Optimal Interpolation of Satellite Derived Irradiance and Ground Data

Antonio T. Lorenzo*, Matthias Morzfeld†, William F. Holmgren‡, Alexander D. Cronin§

*College of Optical Sciences, University of Arizona, Tucson, AZ, 85721, United States
†Department of Mathematics, University of Arizona, Tucson, AZ, 85721, United States
‡Department of Atmospheric Sciences, University of Arizona, Tucson, AZ, 85721, United States
§Department of Physics, University of Arizona, Tucson, AZ, 85721, United States

***Abstract*—We describe how Bayesian data assimilation can be used to improve nowcasts of irradiance over small, city-scale, spatial areas. Specifically, we use optimal interpolation (OI) to improve satellite derived estimates of global horizontal irradiance (GHI) using ground truth data that was collected sparsely over Tucson, AZ. Our results show that the local data indeed improves the satellite derived estimates of GHI. A key to success with OI in this context is to prescribe correlations based on cloudiness, rather than spatially. OI can be used with a variety of data, e.g., rooftop photovoltaic production data or irradiance data, as well as with several different satellite derived irradiance models.**

***Index Terms*—data assimilation, optimal interpolation, remote sensing, solar irradiance**

I. Introduction

Accurate estimates of the global horizontal irradiance (GHI) are crucial to the deployment and grid integration of photovoltaic (PV) systems. Satellite derived GHI estimates are used to design and site PV power plants, to forecast the power output of a fleet of PV generators, and to provide electric utilities real-time estimates of the distributed generation (DG) or "behind the meter" generation of rooftop PV systems. Data assimilation provides a framework to combine the large area coverage of GHI estimates derived from satellite imagery with the more accurate data of ground sensors.

Previous work has also explored using data assimilation techniques to improve solar radiation estimates [1], [2]. We use optimal interpolation (OI), which can be thought of as a generalized least squares approach [3]. OI is a statistical method to combine prior information about some parameter (the background) with observations based on the errors and correlations in the background and observations. The background is computed from satellite estimates, the observations come from a mix of GHI sensors and rooftop PV systems. OI is also used by [2], where numerical weather prediction (NWP) solar radiation data are combined with ground sensors. A key difference and innovation in our paper is that correlations used for OI are prescribed based on differences in cloudiness between locations, rather than spatial distance.

Our paper is organized as follows. We describe OI in Sec. II and apply it to satellite GHI estimates in Sec. III. We discuss the results in Sec. IV and future work in Sec. V. Finally, a summary is provided in Sec. VI.

II. Optimal Interpolation Procedure

A. Method

We briefly describe the OI method; the derivation can be found in many data assimilation textbooks, e.g. [3]. The output of an OI routine (known as the analysis), $\hat{\mathbf{x}}$, is a vector of length N and is a weighted sum of the background (the prior information represented as an N vector), $\mathbf{x}_b$, and the measurements, $\mathbf{y}$ (M vector):

$$\hat{\mathbf{x}} = \mathbf{x}_b + \mathbf{W}(\mathbf{y} - \mathbf{H}\mathbf{x}_b). \tag{1}$$

In this study, $\mathbf{x}_b$ is composed of satellite derived clear-sky indices and $\mathbf{y}$ is composed of clear-sky indices from a number of ground irradiance sensors. The observation matrix, $\mathbf{H}$ ($M \times N$ matrix), maps points in the background space to points in observation space. We construct $\mathbf{H}$ using the nearest neighbor approach of selecting the background points that are closest to the observation locations. The weight matrix, $\mathbf{W}$ ($N \times M$ matrix), is constructed from the error covariance matrix of the background, $\mathbf{P}$ ($N \times N$ matrix), and the error covariance of the observations, $\mathbf{R}$ ($M \times M$ matrix) as

$$\mathbf{W} = \mathbf{P}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{P}\mathbf{H}^T)^{-1}. \tag{2}$$

We also compute the error covariance matrix of the analysis, $\hat{\mathbf{P}}$ ($N \times N$ matrix), as

$$\hat{\mathbf{P}} = (\mathbf{I} - \mathbf{W}\mathbf{H})\mathbf{P}, \tag{3}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix.

