M2-SCREAM: A Stratospheric Composition Reanalysis of Aura MLS data with MERRA-2 transport

Krzysztof Wargan[1], Brad Weir[2], Gloria L Manney[3], Stephen E. Cohn[4], K. Emma Knowland[5], Pamela A Wales[6], and Nathaniel J Livesey[7]

[1]Science Systems and Applications, Inc.
[2]USRA / NASA Goddard
[3]Northwest Research Associates
[4]Data Assimilation Office
[5]Morgan State University
[6]University of Maryland, College Park
[7]Jet Propulsion Laboratory

November 24, 2022

Abstract

MERRA-2 Stratospheric Composition Reanalysis of Aura Microwave Limb Sounder (M2-SCREAM) is a new reanalysis of stratospheric ozone, water vapor, hydrogen chloride (HCl), nitric acid (HNO_3) and nitrous oxide (N_2O) between 2004 and the present (with a latency of several months). The assimilated fields are provided at a 50-km horizontal resolution and at a three-hourly frequency. M2-SCREAM assimilates version 4.2 Microwave Limb Sounder (MLS) profiles of the five constituents alongside total ozone column from the Ozone Monitoring Instrument. Dynamics and tropospheric water vapor are constrained by the MERRA-2 reanalysis. The assimilated species are in excellent agreement with the MLS observations, except for HNO_3 in polar night, where data are not assimilated. Comparisons against independent observations show that the reanalysis realistically captures the spatial and temporal variability of all the assimilated constituents. In particular, the standard deviations of the differences between M2-SCREAM and constituent mixing ratio data from The Atmospheric Chemistry Experiment Fourier Transform Spectrometer are much smaller than the standard deviations of the measured constituents. Evaluation of the reanalysis against aircraft data and balloon-borne frost point hygrometers indicates a faithful representation of small-scale structures in the assimilated water vapor, HNO_3 and ozone fields near the tropopause. Comparisons with independent observations and a process-based analysis of the consistency of the assimilated constituent fields with the MERRA-2 dynamics and with large-scale stratospheric processes demonstrate the utility of M2-SCREAM for scientific studies of chemical and transport variability on time scales ranging from hours to decades. Analysis uncertainties and guidelines for data usage are provided.

M2-SCREAM: A Stratospheric Composition Reanalysis of Aura MLS data with MERRA-2 transport

Krzysztof Wargan[1,2], Brad Weir[2,3,†], Gloria L. Manney[4,5], Stephen E. Cohn[2], K. Emma Knowland[2,3,†], Pamela A. Wales[2,3,†], and Nathaniel J. Livesey[6]

[1] Science Systems and Applications Inc., Lanham, Maryland, USA

[2] Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

[3] Universities Space Research Association, Columbia, Maryland, USA

[4] NorthWest Research Associates, Socorro, New Mexico, USA

[5] Department of Physics, New Mexico Institute of Mining and Technology, Socorro, New Mexico, USA

[6] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA

[†] Now Morgan State University, Baltimore, Maryland, USA

Corresponding author: Krzysztof Wargan (krzysztof.wargan-1@nasa.gov)

Key Points:

- New composition reanalysis of the stratosphere is introduced.
- Microwave Limb Sounderr (MLS) ozone, H_2O, HNO_3, HCl, and N_2O are assimilated for 2004–2021 and will be extended to the present.
- Asimilated species agree well with MLS and with independent data.

Abstract

MERRA-2 Stratospheric Composition Reanalysis of Aura Microwave Limb Sounder (M2-SCREAM) is a new reanalysis of stratospheric ozone, water vapor, hydrogen chloride (HCl), nitric acid (HNO_3) and nitrous oxide (N_2O) between 2004 and the present (with a latency of several months). The assimilated fields are provided at a 50-km horizontal resolution and at a three-hourly frequency. M2-SCREAM assimilates version 4.2 Microwave Limb Sounder (MLS) profiles of the five constituents alongside total ozone column from the Ozone Monitoring Instrument. Dynamics and tropospheric water vapor are constrained by the MERRA-2 reanalysis. The assimilated species are in excellent agreement with the MLS observations, except for HNO_3 in polar night, where data are not assimilated. Comparisons against independent observations show that the reanalysis realistically captures the spatial and temporal variability of all the assimilated constituents. In particular, the standard deviations of the differences between M2-SCREAM and constituent mixing ratio data from The Atmospheric Chemistry Experiment Fourier Transform Spectrometer are much smaller than the standard deviations of the measured constituents. Evaluation of the reanalysis against aircraft data and balloon-borne frost point hygrometers indicates a faithful representation of small-scale structures in the assimilated water vapor, HNO_3 and ozone fields near the tropopause. Comparisons with independent observations and a process-based analysis of the consistency of the assimilated constituent fields with the MERRA-2 dynamics and with large-scale stratospheric processes demonstrate the utility of M2-SCREAM for scientific studies of chemical and transport variability on time scales ranging from hours to decades. Analysis uncertainties and guidelines for data usage are provided.

Plain Language Summary

Earth's stratosphere contains a number of trace gases of various origins, chemical properties and lifetimes. In addition to their importance for stratospheric chemistry, including those affecting the ozone layer, and for the planet's radiative budget, the complex geographical and vertical distributions of atmospheric constituents provide invaluable information about stratospheric dynamics and transport in a changing climate. This paper introduces and evaluates a new high-resolution composition data set produced at NASA's Global Modeling and Assimilation Office. Named MERRA-2 Stratospheric Composition Reanalysis of Aura Microwave Limb Sounder (M2-SCREAM), this stratosphere-focused product consists of assimilated global three-dimensional ozone, water vapor, hydrogen chloride, nitric acid, and nitrous oxide fields, all of which are of primary importance to stratospheric chemistry and transport studies. M2-SCREAM uses high quality data from the Microwave Limb Sounder instrument (2004-present) combined with meteorological information from NASA's MERRA-2 reanalysis. Comparisons with independent observations and a process-based analysis of the consistency of the assimilated constituent fields with the MERRA-2 dynamics and with large-scale stratospheric processes demonstrate the utility of M2-SCREAM for scientific studies of chemical and transport variability on time scales ranging from hours to decades.

1 Introduction

The past decade has witnessed growing scientific interest in the quickly developing field of chemical retrospective analyses (*reanalyses*): multiyear records of assimilated observations of atmospheric constituent gases (Errera et al., 2019; Flemming et al., 2017; Hollingsworth et al., 2008; Innes et al., 2013; 2019; Miyazaki et al., 2015, 2020; van der A et al., 2015). While major multidecadal meteorological reanalyses routinely include ozone and water vapor, their data

assimilation systems do not incorporate full chemistry models, and their treatment of stratospheric water vapor is often simplified to the point of rendering that product unsuitable for science (e.g., Davis et al., 2017; SPARC 2021), with exception of the recent European Centre for Medium-Range Weather Forecasts reanalysis, ERA5 (Hersbach et al., 2020), which has been shown to have a scientifically useful stratospheric water vapor product (Wang et al., 2020; SPARC 2021). In contrast, chemical (or *composition*) reanalyses typically use "full" chemistry models (i.e., ones that explicitly model chemical reactions, rather than parametrizing them) with transport driven by assimilated winds and temperature fields from existing meteorological reanalyses. Composition reanalyses typically include a host of atmospheric constituent fields beyond ozone and water vapor. Most chemical reanalyses focus on tropospheric composition and air quality applications, although some of them, e.g. the Copernicus Atmospheric Monitoring Service (Innes et al., 2019), also provide realistic representations of stratospheric ozone. To our knowledge, the only stratosphere-focused global reanalyses to date are the Belgian Assimilation System for Chemical Observations (BASCOE) Reanalysis of Aura MLS (Microwave Limb Sounder) versions 1 and 2 (BRAM and BRAM2; Errera et al., 2019).

This paper introduces a new chemical reanalysis of stratospheric constituents developed at NASA's Global Modeling and Assimilation Office (GMAO). Named "M2-SCREAM", for Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2; Gelaro et al., 2017) Stratospheric Composition Reanalysis of Aura MLS, this product consists of assimilated global three-dimensional fields of stratospheric ozone, water vapor, hydrogen chloride (HCl), nitric acid (HNO_3), and nitrous oxide (N_2O) mixing ratios, while the tropospheric water vapor and meteorological fields are constrained by MERRA-2. While the model used to produce M2-SCREAM simulates a number of other chemical species, only the five assimilated constituents are evaluated and released to the scientific community at present. The reanalysis horizontal resolution matches the MERRA-2 reanalysis of 0.625° longitude by 0.5° latitude on 72 levels and covers the period of MLS observations from September 2004 to 2022 (March 2022 at the time of writing). The assimilated instantaneous fields are produced at a three-hourly frequency. M2-SCREAM assimilates version 4.2 MLS profiles of the five constituents alongside total column ozone from the Ozone Monitoring Instrument (OMI: Levelt et al, 2006; 2018), using the recently developed Constituent Data Assimilation System (CoDAS: Wargan et al., 2020a,b; Weir et al., 2021). M2-SCREAM provides an accurate and dynamically consistent high-resolution data record of the five constituents, all of which are of primary importance to stratospheric chemistry and transport studies. As an illustration, **Fig. 1** compares tropical stratospheric water vapor from the MERRA-2 and M2-SCREAM reanalyses. Dynamically driven variability in the tropical tape recorder signal (e.g., Davis et al., 2017), especially above 70 hPa, is readily apparent in M2-SCREAM, while only hinted at in MERRA-2. Comparisons with independent observations (**Section 6**) reveal that the M2-SCREAM stratospheric water vapor exhibits realistic variability.

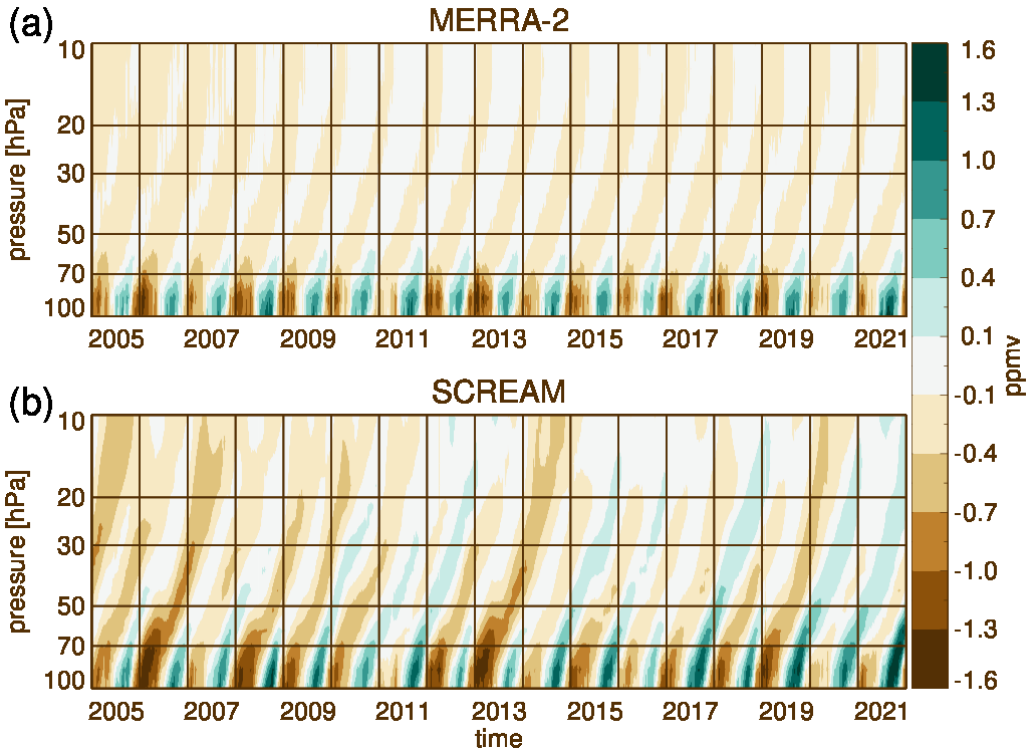

Figure 1: *Water vapor anomalies from MERRA-2 (a) and M2-SCREAM (b) averaged between 15°S and 15°N and plotted as a function of time and pressure. A five-day running mean was applied to the data at every pressure level. The anomalies are computed at every level separately by removing the time average.*

Building on the theory of data assimilation, this study demonstrates that a (composition) reanalysis product is fundamentally data-driven whereby the information content in the assimilated constituent fields is drawn from observations and these fields constitute a near-optimal estimate of the true atmospheric composition given the assimilated data and their uncertainties (**Section 2**). The utility of composition reanalyses for scientific studies, therefore, lies in the global, high-frequency representation of these constituent fields consistent with the underlying dynamics and chemistry of the real atmosphere.

As our focus is on the scientific utility of this reanalysis product, we have taken several steps intended to guide the users. Specifically, we derive and provide monthly sets of uncertainties for each of the assimilated species and flag the areas where the confidence in the assimilation output is low. Furthermore, we evaluate and discuss the suitability of the reanalysis fields for scientific applications by focusing not only on comparisons with independent observations but also on process-based analysis.

The remainder of the sections are organized as follows. Description of the GEOS Constituent Data Assimilation System (CoDAS) is provided in **Section 3**. Data sources assimilated in M2-SCREAM and those used for evaluation of the reanalysis are described in **Section 4**. **Section 5** discusses the internal consistency of the reanalysis. **Section 6** describes and discusses validation of the reanalysis against independent observations as well as process-based evaluation. **Section 7** discusses confidence in the M2-SCREAM output and contains some recommendations for users.

A summary of this study is given in **Section 8**. Three appendices discuss (A) the uncertainty estimation, (B) a technical correction applied to the reanalysis output, and (C) the contents of the M2-SCREAM output provided to the users.

2 Theoretical motivation for constituent data assimilation

This section provides a high-level overview of the theory of constituent data assimilation with emphasis on those of its aspects that motivate the methodological choices of M2-SCREAM and their scientific application. In what follows the term *data assimilation* is taken to mean constituent data assimilation. We do not consider meteorological assimilation in this study other than in the context of driving the model by assimilated winds and temperatures. Furthermore, by *model* we mean a specified dynamics general circulation model (GCM) forced by assimilated meteorology and integrated with a chemistry module. Throughout the paper we will use the terms *analysis* and *analyzed* for constituent fields obtained through the procedure described in this section; we will use the words *assimilation* and *assimilated* for the final product computed by CoDAS using the Incremental Analysis Update (IAU: Bloom et al., 1996) described in Section 3. Only the assimilated fields are archived and distributed to users.

Data assimilation seeks to estimate the probability distribution of global gridded constituent fields x_n at time t_n given a set of observations $y_{1:n} = \{y_1, y_2, \ldots, y_n\}$ of functions H_1, H_2,..., H_n of the states $x_1, x_2, \ldots, x_n$. A data assimilation system (DAS) does this by tracking and updating the probability density function (pdf) $p(x_n \mid y_{1:n})$ of the state x_n conditioned on all current and past observations $y_{1:n}$ as new data arrive. Using Bayes' theorem,

$$p(x_n \mid y_{1:n}) \propto p(y_n \mid x_n)\, p(x_n \mid y_{1:n-1}), \qquad (1)$$

where $p(y_n \mid x_n)$ is the probability density of observing y_n given x_n, the prior is the pdf $p(x_n \mid y_{1:n-1})$ of x_n given *all previous observations*, and the posterior is the pdf $p(x_n \mid y_{1:n})$ after conditioning on the new observations y_n. Sequential filtering methods further decompose the prior into its forecast and initialization components using the assumed Markov property of the state to give the recursion relation

$$p(x_n \mid y_{1:n}) \propto p(y_n \mid x_n) \int p(x_n \mid x_{n-1})\, p(x_{n-1} \mid y_{1:n-1}) dx_{n-1}$$

for the posterior $p(x_n \mid y_{1:n})$ in terms of its value at the previous time t_{n-1}, $p(x_{n-1} \mid y_{1:n-1})$. See Jazwinski (1970), Theorem 6.4 for more details.

Assuming all the distributions are Gaussian, we can define the cost function

$$J_n(x) = (y_n - H_n x)^T R^{-1} (y_n - H_n x) + (x - x_n^b)^T B^{-1} (x - x_n^b), \quad (2)$$

such that $p(x_n \mid y_{1:n}) \propto exp[-\frac{1}{2} J_n(x_n)]$, where R is the error covariance of $p(y_n \mid x_n)$, B the error covariance of $p(x_n \mid x_{n-1})$, and $x_n^b = M(x_{n-1}^a)$ is the *background (*or *forecast*) *state* from the mean (or mode) x_{n-1}^a of the previous pdf $p(x_{n-1} \mid y_{1:n-1})$ propagated forward by the model M. For notational convenience, from here on we will drop the time index subscript n when not

needed and treat the observation operator as its Jacobian matrix at x_n^a. The *analysis*, x^a, is defined as the state that minimizes the cost function $J(x)$. It is thus the mean/mode of the posterior pdf (Nichols et al., 2010; Lahoz and Schneider 2014) and usually serves as the initial condition for the background x^b at the next timestep. However, in M2-SCREAM the initial condition is, instead, the assimilated state obtained from background and analysis through the IAU procedure described in the next section. In either case it follows that x^a is the maximum likelihood estimate of the true state of the constituent fields given available data valid at steps t_n, t_{n-1}, t_{n-2}, …. By definition, it satisfies the equation

$$x^a = x^b + BH^T(HBH^T + R)^{-1}(y - Hx^b), \quad (3)$$

On digital computers, Equation 3 can be solved for by storing x^b and x^a as vectors of constituent mixing ratios defined on the three-dimensional model grid and y as a vector whose length is the number of observations. Since at half-degree resolution over the entire globe the matrix B would require over 2 petabytes of memory to store outright, we solve for x^a using a conjugate gradient method that needs only to multiply by R and B. The sequential filtering approach has the added advantage that it does not need to store previous values of the state and observations. For further discussion, including typical assumptions and simplifications made in a DAS see, e.g. Jazwinski (1970), Cohn (1997), Nichols et al. (2010), Weir et al. (2013), Lahoz and Schneider (2014), and Reich (2019). The remainder of this section explores several important points that follow from the formulation outlined above.

It follows from Equation (1) that data assimilation is fundamentally an observation-driven methodology. The role of the model is limited to the time propagation of the cumulative information from past observations in accordance with the governing equations of motion. If spatio-temporal data coverage is sufficient (i.e., if the state is *completely observable* with respect to the given observations; see Jazwinski 1970 pp. 231–234) then the initial condition for the model integration over each assimilation cycle (six-hourly in our case) is purely the result of previously assimilated observations up to unavoidable uncertainties, and consequently, approximates the *true* constituent fields in contrast to a *possible realization* of that state as is the case in pure model simulations. In fact, assimilation is useful precisely because, and to the degree that, its output's information content is derived from observations. We emphasize this point because it is sometimes stated that assimilated fields are a "blend" of data and model output, which leads to concerns about what assimilated products represent. **Figure 2** illustrates the concept of data assimilation as driven by observations. Shown are the global ozone and HNO_3 analyzed fields interpolated from the M2-SCREAM native level output onto the 500-K potential temperature surface. The circles show the assimilated MLS observations within the six-hourly time window around the analysis time. It is readily seen that the ozone observations agree very well with the assimilated field (**Fig. 2a**), as expected from relatively small observation uncertainties. Precision and accuracy of the constituent field away from the current observations is measured by observation minus "forecast" (O-F) residuals and is shown to be high given the assumed uncertainties (**Section 5**). The considerable advantage of having a high-resolution global gridded field constructed from past observations propagated with the model is evident from the high level of detail seen in the figure, including complex dynamically driven features.

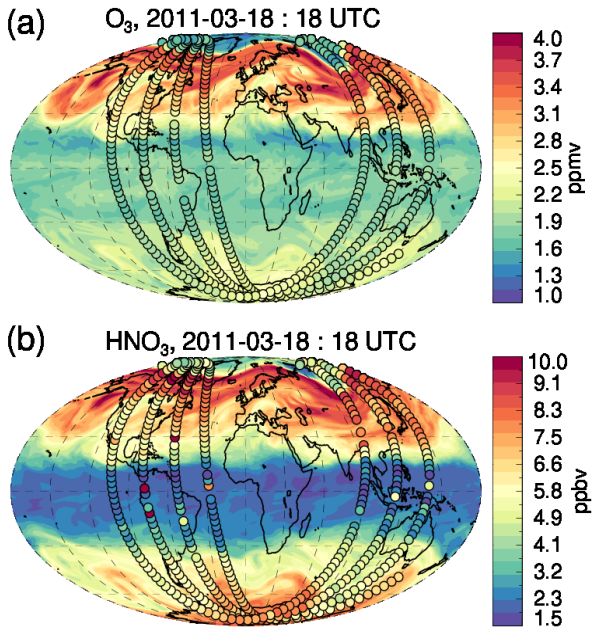

Figure 2: *Assimilated ozone (a) and HNO_3 (b) on 18 March 2011, 18 UTC, interpolated to the 500-K potential temperature surface. MLS observations are overplotted using the same color scheme as circles.*

Data assimilation is a probabilistic methodology. Observation and background (prior) uncertainties play a critical role. In **Fig. 2(b)**, the agreement between the observations and the assimilated HNO_3 field is less close than for ozone: several observed values depart from the assimilated field and appear inconsistent with the average concentrations in the adjacent regions, particularly in the tropics. This is expected from the relatively high observation uncertainties for MLS HNO_3 (Livesey et al., 2020), that is, a large spread in $p(y_n \mid x_n)$. Noise in the data is effectively filtered out as a result. The posterior pdf, $p(x_n \mid y_{1:n})$, can be estimated from the internal statistics of the DAS under suitable assumptions (Desroziers et al., 2005). Using this method, we calculate monthly standard deviations of the posterior pdfs as well as overall estimates of uncertainties for M2-SCREAM (**Appendix A**). These diagnostics are provided to the users in the form of monthly estimates.

In regions of the atmosphere where observations are not assimilated (for example in the middle and lower troposphere unobserved by MLS or during infrequent prolonged data outages) the output fields produced by the DAS can no longer be thought of as data-driven and are more akin to the results of a chemistry model simulation forced by assimilated meteorology. Even in those situations it is expected that observational information content is propagated to some extent into unobserved regions by model transport. Our previous work (Wargan et al., 2020a; 2020b) indicates that assimilation results are valid at all latitudes including the poles despite the MLS coverage boundaries at 82°. Additionally, we demonstrate in **Section 6.1** good qualitative agreement with independent data several kilometers below the tropopause. However, except in the cases delineated above we do not evaluate M2-SCREAM over unobserved regions and do not recommend its use for scientific studies far outside of areas covered by MLS observations. We provide monthly

gridded files that contain uncertainty estimates and that flag the regions of the atmosphere not covered by observations. We also provide some additional recommendations in **Section 7**.

3 GEOS Constituent Data Assimilation System (CoDAS)

The GEOS Constituent Data Assimilation System (CoDAS) is an extension of the GEOS Atmospheric Data Assimilation System (ADAS; Todling and El Akkraoui, 2018) and is capable of assimilating observations of any trace gas simulated by any of several GEOS-compatible chemistry modules. Originally derived from the ozone assimilation code of MERRA-2 (Wargan et al., 2017), CoDAS generalizes those capabilities to arbitrary collections of trace gases with generic averaging kernel and in situ observation operators capable of ingesting nearly every known space-based trace gas retrieval. Current applications include stratospheric (Wargan et al. 2021a,b) and carbon (Weir et al., 2021) constituent assimilation systems with ongoing research incorporating reactive tropospheric gases that determine air quality from a full tropospheric and stratospheric chemistry module. Through its ADAS lineage, CoDAS inherits a suite of assimilation methodologies including three- and four-dimensional variational, ensemble, and hybrid methods for estimating the posterior pdf (Equation 1). For simplicity, we use three-dimensional variational (3DVar) formulation and Gridpoint Statistical Interpolation (GSI; Wu et al., 2002) to discretize the atmospheric state x onto a regular, horizontal grid and η-level vertical coordinates, resulting in the cost function in Equation (1).

The theoretical success of data assimilation owes to two fundamental factors: 1) the power of recycling previous data into each new background as described in **Section 2**, a result that follows from Bayes' theorem of the 1700s, and 2) that data differences from a prior background tend to have much simpler statistics than the data values themselves, a result known since at least the 1950s (Bergthórsson and Döös, 1955). The latter follows from the fact that differencing data and a model can produce a random variable whose statistics are far more smooth in space and regular in time than the data themselves, which can have complex and chaotic behavior. Nevertheless, the practical success of data assimilation relies heavily upon the modeling of the error statistics, namely the covariance B of the background errors and R of the observation errors, which are assumed here to be either additive or multiplicative for simplicity. Since these matrices are far too large to store in memory, we represent them as transformation operators. Simplifying assumptions about error statistics then translate to simpler, and faster, representations as transformations. We use a background error covariance B whose variances and horizontal correlation lengths are constant horizontally. With the exception of ozone, these values are constant vertically as well. Vertical background error correlation is estimated from the vertical correlation length of the modeled values and thus varies in space and time. CoDAS supports log-normal, i.e., multiplicative, error distributions which are used for stratospheric water vapor, HCl, and HNO3 (see **Table 1**). Using multiplicative errors introduces some "flow-dependence" in the error statistics since they are proportional to the background values. We use an observation error covariance R that is the reported retrieval error (precision and accuracy combined) multiplied by a scaling factor that varies by level. These numbers are all tuned using repeated applications of the Desroziers et al. (2005) diagnostics.

Data are assimilated over 6-hour windows centered on "off-synoptic" times (e.g. 21Z) and are cycled back into the model by adding the average analysis minus background increment to simulated tracer values over the window length. This is similar to the IAU used for the

meteorological variables of MERRA-2. While there are advantages to 4DVar and ensemble capabilities in the ADAS (e.g., Skachko et al. 2016), here we use 3DVar and note that refinement of its window length, e.g., to an hour, can impart any desired "flow-dependence" to the increments; furthermore, with the computational costs of simulating full chemistry, it is not currently within the scope of this product to use 4DVar. Research is underway to evolve background error standard deviations rather than the simplified approach of taking them to be proportional to the state as done here (Ménard et al. 2021; Gilpin et al. 2022).

The model used in the present configuration of CoDAS is a version of GEOS GCM (Icarus-3_2_p9) integrated with the stratospheric chemistry module, StratChem (Nielsen et al., 2017 and references therein). This model configuration is the same as that used by Wargan et al. (2020b). The meteorology in the GCM is constrained by the MERRA-2 reanalysis output of temperature, surface pressure, tropospheric (but not stratospheric) water vapor, and winds (GMAO 2015) via the *replay* methodology unique to GEOS and described in detail by Orbe et al. (2017). The dynamical and temperature fields in M2-SCREAM are, therefore, very similar to those in MERRA-2 (not shown). Differences that arise from the upgrades to the GCM used to produce M2-SCREAM since MERRA-2 and from the radiative impacts of assimilated ozone and water vapor are small in the stratosphere (not shown). Below the tropopause, water vapor is replayed to the MERRA-2 analysis, and thus constrained by MERRA-2 within the troposphere but not above it.

The StratChem module is a family chemistry scheme that simulates 125 gas-phase and 35 photolysis reactions important in the middle atmosphere. These include gas-phase and heterogeneous chemistry of the chlorine, bromine and nitrogen families. The reaction rates follow Burkholder et al. (2015). There are 51 transported and 18 inferred species. Polar stratospheric clouds (PSCs) are parameterized following Considine et al. (2000). We refer the reader to Wargan et al. (2020b) for further details of the model setup.

