Point-by-point response to Referee #1

We thank Referee #1 for the thoughtful comments and suggestions, which certainly helped to improve the clarity of the manuscript. Please find below our detailed point-by-point reply to the comments, which we hope have addressed all satisfactorily.

This paper is an important and substantive contribution that, subject to my comments below, certainly merits publication in ACP. Its conclusions are credible and will be valuable for modelers.

Thank you very much!

I do have lots of comments; their overall intent is to improve the paper and make it more useful to the community. An important editorial comment is that the paper is simply too long. There is a tradeoff that must be made between comprehensiveness and readability. These two parameters often anti-correlate. In this case, I believe the authors have leaned too far in the direction of comprehensiveness at the expense of readability. I also think there are some references that they should add- these are noted in the text below.

Reply: See detailed responses below.

For example, Section 2 is too long by probably a factor of two. These datasets are mostly all mature and have been used for this problem in the past. For example, we do not need to be reminded that (line 31 on page 8) that SABER temperatures require non-LTE calculations in the retrievals or the details of when SABER is yawed north or that SABER's duty cycle is still nearly 100% (line 24-25). Rather, there is a rich literature of observational studies of these elevated stratopause events that could and should easily substitute for much of the text in Section 2. For example, although the paragraph on MLS is not too long, I am quite surprised that none of Gloria Manney's work was cited- for instance, her 2009 ACP paper which used the MLS data for the 2006 event much in the same fashion as presently done. Similarly, is there any usage here of ACE-FTS data that differs significantly from Randall et al., GRL, 2009 (also missing from the reference list)? And given that the Funke et al., 2014 JGR papers are cited, the usage of the MIPAS data should just cite those works; again, it is not necessary to tell us how non-LTE vibrational distributions were modeled (line 25, page 7). Finally, there must be dozens of papers which discuss SABER temperatures; two that might be useful to cite here are Siskind et al., GRL, 2007 for the 2006 event and Yamashita et al, JGR, 2013 (see their Figure 7 which shows 4 winters compared with MERRA).

Reply: We agree that Section 2 was rather long and has therefore been shortened. In the revised version we have removed those details on the instruments and retrievals that are available elsewhere. However, we think that some details are relevant for this paper and should be maintained even if provided in previous work. This is the case for the information on sampling patterns and data gaps during the period of interest, as well as on accuracies and known biases.

We have further added the mentioned references for previous observational studies about elevated stratopause events, based on the same instruments, in order to better put our work into the context of existing work. Note, however, that these studies employed different data versions compared to those used in the present work.

A similar comment concerns the models, although the situation is not as bad and

Table 1 is useful. Nonetheless,, given the recent detailed discussions of HAMMONIA by Meraner et al and the discussions of WACCM by Randall et al. (2015), they should probably cut back their descriptions. For example, the citation to Meraner et al is sufficient to discuss HAMMONIA's gravity waves; all the information presented here on the source spectrum etc is superfluous and detracts from readability.

Reply: The model descriptions have been shortened (in particular, the information already listed in Table 1 is not repeated again in the text). However, we think that detailed information on the non-orographic GW parameterizations should be maintained, as it is relevant for the paper and it helps to understand the individual model results.

Other comments

1. I simply do not understand Figure 1. I don't see any symbol in Figure 1 which says "UBC". So how can the figure be showing it when none of the symbols do? What do the arrows show? I don't understand what the deviations are. Is this variance, standard deviation? What is the mathematic expression they are using? They should have two panels- one for absolute values, and one for whatever these deviations are. Then, the text needs to discuss this carefully, not simply with some parenthetical reference.

Reply: We apologise that Figure 1 of the discussion paper is difficult to understand. The plot shows daily averaged NOx mixing ratios from satellite observations and those of the upper boundary condition sampled at the respective observations' time and location. We have improved both figure and caption for the revised version to make this clearer (see new figure below).
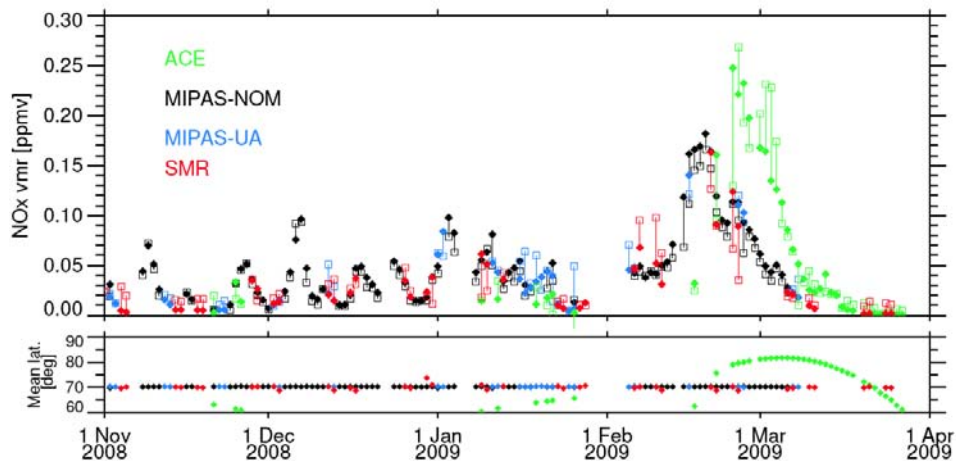
Figure 1. Upper panel: Daily averaged NO_x mixing ratios from satellite observations (open squares) at 0.022 hPa within 60–90°N (black=MIPAS-NOM, blue = MIPAS-UA, red= SMR/Odin, green = ACE-FTS) and those of the upper boundary condition (filled diamonds) sampled at the respective observations' time and location. Lower panel: average latitude of observations. All averages are area-weighted.

2. Figure 3 should be deleted. It adds no new information that is not already clearly shown in Figure 2. I realize that the intent is to illustrate something about timing after the SSW- all I see is a jumble of points. Certainly the spread increases after Feb 1, but I cannot discern anything else.

Reply: Figure 3 has been deleted in the revised version.

3. One misunderstanding that I think I have concerns when exactly during the season does EPP-IE couple most strongly with the stratosphere? From reading Funke et al's two 2014 papers, it appears that significant NOx flux can penetrate into the stratosphere early in the winter, for example, November or December. Indeed in Funke et al., 2014, figure 10, one can see a tongue of EPP-NOy down dipping down to below 40 km on January 1st, much lower in altitude than the descending tongue in the post-SSW period. But in the present paper (page 30), it states that descending NOx can only be distinguished down to 0.3-0.5 hPa. This seems inconsistent and I think bears some explanation.

Reply: The Funke et al. (2014) paper discusses EPP-NOy (the contribution of EPP-generated NOy to total NOy), while in the present study we analyse NOx. Since NOx is converted to other NOy species (HNO3) below approximately 45 km, the NOy descent below this altitude cannot be traced by NOx. Further, dilution of descending NOx with the stratospheric background masks the descent below 45 km in the present analysis, while this is not the case when looking at EPP-NOy. This is now mentioned in the revised version.

Furthermore, if the pre-SSW NOx is more important for its contribution to the stratospheric NOx budget, then isn't the implication of the 2014 papers that the present focus on the post SSW descent is misplaced and of less relevance?

Reply: The higher relevance of the pre-SSW NOx descent for the stratospheric NOy budget is particularly the reason why we included it in our analysis at a similar level of detail as the post-SSW descent, though larger deviations between models and observations in the latter case made it necessary to extend the discussion of post-SSW descent.

4. Section 7.1: It would be useful to discuss and justify the selection of CO as a tracer more. While I realize this is popular because of the low stratospheric values (page 27, line 5), I think it should also be stated that CH4 might be easier to simulate since it would not require the details of CO2 dissociation or reaction with OH to be handled so carefully. Indeed the present author has used CH4 (cf. his 2014 paper) as did Siskind in 2015 and Randall has used this for SH studies (her 2007 paper). Furthermore, with the selection of both CO and NOx we have two tracers that are being transported downgradient. Thus how can we know whether the transport we see is advective or diffusive? Is diffusion important in any of the simulations or in any of the model-model differences?

Reply: Indeed we have analysed CH4 in a similar manner as CO (see Figure below), however, we decided not to include these results in the revised paper for the following reasons:

- *Modelled CH4 distributions deviate larger from each other and from the observations than CO does, likely because of differences in the simulated chemical losses, as well as in the transport by the Brewer Dobson circulation. Despite of this, the evolution of CH4 during the ES event behaves qualitatively similar (though with opposite vertical gradient) as CO such that not much new is learned.*

- *CH4 concentrations drop rapidly towards higher altitudes such that observations are getting typically below the noise level above 0.1 hPa. The*

useful vertical range is thus much reduced compared to CO.

- *The use of a second tracer that is being transported downgradient (in addition to NOx) is intentional. Since CO has no EPP source, it allows us to assess whether model biases are caused by deficiencies in the transport scheme or in the NOx production scheme (or UBC implementation, in the case of medium-top models).*

Regarding the question whether diffusive or advective transport dominates, we would expect that in the case of a significant diffusive contribution the CO peak would be shifted slightly downwards while the CH4 minimum would be shifted upwards. A qualitative comparison of the figure below and Figure 10 of the manuscript suggests that this is not the case, neither in the observations nor in the models.
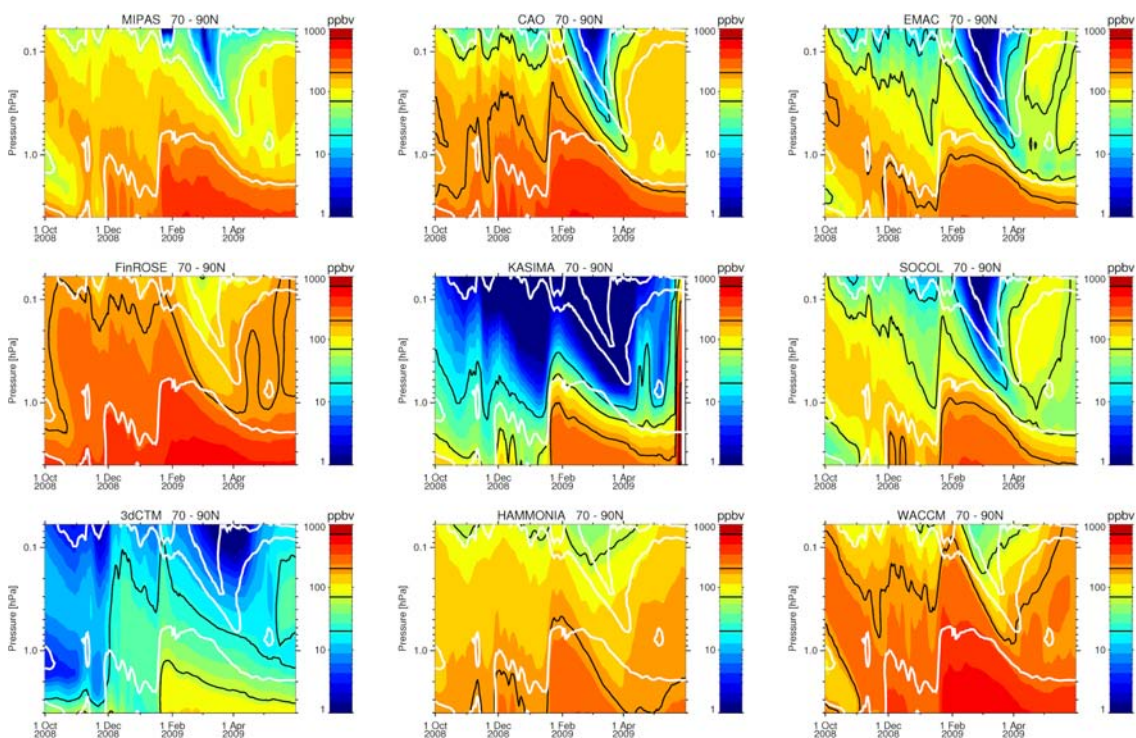
5. There is some discussion of the upper boundary that is used for the medium-top models. But the 2016 Funke paper states that one of its main objectives was to construct such a boundary. How do the values adopted here compare with what is presented in that paper?

Reply: The UBC employed in the present study is based on the same MIPAS observations as used in Funke et al. (2016) for the construction of their semi-empirical model. Their Figure 8 already compares the semi-empirical UBC with the observed NOy (used here as UBC). The purpose of the semi-empirical model is to provide a NOy UBC for long-term climate simulations or for shorter simulations of periods not covered by MIPAS observations. Since this is not the case in the present study, we decided not to use the semi-empirical model but to rely on the observed NOx.

6. Page 39: lines 12-13. I do not see where the consideration of the sampling

patterns has been so important. Figure 19 shows that the temperatures are pretty similar. The only way this sentence can be justified if there were a figure which show a case where the sampling pattern was not considered vs. a case where it was. I don't think they've done this. It would not detract from any of their conclusions if this sentence were deleted.

Reply: We do not fully agree, particularly with respect to the sampling impact on the NOx comparisons. In our opinion, there are significant differences, especially for the ACE sampling. For instance, we show the same model NOx sampled at different locations and times in Figure 2. Here, the ES-related tongue as observed by ACE has apparently a different timing compared to the other instruments. This is caused by the seasonal propagation of latitudes sounded by ACE (see also Fig.1, lower panel) and needs to be considered when comparing to other observations or models.

Editorial comments

1. Line 18, on page 7 seems strange. "MIPAS passes the equator in a southerly direction at 10:00 AM. . . .. . . .observing the atmosphere day and night". Presumably the nighttime data are acquired when MIPAS passes the equator in a northerly direction? This is all phrased more tersely and more clearly in their 2014 JGR paper.

Reply: Following the major first comment above, this sentence has been removed.

2. Page 25, line 1. The proper reference should be Siskind et al., JGR 2010 (not GRL, 2015), which discussed non-orographic drag in great depth. Likewise, consideration should be given to citing Chandran et al., GRL, 2011 who make this point as well.

Reply: Both references are now cited adequately.

3. Page 36, line 10, more grammar: Encountered is a verb and not an adjective and thus does not appear before the noun. It should read: "cold bias encountered at 1 hPa". Likewise page 38, line 3 "spread of the . . .. encountered below 0.1 hPa". And again on page 40, line 3: ". . . encountered during the perturbed. . ." And finally on line 11, page 40.

Reply: Thank you very much for the grammar's corrections. These sentences have been corrected.

Point-by-point response to Referee #2

We thank Referee #2 for the thoughtful comments and suggestions, which certainly helped to improve the clarity of the manuscript. Please find below our detailed point-by-point reply to the comments, which we hope have addressed all satisfactorily.

1) The authors raise the issue of assimilated data biases as a factor in the poor model performance at capturing transport in the wake of the major SSW in January. But another important issue is the use of nudging restricted below 1 hPa. The mesospheric circulation is not a linear problem and is not fully determined by the dynamical boundary condition at 1 hPa. In fact, multiple solutions are possible for the mesospheric state given the same set of stratopause conditions. In other words, it is not really possible to approximate nudging of the mesospheric circulation by restricting the circulation below 1 hPa. In the case of a CTM like FinROSE, the circulation needs to be specified in the mesosphere and the quality of the ECMWF data used is not clear and likely to be deficient (for example, gravity waves become increasingly important with altitude but are treated as noise in standard data assimilation systems.)

Reply: We agree that restricting the nudging of the models to below 1 hPa leaves the mesospheric circulation largely unconstrained, thus making it more difficult to compare to observations. However, the major aim of the paper is to evaluate the capacity of climate models to simulate a realistic polar winter NOx descent. This could hardly be assessed if the mesospheric circulation was entirely constrained by nudging. The restriction of specified dynamics to the stratosphere is a compromise that is hoped to provide a realistic evolution of mesospheric meteorology by upward control, while still allowing for the assessment of self-generated tracer descent in the models. This has now been clarified in the Introduction.

I would think that nudging restricted to below 1 hPa is a bigger problem than the biases in the assimilated data in the stratopause region. In particular, the non-orographic gravity wave drag schemes are essentially running in free mode in the mesosphere and impose biases since there is no local nudging to offset them. Non-orographic gravity wave drag cannot be constrained to the observed regime by imposing the right winds in the stratosphere (e.g., Scinocca and McLandress, 2005). Various built-in assumptions such as a constant source spectrum will still impact the response.

Reply: Yes, we agree. But, as mentioned by the reviewer, the GWD schemes are running basically in a free-running mode and therefore one would expect a more random-like spread of model results. However, our results suggest a systematic model bias with respect to the temperature distribution and the timing of NOx descent in the lower mesosphere. How can such a systematic behaviour be explained by the lack of nudging above 1 hPa?

The basically free mode simulation of the mesosphere also affects the radiative transfer calculations. Radiatively active trace gas distributions are not likely to be ideally distributed to conform to observations. So the radiative damping impact on the evolution of the state in the wake of the major SSW will deviate from observations.

Reply: This is a good point! The potential impact of tracer distributions on the

radiative cooling/heating is now mentioned in the revised version.

In addition the evolution of the highly nonlinear SSW in the mesosphere is not guaranteed to follow observations even if the wave amplitudes at the stratopause conform to observations. The progression of SSW events is rather complex (Matthewman et al., 2009) and the mesospheric component of a major SSW cannot be trivially constrained by the stratospheric part. The SSW exhibits a high degree of top-down evolution with the mesosphere being an important layer of the atmosphere for this phenomenon.

Reply: We agree that SSWs exhibit a high degree of top-down evolution. However, we'd like to mention in this context that it is presently not known if the mesosphere has any control on the stratosphere or, if the mesospheric SSW precursors are triggered from below.

It appears that relatively good performance of EMAC (Fig. 13, 14, 15, 16) and its application of assimilated data up to 0.2 hPa may not be a coincidence since every scrap of nudging in the mesosphere counts.

Reply: EMAC applies the NOx UBC down to 0.1 hPa. This represents a trivial reason for a better performance with respect to the representation of NOx in the region above this level. Note that this model significantly overestimates the descending NOx amounts in the post-SSW phase below 0.2 hPa (see Figure 17) despite the application of assimilated data in this region.

FinROSE appears to do a better job as well (Fig., 13, 14) but performs worse than EMAC apparently due to issues with the vertical upwelling.

Reply: We'd like to point out that the model performance with respect to NOx descent in the pre-SSW phase (Figures 13 and 14) is generally good (except for those models with identified issues regarding the UBC implementation or missing unresolved GWD).

CTMs use geostrophic balance to infer the upwelling from the temperature and horizontal winds, but this fails to capture the ageostrophic part which is not negligible in the mesosphere due to the amplification of gravity waves and general deviation from geostrophic balance.

Reply: We agree with this view.

In addition, several models linearly ramp nudging above 40 km so that they are not imposing as much of a constraint on the upper stratosphere and lower mesosphere dynamical state as models that fully nudge up to 1 hPa.

Reply: Also agreed. We now mention this explicitly in the revised version.

The authors mention the study of Siskind et al. (2015) which highlighted the need for nudging to much higher altitudes (92 km) to improve simulations of NO descent compared to observations. But there is no focused discussion of the limitations of the nudging approach adopted by most models in this intercomparison. Such a discussion has to be included either in the introduction or in the conclusions sections to put the results from these models in a better context.

Reply: Such a discussion has been added.

References:

N. J. Matthewman, J.G. Esler, A.J. Charlton-Perez and L.M. Polvani: A new look at

stratospheric sudden warmings. Part III. Polar vortex evolution and vertical structure, J. Climate, 22, 1566-1585 (2009).

J. Scinocca, and C. McLandress: The GCM Response to Current Parameterizations of Nonorographic Gravity Wave Drag, J. Atmos. Sci., 62(7), 2394-2412, 10.1175/JAS3483.1, 2005.

HEPPA-II model-measurement intercomparison project: EPP indirect effects during the dynamically perturbed NH winter 2008–2009

Bernd Funke[1], William Ball[2], Stefan Bender[4], Angela Gardini[1], V. Lynn Harvey[5], Alyn Lambert[6], Manuel López-Puertas[1], Daniel R. Marsh[7], Katharina Meraner[8], Holger Nieder[4], Sanna-Mari Päivärinta[3,9], Kristell Pérot[10], Cora E. Randall[5], Thomas Reddmann[4], Eugene Rozanov[2,11], Hauke Schmidt[8], Annika Seppälä[3], Miriam Sinnhuber[4], Timofei Sukhodolov[2], Gabriele P. Stiller[4], Natalia D. Tsvetkova[12], Pekka T. Verronen[3], Stefan Versick[4,14], Thomas von Clarmann[4], Kaley A. Walker[13], and Vladimir Yushkov[12]

[1]Instituto de Astrofísica de Andalucía, CSIC, Apdo. 3004, 18008 Granada, Spain.
[2]Physikalisch-Meteorologisches Observatorium, World Radiation Center, Davos, Switzerland
[3]Earth Observation Unit, Finnish Meteorological Institute, Helsinki, Finland
[4]Karlsruhe Institute of Technology (KIT), Institute of Meteorology and Climate Research (IMK-ASF), P.O. Box 3640, 76021 Karlsruhe, Germany.
[5]Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, USA
[6]Jet Propulsion Laboratory, California Institute of Technology, Pasadenia, California, USA
[7]National Center for Atmospheric Research, Boulder, Colorado, USA
[8]Max Planck Institute for Meteorology, Hamburg, Germany
[9]Department of Physics, University of Helsinki, Helsinki, Finland
[10]Chalmers University of Technology, Göteborg, Sweden
[11]Institute for Atmospheric and Climate Science ETH, Zurich, Switzerland
[12]Central Aerological Observatory, Moscow region, Russia
[13]Department of Physics, University of Toronto, Toronto, Ontario, Canada
[14]Karlsruhe Institute of Technology (KIT), Steinbuch Centre for Computing (SCC), Karlsruhe, Germany.

Correspondence to: B. Funke (bernd@iaa.es)

Abstract.

We compare simulations from three high-top (with upper lid above 120 km) and five medium-top (with upper lid around 80 km) atmospheric models with observations of odd nitrogen ($NO_x = NO + NO_2$), temperature, and carbon monoxide from seven satellite instruments (ACE-FTS on SciSat, GOMOS, MIPAS, and SCIAMACHY on Envisat, MLS on Aura, SABER on TIMED, and SMR on Odin) during the Northern Hemisphere (NH) polar winter 2008/2009. The models included in the comparison are the 3d Chemistry Transport model (3dCTM), the ECHAM5/MESSy Atmospheric Chemistry (EMAC) model, FinROSE, the Hamburg Model of the Neutral and Ionized Atmosphere (HAMMONIA), the Karlsruhe Simulation Model of the Middle Atmosphere (KASIMA), the modeling tools for SOlar Climate Ozone Links studies (SOCOL and CAO-SOCOL), and the Whole Atmosphere Community Climate Model (WACCM4). The comparison focuses on the energetic particle precipitation (EPP) indirect effect, that is, the polar winter descent of NO_x largely produced by EPP in the mesosphere and lower thermosphere. A particular emphasis is given to the impact of the sudden stratospheric warming (SSW) in January 2009 and

the subsequent elevated stratopause (ES) event associated with enhanced descent of mesospheric air. The chemistry climate model simulations have been nudged toward reanalysis data in the troposphere and stratosphere while being unconstrained above. An odd nitrogen upper boundary condition obtained from MIPAS observations has further been applied to medium-top models. Most models provide a good representation of the mesospheric tracer descent in general, and the EPP indirect effect

5 in particular, during the unperturbed (pre-SSW) period of the NH winter 2008/2009. The observed NO_x descent into the lower mesosphere and stratosphere is generally reproduced within 20%. Larger discrepancies of a few model simulations could be traced back either to the impact of the models' gravity wave drag scheme on the polar wintertime meridional circulation or to a combination of prescribed NO_x mixing ratio at the uppermost model layer and low vertical resolution. In March–April, after the ES event, however, modelled mesospheric and stratospheric NO_x distributions deviate significantly from the obser-

10 vations. The too fast and early downward propagation of the NO_x tongue, encountered in most simulations, coincides with a temperature high bias in the lower mesosphere (0.2–0.05 hPa) being likely caused by an overestimation of descent velocities. On the other hand, upper mesospheric temperatures (at 0.05–0.001 hPa) are generally underestimated by the high-top models after the onset of the ES event, being indicative for too slow descent and hence too low NO_x fluxes. As a consequence, the magnitude of the simulated NO_x tongue is generally underestimated by these models. Descending NO_x amounts simulated

15 with medium-top models are on average closer to the observations but show a large spread of up to several hundred percent. This is primarily attributed to the different vertical model domains in which the NO_x upper boundary condition is applied. In general, the intercomparison demonstrates the ability of state-of-the-art atmospheric models to reproduce the EPP indirect effect in dynamically and geomagnetically quiescent NH winter conditions. The encountered differences between observed and simulated NO_x, CO, and temperature distributions during the perturbed phase of the 2009 NH winter, however, emphasize

20 the need for model improvements in the dynamical representation of elevated stratopause events in order to allow for a better description of the EPP indirect effect under these particular conditions.

1   Introduction

The potential impact of energetic particle precipitation (EPP) on regional climate is nowadays becoming recognized. Solar forcing recommendations for the recently launched Climate Model Intercomparison Project Phase 6 (Eyring et al., 2016)

25 include, for the first time, the consideration of energetic particle effects (Matthes et al., 2016). EPP is strongly linked to solar activity and hence to the solar cycle, either directly by coronal mass ejections producing sporadically large fluxes of solar energetic particles or indirectly by the quasi-continuous impact of the solar wind on the Earth's magnetosphere. In the mesosphere and lower thermosphere (MLT), EPP-induced ionisation initiates the production of odd nitrogen and odd hydrogen (the latter below ∼85 km), both of them destroying ozone via catalytic cycles. Odd nitrogen ($NO_x = NO + NO_2$) is long-lived

30 during polar winter and is then regularly transported down from its source region into the stratosphere, to altitudes well below 30 km (e.g., Randall et al., 2007; Funke et al., 2014a). This so-called EPP indirect effect contributes significant amounts of NO_x to the polar middle atmosphere during each winter. EPP-induced ozone changes are thought to modify the thermal structure and winds in the stratosphere which, in turn, modulate the strength of the Arctic polar vortex. The introduced signal could then

2

propagate down to the surface, introducing significant variations of regional climate, particularly in the Northern Hemisphere (NH) (Seppälä et al., 2009; Baumgaertner et al., 2011; Rozanov et al., 2012; Seppälä and Clilverd, 2014; Maliniemi et al., 2014).