An essential part of the OI routine is choosing appropriate error covariance matrices, $\mathbf{R}$ and $\mathbf{P}$. The standard method, that we also follow is to assume that the errors between sensors are uncorrelated so that $\mathbf{R}$ is a diagonal matrix. Each diagonal element of $\mathbf{R}$ is the variance of the observations at each location over a given period (in the results that follow we used the entire study period).

The method we use to obtain $\mathbf{P}$ is novel, in fact it is the primary difference between our work and [2]. First, we separate $\mathbf{P}$ into a correlation matrix $\mathbf{C}$ and diagonal variance matrix $\mathbf{D}$:

$$\mathbf{P} = \mathbf{D}^{1/2}\mathbf{C}\mathbf{D}^{1/2}. \tag{4}$$

We obtain $\mathbf{D}$ in a similar manner as $\mathbf{R}$: we take the variance of each pixel in the satellite image over some period of time.

Care must be taken when estimating the background correlation matrix $\mathbf{C}$. A standard method is to assume the correlation decays exponentially with distance between points and this approach is taken in [2]. This method works well for resource assessment with daily or longer integration times and for nowcasts at locations with sensors nearby. The method we use depends on the actual distributions of clouds as seen by the satellite. The idea is that pixels in the background that have similar cloudiness have high correlation and those with very different cloudiness have low correlation. In the final analysis, this translates to only adjusting the cloudy areas with observations that are also cloudy and leaving the clear areas to be adjusted by observations of the clear sky.

To construct the correlation matrix $\mathbf{C}$ ($N \times N$), we first define the distance, d_{ij}, as the difference between pixel i and pixel j of an image $\mathbf{v}$ (N vector) that defines the cloudiness,

$$d_{ij} = |v_i - v_j|. \tag{5}$$

To obtain the elements of $\mathbf{C}$, c_{ij}, we apply a known correlation function, k, to each distance so that

$$c_{ij} = k(d_{ij}). \tag{6}$$

Any one of a number of covariance functions could be chosen for k; see [4] for a partial list. In this work, we studied piecewise linear correlation functions,

$$k(r) = \begin{cases} 1 - \frac{r}{l} & r < l \\ 0 & r \geq l \end{cases}, \tag{7}$$

where l is a characteristic length that must be specified. The choices of k and l need to be tuned to the area that the algorithm is applied to. Once the error covariance matrices are defined, one can compute an analysis estimate using the above equations.

B. Data used for optimal interpolation

This study applies OI to observations and geostationary satellite data from April, May, and June 2014 in Tucson, AZ. The observation data were collected from 22 diverse sensors including a calibrated NREL MIDC sensor [5], custom irradiance sensors [6], and data from rooftop PV systems. Irradiance observations were averaged to 1 minute and PV data are reported as 5 minute averages. We note that all data sources (observations and satellite images) are available in near real-time so that the OI corrected GHI images can be used as a basis for forecasts. To simplify the computation, all data were converted to clear-sky index data using clear-sky expectations for each sensor. Five sensors, including the calibrated NREL MIDC GHI sensor, were not used in the OI process for validation and error statistics are only presented for these withheld sensors. The remaining 17 sensors are used as the observations, $\mathbf{y}$, in the OI routine.

The satellite data were obtained from the GOES-W geostationary satellite, which was GOES-15 for the period of interest. To obtain the background error correlation, we estimate the cloudiness image, $\mathbf{v}$, from the 1 km resolution, visible band of the satellite as follows. We convert the raw visible brightness counts, b_i, to visible albedo, divide by the cosine of the solar zenith angle, ϕ, to correct for the time of day, and arrive at an adjusted visible albedo,

$$v_i = \left(\frac{b_i}{255}\right)^2 / \cos(\phi_i). \tag{8}$$

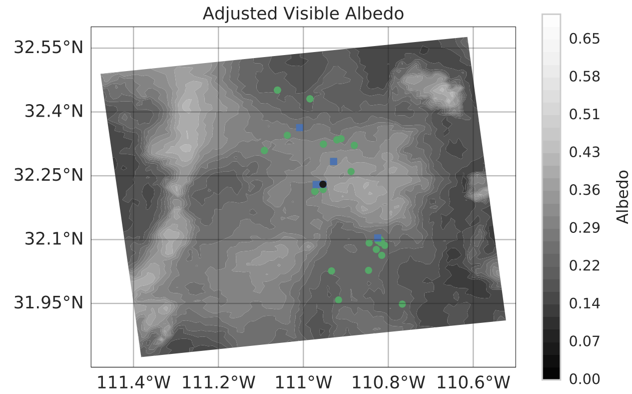

Fig. 1. Adjusted visible albedo image derived from the GOES-W visible reflectance image on 2014-04-18 18:30Z over Tucson, AZ. The lighter/high albedo areas indicate cloudy areas. The green circles are the sensors used for OI, the blue squares are the sensors used for error analysis, and the black circle in the center is the calibrated NREL MIDC sensor.

