Atmos. Meas. Tech. Discuss., 8, C4258–C4274, 2015
www.atmos-meas-tech-discuss.net/8/C4258/2015/

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.

S. S. Kulawik et al.

susan.kulawik@jpl.nasa.gov

Received and published: 10 December 2015

Thank you to the reviewers for their comments on this paper. Each reviewer section starts with a ">" and the response does not. The PDF attachment has color coding for reviewer/responses in case there is confusion. Note also that during the review process I re-ran the analysis with the latest GGG2014 TCCON data. The previous version went through 10/2013 which cut out some GOSAT data (ACOS-GOSAT goes through 5/2014), CT (which goes through 12/2013, and MACC (which goes through 12/2014). Also, in response to a comment by reviewer 2, TCCON is now averaged in 90 minute intervals after rather than before the match-ups to satellites.

> 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.

The reviewer elaborates details in the subsequent comments, so we respond to each below.

> 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)?

Updated caption and text to address all the above. The caption now reads

Figure 2. Time series for matches of CT2013b, MACC, SCIAMACHY, and GOSAT versus TCCON at Lamont (panels a-d) and Lauder125 (for models) or Wollongong (for satellites) (panels e-h). The top plot of each set (panels a1, b1, c1, . . .) shows a time series of all geometric matching pairs. The middle panel of each set (panels a2, b2, c2, . . .) shows the difference versus TCCON, with the blue line the 30-day average difference. The bottom panel of each set (panels a3, b3, c3,. . .) shows a histogram of the differences, indicating an approximate error and bias.

I think that the text (panels a2, b2, c2, . . .) is extraneous to the description, "the middle panel of each set." however the reviewer indicates this is needed to describe the figure.

The text has added text to address the TCCON differences seen in the plots and two different SH sites in section 3, "Lauder is shown for the models to show the phase lag for CT2013b seen at this site in Table 7. Since there are not enough coincidences for the satellites at Lauder, we show Wollongong for satellites."

"As only TCCON/satellite matched pairs are shown, different subsets of TCCON are

AMTD
8, C4258–C4274, 2015

included for models and the two satellites."

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

Updated label to, "stdev (satellite - TCCON) (ppm)". The caption was also updated to read, "Standard deviation of SCIAMACHY and GOSAT minus TCCON for different coincidence criteria and number of satellite observations averaged, n, in the northern hemisphere."

> 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).”

Took out Fig 9.

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

Agree this is redundant but a useful resource for the reader to refer to. Moved to appendix A.

> 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_ _ 2_), it doesn’t seem like it would be high enough resolution to resolve any of the variability between the satellite location and TCCON location. . .

Added text describing Fig 6 and (now) Table 2. "The purple dashed line represents the standard deviation of CT2013b at the satellite time and location vs. CT2013b at the TCCON time and location and represents spatio-temporal mismatch error (co-location error). As expected, this value is much smaller for geometric than for dynamic coin-

AMTD
8, C4258–C4274, 2015

cidence criteria. The value for large n is shown in the CT-CT columns of Table 2 with N.H. averages of 0.3 ppm, 0.7 ppm, and 0.7 ppm, for SCIAMACHY geometric, SCIAMACHY dynamic, and GOSAT dynamic coincidences, respectively. There is not much difference between the estimated co-location error at North American sites of Lamont and Park Falls where the CT2013b model is at 1x1 degrees versus other N.H. sites where the CT2013b model is at 3x2 degrees. The co-location error is subtracted from the a value in quadrature to estimate results without co-location error."

> Table 6: Seems unnecessary.

Took out Table 6. Section 4.2 now has this additional text in place of Table 6. " To see how much of the observed variability in the growth rate is temporal vs. spatial variability in the growth rate, we compare to the global annual increase (growth rate) from surface measurements (http://www.esrl.noaa.gov/gmd/ccgg/trends/global.html) which are: 1.74, 2.1012, 1.7877, 1.6567, 2.4439, 1.7170, 2.4340, 2.5351, 1.86 89 ppm/year for global yearly increases 2006-2014, and 1.76, 2.22, 1.60, 1.8889, 2.4542, 1.8486, 2.6663, 2.0506, 2.13 17 ppm/year for Mauna Loa yearly increases 2006-2014, with error bars 0.05-0.09 ppm/year for Global and 0.11 ppm/year for Mauna Loa. ."