At present, many chemistry climate models account for EPP-induced ionization and its chemical impact on the neutral atmosphere, which is required for the simulation of atmospheric EPP effects and ultimately for the investigation of potential EPP-climate links. A comprehensive evaluation of these models' capacity to reproduce observed EPP effects by means of coordinated intercomparison studies is a necessary step towards this goal. The High Energy Particle Precipitation in the Atmosphere (HEPPA) model vs. data intercomparison initiative (Funke et al., 2011) evaluated the chemical response of nitrogen and chlorine species in nine atmospheric models to the "Halloween" solar proton event in late October 2003 with observations taken by the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) on Envisat. Reasonable agreement of observed and modelled reactive nitrogen and ozone changes was found, demonstrating the models' overall ability to reproduce the direct EPP effect by solar protons. However, most models failed to adequately describe the repartitioning of nitrogen compounds in the aftermath of the event which could be attributed to deficiencies in the representation of the D-region ion chemistry and motivated recent model developments (Egorova et al., 2011; Verronen et al., 2016; Andersson et al., 2016).

The observation-based evaluation of the simulated atmospheric effects of magnetospheric particles, which are thought to be of higher relevance for climate, is more challenging because of the quasi-continuous flux of electrons compared to protons, and the difficulty in separating between local production and downward transport of NO_x during polar winter. Although a pronounced dependence of reactive nitrogen enhancements in the polar winter stratosphere and mesosphere on the geomagnetic activity levels has been demonstrated (Funke et al., 2014b), dynamical variability, particularly in the NH, can mask out this effect. In particular, the occurrence of elevated stratopause (ES) events following sudden stratospheric warmings (SSWs) during Arctic winters often causes much larger mesospheric NO_x enhancements than expected from the actual geomagnetic activity level, after a brief NO_x depletion related to the weakened vortex during the SSW. The ability of climate models to adequately simulate tracer transport in Arctic winters, including perturbed winters characterized by SSW/ES events, is therefore crucial to accurately model EPP effects and their possible NH regional climate impacts.

Simulations of mesospheric tracer descent during dynamically perturbed NH winters have been compared with observations in several studies. Using the KArlsruhe SImulation Model of the middle Atmosphere (KASIMA) with specified dynamics below 48 km and prescribed NO_x concentrations from MIPAS night time NO_2 observations above 55 km, Reddmann et al. (2010) calculated the amount of EPP-NO_x entering the stratosphere from July 2002 to March 2004. KASIMA reproduced the MIPAS observations of NO_x entering the stratosphere reasonably well, even during the SSW winter 2003/2004. However, the ability of the model to adequately simulate mesospheric tracer transport could not be tested because of the constrained NO_x in the mesosphere. Salmi et al. (2011) and Päivärinta et al. (2016), in turn, used FinROSE with constrained NO_x at the upper boundary (∼80 km) for both early 2009 and 2012. Their results show that FinROSE is able to qualitatively reproduce the downward descent of NO_x from the MLT region into the stratosphere, but the actual NO_x amounts can differ significantly from the observations. In the case of CTMs, the results are strongly affected by the meteorological data, i.e., a source of uncertainty, used to drive the model. McLandress et al. (2013) used a version of the Canadian Middle Atmosphere Model (CMAM)

that was nudged toward reanalysis data up to 1 hPa to examine the impacts of parametrised orographic and non-orographic gravity wave drag (GWD) on the zonal mean circulation of the mesosphere during the perturbed NH winters 2006 and 2009 in comparison with temperature and carbon monoxide (CO) observations from the Microwave Limb Sounder (MLS) on Aura. They found that non-orographic GWD is primarily responsible for driving the circulation that results in the descent of CO from

5 the thermosphere following the warmings. Randall et al. (2015) investigated the NO_x descent during the Arctic winter/spring of 2004 with Whole Atmosphere Community Climate Model (WACCM) simulations that were nudged to Modern-Era Retrospective Analysis for Research and Applications (MERRA) data. They found that their simulated NO_x, although qualitatively reproducing the enhanced descent after the ES event, was up to a factor of 5 too low compared with satellite observations. This underestimation was attributed to missing NO production by high-energy electrons in the mesosphere in combination with an

10 underestimation of mesospheric descent during the recovery phase after the SSW. Siskind et al. (2015) compared simulations of mesospheric tracer descent in the winter and spring of 2009 with two versions of WACCM, one constrained with data from MERRA which extends up to 50 km and the other constrained to the Navy Operational Global Atmospheric Prediction System-Advanced Level Physics High Altitude (NOGAPS-ALPHA) which extends up to 92 km. By comparison with Solar Occultation for Ice Experiment (SOFIE) data they showed that constraining WACCM to NOGAPS-ALPHA yields a dramatic

15 improvement in the simulated descent of enhanced NO_x and very low methane.

Most of these studies suggest that the model representation of the perturbed dynamics during NH winters with SSWs and ES events has a crucial impact on the simulated amount of NO_x transported into the stratosphere and that a proper parametrisation of unresolved GWD is key to achieving agreement with observations. However, previous studies focused on individual models, making it difficult to assess the overall ability of state-of-the-art atmospheric models to reproduce the EPP indirect effect in

20 NH winters. Comprehensive multi-model intercomparisons addressing dynamically perturbed NH winters, however, have so far been restricted to the assessment of the temperature zonal mean, planetary wave, and tidal variability during the 2009 SSW event in the middle and upper atmosphere (Pedatella et al., 2014), as well as on the impacts on the ionosphere variability (Pedatella et al., 2016). Further, although our knowledge of temperature and tracer distributions in polar winters has dramatically increased with the advent of atmospheric satellite observations, specific intercomparisons and validation efforts focussing on

25 such conditions are sparse. A systematic assessment of this knowledge is therefore essential to quantitatively diagnose the model performance with respect to mesospheric tracer transport under perturbed (and unperturbed) polar winter conditions.

A coordinated intercomparison project focussing on tracer descent and the EPP indirect effect during such a winter was therefore initiated in the frame of the SPARC/WCRP's SOLARIS-HEPPA activity. In this so-called HEPPA-II project, simulations of the NH polar winter 2008/2009 from eight atmospheric models have been compared with observations of temperature

30 and concentrations of NO_x and CO from seven satellite instruments including the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS) on SciSat, the Envisat instruments Global Ozone Monitoring by Occultation of Stars (GOMOS), MIPAS, and the SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY), as well as MLS on Aura, the Sounding of the Atmosphere using Broadband Emission Radiometry (SABER) instrument on the Thermosphere, Ionosphere, Mesosphere, Energetics and Dynamics (TIMED) satellite, and the Sub-Millimetre Radiometer

35 (SMR) on Odin. The 2008/2009 winter was chosen for this intercomparison exercise not only because of its peculiar dynam-

ical conditions, characterized by the pronounced SSW in January and the unusually strong descent of odd nitrogen despite the low geomagnetic activity level around solar minimum, but also because of the availability of a large number of observations from different satellite instruments that allowed for a detailed evaluation of the model simulations. The models included in the comparison are the 3d Chemistry Transport model (3dCTM), the ECHAM5/MESSy Atmospheric Chemistry (EMAC)

5 model, FinROSE, the Hamburg Model of the Neutral and Ionized Atmosphere (HAMMONIA), KASIMA, the modeling tools for SOlar Climate Ozone Links studies (SOCOL and CAO-SOCOL), and WACCM (Version 4). Only three of these models (3dCTM, HAMMONIA, and WACCM) extend up into the lower thermosphere where a large fraction of EPP-induced odd nitrogen production occurs. All other models have their upper lid in the mesosphere and require an odd nitrogen upper boundary condition, accounting for EPP production higher up, for simulating the introduced EPP indirect effect in the model domain.

10 This upper boundary condition (UBC) has been constructed from NO_x observations of the MIPAS instrument taken during the Arctic winter 2008-2009. The chemistry climate model simulations have been nudged toward reanalysis data below 1 hPa while being unconstrained above. The restriction of specified dynamics to the stratosphere is a compromise that is hoped to provide a realistic evolution of mesospheric meteorology by upward control, while still allowing for the assessment of self-generated tracer descent in the models.

15 In this study we report results from the HEPPA-II intercomparison project. A major aim is the identification and characterisation of model biases and their uncertainties in the simulations of the perturbed 2008/2009 NH winter by systematically comparing to the suite of satellite observations. For this purpose, common diagnostics are applied in all comparisons, and the sampling characteristics of the instruments are taken into account. Since the study focusses on the evaluation of the ability of the models to simulate the source and transport of MLT tracers by means of observed quantities (i.e., temperature and trace gas abundances), any more sophisticated analysis, e.g., qualifying the different GW drag parametrisations etc., is outside the scope of this comparison. On the other hand, our analysis should motivate such studies to identify the deficits in key processes of this vertical coupling.

The paper is organised as follows: Section 2 gives an overview on the satellite observations and data products used in this study. Section 3 describes the participating chemistry climate and transport models. The NO_x UBC employed in the medium-

25 top models is described in Sec. 4, and Sec. 5 introduces the intercomparison method. Results of the intercomparisons are discussed in Sec. 6 with focus on the representation of the EPP indirect effect by the high-top models in the upper mesosphere and lower thermosphere and, in Sec. 7, with focus on the upper stratospheric and mesospheric representation in all models.

2   Satellite observations

2.1   ACE-FTS/SciSat

30 The Fourier transform spectrometer ACE-FTS has performed infrared solar occultation measurements from the SciSat satellite since 21 February 2004 following its launch on 12 August 2003 (Bernath et al., 2005). The instrument is a Fourier transform spectrometer operating between 750 and 4400 with a spectral resolution of 0.02cm^{-1}. The SciSat satellite is in a highly-inclined circular orbit (74°) and thus provides measurements from 85 °N to 85 °S over each year with a significant focus on

polar measurements. Up to 30 measurements are made each day by ACE-FTS extending from the cloud tops to ~150 km~~and, from these sets of spectra, profiles of over 30 trace gases and isotopologues, temperature and pressure are retrieved. For this study~~. ACE-FTS observations of temperature, CO and NO_x during Arctic ES winters have been analysed in several previous studies (e.g., Randall et al., 2009, 2015). Here, version 3.0 of the ACE-FTS data set was used which covers 21 February 2004
5 to 30 September 2010.

The ACE-FTS retrieval algorithm is described in Boone et al. (2005) and the specific details of version 3.0/3.5 are provided in Boone et al. (2013). ~~Briefly, an unconstrained non-linear least squares global fitting approach is used to fit the measured and forward modelled spectra. Selected lines in the spectra are used to retrieve pressure and temperature as a function of altitude and then these results are used to retrieve volume mixing ratio (VMR) profiles of the various trace gases from microwindows~~
10 ~~selected for each of the target molecules. The vertical resolution of the ACE-FTS measurements is ~3 km, based on the instrument field-of-view (Boone et al., 2005).~~ NO_x is provided from ACE-FTS using the retrieved NO (6–107 km) and NO_2 (7–52 km) profiles. Above 52 km, where both sunset and sunrise NO_2 concentrations are very small and hence not detectable, the scaled a priori NO_2 profile has been used to extend the NO_x profiles to the higher altitudes. The CO profiles extend from 5–110 km and temperature is retrieved from 15–125 km. The vertical resolution of the ACE-FTS measurements is ~3 km,
15 based on the instrument field-of-view (Boone et al., 2005).

The version 3.5 NO profiles differ from HALOE by -15 to +6% between 27 and 53 km and from summertime MIPAS measurements by -9 to +2% between 36 and 52 km (Sheese et al., 2016a). For NO_2, the bias found between ACE-FTS and a suite of other limb and occultation sounders is better than 18% from 17-27 km and -15% from 28-41 km (Sheese et al., 2016a). For both of these species, a box model was used to apply a diurnal scaling to the ACE-FTS profiles before the comparisons.
20 ACE-FTS CO has been compared with MIPAS and MLS by Sheese et al. (2016b). On average, there is a -11% bias between 28 and 50 km with respect to MIPAS and a bias of ±10%. Based on comparisons with coincident satellite observations (within 350 km and 3 hours), it has been found that ACE-FTS v3.5 temperatures agree to within ±2 K between 15 and 40 km, within ±7 K between 40 and 80 km and within ±12 K between 80 and 100 km (P. Sheese, personal communication).

2.2 GOMOS/Envisat

25 GOMOS was a stellar occultation instrument on the polar orbiting Envisat satellite, operating between 2002–2012 (Bertaux et al., 2010). This satellite has been flying in a sun-synchronous orbit at approximately 800 km altitude~~and crossing the equator at 10:00/22:00 local time. Unfortunately, the communication to the satellite was lost in April 2012.~~ GOMOS consisted of a UV-visible spectrometer~~for wavelengths 250–675 nm~~, two IR channels, and two photometers, measuring the stellar flux through the atmosphere at high sampling frequency. GOMOS measured vertical profiles of O_3, NO_2, NO_3, H_2, O, O_2, and aerosols
30 in the middle atmosphere. ~~The altitude range varies for each constituent. For example, for ozone the altitude range is about 16–100 km, whereas for the altitude range in non-polar conditions is 20–50 km and extends up to 70 km in polar winter when enhanced amounts of are present in the atmosphere (Seppälä et al., 2007; Hauchecorne et al., 2007). The altitude sampling resolution of GOMOS varies between 0.4 km and 1.7 km depending on measurement geometry (Tamminen et al., 2010).~~

After application of Tikhonov type smoothing, the target resolution of the ozone observations becomes 2 km below 30 km and 3 km above 40 km. For all other gases the target resolution is 4 km (Kyrölä et al., 2010) .

Here, we have used GOMOS NO_2 profiles (version GOPR_6.0c_6.0f) measured in night time conditions (solar zenith angle at tangent point location $> 107°$; solar zenith angle at spacecraft location $> 90°$ to avoid stray light). In addition to the night time condition, only occultations where the temperature of the star was > 6000 K were selected. This mainly influences the precision of the ozone observations, while The altitude range for NO_2 is less affected in non-polar conditions is 20–50 km and extends up to 70 km in polar winter when enhanced amounts of NO_2 are present in the atmosphere (Seppälä et al., 2007; Hauchecorne et al., 2007) . The typical precision of the NO_2 measurements is 5-20% while the systematic error of the NO_2 observations is estimated to be of the order of few percent (1-5%) (Tamminen et al., 2010; Verronen et al., 2009). Vertical resolution is 4 km (Kyrölä et al., 2010) . As NO is quickly converted into NO_2 by reaction with O_3 after sunset, the night time GOMOS NO_2 measurements used here are a reasonable representation of stratospheric and lower mesospheric NO_x.

Because stars are used as the light source, the locations of the observations change with time. A representative distribution of the latitudes sampled during the course of a year can be seen in Figs. 7–9 of Bertaux et al. (2010). Due to this sampling, for the NH polar region in winter 2008–2009, GOMOS night time NO_2 observations were available for the period of December 2008–January 2009. GOMOS measurements provide the constituent profiles as number densities. For the purpose of this study these were converted to volume mixing ratios using temperature and pressure profiles from the WACCM model (see below).

2.3 MIPAS/Envisat

The MIPAS instrument (Fischer et al., 2008) on Envisat provided global stratospheric and mesospheric measurements of temperature (García-Comas et al., 2014), NO and NO_2 (Funke et al., 2014a), CO (Funke et al., 2009), as well as numerous other trace species (e.g., von Clarmann et al., 2009, 2013) during 2002–2012. Here, we use observations taken in the nearly continuous nominal observation mode (scanning range 6–70 km, hereinafter referred to as MIPAS-NOM), as well as occasional special mode observations (middle and upper atmospheric observation modes covering 20–100 km and 40–170 km, respectively, hereinafter referred to as MIPAS-UA), the latter taken with a frequency of about 1 out of 5 days. We also use special mode UA observations which include three orbits per day passing the 20°W–70°E and 160°E–110°W sectors during 14-–18 January and 21—27 January 2009 and which were taken as support for the Dynamics and Energetics of the Lower Thermosphere in Aurora 2 (DELTA-2) campaign (Abe et al., 2006). MIPAS passes the equator in a southerly direction at 10:00 a.m. local time 14 to 15 times a day, observing the atmosphere during day and night with global coverage from pole to pole and a horizontal along-track sampling between 275 and 410km depending on the observation mode.

Temperature and trace gas profiles have been retrieved from calibrated geolocated limb emission spectra with the scientific MIPAS level 2 processor developed and operated by the Institute of Meteorology and Climate Research (IMK) in Karlsruhe together with the Instituto de Astrofísica de Andalucía (IAA) in Granada. The general retrieval strategy, which is a constrained multi-parameter non-linear least squares fitting of measured and modelled spectra, is described in detail in von Clarmann et al. (2003) . Its extension to retrievals under consideration of non-LTE (i.e. , , and) is described in Funke et al. (2001) . Non-LTE vibrational

~~populations of these species are modeled with the Generic RAdiative traNsfer AnD non-LTE population Algorithm (GRANADA) (Funke et al., 2012) within each iteration of the retrieval.~~

MIPAS-NOM NO_x data have been built from NO and NO_2 data versions V5r_NO_220 and V5r_NO2_220, respectively. MIPAS-UA NO_x data is based on data versions V4o_NO_501/611 and V4o_NO2_501/600. In the middle- to high-latitude polar winters, typical vertical resolutions are 4—6 km below 50 km and 6—9 km above, while the single measurement precision is on the order of 5–15%. Systematic errors, dominated by non-LTE-related uncertainties of NO and NO_2, have been estimated to be less than 10%. CO data (version V5r_CO_220) used here have a single measurement precision ranging from 20–30% above 45 km to 70–80% in the lower stratosphere. The vertical resolution is 6—12 km. The single measurement precision of temperature data (versions v5r_T_220 and v5r_T_521/621 for MIPAS-NOM and MIPAS-UA, respectively) is 0.5—2 K below 70 km and 2—7 K above. The systematic error is typically 1—3 K below 85 km and 3–11 K above. The average vertical resolution is 3—6 km below 90 km, and 6—10 km above.

2.4 MLS/Aura

The Microwave Limb Sounder (MLS) instrument (Waters et al., 2006) was launched on 15 July 2004 and measures thermal microwave emission from Earth's limb. On each day MLS provides ∼3500 vertical profiles of temperature and trace gases between 82°S and 82°N spaced ∼1.5° apart along great circles following the orbit track. ~~Version~~ Manney et al. (2009) employed MLS data version 3.2 to analyse tracer transport during the Arctic ES winter 2006. Here, we use version 4.2 temperature and CO ~~are used here~~. Temperature is deemed useful for scientific studies between 316 hPa and 0.001 hPa. The vertical resolution is 5 km near 40 km and increases to ∼10 km near 90 km (Livesey, 2016). In the mesosphere, systematic and random errors are 2.5 K and comparisons with correlative measurements show a 0–7 K cold bias (Schwartz et al., 2008). CO is recommended for scientific use from 215 hPa to 0.0046 hPa (Pumphrey et al., 2007). The vertical resolution is 4–5 km in the stratosphere and 6-7 km in the mesosphere. Froidevaux et al. (2006) indicate that the CO data have a 25–50% positive bias in the mesosphere. Estimates of absolute accuracy are 10% (Filipiak et al., 2005). For this work, ~~version 4.2~~ temperature and CO data have been filtered using the precision, status, quality, and convergence values provided by the MLS science team (Livesey, 2016).

2.5 SABER/TIMED

The SABER ~~(Sounding of the Atmosphere using Broadband Emission Radiometry)~~ instrument is a 10 channel limb scanning radiometer ~~. It scans the Earth's limb continuously recording profiles of infrared radiance from the atmosphere in discrete spectral intervals (Russell et al., 1999) . The specific wavelength bands observed by SABER were chosen so that a variety of data products could be retrieved or derived, including kinetic temperature, ozone, water vapor, carbon dioxide, atomic oxygen, atomic hydrogen, rates of energy deposition, rates of energy loss, and rates of radiative heating and cooling. The instrument was launched~~ (Russell et al., 1999) , launched in December 2001 onboard the NASA ~~Thermosphere-Ionosphere-Mesosphere Energetics and Dynamics (TIMED) mission~~ in December 2001. The TIMED satellite is in an orbit inclined 74° with respect to the equator. In order to keep SABER within its allowable range of operating temperatures, the spacecraft executes a rotation about its yaw axis every 60 days. While~~ TIMED mission. SABER was measuring in the "northward" viewing mode ~~the~~

8

instrument views from approximately (83°N to 52°S~~, after the yaw maneuver the instrument views from approximately 52°N to 83°S. SABER began acquiring scientific data in January 2002 and is still operating with nearly 100% duty cycle. Here we use the temperature measured from 1 October 2008 until 31 May 2009. In that period SABER was measuring in the "northward" viewing mode~~) during the sub-periods: 1 October–17 November in 2008; 11 January–15 March and 18 May–31 May in 2009. The rest of the days, i.e., 17 November 2008–15 January 2009 and 15 March–19 May in 2009, it was observing in the "southward" viewing mode ~~. During those 60-day yaw periods SABER data are available during day and nighttime at almost all local times.~~

SABER's temperatures are ~~derived from measurements of the CO_2 limb radiance at 15 μm, in the altitude range from 20 to 100 km, and thus require non-LTE calculations in the retrievals. The inversion of temperature and pressure (version~~ (52°N to 83°S). There is a rich literature dealing with the analysis of SABER temperature version 1.07 ~~) are described in detail by Remsberg et al. (2008) . SABER vertical sampling is 400m and its vertical resolution is about 2km. Typical SABER single measurement estimates of v1.07 temperature~~ (Remsberg et al., 2008) in the context of NH polar winter dynamics (e.g., Siskind et al., 2007; Our study uses data from the Level 2A files of version 2.0. Typical single measurement random errors are <0.5 K below 55 km, ~~1~~1–2 K ~~at 70km, 2K at 85km and~~ in the mesosphere and ∼7 K ~~at 100km~~above. The systematic errors are <1.5 K below ~~55~~75 km, ~~0.5K at 70km,~~ 4 K at 85 km and 5 K at 100 km (Remsberg et al., 2008; García-Comas et al., 2008). ~~This study uses data from the Level 2A files of version 2.0. Nevertheless, the v2.0 estimated systematic and random errors are not expected to change much from v1.07 since the uncertainties of their sources are the same.~~The vertical resolution is about 2 km. A thorough comparison of these temperatures with those measured by other satellites, MIPAS, ACE-FTS, MLS, OSIRIS, SOFIE and by lidar measurements, has been recently carried out by García-Comas et al. (2014) in ~~the~~ a study about the validation of MIPAS vM21 temperatures. The comparison of SABER v2.0 with MIPAS vM21 is remarkably good, with differences smaller than 2 K at all altitudes and seasons, except for high-latitude summers above 65 km where they are 3–4 K at 65-80 km (MIPAS colder) and 5-7 K around the mesopause (MIPAS warmer).

2.6 SMR/Odin

The SMR (Sub-Millimetre Radiometer) instrument is a limb emission sounder aboard Odin, a Swedish-led satellite launched in 2001 in cooperation with the Canadian, French and Finnish space agencies (Murtagh et al., 2002). ~~It was initially a joint astronomy and aeronomy mission and, until 2007, the observation time was equally divided between the two disciplines. The satellite became a European Space Agency (ESA) third-party mission in 2007, and is entirely dedicated to atmospheric measurements since the same date.~~ Odin is orbiting the Earth in a sunsynchronous orbit at an initial altitude of 580 km and at Equator-crossing times varying between 06:00 and 07:00 am/pm local time. These parameters are slightly changing with time due to the drifting orbit.

SMR is measuring globally a variety of trace gases and the temperature from the upper troposphere to the lower thermosphere ~~. For this purpose, it uses four sub-millimetre channels (486.1–503.9, 541.0–558.0, 547.0–564.0, 563.0–581.4 GHz) and one millimetre-wave channel (118.25–119.25 GHz) (Merino et al., 2002) . The observation of different species requires switching from one channel to another.~~ (Merino et al., 2002) .

Nitric oxide is retrieved from the observation of thermal emission lines in a band centred around 551.7 GHz. The version 2.1 of NO data is used in this study. The overall vertical coverage is from 7 to 115 km, and in the altitude range considered here the vertical resolution is about 7 km (Pérot et al., 2014). NO data is available approximately four days per month after 2007, on an irregular basis of two observation days in a 14-day cycle. Systematic errors amount to 3% from spectroscopic parameters, 2% from calibration, and 3–6% from sideband suppression (Sheese et al., 2013). The single measurement retrieval error amounts to 44-48%, in the case of Antarctic night time mesosphere-lower thermosphere, as studied by Sheese et al. (2013). A comparison study performed by Bender et al. (2015) showed that SMR NO measurements were consistent with NO measurements by SCIAMACHY, MIPAS and ACE-FTS, despite the different measurement methods and retrieval strategies used for these four instruments.

2.7 SCIAMACHY/Envisat

The SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY ~~(SCIAMACHY, see Burrows et al. (1995); Bovensmann~~ (SCIAMACHY, see Burrows et al., 1995; Bovensmann et al., 1999) is a limb-sounding UV–vis–NIR spectrometer on Envisat. ~~SCIAMACHY comprises eight spectral channels from 214nm to 2380nm with spectral resolutions ranging from 0.22nm to 1.48nm.~~ Among the main measurement modes, the nominal limb mode carried out limb measurements from ground to 105 km until mid-October ~~2003. After~~ 2003, and after 15 October 2003 ~~, the nominal mode was restricted~~ up to 91 km~~top altitude~~. From July 2008 until ~~the end of Envisat in~~ April 2012, SCIAMACHY carried out a special mesosphere–lower thermosphere mode (MLT), scanning from 50 km to 150 km for one day every two weeks. ~~The average horizontal distance between the individual limb scans was about seven degrees in both cases.~~

Nitric oxide is retrieved from the NO gamma bands ~~observed with SCIAMACHY's~~ (UV channel 1~~(,~~ 230–314 nm) ~~(Bender et al., 2013, 20~~ ~~The tomographic orbit retrieval was carried out from 60~~(Bender et al., 2013, 2017) in the 60–160 km ~~to 160km and from 90S to 90N on a fixed 2km×2.5altitude–latitude grid~~range using a tomographic approach. The retrieval from the MLT mode ~~delivered~~ yields the NO number densities with a vertical resolution of 5–10 km ~~at altitudes from~~ between 70 km ~~to~~ and 150 km. With the nominal mode, ~~this~~ the same resolution is achieved between 65 km and 80 km. The average single orbit measurement error amounts to about 30%. Systematic errors amount to 7% from uncertain spectroscopic data, 3% from uncertainties in the solar spectrum (Chance and Kurucz, 2010), and about 10% from temperature uncertainties. ~~As~~ Because the NO gamma bands are ~~excited by absorption of solar light~~fluorescent emissions, the retrieval of NO is restricted to daylight observations. Polar winter data are therefore restricted to latitudes equatorward of the polar night terminator (around 70° in the mesosphere/lower thermosphere at winter solstice).

