CSI RD&D (Solicitation 4) Task 2 Final Report

Advanced Operational Ensemble Solar Forecasting System

Submitted to:

California Solar Initiative RD&D Program

July 2016

Submitted by:

Dipak Kumar, Elynn Wu, Handa Yang, Jan Kleissl

Center for Renewable Resource Integration and Center for Energy Research

Department of Mechanical and Aerospace Engineering

University of California, San Diego

9500 Gilman Drive

La Jolla, CA 92093-0411

Abstract

An operational solar forecasting system was demonstrated for San Diego Gas & Electric Company. The purpose of the system was to provide solar generation inputs into system-wide net load forecasts and operational load flow analysis. Analysis of initial condition impacts and physics parameterizations pointed towards biases in the forecast related to the thickness of the marine boundary layer. Analog ensemble post processing was implemented and a 3D data assimilation system were developed to reduce forecast errors. The system was run from May 2014 to present and continues to be used by SDG&E. Error analysis and operational experiences are presented.

Table of Contents

Abstract ii
Table of Contents iii
1. Introduction 1
2. Solar Forecast Requirements 1
3. Solar Forecast System 1
4. Validation sites and metrics 2
   4.1. Ground Stations 2
   4.2. Satellite solar resource data 3
   4.3. Forecast Metrics 3
5. Validation Results 4
6. Forecast improvements through Data Assimilation 4
   6.1. Data Assimilation Setup 4
   6.2. Validation 5
7. Operational Reliability 8
8. Conclusions 8
Appendix A 9

1. Introduction

Solar forecasting is an important tool for power grid operations. Distributed solar power represents a negative load that masks true load on the distribution system. Knowledge of the resulting net load can help to operate the distribution system more efficiently, for instance in scheduling infrastructure repairs, feeder switching, and, emergency operations.

San Diego Gas & Electric Company (SDG&E) is moving towards deep situational awareness on their distribution system. With rapid growth of solar PV in the SDG&E service territory, a solar forecast solution was desired.

Within the SDG&E territory the great challenge for solar forecasting is the accurate prediction of coastal stratocumulus (Sc, commonly termed the marine layer), as Sc is the dominant cloud type in the region, and their presence greatly attenuates the solar resource in the area with the majority of rooftop solar installations. Numerical weather prediction (NWP) models become increasingly important for hours-ahead to days-ahead forecasts, since non-linearities and global coupling in the atmospheric processes make other extrapolation methods less accurate for longer forecast horizons. Furthermore, the mesoscale and global processes which develop in these forecast horizons are amenable to modelled by physics-based models.

2. Solar Forecast Requirements

For distribution system operations the primary time horizon of interest is intra-day forecasts. Forecasts were therefore generated for delivery prior to 0600 Pacific Standard Time (PST = 1400 Universal Coordinated Time, UTC). The forecast horizon was 48 hours providing solar irradiance for the current and the following day at 15 min time steps. In this first operational implementation an intuitive solar irradiance quantity was desired. The Solar Potential Index (SPI) is defined as the forecast global horizontal irradiance (GHI) divided by the peak global horizontal irradiance which would be expected at solar noon on the summer solstice for a clear day. This normalized scale from 0 to 1 provides grid operators a better understanding of the relative importance of solar power generation on a particular day and hour.

Spatially, the SDG&E territory is divided into climate zones that map contiguous areas of similar solar resource. The solar potential index is averaged over all simulation grid cells located within a climate zone. While the SPI was used operationally, in this report we validate solar forecasts of GHI, which is a more commonly used metric to characterize the solar resource.

3. Solar Forecast System

An open-source NWP model that is continuously updated by the National Center for Atmospheric Research (NCAR) is the Weather Research and Forecasting (WRF) model. Two one-way nested domains at 8.1 km for the larger region and 2.7 km resolution for the small region around San Diego were initialized at 12 UTC (= 0400 PST) in WRF version 3.6. Each domain contains 100 grid points in both horizontal directions and 75 vertical levels, with 50 levels below 3 km; their locations are shown in Figure 1. Output was recorded at 15 min intervals, and the time step was 10 s. The model was run for 48 h in forecast mode.

