Technical Memo

ECMWF
European Centre for Medium-Range Weather Forecasts

877

Stratospheric modelling and assimilation

I. Polichtchouk, P. Bechtold, M. Bonavita, R. Forbes, S. Healy, R. Hogan, P. Laloyaux, M. Rennie, T. Stockdale, N. Wedi, N. Byrne (U. Reading), M. Diamantakis, S. English, J. Flemming, S. Gisinger (DLR), L. Isaksen & F. Vána

(Research & Copernicus departments)

January 2021

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

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

Abstract

Accurate representation of the stratosphere in NWP models is important for accurate analysis/reanalysis products. It is also important for extended-range and seasonal forecasting because stratospheric variability can influence tropospheric weather patterns on these timescales. This paper reviews the current status and recent modelling and assimilation advances with regards to the stratosphere at ECMWF. The future evolution of ECMWF's Earth System model and assimilation are outlined with a particular focus on the upper atmosphere. Examples are also shown of stratosphere-troposphere coupling for extended-range timescales, while discussing selected processes and their simulation that could improve this aspect of longer-range predictive skill.

1 Introduction

Accurate representation of the stratosphere at ECMWF is important for several reasons: i) for getting the correct background information into the data assimilation system, given the deep weighting functions of the operational nadir temperature sounders; ii) for analysis and reanalysis products; and iii) for tropospheric predictability on monthly and seasonal timescales, as a growing body of evidence suggests that stratospheric variability has an impact on tropospheric weather patterns in both the Northern and Southern hemispheres (e.g. Baldwin and Dunkerton, 2001; Douville, 2009; Sigmond et al., 2013; Karpechko, 2015; Domeisen et al., 2015; Butler et al., 2016; Scaife et al., 2016; Byrne et al., 2019; Büeler et al., 2020). The aim of this paper is to review recent advances in, and to set out future directions for, representing the stratosphere and understanding its influence on the troposphere in the ECMWF Integrated Forecast System (IFS).

Given the comparatively low data availability in the stratosphere, a realistic model is an underlying basis not only for prediction but also analysis/reanalysis. Therefore, ECMWF's focus over the last three years has been on improving the physical realism of the model behaviour and reducing model biases. This has largely been achieved as a result of a successful collaboration — funded by Copernicus — with Ted Shepherd's group at the Univeristy of Reading and much of the progress has already been summarized by Hogan et al. (2017); Polichtchouk et al. (2017) and Shepherd et al. (2018). Therefore, we will only review model advances here that followed on from these works and refer the interested reader to the above mentioned ECMWF Technical Memoranda.

The ECMWF analysis and reanalysis temperature in the lower- and mid- stratosphere (up to 30 km altitude) is reasonably well constrained by assimilation of soundings from radiosondes and space-based infrared and microwave radiances. In the upper stratosphere (above 30 km altitude), constraint on temperature is mostly provided by the highest sounding channels of the AMSU-A and ATMS microwave radiometers and GNSS radio occultation (GNSS-RO) data. Above z$\sim$35 km (p$\sim$5 hPa), the strong dependence of temperature analysis/reanalysis on the model poses a challenge for model verification. Therefore, independent temperature retrievals from e.g., Aura microwave limb sounder (MLS) are vital for model and analysis verification. Winds in the stratosphere are indirectly constrained by the radiances via mass-wind balance, especially in the extra-tropics. Up until the recent launch of ESA's Aeolus Doppler wind lidar mission, the direct constraint on winds was only via radiosonde soundings in the lower- to mid-stratosphere. The positive impact of assimilating Aeolus wind observations (Rennie and Isaksen, 2020b,a) on the analysis and forecasts is reviewed in this paper. Moreover, recent availability of extra GNSS-RO observations has also helped to constrain winds in the tropics as shown below. Specific humidity analysis in the stratosphere remains poorly constrained as no observations are assimilated above the tropopause. This has several consequences, which are discussed in this paper.

Extended range (up to 45 days ahead) and seasonal (up to 6 months ahead) forecasting has gained a much

higher profile at ECMWF in recent years. Because the stratosphere potentially plays a role in enhancing the extended range and seasonal forecast skill in the extra-tropical troposphere following stratospheric sudden warming (SSW) events (e.g., Baldwin and Dunkerton, 2001; Sigmond et al., 2013; Hitchcock and Simpson, 2014), strong polar vortex events (e.g., Baldwin and Dunkerton, 2001; Büeler et al., 2020), and polar vortex breakdown events in the spring season (e.g., Black and McDaniel, 2007; Byrne et al., 2019), there is an impetus to accurately represent polar vortex variability in models. Accurate representation of equatorial wind variability is also important due to the influence of the quasi-biennial oscillation (QBO) of the stratospheric equatorial zonal winds on polar vortex variability via the Holton-Tan effect (Holton and Tan, 1980, 1982; Anstey and Shepherd, 2014). Over the last 10 years, the skill in SSW prediction has improved in ECMWF extended range forecasts mostly due to an increase in vertical resolution in conjunction with a higher model top (Vitart et al., 2019). Considerable progress in assessing stratospheric predictability and stratosphere-troposphere coupling in ECMWF's extended range forecasts has been achieved as part of a large international WWRP/WCRP subseasonal to seasonal (S2S) effort (Domeisen et al., 2020a,b; Vitart et al., 2019) with the overall conclusion that ECMWF's forecasts perform competitively in this regard. However, it is difficult to cleanly separate the influence of the stratospheric variability on the extratropical troposphere from other remote events such as the Madden-Julian oscillation (MJO) (e.g. Garfinkel et al., 2014), El Nino Southern Oscillation (ENSO) (see Domeisen et al., 2019, for a review), and sea ice anomalies (e.g. Sun and Tan, 2013). The availability of large ensemble hindcasts at ECMWF provides a unique opportunity to quantify the predictability from the stratosphere and to assess the impact of stratospheric variability in comparison to other remote events via the use of conditional probabilities and causal link algorithms. Examples of this are given below.

This paper is structured as follows: In section 2, a description of IFS is first given followed by a review of recent model developments in the middle atmosphere (i.e., the stratosphere and mesosphere). Section 3 reviews few recent four dimensional variational data assimilation (4D-Var) developments. In section 4, the impact of observing system changes on stratospheric analysis and forecasts is described with a focus on information content from GNSS-RO and Aeolus wind observations. Section 5 presents examples of stratosphere-troposphere coupling and discusses the methodology to better separate the influence of the stratosphere on the troposphere. Section 6 provides an outlook for future developments relevant to the middle atmosphere.

2 Model developments

The troposphere may be simply thought of as representing a turbulent boundary layer for a comparatively isolated stratosphere. To a first approximation, the global-mean stratosphere is in radiative equilibrium, with long-wave cooling balancing solar heating via ozone. This property makes the global-mean temperature an excellent first diagnostic for model evaluation and the radiation and ozone representation as the first-order model processes controlling global-mean temperature biases. On the other hand, the latitudinal temperature structure is affected by the meridional circulation (i.e., the Brewer-Dobson circulation), which in the stratosphere is driven by breaking and dissipating planetary and gravity waves, which are also responsible for stratospheric variability (e.g., strong and weak polar vortex, QBO, SSWs, final warmings) and therefore can influence tropospheric weather patterns. Thus, it is important to evaluate the representation of waves in the model, which at typical ECMWF model resolutions are now largely resolved but some also parametrized. As most of the wave spectrum is resolved, the design of the dynamical core plays an important role in the stratosphere. While the stratosphere is very dry in comparison to the troposphere, water vapour is important because of its effect on the radiation. In the lower stratosphere, radiative time scales are very long (1-2 months). Because of this, any deficiencies in

the model are readily exposed.

Thus, after a brief IFS model description we begin this section by first discussing temperature biases together with radiation and ozone representation developments in 2.2 and 2.3, followed on by the representation of waves (mostly gravity waves) in the model in 2.4. We then proceed to discuss the impact of vertical diffusion parametrization on the equatorial winds in 2.5; and the representation of water vapour in 2.6.

2.1 Current IFS model description

The current (Cy47r1) operational IFS is based on a semi-implicit semi-Lagrangian dynamical core using the spectral transform method in the horizontal and vertical finite element (VFE) with hybrid pressure coordinate in the vertical (IFSdoc, 2019). The top of the model is at 0.01 hPa (∼80 km) and the current vertical resolution is 137 levels (L137) for high resolution (HRES) medium-range forecasts and analysis (TCo1279 spectral truncation, or 9 km on a cubic octahedral "TCo" grid) and 91 levels (L91) for all ensemble forecasts. As discussed elsewhere in the SAC documents, ECMWF plans to increase the vertical resolution to L137 for all its applications in the next operational implementation. The vertical level spacing is stretched with altitude; in the L137 configuration, the vertical resolution is 300 m in the upper troposphere, 400 m at 50 hPa, 1000 m at 5 hPa and 3 km near the model top. This degradation in vertical resolution poses a challenge for accurate representation of stratospheric processes, especially at high horizontal resolution (section 2.2.2). To prevent wave reflection at the model top, upward propagating resolved waves are damped by a numerical wave absorbing layer ("sponge") in the form of a fourth-order hyperdiffusion applied on the vertical component of vorticity, horizontal divergence and temperature from 10 hPa (∼30 km) upwards. An additional strong sponge is applied on horizontal divergence from 1 hPa (∼45 km) upwards.

The radiation scheme "ecRad" (Hogan and Bozzo, 2018), which includes longwave scattering, is called hourly. In Cy46r1 new aerosol and ozone climatologies from The Copernicus Atmosphere Monitoring Service (CAMS) were introduced. Other physical parametrizations acting in the stratosphere include a spectral non-orographic gravity wave drag (NOGWD) parametrization for hydrostatic waves by Scinocca (2003) (see also Warner and McIntyre (2001) and Orr et al. (2010)), an orographic gravity wave drag (OGWD) parametrization by Lott and Miller (1997), a vertical diffusion parametrization relying on first-order turbulence closure, and methane oxidation (section 2.6).

2.2 Understanding and alleviating temperature biases

Over recent years, significant progress has been made in understanding stratospheric temperature (T) biases. A schematic of the currently prevailing zonal-mean T biases in the free-running model is shown in Fig. 1. The real model bias examples can be found in Hogan et al. (2017); Polichtchouk et al. (2017) and Shepherd et al. (2018). To complement the zonal-mean schematic, Fig. 2a shows global-mean T biases from a 4 year ensemble of 1-year forecasts at two different horizontal resolutions, verified against ERA5 climatology below 5 hPa and Aura MLS climatology above 5 hPa. Several clear biases can be identified:

1. A global-mean warm bias at the stratopause, which also emerges when verified against Aura MLS and extends to the model top (Fig. 2a). This bias is discussed in section 2.2.1. In the upper stratosphere, this bias is alleviated at higher horizontal resolution due to the unphysical (numerical) global-mean cooling present in the stratosphere at higher horizontal resolution.

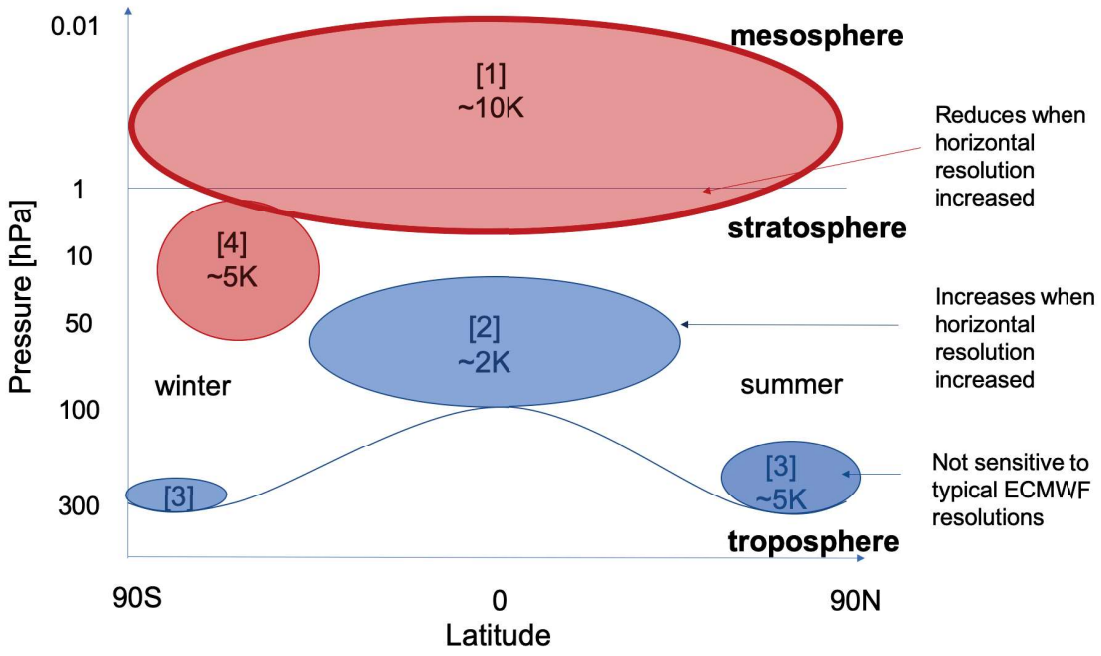

Figure 1: Schematic of zonal-mean IFS temperature biases in the middle atmosphere. Red ovals represent a warm bias and blue ovals a cold bias. The numbers correspond to different biases discussed in the text.

2. A mid- to lower stratosphere cold bias, which maximizes in the tropics. This bias gets worse as horizontal resolution is increased due to unphysical (numerical) global-mean cooling in the model that becomes progressively worse as the horizontal resolution is increased beyond 80 km (cf. red to blue lines in Fig. 2a). This is discussed in section 2.2.2.

3. Cold $\sim$ 5K lowermost stratosphere (LMS) bias over the pole maximizing in the NH summer. This bias is common to many models (Boer et al., 1992) and is most likely due to a moist bias in this region resulting in too strong longwave cooling (Stenke et al. (2008), also see section 7 in Hogan et al. (2017)). Earlier comparisons of ECMWF analysis to coarse vertical resolution Aura MLS (Hogan et al., 2017) and to CARIBIC aircraft observations (Dyroff et al., 2015) suggest a factor of 2-3 too moist analysis bias in this region. Recent collaboration with DLR and KIT has enabled a comparison of ECMWF analysis to very high vertical resolution water vapour data from air-borne infrared limb imager GLORIA, confirming the moist bias in ECMWF analysis (Woiwode et al., 2020). Since no specific humidity observations are assimilated to constrain the analysis in LMS, the moistening occurs because of model errors. This bias is discussed further in section 2.2.3.

4. Warm polar mid- to upper stratosphere bias in the winter hemisphere. At present the reason for this bias is unknown, though it is likely related to deficiencies in the radiation scheme or ozone climatology, or the resolved waves breaking at too low an altitude due to too deep a sponge layer. This bias is briefly discussed in section 2.4.4.

The behaviour of T biases shown in Figs. 1 and 2 carries over to forecasts at all lead times (i.e., for medium-range, extended-range and seasonal forecasts). Therefore, any model improvements discussed here target all ECMWF forecast ranges.

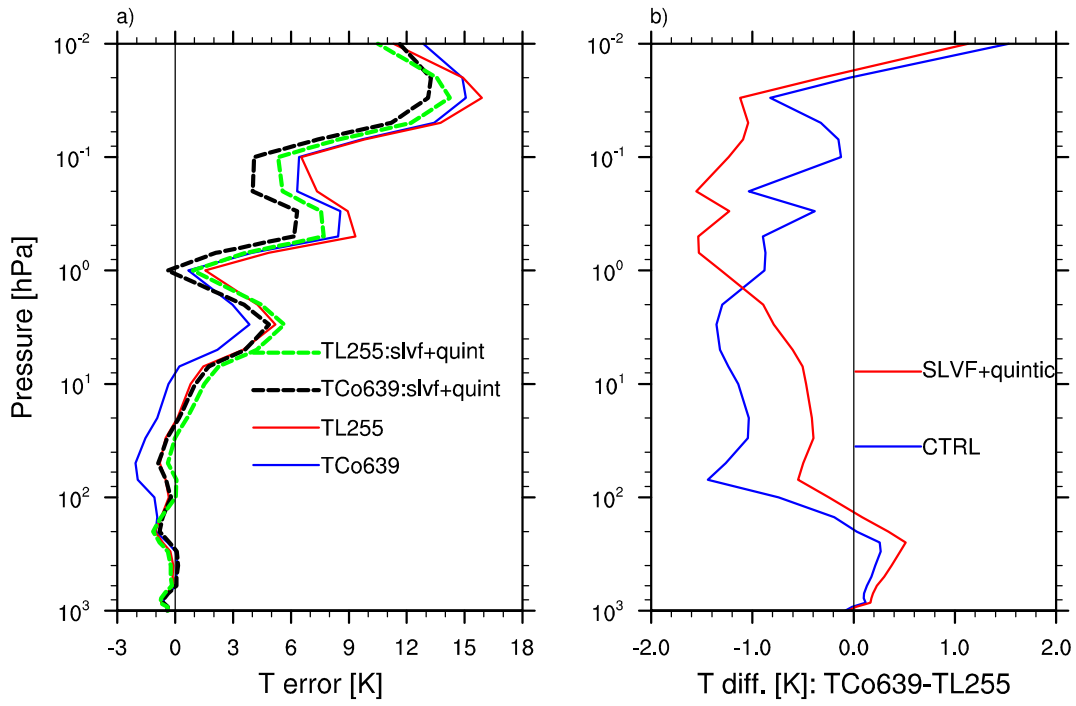

Figure 2: (a) Global-mean annual-mean temperature bias measured against a blend of ERA5 climatology below 5 hPa and Aura MLS climatology above 5 hPa. The bias is shown at TCo639 (blue) and TL255 (red) resolutions for the control Cy46r1 experiments, and at TCo639 (black) and TL255 (green) resolutions for experiments with semi-Lagrangian vertical filter (SLVF) and quintic vertical interpolation (see section 2.2.2 for details). (b) Difference in the global-mean temperature between TCo639 and TL255 resolutions for the control experiments (blue) and SLVF with quintic vertical interpolation experiments (red). Note that the resolution sensitivity of biases does not depend on whether a linear (TL) or a cubic octahedral (TCo) grid is used.

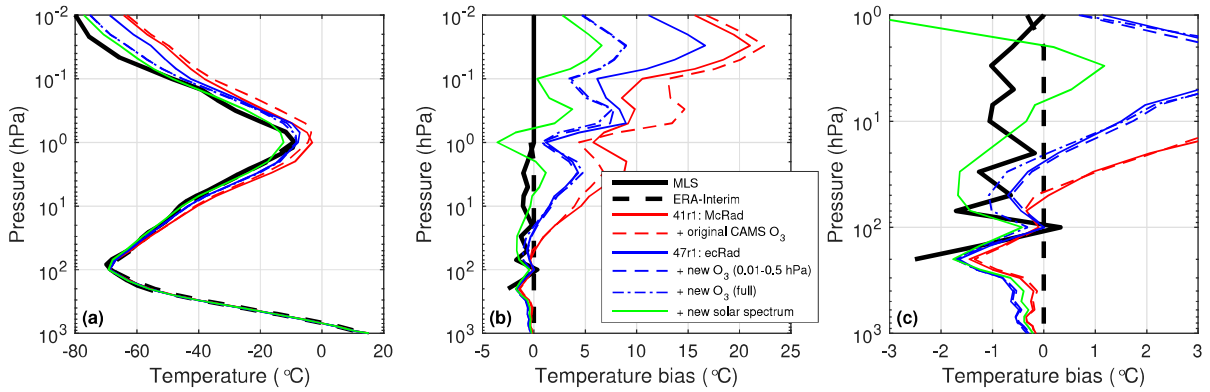

Figure 3: (a) Annual-mean temperature from four free-running 1-year TL255L137 simulations using various versions of the IFS, as well as the Aura MLS and ERA-Interim up to 1 hPa. (b) Bias in the climate simulations versus ERA-Interim in the troposphere and stratosphere, and MLS in the mesosphere. (c) As panel b but zoomed into the stratosphere. The 41r1 simulations are from Hogan et al. (2017) and used prescribed sea-surface temperatures with the radiation scheme called every 3 h, while the 47r1 simulations were coupled to the ocean and called the radiation scheme every 1 h. Note also that quintic vertical interpolation (section 2.2.2) is applied in 47r1 experiments.

2.2.1 *Global-mean temperature bias: Impact of radiation and ozone*

There have been a number of changes to the IFS radiation scheme and ozone climatology in the last five years that have led to improvements in global-mean stratosphere and mesosphere temperature, as summarized in Fig. 3. The red solid line shows a configuration very close to IFS Cy41r1, operational in 2015, which used the older "McRad" radiation scheme and the "MACC" ozone climatology. The radiation scheme was called only every 3 h in this simulation. The red dashed line shows that the warm bias of up to +8.5 K in the upper stratosphere was reduced to +6.5 K when the ozone climatology was updated to use a climatology from CAMS in Cy41r2.

As documented by Hogan et al. (2017), this was followed by two improvements to the treatment of radiative transfer in the stratosphere. Hogan and Hirahara (2016) revealed that the 3-h radiation timestep could explain around 3 K of the upper-stratospheric warm bias, and this could be largely mitigated by improving the way that the sun angle is averaged in time. Now that the radiation scheme is called every 1 h in all operational model configurations, their improvement has less work to do. A further reduction in the bias of around 0.5 K resulted from the introduction of the ecRad radiation scheme (Hogan and Bozzo, 2018) in Cy43r3, due to the longwave two-stream equations being solved exactly. The solid blue line in Fig. 3 corresponds to the current operational cycle and combines these two improvements, with a warm bias of up to +4.5 K in the upper stratosphere but still exceeding +15 K in the middle mesosphere.

There are three limitations with the current ozone climatology: i) It is taken from the CAMS reanalysis when the model had only 60 layers with a top at 0.1 hPa, and the concentrations have been held fixed to the 0.01 hPa top of the current L137 model. ii) The observational constraints on the ozone concentrations are weak throughout the mesosphere. iii) Ozone in the mesosphere is characterized by a pronounced diurnal cycle caused by the fast photo-dissociation of ozone, which reduces the amount of available ozone for the absorption of short-wave radiation. To correct these limitations the current ozone climatology was vertically extended above 0.5 hPa with a modelled day-time ozone climatology constructed from merged results of the CMAM and WACCAM models prepared for CMIP6 (Hegglin et al., 2016). While the correction of the current ozone climatology in the mesosphere with the CMIP6 day-time ozone climatology is a clear improvement (see blue dotted and dash dotted lines in Fig. 3), it does not fully capture the strong

diurnal ozone variability in the mesosphere. Introducing a diagnostic day-night ozone parametrization, which acts on average ozone values seems therefore a worthwhile development goal. Such a diagnostic parametrization could also be used to calculate day-time values from prognostic ozone simulated with the Cariolle scheme or its update the hybrid linear ozone (HLO) scheme. Both approaches are not able to simulate the diurnal cycle of ozone in the mesosphere caused by photo-dissociation. The ultimate aim, however, is to replace ozone climatology with a radiatively interactive HLO scheme, which is currently under the development and is discussed further in section 2.3.

We now review the evaluation of gas optical properties in the IFS. Clear-sky radiative transfer is a challenge because the effects of a very large number of spectral lines need to be represented. Since 2007, the radiation scheme in the IFS has incorporated the "RRTMG" scheme for representing gas absorption in both the longwave and shortwave. Along with almost all state-of-the-art gas optics scheme for global models, it uses the correlated k-distribution (CKD) technique to represent the effects of many spectral lines. Recently, ECMWF has initiated the CKD Model Intercomparison Project "CKDMIP" (Hogan and Matricardi, 2020), with over 10 international participating groups, to evaluate existing CKD models against reference line-by-line calculations, and to aid in the development of new gas optics schemes.

Figure 4 presents an evaluation of the longwave and shortwave heating rates predicted by the RRTMG gas optics scheme in ecRad using the 50 present-day evaluation profiles from CKDMIP, which span a wide range of temperature, water vapour and ozone concentrations. In the longwave, the stratospheric heating-rate bias exhibits "wiggles" of either sign, with an amplitude around 0.2 K d^{-1} in the upper stratosphere and lower mesosphere. The blue line in Fig. 4d shows the much larger shortwave bias in the version of RRTMG used in the IFS, with a systematic heating-rate overestimate peaking at 0.9 K d^{-1} at 2 hPa. The red line shows the result of correcting the fluxes in each band to use the more accurate solar spectrum of Coddington et al. (2016); the heating rate is reduced by around 0.5 K d^{-1} in the upper stratosphere and lower mesosphere, but still leaving a positive bias of up to 0.4 K d^{-1} at 2 hPa.

The green line in Fig. 3 shows the impact on the stratospheric climate of the IFS of scaling the fluxes in each shortwave band to match the improved solar spectrum. The warm bias in the upper stratosphere and lower mesosphere is substantially reduced, but unfortunately we see that it also exacerbates the cold bias 2) (see Fig. 1), where the bias is also sensitive to the horizontal resolution of the model (see ahead to section 2.2.2). This degradation, likely to be affected by other sources of bias in the model, is why we have not yet been able to make the improved solar spectrum operational in the IFS.

In a parallel activity to CKDMIP, we are developing an ECMWF tool "ecCKD" to generate new CKD gas optics models, and the results so far show that models can be generated with significantly lower heating-rate biases in both the longwave and shortwave, and also using significantly fewer spectral intervals (and hence lower computational cost). Since we will be using the improved solar spectrum, it seems likely that the cold bias 2) would still be present (but see the next section for solutions to this).

