**Response to Anonymous Referee #1**

We would like to thank the reviewer for their comprehensive and thoughtful review, and helpful comments which are addressed individually in the response below. The reviewer's comments are included in blue & italics.

*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 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.*

This comment touches on a very important topic and most of the choices were motivated by findings of the first CINDI campaign and MADCAT. There definitely are reasons why some aspects of the intercomparison exercise were prescribed (such as the measurement schedule and the retrieval settings) while others were not (the analysis code and some of the calibration procedures). The organisers of the CINDI-2 intercomparison were aiming at providing a procedure that (1) forced every participating instrument to look simultaneously in the same direction (and to do this as precisely as practically achievable) and hence sample the same airmass and (2) to prescribe the use of analysis settings that were as consistent as realistically possible to enforce a more coordinated analysis.

One step further would have been to also prescribe the analysis software but allowing the individual groups to stick with their own preferred analysis software (which most participants would continue to use after CINDI-2 anyway) led to a more realistic intercomparison, and hence to a more realistic assessment of the participating instrument/group by using the combination of individual instrument plus individually used analysis software but prescribing all other settings and procedures.

Main reasons for not enforcing strict guidelines for the calibration steps were that (1) some of the key calibration steps (wavelength registration and slit function determination) can be obtained in the field using solar lines and dedicated software, (2) calibration facilities were not available to analyse other key instrumental responses, such as stray-light level, detector linearity response or polarization response, and (3) some neglected calibration steps are of minor importance for DOAS-type retrievals (e.g. radiometric response). However, the possibility to address better the missing aspects, and in

particular calibration related issues, will be considered when preparing future campaigns. A short paragraph has been added at the end of Section 2.2 (Campaign design) to motivate better why a lot of effort was spent on certain aspects.

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

This is also a very important comment and helpful feed-back for us. The NDACC UV/Vis Working Group provides recommendations for measurements and data analysis which are mandatory for the inclusion of an instrument (and station) into the NDACC network. These recommendations (referred to as NDACC UV/Vis Appendix) have been substantially updated to also include guidelines for MAX-DOAS measurements and data analysis, and they will be published on the NDACC web page by the end of December 2019. A short statement has been added to address this, which is quoted under item 5 further below.
We have also added a separate section entitled 'Recommendations for network operation and future campaigns' before the conclusions. A part of the conclusions has been moved into this section and some addition text addressing this comment has also been added.

*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 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.*

We have clarified the use of reference data sets in the text. Reasons for why we used the median of the participating instruments rather than one single instrument (or a small group of instruments) is to keep the comparison fairer and not to 'favour' a couple of instruments. If an instrument with an absolute calibration would have been available then using that instrument would certainly have made sense, but there is no absolute reference for such measurements. The approach adopted here is similar to what was used in previous UV-Vis intercomparisons, i.e. identifying a group of mutually consistent instruments and use the median from their measurements as a best estimate ('most probable') of the 'true' value. For $NO_2$, it appeared that a large number of instruments were found to be in mutual agreement within limits derived from previous campaigns, for HCHO only a small sub-group presented a satisfactory level of agreement.

*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.*

Efforts were made to substantially improve the spatial and temporal coincidence between measurements, in comparison to what was done in previous campaigns. However, practical limitations (related to the large variety of participating instruments) also had to be considered. It was found – a posteriori – that the scale of variability of $NO_2$ was in fact small enough to still dominate the variance of the measurements (despite the fact that these measurements were synchronized to better than one minute in time and all telescope pointing in the same azimuthal direction within a few degrees of accuracy, and in the same elevation to better than 1 degree). And this has also been stated in Section 3.7. E.g. the following sentence has been added: 'This means that in this intercomparison, atmospheric

variability limits the reproducibility and representativeness of individual MAX-DOAS measurements for species such as $NO_2$.'

