Technical Memo

ECMWF
European Centre for Medium-Range Weather Forecasts

893

Evaluation and optimization of orographic drag in the IFS

Takafumi Kanehama[1], Irina Sandu, Anton Beljaars, Annelize van Niekerk[2], Nils Wedi, Souhail Boussetta, Simon Lang, Stephanie Johnson, Linus Magnusson

[1] Japan Meteorological Agency, Tokyo

[2] Met Office, Exeter

April 2022

Series: ECMWF technical memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/en/publications/

Contact: library@ecmwf.int

Abstract

In this study we validate and optimize the orographic low-level flow blocking and gravity wave drag parametrizations used in the Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF), across a range of horizontal resolutions (10 to 100 km). We make use of km-scale simulations, that are able to largely resolve these orographic drag processes, building upon the protocol developed for the WGNE/GASS COORDE (Constraining Orographic Drag Effects) model comparison project. Km-scale simulations over the Middle East mountain chains were used in COORDE to evaluate the performance of the flow blocking and gravity wave drag parametrizations in a variety of models including the IFS. COORDE revealed, for example, that the IFS, along with several other models, underestimates orographic drag due to gravity wave breaking, particularly in the lower stratosphere. By constraining orographic drag associated with mountains with scales larger than 5 km, the COORDE km-scale simulations also helped to constrain the partition between low-level drag processes.

The insights gained from COORDE and from recent work using km-scale simulations at the UK MetOffice and ECMWF were used to revise the representation of orographic drag processes in the IFS. This consisted of a revision of the subgrid orography fields, and a repartitioning of the contributions of the low-level orographic drag processes (turbulent orographic form drag and flow blocking) to the surface drag. The revised subgrid orography fields now represent orographic features with scales smaller than the effective orographic resolution of the model rather than the grid-length. Additionally, the strength of the turbulent orographic form drag was decreased while low-level flow blocking drag was enhanced. With these revisions, the impact of the parametrized orographic drag on the zonal winds becomes more similar to the impact of the explicitly resolved drag from the km-scale simulations over the Middle East region, compared to the current version of the IFS. Moreover, with the new subgrid orographic fields and the retuned orographic drag parametrizations, the IFS exhibits a zonal momentum budget that is consistent across a range of resolutions. The impact of the revised representation of orographic drag on forecast performance was evaluated extensively across various applications used at ECMWF from data assimilation and medium-range forecasts to seasonal prediction. Improvements were found in key aspects of the Northern Hemisphere winter circulation at all timescales and resolutions used for applications at ECMWF.

1 Introduction

Orographic drag is known to affect the global atmospheric circulation and, more specifically, key aspects of the Northern Hemisphere (NH) winter circulation (White et al. 2021). In numerical models of the atmospheric flow, a set of governing equations are discretised to obtain numerical solutions. Therefore, boundary conditions such as orography and other characteristics of the land surface also need to be represented discretely. Numerical discretisation inevitably introduces scale truncation, resulting in numerical models that cannot explicitly represent scales smaller than their resolution. The effects of processes taking place at scales smaller than the model resolution are represented using so-called parametrizations. Orographic drag parametrizations are used to represent the drag exerted on the atmosphere by unresolved orographic features. Because orography plays an important role in the evolution of the atmospheric flow, orographic drag parametrizations constitute key ingredients in the accurate prediction of weather and climate (Sandu et al. 2019).

The effects of unresolved orographic scales were first accounted for in a global circulation model through the introduction of an envelope orography, which increases the model mean (or resolved) orography in proportion with the standard deviation of the subgrid scale orographic heights (Slingo and Pearson 1987, Wallace et al. 1983). Later, the concept of envelope orography was replaced by more direct parametriza-

tions of subgrid scale orographic effects, using theoretical principles. These represent gravity wave drag (Iwasaki et al. 1989, McFarlane 1987, Palmer et al. 1986), low-level flow blocking (Lott and Miller 1997, Scinocca and McFarlane 2000) and turbulent orographic form drag. Form drag effects are represented either with an effective roughness length approach (Wood and Mason 1993), which enhances the roughness proportionally to the topographic height, or with an explicitly distributed form drag (Beljaars et al. 2004). Along with the evolution of orographic drag parametrizations, the grid-length (or resolution) of weather and climate models has also continuously increased over time, meaning that more and more of the orographic scales are now resolved explicitly (Bauer et al. 2015, Kanehama et al. 2019).

Despite these advances, the representation of orographic drag processes in weather and climate models remains uncertain and poorly constrained by both theory and observations. Recent community efforts have aimed at assessing and quantifying these uncertainties, as well as finding avenues for improving the representation of these processes (Sandu et al. 2019). Under the framework of the Working Group on Numerical Experimentation (WGNE) of the World Meteorological Organization, a first drag project compared the subgrid surface stresses from major operational global numerical weather prediction (NWP) systems and revealed that these differ substantially, particularly over land (Zadra 2013). The NWP systems differ not only in their total subgrid surface stress, but also in their partition between different parametrized processes. A second model comparison, which focused on the subgrid orographic fields used in the low-level blocking and gravity wave parametrizations (Elvidge et al. 2019), showed that inter-model differences in these fields are of first order importance for the differences in subgrid surface stress found in the initial WGNE drag project. More recently, the WGNE/GASS project on COnstraining ORographic Drag Effects (COORDE, van Niekerk et al. (2020)) compared various orographic drag parametrizations used in global NWP systems, and demonstrated that km-scale simulations can provide useful insights into the deficiencies of low-level flow blocking and gravity wave drag parametrizations. In addition, the implications of the effective resolution (Abdalla et al. 2013, Lander and Hoskins 1997, Skamarock 2004) for subgrid orographic drag parametrizations have been discussed (Vosper 2015), and an approach to estimate the effective orographic resolution has been proposed (Kanehama et al. 2019).

In this paper, we evaluate the orographic drag parametrizations in the Integrated Forecasting Systems (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) and optimise particular uncertain parameters therein, building on the methodology of the COORDE project (van Niekerk et al. 2020) and on lessons learned from recent community efforts on orographic drag processes (Elvidge et al. 2019, Kanehama et al. 2019, van Niekerk et al. 2018, Vosper 2015, Vosper et al. 2019). After a brief description of the orographic drag parametrizations used in the IFS (Section 2), we illustrate their deficiencies using COORDE-like experiments and we motivate recent changes made to alleviate them (Section 3). These consist of generating the subgrid orographic fields so that they represent scales smaller than the effective orographic resolution, rather than the grid-length, and retuning uncertain parameters within the low-level blocking and turbulent orographic drag parametrizations. Details of these changes, along with a first evaluation of their impact on the circulation will be described in Section 4. The impact of the revised representation of the orographic drag on the forecast performance at various time-scales, ranging from day to seasons, is then discussed in Section 5, before drawing conclusions (Section 6).

2 Orographic drag parametrizations in the IFS

To help the understanding of the proposed changes to the representation of orographic drag processes within the IFS, we first briefly describe the orographic drag parametrizations used.

