Atmos. Chem. Phys. Discuss., 14, C3145–C3155, 2014
www.atmos-chem-phys-discuss.net/14/C3145/2014/

Atmospheric Chemistry and Physics
Open Access
Discussions

Interactive comment on "Air quality simulations of wildfires in the Pacific Northwest evaluated with surface and satellite observations during the summers of 2007 and 2008" *by* F. L. Herron-Thorpe et al.

Anonymous Referee #1

Received and published: 3 June 2014

General:

The study represents a lot of useful work comparing the output of a family of fire models with pollution observed both at surface sites and from satellites. It is of interest to ACP readers. Thus, I encourage its publication. I read the paper once and made specific or general comments in chronological order. I added some general comments on the front end here. Normally I would read the paper again and organize the review better, but I will save that for after the major revisions that are needed. Therefore, this review

is not comprehensive and just gives a flavor of the type of revisions needed. Much of the language is casual or informal to the point of being unclear and the paper needs a few more iterations to result in a more precise scientific presentation and with more attention to the big picture.

The authors might want to include some examples of the severe wildfire pollution impacts that have occurred in the NW US in the introduction (plenty were available from the media during the summer of 2012 and 2013). I get the sense that people's mitigation options are somewhat limited so that is a reason to possibly emphasize the broader significance of this effort beyond wildfire AQ forecasting. Part of the goal of the system is to prevent excessive use of prescribed fire at sensitive times. That is not as relevant in summer, but could be mentioned. In the most general sense, it is one test of the ability to represent one of the most important phenomena on Earth with a reasonably complex model infrastructure. Much of the AIRPACT input data is only available in the US, so one can also compare the results from input available globally to results from more detailed input only available (or only reasonably accurate) in the Western US. As a member of the broader atmospheric community, the more general model-measurement comparisons are actually what interest me the most.

A few questions pop up naturally that might be clarified easily with a few words at various points in the text: Why 2007-8? Why AIRS and not other CO? Why not AERONET? Why not O3 from OMNI satellite? Any ideas for future use of VIIRS, GOES-R, or assimilation? The uncertainty in satellite retrievals needs to be factored into the overall philosophy. It should be clarified if the model runs discussed here are the original forecast (it seems not) or a reanalysis with improved fire data that became available later (my current understanding). Is the meteorology the original forecast or was the model run again with the "actual" meteorology? All these choices are of interest and the difference the choices make is also of interest. Specific/general comments in order.

P2, L3: "a suite"

P2, L25-27: Are these "biases" significant given reasonable estimated uncertainty in the remote sensing products?

P3, L5: Wildfires are not just forest fires. Much or most of the PNW is grassland, which also has large fires.

P3, L7: change "respiratory" to "health" since cardiovascular impacts actually dominate.

P3, L7: I don't think the goal is to "alert" people that the AQ is bad, but too forecast bad AQ ahead of time.

P3, L14: Maybe change "potential health" to "air quality" - the actual health impacts from a given air quality adds another much larger layer of uncertainty. L15: not just PM so maybe the text after the comma should just be: "but the task is challenging." P3, L16, Column measurements from space are useful to compare with models, but they have uncertainties and because they are column measurements there is really no such thing as satellite retrievals of AQ yet. See Crumeyrolle, S., Chen, G., Ziemba, L., Beyersdorf, A., Thornhill, L., Winstead, E., Moore, R. H., Shook, M. A., Hudgins, C., and Anderson, B. E.: Factors that influence surface PM2.5 values inferred from satellite observations: perspective gained for the US Baltimore–Washington metropolitan area during DISCOVER-AQ, Atmos. Chem. Phys., 14, 2139-2153, doi:10.5194/acp-14-2139-2014, 2014. Surface measurements are where the people live, but the column satellite data is useful to "connect the dots" between surface observations and evaluate overall model performance. I'd maybe express this as something like: "Satellite-based column measurements enhance the coverage available from surface networks and are useful to evaluate model performance."

P4, L5: change "led to" to "combined with" don't think a dry spring causes a summer drought

P4, L18: What is meant by "the south" and should ID/MT also be exceptions given the

text on L23?