As the numbers in Table (now) 4 are derived from the above numbers it is important to record them as the values on the esrl website could be revised at a future date (as in fact the numbers have been revised between the submission of this paper and this revision).

> 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

AMTD
8, C4258–C4274, 2015

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.

Agree this is confusing. Simplified the conclusions from this brief analysis as no statistical analysis was done on the significance of the findings. This section now reads, "We test whether the biases seen in Figs. 3 and 4 are persistent from year to year. When at least two full-year averages exist for a station, the standard deviation of the yearly bias is calculated. The average over all stations of the yearly bias standard deviation is 0.3 ppm for all sets (CT2013b, MACC, SCIAMACHY, GOSAT)."

> 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.

The caption for (now) Table 2 is simplified with the description of the steps to calculate the co-location error and TCCON error added to the main text. The co-location error addition is discussed above. The TCCON error is now better described in the text, " Subtracting (in quadrature) the co-location error (CT-CT column in Table 2) and TCCON error of 0.35 ppm (Appendix B) results corrected correlated error, a, shown in the "Mean NH: subtr co-location error" row of Table 2. "

> 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 _1.2 ppm.").

Section 3.5 looks at the biases for 3 month intervals, a relatively simple analysis. Section 4 and Table (now) 3 fit the seasonal cycle amplitude, a less direct analysis. The seasonal biases are important for understand the findings in Table (now) 3. The "looking ahead" text is taken out with the findings described from Fig. 8 and now referred to from Section 4. Section 4 are the results using the CCGCRV fitting software. However, 3.5 and 4.1 results are combined for the conclusions.

AMTD
8, C4258–C4274, 2015

> 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: _2 = a2 + b2=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.

Added additional text in the discussion/conclusions, " For example, in Basu et al. (2013) observations within 500 km and 1 hour are assumed to have 100% correlated errors, and are inflated by a factor such that when observations are later treated as if the errors were random, the final error of the average is the same as the error of one observation. This can be improved by setting the inflation factor so that the average observation error is aˆ2+bˆ2/n with a and b set by the geometric values from Table 2, which should result in a lower error than assuming 100% correlation. "

> 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 [TC-

AMTD
8, C4258–C4274, 2015

CON 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. . . ".

The analysis was updated to include the t-test, as all analysis uses matched pairs with only two datasets compared at a time and in most cases small numbers of comparisons.

In section 3.2, the following analysis and text were added, "The gray bars labeled "TCCON Bias Uncertainty" in Fig. 3 are the overall calibration uncertainty in TCCON which is estimated to be 0.4 ppm (Wunch et al., 2010, 2011). The significance of the bias versus TCCON is estimated by the t-test (looking up bias / standard deviation * sqrt(number of comparisons) in a t-test table). The gray box is the uncertainty in the TCCON station-to-station bias and all comparisons with |bias| > 0.4 were found by the t-test to be significant. So from the above two pieces of information, all biases differing from TCCON more than 0.4 ppm are significantly different than TCCON. "

Later in section 3.2, any claims of significance on the year-to-year biases are withdrawn and the result is only presented, "We test whether the biases seen in Figs. 3 and 4 are persistent from year to year. When at least two full-year averages exist for a station, the standard deviation of the yearly bias is calculated. The average over all stations of the yearly bias standard deviation is 0.3 ppm for all sets (CT2013b, MACC, SCIAMACHY, GOSAT)."

The t-test was applied in section 3.5 to determine the significance of the Figure 8 averaged results.

In section 4.1 the errors in Table (now) 3 were redone to be more transparent, and the

AMTD
8, C4258–C4274, 2015

significance is according to the t-test. "When n>1, results are considered significant using the t-test (comparison of the mean difference / standard error difference to a standard look up table)."

> 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.

This is true for GOSAT as well once the co-location error is taken out. This was poorly presented in the previous version. In the previous version, which had 2 stations, the correlated raw error was 0.9 for dynamic and 0.7 for geometric coincidence criteria. The correlated errors were 0.7 for dynamic and 0.3 for coincidence. Subtracting these error in quadrature resulted in smaller correlated errors.

