Energy Research and Development Division
FINAL PROJECT REPORT

VALIDATION OF THE NATIONAL SOLAR RADIATION DATABASE IN CALIFORNIA

Prepared for: California Energy Commission
Prepared by: University of California, San Diego

APRIL 2012
CEC-500-2014-017

PREPARED BY:

Primary Author(s):
Anders Nottrott
Jan Kleissi

University of California, San Diego
9500 Gillman Dr.
La Jolla, CA 92093

Contract Number: 500-08-017

Prepared for:

California Energy Commission

Prab Sethi
Contract Manager

Linda Spiegel
Office Manager
Energy Generation Research Office

Laurie ten Hope
Deputy Director
ENERGY RESEARCH AND DEVELOPMENT DIVISION

Robert P. Oglesby
Executive Director

DISCLAIMER

This report was prepared as the result of work sponsored by the California Energy Commission. It does not necessarily represent the views of the Energy Commission, its employees or the State of California. The Energy Commission, the State of California, its employees, contractors and subcontractors make no warranty, express or implied, and assume no legal liability for the information in this report; nor does any party represent that the uses of this information will not infringe upon privately owned rights. This report has not been approved or disapproved by the California Energy Commission nor has the California Energy Commission passed upon the accuracy or adequacy of the information in this report.

ACKNOWLEDGEMENTS

The authors thank Bekele Temesgen and Kent Frame of the California Department of Water Resources for their technical support. The authors are indebted to Richard Perez and his group at the State University of New York – Albany for providing comments and revisions, and to Stephen Wilcox, David Renne and Ray George of the National Renewable Energy Laboratory (NREL) for their valuable advice and comments. Additional thanks are due to David Sandwell of the Scripps Institute of Oceanography for his advice regarding technical issues related to satellite remote sensing and to Bryan Urquhart (UCSD).

PREFACE

The California Energy Commission Energy Research and Development Division supports public interest energy research and development that will help improve the quality of life in California by bringing environmentally safe, affordable, and reliable energy services and products to the marketplace.

The Energy Research and Development Division conducts public interest research, development, and demonstration (RD&D) projects to benefit California.

The Energy Research and Development Division strives to conduct the most promising public interest energy research by partnering with RD&D entities, including individuals, businesses, utilities, and public or private research institutions.

Energy Research and Development Division funding efforts are focused on the following RD&D program areas:

- Buildings End-Use Energy Efficiency
- Energy Innovations Small Grants
- Energy-Related Environmental Research
- Energy Systems Integration
- Environmentally Preferred Advanced Generation
- Industrial/Agricultural/Water End-Use Energy Efficiency
- Renewable Energy Technologies
- Transportation

Validation of the National Solar Radiation Database in California is the final report for the California Solar Energy Collaborative project (contract number 500-08-017) conducted by University of California, San Diego. The information from this project contributes to Energy Research and Development Division's Renewable Energy Technologies Program.

For more information about the Energy Research and Development Division, please visit the Energy Commission's website at www.energy.ca.gov/research/ or contact the Energy Commission at 916-327-1551.

ABSTRACT

The National Solar Radiation Database is often applied to quantify the amount of energy available from the sun, but its accuracy has not been validated in California. Satellite-derived global horizontal solar irradiance from the National Solar Radiation Database was compared to measurements from 27 weather stations in California during the years 1998-2005. The statistics of spatial and temporal differences between the two datasets were analyzed and related to meteorological phenomena.

The average mean bias error of the global horizontal solar irradiance data related to the National Solar Radiation Database indicated an overprediction of five percent if ground measurements were considered accurate. Year-round systematic positive mean bias errors in the database increased to 18 percent in proximity to the ocean. These errors increased up to 54 percent at coastal sites in the summer mornings. These differences were explained by a tendency for the database to overestimate global horizontal solar irradiance under cloudy conditions during the morning. A persistent positive evening mean bias error that was independent of site location and cloudiness occurred at all stations and was explained by an error in the time-shifting method applied in the database. A correction method was applied and a corrected database for California was published online.

Keywords: National Solar Radiation Database, solar irradiance, global horizontal irradiance

Please use the following citation for this report:

Nottrott, Anders; Jan Kleissi. (University of California, San Diego). 2012. ***Validation of the National Solar Radiation Database in California.*** California Energy Commission. Publication number: CEC-500-2014-017.

TABLE OF CONTENTS

Acknowledgements i

PREFACE ii

ABSTRACT iii

TABLE OF CONTENTS iv

LIST OF FIGURES v

LIST OF TABLES v

EXECUTIVE SUMMARY 1

Introduction 1

Project Purpose 1

Project Results 1

Project Benefits 2

CHAPTER 1: Introduction 3

CHAPTER 2: Data and Data Quality Control 5

2.1 Satellite Irradiance Data 5

2.2 Surface Irradiance Data 5

2.3 Cloud and Topographic Data 7

CHAPTER 3: Methodology 9

3.1 Statistical Error Metrics 9

3.2 Spatial and Geographic Analysis of Irradiance Data 11

3.3 The Clear-Sky Index (kt) 11

CHAPTER 4: Results 12

4.1 Annual Trends of Statistical Error Metrics 12

4.2 Monthly/Hourly Climatologies of Mean Bias Error (MBE) and Cloudiness (SCF) 13

4.3 Satellite-Ground Differences in Cloudy Conditions 15

4.4 Correcting the Satellite Derived Irradiance Data 16

CHAPTER 5: Discussion and Conclusions 20

Appendix A: Mean Bias Error Correction 1

LIST OF FIGURES

Figure 1: Map of California Illustrating the Geographic Distribution of the CIMIS Stations (Red) and Airports (Green) Included in the Analysis of this Report 6
Figure 2: Dependence of GHI and SCF on Site Distance from the Ocean Based on Annual Averages at Individual Sites 12
Figure 3: Annual Average of Hourly MBE between the SUNY Model and CIMIS Data against Station Distance from the Ocean (Same as 'MBE [%]' Column in Table 1) 13
Figure 4: MBE [%/100] between SUNY Model and CIMIS Data Averaged Hourly by Month for: (a) CIMIS Station #173, 0.5 km from the Ocean; (b) CIMIS Station #111, 7.6 km from the Ocean; (c) CIMIS Station #008, 93.0 km from the Ocean; (d) CIMIS Station #125, 111.7 km from the Ocean. 14
Figure 5: Mean SCF at (a) North Island NAS which Experiences Similar Weather Conditions to CIMIS Station #173; (b) Watsonville Muni Airport Near CIMIS Station #111; (c) Red Bluff Muni Airport Near CIMIS Station #008; (d) Meadows Field Near CIMIS Station #125 15
Figure 6: Scatter Plot of Hourly CIMIS kt versus SUNY kt Predicted for the Same Location at: (a) CIMIS #173 Jun-Sep; (b) CIMIS #125 Nov-Feb 16
Figure 7: Instantaneous Hourly MBE [%/100] as a Function of Cos(SZA)*Sign(AZ-180°) and ktSUNY Averaged for 25 Ground Stations 17
Figure 8: MBE [%/100] between SUNY Model and CIMIS Station #173 Averaged Hourly by Month (a) before Applying the Correction Algorithm; (b) after Applying the Correction Algorithm 18

LIST OF TABLES

Table 1: Comparison of Satellite Derived GHI Data and GHI Data Measured at Ground Station Sorted by Site Distance from the Ocean 10
Table 2: Effect of the Correction Algorithm on SUNY Model Annual MBE. Negative Values in the "Percentage Point Δ" Columns Indicate That the Correction Made the Error Worse for a Particular Site 19

EXECUTIVE SUMMARY

Introduction

Understanding the energy potential from solar resources is critical to determining the expected energy output and financial viability of solar energy projects. The National Solar Radiation Database is often applied to quantify the amount of energy available from the sun. The database contains a high-resolution, satellite-derived gridded image of the radiation reaching the earth's surface in a number of different ways. Global horizontal irradiance is the total amount of shortwave radiation received from above by a surface horizontal to the ground. This value is critical to photovoltaic installations and includes both direct normal irradiance and diffuse horizontal irradiance. Direct normal irradiance is the solar radiation that comes in a direct line from the direction of the sun at its current position in the sky. Diffuse horizontal irradiance is solar radiation that does not arrive on a direct path from the sun, but has been scattered by molecules and particles in the atmosphere and comes equally from all directions. The data collected in the National Solar Radiation Database covers the entire United States from 1997 – 2005.

Project Purpose

The accuracy of the National Solar Radiation Database has never been validated in California. The goal of this project was to compare satellite derived global horizontal irradiance from the National Solar Radiation Database to measurements from 27 weather stations in California during the years 1998-2005 and to analyze the statistics of spatial and temporal differences between the two datasets and relate them to meteorological phenomena.

