Atmos. Chem. Phys. Discuss., 12, C8818–C8826, 2012
www.atmos-chem-phys-discuss.net/12/C8818/2012/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "A multi-instrument comparison of integrated water vapour measurements at a high latitude site" *by* S. A. Buehler et al.

S. A. Buehler

sbuehler@ltu.se

Received and published: 31 October 2012

We thank the reviewer for the constructive comments. Below, reviewer comments are in italics and marked by '***R:***', our response is in normal font and marked by '**A:**'.

General comments

R: *This manuscript compares integrated water vapour (IWV) measurements obtained from five different sensors and the ECMWF reanalysis (ERA-Interim). The authors highlight two main objectives. The first is to characterize the systematic and random errors of the sensors used at their subarctic Observatory (Kiruna, Sweden). Their*

aim is to built a long-term data record for climate research. The second objective is to assess the impact of two specific issues arising in such an intercomparison study: the difference in lower altitude limit for the IWV integration and the representativeness error. These objectives are of special interest to the author's work based on the Kiruna datasets. They are also of general interest to the climate community interested in monitoring water vapour. This manuscript hence fits well the topics of the WAVACS special issue hosted by ACPD journal.

Better quantifying the accuracy of the different observing techniques will give insight into the quality of present climate diagnostics derived from those measurements and help improving the observing systems. To this respect, the manuscript provides valuable results for the Kiruna site which could probably be transposed to some extent to other global climate observing sites (at least in similar subarctic climates). The methodology used therefore is sound in using a regression algorithm that takes into account errors on both datasets and considering specific temporal and spatial scales for each dataset for the temporal matching.

As for the secondary objectives, they are not new and the manuscript does not improve our knowledge about them. The impact of lower altitude limit is a well know limitation which has been corrected for in most past studies (at least for the past 10 years, to my knowledge).

A: We disagree with the reviewer here. Of the 10 studies that we review in the article (see Table 6), only one (Bock et al. 2007) mentions this issue. It therefore seems to be far from general knowledge. Furthermore, the other reviewer explicitly commended that aspect of the article. We would therefore like to keep the discussion on the lower integration altitude in the article.

R: *On the other hand, the representativeness error is a major limitation when point measurements and areal averages are compared. This is the case in this work and the authors propose two methods to assess this error source. First they compute statis-*

tics for the difference of distant IWV values extracted from a high resolution (3.5 km horizontal) cloud resolving model. This approach might provide an estimate for the representativeness error when to non-collocated IWV measurements are compared if computed over a statistical representative ensemble. Unfortunately, the authors use a single model simulation and extract IWV from a global latitude band which is certainly not representative of the time variability of the local climate. The second method is based on a similar approach (though this is not very clear from the manuscript) using AMSU-B data. Both methods give consistent rather high estimates of representativeness errors which lead the authors to the conclusion that representativeness error dominates the random errors between theirs datasets. To my opinion, these results of representativeness errors are not accurate enough to draw such a conclusion.

Therefore I recommend that the discussion of lower altitude limit be reduced and that the discussion of representativeness error be revised according to the specific comments given below. Some minor corrections and remarks are added as well for the revision.

A: We have followed most of the reviewer's suggestions. However, we would like to keep some of the content that the reviewer suggests to reduce. See text below for point by point responses.

Specific comments

Lower integration altitude limit:

R: *The two figures (fig. 3 and 4) are probably not necessary as the method is not new and can be well understood from the text alone.*

A: As argued above, we would like to keep the figures, since we believe that both the lower integration limit issue and the correction method that we use are far from generally known.

R: *The text of section 3.4 could be easily shortened by limiting it to the description of the method and giving the empirical relationship used to correct the data.*

A: We feel that the details here are necessary for those readers that are not already familiar with the correction. However, besides mentioning that Bock et al. 2007 made a similar correction (which we already do in the ACPD version), we now explicitly compare our result to theirs. Text added was: "Using the terminology of Bock et al 2007, we can say that the relative bias ΔIWV/IWV is -3.5% per 100 m altitude difference.". At the end of the section, where the possible applicability of the correction to other locations is discussed, we added: "For example, Bock et al. 2007 find a value of -4.0% per 100 m for Africa, a five percent stronger correction than in our case.".