*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.*

To address and clarify this point, we have added a statement in Section 1, paragraph 4, that the aim of the intercomparison is '… to assess the participating instruments in their ability to retrieve the same geophysical quantities (i.e. slant columns of $NO_2$, $O_4$, HCHO and $O_3$) when measured and processed in a controlled way (i.e. using a prescribed measurement protocol and retrieval settings)'.

We have also added/changed the following statement in the conclusion so that is now reads: ' This assessment process, undertaken as part of the CINDI-2 intercomparison campaign, provides the UV-visible absorption spectroscopy research community with guidelines and a procedure on how to assess the performance of MAX-DOAS and DOAS instruments, in particular for the inclusion into NDACC (see NDACC webpage for access to the UV/Vis Appendix describing these recommendations). It is expected that a similar level of consistency, as seen during CINDI-2, can be obtained in the field if recommended settings are implemented and used by each participant of the network. More control in this aspect of homogeneity can be obtained through centralized processing, which is the aim of the currently developed ESA FRM4DOAS project (see http://frm4doas.aeronomie.be/).'

*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).*

We agree with the reviewer (thanks very much for picking this up) and have changed the text in the abstract and in Section 2.3 to be consistent.

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

For trend analysis, the measurement precision and its stability in time (i.e. making sure that measurements are not affected by drifts or discontinuities of any type) should be most important. This means that suitability for trend analysis cannot be determined from a campaign in isolation, since an instrument showing a perfect behavior during two weeks can always be affected by longer term drifts or biases once in operation. However, successful participation to successive campaigns is one way to verify stability. This is e.g. the approach used in the Dobson/Brewer network communities. Another possible approach is to regularly operate traveling standard instruments at the different sites of a network.

*Line 20 The word "unprecedented" seems over hyped*

We have changed this to 'unique'.

*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.*

We agree and have changed the text to '…. the selected refence (which is the median of either all data or a subset), …'

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

As far as I understand, this is correct. However, we used here the same mathematical terms previously used in UV/Vis instrument intercomparisons to be consistent with the analysis performed e.g. during the first CINDI or earlier intercomparisons.

*Line 28 ". . . a quantitative assessment of the measurement performance" – it seems to me more like the "consistency" ?*

We have changed the text to: 'It introduces a quantitative assessment of the consistency between all the participating instruments for the MAX-DOAS and zenith-sky DOAS techniques.' If an instrument was not performing well, this could be clearly identified.

*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.*

Done.

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

This phrase has been deleted.

*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".*

See response to the corresponding comment above.

*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.*

We agree that this wasn't worded well and the text has been changed to accommodate the comment. The part of the sentence "… and to contribute to a harmonisation of the measurement settings and retrieval methods." has been deleted.

*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?*

Yes, we did and this has been incorporated in the updated NDACC UV/Vis Appendix (Protocol for NDACC UV/Vis instrument operation and data analysis) which will be published on the NDACC web page later this month (Dec 2019). This has been added under Conclusions (paragraph 5).

*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?*

Yes, it actually is. Since many of the instruments can differ in the detail of their particular setup, it would have been difficult to fairly assess the instruments performance on grounds of pre-processing without really looking thoroughly at each of the instruments and its pre-calibration features. However, it would certainly be valuable if future campaigns would look into the pre-processing and calibration of the instruments in a more coordinated way (e.g. through organization of a calibration campaign ahead of the field campaign) and this has now also been added in a new section dealing specifically with recommendations based on the CINDI-2 results and experiences.

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

That is a good point and we have addressed this by adding additional text under the new Section 5 (Recommendations for network operation and future campaigns), last paragraph (bullet #2).

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

Yes, it certainly can and the NDACC UV/Vis Appendix also contains information on further documentation containing guidelines for calibrations which will shortly also be available on the NDACC UV/Vis working group web site as well and is currently available here: http://frm4doas.aeronomie.be/ProjectDir/Deliverables/FRM4DOAS_D4_MAXDOAS_Best_Practices_Document_20180110_v1_0.pdf

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

We have changed this to: '… mobile units (similar to shipping containers) were temporarily installed for the campaign period.' The containers are not strictly speaking shipping containers but 'mobile units' which look similar to shipping containers.