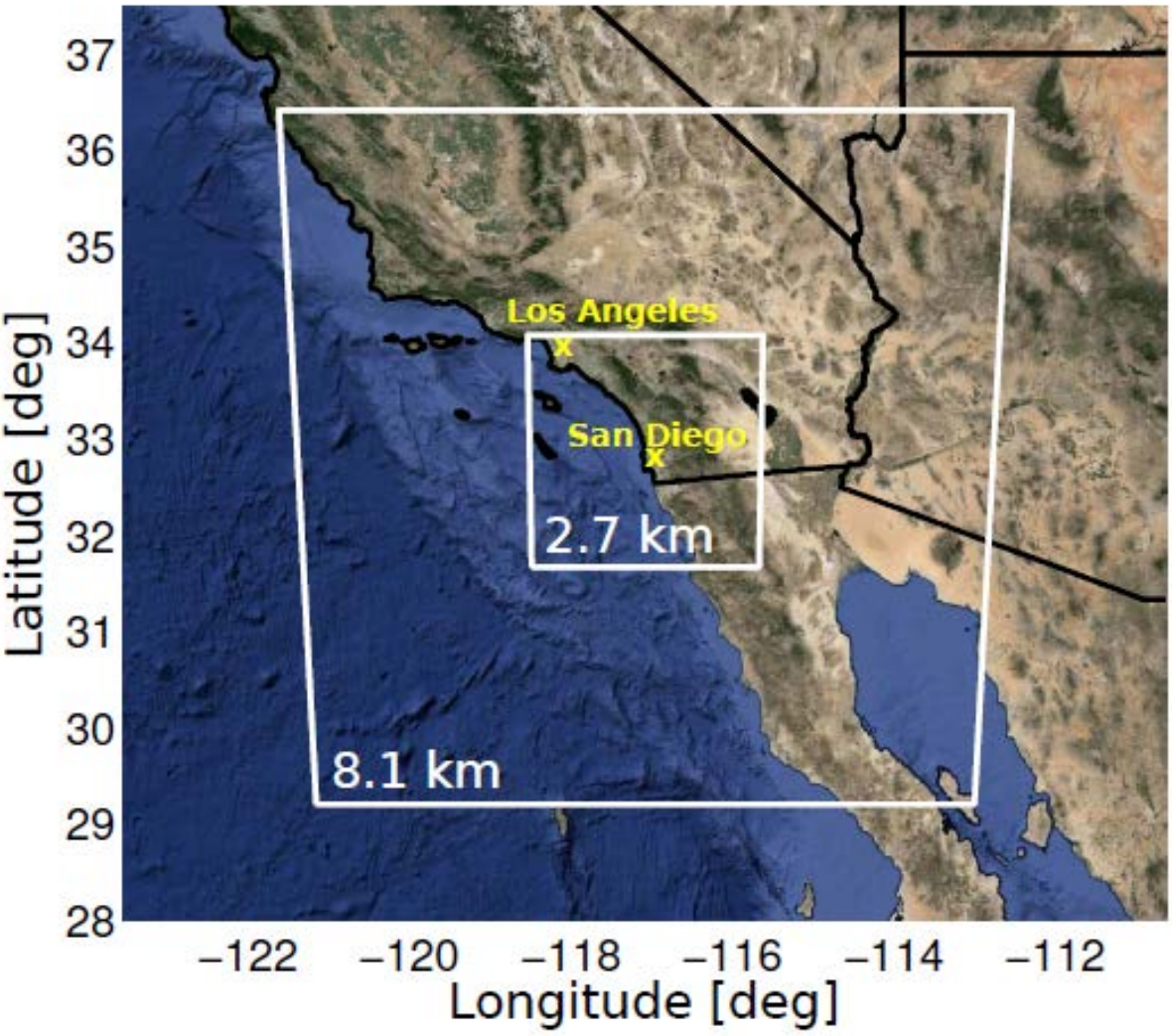

Figure 1: Outer and inner domains at 8.1 km (100x100x75 grid points) and 2.7 km (100x100x75) resolution. Satellite image © 2015 Google.

The model initial and boundary conditions are derived from the 0000 UTC North American Mesoscale (NAM) model. Here 0000 UTC NAM data is used for model initialization instead of the most recent 0600 UTC NAM data, because it was found to deliver lower errors. The physics parameterizations used in this study include Morrison 2-moment microphysics scheme, Goddard scheme for shortwave and longwave radiation parameterization, Kain-Fritsch New Eta scheme for cumulus parameterization, unified Noah land surface model and MYNN 2.5 scheme for planetary boundary layer parameterization.

4. Validation sites and metrics

4.1. Ground Stations

Measurements from ground stations maintained by SDG&E were used for validation. Four weather stations with LI-COR LI200 pyranometers were selected along a line from west to east. Station locations are shown in Figure 2. All stations sample GHI every 10 s and record 5 min averages.