R: *It is not obvious to what range of altitude difference the empirical relationship would work.*

A: Yes. We added the following text at the end of the section: "It is also not obvious for what range of altitude differences our simple correction scheme would work, and also this may be location dependant. For Kiruna our results in Figure 4 show that the correction works at least for altitude differences up to 250 m."

Representativeness error:

R: *Section 3.5 refers to O'Carroll et al., 2008, but while reading I could not find where the authors follow the methodology proposed by O'Carroll for analyzing the error sources of their datasets.*

A: Yes, the early reference to O'Carroll was confusing and has been removed. However, there still is a reference later in the text, see answer to next question.

R: *First, O'Carroll's method is based on a three-way comparison of datasets for deriving estimates of the variance of random error for each of the datasets. The authors of the* *present manuscript do not attempt this though I guess it would be very valuable to quantify individual errors in addition to the differences of datasets.*

A: We did in fact perform a three way error analysis on our data. We systematically tried all three-way dataset combinations, and even combinations of more than three datasets. But the results were non-conclusive. For example, the same dataset could be assigned a small error in one three way combination, and a large error in another combination. According to our analysis, the problem is that the errors in the different observations are correlated, largely due to the large errors of representativeness. Our internal conclusion is that three way error analysis only makes sense if either all three datasets are on the same spatial and temporal scale, or if one has a reliable model for the error correlations. Interestingly, these fundamental issues with the three way error analysis method will not be noticed if one only has exactly three datasets, since it is then not possible to do any consistency checks.

All in all, the results with the multi way error analysis were so discouraging that we decided to not included them in the manuscript. However, triggered by the reviewer questions, we now added at least a bit of new text on this. The new text reads:

"Note, that our concept of representativeness error is slightly different from how it is presented for example by O'Caroll et al. (2008). In their three way error analysis, they assume that each dataset has a representativeness error relative to a common 'truth'. In contrast to this, in our analysis we define the representativeness error of all datasets relative to the GPS dataset, so the GPS dataset itself has $\sigma_{\mathsf{RepXY}} = 0$ by definition.

Representativeness errors for different datasets may be highly correlated, if they have similar sampling characteristics. For example, if the truth is a point value, then all areal estimates will have highly correlated representativeness errors. The presence of significant representativeness errors for our study leads to the consequence that we cannot carry out a formal multi way error analysis in order to separate the errors of the individual datasets. Instead, we simply present the results relative to the GPS

dataset. However, at least we make an attempt to estimate the representativeness errors relative to the GPS, as described below."

***R:** Second, O'Carroll introduces a strong assumption on the independence of errors which is difficult to prove in the presence of representativeness errors. In view of these two points I recommend to remove that part of the text which tries to make a parallel with O'Carroll's publication.*

A: We agree with the reviewer that the errors here are not independent, and therefore three-way error analysis is not appropriate. As described in the answer to the previous comment, we now included a short discussion of this in the text.

***R:** Otherwise, the introduction of representativeness error (lines 4 to 24, page 21030) is satisfying.*

A: Thanks.

***R:** The part of section 3.5 using the NICAM simulations is very debatable in its present state of achievement and would be better removed for now. Moreover, nothing is said about the realism of the IWV field simulated by the model.*

A: The NICAM simulations provide only a very rough estimate of the representativeness error, which we clearly state in the article. (Quoting from the article: "To summarize, we made rough estimates of the representativeness error between the GPS and the other instruments from NICAM model data. The method used a lot of assumptions, the most important ones being [...]. None of these assumptions are strictly valid, so a rough estimate is all that the method can provide. However, we did validate the NICAM-based method with AMSU data for one case (horizontal distance) and found encouraging consistency.".)

