Assimilation of stratospheric and mesospheric temperatures from MLS and SABER into a global NWP model

K. W. Hoppel\textsuperscript{1}, N. L. Baker\textsuperscript{2}, L. Coy\textsuperscript{1}, S. D. Eckermann\textsuperscript{3}, J. P. McCormack\textsuperscript{3}, G. E. Nedoluha\textsuperscript{1}, and E. Siskind\textsuperscript{3}

\textsuperscript{1}Remote Sensing Division, Naval Research Laboratory, Washington DC, USA
\textsuperscript{2}Marine Meteorology Division, Naval Research Laboratory, Monterey, CA, USA
\textsuperscript{3}Space Science Division, Naval Research Laboratory, Washington DC, USA

Received: 17 March 2008 – Published in Atmos. Chem. Phys. Discuss.: 7 May 2008
Revised: 10 July 2008 – Accepted: 8 September 2008 – Published: 22 October 2008

Abstract. The forecast model and three-dimensional variational data assimilation components of the Navy Operational Global Atmospheric Prediction System (NOGAPS) have each been extended into the upper stratosphere and mesosphere to form an Advanced Level Physics High Altitude (ALPHA) version of NOGAPS extending to \(\sim 100 \text{ km}\). This NOGAPS-ALPHA NWP prototype is used to assimilate stratospheric and mesospheric temperature data from the Microwave Limb Sounder (MLS) and the Sounding of the Atmosphere using Broadband Emission Radiometry (SABER) instruments. A 60-day analysis period in January and February 2006, was chosen that includes a well documented stratospheric sudden warming. SABER and MLS temperatures indicate that the SSW caused the polar winter stratopause at \(\sim 40 \text{ km}\) to disappear, then reform at \(\sim 80 \text{ km}\) altitude and slowly descend during February. The NOGAPS-ALPHA analysis reproduces this observed stratospheric and mesospheric temperature structure, as well as realistic evolution of zonal winds, residual velocities, and Eliassen-Palm fluxes that aid interpretation of the vertically deep circulation and eddy flux anomalies that developed in response to this wave-breaking event. The observation minus forecast (O-F) standard deviations for MLS and SABER are \(\sim 2 \text{ K}\) in the mid-stratosphere and increase monotonically to about 6 K in the upper mesosphere. Increasing O-F standard deviations in the mesosphere are expected due to increasing instrument error and increasing geophysical variance at small spatial scales in the forecast model. In the mid/high latitude winter regions, 10-day forecast skill is improved throughout the upper stratosphere and mesosphere when the model is initialized using the high-altitude analysis based on assimilation of both SABER and MLS data.

1 Introduction

The extension of numerical weather prediction (NWP) models to higher altitudes has been motivated by both the desire to improve extended-range weather forecasts, and the goal of improving understanding of the middle atmosphere. Incorporating a realistic stratosphere has resulted in some gains in extended-range forecasts (Jung and Leutbecher, 2007), is expected to benefit the assimilation of new microwave measurements (Han et al., 2007), and has served as the basis for reanalysis (Uppala et al., 2005, Bloom et al., 2005) used for trend studies and transport calculations. Research NWP models such as the Canadian Middle Atmosphere Model (CMAM) have added a full mesosphere, and have been used to characterize the impact of assimilation schemes on the mesospheric forecast (Polavarapu et al., 2005; Sankey et al., 2007; Ren et al., 2008). In these studies the assimilated measurements were confined to altitudes below \(\sim 1 \text{ hPa}\), a limit defined by the altitude range of the thermal channels of the Advanced Microwave Sounding Unit-A (AMSU-A ) instrument (see Fig. 1).

In this paper, we report on the assimilation of stratospheric and mesospheric temperature measurements from the Sounding of the Atmosphere using Broadband Emission Radiometry (SABER) (Russell et al., 1999) and the Microwave Limb Sounder (MLS) (Waters et al., 2006) instruments. These research limb-sounding instruments provide measurements at altitudes well above those currently available from sounders whose data are assimilated operationally by NWP centers. The temperature retrievals from MLS and SABER are only weakly dependent upon the assumed background state, allowing the direct assimilation of temperature profiles rather than radiances, as opposed to nadir-sounding instruments such as AMSU-A and the Special Sensor Microwave Imager and Sounder (SSMIS). Neither the SABER nor MLS instruments currently provide data that meet operational time

Correspondence to: K. W. Hoppel (karl.hoppel@nrl.navy.mil)
requirements, although the MLS team is working on nearreal-time retrieval algorithms (Nathaniel Livesey, personal communication, 2008).

Here the MLS and SABER data are assimilated into a high-altitude version of the Navy’s Operational Global Atmospheric Prediction System (NOGAPS). NOGAPS consists of a global spectral forecast model (Hogan and Rosmond, 1991) plus the Naval Research Laboratory (NRL) Atmospheric Variational Data Assimilation System (NAVDAS) (Daley and Barker, 2001), and currently runs operationally from the ground up to \( \sim 1 \) hPa. The high-altitude extension of NOGAPS, which is designated NOGAPS-ALPHA (Advanced Level Physics-High Altitude), extends the top of the system from the mid stratosphere up to \( \sim 100 \) km altitude.

