

## Report on Stratosphere Task Force

Theodore G. Shepherd<sup>1</sup>,  
Inna Polichtchouk<sup>1,2</sup>, Robin J. Hogan<sup>2</sup>  
and Adrian J. Simmons<sup>3</sup>

<sup>1</sup> Department of Meteorology, University of Reading, UK

<sup>2</sup> Research Department

<sup>3</sup> Copernicus Department

June 2018

This paper has not been published and should be regarded as an Internal Report from ECMWF.  
Permission to quote from it should be obtained from the ECMWF.



European Centre for Medium-Range Weather Forecasts  
Europäisches Zentrum für mittelfristige Wettervorhersage  
Centre européen pour les prévisions météorologiques à moyen

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:

<http://www.ecmwf.int/en/research/publications>

Contact: [library@ecmwf.int](mailto:library@ecmwf.int)

© Copyright 2018 Theodore G. Shepherd and ECMWF

European Centre for Medium Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG2 9AX, England

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.

## Abstract

Recognising the importance of the stratosphere for skilful seasonal and sub-seasonal prediction, the Stratosphere Task Force was set up in 2016 to improve the representation of the stratosphere in ECMWF forecast and analysis systems. This report synthesizes the most notable findings of the Task Force and provides recommendations for the way forward. The main focus is on: 1) Global-mean temperature biases; 2) Horizontal resolution sensitivity of the mid- to lower stratospheric temperatures; 3) Stratospheric meridional circulation and polar vortex variability; 4) Extratropical lower stratospheric cold temperature bias; 5) New sponge design; and, 6) Representation of tropical winds.

## 1 Introduction

The goal of the Stratosphere Task Force, which met between November 2016 and December 2017, was to improve the representation of the stratosphere in ECMWF forecast and analysis systems. Over the years, the stratosphere in the IFS had been somewhat neglected. Different researchers at ECMWF had been dealing with stratosphere issues as they arose but in different ways for different applications, and this had led to a patchwork situation. In line with ECMWF's strategic goal of improving tropospheric predictions particularly on monthly and seasonal timescales, there is a renewed impetus to carefully study all potential sources of predictive skill, of which the stratosphere is one. The motivation for the Task Force was to achieve a more coordinated treatment of the stratosphere across the different applications, and provide a concerted effort to improve the representation of the stratospheric state both in analyses and reanalyses, and in forecasts.

The Task Force focused principally on atmospheric modelling, since a realistic model is the foundation of both analysis and prediction. This is especially the case in the stratosphere, where observations are comparatively limited and there are few reliable anchoring data sets. However, there was also strong involvement of satellite and data assimilation scientists, and some exploration of data assimilation issues. The impact of the stratosphere on tropospheric forecasts is expected primarily at monthly and seasonal timescales, for which a large ensemble of hindcasts is required to demonstrate statistically significant changes to forecast skill. Therefore, the approach taken was to improve the physical realism of the model behaviour, and reduce model biases, before worrying about whether forecast scores were improved, as experience says that this is the best strategy for long-term progress.

The Task Force met approximately once per month, with the meetings chaired by Robin Hogan. It involved scientists from the Research, Forecast and Copernicus Departments of ECMWF, along with Ted Shepherd and Inna Polichtchouk from the University of Reading. It operated on a voluntary basis, with researchers presenting recent findings followed by discussion. Typically there were about half a dozen presenters and about 20-25 participants at each meeting. Meeting summaries and presentations were recorded on the ECMWF intranet pages, and interim results were reported by Polichtchouk et al. (2017). This report collects and synthesizes some of the most notable findings, and makes a number of recommendations.

The conclusions and recommendations from each section are provided at the end of the section, and then collected together at the end of the document in a slightly simplified form. A number of suggestions for additional independent validation data sets are also provided in an Appendix.

## 2 Global-mean temperature

One can think of the troposphere as providing a (large-scale) turbulent boundary layer for the atmosphere, and the stratosphere as being comparatively isolated from the surface of the Earth. Thus, to a first approximation the global-mean stratosphere is in radiative equilibrium, with long-wave cooling balancing solar heating through ozone, and a negligible role for vertical turbulent energy fluxes (Fomichev et al. 2002). This property makes global-mean temperature an excellent diagnostic for model evaluation. Figure 1 shows the observed stratospheric cooling over the last 30 years, which has resulted from a combination of CO<sub>2</sub> increase and (for the first part of the record) ozone depletion, punctuated by warming from volcanic aerosol. The fact that the free-running CMIP5 models can track the observed anomalies so closely demonstrates the strength of this radiative control on global-mean temperature. Latitudinally dependent temperature biases in the stratosphere are more difficult to interpret, since they depend not only on radiative processes but also on the meridional circulation (see Section 4).

The ERA5 reanalysis, which is being produced using IFS cycle 41r2 (the cycle used for operational forecasting in 2016), exhibits several symptoms of global-mean temperature bias in the underlying model. Figure 2 shows the global-mean differences with radiosondes at several lower stratospheric layers, as well as the corresponding differences for ERA-Interim. The differences are generally much larger in the case of ERA5, and exhibit persistent cold biases of up to 0.5 K that are especially severe around 70 hPa. The inference is that the global-mean lower stratospheric temperature biases in the version of the IFS used in ERA5 are larger than they were in the version used in ERA-Interim. This is confirmed in Figure 3, which shows 3-day 50 hPa temperature forecast errors for the extratropical northern and southern hemispheres. The forecast errors for ERA5 show a persistent cold bias, which is much larger than for ERA-Interim. Examination of “climate runs” (ensembles of year-long free-running simulations) of the IFS model cycles used by ERA-Interim and ERA5 reveal that the former was an unusually unbiased cycle in the stratosphere, and that the patterns of temperature bias in ERA5 matched the patterns of bias in the free-running model from the same cycle, but with reduced amplitude.

The differences with radiosondes shown in Figure 2 also exhibit strong temporal inhomogeneities. In particular, ERA5 does not sufficiently capture the lower stratospheric warming in the early 1990s following the eruption of Mt Pinatubo (see Figure 1). These issues were much less apparent in ERA-Interim. The differences with radiosondes are much reduced after the introduction of GPS radio occultation (RO) observations in 2006, which are much more plentiful than the radiosonde observations. The implication is that the radiosondes are much less effective at correcting the lower stratospheric biases in ERA5 than in ERA-Interim. This is in part due to narrower structure functions in the Cy41r2 J<sub>b</sub> (presumably because the model has a much more active mesoscale spectrum than for ERA-Interim) and larger specified radiosonde errors than in ERA-Interim, which cause the analysis to make a smaller adjustment of larger scales when presented with radiosonde data. Use of the Cy41r2 J<sub>b</sub> gives particularly poor fits to radiosonde data in the early 1990s when information in the radiosonde data on the lower stratospheric warming due to the eruption of Mt Pinatubo is not utilised, and the corresponding information in the MSU radiance data is dismissed as a bias in that radiance data.

For separate reasons, the J<sub>b</sub> based on Cy41r2 was found to be unsatisfactory in the early part of the ERA5 record, and was replaced by a different J<sub>b</sub> estimated using data assimilation during 1979. Figure 2 shows that the global-mean lower stratospheric temperature differences between ERA5 and radiosondes are much smaller, and comparable to those for ERA-Interim, during the segments in the first half of the record where the 1979 J<sub>b</sub> was used. (The downward spike in the ERA5 radiosonde data

fits for December 1986 — produced using the 1979  $J_b$  — is due to assimilating warm-biased MSU-4 data during the first month of availability of data from the NOAA-10 satellite, when the variational bias adjustment was spinning up from a poor initial estimate. This will be repaired in a short rerun prior to general data release.)

The improved analysis using the 1979  $J_b$  is confirmed by Figure 4, which shows time series of the differences between ERA5 and ERA-Interim estimates of radiance biases for relevant MSU and AMSUA channels and instruments. (Differences are not shown for the AMSUA instrument on the EOS-Aqua satellite as its data were subject to recalibration prior to their use in ERA5.) Differences are larger in those pre-2006 periods when the Cy41r2  $J_b$  was used. The inadequate weight given to radiosonde data by this  $J_b$  means that prior to the availability of GPSRO data, its use prevents radiosonde data from providing a strong anchoring of the radiance bias estimation. The anchoring is instead provided by the cold-biased model; the satellite data are thus wrongly estimated to be biased warm. The 1979  $J_b$  is now also being used for production in the 1990s, and the analyses for the early and mid 1990s already carried out using the Cy41r2  $J_b$  will be rerun using it.

There are also indications from ERA5 of global-mean temperature biases in the upper stratosphere. This region lies above the altitude range of both radiosondes and GPSRO, hence they provide only limited anchoring for the nadir sounders. The latter were never designed for climate monitoring, and homogenizing the data from different operational satellites, with rapidly drifting orbits, is a challenge (Nash and Saunders 2015). Indeed, ERA-Interim exhibited some significant temporal inhomogeneities in upper-stratospheric global-mean temperature (Dee and Uppala 2008; McLandress et al. 2014a). The comparison between ERA5 and ERA-Interim is more complicated than in the lower stratosphere, as there are also differences due to the use of revised fast radiative transfer calculations for data from the SSU instruments in the ERA5 data assimilation, and to the use of unadjusted SSU-3 as well as AMSUA-14 data as an anchor for the bias adjustment of other radiance data during the period when both SSU and AMSUA data are available. Time series of global-mean temperature analyses for the upper stratosphere nevertheless show shifts associated with the changes in  $J_b$ , particularly at 5 hPa, as well as with the introduction of GPSRO data (Figure 5). The solution at these altitudes may therefore be to explicitly bias-correct to the model, as is done in JRA-55, which puts a premium on minimizing the biases in the model.

Thus, global-mean temperature biases in the model create significant challenges for the representation of the stratosphere in reanalyses. Since these biases are under radiative control, an early focus of the Task Force was on improving the representation of radiative processes in the model, which are well understood. It is not therefore a question of tuning, but rather of ensuring that key processes are accurately represented. Examples of such processes include an improved solar spectrum with 7-8% less ultraviolet radiation, diurnally varying ozone (solar heating occurs during the day, so daily average ozone is not relevant for ozone heating), better treatment of solar zenith angles (the stratosphere can be sunlit even when the ground is not), better ozone climatologies, etc. None of these improvements have significant implications for computational cost. A series of such fundamental improvements was made, which are documented by Hogan et al. (2017) and illustrated in Figure 6. For validation, limb-sounding data (which has relatively high vertical resolution) was used from the Aura MLS instrument.

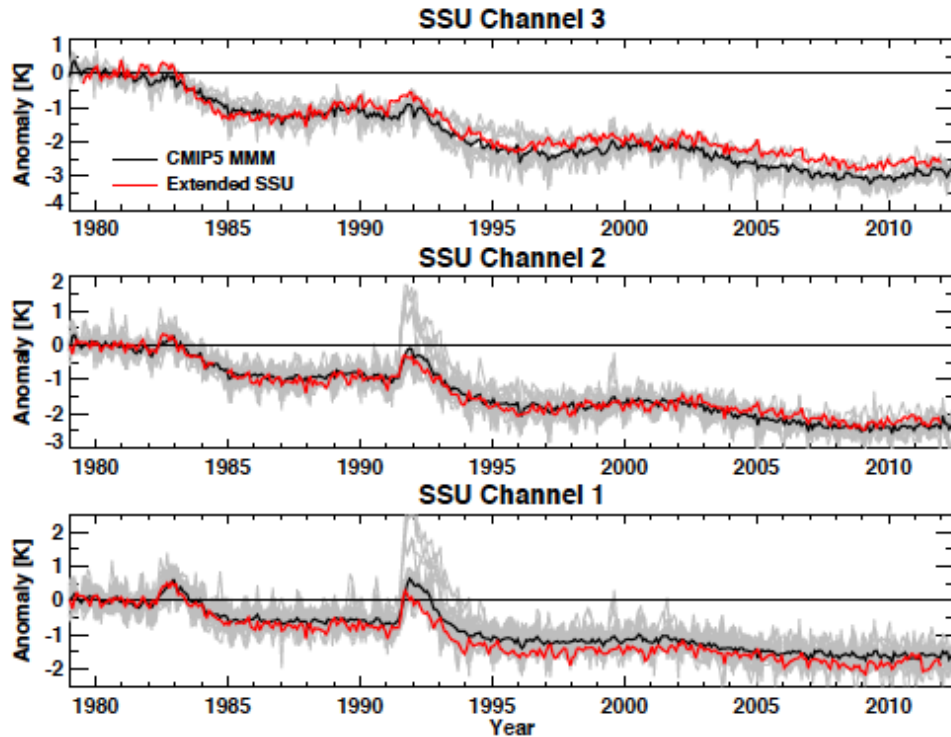
When run in climate mode, IFS Cycle 41R1 (represented by the red line in Figure 6) generally exhibited a warm bias above about 50 hPa, which increased more or less continuously with altitude to values of nearly 10 K in the upper stratosphere and 20 K in the upper mesosphere. Each of the improvements

contributed to a reduction of this bias, to the extent that the final version of the model is essentially unbiased throughout most of the stratosphere. The dark blue dashed line corresponds to a configuration close to that used by ERA5, while the light blue solid line shows the current operational cycle (43R3). The main additional change beyond this is to update the solar spectrum, and indeed this is expected to be used in the next version of EC-Earth.