The retrieved NO number densities from the MLT mode have been compared to ACE-FTS, MIPAS, and SMR (Bender et al., 2015). The measurements were found to be consistent among all instruments with SCIAMACHY retrieving slightly lower densities compared to the other instruments during polar winter but higher values in mesospheric polar summer and mid-to-low latitudes.

Table 1. Summarized description of the atmospheric models involved in this study.

High-top model	vertical domain (km)	horizontal resolution	vert. res. (km)	meteorological data nudging[a]	family approach[a]	kinetic data[b]	EPP-NO_x production
3dCTM	~10–150	2.5°×3.75°	~1–3	LIMA (ECMWF < 1 hPa)	no	S06	AIMOS 1.2
HAMMONIA	~0–250	1.9°×1.9°	~3	ERA-I (<1 hPa)	no	S06	AIMOS 1.6
WACCM	~0–140	1.9°×2.5°	~1.5	MERRA (<50 km)	no	S11	auroral prod.
Medium-top model							NO_x UBC range (hPa)
CAO-SOCOL	~0–80	3.75°×3.75°	~2	ERA-I (<1 hPa)	no	S06	0.01
FinROSE	~0–80	6°×3°	~2–7	ECMWF (whole model domain)	no	S06	0.03–0.01
KASIMA	~7–120	2.8°× 2.8°	0.75–3.8	ERA-I (<1 hPa)	no	S03	0.03
EMAC	~0–80	2.8°×2.8°	~1–4	ERA-I (<0.2 hPa)	reduced	S11	0.09–0.01
SOCOL	~0–80	2.8°×2.8°	~2	ERA-I (<1 hPa)	no	S11	0.01

[a] see model descriptions in Sec. 3 for details.

[b] S11: Sander et al. (2011b), S03: Sander et al. (2003), S06: Sander et al. (2006),

3 Chemistry climate models

In the following, the participating atmospheric models are described and details on the setup of the simulations are provided. Since the dynamical evolution in the mesosphere is strongly constrained by the behaviour of the lower atmosphere, particularly during a perturbed NH winter, model simulations have been either nudged to or rely entirely on meteorological reanalysis data in order to allow for comparisons to observations. High-top models, having their upper lid above 120 km and including explicit schemes for consideration of NO_x production by particle induced ionization, are described in Sec. 3.1. Medium-top models, having their upper lid around 80 km, are described in Sec. 3.2. These models applied a common odd nitrogen UBC in order to account for EPP production above the model domain (see Sec. 4). A summary of the different model settings and characteristics is given in Table 1.

3.1 High-top models

3dCTM

3dCTM is a global 3-dimensional chemistry-transport model developed based on the chemistry scheme of the SLIMCAT model (Chipperfield, 1999) and the transport scheme of the CTM-B Sinnhuber et al. (2003) for use in the middle atmosphere up to the lower thermosphere. It runs on fixed pressure surfaces from 300 hPa (about 10 km) to 4.96×10^{-6} hPa (about 150 km), with a vertical spacing ranging from 1 km in the lower stratosphere and at the mesopause, to about 3 km in the upper stratosphere,

~~lower mesosphere, and lower thermosphere. The horizontal resolution is about 3.75° in longitudinal, about 2.5° in latitudinal direction.~~ Temperature as well as horizontal and vertical wind fields are prescribed by data from the LIMA general circulation model (Berger, 2008), and the model upper boundary is defined by the availability of these data. For the version used here, LIMA is nudged to ($1°\times1°$) ECMWF operational data with a constant nudging of temperature, zonal and meridional winds between the surface and 35 km, and a linear decrease in nudging strength to 45 km, the upper limit of the nudging area. No parametrisation of the gravity wave drag is implemented either in LIMA or in 3dCTM. Only waves with horizontal scales of $\geq$ 500 km and a temporal resolution of 2–12 hours are represented Berger (2008). A comparison of momentum flux climatologies provided in Figure 7 of Berger (2008) with common gravity wave drag schemes as shown, e.g., in Figure 5 of Holton and Zhu (1984), shows that the gravity wave momentum flux in the mesosphere is underestimated by LIMA by about a factor of 2–3 in both the summer and winter hemisphere. In the winter hemisphere, also the vertical structure of the GW momentum flux is somehow different; while Holton and Zhu (1984) essentially show one broad peak at $\sim$65–95 km altitude, varying in strength from -80–120 $\mathrm{ms^{-1}d^{-1}}$, the LIMA profile shows a double peak structure with a broad peak of -40–60 $\mathrm{ms^{-1}d^{-1}}$ at $\sim$70–90 km altitude, a minimum in 90–100 km, and a secondary peak above 100 km. This means that the vertical downward motion throughout the mesosphere will be underestimated during winter.

The model chemistry scheme ~~is based on the JPL-2006 recommendation (Sander et al., 2006) but~~ has been adapted from the original SLIMCAT code for use in the mesosphere and lower thermosphere as described in Sinnhuber et al. (2012): the model considers the photolysis of O_2, CO_2, CH_4, and H_2O in the far-UV wavelength range down to the Lyman α line. Also, in the mesosphere and lower thermosphere, chemical families are not considered for NO_x and O_x species, and H_2O, O_2, and H_2 are now integrated as active chemical species in the model. Additionally, parametrisations for the impact of atmospheric ionization from particle impact and photo-ionisation are considered based on ion-chemistry model studies (Nieder et al., 2014). The photo-ionisation rate is based on the parametrisation of Solomon and Qian (2005); particle impact ionization rates are prescribed using the four-dimensional field provided by the AIMOS model (Wissing and Kallenrode, 2009) version 1.2. Model data are output every 15 minutes, and interpolated onto the satellite geolocations from this.

HAMMONIA

The Hamburg Model of the Neutral and Ionized Atmosphere (HAMMONIA) is an upward extension of the ECHAM5 atmospheric general circulation model (Roeckner et al., 2006). The model's dynamics and radiation are fully coupled to the chemical Model of Ozone and Related Tracers (MOZART, Kinnison et al., 2007). A detailed description of the model is given by Schmidt et al. (2006). To simulate the effects of EPP, HAMMONIA is modified to incorporate the ion chemistry of the E- and F-region as described in Kieser (2011) and Meraner and Schmidt (2016). The ion chemistry treats 5 ion-electron recombinations and 12 ion-neutral reactions including 50 neutral and 6 charged (O^+, O_2^+, N^+, N_2^+, NO^+, e^-) components. Additionally, five reactions directly involving energetic particles are considered. The corresponding reaction rates are calculated using the particle induced ionization rates provided by Atmospheric Ionization Module Osnabrück (AIMOS version 1.6) (Wissing and Kallenrode, 2009). The explicit simulation of energetic particle effects on chemistry is limited to above 10^{-3} hPa, whereas below this altitude the production of $N(^2D)$, $N(^4S)$ and HO_x is parametrised following Jackman et al. (2005a). Pho-

tochemistry includes six reactions involving radiation at wavelengths shorter than Lyman-α. Therefore the parametrisation of Solomon and Qian (2005) and the observed 10.7 cm solar radio flux is used. Orographic gravity waves are parametrised according to Lott and Miller (1997), while non-orographic gravity waves are parametrised according to the Doppler-spread theory from Hines (1997). A geographically uniform isotropic gravity wave source spectrum with a constant root-mean-square (rms)

5   wave wind-speed of 0.8 m/s launched at 830 hPa is used. Additional to the homogeneous source of gravity waves, HAMMONIA considers the generation of gravity waves from tropospheric fronts following Charron and Manzini (2002). At locations where frontogenesis occurs the gravity wave spectrum is launched with an rms wave wind-speed of 2 m/s instead of 0.8 m/s. A more detailed description of the gravity wave scheme used in HAMMONIA is given in Meraner et al. (2016). Note also that this setting of the gravity wave parameters differs from the simulation of the same winter analysed in Pedatella et al. (2014)

10   where the waves were launched at about 650 hPa and no frontal sources were used. ~~In this study, HAMMONIA is run with 119 vertical levels and with a triangular truncation at wavenumber 63 (T63), corresponding to a resolution of about 1.9°in latitude and longitude.~~ Sea surface temperature and sea ice cover are taken from the Atmospheric Model Intercomparison Project 2 (AMIP2) climatology. Output is provided every 2 hours and afterwards interpolated to the satellite geolocations. The model is nudged from 850 hPa to 1 hPa with an upper and lower transition zone to the 6-hourly values of the ERA-Interim reanalysis

15   data (Dee et al., 2011). The 'spin-up' time is one year starting on January 1, 2008.

WACCM

For the simulations presented here, the NCAR Community Earth System Model (http://www.cesm.ucar.edu/, Hurrell et al. (2013)) is used with the Whole Atmosphere Community Climate Model as its atmospheric component (Marsh et al., 2013) (hereinafter referred to as WACCM4). The model is forced with meteorological fields from the Modern Era Retrospective

20   Analysis for Research and Applications (MERRA), a NASA reanalysis using the Goddard Earth Observing System Data Assimilation System Version 5 (Rienecker et al., 2011). The forcing is achieved by relaxing temperature, zonal and meridional winds and surface pressure with a time constant of 50 hours from the surface to 40 km. Above that level the forcing is reduced linearly, so that the model is free-running between 50 km and the model top at approximately 140 km (4.5×10^{-6} hPa). ~~In this 'specified dynamics' version of WACCM4, there are 88 vertical levels and the horizontal resolution is 1.9°latitude by~~

25   ~~2.5°longitude.~~ Heating rates and photolysis are calculated using observed daily solar spectral irradiance based on the empirical model of (Lean et al., 2005) and geomagnetic activity effects in the auroral region are parametrised in terms of the Kp index (Marsh et al., 2007). The standard WACCM chemistry is described and evaluated extensively in WMO (2010). Reaction rates are from Sander et al. (2011b). For these simulations we have modified the $N+N_2$ reaction to include two additional pathways as described in Funke et al. (2008). It should be noted that both WACCM and HAMMONIA use the same chemical solver

30   based on the MOZART3 chemistry (Kinnison et al., 2007), include the same set of ionized species, and use the parametrised EUV ionization rates from Solomon and Qian (2005). For these simulations the latter parametrisation has been extended to include the photoionisation of CO_2 in the EUV. Above 5×10^{-4} hPa (~100 km) ionization from electrons is calculated by the WACCM parametrised aurora. It is assumed that 1.25 N atoms are produced per ion pair and divide the N atom production between ground state, $N(^4S)$, at 0.55 per ion pair and excited state, $N(^2D)$, at 0.7 per ion pair (Jackman et al., 2005b; Porter

et al., 1976). This simulation followed the "REFC1D" protocol of the Chemistry Climate Model Initiative (Eyring et al., 2013) for the specification of time-dependent greenhouse gases and ozone depleting substances. WACCM constituent and temperature profiles were saved at the model grid point and time-step (model time-step is 30 minutes) closest to each of the MIPAS observation locations. Eddy diffusion created by the dissipation of parameterized gravity waves in WACCM depends
5 on the value assumed for the Prandtl number, Pr, which describes the ratio of the eddy momentum flux to the eddy flux of potential temperature or chemical species. In these simulations Pr = 4, as in the study of Garcia et al. (2014).

3.2 Medium-top models

CAO-SOCOL

Since HEPPA-I (Funke et al., 2011) the CCM SOCOL (modeling tool for studies SOlar Climate Ozone Links) has been
10 upgraded to version 3 with substantial changes related to the advection of the species. These changes and the detailed evaluation of the new version performance were documented by Stenke et al. (2013). The CCM SOCOL v.3 consists of the MEZON chemistry transport model (Egorova et al., 2003) and MA-ECHAM5, the middle atmosphere version of the ECHAM general circulation model (Roeckner et al., 2006)~~, with 39 vertical levels between Earth's surface and 0.01 hPa (80 km)~~. Dynamical and physical processes in SOCOL are calculated every 15 minutes within the model, while full radiative and chemical calculations
15 are performed every two hours. Chemical constituents are transported using a flux-form semi-Lagrangian scheme (Lin and Rood, 1996), and the chemical solver is based on a Newton-Raphson iterative method taking into account 41 chemical species, 140 gas-phase reactions, 46 photolysis reactions, and 16 heterogeneous reactions. The ~~rate constants of the gas phase and heterogeneous reactions are taken from Sander et al. (2006) . The~~ CCM SOCOL v.3 was installed in CAO (Central Aerological Observatory, Moscow, Russian Federation) and modified to use assimilation of the meteorological fields from the ERA-I
20 reanalysis, which is necessary to reproduce the considered SSW and ES events in January 2009. The model is nudged from 850 hPa to 1 hPa using the Jeuken et al. (1996) approach. Orographic gravity waves are parametrised according to Lott and Miller (1997). Non-orographic gravity waves are parametrised using Hines (1997) scheme implemented to ECHAM5 with a constant root-mean-square wave wind-speed of 1.0 m/s introduced at 830 hPa for all geographical locations. The daily mean NO_x mixing ratio at 0.01 hPa from MIPAS measurements (see Sec. 4) was used as the upper boundary condition at the
25 uppermost model layer. The NO_x mixing ratio was divided between NO and NO_2 according to their ratio in the model for any particular time step at the second layer from the model top. Model output was interpolated in time and space to the provided satellite geolocations. ~~For the HEPPA-II experiment, the CCM SOCOL was run with T31 horizontal spectral truncation, which corresponds approximately to 3.75° × 3.75°.~~

EMAC

30 The ECHAM/MESSy Atmospheric Chemistry (EMAC) model is a numerical chemistry and climate simulation system that includes sub-models describing tropospheric and middle atmosphere processes and their interaction with oceans, land and human influences (Jöckel et al., 2010). It uses the second version of the Modular Earth Submodel System (MESSy2) to

link multi-institutional computer codes. The core atmospheric model is the 5th generation European Centre Hamburg general circulation model (ECHAM5, Roeckner et al., 2006). For the present study we applied EMAC (ECHAM5 version 5.3.02, MESSy version 2.50) in the T42L90MA-resolution~~, i. e., with a spherical truncation of T42 (corresponding to a quadratic Gaussian grid of approx. 2.8 by 2.8 degrees in latitude and longitude) with 90 vertical hybrid pressure levels up to 0.01hPa.~~

5 ~. The model is nudged to ERA-Interim reanalysis data from the surface to 0.2 hPa (with decreasing nudging strength in the transition region in the five levels above) using the nudging coefficients suggested in Jeuken et al. (1996). The upper boundary condition for NO_x is prescribed in the top 4 layers (0.01 hPa to 0.09 hPa) of the model. For gravity waves we used the submodel GWAVE which contains the original Hines non-orographic gravity wave routines (Hines, 1997) from ECHAM5 in a modularised structure. We tuned the parameter rmscon (root-mean-square wind-speed at bottom launch level of 642.9 hPa),

10 which controls the dissipation of gravity waves, to 0.8 m/s. For gas phase reactions we used the submodel MECCA (Sander et al., 2011a) and for photolysis the submodel JVAL (Sander et al., 2014). 110 gas phase reactions and 44 photolysis reactions were included. ~~Kinetical constants and absorption cross-sections have mainly been taken from (Sander et al., 2011b) .~~ The NO_x family was reduced to NO and NO_2. The chemical tracers were initialized from a multi-annual EMAC model run. Model output was done for each time step (10 minutes) which afterwards was interpolated to the satellite geolocations.

15 **FinROSE**

FinROSE is a global 3-dimensional CTM (further developed model version of the one described by Damski et al. (2007)). The model dynamics for the whole model domain is forced with external meteorological data, whereas the vertical wind is calculated inside the model by using the continuity equation. In this study FinROSE is nudged with ECMWF operational analysis data. This means that changes in the atmospheric composition do not affect the model dynamics, and gravity wave

20 parameterization is included already in the meteorological forcing data. FinROSE reproduces the distributions of 41 species from the stratosphere up to the mesosphere and lower thermosphere, and includes also about 120 homogeneous reactions and 30 photodissociation processes. ~~Chemical kinetic data, reaction rate coefficients, and absorption cross sections are taken from look-up-tables based on the Jet Propulsion Laboratory compilation by Sander et al. (2006) and regularly updated from the available supplements.~~ Photodissociation frequencies are calculated using a radiative transfer model (Kylling et al., 1997).

25 In addition to homogeneous chemistry, the model also includes heterogeneous chemistry, i.e., formation and sedimentation of polar stratospheric clouds (PSCs) and reactions on PSCs. The model is designed for middle atmospheric studies and thus the chemistry is not defined in the troposphere, but the tropospheric abundances are given as boundary conditions. For this study, the UBC for NO_x (i.e. NO + NO_2) was implemented in the MLT region at about 0.03–0.01 hPa (the top two model layers). ~~The model was run with 41 vertical levels (~0–80 km) with a horizontal resolution of 6°×3° (longitude×latitude).~~ Output in

30 the satellite geolocations was composed already during the model run by finding the closest model grid point and time step to every geolocation.

KASIMA

The KASIMA model is a 3D mechanistic model of the middle atmosphere including full middle atmosphere chemistry (Kouker et al., 1999). The model can be coupled to specific meteorological situations by using analysed lower boundary conditions and nudging terms for vorticity, divergence and temperature. Here the version used for the HEPPA I experiment has been applied
5 (Funke et al., 2011) but with a horizontal resolution of about $2.8° \times 2.8°$(T42). The ~~model has 63 pressure levels between 7 and 120 km (chemistry up to 90 km) and a vertical resolution in the lower stratosphere of 750 m, gradually increasing to 3.8 km at the upper boundary. The~~ frequency of output is every 6 hours. The model is nudged to ERA-Interim analyses below 1 hPa. A numerical time step of 6 min was used in the experiments. The model uses a Lindzen-type parametrisation (Holton, 1982) to include the effect of breaking gravity waves , but no specific parametrisation of orographic gravity waves. Further details of
10 the model are found in Funke et al. (2011). The UBC for NO_x was set at the 0.3 hPa level, and not above. This occasionally causes deviations between the observations and the model above this level.

SOCOL

The applied version of the CCM SOCOL improves upon CAO-SOCOL, and was prepared for participation in the IGBP/SPARC CCMI project. The tropospheric chemistry component was extended by adding the Mainz Isoprene Mechanism (MIM-1),
15 which comprises 16 organic species and a further 44 chemical reactions (Poeschl et al., 2000). The cloud influence on photolysis rates was introduced using a cloud modification factor (Chang et al., 1987). Interactive lightning source of NO_x was introduced following the Price and Rind (1992) approach and adopting local scaling factors based on satellite measurements. The kinetic constants and absorption cross-sections were updated following Sander et al. (2011b). The new parametrisation of the UV heating rates (Sukhodolov et al., 2014) as well as NO_x and HO_x production by energetic particles (Rozanov et al., 2012) were
20 adopted. For HEPPA-II the model was run with T42 horizontal resolution, which corresponds approximately to $2.8° \times 2.8°$, and 39 vertical levels between the ground and 0.01 hPa. The nudging set-up and UBC for NO_x are the same as in CAO-SOCOL.

4 NO_x upper boundary condition for medium-top models

The UBC for NO_x mixing ratio has been constructed from MIPAS-NOM observation data versions v4o_NO_200 and v4o_NO2_200 by projecting individual observations onto a regular grid in longitude, latitude, pressure level, and time with daily cadence us-
25 ing a distance-weighting algorithm. All observations taken within ± 12 h time difference, $\pm 10°$ latitude, and $\pm 25°$ longitude have been considered at each grid point (weighted by the inverse distance squared) and have been vertically interpolated to a fixed pressure grid. Data gaps in space and time have been filled by linear interpolation. Note that in the model-measurement intercomparisons a newer version of MIPAS NO_x is used, which was not available when the upper boundary condition was generated prior to the model runs. The horizontal resolution of the NO_x UBC is $1.25° \times 2.5°$(latitude $\times$ longitude). Thirteen
30 vertical pressure levels within 1–0.01 hPa are covered to allow for interpolation to the respective upper lid of the models. The NO_x UBC has been evaluated by comparing with available satellite observations (see Fig. 1). To avoid sampling errors in the
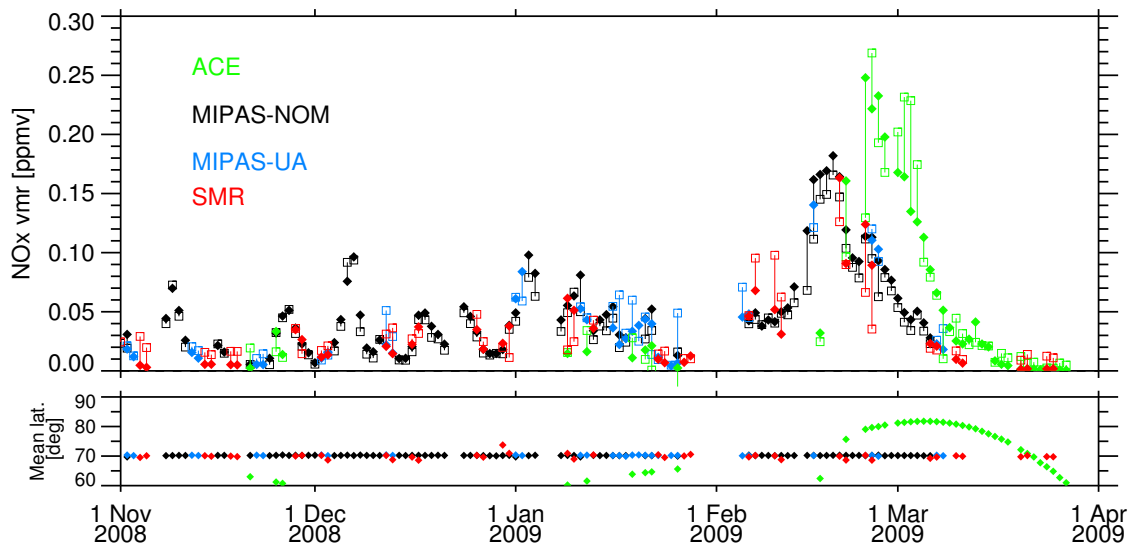
Figure 1. Upper panel: Daily averaged NO$_x$ mixing ratios ~~of the upper boundary condition~~ <u>from satellite observations (open squares)</u> at 0.022 hPa ~~and deviations from daily averaged satellite observations~~ within 60–90°N (black=MIPAS-NOM, blue = MIPAS-UA, red= SMR/Odin, green = ACE-FTS) ~~, represented by arrows (arrow head:~~ <u>and those of the</u> upper boundary condition <u>(filled diamonds)</u> sampled at the <u>respective</u> observations' time and location~~, arrow origin: observation, with the arrow length representing the bias between both)~~. Lower panel: ~~average~~ <u>mean</u> latitude ~~of~~ <u>averaged over all</u> observations <u>of the individual instruments within 60–90°N</u>. All averages are area-weighted. ~~Note that the arrows show up as as triangles in the case that the upper boundary condition coincides with the observations.~~

comparisons, the UBC field has been sampled at the measurements' locations of each day before averaging over the polar cap region. In general, there is very good agreement (within 10–20%) with independent NO$_x$ observations. However, larger differences up to 20–50% occur sporadically for observations close to the vortex edge (e.g., when comparing to ACT-FTS at the end of February) where horizontal gradients are very pronounced.

5   Intercomparison strategy

The discrete horizontal sampling of satellite observations can cause large uncertainties in intercomparisons of observed and modelled averaged quantities, particularly if the sampling is sparse, irregular, or variable in time (Toohey et al., 2013). To reduce the impact of sampling errors, we follow the same approach that was successfully applied in the first HEPPA intercomparison study (Funke et al., 2011): the model output has been sampled at the locations and times of the individual observations and has been vertically interpolated to the observed pressure levels. If available (i.e., in the case of MIPAS and MLS) , averaging kernels have been applied to the model results as described in Funke et al. (2011). Profiles have only been considered in the vertical range where the instruments' sensitivity is high enough to provide meaningful data; the remaining profile regions have been excluded in both observations and model results.

17

Model-measurement comparisons were performed on basis of daily and/or quasi-monthly averaged zonal mean data, which have been calculated in the same way for both observations and simulations. For most comparisons, data has been further binned within 70–90°N applying area-conserving $(\cos(\theta))$ weights. Note however, that the sampled portion of this latitude bin varies from instrument to instrument, making a direct comparison of the observational results difficult. However, the comparison of model biases with respect to different observational datasets is mostly unaffected. The binning has been extended to 60–90°N in the comparisons to ACE-FTS data in order to allow for evaluations prior to February 2009. We recall that ACE-FTS has a discrete but time-varying latitude coverage (see Fig. 1) such that the resulting averages represent only a small fraction of the entire bin.

6   Upper mesosphere and lower thermosphere

In this section NO_x, CO, and temperature fields of the high-top models 3dCTM, HAMMONIA, and WACCM are compared to the observations in the MLT, the source region of odd nitrogen produced by EPP. Although, strictly speaking, temperature is not a tracer of vertical motion, the adiabatic warming during periods of strong descent introduces observable changes of the thermal structure of this region which can be used as diagnostics of vertical transport in the models. The simultaneous evaluation of modelled NO_x, CO, and temperature distributions allows then to attribute model biases to deficiencies in the simulation of either particle-induced NO_x production or of dynamics.

Figure 2 shows the vertical distribution of NH polar NO_x over time in the simulations and MIPAS-UA, ACE-FTS, and ODIN-SMR observations at 0.1 to 2×10^{-4} hPa. SCIAMACHY observations of NO densities have not been included in this figure because NH polar observations are only available after the beginning of February. Note that MIPAS-UA and ACE-FTS provided NO_x volume mixing ratios (VMR), while SMR observed NO VMR, only. This, however, introduces differences only below approximately 0.01 hPa since NO_x is entirely in the form of NO above. The comparisons with the three instruments provide a consistent picture of model biases. While WACCM and HAMMONIA reproduce the observations fairly well during the whole time period in the upper mesosphere and lower thermosphere (above the 0.01 hPa level), 3dCTM exhibits too small NO_x abundances in this vertical region. Below the 0.01 hPa level and during the pre-SSW phase of the winter (November–January), WACCM and HAMMONIA agree well with the observations while 3dCTM overestimates NO_x in this vertical region during most of the pre-SSW phase.