The initial extension and performance of the forecast model component of NOGAPS-ALPHA running without NAVDAS have been progressively documented in a number of recent studies (e.g., Eckermann et al., 2004; McCormack et al., 2004; Coy et al., 2005; Allen et al., 2006; McCormack et al., 2006; Eckermann et al., 2007; Siskind et al., 2007). In this study we run NOGAPS-ALPHA for the first time with NAVDAS to allow for the assimilation of higher altitude data provided by MLS and SABER.

We will show the results of assimilating MLS and SABER temperatures into NOGAPS-ALPHA up to 0.01 hPa during January–February 2006, a time period corresponding to a well documented stratospheric major warming. This time period exhibits very strong vertical coupling between the troposphere, stratosphere and mesosphere via gravity wave drag (Siskind et al., 2007) and illustrates the importance of having a system which extends from the ground to the upper mesosphere. These results also demonstrate the impact and challenges of assimilating upper stratospheric and mesospheric temperatures in NWP models.

The paper is organized as follows. Section 2 presents an overview of the NOGAPS-ALPHA forecast model component. Section 3 describes the new high-altitude version of NAVDAS used in NOGAPS-ALPHA and the satellite data sets to be assimilated. Section 4 describes results from the assimilation run for the January–February 2006 period. Section 5 summarizes these results and outlines future research directions.

### 2 NOGAPS-ALPHA global forecast model

The operational NOGAPS global forecast model is described in detail by Hogan and Rosmond (1991) and Hogan et al. (1991). Briefly, the dynamical core is Eulerian, hydrostatic, spectral in the horizontal with an energy and angular-momentum conserving finite-difference formulation in the vertical based on a generalized vertical coordinate (Simmons and Burridge, 1981). For the experiments reported here, the forecast model was run using a triangular spectral truncation at wavenumber 79 (T79), corresponding to a grid point resolution on the quadratic Gaussian grid of 1.5°.

The model’s dynamical variables are relative vorticity, divergence, virtual potential temperature, specific humidity, and terrain (surface) pressure. The model is central in time with a semi-implicit treatment of gravity wave propagation, implicit zonal advection of moisture and vorticity, and Robert (Asselin) time filtering (Simmons et al., 1978; Simmons and Jarraud, 1983). The operational model includes physical parameterizations of vertical diffusive transport in the planetary boundary layer (Louis, 1979; Louis et al., 1982) coupled to a land surface model (Hogan, 2007), orographic gravity-wave and flow-blocking drag (Webster et al., 2003), shallow cumulus mixing (Tiedtke, 1984), deep cumulus convection (Emanuel and Zivkovic-Rothman, 1999; Peng et al., 2004), convective, stratiform and boundary layer clouds and precipitation (Slingo, 1987; Teixeira and Hogan, 2002), and shortwave and longwave radiation (Harshvardhan et al., 1987).

NOGAPS runs operationally at T239L30 with only a few thick highly-diffused stratospheric levels above \( \sim 25 \) hPa. While seeking to retain most of the features of the operational model, the ALPHA version of the forecast model incorporates a number of additions and modifications. One such addition is prognostic ozone with parameterized photochemistry (McCormack et al. 2004, 2006; Coy et al., 2007).
2.1 Vertical model levels

NOGAPS-ALPHA can be run with a variety of vertical level spacing and top boundary levels. The NOGAPS-ALPHA forecast model also replaces the $\sigma$ coordinate used in NOGAPS (Hogan and Rosmond, 1991) with a hybrid $\sigma-p$ coordinate that transitions smoothly from terrain-following levels at the surface to isobaric levels in the lower stratosphere and higher (Eckermann et al., 2004; Eckermann, 2008a). The version used here contains 68 model layers (L68), with a model top at 0.0005 hPa (~96 km). The lowest levels are identical to the operational L30 setup, but then transition to isobaric layers at altitudes above ~87 hPa, with a height thickness of $\Delta Z \approx 2$ km throughout the middle atmosphere. Isobaric models levels in the middle atmosphere should aid the assimilation of satellite temperature and constituent retrievals which are provided on pressure levels (e.g., Simmons et al., 1989; Trenberth and Stepaniak, 2002). The top two model layers constitute a “sponge layer" and, for the L68 model formulation adopted here, span the range of 0.0005–0.00089 hPa. Damping is achieved through an increase in the spectral diffusion coefficient from the background values applied lower down. Additional damping is applied to the virtual potential temperature in the sponge layer in a way that relaxes temperatures towards an isothermal state.