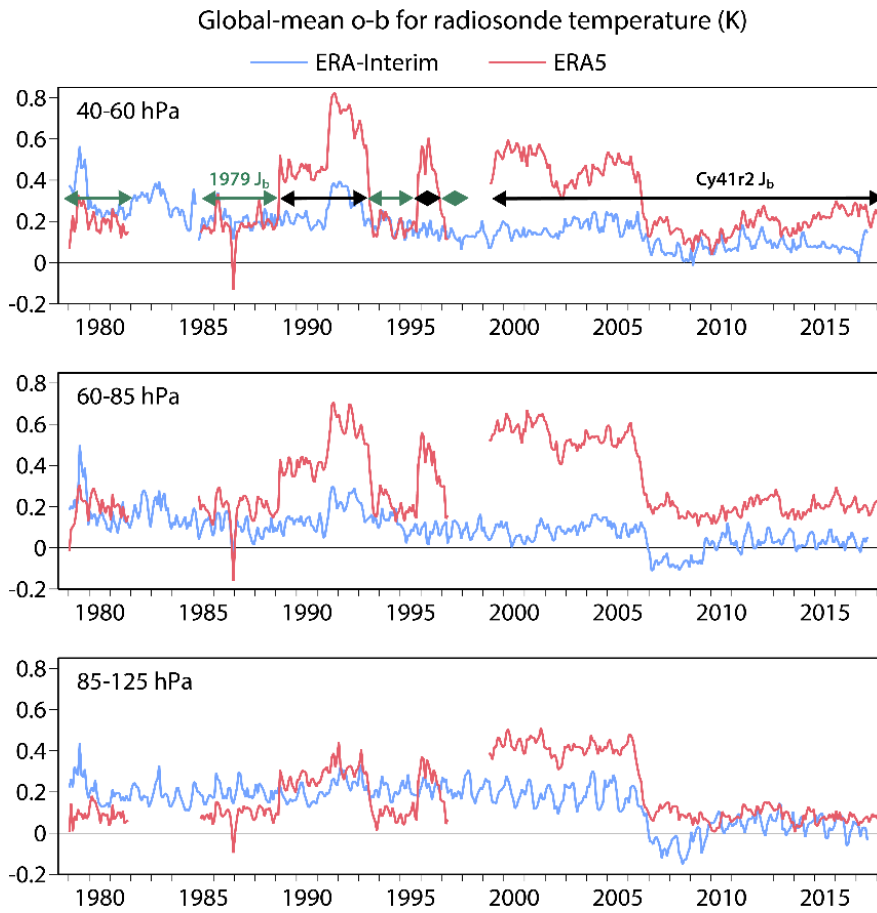
However, an obstacle to implementing the updated solar spectrum in operational forecasts is the resolution dependence of the lower stratosphere temperature. As discussed in Section 3 immediately below, increasing horizontal resolution from TL255 to TCo1279 results in a 1–2 K cooling at 70 hPa unless it is also accompanied by a modest increase in vertical resolution (e.g. 137 to 162 levels). Therefore the change to the solar spectrum is actually to worsen the lower stratosphere cold bias in the high-resolution model with 137 levels. This resolution dependence may also explain why ERA5, which is produced at TL639L137 resolution, has a cold bias in the lower stratosphere (Figure 2), whereas the dark blue dashed line in Figure 6, which is produced at TL255L137, suggests a slight warm bias in this region.

There remain biases around the stratopause region, which will affect radiances from AMSU channels peaking lower down. These require further attention. Note that global-mean temperature biases can arise from errors in the abundance or spatial distribution of radiatively active species, so transport and chemistry are relevant, not just radiative processes *per se*.

**Conclusions:** *The vertical profile of global-mean temperature is a key model diagnostic in the stratosphere. A number of improvements to the radiation scheme and the treatment of ozone were made during 2016 and 2017, and have the capability to eliminate most of the global-mean temperature bias in the stratosphere in the IFS at TL255 resolution, which was quite substantial. Some of these changes have been migrated to the current operational cycle (43R3), but it has not yet been possible to implement the improved solar spectrum due to the cooling of the lower stratosphere when horizontal resolution is increased. Time series of global-mean temperature in the ERA5 reanalysis exhibit a number of problematical features in the stratosphere. Before the next reanalysis, a minimum requirement for the model must be an essentially unbiased stratospheric global-mean temperature. Further attention should be paid to remaining global-mean temperature biases around the stratopause region.*

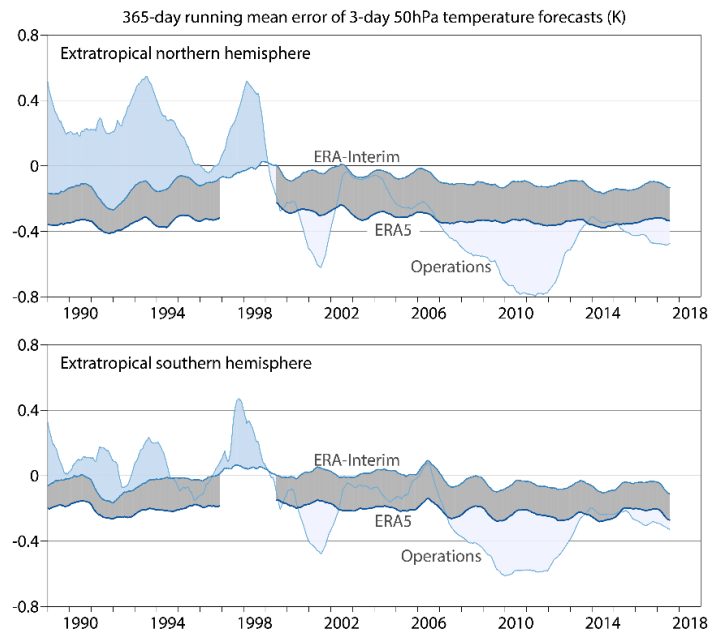


**Figure 1.** Deseasonalized monthly mean near-global-mean ( $75^{\circ}\text{S}$ – $75^{\circ}\text{N}$ ) temperature anomalies for an extension of SSU Channels 1, 2 and 3 using AMSU data (red) and for the CMIP5 multi-model mean (black). SSU Channels 1, 2 and 3 have broad weighting functions, which peak respectively at approximately 30, 39 and 44 km altitude. The light-grey curves are the time series of the individual CMIP5 models used to compute the multi-model mean. Anomalies are computed with respect to 1979–1982; thus the time mean anomaly over this period is zero. From McLandress et al. (2015).

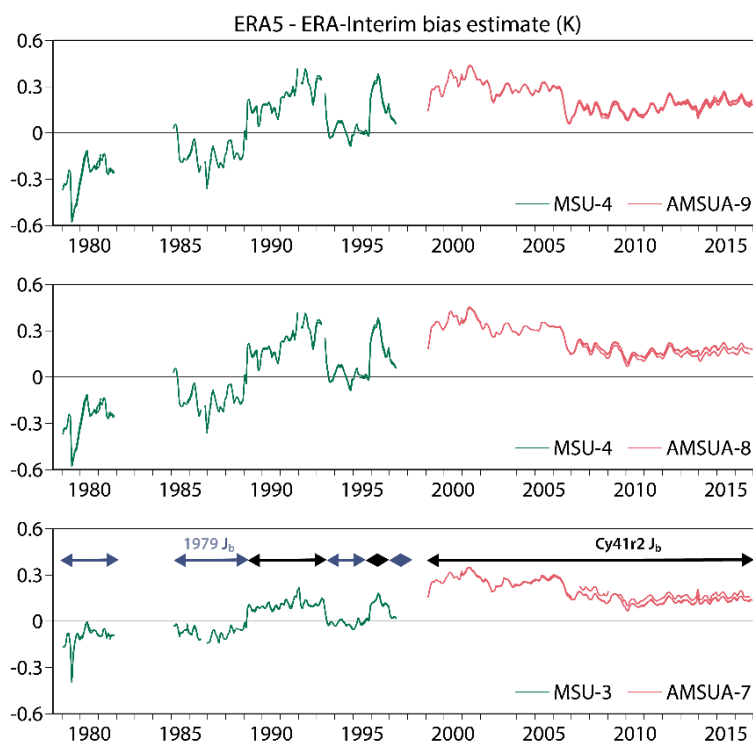


**Figure 2.** Monthly averages of differences between radiosonde temperature observations and ERA-Interim and ERA5 background equivalents. The periods of ERA5 prior to the year 2000 run with the Cy41r2 J<sub>b</sub> are being rerun using the 1979 J<sub>b</sub>.

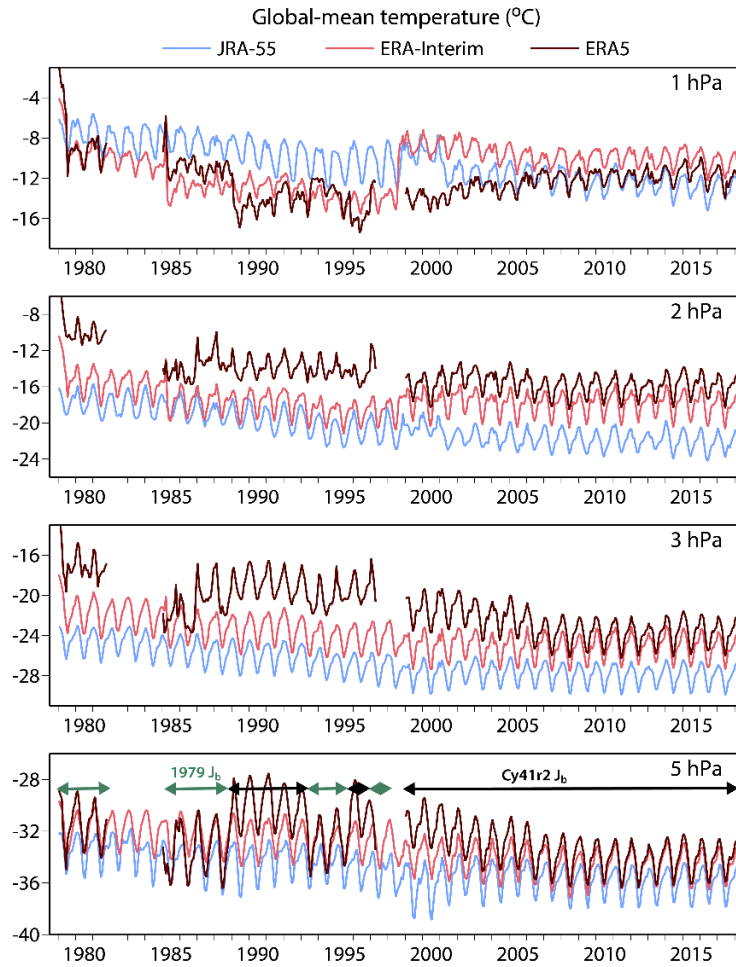




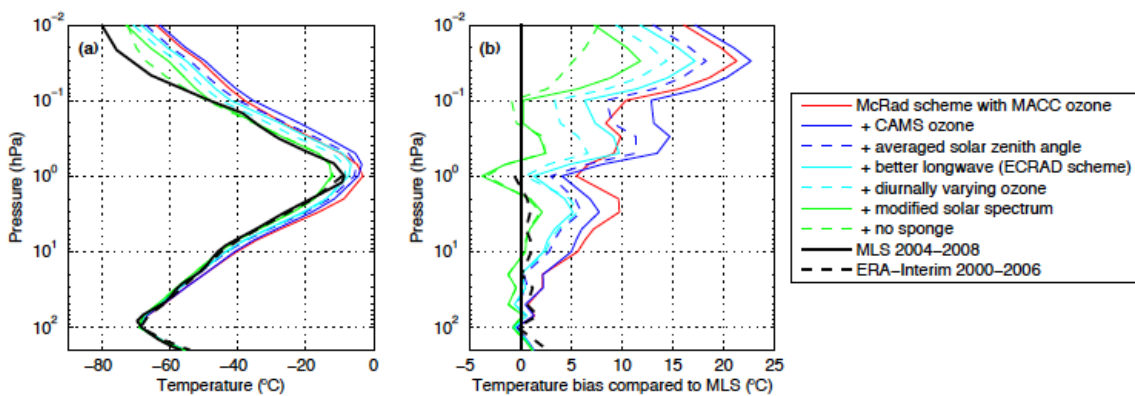
**Figure 3.** 365-day running mean of three-day 50 hPa temperature forecast errors from ERA-Interim, ERA5 and ECMWF operations, for the extratropical northern and southern hemispheres.



**Figure 4.** Monthly averages of differences between ERA5 and ERA-Interim bias estimates for several MSU and AMSUA channels on various satellites. The periods of ERA5 prior to the year 2000 run with the Cy41r2  $J_b$  are being rerun using the 1979  $J_b$ .



**Figure 5.** Monthly global mean temperatures at 1, 2, 3 and 5 hPa from JRA-55, ERA-Interim, and ERA5. The periods of ERA5 prior to the year 2000 run with the Cy41r2  $J_b$  are being rerun using the 1979  $J_b$ .