Project Results

The National Solar Radiation Database satellite data was generally accurate and provided high quality irradiance data with an average uncertainty level of five percent based on global horizontal irradiance data collected in California. Global horizontal irradiance near coastal stations was overestimated, particularly in the mornings during the summer when errors reached up to 54 percent. On summer mornings a coastal marine inversion layer that creates overcast conditions was present nearly every day. The National Solar Radiation Database was accurate under clear sky conditions but in broken or overcast cloud cover global horizontal irradiance was often overestimated.

Inaccuracies in the National Solar Radiation Database during the late afternoon were also found state-wide. Since this error was not observed in the morning it was not related to problems with computing the diffuse component of the irradiance at low sun altitude. The error was related to a programming problem that was corrected and applied to the database to mitigate the observed error patterns. The correction improved the database at the majority of the sites, but further validation of the correction model must be conducted before it can be applied universally.

The National Solar Radiation Database satellite-derived irradiance model provided accurate estimations of surface radiation for most sites even though errors were found under certain conditions. The National Solar Radiation Database was also well-suited for assessing solar

resources because of its complete coverage and high spatial resolution. Nevertheless there is room for improvement and before this database is applied it should be verified against available ground measurements of irradiance, especially in areas where persistent cloudiness or unusual weather or ground conditions create errors.

Project Benefits

Understanding the nuances of the radiation available from the sun is important to Californians as more renewable energy technologies gain market support. Some technologies are better suited for particular areas than others and knowing in advance the energy resources available to a particular region is critical to renewable energy planning and siting.

CHAPTER 1: Introduction

Rapid coastal urbanization coupled with increasing energy demands, and the desire for environmentally sustainable solutions to meet these challenges, necessitate the expansion of renewable energy production in load centers. In low latitude urban areas solar photovoltaic (PV) is generally the most attractive renewable energy option due to large resources and peak capacity factors. Obtaining accurate irradiance data at high spatial resolution is particularly important in California where a large solar resource is collocated with high population density, stringent air pollution standards and ever increasing energy demands and costs. These factors make PV energy production both economically viable and vital to sustainable development. However, large PV penetration requires accurate, site specific estimates of power output for transmission and distribution planning and economic reasons. Preferably measured solar radiation and meteorological data over long time periods would be available on site to evaluate the solar resource. However, solar radiation is typically not measured at standard weather stations. Although a few specialized sensor networks that measure solar irradiance exist, they are often too sparse and interpolating data between sites may not be appropriate[1 2].

Satellite derived irradiance measurements overcome the poor spatial resolution of ground stations providing continuous coverage over large geographic areas at high resolution relative to in situ sensor networks[1 2 3 4 5 6]. In this report satellite derived irradiance data from the SUNY model[2] in the National Solar Radiation Database (NSRDB) are compared to data collected from ground stations throughout the state of California. This analysis is based on similar work conducted in other geographic regions[2 7 8 9]. The SUNY predictions were attributed a minimum

---

1 Cano, D., Monget, J. M., Albuisson, M., Guillard, H., Regas, N. and Wald, L. 1986. A Method for the determination of the global solar radiation from meteorological satellite data. *Solar Energy* **37**(1), 31-39.

2 Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., and Vignola, F. 2002. A new operational model for satellite-derived irradiances: Description and validation. *Solar Energy*. **73:5**, 307-317

3 Schmetz, J. 1989. Towards a surface radiation climatology: retrieval of downward irradiances from satellites. *Atmospheric Research* **23**, 287-321.

4 Zelenka, A., Perez, R., Seals, R., Renné, D. 1999. Effective accuracy of satellite-derived hourly irradiances. *Theoretical and Applied Climatology* **62**(3-4), 199-207

5 Pereira, E.B., Martins, F.R., Abreu, S.L. and Rüther, R. 2006. Brazilian atlas for solar energy. *Artes Gráficas Editoria* **67**, 60.

6 Martins, F.R., Pereira, E.B., Abreu, S.L. 2007. Satellite-derived solar resource maps for Brazil under SWERA project. *Solar Energy* **81**(4), 517-528

7 Muneer, T. 1997. Perez slope irradiance and illuminance models: evaluation against Japanese data. *Lighting Research and Technology* **29**(2), 83-87.

uncertainty of 12 percent for global horizontal irradiance (GHI) in the NSRDB manual[10]. The results of previous validation surveys indicate that the SUNY model provides 5 percent[11], ±10 percent[12] and -21 percent to 31 percent[13] accurate estimations of surface GHI (based on the mean bias error), but it has a tendency to produce large errors in regions where unusual meteorological phenomena exist. Examples are persistent clouds as in the "Eugene, OR, Syndrome" found by Gueymard and Wilcox[12]; low clouds and snow cover found by Vignola *et al.*[11]; and a 31 percent MBE at a coastal Florida site found by Perez *et al.*[13], which was related to the humid subtropical climate at the site and the fact that the satellite pixel covers both ocean and land surfaces with very different albedos. In the United States, California has the largest installed PV capacity, but coastal meteorology creates large spatial gradients in cloudiness motivating this validation of the SUNY model. Our study takes a unique approach to evaluate systematic errors that exist in the SUNY model. For the first time comprehensive climatologies of the SUNY error for different times of day are presented and corrections to systematic errors observed in the SUNY data are applied.[14]

---

8 Vignola, F., Harlan, P., Perez, R., Kmiecik, M. 2007. Analysis of satellite derived beam and global solar radiation data. *Solar Energy* **81**(6), 768-772.

9 Gueymard, C. and Wilcox, S. 2009. Spatial and temporal variability in the solar resource: assessing the value of short-term measurements at potential solar power plant sites. *Solar 2009 Conf., Buffalo, NY, American Solar Energy Soc. (ASES)*

10 Wilcox, S. 2007. National solar radiation database 1991-2005 update: user's manual. NREL/TP-581-41364. <http://rredc.nrel.gov/solar/old_data/nsrdb/1991-2005/> Accessed July 30th, 2009.

11 Vignola, F., Harlan, P., Perez, R., Kmiecik, M. 2007. Analysis of satellite derived beam and global solar radiation data. *Solar Energy* **81**(6), 768-772.

12 Gueymard, C. and Wilcox, S. 2009. Spatial and temporal variability in the solar resource: assessing the value of short-term measurements at potential solar power plant sites. *Solar 2009 Conf., Buffalo, NY, American Solar Energy Soc. (ASES)*

13 Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., Vignola, F. 2002. A new operational satellite-to-irradiance model. *Solar Energy* **73**(5), 307-317.

14 Nottrott A, Kleissl J, Validation of the SUNY NSRDB global horizontal irradiance in California, *Solar Energy*, 84:1816–1827, 2010

CHAPTER 2:
Data and Data Quality Control

2.1 Satellite Irradiance Data

Two independent datasets were used in this analysis. Satellite derived GHI values from the National Solar Radiation Database (NSRDB-SUNY) were compared with ground measurements from meteorological stations in the California Irrigation Management and Information System (CIMIS). Both datasets were given with a precision of 1 W m^{-2}. The NSRDB-SUNY dataset is based on a model developed at the State University of New York – Albany[15]. The model uses visible images from Geostationary Operational Environmental Satellites (GOES) to develop estimates of the cloud index (CI) for each pixel. The CI is then used in a transmittance function that is applied to the modeled clear sky irradiance for each pixel. The SUNY model also accounts for effects of atmospheric turbidity, ground snow cover, ground specular reflectance characteristics and individual pixel sun-satellite angle effects. Atmospheric turbidity is quantified in terms of the Linke Turbidity coefficient which is a function of monthly average atmospheric aerosol content, water vapor and ozone[16]. The model was run between 1998 and 2005 to generate hourly global horizontal, diffuse and direct irradiance values for the entire United States on a 0.1° node registered grid, corresponding to a grid spacing of about 10 km in California. This analysis used hourly GHI values with an hour ending timestamp from the 'Sglo' column in the NSRDB-SUNY database. These data are modeled from on the hour (e.g. 1200) irradiance "snap shots" derived from GOES visible images[17].

2.2 Surface Irradiance Data

Data from 27 ground stations (see Fig. 1 and Table 1) across the state of California were used for validation. The California Department of Water Resources operates the CIMIS network of meteorological monitoring stations distributed throughout California. Only CIMIS data concurrent to the SUNY dataset (i.e. 1998-2005) were included. GHI is measured at 26 CIMIS stations using a Li-Cor LI200SZ silicon photodiode pyranometer and recorded as the hourly average (with an hour ending time stamp) of 60 GHI measurements made within the hour. The sensors are recalibrated annually by the manufacturer with an expected maximum absolute

---

15 Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., Vignola, F. 2002. A new operational satellite-to-irradiance model. *Solar Energy* **73**(5), 307-317.