2.2 Radiative heating and cooling rates

The Harshvardhan et al. (1987) radiation schemes used in the operational NOGAPS have been replaced by the NASA CLIRAD (climate radiation) shortwave (SW) and longwave (LW) radiation parameterizations of Chou and Suarez (1999) and Chou et al. (2001), respectively, which both extend through the stratosphere to ~0.01 hPa. LW cooling rates are also computed using the scheme of Fomichev et al. (1998) to account for the effects of non-local thermodynamic equilibrium (non-LTE) on infrared (IR) CO$_2$ emissions at higher altitudes. The final LW cooling rate profile blends CLIRAD cooling rates at lower altitudes with Fomichev et al. (1998) rates at high altitudes using a ramped linear weighting centered at ~75 km altitude (see Eckermann et al., 2008b for details). CLIRAD also includes a heating rate contribution from near-IR CO$_2$ absorption that becomes unrealistically large in this scheme near 0.01 hPa (see Eckermann et al., 2007). These rates are overestimated at high altitudes due to omission of non-LTE effects, and thus this band’s contribution is deactivated in the experiments reported here. We are currently testing the non-LTE near-IR CO$_2$ heating rate parameterization of Fomichev et al. (2004) in NOGAPS-ALPHA as a potential replacement for CLIRAD near-IR heating rates at high altitudes.

These radiation schemes use the model’s specific humidity fields from the surface to 200 hPa: above this level specific humidities are specified using the zonal-mean observational climatology described in Eckermann et al. (2007). Ozone mixing ratios in the radiation calculation here use the zonal-mean observational climatology described by Eckermann et al. (2007), which uses only daytime ozone values at altitudes above 0.3 hPa where ozone varies diurnally. The radiation schemes can also use the model’s three-dimensional prognostic ozone fields, but that option is not used in the experiments reported here.

To reduce the computational burden, the radiative heating and cooling rates are updated in the model every two hours, and the longwave cooling rates can be computed on a reduced horizontal grid then re-interpolated back onto the model grid, though the latter option was not used here.

2.3 Middle atmospheric gravity wave drag

We parameterize nonorographic gravity wave drag (GWD) here using the Whole Atmosphere Community Climate Model (WACCM) scheme, described in Appendix A of Garcia et al. (2007). As implemented here, we apply only the gravity wave momentum flux divergence tendencies to the model. GWD-induced vertical diffusivities, while calculated, are not at present used to mix momentum, heat and constituents. Benchmarking and optimal tuning of this scheme in NOGAPS-ALPHA is underway through a series of multi-year forecast and climate simulations which will be reported elsewhere (see, e.g., Eckermann et al., 2008b). Here, we choose parameter values similar to those currently used in WACCM. In every grid box we launch at 500 hPa 65 gravity waves whose momentum fluxes have a Gaussian distribution as a function of intrinsic phase speed, centered at zero with of width of 30 m s$^{-1}$. The 65 waves sample this spectrum evenly between intrinsic phase speed limits of $\pm 80$ m s$^{-1}$, with all waves co-aligned with the source-level wind direction. We use the same latitudinal and seasonal variation of the source spectrum as in Garcia et al. (2007) with a background stress $\tau_0^q=0.007$ Pa. To yield a realistic polar summer mesopause temperature in the model, we reduced the gravity wave drag efficiency $e$ from its nominal WACCM value of 0.125 to 0.050. While available, we do not use the WACCM orographic gravity wave drag parameterization. Instead, we use the Palmer et al. (1986) scheme which has been found to capture the interannual variations of the Arctic winter stratopause temperatures during 2006 in previous NOGAPS-ALPHA runs (Siskind et al., 2007). Parameterized gravity waves deposit all their remaining flux in the top sponge layer to conserve column momentum. Section 4.3 further describes the effectiveness of the GWD scheme in maintaining the observed zonal mean temperatures.

3 Assimilation setup

3.1 NAVDAS

The NRL Atmospheric Variational Data Assimilation System (NAVDAS) is a three-dimensional variational (3DVAR)
data assimilation system (Daley and Barker, 2001), designed for use with both global and mesoscale NWP models. NAVDAS became operational in NOGAPS in October 2003. NAVDAS solves the 3DVAR equation in observation space, i.e.:

\[ x_a - x_b = P_b H^T [HP_b H^T + R] [y - H(x_b)] \]  

where \( x_a \) is the analysis vector, \( x_b \) is the background vector, \( P_b \) is the background error covariance, \( y \) is the observation vector, \( R \) is the observation error covariance, and the superscript \( T \) denotes transpose. In general, the application of the observation or forward operator \( H \) represents any necessary spatial and temporal interpolations from the forecast model background to the observation location and time. If the observed quantity is not directly related to the model state variables, then \( H \) also represents the transformation from the forecast values to the observed quantity. The matrix \( H \) is the Jacobian matrix corresponding to the forward operator \( H(x_b) \) linearized about the background state vector. The analysis vector consists of the gridded fields of temperature, winds, geopotential height, and pseudo-relative humidity (e.g. Dee and da Silva, 2003).

For the applications discussed in this paper, the analyses are computed using a 6-h update cycle, and \( x_b \) is the 6-h forecast from the previous update cycle. However, the innovations, \( y - H(x_b) \), are calculated using the 3-, 6- and 9-h NWP forecasts interpolated to the observation location and time (linear in time; bicubic in horizontal space; log pressure in the vertical). This makes NAVDAS a low-time resolution 3DVAR-FGAT (first guess at appropriate time) algorithm. The innovations (also called the observation minus forecast, or O-F) represent the deviation of the forecast from the observations, in observation space. The quantity \( x_a - x_b \) is the correction vector in model grid space.