**Figure 6.** Annual-mean, global-mean temperature (left) and temperature bias with respect to Aura MLS measurements (right), from four 1-year uncoupled TL255L137 climate simulations using different configurations of the radiation scheme. From Hogan et al. (2017).

### 3 Horizontal Resolution

Many of the model investigations discussed in this report concern the IFS at TL255 resolution, where the model can be run for long enough to reliably determine its biases. However the operational resolution is much higher, currently TCo1279. It is thus important to understand how the model biases differ at the different resolutions. Global-mean temperature is a natural first metric to target, because of its low amount of internal variability. The surprising discovery is that global-mean temperature biases in the stratosphere are quite different between the two horizontal resolutions (Figure 7a). These differences are even larger in runs with the physics turned off (Figure 7b), suggesting they arise from dynamics. Since global-mean temperature in the stratosphere is largely under radiative control (Section 2), such sensitivity is puzzling. Whilst the lower temperatures produced by the higher resolution model are not necessarily a problem over most of the stratosphere, they exacerbate the overall cold bias in the lower stratosphere, between about 100 and 50 hPa, and this is a problem (Section 2). Note that in this region there is some dynamical control of global-mean temperature, partly through ozone feedbacks and partly through variations in static stability (Fueglistaler et al. 2011).

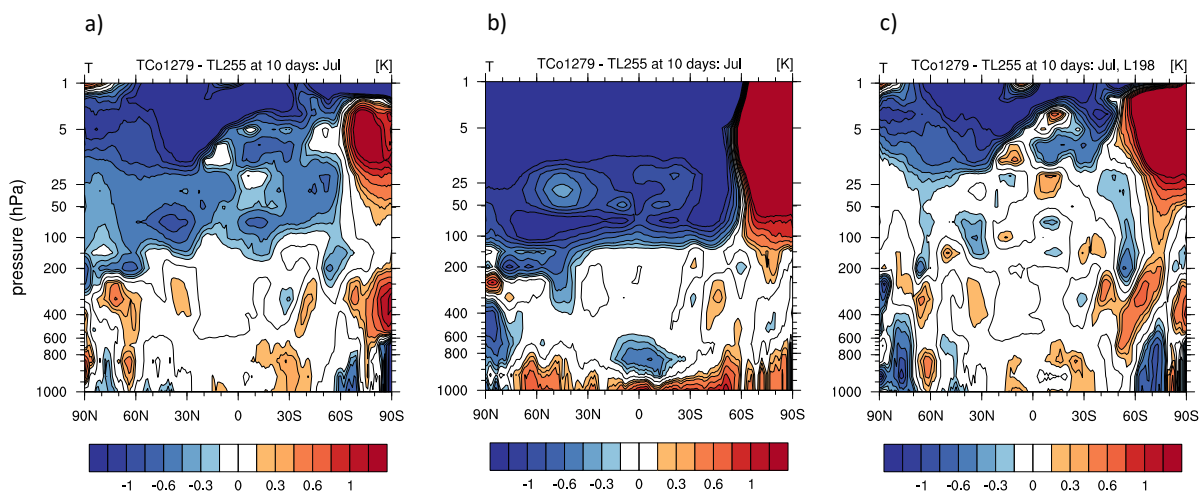
Theoretical arguments suggest that the horizontal/vertical aspect ratio should be roughly  $N/f$  in the extratropics (roughly 200 in the stratosphere), which the high horizontal resolution model certainly does not satisfy at L137 (i.e. there is insufficient vertical resolution). Moreover the required vertical resolution is even more demanding in the tropics (Lindzen and Fox-Rabinovitz 1989). Increasing the vertical resolution does indeed cure the horizontal-resolution sensitivity problem (Figure 7c). However, both the TCo199L91 and TCo319L91 versions of the IFS show a very similar cold bias in the lower stratosphere, especially across the tropics and subtropics. Focusing on the global mean 70 hPa temperature bias, increasing vertical resolution systematically reduces the model bias relative to reanalysis, with the higher horizontal resolution model having always a larger bias than the lower horizontal resolution model, but the difference disappearing (i.e. convergence) as the vertical resolution increases (Figure 8). For TCo199, 200 m resolution in this region (via L198) seems to be enough. However, already 250 m vertical resolution in this region (via L162) considerably improves the problem. Moreover L162 leaves the lower troposphere vertical resolution unchanged, eliminating the need to retune physics in the troposphere to the new vertical resolution.

The fact that a similar vertical resolution seems to work for both TCo199 and TCo1279 would seem to argue against the relevance of the  $N/f$  scaling. The  $N/f$  scaling applies to balanced dynamics, and it may well be that the dynamics is largely unbalanced at resolutions finer than those resolved at TCo199. This hypothesis is supported by high-altitude research aircraft measurements (Bacmeister et al. 1996), which exhibit a shallow,  $-5/3$  slope in their kinetic energy horizontal wavenumber spectrum around 20 km altitude (approximately 50 hPa), for wavelengths shorter than 600 km ( $n=60$ ). At these altitudes, the unbalanced dynamics will consist mainly of upward-propagating gravity waves, supplemented by parameterized gravity waves. These waves carry energy as well as momentum. Most attention is generally focused on the momentum deposition associated with gravity waves, which drives meridional circulations (Section 4), since the energy deposition can be balanced by thermal emission to space. The exception is the upper mesosphere, where the energy deposition from gravity waves and thermal tides is known to be a significant contributor to the thermodynamic balance. Because the resolved gravity-wave spectrum will depend sensitively on model settings, there can be a strong sensitivity of global-mean upper-mesospheric temperature to those settings (Sankey et al. 2007). Indeed, Figure 6 shows a visible impact of the removal of the sponge on global-mean temperature in the upper mesosphere.

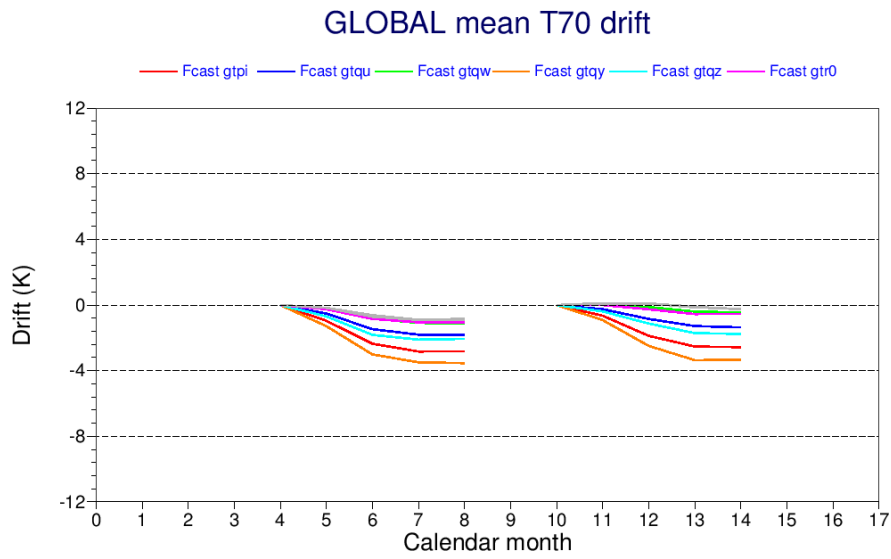
It may be that a similar phenomenon is behind the resolution sensitivity in the lower stratosphere. At these altitudes the radiative timescales are long (Hitchcock et al. 2010) and thus even small changes in heating rates can lead to discernible changes in temperatures. The energy deposition from parameterized non-orographic gravity-wave drag at TL255 provides a heating of 1-2 K/day in the subtropical lower stratosphere. At TCo1279 this parameterized energy deposition is much reduced and has to be provided instead by resolved gravity waves. Simulations where the resolved gravity waves at TCo1279 were strongly damped above 100 hPa, making them more comparable to what is represented at TL255, largely eliminated the resolution sensitivity in the 50-100 hPa region. A possible interpretation of these results is that whilst damping the resolved gravity waves forces energy deposition, allowing the waves to propagate makes the energy deposition sensitive to vertical resolution, with the energy lost to numerical dissipation at L137 but captured at higher vertical resolutions. More work on this problem is needed.

Lindzen and Fox-Rabinovitz (1989) discuss the resolution requirements for resolved gravity waves, but do not come to clear conclusions. It would be timely to revisit the vertical resolution question and update this classic study in view of the latest horizontal resolutions affordable with the IFS.

**Conclusions:** *The global-mean cold bias in the lower stratosphere (between 100 hPa and 50 hPa) was found to get worse as horizontal resolution increases. Such a sensitivity is surprising, but can be understood if the energy deposition from upward-propagating gravity waves is a significant contributor to the thermodynamic budget at these altitudes. Preliminary results suggest this is indeed the case, at least for the IFS. The problem does get better as vertical resolution increases, and 200 m vertical resolution in this region (via L198) seems to be enough to eliminate the difference in bias between TCo199 and TCo1279. Already 250 m (via L162) considerably improves the problem, and would avoid having to retune the physics in the troposphere. Further investigation of this issue is warranted. More generally, it would be timely to revisit the classic study of Lindzen and Fox-Rabinovitz (1989) concerning vertical resolution requirements, in view of the latest horizontal resolutions affordable with the IFS.*



**Figure 7.** Latitude-pressure cross-sections of zonal-mean temperature difference between TCo1279 and TL255 horizontal resolutions for an ensemble of 31 forecasts (ensemble mean shown) valid at 10 days in July. (a) 137L; (b) 137L with physics turned off; (c) 198L.



**Figure 8.** 70 hPa global-mean temperature bias as a function of horizontal and vertical resolution. Red and orange are L91 (TCo199 and TCo319 respectively); dark blue and light blue are L137; green and pink are L198; both grey lines are L320. Figure courtesy of Tim Stockdale.

## 4 Stratospheric meridional circulation and polar vortex variability

Whilst global-mean stratospheric temperature is largely under radiative control (Section 2, with the caveat noted in Section 3), its latitudinal structure is also affected by the meridional circulation. In contrast to the troposphere, where the meridional circulation can be viewed as thermally driven, the stratospheric meridional circulation is mechanically driven by the momentum transfer associated with dissipating waves, known as ‘wave drag’ (e.g. Shepherd 2000). From this perspective, the adiabatic cooling or warming associated with the upwelling or downwelling driven by the meridional circulation induces temperature departures from radiative equilibrium. (Radiative timescales can be up to several weeks in the stratosphere, so there is also a non-negligible transient component of the circulation.) The wave drag arises from both resolved and parameterized waves, through their interaction with the zonal-mean zonal wind. Since the latter is in thermal-wind balance, there are potential feedbacks to this wave-mean interaction.

The meridional circulation is not directly observable, nor is the wave drag associated with gravity waves. The primary observational constraints are the wave drag from Rossby waves (represented by the Eliassen-Palm flux convergence), and zonal-mean temperature. There are also indirect constraints from the transport of chemical tracers (Linz et al. 2017), but these reflect the combined effects of the meridional circulation and eddy mixing so are not easy to interpret (Miyazaki et al. 2016). Comparisons between different reanalyses (see also the SPARC S-RIP web site) generally show that the meridional circulations in more modern reanalyses (ERA-Interim, MERRA, JRA-55) broadly agree with each other in the lower stratosphere (Abalos et al. 2015) but diverge widely in the upper stratosphere, whilst the earlier reanalyses showed inconsistent behaviour throughout the stratosphere.

It is possible to diagnose the impact of unobserved or parameterized wave drag on the meridional circulation in a model, through what is called the ‘downward control’ principle (Haynes et al. 1991): namely the relation between wave drag (or other torque) and vertical mass flux via the zonal momentum balance and the mass continuity equation. This is then a model-sensitivity rather than a model-validation diagnostic.

Figure 9 shows the zonal cross-section of temperature biases relative to Aura MLS for two of the model versions shown in Figure 6. The various improvements to the radiation scheme discussed in Section 2 largely remove the global-mean biases, but there remain significant temperature biases at high latitudes, especially in the seasonal means. There is a particularly strong warm bias of up to 10 K evident in the SH winter upper stratosphere. Although the simulations shown in Figure 9 are relatively short, similar biases were seen in 32-year simulations in Polichtchouk et al. (2017), so they are believed to be robust. Such high-latitude temperature biases that have no imprint on the global mean point to biases in the meridional circulation, although there could potentially also be radiative contributions.

A major focus of the Task Force was to quantify the impact of various modelling choices on the stratospheric meridional circulation in IFS cycle 43R1, at TL255L137 resolution. A detailed discussion is provided in Polichtchouk et al. (2017, 2018), and only a few highlights are reprised here. The main sensitivity was found to be to the parameterization of non-orographic gravity-wave drag (NOGWD). It has long been known that gravity-wave drag is a significant contributor to stratospheric circulation, and in climate models, both orographic and non-orographic gravity wave drag are key aspects of the parameterization suite. Together they generally contribute a substantial fraction (one-third would be a typical value) of the wave drag driving both tropical upwelling and polar downwelling, which together represent the Brewer-Dobson circulation. However, the IFS is run at much higher horizontal resolution

than most climate models, which means that a considerable fraction of the gravity-wave spectrum (even at TL255) is resolved rather than parameterized. This is particularly the case with orographic drag, as the parameterization is explicitly tied to the unresolved topography.