16 Ineichen P. and Perez R. 2002. A new airmass independent formulation for the Linke turbidity coefficient. *Solar Energy* **73**(3), 151–157.

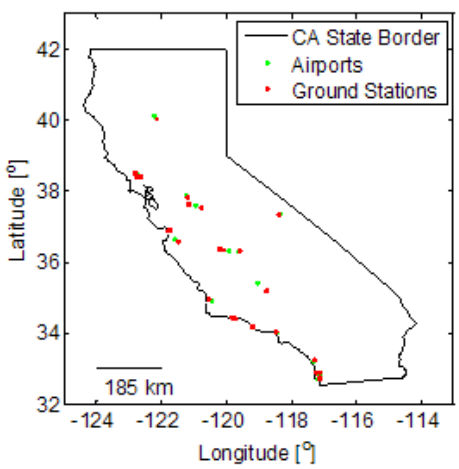
17 Wilcox, S. 2007. National solar radiation database 1991-2005 update: user's manual. NREL/TP-581-41364. <http://rredc.nrel.gov/solar/old_data/nsrdb/1991-2005/> Accessed July 30th, 2009.

error of ±5 percent[18]. In a recent study of pyranometer calibration accuracy Myers[19] found that these instruments have a maximum absolute error closer to ±8 percent.

Some additional limitations apply to photodiode type pyranometers. The Li-Cor 200SZ pyranometer has a non-linear cosine response for incident angles greater than 80° which creates errors in the irradiance measurements when the SZA>80°[20]. For this reason all ground measurements made in the early morning when SZA>80° at the top of the hour and in the late evening when SZA>80° at the end of the hour have been removed. The Li-Cor pyranometer does not have a broadband spectral response and the factory calibration extrapolates irradiance measurements to cover the entire solar spectrum. The error of the sensor will increase by a few percent under cloudy conditions because the spectrum of incident solar radiation is different from that under clear sky conditions. However, the much larger errors observed in this study cannot be explained by the spectral response under cloudy conditions.

Fifteen of these stations were located at coastal sites (<15 km from the ocean) and 12 were located at inland sites (>15 km from the ocean). Data from the Hanford Muni station (see Table 1) which is operated by the National Oceanic and Atmospheric Administration (NOAA) under the Integrated Surface Irradiance Study (ISIS) were also analyzed. Data from the Hanford ISIS station were used during the production of the NSRDB dataset to validate SUNY derived irradiances for the western United States.

Figure 1: Map of California Illustrating the Geographic Distribution of the CIMIS Stations (Red) and Airports (Green) Included in the Analysis of this Report



18 CIMIS 2008a. *Sensor Specs – Total Solar Radiation (pyranometer).* <http://wwwcimis.water.ca.gov/cimis/infoStnSensorSpec.jsp> Accessed July 20th, 2009.

19 Myers, D. R. 2010. Seasonal variation in the frequency distributions of differences between radiometric data for solar resource assessment applications. *Solar 2010 Conf., Phoenix, AZ, American Solar Energy Soc. (ASES)*

20 LI-COR 2005. LI-COR Terrestrial Radiation Sensors Instruction Manual. © 2005, LI-COR Biosciences, Inc.

Quality control (QC) for all CIMIS stations is conducted based on the methods described by Meek and Hatfield[21]. QC output for each CIMIS station is monitored by local personnel to detect potential sensor malfunction[22]. CIMIS data were also tested by comparing the ratio of measured irradiance to the solar constant (RQC) and the solar zenith angle (SZA) for each hour. Data is flagged if any one of the following criteria is met: SZA<80° and either RQC>1.0 or GHI=0, SZA>80° and either RQC>0.85 or GHI≥6 W m^{-2} [23]. A detailed description of CIMIS data QC procedures can be found in the CIMIS technical manual[24]. All flagged CIMIS data were excluded from this analysis.

Since the CIMIS data QC alone is insufficient to ensure accuracy of the data for solar resource applications, careful additional QC was conducted by the authors. CIMIS data were also removed if the "upper envelope" of the maximum hourly values of the CIMIS measurements differed from the upper envelope of the maximum value of the modeled clear sky irradiance by more than 8 percent for a period of two months or more. The clear sky irradiance is derived using a geometric model after Synder and Eching[25]. A careful examination of the diurnal cycles of GHI showed that station #66 was temporarily shaded until 1100 PST year-round. Station #107 was shaded only during the winter mornings. This shading was likely caused by nearby large obstacles. In both cases data collected during these times of day were excluded from the analysis. The minimum length of a CIMIS QCed data timeseries was required to be one year. On average five years of QCed data were available at each station. SUNY data were excluded when CIMIS data were missing or excluded, so that a comparison could be conducted on two data vectors of identical length and time stamp. Despite careful data QC, the data quality issues do not allow firm conclusions from comparisons of individual CIMIS sites with SUNY data. However, consistent trends at several sites and especially relative trends over a year or time of day are expected to indicate fundamental problems with the SUNY data.

2.3 Cloud and Topographic Data

The National Climatic Data Center (NCDC), Integrated Surface Dataset (DSI-3505) was used to analyze the temporal variability of sky cover fraction (SCF) at airports near the CIMIS stations. Hourly SCF is provided using four descriptors that correspond to the amount of sky that is covered by opaque clouds measured in octas. Clear (CLR, SCF=0) indicates no cloud cover,

---

21 Meek, D.W. and Hatfield, J.L. 1994. Data quality checking for single station meteorological databases. *Agricultural and Forest Meteorology* **69**(1-2), 85-109.

22 CIMIS 2008b. *QC overview*. <http://wwwcimis.water.ca.gov/cimis/dataQc.jsp> Accessed July 20th, 2009.

23 CIMIS 2008b. *QC overview*. <http://wwwcimis.water.ca.gov/cimis/dataQc.jsp> Accessed July 20th, 2009.

24 Eching, S.O and Moellenberndt, D. 1998. Technical elements of CIMIS, the California Irrigation Management Information System: State of California, Resources Agency, Dept. of Water Resources, Division of Planning and Local Assistance, 66.

25 Snyder, R.L. and Eching, S. 2002. Penman-Monteith (hourly) reference evapotranspiration equations for estimating ETos and ETrs with hourly weather data. <http://biomet.ucdavis.edu/Evapotranspiration/PMhrXLS/PMhrDoc.pdf> Accessed January 13th, 2010.

while scattered (SCT, 1/8 to 4/8, SCF=0.31), broken (BKN, 5/8 to 7/8, SCF=0.75) and overcast (OVC, 8/8, SCF=1.0) indicate fractional sky coverage. The average SCF was assigned to each indicator to obtain a numerical dataset. The resulting data contain more than four discrete values because the data were interpolated in time to correspond to the time stamp of the CIMIS data. The data in the DSI-3505 dataset have undergone extensive automated quality control. Additional manual quality control is performed at all US Air Force, US Navy and US National Weather Service stations[26]. Only unflagged cloud cover data for the period 1998-2005 coinciding with CIMIS time stamps were used in this analysis. The SCF analysis was conducted for every CIMIS station using sky cover data from the nearest airport to the CIMIS site.

26 NCDC 2008. *Data documentation for data set 3505 (DSI-3505) integrated surface data*. Ashville, NC. <http://idn.ceos.org/> Accessed November 17th, 2009.

CHAPTER 3: Methodology

3.1 Statistical Error Metrics

Statistical quantities describing spatial and temporal variability were used to compare SUNY GHI values with CIMIS measurements, in particular mean absolute error (MAE, Eq. 1), mean bias error (MBE, Eq. 2), root mean square error (RMSE, Eq. 3) and correlation coefficient (r, Eq. 4).

$$MAE = (1/N)\sum_{n=1}^{N}\left|GHI_{SUNY,n} - GHI_{CIMIS,n}\right| \tag{1}$$

$$MBE = (1/N)\sum_{n=1}^{N}\left(GHI_{SUNY,n} - GHI_{CIMIS,n}\right) \tag{2}$$

$$RMSE = \left((1/N)\sum_{n=1}^{N}\left(GHI_{SUNY,n} - GHI_{CIMIS,n}\right)^2\right)^{0.5} \tag{3}$$

$$r = \frac{\sum_{n=1}^{N}(GHI_{SUNY,n} - <GHI_{SUNY}>)(GHI_{CIMIS,n} - <GHI_{CIMIS}>)}{\left(\sum_{n=1}^{N}(GHI_{SUNY,n} - <GHI_{SUNY}>)^2 \sum_{n=1}^{N}(GHI_{CIMIS,n} - <GHI_{CIMIS}>)^2\right)^{0.5}} \tag{4}$$

In Eqs. 1-4 *GHI* denotes an hourly GHI value from the specified dataset, *N* is the total number of data points, and <…>denotes temporal averaging. Relative MAE, MBE, and RMSE were also computed by normalizing Eqs. 1-3 by $<GHI_{CIMIS}>$, where the average is computed over the same period for which the error is computed. For example, for the yearly error all CIMIS GHI values are averaged to normalize by $<GHI_{CIMIS}>_{year}$ (e.g. Table 1) and for the error for an hour of a month (e.g. Fig. 4) all CIMIS GHI for that hour are averaged to normalize by $<GHI_{CIMIS}>_{hour}$. In all cases individual measurements from the two datasets were compared at the same time rather than evaluating long term average GHI. This prevents potential year-over-year bias in comparing the databases.