For technical reasons related to the current implementation of the PSC scheme in StratChem discussed in further detail in Section 5, HNO_3 is not assimilated in regions where model-generated PSCs are present. We recommend that lower-stratospheric HNO_3 from M2-SCREAM during polar night be avoided in scientific studies. Further recommendations regarding HNO_3 are given in **Section 8**. This work also used an MLS observation operator with a bug that mistook model layer centers for layer edges which slightly offset the placement of the observations. Given the vertical resolution of the model compared to MLS profiles, the impact was small and has been successfully corrected in post-processing (**Appendix B**).

4 Data

4.1 Assimilated observations

MLS on NASA's Aura satellite (Waters et al., 2006) is a microwave limb sensor that measures thermal emission of the atmosphere in a range of spectral bands allowing retrieval of the profile information of many atmospheric constituents from the upper troposphere through the mesosphere. The instrument makes day and night measurements between 82°S and 82°N along 15 orbits per day. At the time of writing, the MLS mission covers 17 years of nearly uninterrupted measurements (**Section 7**) since late 2004. The MLS observations have provided and continue to provide invaluable information on stratospheric composition, its changes and variability.

M2-SCREAM assimilates version 4.2 ozone, water vapor, HCl, HNO_3 and N_2O data from MLS (Livesey et al., 2020). **Table 1,** similar to that provided in Wargan et al. (2020b), specifies the vertical extent and resolution of the assimilated profiles. As many details of the MLS data treatment are the same as in Wargan et al. (2020b), here we summarize it briefly and focus on the most important facts and the long-term behavior of these data. The observation uncertainties for ozone are as those used in MERRA-2 (Wargan et al., 2017) and those for the other species were tuned using the method described in Desroziers et al. (2005). Recommended quality screening (Livesey et al., 2020) is applied to all MLS observations prior to assimilation.

Table 1. *Treatment of MLS observations and the model background and observation errors. Values in column "B" for all species except N_2O are factors used to scale the background state, x^b, to obtain the background uncertainty standard deviation. For N_2O we use a constant background uncertainty of 8ppbv. The column "R" lists the ranges of scaling factors applied to the reported MLS uncertainties.*

Constituent	Vertical range	Vertical resolution in lower to middle stratosphere	B	R	Remarks
Ozone	216 – 0.1 hPa	2.5 – 3 km	5%	100%	
HCl	100 – 0.32 hPa	3 km	10%	30–100%	
N_2O	68 – 0.46 hPa	4 – 6 km	8 ppb	23–75%	Significant drift exists (Livesey et al., 2021)
Water vapor	261 – 0.01 hPa (model top)	1.5 – 3 km	6%	37–760%	Drift and wet bias in the stratosphere (Livesey et al., 2021)
HNO_3	216– 1.5 hPa	4 – 4.5 km	10%	30–70%	HNO_3 is not assimilated in regions where model-generated PSCs are present due to technical reasons related to the current implementation of the PSC scheme in GEOS.

The MLS version 4.2 N_2O and water vapor, both retrieved from the 190 GHz band, suffer from known altitude dependent drifts in the period after 2010 (Livesey et al., 2021). Evaluated against other satellite data, the drift in water vapor is positive and ranges from 2–3% per decade in much of the stratosphere to 5% and more at 50 hPa. Additionally, the version 4.2 water vapor is found to be biased high with respect to the latest version (version 5) by about 10% (Livesey et al., 2022).

The post-2010 drift in N_2O is negative. It is confined to the lower stratosphere (LS) but it is larger than the drift in water vapor, up to 15%. The other three assimilated species, ozone, HCl, and HNO_3, have been found to be stable for the duration of the MLS mission (Livesey et al., 2020).

M2-SCREAM was already well in production when version 5 of MLS data became available (Livesey et al., 2022). The water vapor drift has been significantly reduced in version 5, but some bias in N_2O remains (Livesey et al., 2021). Comparisons with independent data suggest that the moderately high bias in stratospheric water vapor seen in version 4.2 has been eliminated in version 5. We compare M2-SCREAM with MLS version 5 observations in **Section 5**.

As described in Wargan et al. (2020a,b), M2-SCREAM also assimilates total ozone observations from the OMI sensor (Levelt et al., 2006; 2018). These observations provide an additional constraint on the analyzed ozone, as demonstrated in the previous GEOS DAS systems described by Wargan et al. (2015; 2017; 2020a,b) and Ziemke et al. (2015).

4.2 Independent observations

This subsection describes the independent (that is, not assimilated) data used to evaluate M2-SCREAM. The Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS) on Canada's SCISAT-1 satellite (Bernath et al., 2005; Bernath 2017) is a solar occultation sensor that provides sunrise and sunset measurements in multiple infrared channels, allowing accurate retrievals of many trace gases including ozone, water vapor, HCl, HNO_3, and N_2O. The instrument provides 30 high vertical resolution profiles per day. The coverage varies significantly with season, with most measurements taken at mid- and high latitudes. However, annually aggregated ACE-FTS data provide near-global coverage that is sufficient for reanalysis evaluation. We use version 4.1 ACE-FTS retrievals (Boone et al., 2020) with additional screening applied using sets of quality flags provided by the instrument team (Sheese et al., 2015; Sheese et al., 2020). Version 3.5 ACE-FTS had an average estimated dry bias of ~5% in the middle-to-upper stratosphere and the lower mesosphere (Sheese et al. 2017); their Fig 3 suggests a bias close to 10% near the stratopause. Our results in **Section 6.3** suggest that some of this bias persists in version 4.1.

Another solar occultation sensor, the Stratospheric Aerosol and Gas Experiment III instrument, was installed on the International Space Station (SAGE III/ISS) in February 2017. The SAGE III/ISS is in a low earth orbit with an inclination angle of 51.6° and a measurement range from about 70°S to 70°N (Wang et al., 2020). Ozone, water vapor, aerosols and other trace gases are retrieved from solar occultation measurements on a 1-km grid for both sunrise and sunset. In this study, version 5.2 of the SAGE III/ISS water vapor product is used, interpolated to a 0.5-km vertical grid (Davis et al., 2021). SAGE III/ISS water vapor profiles are filtered according to Davis et al. (2021) to remove cloud interference in the troposphere and retrieval anomalies in the upper stratosphere and mesosphere. Unlike previous versions of SAGE data, version 5.2 water vapor profiles are not smoothed vertically, resulting in substantial noise within an individual profile.

We use observations of water vapor, ozone, and HNO_3 retrieved from the measurements made by the Gimballed Limb Observer for Radiance Imaging of the Atmosphere (GLORIA) instrument flown on the German High Altitude and Long Range Research Aircraft (HALO) during the joint Polar Stratosphere in a Changing Climate, Gravity Wave Life Cycle Experiment, and Seasonality of Air mass transport and origin in the Lowermost Stratosphere using the HALO Aircraft

campaigns (hereafter, PGS, Johansson et al., 2018). The GLORIA sensor (Friedl-Vallon et al., 2014) is an imaging Fourier transform spectrometer measuring thermal emissions of the atmosphere in the infrared. The data used in this study are made possible by high spectral resolution measurements taken every 13 seconds at the nominal vertical resolution of 250 m. The actual vertical resolution ranges from about 0.5 km to 1 km, finer than that of the assimilated fields (~1.1 km in the upper troposphere and lower stratosphere, *UTLS*). The PGS campaign consisted of 15 HALO flights in the North Atlantic region between 21 December 2015 and 18 March 2016. The timing of the campaign fortuitously coincided with one of the coldest Arctic polar vortex seasons that featured significant denitrification, dehydration, and ozone depletion (Manney and Lawrence 2016; Khorsawi et al., 2017), making it a good test case for the reanalysis.

Frost Point Hygrometers (FPH) are balloon-borne instruments that measure atmospheric water vapor content by maintaining a stable thin layer of ice on a temperature-controlled mirror monitored by an infrared LED beam. The moisture content is derived from the temperature of the mirror at thermodynamic equilibrium (Hurst et al., 2014). This technology allows accurate measurements of specific humidity in the stratosphere well above the altitudes where humidity measurements from radiosondes are suitable for scientific use. We show comparisons between the M2-SCREAM water vapor and National Oceanic and Atmospheric Administration's FPHs stations that provide multidecadal record: Lauder, New Zealand (169.68°E, 45.04°S), Hilo, Hawaii (155.05°W, 19.72°N), and Boulder, CO, USA (105.2°W, 39.95°N). 181, 120, and 277 profiles of specific humidity are available from Lauder, Hilo, and Boulder, respectively, between 2005 and 2020. The vertical resolution of the FPH measurements is 5-10 m. Here, we use data sets averaged in 250-m layers and further map them onto a 1-km vertical grid by averaging within each layer.

All comparisons of M2-SCREAM against MLS and the independent data sets are presented in terms of mixing ratios rather than as relative to the observed values. While relative differences expressed in percent can be useful in other contexts, we find that they are often misleading when used in discussions of constituents. As trace gas variability spans more than one order of magnitude, areas of very low concentrations are dominated by random noise, producing exceedingly large relative differences when the actual differences simply reflect the instrument precision.

5 Internal statistics and agreement with MLS observations

For every observation CoDAS calculates the corresponding O-F, i.e., $y - H(x^b)$, that is the discrepancy between the observed constituent mixing ratio and the background value from a six-hourly integration of the model initialized with the result of the previous assimilation cycle. Although inconsistent with the superscript "b" for background, we use here the traditional "F" for "forecast" in the term "O-F". At the end of a given cycle the system computes the observation minus analysis (O-A) departure, $y - H(x^a)$. On average, the latter are expected to be closer to zero than the corresponding O-Fs. The mean and standard deviation of the O-Fs represent the combined uncertainty of the background field and the uncertainty of the observations (Desroziers et al., 2005, their equation 1). The success of assimilation critically depends on the ability of the model to propagate information from observations forward in time over the length of the assimilation window, that is, the model error accumulated over the integration period should be small. Therefore, O-Fs represent a valuable diagnostic of the performance of the DAS.

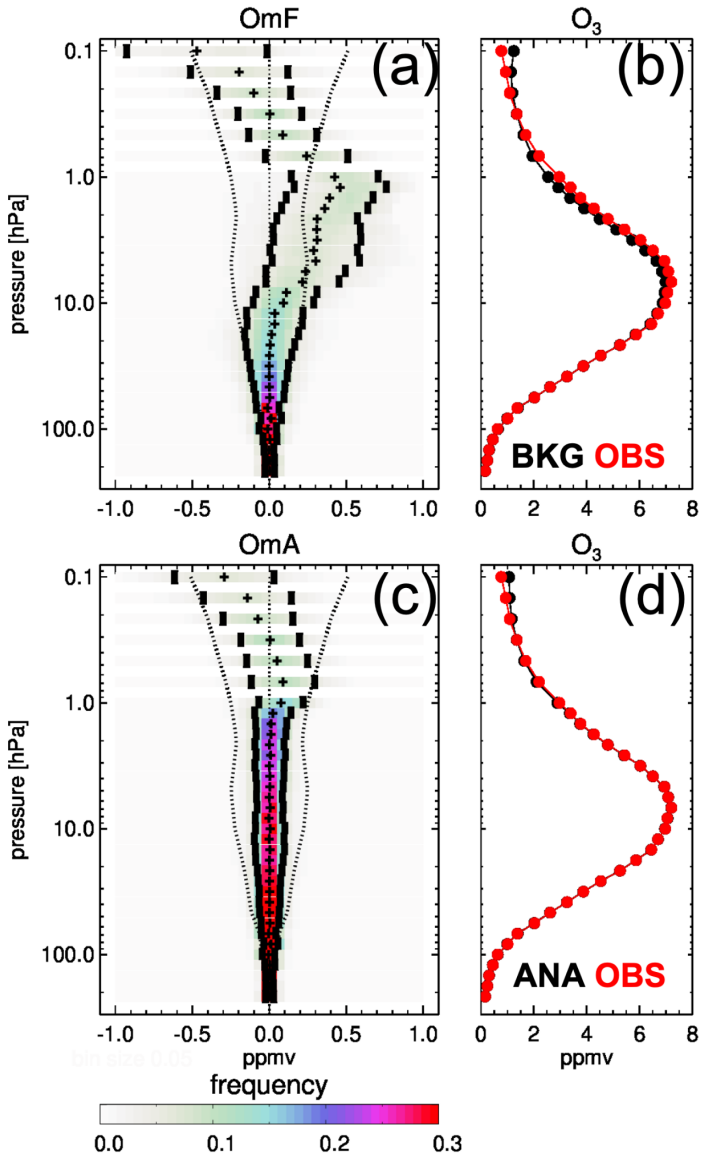

Figure 3: *Global M2-SCREAM O-F (a,b) and O-A (c,d) statistics for MLS version 4.2 ozone. Panels (a) and (c) show the O-F and O-A statistics: the mean (plus signs), standard deviations around the mean (short vertical bars), probability density functions (colors) at the MLS levels from 216 to 0.1 hPa. The dotted lines are plus/minus MLS uncertainty estimates. Panels (b) and (d) show mean background, "BKG" (b) and analysis, "ANA" (d) profiles (black) and mean MLS observed profiles (red) at the same pressure levels.*

Figure 3 illustrates the O-F and O-A statistics for M2-SCREAM ozone calculated globally in January 2005. The statistics do not vary significantly from month to month. O-F and O-A pdfs are shown along with their mean, median, and standard deviation around the mean at each MLS pressure level. Additionally, the figure plots MLS uncertainty estimates derived from the relevant table in Livesey et al. (2020) as the square root of the sum of squares of the reported precision and accuracy, the latter multiplied by 0.5. For pressures greater than 10 hPa, the O-Fs show very little bias and approximately Gaussian distributions, with standard deviations comparable to the

observational uncertainty estimates. There is, however, a pressure dependent bias in the upper stratosphere and the mesosphere. Also evident is an increase in the O-F spread with altitude. In contrast, the O-As exhibit a much smaller bias (near-zero throughout the stratosphere, up to 1 hPa) and standard deviations are within the observational uncertainties. A comparison of the two panels of **Fig. 3** reveals that while assimilation updates bring the assimilated ozone mixing ratio close to the observed values, that information is not fully retained during the six-hourly model integration in the upper atmosphere. This behavior arises because characteristic timescales for stratospheric ozone chemistry rapidly decrease with altitude and become on the order of one hour or less at the higher levels (Brasseur and Solomon, 2005). Short constituent lifetimes pose a challenge to data assimilation, which relies on a cumulative effect of observations as noted in **Section 2**. For this reason, the BRAM2 reanalysis does not assimilate ozone at pressures smaller than 4 hPa (Skachko et al., 2016; Errera et al., 2019). We have made a choice to assimilate MLS ozone at pressures of 0.1 hPa or greater in M2-SCREAM. However, users should treat upper-stratospheric ozone from the reanalysis with caution. Unlike ozone, the lifetimes of water vapor, HCl, HNO_3, and N_2O are sufficiently long for the observational information to propagate and accumulate. The O-Fs for all four constituents exhibit negligible bias and their standard deviations are within the MLS uncertainty estimates (**Figure S1**). Note that the MLS uncertainties shown in these figures are global estimates, not the observation-by-observation estimates provided with the data, and therefore are not necessarily strictly larger than the observational uncertainties even though the O-Fs include a contribution from the background as well as from observation uncertainties. Furthermore, the uncertainty estimates derived using Desroziers' formula and used in CoDAS are typically 50 to 70% of those in the MLS data files. As expected, the O-As (not shown) are reduced compared to the O-Fs, indicative of the internal consistency of the data assimilation system. Compared to the mean observed values, the MLS uncertainties (and the O-F standard deviations) for HCl, HNO_3, and N_2O are substantially larger than those for ozone and H_2O, suggesting more "noisy" observations. That is consistent with our discussion of **Fig. 2(b).** Overall, the O-F results, except for ozone in the upper stratosphere and above, demonstrate excellent performance of the reanalysis in terms of the agreement with the assimilated data and the ability of CoDAS to retain and propagate information from observations forward in time.

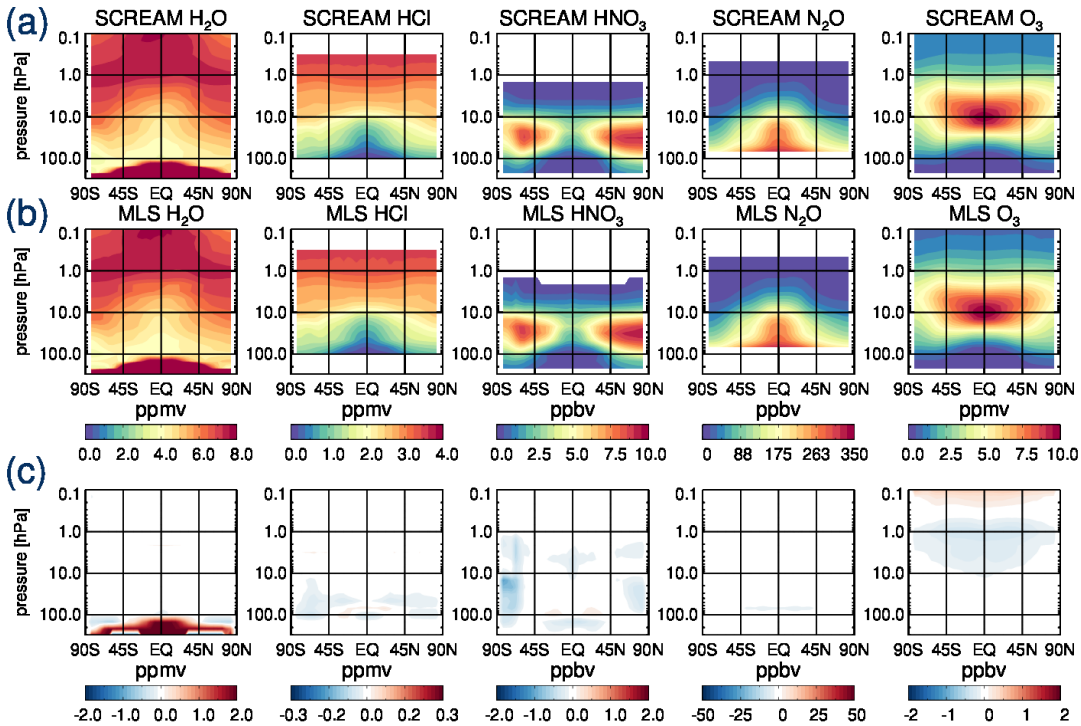

Figure 4: *Zonal mean mixing ratios of the assimilated constituents calculated from M2-SCREAM sampled at the MLS observation locations (a) and MLS version 4.2 data (b). The M2-SCREAM minus MLS differences are shown in (c). All quality-screened MLS data between 2005 and 2021 are used.*

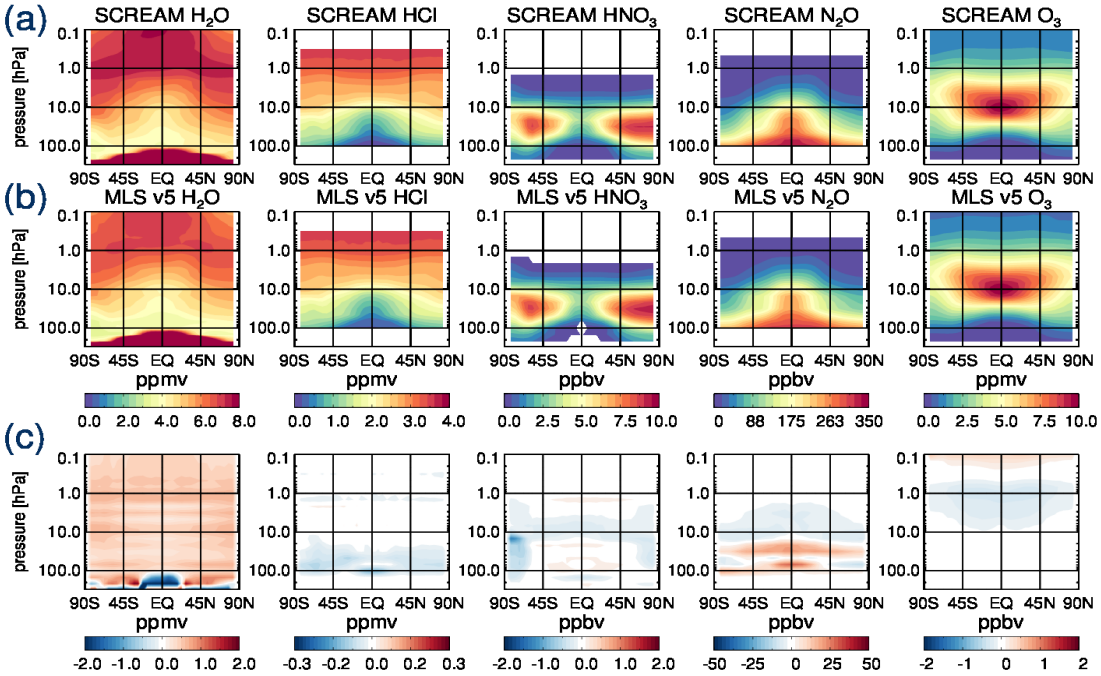

Figure 5: *As in Figure 4 but with MLS version 5.0 data.*

Next, we compare the zonal means of the M2-SCREAM constituents with observations from MLS version 4.2 (assimilated, **Fig. 4**) and version 5 (**Fig. 5**). Both comparisons use quality-filtered MLS

observations between 2005 and 2021 and M2-SCREAM profiles subsampled at the observation locations within three hours of observation times. As expected, the differences between M2-SCREAM and MLS version 4.2 are small overall except for water vapor in the troposphere and HNO_3 in the polar regions. Water vapor below the tropopause is constrained by MERRA-2 through replay (**Section 3**), thus less constrained by MLS. In addition, discrepancies among different data sources are typically large in the upper troposphere because of very sharp vertical gradients of specific humidity there. As explained in Section 3, CoDAS does not assimilate HNO_3 in the presence of (model) PSCs, and there is a tendency in the model to overestimate nitric acid condensation, leading to the high latitude biases seen in **Fig. 4**. Despite deficiencies in the StratChem representation of HNO_3 under PSC conditions, other constituents are modeled with sufficient skill to serve as background conditions during PSC and ozone hole conditions. Small differences in HNO_3 between 10 hPa and 1 hPa are consistent with larger MLS uncertainties. There is a small pressure-dependent ozone bias at pressures below 10 hPa. That may seem surprising given that there is no significant bias in the O-As (**Fig. 3)**. However, this bias results from the bias in O-Fs discussed above. The application of the IAU (**Section 3**) results in only one half of the analysis increment being applied at the center time of an assimilation window. Consequently, the systematic differences seen in **Fig. 4**(c) can be thought of as being half way between the mean O-F and O-A.

Most differences between **Figs 4** and **5** reflect known differences between the two versions of the MLS retrievals and are small for ozone, HCl, and HNO_3. The stratosphere is overall drier in version 5.0 than in version 4.2, with the latter understood to be too wet (Livesey et al., 2021). The largest differences between the two figures are seen in N_2O. Recall that N_2O observations are assimilated down to 68 hPa and we do not advise scientific use of the reanalysis N_2O at pressures larger than about 70 hPa. There are also significant differences in MLS N_2O between the two versions that result in part from a reduction of an unphysical drift (toward lower values) present in the older version assimilated here.

We have also examined the differences between M2-SCREAM and BRAM2 (**Fig. S2**). Overall, the zonal means agree very well between the two reanalyses, both of which assimilate the same data. The differences in HCl and HNO_3 are similar in spatial pattern and in magnitude to those shown in **Fig. 4**. The only noteworthy difference of up to 0.5 ppmv between the two reanalyses is seen in ozone at pressures lower than 4 hPa, where BRAM2 is not constrained by data and M2-SCREAM is only weakly constrained by MLS due to the fast ozone chemistry at those pressures.

Overall, the results presented above demonstrate internal consistency of the reanalysis and its close agreement with the assimilated MLS data within observational uncertainties. Comparisons against the improved version 5 MLS retrievals help diagnose spatially varying biases, especially in assimilated water vapor and N_2O.

6 Comparisons with independent observations

6.1 Representation of small-scale structures

We begin the evaluation of M2-SCREAM with a qualitative comparison of the reanalysis HNO_3, water vapor, and ozone with the observations made by the GLORIA instrument during a single flight on 9 March 2013 (**Fig. 6,** more examples are given in **Figs S3–S16**). The reanalysis

constituents are interpolated to the GLORIA vertical levels for each measurement and each level separately because the geolocations of the measurement tangent points vary significantly with altitude. The UTLS is a particularly challenging region for the reanalysis because this is where MLS uncertainties are typically larger than at higher altitudes and no observations are assimilated below about 10 km. While we assign significantly lower confidence to the constituent profiles below the bottom of the MLS profiles (marked as dashed lines in **Fig. 6**), we still expect the constituent fields' features to be consistent with the real dynamics of the atmosphere at those altitudes because the model is driven by assimilated MERRA-2 meteorology. Since the spatial and temporal resolution of the aircraft data is significantly higher than that of M2-SCREAM, we do not expect the reanalysis to capture many of the small-scale features seen in GLORIA observations such as gravity wave signatures in the constituent fields.

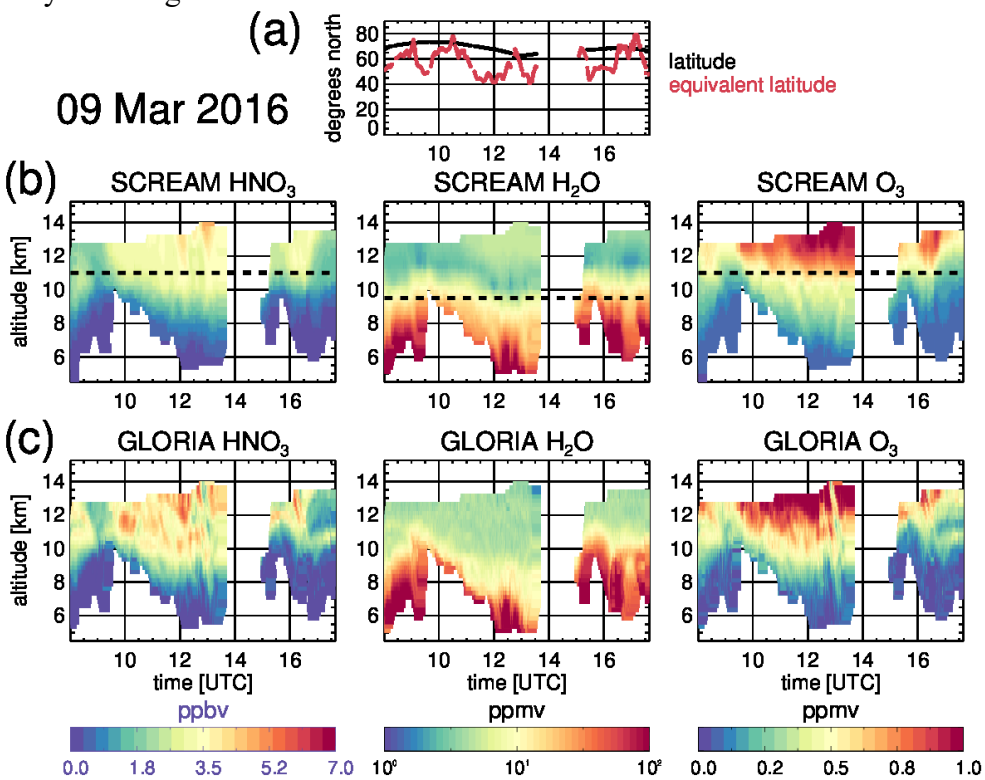

Figure 6: (a) *Latitudes and equivalent latitudes of the GLORIA measurements at 10 km. (b) M2-SCREAM HNO_3, water vapor, and ozone collocated with GLORIA measurements during a single flight on 9 March 2016. The dashed lines mark the lowest altitudes of MLS observations assimilated in M2-SCREAM. (c) GLORIA observations.*

The overall morphologies of the M2-SCREAM HNO_3, water vapor, and ozone profiles agree very well with the observations and are dynamically consistent with the flight trajectory in equivalent latitude space (Butchart and Remsberg, 1986) that ranges between 40ºN and 80ºN (**Fig. 6a**). A particularly interesting detail is the drop in the HNO_3 and ozone mixing ratios between 11 and 13.5 km at the end of the flight, coincident with a sharp increase of the equivalent (but not geographical) latitude, consistent with the aircraft crossing the boundary between different air masses. This feature is seen in M2-SCREAM and (in much greater detail) in the aircraft data. M2-SCREAM underestimates HNO_3, which is qualitatively consistent with the behavior at high northern latitudes in **Fig. 5**. The ozone fields are qualitatively very similar between the reanalysis and the

observations, although the smallest scale features are absent in M2-SCREAM as expected. Similar conclusions are drawn from examining other flights (**Figs S3–S16**).

