Anonymous Referee #1:

We greatly appreciate the positive feedback from the referee and the constructive comments. As described below, we have modified the manuscript according to suggestions and clarified where necessary. We hope that the revised manuscript has improved in respect to the original paper. Please find a rebuttal against each point below.

Black, bold, italic: Referee's comments

Black: Author's reply

Changes in the original discussion paper are highlighted in yellow and attached below

1) There is concern about the assumption of vertical NO2 profile in the APEX retrieval as well mixed profile of NO2 through the boundary layer. There have been many observations and analysis in the literature proving that NO2 is rarely 'well-mixed' in an urban environment (e.g., http://dx.doi.org/10.1002/2015JD024203). (1) There appears to be modeled high resolution model data available from the regional CAMS model that likely at last has some more realistic weighting of NO2 nearer to the surface (negative vertical gradient in the boundary layer). The analysis would be strengthened if results were also shown with those a priori in the APEX retrieval.

It is true that in most studies assumptions are made on the profile shape in the boundary layer as high resolution model profiles are not always available (and also can have significant errors). Most campaigns involving airborne spectrometers are also lacking measurements of the vertical gas distribution as it requires an additional set of in-situ instruments and specific flight patterns. Note that we are involved in a project to address/study this problem by combining a spectrometer and in-situ instruments in one aircraft (RAMOS - http://environment.inoe.ro/article/179/about-ramos). The aircraft will also execute flights over Bucharest, Romania in 2020-2021 in the context of TROPOMI validation. This data set will allow us to better assess the impact of measured, modelled or assumed well-mixed profiles.

As indicated in Sect. 4.3.1, the decision was taken to use box profiles for the reference APEX retrievals in order to be independent from both the standard TROPOMI product based on TM5-MP profiles, and the TROPOMI product based on CAMS profiles. In the paper a sensitivity study was already included where the box profiles were replaced by interpolated TM5-MP profiles. We have followed your suggestion and also assessed the impact of replacing the box profiles by CAMS profiles. The findings are in line with a previous study (Tack et al. 2017) where we assessed the impact on the APEX retrievals of using high resolution a priori NO_2 profiles from the 1 km x 1 km AURORA model instead of box profiles.

In Sect. 4.3.1, we have changed the paragraph accordingly:
"For the APEX retrievals, AEPs and a priori NO_2 profiles were constructed from the AOT and PBL height observations, as discussed in Sect. 4.2. In order to yield retrievals independent from the satellite, box profiles were used instead of the TROPOMI TM5-MP profiles, as displayed in Fig. 3a. When TM5-MP or CAMS profiles would be applied as a priori for the APEX retrievals, the AMF would increase with respectively 9% and 10% on average, which is largely consistent with a similar sensitivity study reported in Tack et al. (2017). For the APEX retrievals, we assumed a well-mixed NO_2

and aerosol box profile scenario and urban aerosols with a high single-scattering albedo (SSA) of 0.93. This causes a multiple scattering scenario and an enhancement of the optical path length in the NO_2 layer, and results in an increase in the AMF. When instead considering a no aerosol scenario for the APEX retrievals, the AMF drops by 10% on average. We assume that the opposing effects of using (1) a priori profile shape assumptions different from the TROPOMI retrievals and (2) different aerosol assumptions tend to cancel each other out in the APEX retrievals."

2) Alternatively or in addition, the analysis would also be strengthened if there was some background on the validation of APEX NO2 observations or perhaps independent validation with measurements from the MAX-DOAS measurements mentioned in this analysis. It is hard to evaluate TROPOMI bias if the reference measurement is not validated itself.

Validation implies that the reference data has a better accuracy than the data set to be validated. This is indeed the case for the MAX-DOAS data when compared to the airborne APEX data. However, there is the issue of differences in horizontal representativity and potential sampling of different air masses.

For the overpasses over the MAX-DOAS station on 26 and 28 June we have compared the MAX-DOAS and APEX retrievals. We have only two overpasses in this data set, but we hope to include more (MAX-)DOAS instruments during the follow-up campaign in summer 2021.

Note as well that APEX NO_2 VCD retrievals have been assessed and validated by comparison with other airborne imagers, as well as GB DOAS measurements during the AROMAPEX intercomparison campaign reported in https://doi.org/10.5194/amt-12-211-2019. This is mentioned in the introduction of the study under review.