2.2.2 *Mid- to lower-stratosphere cold bias: Resolution sensitivity*

Bias 2) in Fig. 1 arises due to the global-mean cooling at higher horizontal resolution (cf. red and blue lines in Fig. 2). Such horizontal resolution sensitivity is undesirable for model development and for 4D-Var where each minimization loop is performed at a different horizontal resolution. The global-mean cooling arises because of vertical discretization errors (Polichtchouk et al., 2019). As the horizontal resolution increases, smaller-scale gravity waves are resolved in the horizontal direction. Some of these waves have vertical wavelengths that can not be resolved with the existing vertical resolution, especially in the stratosphere where the vertical resolution decreases and stratification increases. These short

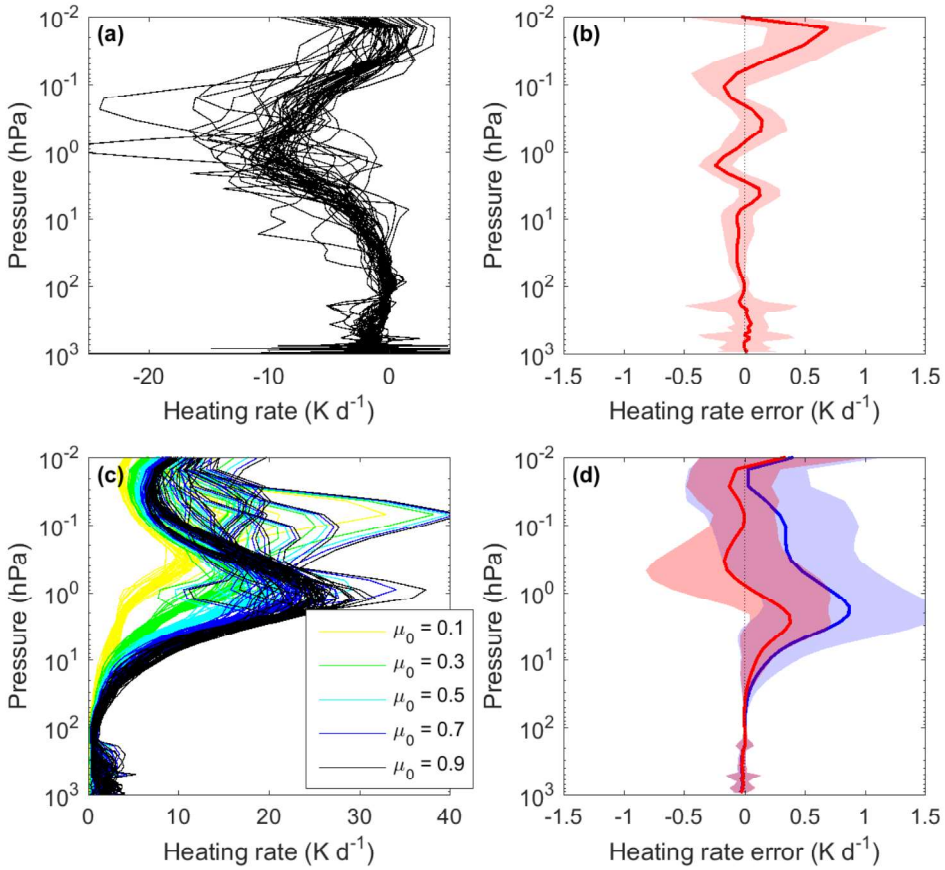

Figure 4: Evaluation of clear-sky heating rates predicted by the RRTMG gas optics scheme used by the ecRad radiation code in the IFS. (a) Reference line-by-line calculations of longwave heating rate for the 50 CKDMIP evaluation profiles (Hogan and Matricardi, 2020), and (b) bias in RRTMG using the same 8-stream radiative transfer solver, with the shaded area encompasing 95% of the data. (c) As panel a but for the shortwave at five values of the cosine of the solar zenith angle, μ_0. (d) As panel b but evaluating the RRTMG shortwave scheme averaging over the full diurnal cycle, where the blue line uses the original solar spectrum as in the IFS, and the red line corrects the fluxes in each band to use the more recent solar spectrum of Coddington et al. (2016).

vertical wavelength waves alias onto the $2\Delta z$ grid-scale noise in the T field, supported by the vertical finite element (VFE) scheme in the IFS especially at high horizontal resolution. This unphysical $2\Delta z$ noise can result in spurious cooling/heating (Hollingsworth, 1995). From a historic perspective, the VFE scheme — introduced into the hydrostatic IFS in 2002 — has greatly damped the unphysical $2\Delta z$ T-noise and therefore reduced the global-mean cooling in the IFS when compared to the previously used finite-difference scheme with Lorenz staggering (Untch and Hortal, 2004) [1].

A simple solution is to increase the vertical resolution when horizontal resolution is increased. In partic-

[1] It is worth mentioning that a successful collaboration in 2019 with Jozef Vivoda from SHMI led to an extension of VFE to the nonhydrostatic dynamical core, which previously only supported finite-difference discretization with Lorenz staggering and thus suffered from more severe unphysical cooling in the stratosphere. This is an important step, enabling more meaningful future comparisons between the hydrostatic and non-hydrostatic options in the IFS. Note that the one way to completely eliminate the unphysical $2\Delta z$ noise is to use Charney-Phillips staggering with finite difference vertical discretization (Charney and Phillips, 1953)(see also Untch and Hortal, 2004; Hollingsworth, 1995) such as done at e.g., the UK Met Office. However, this has the drawback of complicating the semi-Lagrangian advection and the semi-implicit solution procedure in IFS, the coupling to the physics, and potentially impacting an efficient parallel implementation of the model.

ular, it is found that 200 m vertical resolution in the stratosphere is enough to eliminate the unphysical global-mean cooling completely at high horizontal resolution (up to TCo1279). The impact of increased vertical resolution on the zonal-mean T in medium-range HRES forecasts is shown in Fig. 5a: The higher vertical resolution warms the stratosphere and therefore reduces bias 2). Given this, the planned vertical resolution upgrade for ensembles (ENS) from L91 to L137 (made possible by performing forecasts in single precision from Cy47r2) not only unifies the vertical resolution across different ECMWF applications but also improves forecast skill in the lower stratosphere and troposphere (see Figs. 28 and 29 in Vitart et al., 2019).

Increasing vertical resolution further is computationally expensive and experiments with higher vertical resolution than L137 – while reducing the global-mean cooling in the stratosphere – did not show increased tropospheric skill in either medium-range or extended-range forecasts, implying that this bias is not directly affecting tropospheric predictability. This perhaps is not surprising as the elimination of the bias does not significantly alter meridional temperature gradients and therefore polar vortex variability [2]. Therefore, cheaper alternatives were explored. It was found that increasing the order of vertical semi-Lagrangian interpolation from cubic to quintic for the T field greatly alleviates the global-mean cooling in the stratosphere on medium- and extended-range timescales (see Fig. 5b for medium-range results). Therefore, quintic vertical interpolation for T and closely thermodynamically related specific humidity field is now implemented operationally (from Cy47r1).

Another solution to the global-mean cooling is to filter the unphysical $2\Delta z$ T-noise by a Laplacian vertical filter during the semi-Lagrangian interpolation (SLVF, described by Váňa et al., 2008). The SLVF filter can be though of as a vertical analog of the hyperdiffusion. Application of SLVF ensures that the global-mean cooling can be further eliminated, as shown in Fig. 2b. The impact of the SLVF on the zonal-mean T in medium range HRES forecasts is shown in Fig. 5c. The use of SLVF for the potential future operational implementation is currently being explored. The SLVF should reduce the cold bias 2) further and hopefully enable the use of the improved solar spectrum (section 2.2.1).

Note that the vertical resolution increases, quintic vertical interpolation and SLVF all have a larger impact at high horizontal resolution and therefore lead to a reduction of horizontal resolution sensitivity of T biases (Polichtchouk et al., 2019). As the vertical resolution increase, quintic vertical interpolation and SLVF all warm the stratosphere, the forecast skill is improved in the lower to mid-stratosphere where there is a cold bias 2), and degraded in the uppermost stratosphere where there is a warm bias 1) (for skill score figures see Polichtchouk et al., 2019, 2020)[3].

2.2.3 Cold polar lowermost stratosphere bias

It is found that increasing vertical resolution beyond L137 (or 300-400 m in UTLS) or horizontal resolution beyond TCo319 (or 36 km) does not significantly improve the LMS cold bias 3) in Fig. 1 (not shown, but see Fig. 5). In particular, numerical diffusion in semi-Lagrangian dynamics can lead to too strong water vapour leakage across the sharp gradient between the very dry stratosphere and a very moist troposphere; As model resolution increases (both horizontal and vertical), the numerical diffusion decreases as the sharp gradients are better resolved. The insensitivity of this bias to increases of horizontal

[2] While no benefit of increased vertical resolution beyond L137 has yet been identified for tropospheric skill on medium- and extended-range timescales, vertical resolution plays an important role in representation of the QBO in the model. It is therefore possible that vertical resolution increase beyond L137 improves stratosphere-troposphere coupling in seasonal forecasts. This remains to be evaluated.

[3] The warm bias 1) is also affected by the sponge layer: If the sponge layer depth is reduced or the damping weakened, more $2\Delta z$ noise is present that leads to stronger cooling and hence alleviation of the warm bias.

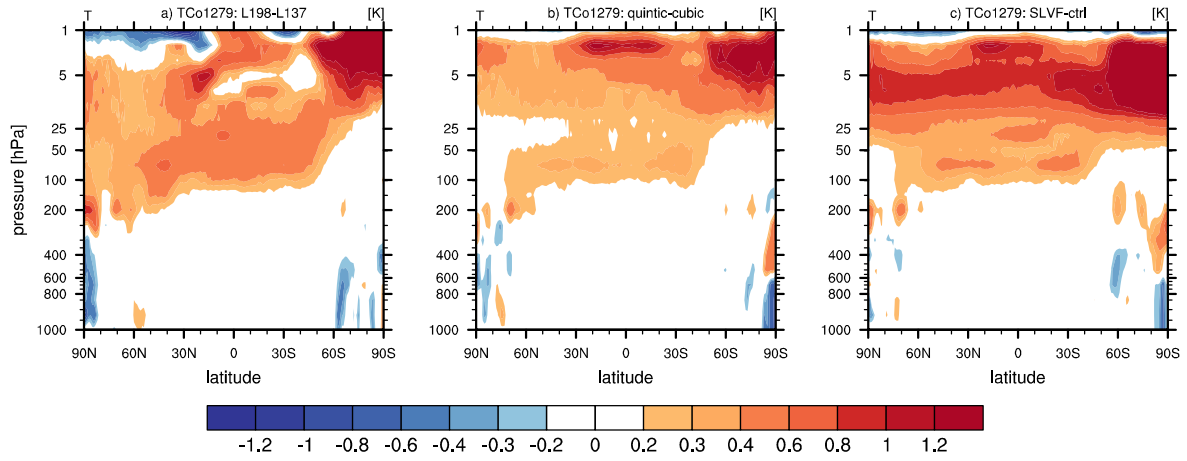

Figure 5: Temperature response at TCo1279 horizontal resolution to (a) increasing vertical resolution from L137 to L198; (b) applying quintic vertical interpolation on temperature and specific humidity at L137 vertical resolution; and (c) applying semi-Lagrangian vertical grid-point filter (SLVF) at L137 vertical resolution. Results are shown for forecasts at a lead time of 10 days for July 2016. The mean over 31 forecasts is shown.

resolution beyond TCo319 and vertical resolution beyond 300-400 m, however, suggests that at these resolutions the hydropause gradients are well resolved.

It is possible that the dynamics around the tropopause may involve a complex mesoscale spectrum of moist processes (e.g. moist conveyor belts, overshooting convection) and unbalanced motion that might not converge until much higher spatial resolution. However, as part of the Horizon 2020 INCITE project, a 1.4 km horizontal resolution four-month free-running IFS simulation was performed covering winter NDJF 2018/2019. This simulation also shows a persistent cold bias in this region despite well-resolved mesoscale motions. It is therefore possible that other model processes contribute to the moist bias and the concomitant cold bias at high resolution. An investigation is underway to characterize the moisture bias in the extratropical LMS in more detail and analyse short forecast tendencies to pinpoint the model processes that result in moistening of the LMS polar region. The moisture bias is also present in the analysis and persists into the forecast. Thus improving the analysis of moisture in the LMS is an area for future improvement. It should be noted that forecasts initialized from a moist-bias corrected analysis maintain cold-bias free LMS initially, with the moist and cold bias starting to develop after $\sim$5 days.

For horizontal resolutions lower than TCo319 it is found that switching off the default 2D quasi-monotone limiter — applied in the IFS to ensure positivity and "shape preservation" of cubic interpolation in the semi-Lagrangian scheme — reduced the cold bias by reducing moisture leakage across the hydropause. The limiter ensures that the new interpolated value at the departure point is no lower or higher than the neighbouring grid points in the horizontal to avoid overshoots or undershoots and is therefore a desirable feature. However, it is probably too restrictive and a 3D version of the limiter, that also checks neighbouring grid points above and below, is less restrictive and also reduces the cold bias, although to a smaller extent. The 3D version of the quasi-monotone limiter will be a candidate for an upcoming IFS cycle. For both limiters, the main effect is from the departure point interpolation in the horizontal, rather than in the vertical.

Finally it is important to emphasize that different from the global-mean cold bias in the stratosphere, there is a strong incentive to eliminate the LMS cold bias: Targeted IFS experiments where water vapour seen by the radiation scheme was reduced around the extratropical tropopause eliminated the cold bias

and improved forecast scores in the troposphere (see Figs. 27 and 28 in Hogan et al., 2017).

2.3 Towards a prognostic ozone scheme

The development of the prognostic hybrid linear ozone scheme (HLO) is now briefly discussed. The HLO scheme uses broadly the same linearization approach as the existing Cariolle scheme to represent the impact of chemistry on ozone tendencies:

$$\frac{\partial O_3}{\partial t}|_{chem} = a + b(O_3 - \overline{O_3}) + c(T - \overline{T}) + d(TCO - \overline{TCO}) \quad (1)$$

In the case of Cariolle, all coefficients are derived from a 2D chemistry model. The hybrid approach instead uses a blend of inputs taken directly from analyses (climatological mean ozone and overhead column ozone from CAMS and temperature from ERA5) or adapted from chemistry model calculations (the rate sensitivity estimates b for ozone, c for temperature and d for overhead column ozone, with internal consistency adjustments, important in regions where heterogenous chemistry is active). The key term in the scheme is the climatological mean production rate a, which is diagnosed from a multi-year IFS run nudging the ozone field towards CAMS analyses. This scheme enables the IFS to reproduce an ozone field which is consistent with the CAMS reanalyses as regards to the mean climatology, and in general does a very good job with interannual and synoptic variability. The major limitation of the scheme is that it can only reproduce the climatological effect of heterogenous chemistry. This is not usually a problem, since most interannual ozone variability is driven by dynamics and is captured well. However, in some circumstances (as occurred in the Arctic spring in 1997, 2011 and 2020) ozone destruction is missed. Accurate modeling in such situations would need a more complex model, at a minimum incorporating a prognostic cold tracer, but quite possibly needing full ozone chemistry. Nonetheless, the overall accuracy of the ozone fields produced by the IFS is greatly improved in all regions, including the Arctic and Antarctic (not shown).

In CAMS, testing of the HLO scheme was sufficiently positive that it was decided to switch to using it operationally from Cy47r1, which shows reduced biases because of the change, especially in the 10-20 hPa region. CAMS testing also confirmed that using HLO in free running model simulation leads to generally greatly reduced biases in the stratosphere. CAMS data assimilation uses climatological (fixed) ozone background errors with no flow dependency, and these were not changed for HLO nor was there any investigation of the possible impact of ozone radiative feedback on NWP scores. Also, ozone radiance assimilation in the infra-red, which could lead to more feedbacks with NWP, is not applied in CAMS (but it is in the NWP operational system). CAMS also has a separate chemistry scheme in the troposphere (CB05), so HLO is only used in the stratosphere, while in NWP operations it is envisaged that HLO will be used throughout the atmosphere. The data assimilation systems for NWP is thus sufficiently different from CAMS that separate testing of the HLO scheme is required, which is now taking place for possible operational use. A key question is whether the radiation calculations can be made using the model ozone values instead of the specified ozone climatology. Experiments are still ongoing, but initial results suggest that allowing this interaction between ozone and radiation gives improvements in the medium-range forecasts in terms of circulation and variability in the stratosphere, but at the expense of a slight deterioration in temperature bias in some regions.

2.4 Resolved and parametrized waves

2.4.1 Sponge layer

The sponge layer is needed in the model to avoid unphysical wave reflections from the model top. Non-hydrostatic models typically apply a sponge on the small-scale vertical velocity (Klemp et al., 2008), which is not possible in hydrostatic models. Historically, the sponge layer in the IFS used to be represented by the Rayleigh drag, which was also used as a surrogate for nonorographic gravity wave drag (NOGWD) to control too strong winds near the model top. Hence, it was used to tune the middle atmosphere circulation. When NOGWD parametrization was introduced at ECMWF (Orr et al., 2010)[4], Rayleigh drag could largely be switched off (with the caveat discussed below for L91 configuration) and the sponge replaced by hyperdiffusion (section 2.1).

It is important to note that a sponge layer is a purely numerical device and should be placed above the region of interest. For the stratosphere this region is below z∼65 km. To replace the need for a sponge layer, the model top could be extended higher into the mesosphere: This way resolved wave dissipation naturally occurs below the model top (e.g., Prusa et al., 1996), but this is expensive as many vertical layers would be wasted. As an alternative to the sponge layer, non-reflective boundary conditions have also been tried in the past (Bougeault, 1983) but with limited success in global models.

The current model sponge formulation in IFS is deeper in comparison to other NWP models, such as ICON and Met Office UM (see ahead to Fig. 9; in IFS the sponge starts already at p=10 hPa or z∼30 km) and applies damping on the zonal-mean component (i.e., on the $m = 0$ zonal wavenumber), which i) violates angular momentum conservation; ii) can introduce torques that are unphysical leading to erroneous meridional circulations (Shepherd and Shaw, 2004; Shaw et al., 2009); and iii) is not necessary since vertically propagating waves have no zonal-mean component (Shepherd et al., 2018). Moreover, additional equatorial zonal-mean Rayleigh drag in the sponge layer is applied in the L91 ECMWF configurations. This is necessary to avoid too strong winds near the model top and can be switched off at higher vertical resolution (i.e., L137) when vertically propagating waves are better resolved and can therefore decelerate the winds. Therefore, the upcoming ENS and extended-range upgrade to L137 is also switching off the undesired zonal-mean Rayleigh drag in the model.

Having a sponge layer that is too deep has several consequences: it can i) introduce spurious momentum and heat sources in the upper stratosphere, and ii) necessitate the use of too strong parametrized gravity wave drag near the model top in order to decelerate the polar night jets. Both arise because the resolved waves damped by the sponge are forced to deposit their momentum and energy at or below the sponge layer. Therefore, an effort was made to reduce the depth of the sponge layer and remove the damping on the $m = 0$ mode in the L137 configuration. In practice, it was found that starting the sponge layer at z∼55 km (or 0.4 hPa) and not damping the zonal-mean flow resulted in neutral impact on medium-range and ENS skill in the troposphere. It is also possible to remove the sponge layer altogether and maintain stability in the free-running IFS even at TCo1279 horizontal resolution.

Resolved gravity wave (GW) amplitudes in ECMWF analysis/short forecasts are generally underestimated in the stratosphere when compared to satellite observations (Hoffmann et al., 2017; Schroeder et al., 2009, Corwin Wright, personal communication), long-duration superpressure balloon data (Jewtoukoff et al., 2015), and Rayleigh lidar (Ehard et al., 2018). To what extent the "too deep" sponge layer in IFS contributes to too low GW activity in the stratosphere (above z=30 km) has been quantified via a collaboration with DLR; The unique over a year long continuous dataset from DLR Rayleigh lidar

[4]NOGWD was found to improve the meridional tilt of the polar night jets in the middle atmosphere.

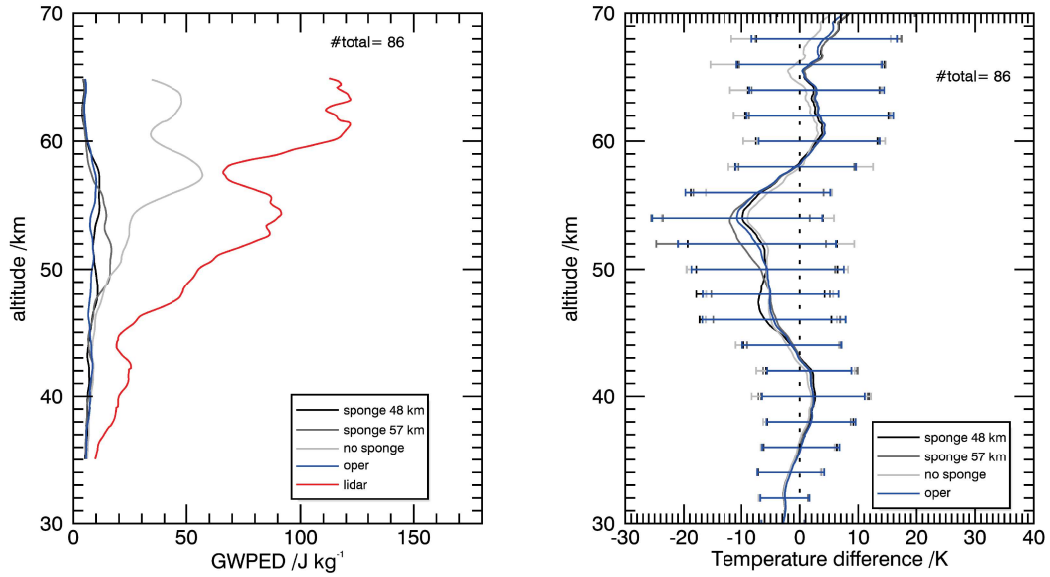

Figure 6: (Left) Comparison of gravity wave potential energy density (GWPED) calculated from temperature perturbations that were determined with a 15 km Butterworth filter (Ehard et al., 2015) and vertically integrated in a 2 km sliding window (Baumgaertner and McDonald, 2007) for lidar and short (0-24h) IFS TCo1279L137 forecasts with different sponge designs at Rio Grande, Argentina for August 2018. Lidar is shown in red; operational IFS sponge design is shown in blue; a shallower sponge starting at z=48 km and z=57 km are shown in dark grey, and grey respectively; and the "no sponge" setup is shown in light grey. (Right) Temperature difference between the lidar and the different sponge configurations. The mean for August is shown and the vertical bars represent +/- one standard deviation.

(Kaifler and Kaifler, 2020) located at Rio Grande, Argentina allows for an assessment of GW momentum flux amplitudes in IFS (Kaifler et al., 2020). The lidar autonomously measures temperature in the $25< z <80$ km altitude region at very high vertical and temporal resolution. GW momentum flux is proportional to potential energy density (GWPED, Ern et al., 2004), which can be retrieved from the lidar measurements. Fig. 6 (left) shows GWPED for August, during the time of strong GW activity in the SH, for lidar data (in red) and short HRES ECMWF forecasts with the default sponge layer (in blue), the new formulation of the sponge layer (in black), and no sponge layer altogether (in light grey). While the IFS underestimates GW amplitudes even with no sponge, the new shallower sponge improves the representation of GW activity in the $45< z <55$ km range. Note also that the new sponge design does not degrade temperature fit of the forecasts to lidar data (Fig. 6 right). The new shallower sponge is currently being tested for operational implementation, but see the caveat discussed in section 3.3 below.

2.4.2 *Resolved gravity waves*

While a shallower sponge will only change the altitude where the GW momentum flux is deposited, it will not change the representation of GWs generated below the sponge (unless spurious wave reflections are present). Representing GWs poses numerous challenges to NWP and the stratosphere in particular due to their high characteristic horizontal phase speeds and horizontal and vertical wave lengths of 10-1000 and 1-10 km, respectively (Ern et al., 2017; Žagar et al., 2017). Furthermore, sources of GWs such as convection or dynamic adjustments in jets and density discontinuities are often under-resolved (Hankinson et al., 2014), GWs are intermittent (Hertzog et al., 2012) and regions of GW breaking are highly turbulent.

Indeed, in the current IFS analysis system the largest first guess temperature departures (i.e., forecast errors) occur over the GW active regions in the stratospheric satellite infrared sounding channels. The errors are particularly pronounced in the SH autumn and winter season east of the Andes between 30-70°S, as shown in Fig. 7a for May 2018. That this region is characteristic of large amplitude GWs can be seen in satellite (AIRS, HIRDLS and SABER) GW climatologies (see e.g., Ern et al., 2018, 2017): Over the southern tip of South America and the Antarctic Peninsula large amplitude orographic GWs are present. Over the South Atlantic at ∼60°S, some of the GWs might have sources in jet-front systems in both the troposphere (Plougonven and Zhang, 2014) and stratosphere (Polichtchouk and Scott, 2020), but a large portion of these GWs is likely to have originated from South America and the Antarctic Peninsula and propagated obliquely large distances downstream (e.g., Dunkerton, 1984; Sato et al., 2012). As shown in Fig. 7b for May 2018, IFS at TCo1279L137 resolution is able to reproduce the characteristic distribution of GWs derived from satellite observations in the lower stratosphere (compare to Figs. 13e and 14e in Ern et al., 2018). In the figure, a high-pass spectral filter retaining total wavenumbers $N >100$ has been applied to the daily T field. Interestingly, the GW active convective regions in the tropics and mid-latitudes in Fig. 7b do not experience large first guess departures. The results clearly suggest that the large first-guess T errors are due to errors in either the amplitude, the phase or the orientation of the high amplitude GWs in the SH. Indeed, the large-scale equatorial wave representation in the IFS is already quite good at TL255 (or 80 km) horizontal resolution. This can be seen in Fig. 8, which shows the wavenumber-frequency spectrum for the outgoing longwave radiation observed by the National Oceanic and Atmospheric Administration (NOAA) polar-orbiting satellites (Wheeler and Kiladis, 1999) (top panels) and from the free-running IFS model (bottom panels).