There is a distinct layer of minimum water vapor mixing ratio above the tropopause in M2-SCREAM that is not present in the GLORIA observations. This is a persistent feature in all the aircraft comparisons (**Figs 6 and S3–S16**). The occurrence of similar lowermost stratosphere minima is also frequent in the reanalysis at other latitudes and periods (not shown). Close examination of available FPH water vapor profiles from the Lauder and Boulder locations reveals a frequent occurrence of such deep minima located between about 2 and 4 km above the tropopause (**Movies M1 and M3** in Supplementary Information). These minima are also present in the station data but are less deep and less frequent than those in M2-SCREAM. One possible explanation is a low bias in MLS water vapor in a shallow layer above the tropopause. Such a bias was identified in version 3 MLS retrievals with respect to a multi-instrument mean by Hegglin et al. (2013) and with respect to HIRDLS data by Schwartz et al. (2015), and v4 retrievals can be expected to be similar in this regard.

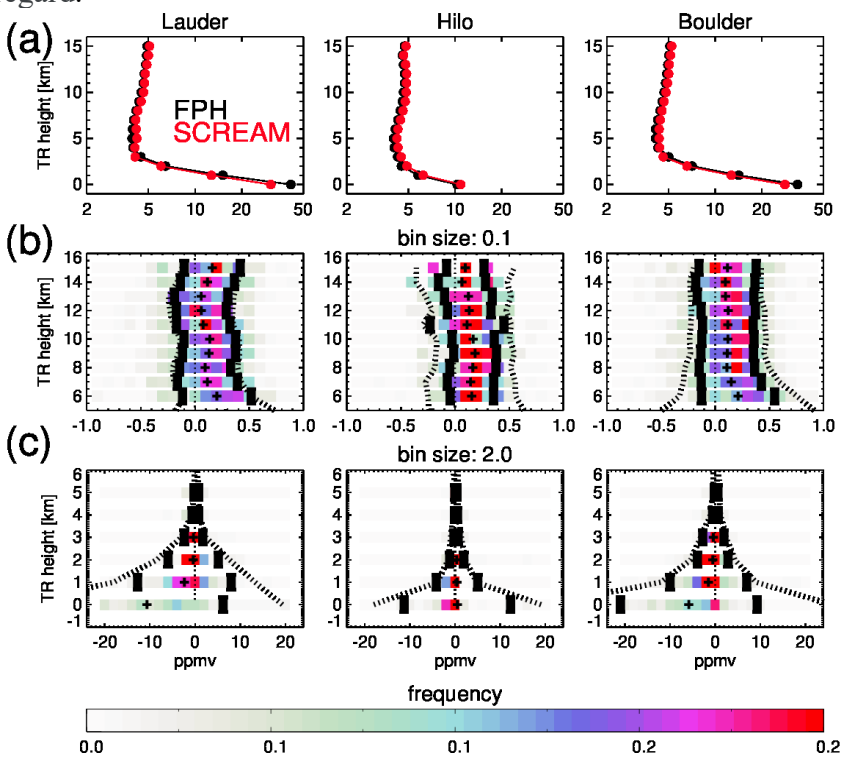

Figure 7: *Statistical comparisons for 2005–2021 of the M2-SCREAM water vapor with NOAA's frost point hygrometer (FPH) measurements at Lauder (169.68°E, 45.04°S) , Hilo (155.05°W, 19.72°N), and Boulder (105.2°W, 39.95°N) in units of ppmv. (a) Mean profiles from FPH (black) and M2-SCREAM (red). (b) and (c) M2-SCREAM minus FPH statistics: mean (plus signs), standard deviations around the mean (short vertical bars), probability density functions (colors) as functions of tropopause-relative (TR) height. The dotted lines are the mean difference plus/minus standard deviation of the FPH observations. Note that the abscissa shows the natural logarithm of H_2O mixing ratio in ppmv. Vertical ranges of 5–15 km and 0–5 km are shown in (b) and (c), respectively. Note the different bin sizes used in (b) and (c).*

Comparisons of the reanalysis water vapor against the FPH data at these three locations are shown in **Fig. 7**. These figures are similar in design to **Fig. 3**. Here, however, the vertical coordinate is the distance from the tropopause ("tropopause-relative height") defined as the 2 potential vorticity (PV) unit surface (1 PVU: 1.0×10^{-6} m^2 s^{-1} K kg^{-1}) or the 380-K potential temperature surface, whichever is located at a lower altitude. The difference averages are calculated as M2-SCREAM minus FPH sonde observations. Additionally, the dotted lines in **Fig. 7(b,c)** do not represent uncertainties but rather the standard deviations of the observed FPH mixing ratios added and subtracted from the mean M2-SCREAM minus FPH difference. The agreement between the two data sets is very close between about 4 and 15 km above the tropopause. Below 4 km there is more spread in the differences but the standard deviations of the difference are generally smaller than the observed variability as the latter becomes very large near the tropopause. These standard deviations range between 17 ppmv at the tropopause (Lauder) and about 0.2 ppmv above 4 km at all three stations. Most of them are smaller than the standard deviations of the observed water vapor mixing ratio also above 4 km but often not by much as water vapor variability is very small in much of the lower and middle stratosphere except close to the tropopause. Correlations between M2-SCREAM and FPH (not shown) range between 0.55 and 0.9. The reanalysis exhibits a positive bias of up to 0.2 ppmv at altitudes higher than the 4 km above the tropopause, consistent with a MLS version 4.2 wet bias compared to the FPH record at these altitudes (e.g., Figure 1 of Livesey et al., 2021). The shapes of the average water vapor profiles are very similar in FPH data and in M2-SCREAM. The minimum is located between 3 and 4 km above the tropopause, indicative of the existence of a tropopause transition layer where moist tropospheric air mixes into the LS. This is consistent with the findings of Hegglin et al. (2009) who identified such a transition layer in the extratropics. Overall, the performance of M2-SCREAM water vapor measured against the FPH is extremely good. We combined the individual sonde profiles and collocated M2-SCREAM profiles into three short animations to facilitate convenient viewing. These animations are included in the Supplementary Information (**Movies S1–S3**). We encourage the reader to watch them as they help to qualitatively assess this agreement on a profile-by-profile basis. Those animations also show assimilation uncertainty envelopes around the assimilated profiles.

6.2 Australian New Year's busfires

Severe bushfires in south-eastern Australia in December 2019 and January 2020 resulted in exceptionally strong pyro-cumulonimbus ("PyroCb") outbreaks that injected plumes of smoke and tropospheric air into the LS (Allen et al., 2020; Kablick et al., 2020; Khaykin et al., 2020; Schwartz et al., 2020). Because of the presence of sunlight-absorbing smoke the largest plumes acted as synoptic-scale heat sources whose thermal expansion spawned localized regions of anticyclonic circulation (Allen et al., 2020). The air trapped in these anticyclones remained relatively isolated and retained signatures of tropospheric composition for up to several months. Using MLS data, Schwartz et al. (2020) tracked several large plumes as they traveled through the southern hemisphere (SH) stratosphere and identified that at least one of them circled the Earth three times before dispersing. The convective plumes associated with the Australian New Year's (ANY) fires are not simulated in the GEOS model. The presence of the plume-induced anticyclones and their chemical composition in the reanalysis can, therefore, arise only from assimilation of radiance data in MERRA-2 and MLS constituent observations in CoDAS.

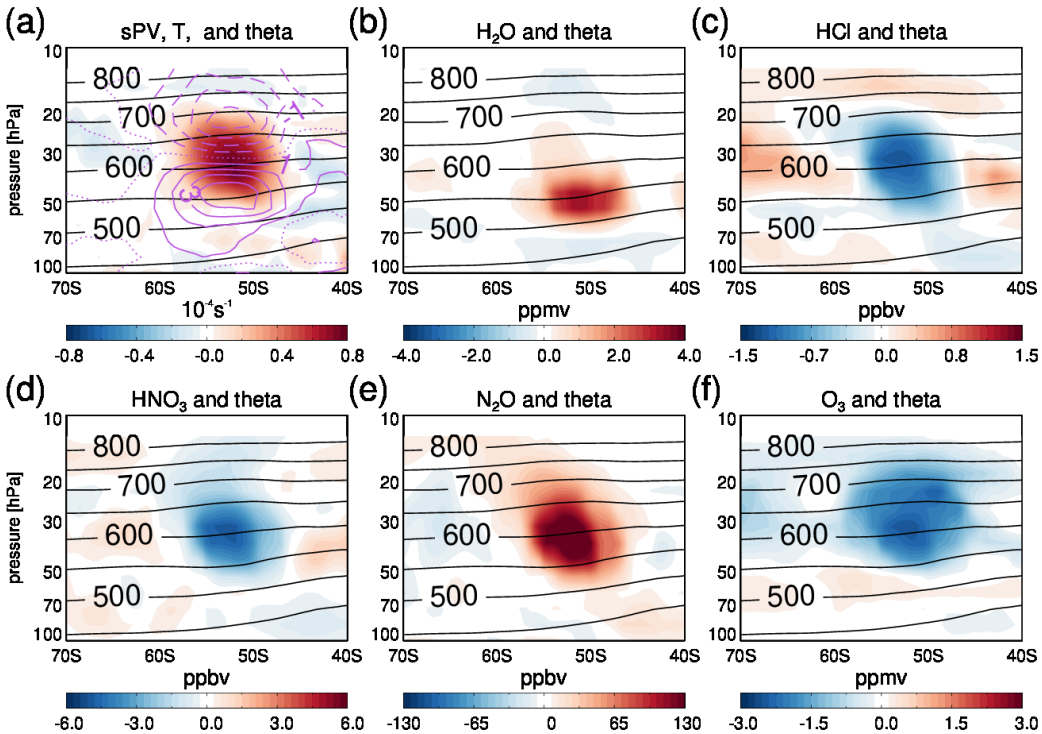

Figure 8: *Departures from the zonal mean for sPV (shading) and temperature (magenta contours, 1 K spacing) (a), water vapor (b), HCl (c), HNO_3 (d), N_2O (e), and ozone (f) at 80 °W on 31 January 2020. Also shown are potential temperature contours (50 K intervals).*

Figure 8 shows the departures from the zonal mean of scaled PV (sPV), temperature, and the five M2-SCREAM constituents near the center of the largest ANY plume at 80°W between 40°S and 70°S on 31 January 2020. These plots are similar to Figures 1 and 2 in Allen et al. (2020). For display purposes we use sPV, defined as PV divided by a standard static stability value at each potential temperature (Dunkerton & Delisi, 1986; implementation as in Manney et al., 1994). This method was chosen rather than simply PV because the rapid increase of the absolute values of the latter with height tends to produce an apparent upward shift in the location of the dynamical anomaly with respect to the temperature anomaly dipole. The anticyclone is seen in **Fig. 8** as a positive sPV anomaly. The associated positive and negative temperature departures at the bottom and top of the plume, respectively, are consistent with the dynamical response to the presence of a localized heat source and were observed in previous studies (Allen et al., 2020; Khaykin et al., 2020). The positive anomalies in water vapor (up to about 3 ppmv) and N_2O (in excess of 130 ppbv) and negative anomalies for HCl (over 1 ppbv), HNO_3 (up to 0.4 ppbv), and ozone (2 ppmv) indicate partly tropospheric air and, for HCl, chemically altered composition of the plume that extended between about 50 and 20 hPa at that time, in agreement with previous studies. Preliminary comparisons with individual MLS profiles (not shown) suggest that M2-SCREAM reproduces the position of the HCl, HNO_3, N_2O, and ozone departures from the zonal mean, although it tends to slightly underestimate the magnitudes of the first three. However, M2-SCREAM fails to capture the magnitude and vertical extent of the water vapor anomaly. MLS observed water vapor mixing ratios as high as 18 to 20 ppmv at the center of the plume, several times larger than those seen in M2-SCREAM. The M2-SCREAM water vapor anomaly is limited to the bottom portion of the plume. A much larger maximum above is not present in the reanalysis.

An analysis of the assimilation statistics revealed that, in the absence of the plume in the GEOS background state, the observations of the extremely large water vapor mixing ratios were rejected by the CoDAS internal quality control. This initial analysis of the ANY plume in M2-SCREAM demonstrates that the reanalysis is capable of reproducing most of its features realistically. It also demonstrates the trade-off between simple assumptions meant to reduce noise in the assimilated product (additional quality control) and the ability to capture short-lived anomalies. The failure to reproduce the maximum water vapor anomalies warrants future improvements to the GEOS assimilation system. A preliminary analysis of the LS water vapor enhancement following the eruption of the Hunga Tonga volcano in mid-January 2022 (not shown) suggests that relaxing the internal quality screening may be necessary to reproduce the observed constituent fields during extreme anomalous events.

6.3 Global comparisons

In this subsection, we perform global comparisons of M2-SCREAM with independent observations from ACE-FTS and SAGE III/ISS. **Figure 9** shows the zonal mean differences between the five M2-SCREAM constituents and all available ACE-FTS data between 2005 and 2020. The reanalysis fields are sampled at ACE-FTS observation locations and within 1.5 hour of observation times and both data sets are interpolated to the same pressure grid. The color scales used in this figure are the same as those in the MLS comparisons to help the reader assess the magnitude of the reanalysis differences with independent data compared to the differences with assimilated observations. As expected, the former are significantly larger than the latter as the reanalysis is tightly constrained to MLS. It follows that **Figs 4** and **9** reflect the relative biases between the two data sources as well as points of agreement between them and potentially any seasonal sampling bias from the ACE-FTS orbit.

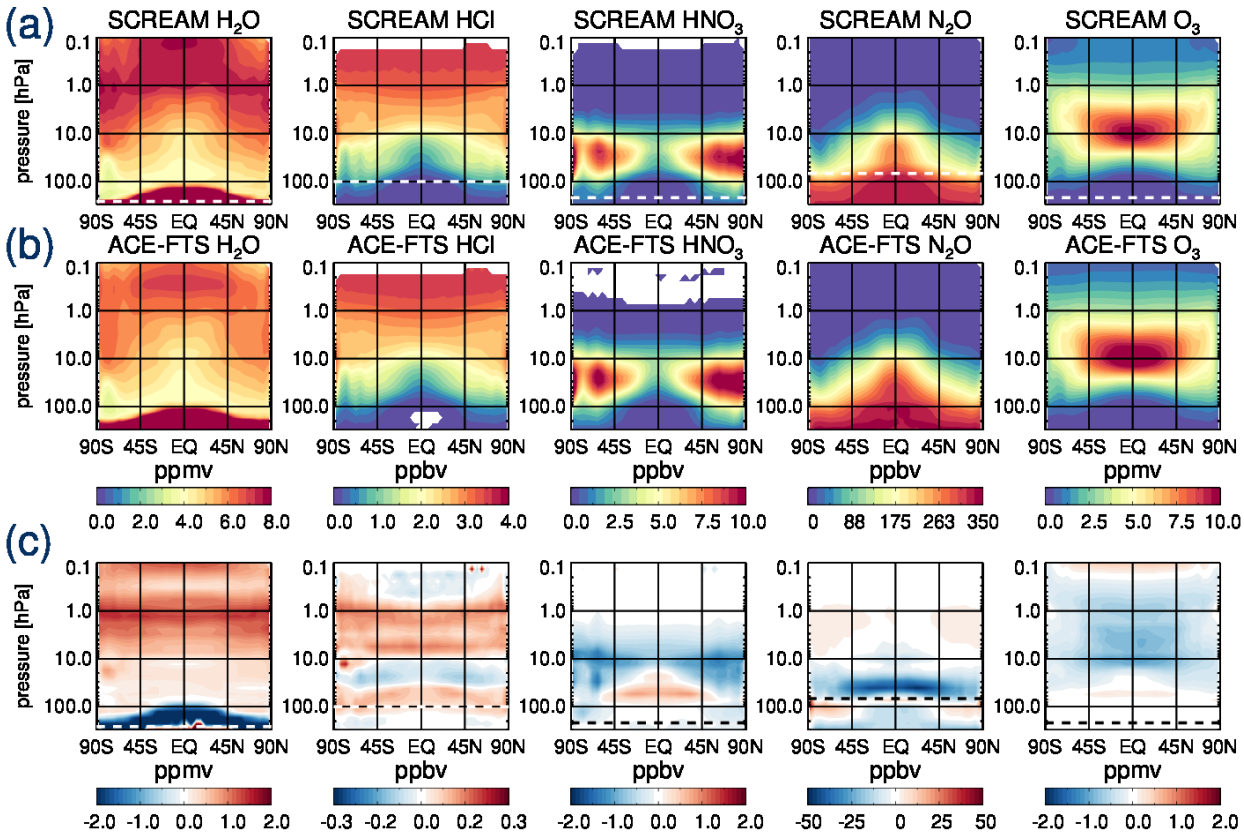

Figure 9: *Zonal mean mixing ratios of the assimilated constituents calculated from ACE-FTS data (a) and M2-SCREAM sampled at the ACE-FTS observation locations (b). The M2-SCREAM minus ACE-FTS differences are shown in (c). The dashed lines (white, (a); black, (c) except HCl where white is used) mark the lowest altitudes of MLS observations assimilated in M2-SCREAM. All quality-screened ACE-FTS data between 2005 and 2021 are used.*

Even a cursory examination of **Fig. 9** (a) and (b) reveals overall very good agreement between the reanalysis and ACE-FTS in terms of the climatological structures in the five constituent fields, with the differences being small relative to the average mixing ratios. M2-SCREAM water vapor is higher than that from ACE-FTS, with the magnitude of the difference increasing with altitude and reaching almost 2 ppmv at the stratopause. This is qualitatively consistent with an overall wet bias in MLS version 4.2, but the magnitude of the differences is much larger than that between M2-SCREAM and MLS version 5 (Figure 5), FPHs (Figure 7), and SAGE III/ISS **(Fig. 11)**. The HCl differences are within about 0.1 ppbv and exhibit an alternating layered pattern with little latitudinal variation. For HNO_3 we see a positive difference (M2-SCREAM greater than ACE-FTS) of up to 0.5 ppbv between 45°S and 45°N and below 10 hPa and larger negative differences in excess of 1 ppbv around 10 hPa. This pattern resembles that of the M2-SCREAM minus MLS differences, but its magnitude is larger. The largest N_2O differences (of about 40 ppbv) are seen between 40 and 50 hPa in the tropics and subtropics. As seen in **Figs 4** and **9,** tropical N_2O is approximately constant between 50 and 20 hPa, in agreement with MLS (see also **Fig. S2**). A close examination of MLS N_2O profiles in the tropics (not shown) confirms a near-zero vertical gradient in that region. As there is no known dynamical mechanism that would produce this feature, we suggest that it results from MLS retrieval errors. Finally, the M2-SCREAM ozone is up to 1 ppmv lower than that reported by ACE-FTS in most of the stratosphere and in the lower mesosphere.

As expected from our M2-SCREAM minus BRAM2 comparison, the biases between M2-SCREAM and ACE-FTS are qualitatively consistent with those between BRAM2 and ACE-FTS calculated in Errera et al. (2019) and shown in their Figure 5 for the 30ºN to60ºN latitude band. Their results also demonstrate that these differences agree with those between collocated MLS and ACE-FTS observations.

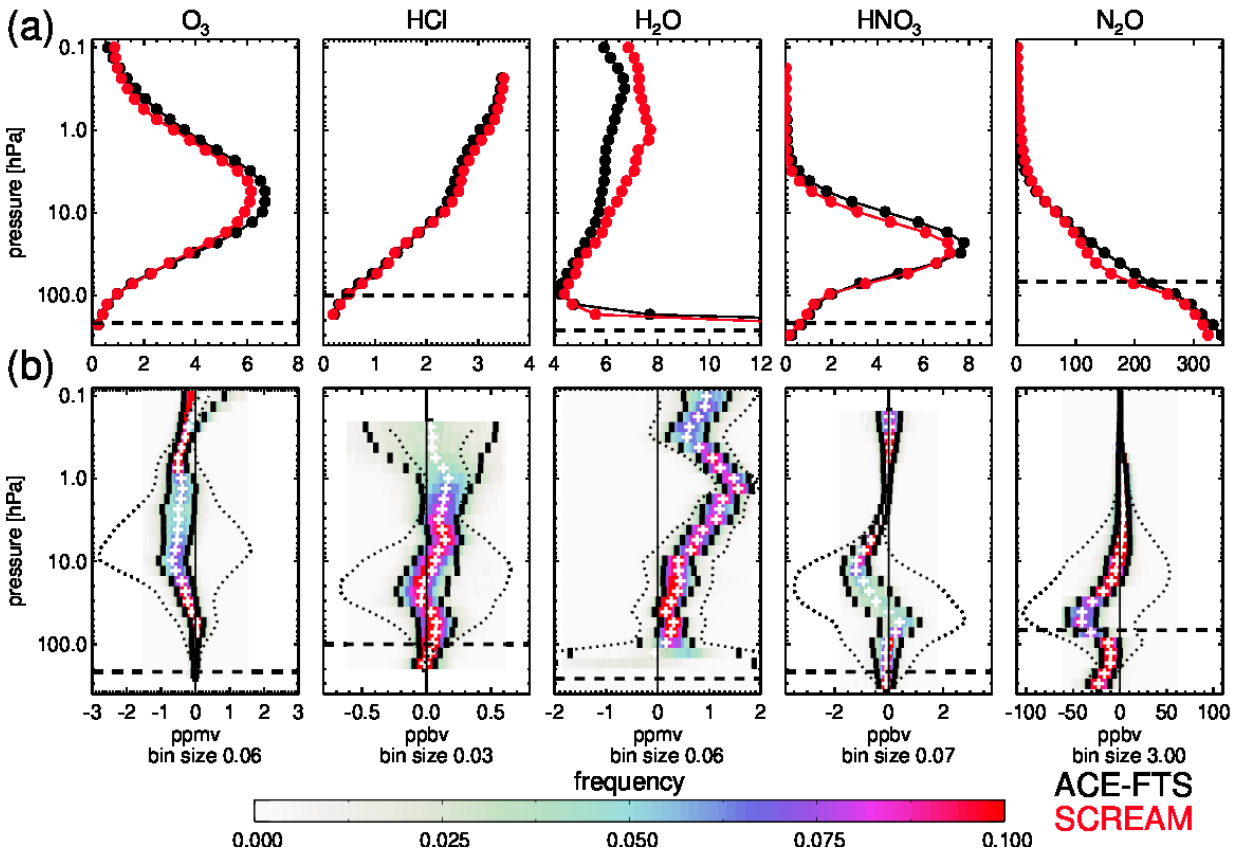

Figure 10: *Statistical comparisons of the M2-SCREAM ozone, HCl, water vapor, HNO_3, and N_2O with ACE-FTS observations for 2020. (a) mean profiles (M2-SCREAM in red and ACE-FTS in black). (b) difference statistics; specifically, difference mean (plus signs), standard deviations around the mean (short vertical bars), probability density functions (colors) at prescribed pressure levels. The dotted lines are the mean difference plus/minus standard deviation of the ACE-FTS observations. All available 2020 ACE-FTS data are used for ozone, HCl, and water vapor. Only data between 60ºS and 60ºN are used for HNO_3. The dashed lines mark the lowest altitudes of MLS observations assimilated in M2-SCREAM.*

Globally aggregated statistics of M2-SCREAM minus ACE-FTS differences along with average constituent profiles are shown in **Fig. 10** (results limited to the 30ºN to 60ºN for December 2005–to January 2020 are shown in **Figs S17** and **S18** for direct comparison with the results of Errera et al., 2019 for BRAM2). Since the global mean differences were discussed above (**Fig. 9**). Here we focus on the difference pdfs and their standard deviations. As in **Fig. 7** the dotted lines show the mean difference plus/minus the standard deviations of the observed mixing ratios. These standard

deviations provide a measure of constituent variability that includes seasonal and geographical variations. For ozone (**Fig. 10**) the differences are approximately normally distributed in the stratosphere. The difference standard deviations (short vertical bars) are within 0.5 ppmv, much smaller than the ozone variability. Above the stratopause the difference pdfs are bimodal. This is likely related to the bimodality of the ozone concentrations around twilight (the local time of ACE-FTS observations). The M2-SCREAM minus ACE-FTS differences for HCl have standard deviations of up to 0.25 ppbv in the stratosphere. The global mean HCl mixing ratios vary between about 0.1 and 3 ppbv. The variability of the observed mixing ratios is typically much larger in the middle stratosphere, reaching about 0.6 ppbv between 20 and 10 hPa. Above the stratopause, the differences become larger than the variability of the observations. This is also where MLS uncertainties increase with altitude (**Fig. S1b)**. The most pronounced feature of the water vapor comparisons in **Fig. 10** is the altitude dependent bias resulting from a relative difference between MLS and ACE-FTS bias that was discussed above. The difference standard deviations are within the climatological envelope of variability. It should be noted that both the M2-SCREAM minus ACE-FTS differences and variability of the observations are relatively small compared to the average stratospheric water vapor. For HNO_3 and N_2O the difference standard deviations are considerably smaller than the observation standard deviations in the deep layers where each of these species exhibits relatively large mixing ratios (for pressures greater than about 5 hPa for HNO_3, and 20 hPa for N_2O).