We plot the adjusted visible albedo as a map over Tucson, AZ, in Fig. 1. The lighter areas in Fig. 1 correspond to areas of high albedo which indicates that the area is cloudy. This adjusted visible albedo is used to obtain the background error correlation matrix via eqs. (5)–(7) with a correlation length of $l = 0.2$. However, other quantities, such as cloud fraction, could also be used to estimate the cloudiness at each satellite pixel.

C. Satellite derived irradiance models

We studied two satellite image to GHI models to generate the background image, $\mathbf{x}_b$, which was also converted to clear-sky index before applying OI.

One satellite to GHI model to generate $\mathbf{x}_b$ is a physically based model called the University of Arizona Solar Irradiance Based on Satellite (UASIBS) model [7]. UASIBS uses the visible and infrared images from the GOES-W satellite to generate a cloud mask. Then, parameterized cloud properties determined from the infrared images are used in a radiative transfer model to determine the surface GHI. This GHI estimate has the same resolution as the visible channel of the GOES-W satellite (approximately 1km).

The second model to generate $\mathbf{x}_b$ is a semi-empirical model, which we refer to as the EM model. This model is based on the SUNY model which applies a regression to the visible channel of the GOES-W satellite [8]. The only differences between the EM model and the SUNY model are that the dynamic range is set only with the 3 months of data used in this study instead of the recommended 60 day window with seasonal correction and that the specular correction factor was neglected.