Apart from comparing GWs in IFS to lidar in Patagonia (section 2.4.1) and satellite climatologies, an ongoing evaluation of GWs in the Southern Andes and Antarctic Peninsula against observations (satellite and super-pressure balloon) and high-resolution (3 km) limited area models (Met Office UM, and WRF)

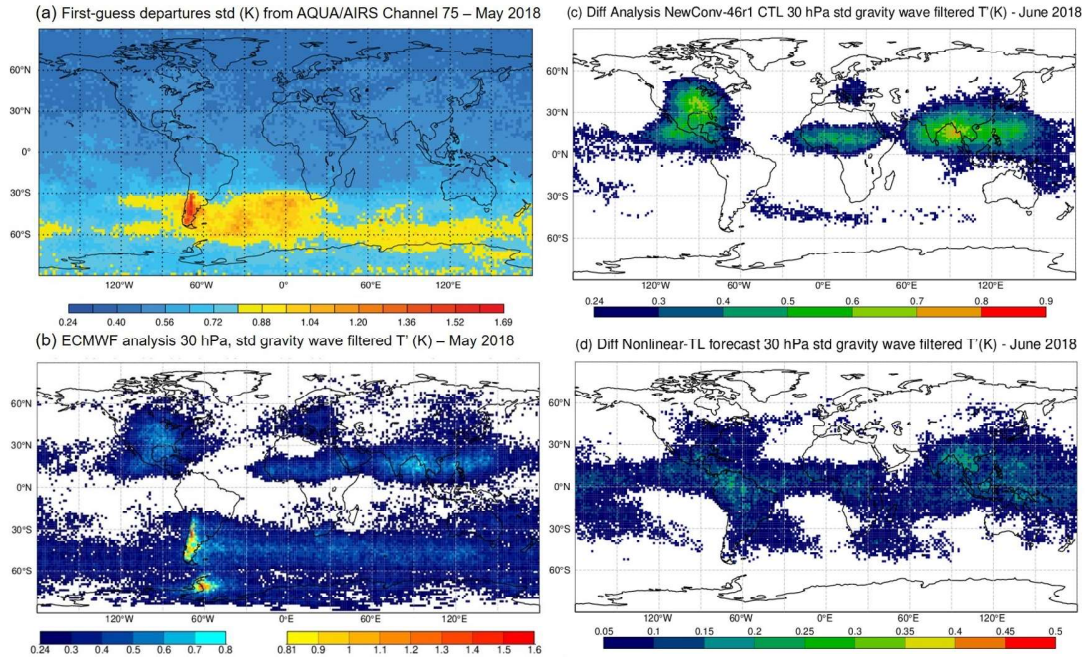

Figure 7: Standard deviations of lower stratospheric temperature departures during May and June 2018 as proxy for gravity wave amplitude: (a) First-guess temperature departures of the IFS short range forecast against the Atmospheric Infrared Sounder (AIRS) channel 75 onboard NASA's AQUA satellite and (b) standard deviation of the gravity wave (total wavenumber > 100) filtered IFS HRES temperature analysis $T'(K)$ at 30 hPa. The impact on the analysis of a more active convection scheme foreseen for implementation in Cy48r1 is shown in (c) and (d) using wavenumber > 60 filtered std of T': (c) differences between the analysis with the new convection and the TCo399 control analysis, (d) impact on the TL approximation that is the difference between the nonlinear forecast and the evolved TL.

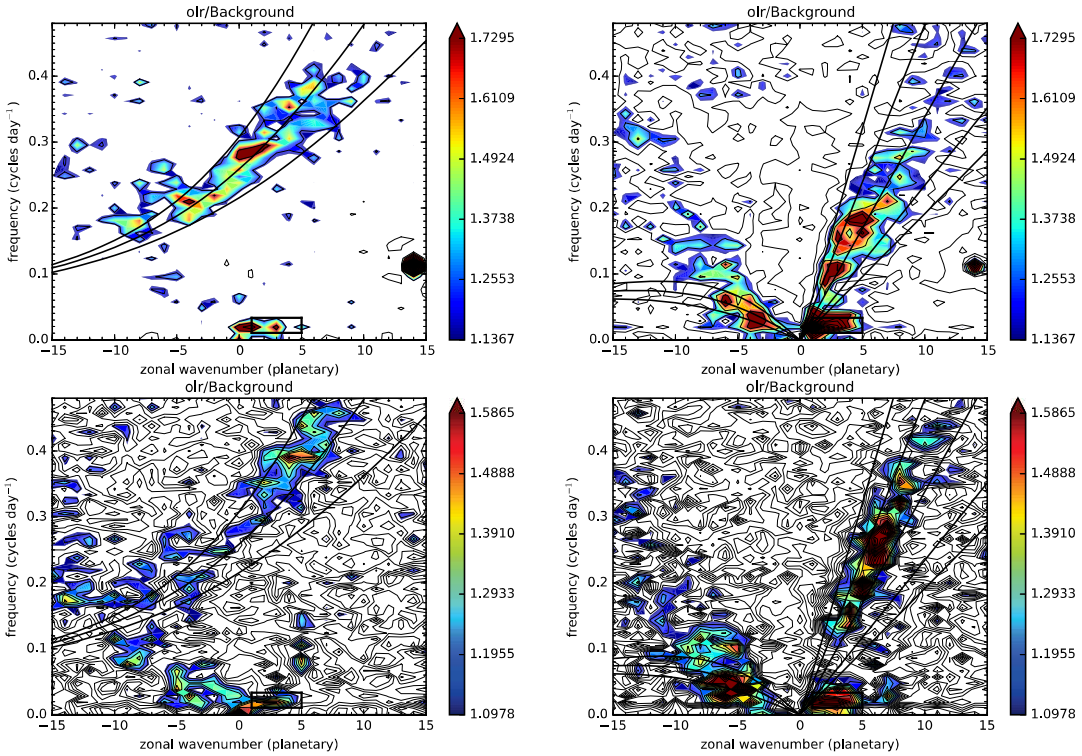

Figure 8: Antisymmetric (left panels) and symmetric (right panels) mode wavenumber-frequency spectra of the outgoing longwave radiation in the equatorial region for (top) control and (bottom) IFS Cy47r1 from a 4-member ensemble of 1 year TL255 free-running forecasts. The black lines in the antisymmetric (symmetric) spectrum show the theoretical dispersion relations for even (odd) meridional mode-numbered equatorial waves for three equivalent depths of 12 m, 25 m, and 50 m.

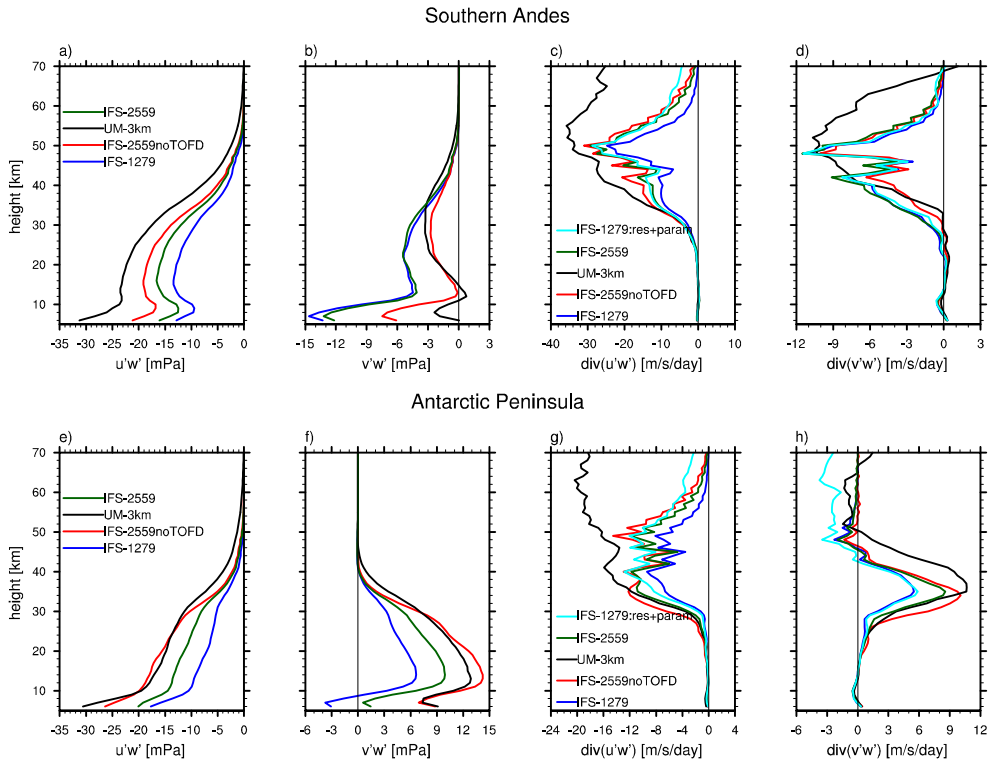

Figure 9: Vertical gravity wave momentum fluxes (a-b, and e-f) and drag (c-d, and g-h) averaged over the Southern Andes (a-d) and Antarctic Peninsula (e-h) in October 2010 for different models: TCo1279 IFS (blue), TCo2559 IFS with (green) and without (red) turbulent orographic form drag parametrization, and 3km Met Office Unified Model (black). Panels c) and d) and g) and h) show the total gravity wave drag (resolved + parametrized OGWD and NOGWD) for IFS in cyan.

is being carried out as part of the International Space Science Institute (ISSI) working group "New quantitative constraints on orographic stress and drag", involving scientists from Germany, France, UK, US and Japan. The comparison to high-resolution limited area models shows that the momentum flux is a factor of 2-3 too small in TCo1279L137 24-48h forecasts, resulting in too little GW drag in the upper stratosphere and mesosphere (see Fig. 9). This is so even if parametrized OGWD and NOGWD contribution — which is small in IFS at TCo1279 resolution — is added to the resolved drag (cf. cyan lines to black lines in Fig. 9). When comparing to retrieved background T perturbations from AIRS over-passes, IFS at TCo1279L137 resolution has an amplitude error between -23% and -40% in the 20-45 km altitude region (not shown). GW momentum flux from higher resolution TCo2559L137 (or 5 km in the horizontal) IFS forecasts was also compared to limited area models: While the resolved IFS GW momentum flux increases somewhat, it is much lower than in the limited area models as shown in Fig. 9. A key parametrization that controls the amplitude of the resolved momentum fluxes in IFS is turbulent orographic form drag (TOFD), which represents near surface orographic drag from hills and individual mountains of scale 5 km to a few hundred meters (Beljaars et al., 2004). In IFS, TOFD has a much larger contribution to total surface stress than in other models, including the Met Office UM (Zadra, 2015). If TOFD is switched off at TCo2559L137 resolution, GW momentum flux approaches what is found in 3 km UM and WRF simulations (cf. red and black lines in Fig. 9). It remains to be understood if TOFD in IFS is too strong at TCo2559L137 resolution or if the limited area models at 3 km resolution are underestimating the low level drag and therefore generating momentum fluxes that are too large. As satellite measurements can only provide GW information for scales larger than 50 km, it is difficult to put a quantitative constraint on the fluxes. It is, however, known that at TCo1279L137 resolution, switching off TOFD considerably degrades forecast skill.

Apart from high-latitude orographic GWs, the impact of convection on stratospheric GW momentum fluxes has also been examined by Stephan et al. (2019). They found that the standard convection parametrization in the ICON model (which is the same as in the IFS Cy47r1) underestimates the mean and the tail in the probability distributions of GW momentum fluxes compared to observations while explicitly simulated deep convection overestimates the fluxes at 5 km horizontal resolution. Similarly, it is found that 9 km horizontal resolution IFS simulations underestimate resolved convective GW momentum flux with the standard convection parametrization but overestimate convective GW momentum flux with explicitly simulated deep convection when compared to 1.4 km IFS simulations (from the Horizon 2020 INCITE project) with resolved deep convection.

ECMWF is currently preparing a moist physics upgrade for Cy48r1 that contains a more active deep convection with a larger tail in the mass flux and precipitation distributions. This has a significant impact on resolved stratospheric GW activity as shown for June 2016 in Fig. 7c. The figure shows the standard deviation of differences in T' (obtained by retaining total wavenumbers $N > 60$) between the TCo399 analysis with the new moist physics and the CTRL 46r1 analysis. Interestingly, the intense summertime convective regions over the continents stick out with little impact on GWs over high latitudes. This is expected, as the changes target deep convection which is most active in the tropics. A degradation in the tangent linear approximation (the fit between the nonlinear model and the evolved tangent linear model) with the new physics package is also noticeable in the stratosphere. Plotting in Fig. 7d, similiar to Fig. 7c, the standard deviation of T' differences between the nonlinear and tangent linear model reveals again that the main differences are in convective regions where GWs are highly variable in space and in their spectral characteristics.

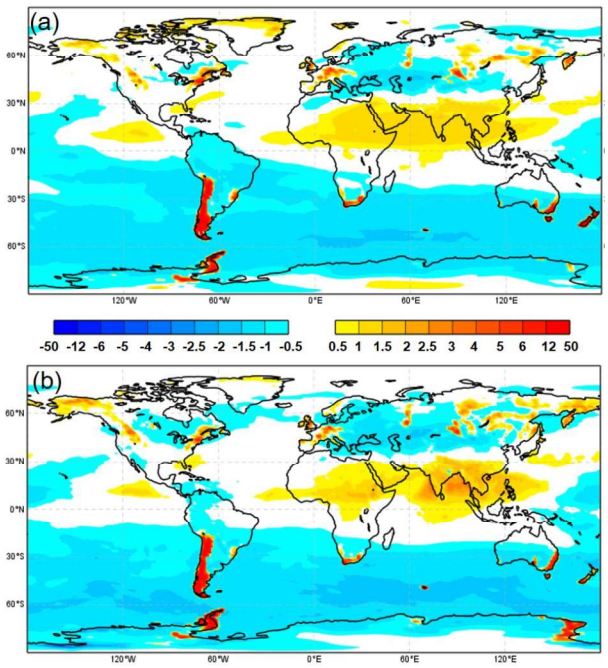

Figure 10: Total (orographic + non-orographic) parametrized zonal momentum flux (mPa) at 100 hPa during August 2000 from (a) the operational model configuration with a globally uniform non-orographic launch momentum flux and (b) with a re-tuned NOGWD scheme with an additional convective mass flux dependent contribution to the launch momentum flux. Note that the orographic fluxes are highlighted in red with the inverse sign.

2.4.3 Nonorographic gravity wave parametrization: building in convective sources

It has been widely suggested that NOGWD parametrizations should include flow-dependent local sources due to convection to improve GW intermittency. While Lott and Guez (2013) developed a method to include convective sources with both varying amplitude and spectral composition, the method employed by Bushell et al. (2015) only applies a locally varying amplitude of the standard subgrid launch momentum flux depending on the total precipitation. We have also experimented with a method similar to Bushell et al. (2015) by adding to the standard IFS NOGWD scheme (Orr et al., 2010) a contribution — to the otherwise globally constant subgrid momentum flux — that depends on the convective mass flux at 500 hPa times the updraft vertical velocity. This source naturally has units of momentum flux (Pa). The results are illustrated in Fig. 10, which shows the total (OGWD+NOGWD) zonal momentum flux at 100 hPa for August 2000 for the control simulation with (a) the standard (i.e., source-independent) NOGWD scheme, and (b) with the additional convective sources. Clearly, the convective sources provide more variability in the convective regions as quantified by the broader exponential probability distribution of the momentum fluxes (not shown). However, the overall results in Fig.10a,b are rather similar as the filtering effect on the waves by the background winds dominates. Furthermore, including locally varying sources made the standard parametrization strongly resolution dependent and required tedious re-tuning. Overall, no clear benefit was seen compared to the simple operational scheme and it was decided not to pursue these developments for now. To the best of our knowledge, current operational or high resolution models do not include a NOGWD scheme with flow-dependent sources. Therefore, we believe that the most important and consistent approach for improving GWs is to increase the model resolution (both horizontal and vertical), to better represent convection, and to reduce the depth of the sponge layer.

2.4.4 *Nonorographic gravity wave parametrization: impact on stratospheric circulation*

The meridional structure of stratospheric temperatures is affected by the Brewer-Dobson circulation (BDC), which is mechanically driven by wave drag from momentum transfer associated with dissipating waves (e.g. Shepherd, 2000). Adiabatic cooling associated with upwelling and adiabatic warming associated with downwelling induces departures of temperature from radiative equilibrium. The wave drag arises from both resolved and parametrized waves: Because of this, it is hard to gain a quantitative constraint on the BDC strength from reanalyses to validate the IFS, especially if the IFS is run at a higher resolution than the reanalyses. As small scale waves in the reanalyses are not resolved and need to be parametrized, there are uncertainties in their representation. Observational constraints on the BDC based on the age of air diagnostics from observations of carbon dioxide or sulfur hexafluoride (SF_6) also have large uncertainties and are hence also not very reliable (for a review see Butchart, 2014). The tropical tape recorder signal of water vapour (Mote et al., 1996) is understood to be the result of vertical advection by the BDC, and therefore has been used as a proxy for the tropical upwelling speed of the BDC. A comparison of the tape recorder signal in the IFS to MLS, suggests that the tropical upwelling is a factor of two too fast in the IFS and a factor of four too fast in ERA-Interim (Polichtchouk et al., 2017).

To investigate if the warm polar mid- to upper stratosphere bias 4) in the winter hemisphere or the cold tropical bias 2) (see Fig. 1) are due to a too strong BDC, the role of NOGWD on the BDC was investigated in Polichtchouk et al. (2018, 2017) with the 32-year TL255L137 resolution IFS simulations. The "downward control" principle (Haynes et al., 1991) was used to separate the role of different parametrized and resolved wave drags on the BDC. In summary, it was found that NOGWD dominated the sensitivity over OGWD, especially in the SH. As OGWD is explicitly tied to the unresolved topography, already at TL255L137 resolution a lot of the gravity wave spectrum is resolved rather than parametrized and therefore OGWD has less of an impact. By changing the amplitude of the NOGW launch flux, the total annual-mean tropical upwelling (and therefore tropical temperatures) in the lower stratosphere is nearly unchanged due to a compensation between the resolved and parametrized wave drag driving lower-stratospheric tropical upwelling. However, for the polar downwelling such strong compensation was not observed. Instead, complex interactions between NOGWD and resolved wave drag were observed over both poles, which does not allow to correct bias 4) simply via NOGWD amplitude changes.

The impact of NOGWD on polar vortex variability, namely SSWs in the NH and final warmings in the SH, was also investigated in Polichtchouk et al. (2018). It was found that the current NOGWD tuning is optimal for polar vortex variability at TL255L137 resolution (see section 4 in Shepherd et al., 2018) but an increase in NOGWD leads to an increase in SSW frequency, reduction in amplitude and persistence, and an earlier recovery of the stratopause following an SSW event. The SH final warming date is also brought forward when NOGWD is increased but the inter-annual variability in polar vortex breakdown dates was unchanged by changes in NOGWD (see Fig. 6 in Polichtchouk et al., 2018).

2.5 Impact of reduced vertical diffusion in the tropical stratosphere

Podglajen et al. (2014) showed that the 2010 operational ECMWF analysis and ERA-Interim had great difficulties representing winds in the tropical lower stratosphere near a critical level. This was corroborated by experiments at ECMWF, using the 2016/2017 operational IFS (Cy43r1), assimilating the same super-pressure balloons (SPB) from the 2010 pre-ConcordIASI campaign. The analysis was able to pull to the SPBs, but after only 12 hours of forecast the information was to a significant degree lost. However, the performance was still a great improvement to the analysis, compared to not assimilating the SPB winds (Dharmalingam et al., 2019). Individual tests assimilating SPBs, radiosondes and GNSS-RO data,

respectively, showed that all these data sets resulted in similar increased wind shear in the tropical lower stratosphere, pointing to the forecast model having problems correctly representing vertical wind shear. Indeed, from a detailed comparison of vertical wind shear between high vertical resolution radiosonde observations and IFS, it is known that the wind shear is a factor of 2-3 too low in the tropical stratosphere (Houchi et al., 2010). Moreover, 10-15 days after initialization IFS shows weakened QBO amplitude (Garfinkel et al., 2018).

Seasonal forecasts of the QBO have also demonstrated problems with winds in the lower stratosphere, and it was discovered some time ago that the only way to obtain both a realistic QBO period and amplitude in the lower stratosphere was to reduce vertical diffusion. An ad-hoc method was developed to do this, whereby the "long-tail" Richardson number dependent formulation used in the IFS was modified to make it more "short-tail". Vertical diffusion is unchanged at low Richardson numbers (when vertical mixing is expected), but is more strongly suppressed for the high Ri-numbers typical of the stably stratified stratosphere, allowing the relatively weak shears of the QBO to persist for many months instead of being damped away. The modification is applied only within the range of pressure levels (nominally 100 to 10 hPa) where it is important for the QBO, and only within the tropics. The reduction is blended smoothly with a quadratic function between 120 and 80 hPa.

The impact of this modification on the QBO is illustrated in Fig. 11, which shows the QBO at 30 hPa from two 5-year free-running integrations. The period of the QBO can easily be tuned by adjusting the strength of the NOGWD parametrization in the model (see e.g., Polichtchouk et al., 2017), which here has been done separately for the two vertical diffusion settings. Much more difficult to reproduce is the observed amplitude and vertical structure of the QBO, and here the modified vertical diffusion (red) gives a large improvement to what is possible with the standard version (blue). Although reduced vertical diffusion allows much improved simulation and forecasting of the QBO, there is still a tendency for too much eastward momentum in the lower equatorial stratosphere, which is believed to be due to errors in resolved wave forcing in the model. Higher horizontal resolution and stochastic physics both exacerbate this problem to some extent. Vertical resolution increases above L137 do not benefit the QBO structure at lower levels, but have been found to give clear improvements to the QBO at 5-10 hPa. Although it is helpful to know that vertical resolution remains an issue at these levels, the cost of a resolution increase above L137 means that it is not yet feasible.

This parametrization modification was also tested in the analysis and forecast model in the medium range. These experiments used IFS Cy43r1 with outer loop and forecast model running at TCo399. The reduced vertical diffusion had a large positive impact in the tropical lower stratosphere. The standard deviation in background departures of tropical radiosonde u-winds at 50 hPa was reduced by 7% and the bias was also significantly improved (Fig. 12). Strong improvements (3-4%, not shown) were also seen in the tropical lower stratosphere relative to radiosonde temperature and GNSS-RO data. Radiosondes and GNSS-RO have high vertical resolution, so they capture the true vertical wind shear well.

The improvement, verified against operational analyses, was maintained for the whole forecast range up to 10-days above 100 hPa, as shown in Fig. 13. However a slight degradation was found around the tropical tropopause ($\sim$100 hPa). It appears that the change is beneficial for the free stratosphere but not around the tropical tropopause level. The method could almost certainly be improved by making the activation level for the reduced diffusion a function of tropopause height, rather than the imposed hard-coded pressure level limits used here, which in any case intrude too far down into the upper tropical troposphere. Note that it is common for other models to reduce or remove vertical diffusion in the stratosphere. However, the ECMWF strategy is instead to seek a more thorough future overhaul of vertical diffusion to allow a unified approach throughout the troposphere and stratosphere.

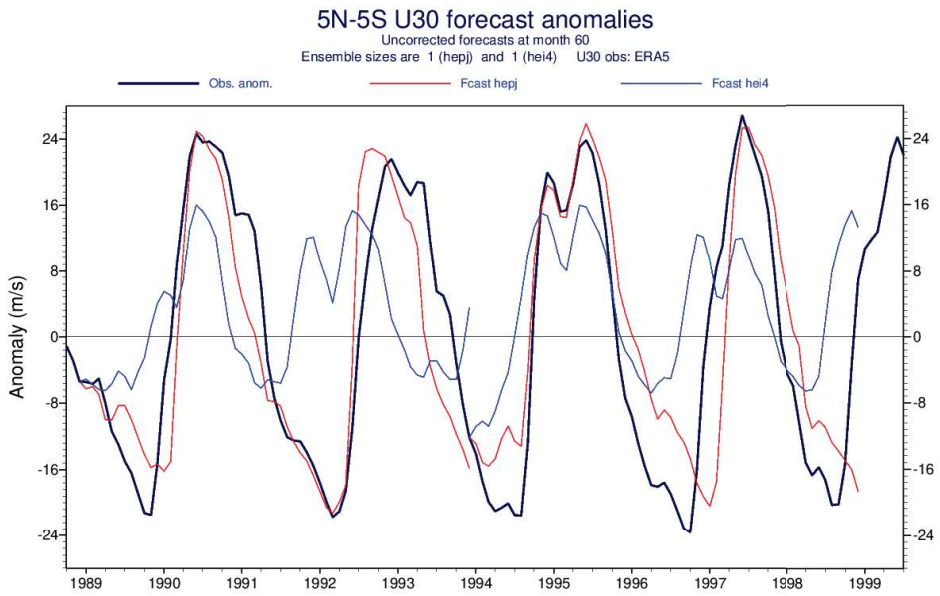

Figure 11: 5N-5S zonal-mean zonal winds at 30 hPa from 5-year free-running integrations of the IFS starting in 1989 and 1994. Blue is the standard and red is the modified vertical diffusion. Model resolution is TCo319L137, and the cycle is 47r1. Model winds are plotted as anomalies relative to the ERA5 seasonal cycle, with no bias corrections applied.

Note that the QBO is believed to be driven by low-frequency equatorial waves – that are partially forced by parametrized processes such as tropical convection – and inertia-gravity waves. Therefore, in addition to the vertical diffusion and NOGWD, resolution (horizontal and vertical), parametrized deep convection (via its effect on resolved GWs as shown in section 2.4.2), and stochastic physics parametrization all play a role in the amplitude and period of the QBO. Therefore, the optimal tuning will depend on the resolution, vertical diffusion and deep convection setup and whether a stochastic physics parametrization is used or not.

2.6 Stratospheric water vapour and methane oxidation

Water vapour in the stratosphere is not constrained in the assimilation system because of a lack of long-term reliable observational data available in near real time. It is therefore critical that realistic parametrizations for stratospheric water vapour processes are included in the IFS. Water vapour enters the stratosphere from the tropical troposphere and the source is limited by freeze drying of the ascending air in the cold temperatures of the tropical tropopause layer. As the air ascends through the stratosphere, methane oxidation produces water vapour and the stratospheric humidity increases. The BDC transports the higher values of water vapour towards the stratopause and descends back towards the lower stratosphere at higher latitudes. To accurately represent the distribution of water vapour in the stratosphere it is therefore important to include a parametrization of the methane oxidation process.

Methane is produced by natural and anthropogenic sources at the Earth's surface and is well-mixed in the troposphere. It is transported upwards across the tropical tropopause and subsequently decreases due to oxidation as it ascends. The methane oxidation process produces two water vapour molecules for each methane molecule. Given this is the dominant source of water vapour within the stratosphere,

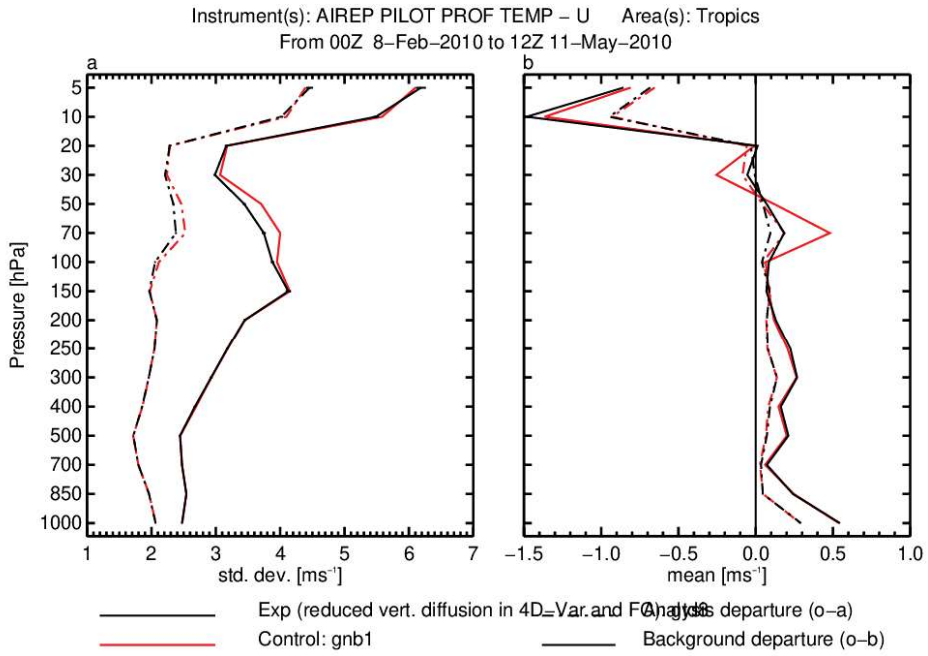

Figure 12: The mean (left) and standard deviation (right) of conventional data u-component winds (mainly radiosondes above 150hPa) departures in the tropics. Red is the control experiment and black is the experiment using reduced vertical diffusion in the tropical stratosphere. The experiments cover the pre-ConcordIASI period 8 Feb 2010 to 11 May 2010.

the sum of twice the mass of methane added to the mass of water vapour ($2[CH_4] + [H_2O]$), represented by M, is therefore expected to be approximately uniform throughout most of the stratosphere. This is confirmed with satellite observations (Jones et al., 1986; Bithell et al., 1994; Randel et al., 1998), except for the region below 10 hPa in the high-latitude Southern hemisphere around the spring equinox, due to condensation at the very cold temperatures in the Antarctic polar vortex (Bithell et al., 1994).

