to it for most meteorological applications. It has been recognized that highly polluted conditions can lead to biases, however (Barbaro 2015).

The standard WRF model since version 3.5 has been capable of using a climatology-based aerosol parameterization developed at the European Center for Medium Range Weather Forecasts (ECMWF). The parameterization uses model results from Tegen et al. (1997) to derive monthly climatological means of the aerosol optical properties. Different models were used to simulate the transport of soil dust (Tegen and Fung 1995), sea salt, sulfate (Chin et al. 1996), and carbonaceous aerosols (Liousse et al. 1996). The 3-D aerosol optical depth (AOD) for each species is computed on a grid with horizontal spacing of 5º longitude by 4º latitude, and 12 pressure levels from 959 hPa to 20 hPa. By activating the parameterization, which by default is turned off, the RRTMG shortwave parameterization infers the AOD from the combined effects of the species with their own assumed properties, and uses them to solve for diffuse and direct irradiance. The AOD values remain constant, or change slowly, throughout the simulation period. This climatology and simple parameterization of aerosol effects are expected to allow a better representation of the direct and diffuse radiation components; no known previous attempts have been made to assess its performance before this work because the direct and diffuse irradiances were not previously available as output. Monthly mean aerosols are potentially useful, but do not exploit current observations from satellites, surface networks, or from faster evolving analysis/re-analysis products.

WRF-Solar™ allows the user to impose evolving aerosol optical properties in the simulation (Ruiz-Arias et al. 2014), and considers humidity effects, to account for the aerosol direct effect. Parameterization of the effects of the aerosol optical properties on the radiation has been added to the Goddard (Chou and Suarez 1999; Shi 2010) and the RRTMG (Iacono et al. 2008) shortwave radiation codes. The parameterization requires the total AOD at 550 nm (visible), and specification of the type of predominant aerosol. Knowledge of the predominant aerosol type allows for estimation of the remaining aerosol optical properties, including the single-scattering albedo and the asymmetry factor. It also permits modeling the spectral variability with estimations of the Ångstrom exponent. The user can optionally provide the single-scattering albedo, asymmetry factor, and Ångstrom exponent rather than allowing the parameterization to infer them based on the predominant aerosol. Ruiz-Arias et al. (2013) present a comparison of the parameterization to observations, which indicates that the parameterization produces accurate estimations of surface irradiance given accurate aerosol optical properties.

3.1.3 Cloud-Aerosol Feedbacks

Prior to version 3.6, the standard WRF lacked representation of aerosol interactions with cloud processes. To enable cloud-aerosol feedbacks and maintain computational affordability for operational applications, WRF-Solar™ uses the simplified representation of the aerosol interaction with the Thompson microphysics scheme (Thompson and Eidhammer 2014). Aerosol species are classified into hygroscopic (attracting water) and non-hygroscopic (avoiding water) aerosols. The two species are currently initialized either from 1) a 3-D monthly climatology of the aerosol number concentrations generated from the Goddard Chemistry Aerosol Radiation and Transport (GOCART) model (Ginoux et al. 2001; Colarco et al. 2010); or from 2) instantaneous aerosols concentrations from chemistry models. The surface emission flux is represented by a variable lower boundary condition based upon the starting aerosol conditions and an assumed mean surface wind.

As of the time of this writing, the combination of the RRTMG radiation scheme and the Thompson and Eidhammer (2014) microphysics scheme fully incorporates the first and second aerosol indirect effects (Twomey 1974; Albrecht 1989). This adaptation of the standard WRF was aimed at NWP and WRF-Solar™ applications, and resulted in minimal computational cost increase compared to running a full chemistry model, with a 16% increase in simulation time compared to the previous Thompson et al. (2008) scheme. Simulations of a large winter cyclone indicate the necessity of quantifying the benefits of the parameterization using long term simulations to robustly quantify model forecast errors and observational uncertainty (Thompson and Eidhammer 2014).

3.1.4 Cloud-Radiation Feedbacks

Three further improvements in WRF close the aerosol-cloud-radiation feedback. First, consistency of the cloud particle distributions in the microphysics and radiation schemes is enforced. Historically, the cloud particle size for shortwave radiation calculations is imposed (i.e., the cloud effective radius is forced to remain constant) internal to a particular radiation scheme (Stensrud 2007). This implicitly assumes that all clouds are homogeneous in terms of their radii. To provide a more physically consistent representation of the cloud-radiation feedbacks, WRF-Solar™ adopts the novel approach of passing the effective radius of the cloud droplets, ice and snow particles

from the microphysics to the radiation (both shortwave and longwave) parameterization schemes (Thompson and Eidhammer 2014; Thompson et al. 2015). This affects the cloud albedo and activates the aerosol indirect effects.

Second, the AOD from combined hygroscopic and non-hygroscopic aerosol number concentrations in the aerosol-aware microphysics can be passed to the radiation scheme. The aerosols are advected with the model dynamics and, at a time step corresponding to a call to radiation physics, the extinction coefficient is computed and passed to the aerosol parameterization (Ruiz-Arias et al. 2014) for radiation. In this way, WRF-Solar™ provides a fully coupled representation of the aerosol-cloud-radiation system (Figure 3-1).

The last development upgrades the feedbacks that sub-grid scale clouds produce in the shortwave irradiance. This effect is implemented in the shallow cumulus parameterization. The standard WRF model does not typically provide a cloud fraction from its shallow sub-grid convection parameterization options. WRF-Solar™ includes one shallow cumulus scheme previously implemented in the PSU-NCAR Fifth-generation Mesoscale Model (MM5) (Deng et al. 2003), which provides a cloud fraction for radiation. The Deng et al. (2003, 2014) shallow convection scheme is a mass-flux based scheme. It includes a cloud entraining/detraining model to represent updrafts, and it is triggered by factors including planetary boundary layer depth and turbulent kinetic energy (TKE). It uses a hybrid closure combining TKE and convective available potential energy, depending on the updraft depth. In addition to the updraft formulation, the scheme also contains two predictive equations for cloud fraction and cloud liquid/ice water content for neutrally-buoyant clouds (inactive clouds detrained from the active updraft core). Deng et al. (2014) shows that the scheme is able to produce reasonable cloud fractions, and reduce surface temperature bias. The benefits of the scheme for solar-energy applications are discussed on section 2.5

3.2 WRF-Solar™ Assessment under Clear Sky Conditions

3.2.1 Experiment Details

Six WRF-Solar™ experiments were completed to measure the importance of the aerosol direct effect on predictions of surface irradiance, and to investigate the sensitivity to aerosol optical property source and treatment. The first experiment (NO-AEROSOL) lacks any aerosol treatment. The second experiment (ECMWF-CLIM) activates the ECMWF monthly climatology. The third experiment (SCS-CLIM) adds two further potential improvements: a monthly aerosol dataset covering North America at high spatial resolution (0.05º latitude by 0.05º longitude) developed by Solar Consulting Services (SCS), and a more sophisticated aerosol property parameterization for the aerosol direct effect. The fourth experiment (GOCART-CLIM) uses the GOCART climatology to activate the feedback to radiation. GOCART-CLIM is the only experiment with aerosol advection, and the only experiment where the Thompson microphysics aerosols are also used for the direct effect. SCS-CLIM and GOCART-CLIM impose the total AOD at 550 nm using data from models that explicitly predict evolving atmospheric chemistry. The fifth experiment (MACC-AOD) uses data from the ECMWF Monitoring Atmospheric Composition and Climate (MACC) reanalysis (Inness et al. 2013). MACC AOD is available globally every three hours at 1.115º latitude by 1.115º longitude, and here is linearly interpolated to hourly input to the WRF-Solar™.

The total AOD is a forecast variable starting from a 0000 UTC analysis and extending to 24 hours. The last (sixth) experiment (GEOS5-AOD) uses NASA's Goddard Earth Observing System model version 5 (GEOS-5; Rienecker et al. 2000) analysis. The GEOS-5 product is global and available every three hours, and is also interpolated to hourly AOD. The horizontal resolution is greater than the MACC, at 0.5º latitude by 0.65º longitude. The experiments prescribing the AOD via the aerosol parameterization of WRF-Solar™ (i.e. SCS-CLIM, MACC-AOD, GEOS5-AOD) impose a rural-type (Shettle and Fenn 1979) predominant aerosol.

The high-resolution aerosol dataset for SCS-CLIM is composed of the AOD at 550 nm and the Ångstrom exponent for each month of the period 2000-2014, and of the mean monthly single scattering albedo. The methodology is similar to that used previously by other authors (Kinne et al. 2003, 2006; Kinne 2009; Kinne et al. 2013; Pappas et al. 2013) to develop climatologies including AeroCom, Hamburg (Pappas et al. 2013), and Max-Planck-Institute Aerosol Climatology (MAC; Kinne et al. (2013)). They are composites based on remote-sensing observations (spaceborne MODIS spectrometers and ground-based AERONET sunphotometers) and predictions from various aerosol transport models. Here, special attention was devoted to large parts of western North America where the Dark Target (DT) MODIS algorithm was found to considerably overestimate AOD. Corrections to remove this bias over high-albedo areas (delineated from MODIS albedo data) were derived from a regional comparison with AERONET Level-2 data, Deep Blue MODIS data, MACC reanalysis, and the Hamburg climatology. The coarse spatial resolution (1º by 1º) of the original monthly gridded data was improved by combining a bi-linear interpolation at sea level, and a correction to account for topographic effects on the vertical aerosol profile, using an exponential scale height of 2.5 km. The Ångstrom exponent was obtained in a similar way, but without topographic correction. Among other differences with the MAC, the SCS dataset does not incorporate any specific AERONET data, and does not force local agreement with ground truth.

Figure 3-3 compares the spatial variation of the AOD at 550 nm as obtained from monthly MODIS Terra (DT algorithm v5.1) and the SCS dataset at their respective spatial resolutions. This comparison is made for July in a climatological sense, i.e. considering the long-term 2000-2014 mean. In Figure 3-3c, the time series of the monthly AOD obtained by AERONET (after appropriate spectral correction) is compared to the MODIS-Terra and the SCS data for Maricopa, Arizona. Maricopa is one of the areas where the MODIS DT retrievals show a notable high bias. The SCS climatology reduces the high AOD bias, providing a better comparison with the ground observations.

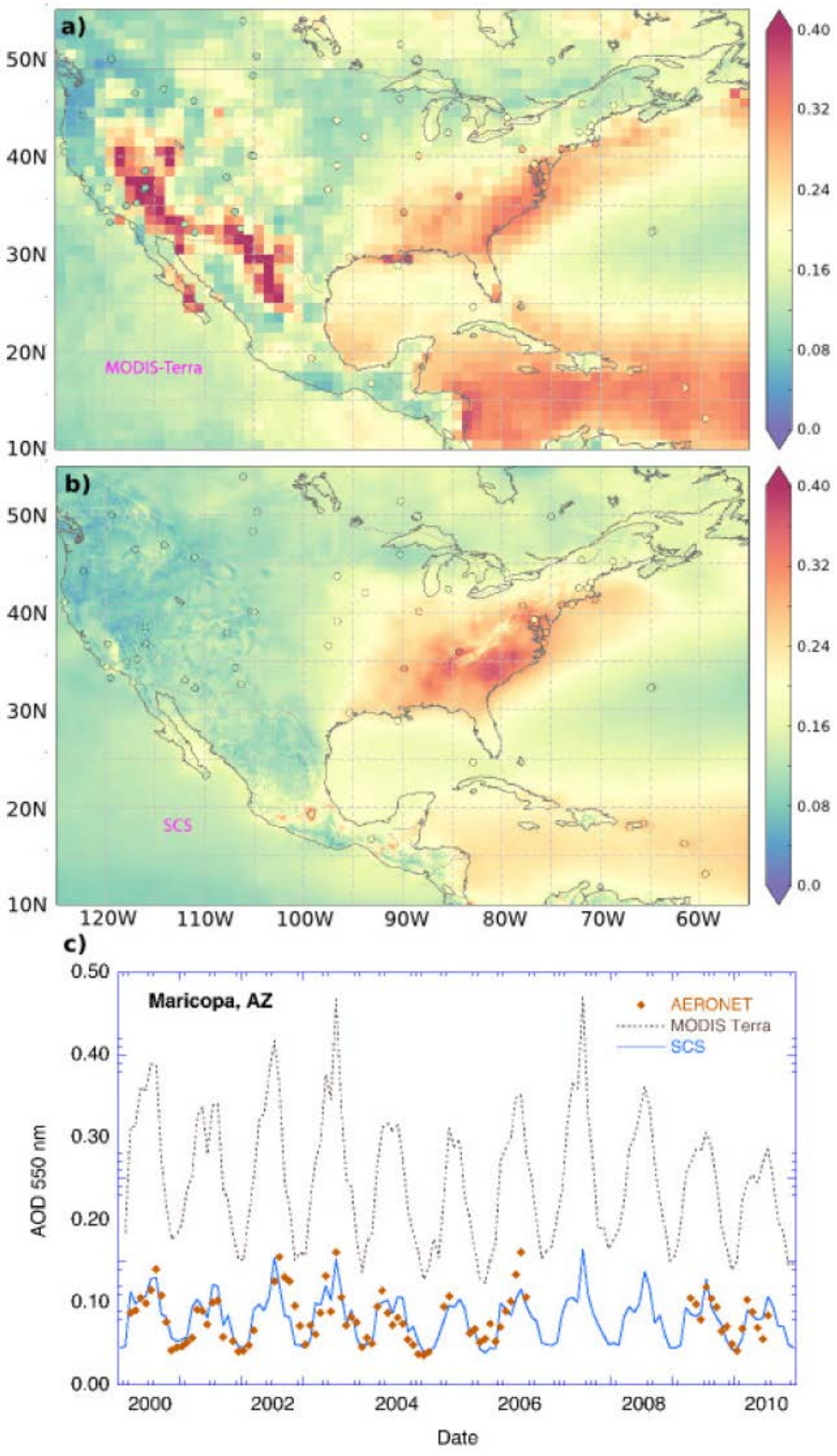


Figure 3-3. AOD at 550 nm from (a) MODIS and (b) the SCS dataset for the July climatology. The time series at Maricopa, AZ, are also shown together with the ground AOD observations in (c). This figure demonstrates that the SCS dataset successfully removes the anomalous high bias in the western U.S.

WRF-Solar™ was configured similarly to NOAA's High-Resolution Rapid Refresh (HRRR) system, with the same physics schemes and on a domain covering the CONUS at 3-km horizontal grid spacing. One key difference from the HRRR is that our implementation calls the radiation code every 5 minutes compared to 30 minutes in the HRRR. Four periods of five consecutive days were selected for analysis in each season, for a total of 20 days. All days were during 2012 because this is the most recent year with MACC reanalysis products available. Analyses from the Rapid Refresh (RAP) model run by NCEP, with 13-km grid spacing, provided initial and boundary conditions every 3 hours. Analyses were used to limit forecast error growth in this evaluation of aerosol direct-effect treatment. Initialization was at 0000 UTC each day, and simulations proceeded for 30 hours to ensure a continuous daytime period in the simulations.

Verification is performed against irradiance observations from the SURFace RADiation budget network (SURFRAD; Augustine et al. 2000, 2005). GHI, DNI, and DIF measurements are recorded every minute at seven geographically diverse sites across the CONUS. WRF-Solar™ was configured to output irradiance components at the SURFRAD sites at every model time step (20 s), and then these 20-s values were averaged to 1-minute values. With the present focus on clear-sky irradiance, the cloud-radiation feedback was deactivated. Verification samples were formed from all daytime minutes corresponding to clear skies in the observations, following Long and Ackerman (2000) to identify clear skies from observed GHI and DIF.

3.2.2 Results

An example demonstrates the expected effects of aerosols on clear sky irradiance. Aerosols absorb and scatter the incoming solar beam, reducing DNI and increasing DIF. Ignoring aerosols leads to systematic over-prediction of DNI and under-prediction of DIF. Figure 3-4a shows the observed irradiance during a nearly clear day at one SURFRAD site. Figure 3-4b shows the corresponding average errors (bias) for the NO-AEROSOL experiment. The simplest treatment for aerosols, ECMWF-CLIM, nearly eliminates the bias in DIF, and reduces the DNI bias magnitude (Figure 3-4c). GHI is slightly overestimated, while DNI is slightly underestimated in the ECMWF-CLIM experiment. Bias in DIF is negligible. The impact on GHI is much weaker because absorption is small and scattering is highly peaked in the forward direction, but the GHI bias drops by more than 10 W m^{-2}. Bias in GHI should be the sum of the DNI and DIF biases, but is not because of observational errors.

With few exceptions, the aerosol effect demonstrated in Figure 3-4 generalizes to the full clear-sky data set and other aerosol treatments (Figure 3-5 and Figure 3-6). The GEOS-5 aerosols lead to the overall lowest bias magnitudes. Simulation at the SURFRAD site labeled GCM is especially challenging for most of the experiments; GEOS5-AOD is the only experiment providing results comparable to those at the rest of SURFRAD sites. It will be shown below that this is a consequence of accurate AOD at GCM in the GEOS-5. Although they improve on NO-AEROSOL, the SCS-CLIM, GOCART-CLIM, and MACC-AOD experiments show too little DNI and too much DIF, thus indicating a high bias in the aerosol effect. A smaller magnitude impact on GHI is clear (Figure 3-5c), but the relative improvement from the aerosol effect is large (Figure 3-6c). Results from GEOS-5 suggest that computationally expensive models that explicitly and accurately solve the atmospheric chemistry can be competitive with, or superior to, high-quality climatology products.

Table 3-1. RMSE in the surface irradiance components (W m^{-2}). The relative improvement with respect to the NO-AEROSOL experiment is shown in parenthesis.

Irradiance	*NO – AEROSOL*	*ECMWF – CLIM*	*SCS – CLIM*	*GOCART – CLIM*	*MACC – AOD*	*GEOS5 – AOD*
GHI	21	16 (23 %)	16 (23 %)	16 (23 %)	20 (5 %)	15 (28 %)
DIF	44	20 (54 %)	19 (57 %)	26 (41 %)	42 (4 %)	12 (73 %)
DNI	103	66 (36 %)	52 (50 %)	58 (44 %)	120 (-16 %)	41 (60 %)

RMSE improvements show that bias reductions from including aerosols are responsible for most of the error reductions (Table 3-1). Again, runs imposing climatological properties of the aerosol provide similar results, which are superior to the MACC-AOD and inferior to the GEOS5-AOD experiments. Almost all the experiments that include atmospheric aerosols reduce the RMSE compared to the NO-AEROSOL run, agreeing with the bias improvements (Figure 3-5 and Figure 3-6). The only exception is the RMSE of DNI for the MACC-AOD experiment. Poor performance at DRA, where MACC strongly overestimates AOD (Figure 3-5 and Figure 3-6), is responsible. Again, the largest improvements are found using the GEOS5-AOD, which shows improvements with respect the ECMWF-CLIM, the standard representation of aerosols in NWP models, of 38% and 40% for the DNI and DIF, respectively.

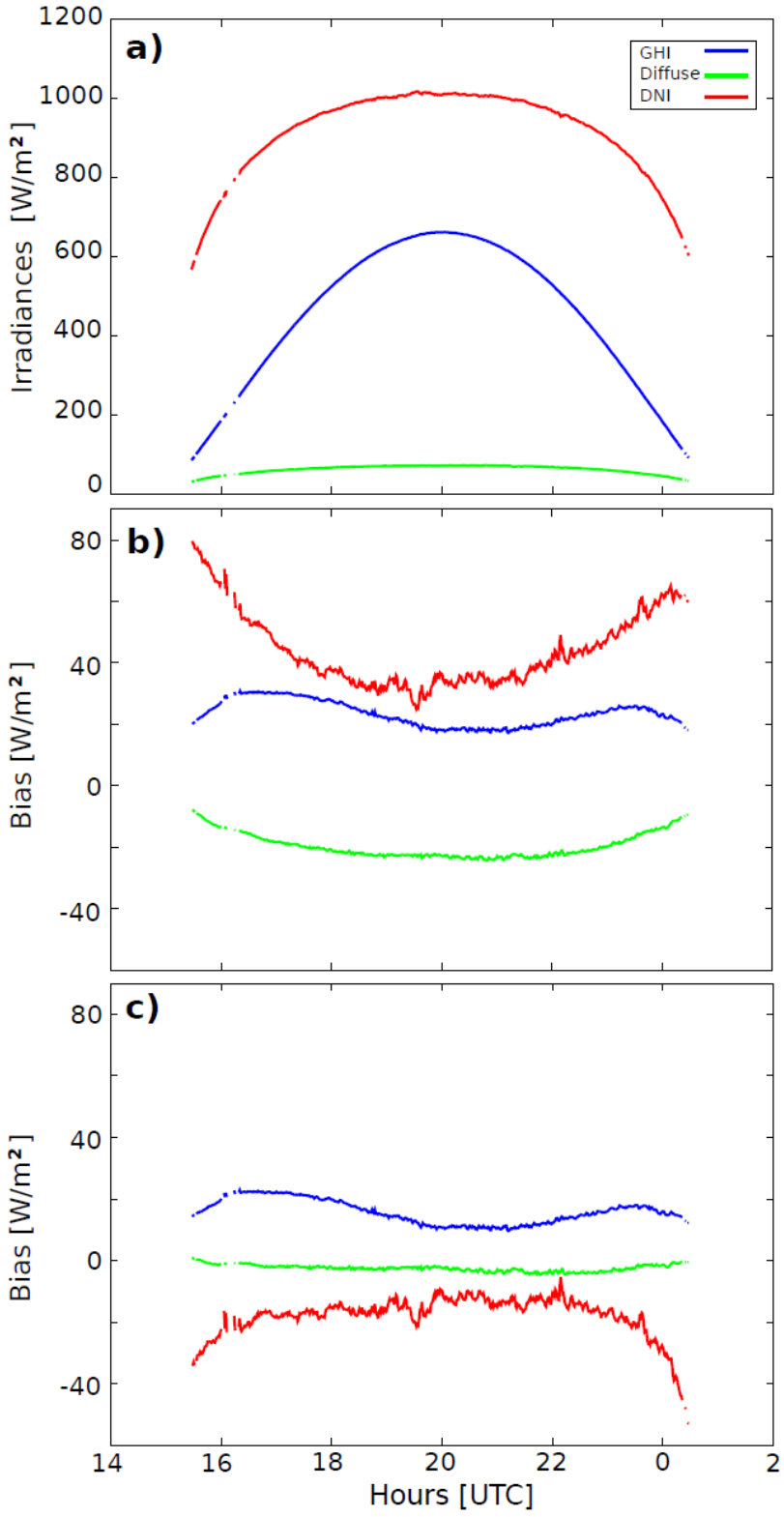


Figure 3-4. a) Observed surface irradiance components and their biases from the b) NO-AEROSOL and c) ECMWF-CLIM experiments (see key on panel a).

Comparing temporal error variability from each experiment allows us to quantify the improvements associated with the AOD variability. Table 3-2 shows the standard deviation of the error defined as the differences between the simulation and the observations. Results from ECMWF-CLIM and GOCART-CLIM indicate that these experiments do not reduce the variability of the error, which indicates that the improvement in the errors (Table 3-1) is associated with a bias reduction. The variability of the error is higher in MACC-AOD than in the NO-AEROSOL experiment. On the contrary, SCS-CLIM and GEOS5-AOD reveal a reduction of the error variability. The GEOS5-AOD experiment shows the most error reduction compared to the standard approach, the ECMWF-CLIM experiment, by 35% and 37% for the DNI and DIF, respectively. The relative improvement of each model experiment is directly associated with accuracy in the aerosol optical properties. To verify this, the AOD recorded by the five multi-filter radiometer channels available at the SURFRAD sites (413.5 nm, 497.4 nm, 615.0 nm, 672.7 nm, 869.8 nm) was interpolated to the primary wavelength of 550 nm using the observed Ångstrom exponent. The estimates from the five channels were then averaged to derive a single value. Comparing these observations to the values extracted from the four data sets under scrutiny here shows that the AOD skill among the various experiments is ranked similarly to the irradiance prediction skill (Figure 3-7).

The SCS-CLIM aerosol reproduces noticeable characteristics in the observed AOD, and agrees well with observations at certain sites (e.g. Figure 3-7e). The MACC-AOD and GEOS5-AOD both show temporal variability exceeding the observed AOD variability, with MACC-AOD additionally revealing a high bias. The GEOS5-AOD and SCS-CLIM both agree with observations better than MACC-AOD. This order is in agreement with the ability of the experiments to reproduce the observed surface irradiance (Figure 3-6 and Table 3-1), indicating that RRTMG properly accounts for aerosol effects when supplied with appropriate inputs. This confirms other recent results (Gueymard and Ruiz-Arias 2015).

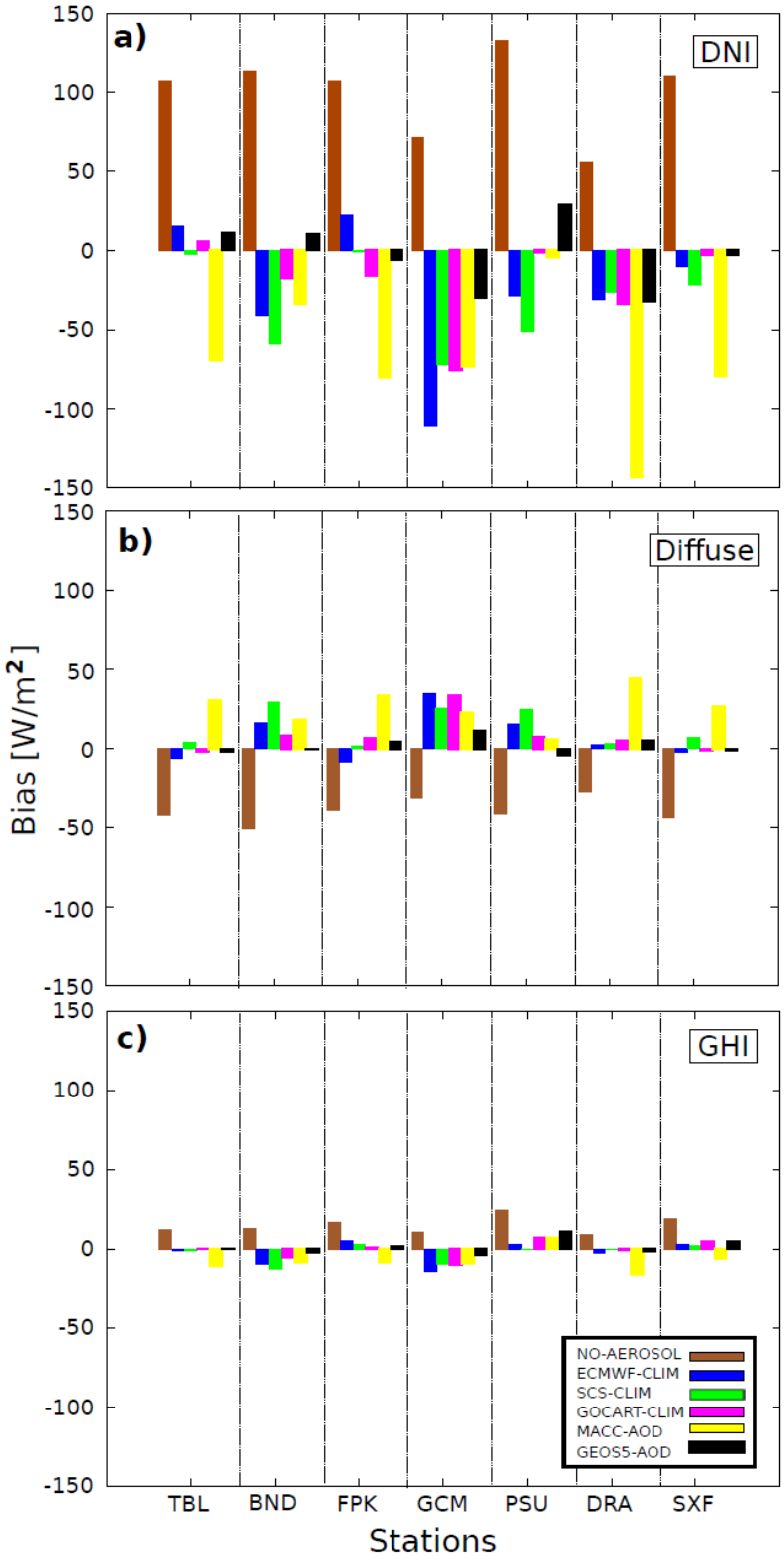


Figure 3-5. Biases in the surface irradiance components over all clear sky minutes at each SURFRAD site. Results for the six numerical experiments are shown (see legend).

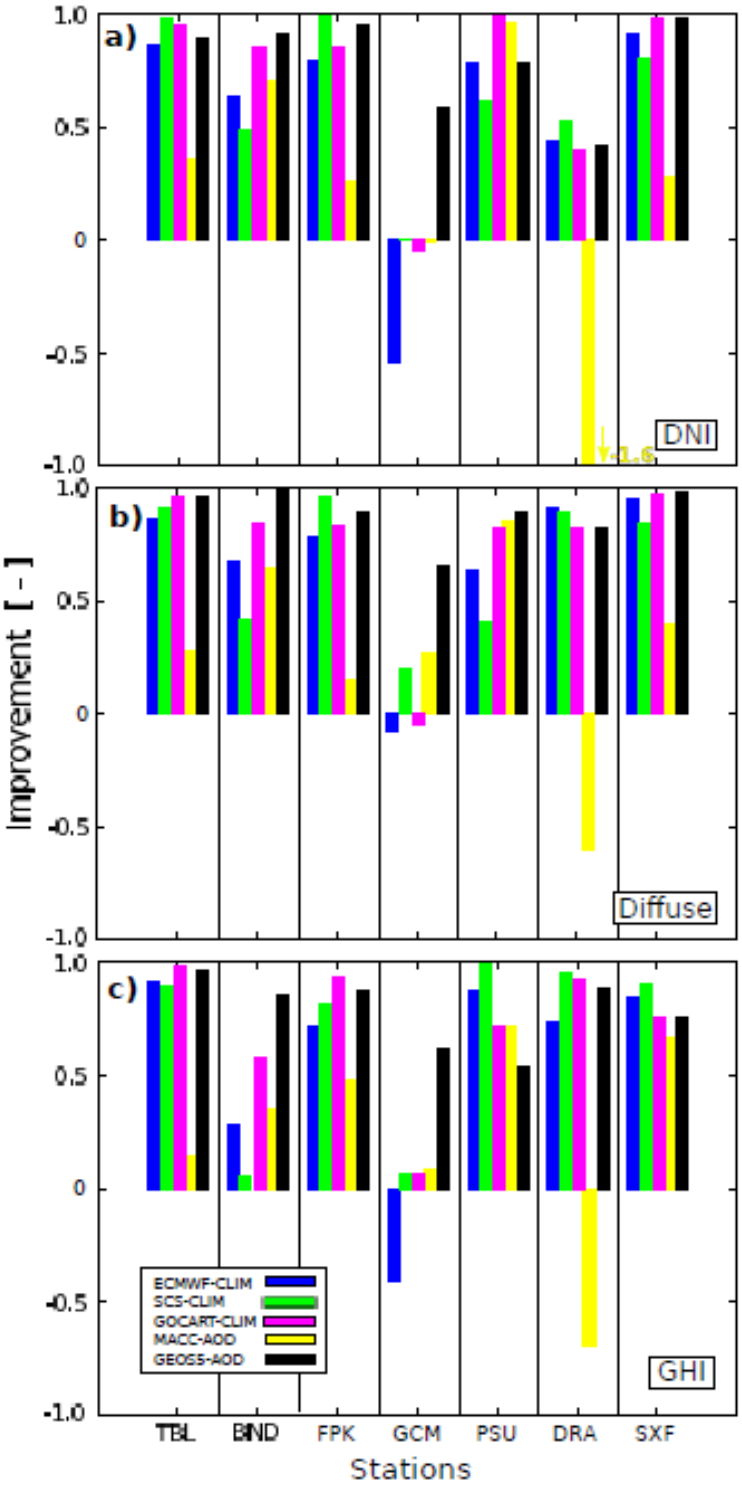


Figure 3-6. Improvements on the surface irradiance biases with respect to the NO-AEROSOL experiment for all clear-sky minutes.

Table 3-2. Standard deviation of the error in the surface irradiance components (W m^{-2}). The relative improvement with respect to the NO-AEROSOL experiment is shown in parentheses.

Irradiance	*NO – AEROSOL*	*ECMWF – CLIM*	*SCS – CLIM*	*GOCART – CLIM*	*MACC – AOD*	*GEOS5 – AOD*
GHI	16	16 (0 %)	16 (0 %)	16 (0 %)	17 (-6 %)	15 (6 %)
DIF	23	19 (17 %)	17 (26 %)	25 (-9 %)	26 (-13 %)	12 (48 %)
DNI	52	62 (-14 %)	45 (13 %)	54 (-4 %)	77 (-16 %)	40 (23 %)

3.2.3 Comparison of WRF-Solar™ to Standard WRF

To conclude the clear-sky assessment, a seventh numerical experiment summarizes the total effect of WRF-Solar™ (GEOS5-AOD) compared to the previous version of WRF that did not output separate irradiance components. The standard (baseline) WRF here is now defined as WRF-ARW version 3.5.1 with two modifications. First, the correction to the sun position algorithm was removed because it is part of the WRF-Solar™ effort. Second, code to output surface irradiance at the SURFRAD sites and at every model time step is added to enable direct comparison to WRF-Solar™. WRF-ARW version 3.5.1 already included the WRF-Solar™ capability to output the direct and diffuse radiation components. Another difference with respect to the NO-AEROSOL experiment is that the standard WRF v3.5.1 does not include the interpolation of the modeled irradiance between radiation calls.

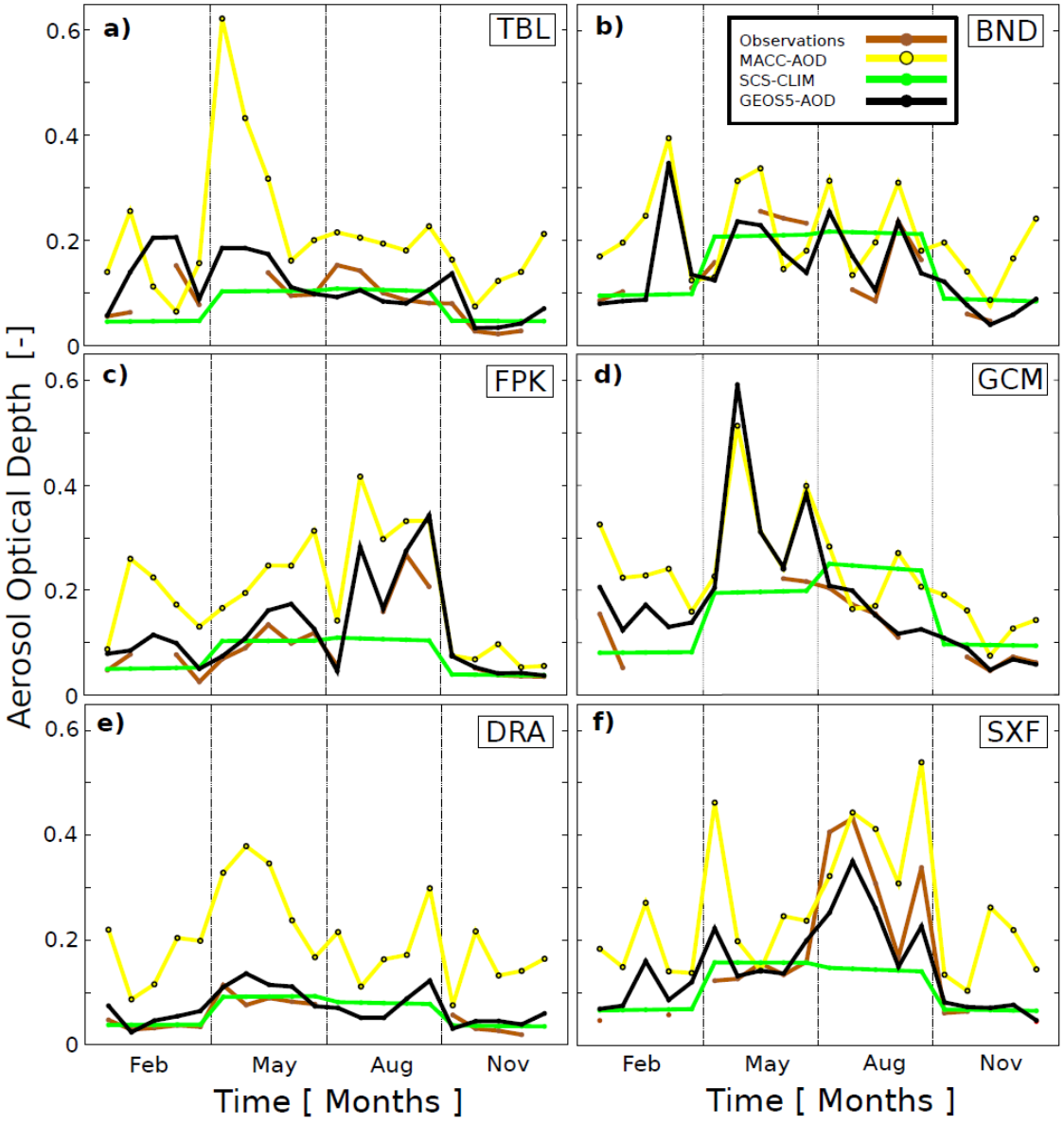


Figure 3-7. Total AOD at 550 nm from the different experiments at the SURFRAD locations with AOD records available.

Figure 3-8 shows the decrease in RMSE in WRF-Solar™ simulations compared to those from the standard WRF. Consistent with earlier results, WRF-Solar™ improves the GHI, DNI, and DIF under clear-sky predictions at all SURFRAD sites. On average, GHI is improved by 46%, DNI by 60%, and DIF by 70%.

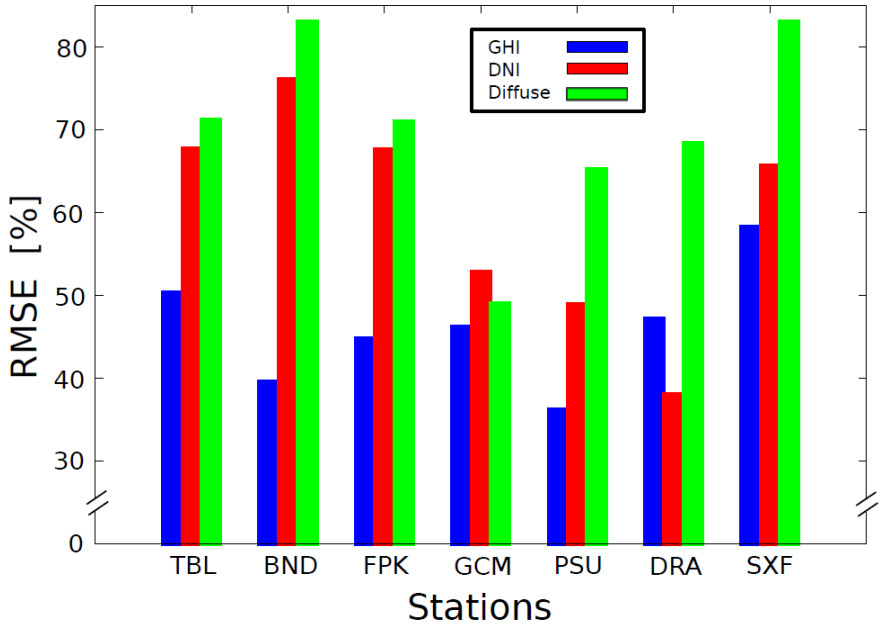


Figure 3-8. Improvements introduced by WRF-Solar™ (experiment GEOS5-AOD) in the estimations of the clear-sky surface irradiance components at the SURFRAD sites. The standard WRF simulations are used as a baseline for comparison.

3.3 Preliminary Assessment of Cloud-Aerosol Feedbacks

The potential benefit of activating the cloud-aerosol feedbacks were examined in a preliminary sensitivity study. Two WRF-Solar™ experiments were performed. The first one is the reference experiment (EXP1). EXP1 does not account for the aerosol indirect effects (cloud-aerosol feedbacks) and represents the aerosol direct effect via the ECMWF monthly climatology (Tegen et al. 1997). EXP1 is therefore the standard representation of aerosols in NWP models.

The second WRF-Solar™ experiment (EXP2) activates both the aerosol direct and indirect effects using the WRF-Solar™ augmentations described in above. Based on the good performance of the GEOS5-AOD experiment in clear skies we initialized the aerosol number concentration using the Modern Era Retrospective analysis for Research Applications Aerosol Reanalysis (MERRAero; Buchard et al. 2014, 2015, 2016) that uses the GEOS-5 model. In a first step, we replicated the clear sky experiment described in the previous section using a lower horizontal resolution (9 km). The comparison of the biases in the surface irradiances for the previous NO-AEROSOL and GOCART-CLIM experiments together with the new MERRAero experiment are shown in Figure 3-9. The MERRAero experiment also reduces the biases of the NO-AEROSOL experiment and shows similar performance that the GOCART-CLIM experiment. This indicates that WRF-Solar™ coupled with MERRAero reproduces the clear sky surface irradiances and motivated the analysis under all sky conditions.

The performance of WRF-Solar™ in all sky conditions is investigated with two one-way nested domains with a three-to-one spatial refinement in order to reach a grid spacing of 3 km over CONUS. Data from the ERA-Interim reanalysis (Dee et al. 2011) was used to generate the initial conditions and the boundary conditions that were updated every 6 hours.

The WRF-Solar™ simulations in EXP1 and EXP2 span the period from 12 February 2015 to 27 August 2015. There are 29 simulations per experiment. The first simulation is initialized on 12 February 2015 at 0000 UTC and WRF-Solar™ is run for 30 hours in order to have a simulation of a complete diurnal cycle after several hours of model spin-up. There is an additional simulation every week in order to span the complete period under investigation.

The simulations are evaluated against irradiance observations from the SURFRAD (Augustine et al. 2000, 2005) and the Integrated Surface Irradiance Study (ISIS; Hicks et al. 1996) networks. In order to mimic the observational spatio-temporal coverage, the WRF-Solar™ irradiances are recorded every model time step (20 s) at the 14 observational sites, and are subsequently averaged to match the 3-minute data from the ISIS network. The 1-minute data from the SURFRAD network is also averaged to 3-min.

Preliminary results are presented in Figure 3-10 that shows the bias in the surface irradiances calculated with the observations/simulations at the 14 sites. Both EXP1 and EXP2 show similar performance with a positive bias in the DNI and GHI and a negative one in DIF. This result is encouraging given the complexity of modeling the cloud-aerosol-radiation system.

We are in the process of extending the simulation to span the complete calendar year of 2015 in order to have representation of winter months. It is also under consideration to span the simulation another year (possibly 2014) in order to enlarge the dataset. In addition, we would like to investigate the impact of the aerosol aware scheme for days with heavy loads of aerosols wherein the impact of the aerosol effects is larger.

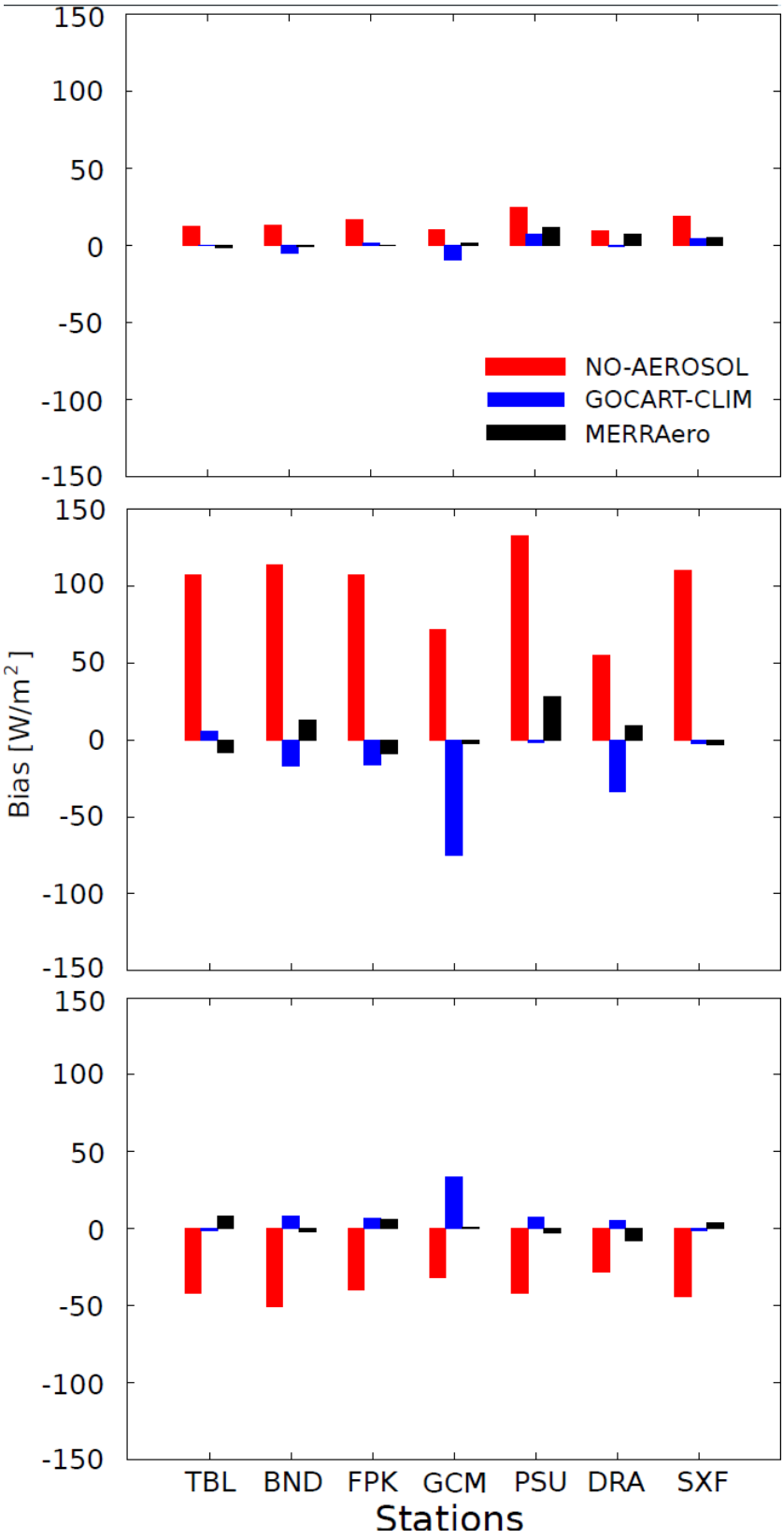


Figure 3-9. Biases in the a) GHI, b) DNI and c) DIF for different experiments (see legend in panel a).

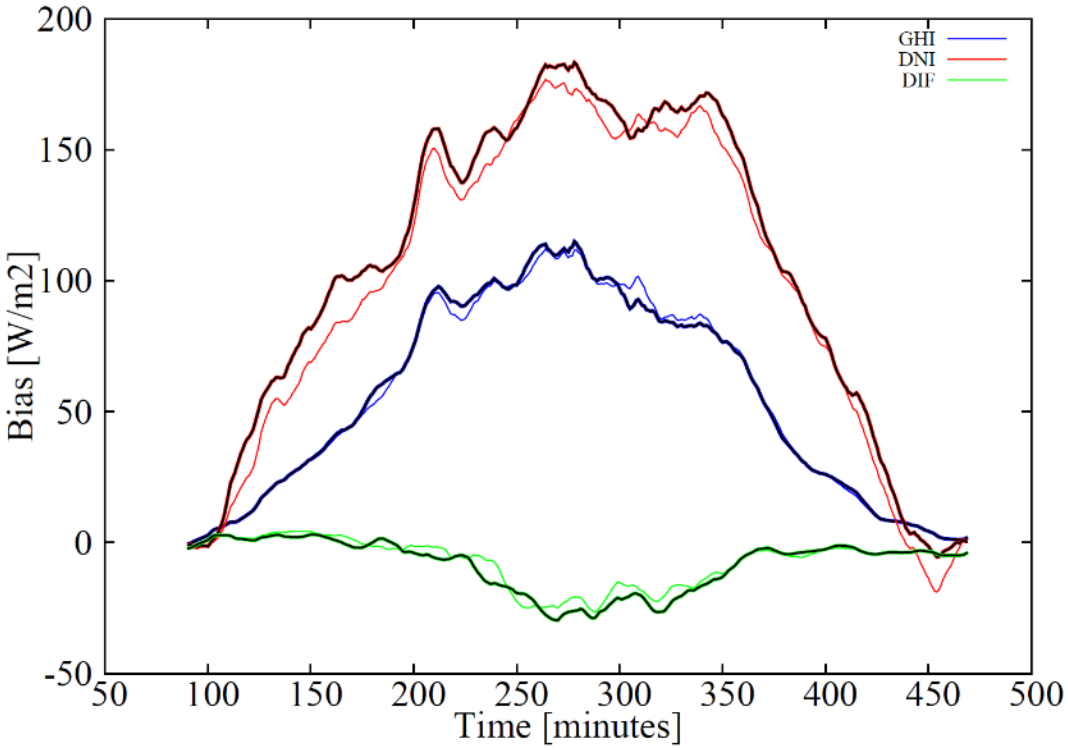


Figure 3-10. Bias in the surface irradiances (see legend) from EXP1 (thin colored lines) and EXP2 (thicker colored lines highlighted in black) as a function of the lead time.

3.4 QUASI-OPERATIONAL WRF-SOLAR™ FORECASTS

WRF-Solar™ has been run quasi-operationally over CONUS since 2015. There is one domain of 3 km of horizontal resolution and two domains of 1 km over regions with solar farms, the San Luis Valley in Colorado and SMUD in California. There is one run per day configured to meet operational needs of the private partners of the project. The run targets the day-ahead forecast. The model is initialized at 0000 UTC and is run for 54 hours in order to provide a day-ahead forecast available at the beginning of the morning. The model therefore runs during the night period and thus the latency of the system is not relevant. The computational cost of activating the 1-km domains is high so these high-resolution domains are only activated during the daytime of the second day of the simulation. The model is initialized with RAP forecasts and it uses both RAP and GFS forecasts to create the boundary conditions. If the RAP forecast is not available at the beginning of the simulation the system only uses GFS forecast data. The model output is saved every hour and passed to the DICast® integrator (Chapter 4).

3.5 CONCLUSIONS

The WRF-Solar™ augmentations to the WRF model, described here, have resulted in the first NWP model specifically designed to meet the growing demand for specialized forecasting products associated with solar power applications. The model includes representation of aerosol-

cloud-radiation feedbacks, and efficient numerical approaches to support operational forecasting. The clear sky assessment reveals large improvements compared to irradiance from the standard WRF. Sensitivity to details of the source of aerosol information emphasizes the importance of accurate aerosol optical properties for accurate estimates of surface irradiance. Models that explicitly solve atmospheric chemistry equations, and are initialized with an aerosol data assimilation process (i.e., GEOS-5), appear the most useful for clear-sky solar irradiance. In particular, imposing the temporal variability of the AOD produces large improvements in DNI and DIF with respect to the more typical use of aerosol climatologies.

Current developments focus on comparing forecasts and actual solar power production to precisely evaluate the model performance under all sky conditions (including cloudy periods such as the cloud-aerosol feedbacks). Further modeling advances in WRF-Solar™ are expected from these efforts, and should enhance the specific clear sky improvements highlighted here. Combined with the ability to output high-frequency irradiance time series, the new WRF-Solar™ should prove helpful to the growing solar industry in general, and contribute to better cloud, aerosol, and solar forecasts in general.

A comprehensive evaluation of the WRF-Solar™ model performance during the quasi-operational forecast is presented in chapter 5.

3.6 Publications Based on this Work

Journal Papers

Jimenez, P.A., S. Alessandrini, S.E. Haupt, A. Deng, B. Kosovic, J.A. Lee, and L. Delle Monache, 2016: Role of Unresolved Clouds on Short-Range Global Horizontal Irradiance Predictability, submitted to *Monthly Weather Review*.

Jimenez, P.A., J.P. Hacker, J. Dudhia, S.E. Haupt, J.A. Ruiz-Arias, C.A. Gueymard, G. Thompson, T. Eidhammer, and A.J. Deng, 2016: WRF-Solar™: An Augmented NWP Model for Solar Power Prediction, *Bull. Amer. Met. Soc,* in press.

Ruiz-Arias, J. A., J. Dudhia, and C. A. Gueymard, 2014: A simple parameterization of the shortwave aerosol optical properties for surface direct and diffuse irradiances assessment in a numerical weather model. *Geoesci. Model Dev.*, 7, 1159–1174.

Thompson, G., and T. Eidhammer, 2014: A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *J. Atmos. Sci.*, 71, 3636–3658.

Xie, Y., M. Sengupta, and J. Dudhia, 2016: A Fast All-sky Radiation Model for Solar Applications (FARMS): Algorithm and performance evaluation. Solar Energy (Under review).

Conference Presentations

Jimenez, P.A, and S.E. Haupt, 2016: WRF-Solar enhancements of the aerosol-cloud-radiation system in support of solar power forecasting. Atmospheric Radiation Science Workshop, Boulder, CO, March 9.

Jimenez, P., S. Alessandrini, S.E. Haupt, and A. Deng, 2016: Accounting for the Effects of Unresolved Clouds in the Shortwave Irradiance Forecast of the WRF-Solar Model to Improve Solar Power Forecasts, Seventh Conference on Weather, Climate, Water, and the New Energy Economy, AMS Annual Meeting, New Orleans, LA, Jan. 11.

Jimenez, P.A., S.E. Haupt, J.P. Hacker and J. Dudhia: WRF-Solar: Upgrading the WRF representation of aerosol-cloud-radiation feedbacks in support of solar energy forecasting. Poster. AGU Fall Meeting, San Francisco, CA, December 14.

Jimenez, P.A., S.E. Haupt, J.P. Hacker, J. Dudhia, 2015: WRF-Solar: An Augmented NWP Model for Solar Power Prediction. International Conference on Energy and Meteorology, Boulder, CO, June 24.