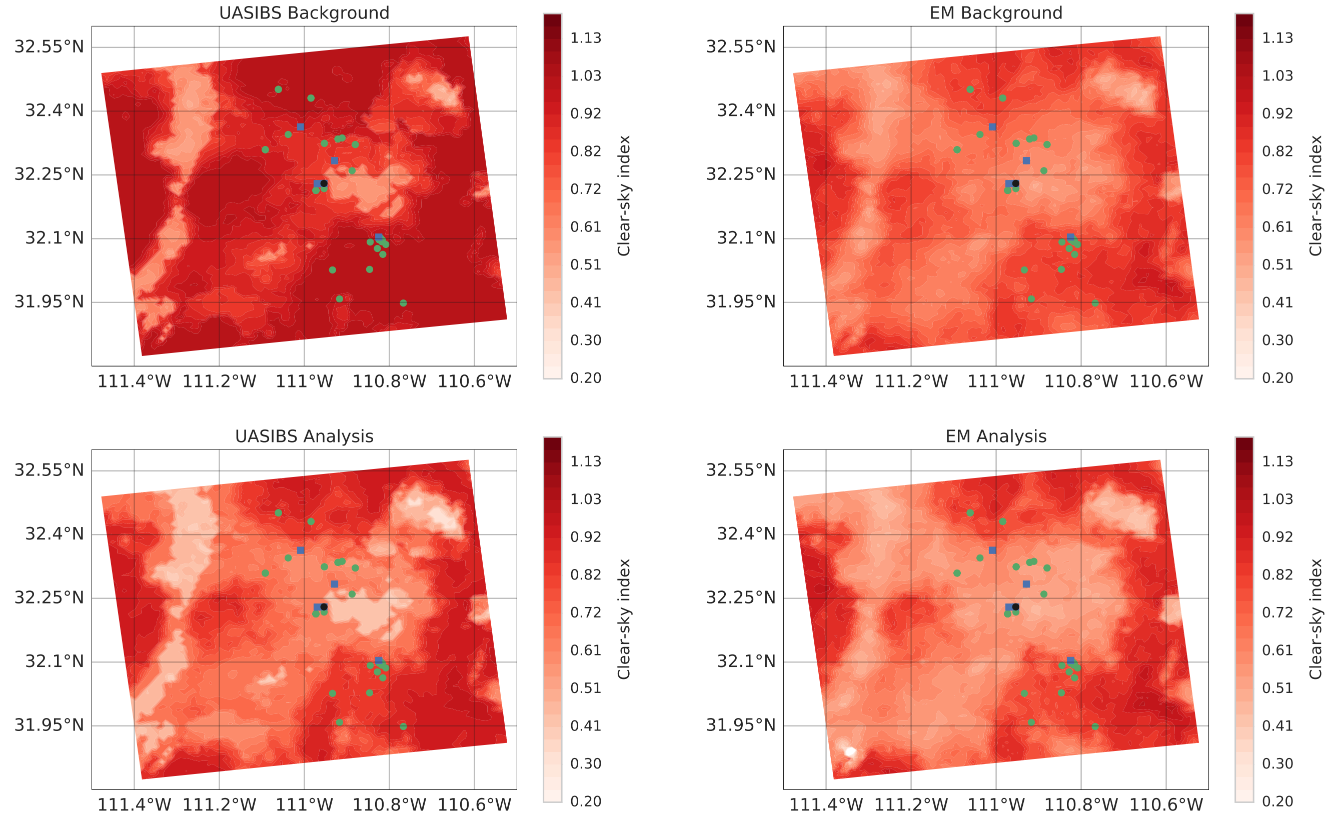

Fig. 2. Example OI results for one image/data taken on 2014-04-19 18:30Z (11:30 AM local time). The top row are the background satellite derived clear-sky index estimates before OI. The lower row are the clear-sky index analysis after performing OI. Satellite derived estimates using the UASIBS model are on the left and those using the EM model are on the right. Lighter shades indicate thicker clouds. The green circles are the sensors used for OI, the blue squares are the sensors used for error analysis, and the black circle in the center is the calibrated NREL MIDC sensor. Comparing the background and analysis, one can see that the thin cloud near the center of the image is made slightly thicker in the analysis.

III. Results

In this section, we present the results of the analysis of roughly 1200 satellite images taken over Tucson, AZ. Both the UASIBS and EM models described above were used to convert the raw satellite images to estimates of GHI at the surface and then used as the background field for the OI algorithm.

Figure 2 shows uncorrected background maps of clear-sky index derived from the UASIBS and EM models and the OI corrected analysis. Notice that in the UASIBS analysis image, the thin clouds in the center of the image are thicker than in the background based on the information of the sensors near the top of the image that measure other parts of the cloud. For the EM model, OI adjusts the cloudy areas to be more cloudy and the clear areas to be more clear. We also see that the analysis images in Fig. 2 are similar suggesting that OI works robustly with different satellite image to GHI models that define the background.

Figure 3 shows an example of the errors of the background and analysis images as compared to sensor observations for a single satellite image/time processed with the UASIBS model. The errors shown are computed from sensors that are *not* used during OI. We see that the absolute error was reduced for all sensors, including the calibrated MIDC sensor and rooftop PV systems. This suggests that the OI correctly propagates information from data to unobserved locations.

We calculated empirical cumulative distribution functions (CDF) for the observations, background, and analysis. Figure 4 shows these CDFs for the UASIBS model. The slope of zero around 0.8 in the CDF of the UASIBS background (red dashed-dotted line) indicates that the UASIBS model does not predict clear-sky indices of 0.8. The analysis (blue dashed line) does predict clear-sky indices in that range and even extends the range over 1.0 to more closely match the observations.

The empirical CDF for the EM model is shown in Fig. 5. We see that the EM model tends to over-predict clouds, but that the OI then removes much of this bias. On the other hand, the figure suggests that the analysis could be improved at smaller clear-sky indices to better match the observations.

For the 1200 images analyzed, root-mean squared errors (RMSE), mean absolute errors (MAE), and mean bias errors (MBE) decreased on average. Error statistics for the EM and UASIBS models in terms of the clear-sky index calculated over all the withheld sensors and for clear, cloudy, and all days are presented in Table I. Error statistics in units of GHI for the calibrated MIDC irradiance sensor are presented in Table II.