Prior to IFS Cy25r1, the methane oxidation parametrization specified 6 ppmv for M and this was used in production of the ERA-40 reanalyses. However, Randel et al. (1998) found the mid-to-upper stratospheric humidity in ERA-40 to be drier than measurements from UARS and suggested a higher value of 6.8 ppmv for M was closer to the observations at the time. This was changed in IFS Cy25r1 for operations in 2002 and used for both the ERA-Interim and ERA5 reanalyses, which therefore have higher humidity in the mid-to-upper stratosphere than ERA-40.

Since then, the amount of methane in the troposphere has continued to increase leading to an associated increase in water vapour in the stratosphere. A recent reassessment of the latest observations of water vapour from the Aura MLS led to a further increase of the value for M by 13% to 7.7 ppmv in IFS Cy45r1, operational from June 2018. Figure 14 shows the stratospheric water vapour from the IFS before and after the change as well as from MLS, highlighting the improved agreement with recent observations.

Given the past changes and future projections for methane in the troposphere (Riahi et al., 2007), it is important to capture the evolving impact on the stratospheric water vapour from methane oxidation. Going forward, one option is to introduce a time-varying function for M in the IFS that would represent the evolving stratospheric water vapour changes over time, both for operational forecasts, historical simulations and future versions of the reanalysis system at ECMWF. Currently the methane oxidation in the reanalysis is fixed over time to values representative of the late 1990s. This is likely to overestimate

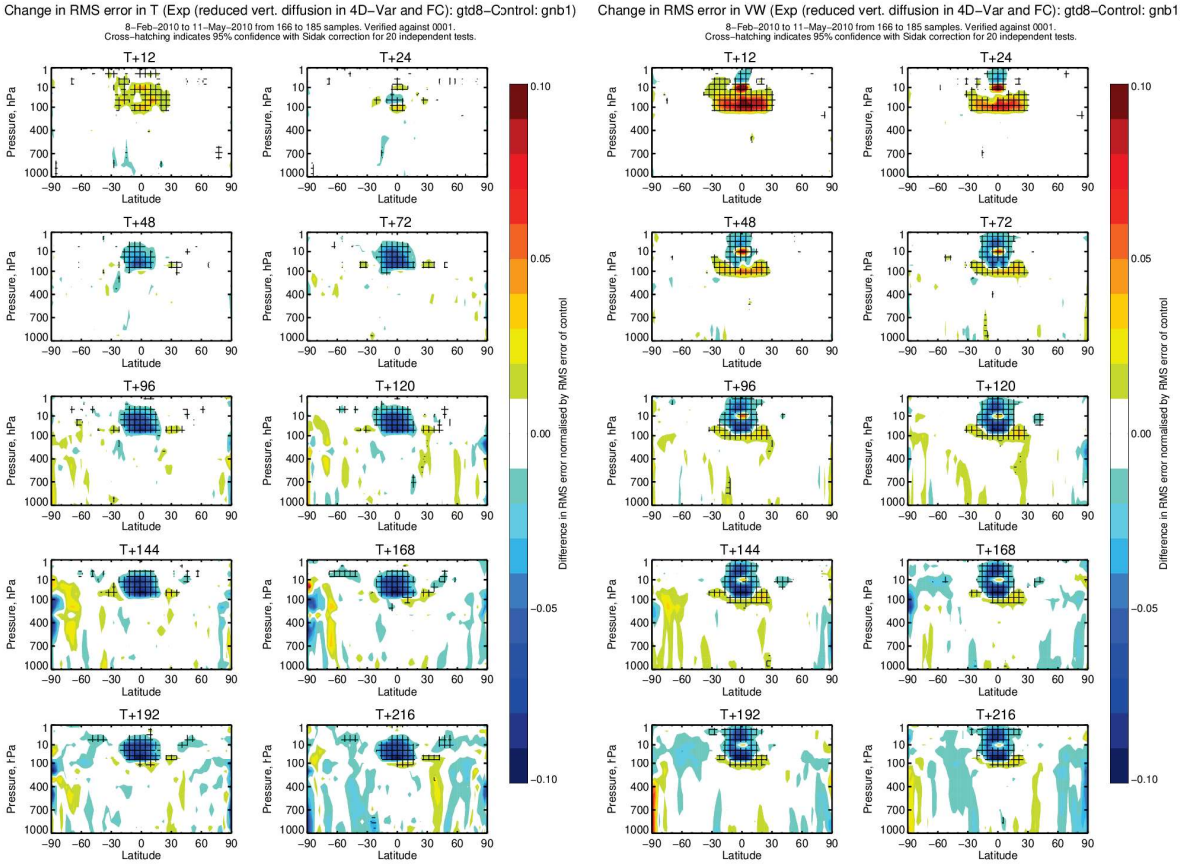

Figure 13: Zonal average normalised change in the root mean square error of (left) temperature and (right) vector wind from reducing the vertical diffusion for the period 8/2/2010 to 11/5/2010. Negative values (blue) indicate a reduction in error from the vertical diffusion reduction and red values an increase in error. Verified against operational analyses. Cross-hatching indicates 95% confidence.

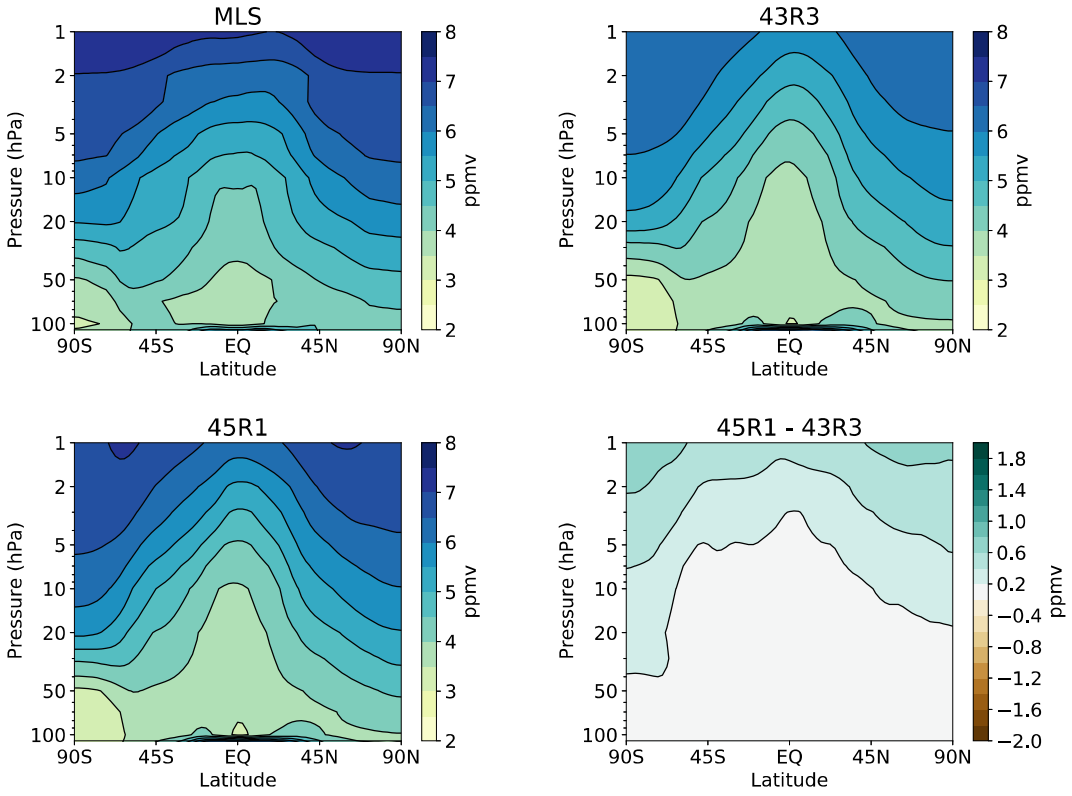

Figure 14: Zonal-mean annual-mean cross section of stratospheric water vapour for (a) Aura MLS data for 2005-2013, (b) IFS Cy43r3, (c) IFS Cy45r1 with increased methane oxidation and water vapour, (d) the difference between 45r1 and 43r3. IFS fields are an average of a 4-member ensemble of 1 year TL255 free-running forecasts. Units are parts per million by volume (ppmv).

the moisture for earlier times and underestimate at later times. Simmons et al. (2020) highlight the underestimate of stratospheric moisture in both ERA-Interim and ERA5 reanalyses compared to the Aura MLS dataset during the period 2007-2011.

CAMS provides an analysis and medium-range prediction of methane in the IFS and a stratospheric chemistry parametrization explicitly represents the methane oxidation process to form water vapour. Verma et al. (2017) evaluate the stratospheric methane from CAMS and find good agreement with the MIPAS data (overall bias of 9 ppb and random error of 6.5 ppb). Huijnen et al. (2016) includes an evaluation of stratospheric water vapour from the CAMS system and shows a reasonably accurate representation of water vapour compared to MLS data. There is therefore potential to use the CAMS fields in the future to evaluate and improve the methane oxidation parametrization for a more accurate representation of stratospheric water vapour in the IFS.

2.7 Key summary points

- Global-mean temperature is the first order quantity to represent correctly in the middle atmosphere. Due to the lack of observational constraints on temperature above 2.5 hPa and a comparatively small number of observations in the stratosphere, accurate representation of global-mean temperature in the model is also important for a good analysis/reanalysis product. Global-mean middle

atmosphere temperature biases 1) and 2) in Fig. 1 have been substantially improved largely by radiation and ozone climatology updates. However, global-mean temperature biases in the IFS are also sensitive to horizontal resolution if vertical resolution is not suitably adjusted and/or a higher order vertical interpolation and vertical filtering is not used.

- Planetary and gravity waves exert a second order effect because they drive: i) middle atmosphere temperature away from radiative equilibrium and thus determine meridional temperature structure, and ii) stratospheric variability (SSWs, QBO, SAO, final warmings). Because of lack of GW resolving assimilated observations in the upper stratosphere/mesosphere it is difficult to put quantitative constraints on GW momentum fluxes and therefore verify the realism of modelled waves. However, resolution, sponge layer design, parametrized deep convection, turbulent orographic form drag, and NOGWD all have an impact on the simulated spectrum of waves in the stratosphere. The complex interaction of NOGWD with the resolved waves means that bias 4) in Fig. 1 can not be fixed via NOGWD strength changes.

- In addition to resolved and parametrized waves, the vertical diffusion parametrization exerts a strong control on the QBO in the lower stratosphere: In stable conditions the vertical diffusion appears too strong and unduly damps equatorial wind shears; reduced vertical diffusion in the tropics markedly improves tropical winds in analysis and forecasts at all time ranges.

- While the stratosphere is very dry in comparison to the troposphere, representation of water vapour is important due to its radiative effect. In the lowermost polar stratosphere, presence of excessive moisture leads to a persistent cold bias 3) there (Fig. 1). The lack of assimilated water vapour observations means that the moist bias is there already in the analysis and therefore the cold bias develops already in short forecasts. This bias is not sensitive to horizontal resolutions beyond TCo319 and vertical resolutions beyond L137. Lack of assimilated moisture observations in the stratosphere also necessitates the need for a methane oxidation parametrization.

- Not all stratospheric biases are important for tropospheric predictability. Alleviating global-mean biases in the middle atmosphere does not appear to impact tropospheric skill. However, biases that affect meridional temperature gradients and influence stratospheric variability appear important for tropospheric influence (see also ahead to section 5.2).

3 4D-Var developments

Here we review two recent developments that specifically target 4D-Var in the middle atmosphere. It should be emphasised, however, that any model changes will also influence 4D-Var analysis, especially in the stratosphere with comparatively little observations (e.g., section 2.5 and Fig. 12) and sometimes changes that are beneficial to the non-linear model do not perform well in the 4D-Var, as is also discussed here.

3.1 Weak-constraint 4D-Var

The standard, strong-constraint formulation of 4D-Var is designed to correct for random, zero-mean errors from the model forecast and observations. However, as discussed in section 2.2, significant systematic errors are generated in the IFS model. To deal with these types of errors a modification of the standard 4D-Var algorithm, weak-constraint 4D-Var, has long been proposed and implemented in operations

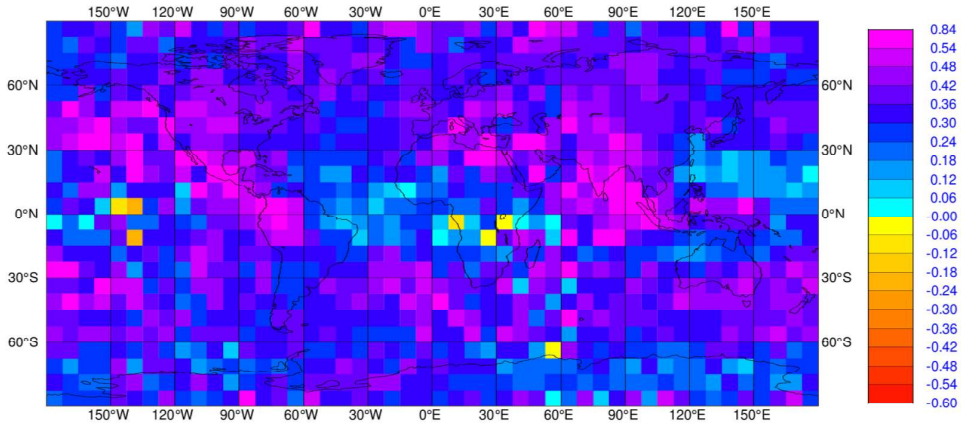

Figure 15: Difference between GNSS-RO temperature retrievals and first-guess temperatures from ECMWF operations between 70 and 100 hPa over the period 31 August 2018 to 31 January 2019.

at ECMWF in 2009. In the original implementation, the model error covariance matrix (which describes the error statistics of model bias) was built using differences between 12-h ECMWF ENS forecasts. The forecasts were initialised from the same initial states but were run with model error perturbations based on the Stochastically Perturbed Parametrization Tendency scheme (SPPT) and the Stochastic Kinetic Energy Backscatter scheme (SKEB). These model error parametrizations produce perturbations whose energy is higher at synoptic and sub-synoptic scales than at larger scales. Consequently, the resulting model error covariance matrix contains length scales similar to the ones that dominate the background error covariance matrix. This situation made it impossible for weak-constraint 4D-Var to attribute model and background error to the correct source. For this reason, the original version of weak- constraint 4D-Var was only implemented in the upper stratosphere, as a full application of the algorithm would have led to an unacceptable degradation. Thus, the impact of weak-constraint 4D-Var on ECMWF analyses and forecasts has always been small.

New model error diagnostics have been developed to compare the first-guess trajectory with GNSS-RO measurements. They provide near-homogeneous coverage that can be used to study the spatial structure of the model bias. Figure 15 shows the difference between temperature retrievals and first-guess trajectories from ECMWF operations between 70 and 100 hPa from 1 October 2018 to 1 February 2019, when IFS Cy45r1 was operational. The diagnostic presented in Fig. 15 suggests that model error which develops during the 4D-Var assimilation window contains identifiable large-scale structures, which are distinct from the small-scale errors in the first guess initial conditions (the background) diagnosed by the Ensemble of Data Assimilations (EDA) system. The cold model bias maximizing in the tropics is due to inconsistent vertical to horizontal resolution aspect ratio (see section 2.2.2). A new estimate of the model error covariance matrix has been computed from a climatology of the model error vectors estimated by the current weak-constraint 4D-Var (Laloyaux et al., 2020). The correlation length scale from this new model error covariance matrix is more consistent with the patterns observed in Fig. 15. In this new implementation of weak-constraint 4D-Var, which became operational in Cy47r1, first-guess trajectory adjustments deal with large-scale temperature errors which vary on timescales longer than the assimilation window (i.e. $>$12 h), while state increments correct smaller-scale and more transient background errors. This scale separation between the two types of errors is a necessary condition for weak-constraint 4D-Var to attribute model and background errors to their correct sources. This has been corroborated through theoretical considerations and experiments with a simplified quasi-geostrophic model (Laloyaux et al., 2020).

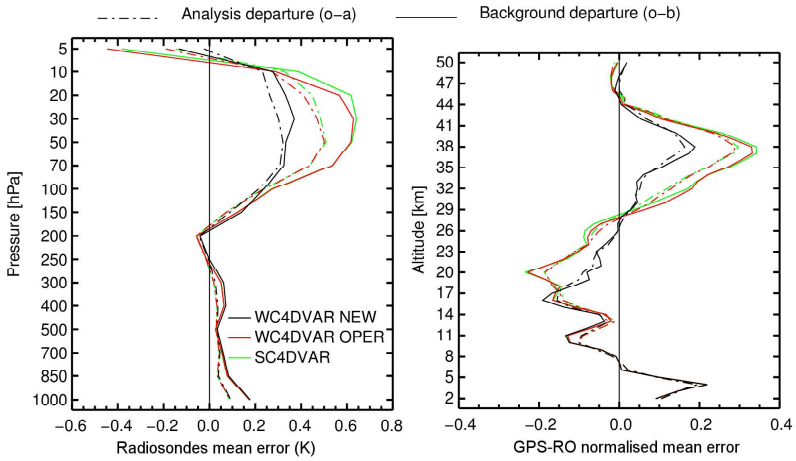

Figure 16: Vertical profiles of analysis departures (O-A) and background departures (O-B) with respect to (a) radiosonde observations and (b) GNSS-RO observations for strong-constraint 4D-Var (SC-4DVAR), the weak-constraint 4D-Var in Cy46r1 (WC-4DVAR-OPER) and the new weak-constraint 4D-Var in Cy47r1 (WC-4DVAR-NEW). The statistics were computed for the period from 1 October 2018 to 1 April 2019.

Experiments have been performed at operational HRES resolution (TCo1279L137) to compare weak-constraint 4D-Var based on the new model error covariance matrix to the weak-constraint 4D-Var configuration used in Cy46r1 and to the strong-constraint 4D-Var configuration, in which no model error forcing is applied. The left panel of Fig. 16 shows vertical profiles of the mean analysis departures and background departures with respect to radiosonde data for strong-constraint 4D-Var (SC-4DVAR), the weak-constraint 4D-Var operational in Cy46r1 (WC-4DVAR-OPER) and the new weak-constraint 4D-Var operational from Cy47r1 (WC-4DVAR-NEW). In the weak- constraint experiments, the first-guess trajectory used to compute the background departures is corrected using the model error estimate from the previous assimilation. A smaller mean error in the background departures shows that the assimilation system is able to estimate and correct a significant amount of model bias. The weak-constraint 4D-Var used in Cy46r1 corrects only a small fraction of the model bias above 40 hPa, while the new model error covariance matrix better corrects the diagnosed temperature biases of the model above 100 hPa, reducing the mean error by up to 50%. Since radiosonde observations are available only up to 5 hPa, the right panel of Fig. 16 shows analysis and background departures with respect to GNSS-RO bending angles to assess the upper stratosphere. The results show that bias in the upper stratosphere located between 30 km and 45 km (11 hPa to 1.5 hPa) is also significantly reduced in Cy47r1.

3.2 Matching time step in outer/inner loops of 4D-Var

One puzzling result of research work into the nonlinear effects in 4D-Var reported in Bonavita et al. (2018), was the different response in the troposphere and stratosphere to increasing the number of IFS outer loop iterations. Namely, going beyond three outer loop re-linearisations was shown to bring about consistent improvements to the analysis and forecast skill in the troposphere, but to be marginally detrimental in the stratosphere. In particular, with five outer loops there were clear signs that the incremental 4D-Var analysis was starting to diverge from stratospheric-peaking radiance observations and increasing the number of outer loops only made the problem worse.

Investigation into this issue, has led to a re-discovery of a 4D-Var divergence mechanism, which was first identified in Trémolet (2007). Specifically, the speed of GW propagation is sensitively dependent on the

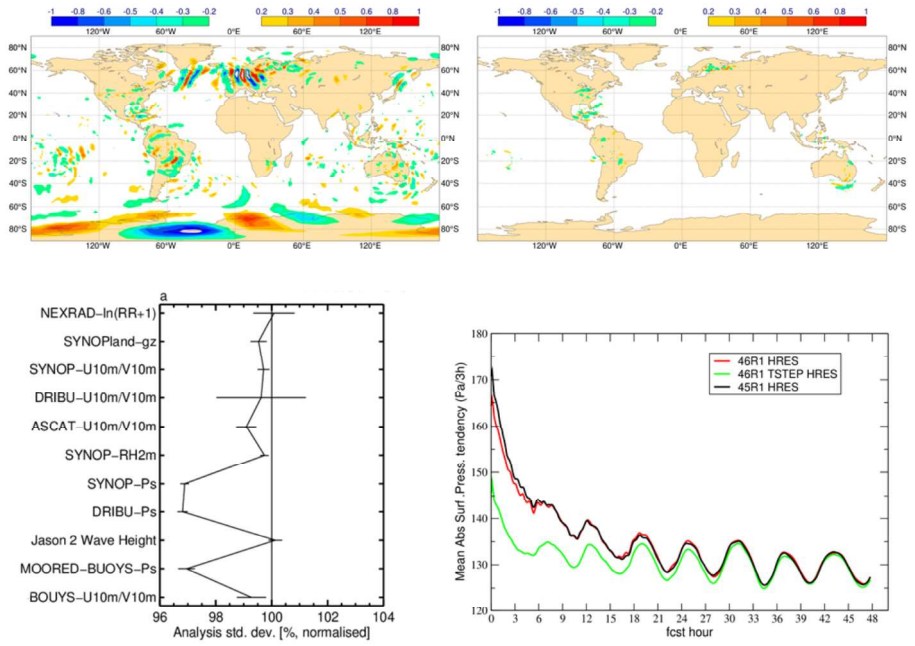

Figure 17: Top row: difference of the 5hPa temperature analysis increments computed by the nonlinear IFS model and the TL IFS model in the last minimisation for an experiment using the current operational time steps (left panel; 450s for the nonlinear model, 900s for the TL model) and an experiment using matching time steps (right panel; 450s for NL and TL models). Bottom row, left: Normalized difference in the standard deviation of the analysis fit of surface observations for timestep matching experiment vs a control. Bottom row, right: Forecast evolution of the global mean absolute surface pressure tendency for IFS Cy45r1 (black), IFS Cy46r1 (red) and IFS Cy46r1 with matching timestep (green).

timestep used in the model. This is because all semi-implicit time-stepping methods reduce the phase speeds of the fastest high-wavenumber modes (if these are not filtered by other means such as a digital filter), with the reduction in phase speed being proportional to the time step size (Simmons et al., 1978; Simmons and Temperton, 1997). In the previous HRES 4D-Var assimilation cycle (Cy46r1), the inner loop timestep was twice as large as the timestep used in the outer loop nonlinear trajectory computation (900s vs 450s). This choice was dictated by the desire to save computational resources and to reduce the total time needed to run 4D-Var in operations. However, it also introduced a significant difference in the GW propagation speed in outer vs inner loop integrations, with the result that spurious gravity-wave-like increments are generated during the 4D-Var analysis (see top left panel of Fig. 17 for an example). On the other hand, having the same time step in the last minimisation eliminates this problem (top right panel of Fig. 17).

The solution of having matching timesteps (450s) in the last minimisation has been tested extensively and it is now used in the operational IFS Cy47r1. At the cost of a relatively small (approx. 6%) increase in the computational cost of the analysis, the matching timestep brings: i) clear improvements to stratospheric analyses and forecasts, with a smaller but visible impact on tropospheric skill (Fig. 18); ii) monotonous convergence of incremental 4D-Var in specific atmospheric situations where current incremental 4D-Var fails to converge at all (e.g., SSW events); and iii) clearly improved initial balance of the 4D-Var analysis, as evidenced by diagnostics in observation space (bottom left panel of Fig. 17: notice the improvement in the model fit to surface pressure observations) and model space (bottom right panel).