This analysis was updated to include more stations by backing off the inclusion criteria to at least n=10. The GOSAT geometric result is now fully integrated into the results of this section. The raw GOSAT dynamic errors of 0.9 ppm is significantly reduced by subtraction of the 0.7 ppm co-location error, whereas the 0.9 ppm geometric error (now larger than previously with different stations included) is not strongly changed by the subtraction of 0.4 ppm co-location error.

Added text, " GOSAT geometric versus dynamic averages show larger correlated error for 4/8 stations. The GOSAT dynamic correlated error of 0.9 ppm is significantly reduced by subtraction of the 0.7 ppm co-location error, whereas the 0.9 ppm geometric error is not strongly changed by the subtraction of 0.4 ppm co-location error. Subtracting (in quadrature) the co-location error (CT-CT column in Table 2) and TCCON error of

AMTD
8, C4258–C4274, 2015

0.35 ppm (Appendix B) results in the corrected correlated error, a, shown in the "Mean NH: subtr co-location error" row of Table 2. "

> 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?

Agree. Added a statement in (now) section 2.1, " In the future, either targeted observations or more specialized models could make better use of these TCCON stations. Although we do not use these stations in averaging, we show results from all TCCON stations in this paper."

Took out references to the GOSAT footprint.

> 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.

Done.

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

Took out this abbreviation.

> 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 bottomup. EDGAR, VULCAN, and HESTIA are examples of bottom-up inventories for CO2. They do not rely on transport. How do these findings (characterizing biases in models and satellite observations) apply to a bottom-up flux estimate?

AMTD
8, C4258–C4274, 2015

The reviewer is right this is not clear and does not make sense as worded. The wording is changed to, "This paper also shows a set of comparisons and tests that may be useful for evaluating bottom-up flux estimates or transport schemes in models."

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

Good idea, done.

> Page 6222, Lines 23–24: XCO2 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.

Changed to, “column averaged dry air mole fraction”

> Page 6222, Lines 26–28: “This information is transferred to the CO2 absorption band. . . ” Odd way to phrase it. Why not simply say it’s jointly estimated (or fitted). Rephrase.

Wording updated to, " BESD is a so-called full physics algorithm, which uses a two-band retrieval, with the O2-A absorption band used to retrieve scattering information of clouds and aerosols while the 1580nm band additionally contains CO2 information. Similar to the ACOS three-band retrieval for GOSAT, the explicit consideration of scattering by this approach reduces potential systematic biases due to clouds or aerosols."

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

Took out TCCON acronym definition in the introduction

> 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. . .

Good point. Took out this sentence.

> 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

AMTD
8, C4258–C4274, 2015

coincidence criteria." Why would there be more spatio-temporal error when averaging with dynamical?

There is more error because instead of coincidence of 5 degrees and 1 hour it is 30 longitude, 10 latitude, and 5 days plus some temperature criteria. There will be more mismatch (spatio-temporal error) but less random error.

Added text, " As expected, this value is much smaller for geometric than for our dynamic coincidence criteria (other dynamic coincidence criteria may do better; in our case the dynamic criteria considers points $\pm$30 degrees longitude, $\pm$10 degrees latitude, $\pm$5 days, and $\pm$2K temperature versus 1 hour, 5 degrees for geometric criteria)."

> 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).

This section is updated to take out the locally influenced station (per reviewer 2). When this was recalculated, the mean error is 0.9 ppm, but the co-location error is 0.7 and the TCCON standard deviation is 0.35 ppm. The previous version missed taking out the 0.35 ppm for the TCCON error. When these are taken out, a= 0.5. The geometric is 0.9 ppm, when the 0.4 ppm co-location and 0.35 TCCON standard deviation are taken out, a=0.8 ppm. So the GOSAT dynamic result has lower correlated error if we believe the larger co-location factor for dynamic coincidence.

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

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

> 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.

Agree. The text in 2.1 now describes which stations are considered complex or locally influenced; these stations are not used in averages. Izana is one of these stations.

AMTD

8, C4258–C4274, 2015

> Page 6237, Line 23: “coloumns” should say “columns”. Corrected

> 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. Text is updated.

> Page 6238, Line 9: “(Keppel-Aleks, 2012)” should say “(Keppel-Aleks et al., 2012)”. Updated

> Page 6241, Line 17: “2 _ 3_” should read “3_ _ 2_”. Done

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

Agree that is true for given the same number of matches, however more matches are achieved with dynamic averaging as seen in Fig. 5. Lamont has an unusually high number of geometric matches due to targeting.