We have added a discussion on the comparison with MAX-DOAS at the end of Sect. 5.2.2:
"For the flights over the Brussels region, we have also compared the TROPOMI and APEX NO_2 VCD with the MAX-DOAS NO_2 VCD at the time of overpass and results are provided in Table 5. The TROPOMI NO_2 VCD is provided for the pixel in which the station resides for both the TM5-MP-based and CAMS-based product. The APEX NO_2 VCD is provided for the average within the TROPOMI pixel footprint over the MAX-DOAS station and for the specific APEX pixel over the station. As the MAX-DOAS is performing elevation scans in a fixed azimuth direction (35° N), APEX observations are also averaged along this line of sight (LOS) in order to take into account the instrument directivity and in order to reduce potential mismatches due to differences in spatial representativity. In this case, however, temporal mismatches can occur as APEX pixels, acquired in different flight lines, are averaged. Based on the study of Dimitropoulou et al. (2020), the horizontal sensitivity of the MAX-DOAS is estimated to be in the order of 10 km for measurements in Brussels in summer time and in the visible wavelength range. MAX-DOAS observations are filtered based on the degrees of freedom (DOFs) which should be larger than two. Secondly, the relative root mean square error (RMSE) of the difference between measured and calculated differential slant column densities with respect to the zenith spectrum of each scan should be smaller than 15 % (Dimitropoulou et al., 2020). On 26 June there is clearly a pollution event not seen over the station but further northeast along the MAX-DOAS LOS, as can be observed in the APEX NO_2 VCD grid (see Fig. 7a and Fig. 11). When averaging the APEX pixels along the MAX-DOAS LOS, the difference in MAX-DOAS and APEX NO_2 VCD is reduced from 4.8 to 0.1 x 10^{15} molec cm^{-2}. On June 28, the diurnal variation in the NO_2 field is much smaller. We see a slight underestimation of 0.3 x 10^{15} molec cm^{-2} for the APEX observation above the station

when compared to MAX-DOAS, while the latter is overestimated by 1.2 x 10^{15} molec cm^{-2} when averaging along the LOS."

Table 5. Co-located TROPOMI, APEX and MAX-DOAS observations for the flights over Brussels. The TROPOMI NO_2 VCD is provided for the pixel in which the MAX-DOAS station resides for both the TM5-MP-based and CAMS-based product. The APEX NO_2 VCD is provided for the average within the TROPOMI pixel footprint over the MAX-DOAS station and for the specific APEX pixel over the station. As the MAX-DOAS is performing elevation scans in a fixed azimuth direction (35° N), APEX observations are also averaged along this line of sight in order to take into account the instrument directivity.

	Flight #1 (26-06-2019)		Flight #3 (28-06-2019)	
NO_2 VCD_{TROPO} pixel over MAX-DOAS station [a] (x 10^{15} molec cm^{-2})	8.7		6.8	
NO_2 VCD_{TROPO}-CRE pixel over station [a] (x 10^{15} molec cm^{-2})	9.3		7.7	
NO_2 VCD_{APEX} (x 10^{15} molec cm^{-2})				
Averaged in TROPOMI pixel over station	8.6		7.2	
APEX pixel over station	8.4		6.4	
APEX pixels averaged along MAX-DOAS viewing direction	13.1		7.9	
	TROPOMI overpass (14:56 LT)	**APEX overpass (14:07 LT)**	**TROPOMI overpass (14:19 LT)**	**APEX overpass (14:25 LT)**
NO_2 $VCD_{MAX-DOAS}$ (x 10^{15} molec cm^{-2})	25.0	13.2	6.7	6.7

[a] TROPOMI Pixel ID #2 in Table 7 for Flight #1 and Pixel ID #3 in Table 9 for Flight #3.

3) There are some missing details about the APEX NO2 tropospheric column algorithm. Please add discussion about the reference spectra (i.e., is there one per flight? One overall? Where is it? I saw the comment that it was estimated using a mobile MAXDOAS) also please add some text that discusses how APEX tropospheric vertical columns are computed (e.g., is it similar to Sect. 3.2.2 and 3.3 in Lamsal et al. http://dx.doi.org/10.1002/2016JD025483 ?)