The effects of orographic drag on atmospheric flow are parametrized separately depending on the scales

of the unresolved orography. Thus, it is assumed that only hydrostatic waves are generated by subgrid orography. Based on linear wave theory of flow over orography in typical winds speeds of order 10 m/s and Brunt-Vaisala frequency of 0.01 s^{-1}, only orography of scales of approximately 5 km in the horizontal can generate vertically propagating gravity waves (ECMWF 1997). As a result, both the orographic gravity wave and flow blocking drag components account for scales larger than 5 km. These effects are represented using the Lott and Miller (1997) scheme, referred to hereafter as the Subgrid Scale Orography scheme (SSO). Scales smaller than 5 km are represented using the Turbulent Orographic Form Drag (TOFD) scheme (Beljaars et al. 2004). For more details on these schemes, see IFS documentation (2019).

2.1 Turbulent Orographic Form Drag

The TOFD scheme applies drag explicitly on model levels in the equation for the horizontal wind components, e.g. the change in zonal wind U due to TOFD is:

$$\frac{\partial U}{\partial t} = \frac{\partial}{\partial z}\frac{\tau_x}{\rho} = -C_{\text{tofd}}|\mathbf{U}(z)|U(z),$$
$$C_{\text{tofd}} = -\alpha\beta C_{\text{md}} C_{\text{coor}} 2.109 e^{-(z/1500)^{1.5}} a_2 z^{-1.2} \tag{1}$$

with τ_x being the stress, ρ the density, z the height above the surface, $\alpha = 35$, $\beta = 1$, $C_{\text{md}} = 0.005$, $C_{\text{coor}} = 0.6$, $a_2 = a_1 k^{n_1 - n_2}$, $a_1 = \sigma_{\text{flt}}^2 (I_H k_{\text{flt}}^{n_1})^{-1}$, $k_1 = 0.003\,\text{m}^{-1}$, $n_1 = -1.9$, $n_2 = -2.8$, $k_{\text{flt}} = 0.00035\,\text{m}^{-1}$, $I_h = 0.00102\,\text{m}^{-1}$ and σ_{flt} is the standard deviation of filtered subgrid orography (to remove scales larger than 5 km). A corresponding equation is used for the meridional wind V. For numerical stability, these equations are solved implicitly, with the absolute wind speed $|U|$ taken from the previous time level and the $U(z)$ and $V(z)$ components evaluated at the new time level. In addition, the drag coefficient is passed to the diffusion solver, where the changes to the momentum due to TOFD, blocking and turbulent diffusion are solved together implicitly.

2.2 Subgrid scale orographic drag

The Lott and Miller (1997) SSO scheme used in the IFS relies on the concept of anisotropic orography proposed by Baines and Palmer (1990). The SSO scheme represents two effects: the gravity wave drag exerted by the breaking of orographically generated vertically propagating gravity waves and flow blocking drag due to the deflection of the flow by the orography at low levels.

2.2.1 Blocked flow drag

The SSO scheme assumes that the flow is deflected laterally when encountering a mountain barrier because it does not have enough energy to pass over it, either because winds are too weak or the atmospheric stability is too strong. The blocking height Z_{blk} is defined as the maximum model level below the mountain height for which the non-dimensional mountain height reaches a critical value $H_{n_{\text{crit}}}$, i.e

$$\int_{Z_{\text{blk}}}^{3\mu} \frac{N}{U_p} dz \geq H_{n_{\text{crit}}} \tag{2}$$

where N is Brunt-Vaisala frequency, U_p is the wind speed in the direction of the incident flow, and z is height above mean orography. The statistical mountain height is selected to be $H = 3\mu$, where μ is

the standard deviation of the subgrid orographic heights. $H_{n_{\text{crit}}}$ is set to 0.5. The vertical distribution of blocked flow drag and its dependendence on wind direction are based on the assumption that subgrid orography consists of a collection of ellipse shaped hills with average slope σ, orientation θ and anisotropy γ. The blocked flow drag D_{blk} is directly applied to model levels using:

$$D_{\text{blk}}(z) = -C_d \max\left(2-\frac{1}{r},0\right)\rho\frac{\sigma}{2\mu}\left(\frac{Z_{\text{blk}}-z}{z+\mu}\right)^{1/2}(B\cos^2\psi + C\sin^2\psi)\frac{U|U|}{2} \tag{3}$$

where C_d is a (constant) drag coefficient, r is the aspect ratio of the elliptic mountain as seen from the incident flow, ρ is the density, and ψ is the angle between incident flow and principal axis of the elliptic mountains. Following Phillips (1984), B and C are estimated as:

$$B = 1-0.18\gamma-0.04\gamma^2, C = 0.48\gamma+0.30\gamma^2 \tag{4}$$

In the IFS r is formulated[1] as

$$r = \sqrt{\frac{\cos^2\psi+\gamma^2\sin^2\psi}{\gamma^2\cos^2\psi+\sin^2\psi}} \tag{5}$$

The change in the zonal wind due to the low-level flow blocking drag is given by

$$\frac{\partial u}{\partial t} = \frac{D_{\text{blk}}}{\rho} \tag{6}$$

In the IFS, the blocking coefficient C_d is currently set to 2.

For numerical stability, as for TOFD, eq. 6 is solved implicitly so that the absolute wind speed $|U|$ is taken from the previous time level and the U and V components are evaluated at the new time level. Similarly to TOFD, the drag coefficient is passed to the diffusion solver, where the changes to momentum due to TOFD, blocking and turbulent diffusion are solved together implicitly.

2.2.2 *Gravity wave drag*

The gravity wave drag part of the Lott and Miller (1997) scheme is based on Palmer et al. (1986) and Baines and Palmer (1990). As the flow below Z_{blk} is blocked and is assumed not to be able to cross the mountain barrier, the amplitude of the gravity waves is considered to be equal to the so called effective mountain height $H_{\text{eff}} = H - Z_{\text{blk}}$. Since IFS cycle 32r2, H_{eff} is multiplied by a factor of 2 in order to adjust gravity wave stress, i.e $H_{\text{eff}} = 2(H - Z_{\text{blk}})$. Following Phillips (1984), the gravity wave surface stress (τ_1, τ_2) in the parallel and perpendicular directions to the incident flow is written as

$$(\tau_1,\ \tau_2) = \rho_l u_l N_l \frac{{H_{\text{eff}}}^2}{9}\frac{\sigma}{\mu}G\left\{B\cos^2\psi_l + C\sin^2\psi_l,\ (B-C)\sin\psi_l\cos\psi_l\right\} \tag{7}$$

where G is a function of mountain sharpness and the subscript l denotes low-level mean between μ and 2μ. Note that equation 7 is divided by 9 rather than 4 which was originally proposed in Lott and Miller (1997) to scale equation 7 with the maximum value of H (3μ when Z_{blk} is zero). Upward propagation of the gravity waves, with a critical level filtering and a saturation of the amplitude is treated in the same manner as in Palmer et al. (1986).

[1] it was previously formulated as $r = \frac{\cos^2\psi+\gamma\sin^2\psi}{\gamma\cos^2\psi+\sin^2\psi}$, but this error was fixed in the current study and the impact of this change was shown to be negligible.

2.2.3 *Subgrid orography fields*

The Lott and Miller (1997) scheme assumes that the subgrid orography is represented by elliptic mountains, characterised by four parameters: standard deviation μ, slope σ, anisotropy γ and angle (or orientation) θ. These characteristics of the subgrid scale orography (the SSO climate fields) are derived from an original 30" (approximately 1-km) orography dataset, averaged at a 2'30" resolution (approximately 5 km). In the current version of the IFS, the SSO fields are constructed to represent scales between 5 km and the grid-length, by taking the difference between this 2'30" orography field and a filtered version of it in which scales smaller than the grid-length are removed by applying a 1-dx filter (where dx is the grid-length). These orography differences at 2'30" resolution are then aggregated to derive the four fields at the respective grid-box resolution (see IFS documentation (2019) for more details).