Jimenez, P.A., S.E. Haupt, J.P. Hacker, J. Dudhia, 2015: WRF-Solar: Improvements to WRF for real-time solar energy forecasting applications and its evaluation. WRF Users' Workshop, Boulder, CO, June 16.

Jimenez, P.A., S.E. Haupt, J. Hacker, and J. Dudhia, 2015: WRF-Solar to Advance Solar Power Forecasting, Sixth Conference on Weather, Climate, and the New Energy Economy, AMS Annual Meeting, Phoenix, AZ, Jan. 6.

Deng, A., B. Gaudet, J. Dudhia, and K. Alapaty, 2014: Implementation and evaluation of a new shallow convection scheme in WRF. *26th Conf. on Wea. Anal. and Forecast./22nd Conf. on Numer. Wea. Pred.* at the *94th Amer. Meteor. Soc. Annual Meeting*, Atlanta, GA, 2-6 Feb 2014. Preprint available at https://ams.confex.com/ams/94Annual/webprogram/Paper236925.html.

4 ENGINEERING

4.1 OVERVIEW

The overarching concept for the Sun4Cast® system is to blend various forecasting models to provide a best consensus-based forecast over all timeframes. Using multiple models is an important part of this concept and is a "best practices" approach to modern meteorological prediction (Young 2001; Mahoney et al. 2012; Orwig et al. 2014; Tuohy et al. 2015; Haupt and Mahoney 2015). The Sun4Cast® system (Figure 4-1) has two main forecast modules, a Nowcast track that forecasts at high temporal resolution extendingto 6 hours, and a Dynamic Integrated ForeCast (DICast®) modules that forecasts at coarser temporal resolution out several days. Both these modules apply a consensus forecasting approach. That is, they consider multiple inputs and perform a forecast integration that takes advantage of the strengths of each input. While the consensus forecasting approach has been applied to forecasting more common weather variables (e.g., air temperature), it has not been applied to solar irradiance forecasting in any significant way. No other public systems use a consensus forecasting approach. In the private sector, some companies may use a consensus approach, while others rely on a single source model; much of this is proprietary and not disseminated.

This project sought to demonstrate the advantages of consensus forecasting. Raising the awareness of these basic advances can benefit the industrial forecast developer community and ultimately create more competition in the solar forecasting sector. Through the initial development of this project, one critical hypothesis that needed investigation was, "Does the consensus forecasting approach work for solar irradiance?" Preliminary results indicate that the DICast® consensus forecast can reduce the global horizontal irradiance (GHI) errors about 15% when compared to the best (single model) ingredient forecast.

Experience from forecasting other meteorological variables indicates that the best single model varies seasonally. These consensus forecast systems are developed using machine learning techniques that continually adapt and learn how to produce a forecast that is better than its ingredients. Clearly this reduction in error demonstrates how these systems address the need for improvement in solar forecasting. The nowcast integrator is the core piece of technology that blends multiple input nowcasts. We developed an operational system that demonstrates a reduction in error when compared to single nowcast inputs. While the aforementioned preliminary results apply to the day-ahead forecasts, this is seminal work in the short-range forecast arena.

Through development of other consensus forecast systems, there is a well-established methodology of forecast verification. The verification software that was developed calculates statistics and also identifies outliers. These are forecasts that poorly match the observations. This can either be due to poor forecasts or bad observation data. Both are useful for debugging, developing improved quality control (QC) algorithms for observations, and/or investigating forecast algorithmic improvements. The forecast error statistics evolve seasonally. This makes it difficult to make gross statements, on topics such as forecasts getting better each month.

When blending multiple models, it is common to have missing data from either *in situ* observations or from at least one of the models. It requires detailed systems engineering to assure that such

situations are dealt with appropriately in real-time. To that end, the systems engineering includes quality control/quality assurance algorithms to assure that the model and observational data used in the forecasts are within appropriate bounds. Where data are missing or only partial data or model outputs are available, the algorithms are able to use alternate methods to provide forecasts. This is actually one advantage of using multiple models for a consensus forecast: there are surely at least some model data available and, in the case of failure to obtain any model data, the forecast could be made with the last model data available. Such an approach to engineering a forecast system uses the concept of measured planned degradation when certain portions of the input stream may not be available or fail the QC criteria.

A big portion of the engineering is data management: ingest, formatting, QC, blending, output, and transmittal to the end user. To make the data streams available for the blending process requires substantial system engineering. This process requires careful logistics to process large amounts of model and observational data and to assure that all of it is available to the blending models on request. The model data are integrated on a dedicated server at NCAR. DICast® and the nowcasting blending algorithms have access to these data streams and these forecast streams were also made available to the forecast provider partners. The output of the blending technology was provided to the utilities and ISOs.

The meteorological irradiance values are not the final output variables. Utilities require a power prediction, meaning that an irradiance-to-power conversion step must be added. A model regression tree (Cubist) is used in Sun4Cast® to train the relationship between the measured irradiance value and the coincident power produced. The empirically derived relationship is then applied in real-time to the irradiance forecast to produce a power forecast. A separate power conversion algorithm must be trained for each solar forecast site. Once the training/testing procedure is accomplished, the algorithm's real-time application runs in a matter of seconds.

Finally, probabilistic forecasts have been requested by partner utilities. NCAR applies the Analog Ensemble (AnEn) approach to produce an appropriate probability density function (pdf) of the forecast uncertainty. The Analog Ensemble searches for similar forecasts made in the past under meteorological conditions analogous to today's forecast. Thus, we wish to identify such analogs in those past forecasts in order to: 1) further correct today's forecast, which improves the deterministic forecast, and 2) produce a pdf of multiple analogs to estimate the uncertainty of the forecast. This flow-dependent uncertainty has been shown to reproduce the forecast and its statistical reliability at least as well as the full ensembles of runs produced at the operational centers (see section 4.6 for additional discussion of AnEn).

NCAR produced these probabilistic solar power forecasts in quasi-operational mode hourly and made them available to the utility and ISO partners for a year in order to provide sufficient data for a full assessment. Preliminary results indicate that each component improves upon baseline forecasts and has a "sweet spot" where that component often produces the best forecast and contributes to an improved forecast.

The forecast system has been operating in a quasi-operational status of blending all portions of the systems described above. Just as the input systems continue to improve, the blending systems adapt to that improvement through updating the weights assigned to the various systems as their skill in

matching the observations becomes more apparent with time. This has been a dynamic process that assures continual improvement throughout the project.

The current version of the engineered Sun4Cast® system appears as Figure 4-1. A simplified view is provided as Figure 4-2.

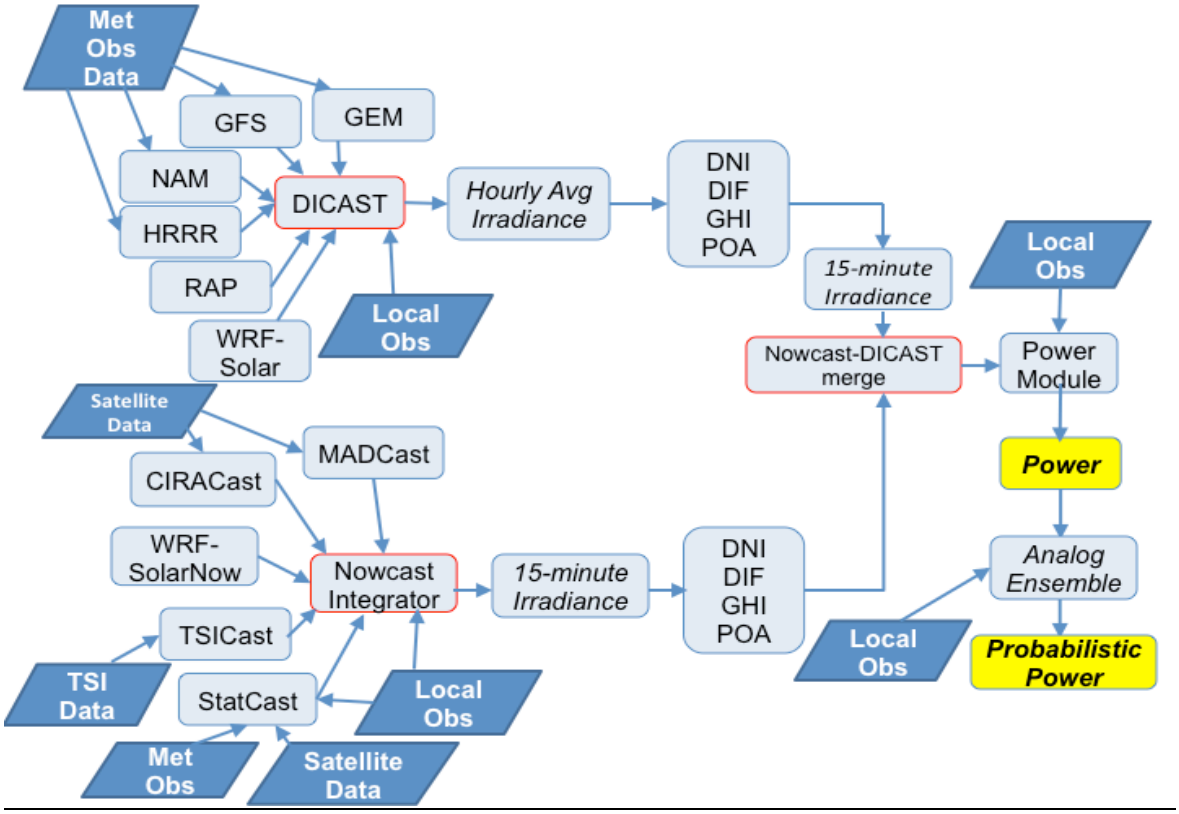


Figure 4-1. The engineered Sun4Cast® system.

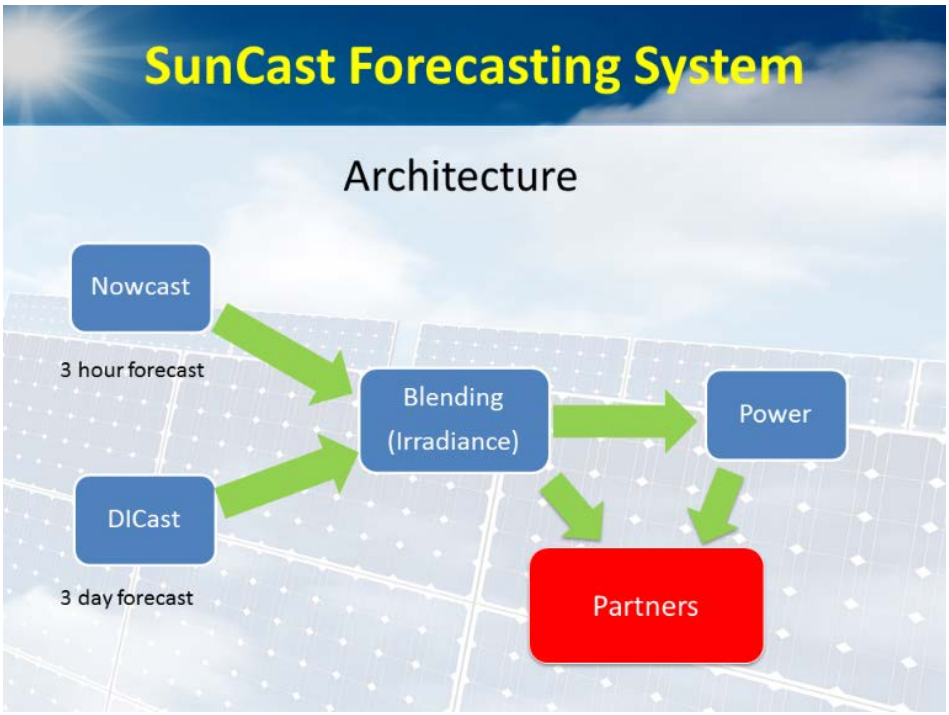


Figure 4-2. Simplified depiction of Sun4Cast® Forecasting System.

4.2 THE NOWCAST SYSTEM

As described in Chapter 2, the Nowcast System is composed of various models and each displays a "sweet spot" in terms of metorological conditions or lead time for producing a most accurate forecast. These nowcasting methods leverage a variety of disparate observational data, statistical and computational intelligence methods, and physical understanding of the atmosphere to produce a "best practices" blended forecast. The Nowcast systems are integrated using a unique Nowcast Expert System Integrator (NESI) that utilizes the recent performance scores of each component model, whether it is based on ground-based cloud observations and computational intelligence methods (StatCast and TSICast), on satellite cloud observations (CIRACast), or includes NWP components (MADCast and WRF-Solar-Now). Although the Nowcast system is currently optimized via an expert system, dynamic automated methods are expected to be useful future applications. Thus, the NESI component of Sun4Cast® is a collection of individual short-term forecasts that use a variety of techniques, each with its own strength, especially over specific forecast lead times. As a consequence, the NESI combines the individual nowcasts into a single forecast for each site as a function of lead time, using a weighted sum of the individual forecasts. The integrated Nowcast GHI forecast (in addition to plane of array (POA), direct normal irradiance (DNI), and diffuse irradiance (DIF) where available) extends to 6 hours at 15-minute resolution. Figure 4-3 illustrated the nowcast blending. NESI is a unique aspect of the Sun4Cast® system.

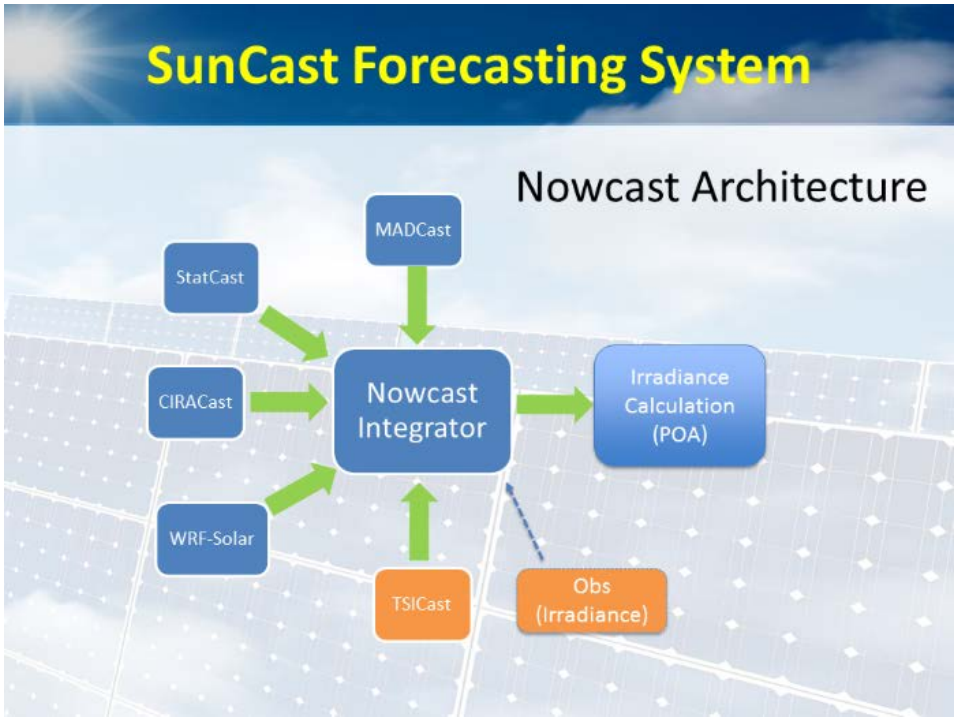


Figure 4-3. The Nowcasting integrated forecast.

4.3 DICAST

NCAR's DICast® system constitutes the core of the day-ahead capability of Sun4Cast® and also supplies value over all time periods (Myers et al. 2011, 2012). DICast® produces automated forecasts using a method that was designed to emulate the human forecast process. It generates forecasts by optimizing the combination of NWP model data through developing empirical relationships gleaned from historical model output and observations. DICast® typically reduces error by about 10-15% as compared to the best individual input model. DICast® employs a two-step process: it first statistically corrects the bias of each input model using Dynamic Model Output Statistics (DMOS) (Glahn and Lowry 1972), secondly, it optimizes the model blending weights for each lead time, producing a consensus forecast. DICast® typically works with up to 90 days of data; this is an advantage because many other methods require a year or more of data for training, during which time some of the models may have been modified or upgraded, making the training process difficult. A variety of weight calculation approaches are possible. These are compared and summarized by Young (2002), Greybush et al. (2008), and Myers et al. (2011, 2012). The adaptive learning approach is computationally simple and robust, can easily adapt to the addition of new input forecast models or the removal of obsolete models, and can easily evolve weights for new sites (Myers et al. 2011, 2012).

DICast creates its consensus as a bias-corrected weighted sum of input forecasts.

$$F = (\Sigma w_i f_i) / (\Sigma w_i) + Bias \quad (4\text{-}1)$$

where f_i are the forecast values and w_i are the weights. Missing forecasts are removed from the consensus. This is a computationally simple, yet effective, forecast combination method. More complex combination schemes are possible, but in our experience, the marginal improvement does not merit the additional complexity.

Each day, the weights are modified in the direction of the gradient in weight space. That is, the vector of weights is nudged in the direction of steepest descent of the error (the difference between the verification v_i and the forecast values):

$$\Delta w_i = S * (\partial / \partial w_i) \left\{ \left(v_i - f_i \right)^2 \right\} \tag{4-2}$$

The step length *S* is a parameter determined by the user to affect how quickly the system adapts. The choice of *S* effectively is a trade off between the initial discovery of the optimal combination and the daily update magnitude. Unless the step size is too large, the updated weights would have, by design, led to a forecast with a smaller error had they been used for the previous day's forecast. There is also a cap on the magnitude of any change so that one day's missed forecast does not completely alter a set of weights that work reasonably well.

In this way, the DICast® integrator never attempts to directly calculate the optimal set of weights. Instead it takes an approach of pursuing the location of the minimum error. The location of the minimum is rarely stationary. It changes daily. In a larger sense, the nexus of the optimal vector changes seasonally to capture the variability in the models' skills. Other weight calculation approaches that examine a longer history require more computational resources. Due to the daily variability in model skill and observational representativeness, this additional computational cost is not merited.

The DICast® system used in Sun4Cast® combines forecasts from multiple numerical weather prediction (NWP) models in an optimal way to produce tuned irradiance forecasts at specific locations and forecast projections. The system uses a history of observations and model runs to determine the performance of each model, then weighs each model based on its relative performance. The resultant integrated forecast produces lower error characteristics than if one used the best-performing individual model over the time period.

DICast® has been applied to many forecasting problems, including transportation (Drobot et al. 2010; Chapman et al. 2010), agriculture, and wind energy (Mahoney et al. 2012; Haupt and Mahoney 2015). For Sun4Cast® , the DICast® system was modified to produce tuned hourly-averaged GHI, DNI, and POA forecasts extending to three days at 15-minute intervals. The DICast® system was modified to handle new irradiance variables from the various input NWP models, and also to ingest irradiance observations from partner solar installations. These observations provide the tuning capability for the system to be able to produce the optimized forecasts. Figure 4-4 illustrates the DICast® consensus system and its inputs employed for Sun4Cast® .

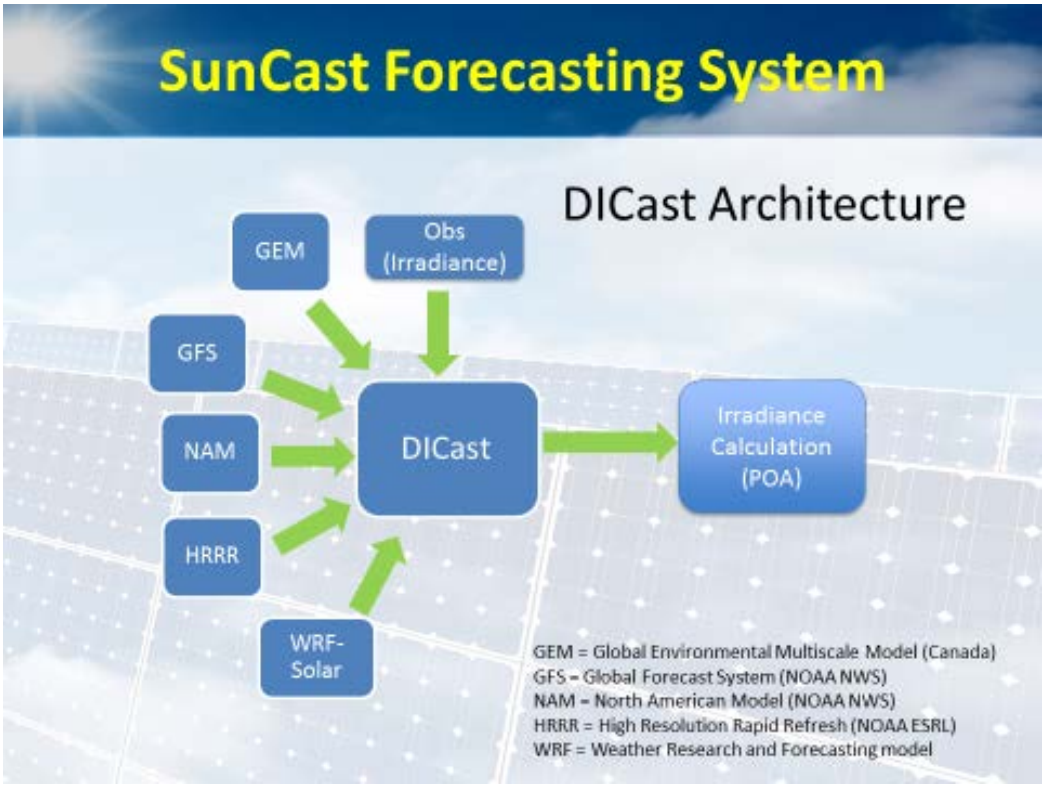


Figure 4-4. Diagram of the inputs to and output from the DICast® integrator in the Sun4Cast® day-ahead system.

4.4 Blending Nowcast and DICast® Systems

The Nowcast forecast is blended with the DICast® forecast to create a unified Sun4Cast® forecast for each site over the whole forecast period, from 0 to 72 hours, and can be extended as far as 168 hours. Before this blending is performed, the DICast® forecast is interpolated from hourly to 15-minute intervals to match the Nowcast forecast interval. This 'smart' temporal interpolation is performed using cloud cover forecasts at the nearby hours to influence each 15-minute interval. Once the two forecast subsystems match temporal resolution, they are blended together over three different forecast periods: 100% Nowcast from 0-3 hours, a decreasing Nowcast/increasing DICast® linear blend from 3-6 hours, and then 100% DICast® from 6-72 hours. Figure 4-5 illustrates the blending concept.

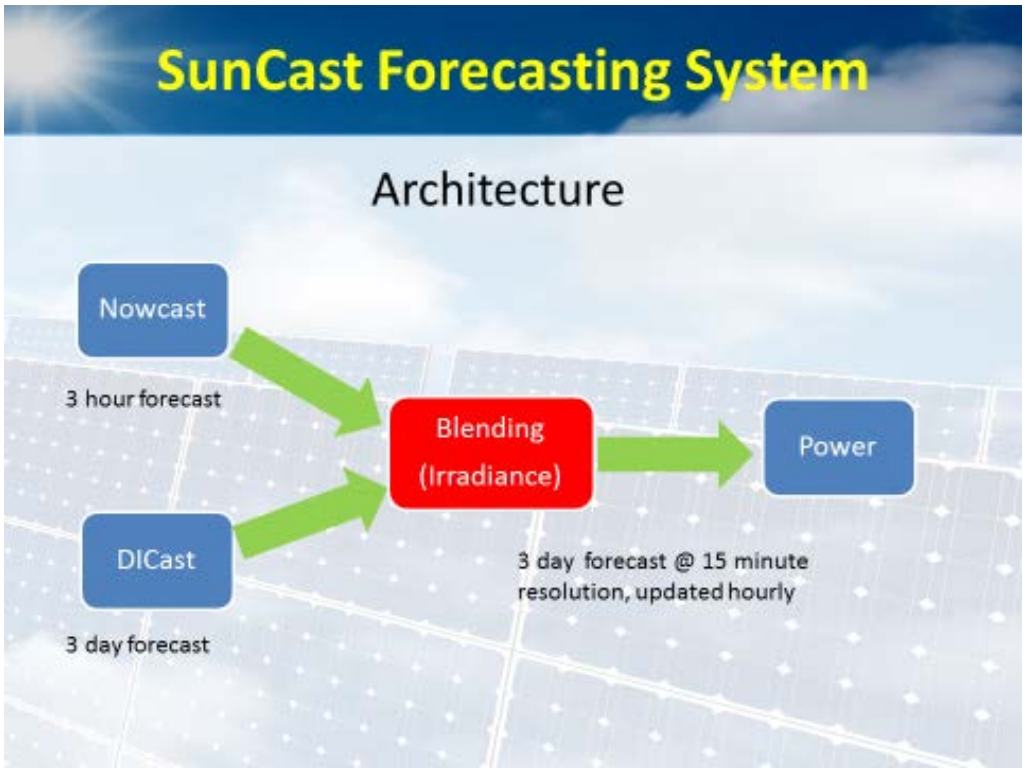


Figure 4-5. Blending the Nowcast and DICast forecasts

4.5 Power Conversion

4.5.1 Irradiance to Power Conversion Model Creation

Predicting irradiance is a step toward predicting power, the variable that the utility and balancing authority actually need in order to plan day ahead unit commitment and balance the grid in real-time. The power conversion depends on the particular type of hardware installed at the solar farm as well as local conditions. There are models such as PVWatts (http://pvwatts.nrel.gov/) that can do this power conversion. In prior work with wind energy, however, we have found that empirical power conversion methods can outperform models (typically power curves for wind energy) because they take into account local effects such as terrain blocking, impact of upstream turbines, density, etc. (Parks et al. 2011). For solar energy, such effect could include dust, shadowing, etc. that cannot be captured in any general model. Our experience in this project also indicates that it may be difficult to obtain the metadata required for PVWatts (such as panel tilt, etc.). Thus, as part of this project, we have developed empirical methods to use across a broad range of solar technologies and geographic locations.

The project team developed such empirical models to convert irradiance to power. The philosophy is that rather than specifying the physics, one obtains a sufficiently long time series of matched irradiance and power output data, and then trains an artificial intelligence model to predict power from irradiance. This approach is straightforward and it is typically superior to more direct prediction methods. Specifically, NCAR has applied regression tree analysis to train conversion algorithms to best match historical observed irradiance/power relationships.

4.5.2 Power Conversion Methodology

The solar power forecasting system forecasts power generation in two stages. The first stage involves forecasting an irradiance component such as GHI. The second stage involves converting the forecasted irradiance to power. In order to convert the forecasted irradiance to power, empirical data consisting of (irradiance, power) pairs are gathered over a sufficiently long period of time for an individual solar farm. These data pairs are then utilized to create an irradiance-to-power conversion model for that farm.

4.5.2.1 Data

The data needed to create an irradiance-to-power conversion model consist of observations of irradiance (GHI, DNI, or POA) matched with observations of power. When matching the observations, we chose to average the values over a fixed time interval. This process ensured that observations were matched to a common time and eliminated issues related to differing data update rates and mismatched observations. It is best to have the observations on the same time scale as the forecast, i.e., for forecasts every 15 minutes, having matching observations of irradiance and power every 15 minutes is ideal. For the majority of the models we created this was possible. However, for a few farms this was not the case and observations were matched on an hourly basis. Data were collected for the various farms over periods of one year and extending to three years, with a longer history of data producing a more robust model.

The model training data set consisted of matched irradiance and power observations along with information about the hour of day, the day and week of the year for the data, as well as values for the solar azimuth and solar elevation for the observation location and time.

Plots of observed irradiance vs. power (Figures 4-6 through 4-10) were created for each farm to get a sense of the data and its quality. The scatter plots on the left provide information regarding the spread of the data, including outliers. Multiple irradiance-power data points in the same location cannot be discerned in these plots, and they do not elucidate how the data is distributed within the covered area. The density plots on the right show the counts of data binned into irradiance-power boxes, therefore indicating how the data within the scatter plot is distributed, making it easier to identify outliers in the dataset. Different farms exhibited different relationships between irradiance and power. This is likely due to the solar panel hardware including the solar panel tracking employed, and the solar farm environment. Figure 4-6 though Figure 4-8 all show irradiance vs. power for farms that have GHI observations, while Figure 4-9 is for a farm with POA observations. Due to the non-linear nature of the relationship shown in Figure 4-7 and Figure 4-8 we realized that irradiance alone would not be a good predictor of power.

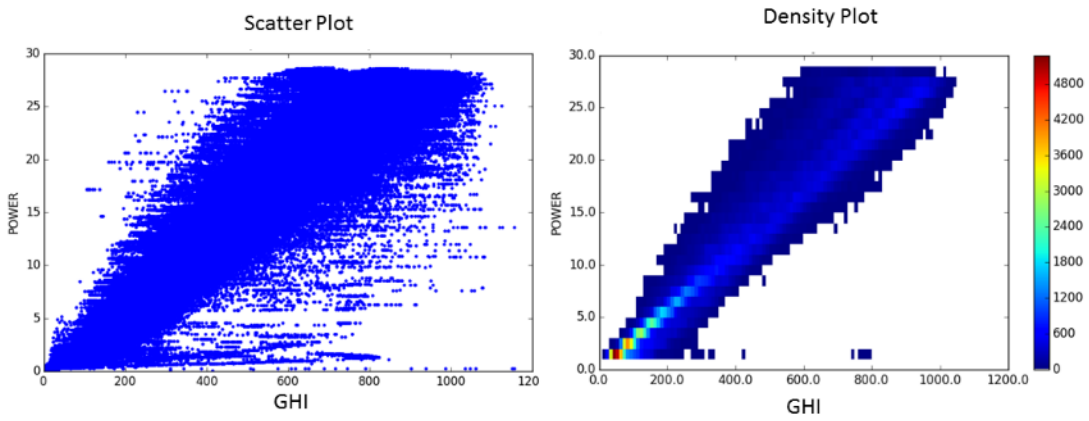


Figure 4-6. Irradiance vs. power observations scatter and density plots for one farm showing GHI relating to power in a somewhat linear way.

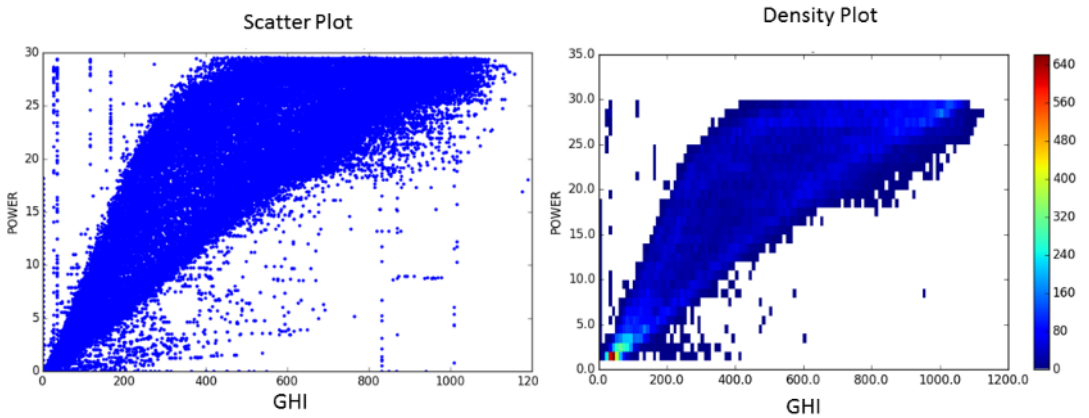


Figure 4-7. Irradiance vs. power observations scatter and density plots for one farm showing GHI relating to power along two distinctly different patterns.

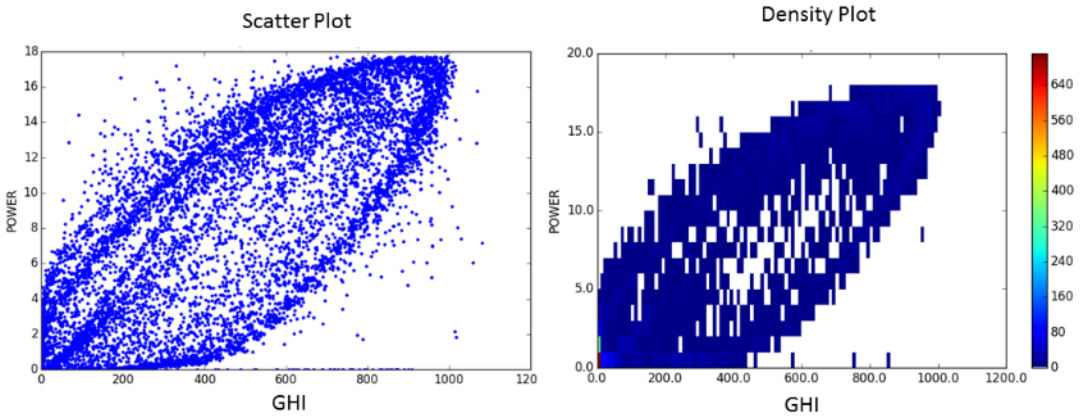


Figure 4-8. Irradiance vs. power observation scatter and density plots for one farm showing GHI relating to power in multiple and distinctly nonlinear ways.

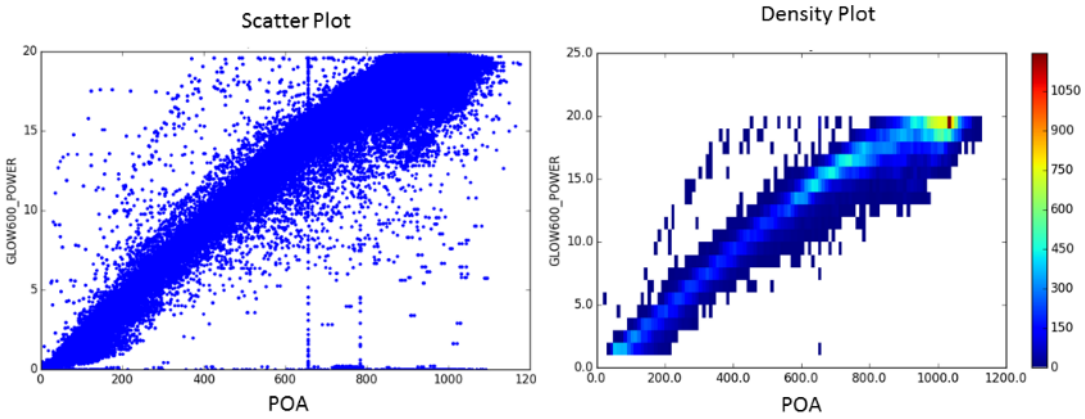


Figure 4-9. Irradiance vs. power observation scatter and density plots for a single farm showing POA relating to power in a mostly linear way.

4.5.2.2 Quality Control

Quality control issues with the data are prevalent and quality controlling the training data set consisting of matched irradiance and power observations is critical. Obvious issues includedcases in which one of the values from either the irradiance or power were stuck on a specific value while the other fields continued to vary. This can be seen clearly in Figure 4-7 and Figure 4-9 above. Another frequent occurrence was power values of 0 with varying irradiance, and vice-versa. Plotting the matched irradiance and power values illuminated these issues but also displayed some values falling outside the standard envelope of observations for each specific farm. Constant values along with any values outside the standard envelope of observations for each particular farm were eliminated from the training set. Figure 4-10 shows the quality-controlled irradiance vs. power observations for the same farm as in Figure 4-9. Plotting power verses solar elevation also showed some observations with large power values but below zero solar elevation. These few observations were also eliminated.

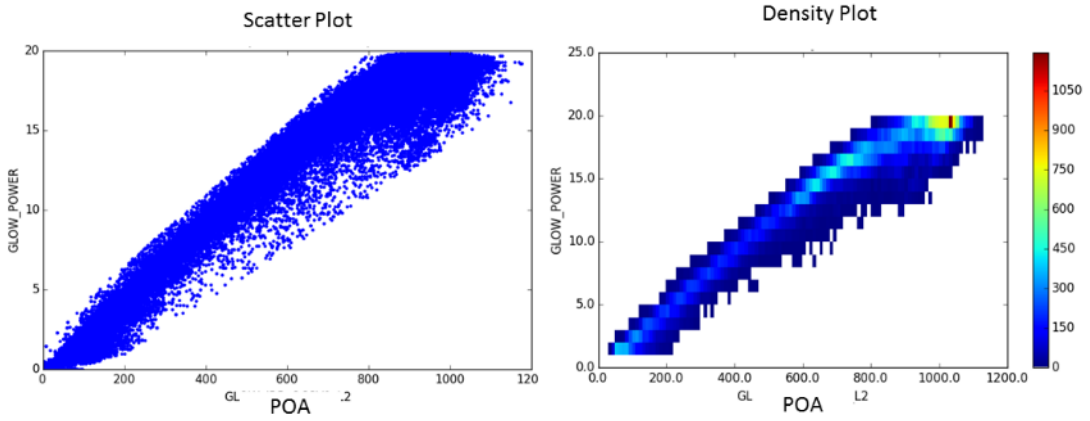


Figure 4-10. Quality-controlled POA vs. power observation scatter and density plots – compare to Figure 4-9.

4.5.2.3 Model Creation

The irradiance-to-power conversion models were created using the rule based regression tree software package Cubist by RuleQuest, which is Quinlan's (1987a,b, 1992, 1996) M5 model tree formatted as a set of rules (Kuhn et al. 2012). Cubist was chosen due to the previous success of using the software to create wind-to-power conversion models for other renewable energy applications, along with familiarity and ease of use. A different model was created for each solar farm, with some models using different predictor variables. In general, irradiance was always used along with variables that identified time and season. Some combination of hour of day, day and week of year, solar elevation and azimuth were used in each model. Our models also took advantage of the Cubist option to create a committee model, which allows for a group of several rule based models to be created that each predict the target value, with the final answer being the average of all the rule based models predictions. Each subsequent committee member model after the initial rule based model attempts to correct the errors from the previous model. The committee models appeared to provide a smoother forecast prediction through time when compared to a basic single rule-based model. Other options, such as using a nearest neighbor from the training set to modify the prediction, were considered but did not seem to improve the final forecast and were ultimately discarded.

4.5.2.4 Software Module

The data mining models developed for the different solar farms are then utilized operationally by the power conversion software system module. The module ingests the latest irradiance forecast input and then matches solar farm site identifiers to appropriate data mining models for the specific solar farms. The forecast irradiance measurements are then converted to power by applying the appropriate data mining models. Note that the software module is applying a perfect prognosis ("perfect prog") approach since the data mining models utilize actual observed irradiance values instead of forecast irradiance values for training.

4.5.2.5 Initial Evaluations

Statistics and case studies were used for the initial evaluations of the power conversion models. In order to evaluate and compare different potential models, the matched irradiance-power data set for each farm was divided into roughly two-thirds training and one-third testing sets using the division of cycling through each day of the year and placing data from every third day in the test set, leaving the remaining days for training. With this division, models could be trained on a disjoint set of data from which they would be evaluated. For the final models used in operation, however, all data were used for the model creation. In general, the power conversion errors for the different models were very low (between 1-3%), and did not differ much between the training and test sets. Table 4-1 shows the training and test set errors for several of the farms. Errors for all data as well as for data with power values above 15% of the farm capacity are shown.

Table 4-1. Training and test set Errors for several farms

Farm	Training set % Capacity MAE	Test Set % Capacity MAE	Training Set % Capacity MAE for observations > 15% Capacity	Test Set % Capacity MAE for observations > 15% Capacity
A	2.6	3.1	4.7	5.7
B	1.5	2.1	3.7	5.2
C	1.0	1.3	2.4	3.0
D	1.1	1.3	2.1	2.7
E	2.6	4.4	2.9	4.8

Case study plots showing forecasted irradiance and power along with the corresponding observations were also used in evaluations of the power conversion models. Comparing the observed and forecasted power for specific cases where the forecasted irradiance closely matches the observed irradiance give a good indication of the skill of the power conversion model. Figure 4-11 and Figure 4-12 are both good examples in which forecast irradiance and power both closely match the observations, alluding to good power conversion model performance. Figure 4-12 shows some difference between the forecast and observed irradiance but, as expected, that same difference is seen in the forecast and observed power. Figure 4-13 shows a slight difference between the forecasted and observed irradiance at the beginning of the day but matches well for the second half of the day. For this farm the power forecast matches observations closer at the second half of the day as well, however, the forecast is not as smooth through time as the power observations. This may be due to this farm's power and irradiance observations not being as linearly related as the farms shown in Figure 4-11 and Figure 4-12, as can be seen in Figure 4-7 above, making the power conversion more challenging. Further research should be performed to determine the exact cause for the conversion error in order to improve this conversion model.

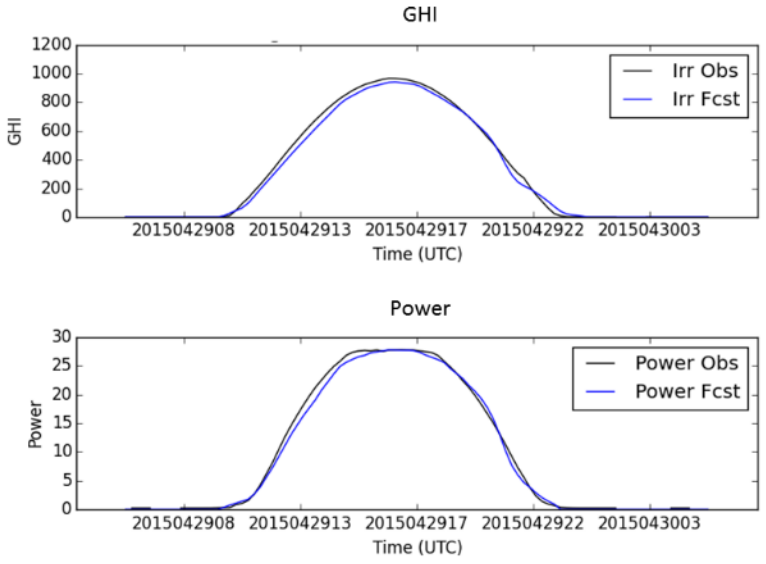


Figure 4-11. Irradiance and power forecasts and observations for a farm with GHI observations.

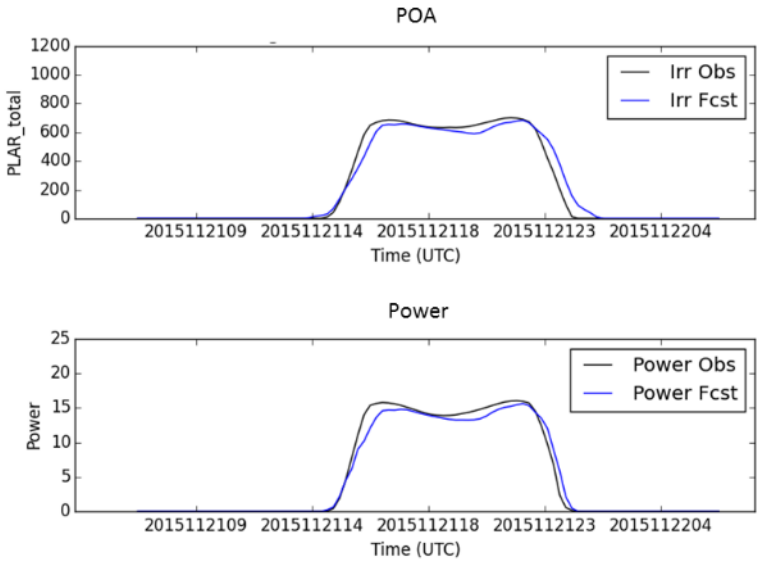


Figure 4-12. Irradiance and power forecasts and observations for a farm with plane of array observations.

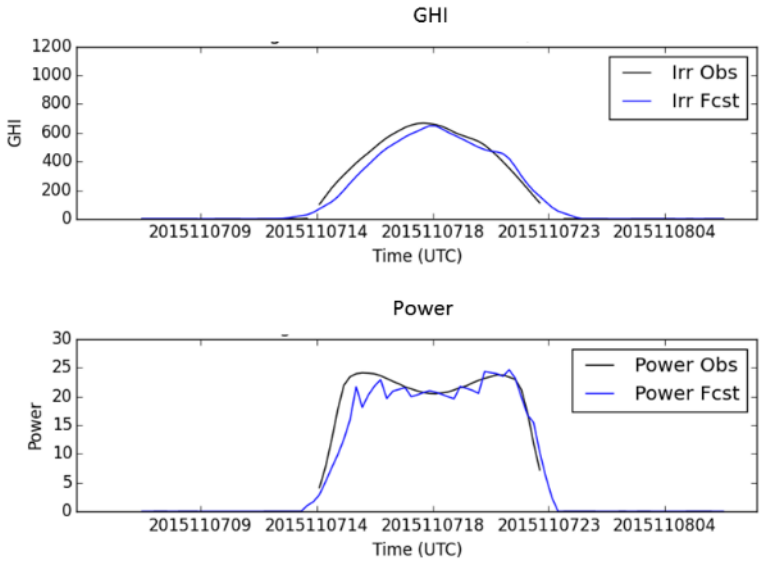


Figure 4-13. Irradiance and power forecasts and Observations for a farm with GHI observations.

4.6 Probabilistic Prediction: The Analog Ensemble (AnEn)

4.6.1 Motivation and Introduction to AnEn

The energy produced by solar photovoltaic farms has a variable nature depending on astronomical and meteorological factors. The former are the solar elevation and the solar azimuth, which are easily predictable without any uncertainty. The amount of liquid water (i.e., clouds) met by the solar radiation within the troposphere is the main meteorological factor influencing the solar power production, as a fraction of shortwave solar radiation is reflected by the water particles and cannot

reach the earth's surface. The total cloud cover is a meteorological variable often used to indicate the presence of liquid water in the troposphere. Cloud cover has a limited predictability, however, which is also reflected in the GHI and, as a consequence, in solar PV power prediction. This lack of predictability makes solar energy integration into the grid challenging. A cost-effective utilization of solar energy over a grid strongly depends on the accuracy and reliability of the power forecasts available to the transmission system operators (TSOs), both for intra-day and day-ahead forecasts (e.g., Feruzzi et al. 2016). Furthermore, several countries have in place legislation requiring solar power producers to pay penalties proportional to the errors of day-ahead energy forecasts, which makes the accuracy of such predictions a determining factor for producers to reduce their economic losses.

Predictions can be categorized into deterministic and probabilistic forecasts. A deterministic forecast consists of a single predicted value of the variable for each prediction time, while probabilistic forecasting provides probability density functions (PDFs) from which probabilities of future outcomes can be estimated. Probabilistic forecasts also provide information about uncertainty in addition to the commonly provided single-valued (best-estimate) power prediction.

There are many examples of how probabilistic predictions can provide a higher value than deterministic ones. For instance, one example is estimating the optimal level of reserves that need to be allocated to compensate for wind and solar power variability and their limited predictability, as discussed in Doherty and O'Malley (2005). Another significant application is when renewable energy is traded in day-ahead electricity markets. In Roulston et al. (2003) it is shown that trading future wind energy production using probabilistic wind power predictions can lead to higher economic benefits than those obtained by using deterministic forecasts alone. Indeed, the maximum income for a producer is obtained by offering in the day-ahead market an amount of energy that can be different from the most expected one.

The AnEn is based on an historical set of deterministic predictions of meteorological variables and solar power (predictors) and observations of the solar power. For each forecast lead time *t*, the AnEn set of solar power forecasts is constituted by solar power observations from the past. These observations are those concurrent with the past forecast at the same lead time, chosen across the past runs most similar to the current forecast. The metric used to rank past forecasts' similarity to the current forecast is defined as

$$\|F_t, A_t\| = \sum_{i=1}^{N_v} \frac{w_i}{\sigma_{f_i}} \sqrt{\sum_{j=-\tilde{t}}^{\tilde{t}} \left(F_{i,t-j} - A_{i,t+j}\right)^2} \tag{4-3}$$

where F_t is the current forecast for the lead time t at a certain location; A_t is an analog forecast for the same lead time and location as F_t; N_v and w_i are the number of physical variables and their weights, respectively, where i is a generic index referring to different variables; σ_{f_i} is the standard deviation of the time series of the past forecasts of a given variable at the same location; $\tilde{t}$ is an integer equal to the half-width of the time window over which the metric is computed (e.g., if $\tilde{t} =$ 1 hour, the distance will be computed over the three forecast lead times corresponding to hours t–1, t, and t+1); and finally, $A_{i,t+j}$ and $F_{i,t+j}$ are the values of the analog and the forecast in the time window for a given variable. The goal is to find past forecasts of the meteorological variables (chosen among the ones with the highest correlation with the quantity to be predicted — solar power) that were predicting similar values and temporal trend (i.e., which have a similar behaviour

as a function of time along the time interval defined by $\tilde{t}$) compared to the current forecast. The predictors commonly used for solar power forecasts are GHI, cloud cover (CC), air temperature at 2 m above the ground (T2M), and solar power forecasts (if available from other post-processing systems).

The assumption is that if similar past forecasts are found, their errors will likely be similar to the errors of the current forecast, which can be inferred from theirs. The main steps of the algorithm can be summarized as follows:

1. Retrieve a historical dataset of predictions (initialized at a given time) issued by a deterministic system (a meteorological model in this case) for the solar farm location.
2. Retrieve an historical dataset of solar power observations at the locations of interest.
3. Choose the physical variables from the meteorological model to be used as predictors in Eq. 1 for the predictand variable (solar power in this application).
4. For each lead time of the current forecast, compute the distance (i.e., Eq. 4.3) from every past forecast issued at the same lead time.
5. For each lead time of the current forecast, rank all the past forecasts and select the *n* forecasts with the lowest distance.
6. The concurrent *n* past measurements are the *n* members that constitute the current AnEn forecast for the lead time considered (usually 15-20 past observations are used).

The AnEn attempts to capture error dependent on a particular process of the atmospheric flow by assigning the observed errors from similar past situations, as described by the high-resolution deterministic model, to the current model forecast. The AnEn has potential advantages and disadvantages as compared to an NWP ensemble. One advantage of the AnEn is the significantly lower real-time computational expense of generating an ensemble, as AnEn only requires a single model forecast, as opposed to the multiple model runs of an NWP-based ensemble. Another advantage is that the forecast uncertainty is based solely upon past observations, thereby eliminating the need to simulate all sources of NWP forecast uncertainty via sophisticated and computationally intensive techniques, and, as will be shown, also avoiding the need for post-processing calibration. One disadvantage is the need for a “frozen” meteorological model in the training data set. Indeed, significant changes in the NWP forecast configuration may prevent the generation of skilful analogs.

4.6.2 Review of Prior Art and Improvements Made

The analog ensemble technique (AnEn) has been originally proposed by Delle Monache et al. (2011, 2013) for deterministic and probabilistic meteorological forecasting, by Alessandrini et al. (2015a) and Junk et al. (2015a,b) for wind power forecasts, and by Vanvyve et al. (2015) for wind resource assessment applications, and by Alessandrini et al. (2015b) for solar power forecasts.

AnEn has improved upon a probabilistic power forecast system based on the quantile regression (QR) method (Bremnes 2006; Nielsen et al. 2006).

Considering a random variable *Y*, *Q(τ)* is defined as the value for which the probability of obtaining values of *Y* below *Q(τ)* is τ. In QR, *Q(τ)*, with $0 < \tau < 1$, is expressed as a linear combination of known regressors and unknown coefficients, exactly as the mean of the random variable *Y* is modelled in (multiple) linear regression. Thus the *τ*-quantile is modelled as:

$$Q(\tau) = \beta_0(\tau) + \beta_1(\tau)x_1 + \cdots + \beta_p(\tau)x_p \tag{4-4}$$

where x_p are the p known regressors, also called explanatory variables, and β_p are unknown coefficients, depending on *τ*, to be determined from *N* observations.

The sample β_p coefficients for the *τ* quantile can be found by minimizing the cost function:

$$\sum_{i=1}^{N} w_i \cdot \rho_\tau\left[y_i - \left(\beta_0 + \beta_1 x_{i,1} + \ldots + \beta_p x_{i,p}\right)\right] \tag{4-5}$$

where w_i is a vector of weights (between 0 and 1) given at the observations y_i of the random variable *Y*, and ρ_τ is the check function defined as:

$$\boldsymbol{\rho_\tau(e)} = \begin{cases} \boldsymbol{\tau e}, & \boldsymbol{e > 0} \\ \boldsymbol{(\tau - 1)e}, & \boldsymbol{e \leq 0} \end{cases} \tag{4-6}$$

The coefficients β_p can be estimated with linear programming techniques. In this case, as in Nielsen et al. (2006), the *R* add-on package *quantreg* has been used.

The AnEn has improved also upon a persistence ensemble (PeEn) forecast (see below). The PeEn ensemble for each of the 72 hours ahead is made of the most recent available 20 measured PV values at the same hour. This set of values can eventually be ranked to define a set of power quantile intervals. The PeEn ensemble forecast can be accurate if clear sky conditions persist for several consecutive days. Indeed, at the same hour of the day similar levels of power are expected as the sun position is quite similar.

Moreover, AnEn has improved predictions when compared to a forecasting system based on a neural network (NN). This NN system was an artificial, feed-forward NN with a single hidden layer. The NN has a simple structure that connects input variables with a response variable, passing by one hidden layer of neurons, which processes the information.

4.6.3 Initial Validation of Solar Forecasting with AnEn

Much of the material in this subsection is derived from Alessandrini et al. (2015b).

Because the AnEn requires a long training period, it was implemented in the Sun4Cast® system toward the end of the project. Chapter 5 includes an evaluation of its performance on multiple solar plants. To research and develop the AnEn used here, the performance of AnEn was compared with QR and PeEn at three solar PV farms in Italy with a one-year dataset of measurements by evaluating important attributes of probabilistic predictions, including resolution, reliability, spread-skill and statistical consistency. In terms of deterministic forecasting, the performance of AnEn and QR were also compared with those obtainable by applying a feed-forward NN to produce a single-valued power forecast. In fact, computing the simple mean or median of an

ensemble distribution allows one to obtain a single-valued deterministic forecast. Also for this comparison PeEn was used as a baseline reference.