APEX NO_2 VCD retrievals are deliberately not discussed in full detail here as this has been done extensively in Tack et al. (2017) (https://doi.org/10.5194/amt-10-1665-2017) and also partly in Tack et al. (2019) (https://doi.org/10.5194/amt-12-211-2019). Tack et al. (2017) focuses on the development of the APEX NO_2 retrieval algorithm (which is indeed similar in concept to Lamsal et al. (2017)) and is applied on data acquired in 2015 over the Antwerp and Brussels region. The developed retrieval algorithm has been applied to the data acquired for the study under review. We prefer to avoid repetition and a too lengthy paper and want to keep the focus on the actual comparison/validation and study on impact of spatial resolution. Having a full discussion again on the APEX retrieval would be out of scope for this paper and it would similarly require a full discussion on the TROPOMI retrievals. We assume that the retrieval algorithms are well documented for both, TROPOMI retrievals in the ATBD and APEX retrievals in Tack et al. (2017). We have adapted Sect. 4.1 and 4.2 in such a way to emphasize why we don't include a full discussion on the retrieval algorithm and highlighted explicit references to the relevant sections in Tack et al. (2017) for the readers, interested in more details about the APEX retrievals.

For each flight, a reference spectrum was selected in a clean background area, upwind of the main sources, and the residual amount of NO_2 in the reference was estimated from co-located mobile-DOAS measurements. This has also been added to Sect. 4.2.

We have updated Sect. 4.2 as follows:

"The APEX NO_2 VCD retrieval scheme is similar in concept to the TROPOMI one and the developed algorithm is well documented in Tack et al. (2017). A full discussion on the retrieval algorithm is beyond the scope of this paper. Therefore, we refer to Sect. 4.1, Sect. 4.2, Sect. 4.3, and Sect. 4.6 in Tack et al. (2017) for all details on the APEX DOAS analysis, reference spectrum, AMF computation, and NO_2 VCD error budget, respectively. The DOAS spectral fit is based on the QDOAS software (Fayt et al., 2016) applied in the 470-510 nm spectral range, optimal for NO_2 retrieval from APEX. Note that interference with unidentified instrumental artefacts or features prevents us from extending the fitting window to wavelengths lower than 470 nm as discussed in Popp et al. (2012) and Tack et al. (2017). Key parameters for the NO_2 SCD retrieval are provided in Table 3. For each flight, a reference spectrum was selected in a clean background area, upwind of the main sources, and the residual amount of NO_2 in the reference was estimated from co-located mobile-DOAS measurements. …"

4) How is sigmaAMF_APEX computed?

Similarly as for comment 3, the APEX NO_2 VCD uncertainty budget is not discussed in full detail here as this has been done extensively in Tack et al. (2017) (https://doi.org/10.5194/amt-10-1665-2017) and also partly in Tack et al. (2019) (https://doi.org/10.5194/amt-12-211-2019). However, we agree more details should be added here, as well as clear references for readers that would like to have a full discussion.

We would like to refer to Section 4.6 in Tack et al. (2017) (https://doi.org/10.5194/amt-10-1665-2017):

"The error in the calculation of the air mass factor σAMF_i is caused by the uncertainties in the assumptions made for the radiative transfer model parameters (See Sect. 4.3.1). The contributing uncertainties can be summed in quadrature to obtain an overall error estimate σAMFi . According to Boersma et al. (2004), the error budget associated with the computation of the AMF is dominated by the cloud fraction, surface albedo and NO2 profile shape: (1) as flights took place under clear-sky conditions, cloud fraction is not considered an error source in this case. (2) Sensitivity tests, performed in Sect. 4.3.2, indicate that the surface albedo has the most significant impact on the effective light path, thus on the AMF. Within the albedo 1σ interval, the AMF variability can be up to 65 %. However, as absolute radiances can be directly derived from the APEX instrument, the albedo can be determined with relatively high accuracy. For a realistic estimate of the uncertainty, the following study was performed: several albedo types were measured in the field with an ASD FieldSpec-4 spectrometer (http://www.asdi.com/products-and-services/ fieldspec-spectroradiometers/fieldspec-4-hi-res) and compared to the APEX surface albedo. For the wavelength 490 nm, the average albedo error over all targets is 10 %, which is assumed to be a realistic estimate of the uncertainty related to the a priori surface albedo. (3) Based on the sensitivity study performed in Sect. 4.3.2, the uncertainty related to the a priori NO2 profile shape is lower than 8 %. (4) According to the performed simulations, the uncertainty related to the assumption of a pure Rayleigh atmosphere is estimated to be less than 10 %. (5) Both the viewing and sun geometry can be determined with high accuracy, thus the impact on the error in the AMF computation is expected to be small. Moreover, the performed sensitivity study, summarised in Table 5, has revealed that varying input for the viewing/sun geometry has a very low impact on the TAMF variability. Therefore it is assumed that the uncertainties related to RAA, VZA and SZA are less than 1 %. Finally, all error sources contributing to the overall error σAMFi are summed in quadrature and an estimate of approximately 15 % is obtained."