Table 1: Comparison of Satellite Derived GHI Data and GHI Data Measured at Ground Station Sorted by Site Distance from the Ocean

Station Location	CIMIS Station ID#	Length of Record [Months]	Station Distance from Ocean [km]	CIMIS Mean Irradiance [W m^{-2}]	SUNY Mean Irradiance [W m^{-2}]	MBE [W m^{-2}]	MBE [%]	MAE [W m^{-2}]	MAE [%]	RMSE [W m^{-2}]	RMSE [%]	r value [-]	CIMIS Latitude [°]	CIMIS Longitude [°]	CIMIS Altitude [m]	Distance Between CIMIS and Airport [km]
Torrey Pines	173	58	0.5	188	202	35	18	66	35	94	50	0.97	32.90	-117.25	102	22.5
Port Hueneme	97	23	2.0	213	212	-1	0	51	24	79	37	0.98	34.17	-119.20	5	3.5
Watsonville West	177	58	2.4	188	197	24	13	60	32	87	46	0.97	36.90	-121.81	65	4.8
San Diego II	184	29	3.2	205	211	16	8	58	28	94	46	0.97	32.73	-117.14	115	4.7
San Diego I	66	21	3.2	275	282	14	5	57	21	92	33	0.98	32.73	-117.14	113	4.7
Santa Monica	99	54	3.5	206	211	12	6	51	25	82	40	0.98	34.04	-118.48	104	3.5
Santa Barbara	107	61	4.1	233	242	20	9	61	26	94	40	0.97	34.44	-119.74	76	9.3
Oxnard	156	49	4.6	202	207	16	8	54	27	79	39	0.98	34.23	-119.20	15	3.5
Pajaro	129	95	5.7	195	198	10	5	55	28	84	43	0.98	36.90	-121.74	20	6.3
Goleta Foothills	94	78	6.0	205	213	23	11	63	31	101	49	0.97	34.47	-119.87	195	5.2
Green Valley Road	111	93	7.6	192	198	17	9	57	30	86	45	0.98	36.94	-121.77	34	1.8
Oceanside	49	59	8.6	206	213	18	9	50	24	77	38	0.98	33.26	-117.32	15	5.2
Guadalupe	120	22	9.2	203	208	13	6	56	28	89	44	0.97	34.96	-120.55	34	11.3
Miramar	150	70	10.5	213	214	3	1	48	23	78	37	0.98	32.89	-117.14	136	2.2
Manteca	70	95	13.1	205	206	7	3	44	21	67	33	0.98	37.84	-121.22	10	5.8
Santa Rosa	83	95	19.5	198	195	-3	-1	47	24	75	38	0.98	38.40	-122.80	24	12.2
Bennett Valley	158	45	23.7	200	200	0	0	42	21	65	32	0.98	38.42	-122.66	82	16.5
Windsor	103	60	26.2	196	196	3	1	47	24	77	39	0.98	38.53	-122.83	26	2.8
Salinas South	89	56	26.2	200	207	18	9	55	28	79	39	0.98	36.61	-121.53	37	9.1
Modesto	71	24	33.8	205	209	15	7	44	21	73	36	0.98	37.65	-121.19	11	22.2
Denair	168	40	61.9	203	206	10	5	43	21	70	34	0.98	37.56	-120.78	43	16.1
Gerber	8	95	93.0	196	200	13	6	48	25	74	38	0.98	40.05	-122.16	76	13.5
Arvin-Edison	125	94	111.7	213	218	12	6	51	24	80	37	0.98	35.21	-118.78	152	35.3
Five Points South West	190	21	118.3	213	214	8	4	46	22	75	35	0.98	36.39	-120.24	82	26.9
Five Points	2	92	122.2	212	216	12	6	46	22	70	33	0.98	36.34	-120.11	87	14.4
Hanford Muni NOAA ISIS	210	83	150.9	212	210	-4	-2	42	20	67	32	0.98	36.32	-119.62	75	1.2
Bishop	35	89	263.0	230	226	-8	-4	58	25	88	38	0.98	37.36	-118.40	1271	2.9
Average Value		61		208	212	11	5	52	25	81	39	0.98				9.9

Statistical descriptors are defined in Eqs. 14. While the mean irradiance columns give the average over a 24 hour day, the other statics were only computed during hours when SZA<80°.

Note that the large mean irradiance values for CIMSIS stations #66 and #107 are a result of filtering out data because of morning shading at these sites.

3.2 Spatial and Geographic Analysis of Irradiance Data

There is a spatial discrepancy between the two datasets because the SUNY data are representative for regularly spaced 0.1° by 0.1° grid cells while the CIMIS data are representative only of the measurement site. In order to compensate for this discrepancy an algorithm was developed based on the approach of Vincenty[27] to determine the straight line distance between SUNY grid points and each CIMIS station. A distance weighted interpolation was then used to interpolate GHI from the four nearest SUNY grid points to each CIMIS station

$$GHI_{SUNY,I} = \sum_{i=1}^{4} \frac{1}{d_i} GHI_{SUNY,i} \Big/ \sum_{i=1}^{4} \frac{1}{d_i} \qquad (5).$$

In Eq. 5 d_i is the distance between the CIMIS station and the SUNY grid point i, $GHI_{SUNY,i}$ is the irradiance at the SUNY grid point and $GHI_{SUNY,I}$ is the SUNY interpolated irradiance at the CIMIS station.

CIMIS station distance from the ocean is used as an independent variable to quantify the effect of the coastal marine layer on different sites. Although distance from the ocean is not the sole factor describing the prevalence of coastal climate at a location, it does provide a relevant and simple metric to organize our results.

3.3 The Clear-Sky Index (kt)

Large differences in cloudiness frequently occur over short distances. Although the NCDC cloud data are expected to give accurate long term average estimations of SCF (i.e. average of SCF over eight years) it will not accurately quantify cloudiness at each ground station at hourly resolution. For this reason cloudiness is also quantified in terms of the clear-sky index (kt), defined as the ratio of actual irradiance to modeled clear sky irradiance

$$kt = \frac{GHI}{GHI_{clear}} \qquad (6).$$

In Eq. 6 GHI is an hourly irradiance from either the SUNY or CIMIS dataset and GHI_{clear} is the modeled clear sky irradiance for the same hourly period[28], thus kt can be computed for both the SUNY and CIMIS data. In principle the value of kt should vary between zero and one, where small kt values indicate cloudy conditions and $kt = 1$ indicates clear sky conditions.

---

27 Vincenty, T. 1975. Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey Review* **22**, 176-183

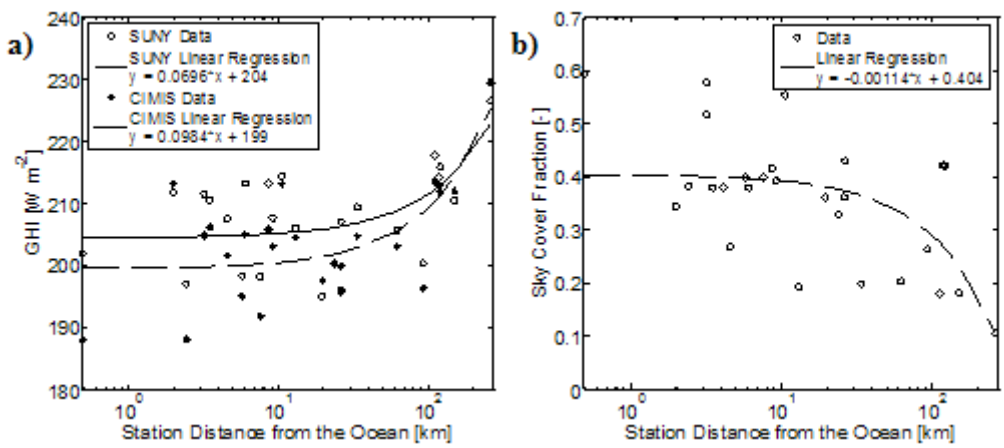
28 Snyder, R.L. and Eching, S. 2002. Penman-Monteith (hourly) reference evapotranspiration equations for estimating ETos and ETrs with hourly weather data. <http://biomet.ucdavis.edu/Evapotranspiration/PMhrXLS/PMhrDoc.pdf> Accessed January 13th, 2010.