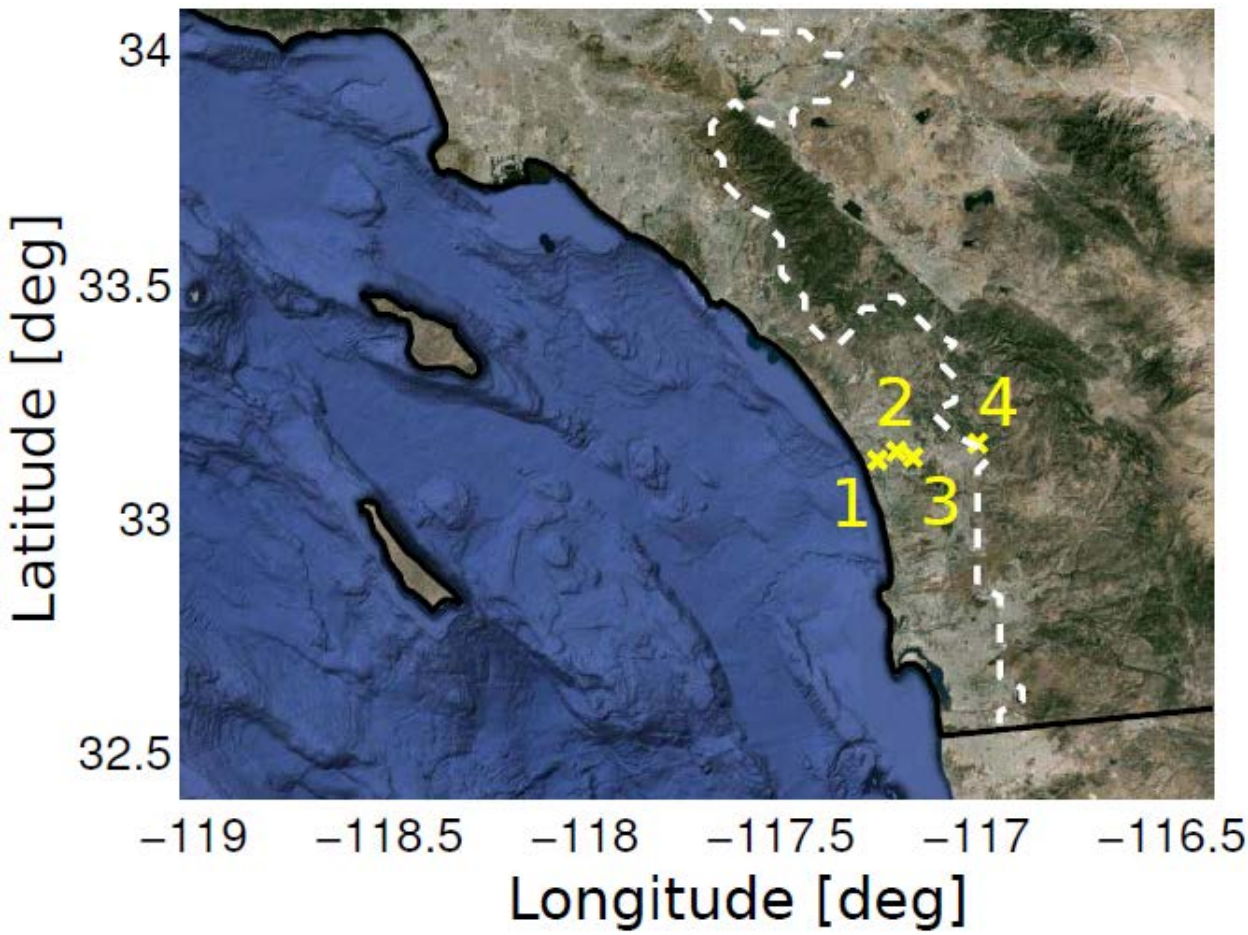

Figure 2: Locations of the four ground stations used for validation. Mount Soledad (MSD) is station 1 and Escondido (ESC) is station 4. The white dashed line marks the border of the coastal marine layer region, defined as the region where land elevation < 375 m MSL and monthly mean observed GHI is < 92% of clear sky GHI. Satellite image © Google.

4.2. Satellite solar resource data

Additional validation was performed against Clean Power Research's satellite data at 0.02° spatial and 30 min temporal resolution. Validation against 52 California Irrigation Management Information System (CIMIS) ground stations at hourly-averaged temporal resolution over the year 2010 showed a previous version of SolarAnywhere to have mean bias error (MBE) of 18.8 W m^{-2} and mean absolute error (MAE) of 45.6 W m^{-2}, an accuracy which is comparable to that of typical ground stations[1].

4.3. Forecast Metrics

The typical forecast error metrics of mean absolute error (MAE) and mean bias error (MBE) are used. The input data are (i) 10 min averages from the SDG&E ground stations MSD and ESC, (ii) 15 min resolution instantaneous WRF forecasts, and (iii) 30 min averages for SolarAnywhere. For the validation and estimation of error metrics all data

1 This work was funded under another CSI RD&D grant and is available here: Validation of SolarAnywhere Enhanced Resolution Irradiation in California. Jun 26, 2012. Jamaly, M. and J. Kleissl. Dept. of Mech. & Aerospace Eng. UC San Diego. Available at http://www.calsolarresearch.org/images/stories/documents/Sol1_funded_proj_docs/UCSD/solaranywh_val_cimis_isis_ed7.pdf

sets are converted to hourly averages. Relative errors such as the Mean Absolute Percentage Error (MAPE) are normalized by the mean daytime value.

Both errors of GHI and kt forecasts are computed. The GHI errors facilitate comparison of forecasts against 24-hr persistence forecasts and other forecasting methods in the literature. On the other hand, kt errors allow visualization of the spatial and temporal patterns of cloud cover in the San Diego region by normalizing out the time of day and locational dependence of solar irradiance. Since marine layer stratocumulus mostly occur during morning periods when the magnitude of GHI is low to medium, kt errors more clearly show cloud modeling issues than GHI errors, which are weighted towards solar noon.

5. Validation Results

Detailed validation results by month and by climate zone are provided in the Appendix. In this section only the annual average results are presented. The annual results in Figure 3 show that the accuracy of the original WRF forecast is on par with persistence near the coast and superior further inland. The correction through an Analog Ensemble (AnEn) create a forecast that is superior to persistence everywhere (Fig. A3). The original WRF is superior to persistence during the winter when synoptic scale frontal systems dominate the weather conditions (Figs. A1, A2). The analogue ensemble improves the most during months when the original WRF forecast is inaccurate compared to persistence. During the winter months at the coast and the summer months at the foothills, when the original WRF forecast accuracy resembles that of persistence, the analog ensemble improvements are small.