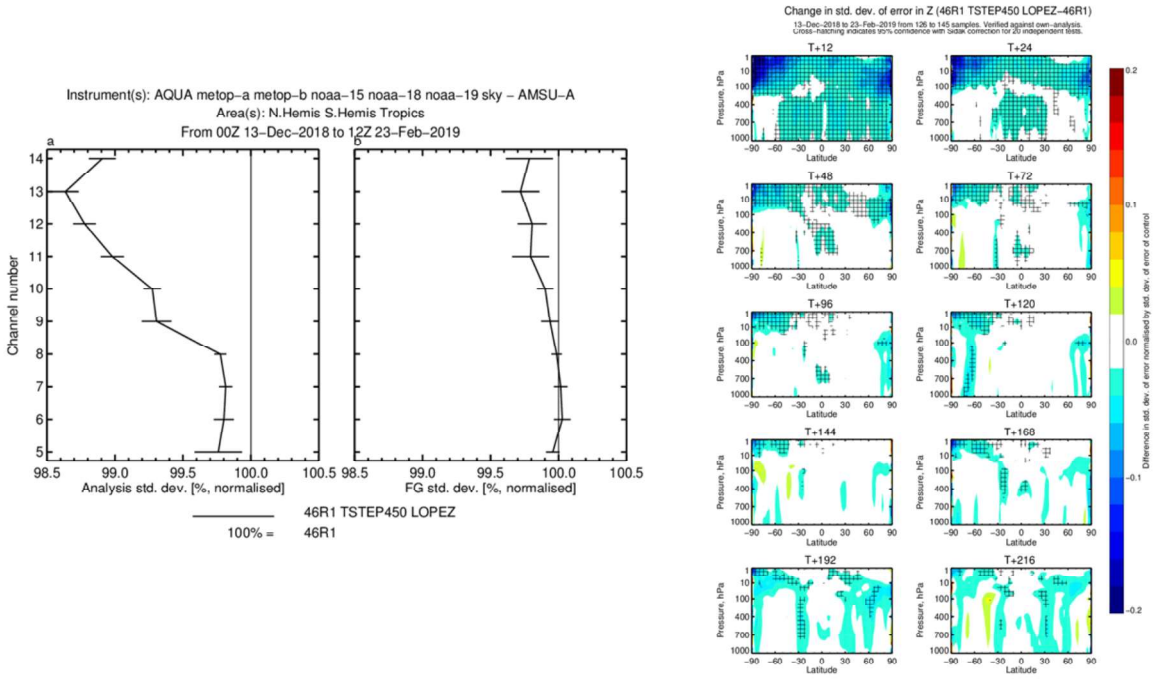

Figure 18: Left panel: Normalised difference in analysis (left) and background forecast (right) observation departures standard deviations for AMSU-A radiances (channels 5 to 8 are mainly sensitive to tropospheric temperatures, channels 9 to 14 to stratospheric temperatures) between a standard Cy46r1 experiment (100% baseline) and an experiment using matching time steps in the last 4D-Var minimisation (values lower than 100 indicate better analysis/forecast fit to observations). Right panel: Normalized difference in the standard deviation of geopotential forecast errors for an experiment using matching timestep vs a control experiment. Blue shading indicate smaller errors, hatching denotes 95% significance.

3.3 Challenges due to resolved gravity waves

The shallower sponge design introduced in section 2.4.1 is clearly beneficial for the resolved wave representation. The same is true for the new moist physics developments discussed in section 2.4.2 that result in convectively generated GWs better matching observations and convection-resolving models. However, better representing GWs in the middle atmosphere poses challenges for 4D-Var. As GW amplitude growth is proportional to the inverse density, GWs reach very large amplitudes in the upper stratosphere/mesosphere and on breaking and/or saturation result in strong nonlinearities (if not artificially damped by the deep sponge). This violates the small-amplitude linearity assumption inherent in 4D-Var (e.g. Tanguay et al., 1995; Polavarapu et al., 2005), making the evolved tangent linear model depart from the non-linear model and generating large increments in the middle atmosphere.

For example, at TL399 resolution the fit of nonlinear model to tangent-linear model degrades in the upper stratosphere by 20% with the new moist physics developments. A more severe degradation of 100% is experienced with the shallower sponge starting at 55 km altitude. If the sponge layer is removed altogether, 750% degradation occurs. Therefore, while the non-linear model is stable and produces good results with a shallow sponge and/or new moist physics updates, the 4D-Var is unstable and with the shallower sponge results in minimization failures at HRES. A possible solution is to decrease the assimilation window, which is not feasible. An alternative setup where a shallow sponge is only applied in the non-linear model integration and the trajectory and not in the minimization of the 4D-Var is currently explored. This way GWs are filtered from the increments but kept in the background state. However, ultimately we want a realistic stratosphere/lower mesosphere and we need a TL model (or an alternative) that can handle this correctly. Otherwise it will not be possible to assimilate observations that can resolve GWs. As already mentioned in section 2.4.2, the analysis system already experiences problems in the SH winter season, when large amplitude orographic GWs are present in the middle atmosphere.

3.4 Key summary points

- Aeolus wind observations and COSMIC-2 temperature observations have a positive impact on tropical wind analysis and medium-range forecasts in the stratosphere. Both observation types have good vertical resolution that better constrains the QBO. While COSMIC-2 data does not directly constrain the winds, its effect on the wind analysis can be understood via equatorial mass-wind balance.

- There are clear gaps in available long-term global middle atmosphere observations that would not only improve the initial conditions (and therefore forecasts) but also provide a good reference to validate model against. These include: i) humidity observations (see also sections 2.2.3 and 2.6); ii) winds observations (both horizontal and vertical); iii) good vertical ($\sim$1 km) and horizontal ($\sim$50 km) resolution radiance observations that would extend to 80 km altitude and would help to constrain temperature as well as GWs (as temperature perturbations are proportional to GW momentum fluxes), and iv) ozone observations.

4 Observing system developments

A range of observations currently constrain the stratospheric analysis in the ECMWF system. Among the largest constraints are provided by microwave and infrared radiances as well as GNSS-RO observations (e.g. Bormann et al., 2019). These observations are primarily sensitive to temperature, but indirectly also

provide information on wind through balance relationships. Radiances and GNSS-RO complement each other well: Microwave radiances provide good coverage at reasonable horizontal resolution, but poor vertical resolution, whereas GNSS-RO adds good vertical resolution, albeit the horizontal resolution is less good. Radio occultation has seen further advances in recent years, with observations being used to higher altitudes, and with far more observations assimilated. This is discussed in detail in section 4.2.

A critical issue for the successful use of radiance observations in the stratosphere is the treatment of observational biases, made more difficult through the presence of significant model biases (section 2). The latter means the analysis needs to separate between model biases on the one hand (now addressed through weak-constraint 4D-Var; see section 3.1), and observational biases on the other (addressed through Variational Bias Correction, VarBC). GNSS-RO as well as radiosonde observations provide vital anchor points to achieve this separation. Near the stratopause, the highest-peaking AMSU-A/ATMS channel is additionally used as a "soft" anchor, that is, larger bias corrections for this channel are penalized through Constrained VarBC (CVarBC, Han and Bormann, 2016). This channel peaks at around 2.5 hPa and the treatment with CVarBC has been found to stabilise the bias corrections for other lower-peaking channels. All these anchors work in tandem with the scale-separation achieved through the design of the model error covariances as described in section 3.1, as well as an appropriate specification of the background error scales.

Other observations are available that are presently not assimilated in the ECMWF analysis. The observation type that observes the highest levels are the upper atmosphere channels of the microwave Special Sensor Microwave Imager Sounder (SSMI/S) close to 60.8 and 63 GHz. These provide observations sensitive to temperature with low vertical resolution, with the highest channel peaking around 75-80 km. The observations are affected by Zeeman splitting which makes their assimilation more difficult and operational centres have been slow to develop capability to use them. The RTTOV model now includes capability for the Zeeman effect, but unfortunately due to lack of operational uptake there is no obvious follow-on mission for these upper stratosphere/lower mesosphere microwave radiances.

Furthermore, limb sounders, such as MIPAS on ENVISAT, or MLS on Aura measure radiances with a limb geometry, hence gaining high vertical resolution and good signal to noise ratio, but on slant paths over hundreds of kilometers. ENVISAT/MIPAS provided observations from 2002 to 2012, and ozone retrievals were assimilated operationally for some time (Dragani et al., 2015). Furthermore, MIPAS radiances were assimilated experimentally at ECMWF, with large influence on temperature, humidity and ozone in the stratosphere (e.g., Bormann and Thépaut, 2007). However, since 2012 no further infrared limb sounder has been available, and attempts to put new technology into space have not received sufficient support (e.g. PREMIER). It is vital to provide this capability to enable operational monitoring and initialisation of the stratosphere. MLS has been providing observations of upper tropospheric and stratospheric water vapour and temperature since 2004, and these observations have been used for evaluation of the IFS (as shown in section 2).

None of these observations directly observe stratospheric wind, and efforts to gain information on stratospheric wind through a 4D-Var tracer mechanism for ozone (equivalent to 4D-Var tracer mechanism for water vapour in the troposphere) have so far proven unsuccessful. Some wind information can be inferred from radiances and GNSS-RO as discussed in section 4.2, but it remained a major gap in the WMO Integrated Global Observing System (WIGOS) until the launch of Aeolus in 2018. Aeolus provided observations of stratospheric winds on a global scale for the first time, and this is discussed in section 4.1. It is very important that the successful Aeolus demonstration mission leads to an operational follow-on. Other concepts for future observations are discussed in section 6, including high altitude balloon observations.

4.1 Impact of Aeolus in the stratosphere

The European Space Agency's Aeolus Doppler wind lidar mission was launched on 22 August 2018. ECMWF has operationally assimilated Aeolus horizontal line-of-sight winds since 9 January 2020 (Rennie and Isaksen, 2020b,a).

The nominally used Aeolus vertical range-bin settings provide clear-air winds (Rayleigh scattering) up to 20-25 km (50-20 hPa) altitude, therefore sampling the lower stratosphere. Aeolus is limited to 24 range-bins, therefore, to keep the vertical resolution reasonable in the troposphere, the maximum possible altitude of 30 km is not applied. The Rayleigh winds become increasingly noisy with altitude due to the exponential drop in the backscatter (proportional to density). These noise aspects and the limited number of range-bins means the stratospheric range-bin resolution typically is set to 1-2 km. The highest cloudy-air winds (Mie scattering) are due to scattering from water/ice clouds at around tropopause level. An exception is polar stratospheric clouds for which decent quality Mie winds can reach $\sim$25 km, particularly over Antarctica. The flexible vertical sampling of Aeolus is raised over the winter poles to sample the polar night jet. An Aeolus follow-on mission should have higher laser power and more range-bins thus allowing the lower stratosphere to be more consistently sampled.

ECMWF's NWP impact from Aeolus in the period August-December 2019 peaks in the tropical upper troposphere ($\sim$150 hPa) and extends into the tropical lower stratosphere. This can be seen in Fig. 19a, which shows the impact for vector winds at 50 hPa ($\sim$20 km). The 200 to 50 hPa region is where ECMWF analysis experiences large model error uncertainty (see Fig. 2 in Žagar et al., 2013), which Aeolus appears to address. The positive impact on wind and temperature persists into the medium range. Positive impact in the extratropical stratosphere relative to in situ wind observations is also evident particularly above 150 hPa (not shown). The tropical impact is corroborated with short-range forecast verification against GNSS-RO (see Fig. 19b) – with impact reaching 35 km. The impact is also positive relative to ATMS (microwave radiances) channels 13 to 15 which peak between 20-40 km (not shown). The positive impact occurs at significantly higher altitudes than Aeolus observes directly. The mechanism for this could be via i) Aeolus changing the mean wind field at lower altitudes leading to a better propagation of GWs from the troposphere to the stratosphere, and/or ii) Aeolus better constraining equatorial waves, which propagate upward and drive equatorial winds on breaking/saturation.

For the same time period, Fig. 20a shows the mean change in the zonal wind due to Aeolus at 70 hPa and Figs. 20b and 20c show the standard deviation of changes in zonal wind due to Aeolus at 250 hPa and 150 hPa (where the effect is largest). The changes are strongest in convectively active areas of the tropics. Large random changes are also present at 1 hPa (Fig. 20d) which must be associated with GWs. Research continues to understand these changes.

4.2 GNSS Radio Occultation

GNSS-RO measurements provide important temperature information from the upper- troposphere to the middle stratosphere, spanning the pressure interval from around 200 to 5 hPa. These measurements complement the information provided by satellite radiances because they can be assimilated without bias correction to the forecast model, and they have good vertical resolution because of their limb geometry. The importance of GNSS-RO in constraining biases in stratospheric temperature reanalyses is now well established (Long et al., 2017; Simmons et al., 2020), and they can also be used for verification and testing (e.g., testing the weak constraint 4D-Var configurations Laloyaux et al. (2020)).

The period 2019-2020 has seen a significant increase in the $\sim$ 3000 GNSS-RO measurements per day

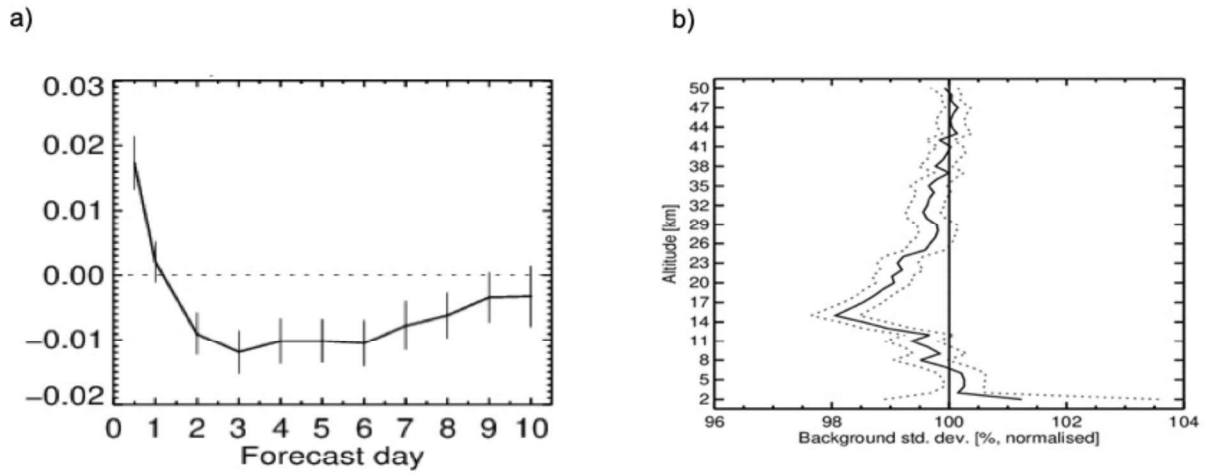

Figure 19: Aeolus impact from an Observing System Experiment for the period 2 August 2019 to 31 December 2019 using IFS Cy46r1 12-h 4D-Var with TCo399 outer loop. a) Normalised change in the root mean square error in vector wind at 50 hPa in the tropics from assimilating Aeolus (Rayleigh-clear and Mie-cloudy) data. Verified against operational analyses, showing the 95% confidence ranges. Negative values indicate a reduction in error from assimilating Aeolus data. b) Change in standard deviation of O-B departures relative to GNSS-RO resulting from assimilating Aeolus, normalised so that the control is 100%. Values below 100% show an improved fit from assimilating Aeolus and above show a degraded fit.

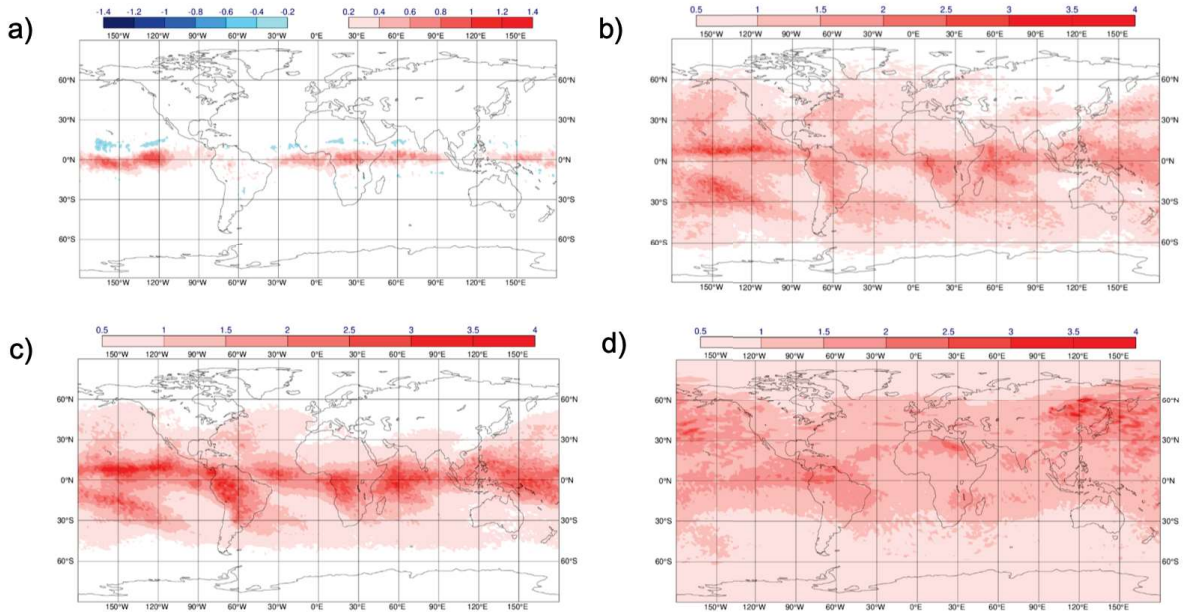

Figure 20: Impact from assimilating Aeolus for the period 2 August to 27 December 2019. a) The mean change in the u-wind analysis (m/s) at 70 hPa. b) Standard deviation of the change in the u-wind analysis (m/s) at 250 hPa. c) Standard deviation of the change in the u-wind analysis (m/s) at 150 hPa. d) Standard deviation of the change in the u-wind analysis (m/s) at 1 hPa.

available for operational applications, with the launch of the joint Taiwan-United States Formosat-7/COSMIC-2 mission (COSMIC-2 hereafter). This constellation of six satellites provides around 4000-5000 more GNSS-RO measurements in the latitude band between $\pm$ 40°. COSMIC-2 was launched on 25 June 2019, declared operational on 25 February 2020, and ECMWF began active assimilation on 25 March 2020 following 6 months of testing with Cy46r1 at TCo639. The assimilation of the COSMIC-2 measurements improves the short-range forecast departure statistics for almost all satellite and in-situ observation types in the tropics, including measurements sensitive to tropospheric humidity. As expected, the COSMIC-2 measurements have a large impact on stratospheric temperatures, reducing both random and systematic errors. A particularly interesting result, given that ECMWF does not currently use an explicit mass-wind balance operator in the stratosphere, is the strong COSMIC-2 impact on the tropical stratospheric winds. Figure 21 shows the improvement in the standard deviation of the short-range forecast and analysis vector wind departures, as a result of assimilating COSMIC-2. The improvements are largest above 200 hPa, but are significant down to 700 hPa. We also find that COSMIC-2 improves the fit to the Aeolus winds. The improvements in the vector winds extend out to day-10 for 30 hPa and 10 hPa, as shown in Fig. 22, where verification is against observations. Note that similar results are obtained when verifying against operational analyses. These results illustrate that the current 4D-Var configuration is retrieving useful tropical stratospheric wind information from mass measurements, without applying the mass-wind balance operator. This is likely a result of the model integration step in the 4D-Var assimilation window, but more work is required to isolate the mechanism. It will be interesting to revisit these COSMIC-2 results if the balance operator is reintroduced in the stratosphere.

More generally, we have also found that zonally averaged GNSS-RO geopotential height climatologies provide useful zonal wind information (Healy et al., 2020). Figure 23 shows the time series of monthly mean zonal-mean zonal winds at 50, 30 and 10 hPa in the $\pm$5° latitude band. The Metop curve is computed by taking the second derivative of the Metop GRAS geopotential height climatology in the meridional direction (Fleming and Chandra, 1989; Scaife et al., 2000). These Metop "balanced" zonal wind estimates are compared with radiosonde climatology produced by Free University Berlin (Naujokat, 1986), ERA5 zonal winds and the balanced winds computed from the ERA5 geopotential heights. These results show that GNSS-RO can provide useful climatological information in the equatorial stratosphere, and it suggests that GNSS-RO climatologies should be useful independent dataset for model testing. A free running model that can reproduce the monthly mean GNSS-RO temperature/geopotential height climatologies should also produce a reasonable QBO.

4.3 Key summary points

- Aeolus wind observations and COSMIC-2 temperature observations have a positive impact on tropical wind analysis and medium-range forecasts in the stratosphere. Both observation types have good vertical resolution that better constrains the QBO. While COSMIC-2 data does not directly constrain the winds, its effect on the wind analysis can be understood via equatorial mass-wind balance.

- There are clear gaps in available long-term global middle atmosphere observations that would not only improve the initial conditions (and therefore forecasts) but also provide a good reference to validate model against. These include: i) humidity observations (see also sections 2.2.3 and 2.6); ii) winds observations (both horizontal and vertical); iii) good vertical ($\sim$1 km) and horizontal ($\sim$50 km) resolution radiance observations that would extend to 80 km altitude and would help to constrain temperature as well as GWs (as temperature perturbations are proportional to GW momentum fluxes), and iv) ozone observations.

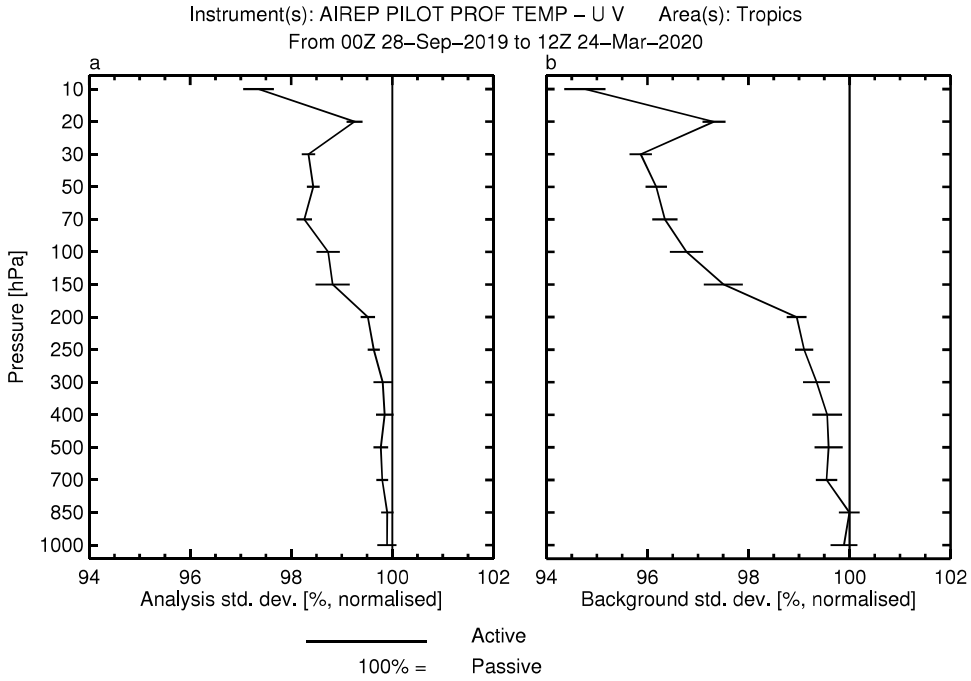

Figure 21: The improvement in the standard deviation of the analysis (left) and short-range forecast (right) vector wind departures as a result of assimilating COSMIC-2 GNSS-RO measurements. The period is 28 September 2019 to 24 March 2020. Values below 100% show that COSMIC-2 is improving the fit to this data.

5 Examples of the influence of the stratosphere on the mid-latitude troposphere

Quantification of the stratospheric influence on the mid-latitude tropospheric weather patterns is essential in order to determine how much effort should be spent in the future on improving the stratosphere at ECMWF. As discussed in the introduction, stratospheric influence on the troposphere occurs only during "windows of opportunity" (i.e., SSWs, strong vortex events, final warmings etc.) and on extended-range and seasonal timescales. The large ensemble (51 members) of IFS seasonal hindcasts from 1981 to the present day allows for a unique opportunity to i) test for sampling uncertainty of the coupling between the stratosphere and the troposphere found in reanalysis data; ii) to assess the predictability of the extratropical troposphere resulting from the stratosphere during the "windows of opportunity" (Vitart et al., 2012); and iii) to tease out stratospheric influence on the mid-latitude tropospheric weather regimes as opposed to other remote events such as ENSO, MJO, Indian Ocean Dipole (IOD), snow cover anomalies etc. For iii) statistically robust conclusions from a relatively short reanalysis record are difficult to draw because conditioning out confounding influences leaves a small sample. In addition to the use of a large hindcast ensemble, ensemble forecasts where the meteorological fields (vorticity, divergence and temperature) are nudged to the reanalysis in the stratosphere can be helpful in determining the influence of the stratosphere on the troposphere. The purpose of this section is threefold: 1) to illustrate how a simulation of selected processes in the middle atmosphere can have an impact on stratosphere-troposphere coupling; 2) to show two clear examples where the stratosphere matters for tropospheric predictability at ECMWF; and 3) to illustrate a successful use of the above approaches at ECMWF. The aim here is not to review a large body of past and ongoing collaborative work at ECMWF in the area of extended-range prediction involving the stratosphere as this topic is too vast to do it proper justice here (but see e.g., Vitart et al., 2019; Domeisen et al., 2020a,b; Charlton-Perez et al., 2019; Büeler et al., 2020).

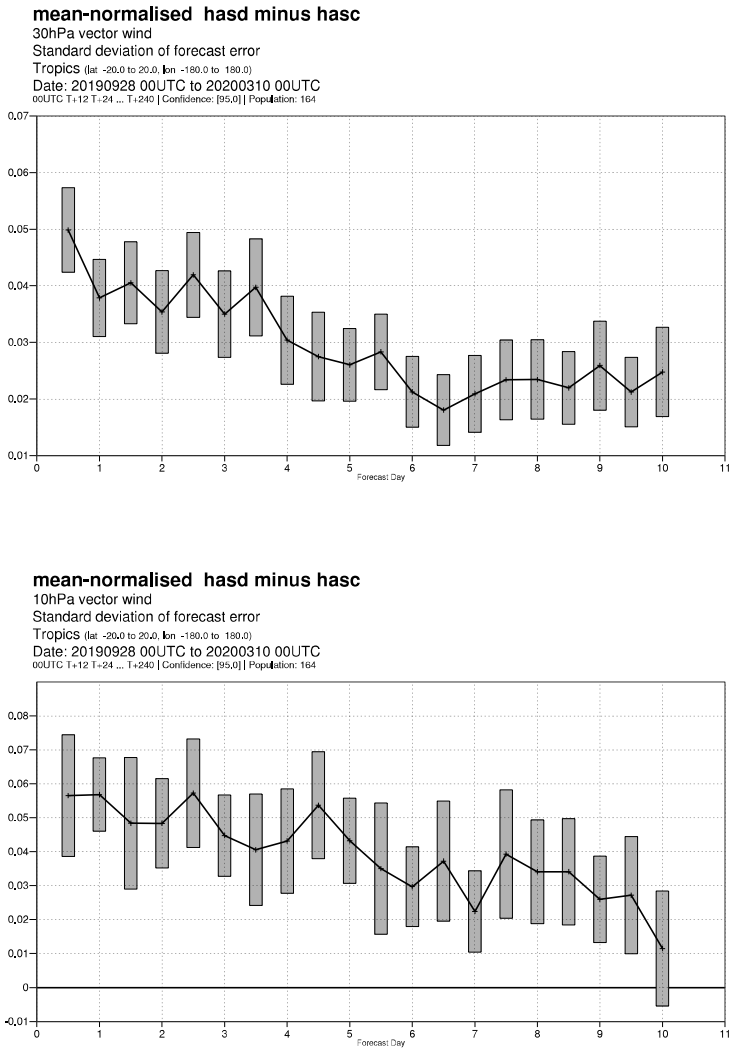

Figure 22: The percentage reduction in the 30 hPa and 10 hPa vector wind forecast errors in the tropics. Above 0 indicates that COSMIC-2 is improving the error statistics. The error bars represent the 95% confidence interval.