CHAPTER 4:
Results

4.1 Annual Trends of Statistical Error Metrics

Table 1 provides yearly averages of the statistical parameters used to compare satellite derived GHI to the ground measurements for the entire dataset. The mean GHI is generally increasing in both SUNY and CIMIS as a function of distance from the ocean. Fig. 2 shows that annual mean irradiance increases and annual mean sky cover fraction decreases with increasing distance from the ocean. While proximity to the ocean would also be expected to impact the aerosol and water vapor column in a way that is not accounted for in the SUNY model, such a difference could not be detected in the clear sky results, presumably due to the limited accuracy of the CIMIS data. Although a high correlation between the two datasets is expected due to the predictable diurnal cycle of insolation, SUNY-CIMIS correlation coefficients of 0.97 or greater indicate that the SUNY model generally follows the GHI trends measured on the ground.

Figure 2: Dependence of GHI and SCF on Site Distance from the Ocean Based on Annual Averages at Individual Sites

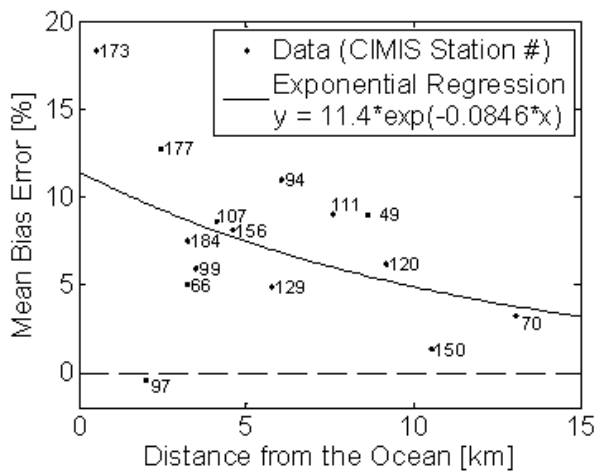


(a) Mean GHI at each CIMIS site and collocated SUNY pixel increases, while (b) mean annual SCF at airports decreases with increasing distance from the ocean.

For all stations the MAE ranges from 34 to 53 W m^{-2} and 17 percent to 28 percent and the RMSE ranges from 60 to 82 W m^{-2} and 29 to 44 percent. There is a weak trend of the largest MAE and RMSE occurring near the ocean, but MAE and especially RMSE are driven by the spatial heterogeneity (or randomness) of the cloud field together with the different location of the satellite pixel and the ground station. A cloud may have obscured much of the satellite pixel but not the CIMIS station resulting in large MAE and RMSE not related to a shortcoming in the SUNY model. The MBE is not sensitive to these effects as they are expected to average out over time. Since SUNY data are most often used to compute climatologies of the solar resource, the MBE is of critical importance. The MBE is generally positive (except for one coastal station and

three inland stations) and less than 27 W m^{-2} throughout. The relative MBE is less than 14 percent for all stations, and there is a trend of decreasing MBE with increasing station distance from the ocean among sites that are <15 km from the ocean (Fig. 3). Since the MBE decreases with distance from the coast a decaying exponential function was fit to the data (legend in Fig. 3).

Figure 3: Annual Average of Hourly MBE between the SUNY Model and CIMIS Data against Station Distance from the Ocean (Same as 'MBE [%]' Column in Table 1)

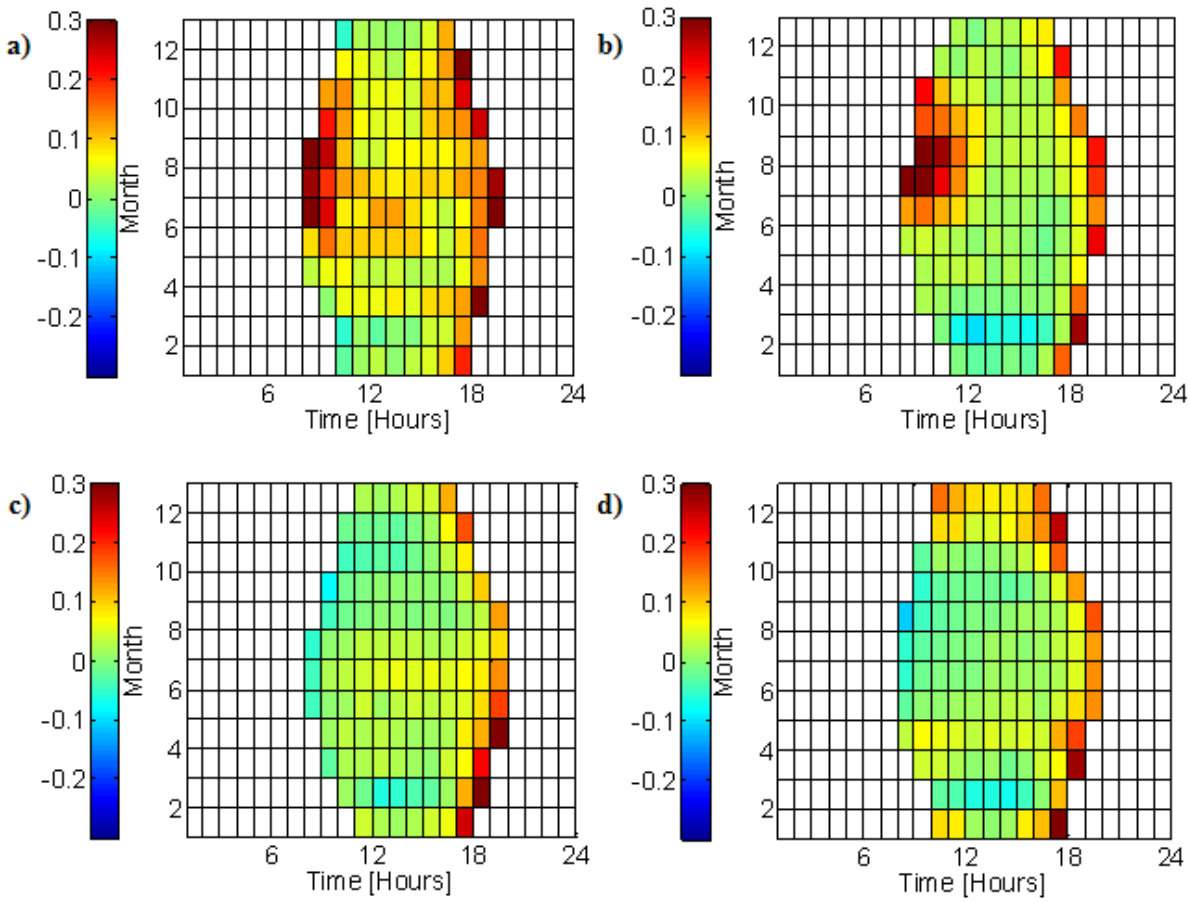


The station number is listed to the right of the corresponding data point.

4.2 Monthly/Hourly Climatologies of Mean Bias Error (MBE) and Cloudiness (SCF)

To further quantify this trend a detailed temporal analysis of the MBE at each station was conducted. At each site the residual between the satellite and ground data was computed for each hour of the day using all the days in the irradiance time series. MBEs for each hour and month of the year were then computed from the 24 hours x 30 days x *n* residual time series, where *n* is the number of years of available data (Fig. 4). There are two important patterns visible in Fig. 4. The first is a trend of large MBE (frequently >25 percent and up to 54 percent) that occurred at all coastal sites during the hours of 0801-1100 PST of June through September (with the exception of site 66 because data during this time period was filtered out due to shading), and it did not occur at any of the inland sites. The second is the occurrence of large positive MBE in the early evening hours (1801-2000 PST) year-round at all sites, independent of distance from the ocean. This year round evening MBE is not caused by the SUNY model, but is related to a programming error that occurred when processing the NSRDB dataset (Perez, personal communication, 2010; see Section 5 for further discussion).