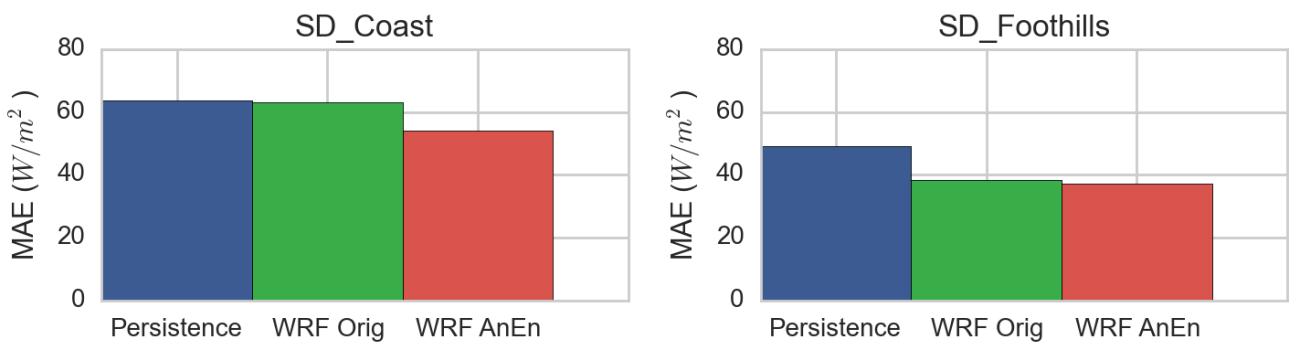

Figure 3: Mean Absolute Error for a coastal climate zone (left) and an inland climate zone (right) over one year of forecasting. WRF Orig is the raw WRF forecast and AnEn is the Analog Ensemble postprocessing.

6. Forecast improvements through Data Assimilation

6.1. Data Assimilation Setup

To improve the simulation results, three-dimensional variational (3DVAR) data assimilation is used within the WRF modelling framework WRF version 3.6.1 to improve the model initialization by assimilating the meteorological observations (such as temperature, dew point temperature, wind speed, wind direction and surface pressure) from the surface and upper-air in-situ observations over the southern California coast. The meteorological observations assimilated in 3DVAR data assimilation are collected from METAR airport weather stations, ships, and radiosonde networks.

3DVAR data assimilation uses observations from various sources and combines them with the short-range model forecast (called first guess) obtained from WRF simulation in order to provide the accurate state of the atmosphere at model initialization called the analysis. In addition to observations and first-guess, background error statistics (BES) also play a major role in a 3DVAR data assimilation system. In variational data assimilation BES are crucial for the quality of the analysis, because they determine to what extent the background field will be corrected to match the observations. Regional background error statistics are used here.

We selected two consecutive days with strong marine layer influence as in 11 and 12 of June 2015 to analyze the impact of 3DVAR data assimilation during model initialization on simulated GHI and meteorological parameters. According to the National Weather Service (NWS), San Diego report at 0800 PDT, 12 June satellite imagery showed marine layer clouds covering the coast and inland valleys. The 12 June 1200 UTC NKX Miramar sounding displayed a 7°C strong marine layer inversion, which supports low cloud persistence along the coast and inland valleys throughout most of the day. However, further inland areas cleared by late morning/early afternoon.

6.2. Validation

The model simulated GHI and associated meteorological parameters such as liquid water content (LWC) and near surface temperature and moisture are analyzed. GHI variations at coastal southern California regions during early summer are mostly affected by stratocumulus cloud cover, which reflects incoming solar radiation. The marine layer cloud optical thickness is directly proportional to the amount of LWC within the cloud layer. Thicker clouds at 1200 UTC generally take longer to dissipate and thus reduce the average GHI.

The NAM parent model to the WRF provides the initial temperature and moisture fields. NAM underestimates the inversion base height and the LWC at model initialization, which reduces marine layer cloud formation under an otherwise suitable synoptic situation. Hence, improved solar forecasting requires improving the initial profiles for the prognostic variables.

Potential temperature and relative humidity profiles at the NKX Miramar radiosonde are shown in Figure 4. The data assimilation runs are much closer towards the observations compared to the original WRF. Simulated profiles obtained from the data assimilation experiment with cyclic mode (WRFDA-cyc) provide the best match with observations. Yet the data assimilation results provide room for improvement with respect to the inversion base height. The data assimilation experiments produced a secondary inversion layer, which might cause thinner cloud layers. This is evident in the relative humidity profiles which are saturated only up to 400 m versus over 600 m in the observations.