3 Motivation for the orographic drag revisions: COORDE and effective orographic resolution

In this study, we exploit two recently proposed approaches to evaluate and revise the representation of orographic drag processes in the IFS. One approach is the framework to constrain orographic drag parametrizations using high-resolution (regional) simulations proposed in the COORDE project (van Niekerk et al. 2018; 2020). The other approach allows us to estimate the effective orographic resolution of a model by analysing the resolved mountain torque obtained from a set of simulations with different atmospheric and orographic resolutions (Kanehama et al. (2019), which follows on from the earlier studies of Brown (2004), Davies and Brown (2001).

The basic idea of the COORDE experimental design is the following. A high-resolution (km-scale) model resolves, to a large extent, orographic effects such as low-level flow blocking and gravity wave drag. The high resolution model is run with two different orographic resolutions, one that is the equivalent to the grid-length and one that is coarser than the grid-length. The difference between these two simulations then provides an estimate of the impact from orographic scales ranging from those close to the grid-length to those of the coarser orography. Such high-resolution simulations (e.g. 1.8 to 10 km), with high (e.g. 1.8 to 10km) and low resolution (100 to 150 km) orography, were performed in COORDE with several models over the Middle East mountain chains (see Fig. 1 of van Niekerk et al. (2018)), and were used to estimate the impact of resolved orography on the flow. This impact of the resolved orographic drag was then compared to the impact of parametrized low-level flow blocking and gravity wave drag. The impact of parametrized drag was deduced by taking the difference between two simulations performed at low (100 km) resolution with and without the flow blocking and gravity wave drag parametrizations turned on. This comparison allows us to constrain the low-level flow blocking and gravity wave drag in the low resolution simulation. It also allows us to indirectly constrain, to some extent, the partition between low-level flow blocking and turbulent orographic form drag. Since multiple international modelling groups have participated in the COORDE model comparison, various interesting characteristics and deficiencies of the orographic drag parametrizations have been found and are summarised in van Niekerk et al. (2020).

In this study we employ the COORDE methodology and experimental design to evaluate the performance of the orographic drag parametrizations used in the IFS over a range of resolutions (10 to 150 km). The high-resolution simulations performed in COORDE only provided constraints on the amount of drag which should be parametrized at resolutions of about 100 km. In order to estimate the amount of drag which should be parametrized at resolutions currently used at ECMWF for medium and extended-range

forecasts (9 to 40 km), we have performed additional km-scale simulations with the Unified Model of the UK Met Office (MetUM) following the COORDE experimental design. We have performed 1.8 km regional simulations over the Middle East region, using different orographic resolutions: 1.8, 9, 40, and 150 km. As in COORDE, these MetUM simulations were performed for 1 to 14 January 2015, and consist of 24 hours forecasts initialised daily at 00UTC from the ECMWF operational analysis. Also as in COORDE, the parametrizations of flow blocking and gravity wave drag, which are considered to be resolved at 1.8 km, were turned off while the parametrization of turbulent orographic form drag, which is still not completely resolved at this resolution, was turned on in these simulations.

The difference between the simulations with 1.8 km orography and the simulations with coarser orographies allows us to estimate the impact of resolving orographic features with scales between 1.8 to 150 km, 1.8 to 40 km and 1.8 to 9 km, respectively, on the atmospheric flow. The impacts of the resolved drag due to these different orographic scales on the zonal winds over the Middle East region are shown in Figure 1. A strong deceleration of the zonal winds can be seen in the lower atmosphere near orography, due to low-level flow blocking, and in the upper troposphere to lower stratosphere, due to gravity wave breaking. The spatial distribution of this deceleration is qualitatively similar for the three ranges of orographic scales, but the magnitude decreases relatively linear as more scales are included in the resolved orography. For example, the deceleration due to resolved drag from orographic scales between 9 and 1.8 km is weaker than that due to drag from scales between 40 and 1.8 km. This corroborates our previous findings that all orographic scales have similar impacts on the large-scale atmospheric circulation during the NH winter (Kanehama et al. 2019).

Figure 2(a-c) shows the impacts of parametrized drag on the zonal wind over the Middle East region in the IFS, at three resolutions which closely match the low resolution orographies used in the 1.8 km MetUM simulations. These impacts were deduced from global short-range (24 hours) IFS forecasts with and without the SSO scheme, initialized daily at 00UTC from 1 to 14 January 2015 from the ECMWF operational analysis. These forecasts were run using a cubic octahedral grid (TCo) and the following spectral truncations: 79, 319 and 1279 (Malardel et al. 2016). The horizontal resolution of these forecasts is about 144, 36 and 9 km in the mid-latitudes. TCo319 is the resolution used for operational ECMWF extended-range and seasonal forecasts and TCo1279 is the resolution used for operational deterministic ten-day forecasts. TCo79 is not used at ECMWF but is equivalent to the resolution used for EC-Earth climate simulations. As in the km-scale MetUM simulations, the TOFD scheme was on in all IFS simulations.

If the SSO scheme was perfectly simulating low-level flow blocking and gravity wave drag effects, the impacts on the zonal winds seen in Figure 2 (a-c) should match those of the resolved orography shown in Figure 1. However, at TCo79, the deceleration of the zonal winds in the stratosphere due to the SSO scheme appears to be too weak in the IFS, and there is no acceleration of the zonal wind on the northward flank of the mountains in the lower troposphere, corroborating the findings of van Niekerk et al. (2018; 2020). These deficiencies of the SSO scheme are also apparent at higher resolutions (TCo319 and TCo1279). Moreover, the low-level deceleration is also too weak at TCo319 and virtually non existent at TCo1279. The low-level flow blocking and gravity wave drag parametrization appears thus to have almost no impact on the zonal winds over this region at TCo1279.

van Niekerk et al. (2020), Vosper et al. (2019) showed that the underestimation of the gravity wave drag in the stratosphere, and the very weak dependence of the gravity wave drag on the horizontal resolution is related to the formulation of the Lott and Miller (1997) scheme. The scheme assumes that all the elliptical mountains within a grid-box have the same length-scale, with this scale being close to the grid-length. Since the average gravity wave surface stress within a grid-box is inversely proportional to this length-scale, this results in an almost constant gravity wave drag when averaged over a particular

region, such as the Middle East. This underestimation of the zonal wind deceleration in the stratosphere, and that close to the surface at higher resolution, may also be exacerbated by the fact that the current scheme only represents effects from orographic scales between 5 km and the grid-length. However, given that the scales credibly resolved by the model's dynamical core are generally larger than the nominal grid-box resolution, there are orographic scales that are neither directly resolved nor parametrized. As discussed in Vosper et al. (2016), these scales lie between the grid-box resolution and the effective orographic resolution. Vosper (2015) has demonstrated that the momentum balance has been improved when orographic drag due to these missing scales was included in regional simulations with the MetUM over South Georgia, and suggested that orographic drag parametrizations should represent scales smaller than the effective resolution rather than the grid-length. In practice, this means that the subgrid scale orographic fields used in the low-level blocking and gravity wave drag parametrizations should be defined such that they represent characteristics of orographic scales smaller than the effective resolution. One issue, however, is that there is no consensus on how to estimate the effective orographic resolution.