P4, L20: Would ARCTAS CARB data be of any value in AIRPACT evaluation for 2008?

P5, top: It's not necessary to name all the fires here or in Figure 1.

P5, L15: I would just show all the burned area in Fig 1 with no names since several smaller fires could be just as important as one big one.

P5, L24: Why project to 2005 instead of 2007/8? Could you evaluate the EGAS software by projecting and then comparing to the 2011 NEI?

P5, L26: change "over" to "from" or say "Canadian anthropogenic emissions are . . ."

P6, L2-3: Maybe a word or two to clarify what is meant by "processing" emissions?

P6, L28: change (jargon) "ICS-209" to "fire"

P7, L2: clarify "well"

P7, L4-6: Here and in general. This sounds like a partial re-analysis – in other words, not testing the original forecast, but testing an improved forecast using updated fire info, but still with the old meteorology? It should be clear what was done and justified why. It would be of interest to know the accuracy of the original operational forecast. From the broader perspective how does actual vs original fire change in magnitude, location, timing, and how does that impact the modeled results? Also, how are fires forecast? In other words SMARTFIRE incompletely tabulates past fires if I understand right. Is that partial fire activity assumed to persist to generate a forecast? A sentence could clarify this.

P7, L7-16: It's my understanding that none of these models have ever been validated, but in any case, the extent to which they have should be provided. For instance, on line 14, combustion "phases" are referred to, which don't actually exist on real fires that burn with a mix of flaming and smoldering.

P7, L16: “short-lived” fuels makes no sense.

P7, L19-20: How about just saying the 60% is fixed in the model, but real fuel consumption can vary about the nominal value?

P7, L27-8: It doesn't seem to make sense to release all smoldering emissions into surface layer when it is well-known that smoldering emissions are entrained into convection columns and can go to any altitude the column does. I guess the paper sort of verifies that, so OK.

P8, L10-11: Change “most” to “much” Aqua retrievals are useful, but they are only offered in areas with no clouds and not so much smoke that the cloud mask thinks it is a cloud. Retreivals with estimated uncertainty above a threshold are rejected, but the remaining ones are known to be biased low compared to AERONET and MISR: T. F. Eck, B. N. Holben, J. S. Reid, M. M. Mukelabai, S. J. Piketh, O. Torres, H. T. Jethva, E. J. Hyer, D. E. Ward, O. Dubovik, A. Sinyuk, J. S. Schafer, D. M. Giles, M. Sorokin, A. Smirnov and I. Slutsker, A seasonal trend of single scattering albedo in southern African biomass-burning particles: Implications for satellite products and estimates of emissions for the world's largest biomass-burning source, Journal of Geophysical Research: Atmospheres, Volume 118, Issue 12, 27 June 2013, Pages: 6414–6432, DOI: 10.1002/jgrd.50500

P8, L19: I'm not questioning the decision to re-grid by grabbing closest value instead of e.g. weighted averaging or more complex re-mapping, but if possible one should estimate the additional contribution of this step to overall uncertainty?

P8, L25: Is it clear that between the uncertainty in re-gridded MODIS AOD and the model calculated AOD that a statistically significant comparison results? It is of interest and value to report and discuss comparisons even if they are not technically “significant,” but it would be helpful to be able to compare uncertainties to biases etc.

P9, L7: Section 2.3 If AIRS CO is one of the many CO products with low sensitivity

to boundary layer CO that should be mentioned. One could consult the Kopacz et al 2010 ACP paper for an idea of the accuracy of individual CO products as opposed to combining them all.

P9-10: Section 2.4 does well to mention that uncertainty occurs due to a potentially inappropriate a-priori vertical profile and air mass factor effects for the OMNI NO2. In addition, the effect of all the data massaging, and the fact that NO2 is rapidly converted to PAN and nitrate in fire plumes, which may not be represented correctly at all times in CMAQ could be mentioned http://www.atmos-chem-phys.net/12/1397/2012/acp-12-1397-2012.html http://www.atmos-chem-phys.net/10/9739/2010/acp-10-9739-2010.html

P10, L27-28: I think the aerosol typing depends on depolarization ratio, but not sure about the attenuated backscatter, altitude, location, and surface type. In general, for the remote sensing instruments, there is too much basic info on things like launch date and principle of measurement and not enough on accuracy and coverage.