TABLE I

ERROR STATISTICS CALCULATED OVER 1200 SATELLITE CLEAR-SKY INDEX ESTIMATES AND OI CORRECTED ANALYSIS. BOTH THE EMPIRICAL (EM) MODEL AND UASIBS MODEL DESCRIBED IN SEC. II-C ARE SHOWN. THE MEAN ABSOLUTE ERROR (MAE), ROOT MEAN SQUARED ERROR (RMSE), AND MEAN BIAS ERROR (MBE) ARE CALCULATED OVER ALL THE WITHHELD SENSORS AND ALL IMAGE TIMES AS A SINGLE TIME-SERIES. STATISTICS WERE CALCULATED FOR ALL DAYS, ONLY CLEAR DAYS (ROUGHLY 700 DAYS), AND CLOUDY DAYS (500 DAYS). ALL NUMBERS ARE IN UNITS OF CLEAR-SKY INDEX WHICH HAS A TYPICAL RANGE OF 0 TO 1.3.

	MAE			RMSE			MBE		
	All	Clear	Cloudy	All	Clear	Cloudy	All	Clear	Cloudy
EM analysis	0.088	0.048	0.149	0.172	0.095	0.245	0.026	0.021	0.033
EM background	0.184	0.152	0.231	0.268	0.213	0.333	0.138	0.140	0.136
UASIBS analysis	0.080	0.039	0.141	0.164	0.088	0.235	-0.005	-0.004	-0.006
UASIBS background	0.094	0.047	0.164	0.190	0.099	0.275	-0.015	-0.003	-0.034

TABLE II

ERROR STATISTICS CALCULATED OVER 1200 SATELLITE GHI ESTIMATES AND OI CORRECTED ANALYSIS FOR THE CALIBRATED NREL MIDC SENSOR. BOTH THE EMPIRICAL (EM) MODEL AND UASIBS MODEL DESCRIBED IN SEC. II-C ARE SHOWN. THE MEAN ABSOLUTE ERROR (MAE), ROOT MEAN SQUARED ERROR (RMSE), AND MEAN BIAS ERROR (MBE) ARE CALCULATED OVER ALL IMAGE TIMES AS A SINGLE TIME-SERIES. STATISTICS WERE CALCULATED FOR ALL DAYS, ONLY CLEAR DAYS (ROUGHLY 700 DAYS), AND CLOUDY DAYS (500 DAYS). UNITS ARE W/M^2.

	MAE			RMSE			MBE		
	All	Clear	Cloudy	All	Clear	Cloudy	All	Clear	Cloudy
EM analysis	56.0	23.4	104.	113.	32.3	174.	16.1	17.3	14.3
EM background	110.	85.7	145.	144.	97.0	194.	75.0	83.8	61.9
UASIBS analysis	50.9	17.5	101.	110.	26.4	171.	2.94	6.96	-3.03
UASIBS background	53.1	16.4	108.	120.	27.9	186.	-12.4	3.02	-35.2

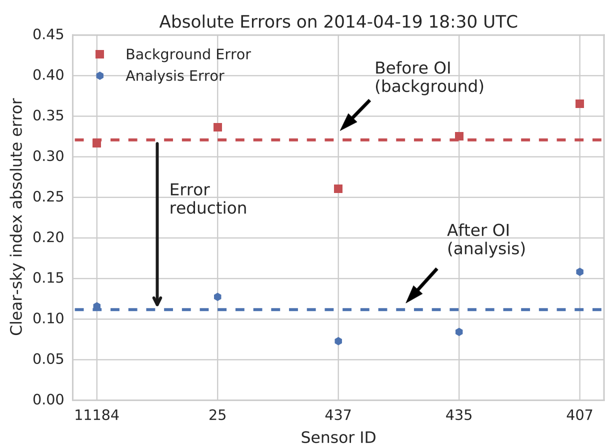

Fig. 3. A plot of the absolute error in the analysis and background images (generated with UASIBS) as compared to observations at some sensor locations showing reduced errors for a single satellite image. The sensors shown were not included in the OI correction routine. Note that sensor 11184 is the MIDC calibrated irradiance sensor and sensors 437, 435, and 407 are rooftop PV systems. The red squares indicate the absolute error in the background image while the blue circles indicate the error in the analysis. The dashed lines indicate the mean absolute errors for the sensors shown.