Following ideas from Davies and Brown (2001), Kanehama et al. (2019) recently demonstrated that the effective orographic resolution of a model can be estimated by analysing the changes in the resolved orographic torque from a series of simulations with varying orographic and atmospheric resolutions. As the atmospheric resolution is gradually increased, while the orographic resolution remains unchanged, the resolved orographic torque increases, since the mountains are more accurately resolved. The effective resolution can then be estimated as the atmospheric resolution at which the resolved orographic torque no longer changes. Figure 3 illustrates that the effective orographic resolution for the IFS, when using a cubic discretization, is around 4 times the grid-box resolution (4-dx). For a more detailed discussion see Kanehama et al. (2019).

In light of these findings, our strategy for revising the representation of orographic drag processes in the IFS is as follows. We regenerate the SSO fields on the effective orographic resolution (4-dx) and use the COORDE-type experiments described above to assess the impact of using these fields, instead of the current fields constructed on the grid-box (1-dx) resolution, on the zonal winds over the Middle East region. We then evaluate the impact on forecast skill at several resolutions and retune some of the constants used in the SSO and TOFD parametrizations.

(a) 1.8 km - 150 km (b) 1.8 km - 40 km (c) 1.8 km - 9 km

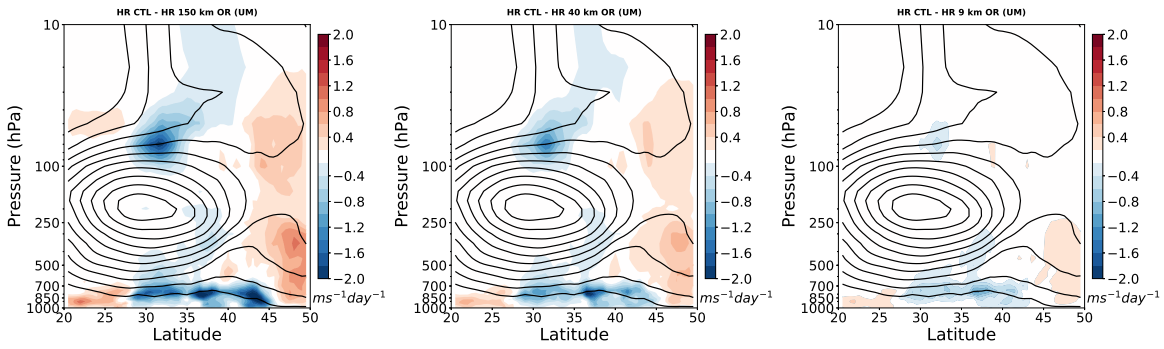
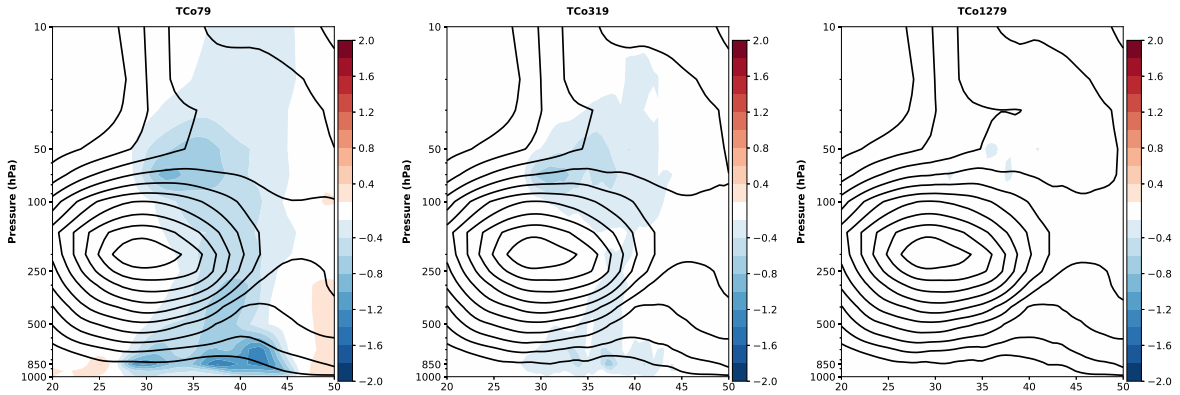
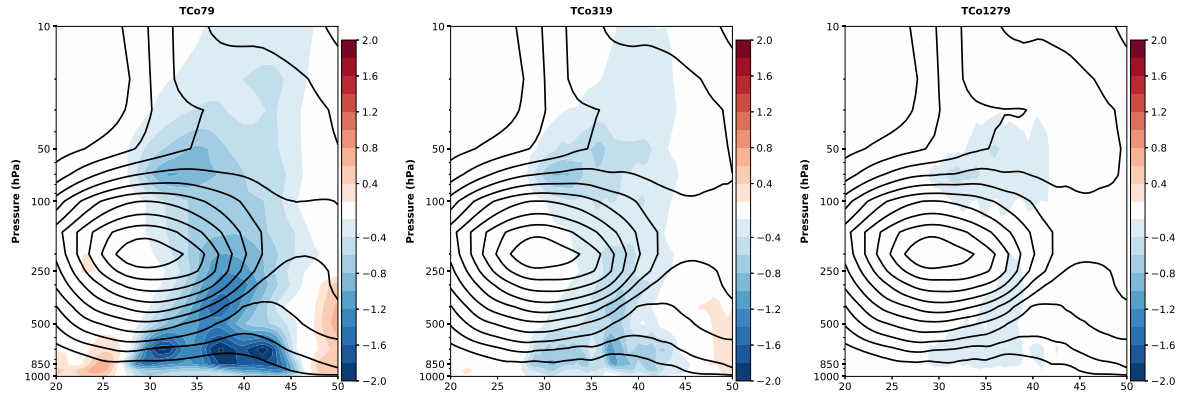


Figure 1: The impact of resolved drag on the zonal winds [m/s/day, shading] over the Middle East region due to orographic scales comprised between (a) 1.8 km - 150 km, (b) 1.8 km - 40 km, and (c) 1.8 km - 9 km, derived from the difference between the 1.8 km MetUM simulations with high (1.8 km) and low resolution (9, 40, 150 km) orographies. The values represent averages over the set of 14 forecasts, at the end of the 24 hours of simulation, and are longitudinally averaged over the Middle East region. Contours show the zonal wind for the runs with 1.8 km orography and contour intervals are 10 m/s.