The SSW event starts with the breakdown of the polar vortex, and the dilution of the mesospheric NO_x by upwelling and and increased horizontal mixing. This is clearly observed by MIPAS and SMR as a decrease of NO_x between roughly 0.01 and 0.001 hPa. This initial NO_x decrease is captured well by WACCM and 3dCTM, though it is too weak in the HAMMONIA simulation. The initial decrease of NO_x during the SSW is followed by strong downwelling of NO_x leading to a pronounced increase of mesospheric NO_x and the development of the characteristic NO_x-"tongue". This is qualitatively captured by all models, however, the amount of NO_x transported into the lower mesosphere (below 0.01 hPa) is significantly underestimated. The timing of the onset of the enhanced descent varies considerably among the models and, compared to the observations,

18
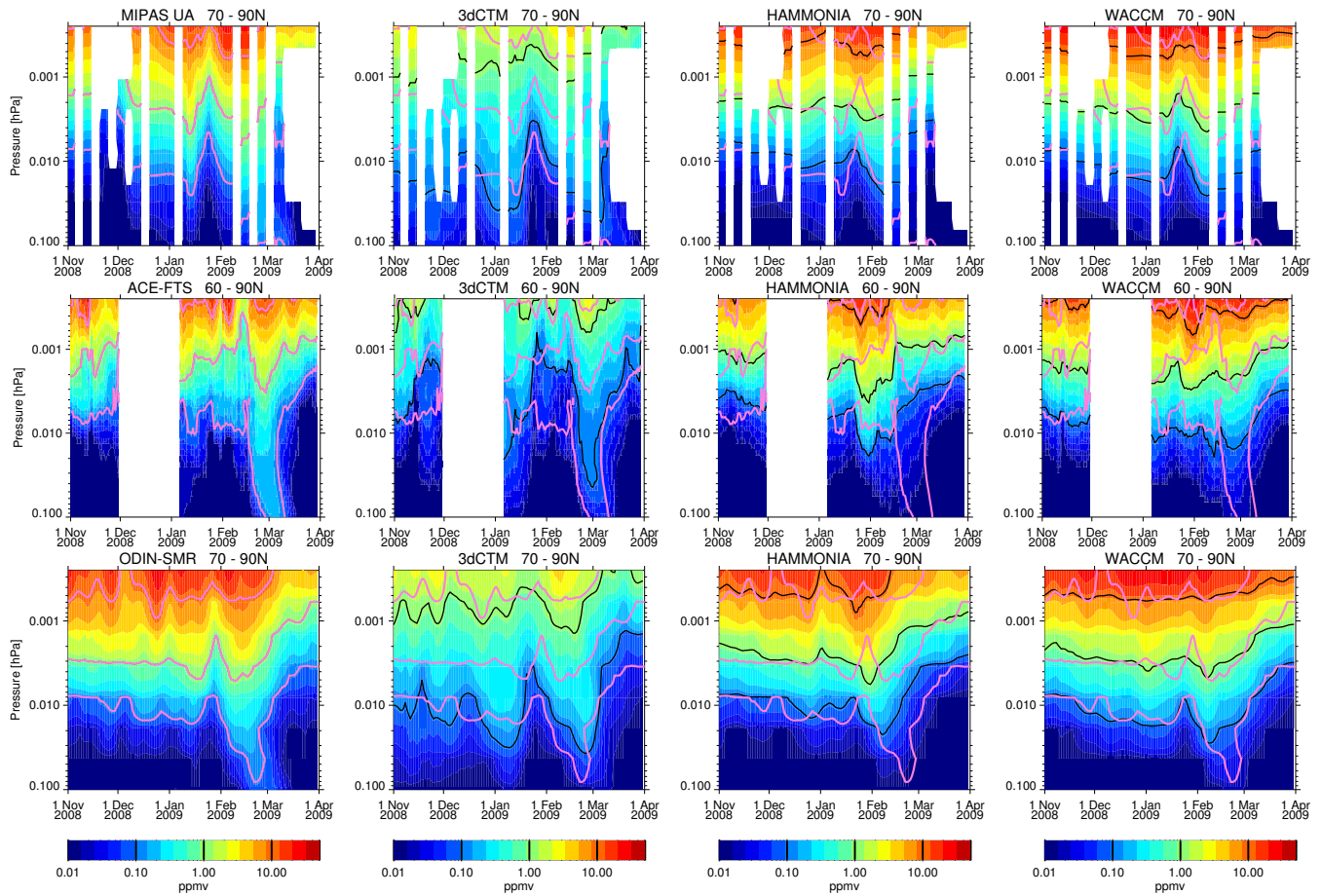
Figure 2. Observed and modelled NO_x VMRs of MIPAS and ACE (upper two rows) and NO of SMR (lower row) in NH polar MLT region during November 2008–March 2009. Model output of the "high-top" models 3dCTM, HAMMONIA, and WACCM has been sampled at the locations and times of the observations (MPAS-UA, ACE-FTS, and SMR) for comparison. Pink lines indicate the observed VMR levels of 0.1, 1, and 10 ppmv. White regions reflect missing or not meaningful data.

occurs slightly too early in HAMMONIA and too late in 3dCTM. The onset of ES-related NO_x increases in WACCM coincides with the observed onset, however, the modelled increases appear to last for a shorter time.

!htDaily averaged observed and modelled at 8×10^{-3}hPa within 70–90°N (60–90°N for ACE-FTS). Blue: 3dCTM, green: HAMMONIA, red: WACCM, black: observations. Model output sampled at the different instrument's locations at times is shown by different symbols: filled diamonds: MIPAS-UA, open squares: SMR/Odin, crosses: ACE-FTS.

These differences in the onset time of enhanced downwelling after the SSW are also visible in Figure **??**, which shows the temporal evolution of the observed and modelled at 8×10^{-3}hPa. The large spread between the concentrations of the same model sampled at different instruments' locations is remarkable during the post-SSW phase (February–March) and is

19
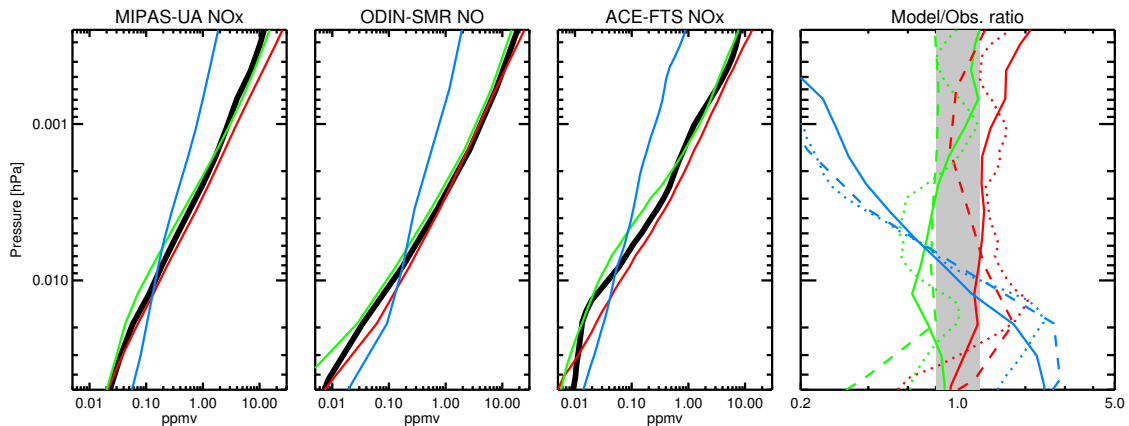
Figure 3. Comparison of observed polar midwinter NO_x mean profiles (thick black lines) a to 3dCTM (blue), HAMMONIA (green), and WACCM (red). Right panel: ratio of model results and MIPAS-UA (solid), SMR/Odin (dashed), and ACE-FTS (dotted) observations. The grey shaded area indicates the $\pm 25\%$ range. Data have been averaged over 70–90°N and 5 December 2008 – 12 January 2009 (60–90°N and 5 November 2008 – 12 January 2009 in the case of ACE-FTS).

~~considerably smaller before the event. This larger spread after the event is caused by the pronounced horizontal and vertical gradients after development of the tongue, highlighting the need to account for instrumental sampling patterns when comparing models to observations during perturbed periods.~~

6.1 Unperturbed early (pre-SSW) phase

5   In the following, the observed and modelled vertical structure of NO_x, CO, and temperature during mid-winter (pre-SSW phase) is analysed in more detail to evaluate the models' ability to reproduce the EPP indirect effect for unperturbed conditions. Figure 3 compares the observed and modelled NO_x mid-winter mean profiles averaged over 70–90°N and 5 December 2008 – 15 January 2009 (60–90°N and 5 November 2008 – 15 January 2009 in the case of ACE-FTS) above the altitude of 0.05 hPa. The observed vertical structure of NO_x is reasonably well reproduced by HAMMONIA and WACCM during

10 this period. Differences with respect to the observations are mostly within 20–50%, with WACCM being overall more on the high side and HAMMONIA more on the low side (particularly at altitudes below 0.002 hPa). As discussed earlier, the 3dCTM simulations show a much less pronounced vertical gradient resulting in a significant (in terms of the observational spread) NO_x underestimation (up to a factor of 8) at altitudes above 10^{-2} hPa and overestimation (up to a factor of 3) below. Figure 4 compares the corresponding mean profiles of CO, observed by MIPAS-UA, MLS, and ACE-FTS above the altitude

15 of 0.5 hPa. Again, WACCM and HAMMONIA show a vertical gradient that is roughly in agreement with the observations. On the other hand, the absolute CO values of WACCM are slightly (up to 40%) higher while HAMMONIA underestimates the CO abundances by a factor of 2—3. The latter can be explained by missing thermospheric production mechanisms in the model, specifically the CO_2 photolysis in the extreme ultraviolet (at wavelengths < 121 nm) and the reaction of CO_2 with the

<center>20</center>
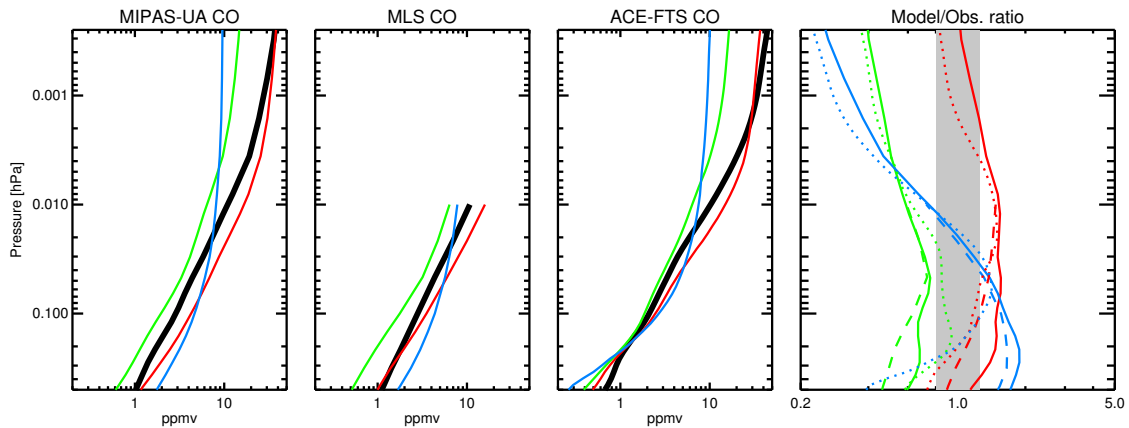
Figure 4. Comparison of observed polar midwinter CO mean profiles (thick black lines) to 3dCTM (blue), HAMMONIA (green), and WACCM (red). Right panel: ratio of model results and MIPAS-UA (solid), MLS (dashed), and ACE-FTS (dotted) observations. The grey shaded area indicates the ±25% range. Data have been averaged over 70–90°N and 5 December 2008 – 12 January 2009 (60–90°N and 5 November 2008 – 12 January 2009 in the case of ACE-FTS).
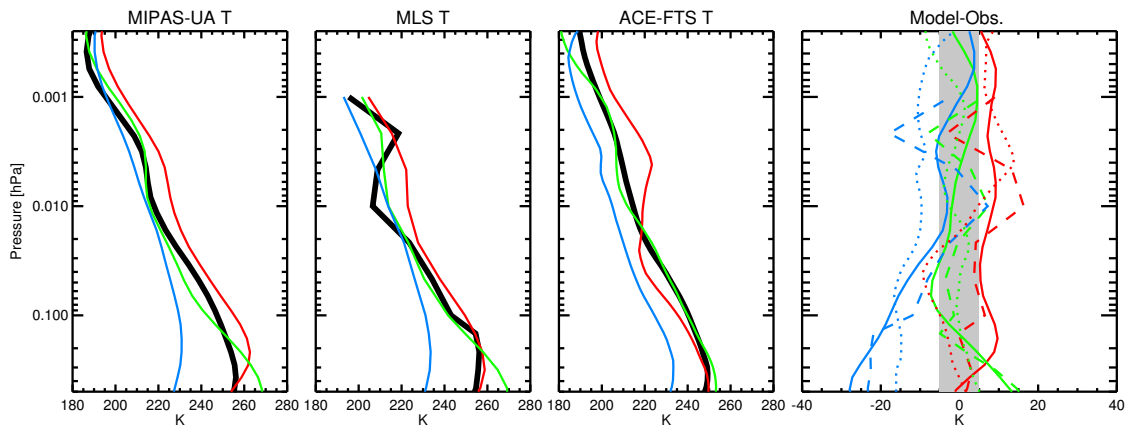
Figure 5. Comparison of observed polar midwinter temperature mean profiles (thick black lines) to 3dCTM (blue), HAMMONIA (green), and WACCM (red). Right panel: temperature difference of the simulations and MIPAS-UA (solid), MLS (dashed), and ACE-FTS (dotted) observations. The grey shaded area indicates the ±5 K range. Data have been averaged over 70–90°N and 5 December 2008 – 12 January 2009 (60–90°N and 5 November 2008 – 12 January 2009 in the case of ACE-FTS).

atomic oxygen ion (Garcia et al., 2014), that act in addition to the photolysis of CO_2 in Lyman-alpha and the Schuman-Runge continuum. The 3dCTM simulations, similarly as for NO_x, show a too weak gradient in the mesosphere compared to the observations, resulting in an underestimation above 0.03 hPa and an overestimation below. The corresponding temperature profiles (see Figure 5), observed by MIPAS-UA, MLS, and ACE-FTS (note that SABER is not included because the observations in

December cover only up to 52°N) indicate good agreement with the observations for HAMMONIA and a slight warm bias of 5–10 K for WACCM. Mesospheric 3dCTM temperatures are systematically too cold by 10–30 K in the middle and lower mesosphere.

The good overall agreement of NO_x, CO, and temperature from HAMMONIA and WACCM with the observations in December suggests that both NO_x sources and dynamical conditions are well represented by these models, allowing for an adequate description of the EPP indirect effect in the MLT during unperturbed conditions early in NH winters. Interestingly, the consideration of mid-energy electron (MEE)-induced ionization in HAMMONIA (via AIMOS) does not introduce noticeable differences in the NO distribution with respect to WACCM, the latter only accounting for auroral electrons. This suggests that the impact of MEE during the solar minimum 2008/2009 NH winter was rather small. 3dCTM simulations, on the other hand, show significant discrepancies with the observations. The similarity of the model bias in the vertical gradients of NO_x and CO suggests that these differences with respect to the observations are due to the representation of dynamics in 3dCTM rather than to the EPP source. The vertical gradient of the 3dCTM CO and NO_x profiles both show too low values in the lower thermosphere, and too high values in the upper to mid mesosphere. The underestimation of lower thermospheric CO is likely due to the model chemistry as, like in HAMMONIA, neither the EUV photolysis of CO_2 nor the production of CO by positive ion chemistry in the lower thermosphere are considered in 3dCTM. The underestimation of thermospheric NO_x could be caused by a too weak NO production or too fast transport out of the (polar) source region, either by horizontal mixing, or across the mesopause. The high values of both CO and NO_x in the mesosphere on the other hand are likely due to the representation of mesospheric dynamics in 3dCTM, which is driven by temperatures and wind fields from the LIMA model. A likely reason seems the neglect of sub-scale ($\leq$500 km) gravity waves in the LIMA model leading to an underestimation of the GW drag throughout the mesosphere, but to an overestimation in the lowermost thermosphere (see Section 2.7). This leads to a suppression of vertical motion in the mesosphere which is also reflected in a negative bias in temperatures, and consequently, to an accumulation of CO and NO_x.

6.2 Perturbed late (post-SSW) phase

Figure 6 compares the observed and modelled NO_x February mean profiles corresponding to the perturbed post-SSW phase of this winter, characterized by enhanced descent of NO_x. This comparison includes also SCIAMACHY NO density averages. Above 0.005 hPa, a larger spread of model-measurement differences compared to December is found, likely related to the enhanced spatial and temporal variability~~(see also Fig. ??).~~ On average, however, these differences are very similar to those encountered during mid-winter. Below 0.005 hPa, all models systematically underestimate the observed NO_x increases associated with the ES event by a factor of 2–3.

Adiabatic heating associated with the enhanced mesospheric descent is responsible for the reformation of the stratopause at a pressure level as high as 0.005 hPa. Figure 7 shows the temporal evolution of the vertical temperature structure at 70–90°N in January–March as observed by SABER and simulated by 3dCTM (LIMA), HAMMONIA, and WACCM. We have chosen this observational dataset for the comparison to the models because of its full temporal coverage in this period and the high vertical resolution in the entire vertical range. The observed elevated stratopause started to develop at the beginning of February and
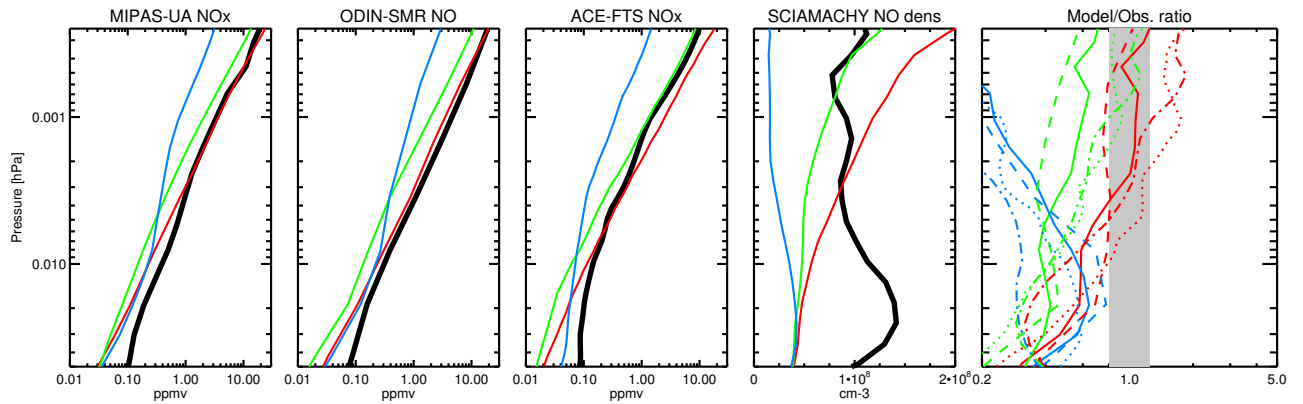
Figure 6. Comparison of observed NO_x mean profiles (thick black lines) for February 2009 (during the ES event) and 70–90°N to 3dCTM (blue), HAMMONIA (green), and WACCM (red). Right panel: ratio of model results and MIPAS-UA (solid), SMR-Odin (dashed), ACE-FTS (dotted) and SCIAMACHY (dash-dotted) observations. The grey shaded area indicates the ±25% range. Data have been averaged over 70–90°N and 1 February 2009 – 1 March 2009 (60–90°N and 1 February 2008 – 15 March 2009 in the case of ACE-FTS).

remained at around 0.01 hPa for a month before it descended to its climatological height in the course of March. The highest stratopause temperatures during the elevated phase were reached around 20 February. Although all models simulate an elevated stratopause, its temporal evolution differs significantly from the observations. HAMMONIA and WACCM show an ES onset and formation level similar to the observed ones, however, highest temperatures at this level are reached immediately after
5   the onset, about 20 days earlier than in the observations. In both models, the ES level starts to descend immediately after its formation, more quickly than observed, and faster in HAMMONIA than in WACCM. During the descent, the modelled stratopauses become increasingly warmer. 3dCTM, in contrast, simulates a much later onset (about 2 weeks after the observed one) and the ES temperatures are much colder than in the observations. However, the modelled ES remains at an elevated level for a longer time (although slightly lower than the observed ES) and the time delay until reaching the maximum ES
10   temperatures is comparable to the observed temperature evolution. These differences between 3dCTM on the one hand, and WACCM, HAMMONIA, and mostly also the observations, on the other hand, highlight the role of subscale gravity waves for the temporal evolution of the ES event. The onset of the SSW event is driven mainly by large-scale planetary waves breaking down the horizontal circulation, and is captured comparatively well by all three models. However, the reformation of the stratopause at upper mesospheric altitudes is driven by small-scale gravity waves reaching up to the upper mesosphere after
15   the event. As these smaller gravity waves are essentially missing in the LIMA data, the build-up of the elevated stratopause is delayed in 3dCTM, and its strength is weaker.

To investigate whether the encountered differences between the models and SABER data are robust with respect to instrumental uncertainties, we extend the analysis to MIPAS-UA, ACE, and MLS temperature observations and compare the model differences to all observations (see Figure 8). Despite minor changes related to the different latitude range covered by the
20   instruments, the encountered model biases are consistent for all instruments, indicating a too cold mesosphere of 3dCTM, and

23
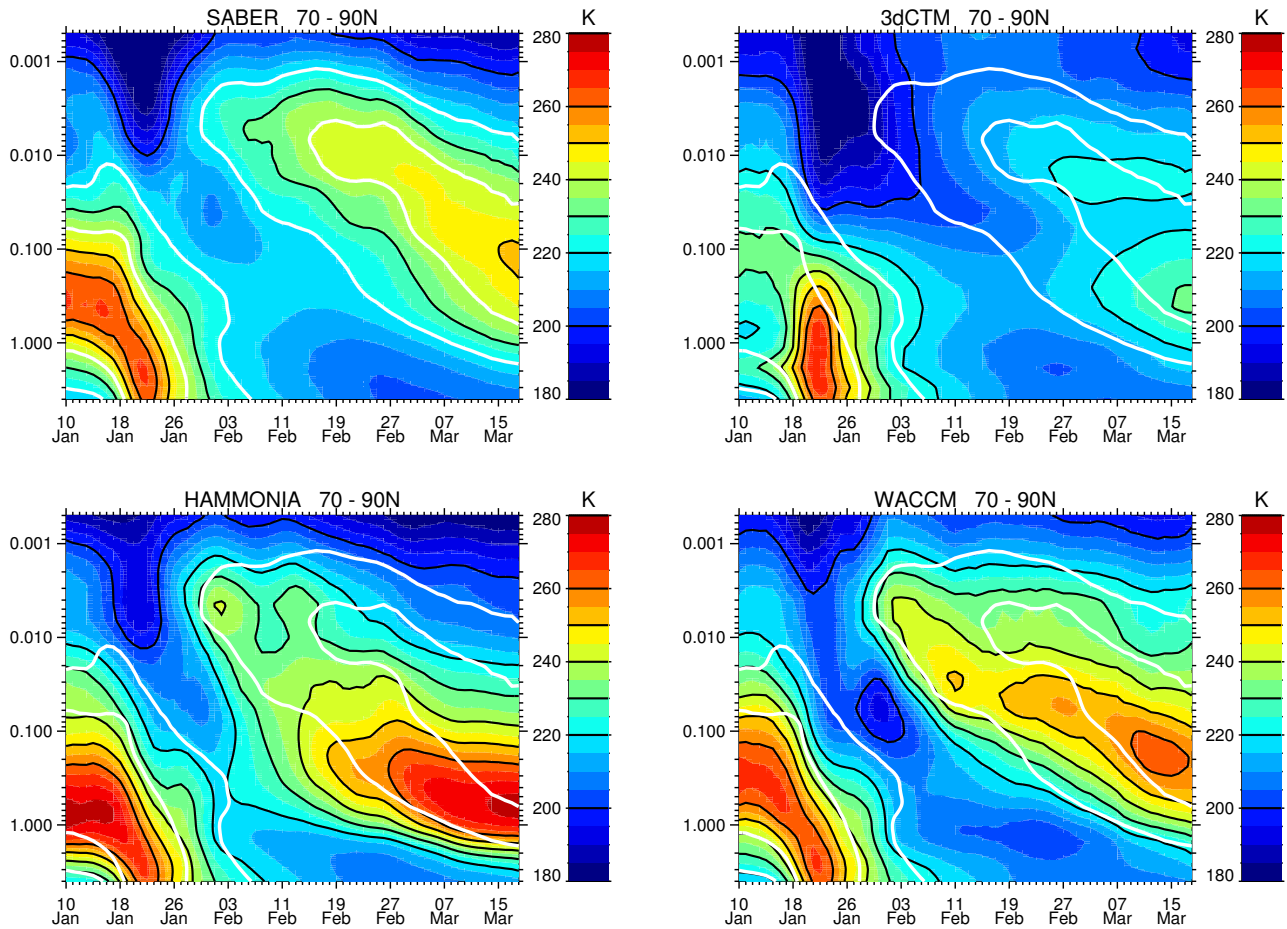
Figure 7. Temporal evolution of daily averaged polar cap temperatures at 4–0.0005 hPa from SABER observations and simulations of 3dCTM, HAMMONIA, and WACCM (from top left to bottom right). The white contours correspond to the observed temperatures of 220 and 240 K.

a dipole type pattern in HAMMONIA and, less pronounced, in WACCM with colder temperatures after the ES onset in the upper mesosphere and warmer temperatures below.