Data collected from three PV farms were considered in this study. They are located in Milano (northern Italy) and Catania (Sicily, southern Italy), both with a nominal power (NP) of 5.21 kW, and in Calabria region with a nominal power around 5 MW. The Milano PV farm is located in the east suburban/industrial area of Milano city. The Catania PV farm has the same set-up as that of Milano, in terms of type of photovoltaic panels and electronic components. It is located in the suburban area of Catania close to the Mount Etna volcano. Hourly averaged power data are available for the periods January 2010-December 2011 (Catania), July 2010-December 2011 (Milano), and April 2011-March 2013 (Calabria).

In order to assess the quality of the AnEn and QR systems without being limited by a specific threshold value, as is the case with reliability diagrams, Brier score, or relative operating characteristic (ROC) skill score, the continuous ranked probability score (CRPS) is computed. The CRPS is the equivalent of the Brier score integrated over all possible threshold values. In other words, it compares a full probabilistic distribution with the observations, where both are represented as cumulative distribution functions (CDF). The more the PDF originating the CDF is sharp and centered on each observation, the lower the CRPS. It is given by

$$CRPS = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{\infty}\left(F_i^{\,f}(x) - F_i^{\,0}(x)\right)^2 dx \qquad (4\text{-}7)$$

where $F_i^{\,f}(x)$ is the CDF of the probabilistic forecast and $F_i^{\,0}(x)$ is the CDF of the observation for the i^{th} ensemble prediction/observation pair, and N is the number of available pairs. The CRPS reduces to the mean absolute error (MAE) for a deterministic forecast (Hersbach 2000). A lower value of the CRPS indicates better performance, with 0 being a perfect score. The index is expressed in the same units as the forecasted variable.

In Figure 4-14 the CRPS is plotted as a function of forecast lead time for the three solar farms and the three probabilistic prediction methods, normalized by both NP (left axis) and mean measured power (MP, right axis). Next to the left vertical axes the CRPS (and its confidence interval) computed with all available lead times (considering only the hours with a positive average solar elevation) is shown. These average CRPS values for QR and AnEn are consistently lower than those of PeEn. When looking at CRPS computed independently for each lead time, in the early morning and late afternoon when the solar elevation is low, QR is at times worse than PeEn and AnEn (see for instance lead time 29 for Catania and lead time 65 for Calabria). This can be explained by the lower correlation between the forecasts and the observations when the solar elevation is low. This lower correlation affects the QR performance more than AnEn, because in QR the regression coefficients are computed considering all the data in the training data set, while in AnEn only a few cases are selected. At Milano, QR shows the lowest total average CRPS and also the best performance during the peak production hours. At Catania and Milano there is no statistically significant difference between QR and AnEn, as the total average CRPS bars overlap. To analyze the relative performances of the different forecast systems at the different plants, the CRPS can be also normalized with the MP to account for different climatic conditions (i.e.,

different MP/NP ratio). For Catania and Calabria, QR and AnEn show an average CRPS/MP of about 15%, i.e., a similar level of accuracy in similar climatic conditions. At Milano, QR and AnEn are close to 20% and 21%, respectively, i.e., they exhibit a worse performance in a climate with more cloudy days, which is expected given the reduced solar power predictability.

A quantitative assessment of the deterministic prediction quality (also referred to as spot power forecasts) has also been carried out. A straightforward way to obtain a deterministic prediction from a probabilistic one is to take, for every issued forecasted PDF, its mean or median.

A common verification framework, as suggested for WPF by Madsen et al. (2005), should be used to evaluate a deterministic forecast. The mean absolute error is used to evaluate the three different deterministic forecasts, which can be expressed as

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|o_i - f_i| \tag{4-8}$$

where o_i is the i^{th} observed value and f_i the i^{th} forecasted value. MAE allows measuring the average error magnitude in the forecasts. As a selection criterion of forecast-observation pairs in Eq. (4-8) we retained only the hours with a positive average solar elevation that are relevant for PV forecasting. The median of every issued forecast PDF is used as the f_i because it provides lower MAE than the mean value of the PDF (Gneiting 2011).

In Figure 4-15 MAE values are normalized by NP (left vertical axis) and reported as a percentage for every lead time separately and over all the lead times together (as in Figure 4-14). The MAE is also normalized by MP (right vertical axis) during the test period. The AnEn and QR show a total MAE/NP statistically significantly lower than PeEn with values close to 6-7% for all three sites. The AnEn is slightly better than QR at Catania but slightly worse at Milano, while they have similar total MAE/NP at Calabria. The QR achieves better accuracy than AnEn at the Milano site for the central hours of the first two forecast days, while again QR performance degrades for early morning and late afternoon. NN shows slightly worse results than either AnEn and QR at Catania and Calabria, exhibiting higher errors during the central hours of each forecast day, while at Milano it performs similarly to AnEn. The different climatology of the three solar farms affects the performances of QR and AnEn particularly in terms of MAE/MP. In fact, at Catania and Calabria, MAE/MP total average values are under 20% while at Milano they are around 25%. Therefore, the impact of more cloudy weather and a polluted environment in terms of predictability in the 1-72 hour range can be quantified for this data set to be 7-10% of MAE/MP. A higher impact on this index may occur when dealing with locations with weather more cloudy and higher power variability than Milano.

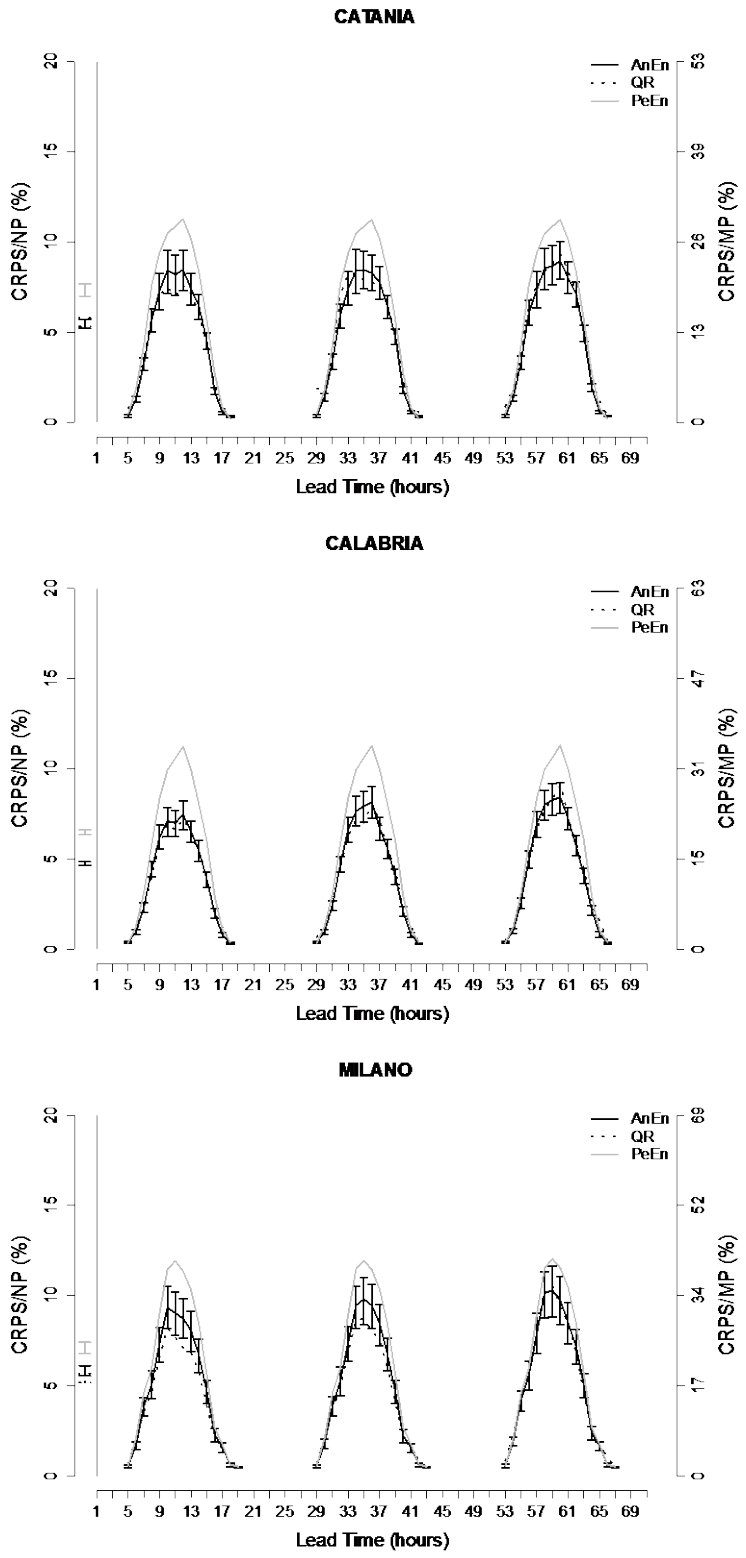


Figure 4-14. CRPS normalized by nominal power (NP, left axis) and mean measured power (MP, right axis) as a function of forecast lead time; 5-95% bootstrap confidence intervals are plotted for AnEn only to reduce clutter. The vertical bars next to the left vertical axis indicate CRPS considering all the lead times together (excluding the hours with a negative average solar elevation).

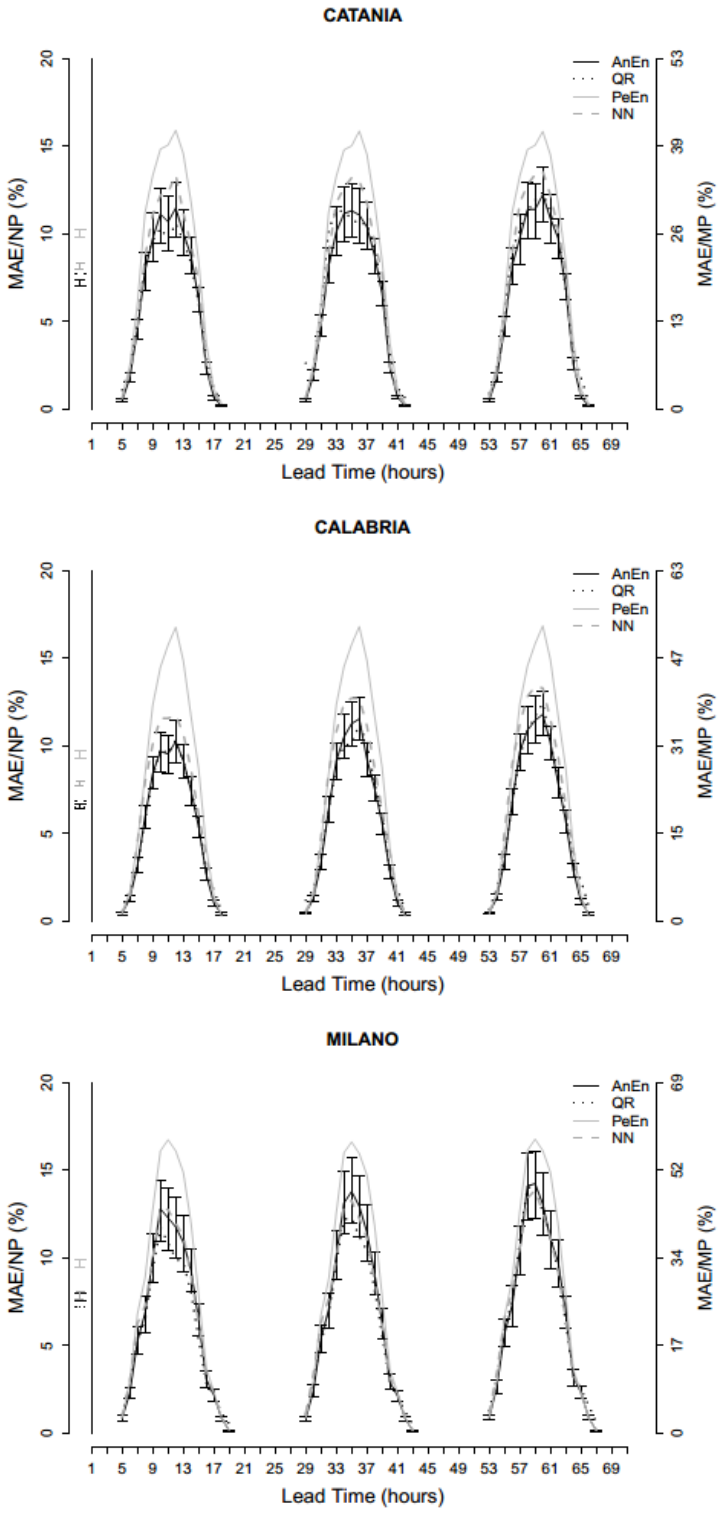


Figure 4-15. Percentage mean absolute error normalized by nominal power (MAE/NP %, left y axis) and by mean power (MAE/MP %, right y axis) of analog ensemble (AnEn), quantile regression (QR), neural network (NN), and persistence ensemble (PeEn) as function of forecast lead time. Bootstrap 5-95% confidence intervals are plotted for AnEn only to reduce clutter. Next to the left y-axis the metric confidence interval is computed by pulling all the lead times together for the hours with a positive average solar elevation.

4.6.4 Summary of AnEn and Application to SMUD Farms

Probabilistic predictions can provide accurate deterministic forecasts along with a quantification of their uncertainty, as well as a reliable estimate of the probability to overcome a certain production threshold. We have applied the AnEn as a novel method for PV power forecasting over the 0-72-hour lead time period.

The AnEn has been applied to generate probabilistic power forecasts made of 15 members over eight PV power farms located in the Sacramento Municipal Utility District (SMUD) in California.

In this application, AnEn has used as predictors the solar power, the GHI, and the T2M forecasts generated by the Cubist power conversion algorithm (section 4.5) at each solar farm. The forecasts were issued at 3:30 AM (local time) for three days ahead with a forecast step of 15 minutes for the period between 18 April 2015 and 09 February 2016. The first 149 days were used as a training period and the remaining days for verification. A specific weight optimization for the three predictors has been performed, searching for the set of weights minimizing the CRPS (Eq. 4-7) over the training period. The analog forecasts were searched over a time window equal to one hour (four time steps). The weights were kept unchanged over the verification period. Figure 4-16 and Figure 4-17 compare the AnEn mean and Cubist in terms of BIAS and RMSE. The AnEn is able to consistently reduce the BIAS of about 1 MW compared to output of Cubist. The RMSE is also reduced by the AnEn of about 1.5 MW during the hours corresponding to the higher power production rates.

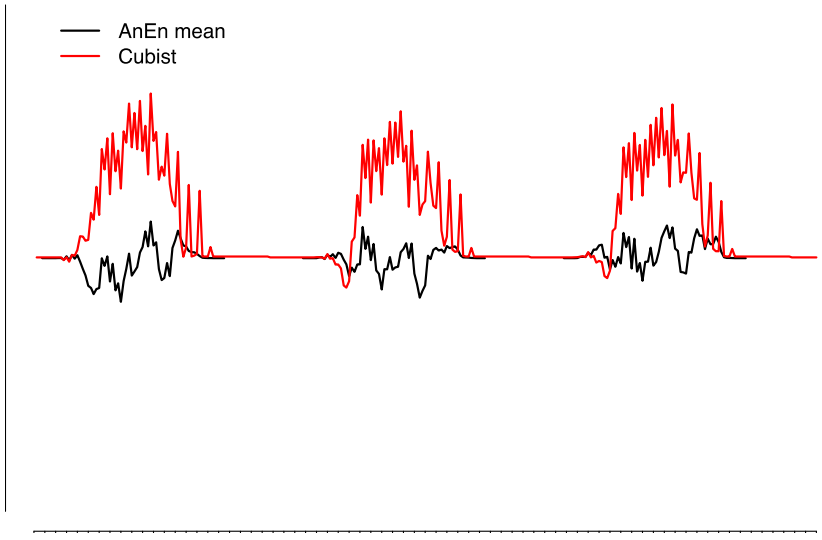


Figure 4-16. Bias as a function of forecast lead time average over the eight solar farms. The mean over the 15 members generated by the AnEn has been compared with spot forecast by Cubist.

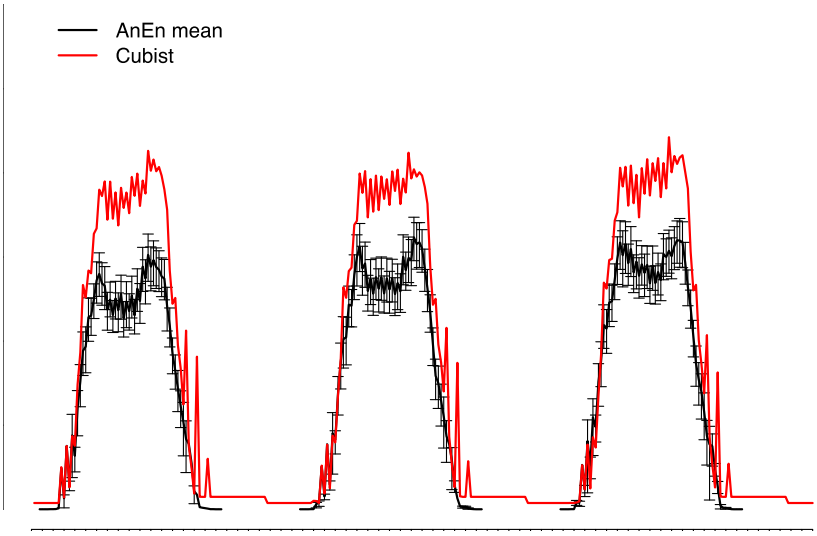


Figure 4-17. RMSE as a function of forecast lead time average over the eight solar farms. The mean over the 15 members generated by the AnEn has been compared with spot forecast by Cubist. Bootstrap 5-95% confidence intervals are plotted for AnEn only to reduce clutter.

The ability of a probabilistic prediction to quantify its uncertainty can be assessed by compiling binned-spread/skill diagrams. In a spread/skill diagram the ensemble spread is compared to the RMSE of the ensemble mean over small class intervals (i.e., bins) of spread, instead of considering its overall average. A good correlation in the spread/skill diagram is an indication that an ensemble system is able to forecast its own error (Pinson et al. 2010). If the observations are indistinguishable from the predictions, i.e., the observations and the ensemble members are samples from the same distribution, it can be shown that the ensemble standard deviation (i.e., the ensemble spread) and the RMSE of the ensemble mean should be equal. However, if the ensemble members and ensemble mean errors do not have a Gaussian distribution, the spread-RMSE matching requirement is only a necessary condition for spread-skill consistency. The latter is satisfied if the two indices match at all values (i.e., the resulting trend lies on the plot's 1:1 diagonal). Binned-spread/skill diagrams for the AnEn system are reported in Figure 4-18. Each bin has the same number of data points, which results in bins of different width. An excellent correlation index between ensemble spread and RMSE is found.

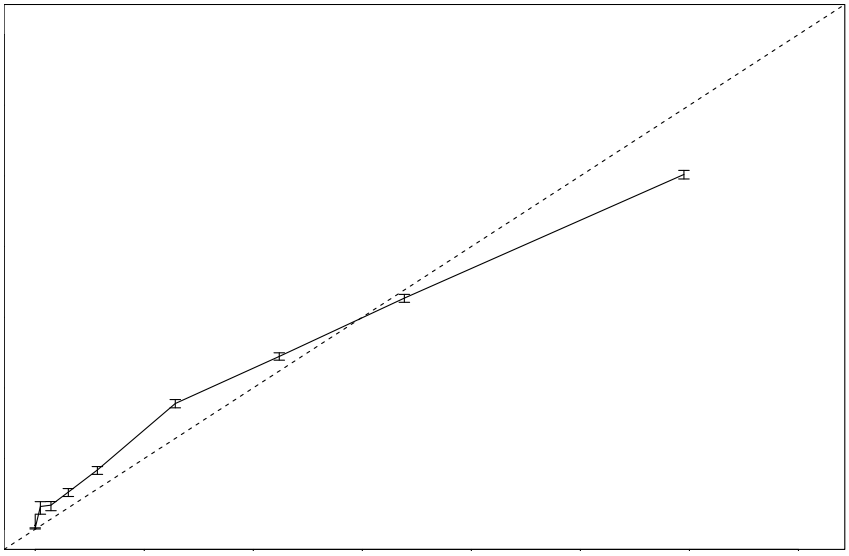


Figure 4-18. Binned Spread/skill plots for the 8 solar farms for the AnEn forecasts. Both root mean square error (RMSE) and spread values are normalized by nominal power (NP); the 5-95% bootstrap confidence intervals are plotted.

Consistency indicates whether the members of an ensemble system are statistically indistinguishable from the observations. If an ensemble system is statistically consistent, an observation ranked among the corresponding ordered ensemble members is equally likely to take any rank in the range of the whole PDF. A rank histogram can be used to assess whether the observations are equally distributed among the forecasted PDF. A perfectly uniform rank histogram would be flat, i.e., with uniform rank probability of $1/(n+1)$ where n is the number of ensemble members.

The vertical bars represent the confidence intervals and are calculated with a quantile function for a binomial distribution. The confidence intervals delimit the 5-95% quantile interval of the binomial distribution. They assess the confidence in the estimated distance from the perfect forecast (the one lying along the horizontal bar), which varies based on the available samples in each bin. In other words, with a limited number of samples a deviation from the perfect forecast could just occur by chance, while with an infinite number samples would reduce the confidence interval length to 0. The missing rate error (MRE), which is the fraction of observations lower (higher) than the lowest (highest) ranked prediction above or below the expected missing rate of $1/(n+1)$, is also shown. A larger positive (negative) MRE reveals a more under-dispersive (over-dispersive) ensemble. Figure 4-19 shows the rank histograms for the AnEn over the eight solar farms. Only the forecast lead times when the hourly average solar elevation is greater than 0 are considered. The general level of statistical consistency is excellent. The AnEn is slightly over-confident for the highest verification probability.

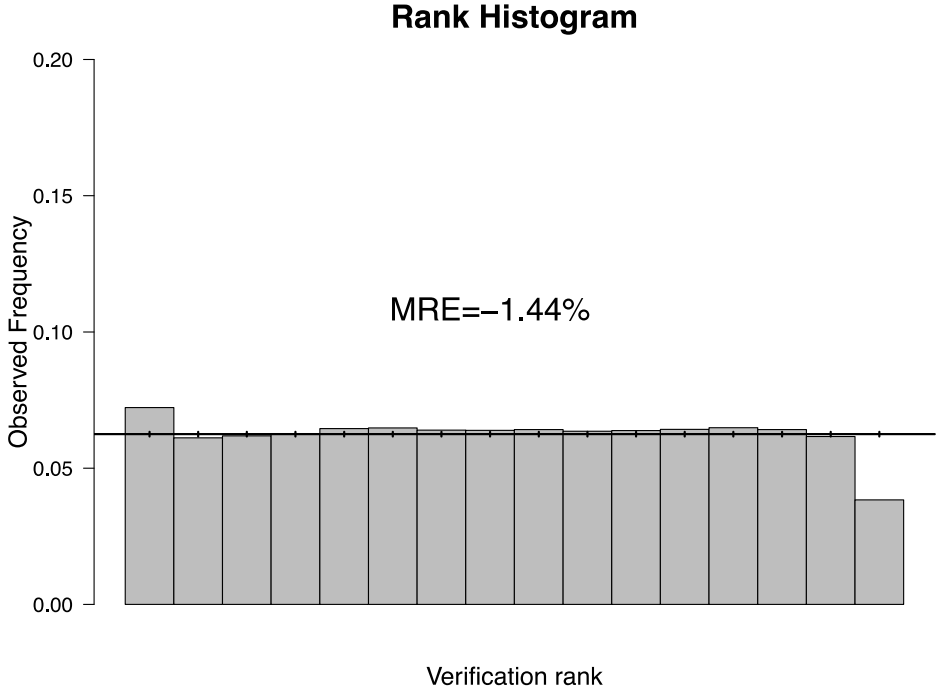


Figure 4-19. Rank histogram for Analog Ensemble (AnEn), computed over all the solar farms and all the lead times with a positive hourly average solar elevation.

4.6.5 Advances Made to AnEn

NCAR has introduced the optimization method that can be applied to define the set of weights (w_i) used in Eq. 4-3. The set of optimal weights is usually defined by choosing the combination that minimizes the CRPS over a training period. All possible combinations defined with the constraint $\sum_{i=1}^{5} w_i = 1$, where $w_i \in [0, 0.1, 0.2, \dots, 1]$, are considered.

We also considered different time shift lengths ($\tilde{t}$) in Eq. 4-3. Specifying the length of the time window on which searching for the analog forecast to be greater than 1 allows for improved statistical consistency of the AnEn forecast, especially with short training datasets.

Publications related to AnEn

Journal Papers

S Alessandrini, L Delle Monache, S Sperati, and G Cervone, Solar forecasting with an analog ensemble. Applied Energy 157, 95–110, 2015

S Sperati, S Alessandrini, P Pinson, G Kariniotakis, The "Weather Intelligence for Renewable Energies" Benchmarking Exercise on Short-Term Forecasting of Wind and Solar Power Generation. Energies, 8(9), 9594-9619, 2015

Sperati, S., Alessandrini, S., and Delle Monache, L., 2016. An application of the ECMWF Ensemble Prediction System for short-term solar power forecasting Solar Energy. Conditionally accepted, Solar Energy.

Davo', F., Alessandrini, S., Sperati, S., and Delle Monache, L., 2016. A Principal component analysis for regional wind and solar power forecasting. Conditionally accepted.

Conference Presentations

Sperati, S., S. Alessandrini, L. Delle Monache, G. Cervone, L. Harding, and S.E. Haupt, 2016: Gridded Probabilistic Forecasts of Weather Parameters with an Analog Ensemble, 23rd Conference on Probability and Statistics in the Atmospheric Sciences, AMS Annual Meeting, New Orleans, LA, Jan. 12.

Delle Monache, L., S. Alessandrini, G. Cervone, C. Junk, D. Rife, J. Ma, S. Sperati, S.E. Haupt, T. Brummet, P. Prestopnik, G. Wiener, J. Nielsen, S. Hawkins, 2015: The Analog Ensemble for Renewable Energy Applications: An Overview. International Conference on Energy and Meteorology, Boulder, CO, June 25

S Alessandrini, Solar Forecasting with an Analog Ensemble, Solar Power International, Anaheim, California, United States of America, 2015.

S Alessandrini, L Delle Monache, S Haupt, An Application of an Analog Ensemble for Short-Term Solar Power Forecasting, AMS annual meeting, Phoenix (AZ), 2015

S Alessandrini, The WIRE solar & wind forecasting benchmark exercise, Renewable Energies Forecasting - State of the art & challenges for the future "WIRE" final workshop, Paris, France, (2014). (invited presentation)

4.7 THE GRIDDED ATMOSPHERIC FORECAST SYSTEM (GRAFS)

Most of the effort in this project has focused on point forecasting because that is most relevant for commercial solar plants. For those, we can leverage those point forecasts to tune for the idiosyncrasies of the specific site and of the particular data. But for the growing deployment of distributed photovoltaic systems, the specific information necessary for the point forecasts and site-specific data is seldom available. For these applications, it is more appropriate to forecast irradiance on grid, tuned to available observation, even if those observations do not match the precise locations of the PV arrays. To that end, NCAR has leveraged some of the cost share funds to develop, tune, and deploy a Gridded Atmospheric Forecast System (GRAFS). Such a gridded system can then become the basis for regional "behind-the-meter" distributed PV power forecasts via upscaling (Lorenz et al. 2014; Haupt et al. 2016). Work developing, adapting, and tuning GRAFS was accomplished by a team of NCAR software engineers and David John Gagne, a Ph.D. student at the University of Oklahoma who was sponsored by NCAR to spend a year of his graduate studies at NCAR. Part of this work will become part of his Ph.D. dissertation.

4.7.1 GRAFS Overview

A gridded forecast system synthesizes numerical, statistical, and sometimes human weather forecasts on a regular grid and typically serves as the "last mile" of the information pipeline to the end user. We assessed the current systems and determined that they are not adequate for the purpose of developing a distributed solar power forecasting system, as reviewed below.

4.7.1.1 Current State-of-the-Art

The National Digital Forecast Database (NDFD; Glahn and Ruth 2003) has been the gridded forecast product for the U.S. National Weather Service (NWS) since the early 2000s. Forecasters in each NWS Weather Forecast Office (WFO) have been free to choose any arbitrary blend of NWP model forecast(s) as an initial gridded state, which would then be modified by hand by the forecaster before submission to the NDFD. When neighboring WFOs use different initial model blends for their grids, discontinuities at WFO boundaries frequently are readily apparent in the NDFD (Craven et al. 2013). To remedy this shortcoming in the NDFD, NWS is currently developing and recently implemented version one of the National Blend of Models, which specifies a calibrated, common blend of deterministic and ensemble forecast model output to initialize grids for all WFOs nationwide, using statistical learning and post-processing techniques (Gilbert et al. 2015, 2016).

For several years the Australian Bureau of Meteorology (BoM) and Bureau National Operations Centre (BNOC) has been using a sophisticated gridded operational consensus forecast (OCF) system (Engel and Ebert 2012; BNOC 2014). In the BoM gridded OCF, information from several component forecast models from international operational centers are blended through interpolation to a common coarse grid, model-dependent bias correction and weighted averaging on the coarse grid by comparing 15 days of historical forecasts against an analysis product, and statistical downscaling. This process yields a high-resolution, hourly, deterministic gridded forecast product over Australia, for several surface and upper-air quantities. The 2-m T and 2-m Td fields also use an “intelligent grid point selection” method near coastlines, which interpolates land points to land points and water points to water points, to preserve realistic land-sea temperature and dewpoint gradients (BNOC 2014). Furthermore, several of the surface variables are on an irregular network of grid points, with higher densities of grid points near coastlines and complex terrain, where sharper gradients are more likely, and near observation locations, which are generally located near population centers.

4.7.1.2 Vision for GRAFS

Against the background of the aforementioned consensus forecasting systems currently in existence and in development, we saw a need to develop an open-source gridded consensus forecasting system that is capable of using several statistical learning algorithms for blending component forecast modules into a single product. Our initial application of GRAFS is to global horizontal irradiance (GHI) forecasts because it is highly correlated with solar power generation, making this product useful to electrical utilities in balancing solar energy generation with the rest of their energy portfolio and demand on the grid. We label this version of the product GRAFS-Solar.

Our vision for GRAFS is to produce a gridded forecast product that blends information from multiple numerical models with smart artificial intelligence techniques, tuning to observations. This product will be modular and customizable, allowing a variety of input data and model sources, blending and optimization algorithms, and output formats. Once developed, this product will be widely accessible and applicable to the broader research community as open source software. When needed for implementation with proprietary data, however, GRAFS can be highly customizable and will leverage existing blending technology, including DICast®.

4.7.2 General Structure of GRAFS

Figure 4-20 depicts the general framework of GRAFS-Solar. NWP models serve as the basis for the gridded product. One can use a single model, tuned to observations, or blend a host of models, such as described above for DICast® above. The base model (initially the North American Model – NAM) is downscaled to a standard grid spacing, with 4-km spacing provided as an example for the plots below. Each model that is input has its own interpretation of averaging time and valid time of the hourly averages of the downward shortwave radiation. Thus, it is imperative to use smart interpolation to make consistent hourly averages.

Observations are also ingested for as many sources as possible, standard MADIS or SURFRAD sites, or from specialized networks like state mesonets, or from specific sites where solar forecasts are desired (such as utility sites). Those observations are also interpolated to the grid and used to compute the difference from the raw forecast. Finally, the interpolated differences are added back to the initial grid forecast at each grid point. The tuned forecast impacts beyond a single grid point and is spread to neighboring points via a natural neighbor approach.

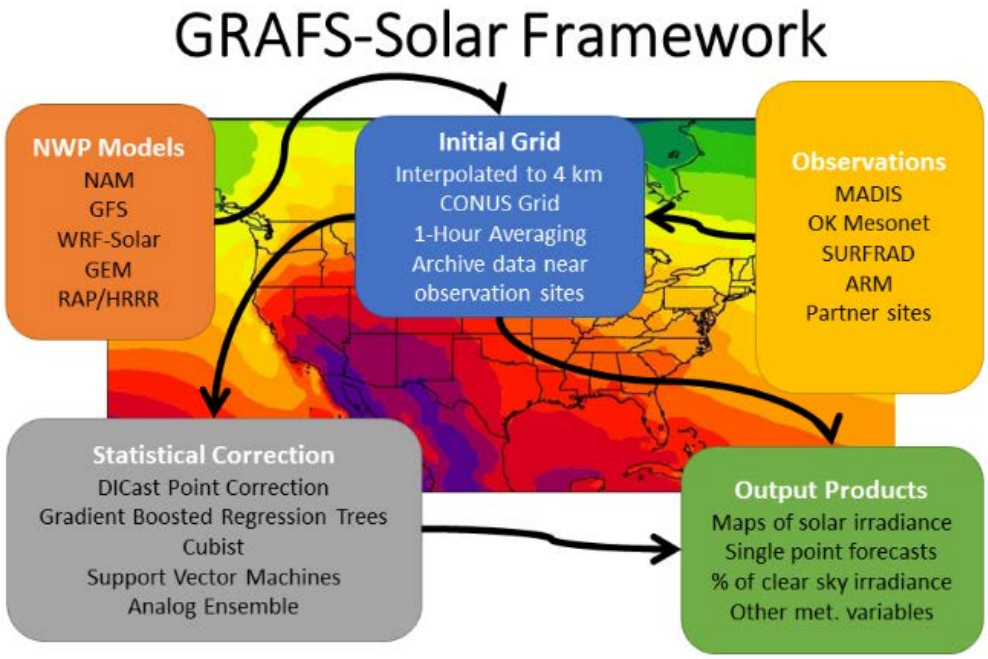


Figure 4-20. General framework of GRAFS-Solar.

Figure 4-21 shows the process of using artificial intelligence or machine learning (ML) to correct the initial forecast. One can derive multiple variables for the model output and correlate the metadata about the location, time, solar position, etc. This becomes the merged training data that is used for training the ML model. Those trained models together with the merged data are used to predict corrections at the observation sites. The corrections are applied and spread to the neighboring grid points. A variety of methods have been applied to produce those corrections, including the Cubist regression tree model that is also used in the power conversion model described above, random forests, gradient boosted regression trees, and multilinear regression.

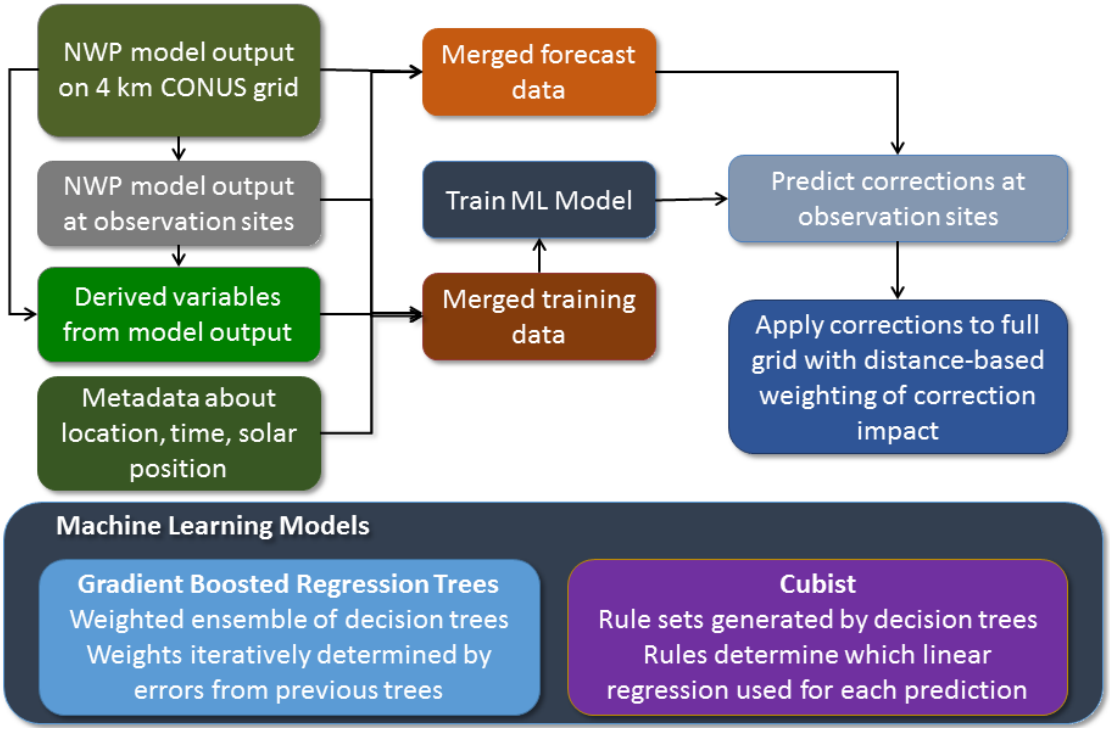


Figure 4-21. Correcting the gridded forecast with machine learning methods.

4.7.3 Exampled of Initial Results with GRAFS

The output of GRAFS can be viewed as time series at a point or as two-dimensional plots of the gridded forecast variables. These views will be expanded in the future. In addition to solar irradiance (GHI), one can also tune and plot temperature, wind speed, and relative humidity fields.

Figure 4-22 displays a line plot of GHI at a particular site in the SMUD network over a 72-h mostly clear period. A gridded plot of GHI appears as Figure 4-23. In that plot, who can discern the blockage of GHI by the clouds due to passing synoptic systems.

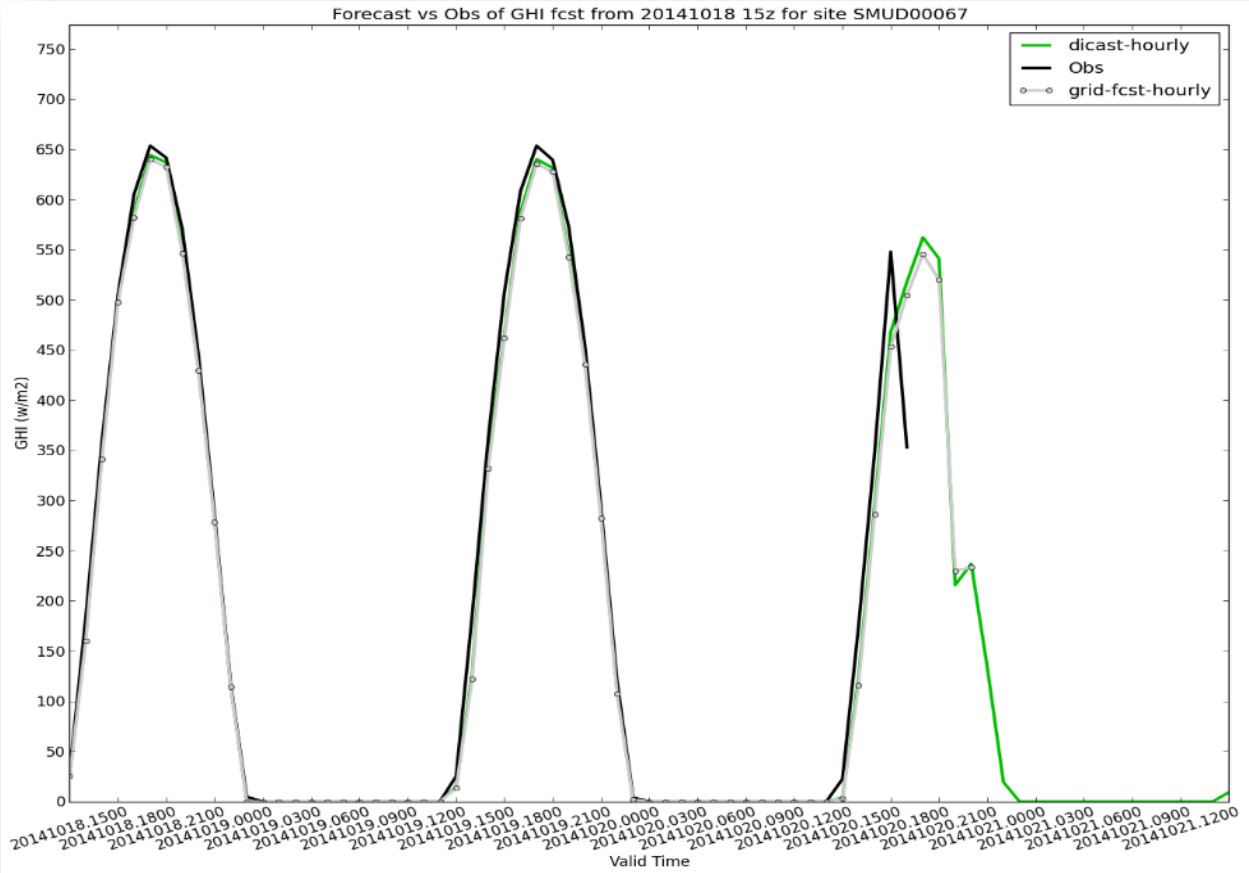


Figure 4-22. Time series plot of GRAFS GHI output at a SMUD site. The tuned forecast (green lines) match well the observations (dark black lines).

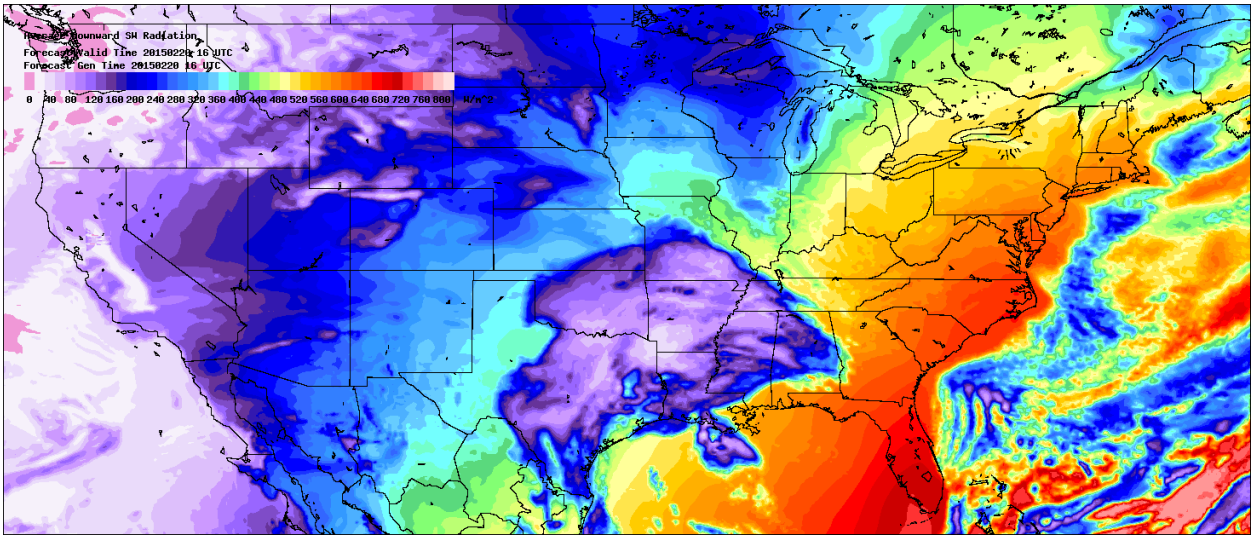


Figure 4-23. Plot of GHI over CONUS valid for 20 Feb 2015 at 1600 UTC.

The value of ML corrections are displayed in Figure 4-24. In that plot the mean absolute error of the raw forecast that was interpolated to 4 km based on the NAM is shown as the blue curve. Two ML techniques are compared, random forest and gradient boosted regression trees. The large improvement by using the ML techniques accounts for up to 60% improvement over the error of the raw forecast.

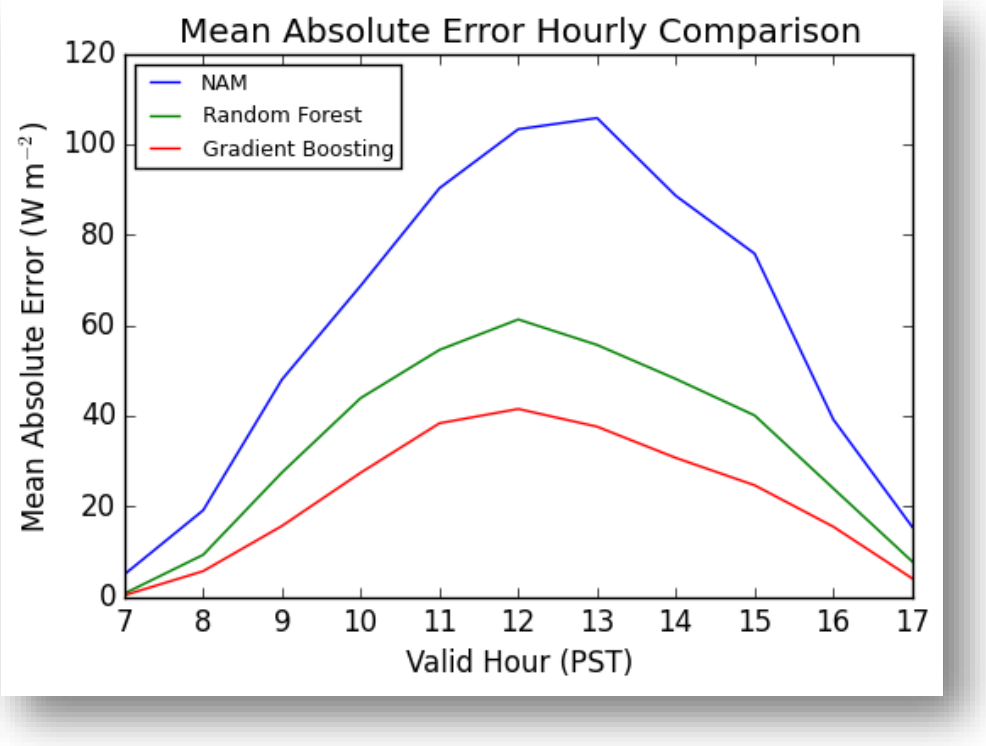


Figure 4-24 Mean absolute error at each forecast hour of the raw NAM forecast compared to that of two ML methods.

The value of GRAFS as a research tool to compare the different ML methods is expanded in Figure 4-25. In this case, we wish to discern the value of training to each site individually (MAES plotted on the ordinate) as compared to merging data from multiple sites (MAEs on the abscissa). The resulting scatter plots compare results for forecasts from the raw Global Forecast System (GFS) model forecast with those from applying linear regression, random forests, and gradient boosted regression trees. The errors for the raw GFS are large and there is no discernable difference between whether the observation points are merged or not. All three tuning methods reduce the MAE of both the single and merged sites. The MAEs of the random forecast are quite reduced and the differences between the single and merged models are small. Errors for the linear regression forecasts exceed those of the random forest forecasts and errors may be a bit higher for the merged site models. The forecast errors using the gradient boosting method are the smallest, but there is a distinguishable higher error when tuned for single sites than when they are merged. This observation indicates that gradient boosting may be an efficient technique for tuning the forecasts, but it works best with a larger amount of observational data than may be available if tuned to each site individually. This example shows how GRAFS can be used for research into the usefulness of different AI techniques for tuning forecasts.

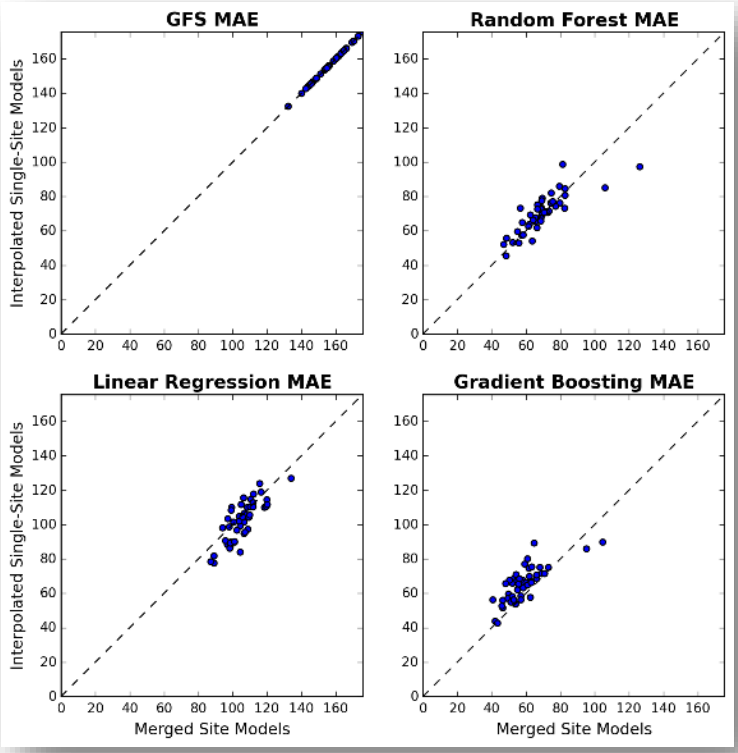


Figure 4-25 Comparison of three ML tuning techniques to the raw GFS forecasts via scatter plots of MAE as computed for a single site model vs. a model built from data merged from multiple sites.

4.7.4 Scientific Output related to GRAFS

Students Sponsored by this work:

David John Gagne, Ph.D., Atmospheric Science, University of Oklahoma, expected August 2016. Dissertation: "Coupling Data Analytics Techniques and Numerical Weather Prediction Models for High-impact Weather Prediction."

Journal Papers based on GRAFS:

Several expected within 6 months of publishing this report.

Conference Presentations Based on GRAFS:

Gagne, D.J., S.E. Haupt, S. Linden, and G. Wiener, 2016: An Evaluation of Statistical Learning Methods for Gridded Solar Irradiance Forecasting, Joint Session between 14th Conference on Artificial and Computational Intelligence and its Applications to the Environmental Sciences and Seventh Conference on Weather, Climate, Water, and the New Energy Economy, AMS Annual Meeting, New Orleans, LA, Jan. 12.

Gagne, D.J., S.E. Haupt, S. Linden, G. Wiener, 2015: A Community Gridded Atmospheric Forecast System for Calibrated Solar Irradiance. International Conference on Energy and Meteorology, Boulder, CO, June 24.

Linden, S., D.J. Gagne, and S.E. Haupt, 2015: Initial Implementation of the Solar Gridded Atmospheric Forecast System (GRAFS-Solar), NCAR retreat, Dec. 8.

Gagne, D.J., S.E. Haupt, S. Linden, J.K. Williams, A. McGovern, G. Wiener, J.A. Lee, and T.C. McCandless, 2015: Scaling Machine Learning Models to Produce High Resolution Gridded Solar Power Forecasts, 13th Conference on Artificial Intelligence, AMS Annual Meeting, Phoenix, AZ, Jan. 7.

5 ASSESSMENT

5.1 SUMMARY OF ADVANCEMENTS

The primary advancements that have been achieved in this public-private-academic partnership have come in the development and improvement of several forecast models. Each of the component models in the Sun4Cast® system, which were described in detail in chapters 2 and 3, represents a step forward in solar energy forecasting, both in the nowcasting and day-ahead time frames.

TSICast (section 2.1) is a short-term solar forecast system utilizing ground-based sky cameras. The system adopts multi-angle observations to undertake the task of cloud tracking based on spatial and temporal correlation, and provides a pipeline to detect multi-layer motions via clustering. The robust feature extraction and irradiance models are then vetted for real production forecasts. Compared with single-camera models, TSICast significantly enlarges the field of view, enables 3-D cloud tracking, and obtains more accurate forecasts. As is shown in Figure 2-20, TSICast improves on persistence GHI forecasts by an average of 25-30% in the 5-15 minute range, and by 30-40% in the 1-5 minute range, in testing performed for the network of 25 pyranometers at BNL.

StatCast (section 2.2), which has several versions that were built for this project, uses machine learning techniques to predict GHI out to 3 h, based on the most recent hour of GHI observations, nearby METAR observations, and a training dataset. StatCast-Persistence is a "smart persistence" technique that predicts persistence in clearness index, automatically adjusting for the continuously changing solar zenith and azimuth angles in providing GHI forecasts. StatCast-Persistence is the baseline against which all the short-range models are evaluated later in this chapter, in section 5.2. StatCast-Cubist uses the Cubist model tree to define a set of rules using the training dataset, which are then applied to yield a GHI forecast at each lead time. Compared to StatCast-Persistence, StatCast-Cubist yielded improvements of generally 35-50% at all lead times on clear days, and 10-50% on cloudy days, with improvements getting larger with increasing lead time (Figure 2-37). StatCast-RD (regime-dependent) breaks up the sky condition into different cloud regimes using k-means clustering, and trains a separate artificial neural network (ANN) for each regime. As Figure 2-39 shows for cloudy days over the SMUD pyranometers, various versions of StatCast-RD frequently improve upon StatCast-Persistence by generally 15-25% for lead times of 1 to 3 hours, with the best performance in the version that incorporates cloud data from the GOES-East satellite.

CIRACast (section 2.3) is a model that identifies cloud features in visible satellite imagery, and then advects those cloud features in a NWP wind field. Much of CIRACast was developed prior to this project, but several advances to CIRACast were made during this project, including accounting for satellite viewing parallax in GOES cloud height retrievals, improved cloud shadowing algorithms, and accounting for differential steering effects. As seen in Figure 2-41 and Figure 2-42, relative MAE values of GHI forecasts computed for two SURFRAD sites were 9.6% for Desert Rock, NV (beating the project target error rate of 10% for "simple" sites), and 21.8% for Table Mountain, CO (nearly achieving the project target error rate of 20% for "complex" sites).

MADCast (section 2.4) is a model that identifies cloud features from IR satellite imagery, and then advects and diffuses those cloud features as tracers in a stripped-down version of WRF that does not have physical parameterization schemes, in order to achieve faster performance. MADCast was developed prior to this project, but enhanced and configured for this project to produce GHI forecasts out to 6 h. Compared to a purely NWP solution (WRF-Solar-Now, discussed next), MADCast has generally 5-20% lower errors over the first 1.5 hours of the forecast (Figure 2-44). The MADCast analysis is generally a good one, although errors steadily increase with time due to the lack of physics schemes to develop and dissipate clouds.

WRF-Solar-Nowcasting (section 2.5) is a version of the WRF-Solar™ NWP model (chapter 3; summarized below) configured for nowcasting (0-6 h) purposes. Among the developments for WRF-Solar-Nowcasting was the implementation of cloud-radiation feedbacks in the model, and improved representation of unresolved clouds. Incorporating these feedbacks is shown in Fig. Figure 2-46 to reduce GHI bias errors (across all SURFRAD and ISIS sites) by as much as 50-70% over forecasts without these feedbacks, especially for lead times in the 3-6 hour range.