(a) 144km, control (b) 36km, control (c) 9km, control



(d) 144km, SSO4dx (e) 36km, SSO4dx (f) 9km, SSO4dx



(g) 144km, final (h) 36km, final (i) 9km, final

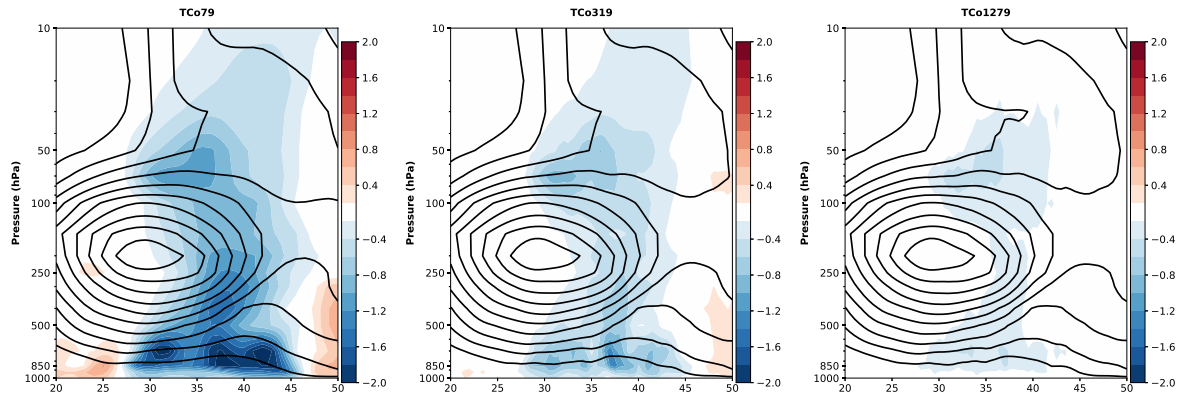


Figure 2: The impact of the SSO scheme on the zonal winds [m/s/day, shading] derived by taking the difference between IFS forecasts with and without SSO scheme over the Middle East region at TCo79 (144 km, left), TCo319 (36 km, center) and TCo1279 (9 km, right) (a,b,c) for the control orographic configuration, (d,e,f) for the forecasts using the 4-dx SSO fields, and (g,h,i) for the final orographic drag configuration. The values represents averages over the set of 14 forecasts, at the end of the 24 hours of simulation and are longitudinally averaged over the Middle East region. Contours show the zonal wind for the runs with SSO scheme and contour intervals are 10 m/s.

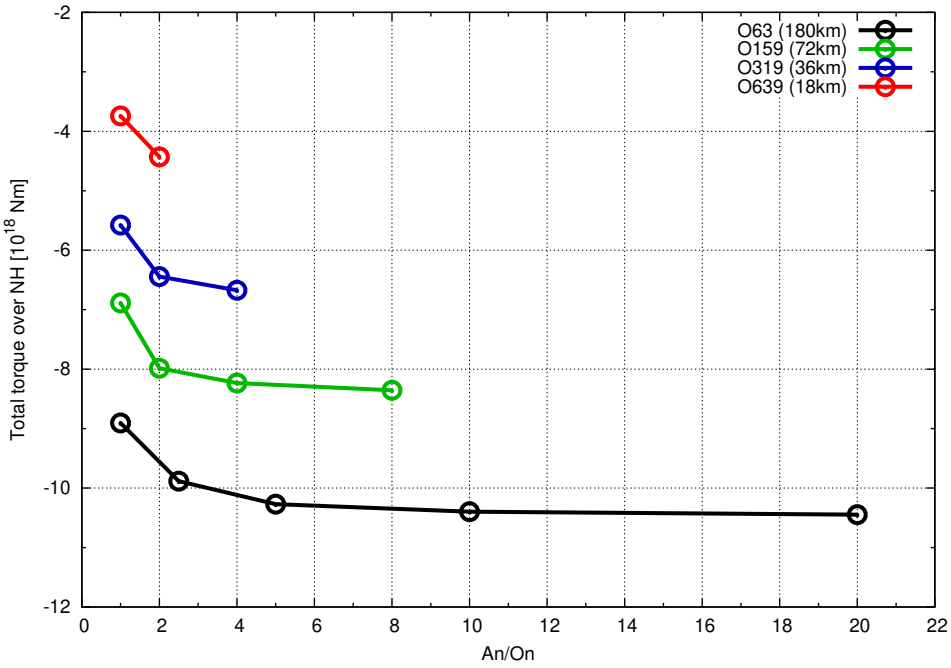


Figure 3: The total resolved torque [N m] over the Northern Hemisphere, averaged over the first 24-hours of the IFS forecasts, as a function of the ratio of atmospheric and orographic resolutions ranging from TCo63 (180 km) to TCo1279 (9 km). This is derived from the experiments with a varying atmospheric resolution and a constant orographic resolution (indicated in the legend), with the SSO and TOFD schemes switched off, described in Kanehama et al. (2019). Adopted from Figure 4. of Kanehama et al. (2019).

4 SSO fields at the effective orographic resolution and their impact

In this section we discuss how the SSO fields were changed to represent scales up to the effective resolution, as opposed to the grid-length. We also illustrate the implications of this change for the subgrid orography characteristics, for the zonal winds over the Middle East region, and more generally for forecast skill in the NH during winter.

4.1 SSO fields at the Effective Orographic Resolution

As shown in Figure 3, the effective orographic resolution can be approximated as 4-dx for the cubic discretisation of the IFS (Kanehama et al. 2019). To regenerate the SSO climate fields on this effective orographic resolution, we used a 4-dx filter instead of a 1-dx filter when creating the filtered 2'30" orography field which is substracted from the original orography (see Sect 2.2.3). This smooths out the filtered orography field, and results in larger subgrid mountain heights. This results in an increase of the standard deviation and slope of the subgrid orography, which, as shown by Elvidge et al. (2019), are the parameters that contribute most to the surface stress of the SSO scheme and impact the large scale circulation.

The increase in the standard deviation of the subgrid orography when using a 4-dx instead of a 1-dx filter is illustrated using a probability density function, which shows increases in the higher values (Figure 4).

Figure 5 also shows increased values of the standard deviation of the subgrid orography, as well as an increased regional extent of these larger values, over the Tibetan Plateau. A similar behaviour is found for the SSO slope (not shown). The changes in the anisotropy and angle fields are less marked, and are less relevant because they have a negligible impact on the overall forecast performance (Elvidge et al. 2019). In the process of changing the SSO fields to represent scales less than 4-dx, several minor bugs were found and fixed in the climate generation software at ECMWF, in particular to the definition of the anisotropy. The impact of these corrections on the large-scale forecast skill was found to be negligible and they are included in the 4-dx version of the SSO fields.

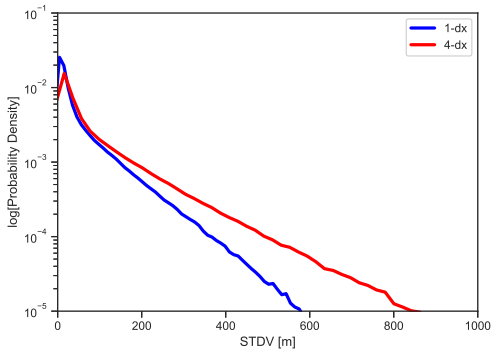


Figure 4: Probability density function of the standard deviation of the SSO height at TCo639 (approximately 18 km resolution in the mid-latitudes, used for operational medium-range ensemble forecasts), constructed with a 1-dx (blue) and a 4-dx filter (red). Note that only land grid points are included.

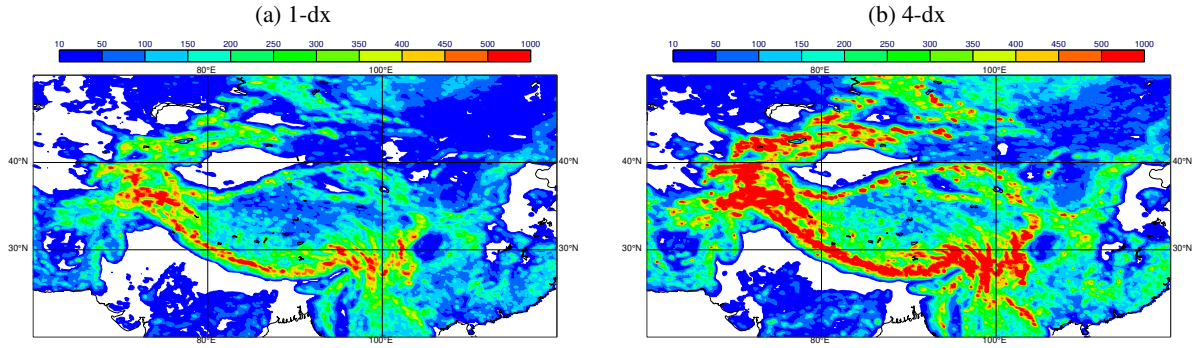


Figure 5: Maps of the standard deviation [m] of SSO height at TCo639, constructed with (a) 1-dx and (b) 4-dx filter over the Tibetan Plateau.