Zonal mean and annual differences between M2-SCREAM and SAGE III/ISS profiles of water vapor are shown in **Fig. 11**. Because of the significant noise in the SAGE III/ISS retrieval (see section 4.2), we chose to show the instrument uncertainties (dotted lines in **Fig. 11d**) rather than the standard deviation of the observed values, shown in **Fig. 10**. Throughout the stratosphere, M2-SCREAM mixing ratios of water vapor are biased high with respect to SAGE III/ISS retrievals, with the differences ranging from about 0.1 to 0.7 ppm. The largest differences, up to about 10%, are seen in the upper stratosphere. SAGE III/ISS water vapor profiles were reported to have a roughly 10% dry bias with respect to MLS version 4.2 (Davis et al., 2021) and close agreement with MLS version 5.0 outside of tropospheric regions (Park et al., 2021). Among the data sets analyzed here, MLS, ACE-FTS, and SAGE III/ISS the low water vapor concentrations reported by ACE-FTS in the upper stratosphere and lower mesosphere emerge as an outlier.

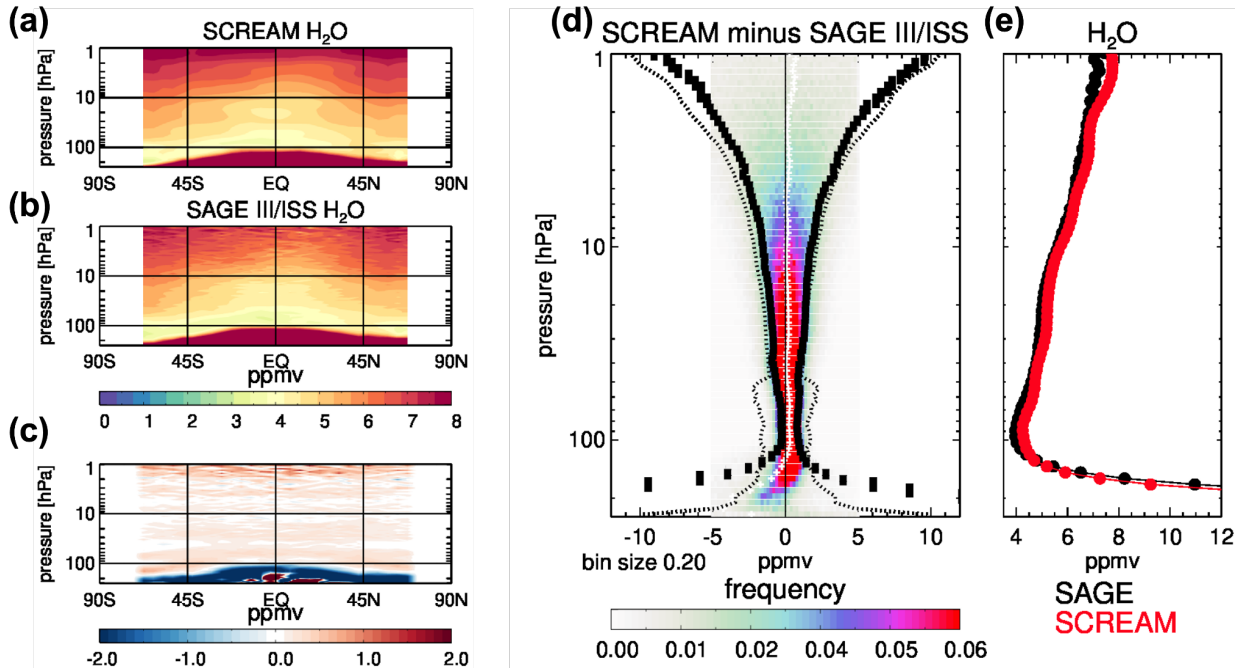

Figure 11: *Comparison of M2-SCREAM water vapor with SAGE III/ISS observations in 2018. Zonal mean annual water vapor from M2-SCREAM and SAGE III/ISS are shown in panels (a) and (b), respectively. Panel (c) shows the reanalysis minus SAGE III/ISS. Various global statistics are plotted in (d) and (e). Panel (d) plots the mean difference (plus signs), mean plus/minus difference standard deviation (short vertical bars), and pdfs of the difference (colors) in the left-hand side panel. The dashed lines are plus/minus estimated SAGE III/ISS uncertainties. Panel (e) displays the average profiles.*

6.4 Long-term behavior and interannual variability

Since ACE-FTS is available for the entire MLS record, we use ACE-FTS observations to assess the performance of M2-SCREAM over the reanalysis period. Comparisons are done for ozone, HCl, water vapor and HNO_3 using annually aggregated data for 2005 to 2020 (70 hPa, 30ºN–60ºN, **Fig. 12)**. We have decided not to evaluate the long-term performance of assimilated N_2O because of the known significant drift in the MLS retrievals of that constituent in the LS, even though this is where N_2O is of particular interest owing to its long lifetime. We do not recommend N_2O from M2-SCREAM for studies of long-term changes. We do however emphasize its utility for transport studies on shorter time scales (see **Section 6.5**). The figure shows pdfs (color), averages (black dots), standard deviations (dashed lines) of annually aggregated M2-SCREAM (top row) where, as always, M2-SCREAM is sampled at observation locations), ACE-FTS (middle row) and their differences M2-SCREAM minus ACE-FTS (bottom row).

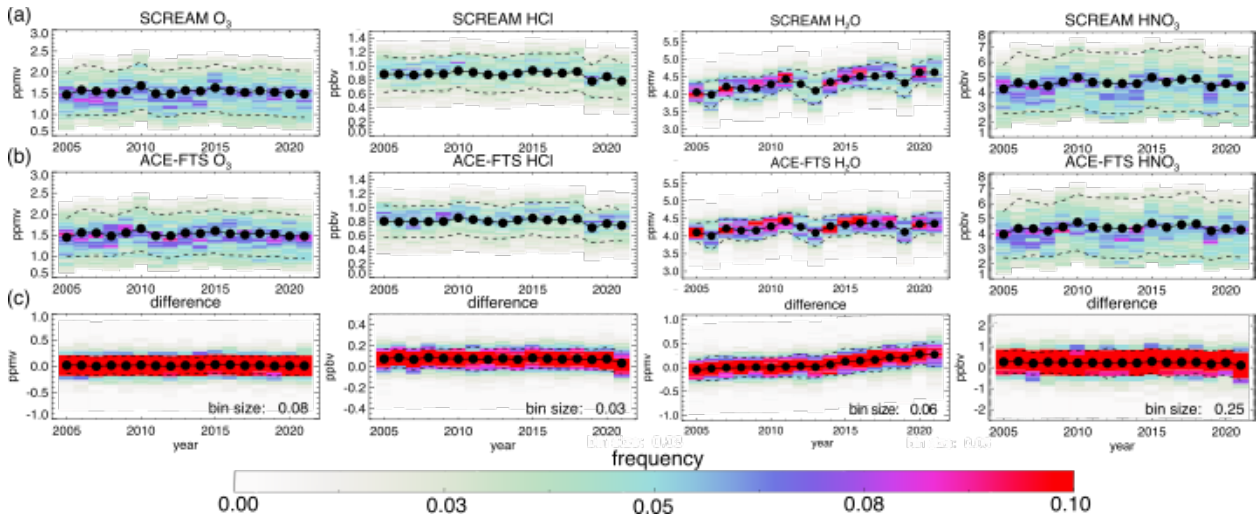

Figure 12: *Time series of annual 70-hPa 30ºN–60ºN pdfs from M2-SCREAM subsampled at ACE-FTS observation locations (a) and from ACE-FTS data (b). Panel (c) shows the pdfs of M2-SCREAM minus ACE-FTS. Results are shown for (from left to right) ozone, HCl, H_2O, and HNO_3. The black filled circles are the annual averages, and the dashed lines mark the 1-sigma envelopes around the mean.*

The distributions of ozone and HCl vary considerably from year to year. There is a very good agreement between the structures of the histograms from ACE-FTS and M2-SCREAM (**Fig. 12** (a, b)), although the reanalysis HCl exhibits a slightly larger spread than that in ACE-FTS observations. The difference pdfs (**Fig. 12c**) use the same bin sizes as those used in the other panels to emphasize the relative magnitudes of the differences and the mixing ratios themselves. For both constituents, the reanalysis minus satellite differences are much smaller than the mixing ratio variability. The reanalysis HCl exhibits a positive bias of about 10 to 12% consistent with the results shown in **Fig. 9**. Linear fits to the differences (not shown) show no evidence of any relative drift between M2-SCREAM and ACE-FTS ozone and HCl.

Figure 12 also shows the results for water vapor and HNO_3. Here again, the details of the pdf shapes are well reproduced by M2-SCREAM. As seen in the ozone and HCl differences, the difference pdfs (panel (c)) are significantly more concentrated (indicative of little spread) and more symmetric around the mean than the mixing ratio distributions. The reanalysis HNO_3 shows a small positive bias of about 0.3 ppbv (compared to the mean mixing ratio of about 4 ppbv) as already seen in **Fig. 9**. There is no evidence of a drift between the M2-SCREAM and ACE-FTS HNO_3. Water vapor in the reanalysis exhibits a drift of approximately 0.2 ppmv per decade, consistent with the known drift in MLS version 4.2 water vapor.

Overall, all four reanalysis constituents agree well with ACE-FTS in the extra-tropical LS in terms of interannual variability of the tracer mixing ratio distributions over the reanalysis period. Apart from water vapor (and N_2O, not shown) there is no evidence of time dependent biases between M2-SCREAM and ACE-FTS observations.

6.5 Dynamically controlled interannual variability

We now turn to the representation of dynamically controlled interannual variability of the five assimilated constituents. Time series of detrended anomalies of N_2O (**Fig. 13**), water vapor, HNO_3 (**Fig. 14**), ozone, and HCl (**Fig. 15**) interpolated to the 520-K potential temperature surface (between about 40 and 70 hPa, the latter value being near the lowest level where N_2O is assimilated) shown as functions of equivalent latitude. The black lines in all three figures are selected PV contours, with all except the lowest value shown indicating the location of the polar vortex edge during fall-winter-spring. Also shown are effective diffusivity anomalies, k_{eff} (green) (Nakamura, 1996; Haynes and Shuckburgh, 2000a,b; Allen and Nakamura, 2001). The latter indicate regions of enhanced (solid) and suppressed (dashed) isentropic mixing. The anomalies are calculated by removing the monthly means calculated over the reanalysis period. For the trace gas fields, rather than subtracting the climatology, a linear fit to the daily time series over all the years is removed. Trace gas distributions on potential temperature surfaces are controlled by the effects of slow vertical diabatic transport and faster isentropic mixing by waves (Plumb, 2007; Shepherd, 2007). These two processes typically act to sharpen and weaken the meridional tracer gradients, respectively. Because potential temperature, equivalent latitude and passive tracers are conserved in adiabatic and frictionless flows, any temporal variability of a constituent in the potential temperature / equivalent latitude space arises from non-conservative processes and/or chemistry. For further discussion of tracer analysis in potential temperature / equivalent latitude coordinates see Manney et al. (2005; 2009), Santee et al. (2011), and references therein.

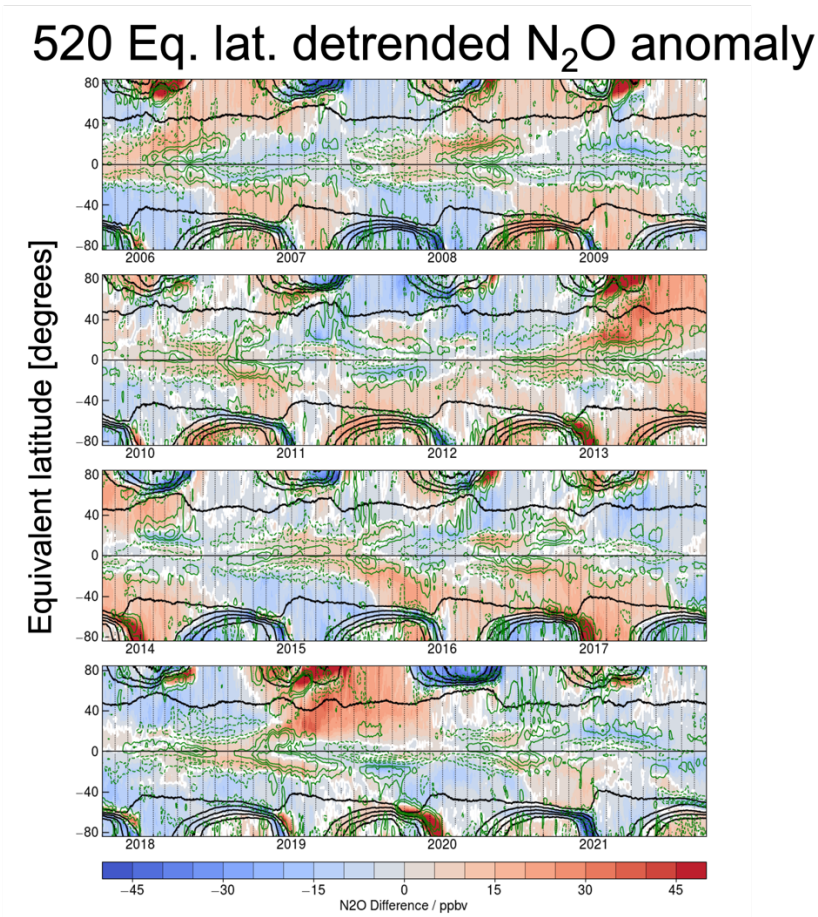

Figure 13: *Equivalent latitude time series on the 520-K isentropic surface of M2-SCREAM N_2O detrended differences from the Aura mission climatology (time series is detrended by removing a linear fit over the mission to the values for each day of year). Black overlaid contours are sPV from 1.0 to 2.6 $\times 10^{-4}$ s^{-1} by 0.4 $\times 10^{-4}$ s^{-1}, with the 1.4 contour representing the outer part of the vortex edge region. Green overlaid contours are anomalies from climatology in effective diffusivity expressed as log-normalized equivalent length (dashed negative contours indicate less mixing than in the climatology, solid contours more mixing than in the climatology).*

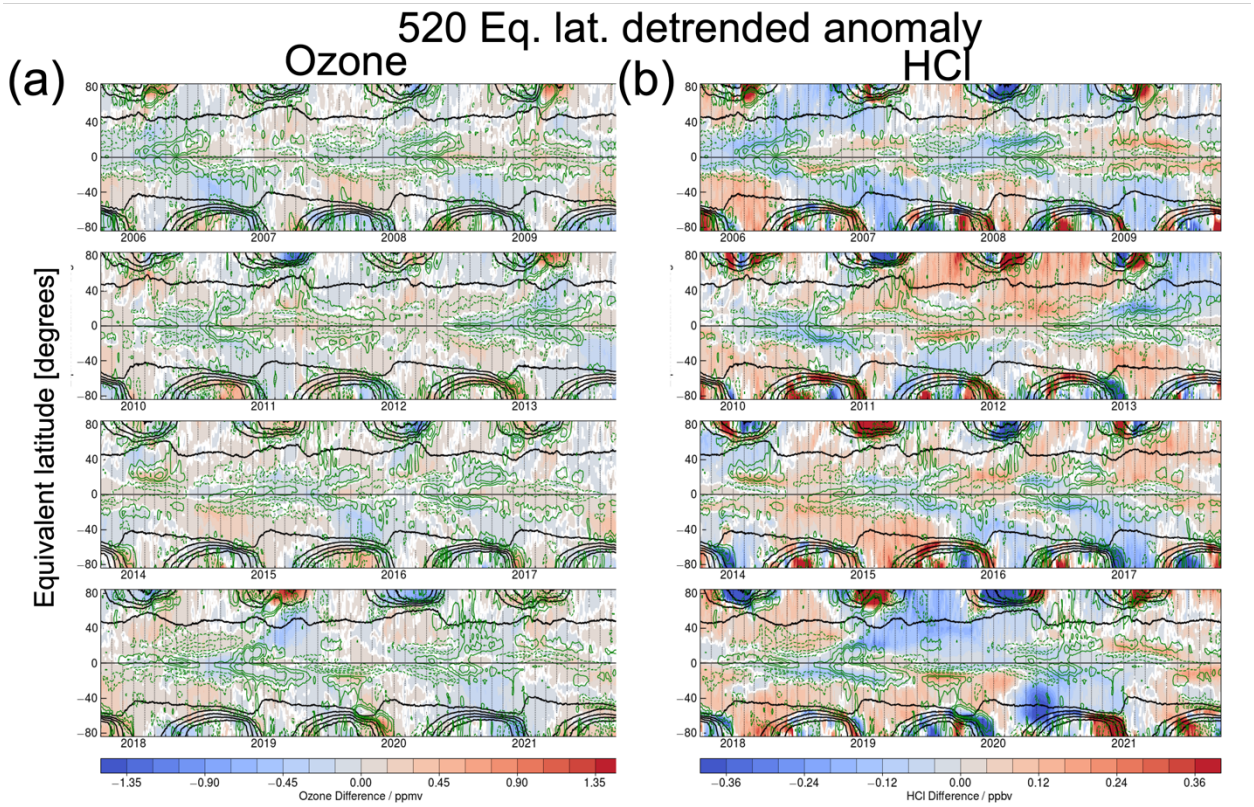

Figure 14: *As in Fig. 13 but for ozone (a) and* HCl (b)

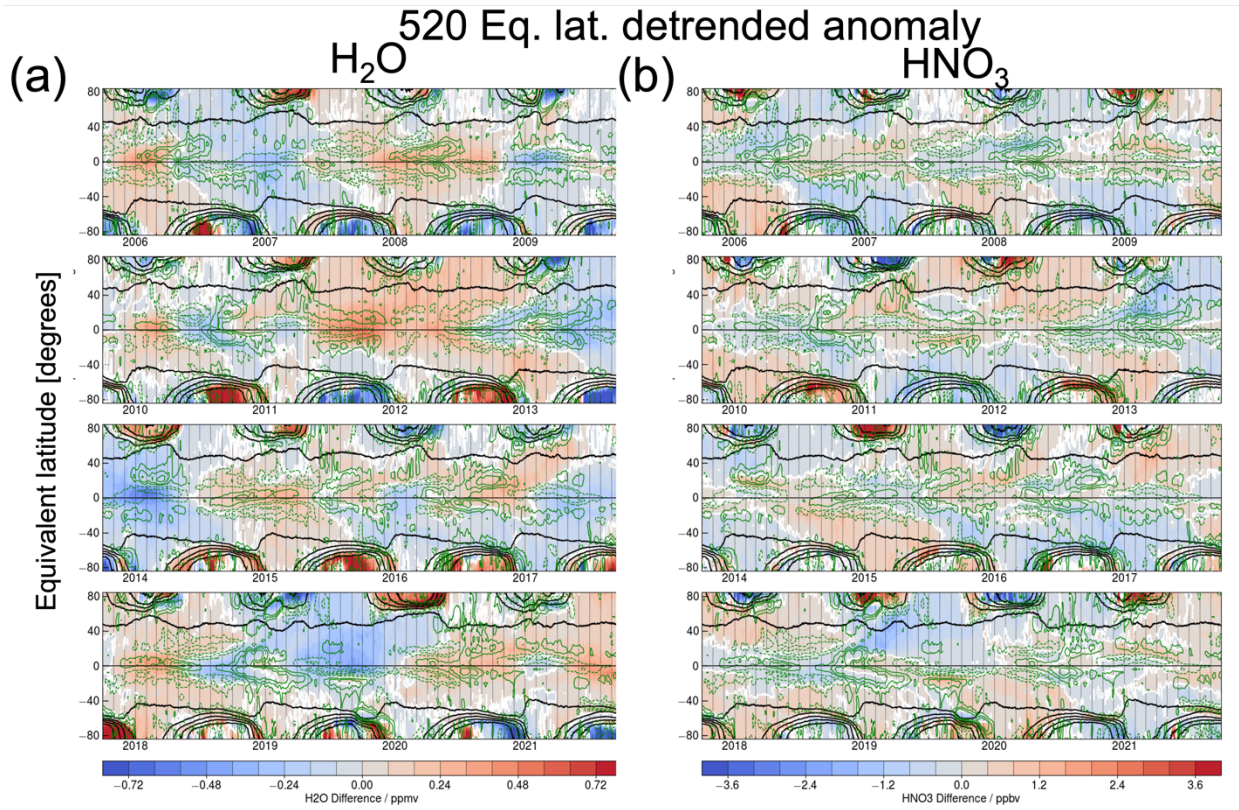

Figure 15: *As in Fig. 13 but for water vapor (a) and* HNO_3 (b)

As a chemically inactive gas in much of the stratosphere, N_2O is an excellent transport tracer. We highlight some of the main features of the dynamical variability deducible from M2-SCREAM N_2O on interannual scales (**Fig. 13**). The polar vortices that form in autumn and break up in spring in each hemisphere are demarcated by a region of strong PV gradients and isolate the air within them. In the absence of mixing, the slow descent of air within an undisturbed polar vortex leads to depressed N_2O concentrations. In contrast, dynamical disturbances such as sudden stratospheric warmings (SSWs) increase the vortex edge permeability (measured by k_{eff}), leading to higher than average N_2O. Signatures of NH SSWs in 2006 (Manney et al., 2008,2009a; Coy et al., 2009), 2009 (Manney et al., 2009b; Harada et al., 2010), 2013 (Coy and Pawson, 2015), 2018 and 2019 (Butler et al., 2020), and 2021, and a SH SSW in 2019 (Hendon et al., 2019; Wargan et al., 2020b; Safieddine et al., 2020) are clearly discernible in **Fig. 13** in the PV contours and enhanced N_2O mixing ratios around the vortex edge in these years. These events are associated with positive k_{eff} anomalies. A signature of the early final warming in March 2016 (Manney and Lawrence, 2016) is also seen as strongly positive N_2O and k_{eff} anomalies. In contrast, the very cold and strong NH polar vortex in 2020 featured record low N_2O in the MLS period (Manney et al., 2020; Wohltmann et al., 2020; 2021). Most of these events are well documented in the studies cited above and references therein. In the tropics, a clear marker of the quasi-biennial oscillation (QBO) is seen in the pattern of positive and negative N_2O anomalies alternating with a period of 26–28 months. The 520-K potential temperature surface is located near the bottom of the extratropical surf zone (McIntyre and Palmer 1983), a region of strong wintertime wave activity and resulting mixing flanked by the edge of the polar vortex and the subtropical mixing barrier on the poleward and equatorward sides, respectively. We note that positive subtropical k_{eff} anomalies in some years

indicate a weakened transport barrier and are concurrent with enhanced N_2O transport into the extratropics. The consistency of N_2O and k_{eff} variability seen in **Fig. 13** demonstrates the utility of this M2-SCREAM data set for transport studies despite the biases and drifts discussed in previous sections.

Unlike N_2O, variability of ozone, HCl, water vapor, and HNO_3 shown in **Figs 14** and **15** depends not only on transport but also chemistry and, in the case of water vapor and HNO_3, also on the thermodynamics of phase transitions. Low HCl and ozone anomalies in the exceptionally cold NH winters of 2011, 2016, and 2020 indicate strong chlorine activation and ozone depletion consistent with previous studies (Manney et al., 2011; Manney and Lawrence 2016; Manney et al., 2020; Wohltmann et al., 2020). In the SH, strong dehydration (through condensation) correlates well with strong ozone depletion because both are associated with below-average temperatures. The situation in the NH HNO_3 is more complex because of extreme interannual variability and temperatures that are commonly close to the thresholds for forming of HNO_3- and ice-containing PSCs, and we do not analyze it here. We also do not discuss HNO_3 in the polar regions further because of its low quality in M2-SCREAM (**Section 3**).

Outside of the polar regions the distribution of all five assimilated species at 520 K is determined mainly by transport. The spatial patterns of the high and low anomalies at extrapolar latitudes are remarkably similar among N_2O, ozone, HNO_3, and HCl. Note that since N_2O has sources at the surface and the other three in the upper atmosphere, the signs of the anomalies are reversed between N_2O and the other tracers. High correlations between long-lived constituents are expected from theory (Plumb et al., 2007) and provide confidence in dynamical consistency of the M2-SCREAM output. Some degree of correlation is also seen between the N_2O and water vapor anomalies, but interannual variability of the latter outside the polar regions is primarily controlled by the temperature variations at the tropical cold point tropopause (e.g., Randel and Park, 2019). One notable exception to the dynamically induced correlation among the tracers is the strong negative HCl anomaly in the SH in 2020, which lacks counterparts in HNO_3 and N_2O, suggesting a chemical origin of the HCl anomaly. Santee et al. (2022) present strong evidence of chlorine activation on smoke particles from the ANY event in early 2020 and show results similar to ours using MLS version 5 data (their Figure 1, see also Rieger et al., 2021). Overall, the features seen in **Figs 13-15** closely track those in analogous plots constructed from MLS data at this and other isentropic levels (not shown).

7 Recommended usage

As noted in **Section 2**, data assimilation is by construction driven by observations. There are at least three types of situations in which the reanalysis output is not sufficiently informed by observations. First, there were several MLS data outages long enough to compromise the assimilation results. These are listed in **Table 2**. We recommend caution when using the reanalysis over those periods. Second, due to short chemical time scales of upper-stratospheric ozone the information content from MLS data cannot be assumed to fully propagate forward during six-hourly model integrations (see **Section 5**). The M2-SCREAM ozone fields at pressures lower than 10 hPa (the upper stratosphere) should be treated with caution. Third, M2-SCREAM does not assimilate HNO_3 at locations where condensed nitric acid is present in the model. Due to model deficiencies, significant low bias exists in those regions. This affects HNO_3 in the polar LS during

winter and spring in both hemispheres. Thus, we do not recommend using the M2-SCREAM HNO_3 in those regions and seasons. The model's condensed HNO_3 is provided in the M2-SCREAM output files. While it is not recommended for scientific use it does provide information indicating the areas where gas-phase HNO_3 is not assimilated.

Table 2 MLS Outages. MLS outages 5 days and longer

From	To	Remarks
30 March 2006	4 April 2006	
13 July 2007	8 August 2007	HCl only
27 March 2011	19 April 2011	
19 February 2012	24 February 2012	
15 March 2012	21 March 2012	
4 June 2018	11 June 2018	
21 June 2018	26 June 2018	
10 July 2018	19 July 2018	
27 January 2019	31 January 2019	

8 Summary

This study describes and evaluates M2-SCREAM, a new reanalysis of stratospheric ozone, water vapor, HCl, HNO_3, and N_2O from assimilation of MLS constituent profiles and OMI total ozone with the GEOS CoDAS system developed at NASA's GMAO. The reanalysis covers almost the entire MLS mission from September 2004 through February 2022 (at time of writing) and is made freely available to the scientific community. We argue that the reanalysis, as produced using data assimilation, is fundamentally an observation-driven product. That is, the information content of the available species is derived exclusively from the data where observations are assimilated. This is supported by the excellent agreement between the M2-SCREAM constituent fields and the assimilated MLS data, as well as by an analysis of the O-F and O-A residuals presented in **Section 5**. We provide several examples that show that the reanalysis faithfully captures small-scale structures in the constituent fields in the LS as compared with aircraft data from the PGS campaign and balloon-borne frost point hygrometers at Boulder, CO and Hilo, HI, USA, and Lauder, NZ. We also show that the composition of one of the plumes from PyroCb events associated with Australian New Year's bushfires is in good qualitative agreement with other studies, though the maximum moisture of the plume is underestimated in M2-SCREAM because many of the MLS observations of high water vapor concentrations were rejected by the CoDAS quality control. The reanalysis agrees well with ACE-FTS and SAGE III/ISS observations. The M2-SCREAM-satellite

differences are near-normally distributed with standard deviations smaller than those of the ACE-FTS data. Relative biases between the reanalysis and ACE-FTS reflect systematic differences between MLS and ACE-FTS. Comparisons with independent satellite data as well as a process-based analysis of the consistency of the assimilated constituent fields with the MERRA-2 dynamics and with large-scale processes documented in the literature demonstrate the utility of M2-SCREAM for scientific studies of chemical and transport variability on time scales ranging from hours to decades.

