Response to Referee 2

We would like to thank the referee for her/his helpful comments and remarks. We expect the revised version to address all comments.

We reproduce comments from the referee in "script" font followed by our answer. A document listing the revisions to the manuscript is also provided.

`Review for Manuscript #gmd-2018-216 "A benchmark for testing the accuracy and computational cost of shortwave top-of-atmosphere reflectance calculations in clear-sky aerosol-laden atmospheres" by Escribano et al.`

`General comments:`

`This paper gives very comprehensive comparisons between a benchmark radiative transfer model and other similar but simplified RT models. RT model is one of the most important components in both remote sensing applications and atmospheric modeling studies. However, just a few papers or reports document the performances of different RT models. The primary object of this study is to provide a benchmark that can help people understand RT model performance in terms of computational cost and accuracy. In my opinion, this work is important, method is solid, results are reproducible and reliable, and the paper is well organized and written.`

`Generally, I just have one concern about the comparison. The authors gave comparisons at different geometries (SZAs, VZAs, and Azimuthal angles), as described in Table 2. However, I believe it would be more important to give comparisons as a function of scattering angle. For aerosols with "rainbow" feature, aer_ss for instance, the accuracy of I and/or Q/U components may be largely influenced by the number of streams when scattering angles are close to 140 deg.`

We appreciate the reviewer's comments and, indeed, the dependence of model errors on the scattering angle is important. We have added two figures to the appendix showing violin plots of the relative errors as function of the scattering angle, and the polarised reflectance computed with the reference model. We have also added a comment Section 4.2.

We have estimated the accuracy of the DISORT model according to the number of streams, but it would be out of scope to do the same for the VLIDORT model which is our reference model, as this model is relatively expensive.

`Other minor points are listed below:`

`1. Table 2, second line: Solar viewing angle -> viewing angle`

We have changed "Solar viewing angle" by "Viewing zenith angle".

`2. You may want to put this paper in the references: Ding S, Xie Y, Yang P, et al. Estimate of radiation over clouds and dust aerosols: optimized number of terms in phase function expansion. 2009;110:1190-8.`

The reference has been mentioned in Section 2.3.

`3. Page 2, Lines 28-32: Please consider add the two papers which describe fast IR and SW RT models and performance comparison against DISORT in the references:`

Wang, C., P. Yang, S. Platnick, A. K. Heidinger, B. A. Baum, T. Greenwald, Z. Zhang, and R. E. Holz, 2013: Retrieval of ice cloud properties from AIRS and MODIS observations based on a fast high-spectral-resolution radiative transfer model. J. Appl. Meteor. Clim., 52, 710-726, doi:10.1175/JAMC-D-12-020.1.

Wang, C., P. Yang, S. Nasiri, S. Platnick, B. A. Baum, X. Liu, and A. Heidinger, 2013: A fast radiative transfer model for visible through shortwave infrared spectral reflectances in clear and cloudy atmospheres. J. Quant. Spectrosc. Radiant. Transfer, 116, 122- 131, doi: 10.1016/j.jqsrt.2012.10.012.

We have added these references in the introduction section.

4. Page 5, vertical profiles: For aerosols with the given vertical profile, is there a minimum AOD value that is considered for each layer? If there is a minimum value, then the computing time of AOD=2 case should be longer than AOD=0.2 since for the latter case, less layers contain aerosol particles, right? The computing time comparison maybe biased.

There is no minimum value for AOD in the computations, and we had not tuned the models to avoid computations in layers with low AOD. By model design, the computing time for 6SV2 is linked with the AOD (shown in Figure 6). For FLOTSAM, it is stated in Section 4.3 that the number of operations is the same in every case, which implies that the computing time is similar for all the cases. It can be seen that the standard deviation of the computing times in Table 4 is small.

For DISORT, there is a difference in computing time for the case AOD=0 or AOD>0. In fact, for the Lambertian cases and computing 20 geometries (the *mult* case), the average computing time for AOD=0 is 0.024 s, while for AOD>0 it is around 0.3 s.

The same test cases are simulated for all the models, and there is no reference model for the computation time, thus a bias in the comparison is hard to identify.

We have modified the manuscript in Section 4.3 (Page 18, L6) as follows:

"... because DISORT efficiently decreases the number of internal computations for symmetric geometries (solar zenith angle equal to zero*, 21% of the cases) and it is faster for AOD=0 (3.4% percent of the cases)*. For the remaining cases of Table 5, there is almost no spread in the DISORT runtime. "