The solution to (1) is calculated in observation space, using the following 3 steps. First, we calculate the observation space matrix and innovation vector:

\[ A = HP_b H^T + R; \quad d = y - H(x_b) \]  

Next, we solve the linear system:

\[ Az = d \]  

Last, we perform the post-multiplication:

\[ x_a - x_b = P_b H^T z \]  

The background error covariance, \( P_b \), is formulated as a separable product of vertical and horizontal functions. The background variances are static, and specified as a function of latitude and pressure. A second-order autoregressive (SOAR) function is used to represent spatial correlations in the vertical and horizontal, with correlation lengths that vary as a function of variable and pressure level. Options are built into NAVDAS for non-separable formulations, but these have not been explored for this work. The multivariate correlations are derived from hydrostatic and geostrophic balance constraints, following the formalism of Daley (1991) and Daley and Barker (2001). The strength of the temperature-wind geostrophic coupling is given by the factor 0.9sin(\( \varphi \)) for latitudes |\( \varphi \)| > 30°. The coupling factor decreases rapidly to zero equatorward of 30°. The background error covariances control how the information is spread from the observation to the surrounding grid points, and to other variables (e.g., wind observations will produce height increments away from the equator).

The research version of NAVDAS used in this study has an extended vertical range with a data top at 0.01 hPa. Satellite observations that are currently assimilated operationally, and used in this study, include AMSU-A radiances (Baker et al., 2005), surface winds and total precipitable water from polar orbiting microwave imagers, atmospheric motion vectors from polar and geostationary satellites, and surface winds from scatterometers. A complete list of assimilated observation types, and typical data counts may be found in Baker et al. (2007). Measurements from the Atmospheric Infrared Sounder (AIRS), Global Positioning System-Radio Occultation (GPS-RO), and SSMIS are not included in this study. For this study, AMSU-A channel 10, which has a weighting function that peaks around 50 hPa, is the highest AMSU-A channel that is used. Higher-peaking channels 11–14 are not used, due to the tendency of the current operational radiance bias correction scheme (Campbell et al., 2005) to reinforce the model bias at these levels.

3.2 MLS data

The Microwave Limb Sounder (MLS) was launched aboard the Aura satellite in July 2004 (Waters et al., 2006). It retrieves atmospheric temperature using limb observations of the 118-GHz O$_2$ and the 234-GHz O$_{18}$O spectral lines. Here retrieval version 2.2 (v2.2) temperatures between 32–0.01 hPa are assimilated into NOGAPS-ALPHA. The precision of the temperature measurement is 1 K or better at altitudes below 0.316 hPa, but degrades to ∼2.2 K at 0.01 hPa (Schwartz et al., 2008). Schwartz et al. (2008) presented comparisons of MLS v2.2 temperatures with correlative data sets. They showed that while the bias in the stratosphere was generally less than 2 K when compared to other observations, at some levels there were persistent MLS temperature biases with ∼3 K peak-to-peak vertical structure. In the mesosphere MLS v2.2 temperatures are ∼0–7 K lower than most other measurements.

The horizontal resolution of the MLS temperature measurements in the stratosphere is about ∼180 km along track and ∼12 km cross-track. Because here the forecast model is being run at a lower resolution (T79) than either the MLS and SABER data resolution, the analysis does not account for the specific limb sampling geometry.
The vertical resolution of the temperature retrievals, expressed as the full-width half-maximum (FWHM) of the averaging kernel is \( \sim 3.5 \text{ km} \) at 31.6 hPa, and degrades at altitudes above 20 hPa to \( \sim 6.2 \text{ km} \) at 3.16 hPa and \( \sim 14 \text{ km} \) at 0.01 hPa (Schwartz et al., 2008). In principle, the assimilation algorithm should incorporate the retrieval’s height-dependent vertical averaging kernel in the observation operator (H in Eq. 1). For MLS temperatures, this is problematic near the top of the analysis domain (0.01 hPa) because the observation temperature is sensitive to temperatures above the top. Thus, for simplicity, in this work NAVDAS uses a Gaussian vertical averaging kernel for MLS with a FWHM of \( \sim 4 \text{ km} \) at all altitudes. This analysis averaging kernel is smaller than the true MLS averaging kernel in the upper stratosphere and mesosphere, and is a possible source of error at altitudes above \( \sim 0.1 \text{ hPa} \).