Table 1 shows the annual-mean tropical upward mass flux, and the extended winter-mean polar cap downward mass flux for the two hemispheres, at 70 hPa (lower stratosphere) and 10 hPa (middle stratosphere). The contributions to these mass fluxes from both parameterized and resolved wave drag, inferred from the downward control principle, are also provided. The results are shown for three simulations: the control simulation, and simulations where the NOGWD source spectrum is either reduced or increased in magnitude by a factor of about four. Looking first at the control simulation, at 70 hPa the parameterized drag provides only 10% of the tropical and NH fluxes, and 20% of the SH flux. In the SH this is all coming from NOGWD. At 10 hPa the relative contributions to polar cap downwelling from parameterized drag increase substantially, as also found in climate models.

When NOGWD is either increased or decreased, the total tropical upwelling at 70 hPa is nearly unchanged. This points to a compensation between the resolved and parameterized wave drag driving lower-stratospheric tropical upwelling, as has been previously seen in a climate model (Sigmond and Shepherd 2014). However such a strong compensation is not seen for polar downwelling (also consistent with Sigmond and Shepherd 2014), which varies between  $13.5$  and  $19.3 \times 10^8$  kg/s in the SH and between  $20.7$  and  $23.2$  in the NH as the NOGWD is changed from reduced to increased values. This shows that even at the high resolution (in a climate-modelling context) of the IFS at TL255, NOGWD can exert quite some leverage on the stratospheric circulation at high latitudes. Moreover, the partial compensation seen in the extended-winter average hides the fact that the resolved wave-drag response to changes in NOGWD is offset within the seasonal cycle (see Polichtchouk et al. 2017, 2018). Thus, the effect of NOGWD on the evolution of the seasonal cycle is even more pronounced.

The effect of NOGWD on the most important aspects of stratospheric polar variability — the final vortex breakdown in the SH, and stratospheric sudden warmings (SSWs) in the NH — are shown in Figures 10 and 11, respectively. These phenomena provide the main mechanisms through which stratospheric variability influences the troposphere, so are worthy of close study from the perspective of prediction. In the SH, the final breakdown can be advanced by several weeks when NOGWD is increased (Figure 10). This is similar to what was found in McLandress et al. (2012), using orographic GWD in a climate model. This sensitivity is pertinent because the timing of the stratospheric vortex breakdown is generally too late in climate models (Butchart et al. 2011), and the timing of the breakdown appears to affect tropospheric summertime circulation (Byrne et al. 2017). In the IFS cycle 43R1 at TL255L137 resolution, the current NOGWD settings seem to be optimal. With regard to SSWs, increased NOGWD reduces the amplitude and persistence of the events, while decreased NOGWD has the opposite effect (Figure 11). Once again, the current NOGWD settings seem to be optimal for this version of the model. Note, however, that this comment applies only to the polar vortex variability; there remain significant mean temperature biases in the polar upper stratosphere, especially during the winter seasons (Figure 9).

Nudging, where the troposphere is nudged to ERA-Interim, is an efficient method for conducting case studies and isolating the impact of various modelling choices on, e.g., a particular SSW, as nudging guarantees that the observed planetary wave fluxes enter the stratosphere, thereby initiating the SSW in the model. By providing such conditioning on the dynamical forcing from the troposphere, which is otherwise chaotically varying, nudging eliminates the need for long integrations and/or large ensemble

sizes. Polichtchouk et al. (2018) used nudging to evaluate the impact of NOGWD on the recovery phase of the long lived 2006 SSW, which had a strong influence on the troposphere. The impact of NOGWD determined in this way was found to be the same as for the SSW statistics from the 32-year free-running model. Nudging could be a useful way of quantifying the effect of radiative changes in high-latitude regions.

The impact of stratospheric polar vortex variability on the tropospheric annular modes — the main indicator of stratosphere-troposphere dynamical coupling, with implications for tropospheric predictability (e.g. Thompson et al. 2002) — is shown in Figures 12 and 13 for the NH and SH, respectively. In the NH (Figure 12), the variability is defined in terms of weak and strong stratospheric polar vortex anomalies. (SSWs are weak vortex anomalies.) The coupling is strengthened when NOGWD is reduced, and weakened when NOGWD is increased, consistent with the effect of NOGWD on SSW amplitude and persistence seen in Figure 11. The comparison suggests that the stratosphere-troposphere coupling in the NH mainly depends on the strength and persistence of the stratospheric anomalies.

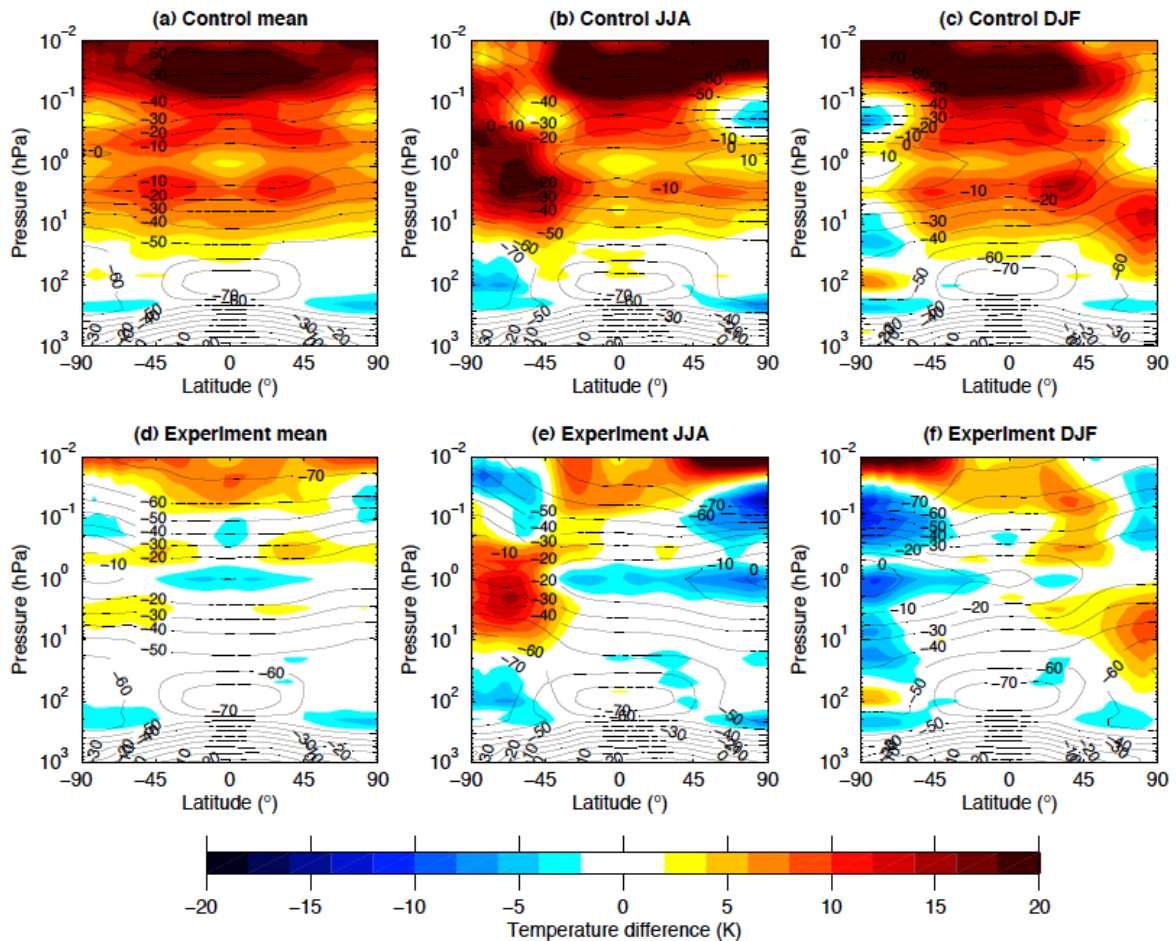
In the SH (Figure 13), the stratospheric polar vortex variability is mainly associated with inter-annual variability in the seasonal cycle leading to the annual vortex breakdown (Byrne and Shepherd 2018). It is defined here in terms of weak and strong polar vortex evolutions, corresponding respectively to early and late vortex breakdowns. In this case, opposite to the situation in the NH, the coupling is weakened when NOGWD is reduced, and strengthened when it is increased. This reflects the primary effect of NOGWD on the seasonal evolution of the vortex; too strong a vortex during the breakdown period reduces the potential for stratosphere-troposphere coupling. Thus, the two hemispheres have quite different sensitivities to NOGWD in terms of stratosphere-troposphere coupling. There is a suggestion that the coupling may be slightly too weak in both cases, for the model version shown.

There is some evidence that SH tropospheric variability during spring, prior to the vortex breakdown, can be predicted from stratospheric initial conditions in late winter (Seviour et al. 2014). Figure 14a shows that in the observations, SH stratospheric polar vortex anomalies persist through late winter and then propagate down to the troposphere during October. Figure 14b shows that the corresponding anomalies in the ensemble members of SEAS5 decay much too rapidly, and fail to couple to the troposphere. As a result, there is essentially no predictability of tropospheric springtime variability from August 1 forecasts in SEAS5 (Figure 14c), in contrast to what is seen in the Met Office GloSea5 system (Seviour et al. 2014). It is interesting that SEAS4 did exhibit predictability during this time of year (Figure 14d). This may be connected with the fact that the polar vortex breakdown in SEAS4 is fairly realistic, whereas it is much too late in SEAS5 (not shown). Further investigation of this issue is warranted.

**Conclusions:** *Even at the relatively high resolution (in a climate-modelling context) of TL255, parameterized gravity-wave drag is an important driver of meridional circulation and polar vortex variability in the IFS. It is less critical for lower stratospheric tropical upwelling because of the compensation between resolved and parameterized drag in this region. NOGWD dominates the SH, and both orographic GWD and NOGWD are important for the NH. As there are no direct observational constraints on GWD, the parameterizations need to be tuned to obtain realistic polar vortex variability and the associated stratosphere-troposphere dynamical coupling. The most important aspects for predictability are the seasonal evolution and timing of the annual vortex breakdown in the SH, and SSWs in the NH. Nudging is a useful way to obtain robust results from short simulations of the recovery*



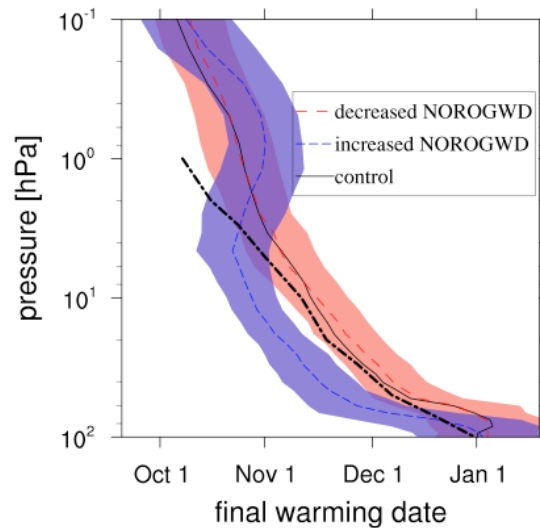
phase of SSWs, which affects stratosphere-troposphere coupling. Nudging could also be a useful way of quantifying the effect of radiative changes in high-latitude regions, where upper-stratosphere temperature biases of up to 10 K remain. Whilst the NOGWD settings in the IFS (for cycle 43R1, at TL255L137) appear to be optimal for polar vortex variability, they need to be monitored closely as the model evolves or is used in other configurations. SEAS5 seems to lack the SH springtime stratosphere-troposphere coupling and associated predictability that was present in SEAS4, presumably because of an unrealistic seasonal evolution of the annual vortex breakdown in SEAS5.