> 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. Took this out.

> Page 6245, Lines 20–23: Seems excessive for the Conclusions. Just restate the findings. Agree. Left the first sentence to give context, "The seasonal cycle phase can detect seasonally dependent biases in satellite data and issues with model fluxes or transport errors. The GOSAT r.m.s. phase difference versus TCCON is . . ."

> 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.” Done âĂČ > Referee 2 > Most of the issues with this paper have already been flagged by Referee #1 and I will try not repeat them here. Suffice to say that I concur with his assessment. Particularly the lack of and, if at all present, the often erroneous representation of confidence intervals on the retrieved data is troublesome. In its present state I cannot deduct from this work

AMTD
8, C4258–C4274, 2015

if any of the attained biases are indeed statistically significant. This is crucial given that the data is often derived from small data samples.

> There is no mention of applying the TCCON column averaging kernels and a priori to the model and satellite data as per the TCCON guidelines(https://tcconwiki.caltech.edu/Network_Policy/Data_Use_Policy/Auxiliary_Data). One should at the very least outline why this was omitted.

This information now added to section 3.1 "Coincidence criteria and other matching details". [. . .] "When comparing models versus TCCON, the TCCON averaging kernel is applied to the model. When using the model to assess satellite coincidence error, the satellite averaging kernel is applied to the model at the satellite and TCCON coincidences. Note that the TCCON averaging kernel cannot be applied to satellite data and vice versa because a profile of higher resolution than the comparison observation is needed to apply an averaging kernel. In the satellite/TCCON comparison, both products have ∼1 degree of freedom. (The satellites do initially retrieve a profile with ∼1.6 degrees of freedom for GOSAT, but the profile is not what we are validating)."

> The Izana station is located at 2370m asl. Chapter 3.2 mentions Garmisch, Four Corners, Bremen and JPL as sites which (due to their location), potentially harbor a significant collocation error impacting the overall bias, yet Izana remains curiously unmentioned here even though its high altitude is bound to generate a bias with surrounding satellite soundings (not to mention that it mostly collocates measurements from the Saharan desert, while Izana itself is located on an island). Later when discussing the seasonality, Izana is suddenly flagged, while stations like Bremen are deemed ok (apparently the authors are confident that the local urban sources at Bremen show no seasonal pattern). I think it would make more sense to predefine a subset of 'clean' TCCON sites on which all parameters are drawn instead of tailoring the subset based on the particular results. Predefining such a set in the TCCON section would also greatly clarify the results section as this is, in its current state, often convoluted and confusing.

AMTD

8, C4258–C4274, 2015

Agree with reviewer. Moved text on TCCON "special stations" from Section 3.2 in the results to Section 2.1 where TCCON is now introduced. Describe the stations that have special circumstances which will result in local influence and clarify that these stations are not used in averages. Took these stations out of averages in the rest of the paper.

> Text moved and expanded on in Section 2.1 " Stations which have special circumstances regarding validation and are considered locally influenced are: Garmisch which is in the midst of complicated terrain [. . .] These stations are not included in averages (e.g. average bias, average seasonal cycle amplitude differences, etc.). In the future, either targeted observations or more specialized models could make better use of these TCCON stations. Although we do not use these stations in averaging, we show results from all TCCON stations in this paper."

> Minor comments: > Title: The paper consistently mentions that it is evaluating GOSAT and SCIAMACHY, while it in fact evaluates one particular GOSAT and SCIAMACHY algorithm (ACOS and BESD). Granted this is mentioned in the paper but I feel it should be part of the actual title. Now I cannot but feel that the reader is left under the impression that these are the definitive GOSAT and SCIAMACHY algorithms.

Title updated. Good point.

> p 6228,line 4: using +-5_ latitude and longitude for the geometric criteria, puts high latitude stations at a disadvantage. Better to use a distance limit to the TCCON site.

I agree, however it is more of a problem that there are no data in the far north for the DJF season, no matter what coincidence is used. I tried using a variable definition of longitude, pegged to 5 degrees at 40N, cutoff = cos(40)/cos(valid0[ii].latitude) * 5, however no new stations were added in the DJF season. Added text to Section 3.1 (coincidence criteria), "Because of the earth's curvature, high latitude sites could have relaxed coincidence in longitude, particularly for geometric coincidence. However stations north of 60N have gaps in the winter months in the satellite record such that relaxed criteria does not add additional stations to the analysis."