M2-SCREAM agrees well with the Belgian reanalysis BRAM2 that also assimilates MLS data. There are, however, several differences between the two reanalyses that, as we hope, make them complement each other from the standpoint of the user. BRAM2 assimilates several more species than M2-SCREAM does and its HNO_3 product is of higher quality in the polar regions. Some advantages of M2-SCREAM are a longer period covered by this reanalysis and higher horizontal resolution of the assimilated fields. In addition, along with the assimilated fields we provide estimates of the reanalysis uncertainties and data quality flags designed to guide the scientific use of this new reanalysis. We note that the latest release of BRAM2 includes standard deviations of the ensemble means of the assimilated species (Q. Errera, personal communication, 2022).

Several issues identified during the preparation, production and evaluation of M2-SCREAM will guide future work on chemical reanalyses at NASA's GMAO. Future work will include the development of a PSC scheme that will be fully compatible with assimilation of HNO_3, strategies for assimilating short-lived constituents (including upper-stratospheric and mesospheric ozone), and elimination of drifts in multidecadal data sets. A significant step of achieving the last goal will be assimilation of version 5 of MLS retrievals in the next composition reanalysis.

Appendix A. Calculation of analysis uncertainties

It is assumed that the analysis state x^a is normally distributed and unbiased with an uncertainty covariance matrix A. Similarly, the background state x^b is assumed unbiased with an uncertainty covariance matrix B. Under these assumptions, Desroziers et al. (2005) derived the following expressions for A and B (their equations (2) and (4)).

$$E(H[x^a - x^b][y - Hx^b]^T) = HBH^T$$
$$E(H[x^a - x^b][y - Hx^a]^T) = HAH^T.$$

Here E and T denotes average and transpose, respectively, and H is the (linear) observation operator. We use these formulas to estimate the background and analysis variances (the diagonal elements of A and B), $Var^b = \frac{N+1}{N}(\sigma^b)^2$ and $Var^a = \frac{N+1}{N}(\sigma^a)^2$ from the reanalysis' internal statistics (see equations (2) and (4) in Desroziers et al. (2005)). Here, N denotes the number of observations and σ^a and σ^b are the diagonal elements of HAH^T and HBH^T, respectively. The calculations are done using O-F, O-A and A-F residuals aggregated within $10^o \times 10^o$ longitude/latitude bins on the MLS pressure levels for each month of the reanalysis separately. Very rarely (a few percent of cases) this procedure produces negative variance estimates. In those cases, we set Var^b to $(\alpha E(x^b))^2$, where α is the proportionally coefficient for background uncertainties (**Table 1,** α=0.1 for N_2O), and we redefine $Var^a = MAX + \frac{N}{N+C}(MIN - MAX)$

with, MAX = Var^b, MIN=0.25MAX, and C=300. Additionally, if the analysis uncertainty variance estimate at any point is greater than the background uncertainty estimate, then we set the former to the latter. These choices represent our attempt to provide reasonable estimates in the rare cases where the Desroziers' procedure fails. The analysis and background variances are interpolated to the geolocation-dependent average model levels, mapped onto the horizontal grid of the reanalysis, and converted to standard deviations. Because the reanalysis output is obtained from analysis increments through IAU (**Section 3**) it is a linear combination of the background and analysis states. While the uncertainties of the two are, in principle, correlated, the conservative choice that we make here is to neglect the correlations and calculate the uncertainty of the assimilated output as

$$\tau^a = \sqrt{\frac{1}{2}[(\sigma^b)^2 + (\sigma^a)^2]}.$$

This quantity is provided in monthly NetCDF files alongside the reanalysis output. Also provided are additional uncertainties associated with the bug fix as described in **Appendix B**. We emphasize that τ^a is an estimate of the standard deviation of the reanalysis constituents' uncertainty conditioned on the assimilated data and derived under assumption of optimality. It, therefore, measures the CoDAS's "confidence" in the assimilation results given the prior (background) uncertainty distribution and the usual assumptions of zero bias and Gaussianity. It tends to be small compared to the constituents' mixing ratios. M**ovies M1–M3** show M2-SCREAM water vapor profiles along with error bars obtained by combining τ^a with the additional error estimate from **Appendix B**.

Appendix B. Vertical shift correction

A coding error, identified after the reanalysis was completed for the period 2004 – April 2020, resulted in an upward shift of the assimilated water vapor, HCl, HNO_3 and N_2O fields by half the model layer, or approximately 0.5 km. Formally, this error can be described as an erroneous application of a vertical shift operator, $(Sx^b)_i = x_{i+1/2}$ to the background constituent profile $x^b = (x_1, x_2, \dots x_k)^T$prior to analysis. The analysis state is given by $x^a = x^b + K(y - Sx^b)$, where y denotes observations, and $K = BS^T(SBS^T + R)^{-1}$ is the gain matrix. Here, B and R denote the background and observation uncertainty covariance matrices, respectively. Assuming that the analyses are sufficiently frequent, and the dynamics at any given level are sufficiently like those ½ grid point away from that level, that these analyses eventually drive the subsequent forecasts so much toward the observations that $x^b = S^T x^t + \varepsilon^b$, where the mean $E[(\varepsilon^b)(\varepsilon^b)^T] = B$ and x^t is the true state, it can be shown that the application of the correction Sx^a to x^a results in the optimal analysis state, i.e. the state one would obtain if the coding error were absent and the background uncertainty were given by $E[(S\varepsilon^b)(S\varepsilon^b)^T] = SBS^T$. We applied the correction above to the reanalysis output and tested it against a three-month long assimilation with a system where the initial error was eliminated ("correct analysis"). The relative bias between the corrected and the correct analyses is negligible. The difference standard deviations between the two runs result from interpolation errors and small departures from the assumptions given above and represent an additional uncertainty in the assimilated M2-SCREAM water vapor, HCl, HNO_3, and N_2O fields. We determined that these uncertainties are of similar magnitude to the assimilation uncertainties discussed in **Appendix A**. Therefore the recommended way to estimate the combined assimilation and additional uncertainties is to multiply the former by $\sqrt{2}$.

Appendix C. output provided to the users

The main reanalysis product is provided in the form of NetCDF files valid at 0, 3, 6, 9, 12, 15, 18, 21 UTC and contains instantaneous assimilated fields at a 0.625° longitude ×0.5°latitude horizontal resolution on 72 GEOS model layers. The layer center pressures and layer pressure thickness are also provided. Water vapor, HCl, HNO_3 and N_2O mixing ratios are given in mol/mol and ozone is provided in ppmv. In addition to the five assimilated constituents the output files also contain temperature, winds, cloud fraction, and PV from the GCM replay (in very close agreement with MERRA-2) and the following non-assimilated constituents: ClO, CO, and condensed nitric acid. These are not evaluated in this study and are provided for reference only. Neither are the tropospheric concentrations evaluated and not advised to use.

The assimilation uncertainties are given in monthly NetCDF files. The uncertainties are mapped onto the full latitude-longitude grid as that of the reanalysis output. Nominal three-dimensional pressure grid is also provided. Fill values indicate regions where no profile data are assimilated. In addition to the assimilation uncertainties (**Appendix A**), the uncertainty files for months prior to May 2021 also contain estimated additional uncertainties arising from the coding error and correction (**Appendix B**). These are calculated from the reanalysis and an additional assimilation experiment valid for March 2020. These fields are identical in all the uncertainty files. All uncertainties are provided as one-sigma.

Acknowledgments

KW, BW, GLM, and SEC acknowledge support for this research by NASA's Modeling, Analysis and Prediction (MAP) grant "A new look at stratospheric chemistry with multispecies chemical data assimilation". KEK and PAW acknowledge support by the NASA SAGE III/ISS Science Team grant 80NSSC21K1197. Work at the Jet Propulsion Laboratory, California Institute of Technology, was done under contract with NASA. Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation at Goddard Space Flight Center (GSFC). MERRA-2 is an official product of the Global Modeling and Assimilation Office at NASA GSFC, also supported by MAP. We thank the MLS team, especially Brian Knosp and Ryan Fuller, for data management and analysis support at JPL. We thank Dr. Dale Hurst and Dr. Sören Johansson for their help with FPH and GLORIA data, and Dr. Kaley Walker for detailed information on the use of ACE-FTS observations.

Open Research

M2-SCREAM (GMAO 2022a; 2022b) is available through the Goddard Earth Sciences Data and Information Services Center (https://disc.gsfc.nasa.gov) and can be accessed via these links: https://doi.org/10.5067/7PR3XRD6Q3NQ (assimilated fields) and https://doi.org/10.5067/7XRIJO9OP8PE (analysis uncertainties). ACE-FTS is the primary instrument on the SCISAT satellite, a Canadian-led mission mainly supported by the Canadian Space Agency. ACE-FTS data were downloaded from http://www.ace.uwaterloo.ca/data.php. SAGE-III/ISS data were downloaded from the NASA Atmospheric Science and Data Center (ASDC; https://eosweb.larc.nasa.gov/project/SAGE%20III-ISS. FPH observations were obtained from ftp://ftp.cmdl.noaa.gov/ozwv/WaterVapor/. GLORIA data are available at https://publikationen.bibliothek.kit.edu/1000086506.

References

Allen, D. R., and N. Nakamura (2001), A seasonal climatology of effective diffusivity in the stratosphere, *J. Geophys. Res*., 106, 7917–7935

Allen, D.R., Fromm, M.D., Kablick III, G.P., and Nedoluha, G.E. (2020), Smoke With Induced Rotation and Lofting (SWIRL) in the stratosphere, *J. Atmos. Sci.*, *77*, doi:10.1175/JAS-D-20-0131.1

Bergthórsson, P., Döös, B. R. (1955), Numerical weather map analysis. *Tellus* 7, 329–340

Bernath, P. F., McElroy, C. T., Abrams, M. C., Boone, C. D., Butler, M., Camy-Peyret, et al. (2005), Atmospheric Chemistry Experiment (ACE): Mission overview, *Geophys. Res. Lett*., 32, L15S01, https://doi.org/10.1029/2005GL022386

Bernath, P.F. (2017), The Atmospheric Chemistry Experiment (ACE), *Journal of Quantitative Spectroscopy and Radiative Transfer*, Volume 186 Pages 3-16, ISSN 0022-4073, https://doi.org/10.1016/j.jqsrt.2016.04.006

Bloom, S., Takacs, L., DaSilva, A., & Ledvina, D. (1996), Data assimilation using incremental analysis updates. *Mon. Wea. Rev*., 124, 1256–1271

Boone, C.D., Bernath, P.F., Cok, D., Jones, S.C., and Steffen, J. (2020), Version 4 retrievals for the atmospheric chemistry experiment Fourier transform spectrometer (ACE-FTS) and imagers, *Journal of Quantitative Spectroscopy and Radiative Transfer,* Volume 247, 106939, ISSN 0022-4073, https://doi.org/10.1016/j.jqsrt.2020.106939

Brasseur, G. P., and S. Solomon (2005), Aeronomy of the Middle Atmosphere, 3rd ed., *Springer,* New York

Burkholder, J. B. , Sander, S. P. , Abbatt, J. Barker, J. R., Huie, R. E. Kolb, C. E., Kurylo, M. J., Orkin, V. L., Wilmouth, D. M. and Wine, P. H. (2015), Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies, Evaluation No. 18. JPL Publication 15-10, Jet Propulsion Laboratory, Pasadena, 2015 http://jpldataeval.jpl.nasa.gov

Butchart, N. and Remsberg, E.E. (1986), The area of the stratospheric polar vortex as a diagnostic for tracer transport on an isentropic surface, *J. Atmos. Sci*., 43, 1319–1339

Butler, AH, Lawrence, ZD, Lee, SH, Lillo, SP, Long, CS (2020), Differences between the 2018 and 2019 stratospheric polar vortex split events. *Q J R Meteorol Soc*.; 146: 3503– 3521. https://doi.org/10.1002/qj.3858

Cohn, S. E. (1997), An introduction to estimation theory. *J. Meteor. Soc. Japan*, 75, 257–288

Considine, D. B., Douglass, A. R., Connell, P. S., Kinnison, D. E. and Rotman, D. A. (2000), A polar stratospheric cloud parameterization for the global modeling initiative three-dimensional model and its response to stratospheric aircraft, *J. Geophys. Res.,* 105(D3), 3955–3974

Coy, L., Eckermann, S., and Hoppel, K. (2009), Planetary wave breaking and tropospheric forcing as seen in the stratospheric sudden warming of 2006, *J. Atmos. Sci.,* 66, 495–507

Coy, L., and Pawson, S. (2015), The Major Stratospheric Sudden Warming of January 2013: Analyses and Forecasts in the GEOS-5 Data Assimilation System, *Monthly Weather Review*, *143*(2), 491-510

Davis, S. M., Hegglin, M. I., Fujiwara, M., Dragani, R., Harada, Y., Kobayashi, C. et al. (2017), Assessment of upper tropospheric and stratospheric water vapor and ozone in reanalyses as part of S-RIP,. *Atmos. Chem. Phys.*, 17, 12743-12778, https://doi.org/10.5194/acp-17-12743-2017

Desroziers, G., Berre, L., Chapnik, B. and Poli, P. (2005), Diagnosis of observation, background and analysis-error statistics in observation space. *Q.J.R. Meteorol. Soc*., 131: 3385-3396. https://doi.org/10.1256/qj.05.108

Dunkerton, T. J., & Delisi, D. P. (1986), Evolution of potential vorticity in the winter stratosphere of January-February 1979. *Journal of Geophysical Research*, *91*(D1), 1199–1208. https://doi.org/10.1029/JD091iD01p01199

Errera, Q., Chabrillat, S., Christophe, Y., Debosscher, J., Hubert, D., Lahoz, W., Santee, M. L., Shiotani, M., Skachko, S., von Clarmann, T., and Walker, K. (2019), Technical note: Reanalysis of Aura MLS chemical observations, *Atmos. Chem. Phys.,* 19, 13647–13679, https://doi.org/10.5194/acp-19-13647-2019

Friedl-Vallon, F., Gulde, T., Hase, F., Kleinert, A., Kulessa, T., Maucher, G., Neubert, T., Olschewski, F., Piesch, C., Preusse, P., Rongen, H., Sartorius, C., Schneider, H., Schönfeld, A., Tan, V., Bayer, N., Blank, J., Dapp, R., Ebersoldt, A., Fischer, H., Graf, F., Guggenmoser, T., Höpfner, M., Kaufmann, M., Kretschmer, E., Latzko, T., Nordmeyer, H., Oelhaf, H., Orphal, J., Riese, M., Schardt, G., Schillings, J., Sha, M. K., Suminska-Ebersoldt, O., and Ungermann, J. (2014), Instrument concept of the imaging Fourier transform spectrometer GLORIA, *Atmos. Meas. Tech*., 7, 3565–3577, https://doi.org/10.5194/amt-7-3565-2014

Flemming, J., Benedetti, A., Inness, A., Engelen, R. J., Jones, L., Huijnen, V., Remy, S., Parrington, M., Suttie, M., Bozzo, A., Peuch, V.-H., Akritidis, D., and Katragkou, E. (2017), The CAMS interim Reanalysis of Carbon

Monoxide, Ozone and Aerosol for 2003–2015, *Atmos. Chem. Phys*., 17, 1945–1983, https://doi.org/10.5194/acp-17-1945-2017

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L. et al., (2017), The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Climate, 30, 5419–5454,* https://doi.org/10.1175/JCLI-D-16-0758.1

Gilpin, S., Matsuo, T., and Cohn, S.E. (2022), Continuum covariance propagation for understanding variance loss in Advective systems. *SIAM/ASA J. on Uncertainty Quantification*, 10:3, 886-914, https://doi.org/10.1137/21M1442449

Global Modeling and Assimilation Office (GMAO)(2015), MERRA-2 inst3_3d_asm_Nv: 3d,3-Hourly,Instantaneous,Model-Level,Assimilation,Assimilated Meteorological Fields V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: **January 2021**, 10.5067/WWQSXQ8IVFW8

Global Modeling and Assimilation Office (GMAO) (2022), M2-SCREAM: 3d,3-Hourly,Instantaneous,Model-Level,Assimilation,Assimilated Constituent Fields,Replayed MERRA-2 Meteorological Fields, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: July 2022, doi:10.5067/7PR3XRD6Q3NQ

Global Modeling and Assimilation Office (GMAO) (2022), M2-SCREAM: Monthly,Model-Level,Assimilated Constituent Fields uncertainties, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: July, 2022. Doi:10.5067/7XRIJO9OP8PE

Harada, Y., Goto, A., Hasegawa, H., Fujikawa, N., Naoe, H., & Hirooka, T. (2010), A Major Stratospheric Sudden Warming Event in January 2009, *Journal of the Atmospheric Sciences*, *67*(6), 2052-2069

Haynes, P. and Shuckburgh, E. (2000a), Effective diffusivity as a diagnostic of atmospheric transport 1. Stratosphere, *J. Geophys. Res*., 105, 22777–22794

Haynes, P., and Shuckburgh, E. (2000b), Effective diffusivity as a diagnostic of atmospheric transport: 2. Troposphere and lower stratosphere, *J. Geophys. Res*., 105, 22,795–22,810

Hegglin, M. I., Boone, C. D., Manney, G. L., and Walker, K. A. (2009), A global view of the extratropical tropopause transition layer from Atmospheric Chemistry Experiment Fourier Transform Spectrometer O3, H2O, and CO, *J. Geophys. Res*., 114, D00B11, doi:10.1029/2008JD009984

Hegglin, M. I., et al. (2013), SPARC Data Initiative: Comparison of water vapor climatologies from international satellite limb sounders, *J. Geophys. Res. Atmos.*, 118, 11,824– 11,846, doi:10.1002/jgrd.50752

Hendon, H.H.,Thompson, D.W., Lim, E.-P., Butler, A.H., Newman, P.A., Coy, L et al. (2019), Rare forecasted climate event under way in the Southern Hemisphere, *Nature*, 573(7775), 495–495. doi:10.1038/d41586-019-02858-0

Hersbach, H, Bell, B, Berrisford, P, et al. (2020) The ERA5 global reanalysis. *Q J R Meteorol Soc*., 146: 1999– 2049. https://doi.org/10.1002/qj.3803

Hollingsworth, A., Engelen, R. J., Textor, C., Benedetti, A., Boucher, O., Chevallier, F., Dethof, A., Elbern, H., Eskes, H., Flemming, J., Granier, C., Kaiser, J. W., Morcrette, J.-J., Rayner, R., Peuch, V.-H., Rouil, L., Schultz, M. G., Simmons, A. J., and The GEMS Consortium (2008), Toward a Monitoring and Forecasting System For Atmospheric Composition: The GEMS Project, *B. Am. Meteorol. Soc*., 89, 1147–1164, https://doi.org/10.1175/2008BAMS2355.1

Hurst, D. F., Lambert, A., Read, W. G., Davis, S. M., Rosenlof, K. H., Hall, E. G., Jordan, A. F., and Oltmans, S. J. (2014), Validation of Aura Microwave Limb Sounder stratospheric water vapor measurements by the NOAA frost point hygrometer, *J. Geophys. Res. Atmos.*, 119, 1612– 1625, doi:10.1002/2013JD020757

Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., Clerbaux, C., Coheur, P., Engelen, R. J., Errera, Q., Flemming, J., George, M., Granier, C., Hadji-Lazaro, J., Huijnen, V., Hurtmans, D., Jones, L., Kaiser, J. W., Kapsomenakis, J., Lefever, K., Leitão, J., Razinger, M., Richter, A., Schultz, M. G., Simmons, A. J., Suttie, M., Stein, O., Thépaut, J.-N., Thouret, V., Vrekoussis, M., Zerefos, C., and the MACC team (2013), The MACC reanalysis: an 8 yr data set of atmospheric composition, *Atmos. Chem. Phys*., 13, 4073–4109, https://doi.org/10.5194/acp-13-4073-2013

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M. (2019), The CAMS reanalysis of atmospheric composition, *Atmos. Chem. Phys*., 19, 3515–3556, https://doi.org/10.5194/acp-19-3515-2019

Jazwinski, A. H. (1970), Stochastic processes and filtering theory. Academic Press, Inc. New York, New York.

Johansson, S., Woiwode, W., Höpfner, M., Friedl-Vallon, F., Kleinert, A., Kretschmer, E., Latzko, T., Orphal, J., Preusse, P., Ungermann, J., Santee, M. L., Jurkat-Witschas, T., Marsing, A., Voigt, C., Giez, A., Krämer, M., Rolf, C., Zahn, A., Engel, A., Sinnhuber, B.-M., and Oelhaf, H. (2018), Airborne limb-imaging measurements of temperature, HNO_3, O_3, $ClONO_2$, H_2O and CFC-12 during the Arctic winter 2015/2016: characterization, in situ validation and comparison to Aura/MLS, *Atmos. Meas. Tech*., 11, 4737–4756, https://doi.org/10.5194/amt-11-4737-2018

Kablick, G.P., Allen, D.R., Fromm, M.D., and Nedoluha, G.E. (2020), Australian PyroCb smoke generates synoptic-scale stratospheric anticyclones, *Geophys. Res. Lett.*, *47*, e2020GL088101, doi:10.1029/2020GL088101

Khaykin, S., Legras, B., Bucci, S. et al. (2020), The 2019/20 Australian wildfires generated a persistent smoke-charged vortex rising up to 35 km altitude. *Commun. Earth Environ., 1***,** 22, doi:10.1038/s43247-020-00022-5

Khosrawi, F., Kirner, O., Sinnhuber, B.-M., Johansson, S., Höpfner, M., Santee, M. L., Froidevaux, L., Ungermann, J., Ruhnke, R., Woiwode, W., Oelhaf, H., and Braesicke, P. (2017), Denitrification, dehydration and ozone loss during the 2015/2016 Arctic winter, *Atmos. Chem. Phys.,* 17, 12893–12910, https://doi.org/10.5194/acp-17-12893-2017

Lahoz, W. A. and Schneider, P. (2014), Data assimilation: making sense of Earth Observation, *Front. Environ. Sci*., 2, https://doi.org/10.3389/fenvs.2014.00016

Lawrence, Z.D., Perlwitz, J., Butler, A.H., Manney, G.L., Newman, P.A., Lee, S.H., and S.H., Nash, E.R. (2020), The remarkably strong Arctic stratospheric polar vortex of winter 2020: Links to record-breaking Arctic Oscillation and ozone loss, *J. Geophys. Res.*, *125*, e2020JD033271, doi:10.1029/2020JD033271

Levelt, P. F., van den Oord, G. H. J., Dobber, M. R., Mälkki, A., Visser, H., Vries, J. D., Stammes, P., Lundell, J. O. V., and Saari, H. (2006), The Ozone Monitoring Instrument, *IEEE Trans. Geosci. Remote Sens*., 44, 1093–1101, https://doi.org/10.1109/TGRS.2006.872333

Levelt, P. F., Joiner, J., Tamminen, J., Veefkind, J. P., Bhartia, P. K., Stein Zweers, D. C., Duncan, B., et al. (2018), The Ozone Monitoring Instrument: overview of 14 years in space, *Atmos. Chem. Phys*., 18, 5699-5745, https://doi.org/10.5194/acp-18-5699-2018

Livesey, N. J., Read, W. G., Wagner, P. A., Froidevaux, L., Lambert, A., Manney, G. L. et al. (2020) Version 4.2x Level 2 and 3 data quality and description document, JPL D-33509 Rev. E. Retrieved from https://mls.jpl.nasa.gov/data/v4-2_data_quality_document.pdf

Livesey, N. J., Read, W. G., Froidevaux, L., Lambert, A., Santee, M. L., Schwartz, M. J., Millán, L. F., Jarnot, R. F., Wagner, P. A., Hurst, D. F., Walker, K. A., Sheese, P. E., and Nedoluha, G. E. (2021), Investigation and amelioration of long-term instrumental drifts in water vapor and nitrous oxide measurements from the Aura Microwave Limb Sounder (MLS) and their implications for studies of variability and trends, *Atmos. Chem. Phys.*, 21, 15409–15430, https://doi.org/10.5194/acp-21-15409-2021

Livesey, N. J., Read, W. G., Wagner, P. A., Froidevaux, L., Santee, M. L., Schwartz, M. J. et al. (2022), Version 5.0x Level 2 and 3 data quality and description document, JPL D-105336 Rev. B. Retrieved from https://mls.jpl.nasa.gov/data/v5-0_data_quality_document.pdf

Manney, G.L., Santee, M.L., Livesey, N.J., Froidevaux, L., Pumphrey, H.C., Read, and W.G., Waters,J.W. (2005), EOS Microwave Limb Sounder Observations of the Antarctic Polar Vortex Breakup in 2004, *Geophys. Res. Lett.*, 32, L12811, doi:10.1029/2005GL022823

Manney, G. L., et al. (2008), The evolution of the stratopause during the 2006 major warming: Satellite data and assimilated meteorological ana- lyses, *J. Geophys. Res*., 113, D11115, doi:10.1029/2007JD009097

Manney, G.L., et al. (2009a), Satellite Observations and Modelling of Transport in the Upper Troposphere through the Lower Mesosphere During the 2006 Major Stratospheric Sudden Warming, *Atmos. Chem. Phys*., 9, 4775–4795

Manney, G. L., Schwartz, M. J., Krüger, K., Santee, M. L., Pawson, S., Lee, J. N., Daffer, W. H., Fuller, R. A., and Livesey, N. J. (2009b), Aura Microwave Limb Sounder observations of dynamics and transport during the record-breaking 2009 Arctic stratospheric major warming, *Geophys. Res. Lett.*, 36, L12815, doi:10.1029/2009GL038586

Manney, G.L., Santee, M.L., Rex, M., Livesey, N.J., Pitts, M.L.C., Veefkind, P. et al. (2011), Unprecedented Arctic ozone loss in 2011, *Nature*, *478*, 469–475, doi:10.1038/nature10556

Manney, G. L. and Lawrence, Z. D. (2016), The major stratospheric final warming in 2016: dispersal of vortex air and termination of Arctic chemical ozone loss, *Atmos. Chem. Phys*., 16, 15371–15396, https://doi.org/10.5194/acp-16-15371-2016

Manney, G. L., Livesey, N. J., Santee, M. L., Froidevaux, L., Lambert, A., & Lawrence, Z. D., et al. (2020), Record-low Arctic stratospheric ozone in 2020: MLS observations of chemical processes and comparisons with

previous extreme winters. *Geophysical Research Letters*, 47, e2020GL089063. https://doi.org/10.1029/2020GL089063

McIntyre, M.E. and Palmer, T.N. (1983), Breaking planetary waves in the stratosphere. *Nature*, 305, 593–600

Ménard, R., Skachko, S., and Pannekoucke, O. (2021), Numerical discretization causing error variance loss and the need for inflation, *Q. J. Roy. Meteor. Soc.*, 147, 3498-3520, https://doi.org/10.1002/qj.4139

