Partial review of resubmission of egusphere-2024-3260:

"Understanding Boreal Summer UTLS Water Vapor Variations: A Lagrangian perspective"

Hongyue Wang, Miejong Park, Mengchu Tao, Cristina Peña-Ortiz, Nuria Pilar Plaza, Felix Ploeger and Paul Konopka

Explanation for this partial review:

This is not a complete review and does not therefore include a recommendation on its acceptability for publication. The paper does appear to present interesting results that are relevant to the topic of the controls on lower stratospheric water vapor. Unfortunately, the paper is not well-written, and as a result I found it was taking more time than I had available to do a fair assessment of the paper's scientific merit.

So, the following includes a brief summary of the paper and a detailed commentary on Sections 1 through 3.1.1 up through line 236.

Summary:

As in their original submission, the authors present results of Lagrangian back-trajectories calculated from ERA5 meteorological fields to reconstruct the horizontal and vertical distributions of water vapor in the tropical upper troposphere and lower stratosphere (UTLS) during boreal summer. The paper poses three questions. First, how well do their simplified Lagrangian back-trajectory method reproduce water vapor values observed by SAGE III/ISS and MLS in the Asian Summer Monsoon (ASM) and North American Monsoon (NAM)? Second, are these values locally controlled or represent freeze drying upstream? Third, is the general tendency to a dry bias in the reconstructions related to particular processes?

The reconstructions are carried out using the CLaMS trajectory module (Konopka et al., 2022). Back trajectories are calculated for 180 days using the CLaMS model's trajectory module and were initiated from the satellite measurement locations and times. Results are presented for comparisons of the reconstructions with SAGE-III and MLS water vapor values at the back-trajectory initiation points. The water vapor reconstructions are based upon coldpoint temperatures, identified either from the local vertical temperature profile or from the back-trajectory minimum temperature (the Lagrangian CPT). Three types of reconstructions are thus done based on the type of CPT: (a) using local CPTs (LOC), (b) using the Lagrangian CPT for every single trajectory (LAG_single), and (c) using the average Lagrangian CPT for a cluster of 51 back trajectories.

The authors find that while both SAGE III and MLS reconstructions produce similar spatial patterns, the SAGE III reconstruction values tend to be higher than those for the MLS.

Reconstructions generally succeed in the ASM but they do not capture the NAM pattern. Reconstructions in the former region appear to be predominantly controlled locally while reconstructed water vapor in the NAM seems to be largely remotely controlled. Finally, the dry bias in the reconstructions over the ASM tends to increase with the intensity of convection.

General comments:

The revised submission addresses a topic that has received considerable attention over the past 20 years or so. And as in the original submission, it is a positive feature of the approach that the authors have undertaken in this work that they directly compare their water vapor reconstructions with SAGE-III ISS and MLS water vapor observations.

Specific comments:

Section 1

L.21: Suggested rewording:

"…**where air masses undergo slow diabatic ascent**…"

"…**into the stratosphere over time scales of weeks to months**."

L36: What exactly is the **tropical stratospheric water vapor anomaly**? An anomaly with respect to what? Likewise for the **summertime NH extratropical water vapor maximum.**

L41: Replace **predict** with **assess**.

L42: The sentence here is confusing. The large-scale temperature and wind fields are no more than a *representation* of the atmosphere that the reanalysis provides. They do not have their own set of processes. The freeze-drying occurring in the actual atmosphere can of course be estimated with reanalysis wind and temperature fields – that's the methodology here after all. But as I read the sentence, it implies that there is a freeze-drying mechanism for the large-scale fields and another at finer resolution.

L50: The phrase beginning with **to reconstruct**… and ending with **…Version 5.0 Lambert et al. (2017)** is a nightmare of parentheses. Since this is only the introductory section, there is no immediate need to specify version numbers of datasets or their specific references. This can and should be the business of Section 2. In any case, by eliminating all those version numbers and references, the profusion of parentheses goes away. Furthermore, identifying the SAGE III and MLS versions here implies that the updates they represent were somehow critical to the outcome of the study, which I am pretty sure is <u>not</u> what you are saying.

L59: This sentence amounts to an awkward juxtaposition inasmuch as you have just pointed out the advantage of the SAGE III data over MLS. The advantage of the <u>MLS</u> data of course is their unmatched temporal and spatial coverage. In this regard, SAGE III/ISS is a very poor cousin indeed.

L60: improper use of **furthermore.** Use another more appropriate conjunctive adverb.

L68: ...are most critical. Critical for what, exactly? Presumably remote control, but that is not stated.

Section 2

L83: It would be helpful here to describe the particular sampling challenge presented by the MLS data. A 10° latitude x 20° longitude grid box is a big piece of real estate but nonetheless at the latitudes of interest in this study, on most days it would be traversed by only a couple of MLS overpasses, one ascending (afternoon) and one descending (nighttime). The limb-viewing geometry also comes into play here, in particular the 200-km swath length along the orbit path and at the lower MLS levels (100 hPa and below) where spatial variability rears its ugly head. In short there are a number of temporal and spatial sampling considerations that almost certainly play a role in a three-dimensional gridding of MLS data, and the same can be said for SAGE III/ISS. Given their very different times at which they sample alone, MLS and SAGE III/ISS "see the world" as it were in different ways. Are these different perspectives significant? I don't know – I didn't do the work! But I think it's a fair question and should be addressed here.

L88: fix reference.

L90: missing terminal parenthesis, period and double space

L97: The latitude range 35°S-35°N is incorrectly referred to as the subtropics.

