We would like to thank the editor for his constructive comments and suggestions.
Below we reply to the raised issues point by point.
Figure numbers refer to the discussion version of manuscript and supplement.

Main comments:

1. One of the strengths of STREAM is its capability of exploiting measurement data also over continents, for estimating Vstrat. The largest benefit with respect to other STS methods would hence be at clear sky continental scenes. Please clarify if that is correct. If so, a dedicated analysis of such scenes would be valuable, eg based on the synthetic data used.

Reply: One of STREAM's features is the inclusion of clouded pixels (via w_{cld}), which provide additional supporting points over (not too polluted) continental regions, and thus avoids interpolation errors or wave fitting artefacts. This is expected to result in more realistic stratospheric estimates over continents and thus more accurate tropospheric residues, not only for clear sky.

In section 5.2.2, STREAM is compared to the GOME-2 GDP 4.7, which is based on a modified RSM with a pollution mask. Figure 12 clearly shows that the GDP 4.7 TRs are biased low over continents, and that STREAM overcomes this bias.

In addition, sensitivity study 4.2.1 clearly demonstrates the improvement due to the application of w_{cld}; if the latter is omitted, TR over potentially polluted regions (which are at large part congruent with continents) is obviously biased low (see Fig. S11 of the supplement).

2. STREAM differs from other STS methods in the sense that scenes with tropospheric NO2 are not excluded but that a low weight is assigned. Please clarify whether this introduces a bias in Vstrat, and - if so - whether the tropospheric-residual-based weight mitigates such a bias. STREAM is tested in a scenario where only the pollution weights and the cloud weights are applied (Section 4.2.5), which would be the ideal test scenario to answer this question.

Reply: Generally, all modified RSM methods estimate the stratospheric column from total column measurements. Thus, the estimated stratospheric field is generally expected to be biased high due to the tropospheric background (if the latter is not explicitly corrected for based on CTMs or other a-priori). Consequently, TR are expected to be biased low. The effect is dampened in STREAM by the high weight of clouded pixels. But still, TR over the Pacific and remote regions is about 0.1 CDU for OMI, and less for GOME 1/2 and SCIAMACHY, while models expect a slightly higher background (see the reply to the first remark of reviewer 2).

This topic is mentioned in the introduction (item (a) on 2nd page), but still we have tried to further clarify this fundamental aspect in the revised manuscript by extending the end of the introduction:

"In particular, clouded observations are weighted high, as they provide direct measurements of the stratospheric field. This approach dampens the small but systematic high bias of stratospheric columns estimated from total column measurements and the resulting low bias of tropospheric columns."

However, the TR weight cannot correct this small systematic bias, as it is only applied if the mean TR has an absolute value > 0.5 CDU (otherwise, noise would introduce severe artefacts via w_{TR}, see section 2.2.3). We have modified the end of section 2.2.3 accordingly:

"Note that due to the threshold of 0.5 CDU (criterion 2), w_{TR} cannot correct small biases such as the expected low bias in TR caused by estimating the stratospheric column from total column measurements."

3. The pollution weight is derived from multi-annual mean tropospheric NO2 column data from SCIAMACHY (Beirle and Wagner, 2012). Please clarify how spatially smooth these data are, in the context of the width of the convolution kernels, and vis-à-vis the application of STREAM to higher spatially resolving instrumentations.

Reply: The definition of the pollution proxy P is provided in section S2.21 of the supplement and displayed in Figure S2. We have clarified this in the revised manuscript:

"Details on the definition of P are given in the supplement (sect. S2.2.1), and P is displayed in Fig. S2d."

As P is intended to indicate *potentially* polluted regions, and the original SCIAMACHY climatology is intentionally smeared out by convolution, SCIAMACHY's spatial resolution is not critical.

We have extended section S2.2.1 accordingly:

"1. Grid pixels with a mean TVCD below 1 CDU are removed (Fig. S2b).

2. The resulting clipped climatology is smoothed by convolution with a 2D-Gaussian with $\sigma = 2°$ (Fig. S2c).

3. For the pollution proxy P, values between 0 and 1 CDU are set to 1 CDU. By this operation, a "safety margin" of potentially polluted areas is created (Fig. S2d).

Note that due to steps 2&3, the initial spatial resolution of SCIAMACHY is fully sufficient for the definition of P (and thus w_{pol}) even for applications of STREAM to instruments with better spatially resolution."

Minor comments:

- The assumptions made for the stratospheric Air Mass Factor should be specified (it is expected to be approximated by the geometric AMF).

Reply: Stratospheric AMFs have been taken from the same data source as the NO_2 columns, as listed in Table 1.

We have extended section 3.1 accordingly:

"Table 1 summarizes the characteristics of the instruments ~~discussed in this study~~ and provides references to the data products used in this study, from which the total SCD, the stratospheric AMF, and the cloud fraction/cloud top height are taken as input for STREAM."

- p4, line 16: Typo: complImentary → complEmentary

Reply: Done.