Review of Miraille et al., 2026

Overall comments

The study by Mirallie et al., 2026 presents a methodologically original and timely adaptation of the BAyeSian Integrated and Consolidated (BASIC) merging framework of Ball et al. (2017) to ground-based partial column ozone records from the NDACC network, deriving regional stratospheric ozone trends for the 2000–2024 period. The work addresses a genuine community need: with the anticipated decline of limb-viewing satellite instruments over the coming years, well-consolidated ground-based composites will become increasingly central to monitoring ozone recovery, and the partial column framework is a sensible and well-reasoned approach to reducing the large noise inherent to individual profile records. The paper is well-structured and clearly written throughout, with a logical flow from data description to methodology and results that makes it accessible to a broad readership. The figures are generally informative and well-chosen, and the inclusion of detailed group composition tables and trend maps in the supplement is a valuable addition that substantially increases the paper's utility as a community reference dataset. The introduction of the alternative partial column set (aPC) to isolate UTLS variability, the correlation-based regional grouping methodology, and the direct comparison with the conventional weighted mean are all meaningful contributions that go meaningfully beyond prior work in this area. I feel the paper contributes significantly to the community and, importantly, the results are largely consistent with the existing literature in the upper stratosphere while offering new regional insight in the lower stratosphere. The significant negative trends found at Lauder and in the European UTLS are inspiring, and the transparency with which the authors discuss known limitations — such as the HILO ozonesonde artifact and the North Canada data quality issues — reflects well on the overall scientific rigour of the manuscript. I recommend this paper for publication with minor revisions. Please address my specific comments below.

Specific comments:

Systematic uncertainties excluded from BASIC weighting (section 2.2.2)

For FTIR retrievals, the paper states that systematic uncertainties (5–7% per partial column from spectroscopy, instrumental line shape, and temperature profile errors) are comparable to or larger than the propagated random uncertainties (~4–6%). Since BASIC only uses random uncertainties for weighting, FTIR instruments are effectively over-weighted relative to their total measurement reliability. The offset removal corrects for mean systematic offsets but not for time-varying systematic biases. This limitation should be acknowledged along with any potential implications.

Ozonesonde partial-column uncertainties (lines 188–191 vs. lines 245–249)

Ozonesonde partial-column uncertainties are obtained by summing up the individual uncertainties at each pressure level and in the the Dobson Umkehr section (lines 245–249) it is stated explicitly that the root-sum-of-squares method is used. This is an inconsistency between instrument types that should be briefly acknowledged. If random errors dominate for ozonesondes, the summation overestimates uncertainty by roughly $\sqrt{N}$ relative to quadrature, systematically downweighting ozonesondes in BASIC. I believe that the study would benefit by adding even a sentence noting the choice and its direction of effect.

CAMS-based region definitions (lines 273–284)

The spatial correlation structure used to define groups is derived from CAMS EAC4, which assimilates satellite ozone data. The groups therefore reflect how a model and satellite-constrained reanalysis represents ozone coherence, rather than how the ground-based observations themselves covary — a structural dependency that is worth acknowledging given the paper's stated aim of providing an independent ground-based reference. Additionally, CAMS data is available only from 2003 onward (as stated in Line 283), meaning the group definitions extrapolate back to the 2000–2002 portion of the trend period. A brief sentence noting this as a minor limitation would be appropriate.

Correlations between the oPC aPC in UTLS (lines 279–281)

The CAMS-based correlations are computed only for oPC and reused for aPC on the grounds that the two sets overlap in altitude range as stated by the authors. The supplement's group maps (Figures S7–S8) confirm the same groups are used across both sets. However, I think that the key difference between oPC and aPC lies precisely within the UTLS layer, which is governed by tropopause dynamics quite distinct from the lower stratosphere. I would ask the authors to add a brief discussion acknowledging that the UTLS groupings carry more uncertainty than the stratospheric ones in this regard and what their implications might be.

The BASIC prior (lines 350–354)

The prior is replaced from Ball et al. (2017)'s empirically-derived version with the McPeters and Labow (2012) ML climatology, to ensure independence from the input datasets. However, the ML climatology is itself built from Aura MLS and ozonesonde data, and several ozonesonde stations in this study (e.g., Lauder, Hohenpeissenberg, Boulder) very likely contributed to it — so the independence is not complete. More importantly, the paper acknowledges directly that the climatology "has a very large variability," making the prior essentially uninformative. I would simply ask the authors to tone down the independence claim slightly, acknowledging that the prior plays a minimal constraining role, rather than citing independence as a primary motivation for the change.

MCMC implementation (line 364)

The posterior is sampled via MCMC but no information is given on number of chains, burn-in length, or convergence criteria. Given that Figure 6c (August 2012) shows an explicitly bimodal likelihood — where convergence to the correct mode is non-trivial — I would ask the authors to add at least one sentence describing what convergence diagnostics were applied.

PCA uncertainty scaling (section 3.2.1)

The PCA-based multiplicative scaling of individual instrument uncertainties is the paper's key novel methodological contribution relative to Ball et al. (2017), however, Ball et al., (2017) used PCA to directly construct monthly uncertainties (their Eq. 5), whereas here it scales pre-existing propagated random uncertainties. These are fundamentally different operations. I would appreciate having a brief discussion on that in the study; why the scaling is multiplicative, and acknowledge that by down-weighting instruments deviating from the PCA consensus, the method could in principle suppress genuine physical signals captured only by a single high-resolution instrument.

Comparison with the most recent satellite literature (section 4, trend results)

The paper benchmarks its results primarily against Godin-Beekmann et al. (2022) and the WMO 2022 assessment. However, Sofieva et al. (2025) provides an updated multi-satellite LOTUS trend analysis extending to 2024, covering the same trend period as this study. Given that both works use the 2000–2024 period and the LOTUS MLR framework, a brief and direct comparison of the trend magnitudes in the upper and middle stratosphere between this paper's results and Sofieva et al. (2025) would considerably strengthen the results section and help the reader assess the consistency between the two approaches.

oPC vs. aPC results presentation in section 4

The results are presented layer by layer across both PC sets, which leads to some repetition and makes it difficult for the reader to grasp what the aPC definition actually adds over oPC in terms of trend detectability. I would encourage the authors to include a brief summary paragraph (maybe even a concise tabular overview) — at the end of Section 4 that directly contrasts the key trend differences between oPC and aPC for the groups where the distinction matters most (e.g., Lauder LS vs. UTLS, European LS vs. UTLS). This would make the value of the aPC contribution considerably clearer.