L97-L100: The previous subsection on MLS does not mention a specific period, so presumably the study period mentioned here, the months of August from 2017 to 2022, applies to both MLS and SAGE III/ISS. But perhaps not, as the second sentence in the paragraph says that data for the years 2020-2022 were <u>added</u> in order to get more spatial coverage. I find it hard to make sense of what's going on here unless there are actually <u>two</u> periods of study here: a 2017-2019 period for MLS and SAGE-III, and a supplemental 2020-2002 period just for SAGE III/ISS. If this is in fact the case, then the trustworthiness of the foregoing comparisons between the MLS and SAGE III/ISS reconstructions is seriously undermined. That it was the case that MLS data were not used in the latter period is strongly implied by the statement that there is no significant difference between the SAGE reconstruction results from the two periods.

This of course begs the question of why MLS data weren't used over the entire 2017-2022 period in the first place. If there is a technical reason for this, one isn't mentioned.

NOTE: The question of the two study periods is answered at **L131**. That this is referred to only obliquely here muddies the narrative unacceptably.

L115: In this paper, the term **anomaly** is used primarily for the difference between a derived value at a given geographical location from its long-term mean. However, here it refers instead to the difference between an (OLR) instantaneous value and its temporal mean. The latter usage is the more natural one as I see it, but in any case the dual usage here leads to some confusion in the text. In particular, back at **L36** the phrase **tropical stratospheric water vapor anomaly** is in the same sentence as the **summertime NH extratropical water vapor maximum.** Is this **maximum** a spatial or temporal one? It's not clear.

I mention this since the spatial features in the water vapor fields that are associated with the ASM and NAM could easily be referred to as local maxima without any ambiguity.

L150: Suggest replacing **along** with **in.** The word "along" is more suitable for horizontal spans or temporal stretches.

L151: The SAGE III/ISS CPTs used in LOC are from the MERRA-2 reanalysis while the trajectory reconstructions use ERA5 reanalysis temperatures. This necessarily introduces a complication into any comparison between the LOC and two Lagrangian experiments.

L172: Here is the first reference in the manuscript to the direction of crossing the tropopause. If it's not significant enough consideration to merit mention in the Introduction, why does it figure into the methodology?

Section 3

Figure 1 (from **L180):**

- The caption should identify the level of the water vapor fields in the panels. The reader should not need to refer to the text.
- The lower four panels are labeled as either **Observation anom** and **Reconstruction anom**. I would simply identify them as maps of the differences between the reconstructions and the satellite observations.
- The concluding clause in the caption is awkward English, the use of the word **strings** in particular.

¶1 (L181-L192) In keeping my preceding comment, the spatial features in the eight panels in Figure 1 are variously referred to as "enhanced", "elevated", "high", "maxima" and "anomalies". If, as I think is the case, all these terms essentially refer to the same thing, a single term should be used. Otherwise the text is ambigious.

L195: A simpler way to put this is that the reconstruction only faintly reproduces the observed NAM water vapor pattern.

L196: change **increase** to **elevation.** The word increase implies a temporal change.

Comment on the discussion of Figure 1: The spatial patterns displayed by the positive-valued reconstruction "anomalies" are a nice result, in particular their similarity to the maximum in the field of SAGE III observation in the vicinity ASM. While the text at **L188** does note the overall dry bias of the reconstructions for each satellite dataset, one might expect that, outside of the regions like the ASM where the water vapor stratospheric entry values of water vapor are controlled locally, the reconstruction anomaly field would be more or less flat. Instead the negative-valued anomalies more or less mimic the patterns in the observations, in much the same way as the positive-valued anomalies.

I don't know if there is a simple explanation for this or not, but it seems whatever is causing coherence of the positive-values reconstruction anomalies and the observed fields may also be contributing to a coherence of the negative-values anomalies.

Figure 2 (from **L197):**

- (caption) replace **concentrations** with **mixing ratio** throughout.
- (caption) replace **For each subplot, it shows** with **Each panel shows.**
- (caption) Properly, biases are not "between" fields/variable, but the difference of a particular estimate of a variable/field from a "reference" or "true" value of the same.
- (caption) The phrase **reconstructed values subtract observed values** needs to be corrected. Suggest **reconstructed minus observed**.
- The lower panels are labeled as either **Observation anom** which suggests that identifying them as maps of the differences between the reconstructions and the satellite data would be preferable. Thus they hange **the portions of TST are shown with upper right strings of c-d and g-h** to **TST fractions noted in panels c-d and g-h.**

L202-206: It is not surprising that the reconstructions do so poorly below the tropopause, as this where the remoistening by clouds will be a factor.

L221, L237: What is the basis of the statement that in the ASM region the tropopause layer is higher and thinner than elsewhere in the tropics? Likewise that the NAM is also thinner but lower?

L224: It's not clear what the phrase **vertical performance** means.

L225-L229: As written, neither of these sentences make sense. What do **one-third of the observed anomalies** and **over two-thirds of the observed values** mean? I assume these mean the magnitudes of the reconstructions that are being referred to, but the phraseology is poorly chosen.

L230:-L236: How is the "consistent behavior of the reconstruction in the ASM compared to the tropics" evidence for the water vapor above the ASM by something called **mechanisms-freeze-drying in the large-scale temperature.** This is its first mention in the paper, and if it is indeed the 'advection-condensation' paradigm of Liu et al. (2010), is there a reason it needs to be given a new name – and a grammatically unpleasant one at that? In any case, what exactly is this **mechanisms-freeze-drying in the large-scale temperature** process and how it is related to the "consistent behavior" above?