P11, L24: It makes sense to compare the plume rise above the deresolved elevation in AIRPACT to the more accurate plume rise a.g.l. measured by CALIOP, but even this is tricky if the terrain is not flat or for whatever reason the actual plume height (a.g.l. or m.s.l) is not constant. But I suspect the CALIOP plume heights are one of the more exact comparisons possible though some vertical uncertainty could be estimated.

P12, Sect 2.6: If archived, the GOES visible and “fire channel” is very helpful for understanding fire timing or the temporal profile of emissions.

P12, L24: Is it better to say Canada is outside the AIRPACT domain rather than AIRPACT has no fire emissions in Canada? Canadian fire emission can impact the US AIRPACT domain and are ostensibly provided by MOZART in the boundary conditions.

P12, L25: If AIRPACT simulated a doubling or more of PM2.5 and a data set showed no increase during “fire events” - this seems like a case that is important to include in the comparison. It also seems like these events are discussed later in the paper? Or is that only at sites where these events were occasional rather than universal?

P13, L1-2: This doesn't make any sense to me. Aren't you comparing the model produced column to the satellite product column (in Fig 2-7) and why would that be restricted to places with surface sites?

P13, L5-8: Interesting and useful idea to select and compare separately only the cases where both model and satellite are strongly elevated. The restriction could conceivably inflate the degree of agreement, but it also potentially selects for higher S:N and lower uncertainty in the satellite product!

P14, L3: Here is one of many places where I wonder if AIRPACT really "under-predicted" or was AIRPACT actually similar within combined uncertainties, or if the satellite over-predicted, etc. It seems better to just consistently refer to differences (like in the figures) or offsets rather than imply value judgments, except maybe against CALIOP? Also there should be some definition here that is relevant throughout the text that specifies what you mean by "agreed well" as opposed to under/over-predicted? E.g. is within +/- 20% OK? Good work, but deserving of more precise terminology.

P14, L5-7: This is a good example of a difference that is not an "over-prediction" by the model, since the modeled fires really happened. Thus, this work is also simultaneously evaluating the remote sensing products.

P14, L9: In light of above; "performance" might be better as "agreement" and over/under prediction as "differences."

P14, L9-10: Were fire locations predicted? Or were they modeled retrospectively? Not sure what is being done here. Were the previous day's fires from SMARTFIRE assumed to persist? If the fire popped up after the satellite, how was it predicted despite the satellite missing it? It seems like you are referring to a model run done after the fires using the updated fire information. L10 What is meant by "intensity"?

P14, L15-17: So this is interesting once you've defined what you mean by: 1) "event," 2) how you compare to an "event," and 3) "well". And "large" could be replaced with a ">X km" definition? It seems 100 km is adopted later?

P14, L26-27: High model NO2 values could result from the model lifetime being too long in some plumes? L28 "under-biased" = "biased low"?

P15, L1-3: Use fractions or percentages, but don't mix in same sentence; and what is the significance of "140 sites" for column and model data??

P15, L3: If NO2 was 39% low on average despite being "over-predicted" 48% of time, then it seems there must have been a few massive "under-prediction" events?

Table 1a footnote: misspelled "source". Regional totals more useful that USA totals.

Table 1b title L4: "approximate" to "approximation" and, in general, what is meaning of a range of ignition dates?

P15, L6-9: Noting that the mismatches for AOD and NO2 are larger when there is more "signal to noise" but CO still seems to agree "pretty well."

P18, L7: Fires can entrain surface soil, dust, and ash into the convection column along with the smoke. L11: Missing a word?

P18, L19: Low O3 due to the CMAQ-SAPRC chemical mechanism might be expected (Alvarado and Prinn 2009 in JGR).