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

We have changed the text accordingly (see above).

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

True, that is strictly speaking correct and we have added 'approximately'. We were working of a table that stated that WNW is associated with angles between 281.25° – 303.75°.

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

Agreed and this has been changed accordingly.

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

*Done.*

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

This was already previously raised (first comment of the 'specific comments section') and has been changed in the text so it is consistent in Section 2.3 and the abstract.

*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?*

The objective is to define a methodology which is then also applied to the CINDI-2 data products. The text has been changed accordingly.

*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?*

We understand were the reviewer is coming from but since all analysis software packages basically solve the same mathematical equations (which are part of the DOAS technique), the differences lie in the details of the implementation (in particular wavelength registration issues) rather than in the

actual analysis software. Hence the approach to harmonize and prescribe the settings as much as possible but allow for individual software packages to be used.

*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.*

It is certainly correct that in the field, it is often not possible to get this kind of feed-back and the semi-blind intercomparison procedure is in this regard a compromise between (1) a strict 'blind' intercomparison which would not allow for any exchange of information between the participants and (2) the opportunity especially (but not only!) for more inexperienced participants to gain a lot of experience and knowledge, and if possible, to have an independent referee intervene if there is an obvious problem with instrumentation that can be fixed (e.g. a problem with the hardware, such as the elevation pointing). The information provided at the daily meetings also encouraged the participants to be more engaged in the intercomparison overall without giving away how well their individual measurements were doing.

*Line 14 "operation" should be "operational"*

Done.

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

We agree and this has been fixed; the explanation is now been provided earlier on under Section 2.2.

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

Fair enough and done.

*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".)*

Agreed & done.

*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?*

We appreciate the comment and have reworded the sentence accordingly.

*Line 13 "synchroncity" -> "synchronisation"*

Done.

*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.*

Good point and we have toned the statement down accordingly.

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

Possibly, but we were not aware of that issue when the measurement schedule was designed, and this would also have meant that it would have affected everybody's schedule not just the Pandora instruments.

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

The full name has been added in brackets.

*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.*

Both, (2) and (3) should be straight forward to implement in a field application. All this is discussed in much more detail in Donner et al., 2019 which has been submitted and is currently under review (the reference has been updated accordingly). It is certainly true that option (1) requires the availability of a strong lamp but for a campaign such as CINDI-2, this was definitely a very valuable additional test and helped each of the groups to find out more about the accuracy of the elevation pointing of their instrument.

*Page 8*
*Line 12 "we used" – until now the manuscript has been written using the traditional third person passive voice.*

Agreed & this has been changed to the passive form.

*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.*

Good point. We have added a short description of the Cabauw measurement site under Section 2:

"In short, the CESAR site at Cabauw is overall a rural site, with only a few pollution sources nearby, but the wider vicinity of Cabauw is densely populated, with the cities of Utrecht, Amsterdam, The Hague and Rotterdam less than 60 km away and a dense highway grid within 25 km, so that the site experiences recurring pollution events, e.g. such as from the daily morning and afternoon rush hours."

*Page 9*
*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?*

Possibly, and CINDI-2 really helped us to appreciate how important it is for the measurement quality to verify the accuracy and stability of the elevation scans. This also means that in the field, it is important to regularly monitor the accuracy of the elevation scans to avoid any drift, bias or discontinuity in data series, hence we made a recommendation to this end (2nd paragraph on conclusions).

*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?*

Yes, that is correct. For a resubmission, the groups had to state clearly what mistakes they made and how they were remedied. Admittedly, in a real word situation (e.g. due to time constraints) we might not always look carefully enough at our data sets but if we would, we should be able to identify and correct the issues which were identified.

*Page 10*
*Lines 6-14 This seems to create a problem though, because in the field, this would not generally be possible?*

We don't quite understand why this is a problem. We don't mean to imply that everything we applied during the intercomparison has to be 100% reproducible in a field situation and we think it is ok that

we create somewhat more idealized conditions which show us how well we can agree if we pay attention and get everything is right as possible.