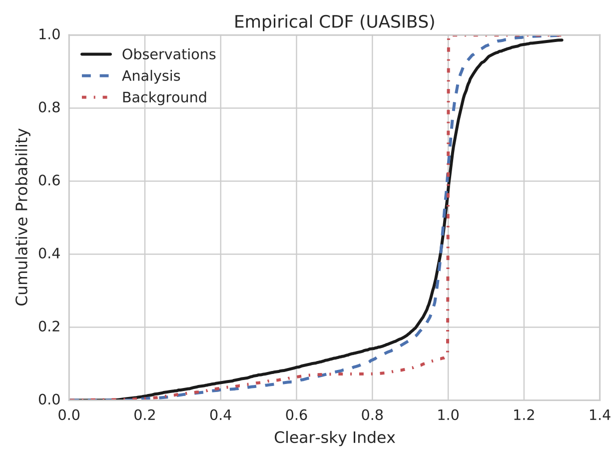

Fig. 4. UASIBS empirical cumulative distribution function. The black line is the CDF of the observations, the red dashed-dotted line is the CDF of the background, and the blue dashed line is the CDF of the analysis. The UASIBS background does not predict clear-sky indices around 0.8 and does not extend beyond 1.0. The analysis shows better agreement with the observed CDF.

IV. DISCUSSION

Our results show significant improvement by the OI for the EM model. Improvements for the UASIBS model are more modest. The reasons for this are as follows. UASIBS is a more sophisticated satellite image to GHI model, so that improvements are harder to obtain. In particular, the average error values shown in the tables above differ from the large improvements we have seen on many days, and illustrated in Fig. 3. We suspect that average errors are likely to be affected by large errors occurring only on some days due to parallax.

Parallax refers to the discrepancy between the actual location of a cloud and the location tagged by a satellite [9]. The GOES-W satellite is located at 135°W on the equator while Tucson, AZ is at roughly 32°N and 110°W, so the satellite is viewing the clouds at an angle. The satellite geolocates each pixel as if it were at the surface. This means that a cloud

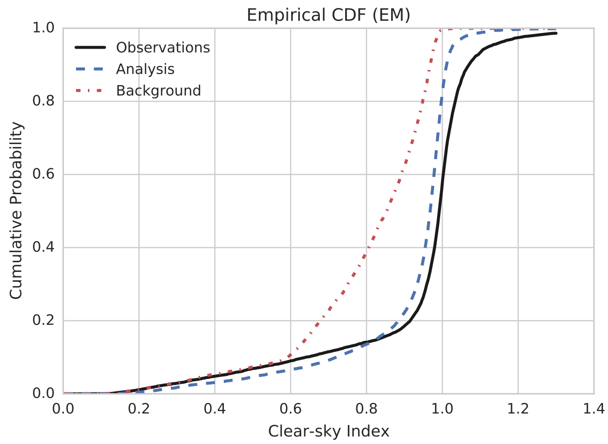

Fig. 5. The empirical cumulative distribution function for the EM model. The black line is the CDF of the observations, the red dashed-dotted line is the CDF of the background, and the blue dashed line is the CDF of the analysis. The background EM model seems to make clouds thicker than they are in reality. The analysis corrects much of this bias and changes the shape of the CDF to more closely match the observations.

obscures a pixel that is to the NE of the cloud, so the actual location of the cloud is to the SW of what the satellite tags the pixel as. Thus, when the OI algorithm tries to compare the observations with the background derived from the satellite image, the observations and background may disagree about whether a cloud is present at all.

This issue is illustrated in Fig. 6 where many sensors are near the edges of the estimated clouds. Some of the sensors locations that are reported as clear in the background are actually cloudy. The OI algorithm tries to rectify this by adjusting the areas that were clear in the background to be cloudy. When we compare this analysis to the background and adjusted visible albedo image, we see the analysis does not look physical. If we shift the satellite image by a small amount to the SW and rerun OI, we see that this shifted analysis looks more like what one would expect given the visible albedo image. This suggests that we first need to correct the parallax issue before performing OI, and that the error statistics calculated over 1200 image times are likely skewed by these errors.

V. Future Work

We plan to improve this work in several aspects. An important task will be to correct the issue of parallax that can cause large errors in the OI analysis. We have experimented with estimating the cloud top height and adjusting for parallax on a pixel by pixel basis, but found this is challenging to do well. In the future, we plan to group classes of clouds together to then determine a height for each cloud group and shift the group appropriately.

