**Improvements to the KFAN air quality post-processing algorithm to more accurately forecast extreme events.**



*Figure 1. Schematic of the standard analog post-processing scheme.*

A description of the standard analog post-processing technique is illustrated in Fig. 1 in schematic form for an ensemble of only two analogs. Here the black curve in the top panel is a time series predicted by a model over an eight day period, and the red curve is the corresponding set of observations. The same hour of the day is considered within each day. The data to the right of the dashed line at t=0 represents the new forecast, while the data to the left represents the previous 7 one-day forecasts. As shown in the top panel, the analog technique searches for previous forecasts at the same lead time that are similar to the current new forecast (blue star) – where the similar forecasts are circled in black, and re-orders the time series with the first closest analog directly preceding the new forecast, and the second closest forecast next (bottom panel). The observations (in red) corresponding to the two-best analog forecasts are then weighted by how closely they resemble the current forecast, and their weighted sum (the ensemble mean) provides the corrected forecast estimate. Although this schematic figure shows only two analogs for a single variable and a one week training period, in practice any number of analogs can be used with multiple search variables, and the training period can extend indefinitely into the past. Once the bias correction is calculated at a given forecast hour for each AIRNow observation site, the bias is interpolated across the CMAQ grid, and then added to the CMAQ gridded forecast.

Next we describe how the analog search algorithm is modified to better account for extreme forecasted events. Figure 2 is a simplified version of the figure above, while Fig. 3 and Fig. 4 show the new algorithm.

In Fig. 2 the current forecast is again shown in blue, and now we show three forecast analogs (in red) and their corresponding observations in black. The ensemble mean value of the three corresponding observations is indicated by the horizontal black line, and the difference between it and the CMAQ forecast is the bias correction, indicated by the blue arrow, leading to the new bias-corrected forecast indicated in green.

Analogs

Obs

CMAQ

KFAN

Time

Ozone, PM2.5

*Figure 2. Standard KFAN correction where Analog mean concentration value is very close to the raw CMAQ forecast value – that is, good analogs have been found.*

Note that in Fig. 2 the ozone or PM2.5 values of the analogs are very similar to those of the current CMAQ forecast. If one has a training period long enough so that every new forecast can find good analogs, this will be the case. However, if a new forecast predicts an extreme value that has never been observed before at the site in question, or has occurred many fewer times than the number of analogs being examined, the analog mean will then have a considerably different value than the current forecast. This is more likely to happen when only a short training period is available, but of course can happen even with long training periods.

Figure 3 shows this situation when the current forecast is a moderately extreme event for which there are no good historical analogous forecasts. In this case the analogs (red symbols) all have lower ozone or PM2.5 values than the current forecast. The standard analog correction in this scenario would again be the ensemble mean of the corresponding observations, indicated by the sold green circle.

Analogs

Obs

CMAQ

KFAN

Time

Ozone, PM2.5

KFAN-New

*Figure 3. KFAN correction where Analog mean concentration value is slightly larger than the raw CMAQ forecast value – that is, moderately good analogs have been found.*

The basic idea of the extreme event correction is very simple. It calculates the difference between the current forecast and the poor analogs, indicated by the dashed red line in Fig. 3, and adds this difference to the standard correction, with the final bias-corrected forecast shown as the open green circle. The assumption behind this correction is that the CMAQ model, when predicting an extreme event, has some skill beyond that of its historical best analogs. This might happen for PM2.5 for example, if there is a forecast fire nearby an AIRNow observations site, with no previous fires at the same location in the training data set. Since CMAQ has fire emissions that are satellite based, it is to be expected that the model will produce somewhat realistically high PM2.5 values, and that these high model values should be believed even if the model has never predicted such high values at that same site in the training data. A measure of the exceptionality of this event is given by the difference between the CMAQ forecast and its previous analogs (dashed re line), which is then added to the final bias-corrected forecast (open green circle).

In our last figure, Fig. 4, we show a similar situation but when the CMAQ forecast is much larger than has ever been observed before. In this case the extreme event correction, again indicated by the dashed red line, is large enough so that the final bias-corrected KFAN value is larger than either the set of analogs or their corresponding observations, coming close to the actual value of the CMAQ forecast.

KFAN-New

Analogs

Obs

CMAQ

KFAN

Time

Ozone, PM2.5

*Figure 4. KFAN correction where Analog mean concentration value is much larger than the raw CMAQ forecast value – that is, only poor analogs have been found.*

One might consider applying this correction only when the CMAQ forecast is at a climatologically extremely high or low value. However, we have found that applying it to all forecast values increases the skill of the bias-corrected forecasts over the entire range of ozone and PM2.5 values.

Figure 5. KFAN correction where the raw CMAQ forecast is much less than both the Analog mean concentration value and the observed values. This is probably unlikely to occur, as there is no obvious mechanism to make the CMAQ forecast so much lower than its analogs. But it leads to KFAN lower than even the low CMAQ forecast.

KFAN-New

Analogs

Obs

CMAQ

KFAN

Time

Ozone