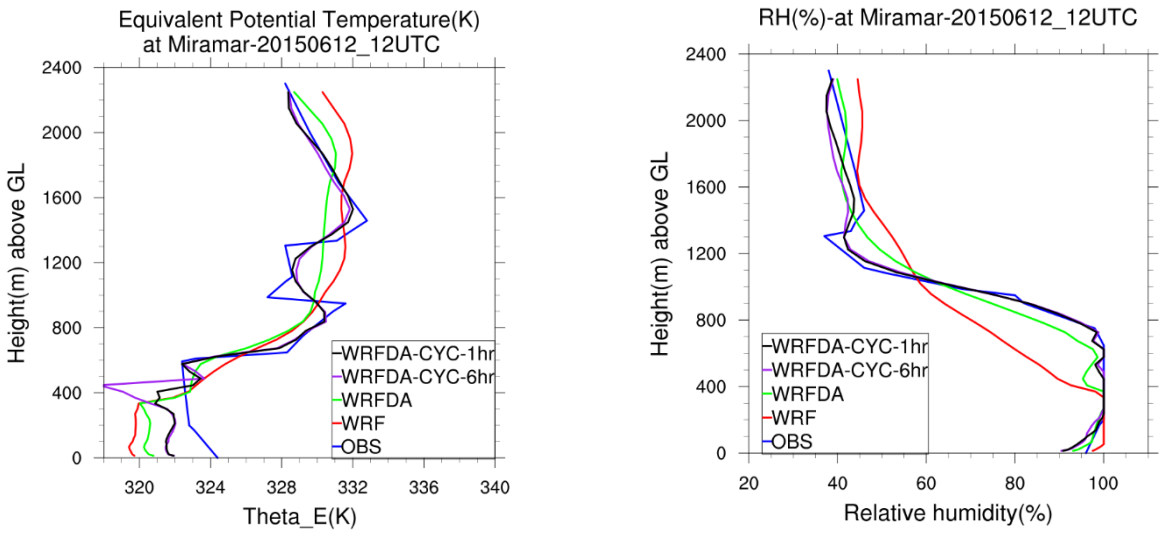

Figure 4: Validation of model simulated initial profiles for potential temperature (left) and Relative humidity (right) at 1200 UTC 12 June 2015 with NKX Miramar sounding. Red is the original WRF run without data assimilation and WRFDA-cyc is the final implementation of data assimilation.

The forecasted GHI from WRF and data assimilation experiments are validated against SDG&E pyranometer station observations and SolarAnywhere (SAW) gridded data. Figure 5 depicts the spatial biases between the forecasted GHI and SAW gridded GHI data sets. The GHI biases over the southern California coast are significantly less in WRFDA-cyc forecast, especially over San Diego County. The GHI biases are reduced by 100-200 W m^{-2} during the morning hours and up to 400-500 W m^{-2} during the noon hours in WRFDA-cyc simulation as compared to WRF simulation. It indicates that the marine layer clouds from the WRFDA-cyc simulation last longer over the coastal and inland valley regions as compared to WRF simulation. However, GHI biases are present in both WRF and WRFDA-cyc simulations during the early morning hours over regions with terrain heights greater than 350-400 m above mean sea level. This is likely a result of the model simulated inversion base heights being comparatively lower than the observation.

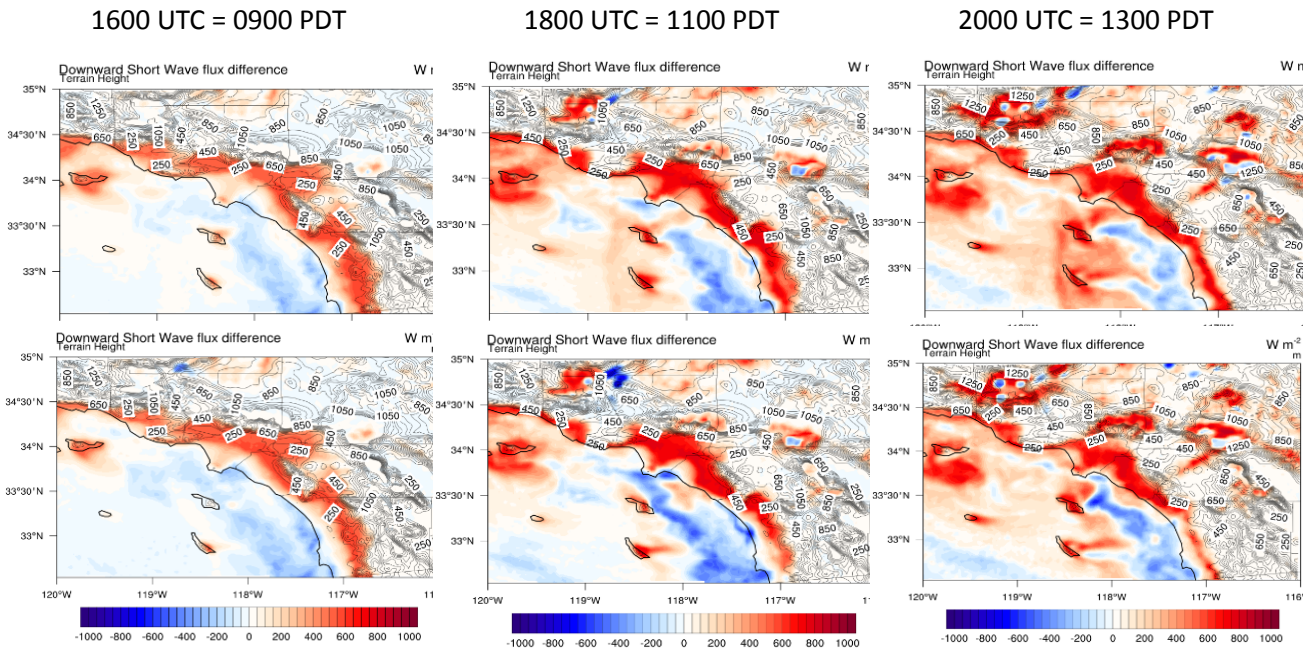

Figure 5: Spatial Biases in GHI between model simulations and SolarAnywhere (Model – SolarAnywhere) at 1600 UTC (left panel), 1800 UTC (center panel) and 2000 UTC (right panel) on 12 June 2015 for (top) WRF and (bottom) WRFDA-cyc simulations. The colorbar shows the GHI bias as red (overprediction) and blue (underprediction). Black contours show terrain height.

The diurnal variations of forecasted GHI from WRF and data assimilation experiments are also validated against SDG&E coastal station observations as shown in Figure 6. WRFDA-cyc outperformed WRF reducing MAE in GHI by 100-110 W m^{-2} as compared to WRF. Additionally, WRFDA-cyc outperformed persistence forecasts.