3.3 SABER data

SABER is a 10 channel broadband, limb-viewing, infrared radiometer which has been measuring stratospheric and mesospheric temperatures since the launch of the Thermosphere Ionosphere Mesosphere Energetics and Dynamics (TIMED) satellite in December 2001. Stratospheric temperature is obtained from the 15 \( \mu \text{m} \) radiation of CO\(_2\). This emission is in local thermodynamic equilibrium (LTE) in the stratosphere and lower mesosphere and has been extensively discussed and validated by Remsberg et al. (2003). In the middle to upper mesosphere and lower thermosphere (MLT), non-LTE conditions prevail. Initial results from a non-LTE temperature retrieval have been presented by Mertens et al. (2004). Here we use retrievals with the non-LTE effects included (Version 1.06 in the SABER database) over the pressure range of 32–0.019 hPa. Remsberg et al. (2003) estimated the precision by calculating the zonal standard deviation at 50° S during the summertime when geophysical variability is low. The estimated precision was \( \sim 1 \text{ K} \) at 32 hPa, and monotonically increases to \( \sim 4 \text{ K} \) at 0.01 hPa. The SABER v1.06 temperatures are known to have a low bias of \( \sim 5–10 \text{ K} \) in the cold polar summer mesopause region (Kutepov et al., 2006), a problem that has been corrected in the most recent retrieval version (Remsberg et al., 2008). This bias is not corrected for in the analysis, and thus ordinarily would lead to analysis errors near 0.01 hPa in the southern polar region for this assimilation experiment. However, SABER views to the side of the spacecraft and during January-February 2006 was preferentially viewing high northern (winter) latitudes only, with data in the summer hemisphere extending only to \( \sim 52° \text{ S} \). The SABER retrieval vertical resolution is \( \sim 2 \text{ km} \) and the along-track profile spacing is \( \sim 3° \). The forecast model vertical resolution in the stratosphere and mesosphere is also \( \sim 2 \text{ km} \). Therefore the analysis observation operator, H, for the SABER observations uses vertical interpolation with no extra smoothing.

3.4 Assimilation of MLS and SABER data

NOGAPS-ALPHA assimilates MLS and SABER temperatures between 32 and 0.01 hPa. Figure 2 illustrates the MLS and SABER measurement locations during one particular 6-h analysis update. It also shows the correction field, \( x_{a} - x_{b} \), and illustrates the horizontal spreading of the observation impact, which is controlled by the background error covariance. The horizontal correlation length of the background temperature is 385 km. At altitudes above 0.01 hPa, the upper-level correction fields are damped over a height range of \( \sim 6 \text{ km} \) before reverting to the free-running forecast model fields up to

---

**Fig. 2.** Example of the MLS (squares) and SABER (crosses) measurement locations during a 6-h assimilation cycle on 5 February 2006 at 00:00 UTC. The color shading shows the analysis-forecast \( (x_{a} - x_{b}) \), in Eq. 1) temperature correction (K) at 0.04 hPa.

**Fig. 3.** (a) The SABER (v1.06)-MLS (v2.2) global temperature bias estimate. (b) The global average O-F for MLS and SABER for the January–February 2006 analysis. The bias profile (a) was subtracted from the SABER temperatures prior to assimilation summarized in (b).
0.0005 hPa. Although no measurements are assimilated at altitudes above 0.01 hPa, we find that the model layers between 0.005 hPa and 0.0005 hPa are still important for capturing the effects of gravity wave breaking on the mesospheric and stratospheric circulations.

The background temperature error variance is specified as a static function of pressure and latitude. Above the 32 hPa level, the error variance is in the range of 1.3 to 16.1 K, increasing with altitude and increasing poleward. The observation errors are taken from the SABER and MLS retrieval files with two additional constraints. The minimum observation error is set to the larger of 2 K or 30% of the O-F magnitude. This latter constraint prevents the large mesospheric innovations from being rejected by quality control filters in the analysis algorithm. With these choices for the error variance, the analysis is tightly constrained to the SABER and MLS observations, with RMS residual differences (A-O) of ∼2 K.

Global mean systematic biases between MLS and SABER temperatures have been removed to prevent the introduction of spurious temperature structures in the analysis. The relative bias between SABER and MLS temperatures was estimated from the globally averaged innovation (O-F) statistics. The difference between the average SABER and MLS innovation, shown in Fig. 3a, was used to modify the SABER data prior to assimilation. This bias estimate is similar to the SABER-MLS differences reported in the MLS temperature validation study of Schwartz et al. (2008). Figure 3b shows the global average O-F for the analysis performed with the bias-corrected SABER data. The MLS and SABER average innovations differ by less than ∼1 K, which suggests that most of the relative bias between the instruments has been removed using this simple procedure.

4 Results

4.1 Analysis of the 2006 stratospheric sudden warming

The analysis period of January–February 2006 encompasses a well-documented stratospheric sudden warming (Manney et al., 2008a; Manney et al., 2008b; Hoffmann et al., 2007; Siskind et al., 2007; Coy et al., 2008). The Northern Hemisphere winter of 2006 was disturbed by a major stratospheric sudden warming (SSW) on 20 January 2006. After the major SSW the lower stratosphere remained warm until the end of February, while during the same time the polar stratopause reformed at an unusually high altitude. Figure 4 compares the daily polar temperature from NOGAPS-ALPHA with the SABER observations. The analysis captures the descent of warm air after the major SSW, the disappearance of the stratopause in late January, and the high-altitude reformation of the stratopause in early February. There are small differences between the NOGAPS-ALPHA and SABER temperatures because the analysis provides a synoptic polar value, while the SABER estimate is from a limited range of longitudes and local times poleward of 80° N. The high stratopause formation on 1 February seen in SABER occurs near the top of the assimilated observations at 0.005 hPa and is lower and cooler in the analysis. Manney et al. (2008b) noted the difficulty of reproducing this high-altitude stratopause in the version 5 Goddard Earth Observing System (GEOS-5) and European Centre for Medium-Range Weather Forecasts (ECMWF) analyses. This difficulty is presumably due to the absence of mesospheric temperature data in these analysis, which must instead rely on the accurate parameterization of subgrid-scale orographic (Siskind et al., 2007) and non-orographic (Ren et al., 2008) GWD in the forecast model to force the stratopause changes.
during the SSW. This is made more difficult by the lower model tops of 0.01 hPa for GEOS-5 and ECMWF.

