Atmos. Meas. Tech. Discuss., 8, C2148–C2157, 2015
www.atmos-meas-tech-discuss.net/8/C2148/2015/
© Author(s) 2015. This work is distributed under
the Creative Commons Attribute 3.0 License.

Atmospheric Measurement Techniques
Discussions

Interactive comment on "Consistent evaluation of GOSAT, SCIAMACHY, CarbonTracker, and MACC through comparisons to TCCON" *by* S. S. Kulawik et al.

Anonymous Referee #1

Received and published: 20 July 2015

1 Overview:

Review of "Consistent evaluation of GOSAT SCIAMACHY, CarbonTracker, and MACC through comparisons to TCCON" by Kulawik et al.

Kulawik et al. present a slew of satellite–TCCON and model–TCCON comparisons. Most notably, they found random errors for of 0.9, 0.9, 1.7, and 2.1 ppm for CT2013b, MACC, GOSAT, and SCIAMACHY through comparison with TCCON. They also attempt to quantify seasonally-dependent biases in the models and satellites. The results of these comparisons could help to inform future flux estimate studies. However,

I think the manuscript needs substantial revisions before it is in a publishable state.

1. Does the paper address relevant scientific questions within the scope of AMT? **Yes**
2. Does the paper present novel concepts, ideas, tools, or data? **Yes**
3. Are substantial conclusions reached? **Yes**
4. Are the scientific methods and assumptions valid and clearly outlined? **No**
5. Are the results sufficient to support the interpretations and conclusions? **No**
6. Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)? **No**
7. Do the authors give proper credit to related work and clearly indicate their own new/original contribution? **Yes**
8. Does the title clearly reflect the contents of the paper? **Yes**
9. Does the abstract provide a concise and complete summary? **Yes**
10. Is the overall presentation well structured and clear? **No**
11. Is the language fluent and precise? **No**
12. Are mathematical formulae, symbols, abbreviations, and units correctly defined and used? **Yes**
13. Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated? **Yes**
14. Are the number and quality of references appropriate? **Yes**
15. Is the amount and quality of supplementary material appropriate? **Yes**

2 Major comments:

Manuscript is confusing and too long
In it's present state, the manuscript is far too long, convoluted, and confusing to be accessible to the scientific community. There are numerous spelling and grammatical errors, acronyms are introduced multiple times or not at all, figures are confusing, and much of the content seems unnecessary. In particular, Sections 3–5 need substantial revisions.

Here are some examples of confusing or unnecessary content (also see the minor comments):

Figure 2: What data is being shown in the different panels? The caption is not helpful. There are multiple things labeled "top" in the caption, why are you showing different SH TCCON sites for models and satellites, why is there a spike in TCCON data in panel "b1" & "b2" but not "a1" & "a2" (isn't the same TCCON data plotted)?

Figure 5: What does the y-axis label mean ("Stddev vs. TCCON")? What are the units? Is this a ratio?

Figure 9: Does this figure add anything? It was only mentioned once in the main text (Page 6236 Lines 6–7: "Plots are individually examined to ensure that there is adequate data (e.g. see Fig. 9)."

Table 2: Redundant. It seems that all of the information is already presented in the main text.

Table 3: Many of the concepts mentioned in the table are not discussed in the main text at all. For example, it seems that Kulawik et al. estimate a "co-location error" in Table 3 but it's totally unclear how this is done. It seems the authors are also sampling CarbonTracker output at satellite and the TCCON locations (not sure though, it's not explained in the text). CarbonTracker is very coarse ($3^\circ \times 2^\circ$), it doesn't seem like it

would be high enough resolution to resolve any of the variability between the satellite location and TCCON location. . .

Table 6: Seems unnecessary.

Page 6231, Lines 9–15: "The year-to-year variability in the bias could be partly attributed to the distribution of data seasonally. Stations which have absolute biases more than 0.3 ppm different than the mean bias therefore have biases that are persistent from year to year. The stations which do not show biases are: GOSAT: Bialystok, Karlsruhe, Lamont, Izana. SCIAMACHY: Lamont. CT2013b: Ny Alesund, Orleans, Izana, Darwin, Wollongong, Lauder (both). MACC: Ny Alesund, Orleans, Park Falls, Lamont, Izana, Darwin, Wollongong, Lauder (both)." Confusing. Are these stations that don't have persistent biases but may have seasonally dependent biases? The authors mention that the seasonal distribution of the data could be an important factor in the bias then don't mention which satellite/sites are adversely affected by this.

Section 3.4: Again, it's unclear exactly how all the error terms were derived. Table 3 lists co-location error and Page 6233 discusses "co-locations error" for the different co-location methods and satellites but does not discuss how that was derived.

Section 3.5: Is it necessary? It's confusing and it seems that the authors do this again later (Page 6235, Lines 12–13: "Looking ahead to Section 4 and Table 4, SCIAMACHY overestimates the seasonal cycle amplitude from 0–45°N by $\sim$1.2 ppm.").