*Page 11*
*Line 25 -"drastically reduced" – that would depend on the temporal and spatial variability though?*

Good point. We have toned the text down somewhat and changed ‘drastically’ into ‘considerably’ and changed ‘should accurately reflect’ to ‘should more accurately reflect’

*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?*

We agree with the reviewer and have deleted ‘and/or atmospheric variability’.

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

Done.

*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.*

A statement to this effect has been added.

*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.*

We agree and we expect that the scale of variability of $NO_2$ is much smaller than the scale of any $NO_2$ satellite measurement. The main issue is therefore to assess the representativeness of correlative measurements for comparison to satellite data.

*Line 39 "keeps" should be "stays"*

Done. We have changed it to ‘remains’.

*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.*

We have reworded the sentence to ‘… for this molecule most of the residual variance from regressions involving good instruments can be explained by instrument shot noise.’

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

Done.

*Line 19 Replace "largest" with "the largest"*

Done.

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

Yes, we agree. This is already covered under Section 3.4 so we didn’t want to repeat it here again.

*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.*

Done.

*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.*

We agree that this is not straight forward and a bit of a delicate issue as well. Even though we tried to introduce some elements of objectivity, the choice of a limit is fundamentally arbitrary (but not totally subjective since it is based on statistical arguments). Since the limits were chosen to exceed the median of the measurements (this has now also been added to Figure 18 and in the text), instruments that exceed them can be seen as "out of the norm". This does not necessarily mean that such measurements are problematic and this is why we are checking several parameters. Failing in one parameter, especially if very close to the limit is not a problem per se but failing in two or more usually is.

*Page 14*
*Line 1 "statistic" should be "statistics" if I've understood the sentence correctly.*

Done.

*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.*

See discussion above.

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

This sentence has been deleted, AIOFM have made a mistake and chose the wrong ozone cross-section. They have re-analysed their data with the correct ozone cross-section and we have updated the figures correspondingly. This did affect Figures 13, 18, 20 and 22, and Fig-S22, Fig-S23 and Fig-S24 from the Supplement, and some text in Appendix B.

*Page 15*
*Line 8 "wavelengths" should be "wavelength"*

Done.

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

Done.

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

Done.

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

Done.

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

Done.

*Line 25 "satisfies" should be "satisfy"*

Done.

*Page 16*
*Line 1 ". . . a reduction in of the atmospheric changes on the intercomparison exercise." A reduction compared to what? (CINDI-1 I assume).*

We agree that this needs to be fixed and have added to the sentence so it reads: '… atmospheric changes on the intercomparison exercise in comparison to CINDI.'

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

Fair enough and we have dropped this phrase.

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

We agree and have changed the text accordingly: '… with a reference data set was performed (see Section 3.5 for details on how the reference data sets were derived) …'

*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.*

To clarify this, we have replaced '… the reference…' with '… the reference data sets …'

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

Done.

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

Agreed and changed.

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

The sentence has been changed and the phrase has been dropped.

*Page 17*
*Line 4 "instruments" should have an apostrophe - "instruments' " or re-word to "the elevation point calibration of instruments".*

Done.

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

Fair enough and this has been toned down in the text and this sentence has been moved to the new Section 5 (Recommendations …).

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

The feedback is much appreciated and this will clearly be considered in the design of the next UV/Vis intercomparison. This has now also been moved into Section 5.

*Page 32 (Figure 3)*
*The individual plots are very small but adequate for qualitative use of the figure.*

That was the intention. If ok with the reviewer, we would prefer to leave the plot as is.

*Page 40 (Figure 11)*
*Delete the unwanted carriage return in the caption.*

Done.

*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.*

Done.

*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. )*

Good point, we have fixed this by replacing pink with black.

*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.*

We agree and have fixed this figure so that the numbers are much clearer to see.