4.2 Impact of the SSO fields at effective orographic resolution

The impact on circulation from using the 4-dx SSO fields, via the changes in the parametrized drag from the SSO scheme, is first analysed in the COORDE framework. For the COORDE period 24 hours forecasts were thus performed using the new SSO fields derived with the 4-dx filter (labelled SSO4dx hereafter) at TCo79, TCo319 and TCo1279. The impact of the parametrized drag on the zonal winds over the Middle East region was deduced by taking the difference with the corresponding IFS runs without the SSO scheme at these resolutions (Figure 2d,e,f). The larger SSO standard deviation and slope in the SSO4dx runs lead to an increase of the parametrized SSO drag (particularly of the low-level blocking component) compared to the control IFS configuration, and result in a larger impact on the zonal winds.

Comparing the middle and top rows of Figure 2, it appears that the deceleration of the zonal winds, both near the surface and in the stratosphere, is stronger in the SSO4dx runs bearing a closer resemblance to that from the resolved drag (Figure 1). At TCo79 and TCo319, there is also a hint of the acceleration on the northern flank of the mountain chains in the SSO4dx runs, which was absent in the control IFS runs. The deceleration of the winds throughout the troposphere is also stronger than in the control runs at all resolutions. This deceleration was already too strong in the control runs compared to the km-scale simulations, particularly at TCo79. As discussed in van Niekerk et al. (2018; 2020), this deceleration of the tropospheric zonal wind in the parametrized runs is due to a response of the dynamics to the use of the SSO scheme. In the SSO4dx runs, the overestimation of this tropospheric deceleration becomes even stronger, particularly at TCo79 and TCo319. This is likely related to the findings of van Niekerk et al. (2016) that the low-level blocking drag varies too rapidly with resolution.

This comparison of the parametrized and resolved orographic drag suggests that the 4-dx SSO fields allow us to better capture key features of the impact of orographic drag on the zonal winds over the Middle East region, i.e. the deceleration near the surface and in the lower stratosphere, across considered resolution range (9 to 144 km). However they lead to an overestimation of the tropospheric impact, particularly at the lowest resolutions evaluated here (TCo79 and TCo319).

The impact of the 4-dx SSO fields was then evaluated by assessing the changes in forecast skill at TCo639 (18 km), the resolution of ECMWF's operational medium-range ensemble forecasts. Forecast only experiments, initialized from the operational ECMWF analysis, were performed at this resolution, both in deterministic mode for the COORDE period and in ensemble mode for a more recent winter period. In the ten-day deterministic forecasts, the use of the 4-dx SSO fields has a positive impact throughout the troposphere and particularly in the stratosphere in the NH mid-latitudes, where the largest mountain chains are situated, but leads to a deterioration of the geopotential height in the Tropics. As illustrated in Figure 6a, both the positive and negative impacts in the mid-latitude NH and Tropics, respectively, are statistically significant in the medium-range. The ensemble forecast skill was also deteriorated in certain regions, and again particularly in the tropics (not shown).

These results suggest that using the 4-dx SSO fields with the current parameter choices for the SSO and TOFD schemes lead to too much orographic drag. Figure 7 illustrates that the use of the 4-dx SSO fields results in a significant increase in the contribution of the SSO scheme to the surface stress over land (100-150%) and an increase in the zonally averaged total parametrized surface stress of about 10-15 % in the NH mid-latitudes. This is commensurate with the differences in surface stress between the IFS and the MetUM over this region found in the WGNE drag project and discussed in Sandu et al. (2016). As shown in Elvidge et al. (2019), these differences can to a large extent be explained by the differences between the SSO fields used in the IFS and the MetUM (e.g. related to differences in filtering, orography datasets).

4.3 Retuning of the TOFD and SSO schemes

This first evaluation of the impacts of using the 4-dx SSO fields in the IFS motivated a retuning of certain parameters in the SSO and TOFD schemes. The aim was to mitigate the deterioration in skill in the Tropics found in both the deterministic and ensemble configurations, while maintaining the improvements in the mid-latitude NH, and the closer match of the impact of the parametrized drag with that of the resolved drag over the Middle East region.

The tuning consisted of reducing the strength of the form drag and enhancing the low-level flow blocking drag. This was motivated by the fact the α coefficient in the TOFD scheme, which controls the strength

of the form drag, is currently set to 35, while a value of 12 had been proposed in the original paper by Beljaars et al. (2004). The coefficient α has been increased over time to optimise forecast skill. Historically, α and C_d (the coefficient which controls the strength of the flow blocking in the SSO scheme), were increased from 27 to 35, and from 1.5 to 2, respectively, in IFS cycle 40r1 to alleviate the negative impacts on forecast skill of a reduction of the turbulent diffusion in stable layers (Sandu et al. 2013). This increase of the coefficients controlling the magnitude of orographic drag was motivated by the fact that both the diffusion in stable layers and orographic drag affect the large-scale circulation in the NH during winter, through their influence on surface stress. A reduction in surface stress associated with the reduction of the turbulent diffusion maintained in stable conditions resulted in a marked (few percent) increase in the root mean square error (RMSE) of the geopotential height in the NH midlatitudes during winter throughout the medium-range. Compensating this decrease in surface stress through an increase of the contributions to the surface stress of the TOFD and the SSO schemes helped overcome this deterioration, and even led to an improved forecast skill in the NH throughout the medium-range (Sandu et al. 2014). In retrospect, the changes made in cycle 40r1 represented a repartition of the contributions to surface stress due to various processes. At the time this repartition was purely made to optimise skill, in the absence of any constraints on the drag processes or on their partition. As later shown by the WGNE drag project, it resulted in IFS having much more stress due to TOFD and less stress due to the SSO scheme than the MetUM (Sandu et al. 2016), and other models (van Niekerk et al. 2020).

The increase in orographic drag due to the use of the new 4-dx SSO fields allows us to reduce to some extent the contribution of the TOFD scheme. A range of smaller values for α were trialed and ultimately a value of 24 was selected based on a classical forecast skill optimization exercise. Even when using the 4-dx SSO fields, a value of 12 for α, as initially proposed by Beljaars et al. (2004), still leads to a deterioration of the forecast skill throughout the troposphere and stratosphere in the NH winter.

In parallel, the SSO code was cleaned to bring it closer to the original Lott and Miller (1997) scheme, by removing certain arbitrary choices made over the years for some parameters or in the computations of some levels. To compensate for a loss in forecast skill due to this code cleaning, the value of the C_d coefficient in the blocked flow drag part of the SSO scheme was increased again from 2 to 3.