We have added more details on this in the manuscript as follows: "A full error budget for APEX NO_2 VCD retrievals has been discussed in Sect. 4.6 in Tack et al. (2017). Like for TROPOMI, the overall error on the retrieved APEX NO_2 VCDs, σVCD_{APEX}, is dominated by uncertainties related to the DOAS fit and AMF computation. The error on the retrieved DSCD or the slant error, $\sigma DSCD_{APEX}$, estimated from the fit residuals in the DOAS analysis, is 3.1 x 10^{15} molec cm^{-2}, on average. The error on the AMF computation, σAMF_{APEX}, depends on uncertainties in the assumption of the RTM inputs with respect to the true atmospheric state and is dominated by systematic errors in the surface albedo, NO_2 profile, and aerosol parameters. An estimate of approximately 15% is obtained for σAMF_{APEX}, following the detailed error budget described in Sect. 4.6 in Tack et al. (2017)."

5) It is interesting in Table 4 how the bias/slopes are different between the two cities. Antwerp has a lower slope for all three column comparisons as well as a larger negative bias. Any comment on this?

Your observation is correct. We checked the individual correlation plots for the different flights and it is hard to give a conclusive explanation based on the current data sets. The main difference between the two data sets is the type of emissions: prevailing industrial emissions in Antwerp and more traffic emissions in Brussels. This leads to a larger dynamic range and heterogeneity in the NO_2 field for the Antwerp region. Even if the APEX measurements are averaged within the TROPOMI pixel footprints, this still might have an effect for example due to the non-perfect time coincidence, point spread function, local albedo variability, etc. However, note that the correlation coefficient does not seem to be affected. It is hard to say as we don't have enough statistics. As new flights over both areas are expected in summer 2021, we hope to be able to check this again if it is a coincidence or really something geophysical.

6) On page 5, there is discussion about AOT measurements. Were any observed in Antwerp or only in Brussels?

Unfortunately no AOT measurements were done in Antwerp. Due to restricted national funding, this was a "lightweight" campaign and we relied on existing ground-based stations like the CIMEL and MAX-DOAS station we have in Uccle. A new S5P validation is scheduled in summer 2021 based on ESA funding which would give is more room to invite other teams and maybe add additional instruments in the two regions.

7) In Figure 15 and Sect. 6.2: why does the color bar go to zero if the background is 3x10ˆ15 and the detection limit is assumed at 5.1x10ˆ15? I am not sure if this is an oversight or if the section needs some clarifying discussion about the interpretation of this figure.

Thanks for pointing this out. We took indeed a standard color bar between 0 and 5 x 10ˆ16 molec cm^{-2} while the data shown is only ranging between 0.51 and 5 x 10ˆ16. Synthetic NO_2 VCDs below the detection limit of 5.1 x 10ˆ15 molec cm^{-2} are masked white and indeed even without masking, the lowest values would be 0.3 x 10ˆ16 molec cm^{-2} and not 0. But note that no VCD values in the plot had the deep blue colors representing 0 to 0.51 x 10ˆ16 molec cm^{-2}. To avoid any confusion we have adapted the colorbar with limits between 0.5(1) and 5 x 10ˆ16.

8) Page 1 Line 31 and generally in the paper: These biases are for these Belgian cities but are stated as general results for 'urban areas'. Could these results perhaps be different in other cities?

Indeed, this can be certainly different for other cases, depending on the amount of heterogeneity in the NO_2 field as well as the satellite pixel size (at nadir or more at edge of the swath). These nuances are well discussed in Sect. 6.1, also with reference to other studies. But indeed the statement in the abstract is "too strong" like this. We have adapted this in the abstract to (also following comment #11 from reviewer #2): "For a case study in the Antwerp region, the current TROPOMI data underestimates localised enhancements and overestimates background values by approximately 1-2 x 10^{15} molec cm^{-2} (10- 20%)."

