Atmospheric
Chemistry
and Physics
Discussions

Interactive comment on "Multi-year objective analyses of warm season ground-level ozone and PM$_{2.5}$ over North America using real-time observations and Canadian operational air quality models." *by* A. Robichaud and R. Ménard

Anonymous Referee #3

Received and published: 26 August 2013

This study by A. Robichaud and R. Ménard presents the results of a multiyear Objective Analysis of the warm season surface ozone and particulate matter of diameter less than 2.5 microns (PM2.5). The first part of the paper describes the background, motivation, data and the methodology in some detail. The final few sections focus on some results: a nice trend analysis, and correlations with meteorological and economical indices. I especially like the trend analysis that shows the decrease of ozone and PM2.5 at high percentiles, indicative of emission reductions and an increase of background (low percentile) values. These results are not entirely new or surprising but they

do corroborate and extend earlier findings (e.g. Cooper et al 2010) using a different methodology. In addition, the paper makes a solid case for Optimal Interpolation. The mathematics appears sound but a few items could be made clearer and there are a few minor mistakes, which I list in Specific comments below.

Because of the methodology and scientifically important results, the paper should be of interest to the ACP readers, however, I have found some areas that need some serious editing.

General comments

I would put more emphasis on the importance of the cross-validation results. Basically what they show is that this OA yields good results in areas where there are no observations (but not too far away from observations). This builds a case for using OA rather than calculating trends directly from measurements. This is mentioned in the text but it should be emphasized more. I would move Section 4 up before current Section 3 and state clearly why this validation is important, in the first paragraph (instead of just saying that it is required).

Since this impacts the entire analysis, please discuss the validity of the Gaussian assumption for errors. For example, the mixing ratios for ozone at the surface can be expected to be comparable to the standard deviations of the errors, in which case the Gaussian assumption fails. Figure 3 shows that the random model error can be over 15 ppbv. Isn't it comparable with the mean background surface ozone in rural areas?

English. There are a lot of ungrammatical sentences and style issues, especially in the first few sections. The manuscript should be seriously edited for language. I don't feel particularly competent here but I point out some of the most glaring problems in Specific comments. There are many more.

Figures. The fonts are too small. Color bars are impossible to read without some serious magnification. Essentially, almost all the figures should be edited and made

more readable

Specific comments

P 139740 L26. 'to provide to the public and health specialists with'. Drop the second 'to'

P 13974. L2. Reference to Menard and Robichaud (2005). This is a seminar talk. It would be better to cite peer-reviewed material if available

P13974, L5: 'analysis matrix' is fine but "analysis vector" would jibe better with the terminology you use later on.

P139745 L20 "Finally, σf and Lc represent respectively the background error variance and the correlation length". Shouldn't it be "standard deviation"?

P13975. Equation (3) needs a reference and some justification, e.g. why this gives a positive semi-definite background error covariance. Maybe Gaspari and Cohn 1998.

P 13978, Eq. (11). Why are you making this assumption? Please, explain, as I don't think this is discussed in Menard 2010. An exponential decay of the model bias like this could result in essentially zero bias in between regions if they are sufficiently far away. Why were these particular values chosen for a and b? If there are only four elliptical regions then a map with the region boundaries would help visualize what's going on.

P 13979, L12 'calculate' $\rightarrow$ calculating

P 13979, L23 'Goddard Space Flight' $\rightarrow$ Goddard Space Flight Center

P 13980, Eqs. (12) & (13). Introducing coefficients alpha and beta seems unnecessary. They just represent the fact that L and sigma can be tuned by scaling. Otherwise, they are the same as Eqs (3) and (4). You could simply state that you are tuning the two parameters in (3) and (4), thus avoiding the repetition and making the presentation more concise.

C6178

P 13980, If Eq. (13) ends p being used after all I think the l.h.s. should read (HB)ˆt, not H(HB)ˆt. Also, 'T' and 't' are used interchangeably to indicate transpose. I would just stick to 'T'.

P 13980, Eqs. (14) - (16):

1) It's not clear to me why applying (14) would produce a Kalman gain that will bring the chi-square statistic closer to 1.0. Please explain. It's intuitively clear why Eq. (16) would do that (if chi_sq > 1 then an increased error variance should bring it down and vice cersa). Equivalently, why does the sequence in (14) converge?

2) 'until there is convergence or until chi-sq $\sim$ 1'. I think convergence will be there if and only if chi-squared converges to 1, so it's not either or. I would like to see more detailed justification for this adaptive scheme

P 13981 L1. 'If needed only if' $\rightarrow$ drop the first 'if'

P 13980, L18. 'more closer' $\rightarrow$ closer

Section 2.4. Expand the description of the two models: the main relevant chemical reactions, the number of species modeled, emission inventories, etc.

P 13984. "the standard deviation of the observation error including the representativeness error is believed to be higher than 5 ppbv". Please, state the typical mixing ratios in regions of high and low concentrations (industrial and rural). The same for PM2.5. This will help judging the validity of the Gaussian errors assumption, especially in areas of low concentrations.

P13985 L17. Avoid using "etc". It's better to spell things out

P 13976, 25. A reference to Wikipedia? Some linear algebra textbook would be more suitable

P13974, L5 "analysis increments" $\rightarrow$ "analysis increment" "could be view" $\rightarrow$ "could be viewed"

C6179

P13986 L5 State what satellite

Figure 7. Can you explain why there is an apparent seasonal dependence of the bias?

Section 5.1.1

P13992, L14 The difference between two years of the analysis cannot be taken as evidence for a trend, even if these years are 'similar'. I suggest starting with actual trend analysis (Section 5.2) and then discussing the 2012 – 2005 difference as an illustration and without calling it itself a trend. That would mean swapping Sections 5.1 and 5.2 and the corresponding tables. It's important to get this right. The trend analysis and the regression on various predictors are the two key results of the study.

P13993 L8. 'positive trend becomes significant' – significant in what sense? Statistically? You can't claim that based on a two year difference.

P13995, L19 onward. Since there are multiple predictors it is appropriate to do multiple regression analysis as it is done on PP13997+. I'm not sure if there is a point in analyzing pairwise correlations with meteorological and economic indices separately. I would start with multiple regression, base the entire analysis in this section on its results, and drop the individual correlations altogether. This would make the section more compact, easier to follow and less redundant. Also note that some indices are not independent (temperature and precipitation) as shown in Table 7. Again, There are some important findings there and they should not be buried under a lot of redundant numbers.

P13997 , L10. 'A multiple regression model using a stepwise-like procedure'. Please be more specific and describe the procedure in some detail, if possible provide a reference, not just the name of the algorithm.

P14002, L6 'estimate of the two main components of smog' I would say 'two of the main components'. There are other main components such as NOx

---

C6180