A similar analysis of NH polar temperature evolution in early 2009 in several whole atmosphere models (including HAMMONIA) and MLS observations has been performed by Pedatella et al. (2014). Their Figure 1 can be directly compared to our

5 Fig. 7. In agreement with our results, most of the investigated models in the study of Pedatella et al. (2014) did not maintain the stratopause height near 0.01 hPa until the end of February as in the observations, except WACCM-X, which was nudged to NOGAPS-ALPHA reanalysis data (assimilating observed temperatures) up to 92 km. Siskind et al. (2015) further showed with WACCM simulations of the same NH winter, that nudging to a more realistic meteorology (with an ES evolution closer to the observations) up to 92 km dramatically improves the simulated NO descent during this event compared to SOFIE observations.
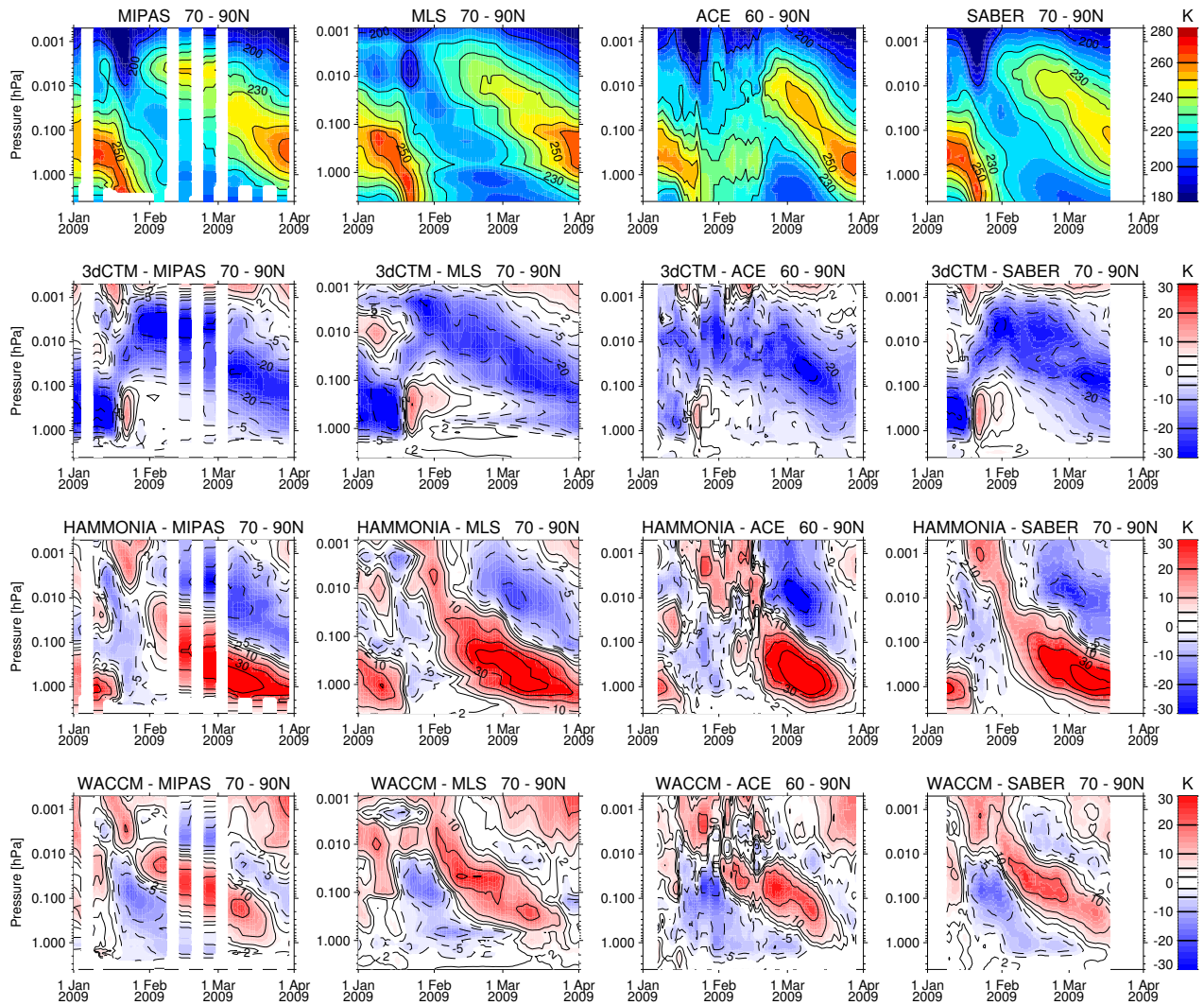
Figure 8. Top: temporal evolution of daily averaged polar cap temperatures at 4–0.0005 hPa observed by MIPAS-UA, MLS/Aura, ACE-FTS, and SABER (from left to right). Bottom: Corresponding differences between temperatures simulated with the "high-top" models (3dCTM, HAMMONIA, and WACCM) and the observations.

Unresolved non-orographic GWD is thought to play a crucial role in the strengthening of mesospheric descent in the vicinity of the NO source region during ES events by providing enhanced westward momentum which forces a poleward and downward residual circulation ~~(McLandress et al., 2013; Siskind et al., 2015)~~ (Siskind et al., 2010; Chandran et al., 2011; McLandress et al., 2013). Motivated by the results of our analysis, Meraner et al. (2016) investigated the sensitivity of the HAMMONIA model to changes in the parametrisation of non-orographic gravity waves. By weakening the amplitude of the gravity waves at the source level, they could substantially improve the modelled temperature and NO_x increases (both in terms of timing and amount) compared
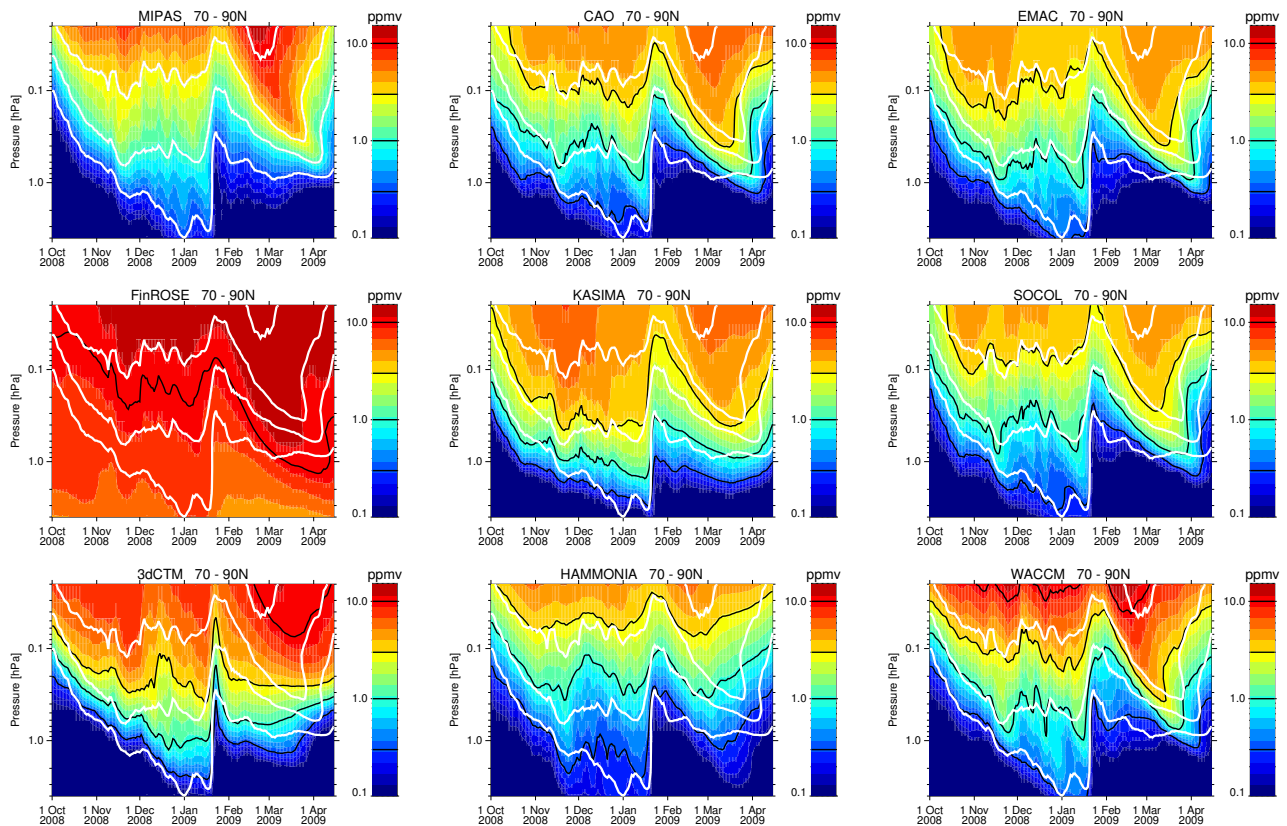
Figure 9. MIPAS-NOM and modeled temporal evolutions of CO at 4–0.02 hPa within 70–90°N. White lines indicate the observed VMR levels of 0.3, 1, 3, and 10 ppmv.

to the MIPAS observations. They found that the amount of transported NO_x depends strongly on the altitude at which momentum is deposited in the mesosphere. Smaller gravity wave amplitudes favour the wave breaking and momentum deposition at higher altitudes, closer to the NO source region. The structural similarities of HAMMONIA and WACCM temperature biases suggest that changes in the non-orographic GWD parametrisation might also improve the representation of NO_x descent during ES events in WACCM.

7 Upper stratosphere and mesosphere

In this section CO, NO_x, and temperature fields of all involved models are compared to the observations in the upper stratosphere and mesosphere (USM). The aim is to evaluate the models' ability to reproduce NO_x transport into the stratosphere during both the unperturbed pre-SSW phase and the ES event, and to identify whether discrepancies with respect to the obser-

26

vations are related to dynamics or chemistry. The latter is of particular concern for the medium-top models applying the NO_x upper boundary condition.

7.1 CO

CO is an excellent tracer of vertical motion in the USM during polar winter because of its pronounced vertical gradient in this

5 region and the long chemical lifetime under dark conditions. Further, the relatively less pronounced gradient at higher altitudes (compared to NO_x) results in a weaker sensitivity to dynamical variability in the MLT, hence allowing to study the descent in the USM separately. In addition, the very low stratospheric CO background concentrations allow to trace mesospheric descent down to altitudes below 30 km without the need to invoke tracer correlations as in the case of odd nitrogen (Funke et al., 2014a).

CO observations are available from MLS, ACE, and MIPAS. As an example, Figure 9 compares the MIPAS-NOM CO

10 temporal evolution with the models. At a first glance, the observed evolution of the CO vertical distribution is qualitatively well reproduced by most models, except for FinROSE which exhibits a very weak vertical gradient all over the winter. This behaviour is caused by a simplified CO_2 representation leading to overestimation of CO production and a largely enhanced CO background in the middle and upper atmosphere. All other models capture the observed polar winter descent down to pressure levels around 3 hPa in the first part of the winter, the sudden reduction of CO during the SSW caused by meridional mixing

15 and upwelling, as well as the enhanced descent during the ES event.

A more quantitative analysis is provided by Figures 10 and 11, comparing the modelled CO evolutions at 0.02 and 0.5 hPa, respectively, to MIPAS-NOM and MLS observations (note that FinROSE is not included here because of the unrealistically high mixing ratios). The comparisons to both instruments provide a very similar picture, hence confirming the robustness of the encountered model biases. Observed CO abundances at 0.02 hPa are around 6–8 ppmv during the pre-SSW phase, decrease to

20 4 ppmv during the SSW, and show a pronounced peak of 12–14 ppmv in February related to the ES event. Medium-top models exhibit slightly lower CO abundances (around 5 ppmv) that do not vary significantly over the winter. This behaviour is expected since transport of lower thermospheric CO into the model domain is typically not considered and, as consequence, dynamically induced variations are mostly absent at this pressure level close to the models' upper lid. As an exception, tracers are transported in KASIMA above the chemical domain at 90 km which causes accumulation effects, resulting in slightly increased abundances

25 during early winter. Further, minor differences in the late winter abundances simulated by KASIMA and CAO on the one hand and EMAC and SOCOL on the other hand can be attributed to the use of different kinetic data in the chemistry schemes, primarily affecting OH involved in the CO loss reaction. The observed CO evolution at 0.02 hPa is qualitatively well captured by WACCM, although the abundances during the pre-SSW phase of about 10 ppmv are overestimated by ∼40% compared to the observations and the ES-related peak occurs earlier than in the observations. HAMMONIA CO abundances are underestimated

30 due to missing thermospheric CO production mechanisms (see previous section) and are very close to the CO amount simulated by the medium-top models (∼5 ppmv). 3dCTM simulates early winter CO abundances that are roughly in agreement with the observations. ES-related CO enhancements in the post-SSW phase, however, are delayed and persist for a longer period than observed.
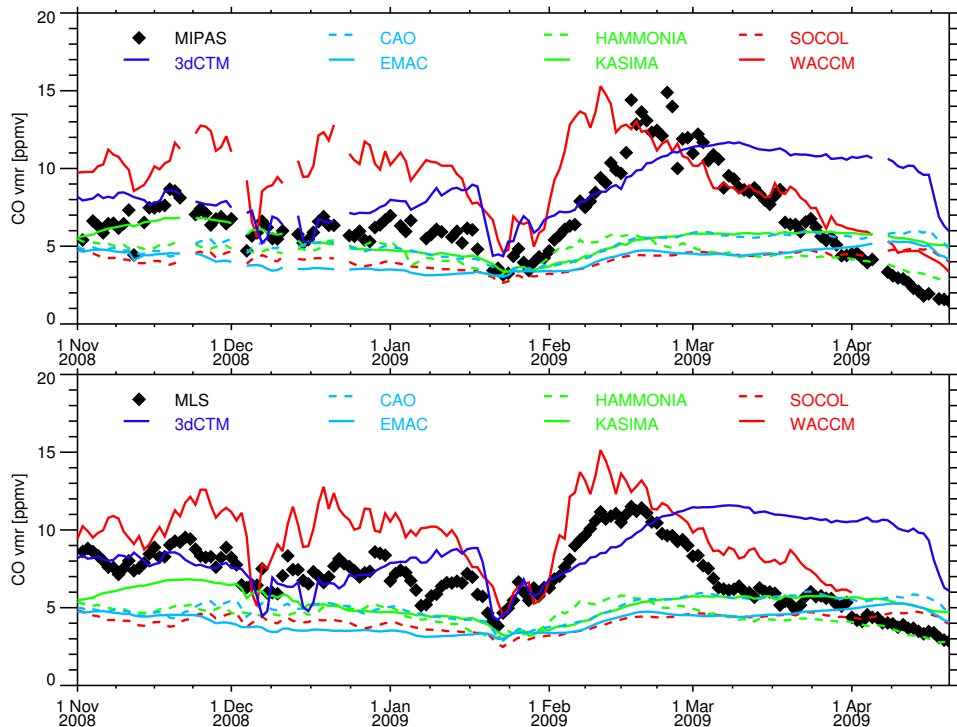
Figure 10. MIPAS-NOM (top) and MLS/Aura (bottom) temporal evolutions of CO VMR in comparison with the model results within 70–90°N at 0.02 hPa.

The observed CO evolution at 0.5 hPa is well reproduced by most medium-top models and WACCM in the pre-SSW phase. KASIMA, and 3dCTM overestimate the CO abundances by a factor of ∼2.5 and ∼1.5, respectively, while HAMMONIA simulates about 50% lower than observed CO abundances. The ES-related CO increases peak in most models too early (around mid March) compared to the observed peak occurrence around 1 April, although the peak magnitude is reasonably well simulated (with exception of HAMMONIA). The CO peak in HAMMONIA occurs even 2 weeks earlier than in the other models. In 3dCTM, the CO tongue does not reach the 0.5 hPa level (see Fig. 9), likely because of the too-late formation of the elevated stratopause discussed in the previous section. The high CO abundances of this model in February, immediately after the SSW, seem to be caused by horizontal mixing, after a short period of localised upwelling during the sudden warming.

The individual impacts of orographic and non-orographic gravity wave drag on the mesospheric CO evolution in the CMAM model has been evaluated by comparing with the same MLS observations during the 2008-2009 NH winter by McLandress et al. (2013). Our Figure 11 can be qualitatively compared to their Figure 8 (although the latter shows the CO evolution at a slightly higher pressure level). The CO evolution in the CMAM simulation, including all gravity wave sources, is very similar to that obtained by most of the models included in our study (note that the apparently smaller time lag of the ES-related peak in the McLandress et al. (2013) study is related to the higher pressure level of their comparison). On the other
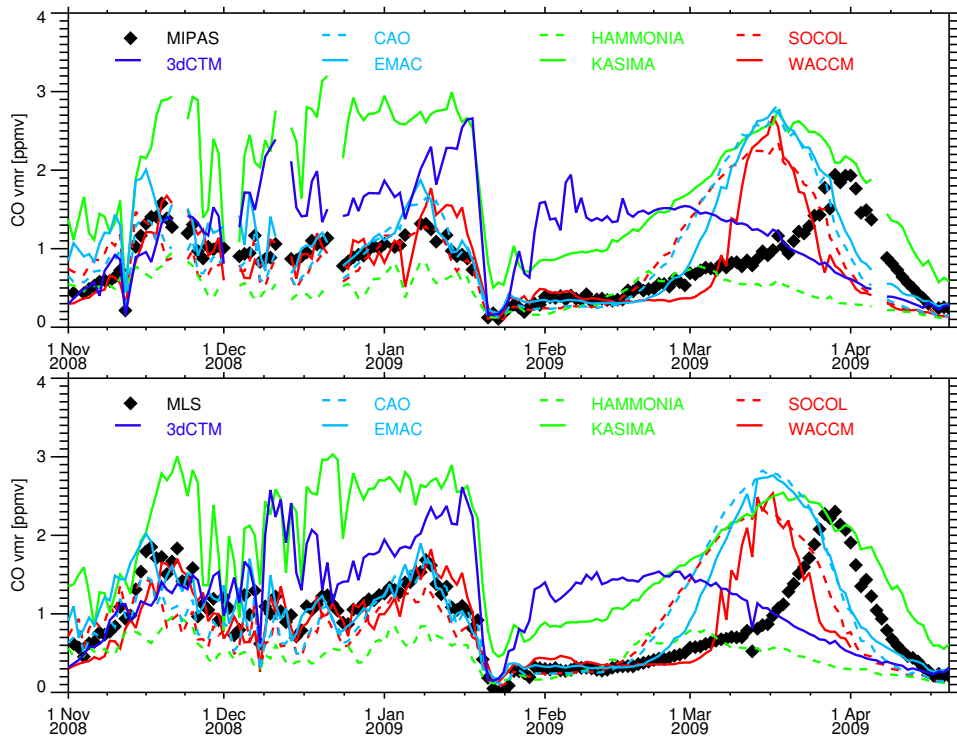
Figure 11. MIPAS-NOM (top) and MLS/Aura (bottom) temporal evolutions of CO VMR in comparison with the model results within 70–90°N at 0.5 hPa.

hand, there are similarities between their simulation without orographic GWD and the KASIMA simulation presented here, particularly regarding the CO overestimation in the pre-SSW phase and the relatively broad CO peak after the ES event. Note, that KASIMA does not employ a specific parameterization for orographic gravity wave drag which may be justified as KASIMA is nudged up to 1 hPa but seems not to be sufficient near the stratopause. This is also seen in the low bias of

5 the stratopause temperature in the pre-SSW phase (see Fig. 18). Further, our 3dCTM results share some characteristics of the CMAM simulation without any GWD. In particular, both simulations exhibit a steady (though fluctuating) increase of CO until the SSW, a short recovery time after the warming, and the absence of an ES-related peak in March/April. This again highlights the importance of the proportion of the gravity wave spectrum not considered in the LIMA model – the sub-scale ($\leq$ 500 km) waves for the mesospheric meridional wintertime circulation, in particular during the recovery phase of the elevated stratopause

10 event as discussed in the previous section, but also for the "undisturbed" pre-event period.

7.2 NO_x in the early (pre-SSW) phase

In the following, the observed and modelled vertical structure of NO_x in the USM during mid-winter (pre-SSW phase) is analysed in more detail to evaluate how well the models reproduce the EPP indirect effect in this region for unperturbed conditions.

29
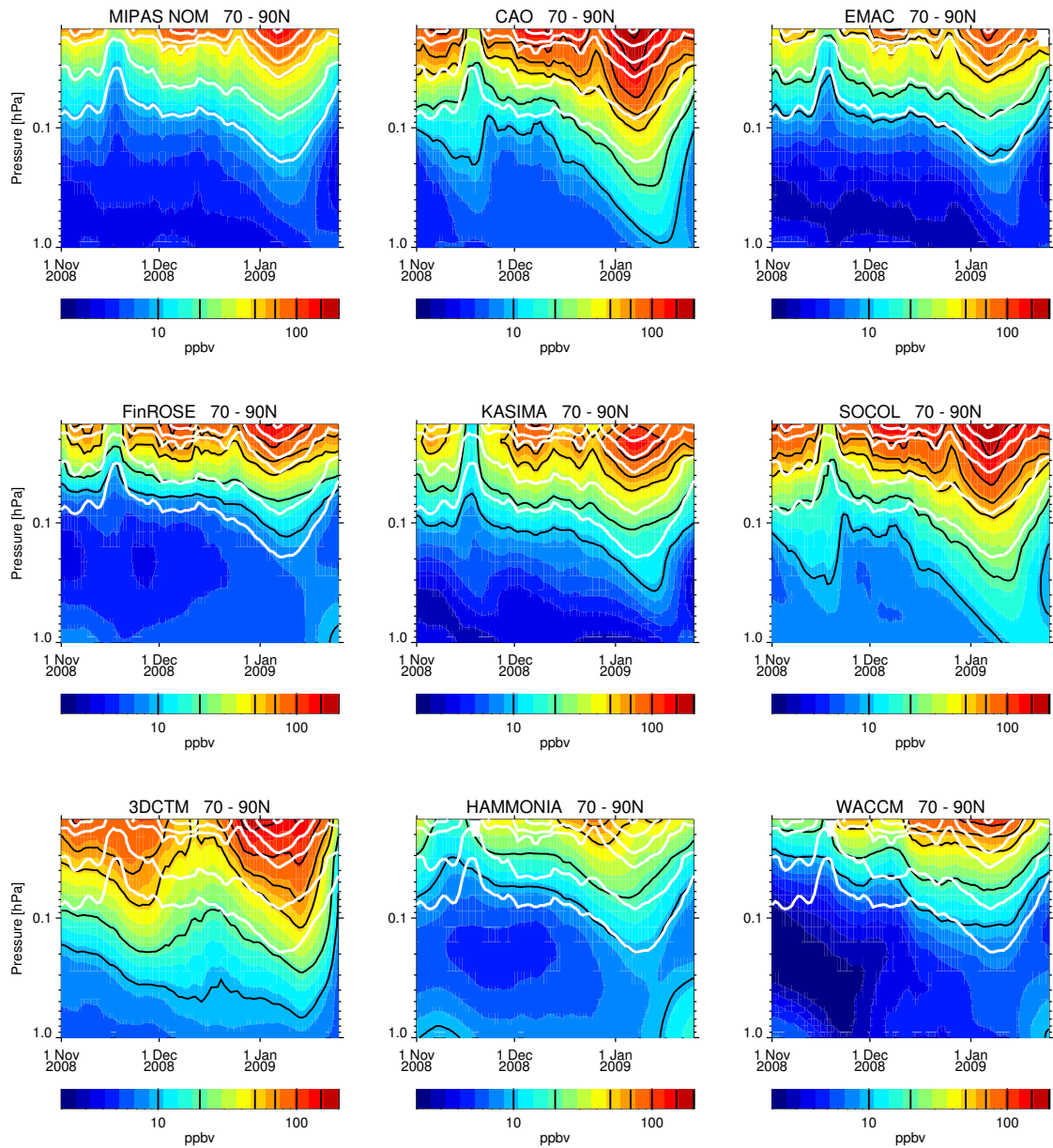
Figure 12. MIPAS-NOM and modelled temporal evolutions of NO_x in the pre-SSW phase of the 2008/09 NH winter at 1–0.02 hPa within 70–90°N. White lines indicate the observed VMR levels of 10, 20, 50, 70, 100, and 150 ppbv. White regions reflect missing or not meaningful data.

Figure 12 compares the NO_x evolution of all models at 1–0.02 hPa with the MIPAS data. All models capture the observed early winter NO_x descent characterized by a quasi-continuous increase of NO_x until the SSW-related disruption in mid-January.

The magnitude of the observed NO_x enhancements is well reproduced by EMAC, FinROSE, KASIMA, HAMMONIA, and WACCM. Descending NO_x can be distinguished from the background in these simulations and in the observations down to pressure levels of 0.3–0.5 hPa. Further descent below this level cannot be traced because NO_x is converted to other reservoir nitrogen species (principally HNO_3) below approximately 45 km. Descent of EPP-generated total reactive nitrogen has been observed down to altitudes as low as 30 km during the pre-SSW phase of Arctic winter 2009 (Funke et al., 2014a).

As discussed in Sec. 6, 3dCTM overestimates the observed NO_x increasingly towards lower altitudes and shows a double peak structure (with a NO_x depletion around mid-December) that is not seen in the MIPAS NO_x data, though a similar feature is also observed in 3dCTM CO, and at least indicated in MIPAS CO, at the same time. Also SOCOL and CAO overestimate substantially the descending NO_x amounts. Since the CO descent is well described by the latter two models, the NO_x overestimation is likely related to the prescription of NO_x at the upper model lid. The NO_x abundances at the upper model level (0.01 hPa) are in agreement with the values specified by the UBC. However, in contrast to the observations and other models, which show a rapid decrease towards lower altitudes, the abundances remain nearly constant in the entire vertical range above 0.03 hPa. This behaviour is caused by a model boundary artefact introducing unrealistically fast vertical propagation of the NO_x caused either by too high vertical velocities at the model lid or low vertical model resolution. Indeed, the descending NO_x amounts are substantially reduced in a test simulation with NO_x prescribed at the second layer from the top (not shown) making the SOCOL results similar to those of EMAC.

A more quantitative view of the modelled midwinter NO_x profiles in comparison with observations of the MIPAS and GOMOS instruments (the latter measuring nighttime NO_2) is provided in Figure 13. Other instruments measuring NO_x species could not be included in this comparison: SMR because they measured only NO but most of NO_x is in the form of NO_2 below 0.1 hPa in dark conditions, SCIAMACHY because it is not sensitive to NO below $\sim$65 km, and ACE-FTS because it did not sample latitudes polewards of 70°N in midwinter. Both MIPAS and GOMOS consistently show VMRs of about 20 ppbv at 0.05 hPa, decreasing to the background values of 5 ppbv at 0.8 hPa. The observed profile is reproduced within 20% by EMAC, FinROSE, HAMMONIA, and WACCM. The KASIMA results are about 50% higher than the observations. 3dCTM, CAO, and SOCOL overestimate the observations by a factor of 2–3.

Overall, most atmospheric models are capable of providing a realistic and consistent picture of NO_x descent in dynamically and geomagnetically unperturbed NH early winters as in 2008/2009. This is the case for high-top models explicitly considering odd nitrogen production by EPP in the MLT region, as well as for medium top models employing a NO_x upper boundary condition. However, some individual models show significant biases in the simulated early winter NO_x descent which could be traced back to deficiencies in either the dynamical or chemical schemes.