In summary, the SSO orographic drag was thus increased through the regeneration of the SSO fields to include scales than the (4-dx) effective orographic resolution and through the cleaning and retuning of the SSO scheme, while the turbulent orographic form drag was decreased. The final orographic drag configuration evaluated hereafter, includes use of the 4-dx SSO fields, the cleaned SSO code, and the revised values of $\alpha = 24$ and $C_d = 3$. As shown in Figure 7, this configuration decreases the boundary layer surface stress (given by the sum of the turbulent stress and TOFD stress) and increases the SSO surface stress, bringing the IFS surface stress partition closer to that in the MetUM. In the NH midlatitudes, the final configuration also increases the total parametrized surface stress (by 4-8 %), but less than when using only the 4-dx SSO fields.

The impact of the parametrized drag on zonal winds over the Middle East region obtained with the final configuration looks broadly similar to that obtained when using only the 4-dx SSO fields (bottom versus middle row in Figure 2). Compared with the current configuration of the IFS, it better matches the deceleration near the topography and in the stratosphere, seen in the km-scale simulations, while it overestimates more markedly the tropospheric deceleration, particularly at TCo79. The overestimation in the final configuration at this lowest resolution will also become apparent in the next section when discussing the atmospheric angular momentum budget. However, given that optimizing parameters in the different drag schemes across resolutions is extremely difficult, especially since these schemes do not necessarily behave well for a wide range of resolutions (van Niekerk et al. 2016, Vosper et al. 2016), our focus here was on an improving performance at resolutions used operationally at ECMWF, i.e. resolutions beyond

TCo319.

Figure 6 suggests that the final configuration mitigates, to a large extent, the deterioration in forecast skill seen in the TCo639 SSO4dx runs in the Tropics and NH high-latitudes. In the next section, we will present the results of an extensive performance evaluation of this configuration across resolutions and timescales.

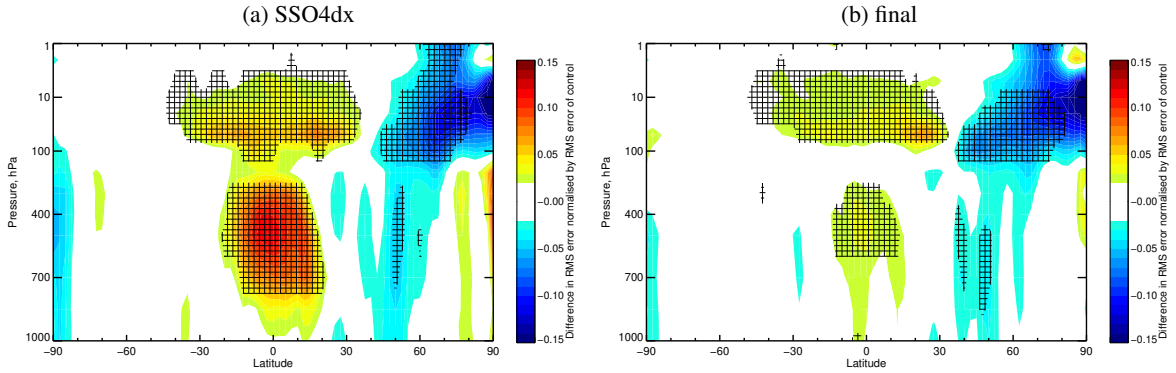


Figure 6: Normalized difference in root mean square error (RMSE) of geopotential height between the TCo639 forecasts using (a) the 4-dx SSO fields and (b) the final configuration and the control TCo639 forecasts for January 2015, at T+120 [hr]. The normalized difference in RMSE is defined as the difference between the RMSE of the sensitivity experiment and the RMSE of the control experiment divided by the RMSE of the control. Statistically significant ($\geq$ 95%) regions are hatched. Warm (cool) colours indicate an increase (decrease) in RMSE in the experiments with modified orographic drag with respect to the control.

5 Performance evaluation of the orographic drag revision

The performance of the final orographic drag configuration described above was evaluated for various applications used at ECMWF. In this section, we first describe the experiments performed and then discuss the results obtained for the different forecast and data assimilation configurations.

5.1 Experiments

The experiments performed with the final and control orographic drag configurations are summarised in Table 1. Although two different versions of the IFS were used for the experiments discussed below, the configurations of the turbulence, SSO and TOFD schemes are the same in these two cycles, which means that the impacts of the changes to the orographic drag are very similar. While for operational forecasts the IFS is coupled to an ocean and sea-ice model (NEMO and LIM2), most of the experiments performed here (except for the medium-range ensemble experiments) were not coupled, to save computational resources. A test in forecast-only mode has shown that, at least in ten-day forecasts, the impact of the drag changes discussed here is similar in coupled and uncoupled mode (not shown).

Ten-day forecast experiments, initialized daily at 00UTC from the operational ECMWF analysis (forecast-only experiments), were performed at various resolutions that are used in standard IFS applications. TCo1279 and TCo639 are the resolutions used for operational high-resolution deterministic ten-day fore-

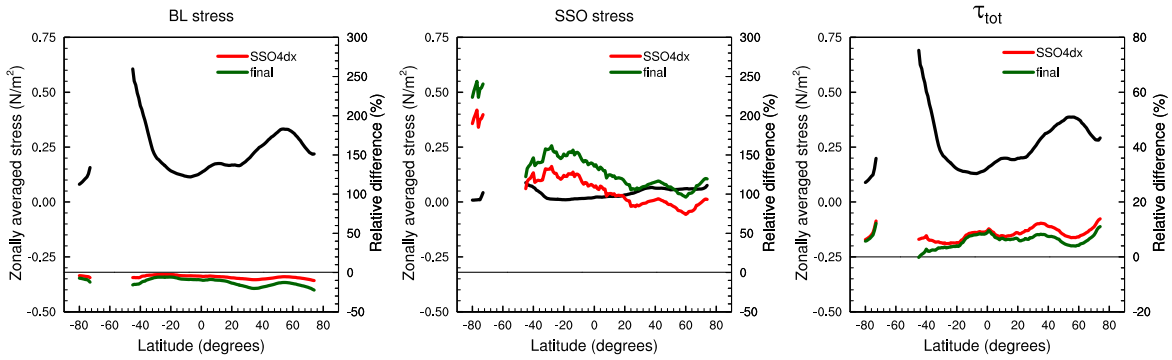


Figure 7: Zonal averages of the magnitude of the total parametrized surface stress τ_{tot} [N/m^2] (left) and of its boundary layer (turbulent+TOFD, middle) and SSO (right) components, for TCo639 forecasts for January 2015 using the control IFS configuration (black lines). Relative differences (%, right axis) in the total stress and in its components between the experiments using the 4-dx SSO fields and the final configuration for orographic drag and the control experiment are overlaid in red and green, respectively. The stresses represent averages over the 31 forecasts and over the first 24 hours each forecast. The zonal averages include only land grid cells and were smoothed using a 5 degree running mean. The lines present gaps at ocean only latitudes.

casts and medium-range ensemble forecasts, respectively. TCo319 is the resolution used for operational extended-range (days 15 to 46) and for seasonal forecasts (up to one year ahead). TCo199 and TCo399 are used for testing the impact of upcoming changes to the IFS on the extended-range forecasts, and on data assimilation and ensemble forecasts, respectively.