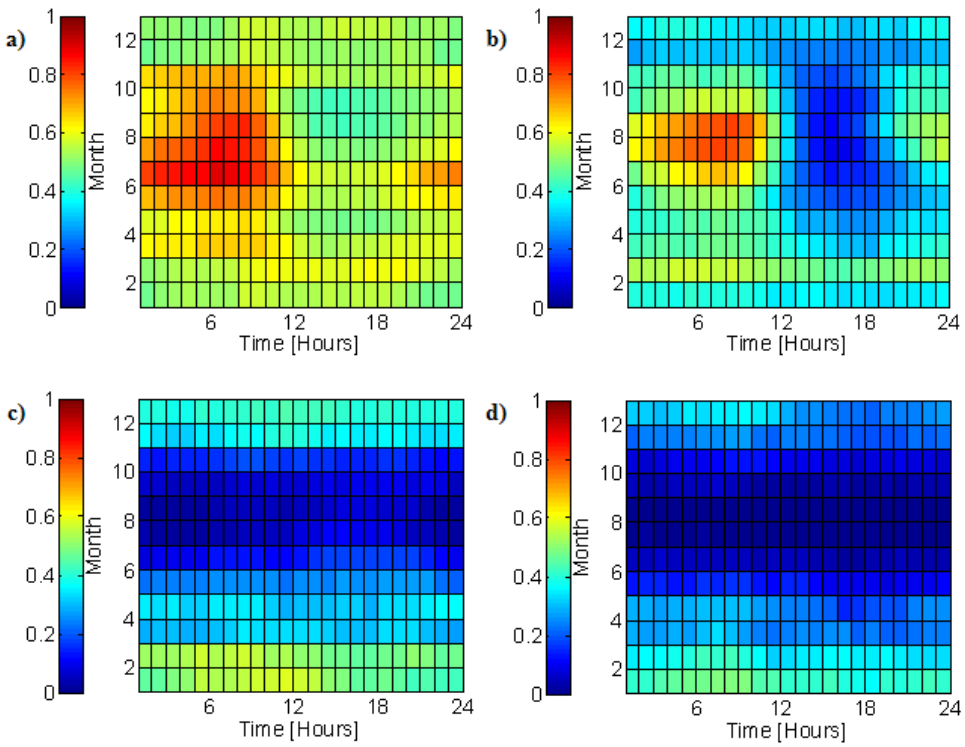
Figure 4: MBE [%/100] between SUNY Model and CIMIS Data Averaged Hourly by Month for: (a) CIMIS Station #173, 0.5 km from the Ocean; (b) CIMIS Station #111, 7.6 km from the Ocean; (c) CIMIS Station #008, 93.0 km from the Ocean; (d) CIMIS Station #125, 111.7 km from the Ocean.



White areas in these figures represent times that were excluded because of sensor cosine response issues when SZA > 80o.

The results of the analysis presented in Fig. 4a-d suggest that positive MBE in the SUNY data on summer mornings only occurs at coastal sites, prompting a search for possible explanations. Fig. 5 shows NCDC DSI-3505 SCF for the airports near the CIMIS stations in Fig. 4 averaged hourly for each month of the year (the same method that is used in Fig. 4). It is important to note that because there are only four discrete cloud cover descriptors (CLR, SCT, BKN, OVC) that describe a range of sky cover conditions there is a large error associated with individual sky cover observations, but this error is expected to average out over long time periods. Dense cloud cover is noted in the morning hours during summer months at all coastal sites. The cloud cover in Figs. 5a,b correlates strongly with the summer morning MBE of Figs. 4a,b, so the SUNY model overestimates GHI at the same times when there is persistent broken to overcast cloud cover. Both summer morning MBE and annual MBE are greater at station #173 than at station #111. Figs. 5a,b indicate that it is usually cloudier at #173 than at #111 which is further evidence that the summer morning MBEs are related to local patterns of cloudiness. There is no apparent correlation between cloud cover and positive MBE that would explain the early evening errors that were observed at all sites in the analysis.

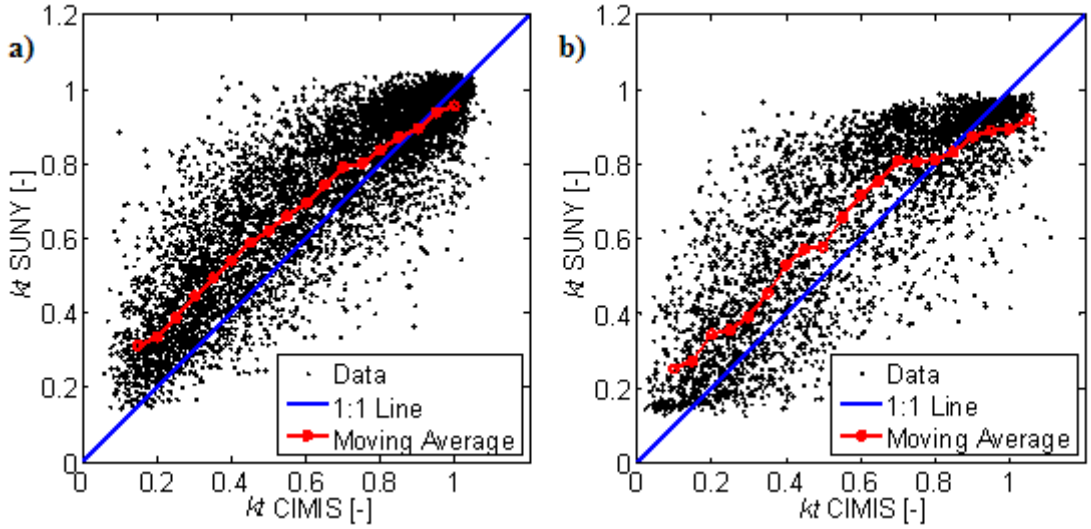
Figure 5: Mean SCF at (a) North Island NAS which Experiences Similar Weather Conditions to CIMIS Station #173; (b) Watsonville Muni Airport Near CIMIS Station #111; (c) Red Bluff Muni Airport Near CIMIS Station #008; (d) Meadows Field Near CIMIS Station #125



4.3 Satellite-Ground Differences in Cloudy Conditions

While the results of Figs. 4 and 5 suggest a relationship between cloud cover and MBE they do not prove causality, i.e. that individual values generated by the SUNY model are inaccurate under cloudy conditions. In order to establish causality between MBE and cloud cover scatter plots were created of SUNY and CIMIS *kt* to investigate the correlation between the two datasets under clear sky and cloudy conditions (Fig. 6). Fig. 6a shows only data for the months of June through September at the coastal station #173 and Fig. 6b shows data for the months of November through February at the inland station #125. The data were filtered for different months of the year for coastal and inland sites in order to include data from the cloudiest times of the year when cloud cover induced MBE would be most likely to occur based on the SCF climatologies in Fig. 5.

Figure 6: Scatter Plot of Hourly CIMIS kt versus SUNY kt Predicted for the Same Location at: (a) CIMIS #173 Jun-Sep; (b) CIMIS #125 Nov-Feb



The red line is a moving average of the data. The blue line is the 1:1 line which would indicate perfect agreement between the datasets.

Figs. 6a,b indicate that during clear sky conditions the SUNY and CIMIS data are in good agreement at both sites. Note that *kt* can be larger than one due to inaccuracies in the clear sky model, sensor inaccuracies, cloud or obstacle reflection of irradiance onto the sensor or an increased diffuse irradiance component produced by shortwave upwelling radiation reflected from the surface and/or obstacles surrounding the ground station. The SUNY model does not account for such effects causing the underprediction of measured GHI for $kt > 1$. Fig. 6b shows that for small clear-sky index ($kt \leq 0.6$) the SUNY model overestimates GHI at both sites.

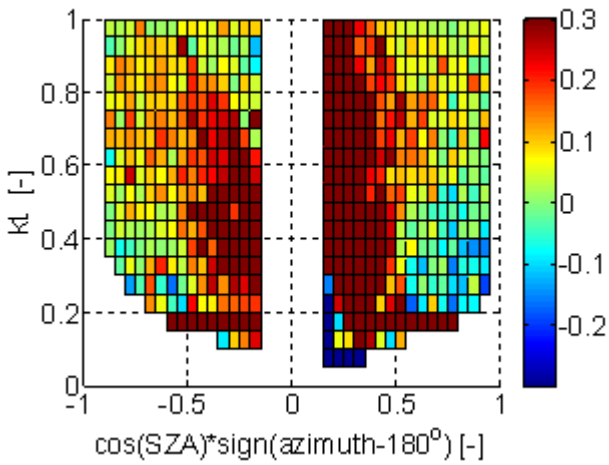
4.4 Correcting the Satellite Derived Irradiance Data

The NSRDB-SUNY dataset is a useful tool for solar resource assessment but errors in the modeled data make its application somewhat problematic. The SUNY model overestimates irradiance resulting in inaccurate sizing of PV systems and cost/benefit analyses which may be too optimistic. As the SUNY model is complex and computationally expensive and the NSRDB irradiance dataset is already freely available it would be useful to develop a post-processing correction for the NSRDB-SUNY dataset. Here such a correction using modeled output statistics (MOS) following the procedure of Lorenz *et al.*[29] was attempted. The combined effect of cloudiness and solar altitude on the MBE of the SUNY model by plotting MBE as a function of *cos(SZA)*sign(AZ-180°)* and *kt* (Fig. 7) was examined. AZ is the solar azimuth angle, defined to be 0° at North and increasing in the clockwise direction. Using this definition allows differentiation of morning and afternoon data since sign(AZ-180°)<0 before solar noon,

29 Lorenz, E., Hurka, J., Heinemann, D., Beyer, H.G. 2009. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2**(1), 2-10.

sign(AZ-180°)>0 after solar noon and sign(AZ-180°)=0 at solar noon. The use of *cos(SZA)*sign(AZ-180°)* is motivated by the fact that the two areas of significant positive MBE in Fig. 5 are not symmetric about solar noon. Our results suggest that these errors were generated are not solely a function of SZA so the parameter sign(AZ-180°) was used to differentiate the correction for these errors. MBE dependence on cos(SZA) was examined rather than SZA so that they both vary on a scale from zero to one. Fig. 7 indicates that positive MBE related to morning clouds depends on both *cos(SZA)* and *kt* while positive evening MBE depends strongly on *cos(SZA)* but is relatively constant in *kt*. This behavior is consistent with the results of Figs. 4 and 5. Fig. 7 also shows that the MBE is only large for morning clouds, but not clouds during other times of the day.