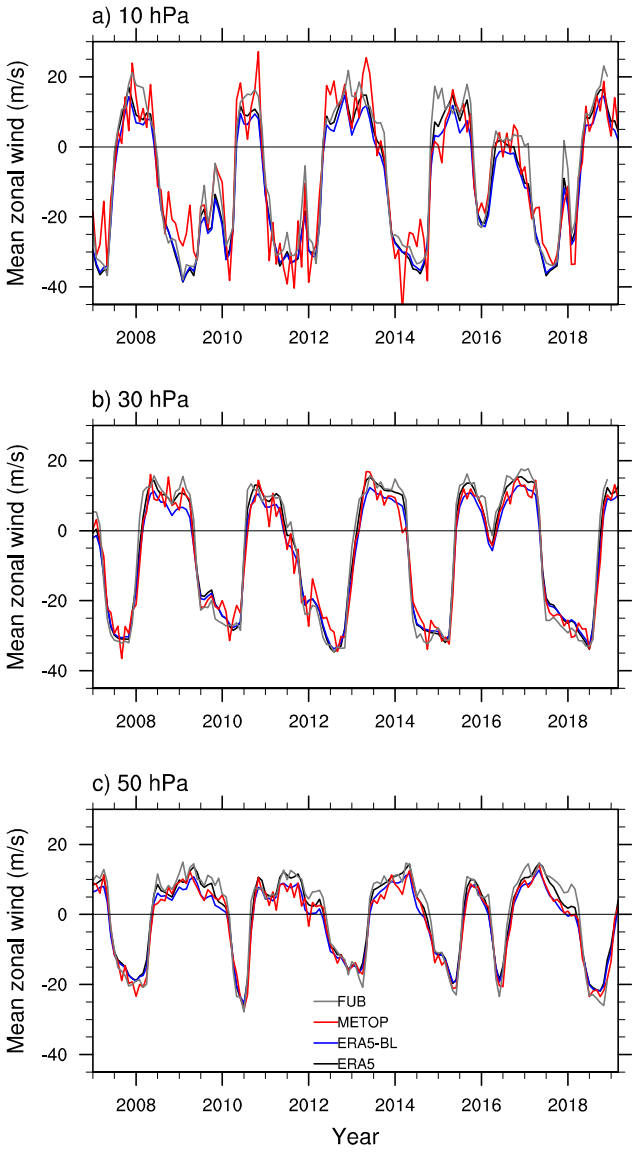

Figure 23: The timeseries of the monthly-mean zonal winds at 10 hPa, 30 hPa and 50 hPa for the Free University Berlin Singapore radiosonde measurements (grey line), the Metop balanced winds computed with the second derivative of the geopotential height fields in the meridional direction (red line), the ERA5 zonal winds (black line) and the ERA5 balanced zonal winds (blue line).

5.1 Impact of parametrized nonorographic gravity wave drag on stratosphere-troposphere coupling

As NOGWD has a strong impact on polar vortex variability (section 2.4.4), it also influences the stratosphere-troposphere dynamical coupling as was explored in Polichtchouk et al. (2018) with 32 years of 4-year free running IFS simulations at TL255L137 resolution. In the NH the "window of opportunity" for stratospheric influence occurs after the onset of SSWs when the tropospheric eddy-driven jet (EDJ) tends to shift equatorward projecting onto a negative phase of the Annular Mode. In the SH, the stratospheric polar vortex variability is mainly associated with interannual variability in the seasonal cycle leading up to the springtime polar vortex break-down (i.e. the final warming): Anomalously strong polar vortex years, which tend to have an anomalously late final warming, are associated with a stronger poleward shift of the SH EDJ in September-November and a delay in the equatorward EDJ shift in early summer (Byrne et al., 2017; Hio and Yoden, 2005; Seviour et al., 2014; Byrne et al., 2019). It was found that IFS is able to represent the above described dynamical coupling in both hemispheres (cf. a to b in Figs. 24 and 25), and that changes in NOGWD strength have a substantial impact on the EDJ, but the sense of the impact is opposite in the two hemispheres. In the NH the impact occurs via changes in the amplitude and persistence of stratospheric anomalies following SSWs; by increasing NOGWD persistence and amplitude of stratospheric anomalies reduces and therefore the strength of stratosphere-troposphere coupling is reduced. This is shown in Fig. 24 with composites of Northern Annular Mode (NAM) index centered on the SSW central date (i.e., the date when zonal-mean zonal wind at 10 hPa and 60°N changes from westerly to easterly). In the SH, the impact occurs via differences in the sensitivity of the EDJ to a given stratospheric anomaly arising from changes in the seasonal cycle leading up to the polar vortex breakdown; therefore by increasing NOGWD and by bringing final warming date forward the strength of the stratosphere-troposphere coupling is increased.

5.2 Predictability of SH eddy-driven jet during austral spring and early summer

As discussed above, the stratosphere-troposphere dynamical coupling in the SH arises mainly from the interannual variability in the timing of the final warmings. The timing influences the interannual variability in the eddy-driven jet (EDJ) and therefore storm track position; in early final warming years equatorward seasonal transition of the EDJ is accelerated (Byrne et al., 2017). Thus the timing of final warmings can give information about extratropical tropospheric circulation later on in the season (from August to January). Using a large ensemble of hindcasts from ECMWF System 4 (Molteni et al., 2011) initialized on 1 August and 1 November, Byrne et al. (2019) investigated the robustness of the above relationship to sampling variability by classifying each year and ensemble member as either having anomalously early or anomalously late final warming date. It was found that the relationship between polar vortex and EDJ in hindcasts is in good agreement with that found in reanalysis, and that this relationship is robust to sampling variability. Forecast skill of the monthly- and weekly-mean tropospheric EDJ was then assessed. A subset of the results is shown in Fig. 26. Perhaps most striking is the re-emergence of monthly-mean forecast skill in the troposphere for forecasts initialised on 1 August. This re-emergence of skill is consistent with an influence from the polar vortex descent on the latitude of the EDJ. The results suggest that ∼20-30% of monthly tropospheric jet stream variability at lead times of 3-4 months is predictable based on knowledge of the stratospheric state. Similarly, the 1 November forecasts show results consistent with an influence from the polar vortex breakdown event on the EDJ. They suggest that ∼20-30% of weekly tropospheric jet-stream variability at week 3 and week 4 is predictable based on knowledge of the stratospheric state.

Previous studies (e.g., L'Heureux and Thompson, 2006) have also suggested a similar relationship be-

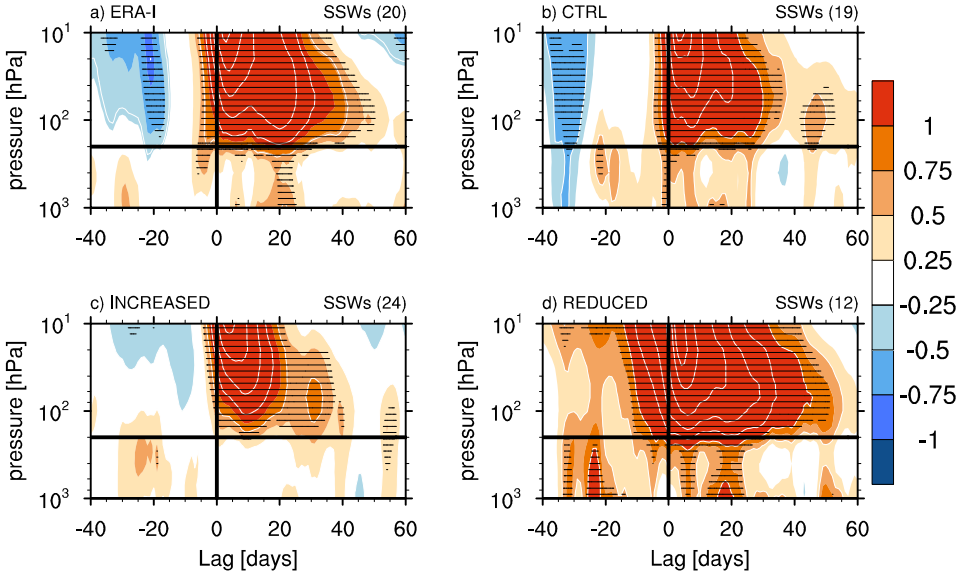

Figure 24: Composites of the Northern Annular Mode index (unit: standard deviation) centered on the SSW central date in (a) ERA-Interim, (b) control, (c) increased, and (d) reduced NOGWD run. Shading interval is 0.25 standard deviations, and contour interval is 0.5 standard deviations. The number in parentheses above the panels shows the number of SSWs in each case. Hatching means different from 0 at 5% significance level. The thick black vertical line shows the central date, and the horizontal line the 200 hPa level. Note that the exact agreement between (a) and (b) is not expected as the forcing employed in CTRL does not match that of the ERA-Interim period. Figure from Polichtchouk et al. (2018).

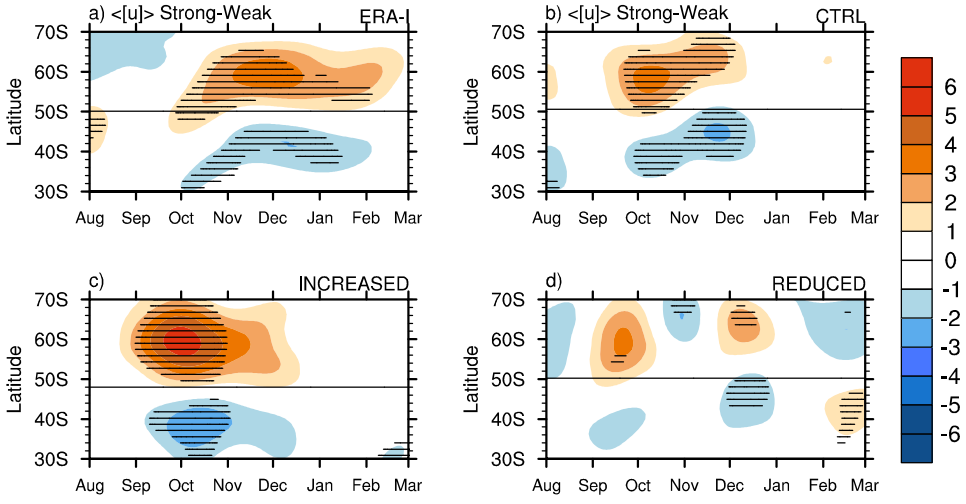

Figure 25: Difference in zonally and vertically averaged zonal wind in the troposphere (shading, in m/s) between the 11 weakest and 11 strongest stratospheric polar vortex years for (a) ERA-Interim, (b) control, (c) increased, and (d) reduced NOGWD runs. Hatching marks where the differences are significance at the 5% level. The horizontal line shows the annual mean eddy-driven jet latitude averaged over all years for each case. Note that the exact agreement between (a) and (b) is not expected as the forcing employed in CTRL does not match that of the ERA-Interim period. Figure from Polichtchouk et al. (2018).

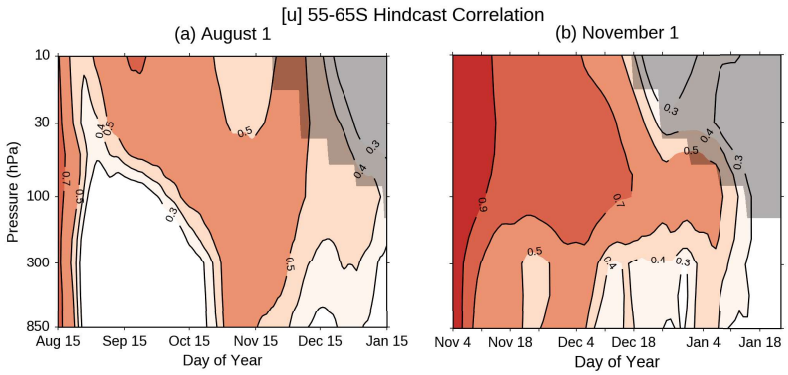

Figure 26: Forecast skill by lead time and pressure level from System 4 seasonal hindcasts. The left-hand panel shows the correlation between 31-day mean ensemble mean winds in the east-west direction from 55°S to 65°S and corresponding values in ERA-Interim as a function of calendar day and pressure level for forecasts starting on 1 August. Values on the x-axis represent the central date of the 31-day mean. The right-hand panel shows the same but for 7-day means for forecasts starting on 1 November. In both panels, shaded contour regions are statistically significant at the 95% level. Figure from Byrne et al. (2019).

tween the EDJ and ENSO in DJFM, whereby in strong El Nino years the EDJ shifts poleward in the zonal mean. System 4 ensemble was used in Byrne et al. (2019) to probe the ENSO–stratosphere–EDJ link. It was found that after removal of stratospheric pathway – by conditioning on either early or late final warming years — the ENSO–EDJ teleconnection largely vanished. The conclusion from this study was that ENSO appears to impact the stratosphere, which in turn impacts tropospheric EDJ. It should be noted that this result relates only to the high-latitude zonal-mean response to ENSO; it does not relate to any potential high-latitude zonally asymmetric response to ENSO arising from Rossby wave response to tropical heating anomalies due to ENSO.

The above described work revealed an important SH stratosphere model bias introduced in the ECMWF's new seasonal forecasting system SEAS5 (Johnson et al., 2019). It was found that the maximum interannual stratospheric variability is too weak and occurs too early in SEAS5 in comparison to reanalysis (see Fig. 27). Such bias was not present in System 4 and is possibly related to the use of ozone climatology in radiation in SEAS5 compared to radiatively interactive prognostic Cariolle ozone scheme (Cariolle and Déqué, 1986) used in System 4. In SEAS5 prognostic Monge-Sanz ozone scheme (BMS scheme, Monge-Sanz et al., 2011) is used, but it is not radiatively interactive. However, SEAS5 experiments with radiatively active BMS scheme do show a reduction in this bias (Fig. 27b) as now the true interannual variability is fully contained in the model's 25-75% percentile. As the timing of the polar vortex breakdown has an impact on the EDJ in late spring/summer season, this bias is potentially linked to larger EDJ bias in austral spring and summer season in SEAS5 than in System 4 (not shown), and the inability of stratospheric anomalies to couple to troposphere in SEAS5 as opposed to System 4 (see Fig. 14 in Shepherd et al., 2018). By nudging the SEAS5 hindcasts above 70 hPa to the observed stratospheric variability, the impact of removing this bias on the EDJ bias can be quantified in Fig. 28 (note that the radiatively interactive BMS scheme performs worse in the troposphere than ozone climatology; therefore radiatively interactive BMS experiments can not be used to cleanly isolate the impact of the stratospheric bias on the EDJ). As expected, the EDJ bias in the spring/summer season in SEAS5 is markedly reduced as a result of alleviating the stratospheric variability bias. This is another example, together with the LMS polar cold bias, where fixing a stratospheric bias improves the model behaviour in the troposphere on S2S timescales.

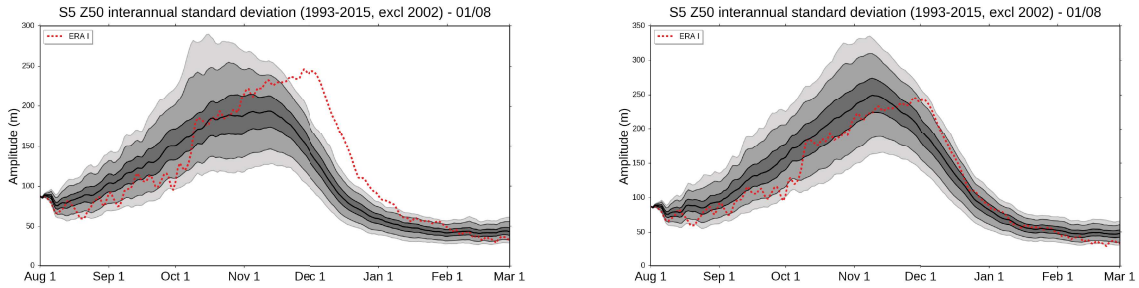

Figure 27: Left: Bootstrap estimate of sampling uncertainty associated with 34-year mean of polar-cap averaged geopotential height standard deviation at 50 hPa using ensemble members from the August 1 initialisations with SEAS5. The bootstrap estimate was generated using 10000 time series of length 34 where an ensemble member has been randomly selected from each year in the 34-year period excluding the year of 2002. Solid lines represent 1, 5, 25, 75, 95 and 99% thresholds respectively. Dashed red line indicates jet latitude from ERA-Interim. Right: Same as (left) but for SEAS5 experiments with radiatively interactive BMS ozone scheme.

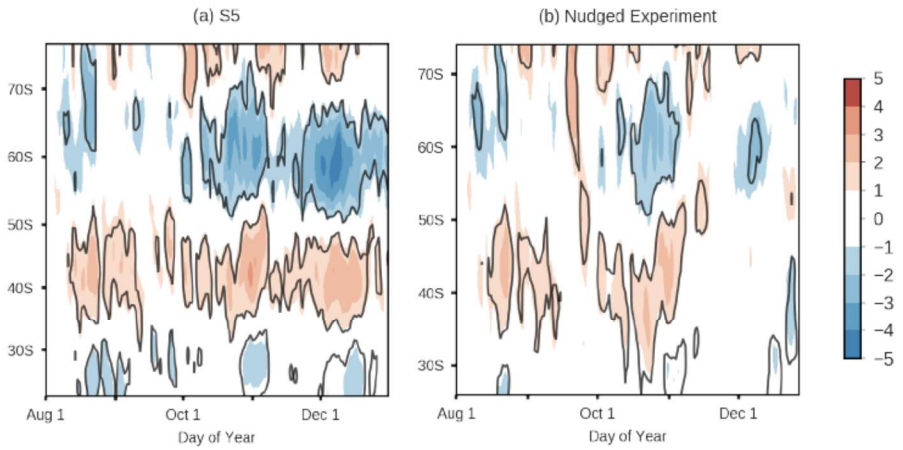

Figure 28: Left: Estimate of zonal wind bias at 500 hPa when verified against ERA-Interim using ensemble members from the August 1 initialisations with SEAS5. Right: Same as (left) but for SEAS5 experiments with stratosphere nudged to ERA-Interim. Note the reduction in biases as a result of correct stratospheric evolution.

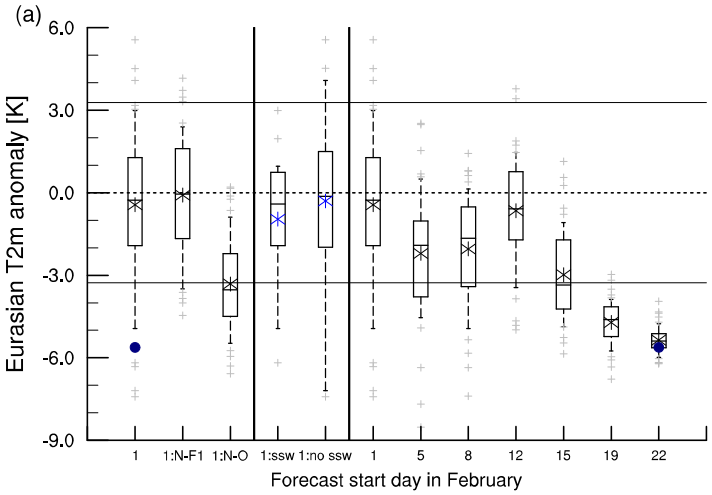

Figure 29: Boxplots of the Eurasian T2m anomaly (K), averaged between February 25 and March 6, 2018 for different initialization dates of the extended-range forecasts. The whiskers show the 10th and 90th percentiles, while the outliers are grey crosses. The boxes labelled "1:N-F1" and "1:N-O" represent the nudged forecasts initialized on February 1, where the stratosphere above 70 hPa is nudged to February 1, 2018 and to the reanalysis, respectively. The box labelled "1:ssw" shows only those members that predicted SSW (within ± 3 days from the central date) in the February 1 forecasts and "1:no ssw" shows those that did not (18 and 23 members, respectively). The solid black horizontal lines show ± one standard deviation obtained from ERA-Interim daily northern Eurasian T2m anomalies between January and March 1979-2018. Blue dots show the observed values. Figure from Kautz et al. (2020).*

5.3 Impact of 2018 SSW on the late winter Eurasian cold spell

Most of the evidence for stratospheric influence on tropospheric weather regimes such as the North-Atlantic Oscillation (NAO), NAM and Southern Annular Mode (SAM) comes from various types of composite analyses. However, as stated in a recent overview article by Butler et al. (2019), "challenges remain in arriving at a set of general unifying principles that could provide a quantitative description of the role of the stratosphere on an event-by-event basis". Moreover, only about two thirds of all SSWs in the reanalysis tend to exhibit a canonical impact of negative NAO (e.g. Domeisen, 2019), with those with extended recovery periods (i.e., longer persistence) exhibiting downward impact (e.g. Hitchcock et al., 2013). Motivated by the above, as part of a member state collaboration (FMI and the University of Reading) Karpechko et al. (2018), and (KIT and LMU) Kautz* et al. (2020) investigated the potential of the stratosphere to enhance extended-range predictability of the late February to early March 2018 extreme Eurasian cold spell, which was associated with a shift of the NAO to a negative phase and followed on from a major SSW event on 12 February 2018. This event preceded a detectable downward NAM signal from the stratosphere into the troposphere (Ayarzagüena et al., 2018).

By analyzing the extended range ECMWF forecast ensemble for different start dates close to the event, Kautz* et al. (2020) found that the extreme cold spell (i.e. 2σ-event) was predicted with the observed strength $\sim$10 days in advance (see Fig. 29). This is consistent with the assumed deterministic predictability limit of the NAO (e.g. Scaife et al., 2014; Ferranti et al., 2015) and suggests that ultimately the extreme cold spell was a result of internal tropospheric synoptic-scale dynamics. However, the probability of the cold spell doubled up to 25 days in advance, when the SSW was accurately predicted in all ensemble members for forecasts initialized from February 5 onwards (see all the boxes to the right of the second vertical line in Fig. 29). In forecasts initialized on February 1, ensemble members that

accurately predicted the onset of the SSW showed slightly enhanced (but not statistically significant) likelihood for the cold spell compared to members that completely failed to predict the SSW (see the boxes in between the first and the second vertical lines in Fig. 29). To allow for a more unambiguous conclusion, dedicated ensemble (50-members) experiments where the stratosphere was nudged (above 70 hPa) to i) the observed evolution, and to ii) a time-invariant state close to climatology were performed for February 1 initialization. Based on these experiments, it was possible to show that the probability of an extreme cold spell was enhanced to 45% by the correct stratospheric evolution, whereas the skill completely disappeared (i.e., was at a climatological value of 5%) when the stratosphere was nudged to a time-invariant state (see boxes to the left of the first vertical line in Fig. 29). This suggests that it is not the occurrence of SSW itself in the forecast, but the subsequent evolution throughout the lower stratosphere that is crucial in coupling to the troposphere. Moreover, it can be seen from Fig. 29 that the forecasts with stratosphere nudged to the observed evolution predicted the cold spell with a higher probability than the subsequent initialized forecasts until February 15. This suggests that tropospheric initial conditions only became important from February 19 onwards and that the slightly enhanced skill in February 15 forecasts is due to the correct stratospheric evolution.

5.4 Key summary points

- Parametrized NOGWD not only influences internal stratospheric variability, but exerts control on stratosphere-troposphere coupling in both hemispheres on extended range timescales.
- Conditional probabilities applied on large hindcast ensemble, and the nudging technique can be effectively used to ascertain stratospheric influence on the troposphere. While only two specific examples are considered here, these techniques can be used to probe other aspects of stratosphere-troposphere coupling on extended-range time scales in ECMWF system, including (but not limited to): SSWs to surface impact (e.g., Baldwin and Dunkerton, 2001; Sigmond et al., 2013; Hitchcock and Simpson, 2014); QBO to polar vortex to surface teleconnection (e.g., Gray et al., 2018); NH final warmings to surface impact (Butler et al., 2019); the relative role of the Indian Ocean Dipole, ENSO and the polar vortex anomaly on the surface impact in the SH (e.g., Lim et al., 2020).
- Lower stratospheric variability following SSWs (rather than the occurrence of SSW itself) appears important for surface response (this is also the conclusion of several previous studies; e.g., Hitchcock et al., 2013; Karpechko et al., 2017). This merits a closer inspection in models and may provide a better metric to assess the influence of the stratosphere on the troposphere than the less discriminating but often used zonal-mean zonal wind at 10 hPa and 60°N metric.
- Inter-annual stratospheric variability and the associated EDJ shift in the SH is important for an accurate tropospheric response (section 5.2, Figs. 27 and 28).

6 Future developments relevant for the middle atmosphere

Considering the next four year planning horizon, plans are presented together with potential future directions beyond. This is followed by key questions and planned diagnostics for evaluating stratospheric processes and stratosphere-troposphere coupling. This section is partly based on the results already presented and partly on the working group discussions with expert scientists during the ECMWF workshop on "Stratospheric predictability and impact on the troposphere" held in November 2019 (https://www.ecmwf.int/en/learning/workshops/workshop-stratospheric-predictability-impact-troposphere).

6.1 Model development

6.1.1 Plans

- While the interactive hybrid linear ozone (HLO) scheme has been successfully implemented in CAMS, testing of the radiatively interactive HLO scheme for an operational implementation is ongoing. In CAMS currently a different ozone chemistry scheme is used in the troposphere (CB05) and HLO is used in the stratosphere only. Using the HLO throughout the troposphere and stratosphere shows improved stratospheric variability forecast performance in medium-range and (single ensemble member) seasonal forecasts, but increases bias 2) (cf. Fig. 1). It is hoped that the radiatively interactive HLO scheme improves on the SH stratospheric inter-annual variability bias present in SEAS5 (section 5.2). This work is part of an effort to further unify the treatment of ozone across different ECMWF applications.

- It is planned to explore the use of CAMS methane fields to improve the methane oxidation parametrization for a more accurate representation of stratospheric water vapour in the IFS (section 2.6).

- It is planned to port the radiation tool "ecCKD" (with better gas optics) that would reduce the global-mean warm bias 1) in Fig. 1 through an improved solar absorption spectrum (section 2.2.1). The "ecCKD" will be tested in conjunction with the SLVF vertical filter (section 2.2.2) to make sure the cold bias 2) (cf. Fig. 1) is not increased.