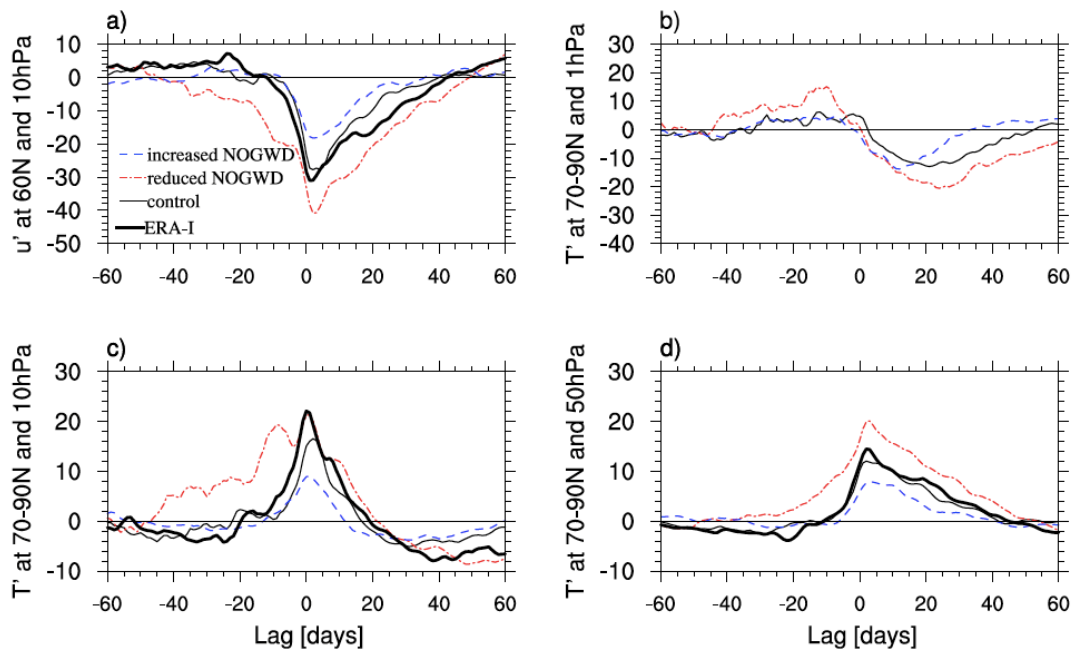
**Figure 9.** Mean temperature from the first and last IFS simulations shown in Figure 6: (top row) McRad scheme with MACC ozone, and (bottom row) after multiple changes as indicated in Figure 6. The black contours show temperature and the colours show the difference against a reference dataset consisting of the Aura MLS climatology at pressures of 100 hPa and less, and ERA-Interim at pressures greater than 100 hPa. The left column shows the annual mean, the middle column the northern-hemisphere summer and the right column the northern-hemisphere winter. From Hogan et al. (2017).

**Table 1.** Resolved and parameterized (OGWD and NOGWD) wave drag contribution (in % of the total) to the annual-mean tropical mass flux and extended winter (Mar-Nov for the SH, and Oct-May for the NH) polar cap downward mass flux for the control, reduced NOGWD and increased NOGWD runs at 10 hPa and at 70 hPa, for the IFS at TL255L137 resolution. Positive percentage denotes tropical upwelling and polar cap downwelling, and negative percentage denotes tropical downwelling and polar cap upwelling. From Polichtchouk et al. (2018).

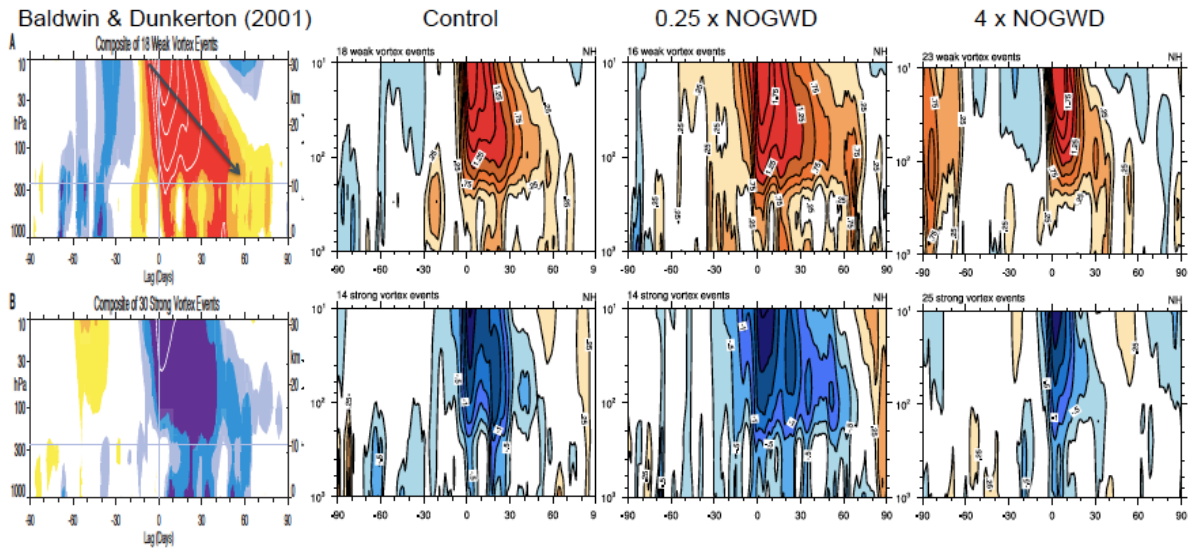
| Experiment      | Region                   | Pressure<br>[hPa] | Parameterized wave drag [%] |       |      | Resolved wave drag<br>[%] | Mass flux<br>$\times 10^8$ [kg/s] |
|-----------------|--------------------------|-------------------|-----------------------------|-------|------|---------------------------|-----------------------------------|
|                 |                          |                   | OGWD                        | NOGWD | All  |                           |                                   |
| Control         | Annual-mean upwelling    | 10                | 2.4                         | -2.4  | 0    | 100                       | 15.5                              |
|                 |                          | 70                | 5                           | 2     | 7    | 93                        | 58.1                              |
|                 | NH polar cap downwelling | 10                | 14                          | 5     | 19   | 81                        | 5.7                               |
|                 |                          | 70                | 7                           | 0     | 7    | 93                        | 22.2                              |
|                 | SH polar cap downwelling | 10                | 0                           | 40.6  | 40.6 | 59                        | 5.8                               |
|                 |                          | 70                | 0                           | 19    | 19   | 81                        | 15                                |
| Reduced NOGWD   | Annual-mean upwelling    | 10                | 2.5                         | -3    | -0.5 | 100.5                     | 16.1                              |
|                 |                          | 70                | 5                           | -3    | 2    | 98                        | 57.1                              |
|                 | NH polar cap downwelling | 10                | 14                          | -9    | 5    | 95                        | 6.2                               |
|                 |                          | 70                | 8                           | -6    | 2    | 98                        | 20.7                              |
|                 | SH polar cap downwelling | 10                | 0                           | 12    | 12   | 88                        | 5.4                               |
|                 |                          | 70                | 0                           | 6     | 6    | 94                        | 13.5                              |
| Increased NOGWD | Annual-mean upwelling    | 10                | 4                           | 6     | 10   | 90                        | 11.8                              |
|                 |                          | 70                | 4                           | 16    | 20   | 80                        | 57.4                              |
|                 | NH polar cap downwelling | 10                | 7.5                         | 38.5  | 46   | 54                        | 5.2                               |
|                 |                          | 70                | 5                           | 14    | 19   | 81                        | 23.2                              |
|                 | SH polar cap downwelling | 10                | 0                           | 88    | 88   | 12                        | 6.7                               |
|                 |                          | 70                | 0                           | 45    | 45   | 55                        | 19.3                              |



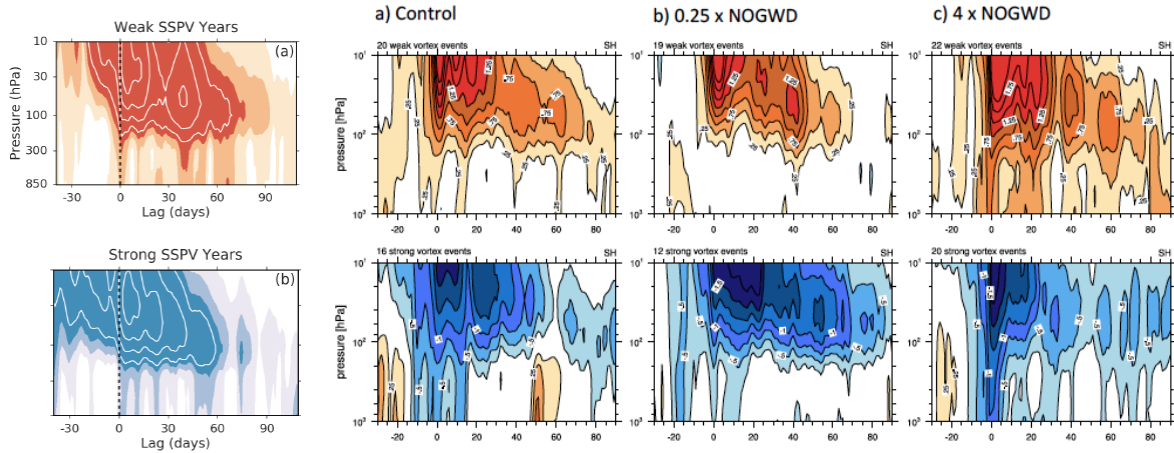
**Figure 10.** Average of the final warming dates in the SH for the control run (solid black), the reduced NOGWD run (long-dashed red) and the increased NOGWD run (short-dashed blue), for the free-running IFS at TL255L137 resolution, Cy43R1. The average of the ERA-Interim final warming dates between 2004 and 2015 is shown in thick dot-dashed black contour. The shading shows the  $2\sigma$  interval for the increased and reduced NOGWD runs only. From Polichtchouk et al. (2018).



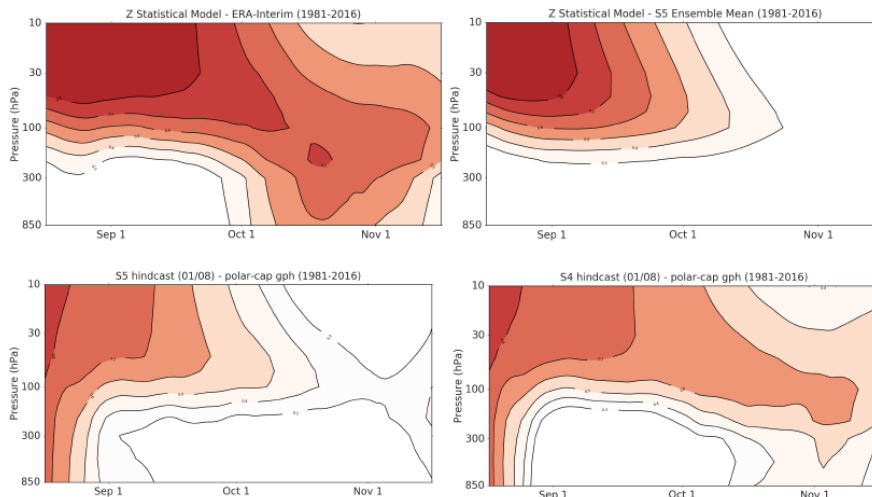
**Figure 11.** Composites of all SSWs for the control run (thin solid black), reduced NOGWD run (dot-dashed red) and increased NOGWD run (dashed blue), for the free-running IFS at TL255L137 resolution, Cy43R1. Thick black line shows composites of SSWs from the ERA-Interim reanalysis between 1979 and 2016. (a) Zonal-mean zonal wind anomaly at  $60^\circ\text{N}$  and 10 hPa (in m/s); polar-cap average (from  $70^\circ\text{N}$  to  $90^\circ\text{N}$ ) zonal-mean temperature anomalies (in K) at (b) 1 hPa; (c) 10 hPa; and (d) 50 hPa. From Polichtchouk et al. (2018).



**Figure 12.** Composite plots of NH Annular Mode indices for weak and strong vortex events as defined using the Annular Mode index at 10 hPa. Shading interval and contour interval are both 0.25 standard deviations. Shading is drawn for values greater than  $\pm 0.25$  standard deviations. Left column shows observations, from Baldwin and Dunkerton (2001). The other columns show results for the free-running IFS at TL255L137 resolution, Cy43R1, for the control, reduced NOGWD, and increased NOGWD runs.



**Figure 13.** Composite plots of SH Annular Mode indices for weak and strong years as defined using the Annular Mode index at 30 hPa for observations and 10 hPa for the model. Shading interval is 0.25 standard deviations and contour interval is 0.5 standard deviations for the observations, and 0.25 for the model. Shading is drawn for values greater than  $\pm 0.25$  standard deviations. Left column shows observations (ERA-Interim), from Byrne and Shepherd (2018). The other columns show results for the free-running IFS at TL255L137 resolution, Cy43R1, for the control, reduced NOGWD, and increased NOGWD runs.



**Figure 14.** Top panels: Correlation between daily polar cap average geopotential height as a function of day of year and pressure level with the value at 10 hPa on August 1, from (left) ERA-Interim over 1981-2016, and (right) the 25 ensemble members of SEAS5. Bottom panels: Correlation with ERA-Interim over 1981-2016 of the ensemble mean hindcast of daily polar cap average geopotential height, for forecasts initialized on August 1, from (left) SEAS5, and (right) SEAS4. Figure courtesy of Nick Byrne, University of Reading.

## 5 Extratropical lowermost stratosphere temperature

Figure 9 also reveals a cold bias of up to around 5 K in the extratropical lowermost stratosphere, in both the NH and SH, peaking around 200 hPa poleward of 60 degrees latitude, and most severe during the summer season. This is a longstanding bias in climate models, and has been noted in the IFS for some time, going back at least to 1990. Attention has long focused on the possible role of water vapour: the lowermost stratosphere is exceedingly dry (as first pointed out by Brewer 1949), and it can be expected that models will fail to maintain a realistically sharp gradient across the tropopause because of limited vertical resolution. Any leakage of water into the lowermost stratosphere would lead to a cold bias in this region, because of the radiative cooling from water vapour. This can be expected to affect forecast scores, through the effect on tropopause height and thus storm dynamics. Indeed, about 20 years ago, there was a dramatic degradation in forecast scores from an inadvertent leaking of moisture into the lower stratosphere.