Figure 7: Instantaneous Hourly MBE [%/100] as a Function of Cos(SZA)*Sign(AZ-180°) and ktSUNY Averaged for 25 Ground Stations



(CIMIS #66 and #107 were excluded due to shading). Sunrise and sunset occur at *cos(SZA)*sign(AZ-180°)*=0. Data at 1159 solar time appear on the far left and data from 1201 solar time appear on the far right of the plot.

The data from Fig. 7 can be used to correct the NSRDB-SUNY directly without using any surface measurements of irradiance as inputs to the correction algorithm. This is desirable because the high spatial resolution of the SUNY dataset means that for many locations no such ground data are available to quantify the correction. The correction is accomplished in the following manner. The MBE data from Fig. 7 were separated about *cos(SZA)*sign(AZ-180°)*=0 (to differentiate morning and afternoon errors) and each part was fit using a 5th order polynomial in *cos(SZA)*sign(AZ-180°)* and *kt* (see Appendix). Separate polynomials were generated from MOS for coastal (<15km from the coast) and inland (>15km from the coast) sites. This polynomial expresses the expected error, MBE_p, for each modeled hourly irradiance using solar geometry (SZA and AZ) and the clear-sky index computed from the SUNY timeseries as inputs. Then the corrected hourly irradiance was determined by rearranging Eq. 2 to read $GHI_{SUNY,c} = GHI_{SUNY}/(MBE_p+1)$, where GHI_{SUNY} is an uncorrected hourly irradiance value and $GHI_{SUNY,c}$ is the corrected hourly irradiance. Testing showed that the 5th order polynomial was

not accurate for small SZAs resulting in erroneous corrections. Therefore the correction was only applied to hourly irradiance values when MBE_p>0.2.

Figure 8: MBE [%/100] between SUNY Model and CIMIS Station #173 Averaged Hourly by Month (a) before Applying the Correction Algorithm; (b) after Applying the Correction Algorithm

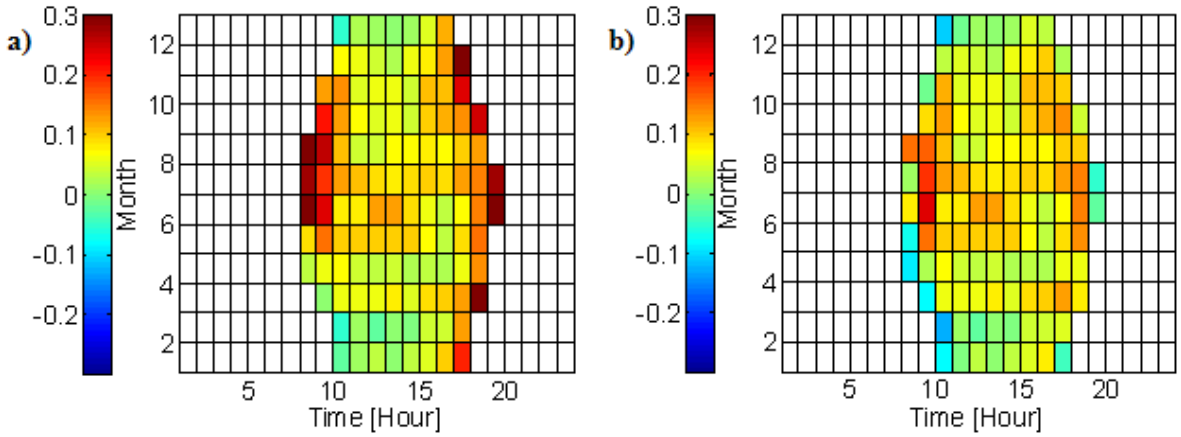


Fig. 8 illustrates the effects of the correction algorithm for coastal CIMIS #173. Positive summer morning MBE is improved dramatically by the correction algorithm but still remains positive. Year-round positive MBE that occurs in the late evening is slightly over-corrected so that when the correction is applied this error is reduced from approximately 30 percent to -5 percent. These results were consistent for other stations. Table 2 quantifies the effect of the correction algorithm in terms of the annual MBE at each ground station. The annual MBE improved at 18 stations and worsened at 9 stations (see "Percentage point Δ" columns in Table 2). In Table 2 Percentage point Δ is defined as $|MBE_{before}| - |MBE_{after}|$. A positive value in this column indicates that the correction improved the error at a particular site, while a negative value indicates that error increased. The average MBE at all sites after the correction was reduced to 1.6 percent and was distributed more evenly around zero. During summer mornings (Jun-Sep, 80°>*SZA*>10°) the average MBE at all sites was reduced to 5.1 percent. Year-round in the evening (80°>*SZA*>65°) the average MBE at all sites was reduced to 2.1 percent.

Table 2: Effect of the Correction Algorithm on SUNY Model Annual MBE. Negative Values in the "Percentage Point Δ" Columns Indicate That the Correction Made the Error Worse for a Particular Site.

Station Name	CIMIS Station ID#	Station Distance from Ocean [km]	Annual MBE Before Correction [%]	Annual MBE After Correction [%]	Percentage Point Δ in Annual MBE [%]	Jun-Sept MBE Before, Morning, 80°>SZA>10° [%]	Jun-Sept MBE After, Morning, 80°>SZA>10° [%]	Percentage Point Δ in Jun-Sept MBE [%]	Year-round MBE Before, Evening, 80°>SZA>65° [%]	Year-round MBE After, Evening, 80°>SZA>65° [%]	Percentage Point Δ in Evening MBE [%]
Torrey Pines	173	0.5	18.4	10.6	7.8	23.3	9.3	14.0	20.8	15.8	5.0
Port Hueneme	97	2.0	-0.4	-5.5	-5.1	10.6	2.0	8.5	-1.3	-5.5	-4.2
Watsonville West	177	2.4	12.8	5.3	7.5	26.4	14.5	11.9	11.6	5.6	6.0
San Diego II	184	3.2	7.6	2.8	4.8	4.9	-2.4	2.5	8.4	5.1	3.3
San Diego I	66	3.2	5.1	2.6	2.5	3.3	2.7	0.6	4.8	2.3	2.6
Santa Monica	99	3.5	6.0	0.6	5.4	11.0	3.7	7.3	6.0	1.7	4.3
Santa Barbara	107	4.1	8.7	4.6	4.0	13.6	8.5	5.1	8.4	4.8	3.6
Oxnard	156	4.6	8.2	3.0	5.1	7.7	0.3	7.4	10.4	6.1	4.3
Pajaro	129	5.7	5.0	-1.1	3.9	22.4	15.2	7.3	2.9	-2.4	0.5
Goleta Foothills	94	6.0	11.1	6.6	4.5	15.8	10.8	5.0	13.9	9.9	3.9
Green Valley Road	111	7.6	9.1	3.7	5.4	25.7	20.1	5.5	7.1	2.0	5.1
Oceanside	49	8.6	9.0	5.0	4.0	14.2	9.5	4.7	9.1	5.6	3.5
Guadalupe	120	9.2	6.2	2.2	4.0	24.8	20.0	4.8	1.8	-2.1	-0.3
Miramar	150	10.5	1.4	-2.0	-0.6	3.4	-0.5	2.9	1.7	-1.2	0.5
Manteca	70	13.1	3.3	-0.8	2.5	-7.3	-7.9	-0.6	7.6	2.8	4.8
Santa Rosa	83	19.5	-1.4	-5.2	-3.8	6.1	4.1	2.0	-3.5	-7.9	-4.4
Bennett Valley	158	23.7	-0.1	-3.4	-3.3	-2.3	-3.8	-1.5	0.1	-3.9	-3.8
Windsor	103	26.2	1.5	-1.5	0.0	8.1	7.0	1.2	1.2	-2.6	-1.3
Salinas South	89	26.2	9.1	6.2	2.9	19.2	17.4	1.8	7.5	4.0	3.5
Modesto	71	33.8	7.1	4.5	2.6	1.4	1.2	0.2	10.3	6.5	3.8
Denair	168	61.9	4.7	2.1	2.7	4.8	4.5	0.2	5.6	1.6	4.0
Gerber	8	93.0	6.4	3.9	2.5	2.6	2.3	0.2	9.3	5.7	3.6
Arvin-Edison	125	111.7	5.6	3.5	2.1	-4.5	-4.6	-0.1	8.8	5.6	3.1
Five Points South West	190	118.3	3.6	1.1	2.5	19.9	19.8	0.1	2.1	-1.7	0.4
Five Points	2	122.2	5.7	3.4	2.3	-4.2	-4.3	-0.1	10.0	6.7	3.3
Hanford Muni NOAA ISIS	210	150.9	-1.8	-4.2	-2.5	-5.3	-5.5	-0.2	-0.8	-4.3	-3.6
Bishop	35	263.0	-3.5	-5.8	-2.3	-6.7	-6.9	-0.2	-0.6	-4.1	-3.5
Average Value			5.5	1.6	2.3	8.8	5.1	3.4	6.0	2.1	1.8

