3. Optimal interpolation

We now describe the OI method. Under wide assumptions, OI is optimal in the sense that it is the best linear, unbiased estimator of a field. Further detail can be found in data assimilation textbooks, e.g. Kalnay (2003).

The result of the OI routine, known as the analysis, $\mathbf{x}_a$, is a vector that is produced by computing a weighted sum of the background (or prior information), $\mathbf{x}_b$, and a correction vector (or "innovation" in OI) that depends on the measurements, $\mathbf{y}$:

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

As discussed in Section 2.1, the N satellite derived clear-sky indices from one image are represented as the background vector, $\mathbf{x}_b$. The measurement vector, $\mathbf{y}$, is a vector of length M of clear-sky indices generated from M ground irradiance sensor and rooftop PV power data observations as discussed in Section 2.2. The observation matrix, $\mathbf{H}$, is an $M \times N$ matrix that maps points in the background space to points in the observation space. We construct $\mathbf{H}$ using the nearest neighbor approach of selecting the satellite pixels that are closest to the observation locations. Another possible approach is to average the points in the background that are within a given radius of each sensor location. Furthermore, $\mathbf{H}$ can contain conversion factors to convert the units of $\mathbf{x}_b$ to the units of $\mathbf{y}$. In our case however, $\mathbf{H}$ is unitless because $\mathbf{y}$ and $\mathbf{x}_b$ are both in units of clear-sky index. Example background and analysis images for the UASIBS and SE models are shown in Fig. 2.

The weight matrix, $\mathbf{W}$, is an $N \times M$ matrix constructed from the error covariance matrices of the background, $\mathbf{P}$, and the observations, $\mathbf{R}$, as

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

Choosing these error covariance matrices must be done with care: they define how information is transferred from sensor locations to other locations in the satellite image, and how much weight is given to any one sensor or satellite pixel.

$\mathbf{R}$ is defined as the error covariance matrix of the observations such that

$$\mathbf{y} = \mathbf{y}_t + \mathbf{e}, \quad \mathbf{e} \sim N(\mathbf{0}, \mathbf{R}) \quad (3)$$

where $\mathbf{y}_t$ is the true value of the observation and $\mathbf{e}$ is a random vector sampled from a multivariate normal distribution with mean 0 and covariance $\mathbf{R}$. On clear days, we assume the true clear-sky index values are 1.0. We also assume that the measurements are unbiased and that the correlations in the *errors* between sensors is negligible, so $\mathbf{R}$ is a diagonal matrix in our case. Thus, we estimate the diagonal elements (sensor error variances) by computing the variance on a set of clear days in the training data set for each sensor individually. Furthermore, we restrict the minimum variance to be 0.001 or about a 3% clear-sky index RMS error to avoid exact interpolation at sensor locations. With $\mathbf{R}$ calculated from the ground sensor data, we describe various ways to parameterize $\mathbf{P}$ next.

4. Covariance parameterization and correlation structure

Choosing an appropriate background error covariance matrix is an important step in the OI method for this application and determines how well OI performs. The background error covariance matrix, $\mathbf{P}$, defines how information is transferred from sensor

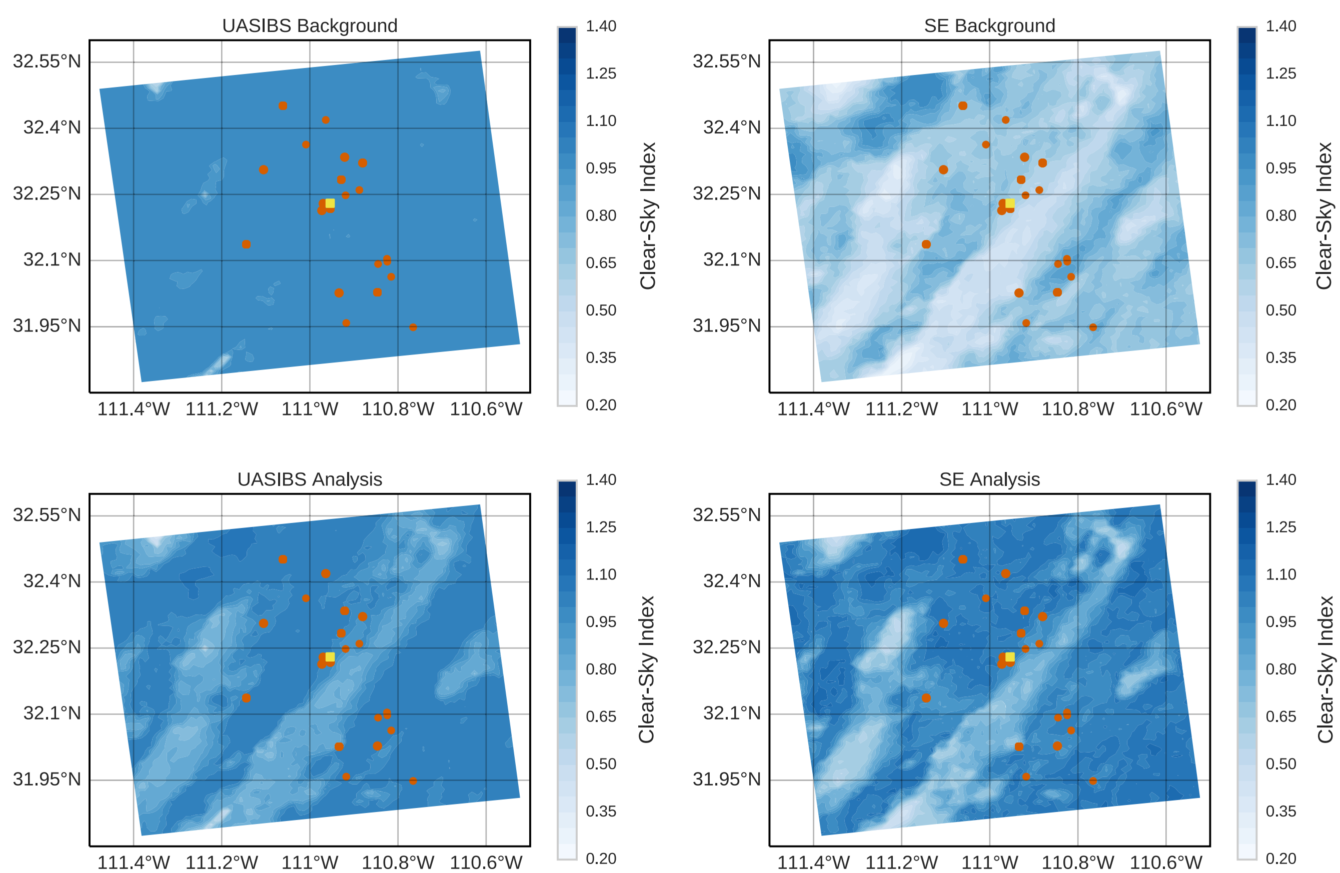

Fig. 2. Example background (top row) and analysis (bottom row) clear-sky index images using the UASIBS (left column) and SE (right column) satellite image to ground irradiance models applied to the visible satellite image shown in Fig. 1. Note that in this case, UASIBS failed to produce many clouds. OI adds clouds to the analysis and also makes the darker, clear areas even more clear. In this case, the SE model overproduces clouds. OI reduces the cloud amount while keeping clouds in suitable locations.