AMTD
8, C4258–C4274, 2015

> p 6228,line 5: Satellite data is paired with 90 minute TCCON averages. It is unclear if this is done prior to the collocation routine or after. It would also be useful to report the TCCON variability within this time frame.

The data previously was analyzed by averaging TCCON and then matching. The analysis was updated to match to TCCON and then average. Added text to Section 3.1, "Coincidence criteria and other matching details" "Satellite measurements, which satisfy the so-called geometric criteria, are within $\pm$1 hour, $\pm$5 degrees latitude and longitude of an unaveraged TCCON observation. Following the match-ups, all TCCON observations matching one satellite observation which are within 90 minutes are averaged, reducing the TCCON random error." [. . .] " Matches are found with un-averaged TCCON data; TCCON observations matching a single satellite observation are averaged within 90-minute intervals."

> p6233,line 12: "The purple dashed line represents spatio-temporal mismatch error and as expected, this value is much smaller for the geometric than dynamic coincidence criteria". This mismatch error is derived from the CT values at the sat retrieval and the TCCON site. Given the inevitable smoothing associated with model output, this value should be describe as an estimate of the spatio-temporal mismatch error or even "a lower-bound estimate of the spatio-temporal mismatch error". Secondly, the fact that it is straightforwardly assessed that the geometric criteria performs better begs the question if other more stringent dynamic criteria would yield other results. Thus replace "than dynamic" with "than our dynamic" coincidence criteria.

Good points. Paper content updated. In section 3.4, "The purple dashed line represents the standard deviation of CT2013b at the satellite time and location vs. CT2013b at the TCCON time and location and represents a lower bound of the spatio-temporal mismatch error (co-location error). As expected, this value is much smaller for geometric than for our dynamic coincidence criteria (other dynamic coincidence criteria may do better)."

AMTD
8, C4258–C4274, 2015

> p6234: section 3.5: This essentially deals with the same issues as section 4.1

They are similar but 3.5 uses a much simpler approach. Section 3.5 calculates seasonally dependent biases by averaging over 3 months and subtracting yearly biases. All of section 4 deals with results using the NOAA CCGCRV software, which is a more opaque method. In response to reviewer 1 and 2 both, more clarity was added as to these two methods. In section 3.5, the following text was added, "This is a simple averaging method which will later be compared to seasonal cycle amplitude fit results." with all comparisons between section 4.1 and the conclusion.

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

Thank you to the reviewers for their comments on this paper. Reviewer comments are indented and in black font; responses are in blue font. Note also that during the review process I re-ran the analysis with the latest GGG2014 TCCON data. The previous version went through 10/2013 which cut out some GOSAT data (ACOS-GOSAT goes through 5/2014), CT (which goes through 12/2013, and MACC (which goes through 12/2014). In response to a comment by reviewer 2, TCCON is now averaged in 90 minute intervals after rather than before the match-ups to satellites.

> 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.

The reviewer elaborates details in the subsequent comments, so we respond to each below.

> 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)?

Updated caption and text to address all the above. The caption now reads

Figure 2. Time series for matches of CT2013b, MACC, SCIAMACHY, and GOSAT versus TCCON at Lamont (panels a-d) and Lauder125 (for models) or Wollongong (for satellites) (panels e-h). The top plot of each set (panels a1, b1, c1, …) shows a time series of all geometric matching pairs. The middle panel of each set (panels a2, b2, c2, …) shows the difference versus TCCON, with the blue line the 30-day average difference. The bottom panel of each set (panels a3, b3, c3,…) shows a histogram of the differences, indicating an approximate error and bias.

I think that the text (panels a2, b2, c2, …) is extraneous to the description, "the middle panel of each set." however the reviewer indicates this is needed to describe the figure.

The text has added text to address the TCCON differences seen in the plots and two different SH sites in section 3.

"Lauder is shown for the models to show the phase lag for CT2013b seen at this site in Table 7. Since there are not enough coincidences for the satellites at Lauder, we show Wollongong for satellites."

"As only TCCON/satellite matched pairs are shown, different subsets of TCCON are included for models and the two satellites."

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

Fig. 1. color coded response to reviewers

AMTD
8, C4258–C4274, 2015