This work focused only on the area around Tucson, AZ. One future experiment could examine how OI can improve background estimates using observations that are very far apart and may experience different weather conditions.

VI. Conclusion

There are a number of models to convert satellite images to ground irradiance, and all are prone to errors. These satellite derived irradiance images are important to many phases of PV integration, from siting to forecasting the output of a fleet. We describe how to improve the irradiance estimates using ground data and optimal interpolation.

The optimal interpolation technique uses satellite derived estimates of GHI, ground observations, and the associated error estimates to produce a GHI estimate that has, on average, better error statistics. An important consideration for the method as described is the specification of the error correlation between pixels in the satellite image. We propose using the (almost) raw visible image from the satellite to correlate pixels based on the cloudiness at each pixel. We apply this method to a physically based satellite image to GHI model and show that the distribution of the estimated GHI more closely matches the data. Similarly, the method applied to an empirical satellite image to GHI model removes a large bias from the GHI estimate.

One limitation in the optimal interpolation method is that errors in the estimated locations of the clouds in the satellite GHI estimate can produce analysis images that are unreasonable. Thus, future work will explore correcting this issue of parallax or recognizing when this issue occurs so that optimal interpolation can be avoided for those times. Other future work includes producing a forecast from these improved satellite derived GHI nowcasts.

Acknowledgment

This project was funded in part by Tucson Electric Power and Arizona Public Service. We thank Technicians for Sustainability, a local solar PV installer, for providing data from rooftop PV systems. ATL thanks the University of Arizona Renewable Energy Network for support and Professor Avelino Arellano for instruction and guidance. WFH thanks the Department of Energy (DOE) Office of Energy Efficiency and Renewable Energy (EERE) Postdoctoral Research Award for support.