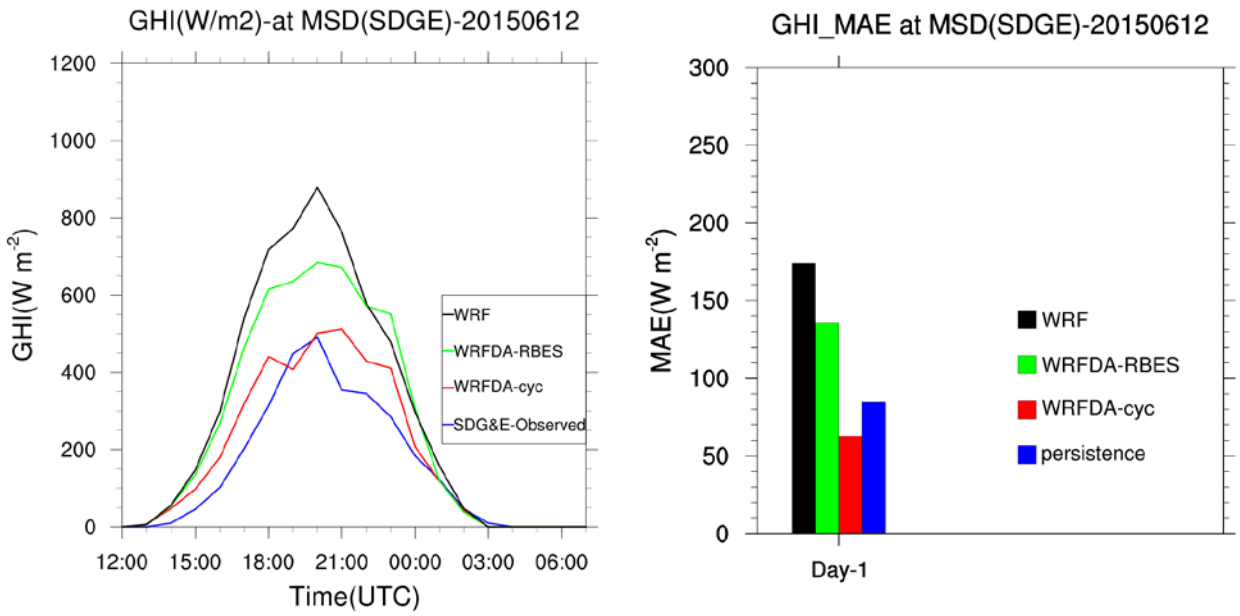

Figure 6: Comparison of diurnal variations of model forecasted GHI along with SDG&E station observations at MSD (Left) Time-series and (right) mean absolute error for diurnal variations. Hourly averages are applied to temporally align both the data sets.

7. Operational Reliability

Like with any operational product that is provided in a time sensitive production environment, one of the key success metrics is the reliable delivery of the product. In this specific case the distribution system operators relied on the timely solar forecast delivery as input to net load forecasting.

The original WRF forecasts were delivered on time 99% of the time (4 failures in one year). Failures occurred when the WRF forecast crashed presumably due to numerical stability issues in certain meteorological conditions. After the implementation of a redundant forecast delivery system that provides persistence forecasts from the previous day in case of a failure of the WRF run, the forecast were successfully delivered 100% of the time since January 2016.

8. Conclusions

An operational solar forecasting system was created for San Diego Gas & Electric Company. The purpose of the system was to provide solar generation inputs into system-wide net load forecasts and load flow analysis. Analysis of initial condition impacts and physics parameterizations pointed towards biases in the forecast related to the thickness of the marine boundary layer. Analog ensemble postprocessing was implemented to reduce forecast errors. The system was run from May 2014 to present. The errors in the original forecasts were typically larger than 24 hour persistence forecasts. The implementation of an analogue ensemble forecast system improved the errors. Data assimilation further reduced the errors and will be implemented in the operational forecast system in the near future.

Appendix A

The appendix shows more detailed validation results by climate zone and month.

