

Interactive comment on “Intercomparison of NO₂, O₄, O₃ and HCHO slant column measurements by MAX-DOAS and zenith-sky UV-Visible spectrometers during the CINDI-2 campaign” by Karin Kreher et al.

Anonymous Referee #1

Received and published: 19 August 2019

Review of Kreher et al. - Intercomparison of NO₂, O₄, O₃ and HCHO slant column measurements by MAX-DOAS and zenith-sky UV-Visible spectrometers during the CINDI-2 campaign

GENERAL COMMENTS

As clearly stated in the title, this manuscript presents results from the 2016 "CINDI-2" intercomparison campaign relating to retrievals of key trace species (NO₂, O₄, O₃ and HCHO) using either MAX-DOAS or zenith sky UV/Visible spectrometers. These types

C1

of measurements have grown to considerable importance in the field of atmospheric composition in recent years, and are expected to continue to increase rapidly in number and range of applications, making a very careful campaign such as CINDI-2 of great interest to a broad community. Importantly, the major types of instruments now in widespread use (such as Pandora, SAOZ, the former EnviMeS MAX-DOAS and the Hoffman mini-DOAS) all participated in the campaign which ensures the relevance of the CINDI-2 results to the actual measurements being made around the world.

The manuscript is comprehensive and clearly written, and many of the author team are among the world experts in this field, and overall, I believe is very suitable for publication in AMT.

I do have a number of general comments and questions. I believe it will help the reader better understand the philosophy and approach of CINDI-2 if each of these could be briefly addressed in either the introduction or the discussion section of the manuscript.

1. It is evident that while great attention was paid to ensure the consistency of certain aspects of the measurements and retrievals, other aspects – which would also affect the results - were left to the individual groups. I am sure the decisions of the organisers in this regard were made with thought but it is not always clear to the reader what the motivation was for the different inclusions and exclusions and how these related to the stated aims.

2. To what extent, can the results of the intercomparison obtained in idealised and tightly co-ordinated conditions be applied to the operational, geographically-distributed real-world measurement sites? Recommendations for the networks seem minimal (elevation scans are mentioned).

3. Limited of course by my own experience, it seems quite unusual for an intercomparison to be carried out without a designated reference instrument or standard, and instead to use the median of the participants as a reference. (Although in the case of formaldehyde a subgroup of better-performing instruments is identified and so this is

C2

closer to an orthodox reference group). As far as I can see, this means there can be no traceability of any of the measurements? I would also add that in places I found the text has the potential to be misleading by referring to "the reference" in the abstract and conclusions, which readers might read in isolation to the rest of the paper.

4. In many places the manuscript notes the efforts made to eliminate spatial and temporal mismatches between the participating instruments, but this does not seem linked to the scales of temporal and spatial variability expected for these species, and indeed, in section 3.7 it seems NO₂ varies on a finer scale.

5. From time to time the stated aims seem to interfere with each other. To really understand the differences between instruments requires a somewhat different approach compared to undertaking a strict performance evaluation, particularly if the aim is to simulate realistic conditions in the field. This point is closely related to (1) about the overall design of the exercise and what is or isn't being evaluated.

SPECIFIC COMMENTS

Page 2

Lines 9-12 The "major aims" don't quite agree with what appears later (Section 2.3, page 5 lines 31-32).

Lines 12-14 I don't see how you can do "trend analysis" without traceability to a standard?

Line 20 The word "unprecedented" seems over hyped

Line 25 "bias and offset of the individual data sets against the reference". I think this is likely to mislead the reader of the abstract because it implies the existence of a reference instrument.

Lines 23-26 This seems like the "reproducibility" in usual metrological terms.

Line 28 "... a quantitative assessment of the measurement performance" – it seems to

C3

me more like the "consistency" ?

Page 3

Line 38 "The interest of ESA for ..." change to either "The interest of ESA in " or "The desire of ESA for" or similar.

Lines 40-41 "planned at the horizon 2022-2023" – I don't know what this phrase means sorry.