Page 20-21: In comparing to MBO PM several factors suggest the AIRPACT-modeled SOA is too low (e.g. the usual lower modeled-AOD in plumes longer than 100 km). However, the MBO PM is based on scattering and scattering can sometimes increase without an increase in mass, most likely due to a change in the size distribution (Akagi et al 2012 link given earlier). In general SOA is highly variable and poorly understood (Vakkari et al., 2014 in GRL).

P21, L21-22: I don't expect the global model to capture spikes. Is it really possible to differentiate between how well the fire emissions are represented in MOZART and how transport, the chemistry mechanism, or resolution effect the comparison? If a problem with the MOZART emissions can be demonstrated it should be described at least semi-quantitatively as a problem with amount, timing, or whatever it is rather than saying "poor."

P22, L15: It would be interesting to try some model runs with the plume rise offset consistent with the CALIOP-measured underestimate. The CALIOP/AIRPACT plume height comparison is not highly correlated, but this could still be tried. Are there plans to fix plume rise by scaling, etc?

P22, L27: This is the first mention of overestimation of plume height a.g.l. Is this referring to just a handful of cases?

P23, L12: change "when detecting" to "near many" since OMNI doesn't "detect fires" and some fires do inject emissions near the surface.

P23, L11-16: A number of recent papers take the OMNI NO2 retrievals in smoke as an accurate basis for global NO2 emissions estimates and so is this accuracy being disputed? Also is the AIRS a-priori CO profile any better suited for fire conditions? AIRS-CO seems consistent with AIRPACT.

P23, L18-19: Does "but there were often similar estimates of column CO over active fire regions" mean good agreement on column CO "often" occurred between AIRPACT and AIRS above fire locations. Also, are there any useful surface observations of CO?

P23, L19: By "The AIRS retrieval is not sensitive to the surface" do you mean literally that it is not affected by land cover type, or that it has low sensitivity in the boundary layer, or something else?

P23, L27-28: Probably good to cite some papers relying on more advanced measurements of PM that find SOA is highly variable: e.g. from none at all to a factor of four (Jolleys et al. 2012 in ES&T; Yokelson et al. 2009 in ACP; Vakkari et al. 2014 in GRL).

P24, L5-9: This seems like an important result and should probably be developed/integrated into text and tables more fully, rather than appearing almost as an afterthought at the end of the paper.

P24, L15: Do you actually mean that some of the fires in the historical SMARTFIRE database don't exist?

P24, L20-24: How would complex terrain or cloud cover cause SMARTFIRE to miss fires when it includes the ICS-209s? Maybe wilderness fires that no report is filed on? How would a lack of dead woody fuels cause a fuel loading underestimate? Maybe change "that completely lack dead woody fuels" to "for which dead woody fuels are omitted"

P24, L27-P25, L1: By under-predicted emissions, do you mean total emissions as opposed to certain species? Why would emissions scale with plume heights? Is buoyancy assumed proportional to amount of fuel burned?

P25, L3-20: This reads like confident conclusions about the benefits of model changes that were not discussed in the paper or tested except for number 4, and it omits the thing you did demonstrate needs fixing: the plume height. Suggest presenting this as a list of additional (in addition to plume height) future avenues to explore for potential improvement.

Table 2: Shouldn't the formulas for percentages require a 100 instead of a 1 as first number? Also, why is the bias and error sometimes computed with respect to the observation and sometimes with respect to the mean of the model and observation? If the model and observation are equally valid then the concept of model performance or "under or over" prediction throughout the text seems less meaningful. In Table 2, the normalized quantities are defined as percentages, but then not used as percentages in Table 3.

Table 4, Title: I thought both the satellite and AIRPACT data are rural only in this comparison. How about instead of “performance” in sentence one and including sentence two, just say “Summary of matched threshold comparison limited to polluted rural sites for 3 July . . ..”?

Figure 1: could be better without fire names.

Figure 2 caption, L4: “and” before “exclusion (also in rest of similar figures).

Figure 9: SMOKE model seems to reduce false positive events. I thought this was showing a single site at first and now wondering if the PM spikes that occur even when averaging over all the sites are due to massive modeled impacts at a few sites closer to fires?

Interactive comment on Atmos. Chem. Phys. Discuss., 14, 11103, 2014.