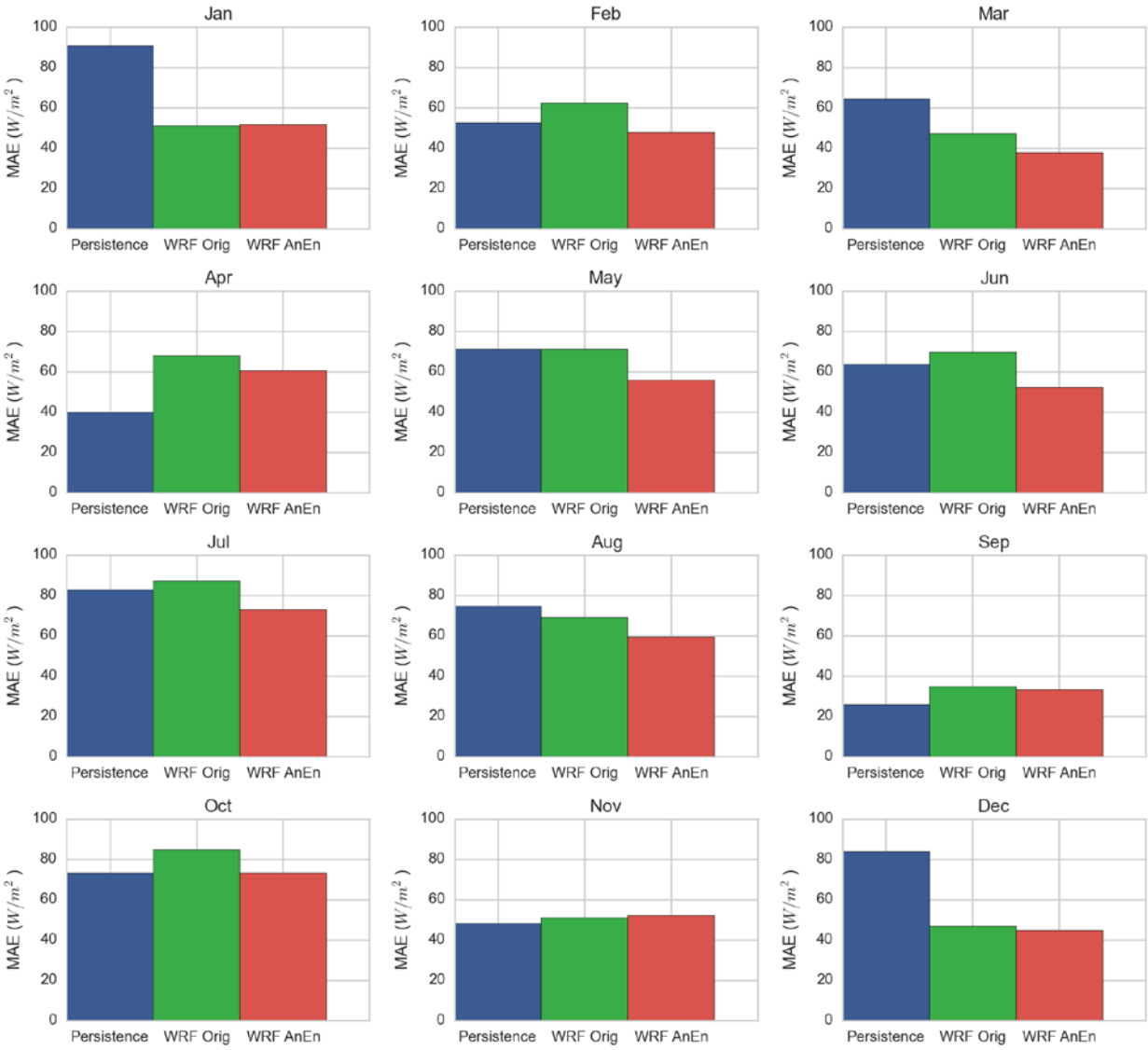

Figure A1: Month-by-month validation results for San Diego coast against SolarAnywhere for May 2014 to April 2015.