We believe that this estimate, even though crude, is still valuable, and much better than giving no estimate at all. We would therefor like to keep it in the article. (See also answer to next question.)

***R:** The part using AMSU-B data might be a good alternative to using the NICAM simulation given that the IWV data used to compute the statistics are representative of the time variability of the local climate.*

A: We agree with the reviewer that it is the more straightforward method, therefore we used it for a sanity check of the NICAM based method. But we cannot use it to test anything else than horizontal distance, because AMSU-B data has significantly poorer horizontal resolution than NICAM data. So, the AMSU based method cannot replace the NICAM based method.

***R:** The text just mentions that data along the satellite track for the year 2008 is used but it is not clear whether the data was limited to a latitude / longitude box around the study site? I recommend that that part of the study be clarified or improved if it is to remain in the manuscript.*

A: The AMSU-B data is exactly centered on Kiruna (Kiruna is at distance 0). We slightly modified the text to make this clearer. (We now write "as a function of the distance from Kiruna" instead of "as a function of the distance".)

Minor corrections and remarks

On GPS results:

***R:** The authors write in section 2.1 that the radome covering the GPS antenna is cleaned when it is covered with snow. To what extent does the accumulation and removal of snow impact the IWV estimates?*

A: We cannot answer this question. The main reason to mention snow specifically in

the text is to make the point that the snow is supposed to be regularly removed by Esrange staff.

***R:** The authors write also that the antenna PCV can be calibrated but quote a possible 1 kg/m2 bias in their dataset because the data were not analyzed with the proper model. Could this explain the bias in the GPS IWV mentioned in the discussion section (5.1)?*

A: In principle yes, or at least partly. However, according to our understanding, an antenna phase center offset should give more or less a constant bias, so the GPS should then hava a moist bias even for high IWV values. But there some other datasets are moister, others drier. So, while it is likely that there is an antenna phase center offset, our comparison data from the other instruments are not consistent enough to really identify it.

***R:** Other limiting factors are mentioned in section 5.2.1. Could the authors give an idea of the order of magnitude of these biases?*

A: We expanded the text with some number for the GPS factors.

On radiosonde results:

***R:** Three different radiosonde types are used in this study, but the authors do not distinguish results as a function of radiosonde type, though it is well known from the literature that radiosonde biases are sensor-dependent as well as dependent on time of launch (day-night difference). Could the authors present separate statistics or justify about their choice in this respect?*

A: We agree with the reviewer that it would be interesting to study this. But we believe that our radiosonde dataset is too small to subdivide it and still get statistically significant results.

On ERA-Interim data:

R: *The authors write that "ERA-Interim grid point data are not area averages, but are valid at the exact location of the grid points indicated by the grid." To my knowledge this is wrong. The model fields are areal values representative of the grid length-scale. Remember that when extracted on a 0.75°x0.75°, model fields are filtered and interpolated from the native grid.*

A: Yes. We removed the confusing statement (in two places, Section 2.6 and Section 3.5).

On FTIR results

R: *The authors mention that "FTIR and microwave data are suitable if and only if instrument and algorithm are kept the same over long time periods" (section 5.3). I wonder if the raw data from these instruments could be reprocessed over long time periods using a fixed algorithm to overcome at least of these limitations?*

A: Yes, this is a very good point. We have modified the text accordingly, the paragraph in question now reads: "FTIR and microwave data are suitable if and only if the instruments are kept the same over long time periods. Besides changes in the instruments themselves, changes in the algorithms for the data processing are also problematic. The different alternative sets that we studied (FTIRa/FTIRb and the different microwave datasets) show that processing changes easily introduce systematic differences in the data (compare the two green curves in Fig. 9 for the FTIR data, or the two dark-blue curves for the microwave data. On the other hand, this issue can be solved by reprocessing the raw instrument data at a later time. It is therefore crucial that raw instrument data and auxiliary data are carefully archived."