The current IFS appears to have a moist bias in the lowermost stratosphere. Figure 15 shows the bias against Aura MLS. Although the vertical resolution of Aura MLS is limited, an earlier comparison of ECMWF analysis with CARIBIC aircraft observations (Dyroff et al. 2015) showed a persistent moist bias in the lowermost stratosphere, even though the upper tropospheric moisture was perfect. Artificially reducing the water vapour seen by the radiation scheme in the IFS around the extratropical tropopause reduces the cold bias (Figure 16), and even seems to improve forecast scores (Figure 17). This Task Force sensitivity experiment suggests that targeting the moist bias would at the same time alleviate the cold bias and improve both analyses and forecasts.

This then raises the question of what is the origin of the moist bias. Blackburn (1997) showed that the mitigation of the cold bias in the IFS at that time that resulted from the inclusion of semi-Lagrangian

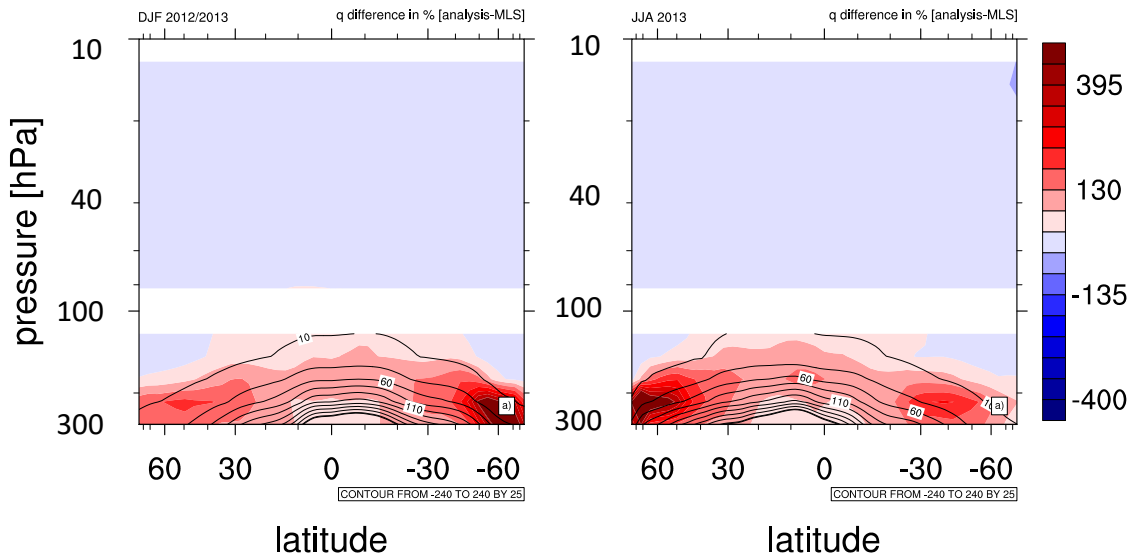
advection was explained by the resulting change in the water vapour. A variety of evidence was presented in Task Force meetings suggesting that leakage of water into the lowermost stratosphere continues to be an issue. First, strengthening the limiter in the semi-Lagrangian advection scheme exacerbates the cold bias via increased moistening of the lower stratosphere (M. Diamantakis). Second, more diffusive numerics makes the bias worse, whilst less diffusive numerics makes it better (R. Forbes). Third, the cold bias is already present (and growing) in 24-hour and 5-day forecasts; removing the humidity bias in the initial conditions controls the temperature bias for 10-day forecasts, but it develops after that (R. Forbes). Fourth, ERA5 is moister than ERA-Interim in the lower-latitude lower stratosphere, mainly because of moistening in the boreal summer. The spatial pattern of moistening seems realistic, but is it too much moistening? Aircraft measurements suggest that the moistening comes from over-shooting convection, which may be too intense in ERA5.

The importance of the initial condition for humidity means that the moist bias could potentially be controlled through data assimilation, at least for short time horizons. Currently, humidity increments are disallowed in the stratosphere, because small biases in the upper troposphere could lead to large increments in the stratosphere. In a sensitivity experiment, turning the increments on led to a pronounced drying in the lower stratosphere, spreading outwards from the tropics (E. Holm). This outward spreading of the signal from the tropics is consistent with transport in the lowermost stratosphere. The effects were seen in the temperature in the analysis (as a warming). Overall, after three months of assimilation, the forecast scores were not degraded but the biases were improved. However, it seems equally likely that this procedure could have made things worse. Thus, this experiment shows the potential benefit of introducing humidity information in the stratosphere, even in the tropics, although the solution must be to assimilate stratospheric measurements. Perhaps even sparse measurements could be used to bring climatological information into the background, since the memory of humidity increments in the lower stratosphere can be expected to persist for months, and to spread from the tropics into the extratropics. For example, water vapour from the ACE-FTS limb sounder has high (roughly 1 km) vertical resolution because of the high precision from the solar occultation technique (Figure 18), but this comes at the cost of very limited sampling. Yet even such sparse data can provide climatological information when the data is considered in context (Figure 18).

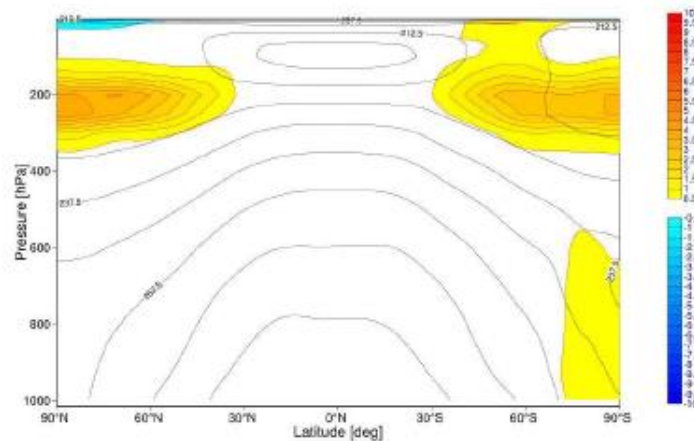
The expectation would have been that leakage of water vapour from the upper troposphere into the lowermost stratosphere in a model would be a result of insufficient spatial resolution and inaccurate transport, and would get better as resolution improved over time. But then why is the problem still there at TCo1279L137 resolution? Moreover, there seems to be no benefit obtained from going from 300 m to 200 m vertical resolution, and no sensitivity to time step. This suggests that the dynamics around the tropopause might not converge with increasing spatial resolution (as it would if it consisted only of stirring by synoptic-scale disturbances), but may involve a complex mesoscale spectrum of moist processes (e.g. moist conveyor belts, overshooting convection) and unbalanced motion. Indeed, there is much active research on such processes. Such a mesoscale spectrum would become more active in the model as resolution increases, but would not be well resolved. In any case, the moist bias and associated cold bias problems remain, with no immediate solution being apparent.

**Conclusions:** *The IFS exhibits a notable cold bias of up to 5 K just above the extratropical tropopause, which maximizes at high latitudes in the summer season. This is a robust feature in models, and has been present in the IFS for a very long time. All evidence points to the cause being too much moisture leaking in to the region from the upper troposphere. The surprising thing is that the problem is not improved by increased spatial resolution. This is consistent with the view that there is a complex*

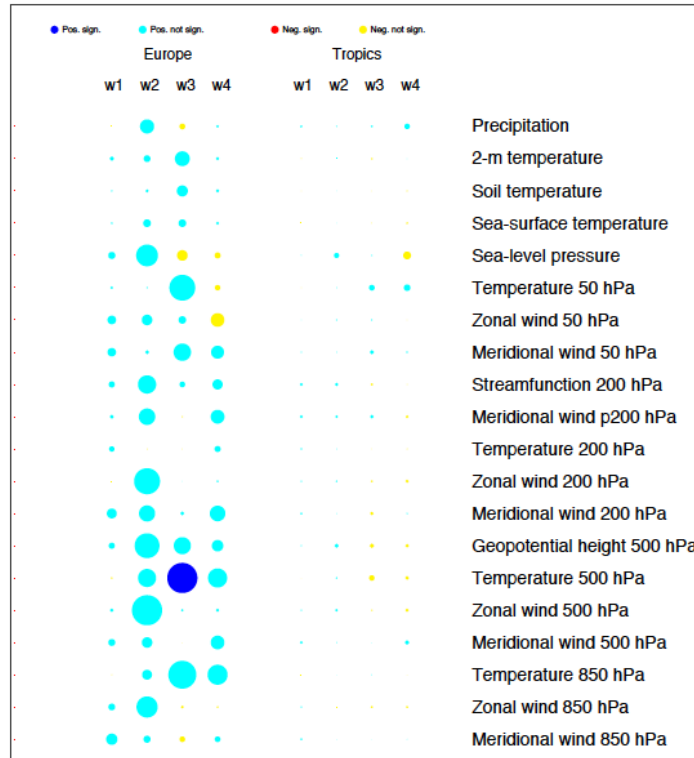
*mesoscale spectrum of moist processes and unbalanced motion around the tropopause, which becomes more active in the model as resolution increases, but is not well resolved. Diagnostics targeting such processes would be useful, and the sensitivity of cross-tropopause water vapour transport to numerics should be explored. High vertical-resolution water vapour measurements should be used for model evaluation, and are available from aircraft campaigns as well as from the ACE-FTS limb sounder. The possibility of assimilating sparse vertically-resolved stratospheric water vapour measurements should be explored.*



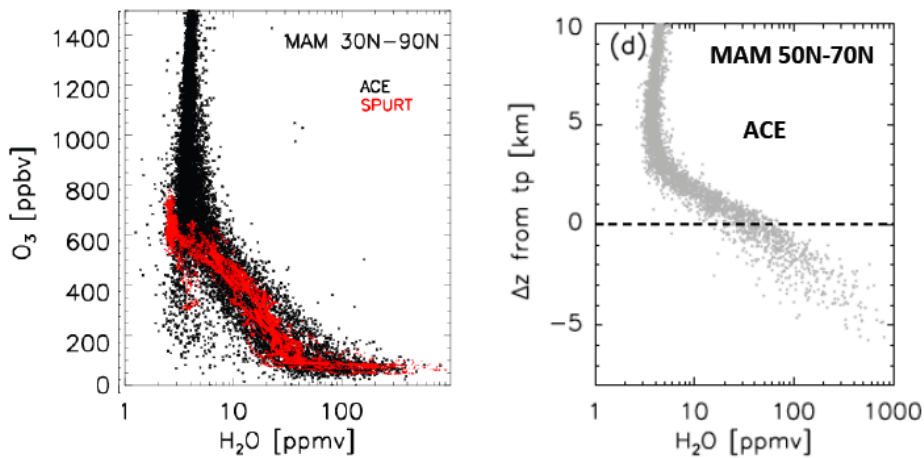
**Figure 15.** Bias in water vapour (in %) of the operational analysis from 2012/2013 (based on Cy38R2) with respect to Aura MLS, for DJF (left) and JJA (right).



**Figure 16.** Impact on zonal-mean temperature of artificial reduction of water vapour above the extratropical tropopause.



**Figure 17.** Changes in forecast scores resulting from the change shown in Figure 16. Figure courtesy of Frédéric Vitart.



**Figure 18.** Data from the ACE-FTS limb sounder. Left panel shows scatterplots of coincident water vapour vs ozone measurements in the NH extratropics during spring. Note that water vapour is plotted on a logarithmic scale. The vertical branch at the top is stratospheric air, the horizontal branch at the bottom is tropospheric air, and there is a transition layer in between. The red points show observations (taken in different years) from the SPURT aircraft campaign, and reveal that the ACE-FTS observations are nearly as precise as the aircraft measurements. From Hegglin et al. (2008). Right panel shows vertical profiles of ACE-FTS water vapour relative to the location of the tropopause, showing that the transition to dry stratospheric air occurs within 2 km above the tropopause. From Hegglin et al. (2009).



## 6 Sponge layer

Atmospheric waves of various types are generated in the troposphere and propagate upwards. Because of the decreasing atmospheric density with altitude, the waves will grow in amplitude and eventually break. However, a model has a lid and there is the potential for wave reflection from the lid. Thus, models need some kind of absorbing upper boundary condition. The normal approach is to use a sponge layer, with either a linear relaxation or an enhanced horizontal diffusivity. As a purely numerical device — a vertical spatial analogue of the dissipation range in a wavenumber spectrum — the sponge region should not be regarded as physically meaningful and should be placed above the region of interest. Sponge layers need to be implemented with care, since if the damping rates vary too rapidly with altitude, then the sponge layers can themselves cause reflection. A general rule of thumb is that a sponge layer should span at least two density scale heights and should switch on over at least one scale height. The model lid in the IFS is at ~80 km (0.01 hPa). Assuming a scale height of 8 km, the sponge should start switching on at ~1 hPa and attain full strength at ~0.3 hPa.

It has been common practice in global atmospheric modelling to apply sponge layers to the zonal mean flow, which has the practical benefit of controlling zonal wind speeds. In this way, sponge layers have frequently been used as a surrogate for gravity-wave drag. However, this then introduces torques that are unphysical, since gravity-wave drag represents momentum flux convergence, and GWD perturbations induced by zonal-wind variations are dipoles rather than the monopoles that would be induced through a zonal-mean sponge (Shepherd and Shaw 2004). In particular, the wave drag applied within a sponge layer by the absorption of upward propagating resolved waves should be driving a meridional circulation, but if there is a zonal-mean component to the sponge, the drag force will be compensated by the sponge rather than by the meridional circulation, nullifying the induced circulation.