MAD-WRF (section 2.6) is a coupled model that was developed late in the project as it became obvious that combining the advantages of WRF-Solar-Nowcasting with MADCast satellite-based cloud assimilation could be beneficial. As Figure 2-44 shows, MADCast has lower errors than WRF-Solar™ in the first 1.5 hours, after which point WRF-Solar™ has the lower errors. Consequently, MAD-WRF has been configured to use MADCast cloud fractions and cloud mixing ratios for 0-1 h, WRF-Solar™ cloud fractions and mixing ratios for 1.5-6 h, and an average blend of the two for 1-1.5 h.

Several case studies of the nowcasting models (section 2.7) were conducted for different sky cover regimes around Sacramento, California. These case studies found that each of the components (StatCast-Cubist, CIRACast, MADCast, and WRF-Solar™) had their strengths in various regimes, times of day, and lead times, further illustrating the need to have this wide array of forecast models in a single system like Sun4Cast® .

The WRF-Solar™ model, which is used to make day-ahead GHI forecasts, was highlighted in-depth in chapter 3. The primary enhancements that set WRF-Solar™ apart from the standard WRF NWP model include improved representation of the aerosol-radiation feedback, incorporation of cloud-aerosol interactions, and improved representation of cloud-radiation feedbacks, in addition to improving the handling of the solar position/equation of time calculation and making all three irradiance components (GHI, DNI, DIF) available for output every model time step as desired. These improvements have led to greatly reduced errors in GHI predictions. Figure 3-8 highlights this well for all the SURFRAD sites, where, for clear skies, errors in GHI predictions from WRF-Solar™ improved upon standard WRF by a remarkable 40-60%.

5.2 Systematic Evaluation

At the beginning of this project, the Department of Energy (DOE), the National Oceanic and Atmospheric Administration (NOAA), the National Center for Atmospheric Research (NCAR), National Renewable Energies Laboratory (NREL), Argonne National Laboratory, and IBM worked together on developing metrics to assess the quality of tools for solar irradiance and solar power production forecasting. These metrics were determined through feedback gathered from

stakeholders during several workshops and discussions over the first 14 months of the project (Budget Period 1 – BP1). An NCAR technical note has been developed to allow interested users access to the team's findings (Jensen et al. 2016). It contains additional background on this effort and the workshops as well as summarizes feedback from stakeholders. The technical note covers only metrics that pertain to the quality of the solar forecasting tool (statistical metrics) and its economic impacts and user benefits (economic metrics). Customized metrics that evaluate specific components of the solar forecasting tool and other processes, such as successful user integration of solar forecasting into grid operations, are not included in this report.

This section focuses on the quality of the Sun4Cast® forecast system. The economic impacts are covered in section 5.3. Table 5-1 summarizes the metrics identified in the first budget period. It includes a set of six base metrics that were proposed to meet the needs of multiple users. Four of these metrics—Mean Absolute Error, Root Mean Square Error, Distribution of Forecast Errors, and Categorical Statistics—provide statistical insights to the performance of forecasts in order to aid the comparison of different forecast models or to validate a single forecast model output against actual values. The two remaining base metrics—Operating Reserves Analysis and Production Cost—translate forecast performance to economic value that will provide insight into the benefits realized with improved accuracy of forecasts. Additionally, enhanced metrics are also proposed to provide more detailed insights into forecast performance under various scenarios. Each proposed metric is defined in that technical note with emphasis on the insight provided by the metric, its calculation process, and its application as illustrated with an example of a use case.

The verification system was developed during the second year of the project and augmented in third year to include sky condition information and ramp identification and scoring. The system uses the Model Evaluation Tools (MET) Statistical-Analysis tool to compute the verification measures and the METViewer database and display system to aggregate the results. Table 5-2 summarizes the dates of forecast availability and summarizes the metrics calculated for the milestones associated with this subtask. Because the partners in the project are located in a range of locations – the eastern (BNL), central (Xcel), and western US (SMUD and SCE) – the results represent extensive geographical diversity. The evaluation period extends from January 1, 2015 to March 20, 2016. Many of the milestones were based on improvements from the beginning of 2015 to the beginning of 2016. The 2015 data extended from January to May 2015. Thus, the dates for 2015 and 2016 do not completely overlap; hence, for milestone performance comparisons the period January – March 2015 was used for 2015 and January – March 2016 was used for 2016. However, results for the intervening months is also provided for completeness. It should be noted that TSICast provides a forecast out to 14 minutes and hence is not included in the full Sun4Cast® system. For this reason, the performance of TSICast was evaluated using a separate system and these results are also included in this document.

Table 5-1. Matrix of Solar Metrics developed during BP1 and broken down into Base and Enhanced Comparison and Economic Value.

	Model-Model Comparison	**Economic Value**
Base	Mean Absolute Error Root Mean Square Error Distribution (including Statistical Moments and Quantiles) Categorical Statistics for Events	Operating Reserves Analysis Production Cost
Enhanced	Maximum Absolute Error Pearson's Correlation Coefficient Kolmogorov-Smirnov Integral Statistical Tests for Mean and Variance Renyi Entropy OVER Metric Brier Score Receiver Operating Characteristic (ROC) Curve Calibration Diagram Probability Interval Evaluation Frequency of Superior Performance Performance Diagram for Events Taylor Diagram for Errors	Cost of Ramp Forecasting

Table 5-2. Month when forecast component became available for assessment.

Component	SMUD	SoCal Edison	BNL	Xcel
StatCast	Apr2014	Jul2014	Jul2014	Apr2015
TSICast	Jan 2015	Jan 2015	Jan 2015	Jan 2015
CIRACast	Jun2014	Jul2014	Jul2014	Jan2015
MADCast	Oct2014	Oct2014	Oct2014	Jan2015
WRFSolar	Sep2014	Sep2014	Sep2014	Jan2015
WRFSolarNow	Sep2014	Sep2014	Sep2014	Jan2015
MAD-WRF blend	Aug2015*	Aug2015*	Aug2015*	Aug2015*
Power Conversion	Apr2014	Jul2014	Jul2014	Apr2015
AnEn (GHI)	Jun2014	Jul2014	Jul2014	Oct2015
NowCast	Apr2014	Jul2014	Jul2014	Apr2015
DICast	Apr2014	Jul2014	Jul2014	Apr2015

NOTE: *Available for 14 and 16 UTC initialization times only.

Several baselines are available for this evaluation, including persistence with knowledge of sky condition for NowCast components (Smart Persistence, as described in section 2.2.5.1) and publicly available numerical weather prediction (NWP) models for both the NowCast and DICast® components. Table 5-3 provides details describing the baselines.

As shown in Table 5-3, the publicly available NWP baselines that were evaluated included the North American Mesoscale model (NAM), the Global Forecast System (GFS), the Global Environment Multiscale Model (GEM), and the High Resolution Rapid Refresh (HRRR). The HRRR is run at two locations: NOAA's National Centers for Environmental Predication (NCEP) and NOAA's Earth System Research Laboratory's Global Systems Division (GSD). The HRRR run at NCEP is the publicly available operational model and is based on WRFv3.5.1 (similar to WRF-std baseline). The HRRR run at GSD is the development version in preparation for transition to operations in the next few months and includes many of the advancements in the WRF-Solar™ package as well as other developments.

Table 5-3. Characteristics of baselines for metrics evaluation.

Baseline	**Abbrev-iation**	**Comparison with:**	**Fcst available**	**Comments**
Smart Persistence based on initial StatCast	SmartP	NowCast	0-6 hr with 15 min increment	Requires past irradiance obs; when not available, constant clearness index of 0.6 used in lieu of obs. See NowCast section for complete description.
Standard Version of WRF	WRF-std	NowCast and DICast	0-48 hr	WRF version 3.5.1 with WRF solar development removed
Power Obs	Obs	Power Conversion	0-72 hr	Observations from partner locations
Publicly available NWP	NAM	NowCast and DICast	0-72 hr	Available via NOAA/NCEP 12 km -– hourly
	GFS	DICast	0-72 hr	Available via NOAA/NCEP 0.25 deg (~30 km) – 3 hourly
	GEM	DICast	0-72 hr	Available via Environment Canada – 25km – 3 hourly
	HRRRops	NowCast	0-15hr	Available via NOAA/NCEP – 3 km -– hourly
	HRRRx	NowCast	0-15hr	Available via NOAA/GSD – includes WRF-Solar™ improvements – 3 km - hourly

Figure 5-1 provides a picture of the seasonal performance for partly cloudy conditions aggregated over three partners, BNL in New York, Xcel in Colorado and SMUD in California. Smart Persistence MAE tends to be lowest during late summer and fall, moderately high in spring and early summer with the highest MAEs reported during the winter. MAE values appear to be slightly higher for the 2015-2016 winter, which is likely caused by increase cloudiness due to El Niño conditions. The components of NowCast follow a slightly different trend with low MAE scores during winter 2014-2015 and a peak similar to Smart-Persistance during spring and early summer 2015. The components have varied responses during the transition to the 2015-2016 winter. WRF-Solar-Now appears to have the most consistent performance amongst the components.

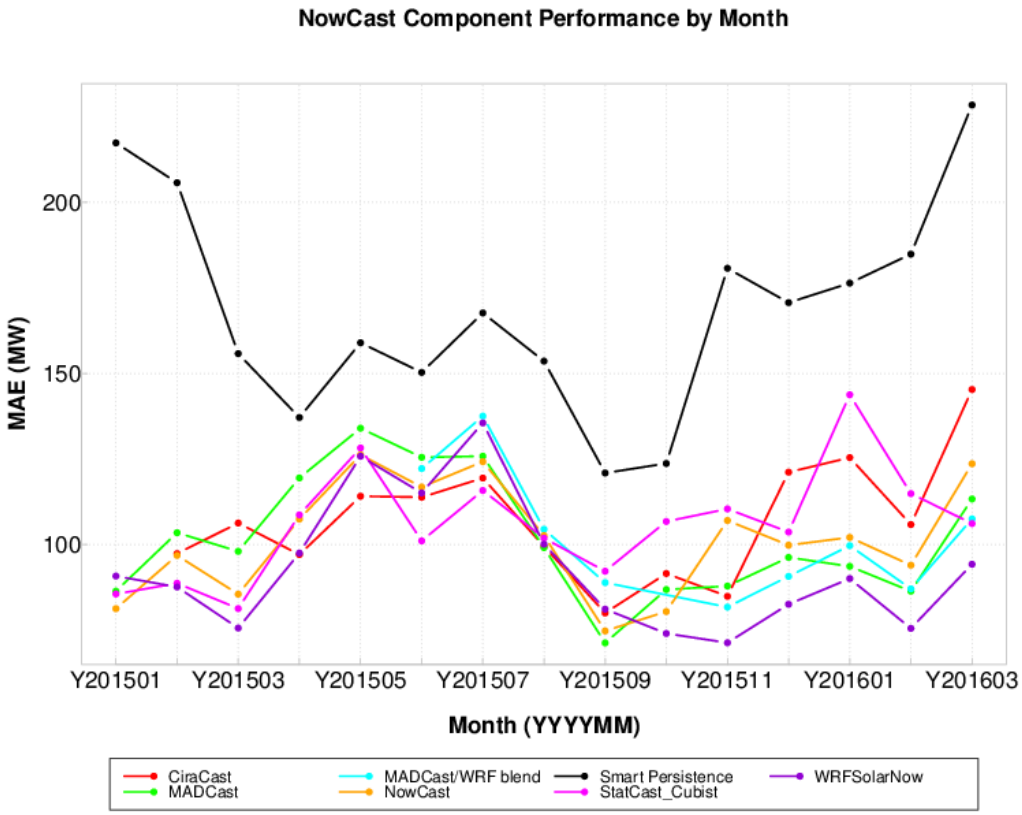


Figure 5-1. Median MAE in W m^{-2} for SMUD stratified by month.

5.2.1 NowCast System and Components

One purpose of exploring multiple nowcast components is that each one is potentially skillful for a different forecast horizon (lead time) and sky condition. Understanding the strengths and weaknesses of each system can lead to development of an expertly blended system. Components of the system include observation-only forecasts such as TSICast, for forecast on the scale of minutes and StatCast and CIRACast, with 15-min forecasts out to 3 h during daylight hours; and NWP-driven components like MADCast, WRF-Solar-Now with 15-min forecasts out to 6 hours. All components are used by the NowCast blending algorithm to develop a statistical forecast. Additionally, the DICast® forecasts derived from the operational NWP was interpolated to 15-min intervals and made available for this analysis (called DICast-Now). The baseline for the NowCast system comparisons is the StatCast Smart Persistence technique developed toward the beginning of the project. It was extended from 3 hours to 6 hours to allow for a complete comparison. This section will discuss the evaluation of each of these components in the temporal resolution order established in this paragraph.

TSICast provides 1-min forecasts out to 12-15 minutes depending on the weather conditions. The MAE values for TSICast provided in Table 5-4 indicate the largest errors are for the 14-min forecasts, with an MAE of 64.1 W m^{-2}. The forecasts were evaluated using the verification library within the R statistics package. The bias, RMSE, skewness and kurtosis listed in Table 5-4 represent the first, second, third, and fourth moments of the error distribution.

Table 5-4. Scores for TSICast error distribution including, bias (or mean error), mean absolute error, root mean square error, inner quartile range, maximum value, minimum value, and skewness. Units for all scores are W m^{-2}.

Lead Min.	Bias	MAE	RMSE	IQR	Max	Min	Skewness	Kurtosis
0	-7.4	35.8	94.3	10.38	1042.8	-1636.4	-0.41	27.4
2	-6.6	43.0	104.2	18.4	1064.6	-2170	-0.23	22.6
4	-6.1	48.4	111.0	26.8	1073.5	-1639.68	-0.14	19.2
6	-7.7	52.6	116.0	34.3	1105.5	-2192	-0.25	17.8
8	-8.9	57.2	121.1	41.4	1113.5	-1662.97	-0.18	15.7
10	-9.3	60.8	125.2	48.3	1147.7	-1636.4	-0.21	14.5
12	-9.0	63.0	126.7	54	1132.8	-2170	-0.20	14.1
14	-8.8	64.1	124.8	61.2	1134.9	-1639.68	-0.28	13.8

From the Metrics Technical Report (Jensen 2016), a positive skewness of the forecast errors leads to an over-forecasting tail while a negative skewness leads to an under-forecasting tail. All values of skewness for TSI are negative and hence suggest an under-forecasting bias, which is supported by the bias (or mean bias error – MBE) shown in the 2nd column of Table 5-4. Positive kurtosis indicates a narrow distribution; while a negative kurtosis value indicates a flat wide distribution. All kurtosis values are positive in Table 5-4, suggesting that the error distribution is narrow. The inner quartile range (IQR) represents the distance between the 25th and 75th percentile of the distribution and generally represents the range of the most likely scores. The IQR increases with time from 10 to 60 W m^{-2} indicating a rapidly expanding error distribution within 14 minutes. The MAE of the TSI forecasts at 14 minutes is 64.1 W m^{-2}. Similarly, the MAE at 15 minutes from the NowCast system for Smart Persistence was 97.3 W m^{-2}. Using this value as the baseline, the percent improvement of TSICast over Smart Persistence is 34%.

5.2.2 Blended NowCast System

Observationally-based methods like StatCast and CIRACast are expected to perform best in the 0-3 h time window. These methodologies are also expected to perform best when the conditions are relatively constant (e.g. clear or cloudy conditions). Like all statistical methods, the skill of the forecasts is also driven by the amount of data upon which to train the method. NWP-based methods like MADCast, MAD-WRF, and WRF-Solar-Now extend out to 6 hours. These methodologies may add skill to variable cloud conditions but may also be subject to double penalty errors due to displacement of clouds within the simulation. The skill of the forecasts is driven by the initial conditions and physics parameterizations of the method. Weights were assigned by initial skill of the component for use in the NowCast blended system. It was intended these weights would be adjusted as the project continued and the components improved but time did not allow for this optimization.

The partners where solar penetration is the largest, for example the California utilities SMUD and SCE, provided data earlier than others. Figure 5-2 through Figure 5-4 show component performance stratified by partner. Scores for BNL (blue), Xcel (orange) and SMUD (green) are included. SCE provided observations in plane-of-array rather than GHI and thus did not have forecasts from many of the components.

In Figure 5-2 it is clear that most scores for each partner tend to cluster. The baseline, StatCast-Smart-Persistence, has large errors with MAE values ranging from 100-250 W m^{-2}. The forecasts for BNL appear to have the least error throughout the 15-month period. StatCast-Cubist was trained late in the project and forecasts were produced retrospectively for all SMUD sites but only one site at BNL and Xcel retrospectively. At SMUD, StatCast-Cubist MAE values are approximately 50% lower than Smart-Persistence during the winter months.

Figure 5-3 indicate that the forecasts for CIRACast and MADCast at BNL and Xcel locations are much higher than SMUD at the beginning of the 15 month evaluation period (data for training models was available later in the project than for SMUD) but equilibrate with it during winter 2016. While CIRACast forecasts were provided for Xcel locations, their time-stamp precluded the forecasts from being included in this evaluation. A similar pattern is depicted in Figure 5-4 MAE at SMUD (green) tends to remain between 50 and 100 W m^{-2} throughout 2015 and into 2016, while at BNL it started between 100-150 W m^{-2} and dropped to similar scores as SMUD after August 2015. Similarly, scores at Xcel started between 150-200 W m^{-2} and also dropped to similar lower values in late summer.

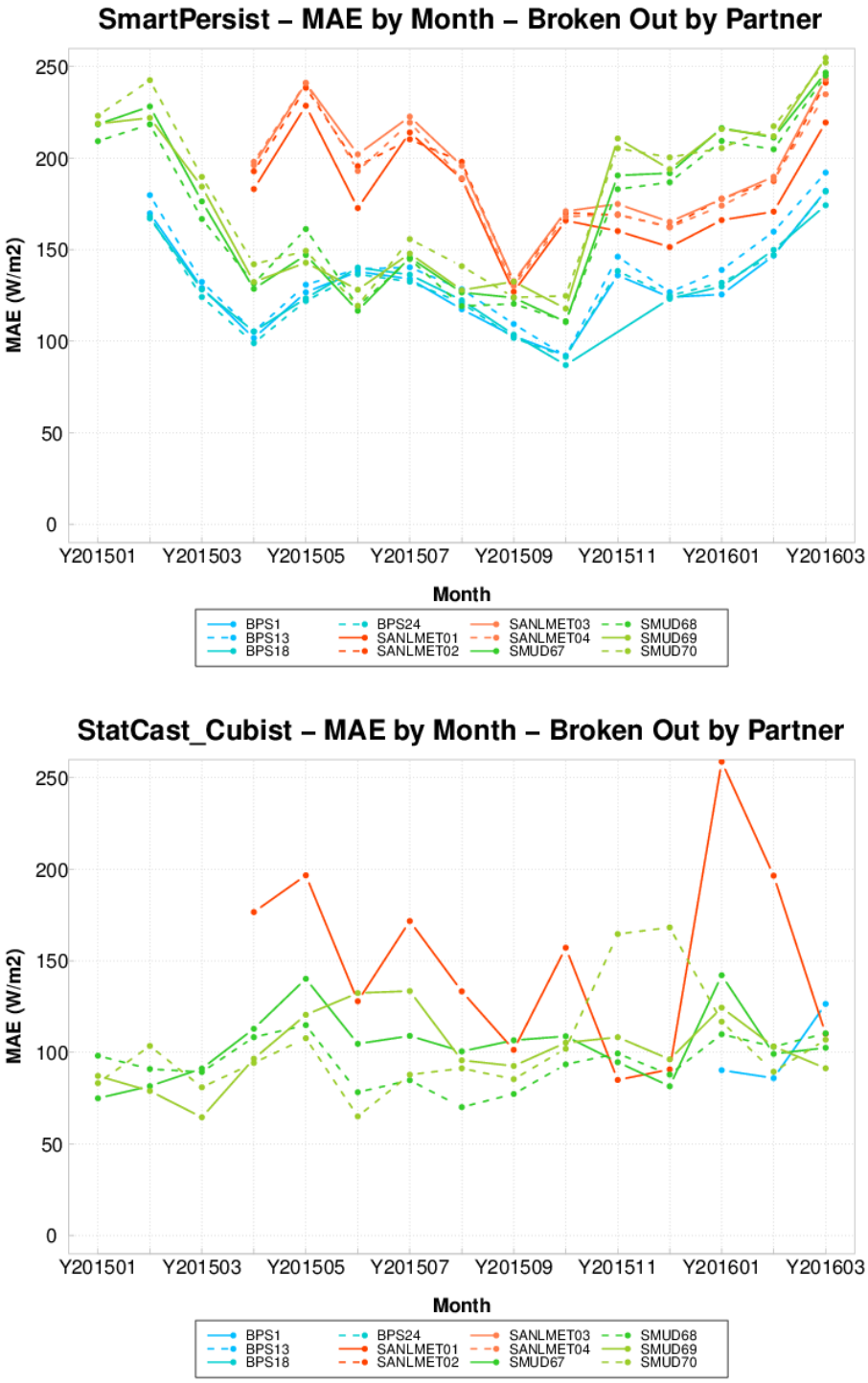


Figure 5-2. MAE in W m^{-2} of Smart Persistence (top) and StatCast-Cubist (bottom) stratified by location. Scores for BNL (blue), Xcel (orange) and SMUD (green) are included. Scores are aggregated over all sky conditions.

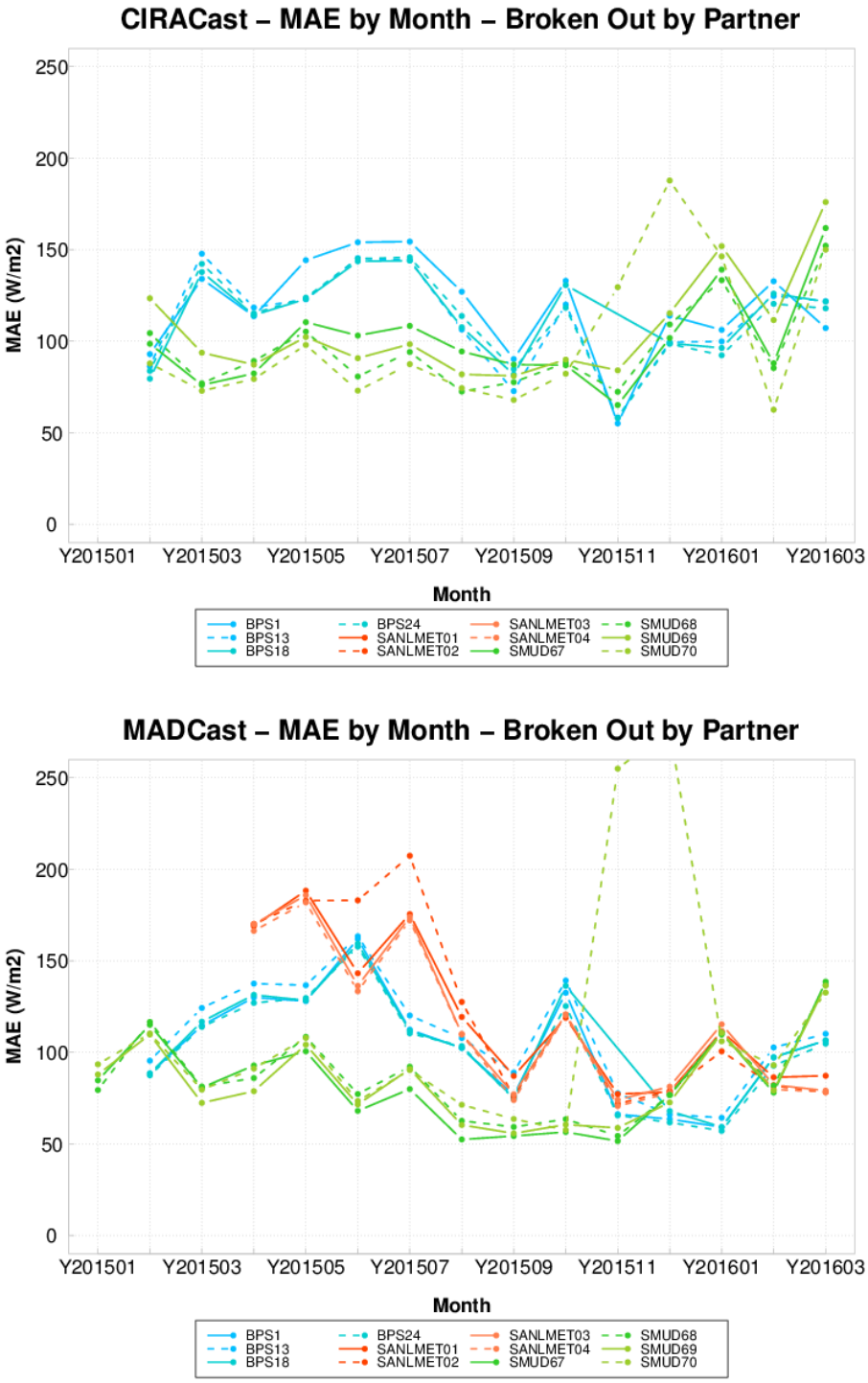


Figure 5-3. MAE in W m^{-2} of CiraCast (top) and MADCast (bottom) stratified by location. Scores for BNL (blue), Xcel (orange) and SMUD (green) are included. Scores are aggregated over all sky conditions.

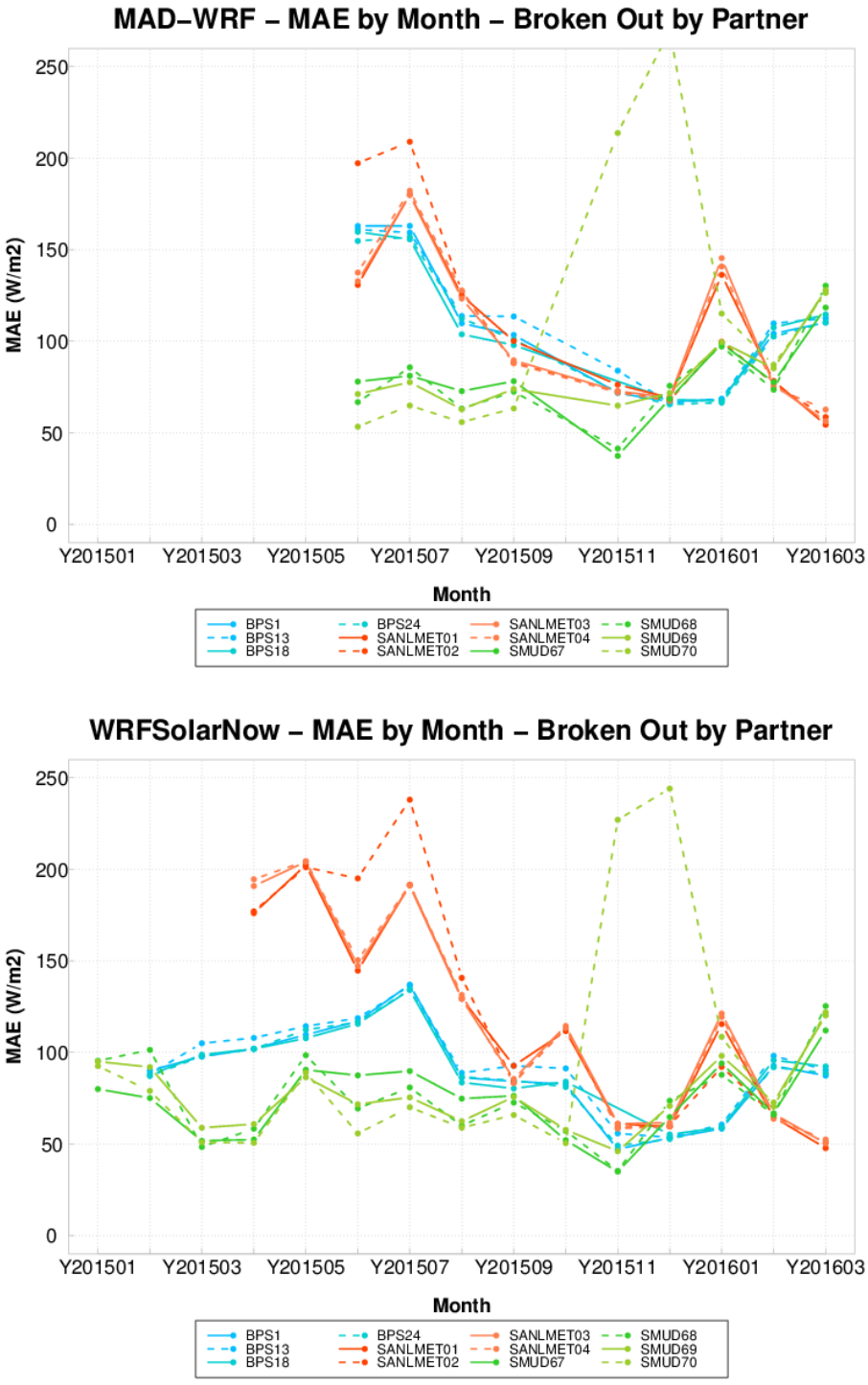


Figure 5-4. MAE in W m^{-2} of MAD-WRF (top), WRFSolarNow (bottom) stratified by location. Scores for BNL (blue), Xcel (orange) and SMUD (green) are included.

The resulting scores for the fully blended NowCast system, shown in Figure 5-5, follow a similar trend with higher scores at the Xcel location as it spun up. The improvement in performance at Xcel between April 2015 and March 2016 represents a 45-48% improvement within the time-frame of the project.

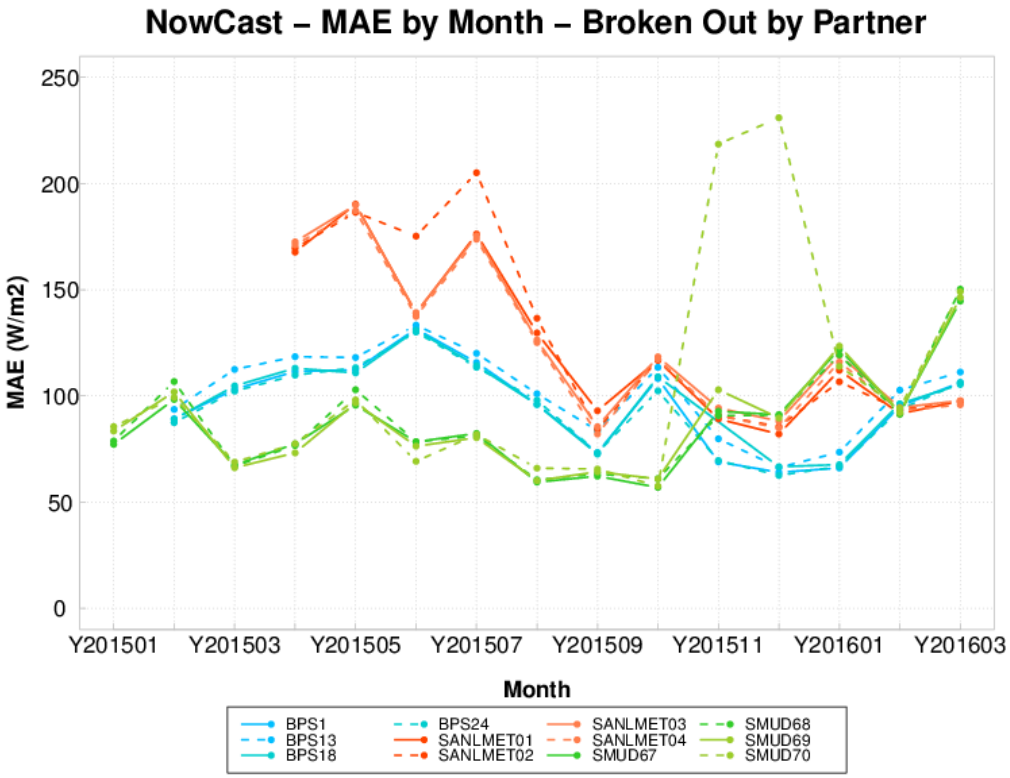


Figure 5-5. MAE in W m^{-2} of blended NowCast stratified by location. Scores for BNL (blue), Xcel (orange) and SMUD (green) are included.

Figure 5-6 provides a measure of each model's skill when these scores are accumulated over all geographic regions for the entire 15-month evaluation period. The scores were aggregated over all issue times. It also shows the skill for clear, partly cloudy and cloudy conditions. During clear conditions, only WRF-Solar-Now and NowCast outperform Smart Persistence to 45 minutes (0.75 hours). After this, all methods have lower MAE than Smart Persistence, with WRF-Solar-Now and CIRACast performing the best out through 2 hours and WRF-Solar-Now and MAD-WRF through 3-6-h lead times. For partly cloudy and cloudy conditions, the performance of the components is much more variable, with NowCast and MADCast providing the best forecasts during partly cloudy conditions and StatCast-Cubist and MAD-WRF giving the better forecasts during cloudy conditions. It is interesting to see that the performance of StatCast-Cubist is fairly poor for clear conditions and quite good for cloudy. It is possible the training dataset was biased with cloudy conditions therefore leading to high skill in cloudy conditions.

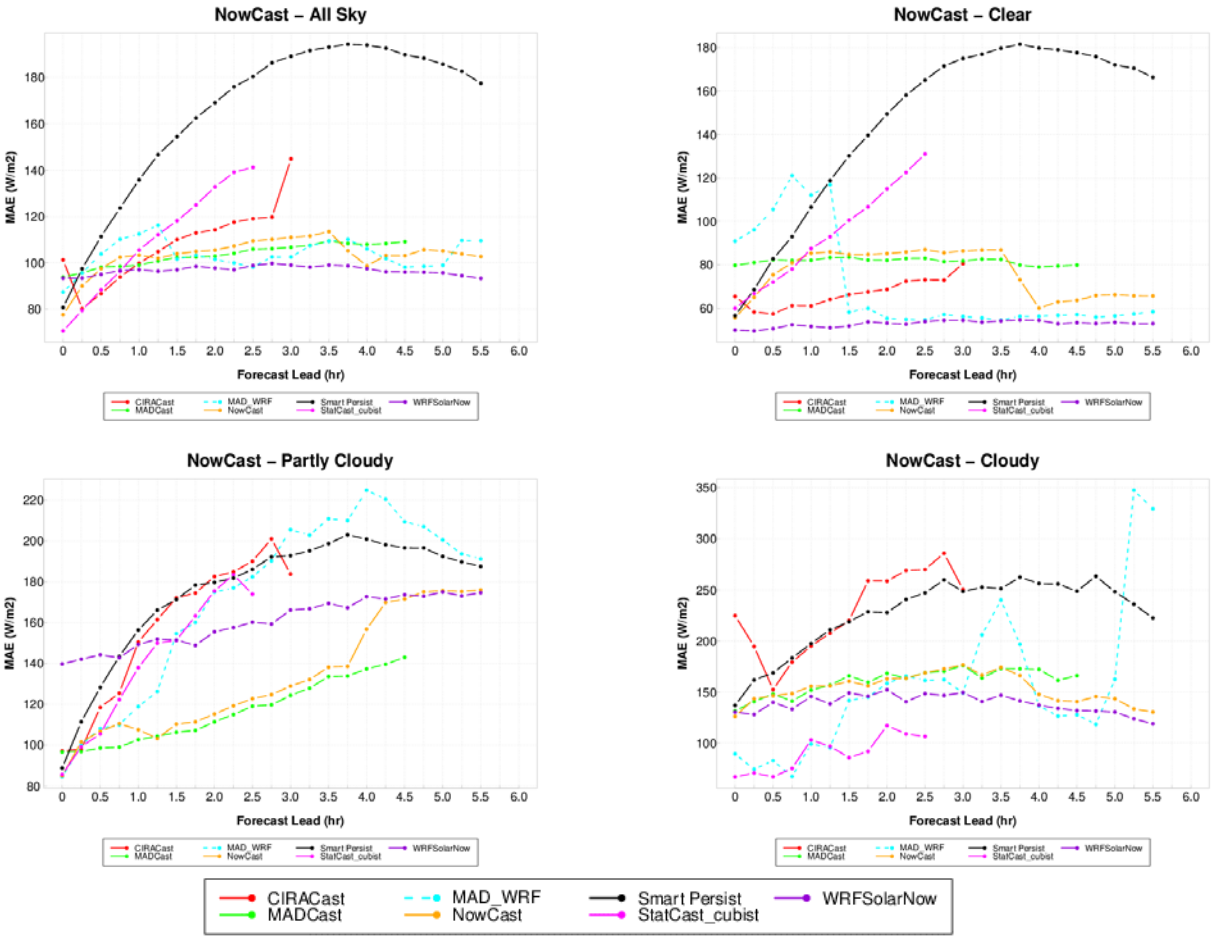


Figure 5-6. MAE in W m^{-2} for all NowCast components aggregated over all partners and all sky conditions (upper left), clear (upper right), partly cloudy (lower left) and cloudy (lower right) sky conditions.

While the components tend to perform better than Smart Persistence when initialized in the early morning, runs initialized in the afternoon struggle to beat Smart Persistence. Figure 5-7 provides plots of MAE for all NowCast components for forecasts initialized at 14, 16, 18 and 20 UTC. Most components outperform Smart Persistence (black line) at the 14 and 16 UTC issue times but struggle to beat Smart Persistence out to 3 hours for 18 UTC and to some extent 20 UTC issue times. Notably, WRF-Solar™ struggles at the beginning of the 18 UTC and 20 UTC forecast cycles, likely due to spin up issues for periods when clouds may be present. As expected, the observationally based methods such at StatCast_cubist, CIRACast and MADCast seem to perform the best during the first 3 hours with the NWP based methods, WRF-Solar-Now and MADCast outperforming Smart Persistence after 4 hours.

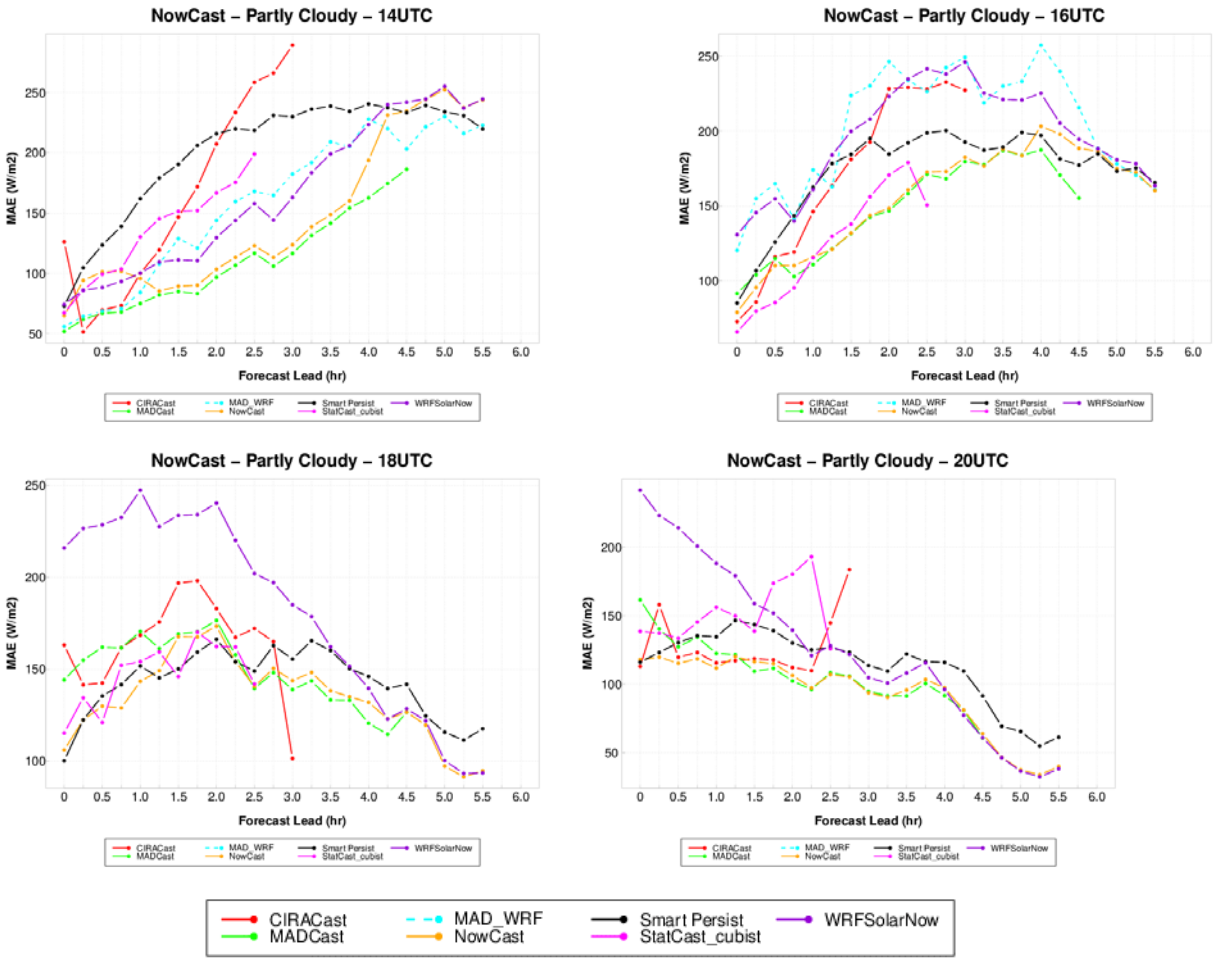


Figure 5-7. MAE in W m^{-2} for each NowCast components aggregated over all partners and partly cloudy conditions for 14 UTC (upper left), 16 UTC (upper right), 18 UTC (lower left) and 20 UTC (lower right) issue times.

To facilitate better weighting of the components, and to help determine which ones are imperative to the Sun4Cast® system, the performance of each component was ranked based on lead times and sky conditions. Table 5-5 and Table 5-6 provide these rankings. When all sky conditions are considered, WRF-Solar-Now is the top performer followed by MAD-WRF when available and MADCast (Table 5-5). However, for the hour ahead (HA: 0-1 hr) forecast, the best methods for weighting are StatCast-Cubist and CIRACast. In Table 5-6, the components are ranks based on performance in sky conditions in the 1-3 h time frame. For clear skies, WRF-Solar-Now and MAD-WRF are the top performers, whilch MADCast and StatCast_Cubist do better in partly cloudy conditions. In cloudy conditions, StatCast-Cubist and WRF-Solar-Now are the best components. If all conditions are rolled together, WRF-Solar-Now and MAD-WRF contribute the most to the skill.

Table 5-5. Rank of each component for all sky conditions stratified by lead-time.

	Rank for the All Sky Conditions			
Component	**0-1hr**	**1-3hr**	**3-6hr**	**All**
CiraCast	2	5	n/a	5
MADCast	5	3	4	4
MAD-WRF	6	2	3	2
Smart Persist	7	7	5	7
StatCast_Cubist	1	6	n/a	6
WRFSolarNow	3	1	1	1
NowCast	4	4	2	3

Table 5-6. Rank of each component for the 1-3hr forecasts stratified by sky condition.

	Rank for the 1-3hr Forecasts			
Component	**Clear**	**Partly Cloudy**	**Cloudy**	**All**
CiraCast	3	5	7	5
MADCast	5	1	5	3
MAD-WRF	2	7	3	2
Smart Persist	7	6	6	7
StatCast_Cubist	6	3	1	6
WRFSolarNow	1	4	2	1
NowCast	4	2	4	4

The GHI forecasts at SMUD were also analyzed for performance during variable conditions. Ramp events varying in magnitudes from 100 W m^{-2} to 350 W m^{-2} change in 30 minutes were identified within each NowCast component. Figure 5-8 shows frequency bias (number of forecasted events divided by the number of observed events) three of the components (CIRACast, MADCast and WRF-Solar-Now) and the NowCast blended forecast. Values of frequency bias greater than 1.0 indicate an over-forecast of events and less than 1.0 an under-forecast of events. If there is no box for a given magnitude, it indicates that there were no ramps of that magnitude identified in the component. The box plots are stratified by lead time to help identify how the components perform with increasing lead.

Generally, CIRACast also only predicts lower intensity ramp events but over-predicts the number. MADCast captures the lower intensity events early on (0-1 h) and at the end of the forecast cycle (3-6 h). It is also able to identify the larger ramps (>200 W m^{-2}) in the 1-3 hour forecast window but under-predicts the number of events. WRF-Solar-Now has similar ability to identify ramp events but over-predicts the number of events out to 3 hours and then gets the number of the lower intensity events just about right. The NowCast system seems to do the best at identifying both low- and high-intensity events after 1 hour even though there tends to be an under-prediction of the frequency.

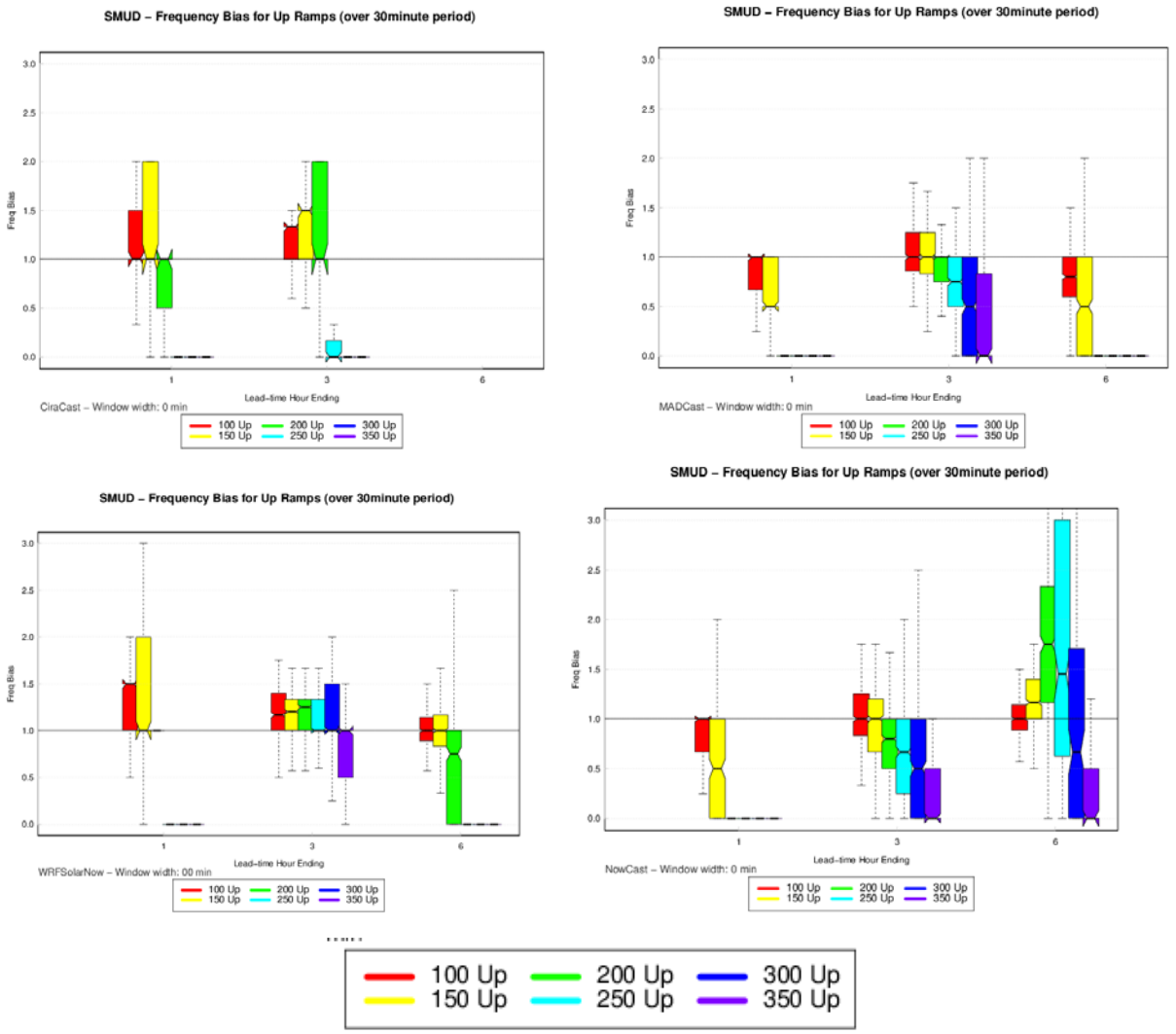


Figure 5-8. Performance of forecasts during ramp events. CIRACast (upper right), MADCast (middle left), WRF-Solar-Now (middle right) and NowCast blended forecast (lower left). Box plots are stratified by forecast lead time (0-1, 1-3, 3-6 hours) and intensity of the ramp (100 W m^{-2} – red to 350 W m^{-2} purple by 50 W m^{-2}).

Because the evaluation period was 15 months, there is a three-month period to assess performance changes between years. Table 5-7 summarizes the differences between clear, cloudy and all sky conditions for all components, including TSICast. Because each year is different, it is not surprising to see that improvements over Smart Persistence are not larger in 2016. StatCast-Cubist is the only method that was not enhanced or changed during the 15-month period. The decrease in scores from winter 2015 to winter 2016 support the theory that 2016 is a more challenging forecast year due to El Nino. Despite this, some components either maintained their median improvement (WRF-Solar-Now and MADCast) or increased it (TSICast).

Table 5-7. Median MAE improvement in % for NowCast components versus Smart Persistence for Clear, Cloudy (Partly Cloudy and Cloudy) and all sky conditions at the beginning and end of the evaluation period.

	Jan-Mar 2015			Jan-Mar 2016		
Component	**Clear**	**Cldy**	**All**	**Clear**	**Cldy**	**All**
TSICast*	-	-	29.0	-	-	34.0
StatCast_Cubist	46.2	61.9	55.1	38.5	37.4	37.1
CIRACast	62.1	16.5	45.9	66.0	12.8	36.3
MADCast	60.9	40.1	48.9	58.9	43.8	50.3
WRFSolarNow	73.7	40.7	55.9	71.4	42.4	55.6
NowCast	65.7	43.6	53.3	54.5	38.6	45.7

5.2.3 DICast® System and Components

The DICast® blending system serves to statistically blend NWP forecast for the Sun4Cast® solar forecasting system (see section 4.3). It provides the sole forecast beyond 6 h, although it also produces forecasts from time t=0 as also assessed in this section. As shown in Table 5-3, several publicly available operational forecast models were statistically blended using DICast®. Additionally, an instance of WRF-Solar™ was also integrated into the system. Similar to the evaluation for NowCast components, scores were computed for each DICast® component. Figure 5-9 displays correlations between the forecast and observed values aggregated by month. It provides a similar picture of the seasonal performance that that in Figure 5-1 in the NowCast section above. In this case, Figure 5-9 is aggregated over all sky conditions at SMUD locations for the 6-12-h forecasts. Correlation is lower during the winter months and higher between March and November with most of the components exhibiting similar scores. Correlations were much lower in January and February 2016 than in 2015. Once again, this may be due to the increased cloudiness, which is difficult to predict precisely, as a result of El Niño.

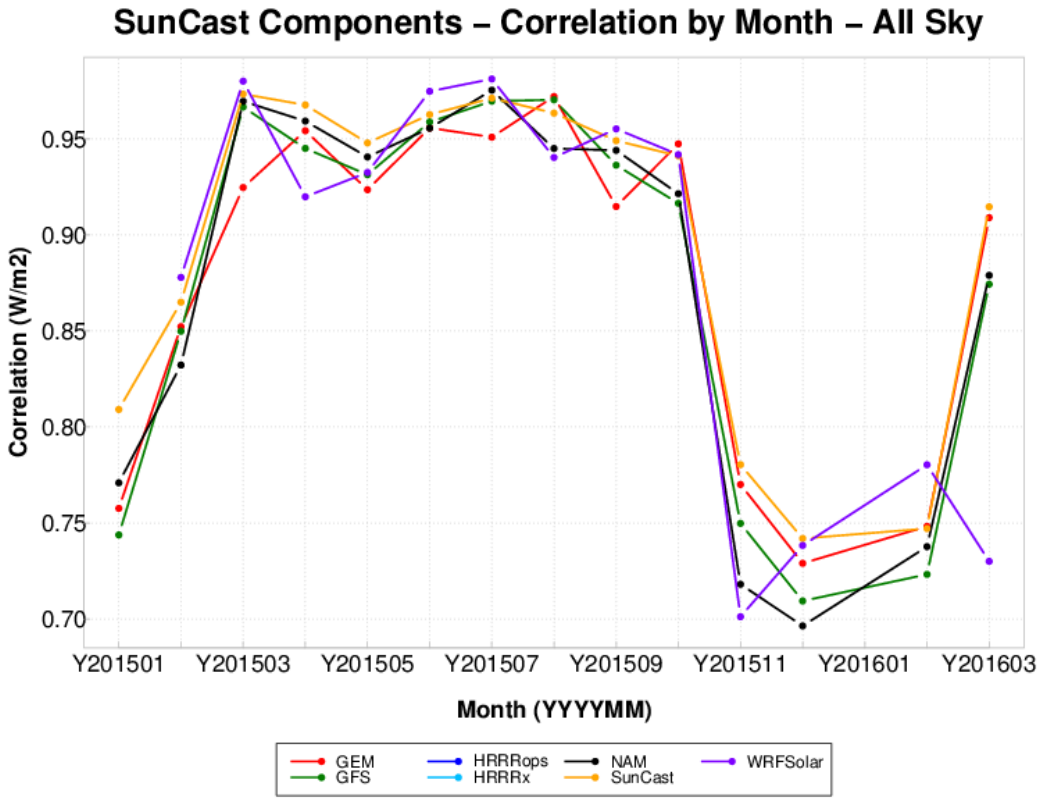


Figure 5-9. Pearson's correlation in W m^{-2} for DICast® components aggregated by month.