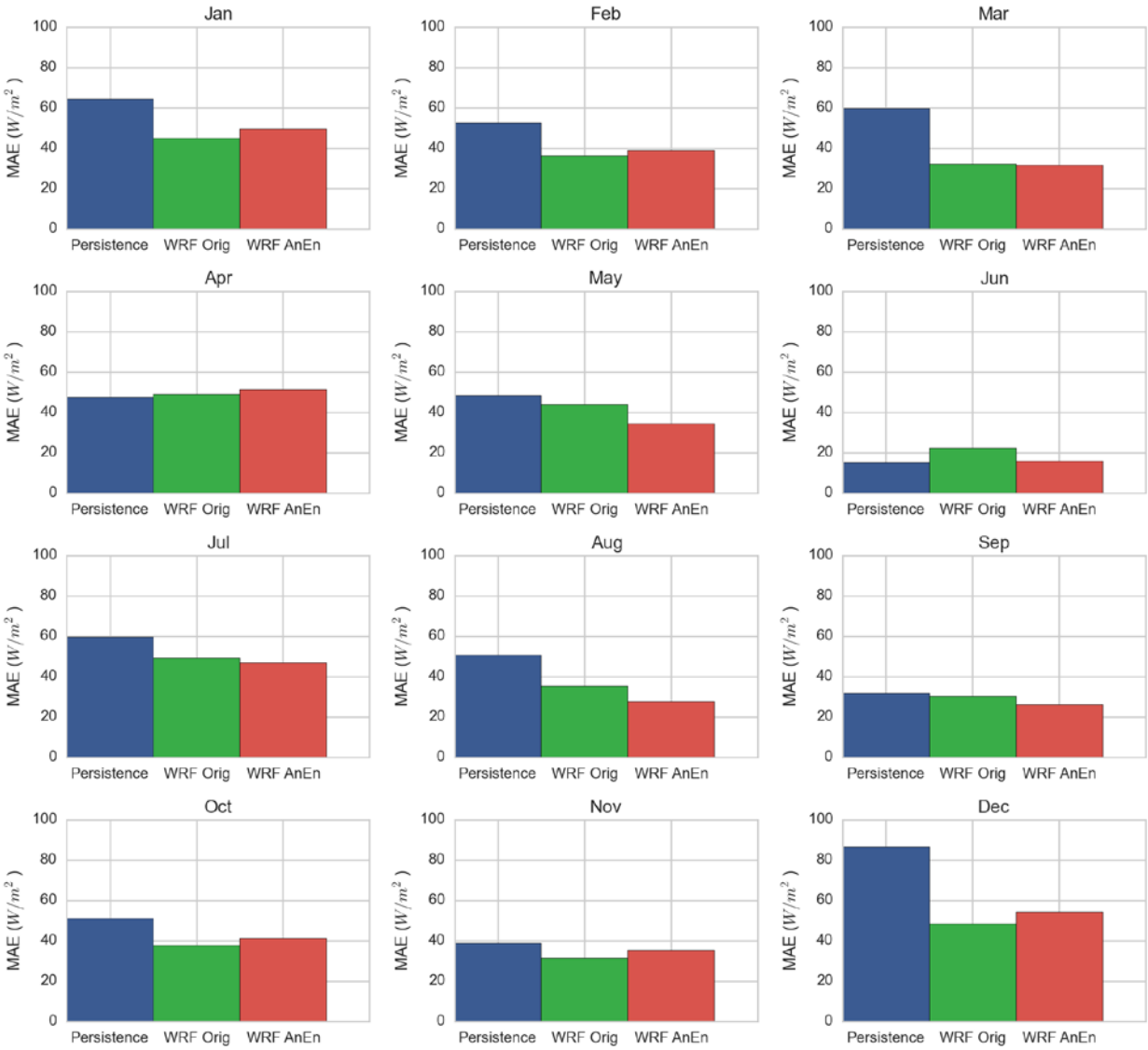

Figure A2: Month-by-month validation results for San Diego foothills against SolarAnywhere for May 2014 to April 2015.

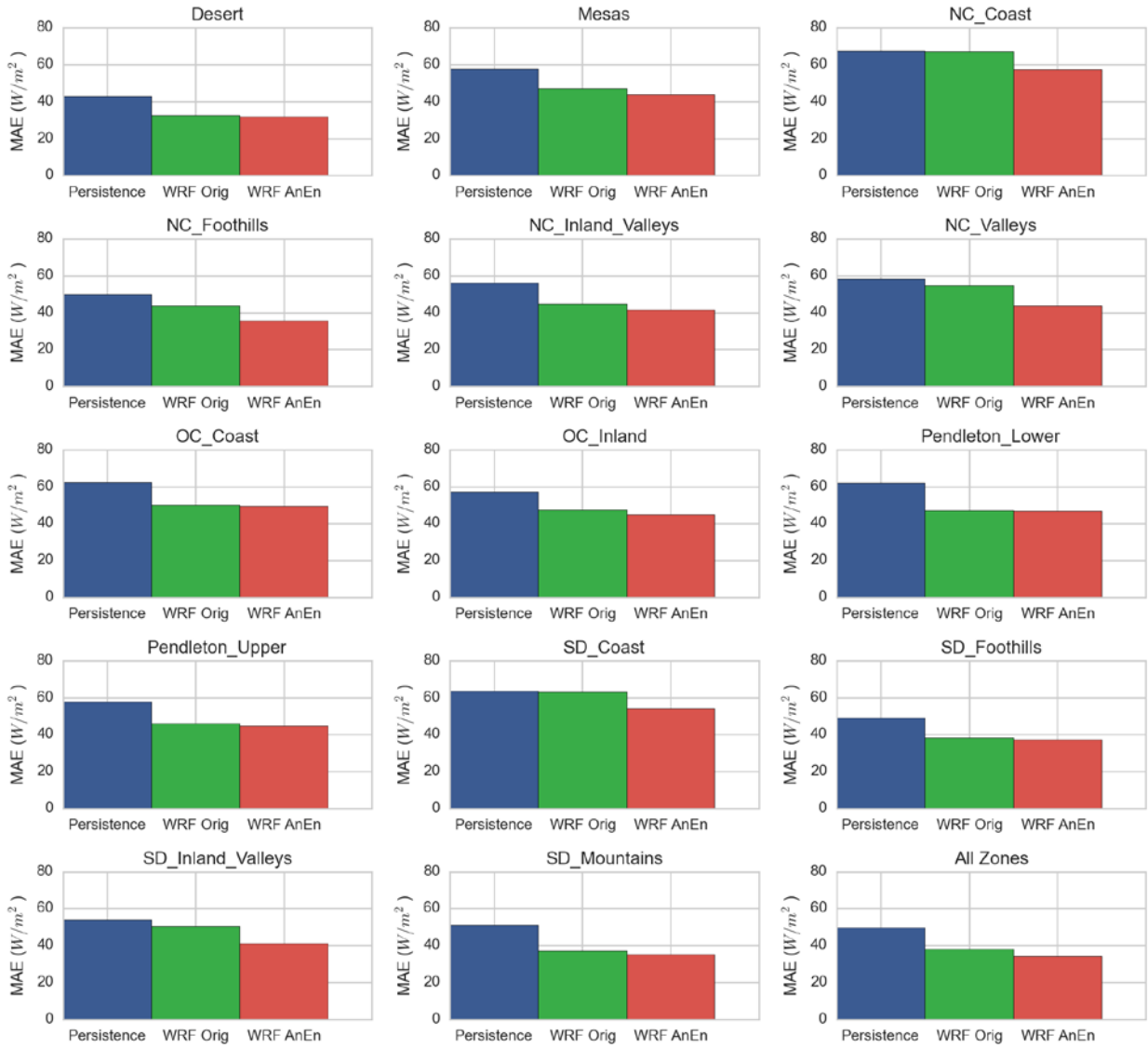

Figure A3: Annual validation results for all climate zones against SolarAnywhere for May 2014 to April 2015.