The major SSW resulted from the rapid advection of tropical, low potential vorticity (PV) air over the pole near 10 hPa (Coy et al., 2008). As this low PV air was transported over the pole, conservation of PV induced an anti-cyclonic circulation along with an associated high pressure system. The development of this high pressure system can be seen in the NOGAPS-ALPHA geopotential height fields (Fig. 5). The breaking wave that initiated the poleward intrusion of tropical air occurs at 10 hPa near the Greenwich Meridian on 18 January 2006, although the developing high is too small to be seen at this time (Fig. 5a), and only the quasi-stationary Aleutian high is apparent. By 20 January 2006 (Fig. 5b) the developing anti-cyclone (near 60° N, 80° E) is already stronger than the weakening Aleutian high. The 10 hPa geopotential height of the developing anti-cyclone continues to increase to over 40.2 km on 22 January as it moves closer to the pole (Fig. 5c), and peaks at over 40.4 km on 24 January (Fig. 5d). The Aleutian High is now gone, having “merged” with (i.e., wrapped around) the developing high, similar to the merging Aleutian and developing highs that occurred in the January 1992 minor warming (see O’Neill et al., 1994). While the high is developing, the low of the polar vortex fills in as the vortex weakens, as can be seen by the shrinking of the purple area over time in Fig. 5. Thus

Fig. 5. Analyzed NOGAPS-ALPHA geopotential heights (km) at 10 hPa on 18–24 January 2006 at 12:00 UTC. Panels (a) through (d) show the fields at 2-day intervals. The Lambert equal area projections are centered on the North Pole and extend to the equator. Zero degrees longitude is at the bottom. The contour interval is 0.2 km. Heights less than 30.4 km have blue/purple shading and heights greater than 30.6 km have yellow/green/light-blue shading and black contours.
the NOGAPS-ALPHA assimilation realistically captures the mid-stratospheric evolution of the major SSW, consistent with operational analyses, in addition to the mesospheric temperature evolution during this highly dynamic time.

A key test of the analysis is the quality of meteorological quantities other than the temperature, which is directly constrained by the assimilation. Figure 6 shows some zonally averaged diagnostics before (Fig. 6a) and after (Fig. 6b) the major SSW, including zonal winds, Eliassen-Palm (EP) flux vectors (a measure of planetary wave activity and propagation) and velocity vectors of the residual mean meridional circulation. Before the SSW the Northern Hemisphere planetary waves are strong (as evidenced by the large upward and equatorward EP flux vectors in Fig. 6a) and the Northern Hemisphere zonal-mean zonal winds are weak. After the SSW, the planetary waves are weak and the zonal winds are strong in the winter mesosphere. In the winter stratosphere the zonal winds remain weak with a prominent zero-wind line near 60°N in the lower stratosphere. This zero-wind line blocks the vertical propagation of planetary waves near 60°N after the SSW. The temperatures after the SSW show the elevated polar stratopause (near 0.02 hPa), with cold air at 1 hPa, the typical winter polar stratopause height (as in Fig. 6a).

The residual mean meridional circulation (red arrows in Fig. 6) shows strong poleward and downward motion north of 60°N at 1 hPa before the SSW, forced mainly by the EP flux divergence of the planetary scale waves. After the SSW, the poleward and downward motion north of 60°N is located at higher altitudes, at and above the 0.1 hPa level, forced mainly by the gravity wave drag parameterization acting in the upper part of the westerly jet where the zonal winds are decreasing with altitude. Note that, after the SSW, the meridional circulation in the mesosphere is strong into the zonal wind westerly jet, and weak north of the mesospheric jet. The unusual mesospheric jet seen in February 2006 has been noted by Manney et al. (2008b) and Siskind et al. (2007) The ability of NOGAPS-ALPHA to produce realistic winds and derived secondary circulations and planetary wave diagnostics in the mesosphere will enable detailed future studies of this highly dynamic region.

4.2 O-F statistics

There are currently few stratospheric and mesospheric temperature measurements that can be used as an independent validation of the analysis. We therefore characterize the quality of an assimilation by examining the observation minus forecast (O-F) statistics generated by the analysis. As described in Sect. 3.1, each O-F value is calculated for a forecast time of 3 to 9 h during an update cycle. For each 6-h update cycle, O-F statistics (mean, standard deviation, correlation coefficient) were calculated in 20° latitude bins. The statistics were then averaged over all update cycles during the analysis period. As Fig. 2 illustrates, the observations in mid- and low-latitude regions generally consist of ~40–100 profiles per latitude bin distributed along 6 north-south oriented tracks for each instrument. Figure 7a–c shows the mean O-F for three latitude bins representing mid-latitude...
summer, equatorial, and mid-latitude winter regions. Although a globally-averaged SABER-MLS bias correction has been applied, residual latitudinally-varying biases are still evident in the differences between the mean SABER and MLS O-F profiles. The source of these residual biases is under investigation, and may be a combination of differences in vertical resolution, systematic differences in local time sampling, and latitudinally varying instrument errors. The largest mean O-F differences occur near the stratopause and mesopause. The structure of the O-F bias in the tropics indicates that the observed stratopause is slightly warmer and occurs at a slightly higher altitude than the forecast stratopause. There is some indication in MLS temperature comparisons with other instruments that a MLS warm bias exists near 1 hPa (Schwartz et al., 2008).