Individual component performance as a function of issue time and measured using MAE is provided in Figure 5-10. The forecasts being evaluated are Day-Ahead forecasts with lead times between 28-39 hours. Day-Ahead decisions are made between 9-10 UTC for BNL; 10-11 UTC for Xcel, and 11-12 UTC for SMUD and SCE locations. Generally, the Sun4Cast® system (orange) provides improved forecasting for all sky conditions and when specifically looking at partly cloudy and cloudy conditions. For clear conditions, WRF-Solar™, NAM and GFS outperform the blended Sun4Cast® system. The poorer performance of the GEM in clear sky conditions appears to be influencing the Sun4Cast® system. Considering that the real challenge in solar irradiance forecasting is capturing the irradiance variations due to cloudy and partly cloudy conditions, the marked improvement by the Sun4Cast® System under these conditions is encouraging. Figure 5-11 provides plots of RMSE and mean error (bias). Similar conclusions may be drawn if RMSE is used for evaluation, however the difference between Sun4Cast® scores and the operational component scores is larger. Interestingly, despite Sun4Cast's lower error scores, it tends to have a low bias by 5 to 10 W m^{-2} in forecasts issued after the sun comes up (shown in Figure 5-11). The Canadian GEM model also has a low bias, while NAM, GFS, and WRF-Solar™ all have high biases, on the order of 10-20 W m^{-2}.

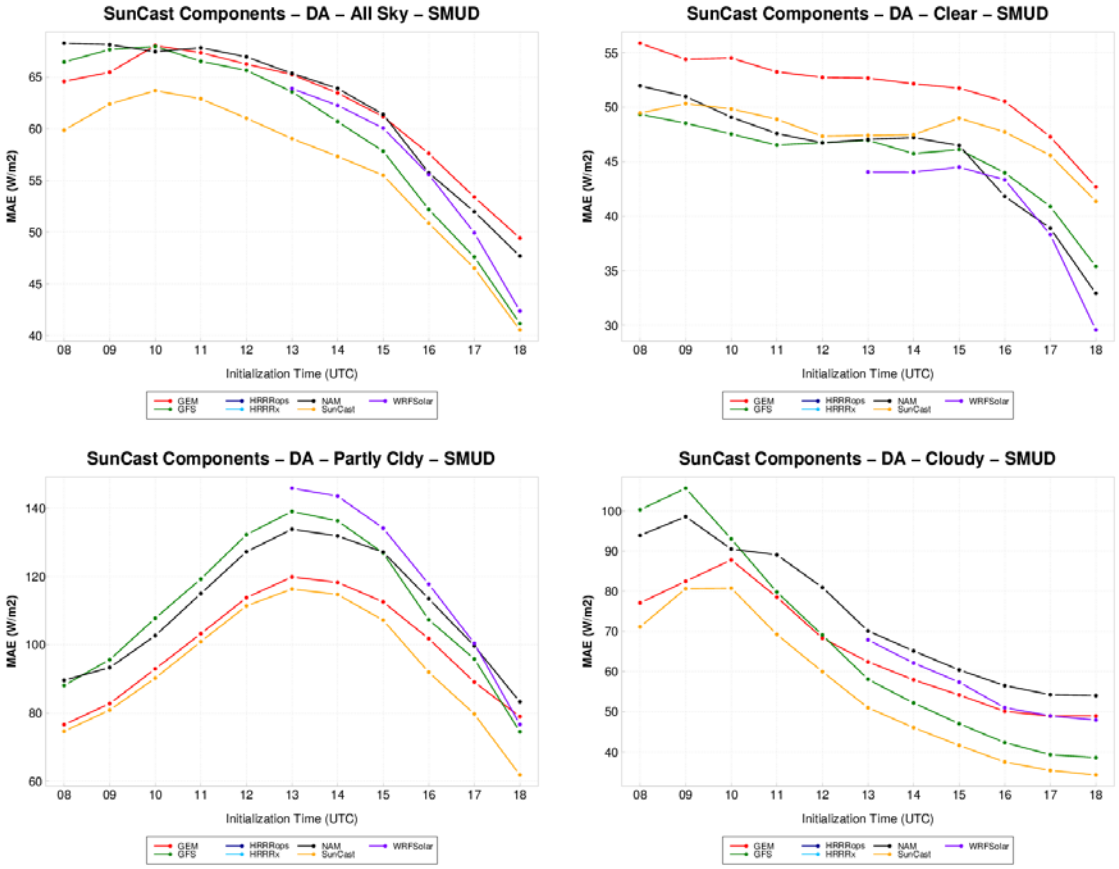


Figure 5-10. MAE in W m^{-2} for all Day-Ahead forecasts from DICast® components and Sun4Cast® system at SMUD locations and all sky conditions (upper left), clear (upper right), partly cloudy (lower left) and cloudy (lower right) sky conditions.

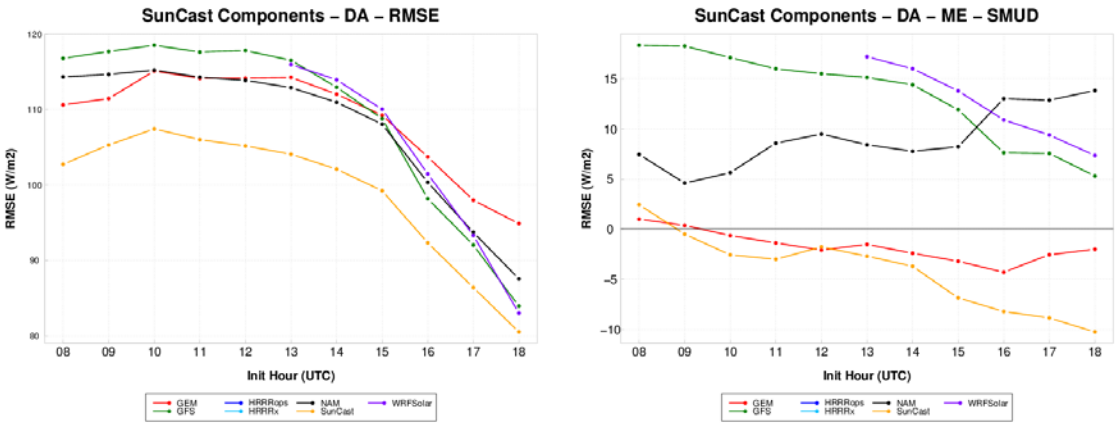


Figure 5-11. RMSE (top) and Bias (bottom) in W m^{-2} for each Day-Ahead forecast from DICast® components and Sun4Cast® system at SMUD locations.

Figure 5-12 indicates that when scores are aggregated over all partners, including BNL, Xcel, SMUD, and SCE, the Sun4Cast® and WRF-Solar™ systems perform better than the operational models for Day-Ahead forecasts. The notches about the median on box plots may be used to

determine if these results are statistically significant. If the notches do not overlap, the results are considered significant. Based on the box plots presented in Figure 5-13, these results are statistically significant at all lead times.

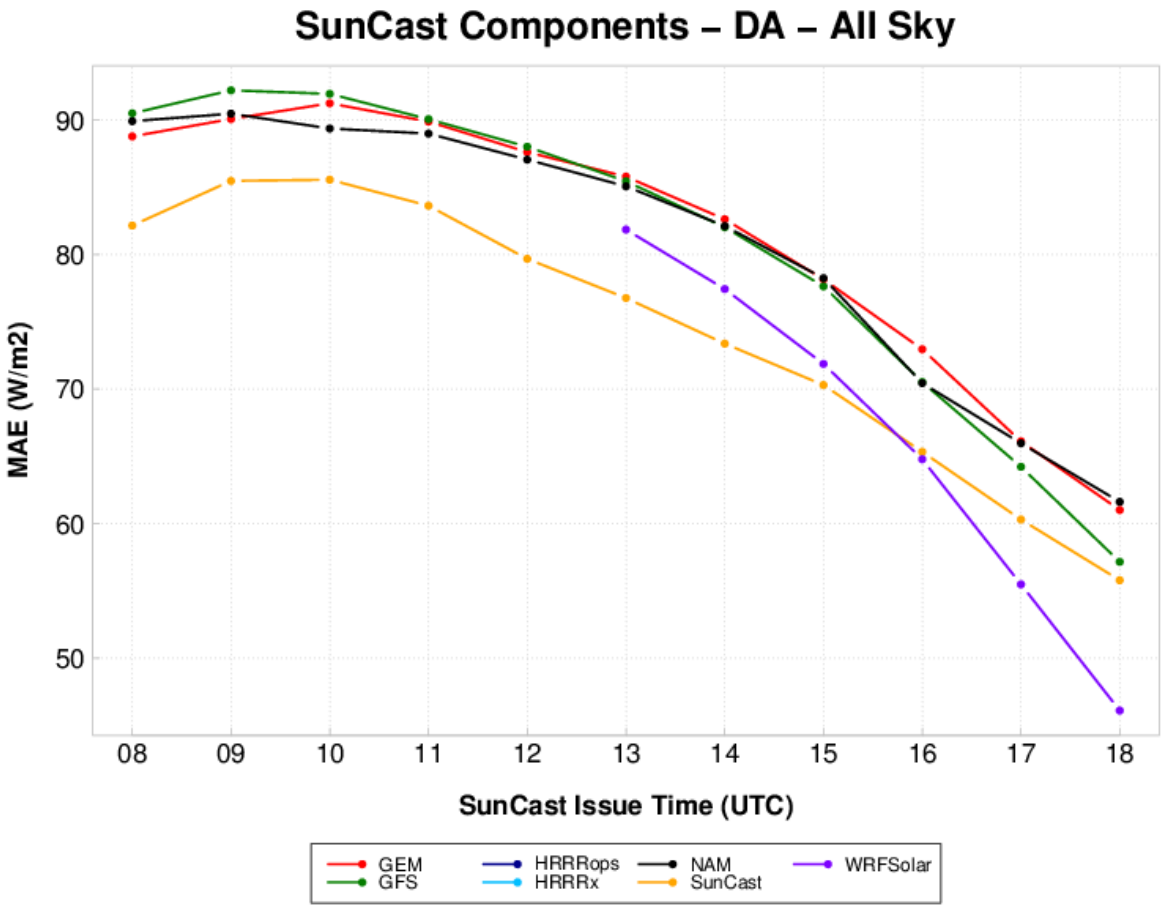


Figure 5-12. MAE in W m^{-2} for Day-Ahead forecasts from DICast® components and Sun4Cast® system at all partner locations and all sky conditions.

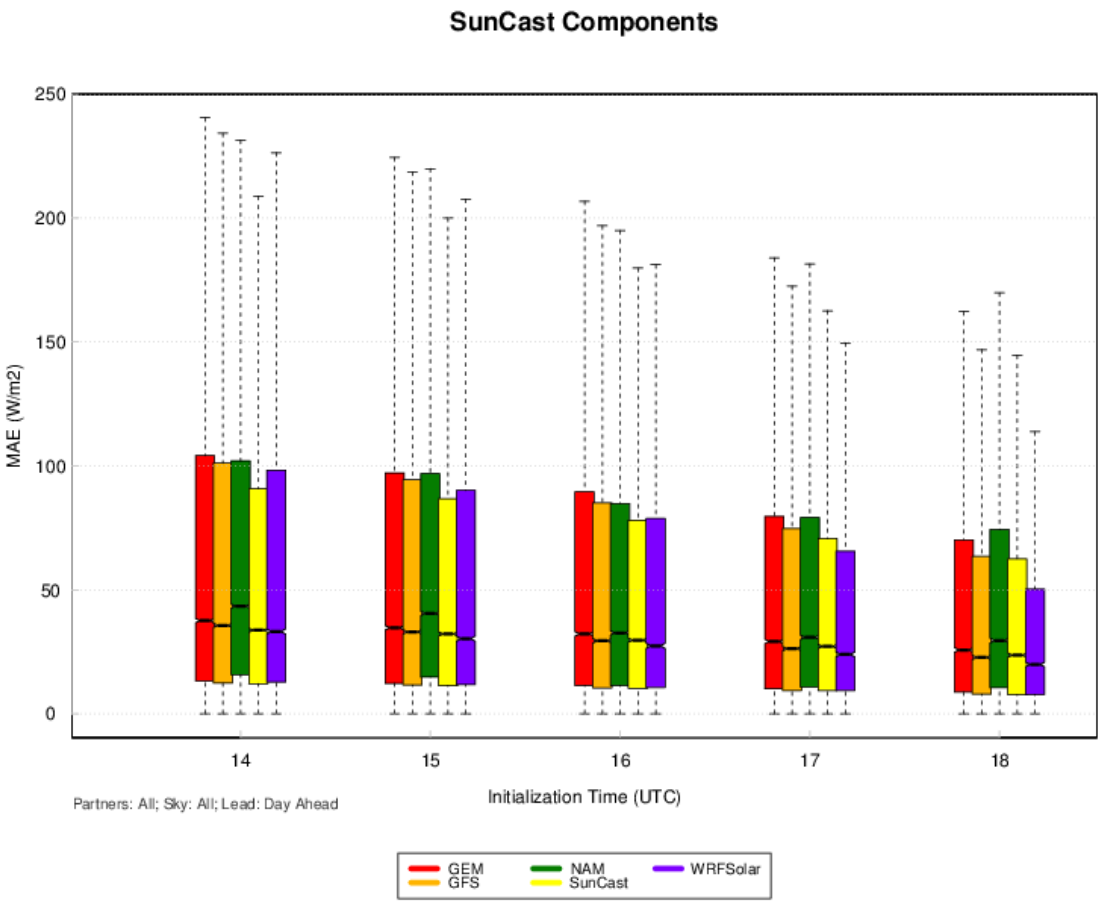


Figure 5-13. Box plots of MAE in W m^{-2} for Day-Ahead forecasts from DICast® components and Sun4Cast® system at all partner locations and all sky conditions.

Shorter lead times must be examined to compare the Sun4Cast® system with the higher resolution operational and research models (HRRRops and HRRRx). Figure 5-14 provides boxplots for the accumulation of scores over the 6-12 hour leads (top). Both HRRRops and HRRRx appear to have the highest MAE values. This is not completely surprising because they are both at 3-km resolution. Higher resolution models tend to be penalized by displacement errors more heavily than the smoother, coarser models that are blended into Sun4Cast® . For these initialization times, a 6-12-hforecast will be valid during the late afternoon or evening hours, on the tails of the daily irradiance profile. Inspection of the results for issue times between 0800 and 1300 UTC are very similar, hence the figure is not provided. For these lead times, it appears that the best performers are the NAM and WRF-Solar™, with similar median values and no statistically significant differences. To gain a sense of how these errors compare to the actual values, box plots of MAE normalized by actual values are shown in the bottom plot in Figure 5-14. The skill of HRRRx appears to be less than that of NAM, WRF-Solar™, and Sun4Cast® .

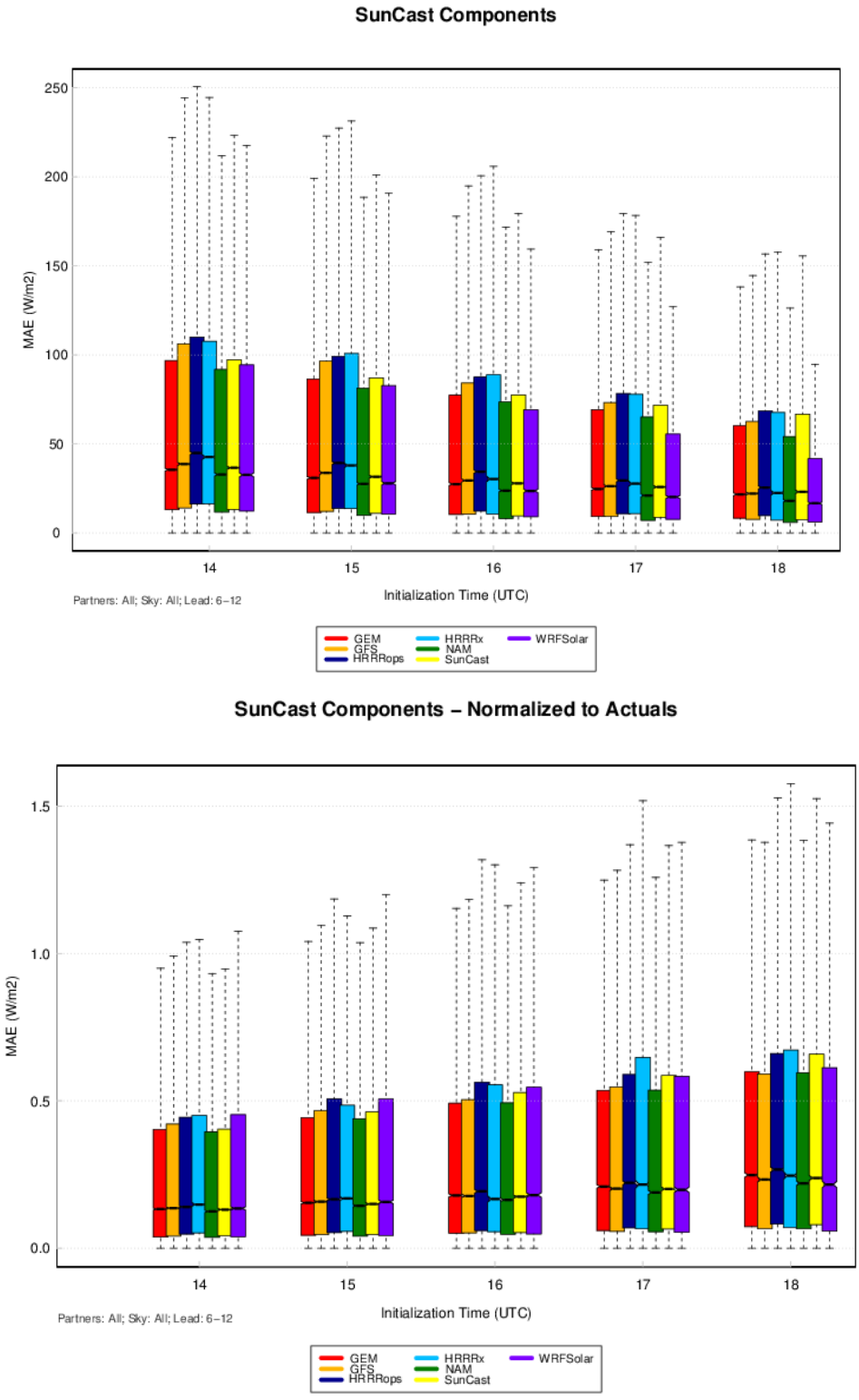


Figure 5-14. Box plots of MAE in W m^{-2} for 6-12-hour forecasts from DICast® components and Sun4Cast® system at all partner locations and all sky conditions.

5.2.4 Expert Sun4Cast® System

Much of section 5.2.3 includes a discussion of how the Sun4Cast® system performs with respect to its components. Of interest is also how Sun4Cast® performs as a blended system throughout a year of forecasting. Figure 5-15 provides a summary of the Sun4Cast® intraday and Day-Ahead forecast performance using MAE stratified into two-month aggregates. These plots reveal an interesting trend that is different from those displayed for only the SMUD location in Figure 5-1 and Figure 5-9. In general, MAE is 50% lower in winter months and 30% lower when Jan-Feb 2015 is compared with Jan-Feb 2016 for all sky conditions. Figure 5-16 shows the performance broken out by partner. The plots were generated by aggregating the Day-Ahead decision times (typically between 0900 and 1200 UTC) and stratified by sky condition. In Figure 5-16, the BNL and Xcel forecasts have lower MAE during fall and winter and higher in summer with performance for SMUD and SCE being the opposite. This is in part due to the cloud climatology at each location. At SMUD and SCE, clouds are more prevalent during the winter months while in New York and Colorado, they are more prevalent in summer. The highest errors tend to be for partly cloudy and cloudy conditions during the summer (May-Jun and Jul-Aug). Considering the convective, hence patchy, nature of clouds during these months, a marked increase in error is expected.

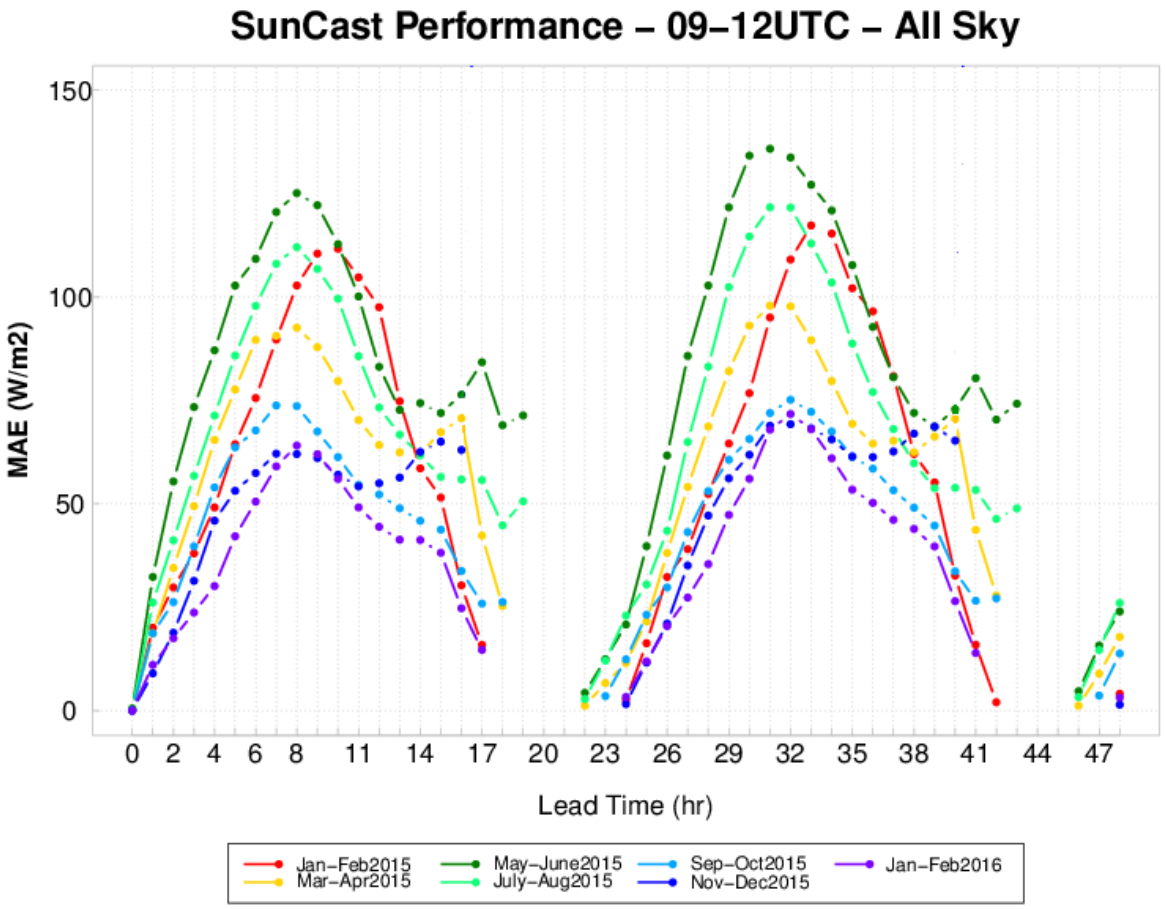


Figure 5-15 Sun4Cast® performance stratified by 2month periods for all partners during all sky conditions.

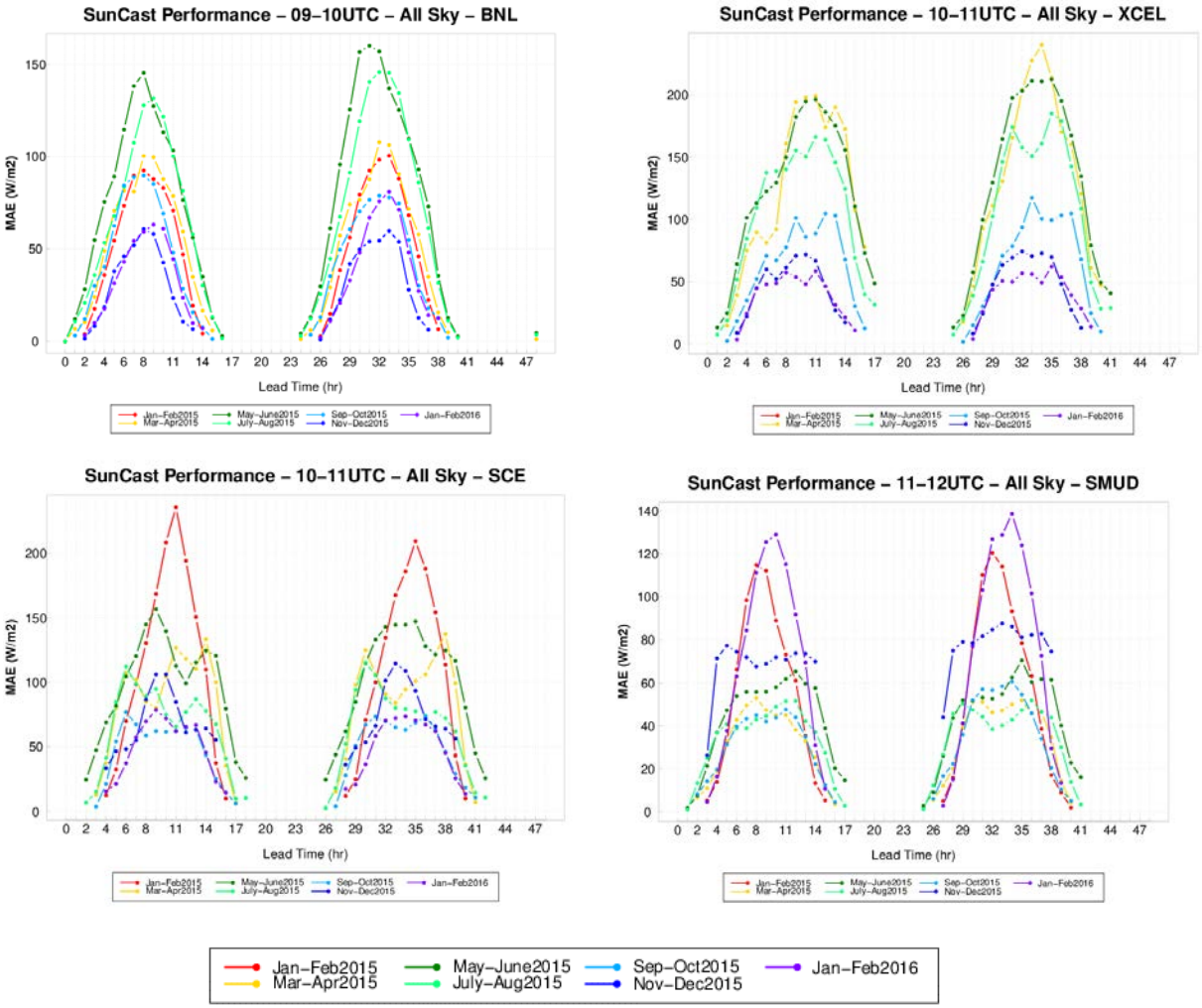


Figure 5-16. MAE in W m^{-2} for Sun4Cast® system at SMUD locations for BNL (upper left), XCEL (upper right), SCE (lower left) and SMUD (lower right) sky conditions.

5.2.5 Improvements Over Baselines

An in-depth analysis was performed for each component of the Nowcast system, including differing issue times, lead times, and sky conditions. Table 5-8 to Table 5-11 provide detailed analysis of the percentage improvements of the NowCasting components over Smart Persistence. Table 5-12 to Table 5-14 show improvements over NWP. This analysis was performed for the period of January – May 2015 but the results are generally applicable throughout the year. Positive values above 5% improvement but below a predetermined success value are shaded in light blue. Values between -5% improvement and 5% improvement are considered neutral and have no shading. Values that fall below -5% are considered degraded skill and are shaded in red. The median across all available forecast cycles, represented by Issue Time is listed directly below the table. The percentage of values above the success value is listed in the second row.

The analysis displayed in Table 5-8 to Table 5-10 suggests that the NowCast components included in this evaluation do, in general, improve upon forecasting with Smart Persistence (SmartP). The

stratification by sky type suggests that the Sun4Cast NowCast system outperforms SmartP much of the time. In the morning hours, this could be because the SmartP system relies on previous observations. If no observations are available, it assumes a clearness index of 0.6, which is the lower bound of the clear definition. However, looking at system performance in the later morning indicates an improvement in clear sky forecasts overall.

Table 5-8. Summary of NowCast component performance versus Smart Persistence (SmartP) baseline for 0-1-h forecast leads.

		0-1 HR Lead				% Improvement over Smart Persistance							
0-1 HR Lead		Clear				Partly Cloudy				Cloudy			
Issue Time	Valid Time	CiraCast	MADCast	WRFSolarNow	NowCast	CiraCast	MADCast	WRFSolarNow	NowCast	CiraCast	MADCast	WRFSolarNow	NowCast
UTC	UTC	%	%	%	%	%	%	%	%	%	%	%	%
09	09-10	-	51	73	36	-	74	77	75	-	62	72	51
10	10-11	-	55	73	61	-	77	58	64	-	62	74	53
11	11-12	-	34	75	46	-	44	11	36	-	44	50	35
12	12-13	-	46	59	24	-	72	61	65	-	10	48	8
13	13-14	-	52	71	36	-	68	64	57	-	64	69	63
14	14-15	76	51	76	49	50	67	58	61	-	77	75	72
15	15-16	70	44	72	44	-5	70	52	65	-40	68	71	56
16	16-17	62	41	66	39	14	53	29	47	-20	26	41	28
17	17-18	52	23	63	40	-104	15	-14	21	-248	-107	-41	-63
18	18-19	51	17	57	28	-177	-41	-73	-16	-365	-241	-142	-154
19	19-20	41	3	57	22	-241	-33	-90	-7	-489	-288	-188	-185
20	20-21	31	6	54	15	-194	-12	-51	7	-294	-66	-10	-37
21	21-22	36	25	56	24	-146	0	-21	4	-338	-20	17	-9
22	22-23	42	6	59	38	-47	23	-13	20	-201	59	68	64
23	23-00	45	-77	47	30	-4	48	1	41	11	72	72	75
	median	48	24	59	33	-75	36	-6	31	-248	18	44	18
	%above	100	73	100	100	20	73	47	73	0	60	73	60
	%positive	100	87	100	100	20	73	53	87	10	67	73	67

Table 5-9. Summary of NowCast component performance versus Smart Persistence (SmartP) baseline for 1-3-h forecast leads.

		1-3 HR Lead				% Improvement over Smart Persistance							
1-3 HR Lead		Clear				Partly Cloudy				Cloudy			
Issue Time	Valid Time	CiraCast	MADCast	WRFSolarNow	NowCast	CiraCast	MADCast	WRFSolarNow	NowCast	CiraCast	MADCast	WRFSolarNow	NowCast
UTC	UTC	%	%	%	%	%	%	%	%	%	%	%	%
09	09-10	-	0	58	7	-	69	74	68	-	44	62	46
10	10-11	-	-6	58	-4	-	73	103	72	-	60	70	60
11	11-12	-	48	79	49	-	65	84	65	-	43	52	43
12	12-13	-	27	70	31	-	63	96	62	-	62	64	60
13	13-14	-	51	73	53	-	67	134	67	-	63	63	61
14	14-15	73	61	78	64	7	61	113	60	-	48	59	47
15	15-16	78	65	79	66	-6	42	57	40	-	34	57	33
16	16-17	79	59	77	60	-10	26	15	24	-	-2	31	1
17	17-18	75	56	75	57	-18	-7	-54	-10	-	-60	0	-53
18	18-19	67	47	71	50	-21	6	-33	3	-	-122	-22	-107
19	19-20	55	36	67	38	-50	10	-19	9	-	-93	-7	-88
20	20-21	58	33	65	36	-145	5	-14	4	-	45	63	46
21	21-22	70	34	67	36	-	24	19	23	-	69	75	70
22	22-23	77	19	43	20	-	58	48	55	-	78	77	77
23	23-00	55	-57	-73	-56	-	65	36	64	-	61	48	59
	median	71	42	70	44	-18	50	42	48	-	47	58	47
	%above	100	80	93	80	0	73	73	73	-	73	80	73
	%positive	100	80	13	87	14	93	73	80	-	73	80	73

Table 5-10. Summary of NowCast component performance versus Smart Persistance (SmartP) baseline for 3-6-h forecast leads.

		3-6 HR Lead	% Improvement over Smart Persistance										
3-6 HR Lead		Clear				Partly Cloudy				Cloudy			
Issue Time	Valid Time	CiraCast	MADCast	WRFSolarNow	NowCast	CiraCast	MADCast	WRFSolarNow	NowCast	CiraCast	MADCast	WRFSolarNow	NowCast
UTC	UTC	%	%	%	%	%	%	%	%	%	%	%	%
09	12-15	-	-67	45	43	-	78	70	66	-	61	78	80
10	13-16	-	28	77	76	-	78	68	67	-	76	82	82
11	14-17	-	49	77	76	-	78	69	70	-	83	81	80
12	15-18	-	36	62	61	-	80	71	71	-	85	84	84
13	16-19	-	61	76	76	-	65	45	47	-	-6	22	26
14	17-20	-	78	84	84	-	41	18	20	-	-74	-19	-5
15	18-21	-	71	84	84	-	-12	-22	-21	-	-41	16	25
16	19-22	-	63	81	81	-	35	29	31	-	-80	-6	1
17	20-23	-	65	83	82	-	26	30	29	-	-42	22	41
18	21-00	-	19	59	58	-	9	5	2	-	-125	-10	21
19	22-01	-	-31	33	31	-	-37	-25	-20	-	52	65	65
20	23-02	-	-41	25	24	-	30	46	50	-	76	79	81
	median	-	42	76	76	-	38	37	39	-	23	43	53
	%above	-	75	100	100	-	75	75	75	-	50	75	83
	%positive	-	75	25	58	-	75	75	75	-	50	75	83

One encouraging outcome of the NowCast component evaluation is the general trend toward improvement in partly cloudy and cloudy conditions. While there are still errors during mid-day in many of the components, the NowCast system generally outperforms Smart Persistence at longer lead times (see Table 5-9 and Table 5-10 for 1-3-h and 3-6-h results.).

Table 5-11. Summary of WRF Solar and DICast® component performance versus NAM baseline for 3-6-h forecast leads.

% Improvement of WRF Solar vs. NWP (NAM)

	Clear					Partly Cloudy					Cloudy				
Issue Time	6-9 HR	10-12 HR	13-15 HR	16-18 HR	DA	6-9 HR	10-12 HR	13-15 HR	16-18 HR	DA	6-9 HR	10-12 HR	13-15 HR	16-18 HR	DA
UTC	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%
06	-4	12	6	-	-29	8	-11	-22	40	-6	-9	-5	6	7	0
07	3	-1	7	81	-17	2	-23	-16	48	-8	-31	3	7	5	12
08	10	-9	7	45	-1	-9	-21	-14	26	-4	-64	6	7	3	25
09	12	-23	-39	11	-	-11	-22	-16	-88	-	-66	-31	-49	-11	-
10	-7	-39	-16	18	-	-19	-21	-28	-111	-	-26	-34	22	-17	-
11	-19	-42	4	-1	-	-20	-20	-45	-	-	-22	-41	60	-17	-
12	-35	-32	10	-	-	-23	-20	-75	-	-	-29	-48	63	-6	-
median	-4.4	-23.4	6.5	18.4	-16.7	-11.4	-20.5	-22.2	26.0	-6.4	-29.0	-31.0	6.6	-6.0	11.7
% above	42.9	28.6	71.4	71.4	14.3	0.0	0.0	0.0	40.0	0.0	0.0	0.0	28.6	0.0	0.0
% positive	42.9	28.6	71.4	71.4	14.3	14.3	0.0	0.0	60.0	0.0	0.0	14.3	85.7	28.6	66.7

% Improvement of DiCast vs. NWP (NAM)

	Clear					Partly Cloudy					Cloudy				
Issue Time	6-9 HR	10-12 HR	13-15 HR	16-18 HR	DA	6-9 HR	10-12 HR	13-15 HR	16-18 HR	DA	6-9 HR	10-12 HR	13-15 HR	16-18 HR	DA
UTC	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%
06	-12	3	1	-	-18	22	21	23	88	15	10	-14	1	1	29
07	-11	-3	1	75	-15	26	25	18	69	12	8	8	1	1	29
08	-8	-2	1	43	-22	21	26	14	52	11	14	7	1	2	29
09	1	3	-12	-28	-7	20	28	6	-70	16	16	25	-23	-24	18
10	-11	4	-14	-19	-9	24	18	-8	3	14	29	15	25	-10	15
11	-10	-6	-15	10	-5	25	12	-33	-	14	30	5	49	9	25
12	-7	-1	-22	-	-3	23	3	-68	-	14	20	-16	60	8	29
median	-9.6	-1.0	-11.5	9.6	-8.9	23.1	20.6	5.9	52.3	14.4	15.6	7.1	1.1	1.4	28.6
% above	14.3	85.7	42.9	42.9	14.3	0.0	0.0	0.0	60.0	0.0	28.6	0.0	28.6	0.0	28.6
% positive	14.3	85.7	42.9	42.9	14.3	100.0	85.7	57.1	42.9	100.0	100.0	71.4	42.9	28.6	100.0

WRF-Solar-Now and the NowCast forecasts are evaluated against the NAM as the baseline in Table 5-12 to Table 5-14. WRF-Solar-Now, a lower resolution model more comparable to the NAM shows a significant number of large improvement values during clear sky conditions. Finally, the performance of the NowCast consensus forecast versus the NAM indicates a need for some additional tuning for clear sky conditions but show remarkable improvements for the toughest forecast conditions, the partly cloudy conditions.

Table 5-15 and Table 5-16 present detailed analysis of WRF-Solar™ and the Sun4Cast® blended forecast with operational NWP. While there are some time periods when WRF-Solar™ improvements exceed 30% improvement during cloudy conditions (the 16-18-hour forecast), there are also times during other parts of the day when 5-29% positive improvements are found. For WRF-Solar™, the heavy weighting of negative improvement between 10-12 UTC issue times suggests that there may be a need to look more closely at the forecasts on a location by location basis. The Sun4Cast® system outperforms the NAM, many times with improvements above 20-30% for partly cloudy and cloudy conditions. The uncertainty of cloudy conditions has been identified by stakeholders as the costliest to their operations, so an overall improvement in forecasting those conditions seems to outweigh the reduced skill for clear sky cases at this time.

Table 5-12. Summary of WRF Solar Now and NowCast versus NAM baseline for 0-1-h forecast leads.

		0-1 HR Lead	% Improvement over NWP (NAM)				
0-1 HR Lead		Clear		Partly Cloudy		Cloudy	
Issue Time	Valid Time	WRFSolarNow	NowCast	WRFSolarNow	NowCast	WRFSolarNow	NowCast
UTC	UTC	%	%	%	%	%	%
09	09-10	-	-	34	-16	-31	-185
10	10-11	23	1	-83	-172	-171	-524
11	11-12	-7	-185	-237	-70	-87	-211
12	12-13	-66	-249	-139	-106	22	-62
13	13-14	-47	-188	17	-21	-68	-76
14	14-15	30	-88	-28	-52	21	10
15	15-16	11	-105	-24	5	-50	-155
16	16-17	-1	-97	-13	12	-82	-109
17	17-18	3	-68	-25	37	5	14
18	18-19	-17	-86	1	41	-14	2
19	19-20	15	-29	7	52	-6	28
20	20-21	29	-21	-20	45	49	33
21	21-22	27	-25	-2	33	-11	-29
22	22-23	23	0	-2	33	74	48
23	23-00	39	24	-7	29	58	62
	median	13	-77	-17	21	-8	-13
	%above	71	21	7	47	20	20
	%positive	71	21	13	60	40	40

Table 5-13. Summary of WRF Solar Now and NowCast performance versus NAM baseline for 1-3-h forecast leads.

		1-3 HR Lead	% Improvement over NWP (NAM)				
1-3 HR Lead		Clear		Partly Cloudy		Cloudy	
Issue Time	Valid Time	WRFSolarNow	NowCast	WRFSolarNow	NowCast	WRFSolarNow	NowCast
UTC	UTC	%	%	%	%	%	%
09	10-12	37	-51	-81	-7	-60	-147
10	11-13	-22	-196	-195	-77	-34	-118
11	12-14	-38	-183	-15	-4	-49	-93
12	13-15	12	-111	-15	-13	2	-24
13	14-16	6	-81	-9	-7	-25	-30
14	15-17	8	-78	-21	0	-28	-55
15	16-18	4	-82	-22	8	-21	-92
16	17-19	-8	-96	-9	15	-20	-56
17	18-20	1	-71	-1	28	7	-14
18	19-21	9	-47	2	31	17	-7
19	20-22	21	-41	-2	36	5	-8
20	21-23	27	-33	-16	26	27	-4
21	22-00	30	-27	-3	21	54	36
22	23-01	27	-3	10	33	63	58
23	00-02	30	16	17	44	41	50
	median	9	-74	-9	18	3	-19
	%above	80	13	0	27	20	20
	%positive	80	13	13	60	47	20

Table 5-14. Summary of WRF Solar Now and NowCast performance versus NAM baseline for 3-6-h forecast leads.

3-6 HR lead		3-6 HR lead % Improvement over NWP (NAM)					
3-6 HR lead Issue Time	Valid Time	Clear WRFSolarNow	Clear NowCast	Partly Cloudy WRFSolarNow	Partly Cloudy NowCast	Cloudy WRFSolarNow	Cloudy NowCast
UTC	UTC	%	%	%	%	%	%
09	12-15	29	7	-3	0	26	2
10	13-16	32	-6	-4	5	3	-11
11	14-17	13	-13	8	8	-17	-11
12	15-18	15	-14	0	5	3	2
13	16-19	-1	-13	11	23	0	-9
14	17-20	-8	-17	0	13	5	-16
15	18-21	3	-10	-10	1	23	-5
16	19-22	6	-11	-6	5	-14	-46
17	20-23	20	1	-12	-3	-14	-59
18	21-00	17	-3	-10	-1	8	-56
19	22-01	21	-11	-16	2	44	15
20	23-02	24	-13	8	12	44	37
	median	16	-11	-3	5	4	-10
	%above	92	25	0	0	17	8
	%positive	92	25	25	58	50	17

Table 5-15. Summary of WRF Solar (top) and Sun4Cast® (bottom) performance versus NAM baseline for clear conditions.

	Clear				
Issue Time	6-9 HR	10-12 HR	13-15 HR	16-18 HR	DA
UTC	%	%	%	%	%
06	-4	12	6	-	-29
07	3	-1	7	81	-17
08	10	-9	7	45	-1
09	12	-23	-39	11	-
10	-7	-39	-16	18	-
11	-19	-42	4	-1	-
12	-35	-32	10	-	-
median	-4.4	-23.4	6.5	18.4	-16.7
% above	42.9	28.6	71.4	71.4	14.3
% positive	42.9	28.6	71.4	71.4	14.3

	Clear				
Issue Time	6-9 HR	10-12 HR	13-15 HR	16-18 HR	DA
UTC	%	%	%	%	%
06	-12	3	1	-	-18
07	-11	-3	1	75	-15
08	-8	-2	1	43	-22
09	1	3	-12	-28	-7
10	-11	4	-14	-19	-9
11	-10	-6	-15	10	-5
12	-7	-1	-22	-	-3
median	-9.6	-1.0	-11.5	9.6	-8.9
% above	14.3	85.7	42.9	42.9	14.3
% positive	14.3	85.7	42.9	42.9	14.3

Table 5-16. Summary of WRF Solar (top) and Sun4Cast® (bottom) performance versus NAM baseline for partly cloudy (left) and cloudy (right) conditions.

% Improvement of WRF Solar vs. NWP (NAM)

	Partly Cloudy					Cloudy				
Issue Time	6-9 HR	10-12 HR	13-15 HR	16-18 HR	DA	6-9 HR	10-12 HR	13-15 HR	16-18 HR	DA
UTC	%	%	%	%	%	%	%	%	%	%
06	8	-11	-22	40	-6	-9	-5	6	7	0
07	2	-23	-16	48	-8	-31	3	7	5	12
08	-9	-21	-14	26	-4	-64	6	7	3	25
09	-11	-22	-16	-88	-	-66	-31	-49	-11	-
10	-19	-21	-28	-111	-	-26	-34	22	-17	-
11	-20	-20	-45	-	-	-22	-41	60	-17	-
12	-23	-20	-75	-	-	-29	-48	63	-6	-
median	-11.4	-20.5	-22.2	26.0	-6.4	-29.0	-31.0	6.6	-6.0	11.7
% above	0.0	0.0	0.0	40.0	0.0	0.0	0.0	28.6	0.0	0.0
% positive	14.3	0.0	0.0	60.0	0.0	0.0	14.3	85.7	28.6	66.7

% Improvement of DICast vs. NWP (NAM)

	Partly Cloudy					Cloudy				
Issue Time	6-9 HR	10-12 HR	13-15 HR	16-18 HR	DA	6-9 HR	10-12 HR	13-15 HR	16-18 HR	DA
UTC	%	%	%	%	%	%	%	%	%	%
06	22	21	23	88	15	10	-14	1	1	29
07	26	25	18	69	12	8	8	1	1	29
08	21	26	14	52	11	14	7	1	2	29
09	20	28	6	-70	16	16	25	-23	-24	18
10	24	18	-8	3	14	29	15	25	-10	15
11	25	12	-33	-	14	30	5	49	9	25
12	23	3	-68	-	14	20	-16	60	8	29
median	23.1	20.6	5.9	52.3	14.4	15.6	7.1	1.1	1.4	28.6
% above	0.0	0.0	0.0	60.0	0.0	28.6	0.0	28.6	0.0	28.6
% positive	100.0	85.7	57.1	42.9	100.0	100.0	71.4	42.9	28.6	100.0

When comparing models, the best way to determine if differences are statistically significant is to use pairwise differencing. This removes the impact of the forecast challenge and allows for a fair comparison between systems. Figure 5-17 through Figure 5-19 provide the MAE error curves and pairwise difference between the Sun4Cast® system and the operational benchmark, NAM for three partner locations (BNL, Xcel and SMUD). In these figures, positive differences indicate improved skill by the Sun4Cast® system and negative differences indicate improved skill by the NAM. The difference lines have boot-strapped confidence intervals applied. They are computed at the 95% confidence level and bolded if the intervals do not intersect zero. The bolded confidence intervals indicate when differences are statistically significant. Finally, issue times vary by partner, depending on the appropriate Day-Ahead decision time.

In Figure 5-17, there are statistically significant improvements in the Sun4Cast® irradiance forecasts for both the intra-day forecast period (first day) and DA periods. Note the 5-18 W m^{-2} improvement during both periods. Improvements are smaller in the morning and evening shoulders of the irradiance profiles. At the Xcel Energy sites, Sun4Cast® appears to struggle with performing better than the NAM during much of both forecast periods (see Figure 5-18). Much of the error appears to be a 10-20 W m^{-2} over-prediction in the morning. It is possible that the observations were influenced by terrain in the area while the forecast was not, leading to an over-forecast in the morning. Finally, the performance of Sun4Cast® is similar to NAM in the morning and evening hours of the intra-day period at SMUD sites (see Figure 5-19). Additionally, at the SMUD sites, there is some statistically significant improvement at hours 8-10 and during the second half of the DA forecast period.

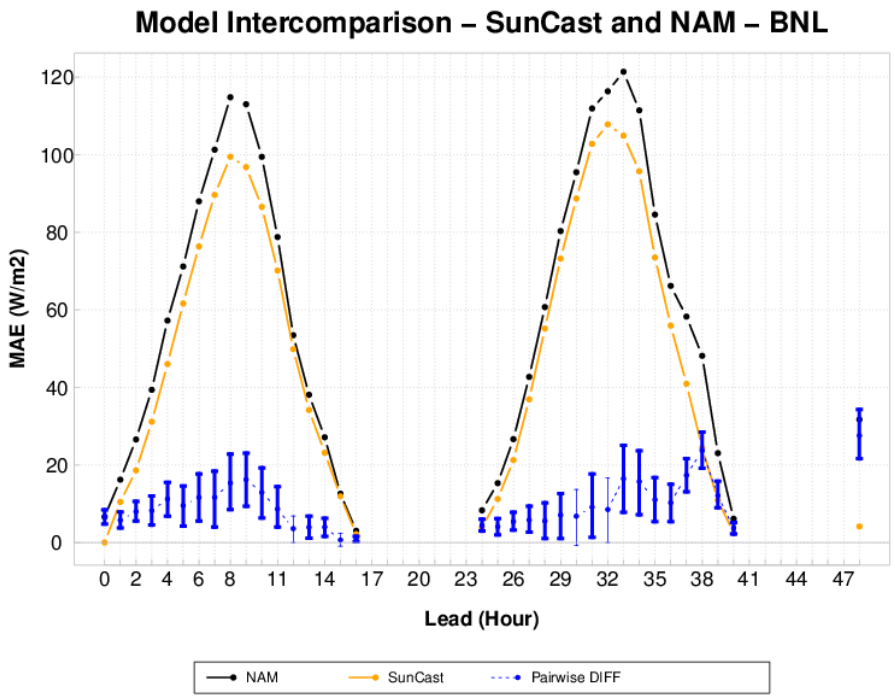


Figure 5-17. Pairwise difference (blue) MAE in W m^{-2} for Sun4Cast® system (orange) and NAM (black) at BNL (tope) for all sky conditions. Difference line has boot-strapped confidence intervals. Bolded confidence intervals indicate statistically significant differences.

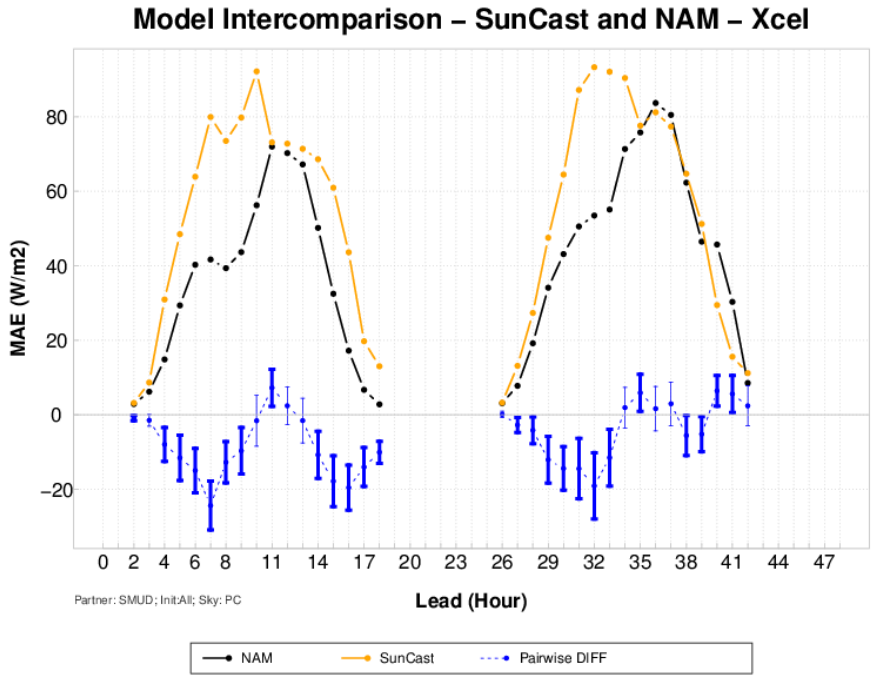


Figure 5-18. Pairwise difference (blue) MAE in W m^{-2} for Sun4Cast® system (orange) and NAM (black) at Xcel for all sky conditions. Difference line has boot-strapped confidence intervals. Bolded confidence intervals indicate statistically significant differences.

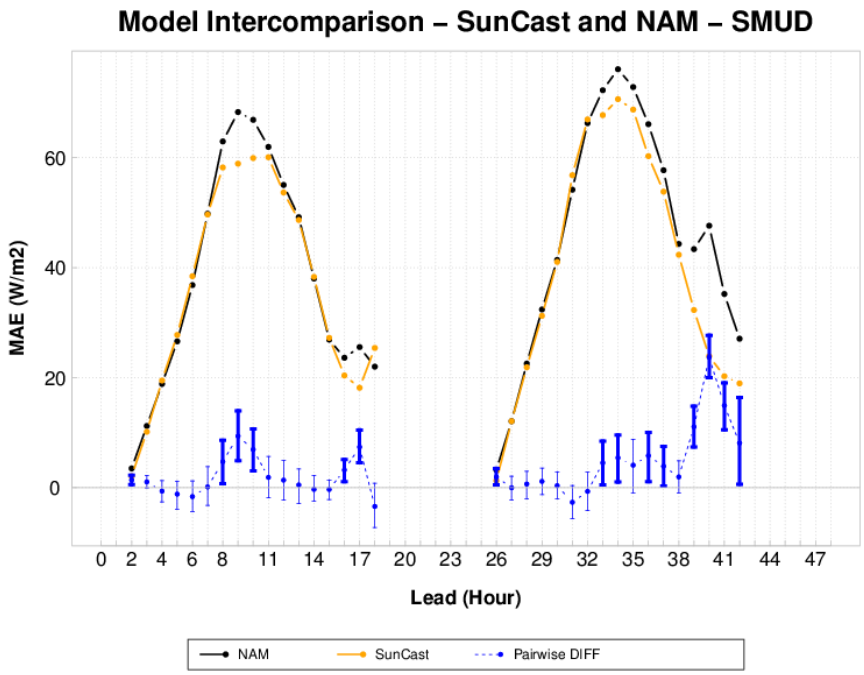


Figure 5-19 Pairwise difference (blue) MAE in W m^{-2} for Sun4Cast® system (orange) and NAM (black) at SMUD locations for all sky conditions. Difference line has boot-strapped confidence intervals. Bolded confidence intervals indicate statistically significant differences.

Table 5-17 summarizes the median percent improvement over the operational baseline, NAM, in MAE for the Sun4Cast® components and the Sun4Cast® irradiance forecasts. The table is for all issue times and forecast hours. In general, WRF-Solar™ was the best performing component, providing a 22-47% improvement over the NAM. In many ways, this improvement is notable because the NAM is at a 4 to 6 times coarser resolution than WRF-Solar™. The typical response of continuous statistics, such as MAE and RMSE, is to doubly penalize the finer-resolution model and hence degrade its score. Therefore, these improvements come despite that double-penalty. Interestingly, the Sun4Cast® irradiance forecast did not match the improvements of WRF-Solar™. Further investigation into why may be necessary to improve the entire system.

Table 5-17. Median MAE improvement in % for Sun4Cast® components versus NAM for Clear, Cloudy (Partly Cloudy and Cloudy) and all sky conditions at the beginning and end of the evaluation period.

