Dear referee #2,

We appreciate your comments to improve the manuscript. Below you will find our response in blue and modifications to the manuscript in green.

General comment:

One of the key points of the parameterization method proposed by Dix et al. is the selection of appropriate reference spectra which are close enough in time (or SZA) with the limb spectra, so that the contributions from above and below the instrument altitude are cancelled out and therefore the resulting limb dSCDs is only representative of the atmospheric layers close to the instrument altitude (the so-called sensitivity range S in the manuscript). However, when applying this method to measurements from the TORERO field experiment, the authors used one fixed reference spectrum per flight (see page 25, lines 23-24). This means that except for the limb spectra recorded close to the reference spectrum, there is always a significant difference in SZA between reference and limb measurements. So, most of the time, we are potentially in conditions where the dSCD contributions from outside the range S is significant, making the retrieval less accurate. The authors should explain why they proceed like this, instead of analyzing each limb spectrum with the closest zenith or EA10_ spectrum as reference. A useful and interesting test would be to compare VMRs retrieved using fixed and 'closest in SZA' reference spectra.

This is a very good point. We have added an explanation on the choice of reference spectra and why the field data is analyzed with one fixed reference per flight.

p7, line 3: Ideally each set of EA 0° dSCD measurements has a reference recorded under the same atmospheric conditions. Actual field data might not contain a sufficient amount of suitable references for a variety of reasons. Therefore a dSCD correction term is included, $dSCD_c^i$, to account for contributions from outside *S* that are not cancelled out by the reference.

p.25, line 24: Using a fixed reference is for practical reasons. The number of reference spectra suitable for the parametrization method is limited in the TORERO data set, particularly because zenith references for BrO and NO_2 need to be close in time to the O_4 EA 10° references. Not all flights provide more than one set of suitable reference spectra. To keep results comparable between flights, each flight is analyzed with one fixed reference.

The comparison of VMRs retrieved using a fixed or "close in SZA" reference is contained in the results of the sensitivity studies that vary SZA and ΔSZA. We have no reason to believe the effects would differ in field data.

Specific comments:

Page 4, line 15: the upper layer of the sensitivity range S should be no more than 3.5km above the altitude layer of the instrument. The authors should justify this upper limit value of 3.5km.

Agreed. We have added further information on the choice of the upper and lower boundaries (see also response to reviewer #1).

p4, line 13: Initial sensitivity studies have shown that the parameterization method typically works best when about 90 % of the sum over the dBox-AMF trace is included in *S*. Therefore the lower boundary n_L is set to 1 km below the instrument layer, n_{instr}, while n_U is set to the altitude layer before the difference between two consecutive dBox-AMFs is smaller than 10 %, and no more than 3.5 km above n_{instr}. Due to the distinct shape of the dBox-AMF peak, the placement of the lower boundary is less critical and remains fixed, while the placement of the upper limit is flexible. The dependency of *S* on altitude, reference and wavelength is shown in Fig. S1.

Page 9, lines 20-23 and figure 2: Where these aerosol extinction profiles come from?

Any reference(s)? If not, the authors should explain how they constructed them.

We have added explanation on how the aerosol extinction profiles were constructed.

p9, line 22: Profile 1 is similar to extinction we found over pristine ocean during the TORERO project. Profiles 2 and 3 are constructed specifically for the sensitivity studies here to investigate the effects of higher AOD (2) and lofted pollution (3).

Figure 8 and Section 5.3: correlations of TORERO AMAX-DOAS BrO, IO, and NO2 VMR data retrieved by parameterization and optimal estimation are shown and discussed only for a selection of flights. Why data from all 17 flights are not plotted? Why the selected flights are different for BrO and IO on one side, and NO2 on the other side?

We have added an explanation on the choice of OE profiles used for comparison and add NO_2 comparisons for RF01, RF04, RF05 and RF17. At the time of manuscript submission NO_2 parameterization data for these four flights was not available (see also answer to following comment).

p.28, line 2: To increase statistics for the correlation, trace gas profiles from more flights are included, i.e. from RF01, RF04, RF05, RF12, RF14 and RF17 for all three trace gases (see also

Wang et al. (2015) and Volkamer et al. (2015)). Profile selection is based on availability of high quality OE profiles.

Figure 9 and Section 5.4: BrO and IO VMR profiles retrieved for all 17 TORERO flights are shown. Why a similar plot for NO2 is neither included, nor discussed in the manuscript?

Good point, we will add an explanation on missing NO_2 parametrization data.

p.26, line 25: For NO_2, a typical TORERO $dSCD_{strat}$ value at 60° SZA is 1.5-3 x 10^{15} molec/cm^2, which is about a factor of 2-4 higher than a tropospheric EA 0° dSCDs measured in pristine background air (e.g. NO_2 c-profile). Here, our EA 10° dSCD fit error of 3-5 x 10^{14} molec/cm^2 is almost on the same order as the tropospheric EA 0° dSCDs, underlining the need for a highly accurate characterization of the stratospheric VCD by EA 10° measurements. Since TORERO specifically targeted very pristine air masses, we refrained from running the NO_2 parametrization retrieval on the complete data set, and focused instead on select case studies with $dSCD_{strat} << $ EA 0° dSCDs. For these cases evaluation of NO_2 is still possible down to as low as 10 pptv, as has been demonstrated in Fig.10 in Volkamer et al. (2015).