ances, **R**, from the observation data and define the background error covariances, **P**, either empirically or by following the above procedure after choosing a distance metric r, the correlation function k, the correlation length l, and the scaling factor d using Eqs. (5)–(7). Finally, Eqs. (1) and (2) can be used to compute the analysis, $\mathbf{x}_a$.

5. Geolocation correction

It is important to consider errors in the tagged location for each satellite pixel compared to the ground sensors and the time-stamp of the image. Furthermore, one must take into account the position of the sun in order to predict the cloud shadow location on the ground. If this cloud shadow location is inaccurate, the optimal interpolation routine may perform poorly, or worse, may invert the cloudy and clear areas of the images. Examples of an inverted analysis and the corrected analysis once these position adjustments are taken into account are shown in Fig. 3.

The first geolocation issue is called parallax, which is the discrepancy between the actual location and the location tagged by a satellite due to the satellite viewing the scene at an angle (Vicente et al., 2010). The GOES-W satellite is located at 135°W on the equator while Tucson, AZ is at roughly 32°N and 110°W. The satellite tags the location of each pixel as if it were at the surface. This means, for our region, that a cloud obscures a pixel that is to the NE of the cloud. Thus, the actual location of the cloud is to the SW of what the satellite tags the pixel as.

Another source of error is a timing issue that arises because the satellite tags each image with a single time, however it may take the satellite 30 min to sweep and capture that image. Thus, there is uncertainty in the time that any part of the image was captured.

Estimating where the cloud shadow falls on the surface due to solar position effects is the final geolocation issue we take into account. If the shape and height of the clouds is known, the correction for both parallax and solar position is a simple geometry problem. However, cloud shape and height are difficult to determine with sufficient accuracy, and we rely only on an estimate of the height of the top of the clouds and ignore the vertical thickness. We also assume that the cloud height is uniform in one image.

Given these limitations, we use a simple strategy to correct for geolocation errors. We find a single optimal cloud height by minimizing the mean squared error (MSE) between the OI analysis and sensors that are not used to perform OI. The sensors not used are the same cross-validation sensors we will discuss next. We perform this correction using a grid search for cloud heights ranging from 0 to 14 km and we shift the entire background image based on that height, perform OI, then calculate the MSE. Once the height that minimizes MSE is found, we perform OI again on the shifted background image and save the analysis as our result for the given time. This technique assumes that there is a single cloud layer, which is not always the case and can be improved in the future.

6. Tuning OI to a specific location

As discussed in Section 4.2, k, l, and d are tunable parameters that determine how information is spread through the image. In order to find suitable values of these parameters, we split the satellite images into a training and a verification set as described

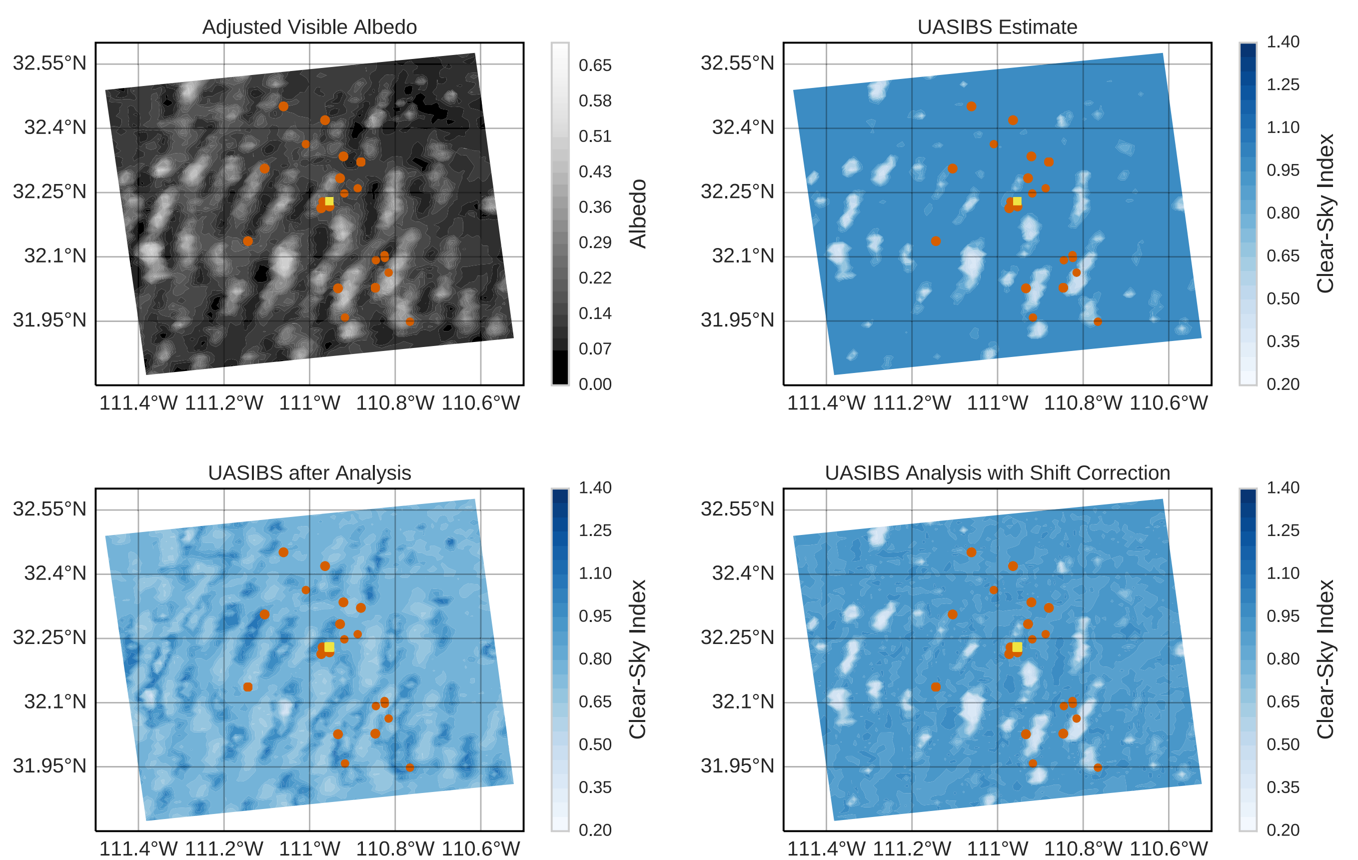

Fig. 3. An example of a time when errors in geolocation of the satellite image result in an analysis that is inconsistent with the actual satellite image. The background estimate in this case (upper right) agrees well with the visible satellite image (upper left). However, after performing OI, the analysis (lower left) has clouds in areas that should be clear according to the visible image and sometimes makes areas that should have clouds clear. After shifting the background image slightly, OI produces an analysis (lower right) that is consistent with the visible image.