For the same reason the related paragraph in the conclusion was adapted to:" The TROPOMI spatial resolution is limited to resolve fine-scale urban NO_2 plumes and can cause a considerable smoothing effect in case of the observation of strongly polluted scenes with steep gradients. This depends both on the instrument pixel size and the amount of heterogeneity in the NO_2 field. The high-resolution APEX retrievals allow to monitor the effective horizontal variability in the NO_2 field at much finer scale. In Sect. 6, the impact of smearing of the effective signal due to the finite satellite pixel size was studied for the Antwerp region based on a downsampling approach of the APEX retrievals. Assuming a pixel size of 25 to 20 km^2, equivalent to the initial 3.5 km x 7 km and new TROPOMI 3.5 km x 5.5 km spatial resolution (at nadir), the TROPOMI data underestimates localised enhancements and overestimates urban background values by approximately 1-2 x 10^{15} molec cm^{-2}, on average, or 10% - 20%, for the Antwerp case study. The average under- and overestimation is further reduced to 0.6-0.9 x 10^{15} molec cm^{-2}, or smaller than 10%, when increasing the pixel size to 1 km^2. Therefore, detailed air quality studies at the city scale still require observations at higher spatial resolution, in the order of 1 km^2 or better, in order to resolve all fine-scale structures within the typical heterogeneous NO_2 field."

Please see also a related comment (comment #8) from reviewer #2.

9) Technical Comments: Page 1: Line 23: You refer to the slope of 0.93 after the introduction of the CAMS profile, however the original slope is not listed. Please add this to the abstract to be consistent.

We suggest to change to "When replacing the coarse 1° x 1° TM5-MP a priori NO_2 profiles by NO_2 profile shapes from the CAMS regional CTM ensemble at 0.1° x 0.1°, R is 0.94 and the slope increases from 0.82 to 0.93. The bias is reduced to -0.1 ± 1.0 x 10^{15} molec cm^{-2} or -1.0% ± 12%."

10) Page 3 Line 1: please add the TROPOMI resolution sooner than is mentioned in page 3 line 15 as it is referenced in relation to other missions.

You are right the resolution should be given here. We have moved the sentence from line 15 (initially 3.5 km x 7 km at nadir observations and 3.5 km x 5.5 km since 6 August 2019) and changed the sentence at line 15 to "The APEX spatial resolution is considerably higher than the typical resolution of spaceborne sensors. For example, one TROPOMI pixel of 3.5 km by 7 km comprises approximately 4000 APEX pixels."

11) Page 3: Please consider swapping the placement of the second and third paragraphs in this page (Paragraph 2 being 'In this study. . .' and Paragraph 3 being 'Richter et al. . .'). It would improve flow as it talks about the challenges then state how this study addresses those challenges

Thank you for the suggestion. We agree swapping the two paragraphs improves the flow.

12) Page 3 Line 31: There is this reference also in AMTD. https://amt.copernicus.org/preprints/amt-2020-151/ Perhaps make the statement more defining to the region studied or other details. Or remove/edit accordingly.

The study https://doi.org/10.5194/amt-2020-**151** was indeed submitted to AMT in the same week as the study under review (https://doi.org/10.5194/amt-2020-**148**). We have adapted the paragraph in the manuscript and we have added a proper reference, now it is available:

This is one of the first publications assessing TROPOMI NO_2 retrievals over strongly polluted regions based on the comparison with airborne remote sensing observations and it is one of the first airborne spectrometer data sets coinciding in space and time with a large amount of fully sampled satellite pixels. At the same time the study of Judd et al. (2020) on the Long Island Sound Tropospheric Ozone Study (LISTOS) campaign in the New York City/Long Island Sound region has been submitted. Earlier studies reporting on the validation of spaceborne observations based on airborne spectrometer data, such as Heue et al. (2005), Constantin et al. (2016), Lamsal et al. (2017), Broccardo et al. (2018), and Merlaud et al. (2020) have shown high potential but are scarce, mainly due to the relatively large pixel footprint of TROPOMI's predecessors with respect to the area that can be covered with an airborne mapping spectrometer.

13) Page 5 Line 30: AURA should be Aura. It is not an acronym. Same with PANDORA– >Pandora.

Thanks for clearing this out. This is corrected throughout the manuscript.

14) Page 7 Line 13-14: 'is based' is used twice in one sentence.