7.3 NO_x in the perturbed late (post-SSW) phase

Limitations of high-top models to reproduce quantitatively the observed NO_x descent from the upper mesosphere during the perturbed part of the 2008/09 NH winter (post-SSW phase) have already been discussed in Sec. 6. An important question is whether medium-top models, prescribing realistic NO_x distributions at the model's upper lid, could provide a better description of ES-induced odd nitrogen transport by bypassing the problem of underestimated descent in the region above 80 km, as
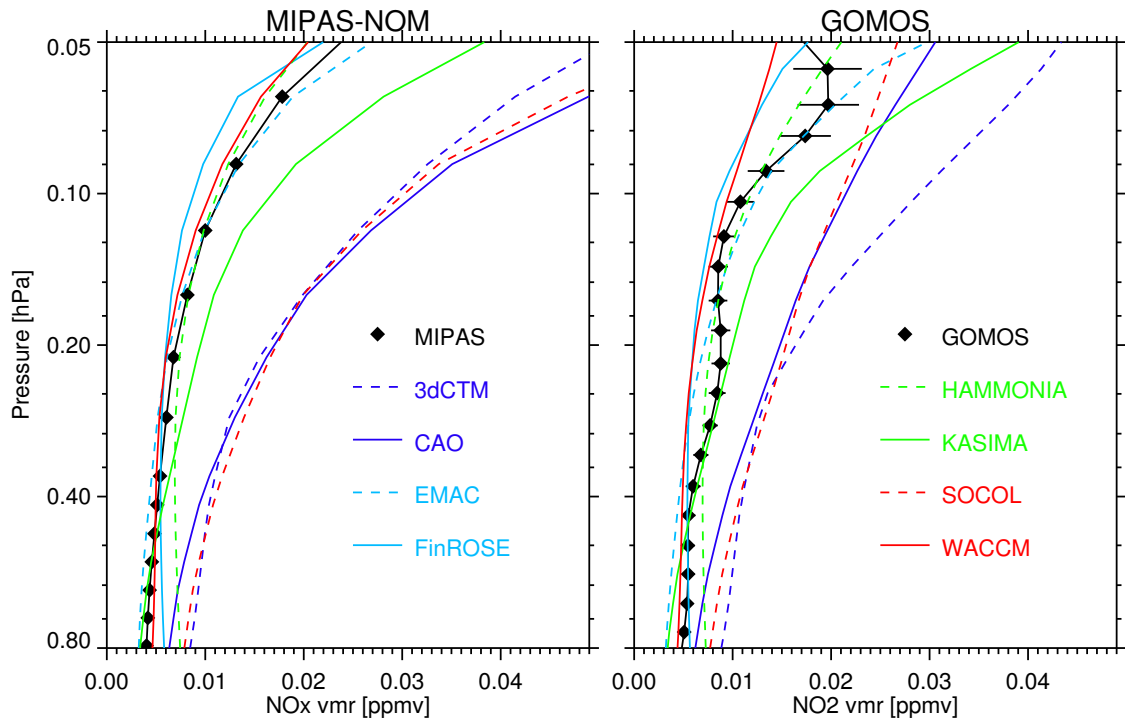
Figure 13. Left: MIPAS-NOM and modelled mean NO_x profile for the period 15 December 2008–12 January 2009 within 70–90°N. Right: GOMOS and modelled mean nighttime NO_2 for the same period within 75–85°N. The error bars indicate random retrieval errors of the averaged observational data.

encountered in the high-top models. Figures 14 and 15 show the temporal evolutions of modelled NO_x during the ES event in comparison with MIPAS-NOM and ACE-FTS observations, respectively. Despite the sampling-related differences, both instruments provide a very consistent picture of model biases. In particular, the time shift (earlier occurrence) of the modelled NO_x tongue (except 3dCTM), also identified in the CO comparisons, is clearly visible in the comparisons with both instruments.

5   Again, SOCOL and CAO overestimate significantly the observed NO_x (about a factor of 5) in the descending tongue (for the reasons already identified in the midwinter comparisons). This overestimation is even more pronounced than in the pre-SSW phase. In the case of HAMMONIA, related to the fast downward propagation of the ES (see Sec. 6), the NO_x peak occurs earlier and the tongue descends faster, merging with the background already in mid-February. In 3dCTM, the NO_x tongue reaches the lower mesosphere (0.02 hPa) later than in the other models and in observations due to the too slow descent rates

10   throughout the mesosphere. Thus, the development of the NO_x tongue in the lower mesosphere is delayed, and it does not reach to stratospheric altitudes.

The NO_x tongue observed by MIPAS reaches the 1 hPa level by the end of April. The reversal of the residual circulation in spring disabled further downward propagation of the tongue. ACE-FTS observed polar latitudes until 25 March, when the tongue reached the 0.3 hPa level in agreement with MIPAS observations at the same time. Compared to the observations, the

32
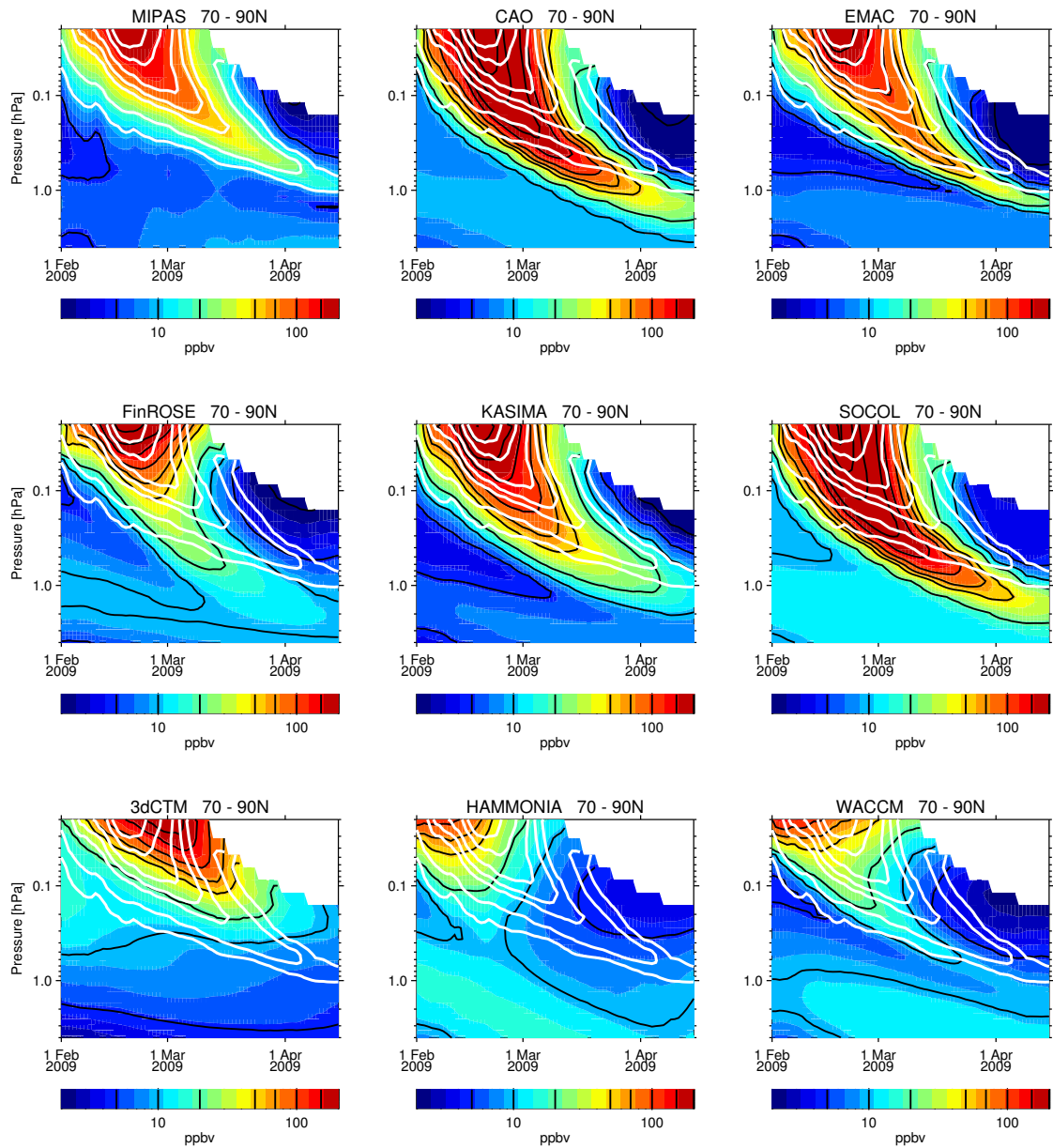
Figure 14. MIPAS-NOM and modelled temporal evolutions of NO_x during the ES event at 1–0.02 hPa within 70–90°N. White lines indicate the observed VMR levels of 10, 20, 50, 70, 100, 150, and 200 ppbv.

NO_x tongue in the model simulations (except HAMMONIA and 3dCTM) penetrates deeper, reaching the 2–3 hPa pressure levels at the end of April.
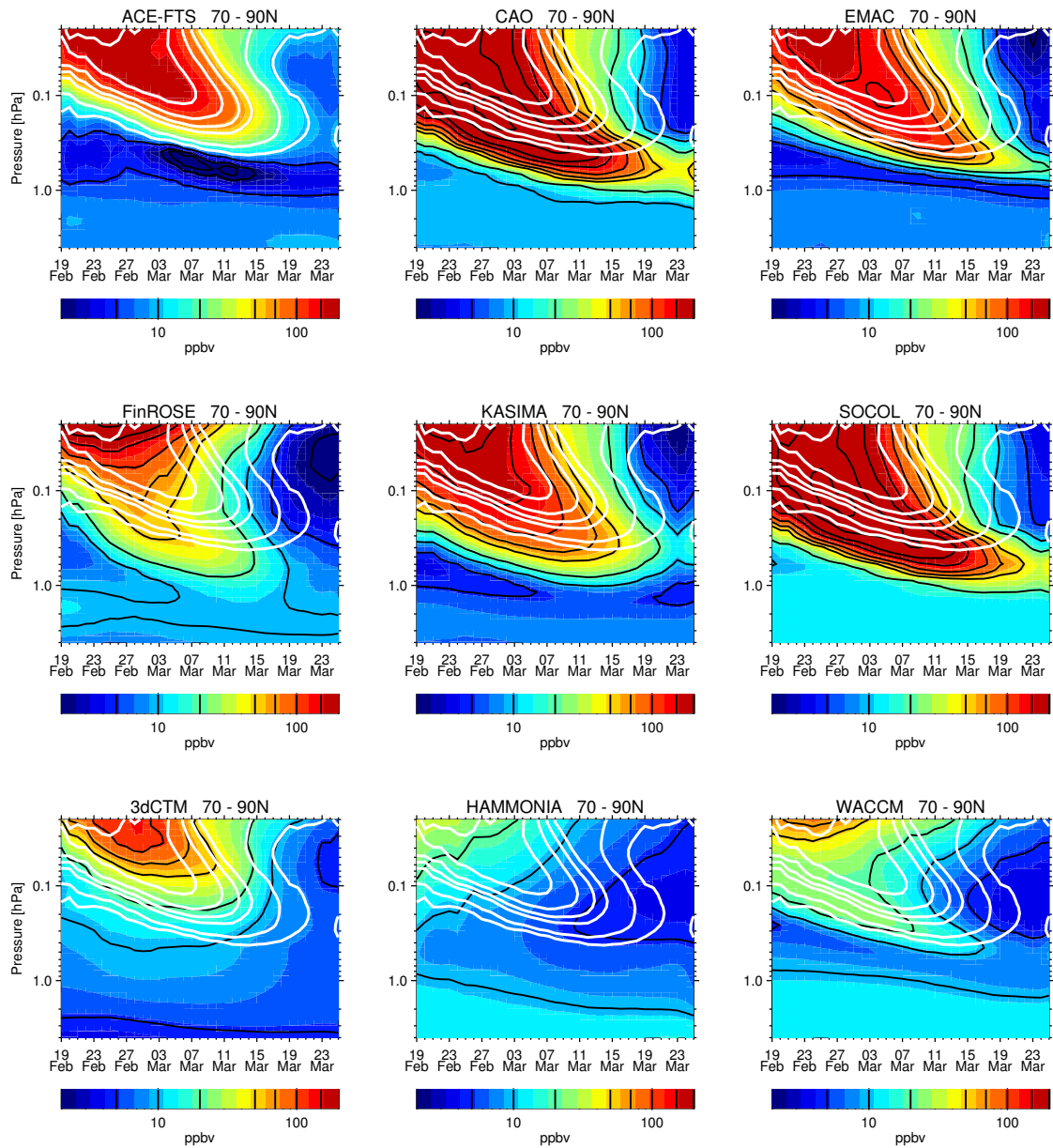
Figure 15. Same as Fig. 14, but for ACE-FTS.

Figure 16 shows more quantitatively the observed and modelled occurrence time and magnitude of the NO_x peak as a function of pressure level. The similar peak timing simulated by all models (except 3dCTM and HAMMONIA), about 2 weeks earlier than the observed peak below the 0.2 hPa level, is surprising. In the WACCM simulation, this time shift with respect to the observations is present over the whole vertical range. Interestingly, the peak occurrence time in the medium-top models, all
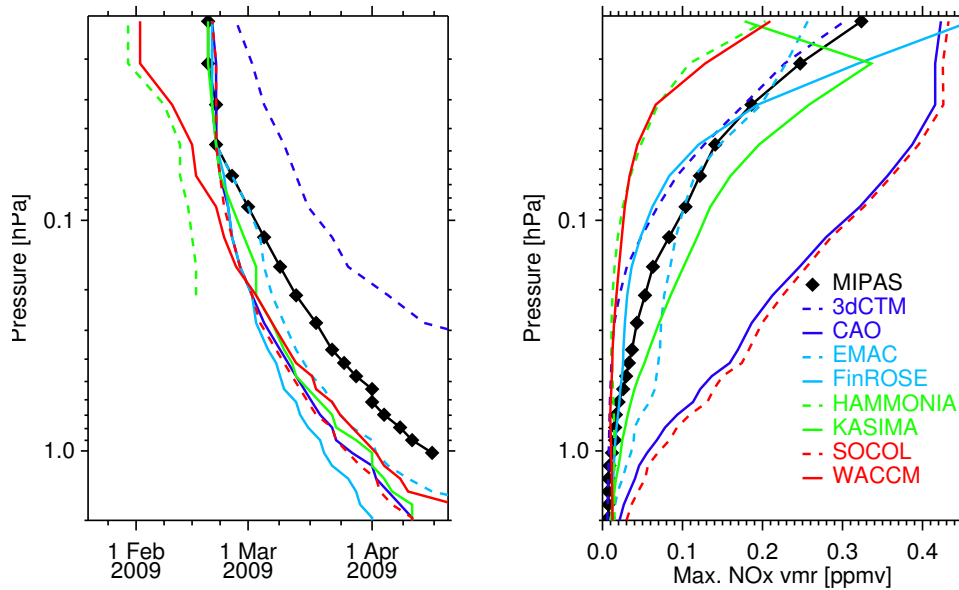
Figure 16. Left: MIPAS-NOM and modelled time evolution of the occurrence of the NO_x peak as function of pressure after the ES event. Right: observed and modelled NO_x peak values, averaged over 70–90°N.

prescribing the observed NO_x evolution at their upper lid, converges with the descent to the same occurrence time as simulated by WACCM at lower altitudes, i.e. earlier than in the observations. It is worth noting that a HAMMONIA simulation (not shown) with reduced non-orographic gravity wave amplitude (Meraner et al., 2016) exhibits both a NO_x peak occurrence time and magnitude in very good agreement with the observations down to pressure levels around 0.3 hPa. Below, however, the peak
5   occurrence time in this particular HAMMONIA simulation converges again to that of most of the other simulations.

Despite the consistency of the models with respect to the timing of the NO_x descent in the lower mesosphere, indicating similar dynamical representations, the spread of the magnitude of the modelled NO_x peaks (right panel of Fig. 16) is very large (within 0.2–3 times the observed magnitude), even when excluding the CAO and SOCOL results. This is particularly surprising in the case of the medium-top models, all of them prescribing the same NO_x obtained from observations, and will
10 be discussed in more detail at the end of this section.

Figure 17 shows the temporal evolution of the MIPAS observations and modelled NO_x at 0.5 hPa together with the temperature evolution slightly above, at 0.2 hPa. There is a clear link between the earlier occurrence of the modelled NO_x peaks and the time shift of the modelled temperature increases after the SSW, occurring systematically about 2 weeks earlier than in the observations (with the exceptions of HAMMONIA and 3dCTM). In order to check if the temperature bias of the simulations
15 with respect to MIPAS is consistent with the other measurements, we show in Fig. 18 the vertical structure of the temperature differences between the medium-top models and MIPAS-NOM, MLS, ACE-FTS, and SABER observations, similarly as done for the high-top models in Sec. 6. All medium-top models show a warm bias of 15–25 K around 0.2 hPa in February and early
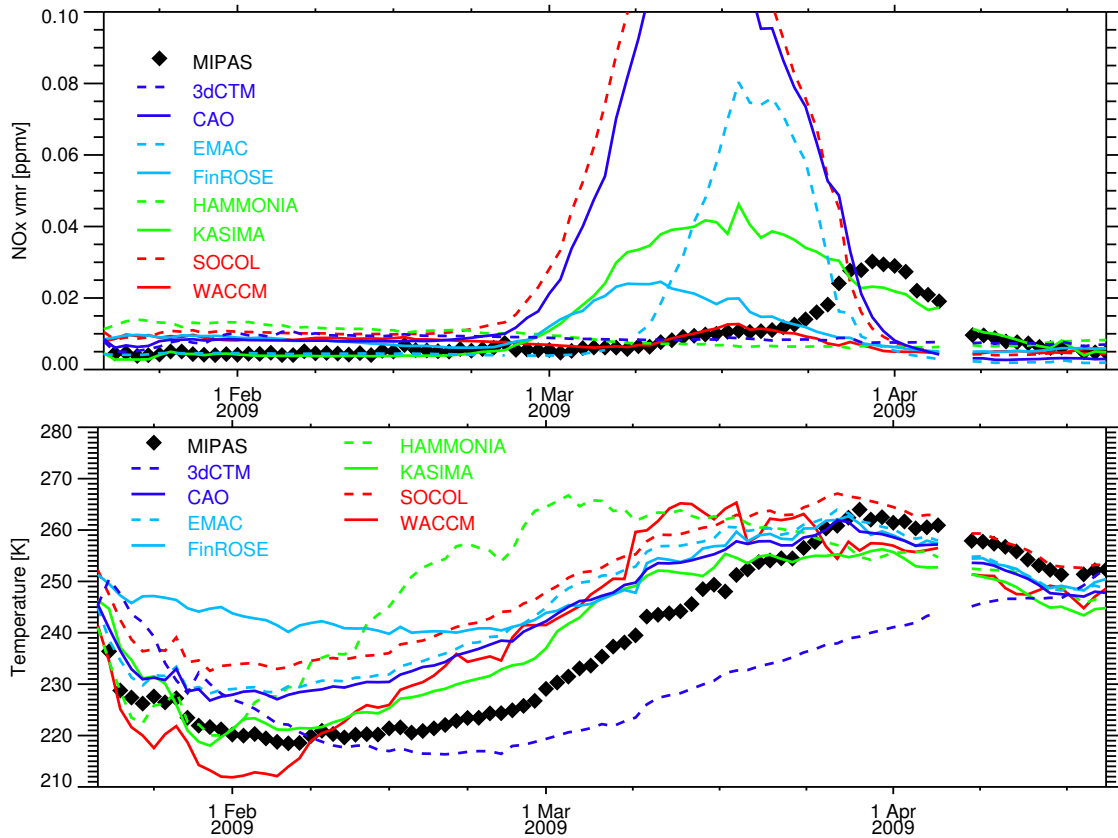
Figure 17. MIPAS-NOM and modelled temporal evolutions of NO_x at 0.5 hPa (top) and of temperature at 0.2 hPa (bottom) within 70–90°N during the ES event.

March, and a cold bias of 5–10 K around 1 hPa during the same period (though slightly less pronounced in KASIMA). Similar biases have been detected in the WACCM simulations (see Fig. 8).

The systematic, dipole-type temperature bias of the high-top model WACCM and all medium-top models, with similar amplitudes and time evolutions, explains the consistently too early occurrence of the NO_x descent encountered in these models.

5 It also hints at a common origin. One plausible reason for the temperature bias could be the meteorological data nudged in most models below 1 hPa. Around this pressure level, a cold bias of these models is observed, including FinROSE which relies entirely on ECMWF operational analysis data, and EMAC which applies the nudging to ERA Interim reanalysis data up to the altitude of 0.2 hPa. This indicates that the cold bias is present already in the ECMWF operational analysis and ERA Interim data. This bias might then likely influence the model dynamics extending above the nudged region. The cold bias around 1 hPa

10 in February is also seen in the WACCM simulation (c.f. Fig. 7), suggesting that it is also present in the MERRA reanalysis. This is confirmed by comparison of MERRA and MLS temperatures (not shown). Only in the HAMMONIA simulation, which shows a pronounced warm bias in the entire 2–0.1 hPa region, the local influence of the nudged meteorology at the edge of the

36
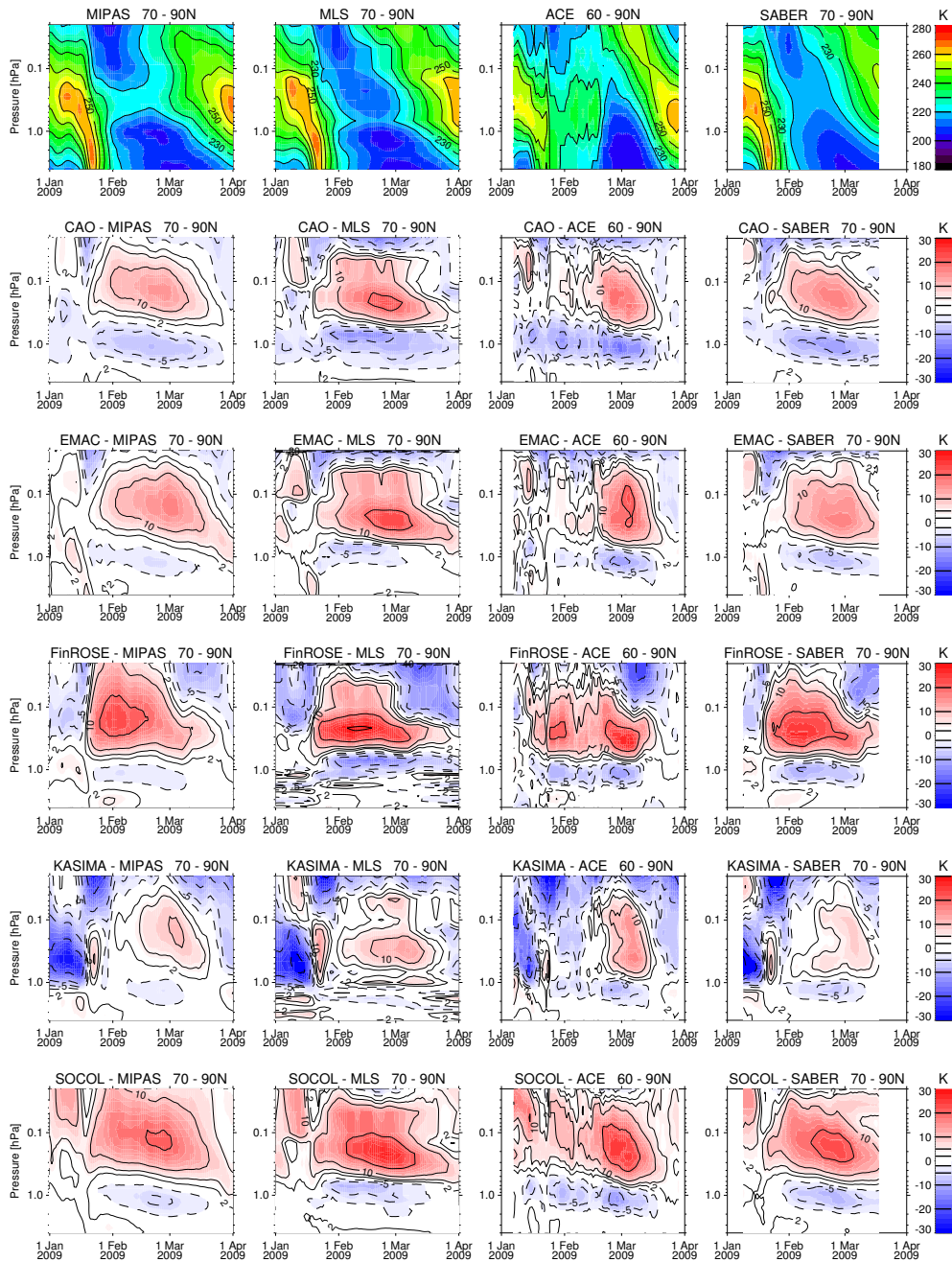
Figure 18. Top: temporal evolution of daily averaged polar cap temperatures at 4–0.02 hPa observed by MIPAS-NOM, MLS/Aura, ACE-FTS, and SABER (from left to right). Bottom: Corresponding differences between temperatures simulated with the "medium-top" models (CAO, EMAC, FinROSE, KASIMA, and SOCOL) and the observations.