	Jan-Mar 2015			Jan-Mar 2016		
Component	**Clear**	**Cldy**	**All**	**Clear**	**Cldy**	**All**
GEM	16.1	8.4	0.5	0.7	18.0	4.2
GFS	6.9	11.7	-3.4	6.1	7.5	3.2
WRFSolar	34.4	35.8	47.2	0	34.9	22.0
Sun4Cast Irr	13.2	28.3	24.5	10.5	19.9	13.0

5.2.6 Sun4Cast® Power

For the power conversion module, from Table 4-1, test values for five farms range from MAE values of 1.3 to 4.4 normalized to capacity (also known as MAPE) were reported with a median value of 2.1%. During BP3, median MAPE was 2.9% of capacity for three farms. The decrease in error may be attributed to different algorithms and longer training on better quality-controlled power data. Power forecasts generated from the Sun4Cast® system for four partners were evaluated to assess skill. Figure 5-20 and Figure 5-21 show the distribution of MAE and mean error (bias) for all plants individually through box plots. Many of the SMUD (blue) and SCE (gold) plants show relatively small errors with narrow error distribution. In contrast, the two Xcel Plants (green and red) show some of the largest range of errors. Much of the difference is likely due to the length of the data record. SMUD and SCE were the first partners for which forecasts systems were established and Xcel was the last, with some substantial data quality control issues involved. In addition, note that the Congentrix plan is concentrated solar and may require more specialized handling than was accomplished in the timeframe available.

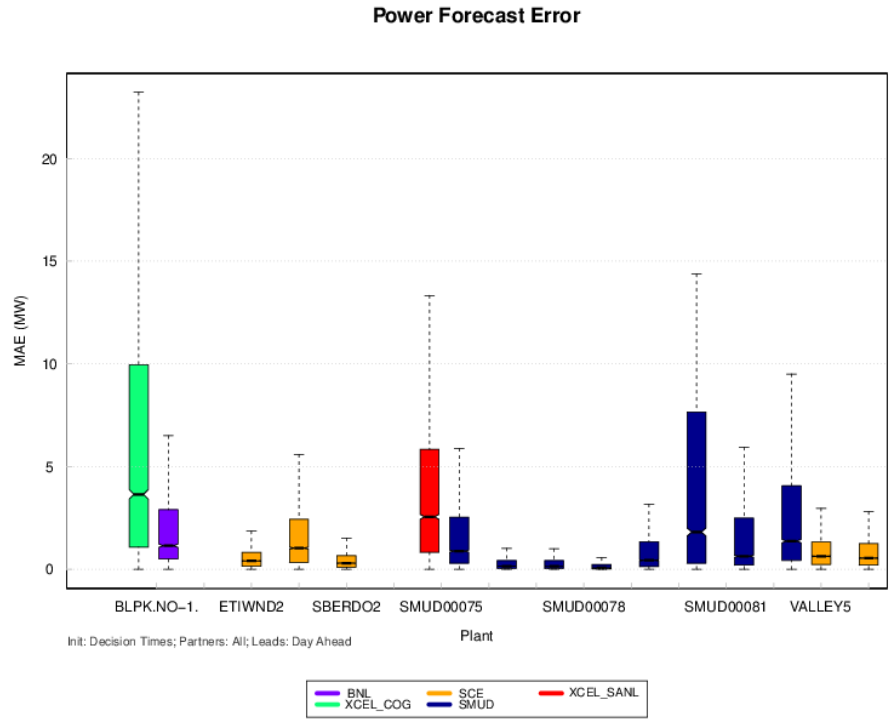


Figure 5-20. MAE in MW for all power forecast sites.

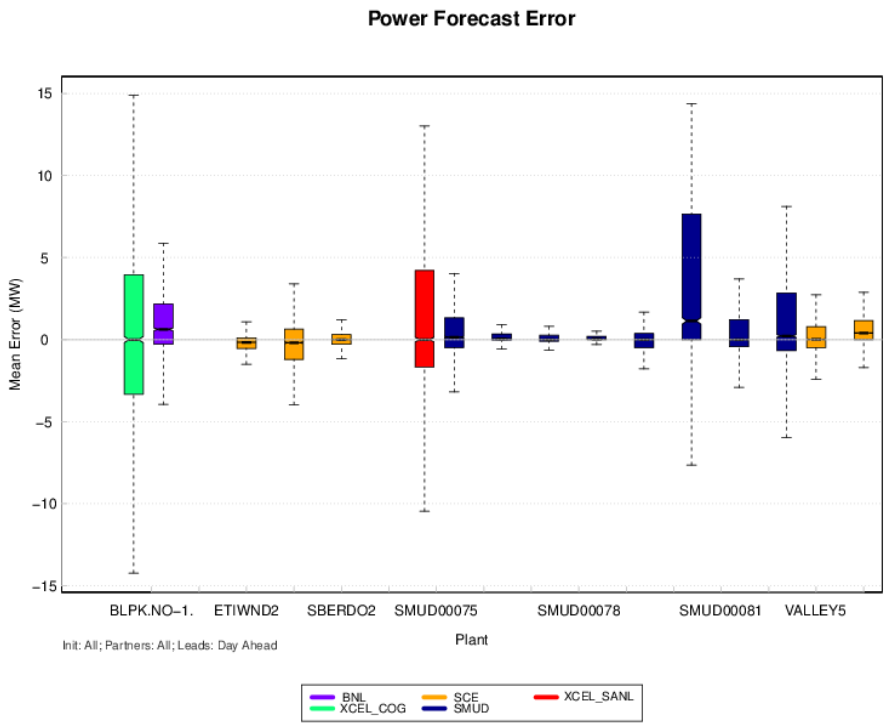


Figure 5-21. Bias (ME) in MW for all power forecast sites.

Table 5-18 provides a summary of some of the base measures along with the maximum absolute error. Forecasts for the Xcel plants tended to have the highest maximum absolute error at 25-29 MW, while those for some of the SMUD plants (76, 77 and 78) were less than 3 MW. Both Figure 5-21 and Table 5-18 indicate that forecasts for most plants were unbiased, however the BNL plant, two SMUD plants (80 and 82), and the Xcel San Luis plant all experienced an over-forecasting bias greater than 1 MW. Half of the locations have MAE normalized to capacity (nMAE) less than 10% and three-quarters of the sites are less than 15% nMAE. The plant with the largest MAE error is a concentrating solar plant with a different configuration than the others. It is clear more work needs to be done to reduce this error. When all the scores for all plants are aggregated and normalized by total capacity, the nMAE is 0.101 or 10.1%.

Table 5-18. Summary of scores for all power forecasting sites.

Partner	Max Abs Error	MAE	nMAE	RMSE	nRMSE	Bias (ME)
BNL	17.1	2.3	0.071	3.6	0.113	1.3
SCE	5.3	0.7	0.081	1.2	0.140	0.0
SCE	19.4	2.2	0.085	3.8	0.148	0.0
SCE	3.1	0.5	0.112	0.8	0.178	0.1
SCE	7.4	1.1	0.103	1.9	0.179	0.4
SCE	5.1	0.9	0.132	1.5	0.221	0.7
SMUD	17.5	2.0	0.111	3.3	0.183	0.7
SMUD	2.4	0.3	0.100	0.5	0.167	0.2
SMUD	2.3	0.3	0.100	0.5	0.167	0.1
SMUD	0.8	0.1	0.100	0.2	0.200	0.1
SMUD	7.6	0.9	0.095	1.5	0.159	0.0
SMUD	14.4	4.2	0.280	6.2	0.413	3.8
SMUD	12.9	1.9	0.127	3.2	0.213	0.8
SMUD	22.7	3.0	0.100	4.8	0.16	1.6
Xcel	29.3	6.3	0.170	9.2	0.249	0.2
Xcel	25.1	4.1	0.137	6.1	0.203	1.7

Scores computed for the Western Wind and Solar Integration Study (WWSIS) and reported in the metrics technical note (Jensen et al., 2016) may be used as a baseline for comparison. Table 5-15 shows nMAE and nRMSE for the SMUD Sun4Cast® system and the WWSIS single plant example. The total capacity forecasted at SMUD is 94.4 MW and for the WWSIS single plant is 100MW. The Sun4Cast® power forecast improvement over WWSIS for the hour-ahead (HA) forecasts is 77.2% and 73.5% in nMAE and nRMSE respectively. For day ahead (DA) forecasts, the improvement is 87.3% and 81.8% respectively.

Table 5-19. Sun4Cast® Improvement over Western Wind and Solar Integration Study (WWSIS) as a baseline.

	Sun4Cast SMUD HA	**WWSIS HA**	**HA % Imp**	**Sun4Cast SMUD DA**	**WWSIS DA**	**DA %Imp**
Capacity	**94.4**	**100**	**-**	**94.4**	**100**	**-**
MAE	2.38	11.1	87.3	1.75	14.8	88.1
nMAE	0.025	0.11	**77.2**	0.019	0.15	**87.3**
RMSE	4.24	17.1	81.2	3.74	22.1	83.1
nRMSE	0.045	0.17	**73.5**	0.040	0.22	**81.8**

5.2.7 Sun4Cast® Analog Ensemble

The Analog Ensemble (AnEn) showed promising results for providing an ensemble mean forecast and uncertainty quantification for GHI forecasts. During the final phase of the project, the technique was applied to power forecasts for SMUD locations. In Figure 5-22, the RMSE of the AnEn mean and Sun4Cast® versus power measurements are plotted versus the 0-72-hour forecast lead for an example application. Overall AnEn provides substantial improvement to the deterministic forecasts as measured by RMSE, MAE, and bias. Improvements in power forecasts are similar to those reported for GHI forecasts with a median improvement over the Sun4Cast® irradiance forecasts of 35% in RMSE and 62% in MAE for the intra-day forecasts. For the day-ahead period, the improvements are lower with 17% in RMSE and 16% improvement in MAE. The largest improvements can be found in the Day 2 forecasts, with 88% and 96% improvements in RMSE and MAE respectively over Sun4Cast® irradiance. It should be noted that the Sun4Cast® errors shown during night-time hours were not included in these calculations. The improvements in HA and DA forecasts were applied to the Sun4Cast® power forecast errors to compute the total Sun4Cast® system improvements over the WWSIS baseline. Table 5-20 summarizes these results and indicates a 90% improvement in MAE for both the HA and DA forecasts periods over the state-of-the-science when this project started (WWSIS). In RMSE, the improvements are on the order of 82-86% for HA and DA respectively.

Probabilistic forecasts were also computed for 10, 25, 50, 75, and 90% exceedance of power capacity. In addition, Brier scores and probability intervals are presented. Figure 5-23 provides a compares the AnEn (black) and Sun4Cast® performance for these thresholds. It shows a marked improvement (lower values for AnEn in black) in Brier score for probabilities of an exceedance of 50% of capacity. The computed Brier Skill Score (BSS) across all lead times is 0.55.

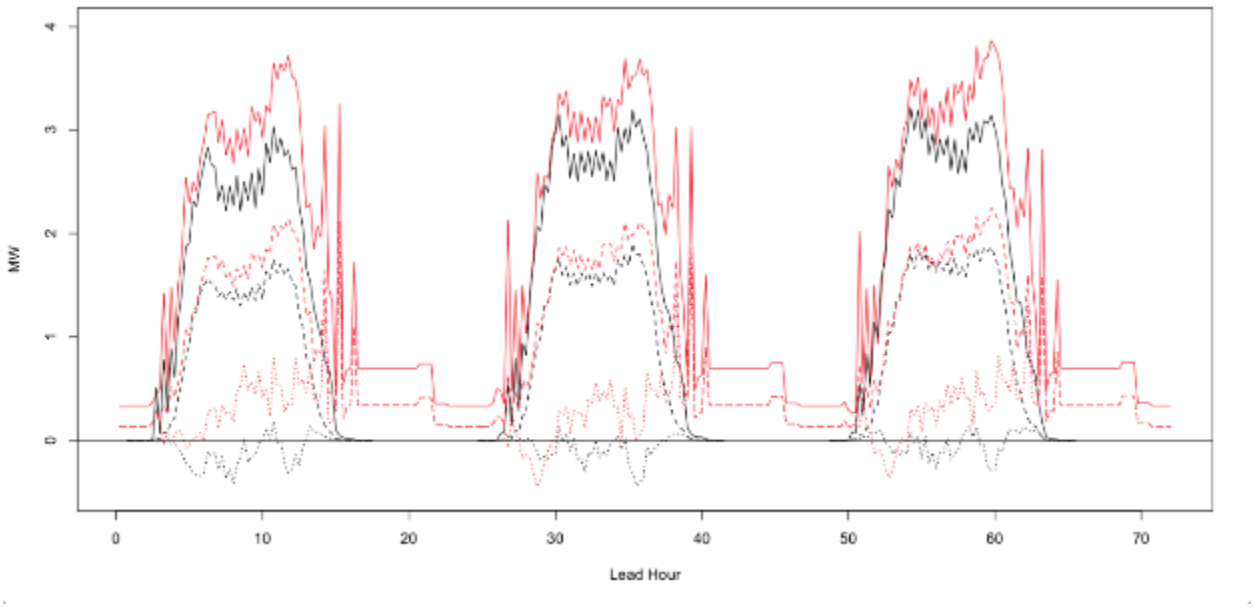


Figure 5-22. RMSE (solid), MAE (dashed), and bias (dotted) for AnEn mean (black) and Sun4Cast® (red) systems.

Table 5-20. Final Sun4Cast® Improvement, after AnEn applied, over Western Wind and Solar Integration Study (WWSIS) as a baseline.

	Sun4Cast SMUD HA	**WWSIS HA**	**HA % Imp**	**Sun4Cast SMUD DA**	**WWSIS DA**	**DA %Imp**
Capacity	**94.4**	**100**	**-**	**94.4**	**100**	**-**
MAE	2.38	11.1	87.3	1.75	14.8	88.1
After AnEn	0.90	-	-	1.45	-	-
nMAE	0.01	0.11	**90.9**	0.02	0.15	**89.7**
RMSE	4.24	17.1	81.2	3.74	22.1	83.1
After AnEn	2.76			3.14		
nRMSE	0.03	0.17	**82.8**	0.03	0.22	**86.8**

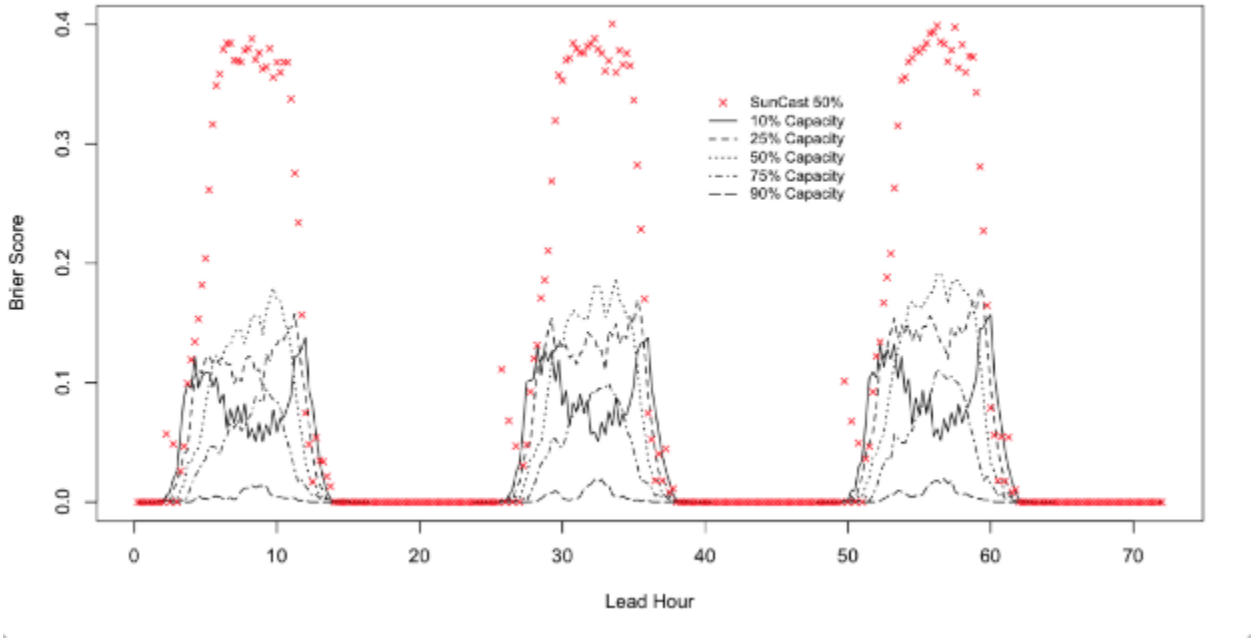


Figure 5-23. Brier Score of AnEn Mean (black) and Sun4Cast® (red) systems for probability of exceeding 10% (solid), 25% (dashed), 50% (dotted), 75% (dash-dot) and 90% (long dash) of capacity for a plant.

Depending on the distribution of the data, two different interval forecasts may capture the same proportion of the observations correctly (i.e., have the same accuracy) but still have different interval widths. To evaluate probability interval forecasts, a set of observed values (at minimum 30, and preferably 100 or more), are accumulated and the percent of observations falling into each category is calculated: 1) falls within the forecasted interval, 2) falls above forecasted interval, and 3) falls below forecasted interval. For a 'perfect' forecast, the correct percentage will fall within the interval and about equal percentages will fall above and below the interval. The AnEn forecasts were evaluated in this manner. Figure 5-24 shows an example of a probability interval evaluation for one of the SMUD power stations. The dashed lines represent to 10^{th} and 90^{th} percentile of power derived from the AnEn members making an 80% interval. The solid black line represents the ensemble mean and the solid red line is the original power forecast from the Sun4Cast® system. The blue points are the observed values.

The percent of observations that fall within the dashed lines over the entire data record is used to assess how well the AnEn predicts power. Table 5-21 summarizes the probability interval percentages for selected lead times in the SMUD Day-Ahead forecasts. A percentage interval of 75-85% is considered. Percentages above 85% indicate that the probability range is too wide and therefore the ensemble is likely over-dispersive. Percentages under 75% indicate the range is too small and the ensemble is likely under-dispersive. All values in Table 5-21 are above 78%. Five lead times over three stations indicate the analog ensemble range was just wide enough. The other seven lead times over four stations are higher than 85%, indicating that the interval is too wide.

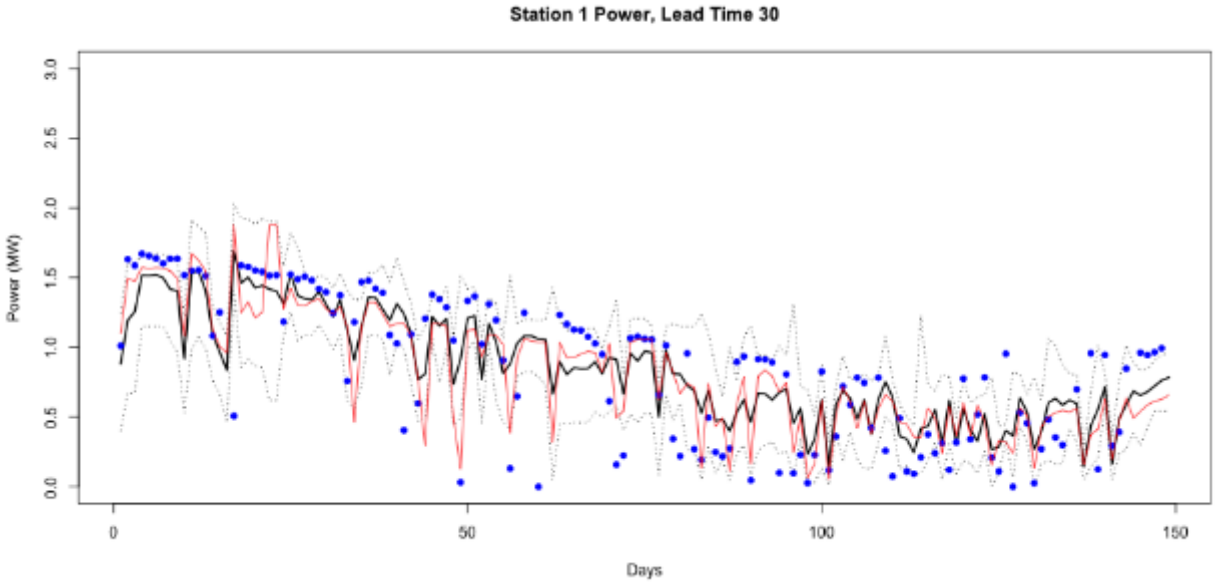


Figure 5-24. Probability Interval plot for SMUD Station 75 including AnEn Mean (black) and Sun4Cast® (red) forecast. 10% probability and 90% probability values represented by thin dashed lines. Percentage of observed values (blue dots) falling with probability interval is summarized in Table 5-21.

Table 5-21. Summary of percent of observations falling between the 10% and 90% probability lines. Green shading indicates percentages near the interval width of 80%.

Lead Time	Stn 1	Stn 2	Stn 3	Stn 4
30	84.7	90.7	88.0	89.3
33	86.0	86.0	78.0	92.7
36	81.3	86.0	80.7	78.0

5.2.8 Metrics Summary

The Sun4Cast® system is a very complex system, affording many opportunities to explore cutting edge research while improving the forecasting benchmark for solar forecast users. The evaluation of such a system is also highly complex. The work completed to define metrics is contained in an NCAR technical note (Jensen et al. 2016). Thirty-two metrics and methods were used during the three years of the project to assess the quality of the evolving forecast system (some not included in this report). These include from the base metrics: MAE, RMSE, bias, standard deviation, skewness, kurtosis, inner quartile range, maximum error, and minimum error. Also from the base metrics, categorical statistics were used to evaluate ramp events (observational base rate, forecast frequency bias, critical success index, Gilbert skill score, probability of detection, false alarm ratio, and Hanssen-Kuiper skill score). From the enhanced metrics, the following were used for evaluation: maximum bias error, correlation, pairwise differences to test variance, normalization, Brier score plus its decomposition, continuous ranked probability score, rank histograms, Brier

skill score, receiver operating characteristic diagrams, reliability (calibration) diagrams, performance diagrams, Taylor diagrams, and frequency of superior performance diagrams.

Over the course of the project, six different components of the NowCast system were developed and systematically evaluated, including TSICast, StatCast-Smartpersistence, StatCast-Cubist, CIRACast, MADCast, WRF-Solar-Now, and a blend using MADCast and WRF-Solar-Now (MAD-WRF). During 2014, basic persistence, where the observation this time period is assumed to persist to the next, was used for the baseline. StatCast-SmartPersistence eventually became the baseline for comparison during the project, which allows for changes in the solar angle as part of a baseline forecast. The following provides a brief summary of the strengths and weaknesses of each component.

TSICast provides forecasts on scales out to 12-14 min, and hence was not incorporated into the NowCast blended system. The range of errors, called inner quartile range, start small at 10 W m^{-2} and grow to 60 W m^{-2} throughout the first 14 minutes. However, throughout the course of the project, the improvement over Smart Persistence remained at approximately 29%.

StatCast provided two components to the system. During 2014, Smart Persistence excelled at improving prediction over basic persistence during the early morning hours but struggled when forecasts were initiated during mid-day. Figure 5-3 indicates that during 2015, errors ranged from 100-220 W m^{-2}. StatCast-Cubist was added to the system in 2015 based on Figure 5-3 improved MAE by 50 W m^{-2}. Figure 5-6 indicates that much of this improvement comes during clear and cloudy conditions.

During 2014, CIRACast showed no improvements reported over basic persistence. The CIRACast team worked to modify their algorithms to handle the challenges provided by this project. Based on Figure 5-4, CIRACast shows generally a 25-40% improvement over Smart Persistence between sunrise and approximately 1600 UTC.

MADCast provided 20-60% improvement over basic persistence in 2014 skill during the first few hours of the forecast but struggled with skill from 3-6 h leadtimes. Figure 5-6 indicates that during 2015, MADCast provided at least 70% improvement over Smart Persistence, with most of that skill being derived during partly cloudy conditions.

WRF-Solar-Now was the top-performing component in 2014 with 20-60% improvements over basic persistence during most leads and across most issue times. Figure 5-7 indicates improvements grew to 50-60% over Smart Persistence for forecasts up to 1600 UTC but Smart Persistence beats out most components when initialized between 1800 and 2000 UTC.

MAD-WRF represents WRF-Solar-Now blended with MADCast fields. It was added to the suite of components to evaluate in August 2015. It was run twice per day at 1400 and 1600 UTC. In Figure 5-6, MAD-WRF appears to decrease (improve) MAE values incrementally over WRF-Solar-Now through 3 hours. MAE improvements for forecasts from 3-6 hours tend to be improved over WRF-Solar-Now by 20%.

NowCast blended forecasts take into account all of these components through an expert system for blending. This expert system was informed by scores computed during 2014. Due to the multiple starts and stops during 2015, the developers were unable to fine-tune the blending for the

evaluations performed during the year. NowCast system scores tend to follow those for MADCast for partly cloudy conditions during all issue (initialization) times reported in Figure 5-6. However, during mid-day forecasts, the NowCast system provides the only forecast with MAE scores lower than Smart Persistence. In Figure 5-5, the NowCast system scores are broken out by partners. When the system was implemented in April 2015 for Xcel Energy, MAE values were 175-190 W m^{-2}. By the end of the project in March 2016, MAE values had dropped significantly to approximately 100 W m^{-2}. This represents a 42-48% improvement over a one-year period at this partner site.

The DICast® system was also established for longer forecasts during the inter-hour forecasting timeline as well as Day-Ahead forecasts. The evaluation of this system focused primarily on the comparison of WRF-Solar™ and the full Sun4Cast® system with the operational baselines. The NAM was established as the baseline for the evaluation. In Table 5-21, WRF-Solar™ and Sun4Cast® systems both tend to improve Day-Ahead forecasting when aggregated over all partners and all issue times. Improvements range from 22-42% for WRF-Solar™ and 13-24% for the blended system. At BNL and SCE partner locations, these improvements were statistically significant at Day-Ahead decision times. At SMUD and Xcel Energy sites, the improvements or lack of improvements are primarily not statistically significant. Finally, for inter-hour forecasts, NAM and WRF-Solar™ have the lowest errors, follow by Sun4Cast® and then the higher resolution operational model, HRRRops, and its parallel research model, HRRRx. Improvements by WRF-Solar™ over HRRRops and HRRRx are on the order of 10-15%. Note that HRRRx already has assumed some of the improvements made to WRF-Solar™.

The Sun4Cast® power conversion module was used to compute power from the irradiance forecasts generated by the Sun4Cast® system. Power conversion errors dropped from 3.1% to 2.1% from 2014 to 2015 based on MAE calculations. Evaluations of power forecasts were performed at four partner locations. In Table 5-18, errors ranged from 0.3 MW to 4.2 MW. Based on Figure 5-14, the larger errors were primarily due to a high bias. Sun4Cast® power conversion improves upon the state-of-the-science WWSIS power forecast performance by 70-80%.

Finally, the Sun4Cast® power forecasts were integrated into an AnEn module to compute ensemble members and ensemble means. The AnEn module decreased RMSE by 17% over the blended Sun4Cast® power forecasts and provided skill in probabilistic forecast at an unprecendented Brier Skill Score of 0.55. It should be noted that the target improvement in Brier Skill Score was 0.1. To satisfy the utility partners' needs, a Probability Interval Evaluation was performed using the AnEn members and developing an envelope representing an 80% band. In Table 5-21, it was shown that the AnEn method provides either the correct width interval or an interval that is slightly too large. With tuning, this method appears to be very promising for adding additional improvements upon the Sun4Cast® system. Once the AnEn is applied, the full Sun4Cast® system improves upon WWSIS power forecast performance by 80-90%.

5.3 Economic Evaluation

5.3.1 Economic evaluation metrics

An important component of the project "A Public-Private-Academic Partnership to Advance Solar Power Forecasting" in building a solar power forecasting system is to help integrate solar energy

into the power mix more efficiently, economically, and reliably, and thus advance higher penetration of renewable energy. One measure of the success of this effort is evaluating the economic benefits of improved solar power forecasts. One dimension of the use of solar power forecasts is in day-ahead (DA) decision making with respect to unit commitment. Working with stakeholder partners, it was determined that production cost modeling (PCM) for use in DA decision-making was a viable approach to evaluate economic benefits of improved solar forecasts in the context of the project.

Although several partners agreed to undertake PCM using their in-house models as a contribution or part of the current project, due to project delays and the stop-start nature of project implementation under DOE control, six of the seven partners who originally agreed to undertake PCM modeling as a contribution to the project dropped out of the economic analysis.

In this report we present analysis of the PCM modeling undertaken by Xcel Energy (Keith Parks) for the Public Service Company of Colorado (PSCo) to derive estimates of the value of reductions in solar power forecast errors. Those model results are then scaled to a national level by NCAR economist Dr. Jeff Lazo.

After some brief background on the economic value of solar forecasts we provide an overview of the methods used in this analysis. Section 5.3.4 presents the analysis followed in section 5.3.5 by a discussion of the results.

5.3.2 Background of Economic Analysis

The decision context for DA planning is determining the likely configuration of generation assets to meet the forecast demand with the ultimate goal of maintaining system reliability. As different types of generation assets require different lead times for start-up and shut-down and face different cost structures in doing so, once DA decisions are made there are constraints in real-time (RT) in meeting actual demand (economic dispatch).

Reductions in total production costs with improvements in solar power forecasting thus represent the avoided costs of forecast errors.

The objective of PCM is to determine the optimal system configuration (e.g., lowest cost) given expected demand (load) while taking into consideration all other relevant factors (e.g., fuel costs, maintenance on facilities, transmission constraints, etc.). Utilities have PCMs set up with the configuration of their systems and use their PCM in day-ahead decision-making, and can also use their PCM for contingency and policy analysis.

PCMs include information on the utility's generation options including coal, natural gas, hydroelectric, nuclear, wind, solar, and other sources. Given primary objectives of maintaining system reliability there is a certain degree of "flexibility" built into the DA decisions. At low levels of solar penetration, solar power variability may fall within built-in ranges of system flexibility, and thus, average errors in solar power forecasts may have no economic impact – and improving forecasts at low levels of penetration may have no economic value.

At higher solar penetration levels, a utility may have to adjust quickly to larger changes in power generation (or load reduction in the case of distributed PV) at higher marginal costs if quick-start

/ stop generation has higher marginal costs than base load generation (e.g., large-scale coal or nuclear). The economic costs of forecast errors may thus increase as solar penetration increases.

This will also depend on the geographic distribution of solar power because with more distributed solar generation, variability may even decrease as non-correlated forecast generation errors cancel out. This is likely in our case study using a scenario of significantly increased solar generation in Colorado but also much more widely distributed than current solar generation (of which all utility-scale solar plants are located now in the San Luis Valley).

In a study similar to the one reported here, Martinez-Anido et al. (2016) use a PCM approach to derive value estimates for day-ahead solar power forecasting improvements for the New England Independent System Operator (ISO-NE) with varying solar power penetrations (4.5%, 9.0%, 13.5%, and 18.0%) and solar power forecasting improvements (25%, 50%, 75%, and 100%). Their analysis indicates that improved solar power forecasting reduced operational electricity generation costs with increasing benefits as penetration levels increase and at higher levels of forecast improvements. We note that from Martinez-Anido et al. Figure 10 (2016, p. 200) at penetration levels of 13.5%, a 50% improvement in solar forecasts generates approximately $13M annual generation cost savings just for NE-ISO.

5.3.3 Methods

5.3.3.1 Modeling Approach

PSCo was undertaking an analysis of increased solar penetration as part of a regulatory requirement with the Colorado Public Utility Commission (PUC) – their Solar Integration Study (SIS). This is based on ongoing policy analysis at PSCo evaluating operations and costs of much larger solar penetration at 1,800 MW, compared to the roughly 300-400 MW of utility scale solar at this time (or rather, the near future). At this time, we do not have additional information on the configuration of the system in the modeling analysis (this is currently proprietary information). Based on phone conversations with Keith Parks (PSCo), the future solar generation used in this analysis is expected to be more distributed geographically across Colorado than current commercial solar generation, which is all in San Luis Valley. Future generation would likely include facilities in northeast and southeast Colorado. The basic approach in this analysis involved:

1) Developed analysis of solar power forecast error (initially undertaken at NCAR by Tara Jensen and Tressa Fowler, based on generation and forecast information in current PSCO solar facilities in the San Luis Valley).
2) “Set up” the PSCo PCM for regulatory analysis (Solar Integration Study) for analysis year 2024 with approximately 1,800 MW solar generation distributed statewide.
3) Developed a profile of solar generation for the year 2024 on an hourly basis (366x24 hours = 8,784 hours).
4) Ran RT PCM for one year for an “actual solution” using the commit from the DA (based on the actuals) and running a RT solution (essentially a perfect forecast). Steam and

combined cycle (CC) generation were fixed at the RT point and combustion turbines (CT) were allowed to commit to meet load and determined operational costs.[1]

5) Added random errors to actuals (based on prior analysis of baseline forecast errors) to generate a "power forecast" (truncated at with no power forecasts below zero) and used that as a solar power forecast to run DA and set up the PCM. Then ran PCM (based on set up with imperfect forecast) and ran in RT mode using the actuals to compute the "actual" cost (to determine costs under forecast error).

6) Ran exact same forecast through with error reduced by exactly 50% at each hour to set up and then ran in RT mode with actual solar power to determine costs.

The output from the runs of the PCM includes hourly solar forecast, solar actual, and costs identified as generation costs, start and shutdown cost, and wind curtailment (also called FO&M Cost in the output spreadsheet) for one year (2024). Keith Parks then calculated total cost for three forecast scenarios (perfect, baseline, and 50% error reduced) costs and calculated aggregate "error costs" and compared these to determine the "cost of forecast errors" for baseline (estimated from the beginning of project at 20%) errors and with 50% reduction in errors (to 10% error, which is roughly what we observed after full Sun4Cast® deployment). Results are shown below after a discussion of the errors used in the modeling.

Figure 5-25 shows four days of solar generation compared to the power forecast (for the first four days of 2024) as run in the analysis scenario with baseline forecast error levels. The analysis was run for the one-year period from January 1, 2024 through December 31, 2024, which includes a leap day, for a total of 366 days or 8,784 hours. The dataset had 4,542 hours with non-zero solar generation. The total MWh solar power generated during the year is 2,900,043.88 MWh. The total solar power forecast was for 3,156,274.70 MWh so an 8.84% over-estimate of power was forecasted over the course of the year.[2]

[1] As stated in the PSCo 2011 Electric Resource Plan Volume 2 Technical Appendix (https://www.xcelenergy.com/staticfiles/xe/Regulatory/Regulatory%20PDFs/PSCo-ERP-2011/Exhibit-No-KJH-1-Volume-2.pdf) about CT generation: "These simple cycle, natural gas fired units are available in a wide range of sizes (25 MW to 300 MW). Combustion turbines are very similar to a jet engine with an electrical generator connected to the turbine shaft. Combustion turbines are typically inexpensive to build but are less efficient sources of generation. The ideal role for CTs is to be run for a few hours of the year typically at times of the highest electric demand."

[2] As actual solar power cannot fall below zero, large negative errors added to actuals are truncated at zero and thus on average there will be a positive bias to the added errors.

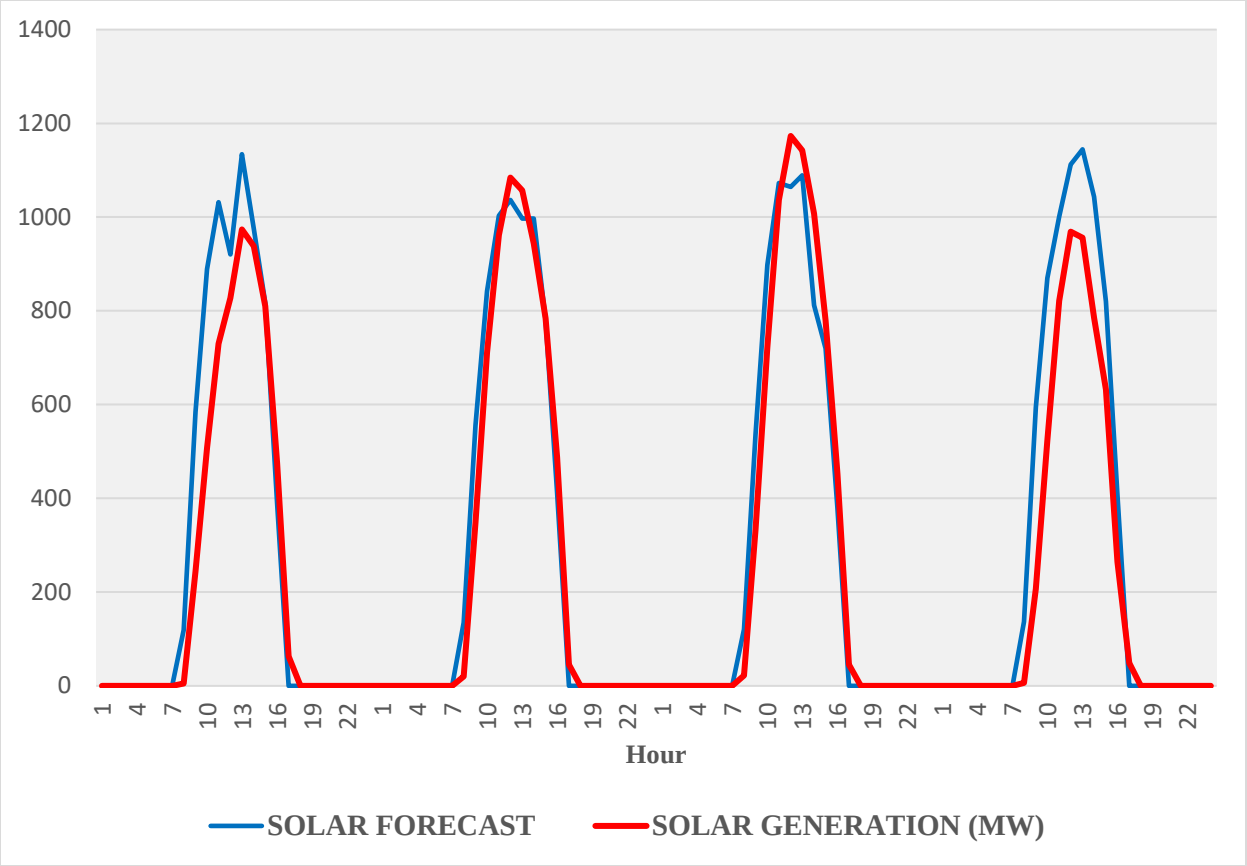


Figure 5-25: Example of solar forecast compared to actual solar generation for a four-day period in January 2024.

5.3.3.2 Error of Solar Power Forecasts

For non-zero hours of actual production (4,542 hours), the average hourly solar power actual is 638.49 MWh and average hourly forecast is for 694.91 MWh so there is a slight upward bias in forecasts. Given that there could be a solar power forecast for a given hour and no actual solar production due to weather conditions, the percent error for any one hour can be considerably larger than 100%. Over the 4,542 actual hours the baseline mean absolute forecast error baseline was 128.03 MWh or roughly 20%. Figure 5-26 shows a histogram of the errors used in the current analysis (note this is the bias error and not the absolute error).

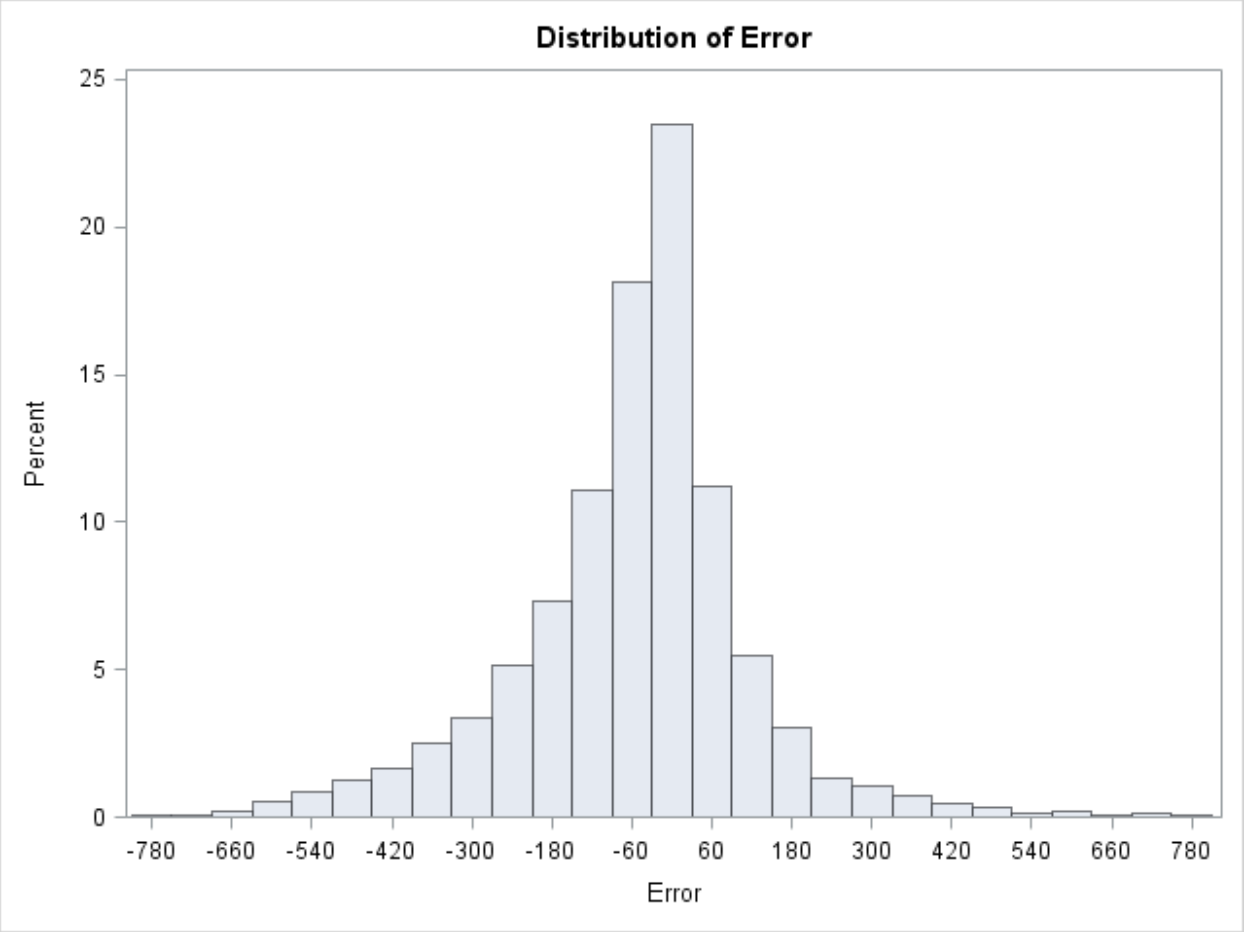


Figure 5-26: Distribution of solar forecast bias errors (MW) using baseline forecast error.

5.3.3.3 Results

Using the basic model output from the PCM scenarios, Keith Parks aggregated component and total annual costs for the different scenarios. We summarize these results in Table 5-22. For each scenario the generation costs, start and shutdown costs, and wind curtailments costs are aggregated to total production cost (for the year).

Table 5-22 PCM output and calculation of cost savings from reduction in forecast error.

Property	Perfect Forecast $(000)	Forecast (1800MW of solar state-wide) $(000)	50% Improvement to Forecast (1800MW of solar state-wide) $(000)
Generation Cost	1,172,848.82	1,173,816.71	1,173,056.56
Start & Shutdown Cost	21,782.52	22,009.18	21,985.44
Wind Curtailment Cost	13,612.59	13,603.99	13,568.06
Total Production Cost	1,208,243.94	1,209,429.88	1,208,610.06
Error Cost (Forecast Cost minus Perfect Cost)		1,185.94	366.12
Value of forecast improvement (Cost with forecast minus cost with 50% improvement to forecast)			819.82

The “error cost” row shows the difference in total production costs between scenarios with forecast error and with perfect forecast for the two forecast improvement scenarios. For the 2024 analysis, the total cost due to baseline levels of forecast error (roughly 20% mean absolute error) is $1.19 million. With a 50% reduction in forecast error the error cost falls to $366,000 – a $819,200 savings or 69% reduction in error costs due to a 50% reduction in forecast errors. With a total reduction in forecast errors over the year to 290,755 MWh, this averages into a production cost savings of $2.8196 per MWh reduction in error.

5.3.4 Analysis

5.3.4.1 Regression Analysis

To take advantage of the significant quantity and variability of the PCM output data we undertook a regression analysis of the relationship between production costs and forecast errors – including other variables from the data set that may also influence production costs to better understand the specific impact of forecast error on costs. The dependent variable in the regression analysis is the difference (delta) in the hourly production cost (PC) with forecast error and with no forecast error (e.g., perfect forecasts) as shown in Equation 5.1.

$$\Delta PC = PC(\text{with forecast error}) - PC(\text{with perfect forecast}) \quad (5.1)$$

A set of linear regression models was fit adding a series of explanatory variables to understand potential influences on production costs and best isolate the effect of forecast errors. Future work could explore a broader range of econometric models, including looking at time lags, as the set-up of the system in any one period is likely affected by prior periods that are interdependent with the forecast error. Equation 5.2 shows the regression model,

$$\Delta PC = f(\text{forecast error, error reduction mode, TC, Solar MWh, month, hour}) + \varepsilon \quad (5.2)$$

where f denotes a functional relationship, TC is total production costs, Solar MWh is the absolute level of actual solar power each hour, month and hour are the month and hour of the observation, and ε allows for noise. As costs are expected to be lower with perfect forecasts than with error in the forecasts, we expect the ΔPC to be positive and increase with larger error. Forecast errors were included as the absolute value of the difference between the forecast's MWh for the hour and the actual MWh.

To take full advantage of the PCM output, we included all model output by "stacking" the data set to include PCM output with baseline error and with 50% reduction in error. As we "stacked" the data from the two PCM model conditions, we included an "error reduction mode" to test for "structural" differences in costs in the two error regimes. This dummy variable is set to "1" for output from the 50% forecast error reduction (and "0" for the baseline forecast error). A significant parameter estimate on this may be an indicator of non-linearities in the response to error reduction.

Including "TC" (or total production costs) as an explanatory variable also examines potential non-linearities in differences in production costs as a function of total production costs. This may be similar to examining the impact of penetration levels or serve as a proxy for total production (e.g., total MWh from all fuel sources that are not currently included in the data set). We expect that the larger the value of TC, the smaller the difference in TC due to changes in absolute forecast error, as such error then has a smaller marginal role in the day-ahead decision.

Including Solar MWh (or the absolute level of actual solar power each hour) as a predictor may capture the relative impact of absolute forecast error as well. The larger the Solar MWh the lower relative impact of any specific level of absolute forecast error (e.g., lower percent error) and thus potentially less impact on overall production costs.

We included dummy variables for month and hour (11 month dummies with December as the excluded month and 23 hourly dummies with 23:00 the excluded hour) to examine if the difference in production costs in the different forecast modes varies by month or hour. We have no *a priori* expectations on these variables but are aware that there are likely seasonal and time-of-day aspects of both solar generation and overall power generation that may influence the difference in costs related to different forecast errors. (At this time we have not included the regression output on hours given no *a priori* expectations on signs or reasonable interpretation of results).

For the monthly dummies we included the dummies as stand-alone explanatory variables and, in a different regression model, interacted the monthly dummies with the absolute error. This interaction term examines different potential impacts of forecast errors during different months. We have no *a priori* expectations on these interactions but, if significant, could suggest future work to better understand how forecast errors play different roles in different times of the year.

We included all hours of the day (including hours without sun) as there were differences in production costs in some of the "dark" hours likely related to differential set-up costs related to forecast errors.

We note that the adjusted R-squared on all of the models is relatively low, but (in part due to the large number of observations) we find a number of significant results. We also note that at this time we have generally only implemented linear regression analysis (ordinary least squares) without any analysis of explicit non-linearities, potential multi-collinearity, heteroscedasticity, or intertemporal correlations. Future work could examine more sophisticated econometric modeling. Table 5-23 displays the results for four different models as described below.

Model 1: The simplest model examines only the potential impact of forecast error on differences in production costs between perfect and imperfect solar forecast conditions. The positive and highly significant parameter estimate (3.936) indicates that a 1.0 MWh increase in absolute forecast error results in a production cost increase of $3.94.

Model 2: The second model adds three explanatory variables Error Reduction Mode (with Current mode = 0 and the 50% Reduction mode = 1), Total Production Cost (the total cost of energy production for that hour across all generation), and Solar Generation (the total MWh actual production in that hour).

The negative and significant parameter estimate on Error Reduction Mode of (-1488.83) indicates that on average, after controlling for the magnitude of the error, production costs in the low error mode are close to $1,500 less per hour than in the baseline error mode. This may be an indication of non-linearities in the relationship between forecast error and costs where there is a "structural" reduction in costs when errors are lower on average.

Table 5-23. Regression analysis. Dependent variable is the difference in hourly production cost between "With Forecast Error" and "With Perfect Forecast" (hourly error cost). Significant estimates highlighted in bold (N=17,568).

	Model 1	Model 2	Model 3		Model 4
Model Fit					
F Value	48.320	117.790	44.830		35.510
Adj R-Sq	0.003	0.026	0.036		0.029
Intercept	**-857.513 *****	**2807.412 *****	**4081.523 *****		**2970.719 *****
T value	**59.588**	**250.990**	**12.650**		**11.710**
Absolute Error of Forecast	**3.936 *****	**4.946 *****	**4.667 *****		**2.752 ****
T value	**0.566**	**0.623**	**7.440**		**1.980**
Error Reduction Mode		**-1488.826 *****	**-1498.076 *****		**-1475.739 *****
T value		**105.875**	**-14.220**		**-13.940**
Total Cost		**-0.018 *****	**-0.023 *****		**-0.019 *****
T value		**0.002**	**-13.670**		**-11.490**
Solar Generation (MWh)		**-1.430 *****	**-1.380 *****		**-1.500 *****
T value		**0.131**	**-10.410**		**-11.280**
Jan (1 for Jan; 0 otherwise)			-63.504	**Jan*AbsErr**	2.474
T value			-0.250		1.060
Feb			-315.314	**Feb*AbsErr**	1.282
T value			-1.230		0.650
Mar			**-765.545 *****	**Mar*AbsErr**	0.149
T value			**-3.030**		0.080
Apr			**-1408.566 *****	**Apr*AbsErr**	1.781
T value			**-5.470**		0.810
May			**-1324.402 *****	**May*AbsErr**	0.265
T value			**-5.150**		0.110
Jun			**-1447.707 *****	**Jun*AbsErr**	0.350
T value			**-5.630**		0.130
Jul			290.886	**Jul*AbsErr**	**11.781 *****
T value			1.150		**4.330**
Aug			365.072	**Aug*AbsErr**	**9.803 *****
T value			1.450		**4.560**
Sep			-179.455	**Sep*AbsErr**	2.530
T value			-0.700		1.050
Oct			**-1828.982 *****	**Oct*AbsErr**	-3.069
T value			**-7.220**		-1.430
Nov			168.272	**Nov*AbsErr**	**6.072 *****
T value			0.660		**2.860**

*, **, and *** represent significant p values at 10%, 5% , and 1% respectively

The negative and significant parameter estimate on Total Production Cost of -0.018 indicates that as total production costs increase the difference in total production costs between imperfect and perfect forecast decreases. This may be an indication that (assuming total costs are monotonically increasing with total energy production) there is a decreasing impact of forecast error as production increases. This may also be an implicit indication of the impact of penetration levels – higher total energy production translates to lower penetration keeping solar production constant. The negative parameter estimate would thus indicate that there is a smaller impact of forecast error at lower penetration levels.

The negative and significant parameter estimate on Solar Generation (-1.430) indicates that as total MWh of solar production increase, the difference in production costs with imperfect and perfect forecasts decreases. This may indicate that proportionally smaller errors (holding absolute error constant, percent error falls as total solar generation increases) have a smaller impact on the difference between production costs with imperfect and perfect forecasts.

Having controlled for these other influences on production costs, the parameter estimate on Absolute Error of Forecast increased to $4.95/MWh in Model 2 from $3.94/MWh in Model 1.

Model 3: This model includes a set of dummy variables for 11 months (December is the excluded dummy and thus the other parameter estimates are interpreted with respect to differential production costs in December). A significant parameter estimate on any of these dummy variables would indicate that there is significant impact of forecasts errors on the difference in production costs with perfect and imperfect forecasts in different months. March, April, May, June, and October all had negative significant parameter estimates indicating that the difference in production costs with perfect and imperfect forecasts was lower in these months compared to December (the excluded month). We have no *a priori* expectation on this or specific explanation and feel it may be worthwhile to examine this in future research.

Model 4: This model included a set of dummy variables for 11 months (December is excluded again) but each is interacted with the absolute forecast error for that observation. Significant parameter estimates here would suggest that the impact of forecast errors on differences in production costs varies by time of year. Positive and significant parameter estimates in July, August, and November indicates that the cost of forecast errors is higher in those months (compared to December). For any given month the change in production cost due to a one-unit change in forecast error is the sum of the parameter estimate on forecast error and the parameter estimate on forecast error month dummy for that month as per Equation 5.3.

$$\Delta PC = \beta_{error}*(\text{Forecast Error})+\beta_{month}*(\text{Forecast Error*Month}) \quad (5.3)$$

Therefore, from Model 4, for the month of July would suggest that the cost of a one-unit (1.0 MWh) error in the forecast is $14.53 – considerably more than a cost of $2.75 for a one-unit forecast error in December.

$$\Delta PC = \beta_{error}*(\text{Forecast Error})+\beta_{july}*(\text{Forecast Error*July}) = 2.752+11.781=\$14.53 \quad (5.4)$$

As before we had no *a priori* expectations on dummy variable sign or magnitude and believe this is amenable to future research given the indication of significant differences month to month of the impact of forecast errors (and thus in reducing forecast errors).

We have not presented the regression results using "hours" in the analysis at this time as there were no *a priori* expectations on the estimates and no specific meaningful interpretation of the results.