Fundamentally, zonal-mean sponge layers violate the momentum conservation that is implicit in any physically-based gravity-wave drag parameterization, and which underlies the mechanisms driving the stratospheric circulation; violating momentum conservation can lead to erroneous meridional circulations, with knock-on effects below (Shaw et al. 2009). There is furthermore no need for a zonal-mean sponge, since vertically propagating waves have no zonal-mean component. Thus, one goal of the Task Force was to work towards the elimination of the zonal-mean component of the sponge; in a spectral model, this move has no computational cost.

The other goal of the Task Force was to rationalize the treatment of the sponge, since in the normal configuration of the IFS it begins at 10 hPa, which is extremely low. It is also not always applied equally to different variables, and is sometimes enhanced in the vicinity of the equator — presumably to control equatorial inertial instability. None of these choices would seem to be physically justified. Currently a sponge that applies an equal amount of damping on vorticity and divergence above 1 hPa and which does not damp the zonal-mean fields is in development. This experimental sponge also does not damp total wavenumbers less than  $n=10$ , although it is not clear that this is appropriate since there will be planetary-scale upward-propagating waves, most notably thermal tides. Preliminary results indicate that whilst the free-running IFS behaves well with this new sponge, stability problems arise when coupled to data assimilation. In particular, the minimisation in the 4DVAR system struggles with the large amplitude wave structures in the mesosphere. However, most tests have been done with cycle 43R1. It would be worth exploring whether the difficulties are still present with the modified background error covariance matrix (see Section 7).

**Conclusions:** *The sponge layer in the IFS seems to have evolved in a very ad hoc way, presumably to control particular problems through damping. But unless damping is designed in a physically appropriate way, it can lead to other problems. To avoid the situation of compensating errors, it is essential that the sponge in the IFS have minimal adverse impact on the atmospheric state in the domain of interest, i.e. below about 0.3 hPa (60 km). However, since unjustified features of the sponge are usually there for a reason, they cannot simply be removed, but must somehow be replaced by something more physical. Some progress was made by the Task Force in this respect, but the effort is unfinished and needs to continue.*

## 7 Tropical zonal winds

In the tropics, the balance between wave drag and meridional circulation that characterizes the extratropics does not apply. Instead, wave drag generically drives oscillating zonal winds, which propagate downward if the waves are propagating upward. This fluid-dynamical phenomenon is well understood theoretically. Its most famous manifestation is the stratospheric Quasi-Biennial Oscillation (QBO), which has a varying period around an average of about 28 months. The QBO is important for forecasts because it is known to affect the variability of the polar vortices at particular times of the year, which is communicated down to the troposphere (Thompson et al. 2002; Anstey and Shepherd 2014). Modelling of the QBO is especially important for seasonal forecasts, which would otherwise lose this source of predictability from the initial conditions.

The QBO is understood to be driven by a combination of low-frequency equatorial waves and inertia-gravity waves, although the observational constraints on the different components of the forcing are limited. This immediately suggests that accurate modelling of the QBO will be challenging. Whilst the low-frequency equatorial waves are in principle resolvable, they are forced by parameterized processes such as tropical convection, and inertia-gravity waves will be partly resolved (though likely inaccurately) and partly parameterized. Atmospheric models with sufficient vertical resolution (generally finer than 1 km) to resolve the wave, mean-flow interaction behind the QBO generally exhibit a “QBO-like” oscillation, but the magnitude and period of the oscillation can depart significantly from observations. This is understandable on theoretical grounds, and it is not expected that a realistic QBO can be simulated from first principles.

The sensitivity of the QBO to details of model specifications was explored in detail in Polichtchouk et al. (2017). As might be expected from the discussion above, the QBO amplitude and phase are sensitive to the launch spectrum magnitude of the NOGWD scheme, as well as to the numerics through the TCo grid and the SPPT scheme. The latter can be expected to increase the magnitude of the resolved wave forcing that helps drive the QBO. In the present state of knowledge, it appears to be necessary to tune the QBO via the NOGWD scheme, for any particular setting of the model.

In an analysis, the lower stratospheric portion of the QBO can be discerned from radiosonde wind measurements, despite their sparseness in the tropics. The model will then generate the upper portion of the QBO through wave, mean-flow interaction, though it need not be realistic and may just be a model construct. Since the zonal-mean zonal wind in the tropics is in thermal-wind balance, the QBO winds are constrained in principle by temperature measurements, although the constraint on zonal wind from temperature is much weaker than in the extratropics. The weak coupling between temperature and winds in the tropics means that zonal wind anomalies can persist for a very long time (i.e. years), because

radiative damping has almost no effect on them (Scott and Haynes 1998). Thus, wind errors can only be controlled by something acting directly on winds.

One curiosity found in the operational model from March 2016 was the development of a very strong westerly equatorial jet (reaching speeds of 160 m/s) centred around 0.1 hPa during October and May. A slightly milder form of this feature, reaching speeds of 100 m/s, seems to have been already present in the operational model from 2013 in cycle 38R2 (H. Hersbach). Its semi-annual appearance suggests that it is connected with the stratopause Semi-Annual Oscillation (SAO), which exhibits westerlies during the equinox seasons (understood to be driven by Kelvin waves) and easterlies during the solstice seasons (driven by advection across the equator from the summer to the winter hemisphere). The easterly phase of the SAO should be a robust feature of any model, but as with the QBO, the westerly phase can be expected to be sensitive to model details, as it is found to be in the IFS (Polichtchouk et al. 2017). The westerly equatorial mesospheric jet in the operational model from March 2016 was not physically implausible, but its amplitude was clearly unrealistic in comparison with the real atmosphere. It did not develop in the free-running version of the model, which suggests it arose from the influence of increments in data assimilation. Unfortunately, the strong mesospheric jet is also present in ERA5, although its intensity varies considerably from year to year (H. Hersbach).

The vertical propagation of information into regions unconstrained by observations (and the equatorial zonal winds at 0.1 hPa will be unconstrained by any observations) is a classic challenge in high-top models. Error variances can be very large in the mesosphere, and the computed error correlations applied to stratospheric observations will inevitably, through insufficient sampling, project into the mesosphere (Polavarapu et al. 2005). Thus, some sort of vertical localization might need to be considered in this region. Indeed, the unrealistically strong westerly mesospheric jet appears to have significantly diminished since the introduction of cycle 43R3 in July 2017 due to a modification of the climatological part of the background error covariance matrix used in the data assimilation system.

***Conclusions:** Tropical zonal winds in the stratosphere are dominated by the SAO in the stratopause region and by the QBO in the rest of the stratosphere. Both phenomena are driven by drag from a combination of low-frequency equatorial waves and inertia-gravity waves, neither of which can be expected to be accurately represented in a model because of their strong sensitivity to parameterized processes and to numerics. The QBO is important for seasonal prediction and needs to be tuned through the NOGWD scheme. The SAO is probably not so important, so the philosophy there should be to ensure that it does not cause detrimental effects. Some attention should be paid to the vertical propagation of wind increments through DA in the tropics, where there can be nothing to limit model error.*

## 8 Other issues

One issue that was only briefly touched on in Task Force meetings is the potential problems associated with noise from resolved gravity waves in the model. Even in coarse-resolution climate models, the mesospheric state is highly variable and dominated by gravity waves, which have a shallow kinetic energy wavenumber spectrum (Shepherd et al. 2000). This will be even more the case in high-resolution models such as the IFS, and will affect background error variances. The importance of this can be seen in the sensitivity of ERA5 lower stratospheric temperature to the  $J_b$  used (Section 2). Gravity waves are also present in observations, especially in radiosonde profiles, and if large enough can lead to rejection of the entire stratospheric profile in the assimilation. In some cases, the gravity waves are orographic and are reasonably well represented in the model. But non-orographic gravity waves will probably not

be well represented in the model. The realism of the resolved gravity-wave spectrum and its effect on data assimilation needs further study.

The diurnal solar (thermal) tide is a prominent feature of the middle atmosphere. Whilst its direct effect on stratospheric dynamics is minimal, its representation in a model is important for data assimilation in order to avoid introducing biases associated with the local time of the measurement. The solar tide is a large-scale, low-frequency phenomenon; there is a propagating component forced by solar heating primarily via water vapour and ozone, and a non-propagating component forced by the diurnal cycle of convection. The former should be representable from first principles in a model; the latter will depend on the convective heating. The tide is also modulated by the zonal-mean winds (McLandress 2002), thus is sensitive to the QBO and SAO. The Task Force did not examine the realism of the tide in ECMWF analysis and modelling systems, but this should be investigated.

Another issue only briefly touched on is the role of stratospheric composition (apart from water vapour, which was discussed in Section 5). In particular, stratospheric ozone has a first-order effect on radiative heating, yet is a highly dynamic field. Use of climatological ozone will thus inevitably introduce state-dependent biases, for which the only remedy is prognostic ozone. However, ozone is strongly slaved to the meteorology — this is the principle behind off-line Chemical Transport Modelling — which means that, at least in principle, a realistic ozone field can be obtained from modelling of ozone chemistry, without assimilation of ozone measurements. In practice, realistic spatial structures in ozone — typically of far higher horizontal and vertical resolution than are present in satellite observations — are readily produced from modelling, and assimilation is mainly needed to correct long-term biases in the model climatology. Chemical data assimilation is thus a very different challenge from that faced in the assimilation of meteorological quantities, and can be left for a second step. With respect to modelling ozone, there are various simplified schemes that can be considered. In the tropical lower stratosphere, where ozone variability is dynamically driven by Brewer-Dobson upwelling and has a significant effect on temperature (Fueglistaler et al. 2011; see also Figure 10 of McLandress et al. 2014b), a parameterization consisting of vertical advection balanced by linear photochemical relaxation could be enough. In polar regions, where ozone is strongly affected by dynamical variability, nudging (see Section 4) could be used to efficiently tune the ozone scheme.

**Conclusions:** *The large amplitude of gravity waves in the stratosphere, both in observations and in models, presents a variety of challenges. The realism of the resolved gravity-wave spectrum in the IFS and its effect on data assimilation needs further study. The realism of the solar (thermal) tide should also be investigated.*

*The large variability of stratospheric ozone, which is radiatively very important, implies that a prognostic ozone field is needed to minimize model biases. The first step is to simply model ozone, without assimilation. Even quite simplified schemes could be effective. Areas of focus could be the tropical lower stratosphere, and the polar vortex. Assimilation of ozone is primarily needed to control model biases, so ozone assimilation schemes must be designed with that as their primary focus.*

## 9 Summary of Conclusions and Recommendations

**The vertical profile of global-mean temperature is a key model diagnostic in the stratosphere.** A number of improvements to the radiation scheme and the treatment of ozone were made during 2016 and 2017, and have the capability to eliminate most of the global-mean temperature bias in the stratosphere in the IFS at TL255 resolution, which was quite substantial. Some of these changes have been migrated to the current operational cycle (43R3). Time series of global-mean temperature in the ERA5 reanalysis exhibit a number of problematical features in the stratosphere. **Before the next reanalysis, a minimum requirement for the model must be an essentially unbiased stratospheric global-mean temperature. Further attention should be paid to remaining global-mean temperature biases around the stratopause region.**

**The global-mean cold bias in the lower stratosphere (between 100 hPa and 50 hPa) was found to get worse as horizontal resolution increases.** Such a sensitivity is surprising, but can be understood if the energy deposition from upward-propagating gravity waves is a significant contributor to the thermodynamic budget at these altitudes. The problem does get better as vertical resolution increases, and 200 m vertical resolution in this region (via L198) seems to be enough to eliminate the difference in bias between TCo199 and TCo1279. Already 250 m (via L162) considerably improves the problem, and would avoid having to retune the physics in the troposphere. **Further investigation of this issue is warranted.** More generally, it would be timely to revisit the classic study of Lindzen and Fox-Rabinovitz (1989) concerning vertical resolution requirements, in view of the latest horizontal resolutions affordable with the IFS.

**Even at the relatively high resolution (in a climate-modelling context) of TL255, parameterized gravity-wave drag is an important driver of meridional circulation and polar vortex variability in the IFS.** It is less critical for lower stratospheric tropical upwelling because of the compensation between resolved and parameterized drag in this region. NOGWD dominates the SH, and both orographic GWD and NOGWD are important for the NH. As there are no direct observational constraints on GWD, **the parameterizations need to be tuned to obtain realistic polar vortex variability and the associated stratosphere-troposphere dynamical coupling.** The most important aspects for predictability are the seasonal evolution and timing of the annual vortex breakdown in the SH, and SSWs in the NH.