Miyazaki, K., Bowman, K., Sekiya, T., Eskes, H., Boersma, F., Worden, H., Livesey, N., Payne, V. H., Sudo, K., Kanaya, Y., Takigawa, M., and Ogochi, K. (2020), Updated tropospheric chemistry reanalysis and emission estimates, TCR-2, for 2005–2018, *Earth Syst. Sci. Data*, 12, 2223–2259, https://doi.org/10.5194/essd-12-2223-2020

Nakamura, N. (1996), Two-Dimensional Mixing, Edge Formation, and Permeability Diagnosed in an Area Coordinate, *Journal of Atmospheric Sciences*, *53*(11), 1524-1537

Nichols, N. K. (2010), Mathematical concepts of data assimilation, in *Data Assimilation: Making Sense of Observations*, eds W. A. Lahoz, B. Khattatov, and R. Ménard (Berlin; Heidelberg: Springer), 13–39

Nielsen, J. E., Pawson, S., Molod, A., Auer, B., da Silva, A. M., Douglass, A. R., et al. (2017), Chemical mechanisms and their applications in the Goddard Earth Observing System (GEOS) earth system model. *J. Adv. Mod. Earth Sys.,* 9, 3019– 3044. https://doi.org/10.1002/2017MS001011

Orbe, C., Oman, L. D., Strahan, S. E., Waugh, D. W., Pawson, S., Takacs, L. L., & Molod, A. M. (2017), Large-scale atmospheric transport in GEOS replay simulations. *J. Adv. Mod. Earth Sys.*, 9. https://doi. org/10.1002/2017MS001053

Park, M., Randel, W. J., Damadeo, R. P., Flittner, D. E., Davis, S. M., Rosenlof, K. H., et al. (2021), Near-global variability of stratospheric water vapor observed by SAGE III/ISS. *J. of Geoph. Res.: Atmospheres*, 126, e2020JD034274. https://doi.org/10.1029/2020JD034274

Plumb, R. A. (2007), Tracer interrelationships in the stratosphere, *Rev. Geophys.*, 45, RG4005, doi:10.1029/2005RG000179

Randel, W. and Park, M. (2019), Diagnosing observed stratospheric water vapor relationships to the cold point tropical tropopause. *J. of Geoph. Res.: Atmospheres*, 124, 7018– 7033. https://doi.org/10.1029/2019JD030648

Reich, S. (2019), Data assimilation: the Schrödinger perspective. *Acta Numerica*, 28, 635-711

Rieger, L. A., Randel, W. J., Bourassa, A. E., & Solomon, S. (2021), Stratospheric temperature and ozone anomalies associated with the 2020 Australian New Year Fires. *Geoph. Res. Lett.*, 48, e2021GL095898. https://doi.org/10.1029/2021GL095898

Safieddine, S., Bouillon, M.,. Paracho, A.-C., Jumelet, J., Tencé, R., Pazmino, A., Goutail, F., Wespes, C., Bekki, S., Boynard, A., Hadji-Lazaro, J., Coheur, P.-F., Hurtmans, D., and Clerbaux, C. (2020), Antarctic ozone enhancement during the 2019 sudden stratospheric warming event, *Geophys. Res. Lett., 47*, e2020GL087810, doi:10.1029/2020GL087810

Santee, M.L., Manney, G.L., Livesey, N.J., Froidevaux, L., Schwartz, M.J., and Read, W.G. (2011), Trace gas evolution in the lowermost stratosphere from Aura Microwave Limb Sounder measurements, *J. Geophys. Res*., 116, D18306, doi:10.1029/2011JD015590

Santee, M. L., Lambert, A., Manney, G. L., Livesey, N. J., Froidevaux, L., Neu, J. L., et al. (2022), Prolonged and pervasive perturbations in the composition of the Southern Hemisphere midlatitude lower stratosphere from the Australian New Year's fires. *Geoph. Res. Lett*., 49, e2021GL096270. https://doi.org/10.1029/2021GL096270

Schwartz, M. J., Manney, G. L., Hegglin, M. I. , Livesey, N. J., Santee, M. L.., and Daffer, W. H. (2015), Climatology and variability of trace gases in extratropical double-tropopause regions from MLS, HIRDLS, and ACE-FTS measurements, *J. Geophys. Res. Atmos.*, *120*, doi:10.1002/2014JD021964

Schwartz, M.J., Santee, M.L., Pumphrey, H.C., Manney, G.L. , Lambert, A.,Livesey, N.J., et al. (2020), Australian New Year's pyroCb impact on stratospheric composition, *Geophy. Res. Lett.*, *47*, e2020GL090831, doi:10.1029/2020GL090831

Sheese, P.E., Boone, C.D., and Walker, K.A. (2015), Detecting physically unrealistic outliers in ACE-FTS atmospheric measurements, *Atmos. Meas. Tech*., 8, 741–750, 2015. doi: doi: 10.5194/amt-8-741-2015

Sheese, P. E., Walker, K. A., Boone, C. D., Bernath, P. F., Froidevaux, L., Funke, B., Raspollini, P., and von Clarmann, T. (2017), ACE-FTS ozone, water vapour, nitrous oxide, nitric acid, and carbon monoxide profile comparisons with MIPAS and MLS, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 186, 63-80, doi:10.1016/j.jqsrt.2016.06.026

Sheese, P.; Walker, K. (2020), "Data Quality Flags for ACE-FTS Level 2 Version 4.1/4.2 Data Set, https://doi.org/10.5683/SP2/BC4ATC, *Scholars Portal Dataverse*, V15

Shepherd, T. G. (2007), Transport in the middle atmosphere, *J. Meteorol. Soc. Japan*, 85B, 165– 191

Skachko, S., Ménard, R., Errera, Q., Christophe, Y., and Chabrillat, S. (2016), EnKF and 4D-Var data assimilation with chemical transport model BASCOE (version 05.06), *Geosci. Model Dev*., 9, 2893–2908, https://doi.org/10.5194/gmd-9-2893-2016

SPARC Reanalysis Intercomparison Project (S-RIP) (2021), Final Report. Masatomo Fujiwara, Gloria L. Manney, Lesley J. Gray, and Jonathon S. Wright (Eds.), *SPARC Report No. 10*, WCRP-6/2021, doi: 10.17874/800dee57d13, available at www.sparc-climate.org/publications/sparc-reports

Todling, R., and El Akkraoui, A. (2018), The GMAO Hybrid Ensemble-Variational Atmospheric Data Assimilation System: Version 2.0. *NASA Technical Report Series on Global Modeling and Data Assimilation*, NASA/TM-2018-104606, 50. https://gmao.gsfc.nasa.gov/pubs/docs/Todling1019.pdf

van der A, R. J., Allaart, M. A. F., and Eskes, H. J. (2015), Extended and refined multi sensor reanalysis of total ozone for the period 1970–2012, *Atmos. Meas. Tech*., 8, 3021–3035, https://doi.org/10.5194/amt-8-3021-2015

Wang, H. J. R., Damadeo, R., Flittner, D., Kramarova, N., Taha, G., Davis, S., et al. (2020), Validation of SAGE III/ISS solar occultation ozone products with correlative satellite and ground based measurements. *J. of Geoph. Res. Atmos.,* 125, e2020JD032430. https://doi.org/10.1029/2020JD032430

Wargan, K., Pawson, S., Olsen, M. A., Witte, J. C., Douglass, A. R., Ziemke, J. R., Strahan, S. E. & Nielsen, J. E. (2015), The global structure of upper troposphere-lower stratosphere ozone in GEOS-5: A multiyear assimilation of EOS Aura data. *J. Geophys. Res. Atmos*., 120: 2013– 2036. https://doi.org/10.1002/2014JD022493

Wargan, K., Labow, G., Frith, S., Pawson, S., Livesey, N., & Partyka, G. (2017), Evaluation of the Ozone Fields in NASA's MERRA-2 Reanalysis. *J. Climate*, 30, 2961–2988, https://doi.org/10.1175/JCLI-D-16-0699.1

Wargan, K., Kramarova, N., Weir, B., Pawson, S., & Davis, S. M. (2020a), Toward a reanalysis of stratospheric ozone for trend studies: Assimilation of the Aura microwave limb sounder and ozone mapping and profiler suite limb profiler data. *Journal of Geophysical Research: Atmospheres*, **125**, e2019JD031892. https://doi.org/10.1029/2019JD031892

Wargan, K., Weir, B., Manney, G. L., Cohn, S. E., & Livesey, N. J. (2020b), The anomalous 2019 Antarctic ozone hole in the GEOS Constituent Data Assimilation System with MLS observations. *Journal of Geophysical Research: Atmospheres*, 125, e2020JD033335. https://doi.org/10.1029/2020JD033335

Waters, J. W., Froidevaux, L., Harwood, R.S., Jarnot, R.F., Pickett, H.M., Read, W.G. et al. (2006), The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura satellite. *IEEE Trans. Geosci. Remote Sens*., 44, 1075–1092. https://doi.org/10.1109/TGRS.2006.873771

Weir, B., Miller,R. N., and Spitz, Y. H. (2013), Implicit Estimation of Ecological Model Parameters. *Bulletin of Mathematical Biology,* 75 (2): 223-257. https://10.1007/s11538-012-9801-6

Weir, B., Crisp, D., O'Dell, C. W. et al. (2021), Regional impacts of COVID-19 on carbon dioxide detected worldwide from space. *Science Advances*, 7 (45): doi:10.1126/sciadv.abf9415

Wohltmann, I., von der Gathen, P., Lehmann, R., Maturilli, M., Deckelmann, H., Manney, G. L., et al. (2020), Near-complete local reduction of Arctic stratospheric ozone by severe chemical loss in spring 2020. *Geoph. Res. Lett.*, *47*, e2020GL089547. https://doi.org/ 10.1029/2020GL089547

Wohltmann, I., von der Gathen, P., Lehmann, R., Deckelmann, H., Manney, G. L., Davies, J., et al. (2021), Chemical evolution of the exceptional Arctic stratospheric winter 2019/2020 compared to previous Arctic and Antarctic winters. *J. of Geoph. Res. Atmos.*, *126*, e2020JD034356. https://doi. org/10.1029/2020JD034356

Wu, W.S., Purser, R.J., Parrish, D.F. (2002), Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Weather Rev.* 130, 2905–2916

Ziemke, J. R., et al. (2014) Assessment and applications of NASA ozone data products derived from Aura OMI/MLS satellite measurements in context of the GMI chemical transport model, J. *Geophys. Res. Atmos*., 119, doi:10.1002/2013JD020914

Journal of Advances in Modeling Earth Systems

Supporting Information for

M2-SCREAM: A Stratospheric Composition Reanalysis of Aura MLS data with MERRA-2 transport

Krzysztof Wargan[1,2], Brad Weir[2,3,†], Gloria L. Manney[4,5], Stephen E. Cohn[2], K. Emma Knowland[2,3,†], Pamela A. Wales[2,3,†], and Nathaniel J. Livesey[6]

[1] Science Systems and Applications Inc., Lanham, Maryland, USA

[2] Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

[3] Universities Space Research Association, Columbia, Maryland, USA

[4] NorthWest Research Associates, Socorro, New Mexico, USA

[5] Department of Physics, New Mexico Institute of Mining and Technology, Socorro, New Mexico, USA

[6] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA

[†] Now Morgan State University, Baltimore, Maryland, USA

Corresponding author: Krzysztof Wargan (krzysztof.wargan-1@nasa.gov)

Contents of this file

Figures S1 to S18

Additional Supporting Information (Files uploaded separately)

Captions for Movies S1 to S3

Introduction

This file contains supplementary figures and movie captions for "M2-SCREAM: A Stratospheric Composition Reanalysis of Aura MLS data with MERRA-2 transport". Figure S1 shows observation minus forecast statistics for water vapor, HCl, HNO_3, and N_2O, analogous to Figure 3a and b in the main text. Figure S2 plots global zonal mean comparisons between M2-SCREAM and BRAM2. Figures S3–S16 show comparisons of M2-SCREAM HNO3, H2O, and ozone with GLORIA data from individual flights during the joint Polar Stratosphere in a Changing Climate, Gravity Wave Life Cycle Experiment, and Seasonality of Air mass transport and origin in the Lowermost Stratosphere using the HALO Aircraft campaigns. These are analogous to Figure 6 in the main text. Figures S17 and S18 show statistical comparisons of M2-SCREAM with ACE-FTS as in Figure 10 but limited to 30ºN–60ºNnand for December–January 2005–2020. Animations S1-S3show all available frost point hygrometer water vapor profile observations between 2005 and 2021 and collocated M2-SCREAM profiles. These are discussed in Section 6.1 of the main text. Figure 7 shows statistical comparisons calculated using the data displayed in the animations.

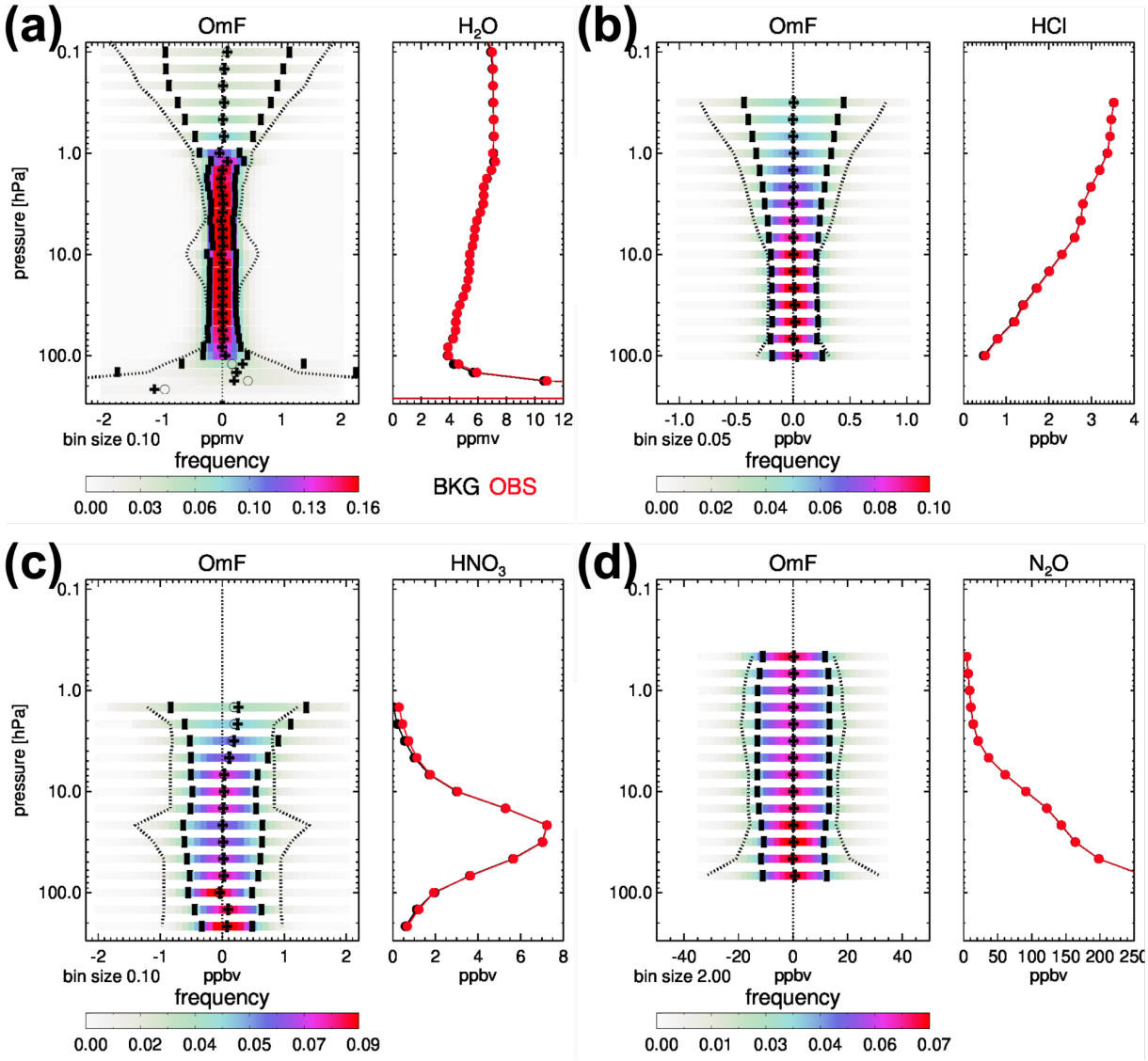

Figure S1: Internal global SCREAM statistics for H_2O (a), HCl (b), HNO_3 ©, and N_2O (d). The right-hand side panels show the mean background and MLS profiles. Observation minus forecast (O-F) statistics are plotted in the left-hand side panels. Shown are the O-F mean (plus signs), median (open circles), standard deviations around the mean (short vertical bars), probability density functions (colors) at prescribed pressure levels. The dotted lines are plus/minus MLS uncertainty estimates. Valid for January 2005.

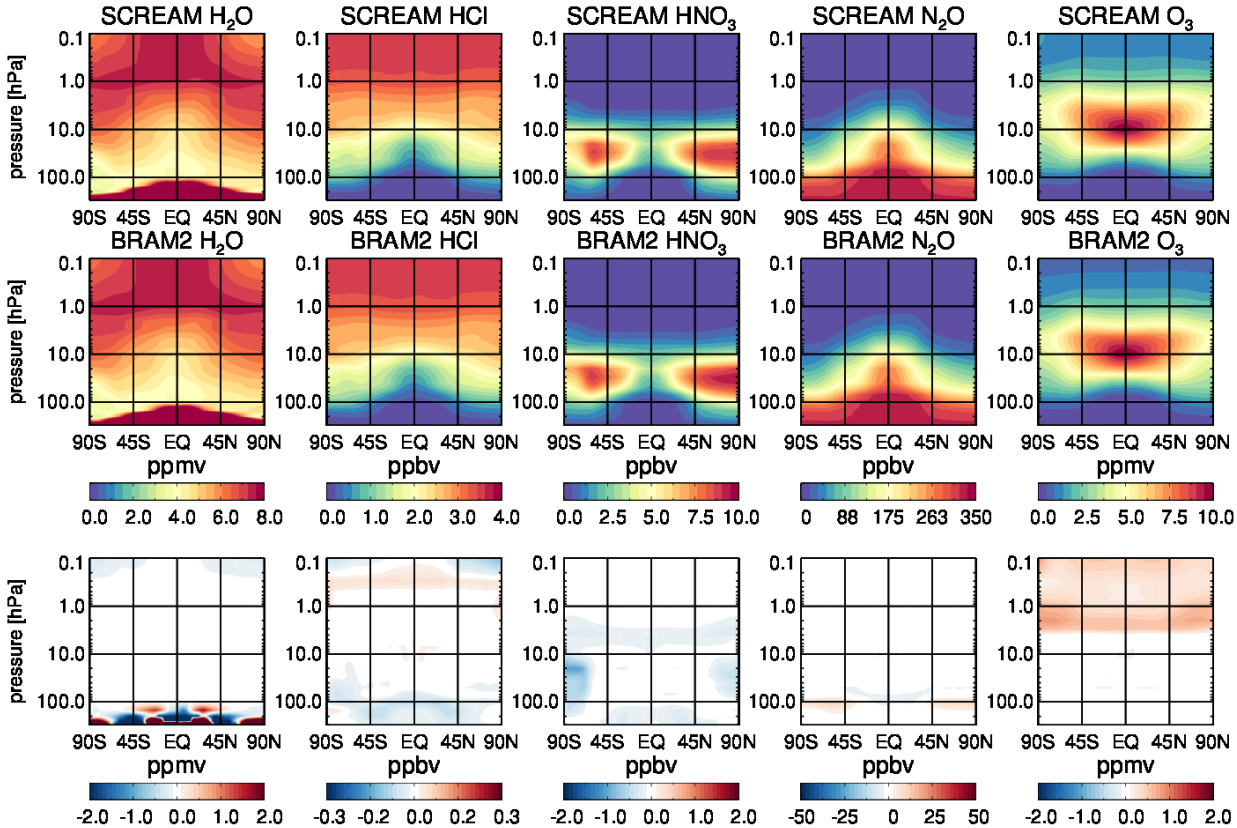

Figure S2: Zonal mean mixing ratios of the assimilated constituents from the BRAM2 reanalysis (a) and SCREAM (b). The SCREAM minus BRAM2 differences are shown in (c).

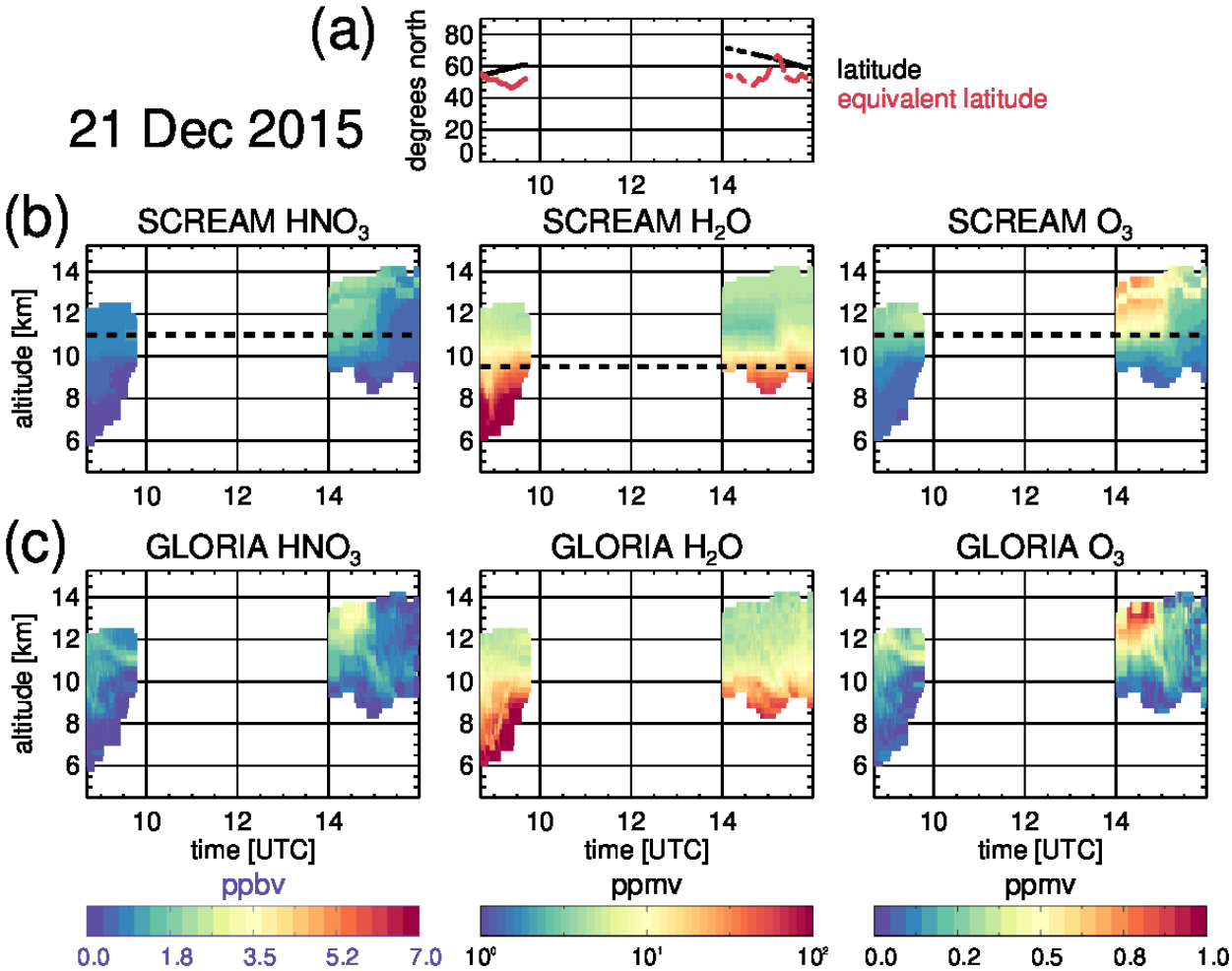

Figure S3: (a) *Latitudes and equivalent latitudes of the GLORIA measurements at 10 km. (b) M2-SCREAM HNO_3, water vapor, and ozone collocated with GLORIA measurements during a single flight on 21 December 2015. The dashed lines mark the lowest altitudes of MLS observations assimilated in M2-SCREAM. (c) GLORIA observations.*

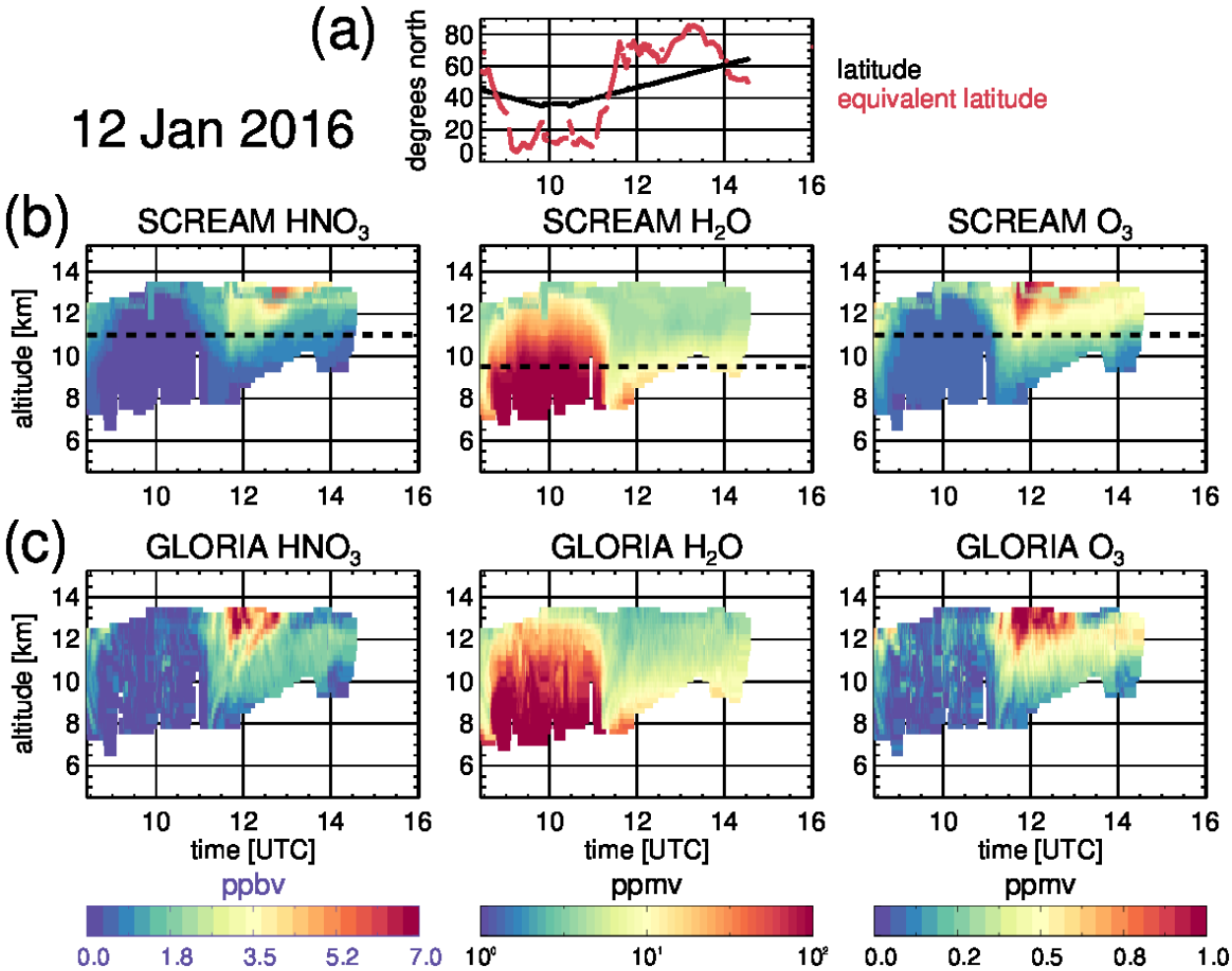

Figure S4: As in Fig. S3 but for 12 January 2016.