Page 4

Line 7 Touching again on the philosophy of CINDI-2, it seems to me just the consistency, there are other aspects of "high quality" needed for "long-term measurements, trend analysis and satellite data validation".

Line 8 "... it is essential ... to contribute to a harmonisation" – it can't be "essential" to "contribute"! These seem to be aims (1) and (3) from the abstract re-worded.

Line 9-10 Did you in fact contribute to a harmonisation of the measurement settings and retrieval methods outside of the intercomparison itself, ie for the networks to use in practice?

Page 5

Lines 5-10 This is very interesting in terms of the philosophy of CINDI-2. It is stated some groups performed more advanced pre-processing, but in general, as far as I can tell, the results from these groups was not weighted any differently from groups that didn't do these steps. Is that logical?

Lines 9-10 Rather than standardise these steps, wouldn't it be more valuable to assess their contribution to better results?

Lines 9-10 Could this be something to recommend to field instruments?

Line 14 "containers". For the first time this word appears, I suggest "shipping contain-

C4

ers", and also the first time it appears in the captions (Figure 1). After the first time, just "container" would be ok. A "container" out of context could be of any size.

Line 14 "temporary containers were rented" – I would prefer "shipping containers were rented and temporarily installed".

Line 24 Strictly, 287 degrees isn't WNW, which is 292.5 degrees from north.

Line 24 Rather than "N=0", it would be clearer to say "north"

Line 29 Change "Like in" to "As in "

Lines 30-32 The objectives don't quite match the three listed earlier (such as in the abstract). Now there are only two.

Lines 31-32 The second objective was previously to "discuss the performance" now it is to "define a robust methodology for performance assessment". Is it to define a methodology or to apply it?

Lines 36-39 It is interesting that the retrieval settings and parameters were specified but not the software. I am struggling to understand the logic of this. I think this decision is worth more explanation. It would be possible to compare a purely raw instrumental quantity, wouldn't it?

Page 6

Lines 1-7 This is another curious feature of the design of the campaign. To me there seems a conflict between the daily meetings which help understand better what is going on, and the strictness of the campaign designed to assess performance. In the field this luxury would certainly not be available.

Line 14 "operation" should be "operational"

Line 27 The sentence "The convention for the azimuth angle . . ." appears in the wrong place

C5

Line 28 "synchronicity" should be "synchronisation" (unless we are talking about Jung or pop music from the early 1980s)

Line 32 I would have thought "an NDACC" rather than "a NDACC" (but this is because I am expecting the reader to read "NDACC" as "en dack".)

Page 7

Line 12 "unprecedented" seems over-hyped to me – don't you really just mean that it was "improved" or "greatly improved" since CINDI-1?

Line 13 "synchronicity" -> "synchronisation"

Line 14 "... the impact of atmospheric noise on the data comparisons could be reduced to a minimum" - How do you know though that the level of co-ordination is enough though? Do you know what time scales and spatial scales you expect the species to vary over? Later on, you imply that actually the co-ordination was not sufficient for N20.

Line 29 Could you have mandated separate times for UV and visible?

Line 34 I don't think "MPIC" has previously been defined.

Page 7 line 30 – Page 8 line 4

Presumably however none of this, except (3), would be available in a field setting? This to me seems a conflict between the different aims of CINDI-2.

Page 8

Line 12 "we used" – until now the manuscript has been written using the traditional third person passive voice.

Lines 25-38 There doesn't seem to be any mention of the type of location Cabauw is in terms of rural versus urban and the expected pollution levels.

Page 9

C6

Lines 18-19 Some of the instruments show a drift over the course of the campaign. Should we therefore expect instruments in the field also to show potentially significant drifts over time?

Lines 35-38 The decision to allow resubmissions is also interesting – I assume the justification is that these types of mistakes would be able to be identified and corrected independently by the instrument operator in a network setting?

Page 10

Lines 6-14 This seems to create a problem though, because in the field, this would not generally be possible?

Page 11

Line 25 -"drastically reduced" – that would depend on the temporal and spatial variability though?