Deriving error statistics for use in flux estimates
The main goal of the manuscript is to derive random errors, systematic errors, and error correlations for use in flux estimates. A major thrust of the manuscript is the derivation of the a and b parameters for the empirical error model: $\epsilon^2 = a^2 + b^2/n$ (Eq. 2 in the manuscript) through comparison of satellite observations "co-located" with TCCON sites. The authors argue that this error model "should help assigning realistic retrieval error correlations in assimilation systems in place of current ad hoc hypotheses (see, e.g., Sect 2.2 in Basu et al., 2013, for an example of such hypotheses)" (Page

6244, Lines 24–26). However it's unclear how this error model could actually assist in specifying error correlations in assimilation systems.

Error correlations used in a flux estimate include contributions from measurement, model, and representation error. The "ad hoc" methods the authors hope to replace are used to characterize the total observational error covariance matrix $\mathbf{R}$ (including contributions from all error types). The "correlated error" only seems to be valid for their simple model used here (the geometric or dynamical colocation model). The error model the authors present here seems to only quantify the diagonal term of the measurement error.

Thus, it doesn't seem that this error model would be substantially better than the "ad hoc" specifications of the $\mathbf{R}$ matrix used by some past work because it only gives an estimate of the diagonal elements for the measurement contribution. It doesn't actually help specify the off-diagonal components of the $\mathbf{R}$ matrix or the other error terms (which are the difficult part to quantify!). It should also be noted that there are methods of using atmospheric data to characterize the $\mathbf{R}$ and $\mathbf{B}$ matrices *a priori* (e.g., Michalak et al. 2005; Miller et al. 2013) or jointly estimate the associated hyperparameters in the inversion (e.g., Ganesan et al., 2014, 2015).

GOSAT & SCIAMACHY comparisons at Eureka, Ny Alesund, and Sodankyla
On Page 6230 Lines 12–14 the authors state, "For the bias we take out stations poleward of 60°N, which have large positive biases for GOSAT and SCIAMACHY, which we note as an issue." Why were these removed? Because there was a bias? Isn't that important? Is this an issue with GOSAT and SCIAMACHY at high latitudes, a sampling issue, or something else?

Statistical tests
There are multiple cases where the authors claim statistical significance without stating a confidence level, p-value, or any other metric of statistical significance. Did the authors perform the appropriate statistical tests? It seems

that the authors are claiming statistical significance when the value is outside the error bars of TCCON, however this is not necessarily indicative of statistical significance (see https://egret.psychol.cam.ac.uk/statistics/local_copies_of_sources_Cardinal_and_Aitken_ANOVA/errorbars.htm). The authors can only claim statistical significance if the null hypothesis is rejected in a t-test, z-test (for sufficiently large sample sizes), F-test, or an ANOVA. Here are a few examples where they seem to have incorrectly claimed statistical significance:

Page 6229, Lines 7–8: "When the measured biases are larger than the gray box [TCCON Bias Uncertainty], they are considered significantly different than TCCON".

Page 6234, Lines 27–28: "Presumably when the bias is larger than the error bar, the bias is significant".

Page 6244, Lines 28–29 & Page 6245, Lines 12–13: "Biases vary by station (see Fig. 3); the station-dependent biases have a standard deviation of ∼0.3 ppm from year to year. Biases larger than ∼0.3 ppm likely represent persistent biases. . . The discrepancies versus TCCON which are statistically significant are that GOSAT has. . . ".

3 Minor comments:

Correlated error with GOSAT and SCIAMACHY
Page 6233 talks about the correlated error in SCIAMACHY and GOSAT. The authors claim that "averaging is more effective [for SCIAMACHY] when it is over a larger spatial/temporal area, probably due to variability in the source of the correlated errors." If this were true, why wouldn't it apply to GOSAT as well? "these stations [Lamont and Park Falls for GOSAT] have smaller correlated error for geometric matches, which is true in all seasons. This could be due to the smaller GOSAT footprint allowing more variability from observation to observation." The speculation about the GOSAT footprint seems highly unlikely.

TCCON sites with complex topography
Page 6241, Lines 1–7 talk about how certain TCCON sites may have topography that makes them bad for comparisons. The author's seem to make the sweeping generalization that This will obviously be resolution dependent (e.g., an LES model could resolve the topography), what resolution would be necessary to safely compare with these sites? Or conversely what resolution models should not compare with these sites?

Reorder Section 2
TCCON description should come before the satellite descriptions because the authors talk about how TCCON is used to evaluate the satellite observations.

Four Corners. The authors give the abbreviation "4C" for Four Corners (Page 6229, Line 13) then don't seem to ever use it.

Page 6221, Lines 11–13: "These findings also apply to bottom-up flux estimates, for example, updates should be made in inventories or transport to correct the model fields at the TCCON stations showing seasonal cycle phase differences." That's not bottom-up. EDGAR, VULCAN, and HESTIA are examples of bottom-up inventories for CO_2. They do not rely on transport. How do these findings (characterizing biases in models and satellite observations) apply to a bottom-up flux estimate?