- The clear benefit of the reduced vertical diffusion to both the analysis and forecasts of the equatorial winds (Figs. 11 to 13) suggests this is highly desirable. The current testing has relied on specifying a fixed vertical level threshold above which reduced diffusion is applied, which likely leads to the temperature forecast skill degradation in the 100 hPa region. Tests where the vertical diffusion in stable conditions is reduced at all latitudes above a time-varying tropopause only are planned. Whether the reduction of the QBO bias in the seasonal to sub-seasonal forecasts as a result of reduced vertical diffusion will improve stratosphere-troposphere teleconnection and therefore tropospheric skill remains to be evaluated. In parallel, there is an effort — as part of a collaboration with Météo-France — to explore the use of the TKE scheme instead of the first-order turbulence closure scheme that the current vertical diffusion parametrization is based on. Thus far, TKE shows problems in stable boundary layers and the UTLS, both of which experience strong stratification and large gradients in static stability that influence turbulent motions and in turn can impact net diffusion on momentum and tracers.

- Efforts to make the sponge layer shallower and less intrusive on the dynamics of the stratosphere (section 2.4.1), while working with 4D-Var (section 3.3), are ongoing. A way forward is to apply the deep sponge in the minimization only, while reducing the depth of the sponge in the trajectory and non-linear model integrations. To what extent the new less intrusive sponge design improves stratospheric variability or stratosphere-troposphere coupling can be evaluated using extended-range or seasonal setup. The use of shallower sponge in the trajectory and model is likely to also increase the spread in the EDA. This might help the analysis over large-amplitude orographic gravity wave dominated regions over Southern Andes and Antarctic peninsula (section 2.4.2).

- Recent developments introduced in the IFS semi-Lagrangian advection scheme, such as using higher order interpolation for the winds during departure point iterations and using high-order Runge-Kutta schemes to compute these points may be advantageous for the stratospheric performance and therefore will be investigated further. Improvements introduced in the past to the

semi-Lagrangian advection scheme resulted in large improvements in sudden stratospheric warming forecasts from Cy41r1 (Diamantakis, 2014). Additionally, resolution of convergence issues in the iterative scheme used to determine the location of the semi-Lagrangian departure points (Diamantakis and Magnusson, 2016), prompted by an increase of horizontal resolution, resulted in further improvements in the stratosphere, in regions over steep orography and locally over tropical cyclones.

- The design of the dynamical core — as discussed in section 2.2.2 — plays an important role in the accurate representation of the stratosphere. The development of an alternative flux-form finite volume dynamical core at ECMWF (IFS-FVM, Kühnlein et al., 2019) offers a unique opportunity to assess the impact of a different dynamical core design on the representation of the stratosphere whilst keeping the physical parametrizations unchanged. It is therefore planned to evaluate the impact of using the FVM dynamical core on the biases sketched in Fig. 1.

6.1.2 Future directions

- The expert consensus is that no known influences on the stratosphere — and therefore the troposphere — exist from processes that occur above the altitude of z=65 km. But to adequately model the middle atmosphere up to z=65 km (such that it is not affected by a sponge layer and/or too coarse vertical resolution), extension of the model top and the increase in the number of layers is required. To avoid getting into the region where non-local thermodynamic equilibrium effects are important but to still get above the mesospheric jets, placing the model lid at z$\sim$110 km could be investigated. To what extent the higher model top enhances predictability can be investigated in sub-seasonal to seasonal forecasts. It is clear, however, that as the model top increases and the sponge layer is pushed to higher altitudes, the amplitude of GWs will increase (as the amplitude increases with decreasing density). This — as already discussed in section 3.3 — will unearth many stability problems and larger discrepancies between the non-linear and tangent linear model used in 4D-Var. Moreover, the highest assimilated observations at ECMWF today come from highest peaking AMSU-A and ATMS observations, which peak at 2.5 hPa. Therefore, to constrain analysis in the mesosphere assimilation of higher probing satellites would be required. The development of non-hydrostatic dynamical cores will put further constraints but also opportunities for the sponge design and more generally for the dynamics of the stratosphere (e.g., Wedi and Smolarkiewicz, 2006).

- The results from the ISSI work (section 2.4.2) and from the COORDE project (van Niekerk et al., 2020) indicate that the parametrized GW drag is still needed at horizontal resolutions of 9 km, despite a large part of the GW spectrum being resolved. It is, however, interesting to note that NOGWD dominates over regions where orographic GWs should dominate (ISSI model intercomparison). This is likely the consequence of assuming column-wise propagation only in the parametrizations: If all the momentum, distributed in reality over a large area due to lateral propagation of GWs (Polichtchouk and Scott, 2020; Dunkerton, 1984; Sato et al., 2012; Polichtchouk and Scott, 2020), is deposited over few model grid points over orography only, the model is likely to experience instability. Due to its prescribed broad phase speed spectrum, NOGWD acts over a broader area in comparison to OGWD. Thus it is likely that most of the required parametrized drag ends up in NOGWD and not OGWD. Understanding what this implies for the middle atmosphere circulation needs to be assessed.

- Stratospheric aerosols are a potential source of forecast degradation on seasonal timescales if not represented accurately after major volcanic eruptions. For example, in the present implementation

of volcanic aerosol in the IFS seasonal system, the vertical distribution of aerosols is fixed. Moreover, the lack of mass conservation by the IFS dynamical core often poses problems for aerosol plumes with large gradients. Therefore more work needs to be done to address these shortcomings.

6.2 4D-Var

6.2.1 Plans

- The developments for 4D-Var described in section 3.2 have brought into sharper focus the related issues of initialisation and stratospheric balance in the ECMWF analysis.

  In terms of initialisation, the most recent changes in the digital filter algorithm used for initialisation of the IFS analysis (Jc-DFI, Gauthier and Thepaut, 2001) were introduced in IFS Cy38r1 as a pragmatic response to problems in the assimilation of surface pressure observations caused by stratospheric instabilities (Fisher, 2012). As shown in Fig. 17, the timestep matching 4D-Var drastically improves the model fit to surface pressure observations and also the initial balance of the analysed fields. It thus makes sense to re-evaluate the utility of Jc-DFI under these changes.

  The balance operators of the Jb term of the cost function were deactivated in the stratosphere in IFS Cy35r3 (September 2009), due to concerns about the noise level present in the analysed fields. In light of the clearer understanding of the origin of the stratospheric noise in the analysis and the concurrent improvements in the weak-constraint part of the 4D-Var cost function (section 3.1), current research is now directed at re-introducing linearised mass-wind balances in the stratosphere in the 4D-Var analysis. The expectation is that this will bring further tangible benefits in the representation of the stratosphere, as in the current global observing system there are very few direct wind observations available to constrain the stratospheric wind field. This is supported by the result (section 4.2), where the monthly QBO winds can be well constrained by the GNSS-RO geopotential observations via equatorial thermal wind balance.

- Work is ongoing to exploit weak-constraint 4D-Var and reduce model biases in the reanalysis. The idea is to use the model bias estimated over a recent well-observed period as a forcing throughout the whole century.

6.2.2 Future directions

- The new model error diagnostics developed to compare the first-guess trajectory with GNSS-RO measurements (section 3.1) could be utilised more generally to rapidly assess model changes affecting the stratosphere in the data assimilation context (and correspondingly modified correlation and error characteristics).

- There is a mathematical similarity between 4D-Var and deep learning methods (Geer, 2020). Therefore, it might be possible to use the science developed under deep learning to extend the weak constraint concept.

6.3 Observations

6.3.1 Plans

- Of great interest is the number of High Altitude Platforms, such as google loon, which has a

potential to significantly increase our ability to monitor the stratosphere. A study to assess the impact of google loon data on analysis and forecasts is planned. Such observations can provide new information on stratospheric wind and temperature, and possibly other variables.

- It was shown in Žagar et al. (2016) that tropical unbalanced modes are associated with significant analysis uncertainties and forecast errors in IFS. Given that both Aeolus and COSMIC-2 data should be better able to constrain the tropical wave modes (due to their good vertical resolution), an analysis similar to that reported in Žagar et al. (2016) is planned with Aeolus and COSMIC-2 observations assimilated to see if the analysis uncertainties and forecast errors due to tropical wave modes reduce.

6.3.2 Future directions

- In section 4, the recent WIGOS evolution to include more GNSS-RO observations (COSMIC-2), as well as Aeolus wind observations, has been very important.

  WMO's Vision 2040 provides a blueprint for a future observing system that will more fully meet the requirements. Alongside increased numbers of GNSS-RO observations and an operational Doppler Wind Lidar, the need for more limb view geometry radiances is noted, in particular infrared and microwave limb geometry building on the successful MIPAS and MLS missions respectively. This will provide full spatial coverage, resolving wave structures that may be difficult to diagnose with a sparser data coverage.

  It would be highly desirable to encourage space agencies to consider providing an upper stratosphere/lower mesosphere nadir viewing mission like SSMIS again, especially now the capability to use this data is enhanced, both through suitable observation operators and model capability. However at the time of writing, prospects for this appear bleak.

- In light of the results discussed in section 2.2.3, the possibility of assimilating sparse water vapour measurements to anchor the analysis in the stratosphere could be explored. ACE-FTS limb sounder would provide high vertical resolution (1 km) but limited sampling.

6.4 Key questions: stratospheric processes and stratosphere-troposphere coupling

6.4.1 Wave characterization

- Observations of GWs, such as from 3D high resolution AIRS retrievals, super pressure balloon, lidar and radar data could be used more for process improvement of sources, which are very uncertain. This is especially so for the coupling of GWs and convection, which also modulates the QBO. Earlier 3D analysis of resolved GWs in short ECMWF forecasts (Preusse et al., 2014) revealed that short-wavelength convectively generated GWs are under-estimated but long-wavelength GWs overestimated. Since then the vertical and horizontal resolution of ECMWF model has increased (from TL799L91 to TCo1279L137) and therefore a repeat of the 3D analysis presented in Preusse et al. (2014) is planned in collaboration with Juelich. The evaluation of orographic GWs in IFS and their comparison to observations is ongoing as part of the ISSI international project.

  The availability of 1.4 km global IFS simulations as part of the Horizon 2020 INCITE project, provides an exciting opportunity to assess the impact of resolved deep convection on GWs. In particular, it is known that the ICON model at 5 km horizontal resolution underestimates GW momentum fluxes when run with the standard convection parametrization (which is the same as in

the IFS Cy47r1), but overestimates the fluxes when run without deep convection parametrization (Stephan et al., 2019). It is likely that at 5 km resolution deep convection is still not well-resolved and is occurring on too large a scales, therefore unduly exciting GWs of too large a scale. This artificially amplifies energy in the divergent part of the kinetic energy spectrum, as is seen in IFS simulations when deep convection is switched off at 9 km and 5 km horizontal resolutions. By comparing GW momentum fluxes from 1.4 km IFS simulations where deep convection is resolved to momentum fluxes from 5 km IFS simulations with and without parametrized convection, this can be quantified. Moreover, the availability of the new high-resolution 1.4 km orography dataset (chapter 11 in ECMWF (2019); Wedi et al. (2020) allows for a better representation of orographic GWs. To what extent the new 1.4 km orography dataset improves orographic GW representation in IFS can therefore be quantified.

6.4.2 Predictability and diagnostics

- Most effort so far has focused on diagnosing, understanding and alleviating stratospheric mean state biases. However, diagnosing and understanding biases in variability is perhaps even more important given its likely impact on predictive skill. As discussed in section 5, the variability of the polar lower stratosphere (in both hemispheres) appears to be important for the coupling of the stratosphere to the troposphere. As the variability of the polar lower stratosphere represents important stratospheric processes (i.e., SSWs, strong vortex events, final warmings, QBO – via Holton-Tan), more efforts should be devoted to quantifying this. A simple metric — suggested by the experts at the workshop — is to use the polar-cap averaged temperature anomaly at 100 hPa, i.e. the "ST100 index" (Baldwin et al., 2018). In this way the impact of model developments, such as radiatively interactive HLO scheme or reduced vertical diffusion (section 2.5), can be routinely measured.

- Tropospheric amplification, as a result of variability in the lower stratosphere, is important to model correctly. To what extent it is captured in IFS is yet to be quantified. This can be assessed either by using the conditional probability approach discussed in section 5.2, or by regressing the ST100 index onto the polar-cap average mean sea level pressure anomalies. Such metric would be useful for both capitalizing on "windows of opportunity" and for more process-based model evaluation. For example, it is known that the timing of surface impact following SSWs is very different depending on the event (e.g., largest surface response for February 2018 SSW occurred in weeks 3-4 following the event, whereas for January 2019 SSW largest response occurred in weeks 4-5 following the event). Moreover, it is not really the occurrence of the SSW itself (as defined by the reversal of zonal-mean zonal wind at 10 hPa and 60° latitude) that matters for the coupling to the surface, but rather the persistence and amplitude of lower stratospheric anomalies following the onset date (sections 2.4.4 and 5.3). Therefore, such metric is better suited for quantifying the model surface response.

- To quantify the strength of stratospheric impact on the troposphere relative to other remote events such as MJO and ENSO, the use of conditional probabilities (section 5.2) or the ST100 index and the regression of this index with a tropospheric field to quantify the coupling strength can be utilized in analogy to what is already done with e.g., MJO, IOD or ENSO indices. The effect of the stratosphere can also be removed by regression, which is analogous to conditioning out influence as discussed in section 5. The use of causal link algorithms provides another attractive path to assess causality and the role of the stratosphere versus other remote events. A desirable feature of causal link algorithms is that the autocorrelation — present in cross and lagged correlations

— is removed, and that the conditioning out of other drivers is done in a more optimized way than just taking a partial linear correlation (Kretschmer et al., 2016; Runge et al., 2019). Work in collaboration with Marlene Kretschmer at the University of Reading is ongoing to implement causal link algorithms on evaluating stratosphere-troposphere coupling on hindcast data.

- In addition to regression and causal link algorithms, nudging experiments are a powerful tool to assess stratospheric impact and to address which stratospheric biases are important, as has already been demonstrated in section 5. Therefore, more nudging experiments are planned as part of a collaboration with FMI and the SNAP activities. In addition to nudging, bias correction (Kharin and Scinocca, 2012) offers an alternative tool to study the impact of stratospheric biases on predictability. In the bias correction procedure the model is first run nudged for several years and/or several ensemble members and the nudging tendencies are saved. Afterwards the daily varying climatology of the nudging tendencies are formed. The model is then run again with the nudging tendencies applied. One can think of the application of the nudging tendencies as an extra parametrization and, unlike the nudging approach, the applied tendencies are independent of the model state and the approach does not damp perturbations. As weak constraint 4D-Var (section 3.1) is similar to the above bias correction procedure, there is a potential to explore the effect of reducing temperature biases (Fig. 1) on predictability using weak constraint formalism.

- Accurate representation of tropospheric precursors in models is important for accurate stratospheric variability. These include blocking events and eddy-jet feedbacks. Blocking events are precursors to many SSWs and it is already known that blocking events in the ECMWF system have lower predictability compared to other weather regimes (Ferranti et al., 2015). It is believed that increases in ocean resolution, and improvements in the ocean-atmosphere coupling – especially in the vicinity of the Golf stream — will help to better represent blocking (Vitart et al., 2019). Eddy-jet feedbacks are important for getting the teleconnections from e.g., MJO into the stratosphere. For example, about a month following MJO in phases 5-6, the polar vortex weakens in comparison to following MJO in phases 1-2. Models including IFS struggle to get the right amplitude and pattern of the response (Vitart et al., 2019, 2017) and this is thought to be due to systematic errors in the extra-tropics (Vitart et al., 2019). It is still an open question as to whether these troposphere-stratosphere teleconnections are well represented in IFS. The large ensemble of hindcasts provides an ideal testbed to assess this via composite analysis or the use of conditional probabilities.

- When testing if model improvements in the stratosphere make a difference for forecast skill, forecasts need to start with "windows of opportunity" when we expect the stratosphere to impact the troposphere. The standard cycle evaluation that assesses previous summer and winter forecasts is not ideal for this. The "windows of opportunity" include e.g., SSWs, final warmings in both hemispheres, unusually strong polar vortex events. Focusing on stratospheric variability from weeks to inter-annual timescales to enhance extended-range predictive skill, as well as targeting transition seasons (MAM, SON), forms part of ECMWF's next 10 year strategy. Part of this effort is to develop robust methods for identifying impact of the stratosphere on the troposphere.

As a closing remark, while a lot of progress has been achieved in recent years regarding stratospheric modelling and assimilation, it is still not always routinely clear to what extent these improvements have had an influence on the operational forecast performance (as a result of cycle updates). Therefore, research into a new regularly monitored "Quarterly Evaluation and Developments regular score" for the stratosphere and its impact on the troposphere should be conducted. An obvious such candidate is the ST100 index and its regression with tropospheric metrics to assess extended- and seasonal- range forecasts.

Acknowledgments

We thank Andy Brown, Magdalena Alonso Balmaseda, Niels Bormann and Florence Rabier for highlighting open points and making useful suggestions to the manuscript. The work described here greatly benefited from collaborations with many colleagues from our Member States and beyond, in particular Christopher Kruse (NCAR), Annelize Van Niekerk (Met Office), Lisa-Ann Kautz (KIT), Thomas Birner (LMU), Hella Garny (DLR), Bill Schreiner (UCAR) and members of the International Space Science Institute (ISSI) team 'New Quantitative Constraints on Orographic Gravity Wave Stress and Drag'.

References

Anstey, J. A. and T. G. Shepherd (2014). High-latitude influence of the quasi-biennial oscillation. *Quarterly Journal of the Royal Meteorological Society 140*(678), 1–21.

Ayarzagüena, B., D. Barriopedro, J. M. Garrido-Perez, M. Abalos, A. de la Cámara, R. García-Herrera, N. Calvo, and C. Ordóñez (2018). Stratospheric connection to the abrupt end of the 2016/2017 Iberian drought. *Geophysical Research Letters 45*(22), 12–639.

Baldwin, M. P., T. Birner, G. Brasseur, J. Burrows, N. Butchart, R. Garcia, M. Geller, L. Gray, K. Hamilton, N. Harnik, M. I. Hegglin, U. Langematz, A. Robock, K. Sato, and A. A. Scaife (2018). 100 Years of Progress in Understanding the Stratosphere and Mesosphere. *Meteorological Monographs 59*, 27.1–27.62.

Baldwin, M. P. and T. J. Dunkerton (2001). Stratospheric harbingers of anomalous weather regimes. *Science 294*(5542), 581–584.

Baumgaertner, A. J. G. and A. J. McDonald (2007). A gravity wave climatology for Antarctica compiled from Challenging Minisatellite Payload/Global Positioning System (CHAMP/GPS) radio occultations. *Journal of Geophysical Research: Atmospheres 112*(D5).

Beljaars, A. C. M., A. R. Brown, and N. Wood (2004). A new parametrization of turbulent orographic form drag. *Quarterly Journal of the Royal Meteorological Society 130*(599), 1327–1347.

Bithell, M., L. J. Gray, J. E. Harries, J. M. Rusell III, and A. F. Tuck (1994). Synoptic interpretation of measurements from HALOE. *J. Atmos. Sci. 51*, 2942–2956.

Black, R. X. and B. A. McDaniel (2007). Interannual variability in the Southern Hemisphere circulation organized by stratospheric final warming events. *Journal of the atmospheric sciences 64*(8), 2968–2974.

Boer, G., K. Arpe, M. Blackburn, M. Déqué, W. Gates, T. Hart, H. Le Treut, E. Rorteckner, D. Sheinin, I. Simmonds, et al. (1992). Some results from an intercomparison of the climates simulated by 14 atmospheric general circulation models. *Journal of Geophysical Research: Atmospheres 97*(D12), 12771–12786.

Bonavita, M., P. Lean, and E. Holm (2018). Nonlinear effects in 4D-Var. *Nonlinear Processes in Geophysics 25*(3), 713–729.

Bormann, N., H. Lawrence, and J. Farnan (2019, 01). Global observing system experiments in the ECMWF assimilation system. (839). Available at: https://www.ecmwf.int/node/18859.

Bormann, N. and J.-N. Thépaut (2007). Assimilation of MIPAS limb radiances in the ECMWF system. I: Experiments with a 1-dimensional observation operator. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography 133*(623), 309–327.

Bougeault, P. (1983, 03). A Non-Reflective Upper Boundary Condition for Limited-Height Hydrostatic Models. *Monthly Weather Review 111*(3), 420–429.

Büeler, D., R. Beerli, H. Wernli, and C. M. Grams (2020). Stratospheric influence on ECMWF sub-seasonal forecast skill for energy-industry-relevant surface weather in European countries. *Quarterly Journal of the Royal Meteorological Society*. Available at: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3866.

Bushell, A., N. Butchart, S. H. Derbishire, D. R. Jackson, G. J. Shutts, S. B. Vosper, and S. Webster (2015). Parameterized gravity wave momentum fluxes from sources related to convection and large-scale precipitation processes in a global atmosphere model. *J. Atmos. Sci. 72*, 4349–4349.

Butchart, N. (2014). The Brewer-Dobson circulation. *Reviews of geophysics 52*(2), 157–184.

Butler, A., A. Charlton-Perez, D. I. Domeisen, C. Garfinkel, E. P. Gerber, P. Hitchcock, A. Y. Karpechko, A. C. Maycock, M. Sigmond, I. Simpson, et al. (2019). Sub-seasonal Predictability and the Stratosphere. In *Sub-Seasonal to Seasonal Prediction*, pp. 223–241. Elsevier.

Butler, A. H., A. Arribas, M. Athanassiadou, J. Baehr, N. Calvo, A. Charlton-Perez, M. Déqué, D. I. Domeisen, K. Fröhlich, H. Hendon, et al. (2016). The Climate-system Historical Forecast Project: do stratosphere-resolving models make better seasonal climate predictions in boreal winter? *Quarterly Journal of the Royal Meteorological Society 142*(696), 1413–1427.

Butler, A. H., A. Charlton-Perez, D. I. Domeisen, I. R. Simpson, and J. Sjoberg (2019). Predictability of Northern Hemisphere Final Stratospheric Warmings and Their Surface Impacts. *Geophysical Research Letters 46*(17-18), 10578–10588.

Byrne, N. J., T. G. Shepherd, and I. Polichtchouk (2019). Subseasonal-to-Seasonal Predictability of the Southern Hemisphere Eddy-Driven Jet During Austral Spring and Early Summer. *Journal of Geophysical Research: Atmospheres 124*(13), 6841–6855.

Byrne, N. J., T. G. Shepherd, T. Woollings, and R. A. Plumb (2017). Nonstationarity in Southern Hemisphere Climate Variability Associated with the Seasonal Breakdown of the Stratospheric Polar Vortex. *Journal of Climate 30*(18), 7125–7139.

Cariolle, D. and M. Déqué (1986). Southern hemisphere medium-scale waves and total ozone disturbances in a spectral general circulation model. *Journal of Geophysical Research: Atmospheres 91*(D10), 10825–10846.

Charlton-Perez, A. J., J. Bröcker, T. N. Stockdale, and S. Johnson (2019). When and where do ECMWF seasonal forecast systems exhibit anomalously low signal-to-noise ratio? *Quarterly Journal of the Royal Meteorological Society 145*(725), 3466–3478.

Charney, J. G. and N. Phillips (1953). Numerical integration of the quasi-geostrophic equations for barotropic and simple baroclinic flows. *Journal of Meteorology 10*(2), 71–99.

Coddington, O., J. Lean, P. Pilewskie, M. Snow, and D. Lindholm (2016). A solar irradiance climate data record. *Bulletin of the American Meteorological Society 97*(7), 1265–1282.

Dharmalingam, S., R. Plougonven, A. Hertzog, A. Podglajen, M. Rennie, L. Isaksen, and S. Kébir (2019). Accuracy of Balloon Trajectory Forecasts in the Lower Stratosphere. *Atmosphere 10*(2), 102.

Diamantakis, M. (2014). Improving ECMWF forecasts of sudden stratospheric warmings. pp. 30–36. Available at: https://www.ecmwf.int/node/17336.

Diamantakis, M. and L. Magnusson (2016). Sensitivity of the ECMWF model to semi-Lagrangian departure point iterations. *Monthly Weather Review 144*(9), 3233–3250.

Domeisen, D. I. (2019). Estimating the Frequency of Sudden Stratospheric Warming Events From Surface Observations of the North Atlantic Oscillation. *Journal of Geophysical Research: Atmospheres 124*(6), 3180–3194.

Domeisen, D. I., A. H. Butler, A. J. Charlton-Perez, B. Ayarzagüena, M. P. Baldwin, E. Dunn-Sigouin, J. C. Furtado, C. I. Garfinkel, P. Hitchcock, A. Y. Karpechko, H. Kim, J. Knight, A. L. Lang, E.-P. Lim, A. Marshall, G. Roff, C. Schwartz, I. R. Simpson, S.-W. Son, and M. Taguchi (2020a). The role of the stratosphere in subseasonal to seasonal prediction Part I: Predictability of the stratosphere. *Journal of Geophysical Research: Atmospheres 125*(2), e2019JD030920.

Domeisen, D. I., A. H. Butler, A. J. Charlton-Perez, B. Ayarzagüena, M. P. Baldwin, E. Dunn-Sigouin, J. C. Furtado, C. I. Garfinkel, P. Hitchcock, A. Y. Karpechko, H. Kim, J. Knight, A. L. Lang, E.-P. Lim, A. Marshall, G. Roff, C. Schwartz, I. R. Simpson, S.-W. Son, and M. Taguchi (2020b). The role of the stratosphere in subseasonal to seasonal prediction Part II: Predictability arising from stratosphere - troposphere coupling. *Journal of Geophysical Research: Atmospheres 125*(2), e2019JD030923.

Domeisen, D. I., A. H. Butler, K. Fröhlich, M. Bittner, W. A. Müller, and J. Baehr (2015). Seasonal predictability over Europe arising from El Niño and stratospheric variability in the MPI-ESM seasonal prediction system. *Journal of Climate 28*(1), 256–271.

Domeisen, D. I., C. I. Garfinkel, and A. H. Butler (2019). The teleconnection of El Niño Southern Oscillation to the stratosphere. *Reviews of Geophysics 57*(1), 5–47.

Douville, H. (2009). Stratospheric polar vortex influence on Northern Hemisphere winter climate variability. *Geophysical Research Letters 36*(18), n/a–n/a. L18703.

Dragani, R., S. Abdalla, R. Engelen, A. Inness, and J.-N. Thépaut (2015). Ten years of ENVISAT observations at ECMWF: A review of activities and lessons learnt. *Quarterly Journal of the Royal Meteorological Society 141*(687), 598–610. Available at: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2380.

Dunkerton, T. J. (1984). Inertia-gravity waves in the stratosphere. *Journal of the atmospheric sciences 41*(23), 3396–3404.