The standard deviation of the temperature O-F, shown in Fig. 7d–f, increases almost monotonically with altitude from \( \sim 1–2 \) K at 10 hPa to \( \sim 6 \) K at 0.01 hPa for all three latitude bins. Some of this increase in standard deviation may be attributable to degradation in the MLS precision in the mesosphere. This MLS O-F standard deviation profile corresponds to approximately twice the estimated MLS precision.
Fig. 8. Comparison of the zonal-mean temperatures from 22 10-day forecasts with MLS measurements. Each 10-day forecast during January–February 2006 was initialized from the analysis. The MLS zonal-mean temperature was calculated using the all MLS measurements within ±1 day of the forecast time.

(Table 2 of Schwartz et al., 2008). The standard deviation profiles for SABER and MLS are very similar to the standard deviation profiles for the coincident SABER-MLS measurements in the validation study of Schwartz et al. (2008). The coincidence criteria used in that study were <220 km and <3 h. This suggests that the O-F standard deviation is a combination of random observation error and geophysical variability over short temporal and spatial scales that is not captured by the forecast. Assimilation in the mesosphere is expected to be more difficult because of increased dynamical variance at small temporal and spatial scales. One potential difficulty is that the NA VDAS multivariate correlation scheme assumes a purely rotational wind with strong geostrophic coupling at high latitudes. Koshyk et al. (1999) have shown that in the mesosphere, the forecast model’s kinetic energy at horizontal wave numbers >∼50 (which includes the spatial scale of the MLS and SABER temperature corrections) is dominated by divergent rather than rotational motions (i.e., gravity waves). While some of this small-scale divergent motion can be resolved in MLS and SABER temperatures, most of it is unresolved (e.g., Preusse et al., 2006; Wu and Eckermann, 2008), potentially explaining some of this increase in O-F standard deviation with height. Further study is needed to determine if the use of unbalanced wind and temperature corrections in the mesosphere or higher resolution mesospheric satellite data (e.g., Alexander et al., 2008) can reduce these O-F standard deviations.

The dotted lines shown with the O-F standard deviations in Fig. 7 are the standard deviation of just the observations (O) that were used for the O-F calculation. This should not be confused with the observation errors, which are not shown. The O standard deviation includes contributions from geophysical variations (zonal and meridional) within the latitude bin and measurement error. If the observation noise is small and the model forecast accurate, we would expect the O-F standard deviation to be smaller than the O standard deviation, because the model should capture some of the true geophysical variability represented in the observations. This is the case for the mid-latitude winter, where the O standard deviation is much larger than that of O-F. In the summer mid-latitudes the O-F standard deviation is somewhat smaller than the O standard deviation, while in the equatorial latitudes the O-F and O standard deviations are similar. Another measure of the quality of the forecast is provided by the correlation coefficient between the observations and forecast, which is shown in Fig. 7g–i. In the mid-latitude winter hemisphere the correlation coefficient ranges from ∼1 in the lower stratosphere to ∼0.8 in the mesosphere. In the summer mid-latitudes the correlation coefficient is ∼0.6–0.8 between 30 and 0.1 hPa, while in the tropics the correlation is ∼0.4–0.6 throughout most of the pressure range. The low correlation in the tropics may reflect both the smaller geophysical variability in this region and absence of stringent balance relations which can be used to constrain winds based on measurements of temperature only.
Comparing medium range forecasts with the analysis has proven to be a useful tool for refining and evaluating the forecast model. Ten day forecasts were found to be sufficient for identifying temperature tendencies and the impact of changing model parameters such as those in the nonorographic GWD scheme (see Eckermann et al., 2008). Fig. 8 shows the average temperature bias from 22 10-day forecasts distributed over the January–February period. Each T79L68 forecast was initialized from the analysis (using MLS and SABER data) and then compared to the zonal-mean temperatures calculated directly from the MLS measurements. The regions with the largest temperature differences are near and just above the stratopause, especially near the summer mesopause and at the equatorial stratopause. The difference near the equatorial stratopause was already apparent in the mean O-F (Fig. 7) and may be due in part to a high bias in the MLS temperatures in this region. Elsewhere the bias after 10 days is less than 5 K. While Fig. 8 suggests that the medium range forecasts produce reasonable zonal-mean temperatures, such forecasts in the mesosphere are expected to be difficult due to the short spatial and temporal correlation scales. Shepherd et al. (2000) showed that on the 1000 K potential temperature surface (~35 km altitude) the correlation time for Eulerian horizontal velocity shear is ~2 days whereas at 4000 K (~70 km), the correlation time drops to ~3 h. This dramatic change with altitude reflects the dominance of gravity-wave motion in the mesosphere (Shepherd, 2007).

