

## ***Interactive comment on “Assessing filtering of mountaintop CO<sub>2</sub> mixing ratios for application to inverse models of biosphere-atmosphere carbon exchange” by B.-G. J. Brooks et al.***

**B.-G. J. Brooks et al.**

bjorn@climatemodeling.org

Received and published: 31 January 2012

I was not able to tell which version of the manuscript was actually seen. However I still believe that the mathematical notation may not meet the requirements outlined in this latest review. So I have once again revised this. I have pasted the revised sections below, which now meet the improvements requested. See below:

### **Method 4: filtering of outliers using a weighted median smoother**

An effective method that has been used in many signal filtering applications is the weighted median smoother (*cf.*, Tukey, 1974). Because our intent is not necessarily to smooth but to reject CO<sub>2</sub> mole fractions that are not regionally representative when

C14912

present in the data, we have modified Tukey’s method. We use a Weighted Median Filter (WM) that rejects an observation if its residual from the daily median is in excess of the summed and weighted inter-day variance for the previous two weeks.

The WM filter slides a backward-looking window over the time sequence of daily medians  $X(N)$ , obtained from hourly measurements  $X(n)$ , to create the related subset of hourly values  $S(n)$ . A range of acceptable mole fractions centered about the daily median value  $\tilde{X}(N)$  is computed dynamically at each step (day)  $N$ . The limits of the range are a function of the sum of differences between each daily median value in the sequence  $[\tilde{X}(N) - \tilde{X}(N - j)]$  over the previous two weeks and are weighted using a geometrical decay function to favor more recent variability. Thus the difference between today and yesterday is weighted at  $1/2$ , and the residual between yesterday and the day before yesterday is weighted  $1/4$ , and so on. The upper or lower limit,  $L$ , is computed using the difference equation:

$$L(N) = \sum_{j=1}^{14} \frac{|\tilde{X}(N - j) - \tilde{X}(N - j + 1)|}{2^j} \quad (1)$$

for which  $\tilde{X}(N)$  is the daily median at day  $N$  in the series. Then for each day,  $N$ , the set  $S(N)$  of hourly data to keep is:

$$S(N) = \{x(N, h) : |x(N, h) - \tilde{X}(N)| < L(N)\} \quad (2)$$

### **Method 5: iterative filtering of outliers from a fitted polynomial (statistical interpolation)**

We also include a statistical interpolation filter (SI) that is duplicated from the method used by (Thoning et al., 1989) at Mauna Loa to identify well-mixed background CO<sub>2</sub> (see the original publication for a complete explanation of the filtering protocol). SI

C14913

considers past and future observations through a sliding window to reject outliers from a fitted spline. The SI filter used here has one key difference from (Thoning et al., 1989), which is that it does not use a low-pass spectral filter.

Following the protocol outlined by Thoning:1989 our SI filter works by passing a ten day sliding window over the original time series of hourly values  $X(n)$  to create subset  $S(n)$  that consists initially of non-afternoon samples (the 15 hourly CO<sub>2</sub> mole fractions for the day excluding hours 11, 12, ..., 19). Daytime values were removed in the first steps by (Thoning et al., 1989) in order to fit the spline to values not strongly influenced by afternoon photosynthesis. For each ten day window a cubic spline  $S(X)$  is fitted through the daily means  $\bar{X}(N)$ , that exclude afternoon samples. In the first phase of filtering if the daily standard deviation ( $\sigma_{X(N)}$ ) exceeds 0.5 ppm the filter will reject the hourly observation  $X(n)$  with the largest residual from spline curve  $S(X)$ , which is described by the expression:

$$s(n) = \{X(n) : \sigma_{X(N)} > 0.5 \text{ and } \max |S(X) - X(n)|\} \quad (3)$$

where the residual is computed as the absolute difference between the spline curve and the hourly measurement.

In the second phase of filtering the window advances across all days, re-fitting a new ten day spline with each new window and rejecting no more than one observation per day with each iteration over the entire time series. After no more than 14 iterations (the maximum number of hourly observations that can be rejected for each day), or when the standard deviation of all daily means is less than 0.5 ppm (e.g.,  $\sigma$  of  $X(N) \leq 0.5$ ) the excluded daytime observations from the original time series (i.e., hours 11, 12, ..., 19) are incorporated back into  $S(n)$ . A final spline is refitted and those observations that are within 0.5 ppm of the spline form the final subset, which is expressed as:

$$S(n) = \{s(n) : |S(X) - s(n)| \leq 0.5\} \quad (4)$$

C14914

As mentioned above to simplify exposition of these filters we also provide a supplementary spreadsheet that can be referred to do duplicate our methods. Figure 3 illustrates the impact of our various filters on one year of measurements for one RACCOON site (Storm Peak).

---

Interactive comment on Atmos. Chem. Phys. Discuss., 11, 25327, 2011.