Dyroff, C., A. Zahn, E. Christner, R. Forbes, A. M. Tompkins, and P. F. J. van Velthoven (2015). Comparison of ECMWF analysis and forecast humidity data with CARIBIC upper troposphere and lower stratosphere observations. *Quarterly Journal of the Royal Meteorological Society 141*(688), 833–844.

ECMWF (2019). *PART IV: PHYSICAL PROCESSES*. Number 4 in IFS Documentation. ECMWF. Available at: https://www.ecmwf.int/node/19308.

Ehard, B., B. Kaifler, N. Kaifler, and M. Rapp (2015). Evaluation of methods for gravity wave extraction from middle-atmospheric lidar temperature measurements. *Atmospheric Measurement Techniques 8*(11), 4645–4655.

Ehard, B., S. Malardel, A. Dörnbrack, B. Kaifler, N. Kaifler, and N. Wedi (2018). Comparing ECMWF high-resolution analyses with lidar temperature measurements in the middle atmosphere. *Quarterly Journal of the Royal Meteorological Society 144*(712), 633–640.

Ern, M., L. Hoffmann, and P. Preusse (2017). Directional gravity wave momentum fluxes in the stratosphere derived from high-resolution AIRS temperature data. *Geophys. Res. Let. 44*, 475–485.

Ern, M., P. Preusse, M. J. Alexander, and C. D. Warner (2004). Absolute values of gravity wave momentum flux derived from satellite data. *Journal of Geophysical Research: Atmospheres 109*(D20).

Ern, M., Q. T. Trinh, P. Preusse, J. C. Gille, M. G. Mlynczak, J. M. Russell III, and M. Riese (2018). GRACILE: A comprehensive climatology of atmospheric gravity wave parameters based on satellite limb soundings. *Earth System Science Data 10*(2), 857–892.

Ferranti, L., S. Corti, and M. Janousek (2015). Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. *Quarterly Journal of the Royal Meteorological Society 141*(688), 916–924.

Fisher, M. (2012). An investigation into intermittent widespread rejection of SYNOP pressure data over Europe throughout 2010 and 1011. *ECMWF Research Memorandum R43/MF/1214*.

Fleming, E. L. and S. Chandra (1989). Equatorial Zonal Wind in the Middle Atmosphere Derived from Geopotential Height and Temperature Data. *J. Atmos. Sci. 46*, 860–866.

Garfinkel, C. I., J. J. Benedict, and E. D. Maloney (2014). Impact of the MJO on the boreal winter extratropical circulation. *Geophysical Research Letters 41*(16), 6055–6062.

Garfinkel, C. I., C. Schwartz, D. I. Domeisen, S.-W. Son, A. H. Butler, and I. P. White (2018). Extratropical atmospheric predictability from the quasi-biennial oscillation in subseasonal forecast models. *Journal of Geophysical Research: Atmospheres 123*(15), 7855–7866.

Gauthier, P. and J.-N. Thepaut (2001). Impact of the digital filter as a weak constraint in the preoperational 4DVAR assimilation system of Météo-France. *Monthly weather review 129*(8), 2089–2102.

Geer, A. J. (2020, 05). Learning earth system models from observations: machine learning or data assimilation? (863). Available at: https://www.ecmwf.int/node/19525.

Gray, L. J., J. A. Anstey, Y. Kawatani, H. Lu, S. Osprey, and V. Schenzinger (2018). Surface impacts of the Quasi Biennial Oscillation. *Atmospheric Chemistry and Physics 18*(11), 8227–8247.

Han, W. and N. Bormann (2016). Constrained adaptive bias correction for satellite radiance assimilation in the ECMWF 4D-Var system. (783). Availabel at: https://www.ecmwf.int/node/16712.

Hankinson, M. C. N., M. J. Redder, and T. P. Lane (2014). Gravity waves generated by convection during TWP-ICE. *J. Geophys. Res. Atmos. 119*, 5257–5268.

Haynes, P. H., M. E. McIntyre, C. J. Marks, T. G. Shepherd, and K. P. Shine (1991). On the "downward control" of extratropical diabatic circulations by eddy-induced mean zonal forces. *Journal of the Atmospheric Sciences 48*(4), 651–678.

Healy, S. B., I. Polichtchouk, and A. Horanyi (2020). Monthly and zonally averaged zonal wind information in the Equatorial stratosphere provided by GNSS radio occultation. *Quarterly Journal of the Royal Meteorological Society*. Available at: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3870.

Hegglin, M., D. Kinnison, J. Lamarque, and D. Plummer (2016). CCMI ozone in support of CMIP6-version 1.0.

Hertzog, A., M. J. Alexander, and R. Plougonven (2012). On the intermittency of gravity wave momentum flux in the stratosphere. *J. Atmos. Sci. 72*, 3433–3448.

Hio, Y. and S. Yoden (2005). Interannual variations of the seasonal march in the Southern Hemisphere stratosphere for 1979–2002 and characterization of the unprecedented year 2002. *Journal of the atmospheric sciences 62*(3), 567–580.

Hitchcock, P., T. G. Shepherd, and G. L. Manney (2013). Statistical characterization of Arctic polar-night jet oscillation events. *Journal of Climate 26*(6), 2096–2116.

Hitchcock, P. and I. R. Simpson (2014). The downward influence of stratospheric sudden warmings. *Journal of the Atmospheric Sciences 71*(10), 3856–3876.

Hoffmann, L., R. Spang, A. Orr, M. J. Alexander, L. A. Holt, and O. Stein (2017). A decadal satellite record of gravity wave activity in the lower stratosphere to study polar stratospheric cloud formation. *Atmospheric Chemistry and Physics 17*(4), 2901–2920.

Hogan, R. J., M. Ahlgrimm, G. Balsamo, A. Beljaars, P. Berrisford, A. Bozzo, F. Di Giuseppe, R. M. Forbes, T. Haiden, S. Lang, et al. (2017). *Radiation in numerical weather prediction*. European Centre for Medium-Range Weather Forecasts. Available at: https://www.ecmwf.int/node/17771.

Hogan, R. J. and A. Bozzo (2018). A Flexible and Efficient Radiation Scheme for the ECMWF Model. *Journal of Advances in Modeling Earth Systems 10*(8), 1990–2008.

Hogan, R. J. and S. Hirahara (2016). Effect of solar zenith angle specification in models on mean shortwave fluxes and stratospheric temperatures. *Geophysical Research Letters 43*(1), 482–488.

Hogan, R. J. and M. Matricardi (2020). Evaluating and improving the treatment of gases in radiation schemes: the Correlated K-Distribution Model Intercomparison Project (CKDMIP). *Submitted to Geosci. Model Devel.*.

Hollingsworth, A. (1995, 02). A spurious mode in the 'Lorenz' arrangement of Y and T which does not exist in the 'Charney-Phillips' arrangement. (211), 12. Available at: https://www.ecmwf.int/node/9943.

Holton, J. R. and H.-C. Tan (1980). The Influence of the Equatorial Quasi-Biennial Oscillation on the Global Circulation at 50 mb. *Journal of the Atmospheric Sciences 37*(10), 2200–2208.

Holton, J. R. and H.-C. Tan (1982). The Quasi-Biennial Oscillation in the Northern Hemisphere Lower Stratosphere. *Journal of the Meteorological Society of Japan. Ser. II 60*(1), 140–148.

Houchi, K., A. Stoffelen, G. Marseille, and J. De Kloe (2010). Comparison of wind and wind shear climatologies derived from high-resolution radiosondes and the ECMWF model. *Journal of Geophysical Research: Atmospheres 115*(22), D22123.

Huijnen, V., J. Flemming, S. Chabrillat, Q. Errera, Y. Christophe, A.-M. Blechschmidt, A. Richter, and H. Eskes (2016). C-IFS-CB05-BASCOE: stratospheric chemistry in the Integrated Forecasting System of ECMWF. *Geoscientific Model Development 9*(9), 3071–3091.

IFSdoc (2019). PART III: DYNAMICS AND NUMERICAL PROCEDURES. Available at: https://www.ecmwf.int/node/19307.

Jewtoukoff, V., A. Hertzog, R. Plougonven, A. d. l. Cámara, and F. Lott (2015). Comparison of Gravity Waves in the Southern Hemisphere Derived from Balloon Observations and the ECMWF Analyses. *Journal of the Atmospheric Sciences 72*(9), 3449–3468.

Johnson, S. J., T. N. Stockdale, L. Ferranti, M. A. Balmaseda, F. Molteni, L. Magnusson, S. Tietsche, D. Decremer, A. Weisheimer, G. Balsamo, et al. (2019). SEAS5: the new ECMWF seasonal forecast system. *Geoscientific Model Development 12*(3).

Jones, R. L., J. A. Pyle, J. E. Harries, A. M. Zavody, J. M. Russell III, and J. C. Gille (1986). The water vapour budget of the stratosphere studied using LIMS and SAMS satellite data. *Q. J. R. Meteorol. Soc. 112*, 1127–1143.

Kaifler, B. and N. Kaifler (2020). A Compact Rayleigh Autonomous Lidar (CORAL) for the middle atmosphere. *Atmospheric Measurement Techniques Discussions 2020*, 1–24.

Kaifler, N., B. Kaifler, A. Drnbrack, M. Rapp, J. Hormaechea, and A. Torre (2020, 09). Lidar observations of large-amplitude mountain waves in the stratosphere above Tierra del Fuego, Argentina. *Scientific Reports 10*, 14529.

Karpechko, A. Y. (2015). Improvements in statistical forecasts of monthly and two-monthly surface air temperatures using a stratospheric predictor. *Quarterly Journal of the Royal Meteorological Society 141*(691), 2444–2456.

Karpechko, A. Y., A. Charlton-Perez, M. Balmaseda, N. Tyrrell, and F. Vitart (2018). Predicting sudden stratospheric warming 2018 and its climate impacts with a multimodel ensemble. *Geophysical Research Letters 45*(24), 13–538.

Karpechko, A. Y., P. Hitchcock, D. H. Peters, and A. Schneidereit (2017). Predictability of downward propagation of major sudden stratospheric warmings. *Quarterly Journal of the Royal Meteorological Society 143*(704), 1459–1470.

Kautz*, L.-A., I. Polichtchouk*, T. Birner, H. Garny, and J. G. Pinto (2020). Enhanced extended-range predictability of the 2018 late-winter Eurasian cold spell due to the stratosphere. *Quarterly Journal of the Royal Meteorological Society 146*(727), 1040–1055.

Kharin, V. and J. Scinocca (2012). The impact of model fidelity on seasonal predictive skill. *Geophysical research letters 39*(18).

Klemp, J. B., J. Dudhia, and A. D. Hassiotis (2008, 10). An Upper Gravity-Wave Absorbing Layer for NWP Applications. *Monthly Weather Review 136*(10), 3987–4004.

Kretschmer, M., D. Coumou, J. F. Donges, and J. Runge (2016). Using causal effect networks to analyze different Arctic drivers of midlatitude winter circulation. *Journal of Climate 29*(11), 4069–4081.

Kühnlein, C., W. Deconinck, R. Klein, S. Malardel, Z. P. Piotrowski, P. K. Smolarkiewicz, J. Szmelter, and N. P. Wedi (2019). FVM 1.0: a nonhydrostatic finite-volume dynamical core for the IFS. *Geoscientific model development 12*, 651–676.

Laloyaux, P., M. Bonavita, M. Chrust, and S. Gürol (2020). Exploring the potentials and limitations of weak-constraint 4D-Var. *QJRMS, submitted*.

Laloyaux, P., M. Bonavita, M. Dahoui, J. Farnan, S. Healy, E. Hólm, and S. T. K. Lang (2020). Towards an unbiased stratospheric analysis. *QJRMS*. Available at: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3798.

L'Heureux, M. L. and D. W. Thompson (2006). Observed relationships between the El Niño–Southern Oscillation and the extratropical zonal-mean circulation. *Journal of Climate 19*(2), 276–287.

Lim, E.-P., H. H. Hendon, A. H. Butler, R. D. Garreaud, I. Polichtchouk, T. G. Shepherd, A. Scaife, R. Comer, L. Coy, P. A. Newman, et al. (2020). The 2019 Antarctic sudden stratospheric warming. *Newsletter n 54 January 2020*, 10.

Long, C. S., M. Fujiwara, S. Davis, D. M. Mitchell, and C. J. Wright (2017). Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP). *Atmospheric Chemistry and Physics 17*(23), 14593–14629.

Lott, F. and L. Guez (2013). A stochastic parameterization of the gravity waves due to convection and its impact on the equatorial stratosphere. *Journal of Geophysical Research: Atmospheres 118*(16), 8897–8909.

Lott, F. and M. J. Miller (1997). A new subgrid-scale orographic drag parametrization: Its formulation and testing. *Quarterly Journal of the Royal Meteorological Society 123*(537), 101–127.

Molteni, F., T. Stockdale, M. Balmaseda, G. Balsamo, R. Buizza, L. Ferranti, L. Magnusson, K. Mogensen, T. Palmer, and F. Vitart (2011). *The new ECMWF seasonal forecast system (System 4)*, Volume 49. European Centre for Medium-Range Weather Forecasts Reading.

Monge-Sanz, B., M. Chipperfield, D. Cariolle, and W. Feng (2011). Results from a new linear O3 scheme with embedded heterogeneous chemistry compared with the parent full-chemistry 3-D CTM. *Atmospheric Chemistry and Physics 11*(3), 1227.

Mote, P. W., K. H. Rosenlof, M. E. McIntyre, E. S. Carr, J. C. Gille, J. R. Holton, J. S. Kinnersley, H. C. Pumphrey, J. M. Russell III, and J. W. Waters (1996). An atmospheric tape recorder: The imprint of tropical tropopause temperatures on stratospheric water vapor. *Journal of Geophysical Research: Atmospheres 101*(D2), 3989–4006.

Naujokat, B. (1986). An update of the observed quasi–biennial oscillation of the stratospheric winds over the tropics. *J. Atmos. Sci. 43*, 1873–1877.

Orr, A., P. Bechtold, J. Scinocca, M. Ern, and M. Janiskova (2010). Improved middle atmosphere climate and forecasts in the ECMWF model through a nonorographic gravity wave drag parameterization. *Journal of Climate 23*(22), 5905–5926.

Plougonven, R. and F. Zhang (2014). Internal gravity waves from atmospheric jets and fronts. *Reviews of Geophysics 52*(1), 33–76.

Podglajen, A., A. Hertzog, R. Plougonven, and N. Žagar (2014). Assessment of the accuracy of (re) analyses in the equatorial lower stratosphere. *Journal of Geophysical Research: Atmospheres 119*(19), 11–166.

Polavarapu, S., T. Shepherd, Y. Rochon, and S. Ren (2005). Some challenges of middle atmosphere data assimilation. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography 131*(613), 3513–3527.

Polichtchouk, I., M. Diamantakis, and F. Vána (2020). Quintic vertical interpolation improves forecasts of the stratosphere. pp. 23–26. Available at: https://www.ecmwf.int/node/19515.

Polichtchouk, I., R. Hogan, T. Shepherd, P. Bechtold, T. Stockdale, S. Malardel, S.-J. Lock, and L. Magnusson (2017). What influences the middle atmosphere circulation in the IFS? (809). Available at: https://www.ecmwf.int/node/17670.

Polichtchouk, I. and R. K. Scott (2020). Spontaneous inertia-gravity wave emission from a nonlinear critical layer in the stratosphere. *Quarterly Journal of the Royal Meteorological Society 146*(728), 1516–1528.

Polichtchouk, I., T. G. Shepherd, and N. J. Byrne (2018). Impact of Parametrized Nonorographic Gravity Wave Drag on Stratosphere-Troposphere Coupling in the Northern and Southern Hemispheres. *Geophysical Research Letters 45*(16), 8612–8618.

Polichtchouk, I., T. G. Shepherd, R. J. Hogan, and P. Bechtold (2018). Sensitivity of the Brewer-Dobson Circulation and Polar Vortex Variability to Parameterized Nonorographic Gravity Wave Drag in a High-Resolution Atmospheric Model. *Journal of the Atmospheric Sciences 75*(5), 1525–1543.

Polichtchouk, I., T. Stockdale, P. Bechtold, M. Diamantakis, S. Malardel, I. Sandu, F. Vána, and N. Wedi (2019, 06). Control on stratospheric temperature in IFS: resolution and vertical advection. (847). Available at: https://www.ecmwf.int/node/19084.

Preusse, P., M. Ern, P. Bechtold, S. Eckermann, S. Kalisch, Q. Trinh, and M. Riese (2014). Characteristics of gravity waves resolved by ECMWF. *Atmos. Chem. Phys 14*(10), 483–10.

Prusa, J. M., P. K. Smolarkiewicz, and R. R. Garcia (1996). Propagation and breaking at high altitudes of gravity waves excited by tropospheric forcing. *Journal of the atmospheric sciences 53*(15), 2186–2216.

Randel, W. J., F. Wu, J. M. Russell III, A. Roche, and J. W. Waters (1998). Seasonal cycles and QBO variations in straospheric CH_4 and H_2O observed in UARS HALOE data. *J. Atmos. Sci. 55*, 163–185.

Rennie, M. and L. Isaksen (2020a). The NWP impact of Aeolus Level-2B Winds at ECMWF. (864). Available at: https://www.ecmwf.int/node/19538.

Rennie, M. and L. Isaksen (2020b). Use of Aeolus observations at ECMWF: Newsletter No. 163 - Spring 2020. Available at: https://www.ecmwf.int/node/19508.

Riahi, K., A. Gruebler, and N. Nakicenovic (2007). Scenarios of long-term socio-economic and environmental development under climate stabilization. *Technological Forecasting and Social Change 74*(7), 887–935.

Runge, J., S. Bathiany, E. Bollt, G. Camps-Valls, D. Coumou, E. Deyle, C. Glymour, M. Kretschmer, M. D. Mahecha, J. Muñoz-Marí, et al. (2019). Inferring causation from time series in Earth system sciences. *Nature communications 10*(1), 1–13.

Sato, K., S. Tateno, S. Watanabe, and Y. Kawatani (2012). Gravity wave characteristics in the Southern Hemisphere revealed by a high-resolution middle-atmosphere general circulation model. *Journal of the Atmospheric Sciences 69*(4), 1378–1396.

Scaife, A., A. Arribas, E. Blockley, A. Brookshaw, R. Clark, N. Dunstone, R. Eade, D. Fereday, C. Folland, M. Gordon, et al. (2014). Skillful long-range prediction of European and North American winters. *Geophysical Research Letters 41*(7), 2514–2519.

Scaife, A., A. Y. Karpechko, M. Baldwin, A. Brookshaw, A. Butler, R. Eade, M. Gordon, C. MacLachlan, N. Martin, N. Dunstone, et al. (2016). Seasonal winter forecasts and the stratosphere. *Atmospheric Science Letters 17*(1), 51–56.

Scaife, A. A., J. Austin, N. Butchart, S. Pawson, M. Keil, J. Nash, and I. N. James (2000). Seasonal and interannual variability of the stratosphere diagnosed from UKMO TOVS analyses. *Q. J. R. Meteorol. Soc. 126*, 2585–2604.

Schroeder, S., P. Preusse, M. Ern, and M. Riese (2009). Gravity waves resolved in ECMWF and measured by SABER. *Geophysical research letters 36*(10).

Scinocca, J. F. (2003). An accurate spectral nonorographic gravity wave drag parameterization for general circulation models. *Journal of the Atmospheric Sciences 60*(4), 667–682.

Seviour, W. J. M., S. C. Hardiman, L. J. Gray, N. Butchart, C. MacLachlan, and A. A. Scaife (2014). Skillful Seasonal Prediction of the Southern Annular Mode and Antarctic Ozone. *Journal of Climate 27*(19), 7462–7474.

Shaw, T. A., M. Sigmond, T. G. Shepherd, and J. F. Scinocca (2009). Sensitivity of Simulated Climate to Conservation of Momentum in Gravity Wave Drag Parameterization. *Journal of Climate 22*(10), 2726–2742.

Shepherd, T., I. Polichtchouk, R. Hogan, and A. Simmons (2018). Report on Stratosphere Task Force. (824). Available at: https://www.ecmwf.int/node/18259.

Shepherd, T. G. (2000). The middle atmosphere. *Journal of Atmospheric and Solar-Terrestrial Physics 62*(17), 1587–1601.

Shepherd, T. G. and T. A. Shaw (2004). The angular momentum constraint on climate sensitivity and downward influence in the middle atmosphere. *Journal of the atmospheric sciences 61*(23), 2899–2908.

Sigmond, M., J. F. Scinocca, V. V. Kharin, and T. G. Shepherd (2013). Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nature Geoscience 6*(2), 98–102.

Simmons, A., C. Soci, J. Nicolas, B. Bell, P. Berrisford, R. Dragani, J. Flemming, L. Haimberger, S. Healy, H. Hersbach, A. Horanyi, A. Inness, J. Munoz-Sabater, R. Radu, and D. Schepers (2020). Global stratospheric temperature bias and other stratospheric aspects of ERA5 and ERA5.1. Technical report. Available at: https://www.ecmwf.int/node/19362.

Simmons, A. J., B. J. Hoskins, and D. M. Burridge (1978). Stability of the semi-implicit method of time integration. *Monthly Weather Review 106*(3), 405–412.

Simmons, A. J. and C. Temperton (1997). Stability of a two-time-level semi-implicit integration scheme for gravity wave motion. *Monthly weather review 125*(4), 600–615.

Stenke, A., V. Grewe, and M. Ponater (2008). Lagrangian transport of water vapor and cloud water in the ECHAM4 GCM and its impact on the cold bias. *Climate dynamics 31*(5), 491–506.

Stephan, C. C., C. Strube, D. Klocke, M. Ern, L. Hoffmann, P. Preusse, and H. Schmidt (2019). Gravity waves in global high-resolution simulations with explicit and parameterized convection. *J. Geophys. Res. 124*, 4446–4459.

Sun, J. and B. Tan (2013). Mechanism of the wintertime Aleutian low–Icelandic low seesaw. *Geophysical Research Letters 40*(15), 4103–4108.

Tanguay, M., P. Bartello, and P. Gauthier (1995). Four-dimensional data assimilation with a wide range of scales. *Tellus A 47*(5), 974–997.

Trémolet, Y. (2007). Incremental 4D-Var convergence study. *Tellus A: Dynamic Meteorology and Oceanography 59*(5), 706–718.

Untch, A. and M. Hortal (2004). A finite-element scheme for the vertical discretization of the semi-Lagrangian version of the ECMWF forecast model. *Quarterly Journal of the Royal Meteorological Society 130*(599), 1505–1530.

van Niekerk, A., I. Sandu, A. Zadra, E. Bazile, T. Kanehama, M. Koehler, M.-S. Koo, H.-J. Choi, Y. Kuroki, M. Toy, S. B. Vosper, and V. Yudin (2020). COnstraining ORographic Drag Effects (CORDE): a model comparison of resolved and parametrized orographic drag. *JAMES*. In revision.

Váňa, F., P. Bénard, J.-F. Geleyn, A. Simon, and Y. Seity (2008). Semi-Lagrangian advection scheme with controlled damping: An alternative to nonlinear horizontal diffusion in a numerical weather prediction model. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography 134*(631), 523–537.

Verma, S., J. Marshall, M. Parrington, A. Agustí-Panareda, S. Massart, M. P. Chipperfield, C. Wilson, and C. Gerbig (2017). Extending methane profiles from aircraft into the stratosphere for satellite total column validation using the ECMWF C-IFS and TOMCAT/SLIMCAT 3-D model. *Atmospheric Chemistry and Physics 17*(11), 6663–6678.

Vitart, F., M. Alonso-Balmaseda, L. Ferranti, A. Benedetti, B. Balan-Sarojini, S. Tietsche, J. Yao, M. Janousek, G. Balsamo, M. Leutbecher, P. Bechtold, I. Polichtchouk, D. Richardson, T. Stockdale, and C. D. Roberts (2019). Extended-range prediction. (854). Available at: https://www.ecmwf.int/node/19286.

Vitart, F., C. Ardilouze, A. Bonet, A. Brookshaw, M. Chen, C. Codorean, M. Déqué, L. Ferranti, E. Fucile, M. Fuentes, et al. (2017). The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society 98*(1), 163–173.

Vitart, F., A. W. Robertson, and D. L. Anderson (2012). Subseasonal to Seasonal Prediction Project: Bridging the gap between weather and climate. *Bulletin of the World Meteorological Organization 61*(2), 23.

Warner, C. D. and M. E. McIntyre (2001). An Ultrasimple Spectral Parameterization for Nonorographic Gravity Waves. *Journal of the Atmospheric Sciences 58*(14), 1837–1857.

Wedi, N. P., P. Dueben, V. G. Anantharaj, P. Bauer, S. Boussetta, P. Browne, W. Deconinck, W. Gaudin, I. Hadade, S. Hatfield, O. Iffrig, P. Lopez, P. Maciel, A. Mueller, I. Polichtchouk, S. Saarinen, T. Quintino, and F. Vitart (2020). A baseline for global weather and climate simulations at 1.4 km resolution. *JAMES*. In revision.

Wedi, N. P. and P. K. Smolarkiewicz (2006). Direct numerical simulation of the Plumb–McEwan laboratory analog of the QBO. *Journal of the atmospheric sciences 63*(12), 3226–3252.

Wheeler, M. and G. N. Kiladis (1999). Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber–frequency domain. *Journal of the Atmospheric Sciences 56*(3), 374–399.

Woiwode, W., A. Dörnbrack, I. Polichtchouk, S. Johansson, B. Harvey, M. Höpfner, J. Ungermann, and F. Friedl-Vallon (2020). Technical note: Lowermost-stratospheric moist bias in ECMWF IFS model diagnosed from airborne GLORIA observations during winter/spring 2016. *Atmospheric Chemistry and Physics Discussions 2020*, 1–11. Available at: https://www.atmos-chem-phys-discuss.net/acp-2020-367/.

Zadra, A. (2015). WGNE drag project: An inter-model comparison of surface stresses. *Environment Canada Tech. Rep 1*, 36.

Žagar, N., M. Blaauw, B. Jesenko, and L. Magnusson (2016). Diagnosing model performance in the tropics. pp. 26–33. Available at: https://www.ecmwf.int/node/17205.

Ẑagar, N., L. Isaksen, D. Tan, and J. Tribbia (2013). Balance properties of the short-range forecast errors in the ECMWF 4D-Var ensemble. *Quarterly Journal of the Royal Meteorological Society 139*(674), 1229–1238.

Ẑagar, N., D. Jelić, M. Blaauw, and P. Bechtold (2017). Energy spectra and inertia-gravity waves in global analyses. *J. Atmos. Sci. 74*, 2447–2466.