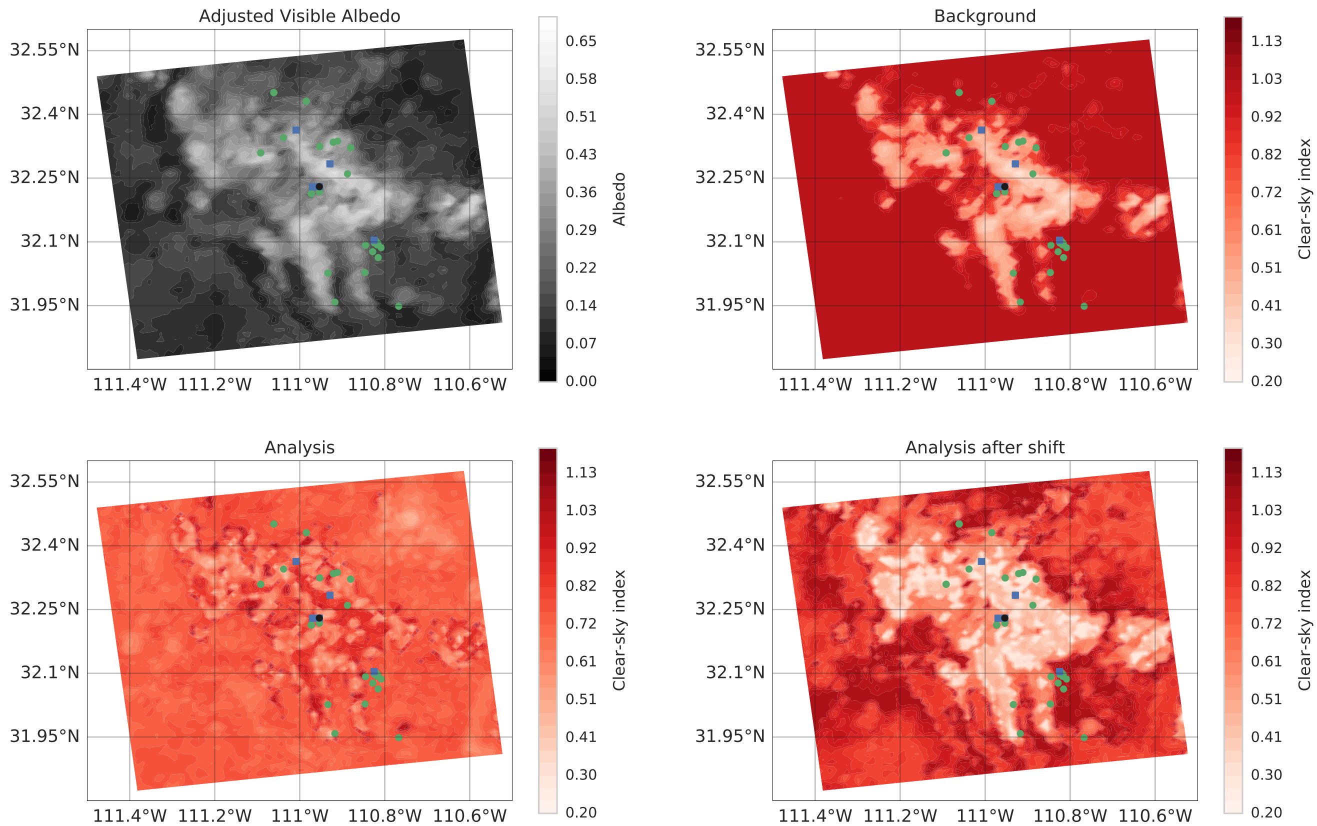

Fig. 6. Illustration of error due to parallax. The adjusted visible albedo image from the GOES-W visible channel, background clear-sky index estimate made with the UASIBS algorithm, analysis after performing OI, and analysis after shifting the satellite image are shown. The green circles are the sensors used for OI, the blue squares are the sensors used for error analysis, and the black circle in the center is the calibrated NREL MIDC sensor. We see that at sensor locations near the edge of clouds in the background, the sensors and background disagree about whether it is cloudy. This causes the OI to fail as it tries to rectify this discrepancy. If we shift the image slightly to the SW and redo the OI, we see that the sensors and background now better agree about whether the area is cloudy so that the analysis after shifting looks more reasonable.

References

[1] L. Linguet and J. Atif, "Estimating Surface Solar Irradiance from GOES Satellite with Particle Filter Model and Joint Probability Distribution," *Canadian Journal of Remote Sensing*, vol. 41, no. 2, pp. 71–85, jul 2015. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/07038992.2015.1040150

[2] J. A. Ruiz-Arias, S. Quesada-Ruiz, E. F. Fernández, and C. A. Gueymard, "Optimal combination of gridded and ground-observed solar radiation data for regional solar resource assessment," *Solar Energy*, vol. 112, pp. 411–424, feb 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X14006021

[3] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 2003.

[4] C. Rasmussen and C. Williams, "Covariance Functions," in *Gaussian Processes for Machine Learning*. MIT Press, 2005, ch. 4, pp. 79–104. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6279030

[5] S. Wilcox and A. Andreas, "Solar Resource & Meteorological Assessment Project (SOLRMAP): Observed Atmospheric and Solar Information System (OASIS); Tucson, Arizona (Data)," nov 2010. [Online]. Available: http://dx.doi.org/10.7799/1052226

[6] A. T. Lorenzo, W. F. Holmgren, M. Leuthold, C. K. Kim, A. D. Cronin, and E. a. Betterton, "Short-term PV power forecasts based on a real-time irradiance monitoring network," in *2014 IEEE 40th Photovoltaic Specialist Conference (PVSC)*, 2014, pp. 0075–0079. [Online]. Available: http://dx.doi.org/10.1109/PVSC.2014.6925212

[7] C. K. Kim, W. F. Holmgren, M. Stovern, and E. A. Betterton, "Toward Improved Solar Irradiance Forecasts: Derivation of Downwelling Surface Shortwave Radiation in Arizona from Satellite," *Pure and Applied Geophysics*, may 2016. [Online]. Available: http://link.springer.com/10.1007/s00024-016-1302-3

[8] R. Perez, P. Ineichen, K. Moore, M. Kmiecik, C. Chain, R. George, and F. Vignola, "A new operational model for satellite-derived irradiances: description and validation," *Solar Energy*, vol. 73, no. 5, pp. 307–317, nov 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X02001226

[9] G. A. Vicente, J. C. Davenport, and R. A. Scofield, "The role of orographic and parallax corrections on real time high resolution satellite rainfall rate distribution," *International Journal of Remote Sensing*, vol. 23, no. 2, pp. 221–230, nov 2010. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01431160010006935