Page 6221, Line 21: Should add Kuze et al. (2009).

Page 6222, Lines 23–24: X_{CO_2} was already explained in Section 2.1. Also the authors called it "column averaged dry air mole fraction" in Section 2.1 and "column-average dry-air mole fraction" in Section 2.2. Be consistent with nomenclature.

Page 6222, Lines 26–28: "This information is transferred to the CO_2 absorption band..." Odd way to phrase it. Why not simply say it's jointly estimated (or fitted). Rephrase.

Page 6223, Line 25: TCCON acronym was already introduced in Section 1.

Page 6224, Line 1: Citation or reference for the GGG software package?

Equation 1: Units seem inconsistent. Adding degrees and temperature? Are latitude and longitude in degrees or km? If it's degrees this could reduce the amount of data available at high latitudes (as the lines of longitude converge at the poles), could that affect the satellite biases at high NH sites?

Page 6229, Lines 4–5: "The black box shows 5 European stations which are very close, geographically, yet have different biases." Doesn't give me much confidence in the "co-location" methods. What would the bias be between the TCCON sites if you applied the dynamical co-location method to nearby sites (ie. use TCCON observations that observe the same airmass)?

Page 6229, Line 24: "Sodannkyla" site is misspelled.

Page 6230, Line 1: I thought CT had an ensemble, couldn't you use that to make the comparison "fair".

Page 6230, Line 5: "northern and Southern Hemispheres". Should use consistent capitalization.

Page 6230, Lines 15–17: "The overall bias is less of a concern than the bias variability in satellite data which indicates regional errors that will translate to regional errors in flux estimates." Couldn't some of that be due to sampling location? A Hi-res model could show the expected representation error.

Page 6232, Line 4: "of 2.3 ppm multiplied by 0.9° with the 2.1 ppm actual error." What is the degree symbol for? Typo?

Page 6232, Lines 7–8: "Figure 5 shows SCIAMACHY and GOSAT standard deviations versus TCCON for geometrical and dynamical coincidence criteria." What is actually plotted? What does versus mean in this context? Is this a slope, ratio, something else?

Page 6232, Line 23: "Errors versus averaging: random and correlated error". I thought you just said you were going to use dynamical for the remainder of the paper. This section then goes on to compare dynamical and geometrical criteria...

Page 6233, Lines 12–14: "The purple dashed line represents spatio-temporal mismatch error and as expected, this value is much smaller for geometric than for dynamic coincidence criteria." Why would there be more spatio-temporal error when averaging with dynamical?

Page 6233, Lines 20–26: "There is more..." From Fig. 6 it looks like geometric coincidence criteria always performs better (lower y-intercept and finishes with a lower error).

Page 6235, Lines 8–14: Use consistent tenses.

Page 6235, Line 22: "DFJ" should say "DJF".

Page 6236, Line 10–13: "The ocean/land behavior..., it does not seem correct to include it,...". Colloquial. Include it or throw it out and give a justification.

Page 6237, Line 23: "coloumns" should say "columns".

Page 6237, Lines 25–26: "Therefore, **the variability of the seen in Table 5** is primarily explained by the time-range of the comparisons." Grammar. Need to fix.

Page 6238, Line 9: "(Keppel-Aleks, 2012)" should say "(Keppel-Aleks et al., 2012)".

Page 6241, Line 17: "$2 \times 3°$" should read "$3° \times 2°$".

Page 6244, Lines 18–22: Fig. 6 made it look like geometric averaging was always better than dynamical averaging.

Page 6245, Lines 5–6: "...TCCON bias uncertainty is on the order of 0.4 ppm; the TCCON team is working to improve this." Unnecessary. Groups are always trying to reduce the uncertainty.

Page 6245, Lines 20–23: Seems excessive for the Conclusions. Just restate the find-

ings.

Page 6246, Lines 11–12: "However, TCCON daily variability has not been validated (there are plans to validate TCCON throughout the day in the near future)". Rephrase to "However, TCCON daily variability has not been validated."

4 References:

Ganesan *et al.*: Characterization of uncertainties in atmospheric trace gas inversions using hierarchical Bayesian methods. *Atmos. Chem. Phys.* **14**, 3355-3864, 2014.

Ganesan *et al.*: Quantifying methane and nitrous oxide emissions from the UK and Ireland using a national-scale monitoring network. *Atmos. Chem. Phys.* **15**, 6393-6406, 2015.

Kuze *et al.*: Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring. *Appl. Optics*, **48**, 6716-6733, 2009.

Michalak *et al.*: Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions. *J. Geophys. Res.* **110**, 2005.

Miller *et al.*: Anthropogenic emissions of methane in the United States. *Proc. Natl. Acad. Sci.* **110**, 20018-22, 2013.

Interactive comment on Atmos. Meas. Tech. Discuss., 8, 6217, 2015.