5.3.5 Results and Discussion

Overall the results here are compatible with those of Martinez-Anido et al. (2016) in that higher levels of solar penetration are associated with higher costs of forecast error and that greater reductions in forecast error save more in terms of reduced or avoided costs – likely in a non-linear manner.

5.3.5.1 Benefit to Forecast Error Reduction in Xcel Energy (2024)

As noted in Table 5-22, the PCM analysis showed a direct savings of $819,200 from a 50% reduction in forecast error. To compare that result to the results of the econometric analysis we use the regression results to calculate a benefit based on 50% error reductions from the current forecasts. Table 5-24 shows the cumulative (total annual) absolute forecast error, the cumulative error with 50% reduction, and the difference in cumulative error (which is simply also 50% of the baseline error. Using the parameter estimate from the regression analysis in Model 1 of $3.94 cost reduction per MWh error multiplied by the total error reduction indicates a total annual value of $1.14M. This is reasonably comparable to the direct approach used in Table 5-24 of $891,820.

Table 5-24. Economic value of a 50% reduction in forecast error for Xcel Energy in 2024.

Cumulative forecast error (baseline) (MWh)	581,511.34
Cumulative forecast error (50% error reduction) (MWh)	290,755.67
Difference in error	290,755.67
Cost of Error per MWh	$3.93589
Total value of 50% reduced forecast error	$1,144,382.33

5.3.5.2 Aggregation to National Values

While the PSCo analysis is unique in that each utility is unique, we use the results from this analysis to suggest aggregate national values for solar power forecast error reduction. The Energy Information Administration projects increased solar energy generation growing at 6.8% annually from 2013 to 2040 (Energy Information Administration, 2015; Table A.16). By 2040 solar energy (measured as net summer generation) is projected to be 110.1 billion kilowatt hours (up from 18.5B kWh in 2013). This would then represent 13.56% of all generation in 2040 compared to 3.8% in 2013 (as per EIA 2015). Figure 5-28 plots the EIA projections for solar power generation indicating a steady increase over the next 25 years.

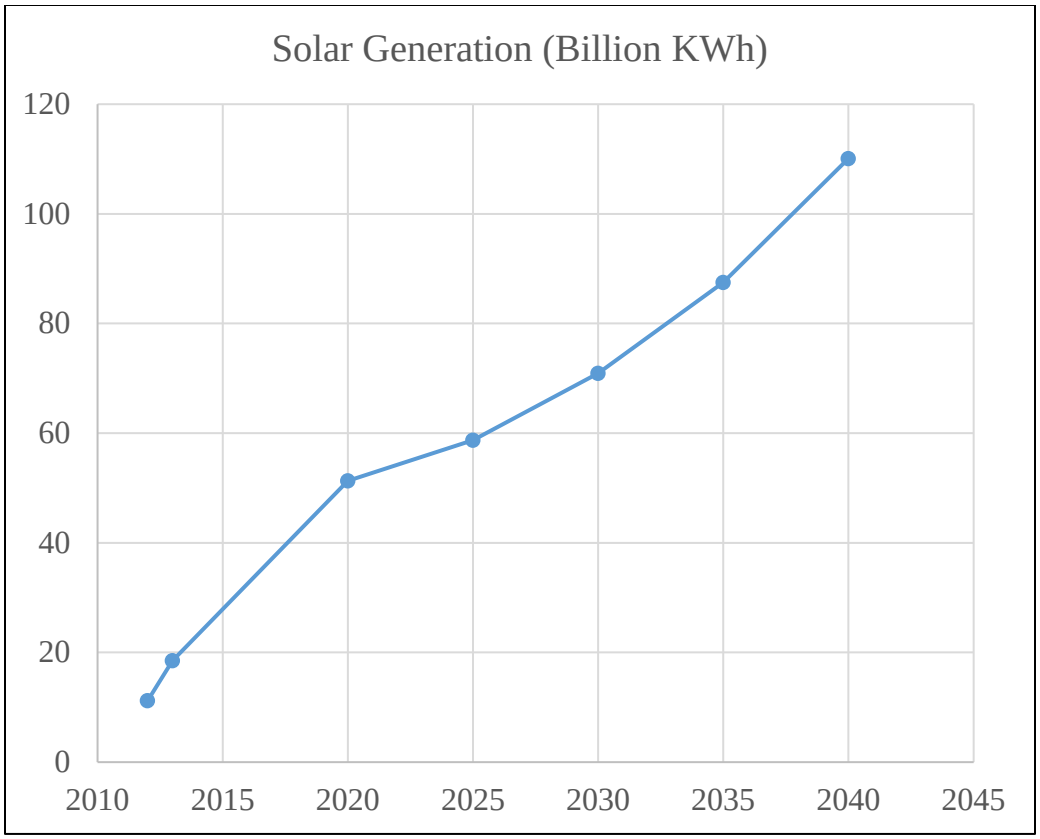


Figure 5-27: Projected solar generation (Source EIA 2015, Table A.16).

Using EIA projections of solar power generation and per MWh savings for forecast error reduction from the baseline analysis, we generate order-of-magnitude estimates of the national value of improved (50% error reduction) solar power forecasts. Using the Billion kWh projections from EIA and a baseline percent absolute error of 20% (the mean baseline error used in the current analysis), we calculate total annual forecast error in terms of MWh. Assuming a 50% reduction in MAE through the solar power forecast improvement program this gives an estimate of the reduction on total forecast error (in MWh). Multiplying this by the "per MWh savings" estimated from the regression model ($3.94 avoided generation cost per MWh reduction in forecast error) we generate annual national benefit estimates as indicated in Table 5-25.

Recognizing, as noted, that this is order of magnitude, there are several caveats to this national aggregate benefit estimate. First we note that the EIA generation projections are for summer generation, which is likely peak generation for solar power, and thus we may be overstating annual total power generation. As noted in Martinez-Anido et al. (2016) and in our analysis, benefits are not linearly related to penetration levels, whereas the current analysis does not account for this. The benefits also depend on baseline levels of error and future levels, as well as adoption of forecasting improvements, and we have not assessed these on national levels (i.e., we are assuming PSCo is nationally representative).

Table 5-25. Aggregation to national benefit estimates at assumed solar photovoltaic penetrations.

Year	2012	2013	2020	2025	2030	2035	2040
Billion KWh	11.2	18.5	51.3	58.7	70.9	87.5	110.1
MWh	11,200,000	18,500,000	51,300,000	58,700,000	70,900,000	87,500,000	110,100,000
20% Absolute Error (MWh)	2,240,000	3,700,000	10,260,000	11,740,000	14,180,000	17,500,000	22,020,000
50% Absolute Error (MWh)	1,120,000	1,850,000	5,130,000	5,870,000	7,090,000	8,750,000	11,010,000
$/GW Error Reduction	$3.94	$3.94	$3.94	$3.94	$3.94	$3.94	$3.94
Benefit	$4,408,197	$7,281,397	$20,191,116	$23,103,674	$27,905,460	$34,439,038	$43,334,149

Also as indicated in Martinez-Anido et al. (2016) (Figure 5-25), a 13.5% level of solar penetration and 50% forecast improvement generated roughly $13M in benefits just for NE-ISO. Given those results, the current aggregation to national benefits may even be an underestimate.

Finally, we note that the benefits from a research program to improve solar power forecasting are ongoing. While there are likely also ongoing costs for observation, modeling, and forecasting systems, we do not have an estimate of those in order to develop benefit-cost estimates. As indicated in Table 5-26, extrapolating linearly between the specific years indicated in the EIA projections, using a 3% rate of discount[3], and a 26-year analysis timeline starting in 2015, we develop an estimate of the present value of benefits from the forecast improvements of $455M in production cost savings.

[3] The discount rate of 3% reflects a relatively conservative approach to discounting given federally suggested nominal rates of 3.2% on 20 year and 3.5% on 30 year projects as per Office of Management and Budget (OMB) Circular A-94 Appendix C. Revised November 2015. (https://www.whitehouse.gov/omb/circulars_a094/a94_appx-c).

Table 5-26. Present value of national benefits.

Year	Current Value Benefit	Discount Rate 3%	Present Value Benefit
2015	$10,969,888	1.000	$10,969,888
2016	$12,814,133	0.971	$12,440,906
2017	$14,658,379	0.943	$13,816,928
2018	$16,502,625	0.915	$15,102,239
2019	$18,346,870	0.888	$16,300,956
2020	$20,191,116	0.863	$17,417,034
2021	$20,773,627	0.837	$17,397,586
2022	$21,356,139	0.813	$17,364,495
2023	$21,938,651	0.789	$17,318,574
2024	$22,521,163	0.766	$17,260,596
2025	$23,103,674	0.744	$17,191,303
2026	$24,064,031	0.722	$17,384,368
2027	$25,024,389	0.701	$17,551,603
2028	$25,984,746	0.681	$17,694,347
2029	$26,945,103	0.661	$17,813,887
2030	$27,905,460	0.642	$17,911,453
2031	$29,212,176	0.623	$18,204,062
2032	$30,518,891	0.605	$18,464,431
2033	$31,825,607	0.587	$18,694,190
2034	$33,132,322	0.570	$18,894,900
2035	$34,439,038	0.554	$19,068,060
2036	$36,218,060	0.538	$19,468,992
2037	$37,997,082	0.522	$19,830,392
2038	$39,776,104	0.507	$20,154,224
2039	$41,555,127	0.492	$20,442,369
2040	$43,334,149	0.478	$20,696,631
	Present Value Total Benefits		**$454,854,415**

5.4 PREDICTABILITY OF SOLAR IRRADIANCE

A tiger team was formed as part of this project to assess the limits of predictability of solar irradiance. Although not much work has been accomplished specific to irradiance, various researchers have addressed the general predictability issue. The first serious study dates to Lorenz (1969), which posed the issue of whether the atmospheric system predictability is constrained or not. That work builds on Lorenz's (1963) seminal earlier work demonstrating that the atmosphere is chaotic with a sensitive dependence on initial conditions, and led to development of a full field of study in the US and abroad as recently reviewed by Bunimovich (2014). Lorenz's 1969 work specifically states that "the uncertainty demanded by Heisenberg's Principle appears not to be very significant, because of the much greater uncertainty resulting from our failure to observe the state of the atmosphere and formulate the governing equations with anything approaching perfection" (pp. 289-90). Lorenz (1969) set up a system of low-order equations of the basic state of the atmosphere to test whether a small error will become much larger than the initial error and whether reducing initial error will commensurately reduce the final error. He shows that each scale of motion exhibits a finite range of predictability.

A yet earlier work by Gleason (1967) proposed theoretical limits to atmospheric predictability. Various studies between then and the present time have sought to further assess this issue. Much of that work was focused on the synoptic scales, which are critical to defining medium-range predictability. An important recent effort is that of Rotunno and Snyder (2007), which used a surface quasi-geostrophic equation as the basis, which is more appropriate for the power spectrum assumed by Lorenz (1969). They found that the specific model was less important than the power of the energy spectrum (-5/3 vs. -3). These studies have advanced our understanding, but say little about predictability at cloud scale. This partnership began the process of investigating this issue in more detail.

One approach to the investigation was to quantify the variability of the resource, and then to forecast it. We note that not only does this approach provide information to the team on the limits of predictability, but it also leads toward forecasting the variability. Utilities have stated that prediction of the variability is important to their operations: that information is useful in determining the amount of reserves that must be carried at a particular time (Bartholomy et al. 2014).

That work to quantify the variability was accomplished by the University of Washington team led by Laura Hinkelman (Hinkelman 2014; Sangupta and Hinkelman 2014; Hinkelman et al. 2015; Schaeffer et al. 2016). This work, reported in section 2.2.2, quantified the variability of the resource due to satellite observations of clouds and well as from ground measurements.

The team followed up by forecasting the variability. The first example of this uses the StatCast approach to predict variability using the Cubist and RD-ANN approaches. The RD-ANN results appear in section 2.2.7. The results using StatCast-Cubist are summarized below as drawn from McCandless et al. (2015). The data used to train Statcast-Cubist to predict variability of irradiance at the SMUD sites was detailed in section 2.2.6. In that data we matched the SMUD GHI observations with the hourly METAR weather observations, where each 15-minute SMUD GHI interval is matched with the corresponding METAR observations at the top of the next hour. For example, the irradiance observation at 26 January 2014 15:15, 15:30, 15:45, and 16:00 would all

be matched with the meteorological data from 1/26/2014 16:00 to form a composite dataset. In addition to the GHI temporal or spatial standard deviations in the last 15-minute interval, the GHI temporal or spatial standard deviations from the previous three 15-minute intervals are also included in the predictor dataset. These four GHI standard deviations for the 15-minute intervals are provided as predictors so that the forecasting technique can model the recent trend in variability. Using the prior four 15-minute intervals is appropriate because SMUD reports data on an hourly basis, so this study matches what would be implemented operationally. The predictors include the previous observations, observed weather data, and time information, which are all fed into the Cubist model regression tree to predict the variability. The temporal standard deviation datasets include 40,127 instances combined for all eight SMUD sites. The spatial standard deviation dataset consists of 4057 instances aggregated at all eight sites. Instances where one or more location had missing data were omitted from the spatial standard deviation training and testing datasets.

Ten-fold cross-validation randomly partitions the data into ten subsets to be used for training and testing the model and provide an assessment of how the model tree generalizes to an independent set of data. The training of the model tree was performed on nine of ten subsets and the remaining subset was used as validation. This process was repeated for all of the ten subsets and the errors were averaged over the ten repetitions to reduce variability in the results. We analyzed the model tree's predictive ability for GHI spatial variability by examining the predictive skill of the standard deviation of the GHI among the SMUD observation sites. The MAE for the GHI spatial standard deviation prediction increases with forecast lead time as did that of the GHI temporal standard deviation prediction (Figure 5-29). However, the error range over the forecast lead times is greater than that for the temporal data. Values of the MAE range from approximately 15 W m^{-2} at 15 minutes to 21 W m^{-2} at 180 minutes.

The relative error of the Cubist model tree compared to climatology (the mean GHI spatial standard deviation computed on the training dataset) is plotted in Figure 5-30. The relative error for the model tree begins at approximately 0.35 for the 15-minute forecast lead time and levels off at about 0.50 for forecast lead times longer than 75 min. This relative error provides evidence that the model tree is able to provide utility companies with at least twice the accuracy as assuming climatological average variability. This is a meaningful result for utility companies that have regional coverage with a range of distributed rooftop solar and solar power farms because the model tree is able to provide a substantial increase in the accuracy of predicting short-term solar radiation variability across a region.

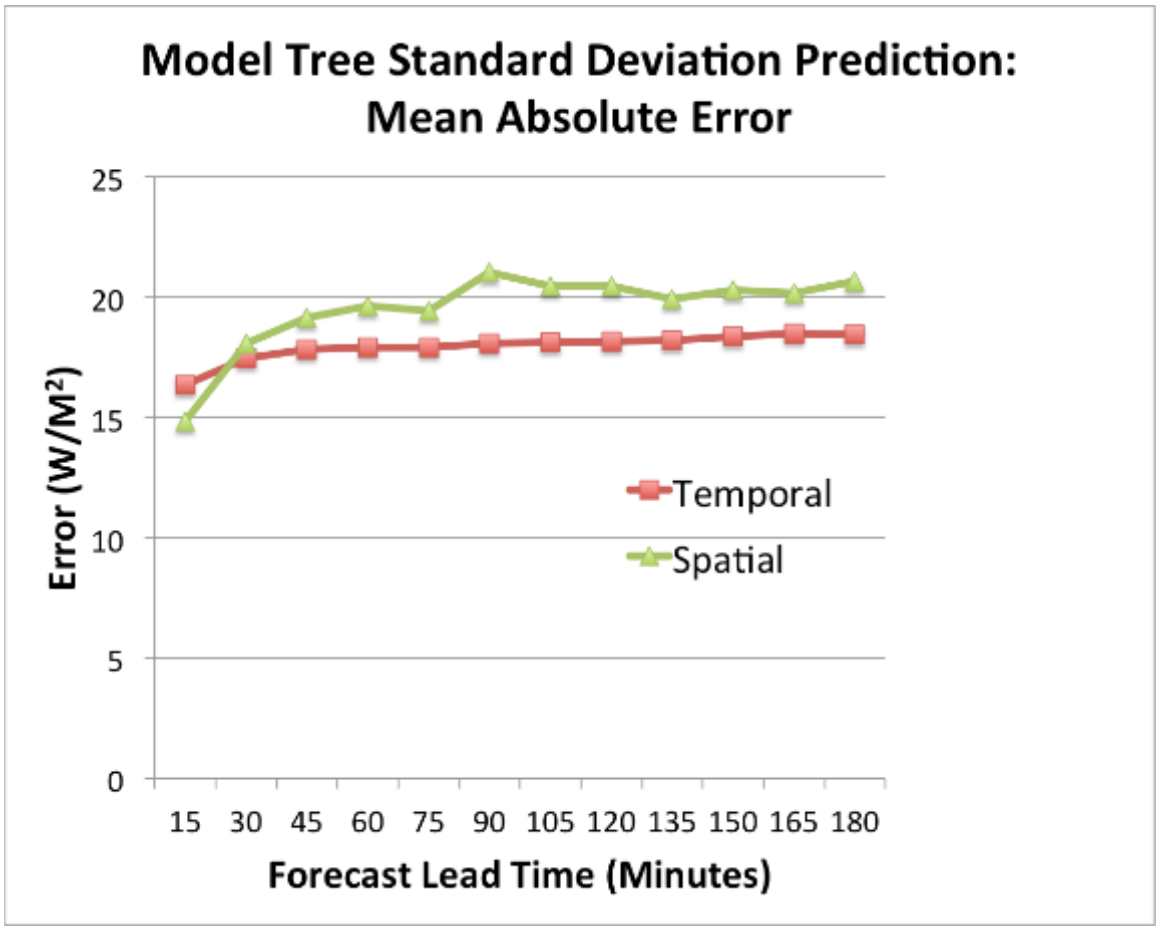


Figure 5-28. Mean absolute error (MAE) for the model tree on the spatial standard deviation of the GHI (green - triangles) and temporal standard deviation of the GHI (red – squares).

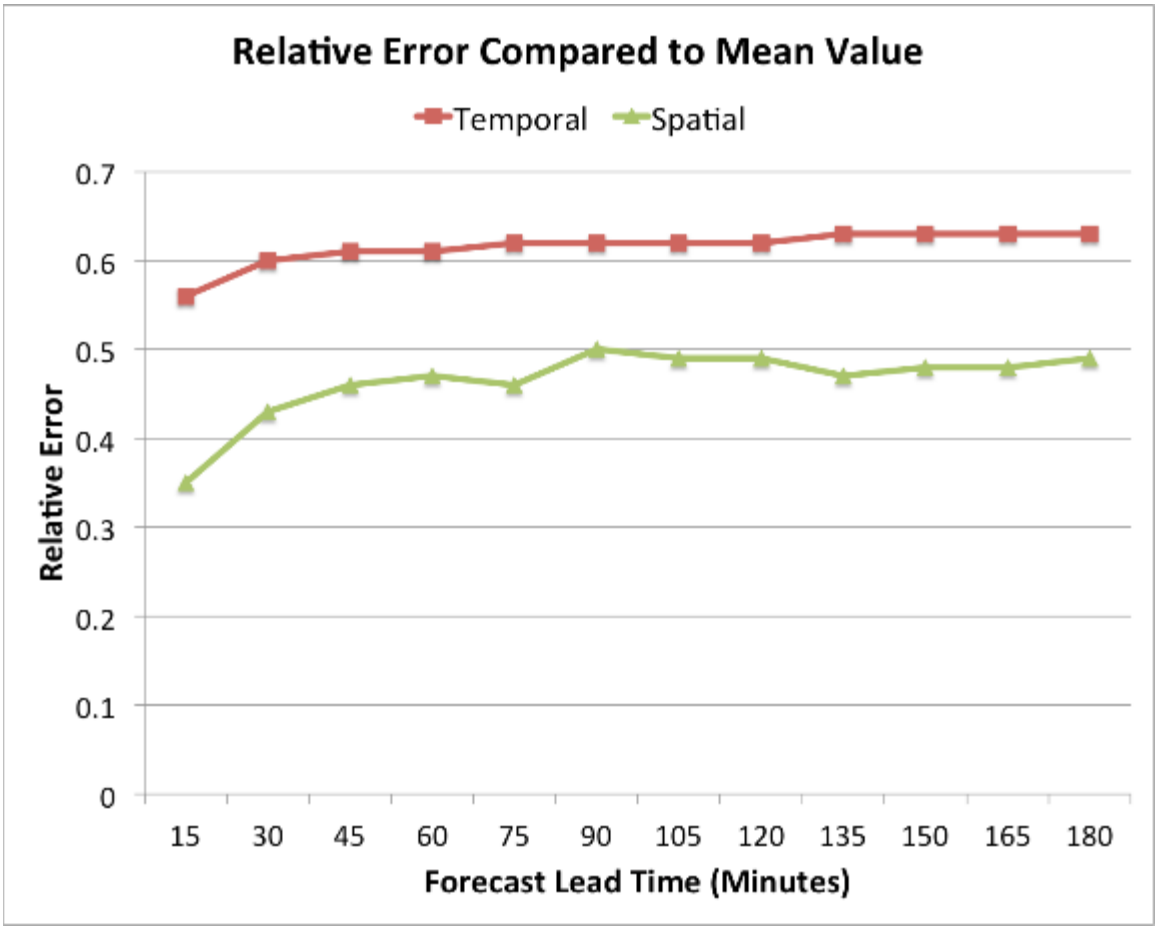


Figure 5-29. Relative error for the model tree on the spatial standard deviation of the GHI (green - triangles) and temporal standard deviation of the GHI (red – squares).

We note that the team also assessed the variability of cloudiness in terms of the short-range predictions made by the Nowcasting version of WRF-Solar™ and reported in section 2.5 based on Jiménez et al. (2016b). In recognition of the stochastic nature of cloudiness as assessed on a grid, that section uses statistical techniques to gauge the effectiveness of the model predictions.

Finally, in recognition of the limits of predictability, this project has sought to quantify the uncertainty using the analog ensemble. Section 4.6 describes that work and shows the uncertainty bounds predicted by our models.

6 SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

6.1 SUMMARY OF SUN4CAST® PROJECT

The DOE-funded Public-Private-Academic Partnership to Advance Solar Power Forecasting project functioned as a collaborative team, with each participant doing their job to provide portions of the Sun4Cast® Solar Power Forecasting System. Chapter 1 of this report laid out the progression of the project, beginning with understanding industry needs and configuring a system to meet those needs based on looking at the problems in terms of a value chain. The end result is a functioning system that is described in chapters 2-4 and thoroughly evaluated as described in chapter 5.

6.1.1 The Nowcast System

The shortest ranges of forecasts must leverage measurements that are available in real-time, typically those from ground-based sensors. The shortest range forecast (0 – 6 hour) is supplied by the NowCast system. The NowCast system consists of several short-range forecasting systems: TSICast, StatCast, CIRACast, MADCast, WRF-Solar-Now, and MAD-WRF. TSICast operates on the shortest time scale, with a latency of only a few minutes and forecasts that currently extend to approximately 15 min. This project facilitated research in improving the hardware and software so that new high definition cameras deployed at multiple nearby locations allow discernment of the clouds at varying levels and advection according to the winds observed at those levels. Pyranometers supply the *in situ* data for initializing the StatCast forecast. During the course of this project, short-range statistical forecasting was advanced by emphasizing regime-dependent forecasting, both implicitly through a regression tree approach, and more explicitly by combining clustering techniques with Artificial Neural Networks. These methods make a substantial improvement (from 15-50%) over short-range persistence forecasts.

A second category of systems employs satellite imagery and uses that information to discern clouds and their motion, allowing the systems to project the clouds, and the resulting blockage of irradiance, in time. The satellite data typically has a latency of 15 – 60 min. This project helped push the NOAA data to a reduced latency while allowing the recovery of higher resolution data. CIRACast was already one of the more advanced cloud motion systems, which is the reason that team was brought to this project. During the project timeframe, the CIRA team advanced cloud shadowing, parallax removal, and implementation of better advecting winds at different altitudes. A second satellite-based system, MADCast, assimilates data from multiple satellite imagers and profilers to incorporate a fully three-dimensional picture of the cloud into the dynamic core of WRF. That allows advection of the clouds via the WRF dynamics directly.

One issue with the observation methods described above is that they do not allow for cloud formation and dissipation, which is in the domain of NWP models. Thus, WRF-Solar™ was adapted for nowcasting, being run at lower resolution more frequently to fill the gap of time where changes in the clouds are most likely. Finally, as the project progressed, it became obvious that combining the advantages of WRF-Solar™ with MADCast, which would assimilate the current cloud observations using the most forefront assimilation techniques, while also allowing for cloud

formation and dissipation. Thus was born MAD-WRF, which was formulated and deployed toward the end of the project.

The Nowcasting system was evaluated systematically as well as with case studies. Both types of evaluations revealed that each component has a "sweet spot" where it is most effective. Thus the blending of the different nowcasting components is an effective method of nowcasting.

6.1.2 WRF-Solar™

Most modern forecasting systems rely on some numerical weather prediction (NWP) model for their base forecasts. Thus, a major emphasis of this project was to improve NWP by developing, testing, evaluating, and improving WRF-Solar™, the first NWP model specifically designed to meet the increasing demand for specialized forecast products for solar energy applications (Jiménez et al. 2016a,b). The WRF-Solar™ upgrades are depicted in Figure 6-1. The first augmentation focused on improving the solar tracking algorithm to account for deviations associated with the eccentricity of the Earth's orbit and the obliquity of the Earth. Because solar energy applications require more frequent calls to the radiation package, inaccuracies in the solar position caused a non-negligible error. Second, WRF-Solar™ added the direct normal irradiance (DNI) and diffuse (DIF) components from the radiation parameterization to the model output in addition to global horizontal irradiance (GHI), parameterizing them (Ruiz-Arias et al. 2010) when needed. Third, efficient parameterizations were implemented to either interpolate the irradiance in between calls to the radiative transfer parameterization, or to use a fast radiative transfer code that avoids computing three-dimensional heating rates but provides the surface irradiance (Xie et al. 2016).

Fourth, a new parameterization was developed to improve the representation of absorption and scattering of radiation by aerosols (aerosol direct effect), including allowing high spatio-temporal variability of aerosols. The treatment of aerosols (Ruiz-Arias et al. 2014) allows for the ingestion of aerosol optical properties with time stamps in order to accurately model the temporal variations in aerosol loading, permitting the ingested aerosol concentration to represent the aerosol optical properties in WRF-Solar™.

A fifth advance was that the aerosols interact with the cloud microphysics, altering the cloud evolution and radiative properties (aerosol indirect effect). This effect has been traditionally only implemented in atmospheric chemistry models, which are significantly more computationally expensive than NWP models without detailed chemistry. WRF-Solar™ uses a simplified treatment of the aerosols (only two aerosol species are allowed) that accounts for changes in the size of cloud hydrometeors in order to represent this aerosol indirect effect (Thompson and Eidhammer 2014) with minimal increase in computational cost (~16%). The aerosols are advected by the model dynamics and the parameterization is linked to the WRF-Solar™ aerosol parameterization in order to have a fully coupled representation of the cloud-aerosol-radiation system.

A sixth development accounts for the feedbacks that sub-grid scale clouds produce in shortwave irradiance as implemented in a shallow cumulus parameterization (Deng et al. 2003, 2014). The scheme includes predictive equations for the sub-grid scale cloud water/ice content and the cloud fraction.

Finally, when coupling WRF-Solar™ with elements of MADCast, MAD-WRF allows assimilation of infrared irradiances from satellites to determine the three-dimensional cloud field (Auligné 2014a,b), resulting in an improved initialization of the cloud field that further increases the performance of short-range forecasts.

These enhancements were made in collaboration with Penn State and NREL. NCAR responded to numerous requests to use beta versions of WRF-Solar™. The community sees it as a way to advance deployment of solar energy by enabling better forecasting of the irradiance resource. NCAR expects to further exercise and improve the WRF-Solar™ in new locales in future projects.

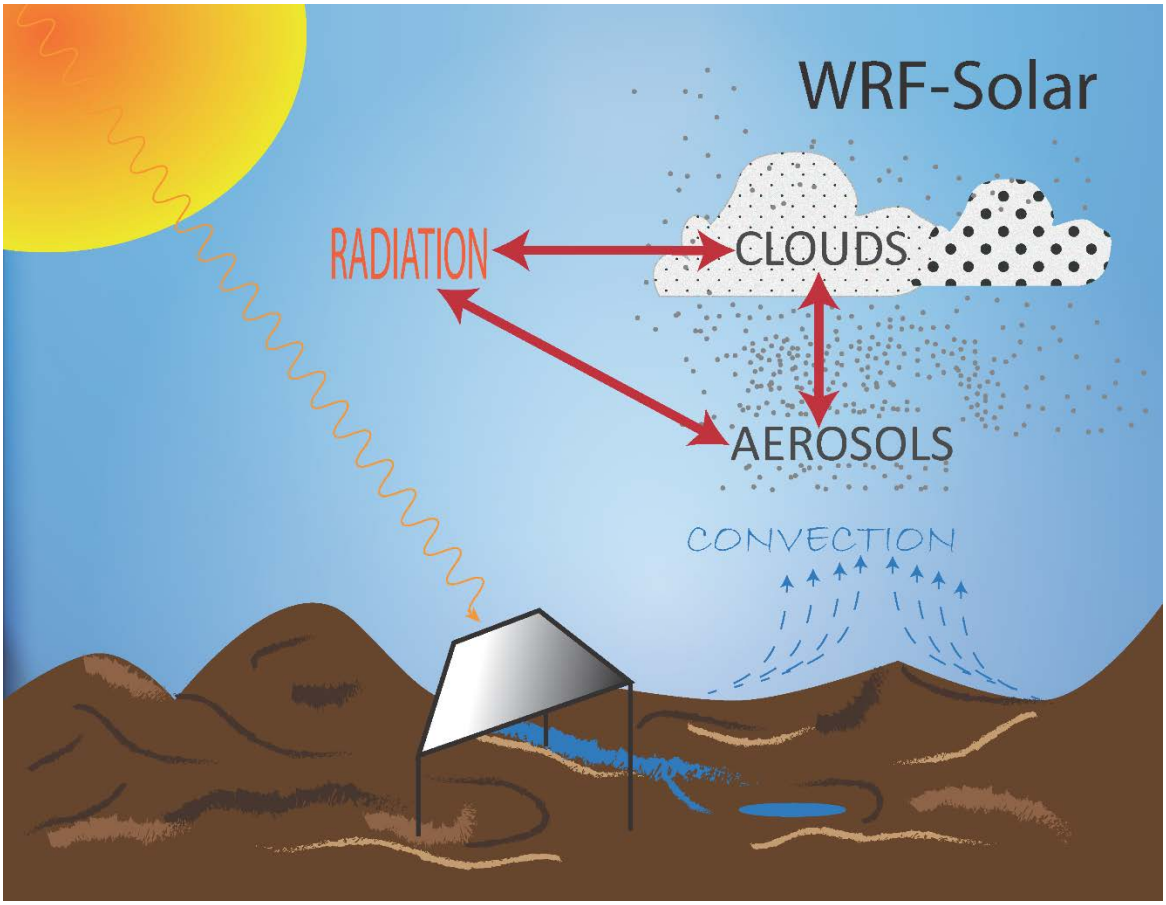


Figure 6-1: Diagram showing the WRF-Solar™ augmentations that now include specific interactions between the radiation, clouds, and aerosols. The radiation scheme was augmented with an improved solar tracking algorithm and explicitly calculates all irradiance components with a capability for high-frequency output of GHI, DNI, and DIF. It additionally allows time variation in the irradiance components between function calls. The aerosols are now allowed to vary in time, as well as to be advected by the winds, in order to allow feedback to the radiation scheme. The aerosols additionally become the microphysics species that determine the cloud condensation nuclei for cloud formation. Clouds in turn scavenge the aerosols. A new shallow convection scheme enhances the capability for the cloud feedback to the radiation. Both aerosols and clouds fully feed back to determining the scattering, absorption, and emission in the radiation scheme.

6.1.3 The Engineered System

Building the individual component models is necessary, but not sufficient, to supplying a high-quality solar power forecast. It is also critical to smartly engineer a system that smoothly handles

data input and output and effectively blends the results of each of the components. This engineered system must allow for missing observations or model results as well as allowing for "graceful degradation" when not all systems are performing optimally. The engineered system was described in detail in chapter 4.

The Nowcasting Expert System Integrator (NESI) uses recent performance information to smartly blend the NowCast components. Although this is currently accomplished using historical statistics, there is evidence that moving to a dynamically blended system in the future could prove advantageous.

The Dynamical Integrated foreCast (DICast®) system smartly blends the NWP models, both our own WRF-Solar™ output as well as publicly available models. Although DICast® has shown a high degree of accuracy for other forecast variables, this project was the first time that it was employed for irradiance forecasts. Development during the project included building algorithms to account for disparate model time frames and consideration of solar angle in blending the model output correctly. The DICast® and NowCast systems must in turn be blended during the overlap periods.

The blended system forecast is in terms of irradiance. Thus it is necessary to convert irradiance to power. This was accomplished by using a Cubist regression tree model that was built on historical irradiance and power observations. One advantage to this empirical approach to power conversion is that it is easier to train with disparate data, as long as the data are consistent. Thus, the training/testing process is equivalent whether the irradiance observations are in terms of GHI, DNI, or plane-of-array (POA).

The last step in the forecast process is the Analog Ensemble (AnEn), which both corrects the forecast and provides probabilistic information to quantify the uncertainty of the forecast. Again, this project was a first opportunity to exercise AnEn for solar power and it performed admirably.

The project additionally leveraged cost-share funds to develop a gridded irradiance forecasting system, the Gridded Atmospheric Forecast System (GRAFS-Solar). GRAFS can be used as a testbed for new techniques or as an operational system that could feed distributed solar forecasts.

6.1.4 System Performance

The system has been fully assessed. The component systems were each thoroughly vetted by their developers during the iterative improvement stage. Those results are blended into the individual discussions of each system. In addition, a formal assessment was accomplished by the NCAR metrics team, who employed the full range of metrics developed with the SunShot, NOAA, and IBM teams during the projects's first year. Those metrics supplied a more complete snapshot of the system than available with traditional MAE and RMSE.

From these metrics, we confirmed that each of the NowCast components has a "sweet spot" where its performance provides value, with the blended system providing a significant improvement over smart persistence. The Day-Ahead system also shows substantial improvement, especially with the inclusion of WRF-Solar™. Improvements range from 22-42% for WRF-Solar™ and 13-24% additional for the blended system.

In addition, an econometric analysis was accomplished based on production cost modeling in collaboration with Xcel Energy. The methodology and intermediate results are described in section 5.3. We found that when scaled to a national level, an estimated savings discounted over a 26-year period, a savings of $455M could be achieved with the sort of improvement in solar power forecasting documented as part of this project.

6.2 Scientific Lessons Learned

The project team was comprised of some of the most-experienced scientists and engineers who predict weather for societal applications. The project funding provided this team an opportunity to advance solar power forecasting. As described above, much was accomplished. But as always happens with any real-world applied projects, challenges were also encountered, which led to unexpected lessons learned.

Some scientific challenges and lessons learned from them include:

- The results of the project confirmed the value of blending multiple types of forecasts by multiple models. The whole produces a much better forecast than any single part. Each system module adds value, but not necessarily equally.
- Blending in the NowCast system showed value. Although this system used the expert system approach to statistically determining weights to use in blending as each component model improved over the course of the project, an automated dynamic blending method would be preferable once the component models are stable.
- As shown in the WRF-Solar™ work, the source of aerosol data is important. Chapter 3 results indicate that real-time forecasts of aerosols produce a much better forecast than simply employing climatology.
- When using sky imagers, the higher resolution equipment substantially helped to improve the forecasts. Using multiple cameras also led to forecast improvements.
- The user-defined metrics were meaningful and a more thorough assessment is available with a broader set of metrics that were defined and used in this project.
- There are limits to predictability due to the chaotic nature of atmospheric flow and sensitivity to initial conditions. Although we have approached these limits in some time frames, we have not yet hit them. Further assessment of these limits would be helpful.
- The availability and quality of data is a critical issue for any forecasting system. Some specific issues include:
    - The quality of the data and the metadata often did not meet our expectations. It was sometimes difficult to discern specific information, such as averaging period and valid time. When these times are not consistent, the forecast will necessarily be degraded. Issues with time zones and standard vs. daylight saving time were not treated consistently in data collection.

- Each utility has its own format for recording measurements of meteorological variables (including irradiance) and power. This fact necessitates customizing data ingest systems for each utility partner. The lesson is that standardized data formatting would greatly benefit all who deal with such data.
- Each utility records a different type of irradiance measurement. Some use GHI, while others use POA, or even DNI for concentrated systems. This fact makes it more difficult to design and build a generally applicable system; instead a system must be adaptable and tailored to ingest and use these different types of irradiances.
- Historical data were often unavailable. Note that statistical learning and artificial intelligence methods require historical data for training the system, so where it does not exist, those techniques cannot be employed.

6.3 Project Management Challenges and Lessons Learned

This was a relatively large project with a substantial budget, but also included a large number of participants. Thus, we encountered some management challenges from which we learned. Organizing and keeping that many partners on track to reach a common goal was at times difficult. It was quite beneficial to include DOE sponsored national laboratories in the project. Their budgets were provided directly to them by SunShot. This was advantageous in that it saved the project from incurring additional overhead charges to manage those laboratories, but at the same time, it was more difficult to include them fully as part of the integrated team. Both laboratories performed admirably and worked closely with the NCAR team to accomplish goals. It was additionally difficult to manage the details of accomplishments across cost share partners, who again were quite beneficial to the project, but they each had their own goals rather than always sharing project goals.

The most beneficial way for integrating the project team was holding the project workshops at NCAR in 2013 and 2014. It is unfortunate that the final workshop was cancelled by SunShot. We would highly support SunShot in configuring and organizing a community partnership around solar power forecasting in the future.

Although we had various major challenges in accomplishing this project, the overarching problem that made the project difficult was a three- to four-month lag in contracting associated with the three contract renewal periods during the course of the project. During each of these periods NCAR and our subcontractors were not authorized to spend funds. This resulted in three periods of STOP WORK that had a large negative impact on the project. These project hiatuses resulted in:

- Loss of personnel who had to be assigned to other projects for the interim, making it difficult to obtain their time when the project resumed. In one case, a talented software engineer on the project could not deal with the uncertainty of the project funding and left NCAR for another company.
- When the project resumed and we recovered the personnel needed for the work, there was a start-up time where everyone had to re-familiarize themselves with details of the work they were accomplishing.

- The quality of the end product is degraded due to the team not being able to iteratively improve the individual components and the blended system during some of the key periods.
- We lost the interest of the industry partners when there were 3-4 month long periods with no new work. Although the system was robust enough to continue to make forecasts, we were not able to analyze and communicate with the end users during those periods, causing a loss of continuity that damaged the working partnership.
- It was difficult to continue to pay personnel. This was particularly problematic for the graduate students involved, who were working on a single project and were relying on continued funding to complete their degree. This makes it difficult to include graduate students in a project that is likely to have breaks during project renewal.
- The team was asked to rewrite the Statement of Project Objectives (SOPO) repeatedly and to write in target metric goals that are impossible to achieve in an atmospheric system. Thus, we added a tiger team to study and quantify the limits of predictability to describe to the SunShot management the impact of the chaotic nature of atmospheric flow (and cloud prediction in particular) on predictability of solar irradiance. Changes in project management personnel at SunShot resulted in these explanations occurring repeatedly.
- Requiring an additional evaluation in the final year that was not part of the original plan was particularly problematic. That was the year of quasi-operational continual forecasting when it was most important to have continuity. Fortunately, most systems continued to operate in the background to collect data and provide forecasts. But since personnel were reassigned, the systems were not actively monitored during the periods with stop work and the systems were not continuously improved during this period. In addition, funds were diverted to prepare and present the continuation review, taking funds away from the science that was needed during that period. This additional review was particularly counterproductive, resulting in a stop work in the middle of the operational period, and substantially affecting fine-tuning, and thus, the final quality of some system components.
- One of our utility partners shared, “It went well in the first phases; however, as the stops and starts began to pile up it seemed a bit disconnected on what was going on. The face-to-face meetings were the best part of the project as people with actual questions could talk and exchange ideas and I think some real discoveries happened for some. I wish we could have been more engaged in the forecast analysis and been able to assist in the fine-tuning and model learning that needs to happen for a forecast to improve. … We could really not get off the ground on receiving and analyzing the forecasts. The starts and stops put a large damper on our ability to stay focused on the forecasts and in the end I had to release folks to other activities.”

6.4 Networks Fostered

This project was built on leveraging a network of experts in all aspects of solar power forecasting to develop, build, deploy, assess, and test the Sun4Cast® System in a wide variety of locations and climatologies. Thus, we brought together university and laboratory researchers, software engineers skilled in Big Data issues, utility and ISO personnel, and forecast providers to assess and help us to iteratively improve the system. As described in the introduction (Chapter 1), the

kick-off workshop and second year workshop were effective in cementing the collaborations and allowing time for deep exchange of ideas. The specific partners that were formally part of the project and their roles include:

National Laboratories

- National Center for Atmospheric Research – lead laboratory for project
- Brookhaven National Laboratory – built TSICast and provided LISF data
- National Renewable Energy Laboratory – built FARMS radiative transfer algorithm and collaborated on statistics
- National Oceanographic and Atmospheric Administration – funded separately – collaborations on observations, satellite data, and modeling advancements

Universities

- The Pennsylvania State University – built shallow convection scheme, collaborated on StatCast versions, built new observation capability
- Colorado State University – built, tested, and deployed CIRACast
- University of Washington – statistical assessment and collaborations on StatCast
- University of Hawaii – forecasting in Hawaii
- University of Buffalo – sky imager research
- Stony Brook University – subcontractor to BNL – helped build TSICast

Utilities

- Xcel Energy Services
- Sacramento Municipal Utility District
- Long Island Power Authority
- New York Power Authority
- Hawaiian Electric Company
- Southern California Edison

ISOs

- New York Independent System Operators
- California ISO
- Hawaiian Electric Company
- Public Service Company of Colorado

Forecast Providers and a Solar Contractor

- Solar Consulting Services
- Atmospheric and Environmental Research
- Global Weather Corporation
- MDA Information Systems
- Schneider Electric/Telvent DTN

In addition, we added several affiliate partners who strengthened our partnership:

Affiliate Partners

- University of Buffalo, Tarun Singh, Professor of Mechanical Engineering, working with sky imagers for campus array.
- Army Research Laboratory, White Sands, Gail Vaucher (primary contact), solar array on-site and doing related research.
- University of Jaén, Spain, José Ruiz-Arias, solar power forecasting.
- Commonwealth Scientific and Industrial Research Organization (CSIRO) of Australia, Alberto Troccoli, building forecasting system.

As we communicated with our partners, they have all emphasized the large importance of solar power forecasting. In most regions, solar power has not yet reached the percentage of capacity that necessitates full forecasting systems. In some areas, however, it has and the utilities and ISOs rely on forecasting. Others, who are not at that capacity yet, see the need coming in the future or have learned from their experience with higher penetrations of wind power.

Finally, as we advanced our system and talked about it at conferences and workshops, a number of people have requested beta-versions of portions of our systems. Most of those related to WRF-Solar™. The WRF-Solar™ users and scientists who we have specifically communicated with regarding its configuration and use, as well as some of their comments, include:

Supplied Code and Configuration Information

- Zack, John: john@meso.com Meso / AWS Truepower March 2015. NCAR provided WRF-Solar™ code. AWS Truepower applied this beta-version of WRF-Solar™ over Hawaii, with nested grids at 18, 6, and 2 km over the Hawaiian Islands. Because they run an ensemble of models, this allowed a direct comparison with these other model configurations. On request, they were able to supply an initial subjective analysis of this comparison. Their impression was that WRF Solar performed better than the other models in some situations when the evolution of the cloud patterns exhibited high predictability among all the models. However, in situations when the overall predictability seemed low, WRF Solar did not demonstrate discernable improved skill relative to the other models. Their interpretation of this is that in the low predictability cases, the use of better physics doesn't improve the forecasts because the uncertainty (sensitivity) is most likely in the initialization process and not in the physics. However, in cases where the sensitivity to the initialization is low, the importance of the physics formulation is increased and in those cases, they report that WRF-Solar™ seemed to exhibit an advantage. Note that they did not have the very recent version that adds in the MADCast satellite data assimilation. It would be interesting for NCAR and AWS Truepower to team on such an analysis over Hawaii.
- Xiaohui Zhong: x6zhong@eng.ucsd.edu San Diego University August 2015. Xiaohui visited NCAR over a several month period in 2015 to work more closely with NCAR WRF researchers. They are using the code, primarily over Southern California for stratocumulus cases and expect to provide more detailed feedback.
- Charlotte Hoppe: ch@eurad.uni-koeln.de Institute of Energy and Climate Research (Germany) September 2015. We provided the WRF-Solar™ code. She used the WRF-Solar™ model to calculate GHI, DNI and DHI for a case study in Southern Spain. They

used AOD and SSA from their chemistry transport model, EURAD-IM, as input to WRF-Solar™. Here, she liked the infrastructure of WRF-Solar™, so that the coupling was relatively easy. They first looked at cloud-free days and have obtained good results. A colleague will present a poster about this work at EGU: http://meetingorganizer.copernicus.org/EGU2016/EGU2016-5691.pdf.
Next, they plan to analyze also cloudy situations and have requested advise on model application.

- Mavromatakis, Fotis: fotis@physics.uoc.gr T.E.I of Crete (Greece) November 2015. We helped him to configure the model and answered many other questions. He is actively using the WRF-Solar™ model. He is new to WRF and is running the model over Crete. He is running WRF-Solar™ and comparing against observations they have. He sent plots of simulations versus observations. He is actively sending feedback with comments/questions on the use of WRF-Solar™. He provided feedback of the model performance.
- Eastman, Joe: joe.eastman@windlogics.com Wind logics December 2015. Helped with options in WRF-Solar™ in the official WRF release.
- Tewari, Mukul mukultewari1@gmail.com IBM January 2016. Actively using the model and helped him with many questions. He provided feedback on the model performance. This group is also actively using WRF-Solar™ and comparing its performance against observations, including sharing his experience with the code. Mukul will likely send more feedback regarding the model performance and to get additional comments/suggestions from our side.
- Mouhamet, Diallo: mouhamet.diallo@ird.fr French Guiana University host at IRD Cayenne February 2016. IRD is the Institute of Research for Development. We provided the code and we did not hear back from him.
- Gordon Huva, Robert: serrgh@nus.edu.sg Solar Energy Research Institute of Singapore (SERIS) April 2016. We provided the code.
- Stevens, Duane: dstevens@hawaii.edu University of Hawaii March 2015. We provided the code and we have not heard back from him.
- Schroedter, Marion: marion.schroedter-homscheidt@dlr.de DLR (Germany) April 2015. We provided the code and we did not hear back from her.
- Michael Palmer: michael.palmer@greenpowerlabs.com Green Power Labs. July 2015. We provided the code and we did not hear back from him.
- José A. Ruiz-Arias jararias@ujaen.es Universidad de Jaén (Spain) July 2015. José's collaborators are probably using the code.

Communications Regarding Aspects of WRF-Solar™:

- Hahmann, Andrea N.: ahah@dtu.dk Technical University of Denmark (DTU) Jan2016 - Brief inquiry about options used in WRF-Solar™ in the BAMS paper.
- James, Eric: james@noaa.gov NOAA Sporadic communication regarding WRF-Solar™ augmentations in the official WRF release.
- Navarro, Jorge: jorge.navarro@ciemat.es CIEMAT (Spain) Sporadic communication regarding WRF-Solar™ augmentations in the official WRF release.
- Many other researchers with inquires about WRF-Solar™ after presentations at conferences.

Finally, we have worked with professional organizations to organize special sessions at conferences and workshops around solar power forecasting. Some of these sessions include:

- Organized and chaired: Joint Session between AMS 15th Conference on Artificial Intelligence and its Applications to Environmental Science and Eighth Conference on Weather, Climate, and the New Energy Economy, AMS Annual Meeting, January 2016.
- Several Team Members on Local Organizing Committee for International Conference on Energy and Meteorology Conference and Local Organizing Committee, including chairing renewable energy sessions, Boulder, CO, June 2015.
- Organized and moderated: Utility Variable Integration Group Tutorial on Stochastic Forecasting Methods and Applications, 16 February 2015.
- Organized and chaired: Joint Session between AMS 13th Conference on Artificial Intelligence and its Applications to Environmental Science and Sixth Conference on Weather, Climate, and the New Energy Economy, AMS Annual Meeting, January 2014.
- Organized and chaired: Joint Session between the AMS 12th Conference on Artificial Intelligence and its Applications to Environmental Science and Fifth Conference on Weather, Climate, and the New Energy Economy, January 2013.
- Panel Moderator: Challenges and Opportunities in Applying AI Techniques to Environmental Science Problems, AMS Annual Meeting, Austin, TX, January 2013.
- Organizing Committee for International Conference on Energy and Meteorology Conference, Toulouse, France, June 2013. Organized Solar Energy session.

6.5 Recommendations for Best-Practice Solar Power Forecasting

A major goal of this project was to draw conclusions about the performance of each component system and make recommendations for best practices in configuring solar power forecasting systems. Some specific recommendations include:

- It is best to blend various component models or systems together. The forecast from blended models/systems is invariably significantly better than those produced by a single model or approach.
- Use a base NWP model tuned for the purpose. We found very significant improvements in forecasting by employing WRF-Solar™.
- Including multiple NWP models improve the blended forecast for time scales from 3 h through the day-ahead forecast and beyond.
- It is possible to improve upon persistence, even at the very short-range by using methods trained on targeted *in situ* observations. StatCast trained to employ pyranometer data was better than persistence, even at short time scales (15 min to 3 h) and TSICast, which uses multiple sky imagers improved upon persistence in the time range less than 15 min.

- Satellite based cloud advection is useful, but tricky. For regions near the mountains or along coasts, it is necessary to include some model physics to account for stationary clouds as well as cloud formation and dissipation. It is important to include the improvements related to correcting for shadowing and parallax as accomplished by CIRA.
- NWP can be combined with satellite data via assimilation to produce a fast-running, short-range (0-6 h) forecast that is helpful for nowcasting (see MAD-WRF, section 2.4). This produced the best forecast on the 1-6-h time scale.
- The analog ensemble approach is helpful for both improving on the deterministic blended forecast as well as for producing a probabilistic prediction that is well calibrated.
- An empirical power conversion method is amenable to training site-specific information, even when missing metadata. Artificial intelligence techniques are capable of predicting directly from an observation to a target value as long as historic training data is available.
- An enhanced series of metrics is helpful for evaluating and tuning individual models as well as the entire system.

Finally, we asked our utility and ISO partners where they see solar power forecasting going in the future and some comments include:

- "…the industry need is still there and it will only get larger as more distributed energy is connected to the grid."
- [Forecasts will be from] "centralized RTO/ISO/BA generated forecasts that will have multiple uses and at varying granularities."

6.6 Final Conclusions, Impact, and Path Forward

This project served to advance the state-of-the-science of solar power forecasting as originally planned. The team that worked on the project included some of the best-known researchers in the field. The team worked extremely synergistically and produced demonstrably better models than existed previously, blended the models to produce improved consensus forecasts, and used forefront post-processing methods to further correct the models as well as convert the irradiance forecasts to power and provide probabilistic forecast information to the utility and ISO partners. Although challenges were encountered, the team rose above those and completed the project admirably. The Sun4Cast® system and its component models were thoroughly assessed using a full range of metrics, some of which were specifically derived for this project. An economic evaluation estimated saving in one particular service region, which was then scaled up to estimate a substantial potential for savings across the US as more solar power is deployed in the future.

Thus, as the capacity of solar power grows, solar power forecasting with systems like Sun4Cast® will provide enabling technologies that will make the economics more feasible, empowering more solar power deployment. Such enhanced deployment has the potential to improve air quality, mitigate climate change, improve energy security, and provide enhanced employment opportunities throughout the renewable energy sector.

The team members have all grown in their research capabilities in solar energy and the collaborative research is expected to continue. A direct point of continuity is continued collaboration among the partners. For instance, the Sun4Cast® system is fully deployed operationally for Xcel Energy's Public Service Company of Colorado (PSCo) for their commercial solar plants, delivered by Global Weather Corporation (GWC). Xcel, PSCo, and GWC are all partners in this project. MDA Federal reports that they are assuming portions of this project's systems as part of their forecasting system. As reported above, WRF-Solar™ has been widely tested and many of its component modules are already part of the public WRF release and being used too widely to document. In addition, NCAR plans to continue to advance the Sun4Cast® system in regions throughout the world with wide partnerships. A specific partnership that is in the process of starting is between NCAR, BNL, and the Electric Power Research Institute (EPRI). That project, led by EPRI, will deploy sky imagers over New York City and provide forecasts using Sun4Cast® systems to utilities in New York. We expect additional collaborations to continue.

In summary, this project not only advanced the state-of-the-science through cutting edge research, but on a grander scale, it enabled a host of partnerships that are in the process improving the economics of solar energy and advancing its deployment.

7 REFERENCES

Albrecht, B., 1989: Aerosols, cloud microphysics, and fractional cloudiness. *Science*, **245**, 1227–1230.

Alessandrini, S., L. Delle Monache, S. Sperati, and J. N. Nissen, 2015a. A novel application of an analog ensemble for short-term wind power forecasting. *Renew. Ener.*, **76**, 768-781. doi:10.1016/j.renene.2014.11.061.

Alessandrini, S., L. Delle Monache, S. Sperati, and G. Cervone, 2015b: An analog ensemble for short-term probabilistic solar power forecast. *Appl. Ener.*, **157,** 95-110. doi:10.1016/j.apenergy.2015.08.011.

Allmen, M.C., and W.P. Kegelmeyer, Jr., 1996 The computation of cloud-base height from paired whole-sky imaging cameras. Journal of Atmospheric and Oceanic Technology, 13(1):97–113, 1996.

Almonacid, F., Pérez-Higueras, P.J., Fernández, E.F., and Hontoria, L., 2014: A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator. *Energy Conversion and Management*, **85**, 389-398.

Anderson, J.L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, **9**, 1518–1530.