Line 26 "and/or atmospheric variability" – I don't understand what you mean here. The sentence seems to contradict itself to me. The sampling and mis-match errors are only small or large relative to the spatial and temporal scale of atmospheric variability. If the comparison noise is caused by atmospheric variability then isn't that a mismatch?

Line 33 "similar as performed" -> "similar to as performed" or "similar to those performed"

Page 12

Lines 22-30 The implication is that the fit residuals should represent a lower bound to the measurement uncertainty, but perhaps another sentence of justification is needed for this.

Line 30 – If the real NO₂ is varying on short scales that in itself is not an error of the measurement, but it would affect the agreement with a given satellite pixel.

C7

Line 39 "keeps" should be "stays"

Page 13

Lines 1-2 ". . . for this molecule most of the residual variance between good instruments can be explained by measurement noise" needs re-wording. I think I know what you mean but the words by themselves don't make much sense.

Line 9 Replace "a couple" by "two".

Line 19 Replace "largest" with "the largest"

Lines 18-21 This must be very relevant for field instruments?

Line 30 ". . . specific limits have been set. . ." You should add something like ". . . specific limits have been set to use for performance evaluation". The way it is now, it takes the reader some time to work out what these limits are all about.

Lines 28-37

Intuitively, I don't find this approach very reasonable. It seems you choose limits somewhat arbitrarily (or at least let's say making use of subjective judgement), and then go through a binary pass or fail evaluation. Especially in figure 19, some of the dots which pass seem to be right on the limit, and some of the failed points fall only just outside it. I appreciate for network use such as NDACC there might need to be a definite threshold, but otherwise the use of pass/fail seems to degrade the information you have gained through the experiment. Perhaps you could discuss this point briefly.

Page 14

Line 1 "statistic" should be "statistics" if I've understood the sentence correctly.

Lines 11-17 Just repeating an earlier comment, the use of green versus orange when the two instruments could be a distance of epsilon other side of an arbitrary line seems odd to me. The use of pink for being four times outside the limit makes more sense.

C8

Lines 37-38 I thought the DOAS settings were all prescribed?

Page 15

Line 8 "wavelengths" should be "wavelength"

Line 9 A better wording might be "and only failed to satisfy one criterion in the O4 .."

Lines 8-10 I suggest breaking this sentence into two parts for easier comprehension.

Lines 23-24 I suggest replacing "at the same time they are meeting" with "at the same time meet"

Line 24 Replace "On the opposite" with a phrase such as "On the other hand" or "Conversely".

Line 25 "satisfies" should be "satisfy"

Page 16

Line 1 "... a reduction in of the atmospheric changes on the intercomparison exercise." A reduction compared to what? (CINDI-1 I assume).

Line 4 "very well coordinated" sounds like boasting to me!

Line 14 "... with a selected reference" seems misleading to me, because it implies a specified reference instrument, which was not part of the intercomparison.

Lines 23-25 "The median bias against the reference is generally low ...". Again I think this might mislead the reader who hasn't read the whole paper, who would assume there was a particular reference instrument.

Line 30 Replace "&" (ampersand symbol) with the word "and".

Line 33 Personally, I don't think you can say "guideline" in the singular like this, but others might disagree.

Line 34 Replace "like the one" with "such as the one".

C9

Page 17

Line 4 "instruments" should have an apostrophe - "instruments' " or re-word to "the elevation point calibration of instruments".

Lines 7-8 "a thoroughly planned and carefully managed campaign" sounds like boasting to me.

Lines 18-26 This sounds really good and would be very valuable to the community.

Page 32 (Figure 3)

The individual plots are very small but adequate for qualitative use of the figure.

Page 40 (Figure 11)

Delete the unwanted carriage return in the caption.

Page 46 (Figure 17)

"The dashed lines indicate the limits ... " For the caption, you need to provide more information, in particular that these limits have been chosen (rather than derived), for the sake of distinguishing outliers.

Page 49 (Figure 20)

In my printed version of the manuscript I find the pink and orange a little bit hard to distinguish. (The green and orange have excellent contrast.)

Page 51 (Figure 22)

The numbers in the green boxes are quite hard to read, and also to some extent those in the red and orange boxes.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2019-157, 2019.