**Nudging is a useful way to obtain robust results from short simulations** of the recovery phase of SSWs, which affects stratosphere-troposphere coupling. Nudging could also be a useful way of quantifying the effect of radiative changes in high-latitude regions, **where upper-stratosphere temperature biases of up to 10 K remain. Whilst the NOGWD settings in the IFS (for cycle 43R1, at TL255L137) appear to be optimal for polar vortex variability, they need to be monitored closely as the model evolves or is used in other configurations.**

**SEAS5 seems to lack the SH springtime stratosphere-troposphere coupling and associated predictability that was present in SEAS4, presumably because of an unrealistic seasonal evolution of the annual vortex breakdown in SEAS5.**

**The IFS exhibits a notable cold bias of up to 5 K just above the extratropical tropopause,** which maximizes at high latitudes in the summer season. This is a robust feature in models, and has been present in the IFS for a very long time. **All evidence points to the cause being too much moisture leaking in to the region from the upper troposphere.** The surprising thing is that the problem is not improved by increased spatial resolution. This is consistent with the view that there is a complex

mesoscale spectrum of moist processes and unbalanced motion around the tropopause, which becomes more active in the model as resolution increases, but is not well resolved. **Diagnostics targeting such processes would be useful and the sensitivity of cross-tropopause water vapour transport to numerics should be explored.** High vertical-resolution water vapour measurements should be used for model evaluation, and are available from aircraft campaigns as well as from the ACE-FTS limb sounder. **The possibility of assimilating sparse vertically-resolved stratospheric water vapour measurements should be explored.**

**The sponge layer in the IFS seems to have evolved in a very ad hoc way,** presumably to control particular problems through damping. But unless damping is designed in a physically appropriate way, it can lead to other problems. **To avoid the situation of compensating errors, it is essential that the sponge in the IFS have minimal adverse impact on the atmospheric state in the domain of interest,** i.e. below about 0.3 hPa (60 km). However, since unjustified features of the sponge are usually there for a reason, they cannot simply be removed, but must somehow be replaced by something more physical. **Some progress was made by the Task Force in this respect, but the effort is unfinished and needs to continue.**

**Tropical zonal winds in the stratosphere are dominated by the SAO in the stratopause region and by the QBO in the rest of the stratosphere.** Both phenomena are driven by drag from a combination of low-frequency equatorial waves and inertia-gravity waves, neither of which can be expected to be accurately represented in a model because of their strong sensitivity to parameterized processes and to numerics. **The QBO is important for seasonal prediction and needs to be tuned through the NOGWD scheme.** The SAO is probably not so important, so the philosophy there should be to ensure that it does not cause detrimental effects. **Some attention should be paid to the vertical propagation of wind increments through DA in the tropics, where there can be nothing to limit model error.**

**The large amplitude of gravity waves in the stratosphere, both in observations and in models, presents a variety of challenges. The realism of the resolved gravity-wave spectrum in the IFS and its effect on data assimilation needs further study.** The realism of the solar (thermal) tide should also be investigated.

**The large variability of stratospheric ozone, which is radiatively very important, implies that a prognostic ozone field is needed to minimize model biases. The first step is to simply model ozone, without assimilation.** Even quite simplified schemes could be effective. Areas of focus could be the tropical lower stratosphere, and the polar vortex. **Assimilation of ozone is primarily needed to control model biases, so ozone assimilation schemes must be designed with that as their primary focus.**

*Acknowledgements: Input was gratefully received from Hans Hersbach, Sylvie Malardel and Beatriz Monge-Sanz.*

## Appendix: Potential observational products for validation

ACE-FTS limb sounding measurements from solar occultation, from 2003 continuing to present-day: high precision, good vertical resolution. Useful for tracer-tracer correlations of long-lived species (including water vapour) and vertical profiles in tropopause-based coordinates (Hegglin et al. 2008, 2009).

SPARC Data Initiative monthly zonal mean annually resolved climatologies of trace gases and aerosol from stratospheric limb sounders: <http://www.sparc-climate.org/data-centre/data-access/sparc-data-initiative/>

SPARC Reanalysis Intercomparison Project publications comparing different reanalyses in the stratosphere: <https://s-rip.ees.hokudai.ac.jp/pubs/intercomp.html>

IGAC/SPARC data from research aircraft: <https://esrl.noaa.gov/csd/globalmodelevel/>

IAGOS data from in-service aircraft: <http://iagos.sedoo.fr/>

12-hour radiosonde temperature differences are a good metric for gravity wave amplitudes.

## References

- Abalos, M., Legras, B., Ploeger, F. and Randel, W.J., 2015. Evaluating the advective Brewer-Dobson circulation in three reanalyses for the period 1979–2012. *J. Geophys. Res.*, **120**, 7534–7554.
- Anstey, J.A. and Shepherd, T.G., 2014. High-latitude influence of the Quasi-Biennial Oscillation. *Quart. J. Roy. Meteor. Soc.*, **140**, 1–21.
- Bacmeister, J., Eckermann, S., Newman, P., Lait, L., Chan, K., Loewenstein, M., Proffitt, M. and Gary, B., 1996. Stratospheric horizontal wavenumber spectra of winds, potential temperature, and atmospheric tracers observed by high-altitude aircraft. *J. Geophys. Res.*, **101**, 9441–9470.
- Baldwin, M.P. and Dunkerton, T.J., 2001. Stratospheric harbingers of anomalous weather regimes. *Science*, **294**, 581–584.
- Blackburn, M., 1997. Advection of water vapour and the cold polar tropopause bias in Eulerian GCM's. WGNE Blue Book.
- Brewer, A.W., 1949. Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere. *Quart. J. Roy. Meteor. Soc.*, **75**, 351–363.
- Butchart, N., et al., 2011. Multimodel climate and variability of the stratosphere. *J. Geophys. Res.*, **116**, D05102, 10.1029/2010JD014995.
- Byrne, N.J., Shepherd, T.G., Woollings, T. and Plumb, R.A., 2017. Non-stationarity in Southern Hemisphere climate variability associated with the seasonal breakdown of the stratospheric polar vortex. *J. Clim.*, **30**, 7125–7139.

- Byrne, N.J. and Shepherd, T.G., 2018. Seasonal persistence of circulation anomalies in the Southern Hemisphere stratosphere, and its implications for the troposphere. *J. Clim.*, **31**, 3467–3483.
- Dee, D. and Uppala, S., 2008. Variational bias correction in ERA-Interim. ECMWF Technical Memorandum No. 575.
- Dyroff, C., Zahn, A., Christner, E., Forbes, R., Tompkins, A.M. and van Velthoven, P.F.J., 2015. Comparison of ECMWF analysis and forecast humidity data with CARIBIC upper troposphere and lower stratosphere observations, *Quart. J. Roy. Meteor. Soc.*, **141**, 833–844.
- Fomichev, V.I., Ward, W.E., Beagley, S.R., McLandress, C., McConnell, J.C., McFarlane, N.A. and Shepherd, T.G., 2002. The extended Canadian Middle Atmosphere Model: Zonal-mean climatology and physical parameterizations. *J. Geophys. Res.*, **107**, 4087, 10.1029/2001JD000479.
- Fueglistaler, S., Haynes, P.H. and Forster, P.M., 2011. The annual cycle in lower stratospheric temperatures revisited. *Atmos. Chem. Phys.*, **11**, 3701–3711.
- Haynes, P.H., Marks, C.J., McIntyre, M.E., Shepherd, T.G. and Shine, K.P., 1991. On the “downward control” of extratropical diabatic circulations by eddy-induced mean zonal forces. *J. Atmos. Sci.*, **48**, 651–678.
- Hegglin, M.I., Boone, C.D., Manney, G.L., Shepherd, T.G., Walker, K.A., Bernath, P.F., Daffer, W.H., Hoor, P. and Schiller, C., 2008. Validation of ACE-FTS satellite data in the upper troposphere/lower stratosphere (UTLS) using non-coincident measurements. *Atmos. Chem. Phys.*, **8**, 1483–1499.
- Hegglin, M.I., Boone, C.D., Manney, G.L. and Walker, K.A., 2009. A global view of the extratropical tropopause transition layer from Atmospheric Chemistry Experiment Fourier Transform Spectrometer O<sub>3</sub>, H<sub>2</sub>O, and CO. *J. Geophys. Res.*, **114**, D00B11, 10.1029/2008JD009984.
- Hitchcock, P., Shepherd, T.G. and Yoden, S., 2010. On the approximation of local and linear radiative damping in the middle atmosphere. *J. Atmos. Sci.*, **67**, 2070–2085.
- Hogan, R., et al. 2017. Radiation in numerical weather prediction. ECMWF Technical Memorandum No. 816.
- Lindzen, R.S. and Fox-Rabinovitz, M., 1989. Consistent horizontal and vertical resolution. *Mon. Wea. Rev.*, **117**, 2575–2583.
- Linz, M., Plumb, R.A., Gerber, E.P., Haenel, F.J., Stiller, G., Kinnison, D.E., Ming, A. and Neu, J.L., 2017. The strength of the meridional overturning circulation of the stratosphere. *Nature Geosci.*, **10**, 663–667.
- McLandress, C., 2002. The seasonal variation of the propagating diurnal tide in the mesosphere and lower thermosphere. Part II: The role of tidal heating and zonal mean winds. *J. Atmos. Sci.*, **59**, 907–922.
- McLandress, C., Shepherd, T.G., Polavarapu, S. and Beagley, S.R., 2012. Is missing orographic gravity wave drag near 60°S the cause of the stratospheric zonal wind biases in chemistry-climate models? *J. Atmos. Sci.*, **69**, 802–818.



- McLandress, C., Plummer, D.A. and Shepherd, T.G., 2014a. Technical Note: A simple procedure for removing temporal discontinuities in ERA-Interim upper stratospheric temperatures for use in nudged chemistry-climate model simulations. *Atmos. Chem. Phys.*, **14**, 1547–1555.
- McLandress, C., Shepherd, T.G., Reader, M.C., Plummer, D.A. and Shine, K.P., 2014b. The climate impact of past changes in halocarbons and CO<sub>2</sub> in the tropical UTLS region. *J. Clim.*, **27**, 8646–8660.
- McLandress, C., Shepherd, T.G., Jonsson, A.I., von Clarmann, T. and Funke, B., 2015. A method for merging nadir-sounding climate records, with an application to the global-mean stratospheric temperature data sets from SSU and AMSU. *Atmos. Chem. Phys.*, **15**, 9271–9284.
- Miyazaki, K., Iwasaki, T., Kawatani, Y., Kobayashi, C., Sugawara, S. and Hegglin, M.I., 2016. Inter-comparison of stratospheric mean-meridional circulation and eddy mixing among six reanalysis data sets. *Atmos. Chem. Phys.*, **16**, 6131–6152.
- Nash, J. and Saunders, R., 2015. A review of Stratospheric Sounding Unit radiance observations for climate trends and reanalyses. *Quart. J. Roy. Meteor. Soc.*, **141**, 2103–2113.
- Polavarapu, S., Shepherd, T.G., Rochon, Y. and Ren, S., 2005. Some challenges of middle atmosphere data assimilation. *Quart. J. Roy. Meteor. Soc.*, **131**, 3513–3527.
- Polichtchouk, I., et al., 2017: What influences the middle atmosphere circulation in the IFS? ECMWF Technical Memorandum No. 809.
- Polichtchouk, I., Shepherd, T.G., Hogan, R.J. and Bechtold, P., 2018. Sensitivity of the Brewer-Dobson circulation and polar vortex variability to parameterized nonorographic gravity wave drag in a high-resolution atmospheric model. *J. Atmos. Sci.*, **75**, 1525–1543.
- Sankey, D., Ren, S., Polavarapu, S., Rochon, Y.J., Nezlin, Y. and Beagley, S., 2007. Impact of data assimilation filtering methods on the mesosphere. *J. Geophys. Res.*, **112**, 10.1029/2007JD008885.
- Scott, R.K., and P.H. Haynes, 1998. Internal interannual variability of the extratropical stratospheric circulation: The low-latitude flywheel. *Quart. J. Roy. Meteor. Soc.*, **124**, 2149–2173.
- Seviour, W.J.M., Hardiman, S.C., Gray, L.J., Butchart, N., Maclachlan, C. and Scaife, A.A., 2014. Skillful seasonal prediction of the Southern Annular Mode and Antarctic ozone. *J. Clim.*, **27**, 7462–7474.
- Shaw, T.A., Sigmond, M., Shepherd, T.G. and Scinocca, J.F., 2009. Sensitivity of simulated climate to conservation of momentum in gravity wave drag parameterization. *J. Clim.*, **22**, 2726–2742.
- Shepherd, T.G., 2000. The middle atmosphere. *J. Atmos. Solar-Terres. Phys.*, **62**, 1587–1601.
- Shepherd, T.G., Koshyk, J.N. and Ngan, K., 2000. On the nature of large-scale mixing in the stratosphere and mesosphere. *J. Geophys. Res.*, **105**, 12433–12446.
- Shepherd, T.G. and Shaw, T.A., 2004. The angular momentum constraint on climate sensitivity and downward influence in the middle atmosphere. *J. Atmos. Sci.*, **61**, 2899–2908.

Sigmond, M. and Shepherd, T.G., 2014. Compensation between resolved wave driving and parameterized orographic gravity-wave driving of the Brewer-Dobson circulation and its response to climate change. *J. Clim.*, **27**, 5601–5610.

Thompson D.W.J., Baldwin, M.P. and Wallace, J.M., 2002. Stratospheric connection to northern hemisphere weather: implications for prediction. *J. Clim.*, **15**, 1421–1428.