Antón, M., J.E. Gil, A. Cazorla, J. Fernández-Gálvez, I. Foyo-Moreno, F.J. Olmo and L. Alados-Arboledas, 2011: Short-term variability of experimental ultraviolet and total solar irradiance in Southeastern Spain. Elsevier: *Atmospheric Environment*. **45**: 4815-4821. doi: 10.1016/j.atmosenv2011.06.020.

Augustine, J.A., J.J. DeLuisi, and C.N. Long, 2000: SURFRAD: A national surface radiation budget network for atmospheric research. *Bull. of the Amer. Met. Soc.*, **81**, 2341–2357.

Augustine, J.A., G.B. Hodges, C.R. Cornwall, J.J. Michalsky, and C.I. Medina, 2005: An update on SURFRAD- The GCOS surface radiation budget network for the continental United States. *J. Atmos. Oceanic Technol.*, **22**, 1460–1472.

Auligné, T., 2014a: Multivariate minimum residual method for cloud retrieval. Part I: Theoretical aspects and simulated observations experiments. *Mon. Wea. Rev.*, **142**, 4383-4398. doi: 10.1175/MWR-D-13-00172.1.

Auligné, T., 2014b: Multivariate minimum residual method for cloud retrieval. Part II: real observations experiments. *Mon. Wea. Rev.*, **142**, 4399-4415. doi: 10.1175/MWR-D-13-00173.1.

Barbaro, E., 2015: Interaction between aerosols and boundary-layer dynamics. Ph.D. thesis, Wageningen University, Netherlands.

Bartholomy, O., T. Vargas, M. Simone, C. Hansen, S. Fitchett, and A. Pohl, 2014: Benchmarking Solar Power and Irradiance Forecasting Accuracy at Sacramento Municipal Utility District. IEEE Photovoltaic Specialist Conference, June 8-13, 63-68.

Benjamin, S.G., S.S. Weygandt, J.M. Brown, M. Hu, C.R. Alexander, T.G. Smirnova, J.B. Olson, E.P. James, D.C. Dowell, G.A. Grell, H. Lin, S.E. Peckham, T.L. Smith, W.R. Moninger, J.S. Kenyon, and G.S. Manikin, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144,** 1669-1694. doi:10.1175/MWR-D-15-0242.1.

Bernecker, D., C. Riess, E. Angelopoulou, and J. Hornegger, 2012: Towards improving solar irradiance forecasts with methods from computer vision. *Pattern Recognition*, 2012.

Bernecker, D., C. Riess, E. Angelopoulou, and J. Hornegger, 2014: Continuous short-term irradiance forecasts using sky images. *Solar Energy*, 110:303–315, 2014.

Berner, J., G. Shutts, M. Leutbecher, and T. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. J. Atmos. Sci., 66, 603–626.

Beyer, H.G., Costanzo, C., and D. Heinemann, 1996: Modifications of the Heliosat procedure for irradiance estimates from satellite data. *Solar Energy*, **56**, 121–207.

Bhardwaj, S., Sharma, V., Srivastava, S., Sastry, O.S., Bandyopadhyay, B., Chandel, S.S., and J.R.P. Gupta, 2013: Estimation of solar radiation using a combination of Hidden Markov Model and generalized Fuzzy model. *Solar Energy*, **93**, 43-54.

Bilionis, I., E. M. Constantinescu, and M. Anitescu, 2014: Data-driven model for solar irradiation based on satellite observations. *Solar Energy*, **110**, 22-38.

Black, M.J., and P. Anandan, 1996: The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. Computer vision and image understanding, **63**(1):75–104, 1996.

BNOC, 2014: Upgrades to the Operational Gridded OCF System. BNOC Operations Bulletin Number 103, 23 pp. Available online at: http://www.bom.gov.au/australia/charts/bulletins/apob103.pdf.

Bosch, J. L., and J. Kleissl, 2013: Cloud motion vectors from a network of ground sensors in a solar power plant. *Sol. Ener.*, **95,** 13-20. doi:10.1016/j.solener.2013.05.027.

Boser, B.E., I.M. Guyon, and V.N. Vapnik, 1992: A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory, pages 144–152. ACM, 1992.

Bouzerdoum, M., A. Mellit, and A. Pavan, 2013: A hybrid model (SARIMA-SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. *Solar Energy*, **98**, 226-235.

Bremnes, J. B., 2006: A comparison of a few statistical models for making quantile wind power forecasts. *Wind Energy*, **9**(1-2), 3-11.

Brown, B.G., E. Gilleland, and E. Ebert, 2012: Forecasts of spatial fields. In Forecast verification: A practitioner's guide in atmospheric sciences, I.T. Jolliffe and D.B. Stephenson, Editors. Wiley.

Brox, T. and J. Malik, 2011: Large displacement optical flow: descriptor matching in variational motion estimation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, **33**(3):500–513, 2011.

Buchard, V., A.M. da Silva, P. Colarco, N. Krotkov, R. R. Dickerson, J. W. Stehr, G. Mount, E. Spinei, H. L. Arkinson and H. He, 2014: Evaluation of GEOS-5 sulfur dioxide simulations during the Frostburg, MD 2010 field campaign. *Atmos. Chem. Phys*., **14**, 1979-1941.

Buchard, V., A. M. da Silva, P. R. Colarco, A. Darmenov, C. A. Randles, R. Govindaraju, O. Torres, J. Campbell, and R. Spurr, 2015: Using the OMI aerosol index and absorption aerosol optical depth to evaluate the NASA MERRA Aerosol Reanalysis. *Atmos. Chem. Phys.*, **15,** 5743-5760. doi: 10.5194/acp-15-5743-2015.

Buchard, V., A.M. da Silva, C. A. Randles, P. Colarco, R. Ferrare, J. Hair, C. Hostetler, J. Tackett, and D. Winker, 2016: Evaluation of the surface PM2.5 in Version 1 of the NASA MERRA Aerosol Reanalysis over the United States. *Atmos. Env.*, **125,** 100-111. doi: 10.1016/j.atmosenv.2015.11.004.

Buizza, R., D.S. Richardson, and T.N. Palmer, 2003: Benefits of increased resolution in the ECMWF ensemble system and comparison with poor-man's ensembles. *Quarter. J. Roy. Met. Soc.*, 5889, 1269–1288.

Bunimovich, L.A., 2014: Short- and Long-term Forecast for Chaotic and Random Systems (50 years after Lorenz's Paper), *Nonlinearity*, **27**, R51-R60.

Cebecauer, T., R. Perez, and M. Suri, 2011a: Comparing performance of SolarGIS and SUNY satellite models using monthly and daily aerosol data. Proc. ISES Conf., Kassel, Germany, International Solar Energy Soc.

Cebecauer, T., M. Suri, and C.A. Gueymard, 2011b: Uncertainty sources in satellite-derived direct normal irradiance: how can prediction accuracy be improved globally? Proc. SolarPACES Conf., Granada, Spain.

Chang, C-C., and C-J. Lin, 2011: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Chapman, M.B., S. Drobot, S. Linden, J. Cowie, W.P. Mahoney, 2010: A Decision–Support System for Winter Maintenance of Roads, Bridges, and Runways. *15th Standing International Road Weather Commission Conference*, Quebec City, Canada, 5–7 February.

Chen, J., J. Tian, N. Lee, J. Zheng, R.T. Smith, and A.F. Laine, 2010: A partial intensity invariant feature descriptor for multimodal retinal image registration. *Biomedical Engineering, IEEE Transactions* on, **57**(7):1707–1718, 2010.

Chen, Z., H. Jin, Z. Lin, S. Cohen, and Y. Wu, 2013: Large displacement optical flow from nearest neighbor fields. In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pages 2443–2450. IEEE, 2013.

Chin, M., D. J. Jacob, G. M. Gardner, P. A. Spiro, M. Foreman-Fowler, and D. L. Savoie, 1996: A global three-dimensional model of tropospheric sulfate. *J. Geophys. Res*., **101**, 18 667–18 690.

Chou, M.-D., and M. J. Suarez, 1999: A solar radiation parameterization for atmospheric studies. Tech. Rep. NASA/TM-1999-104606, **15**, NASA, Greenbelt, Maryland, USA, 38 pp.

Chou,M.-D., 1992: A solar radiation model for use in climate studies. J. *Atmos. Sci*., **49**, 762–772.

Chow, C.W., B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields, and Coauthors, 2011: Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Solar Energy*, **85**, 2881–2893. doi:10.1016/j.solener.2011.08.025.

Chow, C.W., S. Belongie, and J. Kleissl, 2015: Cloud motion and stability estimation for intra-hour solar forecasting. *Solar Energy*, **115**, 645-655.

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

Colarco, P., A. da Silva, M. Chin, and T. Diehl, 2010: Online simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based aerosol optical depth. *J. Geophys. Res*., **115**, D14 207.

Cornaro, C., F. Bucci, M. Pierro, F. Del Frate, S. Peronaci, and A. Taravat, 2013: Solar Radiation Forecast Using Neural Networks for the Prediction of Grid Connected PV Plants Energy Production (DSP Project). Proceedings of 28th European Photovoltaic Solar Energy Conference and Exhibition, Sept 30 - Oct 4, 3992 – 3999.

Corpetti, T., P. Heas, E. Me´min, and N. Papadakis, 2008: Variational pressure image assimilation for atmospheric motion estimation. In Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, volume 2, pages II–505. IEEE, 2008.

Cote, S. and A. Tatnall, 1995: A neural network-based method for tracking features from satellite sensor images. *International Journal of Remote Sensing*, **16**(18):3695–3701, 1995.

Craven, J. P, J. Wiedenfeld, J. Gagan, P. Browning, A. Just, and C. Greif, 2013: The NWS Central Region Extended Forecast Process. Extended Abstract, 38th Natl. Wea. Assoc. Annual Meeting, Charleston, SC, P2.38. Available online at http://www.nwas.org/meetings/nwa2013/extendedabstracts/NWA2013_P2.38_Craven_etal.pdf.

Cros, S., O. Liandrat, N. Sébastien, and N. Schmutz, 2014: Extracting cloud motion vectors from satellite images for solar power forecasting. Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, 4123-4126.

Curtright, A. E., K. Apt, 2008. The character of power output from utility-scale photovoltaic systems. *Progress in Photovoltaics: Research and Applications* **16**, 241-247.

Davis, C.A., B.G. Brown, R.G. Bullock and J. Halley Gotway, 2009: The Method for Object-based Diagnostic Evaluation (MODE) Applied to Numerical Forecasts from the 2005 NSSL/SPC Spring Program. *Weather and Forecasting*, **24**, 1252-1267.

Davy, R. J., and A. Troccoli, 2012: Interannual variability of solar energy generation in Australia. *Sol. Ener.*, **86,** 3554-3560. doi:10.1016/j.solener.2011.12.004.

Dee D. P., S. M. Uppala, A. J. Simmons,P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M.A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. GeerJ, L. Haimberger, S. B. Healy, H. Hersbach , E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi , A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, F. Vitart, 2011: TheERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. Roy. Meteor. Soc.*, **137**, 553–597.

Delle Monache, L., T. Nipen, Y. Liu, G. Roux, and R. Stull, 2011: Kalman filter and analog schemes to postprocess numerical weather predictions. *Mon. Wea. Rev.*, **139**, 3554-3570. doi:10.1175/2011MWR3653.1.

Delle Monache, L., F.A. Eckel, D.L. Rife, B. Nagarajan, and K. Searight, 2013: Probabilistic weather prediction with an analog ensemble. *Mon. Wea. Rev.*, **141**, 3498-3516. doi:10.1175/MWR-D-12-00281.1.

Deng, A., N. L. Seaman, and J. S. Kain, 2003: A shallow-convection parameterization for mesoscale models. Part I: submodel description and preliminary applications. J. Atmos. Sci., 60, 34–56.

Deng, A., B. J. Gaudet, J. Dudhia, and K. Alapaty, 2014: Implementation and evaluation of a new shallow convection scheme in WRF. *26th Conf. on Wea. Analysis and Forecast. 22ndConf. on Numer. Wea. Pred.*, Atlanta, GA, 13 pp. [Preprint available at https://ams.confex.com/ams/94Annual/webprogram/Paper236925.html.]

Descombes, G., T. Auligné, H.-C. Lin, D. Xu, C. Schwartz, and F. Vandenberghe, 2014: Multi-sensor Advection Diffusion nowCast (MADCast) for cloud analysis and short-term prediction. NCAR Tech. Note NCAR/TN-509+STR, 21 pp. doi:10.5065/D62V2D37.

Diagne, M., M. David, P. Lauret, J. Boland, and N. Schmutz, 2013: Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. Renewable and Sustainable Energy Reviews, 27, 65-76. doi:10.1016/j.rser.2013.06.042.

Doherty, R., and M. O'Malley, 2005: A new approach to quantify reserve demand in systems with significant installed wind capacity. Power Systems, IEEE Transactions on, **20**(2), 587-595.

Drobot, S.D., M. Chapman, P.A. Pisano, B.B. McKeever, 2010: Using Vehicles as Mobile Weather Platforms. *Advances in Intelligent and Soft Computing*, Volume **81**, 203–213.

Drucker, H., C.J.C. Burges, L. Kaufman, A. Smola, and V. Vapnik, 1997: Support vector regression machines. Advances in neural information processing systems, pages 155–161, 1997.

Dubus, L., 2014: Weather and climate and the power sector: Needs, recent developments and challenges, in *Weather Matters for Energy* (Troccoli, Dubus, and Haupt, Eds.), Springer, New York. doi:10.1007/978-1-4614-9221-4_18.

Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. J. Atmos. Sci., 46, 3077–3107.

Eckel, F.A., and C. F. Mass, 2005: Aspects of effective mesoscale, short-range ensemble forecasting. *Wea. Forecasting*, **20**, 328–350.

Ela, E, V. Diakov, E. Ilbanex, and M. Heaney, 2013: Impacts of Variability and Uncertainty in Solar Photovoltaic Generation at Multiple Timescales. Technical Report: NREL/TP-5500-58274.

Engel, C., and E.E. Ebert, 2012: Gridded operational consensus forecasts of 2-m temperature over Australia. *Wea. Forecasting*, **27**, 301-322. doi:10.1175/WAF-D-11-00069.1.

Evans, A.N., 2006: Cloud motion analysis using multichannel correlation-relaxation labeling. *Geoscience and Remote Sensing Letters*, IEEE, **3**(3):392–396, 2006.

Fernandez, E., F. Almonacid, N. Sarmah, P. Rodrigo, T.K. Mallick, and P Perez-Higueras, 2014: A model based on artificial neuronal network for the prediction of the maximum power of a low concentration photovoltaic module for building integration. *Solar Energy*, **100**, 148-158.

Feruzzi, G., G. Cervone, L. Delle Monache, G. Graditi, and F. Jacobone, 2016: Optimal bidding strategy in a day-ahead energy market for micro grid under uncertainty in renewable energyproduction. *Energy*, in press.

Fu, C-L., and H-Y. Cheng, 2013: Predicting solar irradiance with all-sky image features via regression. *Solar Energy.* **97**, 537-550.

Gerstmaier, T., M. Bührer, M., Röttger, A. Gombert, C.W. Hansen, and J.S. Stein, 2012: How predictable is DNI? An evaluation of hour-ahead and day-ahead DNI forecasts from four different providers. Proc. CPV-8 Conf., Toledo, Spain.

Gilbert, K.K., J.P. Craven, D.R. Novak, T. M. Hamill, J. Sieveking, D. P. Ruth, and S. J. Lord, 2015: An introduction to the National Blend of Global Models Project. *Spec. Symp. On Model Postprocessing and Downscaling* at the 95th AMS Annual Meeting, 7 Jan 2015, Phoenix, AZ. [Available online at https://ams.confex.com/ams/95Annual/webprogram/Paper267282.html.]

Gilbert, K.K., D.T. Myrick, J.P. Craven, J.P. Gagan, T.M. Hamill, D.R. Novak, D.P. Ruth, J. Settelmaier, J. E. Sieveking, and B. Veenhuis Jr., 2016: The National Blend of Global Models, Version One. *23rd Conf. on Probability and Statistics in the Atmospheric Sciences* at the 96th AMS Annual Meeting, 11 Jan 2016, New Orleans, LA. [Abstract available online at https://ams.confex.com/ams/96Annual/webprogram/Paper285973.html.]

Ginoux, P., M. Chin, I. Tegen, J. M. Prospero, B. Holben, O. Dubovik, and S.-J. Lin, 2001:Sources and distributions of dust aerosols simulated with the GOCART model. J. Geophys. Res., 106(D17), 20 255–20 273.

Glahn, H.R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting, *J. Appl. Meteor*., 11, 1972, pp. 1203–1211.

Glahn, H.R., and D.P. Ruth, 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor. Soc*., **84**, 195-201. doi: 10.1175/BAMS-84- 2-195.

Gleason, T.A., 1967: On Theoretical Limits of Predictability, *J. Applied Meterol*., **6**, 213-215.

Gneiting, T., 2011: Making and evaluating point forecasts. *J. Amer. Stat. Assoc.*, **106,** 746-762. doi:10.1198/jasa.2011.r10138.

Golemati, S., J. S. Stoitsis, A. Gastounioti, A.C. Dimopoulos, V. Koropouli, and K.S. Nikita, 2012: Comparison of block matching and differential methods for motion analysis of the carotid artery wall from ultrasound images. Information Technology in Biomedicine, IEEE Transactions on, 16(5):852–858, 2012.

Grell, G.A., and S. Freitas, 2014: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atm. Chem. and Phys.*, **14**, 5233–5250.

Greybush, S., S.E. Haupt, and G. Young, 2008: The Regime Dependence of Optimally Weighted Ensemble Model Consensus Forecasts of Surface Temperature, *Wea. Forecasting,* **23**, pp. 1146-1161, Dec 2008. doi: 10.1175/2008WAF2007078.1

Gueymard, C. A., and J. Ruiz-Arias, 2015: Validation of direct normal irradiance predictions under arid conditions: A review of radiative models and their turbidity-dependent performance. Renew. Sustain. Energy Rev., (In Press.).

Gueymard, C. A., and S. M. Wilcox, 2011: Assessment of spatial and temporal variability in the US solar resource from radiometric measurements and predictions from models using ground-based or satellite data. *Sol. Ener.*, **85,** 1068-1084. doi:10.1016/j.solener.2011.02.030.

Hall, T.J., C.N. Mutchler, G.J. Bloy, R.N. Thessin, S.K. Gaffney, and J.J. Lareau, 2011: Performance of observation-based prediction algorithms for very short-range, probabilistic clear-sky condition forecasting. *J. Appl. Meteor. Climatol*., **50**, 3–19.

Hamill, T. M., and T. Nehrkorn, 1993: A short-term cloud forecast scheme using cross correlations. *Weather and Forecasting*, 8(4):401–411, 1993.

Hammer, A., D. Heinemann, E. Lorenz, and B. Lückehe, 1999: Short-term forecasting of solar radiation: a statistical approach using satellite data. *Solar Energy*, **67**:1, 139-150.

Han, Y., P. van Delst, Q. Liu, F. Weng, B. Yan, R. Treadon, and J. Derber, 2006: Community Radiative Transfer Model (CRTM) Version 1, NOAA NESDIS. Tech. Rep. 122 Silver Spring, Maryland: Natl. Oceanic and Atmos. Admin, 22 pp.

Hao Huang, H., J. Xu, Z. Peng, S. Yoo, D. Yu, D. Huang, and H. Qin, 2013: Cloud motion estimation for short term solar irradiation prediction. In Smart Grid Communications (SmartGridComm), 2013 IEEE International Conference on, pages 696–701. IEEE, 2013.

Hassanzadeh, M., M. Etezadi-Amoli, and M.S. Fadali, 2010: Practical approach for sub-hourly and hourly prediction of PV power output. North American Power Symposium (NAPS), 1-5, Sept 26-28.

Haupt, S.E., 2013: A public-private-academic partnership to advance solar forecasting. 42nd American Solar Energy Society Annual Conf., Baltimore, MD, 16-20 Apr 2013. Preprint available online at http://proceedings.ases.org/solar-2013/resources/.

Haupt, S.E. and B. Kosovic, 2015: Big data and machine learning for applied weather forecasts: Forecasting solar power for utility operations. IEEE Symp. Series on Comp. Intell., Cape Town, South Africa, 9 Dec 2015. doi:10.1109/SSCI.2015.79.

Haupt, S.E., and B. Kosovic, 2016: Variable generation power forecasting as a big data problem. Submitted to IEEE Trans. Sust. Ener.

Haupt, S.E., and W.P. Mahoney, 2015: Wind Power Forecasting, *IEEE Spectrum*, Nov 2015, pp. 46-52.

He, K., and J. Sun, 2012. Statistics of patch offsets for image completion. In Proceedings of the 12th European Conference on Computer Vision - Volume Part II, ECCV'12, pages 16–29, Berlin, Heidelberg, 2012. Springer-Verlag.

He´as, P., E. Me´min, 2008: Three-dimensional motion estimation of atmospheric layers from image sequences. Geoscience and Remote Sensing, IEEE Transactions on, 46(8):2385–2396.

He´as, P., E. Me´min, N. Papadakis, and A. Szantai, 2008: Layered estimation of atmospheric mesoscale dynamics from satellite imagery. Geoscience and Remote Sensing, IEEE Transactions on, 45(12):4087–4104.

Heidinger, A. K., M. J Foster, A. Walther and X. Zhao, 2013: The Pathfinder Atmospheres Extended AVHRR Climate Data Set. *Bull. Amer. Meteor. Soc*, doi:10.1175/BAMS-D-12-00246.1.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**(5), 559-570.

Hicks, B.B., J.J. DeLuisi, and D.R. Matt, 1996: The NOAA Integrated Surface Irradiance Study (ISIS). A new surface radiation monitoring program. *Bull. of the Amer. Met. Soc.*, **77**, 2857–2864.

Hinkelman, L. M. 2013: Relating Solar Resource Variability to Cloud Type. *NREL Progress Report*.

Hinkelman, L. M., 2014: Statistics of solar resource variability on short time scales, A Public-Private-Academic Partnership to Advance Solar Power Forecasting Project Workshop, Boulder, CO.

Hinkelman, L.M., K.F. Evans, E.E. Clothiaux, T.P. Ackerman, and P.W. Stackhouse Jr., 2007: The effect of cumulus cloud field anisotrophy on domain-averaged solar fluxes and atmospheric heating rates. *Journal of Atmospheric Sciences*, 3499-3520.

Hinkelman, L. M., N. Schaeffer, and T. P. Ackerman, 2015: The character and variability of solar irradiance across the Pacific Northwest American Geophysical Union Fall Meeting, San Francisco, CA.

Hoff, T and R. Perez, 2010: Modeling PV Fleet Output Variability. *Clean Power Research*. p.1-

Hoff, T.E., and R. Perez, 2012: Modeling pv fleet output variability. Solar Energy, 86(8):2177–2189, 2012.

Horn, B.K. and B.G. Schunck, 1981: Determining optical flow. In 1981 Technical Symposium East, pages 319–331. International Society for Optics and Photonics, 1981.

Huang, H., J. Xu, Z. Peng, S. Yoo, D. Huang, and H. Qin, 2013: Cloud motion estimation for short term solar irradiation prediction. Preprint, *IEEE Int. Conf. on Smart Grid Communications*, Vancouver, BC, Canada, 21-24 Oct 2013. doi:10.1109/SmartGridComm.2013.6688040.

Huang, H., S. Yoo, D. Yu, D. Huang, and H. Qin, 2011: Cloud motion detection for short term solar power prediction. In Proceedings of ICML 2011 Workshop on Machine Learning for Global Challenges.

Huang, H., S. Yoo, D. Yu, D. Huang, and H. Qin, 2012: Correlation and local feature based cloud motion estimation. In Proceedings of the Twelfth International Workshop on Multimedia Data Mining, pages 1–9. ACM.

Iacono, M. J., J. S., Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models. *J. Geophys. Res*., **113**, D13 103, doi:10.1029/2008JD009 944.

Inman, R.I., H.T.C. Pedro, and C.F.M. Coimbra, 2013: Solar forecasting methods for renewable energy integration. *Prog. Ener. Combust. Sci.*, **39,** 535-576. doi:10.1016/j.pecs.2013.06.002.

Inness, A., and Coauthors, 2013: The MACC reanalysis: an 8-year data set of atmospheric composition. *Atm. Chem. and Phys*., **13**, 4073–4109.

James, E.P., S. Benjamin, J.B. Olson, M. Marquis, and C.R. Alexander, 2015: Solar Forecasts from the 3-km High-Resolution Rapid Refresh. Presentation 5.4, Sixth Conference on Weather, Climate, and the New Energy Economy, AMS 2015 Annual Meeting, Phoenix, AZ.

Jensen, T.L, T.L. Fowler, B.G. Brown, J. Lazo, S.E. Haupt. 2016: Metrics for evaluation of solar energy forecasts. NCAR Technical Report. In press, report number not assigned yet.

Jiménez, P. A., J. P. Hacker, J. Dudhia, S. E. Haupt, J. A. Ruiz-Arias, C. A. Gueymard, G. Thompson, T. Eidhammer, and A. Deng, 2016a: WRF-Solar: An augmented NWP model for solar power prediction. Model description and clear sky assessment. *Bull. Amer. Meteor. Soc.*, in press. doi:10.1175/BAMS-D-14-00279.1.

Jiménez, P.A., S. Alessandrini, S.E. Haupt, A. Deng, B. Kosovic, J.A. Lee, and L. Delle Monache, 2016b: The role of unresolved clouds on short-range global horizontal irradiance predictability. *Mon. Wea. Rev.*, accepted.

Junk, C., L. Delle Monache, S. Alessandrini, G. Cervone, and L. von Bremen, 2015a: Predictor-weighting strategies for probabilistic wind power forecasting with an analog ensemble. *Met. Z.*, **24,** 361-379. doi:10.1127/metz/2015/0659.

Junk, C., L. Delle Monache, and S. Alessandrini, 2015b: Analog-based ensemble model output statistics. *Mon. Wea. Rev.*, **143,** 2909-2917. doi:10.1175/MWR-D-15-0095.1.

Kassianov, E., C.N. Long, and J. Christy, 2005: Cloud-base-height estimation from paired ground-based hemispherical observations. *Journal of Applied Meteorology*, **44**(8):1221–1233, 2005.

Kinne, S., 2009: Climatologies of cloud-related aerosols. Clouds in the Perturbed Climate System: their relationship to energy balance, atmospheric dynamics and precipitation. MIT press, 608 pp.

Kinne, S., and Coauthors, 2003: A global comparison among models, satellite data and AERONET ground data. *J. Geophys. Res*., **108**, D20, 4634.

Kinne, S., and Coauthors, 2006: An AeroCom initial assessment - optical properties in aerosol component modules of global models. *Atm. Chem. and Phys*., **6**, 1815–1834.

Kinne, S., and Coauthors, 2013: MAC-v1: A new global aerosol climatology for climate studies. *J. Adv. Model. Earth Syst.*, **5**, 704–740.

Kleissl, J. (ed.), 2013: *Solar Energy Forecasting and Resource Assessment*. Academic Press. 462pp ISBN 9780123971777.

Kleist, D.T., D.F. Parrish, J.C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea. Forecast*., **24**, 1691-1705. doi:10.1175/2009WAF2222201.1.

Kohavi, R., 1995: A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'95, pages 1137–1143, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc.

Kuhn, M., S. Weston, C. Keefer, and N. Coulter, 2012: Cubist models for regression. R Project documentation. Available online at https://cran.r-project.org/web/packages/Cubist/vignettes/cubist.pdf.

Kuszmaul, S., A. Ellis, J. Stein, and L. Johnson, 2010: Lanai High-Density Irradiance Senor Network for Characterizing Solar Resource Variability of MW-Scale PV System. 35th IEEE PVSC, Honolulu HI.

Lave, M., and J. Kleissl, 2013: Cloud speed impact on solar variability scaling–application to the wavelet variability model. *Solar Energy*, **91**:11–21, 2013.

Lee, J.A., S.E. Haupt, P.A. Jiménez, M.A. Rogers, S.D. Miller, and T.C. McCandless, 2016: Solar energy nowcasting case studies near Sacramento. *Wea. Forecasting*, submitted.

Leese, J.A., C.S. Novak, and V.R. Taylor, 1970: The determination of cloud pattern motions from geosynchronous satellite image data. *Pattern Recognition*, **2**(4):279–292, 1970.

Lew, D. G. Brinkman, A. Florita, M. Heaney, B-M. Hodge, M. Hummon, and E. Ibanez, 2012: Sub-Hourly Impacts of High Solar Penetrations in the Western United States. *2nd Annual International Workshop on Integration of Solar Power into Power Systems Conference*, Lisbon, Portugal Nov 12-13 2012.

Lew, D., D. Piwko, N. Miller, G. Jordan, K. Clark and L. Freeman. 2010: How do high levels of wind and solar impact the grid? The Western Wind and Solar Integration Study. NREL Technical Report.

Lindeberg, T., 1993: Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention. International Journal of Computer Vision, 11(3):283–318, 1993.

Liousse, C., J. E. Penner, C. Chuang, J. J. Walton, H. Eddleman, and H. Cachier, 1996: A global three-dimensional model study of carbonaceous aerosols. *J. Geophys. Res*., **101**, 19 411–19 432.

Lippmann, R.P., 1987: An introduction to computing with neural nets. IEEE Acoustics, *Speech Signal Process*. **Mag.**, 4, 4–22.

Long, C. N., and T. P. Ackerman, 2000: Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects. *J. Geophys. Res*., **105**,15 609–15 626.

Lopez, G., F.J. Batlles, and J. Tovar-Pescador, 2005: Selection of input parameters to model direct solar irradiance by using artificial neural networks. *Energy*, **30**, 1675-1684.

Lorenz, E., 1963: Deterministic Nonperiodic Flow, *J. Atmos. Sci*., **20**, 130-141.

Lorenz, E., 1969: The Predictability of a Flow Which Possesses Many Scales of Motion, *Tellus*, XXI, 289-307.

Lorenz, E., A. Hammer, and D. Heinemann, 2004: Short-term forecasting of solar radiation based on satellite data. EUROSUN2004, ISES Europe Solar Congress, 841-848.

Lorenz, E., J. Kuhnert, and D. Heinemann, 2014: Overview of Irradiance and Photovoltaic Power Prediction, in *Weather Matters for Energy* (Troccoli, Dubus, and Haupt, Eds.), Springer, New York. doi:10.1007/978-1-4614-9221-4_21.

Lorenz, E., Kuhnert, J., and D. Heinemann, 2012: Overview on irradiance and photovoltaic power prediction. *Weather Matters for Energy*, 429-454.

Lucas, B.D., and T. Kanade, 1981: An iterative image registration technique with an application to stereo vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'81, pages 674–679, San Francisco, CA, USA, 1981. Morgan Kaufmann Publishers Inc.

Madrigal, M. and K. Porter. 2013: Operating and Planning Electricity Grids with Variable Renewable Generation. The World Bank, Washington, DC. 1-35. Print.

Madsen, H., P. Pinson, G. Kariniotakis, H.A. Nielsen, and T.S. Nielsen, 2005: Standardizing the performance evaluation of short-term wind power prediction models. *Wind Engineering*, **29**(6), 475-489.

Mahoney, W.P., K. Parks, G. Wiener, L. Yubao, W. L. Myers, S. Juanzhen, L. Delle Monache, T. Hopson, D. Johnson, and S.E. Haupt, 2012: A Wind Power Forecasting System to

Optimize Grid Integration, *IEEE Transaction on Sustainable Energy*, **3**:4, 670-682, Oct. 2012.

Marquez, R. and C.F.M. Coimbra, 2011: Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database. *Solar Energy*, **85**, 746-756.

Marquez, R., and C.F.M. Coimbra, 2013a: Intra-hour DNI forecasting based on cloud tracking image analysis. *Solar Energy*, **91**, 327-336.

Marquez, R., H.T.C. Pedro, and C.F.M. Coimbra, 2013b: Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs. *Solar Energy*, **92**, 176-188.

Marquez, R., V. Gueorguiev, and C.F.M. Coimbra, 2013: Forecasting of Global Horizontal Irradiance Using Sky Cover Indices, *ASME Journal of Solar Energy Engineering*, **135**, 0110171-0110175.

Martín, L., L.F. Zarzalejo, J. Polo, A. Navarro, R. Marchante, and M. Cony, 2010: Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning. *Solar Energy*, **84**:10, 1772-1781.

Mathiesen, P., J. Kleissl, and C. Collier, 2012: Characterization of irradiance variability using a high-resolution, cloud-assimilating NWP. World Renewable Energy Forum, Denver, 13-17 May, available at http://ases.org/conference/program/wref-detail/.

McCandless, T.C., S. E. Haupt, and G.S. Young, 2014: Short-Term Solar Radiation Forecasts Using Weather Regime-Dependent Artificial Intelligence Techniques, 12th Conference on Artificial Intelligence: Applications of Artificial Intelligence Methods for Energy, Atlanta, GA, *Amer. Meteor. Soc*, J3.5.

McCandless, T. C., S. E. Haupt, and G. S. Young, 2015: A model tree approach to forecasting solar irradiance variability. *Sol. Ener.*, **120,** 514-524. doi:10.1016/j.solener.2015.09.020.

McCandless, T.C., S.E. Haupt, and G.S. Young, 2016a: A regime-dependent artificial neural network technique for short-range solar irradiance forecasting. *Renew. Ener*., **89**, 351-359. doi:10.1016/j.renene.2015.12.030.

McCandless, T.C., G.S. Young, S.E. Haupt, and L.M Hinkelman, 2016b: Regime-Dependent Short-Range Solar Irradiance Forecasting, submitted to *Journal of Applied Meteorology and Climatology*, accepted.

Mellit, A., 2008: Artificial Intelligence Technique for Modeling and Forecasting of Solar Radiation Data: A Review. *Int. Journal Artificial Intelligence and Soft Computing*, **1**:1, 52-76.

Mellit, A., A. Massi Pavan, and V. Lughi, 2014: Short-Term Forecasting of Power Production in a Large-Scale Photovoltaic Plant, *Solar Energy*, **105**, 401-413.

Miller, S. D., M. A. Rogers, A. K. Heidinger, I. Laszlo, and M. Sengupta, 2012: Cloud advection schemes for short-term satellite-based insolation forecasts, *Proc. World Renewable Energy Forum*, 13-17 May 2012, Denver, CO, pp. 1963-1967. [Available online at

https://ases.conference-services.net/resources/252/2859/pdf/SOLAR2012_0385_full%20paper.pdf.]

Miller, S.D. J.M., Forsythe, P.T. Partain, J.M Haynes, R.L. Bankert, M. Sengupta, C. Mitrescu, J.D. Hawkins, and T.H. Vonder Haar, 2014: Estimating Three-Dimensional Cloud Structure via Statistically Blended Satellite Observations. *J. Appl. Meteor. Climatol*., **53**, 437–455.

Miller, S.D., A.K. Heidinger, and M. Sengupta, 2013: Physically-Based Satellite Methods. Chapter 3, Solar Energy Forecasting and Resource Assessment, J. Kleissl, Ed. ISBN 9780123971777.

Miller, S.D., M.A. Rogers, J.M. Haynes, and M. Sengupta, 2016: A Satellite-Initialized Model-Advected Scheme for Short-Term Solar Energy Forecasting. In preparation for Solar Forecasting 2016.

Mills, A. and R. Wiser. 2010: Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power. Ernest Orlando Lawrence Berkeley National Laboratory. http://eetd.lbl.gov/EA/EMP.

Morcrette, J.-J., A. Beljaars, A., Benedetti, L. Jones, L. and O. Boucher, 2008. Sea-salt and dust aerosols in the ECMWF IFS model. *Geophys. Res. Lett*. **35**, doi:10.1029/2008GL036041.

Morf, H., 2014: Sunshine and cloud cover prediction based on Markov processes. *Solar Energy*, **110**, 615-626.

Mori, N. and K-A. Chang, 2007: Introduction to MPIV. User manual and program available online at http://sauron. civil. eng. osaka-cu. ac. jp/˜ mori/softwares/mpiv, 2003.

Morris, V.R., 2005: Total sky imager (tsi) handbook. Technical report, ARM TR-017, 2005.

Muller, M., 1995: Equation of time - Problem in astronomy. Acta Phys. Pol A, 88 Supplement, S49–S67.

Myers, W., G. Wiener, S. Linden, and S. E. Haupt, 2011: A consensus forecasting approach for improved turbine hub height wind speed predictions, in *Proc. WindPower 2011*, Anaheim, CA, May 24, 2011.

Myers, W., S. Linden, G. Wiener, and S.E. Haupt, 2012: A Consensus Wind Forecasting System, World Renewable Energy Forum, Denver, CO, May 17.

Nguyen, D.A., and J. Kleissl, 2014: Stereographic methods for cloud base height determination using two sky imagers. *Solar Energy*, **107**:495–509, 2014.

Nielsen, H. A., H. Madsen, and T.S. Nielsen, 2006: Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wind Energy*, **9**(1-2), 95-108.

Notton, G., C. Paoli, S. Vasileva, M-L Nivet, J-L. Canaletti, and C. Cristofari, 2012: Estimation of hourly global solar irradiation on tilted planes from horizontal one using artificial neural networks. *Energy*, **39**, 166-179.

Orwig, K.D., M. Ahlstrom, V. Banunarayanan, J. Sharp, J.M. Wilczak, J. Freedman, S.E. Haupt, J. Cline, O. Bartholomie, H. Hamman, B.-M. Hodge, C. Finley, D. Nakafuji, J. Peterson, D.

Maggio, and M. Marquis, 2014: Recent Trends in Variable Generation Forecasting and Its Value to the Power System, *IEEE Transactions on Renewable Energy,* **6**(3), 924-933.

Pappas, V., N. Hatzianastassiou, C. Papadimas, C. Matsoukas, S. Kinne, and I. Vardavas, 2013: Evaluation of spatio-temporal variability of Hamburg aerosol climatology against aerosol datasets from modis and CALIOP. *Atm. Chem. and Phys*., **13**, 8381–8399.

Parks, K., Y-H Wan, G. Wiener, Y. Liu, 2011: Wind Energy Forecasting – A Collaboration of the National Center for Atmospheric Research (NCAR) and Xcel Energy, National Renewable Energy Laboratory, Golden, CO.

Pedro, H.T.C., and C.F.M. Coimbra, 2012: Assessment of forecasting techniques for solar power prediction with no exogenous inputs. *Solar Energy*, **86**, 2017-2028.

Peng, Z., D. Yu, D. Huang, J. Heiser, S. Yoo, and P. Kalb, 2015: 3D cloud detection and tracking system for solar forecast using multiple sky imagers. *Sol. Ener.*, **118,** 496-519. doi:10.1016/j.solener.2015.05.037.

Peng, Z., S. Yoo, D. Yu, and D. Huang, 2013: Solar irradiance forecast system based on geostationary satellite. In Smart Grid Communications (SmartGridComm), 2013 IEEE International Conference on, pages 708–713. IEEE, 2013.

Peng, Z., S. Yoo, D. Yu, D. Huang, P. Kalb, and J. Heiser, 2014: 3D cloud detection and tracking for solar forecast using multiple sky imagers. In Proceedings of the 29th Annual ACM Symposium on Applied Computing, 512–517. ACM, 2014.

Perez, R., M. Beauharnois, K. Hemker, S. Kivalov, E. Lorenz, S. Pelland, J. Schlemmer, and G. Van Knowe, 2011: Evaluation of Numerical Weather Prediction solar irradiance forecasts in the US. Proc. Solar 2011 Conf., Raleigh, NC, American Solar Energy Soc.

Perez, R., S. Kivalov, A. Zelenka, J. Schlemmer and K. Hemker Jr., 2010: Improving The Performance of Satellite-to-Irradiance Models using the Satellite's Infrared Sensors. Proc., ASES Annual Conference, Phoenix, Arizona.

Pinson, P., P. McSharry, and H. Madsen, 2010: Reliability diagrams for non-parametric density forecasts of continuous variables: Accounting for serial correlation. *Quarterly Journal of the Royal Meteorological Society*, **136**(646), 77-90.

Quesada-Ruiz, S., Y. Chu, J. Tovar-Pescador, H.T.C. Pedro, and C.F.M. Coimbra, 2014: Cloud-tracking methodology for intra-hour DNI forecasting. Solar Energy, **102**, 267-275.

Quinlan, J.R., 1987a: Generating production rules from decision trees. Proc. 10th Internat. Joint Conf. of Artificial Intelligence, 304-307, Milan, Italy, 23-28 Aug 1987. Preprint available online at http://ijcai.org/Past%20Proceedings/IJCAI-87-VOL1/PDF/063.pdf.

Quinlan, J.R., 1987b: Simplifying decision trees. *Internat. J. Man-Machine Studies*, **27**, 221-234. doi:10.1016/S0020-7373(87)80053-6.

Quinlan, J.R., 1992: Learning with continuous classes. AI '92: Proc. 5th Australian Joint Conf. on Artificial Intelligence, Hobart, Australia, 16-18 Nov 1992. Singapore: World Scientific, 343-348, A. Adams and L. Sterling, Eds. Preprint available online at http://sci2s.ugr.es/keel/pdf/algorithm/congreso/1992-Quinlan-AI.pdf.

Quinlan, J.R., 1996: Improved use of continuous attributes in C4.5. J. *Artif. Intell. Res*., **4**, 77-90. doi: 10.1613/jair.279.

Raftery, A.E., F. Balabdaoui, T. Gneiting, and M. Polakowski, 2003: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155-1174. doi: 10.1175/MWR2906.1.

Randles, C.A., P.R. Colarco, and A. da Silva, 2013: Direct and semi-direct aerosol effects in the NASA GEOS-5 AGCM: aerosol-climate interactions due to prognostic versus prescribed aerosols. *J. Geophys. Res. Atmospheres*, **118**, 149–169.

Rayl, J., G. S. Young, and J. R. S. Brownson, 2013: Irradiance co-spectrum analysis: Tools for decision support and technological planning. *Solar Energy:* **95**, 354-275.

Reed, D.R., and R.J. Marks, 1998: Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks. MIT Press, Cambridge, MA, USA.

Reikard, G., 2009: Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy*, **83**, 342-349.

Rienecker, M. M., and Coauthors, 2000: The geos-5 data assimilation system- documentation of versions 5.0.1, 5.1.0 and 5.2.0. Tech. Rep. NASA/TM-2008-104606, **27**, NASA, Greenbelt, Maryland, USA, 101 pp.

Roesch, A., M. Wild, A. Ohmura, E.G. Dutton, C.N. Long, and T. Zhang, 2011: Assessment of BSRN radiation records for the computation of monthly means. *Atmos. Meas. Tech*, **4**, 339–354.

Rogers, M.A., S.D. Miller, J.M. Haynes, A. Heidinger, S.E. Haupt, and M. Sengupta, 2015: Improvements in satellite-derived short-term insolation forecasting: Statistical comparisons, challenges for advection-based forecasts, and new techniques. 6th Conf. on Weather, Climate, and the New Energy Economy at the 95th AMS Annual Meeting, Phoenix, AZ, 6 Jan 2015. Abstract available online at https://ams.confex.com/ams/95Annual/webprogram/Paper268850.html.

Rosenblatt, F., 1958: The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. In. *Psychological Review*, **65**:6, 386-408.

Rotunno, R. and C. Snyder, 2008: A Generalization of Lorenz's Model for the Predictability of Flows with Many Scales of Motion, *J. Atmos. Sci.*, **65**, 1063-1076.

Roulston, M. S., D.T. Kaplan, J. Hardenberg, and L.A. Smith, 2003: Using medium-range weather forecasts to improve the value of wind energy production. *Renew. Ener.,* **28**(4), 585-602.

Ruiz-Arias, J.A., C.A Gueymard, J. Dudhia, and D. Pozo-Vazquez, 2012: Improvement on the WRF model for solar resource assessments and forecasts under clear skies. Proc. WREF Conf., Denver, CO, Am Solar En Soc.

Ruiz-Arias, J.A., H. Alsamanra, J. Tovar-Pescador, and D. Pozo-Vazquez, 2010: Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions. *Energy Conversion and Management*, **51**, 881-893. doi:10.1016/j.enconman.2009.11.024.

Ruiz-Arias, J.A., J. Dudhia and C.A. Gueymard, 2014: A simple parameterization of the shortwave aerosol optical properties for surface direct and diffuse irradiances assessment in a numerical weather model. *Geosci. Model Dev.*, **7**, 1159-1174. doi:10.5194/gmd-7-1159-2014.

Ruiz-Arias, J.A., J. Dudhia, F.J. Santos-Alamillos, and D. Pozo-Vazquez, 2013: Surface clear-sky shortwave radiative closure intercomparisons in the Weather Research and Forecasting model. *J. Geophys. Res*., **118**, 9901–9913.

Rutledge, G., E. Legg, and P. Menzel, 1991: Operational production of winds from cloud motions. *Global and Planetary Change*, **4**(1):141–150, 1991.

Schaeffer, N., L. M. Hinkelman, and T. P. Ackerman, 2016: Relating solar irradiance variations and weather across the Pacific Northwest, AMS 7th Conference on Weather, Climate, and the New Energy Economy, January 2016, New Orleans, LA.

Schroedter-Homscheidt, M., A. Oumbe, A. Benedetti, and J.J. Morcrette, 2013: Aerosols for concentrating solar electricity production forecasts: Require quantification and EAACMAWAF/MACC Aerosol Forecast Assessment, *Bulletin American Meterol. Soc*., June 2013, 902-914. doi:10.1175/BAMS-D-11-00259.1

Sengupta, M. and L. M. Hinkelman, 2014: Temporal variability of surface solar irradiance as a function of satellite-retrieved cloud properties, American Geophysical Union Fall Meeting, San Francisco, CA.

Sharma, N., P. Sharma, D. Irwin, and P. Shenoy, 2011: Predicting Solar Generation from Weather Forecasts Using Machine Learning: Proceedings of the 2nd IEEE International Conference on Smart Grid Communications, Brussels, 17-20 October, pp. 32-37.

Shettle, I. P., and R. W. Fenn, 1979: Models for the aerosols of the lower atmosphe and the effects of humidity variations on their optical properties. Tech. Rep. AFGL-TR-79-0214, Air Force Geophys. Lab.

Shi, J., 2010: WRF simulations of the 20-22 January 2007 snow events over eastern Canada: Comparison with in situ and satellite observations. *J. Appl. Meteor. & Climatol.*, **49**, 2246–2266.

Shi, J., and J. Malik, 2000: Normalized cuts and image segmentation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22(8):888–905, 2000.

Shusterman, E. and M. Feder. Image compression via improved quadtree decomposition algorithms. *Image Processing, IEEE Transactions* on, **3**(2):207–215, 1994.

Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. Quarter. *J. Roy. Met. Soc*., **131**, 3079–3102.

Singh, M., and M. Glennen, 2005: Automated ground-based cloud recognition. *Pattern Analysis and Applications*, **8**(3):258–271, 2005.

Skamarock, W.C., J.B. Klemp, J. Dudhia, D.O. Gill, D.M. Barker, M.G. Duda, X.-Y. Huang, W. Wang, and J.G. Powers, 2008: A description of the Advanced Research WRF Version 3. NCAR Tech. Note NCAR/TN-475+STR. 113 pp. doi:10.5065/D68S4MVH.

Smola, A.J. and B. Scholkopf, 2004: A tutorial on support vector regression. Statistics and computing, 14(3):199–222, 2004.

Sotiras, A., C. Davatzikos, and N. Paragios, 2013: Deformable medical image registration: A survey. Medical Imaging, IEEE Transactions on, 32(7):1153–1190, 2013.

Stackhouse, P., T. Zhang, W.S. Chandler, C.H. Whitlock, J.M. Hoell, D.J. Westberg, R.Perez, and S. Wilcox, 2008: Satellite Based Assessment of the NSRDB Site Irradiances and Time Series from NASA and SUNY-Albany Algorithms. Proc. ASES Annual Meeting, San Diego, CA.

Stein, J, R. Perez and A. Parkins, 2010: Validation of PV Performance Models Using Satellite-based Irradiance Measurements: a Case Study. Proc. ASES Annual Conference, Phoenix, AZ.

Stensrud, D. J., 2007: Parameterization schemes: keys to understanding numerical weather prediction models. Cambridge University Press, 478 pp.

Stokes, M., M. Anderson, S. Chandrasekar, and R. Motta, 1996: A standard default color space for the internetsrgb, 1996. URL http://www. w3. org/Graphics/Color/sRGB.

Stone, H.S., M.T. Orchard, E-C Chang, and S.A. Martucci, 2001: A fast direct fourier-based algorithm for subpixel registration of images. *Geoscience and Remote Sensing,* IEEE Transactions on, 39(10):2235–2243, 2001.

Sun, D., S. Roth, and M.J. Black, 2014: A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision*, **106**(2):115–137, 2014.

Tapakis, R., and A.G. Charalambides, 2013: Equipment and methodologies for cloud detection and classification: A review. *Solar Energy*, **95**, 392-430.

Tegen, I., and I. Fung, 1995: Modeling of mineral dust in the atmosphere: Sources, transport, and optical thickness. *J. Geophys. Res*., **100**, 18 707–18 726.

Tegen, I., P. Hollrig, M. Chin, I. Fung, D. Jacob, and J. Penner, 1997: Contribution of different aerosol species to the global aerosol extinction thickness: Estimates from model results. *J. Geophys. Res.,* **102**, 23 895–23 915.

Thirion, J-P., 1998: Image matching as a diffusion process: an analogy with Maxwell's demons. Medical image analysis, **2**(3):243–260, 1998.

Thompson, G., and T. Eidhammer, 2014: A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *J. Atmos. Sci*., **71**, 3636–3658.

Thompson, G., M. Tewari, K. Ikeda, S. Tessendorf, C. Weeks, J. Otkin, and F. Kong, 2015: Explicitly-coupled cloud physics and radiation parameterizations and subsequent evaluation in WRF high-resolution convective forecasts. Atmos. Res. (Submitted).

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115.

Troccoli, A. and J.-J. Morcrette, 2012: Forecast assessment of surface solar radiation over Australia. Proc. WREF Conf., Denver, CO, American Solar Energy Soc.

Troccoli, A., L. Dubus, and S. E. Haupt (Eds.), 2014: *Weather Matters for Energy*, Springer, New York. 528 pp. doi:10.1007/978-1-4614-9221-4.

Tuohy, A., J. Zack, S. Haupt, J. Sharp, M. Ahlstrom, S. Dise, E. Grimit, C. Mohlren, M. Lange, M. Garcia Casado, J. Black, M. Marquis, C. Collier, 2015: Solar Forecasting – Method, Challenges, and Performance, *IEEE Power and Energy Magazine*, **13** (6), 50-59. DOI:10.1109/MPE.2015.2461351

Twomey, S., 1974: Pollution and the planetary albedo. *Atmos. Environ.*, **8**, 1251–1256.

Urquhart, B., B. Kurtz, E. Dahlin, M. Ghonima, J. E. Shields, and J. Kleissl, 2015: Development of a sky imaging system for short-term solar power forecasting. *Atmos. Meas. Tech.*, **8,** 875-890. doi:10.5194/amt-8-875-2015.

Urquhart, B., B. Kurtz, E. Dahlin, M. Ghonima, J.E. Shields, and J. Kleissl, 2014: Development of a sky imaging system for short-term solar power forecasting. *Atmospheric Measurement Techniques Discussions*, **7**(5):4859–4907, 2014.

Vanvyve, E., L. Delle Monache, A. J. Monaghan, and J. O. Pinto, 2015: Wind resource estimates with an analog ensemble approach. *Renew. Ener.*, **74,** 761-773. doi:10.1016/j.renene.2014.08.060.

Voyant, C., M. Muselli, C. Paoli, and M-L. Nivet, 2013: Hybrid methodology for hourly global radiation forecasting in Mediterranean area. *Renewable Energy*, **53**, 1-11.

Voyant, C., Muselli, M., Paoli, C., and M-L. Nivet, 2014: Numerical Weather Prediction (NWP) and hybrid ARMA/ANN to predict global radiation. *Energy*, **39**:1, 341-355.

Wang, F., Z. Mi, S. Su, and H. Zhao, 2012: Short-Term Solar Irradiance Forecasting Model Based on Artificial Neural Network using Statistical Feature Parameters, *Energies*, **5**, 1355-1370.

Wedel, A., T. Pock, C. Zach, H. Bischof, and D. Cremers, 2009: An improved algorithm for tv-l 1 optical flow. In Statistical and Geometrical Approaches to Visual Motion Analysis, pages 23–45. Springer, 2009.

Weiss, Y., 1997: Smoothness in layers: Motion segmentation using nonparametric mixture estimation. In Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, pages 520–526. IEEE, 1997.

Wilks, D.S., 2006: Statistical methods in the atmospheric sciences, 2nd Ed., Academic Press, 626 pp.

Wood-Bradley. P., J. Zapata, and J. Pye, 2012: Cloud tracking with optical flow for short-term solar forecasting. In 50th conference of the Australian Solar Energy Society, Melbourne, 2012.

Xie, Y., M. Sengupta, and J. Dudhia, 2016: A Fast All-sky Radiation Model for Solar Applications (FARMS): Algorithm and performance evaluation. *Solar Energy* (Under review).

Xu, D., T. Auligné., X.-Y. Huang, 2015: A Validation of the Multivariate and Minimum Residual Method for Cloud Retrieval Using Radiance from Multiple Satellites. Advances in Atmospheric Sciences, 32, 349-362. doi: 10.1007/s00376-014-3258-5.

Yang, D., P. Jirutitijaroen, and W. M. Walsh, 2012: Hourly Solar Irradiance Time Series Forecasting Using Cloud Cover Index. *Solar Energy*, **86**, 3531-3543.

Young, G. 2001: Combining forecasts for superior prediction, 16th Conference on Probability and Statistics in Atmospheric Science, 13-17 January, Orlando, FL, pp. 107-111.

Zagouras, A., A. Kazantzidis, E. Nikitidou, and A. A. Argiriou, 2013: Determination of measuring sites for solar irradiance, based on cluster analysis of satellite-derived cloud estimations. *Solar Energy*, **97**, 1-11.

Zhang, Q., L. Xu, and J. Jia, 2014: 100+ times faster weighted median filter (wmf). In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pages 2830–2837. IEEE, 2014.