CHAPTER 5:
Discussion and Conclusions

Based on GHI data collected in the state of California the SUNY model is generally accurate and provides high quality irradiance data with an average uncertainty level (MBE) of 5 percent (see Table 1). CIMIS data do not match rigorous surface radiation measurement network standards for sensor recalibration intervals and station maintenance. Consequently despite significant investment in data quality control by the authors to detect shading, misalignment, and miscalibration, especially absolute differences at individual stations may be explained by insufficient data quality. However, consistent trends observed from groups of stations are expected to point to systematic errors in the SUNY model. With no Baseline Surface Radiation Network (BSRN) or Surface Radiation Network (SURFRAD) sites in California, local networks such as the CIMIS network are the only data source available to study the effect of local meteorological patterns on SUNY errors. This situation and approach is an example that could be followed in many other regions where irrigation-related surface weather networks have monitored solar irradiance for many years.

As illustrated by Table 1 and Fig. 3 the errors in the SUNY GHI data are greater near coastal stations than they are at inland stations, particularly in the mornings during the summer (Fig. 4). On summer mornings in California a coastal marine inversion layer is present nearly every day that creates overcast conditions throughout the morning until about 1100 PST. At this time of year sea breezes that drive the marine layer inland (up to 15 km from the coastline) are particularly prevalent because large density gradients exist between cool air over the ocean and warm air over land[30]. Fig. 4 indicates that the SUNY model overestimates the average GHI up to 54 percent MBE on summer mornings. This error is deemed significant because the range of clear sky irradiance during those times is about 500-850 W m^{-2} resulting in a significant overestimation of the total monthly GHI. This large positive MBE occurs because the SUNY model incorrectly parameterizes the effects of cloud cover on surface irradiance. Fig. 5 shows that persistent cloud cover exists at the same times as the large summer morning error. The analysis of Fig. 6 further establishes a relationship between cloud cover and SUNY model error. While the SUNY model is accurate under clear sky conditions ($kt \geq 0.8$), in broken or overcast cloud cover ($kt \leq 0.6$) the SUNY model overestimates the value of surface irradiance.

The inaccuracies in the NSRDB SUNY dataset during the late afternoon ($65^\circ \leq SZA < 80^\circ$) are significant because they occur when the sun is well above the horizon. Since this error is not observed in the morning when $65^\circ \leq SZA < 80^\circ$, it is not related to problems with computing the diffuse component of the irradiance at low sun altitude. This error is not caused by the SUNY model, but is related to a programming error that occurred when on the hour GOES visible images ('Uglo' data) were shifted to the half hour in the NSRDB 'Sglo' column and averaged over the hour (Perez, personal communication, 2010). This shifting process, which uses an interpolation on the clearness index to reconstruct an irradiance value for the half hour

30 Simpson, J. E. 1994. Sea Breeze and Local Winds. Cambridge University Press: New York, 234 pgs.

timestamp, is described in the NSRDB user's manual[31]. The NSRSDB 'Uglo' data are expected to be accurate, but with hourly CIMIS data, this could not be confirmed in our analysis. However, the Uglo data are typically not used since it is not straightforward to integrate the irradiances to obtain total irradiation over a month or a year.

Our attempt to correct the NSRDB-SUNY dataset using the approach proposed by Lorenz *et al.*[32] was only partially successful. Because the data in Fig. 7 are based on the average MBE from 25 stations the correction is not expected to be perfect for individual sites. In addition the data in Fig. 7 are noisy which makes it difficult to fit a polynomial that accurately captures the wide variability in the data. Although the correction did improve the satellite data at the majority of the sites examined in this analysis, extensive validation of such a correction model must be conducted before it can be applied universally to the SUNY data.

Even though errors were found under certain conditions, the SUNY satellite-derived irradiance model provides accurate estimations of surface irradiance for most sites. This conclusion is consistent with previous validation surveys of the SUNY irradiance model[33] [34] [35] [36]. Indeed there are tremendous advantages to using the SUNY model for solar resource assessment because of its complete coverage and high spatial resolution. Nevertheless room for improvement does exist and before the model is applied it should be verified against available ground measurements of irradiance, especially in areas where persistent cloudiness or unusual weather conditions have the potential to create errors in the modeled data. This research has important implications for the future viability of solar energy production in California, and in other regions for which such technologies may otherwise hold great promise in providing sustainable alternative energy sources to growing human population centers.

---

31 Wilcox, S. 2007. National solar radiation database 1991-2005 update: user's manual. NREL/TP-581-41364. <http://rredc.nrel.gov/solar/old_data/nsrdb/1991-2005/> Accessed July 30th, 2009.

32 Lorenz, E., Hurka, J., Heinemann, D., Beyer, H.G. 2009. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2**(1), 2-10.

33 Muneer, T. 1997. Perez slope irradiance and illuminance models: evaluation against Japanese data. *Lighting Research and Technology* **29**(2), 83-87.

34 Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., Vignola, F. 2002. A new operational satellite-to-irradiance model. *Solar Energy* **73**(5), 307-317.

35 Vignola, F., Harlan, P., Perez, R., Kmiecik, M. 2007. Analysis of satellite derived beam and global solar radiation data. *Solar Energy* **81**(6), 768-772.

36 Gueymard, C. and Wilcox, S. 2009. Spatial and temporal variability in the solar resource: assessing the value of short-term measurements at potential solar power plant sites. *Solar 2009 Conf., Buffalo, NY, American Solar Energy Soc. (ASES)*

Appendix A:
Mean Bias Error Correction

The data presented in Fig. 7 were modeled using a robust (bisquare weighting) linear least squares polynomial fit to 5th order. The 5th order polynomial is given in the following form.

$$f(x,y) = \sum_{i=0}^{5} \sum_{j=0}^{5} p_{i,j} x^i y^j \qquad \text{(A.1)}$$

In Eq. A.1 the coefficient $p_{i,j}$=0 if i+j>5. The values of the coefficients used in the correction algorithm are listed in Table A.1.

Table A1: Values of Coefficients used in the 5th Order Polynomial Fit used to Correct the Satellite Derived Irradiance Data

Coastal Data	***AM***	***PM***	***AM***	***PM***	***AM***	***PM***	***AM***	***PM***	***AM***	***PM***	***AM***	***PM***
Degree of Term (i,j)	**j = 0**		**1**		**2**		**3**		**4**		**5**	
i = 0	-1.39	-1.09	4.32	6.66	-15.55	-31.56	23.88	65.05	-16.07	-60.83	4.07	21.25
1	16.08	3.65	-1.30	25.70	17.03	-10.13	8.70	-14.08	-11.49	7.30		
2	-59.28	-16.46	40.94	-92.06	43.15	59.20	-23.30	-5.79				
3	-68.58	43.04	90.04	84.93	-3.34	-32.63						
4	-13.82	-46.04	37.62	-23.22								
5	9.65	16.79										
Inland Data	***AM***	***PM***	***AM***	***PM***	***AM***	***PM***	***AM***	***PM***	***AM***	***PM***	***AM***	***PM***
Degree of Term (i,j)	**j = 0**		**1**		**2**		**3**		**4**		**5**	
i = 0	-1.23	-1.21	7.32	7.67	-33.86	-31.60	68.76	57.50	-63.45	-48.08	21.93	15.21
1	-8.73	5.86	-16.67	21.93	30.75	-10.72	-12.17	-14.18	2.93	8.32		
2	-45.05	-28.05	-16.52	-78.85	56.51	65.58	-10.10	-11.25				
3	-84.90	65.97	26.20	63.64	24.71	-30.80						
4	-62.81	-65.56	26.29	-14.59								
5	-14.36	22.98										

Separate corrections were derived for coastal (<15 km from the coast) and inland (>15 km from the coast) sites. AM and PM denote the coefficients used before and after solar noon.