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 piece-wise 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

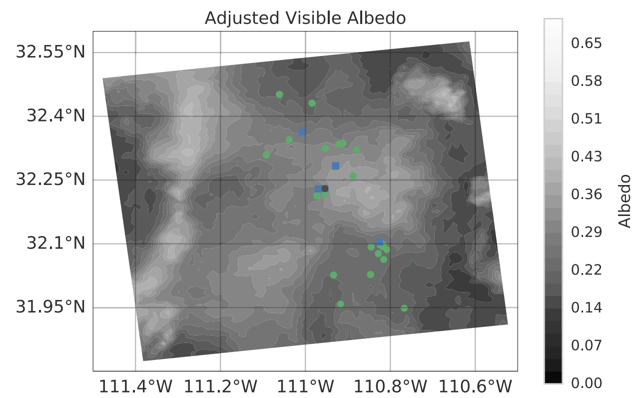

Fig. 1. Adjusted visible albedo image derived from the GOES-W visible reflectance image on 2014-04-18 18:30Z over Tucson, AZ. The lighter/high albedo areas indicate cloudy areas. The green circles are the sensors used for OI, the blue squares are the sensors used for error analysis, and the black circle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

zenith angle, ϕ, to correct for the time of day, and arrive at an adjusted visible albedo,

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

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.