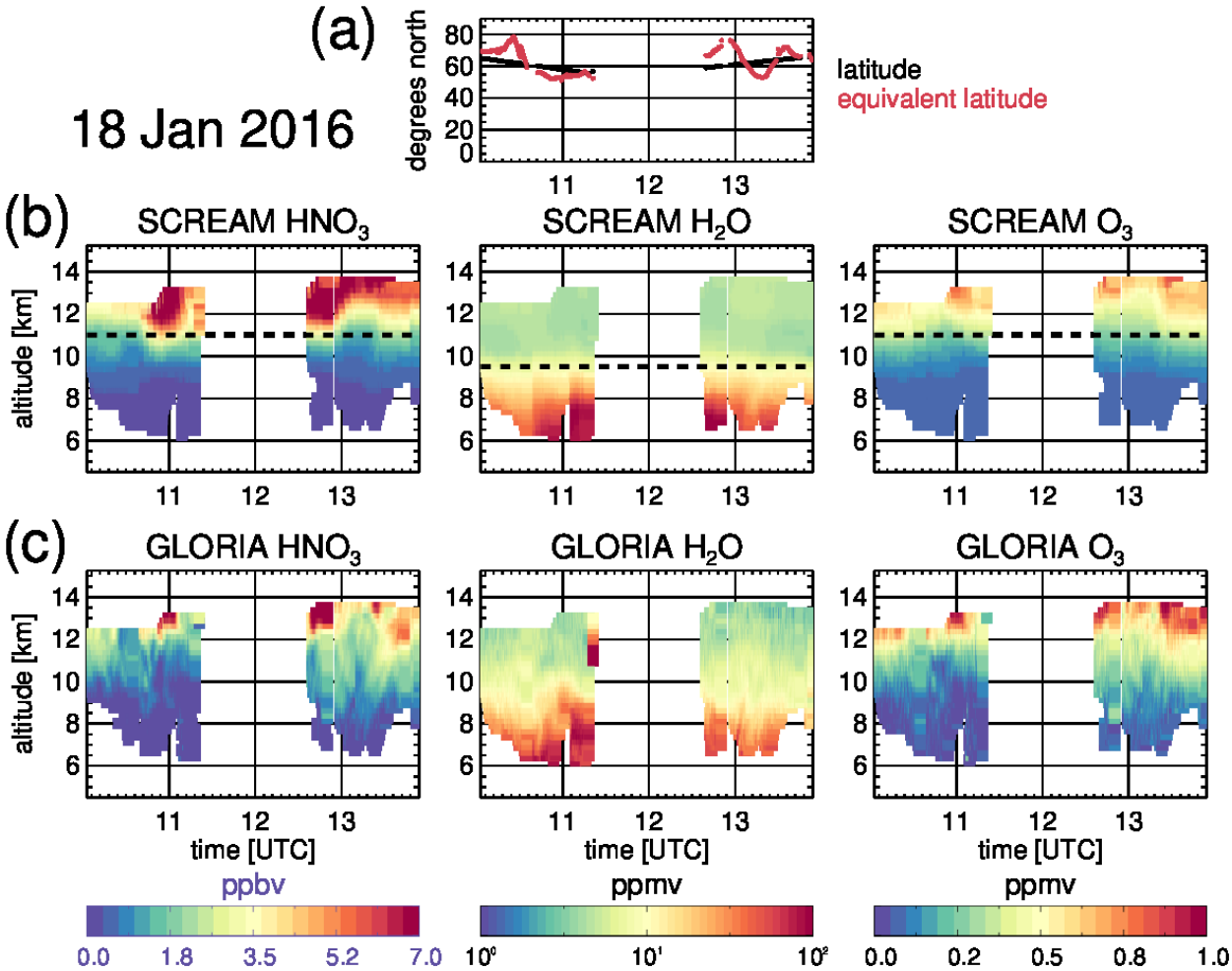

Figure S5: As in Fig. S3 but for 18 January 2016.

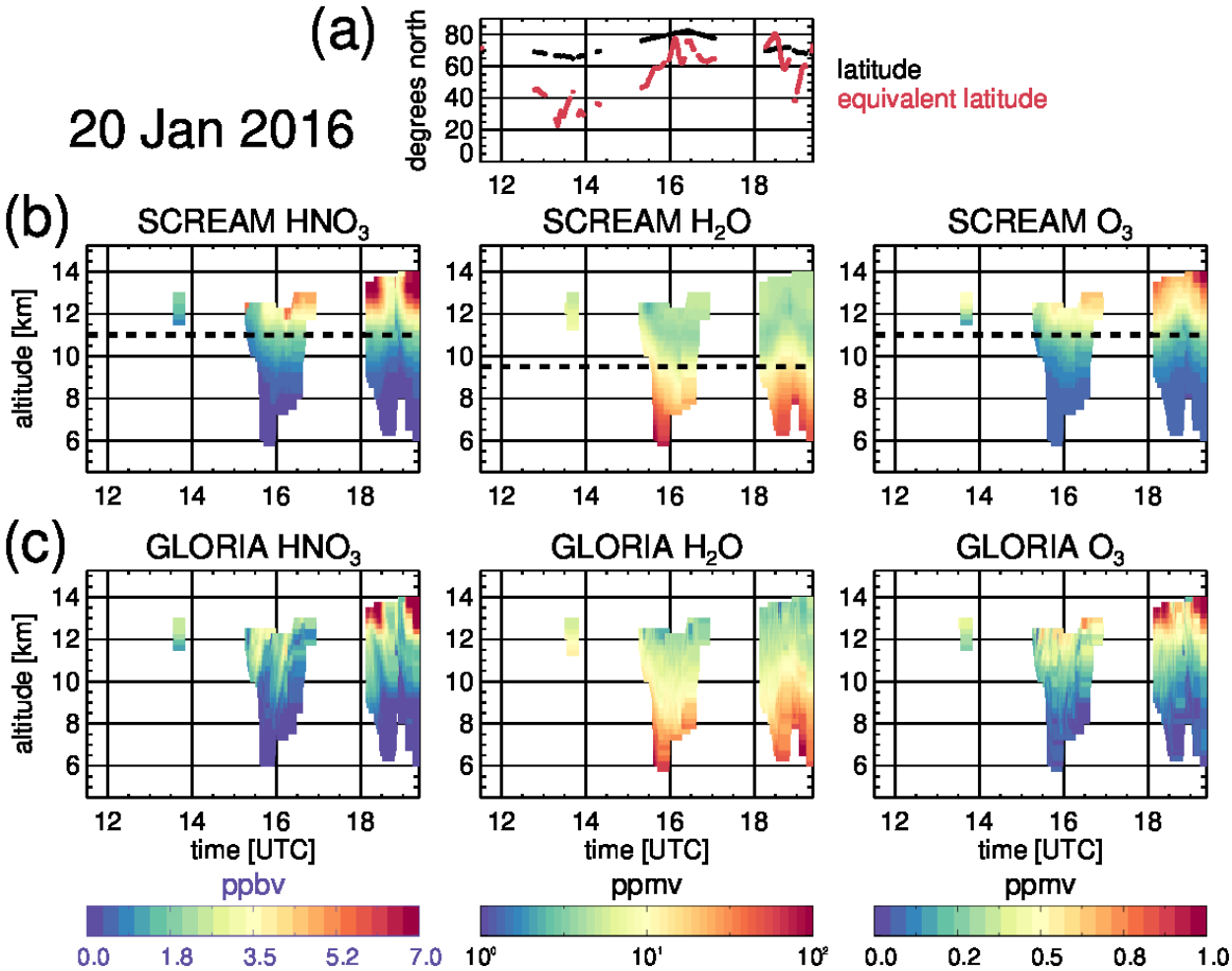

Figure S6: As in Fig. S3 but for 20 January 2016.

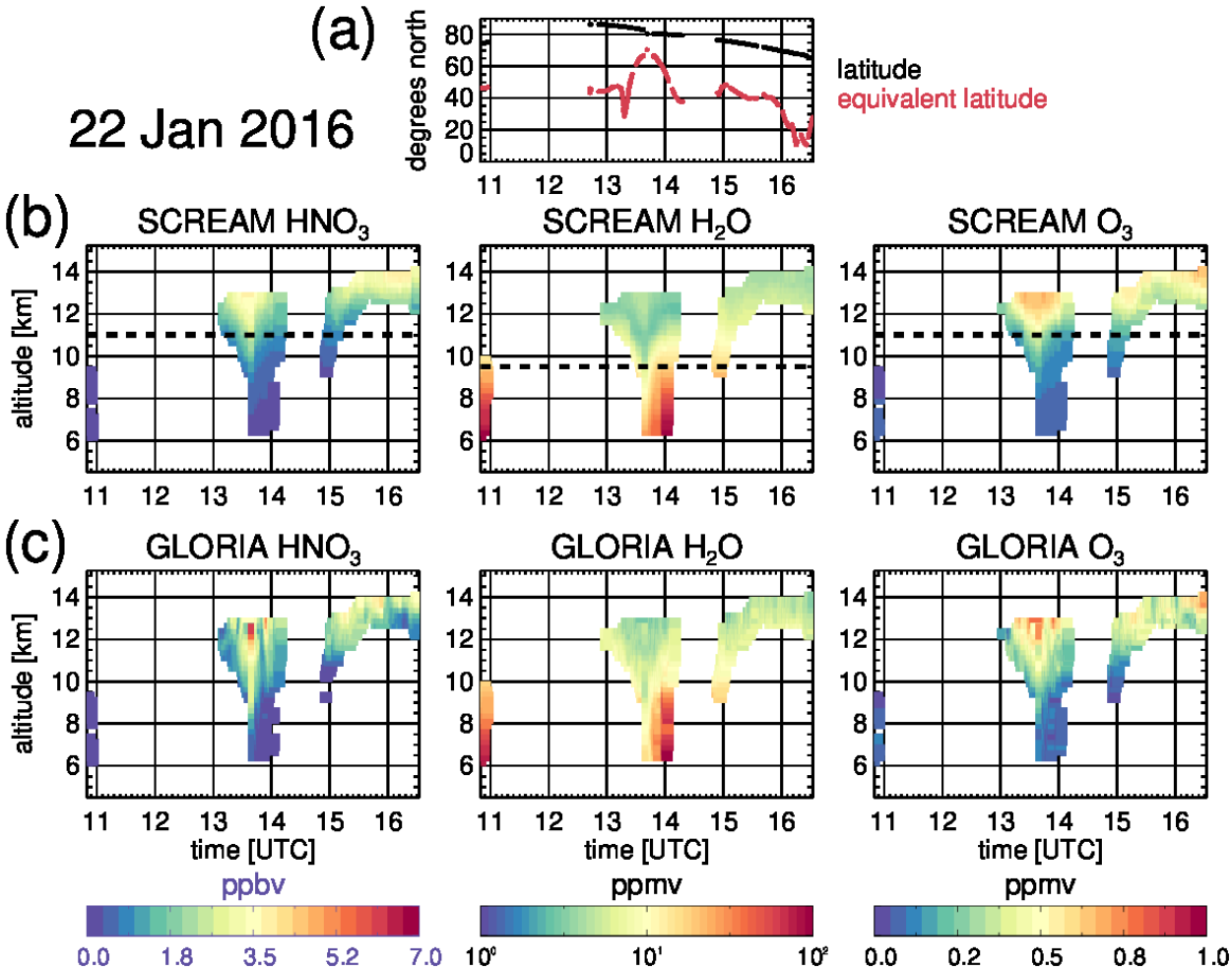

Figure S7: As in Fig. S3 but for 22 January 2016.

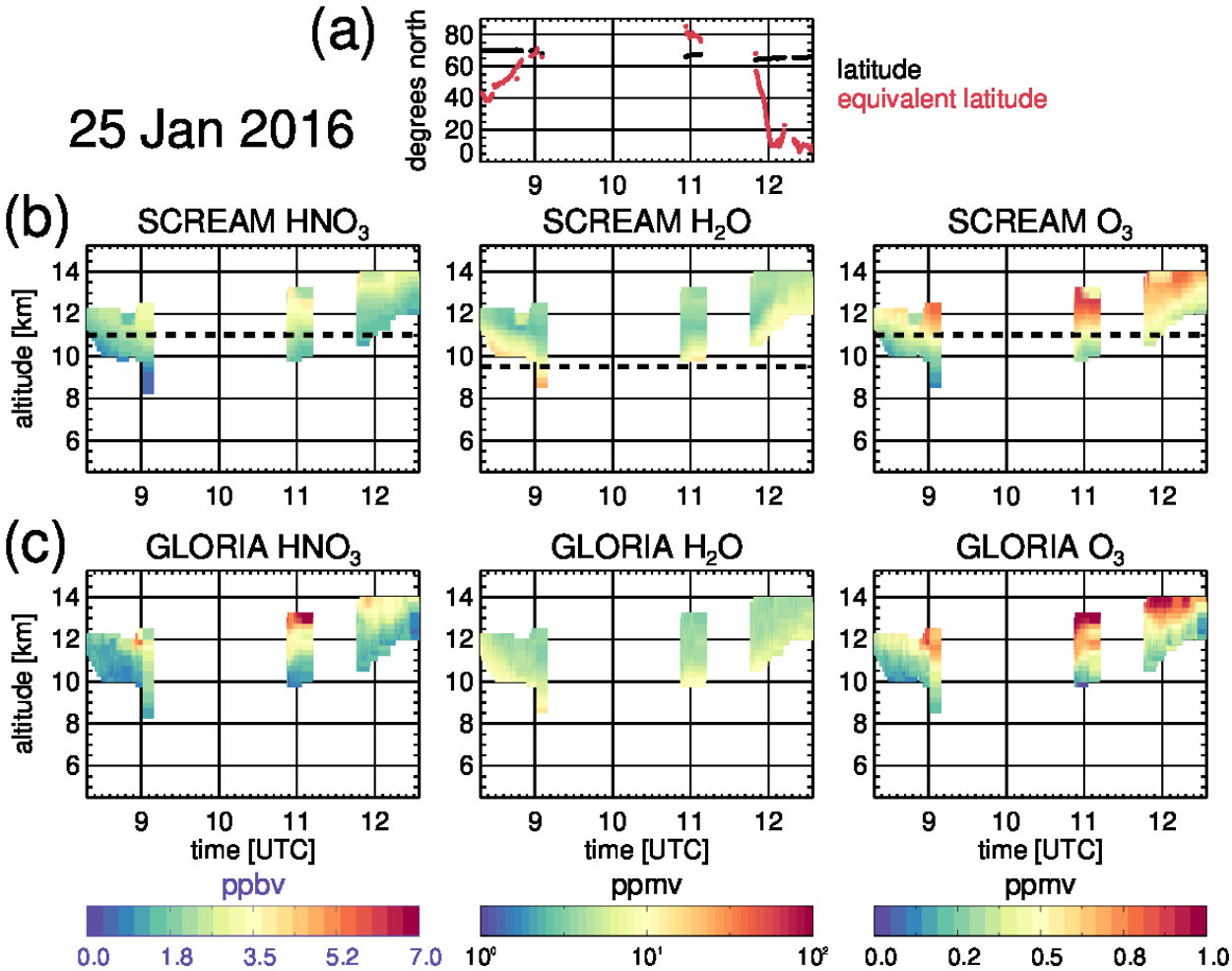

Figure S8: As in Fig. S3 but for 25 January 2016.

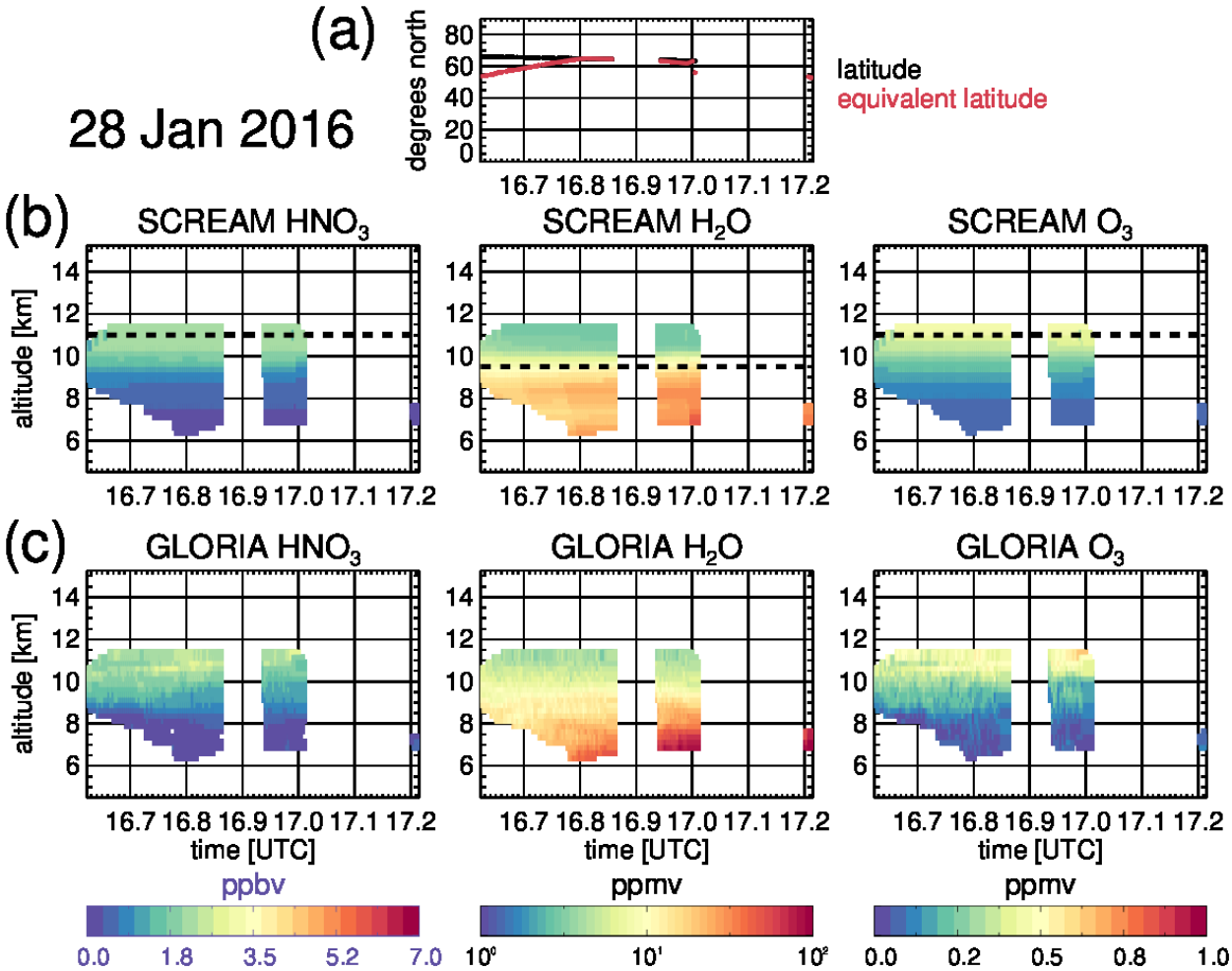

Figure S9: As in Fig. S3 but for 28 January 2016.

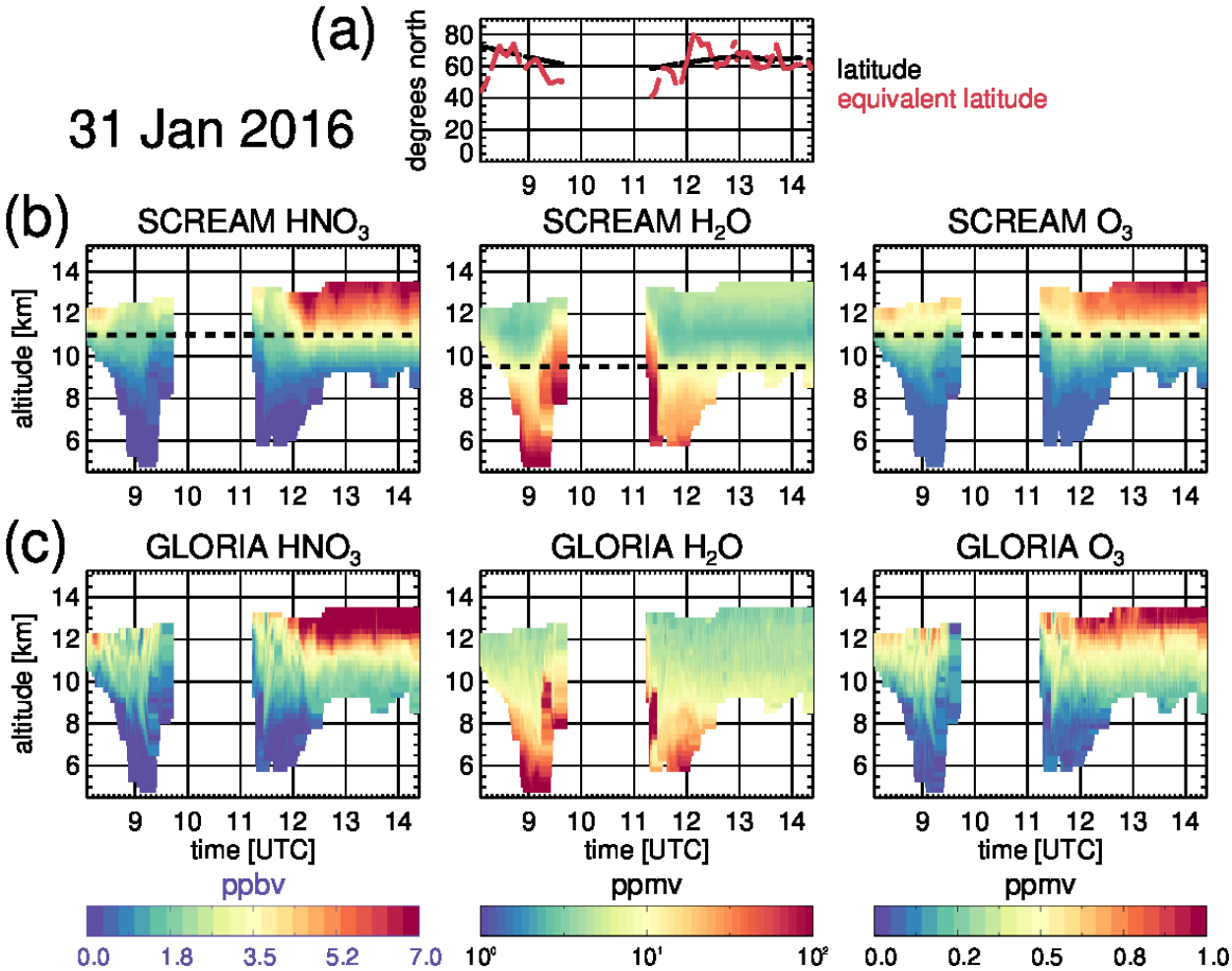

Figure S10: As in Fig. S3 but for 31 January 2016.

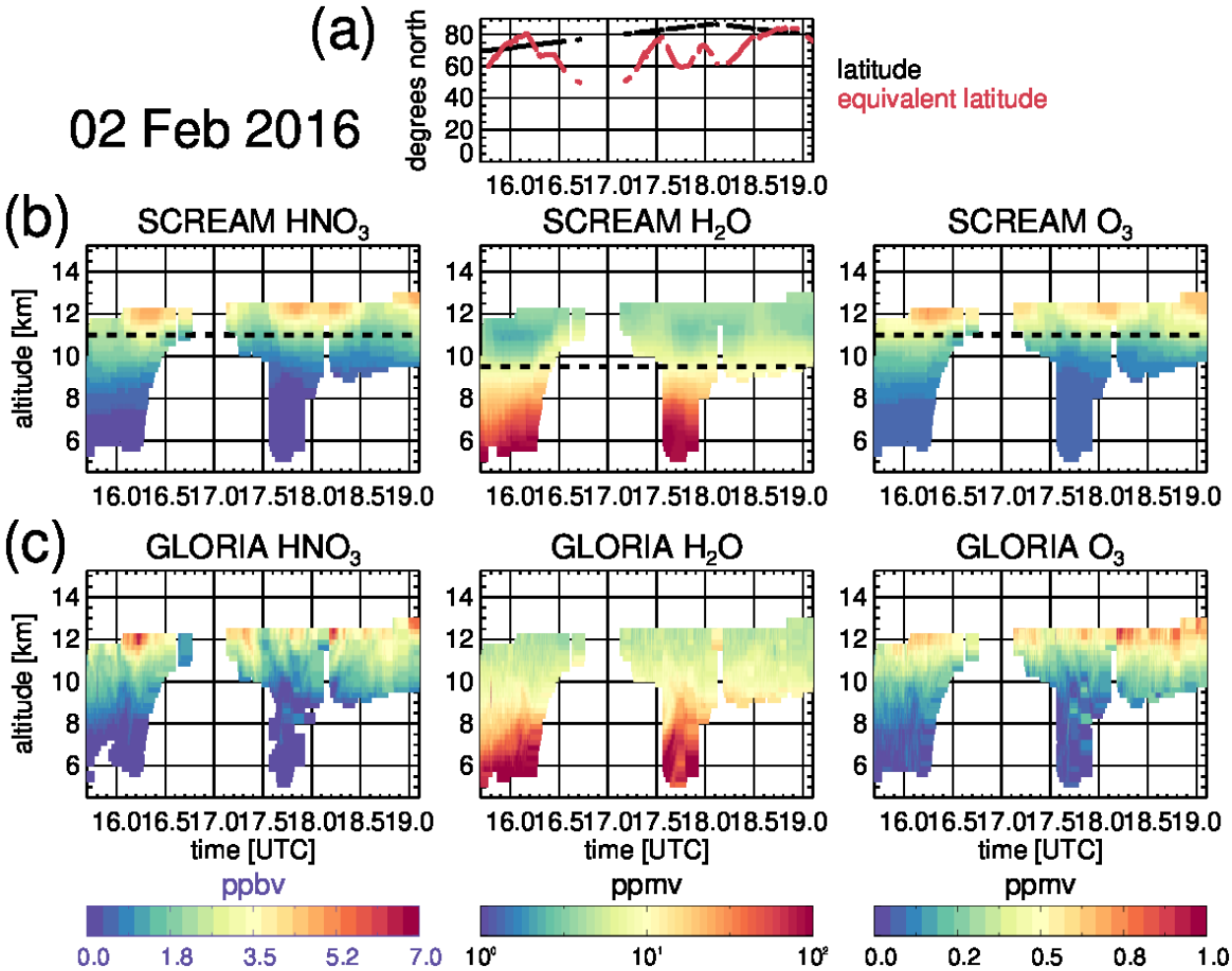

Figure S11: As in Fig. S3 but for 2 February 2016.