Corrected to:

"The processor is based on a retrieval-data assimilation-modelling system using the 3-D global TM5-MP chemistry transport model (CTM) (Williams et al., 2017). It follows a 3-step approach: "

15) Page 9 Final paragraph: This figure shows the difference in Box AMFs based on albedo, and therefore belongs better in the next section rather than Sect. 4.3.1 about A priori NO2 profiles.

We prefer to keep the discussion on the box AMFs (and Figure 3.b) in section 4.3.1 on the NO_2 vertical profiles. They are related as Figure 3.a provides the concentration at each altitude layer while the Box AMF in 3.b provides the vertical sensitivity to NO_2. It is true that we provide the box AMF profiles for two different albedo scenarios, but the key discussion is on the vertical sensitivity. To make this more clear we suggest to change the title of Sect. 4.3.1 from "A priori NO_2 profile" to "NO_2 profile and vertical sensitivity".

16) Page 12: Line 20: Word Choice: refer to Antwerp and Brussels as regions or cities, rather than separate campaigns.

Indeed referring to it as separate campaigns is not appropriate. We suggest to refer to it as regions here

17) Figure 7: please point out the airport for ease of identifying when discussed in the text on Page 13

We have added a white square in Fig. 7 a) and b) and properly referred to it in the caption and text.

18) Page 14 Line 21-22: It is premature to make a statement about the error bars in Figure 8 since the figure is not introduced until a couple pages later. I suggest removing that sentence here.

True, we have removed the sentence in this section. Note that in the next section (Sect. 5.2.2), we added an explicit reference to Eq. 1 and 2: "Vertical error bars indicate the overall error in NO_2

VCD_{TROPO} (Eq. 1), while the horizontal whiskers represent the error in NO_2 VCD_{APEX} retrievals (Eq. 2), averaged over all APEX pixels coinciding with a particular TROPOMI pixel."

19) Page 18 Lines 15-23: Please clarify this discussion on how the temporal variability between TROPOMI overpasses is computed, especially with the differences in pixel footprints. It is hard to follow what those statistics are referring to and how they are computed.

Indeed some details for the comparison were missing here. Prior to the comparison we have regridded the data sets to a common grid of 0.1°. In a next step we compared the absolute and relative differences between the two overpasses (grids) on the same day for the full Belgian domain. So the statistics are the average for all "difference pixels" over Belgium. We have clarified this section as follows: " Both on 26 June and 29 June 2019, there were two early-afternoon S-5P overpasses over Belgium with a time difference between the two orbits of approximately 100 min. To assess the impact of the temporal NO_2 variability, the changes in the NO_2 field have been studied in the subsequent overpasses for the Belgian domain. Prior to the comparison, the data sets have been regridded to a common grid of size 0.1°. On June 26, the absolute value of the differences observed over the full Belgian domain is $3.8 \pm 5.3 \times 10^{14}$ molec cm^{-2} or 12% ± 10%, on average. A maximum difference of 5.8×10^{15} molec cm^{-2} or 57% was observed for a pixel over the harbor of Antwerp, most likely due to a combination of moving air masses in the key plumes and slight changes in the wind pattern. Additionally, the TROPOMI pixel footprints have different sizes and orientations which also has an effect when sampling the effective NO_2 patterns and when regridding to the common grid size of 0.1°. On June 29, the absolute value of the differences observed is $3.6 \pm 3.2 \times 10^{14}$ molec cm^{-2} or 11% ± 8%, on average, with a maximum of 2.0×10^{15} molec cm^{-2}, again seen over the harbour of Antwerp."

20) Page 19 Line 10: delete 'allow to'

Corrected

21) Figure 1: Adding a label for Stabroek as the other ground site where meteorology is measured in Antwerp could be helpful.

Ok, a label was added for Stabroek, Antwerp.

22) Figure 13: Please make the red dots more visible. (Perhaps white like in other Figures). Also in the caption write what they are. And as a suggestion, pull the color bar legend out of panel (a) and make larger since it refers to all four maps.

We have made the red dots larger and white like in Fig. 6 and 7, and described it in the caption. We have extracted the legend from map a) and use it as a general legend for all maps. Note that we have put the different parts of the figure together in the word file. We will make a proper merged figure with the legend more central over the four plots for the final version.

For consistency we have applied the same to Figure 14 and its caption.