37
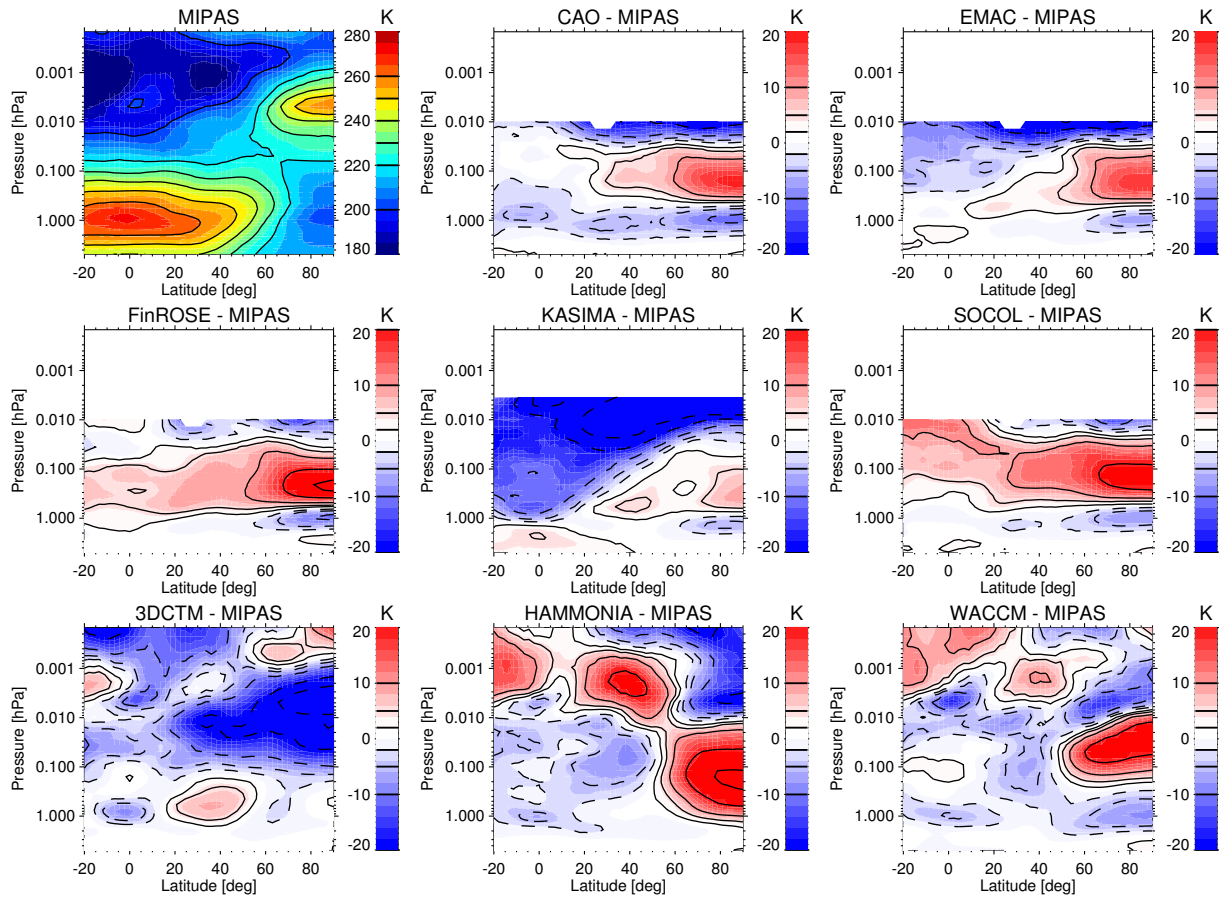
Figure 19. MIPAS-UA and modelled NH zonal mean temperature distribution on 15 February 2009.

nudging region seems to be outweighed by the internal model dynamics. It is beyond the scope of this paper to investigate in detail the possible mechanisms for the vertical propagation of dynamical biases, introduced by the nudging, resulting in a too early descent of mesospheric NO_x. However, since the ~~encountered cold bias~~ cold bias encountered at 1 hPa is restricted to latitudes northward of $60°$ (see Fig. 19) and hence implies a strengthening of the meridional temperature gradient, it is likely

5 to accelerate zonal winds at this level and above, which in turn would lead to changed filtering conditions for the propagation of gravity waves. Another important question which needs to be addressed in upcoming studies is the causes of the cold bias in the employed reanalysis datasets that have been found here.

The ~~encountered~~ spread of the magnitude of the ES-related NO_x tongue encountered below 0.1 hPa in the medium-top models, despite the prescription of a common odd nitrogen upper boundary above, deserves some further discussion. The

10 consistency of simulated temperature evolutions indicates that vertical transport is represented in these models in a similar way. It is therefore unlikely that differences in the descent velocities are the main cause for the spread. Differences in merid-

ional transport and mixing above the vortex edge and subsequent enhanced photochemical loss could also contribute to the differences but would not explain overestimation. A most plausible explanation is the detailed treatment of the upper boundary condition. Prescribing at an altitude with too fast vertical transport, as indicated here at 0.2 hPa, will unavoidably cause a too strong flux of NO_x into the domain below. Therefore, models that use a UBC definition extending to lower pressure levels likely overestimate the NO_x flux. This is, for example, the case of EMAC, which prescribes NO_x in the entire vertical domain above 0.1 hPa: the peak magnitude of the tongue is, as expected, close to the observations in the UBC domain. However, it becomes increasingly larger than the observed magnitude during the descent down to 0.7 hPa, where it is overestimated by a factor of 3. This highlights the importance of a realistic dynamical representation in the UBC domain in models prescribing NO_x concentrations.

8   Conclusions

We have presented the results of the HEPPA-II intercomparison project, conducted in the framework of SPARC/WCRP's SOLARIS-HEPPA activity, which aims at evaluating the simulations of the NH polar winter 2008/2009 from eight atmospheric models by comparison with observations of temperature and concentrations of NO_x and CO from seven satellite instruments. The large number of participating models allowed for a comprehensive assessment of the ability of state-of-the-art chemistry climate models to reproduce the observed EPP indirect effect in a dynamically perturbed NH winter under conditions of very low geomagnetic activity. The use of multi-instrument data for model evaluation not only allowed for the assessment of the significance of identified model biases, but also to estimate the uncertainty range of our current knowledge on tracer and temperature distributions in Arctic winters. It has been shown that the appropriate consideration of the instrument-specific sampling patterns is key to a meaningful multi-instrument analysis, particularly during perturbed dynamical conditions. The high degree of consistency between the comparisons of the models to individual observations has proven the reliability of the currently available satellite record during polar winter conditions.

Most models provide a good representation of the mesospheric tracer descent in general, and the EPP indirect effect in particular, during the unperturbed (pre-SSW) period of the NH winter 2008/2009. Observed NO_x descent into the lower mesosphere and stratosphere is generally reproduced within 20%. Larger discrepancies of a few model simulations, resulting in overestimated NO_x enhancements, could be traced back either to an unrealistic representation of the polar winter dynamics or to an inadequate prescription of the NO_x partitioning at the uppermost model layer leading to boundary artefacts.

In March–April, after the ES event, however, modelled mesospheric and stratospheric NO_x distributions deviate significantly from the observations. The too fast and early downward propagation of the NO_x tongue, encountered in most simulations, coincides with a warm bias in the lower mesosphere (0.2–0.05 hPa) being likely caused by an overestimation of descent velocities. On the other hand, upper mesospheric temperatures at 0.05–0.001 hPa are in general underestimated by the high-top models after the onset of the ES event, being indicative of a too slow descent and hence too small NO_x fluxes. As a consequence, the magnitude of the simulated NO_x tongue is generally underestimated by these models. Descending NO_x amounts simulated by the medium-top models with prescribed NO_x are on average closer to the observations but show a large

spread of up to several hundred percent. This is primarily attributed to the different vertical model regimes where the NO_x upper bounder condition is applied.

In general, the intercomparison demonstrates the ability of state-of-the-art atmospheric models to reproduce the observed EPP indirect effect in dynamically and geomagnetically quiescent early NH winter conditions as present in November 2008 – January 2009. It should be noted, however, that the extrapolation of this result to high geomagnetic activity conditions should be done with caution since mid-energy electron impact in the mesosphere, which was of minor importance during this particular winter, could lead to additional complications. Further, to obtain good agreement between simulated and observed mesospheric tracer descent it is ~~necessary~~ necessary to constrain stratospheric dynamics in the models by (re-)analysed meteorology.

The ~~encountered differences~~ differences encountered between observed and simulated NO_x, CO, and temperature distributions during the perturbed phase of the 2009 NH winter (i.e., February – April), however, ~~emphasize~~ emphasise the need for model improvements in the dynamical representation of ES events in order to allow for a better description of the EPP indirect effect under these particular conditions. Our results reinforce the findings from previous studies that the adequate parametrisation of unresolved GWD, particularly of its non-orographic component, is crucial for achieving such improvements. They also demonstrate that the dynamical boundary condition at 1 hPa, employed in our models, is not sufficient to fully determine the mesospheric circulation, yet crucial for the tracer transport into the stratosphere. Even when the winds are constrained in the stratosphere by observations, the calculated GWD in the mesosphere by different parametrisations can differ strongly. As discussed by McLandress and Scinocca (2005), such differences are more related to the characteristics of the launch spectra rather than to the treatment of the dissipation mechanisms in the parametrisations used. Indeed, Meraner et al. (2016) have shown that by modifying the launch characteristics of the gravity waves it is possible to tune the simulated NO_x descent towards the observations. Depending on the model, heating rates are calculated from coupled fields of radiative active gases, or climatologies are used, adding further to differences and uncertainties. In addition, despite the similar definition of the nudging regime ($<$1 hPa) in all model simulations, the vertical extent of the transition region between fully constrained and free-running mode varies among the models, which could introduce additional model spread. Finally, Smith et al. (2017) have shown that model dynamical fields are prone to errors due to the nudging approach itself, even when data and forcing terms are known exactly and there are no model biases. Interestingly, these intrinsic errors tend to grow with the complexity of the GWD representation employed in the model.

Many of the model-specific issues identified in the course of this project are currently being solved (e.g., Meraner et al., 2016). Lessons learned are hoped to be also of use for future model developments, particularly with respect to the consideration of EPP effects in upcoming coordinated model intercomparison projects. On the other hand, the ~~encountered bias~~ bias encountered in the meteorological reanalysis data in the post-SSW upper stratosphere and lower mesosphere potentially triggered the common tendency of the models to produce a too early descent in the lower mesosphere. These results imply the need to improve data assimilation systems for producing reanalysis data, especially with respect to the representation of the polar winter upper stratosphere and mesosphere. This is particularly important because the use of specified dynamics in atmospheric models is a necessary step to allow for meaningful comparisons to observations on seasonal and shorter time scales.

9 Data availability

All the model and observational data supporting the analysis and conclusions have been archived and are available upon request from the corresponding author.

References

Abe, T., Kurihara, J., Iwagami, N., Nozawa, S., Ogawa, Y., Fujii, R., Hayakawa, H., and Oyama, K.-i.: Dynamics and Energetics of the Lower Thermosphere in Aurora (DELTA)—Japanese sounding rocket campaign—, Earth Planets Space, 58, 1165–1171, doi:10.1186/BF03352006, 2006.

Andersson, M. E., Verronen, P. T., Marsh, D. R., Päivärinta, S.-M., and Plane, J. M. C.: WACCM-D – Improved modeling of nitric acid and active chlorine during energetic particle precipitation, J. Geophys. Res., 121, 10 328–10 341, doi:10.1002/2015JD024173, 2016.

Baumgaertner, A. J. G., Seppälä, A., Jöckel, P., and Clilverd, M. A.: Geomagnetic activity related NO_x enhancements and polar surface air temperature variability in a chemistry climate model: modulation of the NAM index, Atmos. Chem. Phys., 11, 4521–4531, doi:10.5194/acp-11-4521-2011, http://www.atmos-chem-phys.net/11/4521/2011/, 2011.

Bender, S., Sinnhuber, M., Burrows, J. P., Langowski, M., Funke, B., and López-Puertas, M.: Retrieval of nitric oxide in the mesosphere and lower thermosphere from SCIAMACHY limb spectra, Atmos. Meas. Tech., 6, 2521–2531, doi:10.5194/amt-6-2521-2013, http://www.atmos-meas-tech.net/6/2521/2013/, 2013.

Bender, S., Sinnhuber, M., von Clarmann, T., Stiller, G., Funke, B., López-Puertas, M., Urban, J., Pérot, K., Walker, K. A., and Burrows, J. P.: Comparison of nitric oxide measurements in the mesosphere and lower thermosphere from ACE-FTS, MIPAS, SCIAMACHY, and SMR, Atmos. Meas. Tech., 8, 4171–4195, doi:10.5194/amt-8-4171-2015, http://www.atmos-meas-tech.net/8/4171/2015/, 2015.

Bender, S., Sinnhuber, M., Langowski, M., and Burrows, J. P.: Retrieval of Nitric Oxide in the Mesosphere from SCIAMACHY Nominal Limb Spectra, Atmos. Meas. Tech. Discuss., 2016, 1–17, doi:10.5194/amt-2016-233, 2016.

Bender, S., Sinnhuber, M., Langowski, M., and Burrows, J. P.: Retrieval of nitric oxide in the mesosphere from SCIAMACHY nominal limb spectra, Atm. Meas. Tech., 10, 209–220, doi:10.5194/amt-10-209-2017, http://www.atmos-meas-tech.net/10/209/2017/, 2017.

Berger, U.: Modeling of middle atmosphere dynamics with LIMA, J. Atmosph. Sol-Terr. Phys., 70, 1170–1200, doi:10.1016/j.jastp.2008.02.004, 2008.

Bernath, P. F., McElroy, C. T., Abrams, M. C., Boone, C. D., Butler, M., Camy-Peyret, C.and Carleer, M., Clerbaux, C., Coheur, P.-F., Colin, R., DeCola, P., DeMazière, M., Drummond, J. R., Dufour, D., Evans, W. F. J., Fast, H., Fussen, D., Gilbert, K., Jennings, D. E., Llewellyn, E. J., Lowe, R. P., Mahieu, E., McConnell, J. C., McHugh, M., McLeod, S. D., Michaud, R., Midwinter, C., Nassar, R., Nichitiu, F., Nowlan, C., Rinsland, C. P., Rochon, Y. J., Rowlands, N., Semeniuk, K., Simon, P., Skelton, R., Sloan, J. J., Soucy, M.-A., Strong, K., Tremblay, P., Turnbull, D., Walker, K. A., Walkty, I., Wardle, D. A., Wehrle, V., Zander, R., , and Zou, J.: Atmospheric Chemistry Experiment (ACE): Mission overview, Geophys. Res. Lett., 32, L15S01, doi:doi:10. 1029/2005GL022386, 2005.

Bertaux, J. L., Kyrölä, E., Fussen, D., Hauchecorne, A., Dalaudier, F., Sofieva, V. F., Tamminen, J., Vanhellemont, F., Fanton d'Andon, O., Barrot, G., Mangin, A., Blanot, L., Lebrun, J. C., Pérot, K., Fehr, T., Saavedra, L., Leppelmeier, G. W., and Fraisse, R.: Global ozone monitoring by occultation of stars: an overview of GOMOS measurements on ENVISAT, Atmos. Chem. Phys., 10, 12 091–12 148, 2010.

Boone, C. D., Nassar, R., Walker, K. A., Rochon, Y., McLeod, S. D., Rinsland, C. P., , and Bernath, P. F.: Retrievals for the atmospheric chemistry experiment Fourier-transform spectrometer, Appl. Opt., 44, 7218–7231, 2005.

Boone, C. D., Walker, K. A., , and Bernath, P. F.: Version 3 Retrievals for the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS), in: The Atmospheric Chemistry Experiment ACE at 10: A Solar Occultation Anthology, edited by Bernath, P. F., pp. 103–127, A. Deepak Publishing, Hampton, Virginia, U.S.A, 2013.

Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V. V., Chance, K. V., and Goede, A. P. H.: SCIAMACHY: Mission Objectives and Measurement Modes, J. Atmos. Sci., 56, 127–150, doi:10.1175/1520-0469(1999)056<0127:SMOAMM>2.0.CO;2, 1999.

5 Burrows, J. P., Hölzle, E., Goede, A. P. H., Visser, H., and Fricke, W.: SCIAMACHY – scanning imaging absorption spectrometer for atmospheric chartography, Acta Astronaut., 35, 445–451, doi:10.1016/0094-5765(94)00278-T, http://www.sciencedirect.com/science/article/pii/009457659400278T, 1995.

Chance, K. and Kurucz, R.: An improved high-resolution solar reference spectrum for earth's atmosphere measurements in the ultraviolet, visible, and near infrared, J. Quant. Spectrosc. Ra., 111, 1289–1295, doi:10.1016/j.jqsrt.2010.01.036, http://www.sciencedirect.com/science/article/pii/S0022407310000610, 2010.

10 Chandran, A., Collins, R. L., Garcia, R. R., and Marsh, D. R.: A case study of an elevated stratopause generated in the Whole Atmosphere Community Climate Model, Geophys. Res. Lett., 38, doi:10.1029/2010GL046566, http://dx.doi.org/10.1029/2010GL046566, l08804, 2011.

Chang, J., Brost, R., Isaksen, I., Madronich, S., Middleton, P., Stockwell, W., and Walcek, C.: A three-dimensional Eulerian acid deposition

15 model: Physical concepts and formulation, J. Geophys. Res., 92, 14 681–14 700, doi:10.1029/JD092iD12p14681, 1987.

Charron, M. and Manzini, E.: Gravity Waves from Fronts: Parameterization and Middle Atmosphere Response in a General Circulation Model, J. Atmos. Sci., 59, 923–941, doi:10.1175/1520-0469(2002)059<0923:GWFFPA>2.0.CO;2, 2002.

Chipperfield, M.: Multiannual simulations with a three-dimensional chemical transport model, J. Geophys. Res., 104, 1781–1805, 1999.

Damski, J., Thölix, L., Backman, L., Kaurola, J., Taalas, P., Austin, J., Butchart, N., and Kulmala, M.: A chemistry–transport model simu-

20 lation of middle atmospheric ozone from 1980 to 2019 using coupled chemistry GCM winds and temperatures, Atmos. Chem. Phys., 7, 2165–2181, 2007.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Mor-

25 crette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quart. J. Roy. Meteor. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Egorova, T., Rozanov, E., Ozolin, Y., Shapiro, A., Calisto, M., Peter, T., and Schmutz, W.: The atmospheric effects of October 2003 solar proton event simulated with the chemistry–climate model {SOCOL} using complete and parameterized ion chemistry, J. Atmos. Sol. Terr. Phys., 73, 356 – 365, doi:10.1016/j.jastp.2010.01.009, http://www.sciencedirect.com/science/article/pii/S1364682610000428, 2011.

30 Egorova, T. A., Rozanov, E. V., Zubov, V. A., and Karol, I. L.: Model for investigating ozone trends (MEZON), zv. Atmos. Ocean. Phys., 39, 277–292, 2003.

Eyring, V., Lamarque, J.-F., Hess, P., Arfeuille, F., Bowman, K., Chipperfield, M. P., Duncan, B., Fiore, A., Gettelman, A., Giorgetta, M. A., Granier, C., Hegglin, M., Kinnison, D. E., Kunze, M., Langematz, U., Luo, B., Martin, R., Matthes, K., and other athors, .: Overview of IGAC/SPARC Chemistry-Climate Model Initiative (CCMI) Community Simulations in Support of Upcoming Ozone and Climate

35 Assessments, SPARC newsletter, Zurich, Switzerland, 40, 48–66, 2013.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, doi:10.5194/gmd-9-1937-2016, http://www.geosci-model-dev.net/9/1937/2016/, 2016.

Filipiak, M. J., Harwood, R. S., Jiang, J. H., Li, Q., Livesey, N. J., Manney, G. L., Read, W. G., Schwartz, M. J., Waters, J. W., and Wu, D. L.: Carbon monoxide measured by the EOS Microwave Limb Sounder on Aura: First results, Geophys. Res. Lett., 32, L14825, doi:10.1029/2005GL022765, 2005.

5 Fischer, H., Birk, M., Blom, C., Carli, B., Carlotti, M., von Clarmann, T., Delbouille, L., Dudhia, A., Ehhalt, D., Endemann, M., Flaud, J. M., Gessner, R., Kleinert, A., Koopmann, R., Langen, J., López-Puertas, M., Mosner, P., Nett, H., Oelhaf, H., Perron, G., Remedios, J., Ridolfi, M., Stiller, G., and Zander, R.: MIPAS: an instrument for atmospheric and climate research, Atmos. Chem. Phys., 8, 2151–2188, 2008.

Froidevaux, L., Livesey, N. J., Read, W. G., Jiang, Y. B., Jimenez, C., Filipiak, M. J., Schwartz, M. J., Santee, M. L., Pumphrey, H. C.,
10 Jiang, J. H., Wu, D. L., Manney, G. L., Drouin, B. J., Waters, J. W., Fetzer, E. J., Bernath, P. F., Boone, C. D., Walker, K. A., Jucks, K. W., Toon, G. C., Margitan, J. J., Sen, B., Webster, C. R., Christensen, L. E., Elkins, J. W., Atlas, E., Lueb, R. A., and Hendershot, R.: Early Validation Analyses of Atmospheric Profiles From EOS MLS on the Aura Satellite, IEEE Transactions on Geoscience and Remote Sensing, 44, 1106–1121, doi:10.1109/TGRS.2006.864366, 2006.

Funke, B., López-Puertas, M., Stiller, G. P., von Clarmann, T., and Höpfner, M.: A new non–LTE Retrieval Method for Atmospheric Param-
15 eters From MIPAS–ENVISAT Emission Spectra, Adv. Space Res., 27, 1099–1104, doi:10.1016/S0273-1177(01)00169-7, 2001.

Funke, B., García-Comas, M., López-Puertas, M., Glatthor, N., Stiller, G. P., von Clarmann, T., Semeniuk, K., and McConnell, J. C.: En-hancement of N_2O during the October–November 2003 solar proton events, Atmos. Chem. Phys., 8, 3805–3815, www.atmos-chem-phys.net/8/3805/2008/, 2008.

Funke, B., López-Puertas, M., García-Comas, M., Stiller, G. P., von Clarmann, T., Höpfner, M., Glatthor, N., Grabowski, U., Kellmann, S.,
20 and Linden, A.: Carbon monoxide distributions from the upper troposphere to the mesosphere inferred from 4.7 μm non-local thermal equilibrium emissions measured by MIPAS on Envisat, Atmos. Chem. Phys., 9, 2387–2411, http://www.atmos-chem-phys.net/9/2387/2009/, 2009.

Funke, B., Baumgaertner, A., Calisto, M., Egorova, T., Jackman, C. H., Kieser, J., Krivolutsky, A., López-Puertas, M., Marsh, D. R., Red-dmann, T., Rozanov, E., Salmi, S.-M., Sinnhuber, M., Stiller, G. P., Verronen, P. T., Versick, S., von Clarmann, T., Vyushkova, T. Y.,
25 Wieters, N., and Wissing, J. M.: Composition changes after the "Halloween" solar proton event: the High-Energy Particle Precipitation in the Atmosphere (HEPPA) model versus MIPAS data intercomparison study, Atmos. Chem. Phys., 11, 9089–9139, doi:10.5194/acp-11-9089-2011, www.atmos-chem-phys.net/11/9089/2011/, 2011.

Funke, B., López-Puertas, M., García-Comas, M., Kaufmann, M., Höpfner, M., and Stiller, G. P.: GRANADA: a Generic RAdiative traNsfer AnD non-LTE population Algorithm, J. Quant. Spectrosc. Radiat. Transfer, 113, 1771–1817, doi:10.1016/j.jqsrt.2012.05.001, 2012.

30 Funke, B., López-Puertas, M., Stiller, G. P., and von Clarmann, T.: Mesospheric and stratospheric NO_y produced by energetic particle precipitation during 2002–2012, J. Geophys. Res., 119, doi:10.1002/2013JD021404, 2014a.

Funke, B., Puertas, M. L., Holt, L., Randall, C. E., Stiller, G. P., and von Clarmann, T.: Hemispheric distributions and interannual variability of NO_y produced by energetic particle precipitation in 2002–2012, J. Geophys. Res., 119, 13 565–13 582, doi:10.1002/2014JD022423, 2014b.

35 Garcia, R. R., López-Puertas, M., Funke, B., Marsh, D. R., Kinnison, D. E., Smith, A. K., and González-Galindo, F.: On the distribution of CO_2 and CO in the mesosphere and lower thermosphere, J. Geophys. Res., 119, 5700–5718, doi:10.1002/2013JD021208, 2014.

García-Comas, M., López-Puertas, M., Marshall, B. T., Wintersteiner, P. P., Funke, B., Bermejo-Pantaleón, D., Mertens, C. J., Remsberg, E. E., Gordley, L. L., Mlynczak, M. G., and Russell, III, J. M.: Errors in Sounding of the Atmosphere using Broadband Emission Radiom-etry (SABER) kinetic temperature caused by non-local-thermodynamic-equilibrium model parameters, J. Geophys. Res., 113, 2008.

García-Comas, M., Funke, B., Gardini, A., López-Puertas, M., Jurado-Navarro, A., von Clarmann, T., Stiller, G., Kiefer, M., Boone, C. D., Leblanc, T., Marshall, B. T., Schwartz, M. J., and Sheese, P. E.: MIPAS temperature from the stratosphere to the lower thermosphere: Comparison of vM21 with ACE-FTS, MLS, OSIRIS, SABER, SOFIE and lidar measurements, Atmos. Meas. Tech., 7, 3633–3651, doi:10.5194/amt-7-3633-2014, http://www.atmos-meas-tech.net/7/3633/2014/, 2014.

Hauchecorne, A., Bertaux, J. L., Dalaudier, F., Russell, J. M., Mlynczak, M. G., Kyrölä, E., and Fussen, D.: Large increase of NO2 in the north polar mesosphere in January–February 2004: Evidence of a dynamical origin from GOMOS/ENVISAT and SABER/TIMED data, Geophys. Res. Lett., 34, L03 810, 2007.

Hines, C. O.: Doppler-spread parameterization of gravity-wave momentum deposition in the middle atmosphere. Part 1: Basic formulation, J. Atmos. Sol.-Terr. Phy., 59, 371–386, doi:10.1016/S1364-6826(96)00079-X, http://www.sciencedirect.com/science/article/pii/S136468269600079X, 1997.

Holton, J. and Zhu, X.: A further study of gravity wave induced drag and diffusion in the mesosphere, J. Atmos. Sci., 41, 2653–2662, 1984.

Holton, J. R.: The Role of Gravity Wave Induced Drag and Diffusion in the Momentum Budget of the Mesosphere, J. Atmos. Sci., 39, 791–799, doi:10.1175/1520-0469(1982)039<0791:TROGWI>2.0.CO;2, 1982.

Hurrell, J. W., Holland, M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W., Lawrence, D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W., Hack, J., Kiehl, J., and Marshall, S.: The Community Earth System Model: A Framework for Collaborative Research, Bull. Amer. Meteor. Soc., p. 130715075145003, doi:10.1175/BAMS-D-12-00121.1, http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-12-00121.1, 2013.

Jackman, C. H., DeLand, M. T., Labow, G. J., Fleming, E. L., Weisenstein, D. K., Ko, M. K. W., Sinnhuber, M., and Russell, J. M.: Neutral atmospheric influences of the solar proton events in October–November 2003, J. Geophys. Res., 110, A09S27, doi:10.1029/2004JA010888, 2005a.

Jackman, C. H., DeLand, M. T., Labow, G. J., Fleming, E. L., Weisenstein, D. K., Ko, M. K. W., Sinnhuber, M., and Russell, J. M.: Neutral atmospheric influences of the solar proton events in October–November 2003, J. Geophys. Res., 110, A09S27, doi:10.1029/2004JA010888, 2005b.