To examine medium range forecast skill in more detail, we calculated forecast errors by comparing the forecasts with the analyses (F-A). Absent any better estimate of the analysis errors, we use the SH O-F standard deviation of Fig. 7d as the estimated random error of the analysis in the stratosphere and mesosphere. Figure 9 shows the F-A mean and standard deviations as a function of the forecast length for the 22 forecasts. Comparisons are shown for 30–70° latitude for the NH (winter) and SH (summer). The white line denotes the forecast day at which the F-A error exceeds the estimated analysis error. Forecast errors less than the analysis error are not meaningful since the analysis is being used as the truth.

In the SH summer, both the standard deviation and the mean error increase rapidly at altitudes above 0.1 hPa. Between 100–0.1 hPa, the 10-day F-A standard deviation has not increased much above the analysis error, and is similar in magnitude to the zonal standard deviation of just the analysis (not shown). Because there is very little geophysical variability, experiments (not shown) indicate that a forecast based only on persistence yields similar results in the summer at altitudes above ~10 hPa. In the NH winter the forecast error exceeds the estimated analysis error after ~1 day in the mesosphere and ~3 days in the lower stratosphere.

Because the mesosphere is strongly forced by waves propagating from below, the importance of the mesospheric initial conditions to forecast skill may be less than that at lower altitudes. To examine this further, we ran a subset of twelve 10-day forecasts using a zonal mean climatology above 10 hPa for the initial conditions. Between 10 and 1 hPa, the current analysis was transitioned linearly (in log pressure) to the COSPAR International Reference Atmosphere (CIRA) temperature climatology and the UARS reference atmosphere project (URAP) zonal wind climatology (see Eckermann et al., 2004). Figure 10 shows a comparison of forecast RMS error between the two cases. In the SH summer mid-high latitudes, there is no significant difference in forecast RMS error between the two initializations after about 2 days. This is not surprising because zonal-mean MLS temperatures and climatology are similar, and the zonal symmetry of the summertime flow confers little advantage to the analysis. The equatorial regions show little difference in forecast RMS error after about 4 days between the two different initializations. The only exception is a small, persistent improvement near 3 hPa for the forecasts initialized from the analysis. By contrast, in the NH winter mid-high latitudes, using the analysis instead of a climatology for the initial conditions leads to smaller RMS error for the entire 10-day forecast between ~10 and 0.01 hPa. It is important to note that a much larger sample size spanning more than 2 months would be necessary to generalize these results.
5 Discussion and conclusions

For the first time we have assimilated high-altitude temperature measurements from MLS and SABER into NOGAPS-ALPHA and studied the properties of the resulting analyses and forecasts for the period January–February 2006. The resulting high-altitude temperature analyses for January–February 2006 were minimally biased at most heights and latitudes. Furthermore, indirect fields such as zonal winds and highly-derived diagnostic quantities such as EP flux and residual velocity vectors yielded physically sensible results throughout the mesosphere up to \( \sim 0.01 \) hPa. These fields were all useful for understanding the deep circulation, temperature and eddy-flux anomalies that developed in the winter hemisphere during this period. The high-altitude analyses also provided initial conditions that reduced RMS forecast errors in medium range forecasts of the winter upper stratosphere and mesosphere.

In the winter hemisphere, the correlation coefficient between the observations and background (6–9 h) forecasts (O*F) is high (\( \sim 0.8–1.0 \)) from the lower stratosphere to the upper mesosphere. Thus these short 6–9 h background forecasts capture much of the geophysical variance in the MLS and SABER observations. However, the O*F correlation is lower (0.4–0.6) over the same pressure range in the tropics and in the summer hemisphere, where the zonal temperature variance is smaller. The O-F standard deviation increases monotonically with altitude at all latitudes. This increase in the O-F RMS error with altitude is likely due to the progressively greater concentration of dynamical variability at small spatial and temporal scales and larger divergent wind component at high altitudes. The smaller temporal scales are underresolved by the 6-h 3DVAR analysis and the smaller spatial scales are underresolved by MLS and SABER relative to the forecast model.

The temperature measurements have been assimilated here using a coarse time resolution 3DVAR algorithm. A recent study by Sankey et al. (2007) examined the impact of the update method on the wave energy in the stratosphere and mesosphere. They found that the unfiltered update method used here produced significant excess gravity wave energy in the mesosphere due to the propagation of unrealistic model-resolved gravity waves resulting from the data assimilation process at altitudes below 1 hPa. This problem may be further exacerbated here by the use of stratospheric and mesospheric limb measurements which have sparser spatial sampling than most nadir sounders. We plan to investigate the use of both nonlinear normal-mode initialization of analysis increments (Errico et al., 1988; Ballish et al., 1992) and other incremental analysis update methods (Bloom et al., 1996) which produce analysis fields that generate less spurious gravity wave energy in the forecasts. We are also exploring the use of NAVDAS-AR (accelerated representer), a 4DVAR assimilation algorithm that is currently being tested for operational use (Rosmond and Xu, 2006).

Edited by: W. Ward

References


Shepherd, T. G., Koshyk, J. N., and Ngan, K.: On the nature of large-scale mixing in the stratosphere and mesosphere, J. Geo-