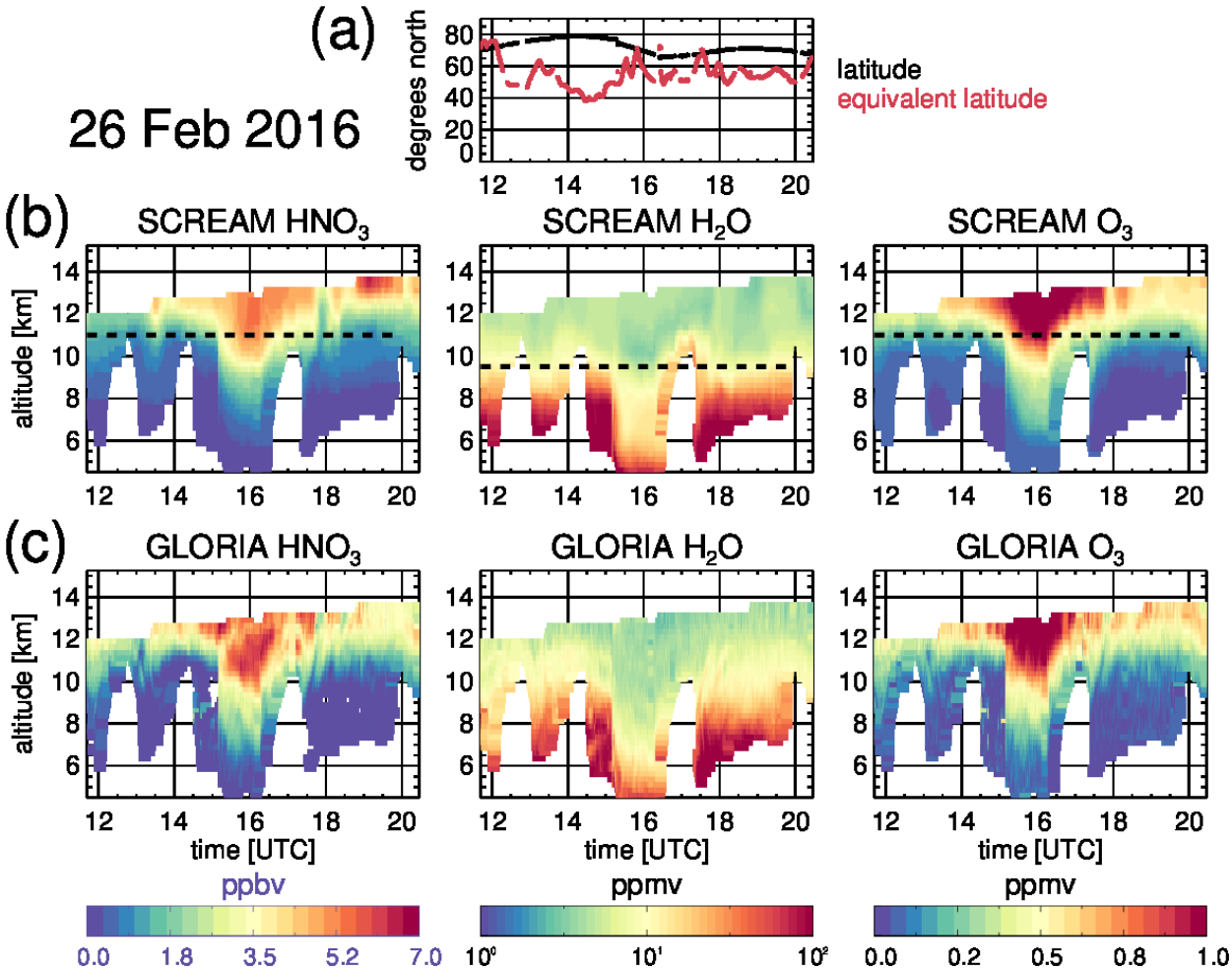

Figure S12: As in Fig. S3 but for 26 February 2016.

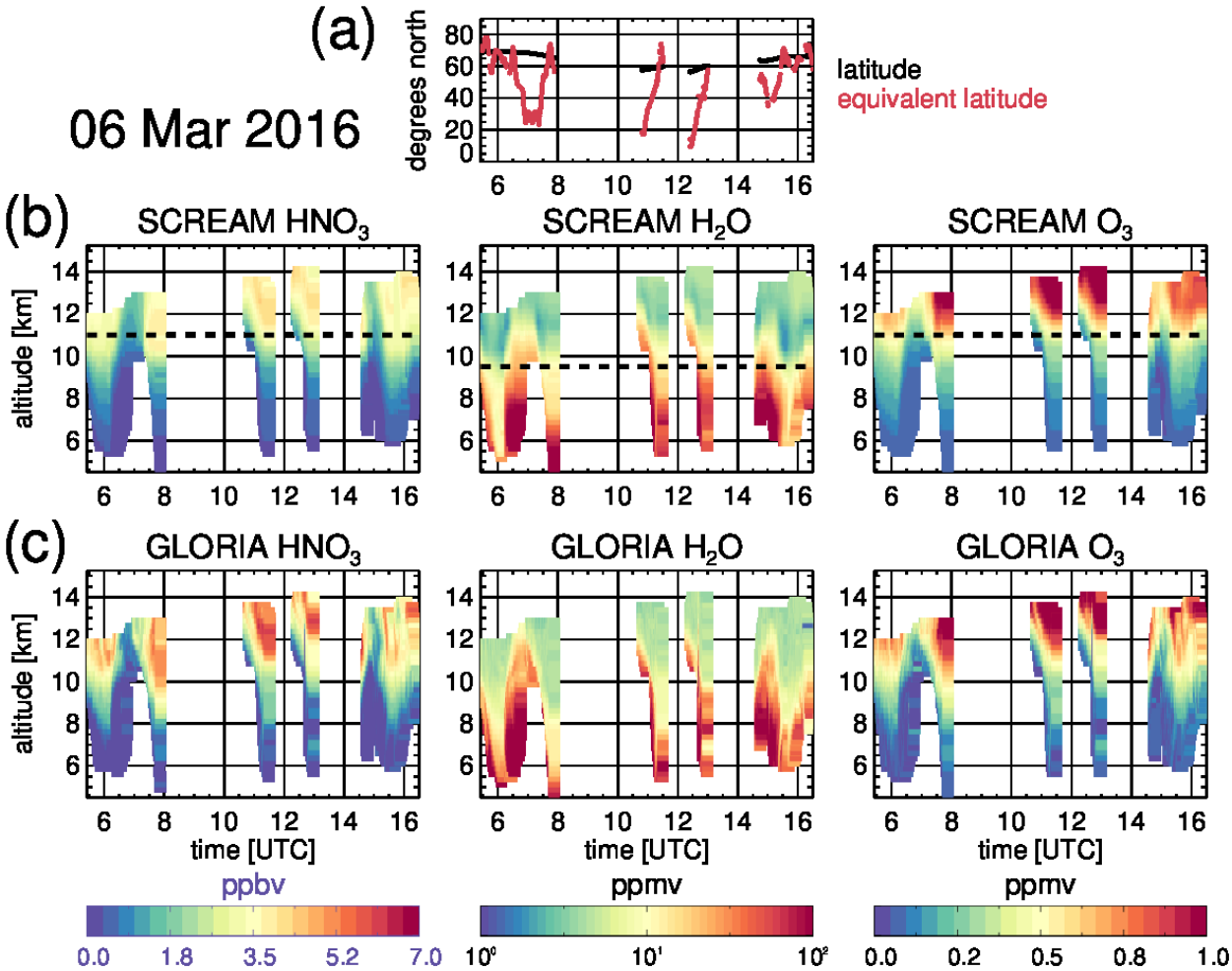

Figure S13: As in Fig. S3 but for 6 March 2016.

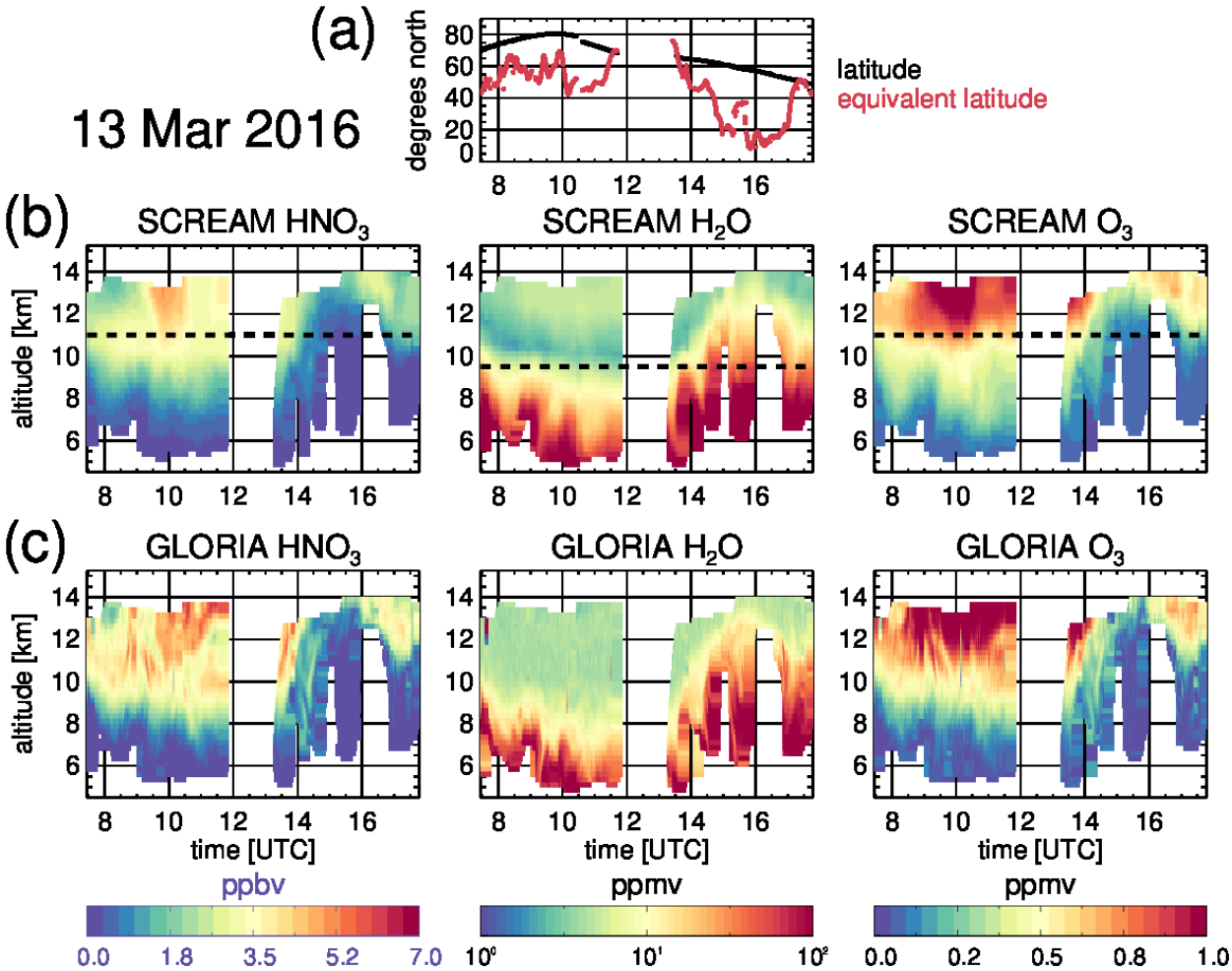

Figure S14: As in Fig. S3 but for 13 March 2016.

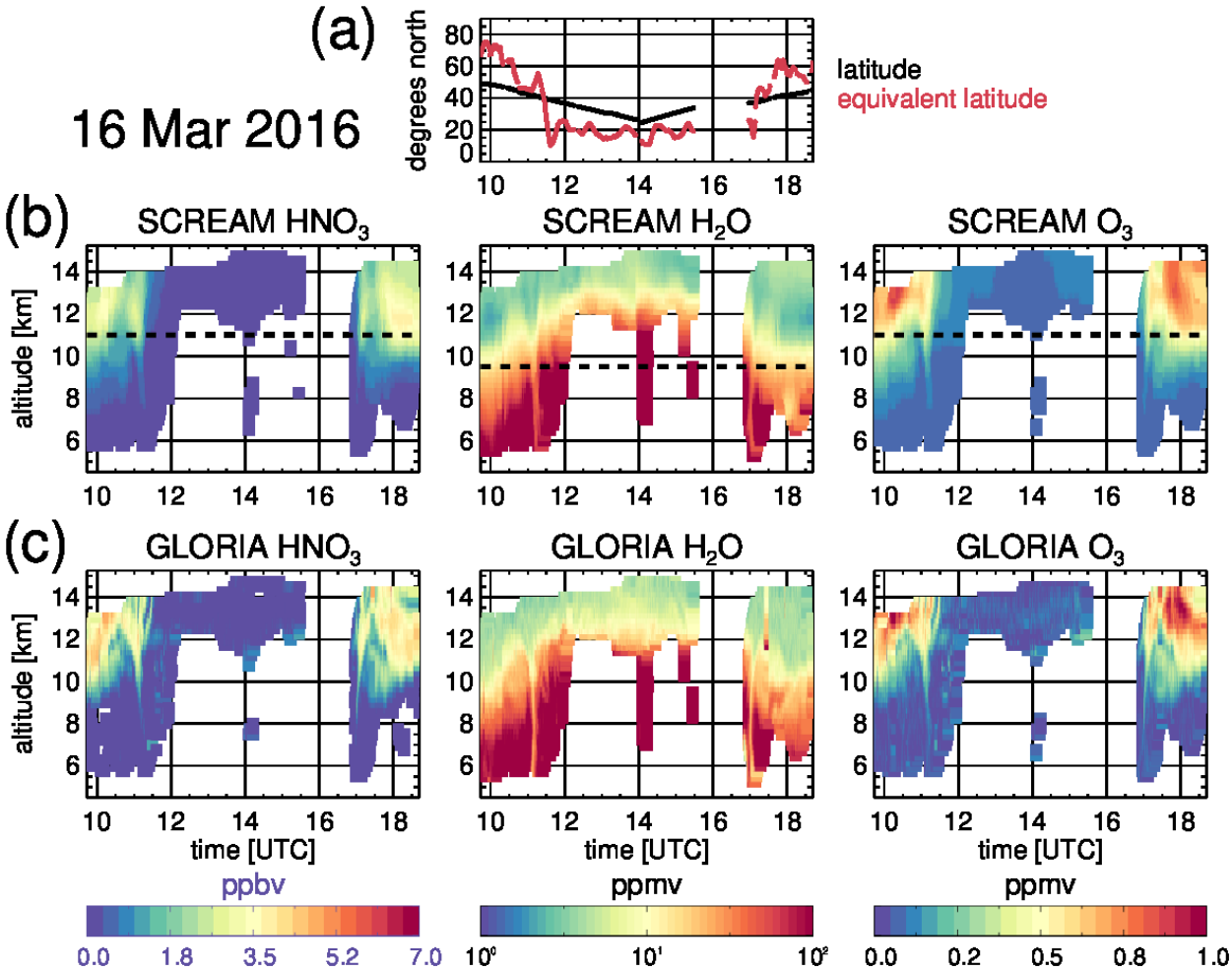

Figure S15: As in Fig. S3 but for 16 March 2016.

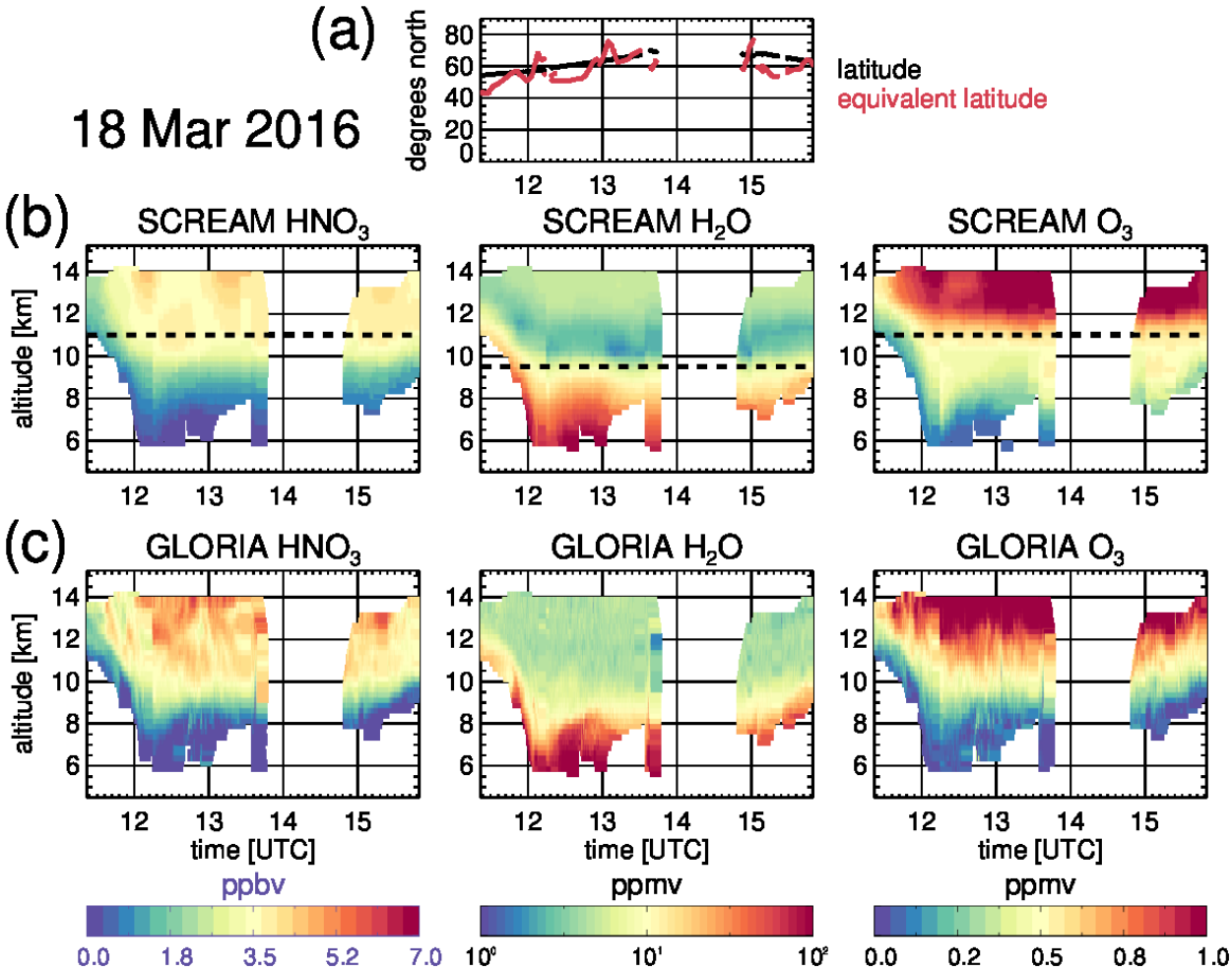

Figure S16: As in Fig. S3 but for 18 March 2016.

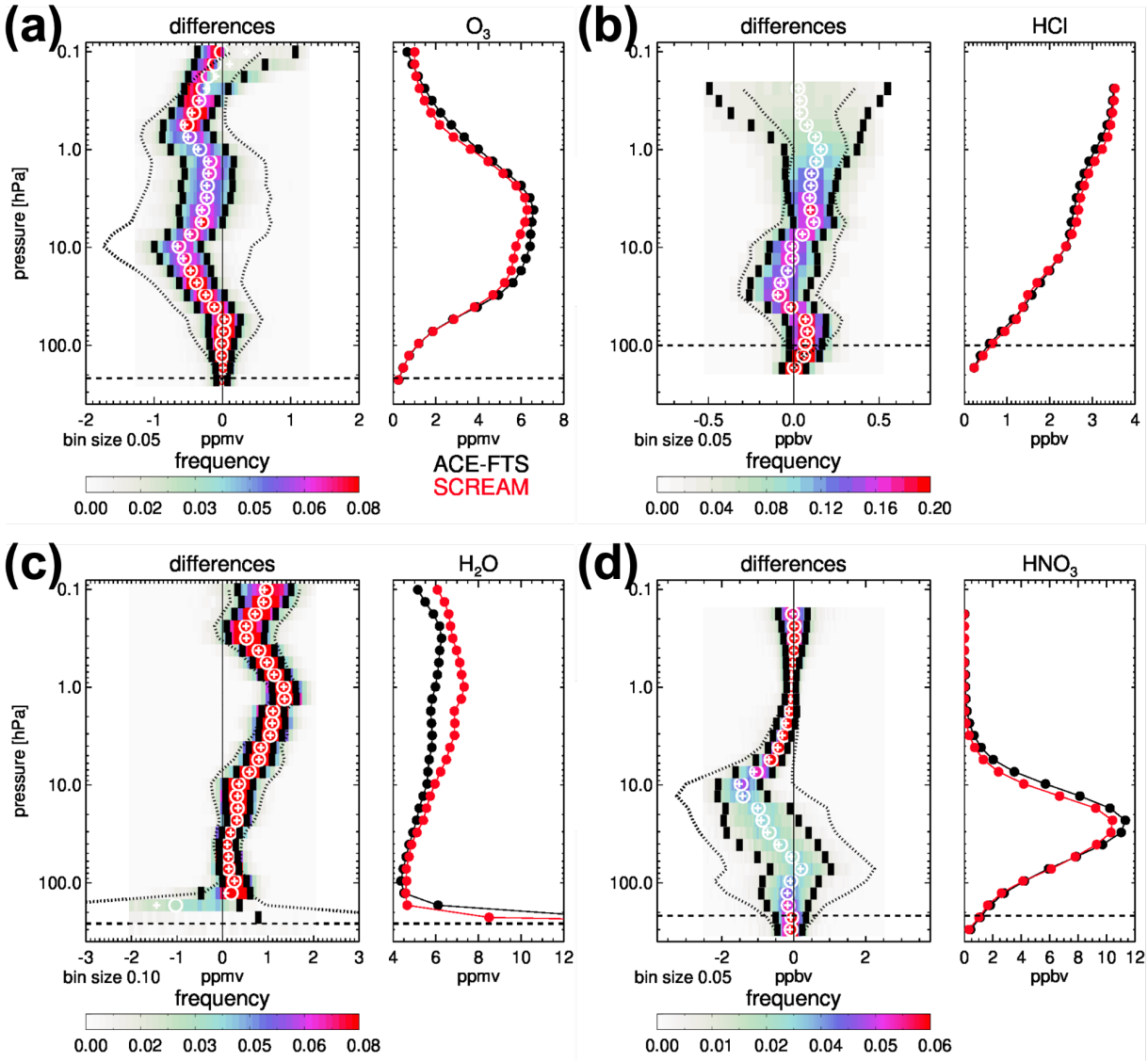

Figure S17: Statistical comparisons of the SCREAM ozone (a) and HCl (b), H_2O (c), and HNO_3 (d). with ACE-FTS observations for December–January, 30ºN–60ºN. The right- and left-hand side panels show mean profiles and difference statistics, respectively. Shown are the difference mean (plus signs), median (open circles), standard deviations around the mean (short vertical bars), probability density functions (colors) at prescribed pressure levels. The dotted lines are the mean difference plus/minus standard deviation of the ACE-FTS observations. All available December-January 2005–2020 ACE-FTS data are used.

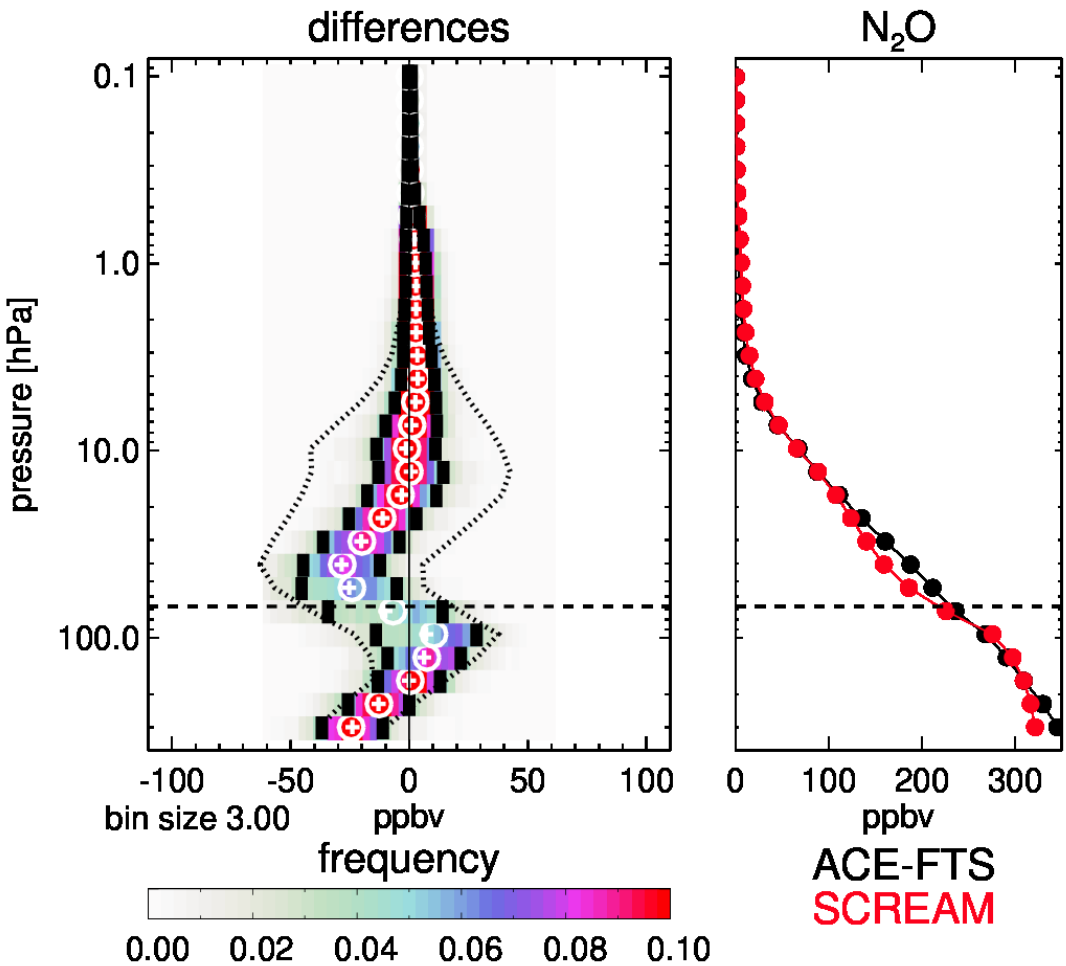

Figure S18: As in Fig. S17 but for N_2O.

Movie S1. File name: M1-Lauder-FPH-SCREAM.mp4. Animation showing all available frost point hygrometer water vapor profiles at Lauder, New Zealand (169.68ºE, 45.04ºS) between 2005 and 2021 (black) and collocated M2-SCREAM profiles (blue).

Movie S2. File name: M2-Hilo-FPH-SCREAM.mp4. As in Movie S1 but for Hilo, Hawaii (155.05ºW, 19.72ºN).

Movie S3. File name: M3-Boulder-FPH-SCREAM.mp4. As in Movie S1 but for Boulder, CO, USA (105.2ºW, 39.95ºN).