Jeuken, A. B. M., Siegmund, P. C., Heijboer, L. C., Feichter, J., and Bengtsson, L.: On the potential of assimilating meteorological analyses in a global climate model for the purpose of model validation, J. Geophys. Res., 101, 16 939–16 950, doi:10.1029/96JD01218, 1996.

Jöckel, P., Kerkweg, A., Pozzer, A., Sander, R., Tost, H., Riede, H., Baumgaertner, A., Gromov, S., and Kern, B.: Development cycle 2 of the Modular Earth Submodel System (MESSy2), Geosci. Model Dev., 3, 717–752, doi:10.5194/gmd-3-717-2010, http://www.geosci-model-dev.net/3/717/2010/, 2010.

Kieser, J.: The Influence of Precipitating Solar and Magnetospheric Energetic Charged Particles on the Entire Atmosphere Simulations with Hammonia, Max-Planck-Inst. für Meteorologie, 2011.

Kinnison, D. E., Brasseur, G. P., Walters, S., Garcia, R. R., Marsh, D. R., Sassi, F., Harvey, V. L., Randall, C. E., Emmons, L., Lamarque, J. F., Hess, P., Orlando, J. J., Tie, X. X., Randel, W., Pan, L. L., Gettelman, A., Granier, C., Diehl, T., Niemeier, U., and Simmons, A. J.: Sensitivity of chemical tracers to meteorological parameters in the MOZART-3 chemical transport model, J. Geophys. Res., 112, D20302, doi:10.1029/2006JD007879, 2007.

Kouker, W., Offermann, D., Küll, V., Reddmann, T., Ruhnke, R., and Franzen, A.: Streamers observed by the CRISTA experiment and simulated in the KASIMA model, J. Geophys. Res., 104, 16,405–16,418, 1999.

Kylling, A., Albold, A., and Seckmeyer, G.: Transmittance of a cloud is wavelength – dependent in the UV-range: Physical interpretation, Geophys. Res. Lett., 24(4), 397–400, doi:10.1029/97GL00111, 1997.

Kyrölä, E., Tamminen, J., Sofieva, V. F., Bertaux, J. L., Hauchecorne, A., Dalaudier, F., Fussen, D., Vanhellemont, F., Fanton d'Andon, O., Barrot, G., Guirlet, M., Mangin, A., Blanot, L., Fehr, T., Saavedra de Miguel, L., and Fraisse, R.: Retrieval of atmospheric parameters from GOMOS data, Atmos. Chem. Phys., 10, 11 881–11 903, 2010.

5 Lean, J., Rottman, G., Harder, J., and Kopp, G.: SORCE contributions to new understanding of global change and solar variability, Sol. Phys., 230, 27—53, doi:10.1007/s11207-005-1527-2, 2005.

Lin, S. J. and Rood, R. B.: Multidimensional flux-form semi-Lagrangian transport schemes, Mon. Wea. Rev., 124, 2046–2070, 1996.

Livesey, N. J.: Earth Observing System (EOS) Aura Microwave Limb Sounder (MLS) Version 4.2x Level 2 data quality and description document,, Jpl d-33509, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 2016.

10 Lott, F. and Miller, M. J.: A new subgrid-scale orographic drag parametrization: Its formulation and testing, Quart. J. Roy. Meteorol. Soc., 123, 101–127, doi:10.1002/qj.49712353704, 1997.

Maliniemi, V., Asikainen, T., and Mursula, K.: Spatial distribution of Northern Hemisphere winter temperatures during different phases of the solar cycle, J. Geophys. Res., 119, 9752–9764, doi:10.1002/2013JD021343, 2014.

Manney, G. L., Harwood, R. S., MacKenzie, I. A., Minschwaner, K., Allen, D. R., Santee, M. L., Walker, K. A., Hegglin, M. I., Lambert, A., 15 Pumphrey, H. C., Bernath, P. F., Boone, C. D., Schwartz, M. J., Livesey, N. J., Daffer, W. H., and Fuller, R. A.: Satellite observations and modeling of transport in the upper troposphere through the lower mesosphere during the 2006 major stratospheric sudden warming, Atm. Chem. Phys., 9, 4775–4795, doi:10.5194/acp-9-4775-2009, http://www.atmos-chem-phys.net/9/4775/2009/, 2009.

Marsh, D. R., Garcia, R. R., Kinnison, D. E., Boville, B. A., Sassi, F., Solomon, S. C., and Matthes, K.: Modeling the whole atmosphere response to solar cycle changes in radiative and geomagnetic forcing, J. Geophys. Res., 112, D23306, doi:10.1029/2006JD008306, 2007.

20 Marsh, D. R., Mills, M. J., Kinnison, D. E., Lamarque, J.-F., Calvo, N., and Polvani, L. M.: Climate Change from 1850 to 2005 Simulated in CESM1(WACCM), J. Climate, 26, 7372–7391, doi:10.1175/JCLI-D-12-00558.1, http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-12-00558.1, 2013.

Matthes, K., Funke, B., Anderson, M. E., Barnard, L., Beer, J., Charbonneau, P., Clilverd, M. A., Dudok de Wit, T., Haberreiter, M., Hendry, A., Jackman, C. H., Kretschmar, M., Kruschke, T., Kunze, M., Langematz, U., Marsh, D. R., Maycock, A., Misios, S., Rodger, C. J., Scaife, 25 A. A., Seppälä, A., Shangguan, M., Sinnhuber, M., Tourpali, K., Usoskin, I., van de Kamp, M., Verronen, P. T., and Versick, S.: Solar Forcing for CMIP6 (v3.1), Geosci. Model Dev. Discuss., 2016, 1–82, doi:10.5194/gmd-2016-91, http://www.geosci-model-dev-discuss.net/gmd-2016-91/, 2016.

McLandress, C. and Scinocca, J. F.: The GCM Response to Current Parameterizations of Nonorographic Gravity Wave Drag, J. Atmos. Sci., 62, 2394–2413, doi:10.1175/JAS3483.1, 2005.

30 McLandress, C., Scinocca, F., Shepherd, T. G., Reader, M. C., and Manney, G. L.: Dynamical Control of the Mesosphere by Orographic and Nonorographic Gravity Wave Drag during the Extended Northern Winters of 2006 and 2009, J. Atm. Sci., 70, doi:10.1175/JAS-D-12-0297.1, 2013.

Meraner, K. and Schmidt, H.: Transport of nitrogen oxides through the winter mesopause in HAMMONIA, J. Geophys. Res., 121, 2015JD024 136, doi:10.1002/2015JD024136, 2016.

35 Meraner, K., Schmidt, H., Manzini, E., Funke, B., and Gardini, A.: Sensitivity of Simulated Meospheric Transport of Nitrogen Oxides to Parameterized Gravity Waves, J. Geophys. Res., in press, 2016.

Merino, F., Murtagh, D. P., Ridal, M., Eriksson, P., Baron, P., Ricaud, P., and de la Noë, J.: Studies for the Odin sub-millimetre radiometer: III. Performance simulations, Canadian Journal of Physics, 80, 357–373, doi:10.1139/p01-154, 2002.

Murtagh, D., Frisk, U., Merino, F., Ridal, M., Jonsson, A., Stegman, J., Witt, G., Eriksson, P., Jiménez, C., Megie, G., de la Noë, J., Ricaud, P., Baron, P., Pardo, J. R., Hauchcorne, A., Llewellyn, E. J., Degenstein, D. A., Gattinger, R. L., Lloyd, N. D., Evans, W. F., McDade, I. C., Haley, C. S., Sioris, C., von Savigny, C., Solheim, B. H., McConnell, J. C., Strong, K., Richardson, E. H., Leppelmeier, G. W., Kyrölä, E., Auvinen, H., and Oikarinen, L.: An overview of the Odin atmospheric mission, Canadian Journal of Physics, 80, 309–319, doi:10.1139/p01-157, 2002.

Nieder, H., Winkler, H., Marsh, D., and Sinnhuber, M.: NOx production due to energetic particle precipitation in the MLT region: Results from ion chemistry model studies, J. Geophys. Res., 119, 2137–2148, doi:10.1002/2013JA019044, 2014.

Päivärinta, S.-M., Verronen, P. T., Funke, B., Gardini, A., Seppälä, A., and Andersson, M. E.: Transport versus energetic particle precipitation: Northern polar stratospheric NOx and ozone in January-March 2012, J. Geophys. Res., 121, 6085–6100, doi:10.1002/2015JD024217, 2016.

Pedatella, N. M., Fuller-Rowell, T., Wang, H., Jin, H., Miyoshi, Y., Fujiwara, H., Shinagawa, H., Liu, H.-L., Sassi, F., Schmidt, H., Matthias, V., and Goncharenko, L.: The neutral dynamics during the 2009 sudden stratosphere warming simulated by different whole atmosphere models, J. Geophys. Res., 119, 1306–1324, doi:10.1002/2013JA019421, 2014.

Pedatella, N. M., Fang, T.-W., Jin, H., Sassi, F., Schmidt, H., Chau, J. L., Siddiqui, T. A., and Goncharenko, L.: Multimodel comparison of the ionosphere variability during the 2009 sudden stratosphere warming, J. Geophys. Res., 121, 7204–7225, doi:10.1002/2016JA022859, 2016JA022859, 2016.

Pérot, K., Urban, J., and Murtagh, D. P.: Unusually strong nitric oxide descent in the Arctic middle atmosphere in early 2013 as observed by Odin/SMR, Atmos. Chem. Phys., 14, 8009–8015, doi:10.5194/acp-14-8009-2014, http://www.atmos-chem-phys.net/14/8009/2014/, 2014.

Poeschl, U., von Kuhlmann, R., Poisson, N., and Crutzen, P. J.: Development and intercomparison of condensed isoprene oxidation mechanisms for global atmospheric modeling, J. Atmos. Chem., 37, 29–52, doi:10.1023/A:1006391009798, 2000.

Porter, H. S., Jackman, C. H., and Green, A. E. S.: Efficiencies for production of atomic nitrogen and oxygen by relativistic proton impact in air, J. Chem. Phys., 65, 154–167, doi:10.1063/1.432812, 1976.

Price, C. and Rind, D.: A simple lightning parameterization for calculating global lightning distributions, J. Geophys. Res., 97, 9919–9933, doi:10.1029/92JD00719, 1992.

Pumphrey, H. C., Filipiak, M. J., Livesey, N. J., Schwartz, M. J., Boone, C., Walker, K. A., Bernath, P., Ricaud, P., Barret, B., Clerbaux, C., Jarnot, R. F., Manney, G. L., and Waters, J. W.: Validation of middle-atmosphere carbon monoxide retrievals from the Microwave Limb Sounder on Aura, J. Geophys. Res., 112, n/a–n/a, doi:10.1029/2007JD008723, d24S38, 2007.

Randall, C. E., Harvey, V. L., Singleton, C. S., Bailey, S. M., Bernath, P. F., Codrescu, M., Nakajima, H., and Russell III, J. M.: Energetic particle precipitation effects on the Southern Hemisphere stratosphere in 1992-2005, J. Geophys. Res., 112, D08308, doi:10.1029/2006JD007696, 2007.

Randall, C. E., Harvey, V. L., Siskind, D. E., France, J., Bernath, P. F., Boone, C. D., and Walker, K. A.: NO_x descent in the Arctic middle atmosphere in early 2009, Geophys. Res. Lett., 36, doi:10.1029/2009GL039706, 2009.

Randall, C. E., Harvey, V. L., Holt, L. A., Marsh, D. R., Kinnison, D., Funke, B., and Bernath, P. F.: Simulation of energetic particle precipitation effects during the 2003–2004 Arctic winter, J. Geophys. Res., 120, 5035–5048, doi:10.1002/2015JA021196, 2015JA021196, 2015.

Reddmann, T., Ruhnke, R., Versick, S., and Kouker, W.: Modeling disturbed stratospheric chemistry during solar-induced NO_x enhancements observed with MIPAS/ENVISAT, J. Geophys. Res., 115, D00I11, doi:10.1029/2009JD012569, 2010.

Remsberg, E. E., Marshall, B. T., García-Comas, M., Krueger, D., Lingenfelser, G. S., Martin-Torres, F. J., Mlynczak, M. G., Russell, III, J. M., Smith, A. K., Zhao, Y., Brown, C., Gordley, L. L., López-Gonzalez, M. J., López-Puertas, M., She, C.-Y., Taylor, M. J., and Thompson, R. E.: Assessment of the quality of the Version 1.07 temperature-versus-pressure profiles of the middle atmosphere from TIMED/SABER, J. Geophys. Res., 113, 17 101, 2008.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, J. Climate, 24, 3624–3648, doi:10.1175/jcli-d-11-00015.1, 2011.

Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornblueh, L., Manzini, E., Schlese, U., and Schulzweida, U.: Sensitivity of Simulated Climate to Horizontal and Vertical Resolution in the ECHAM5 Atmosphere Model, J. Climate, 19, 3771–3791, doi:10.1175/JCLI3824.1, 2006.

Rozanov, E., Calisto, M., Egorova, T., Peter, T., and Schmutz, W.: Influence of the Precipitating Energetic Particles on Atmospheric Chemistry and Climate, Surv. Geophys., 33, 483–501, doi:10.1007/s10712-012-9192-0, 2012.

Russell, III, J. M., Mlynczak, M. G., Gordley, L. L., Tansock, J., and Esplin, R.: An overview of the SABER experiment and preliminary calibration results, in: Proc. SPIE, pp. 277–288, 1999.

Salmi, S.-M., Verronen, P. T., Thölix, L., Kyrölä, E., Backman, L., Karpechko, A. Y., and Seppälä, A.: Mesosphere-to-stratosphere descent of odd nitrogen in February-March 2009 after sudden stratospheric warming, Atmos. Chem. Phys., 11, 4645–4655, doi:10.5194/acp-11-4645-2011, http://dx.doi.org/10.5194/acp-11-4645-2011, 2011.

Sander, R., Baumgaertner, A., Gromov, S., Harder, H., Jöckel, P., Kerkweg, A., Kubistin, D., Regelin, E., Riede, H., Sandu, A., Taraborrelli, D., Tost, H., and Xie, Z.-Q.: The atmospheric chemistry box model CAABA/MECCA-3.0, Geosci. Model Dev., 4, 373–380, doi:10.5194/gmd-4-373-2011, http://www.geosci-model-dev.net/4/373/2011/, 2011a.

Sander, R., Jöckel, P., Kirner, O., Kunert, A. T., Landgraf, J., and Pozzer, A.: The photolysis module JVAL-14, compatible with the MESSy standard, and the JVal PreProcessor (JVPP), Geosci. Model Dev., 7, 2653–2662, doi:10.5194/gmd-7-2653-2014, http://www.geosci-model-dev.net/7/2653/2014/, 2014.

Sander, S. P., Finlayson-Pitts, B. J., Friedl, R. R., Golden, D. M., Huie, R. E., Kolb, C. E., Kurylo, M. J., Molina, M. J., Moortgat, G. K., Orkin, V. L., and Ravishankara, A. R.: Chemical kinetics and Photochemical Data for the Use in Atmospheric Studies. Evaluation Number 14, JPL publication 02-25, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 2003.

Sander, S. P., Friedl, R. R., Golden, D. M., Kurylo, M. J., Moortgat, G. K., Keller-Rudek, H., Wine, P. H., Ravishankara, A. R., Kolb, C. E., Molina, M. J., Finlayson-Pitts, B. J., Huie, R. E., and Orkin, V. L.: Chemical kinetics and photochemical data for Use in Atmospheric Studies, Evaluation Number 15, JPL Publication 06-2, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, USA, 2006.

Sander, S. P., Abbatt, J., Barker, J. R., Burkholder, J. B., Friedl, R. R., Golden, D. M., Huie, R. E., Kolb, C. E., Kurylo, M. J., Moortgat, G. K., Orkin, V. L., and Wine, P. H.: Chemical kinetics and photochemical data for use in atmospheric studies, Evaluation No. 17, JPL publication 10-6, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 2011b.

Schmidt, H., Brasseur, G., Charron, M., Manzini, E., Giorgetta, M., Diehl, T., Formichev, V., Kinnison, D., Marsh, D., and Walters, S.: The HAMMONIA chemistry climate model: sensitivity of the mesopause region to the 11-year solar cycle and CO_2 doubling, J. Climate, 19, 3903–3931, 2006.

Schwartz, M. J., Lambert, A., Manney, G. L., Read, W. G., Livesey, N. J., Froidevaux, L., Ao, C. O., Bernath, P. F., Boone, C. D., Cofield, R. E., Daffer, W. H., Drouin, B. J., Fetzer, E. J., Fuller, R. A., Jarnot, R. F., Jiang, J. H., Jiang, Y. B., Knosp, B. W., Krüger, K., Li, J.-L. F., Mlynczak, M. G., Pawson, S., Russell, J. M., Santee, M. L., Snyder, W. V., Stek, P. C., Thurstans, R. P., Tompkins, A. M., Wagner, P. A., Walker, K. A., Waters, J. W., and Wu, D. L.: Validation of the Aura Microwave Limb Sounder temperature and geopotential height measurements, J. Geophys. Res., 113, n/a–n/a, doi:10.1029/2007JD008783, d15S11, 2008.

Seppälä, A., Clilverd, M. A., and Rodger, C. J.: NOx enhancements in the middle atmosphere during 2003-2004 polar winter: Relative significance of solar proton events and the aurora as a source, J. Geophys. Res., 112, D23 303, 2007.

Seppälä, A., Randall, C. E., Clilverd, M. A., Rozanov, E., and Rodger, C. J.: Geomagnetic activity and polar surface air temperature variability, J. Geophys. Res., 114, doi:10.1029/2008JA014029, 2009.

Seppälä, A. and Clilverd, M. A.: Energetic Particle Forcing of the Northern Hemisphere winter stratosphere: Comparison to solar irradiance forcing, Frontiers in Physics, 2, doi:10.3389/fphy.2014.00025, http://www.frontiersin.org/atmospheric_science/10.3389/fphy.2014.00025/abstract, 2014.

Sheese, P. E., Strong, K., Gattinger, R. L., Llewellyn, E. J., Urban, J., Boone, C. D., and Smith, A. K.: Odin observations of Antarctic nighttime NO densities in the mesosphere–lower thermosphere and observations of a lower NO layer, J. Geophys. Res., 118, 7414–7425, doi:10.1002/jgrd.50563, 2013.

Sheese, P. E., Walker, K. A., Boone, C. D., Bernath, P. F., Froidevaux, L., Funke, B., Raspollini, P., and von Clarmann, T.: ACE-FTS ozone, water vapour, nitrous oxide, nitric acid, and carbon monoxide profile comparisons with MIPAS and MLS, J. Quant. Spectrosc. Radiat. Transfer, p. in press, doi:doi:10.1016/j.jqsrt.2016.06.026, 2016a.

Sheese, P. E., Walker, K. A., Boone, C. D., McLinden, C. A., Bernath, P. F., Bourassa, A. E., Burrows, J. P., Degenstein, D. A., Funke, B., Fussen, D., Manney, G. L., McElroy, C. T., Murtagh, D., Randall, C. E., Raspollini, P., Rozanov, A., Russell III, J. M., Suzuki, M., Shiotani, M., Urban, J., von Clarmann, T., and Zawodny, J. M.: Validation of ACE-FTS version 3.5 NO_y species profiles using correlative satellite measurements, Atmos. Meas. Tech. Discuss., doi:doi:10.5194/amt-2016-69, 2016b.

Sinnhuber, B.-M., Weber, M., Amankwah, A., and Burrows, J. P.: Total ozone during the unusual Antarctic winter of 2002, Geophys. Res. Lett., 30, 2003.

Sinnhuber, M., Nieder, H., and Wieters, N.: Energetic particle precipitation and the chemistry of the mesosphere / lower thermosphere, Surv. Geophys., doi:doi:10.1007/s10712-012-9201-3, 2012.

Siskind, D. E., Eckermann, S. E., Coy, L., E., J. P. M. C., and Randall: On recent inter-annual variability of the Arctic winter mesosphere: Implications for tracer descent, Geophys. Res. Lett., 34, L09806, doi:10.1029/2007GL029293, 2007.

Siskind, D. E., Eckermann, S. D., McCormack, J. P., Coy, L., Hoppel, K. W., and Baker, N. L.: Case studies of the mesospheric response to recent minor, major, and extended stratospheric warmings, J. Geophys. Res., 115, doi:10.1029/2010JD014114, http://dx.doi.org/10.1029/2010JD014114, d00N03, 2010.

Siskind, D. E., Sassi, F., Randall, C. E., Harvey, V. L., Hervig, M. E., and Bailey, S. M.: Is a high-altitude meteorological analysis necessary to simulate thermosphere-stratosphere coupling?, Geophys. Res. Lett., 42, 8225–8230, doi:10.1002/2015GL065838, 2015GL065838, 2015.

Smith, A., Pedatella, N., Marsh, D., and Matsuo, T.: On the dynamical control of the mesosphere-lower thermosphere by the lower and middle atmosphere, J. Atmos. Sc, p. in press, doi:10.1175/JAS-D-16-0226.1, 2017.

Solomon, S. C. and Qian, L.: Solar extreme-ultraviolet irradiance for general circulation models, J. Geophys. Res., 110, A10 306, doi:10.1029/2005JA011160, 2005.

Stenke, A., Schraner, M., Rozanov, E., Egorova, T., Luo, B., and Peter, T.: The SOCOL version 3.0 chemistry-climate model: description, evaluation, and implications from an advanced transport algorithm, Geosci. Model Dev., 6, 1407–1427, doi:10.5194/gmd-6-1407-2013, 2013.

5   Sukhodolov, T., Rozanov, E., Shapiro, A., Anet, J., Cagnazzo, C., Peter, T., and Schmutz, W.: Evaluation of the ECHAM family radiation codes performance in the representation of the solar signal, Geosci. Model Dev., 7, 2859–2866, doi:10.5194/gmd-7-2859-2014, 2014.

Tamminen, J., Kyrölä, E., Sofieva, V. F., Laine, M., Bertaux, J. L., Hauchecorne, A., Dalaudier, F., Fussen, D., Vanhellemont, F., Fanton d'Andon, O., Barrot, G., Mangin, A., Guirlet, M., Blanot, L., Fehr, T., Saavedra de Miguel, L., and Fraisse, R.: GOMOS data characterisation and error estimation, Atmos. Chem. Phys., 10, 9505–9519, 2010.

10 Toohey, M., Hegglin, M. I., Tegtmeier, S., Anderson, J., Añel, J. A., Bourassa, A., Brohede, S., Degenstein, D., Froidevaux, L., Fuller, R., Funke, B., Gille, J., Jones, A., Kasai, Y., Krüger, K., Kyrölä, E., Neu, J. L., Rozanov, A., Smith, L., Urban, J., von Clarmann, T., Walker, K. A., and Wang, R.: Characterizing sampling biases in the trace gas climatologies of the SPARC Data Initiative, J. Geophys. Res., doi:10.1002/jgrd.50874, 2013.

Verronen, P. T., Ceccherini, S., Cortesi, U., Kyrölä, E., and Tamminen, J.: Statistical comparison of night-time NO_2 observations in 2003-
15   2006 from GOMOS and MIPAS instruments, Adv. Space Res., 43, 1918–1925, 2009.

Verronen, P. T., Andersson, M. E., Marsh, D. R., Kovács, T., and Plane, J. M. C.: WACCM-D – Whole Atmosphere Community Climate Model with D-region ion chemistry, J. Adv. Model. Earth Syst., 8, 954–975, doi:10.1002/2015MS000592, 2016.

von Clarmann, T., Glatthor, N., Grabowski, U., Höpfner, M., Kellmann, S., Kiefer, M., Linden, A., Mengistu Tsidu, G., Milz, M., Steck, T., Stiller, G. P., Wang, D. Y., Fischer, H., Funke, B., Gil-López, S., and López-Puertas, M.: Retrieval of temperature and tangent altitude
20   pointing from limb emission spectra recorded from space by the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), J. Geophys. Res., 108, 4736, doi:10.1029/2003JD003602, 2003.

von Clarmann, T., Höpfner, M., Kellmann, S., Linden, A., Chauhan, S., Grabowski, U., Funke, B., Glatthor, N., Kiefer, M., Schieferdecker, T., Stiller, G. P., and Versick, S.: Retrieval of temperature H_2O, O_3, HNO_3, CH_4, N_2O and $ClONO_2$ from MIPAS reduced resolution nominal mode limb emission measurements, Atmos. Meas. Tech., 2, 159–175, http://www.atmos-meas-tech.net/2/159/2009/, 2009.

25 von Clarmann, T., Funke, B., López-Puertas, M., Kellmann, S., Linden, A., Stiller, G. P., Jackman, C. H., and Harvey, V. L.: The Solar Proton Events in 2012 as Observed by MIPAS, Geophys. Res. Lett., 40, 1–5, doi:10.1002/grl.50119, 2013.

Waters, J., Froidevaux, L., Harwood, R., Jarnot, R., Pickett, H., Read, W., Siegel, P., Cofield, R., Filipiak, M., Flower, D., Holden, J., Lau, G., Livesey, N., Manney, G., Pumphrey, H., Santee, M., Wu, D., Cuddy, D., Lay, R., Loo, M., Perun, V., Schwartz, M., Stek, P., Thurstans, R., Boyles, M., Chandra, K., Chavez, M., Chen, G., Chudasama, B., Dodge, R., Fuller, R., Girard, M., Jiang, J., Jiang, Y., Knosp, B.,
30   LaBelle, R., Lam, J., Lee, K., Miller, D., Oswald, J., Patel, N., Pukala, D., Quintero, O., Scaff, D., Van Snyder, W., Tope, M., Wagner, P., and Walch, M.: The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura satellite, IEEE Trans. Geosci. Remote Sens., 44, 1075–1092, doi:10.1109/TGRS.2006.873771, 2006.

Wissing, J. M. and Kallenrode, M.-B.: Atmospheric Ionization Module Osnabrück (AIMOS): A 3-D model to determine atmospheric ionization by energetic charged particles from different populations, J. Geophys. Res., 114, doi:10.1029/2008JA013884, 2009.

35 WMO: SPARC Report on the Evaluation of Chemistry-Climate Models, SPARC Report No. 5, WCRP-132, WMO/TD-No. 1526, 2010.

Yamashita, C., England, S. L., Immel, T. J., and Chang, L. C.: Gravity wave variations during elevated stratopause events using SABER observations, J. Geophys. Res., 118, 5287–5303, doi:10.1002/jgrd.50474, http://dx.doi.org/10.1002/jgrd.50474, 2013.