Reviewer #2

I thank the authors for taking the time to rewrite significant portions of the text. It has made a difference. I find the flow of the paper easier to follow than in the previous version, with more coherency between sections. The authors have added more discussion of the motivations behind the steps they've taken, which makes other unchanged components of manuscript easier to understand. There are several small issues remaining (noted below), but these do not detract from what is now a well-constructed paper.

1. Section 2.3, 1st paragraph. The logic in this paragraph is not clear. The paragraph starts by saying that the SL kernels are too small (too narrow) based on the observation of the stray light row signal increase compared to its calculated increase. The second sentence seems to say that if the kernel was sufficiently large the SL row contents would be much lower than observed. Do the authors instead mean the SL row contents would be much larger than observed? Or is the problem with the use of the word "would" in the second sentence? Perhaps it should be "does".
   Thank you for pointing out the difficulty in the wording. What we intended to convey is that with the original kernel, only a large increase in the signal elsewhere (in Band 2) can lead to an increase in the straylight in Band 1. There is indeed such an increase in signal between 310-320nm due to the bleaching effect, but the small fraction of that extra signal that is redistributed to Band 1 (via the kernel) is far too low to account for the actual measured increase in the Band 1 straylight signal.
   We agree that replacing "would" with "does" in the original paragraph improves clarity (line 175).
2. The description in Section 2.3 of how the SL kernel and SL row information is combined is reasonably clear, though perhaps more complicated than it needs to be. To say that the SL kernel remains static is not really true since the dynamic SL correction effectively increases and decreases the kernel tails in a way that explains the SL row contents. Even though this is not how the correction is implemented in the code, it may help to describe the new correction as a time varying kernel where only the tails are allowed to vary.
   Thank you, this is indeed another way to explain the effective result of the second dynamic correction.
   After line 196, we added: "Alternatively, this extra correction can be seen as a temporally varying adjustment of the tails of the straylight kernel. "
3. Section 2.4. It will help to add a sentence that provides the motivation behind explicitly adjusting the SL kernel. The Section 2.3 discussion gives the impression that the dynamic SL correction compensates for errors in the SL kernel. Why then is it important to get the SL kernel as close to correct as possible? The answer becomes clear in the Section 4 evaluations, but it would help the reader if those justifications were prefaced in this section. Agreed. At the end of Sect 2.4, we added: "Thus, although the dynamic straylight correction ameliorates unwanted effects such as overcorrection, it is nonetheless worthwhile to start with a good initial choice for the straylight kernel (in this case, the elliptical kernel), as it is demonstrate by the results in Section 4."
4. Section 3.2. The authors describe their use of CAMS to seed their RTM calculations, and state that the CAMS total ozone is scaled to that of TropOMI and ozone profiles are scaled to those of MLS. The authors do not indicate if this profile adjustment is a one-time scaling or the rescaling occurs independently for each CAMS-MLS matchup. The former amounts to a time-

independent soft calibration adjustment to MLS ozone, while the latter would tie TropOMI long-term ozone profile changes to those of MLS. This distinction, while not very important when assessing L1 calibration, is very important in understanding the independence of the L2_O3_PR product. Since this paper is not about the L2 profile product, I leave it to the authors to decide whether to clarify this issue or leave it ambiguous.

Thank you for pointing out that this aspect was not clear. We scale the CAMS profile to match the TROPOMI total ozone and MLS profile for each individual observation date used in the RTM calculations. While it is true that this detail is not essential to the manuscript's main focus, we have added clarification in line 316 ("*for each specific observation date*") to make this explicit.

5. Section 4, Line 405. The text references Figure 11l, but there is no such figure. Is this a typo?

   Yes, this is a typo. The text should refer to Figure 11. The manuscript has been corrected accordingly.

6. Section 5, Lines 448-450. The authors reference Figure 16 when stating that the L1 calibration improvements are the main cause of across-track reduction in tropospheric ozone anomalies. No doubt the observed changes were caused by calibration changes, but evidence from Figures 16e and 16f that anomalies have decreased is scant. There is certainly some improvement on the western side of the swath, but it is difficult to see the other improvements the authors describe. Likewise, a reduction in importance of the soft calibration to across-track anomalies, as claimed by the authors, would suggest similarity between Figures 16f and 16h. While the two share some features, they are not obviously closer than the similarity between Figures 16e and 16g. The authors note the smaller latitude dependence of the anomalies in v3.0. This does not necessarily represent an improvement since the v3.0 anomaly is now more recognizable as an across-track bias rather than as geophysical in origin. I find the evidence in this figure rather limited compared to the conclusions the authors draw from it, and would prefer their statements were less expansive in this regard.

   In the text, we have clarified the specific regions where we observe the largest impact, since the reduction is not uniformly visible across the entire swath. We also updated the language to be more explicit without overclaiming. The two key points we intend the readers to take from this figure are:

   - in both retrieval versions, with and without soft calibration correction, the anomalies profiles lie more closely together when using L1 v3.0 than when using L1 v2.1;
   - the relative importance of the soft calibration correction to across-track anomalies appear reduced when using L1 v3.0, such that Figure 16h is closer to 16f, than Figure 16g to 16e

   The original text (Lines 448-450) has been updated to: "With soft calibration applied, the anomaly profiles *tend to show* reduced variability across latitude bands, with individual profiles lying closer together when using L1 v3.0 (Figure 16h vs Figure 16g). *These visual tendencies hint* that improvements in the L1 calibratio*n may be associated with the changes* in the across-track systematic structure. *Moreover, the incremental influence* of the soft calibration on the anomaly metric *appears smaller* when using L1 v3.0 than with L1 v2.1 (Figure 16h-f vs Figure 16g-e). "