Data assimilation experiments were performed at TCo399 due to computational constraints. Apart from being at a lower resolution, these experiments mimic the set-up of ECMWF operational deterministic ten-day forecasts. Ten-day forecasts are initialized each day at 00 and 12UTC from their own analyses. In all data assimilation experiments, the outer loops of the four-dimensional variational data assimilation are using cubic grids (in this case TCo399), while the inner loops are still using linear grids (in this case TL95, TL159 and TL255, see Malardel et al. (2016) for a description of the cubic and linear grids). As the SSO fields were changed to represent scales smaller than the effective orographic resolution (4-dx) only for the cubic grids, this implies that in this experiment the new 4-dx SSO fields are only used in the non-linear model integrations. Tests done to evaluate the impact of also changing the SSO fields for the linear grids (to 2-dx) showed that this has a negligible impact on forecast performance (not shown).

The medium-range ensemble experiments are initialized from the operational ECMWF analysis. Due to computational constraints, they are performed at TCo399 and with 8 members, while the operational medium-range ensemble forecasts are performed at TCo639 with 50 members. The probabilistic performance of the small ensemble is evaluated using a fair continuous ranked probability score (Ferro 2014).

Seasonal integrations are initialized from ERA5 (Hersbach et al. 2020) on November 1st, every year between 1998 and 2016. Again due to computational constraints, this first evaluation at seasonal timescale was carried out in a cheaper configuration than that of ECMWF's operational seasonal prediction system SEAS5 (Johnson et al. 2019). These runs are uncoupled (sea surface temperature and sea ice are prescribed), are shorter (4 instead of 7 months) and have less ensemble members (10 instead of 51) than SEAS5. The horizontal and vertical resolutions are the same as those of SEAS5.

Type	IFS cycle	Horizontal resolution	Period
Forecast-only	45r1	TCo79 (150 km), TCo199 (50 km), TCo319 (32 km), TCo639 (18 km), TCo1279 (9 km)	Jan. 1st – 31st 2015
Data assimilation	46r1	TCo399 (25 km) (outer loops) TL95 – 159 – 255 (inner loops)	Dec. 1st 2018 - Jan. 31st 2019
Medium-range ensemble	46r1	TCo399 (25 km)	Nov. 1st 2018 - Feb. 28th 2019
Seasonal prediction	46r1	TCo319 (32 km)	1st of Nov. 1998 – 2016

Table 1: List of experiments performed with the final and control orographic drag configurations. The resolutions used in each experiment type differ from those in the operational configuration due to limited computational resources. All experiments are performed with 137 levels on the vertical, except for the seasonal experiments which are performed with 91 levels. The seasonal integrations are 4 months long and are performed with 10 ensemble members.

5.2 Ten-day forecast-only experiments at various resolutions

The impact of the final orographic drag configuration in the forecast-only experiments at different resolutions was assessed in terms of both the angular momentum budget and changes in forecast skill. This assessment is presented in the next subsections.

5.2.1 Angular momentum budget

To explore how these these changes to the representation of the orographic drag affect the different terms in the momentum budget, the momentum budget is analysed for the forecast-only experiments conducted at the different resolutions. Given that the SSO and TOFD parametrizations represent the effects of unresolved orographic scales, and that the resolved and unresolved components of the momentum budget should increase and decrease, respectively, as resolution increases, the sum of the orography related terms should be resolution independent. This basic constraint is assessed using the zonally averaged vertically integrated relative angular momentum budget equation (Brown 2004, Sandu et al. 2019)

$$\underbrace{\frac{\partial}{\partial t}\int\left[m_r\frac{dp}{g}\right]}_{\text{TENDENCY}}+\underbrace{\frac{1}{a\cos\phi}\frac{\partial}{\partial\phi}\left[\left(\int m_r v\frac{dp}{g}\right)\cos\phi\right]}_{\text{AMFC}}\underbrace{-fa\cos\phi\left[\int v\frac{dp}{g}\right]}_{\text{CORIO}}$$
$$+\underbrace{\left[p_s\frac{\partial h}{\partial\lambda}\right]}_{\text{RESOLVED}}+\underbrace{a\cos\phi\,[\tau_S]}_{\text{SSO}}+\underbrace{a\cos\phi\,[\tau_T]}_{\text{TOFD}}+\underbrace{a\cos\phi\,[\tau_V]}_{\text{VDF}}=0 \qquad (8)$$

where $m_r = ua\cos\phi$ is the relative angular momentum per unit mass, a is the Earth radius, ϕ is the latitude, u is the zonal wind, v is the meridonal wind, p_s is the surface pressure, f is the Coriolis parameter, g is the gravitational constant, and λ is the longitude. τ_S, τ_T and τ_V are the surface stress terms from SSO, TOFD and turbulence parameterization (VDF). The square brackets [ ] indicate a zonal average. Mathematically, the sum of the left hand side of equation 8 should be zero. However, this is not necessarily the case for a numerical model because of numerical computation errors. The sum of the terms on the left hand side of the equation is in practice close to a small residual.

Figure 8 illustrates the different terms in the zonally averaged vertically integrated relative angular momentum budget equation at TCo639. In winter, the orography related terms (RESOLVED, SSO and

TOFD in Figure 8) peak in the middle latitudes of the Northern Hemisphere because of the large mountain chains such as Himalayas and the Rockies, resulting in a deceleration of the westerly atmospheric flow. Note that a positive (negative) sign indicates a deceleration of westerly (easterly) flow due to surface drag being exerted on the atmospheric flow. Also note that the residual is indeed close to zero.

In Figure 9, we examine how each of the orography related terms in the zonally averaged momentum budget equation varies with resolution in the 40 to 50N latitude band, when the control and final orographic drag configurations are used in the forecast-only experiments from TCo79 to TCo1279. The explicitly resolved torque (red) increases at high resolutions, as more scales of the orography become resolved over this region, and it is very similar in the two orographic drag configurations.

Although both SSO and TOFD account for unresolved orographic effects, their resolution dependence is quite different. The SSO torque decreases rapidly towards high resolutions, as expected, because more of the orographic scales become resolved and less need to be parametrized. This resolution dependence is in the Lott and Miller (1997) scheme by design, through the decrease in the standard deviation and slope of the subgrid orographic features as the resolution increases. As discussed in van Niekerk et al. (2020), van Niekerk and Vosper (2021), Vosper et al. (2019) and mentioned in the previous section, the resolution dependency is much less pronounced for the gravity wave part of the scheme. Previous studies have, however, shown this resolution dependence to be too strong for the blocking component (Sandu et al. 2019, van Niekerk et al. 2016). In contrast, the TOFD torque should be fairly resolution independent, given that it represents effects of orographic features with horizontal scales less than 5 km. Figure 9 shows that, in practice, this is not the case. The TOFD torque increases slightly with resolution from TCo319 to TCo1279, possibly due to the fact that an increasingly resolved orographic variability leads to an increase in wind shear and speed (Beljaars 2020), which then translates to an enhanced form drag. At TCo79, the TOFD and SSO torques are visibly smaller and larger, respectively, than at the other resolutions, corroborating the results of Kanehama et al. (2019) that the two orographic drag schemes may interact in an undesired way at this coarse resolution.

When compared to the control experiment (dotted lines) at the same resolution, the final configuration (solid lines) shows an increase in the SSO torque (blue) because of the use of the 4-dx SSO fields and of an increased blocked flow drag coefficient C_d. It also shows a decrease of the TOFD torque (green) due to the smaller value of the α coefficient in the TOFD scheme and to a deceleration of the zonal winds associated with the increased SSO drag (this is indicated by the decrease of the boundary layer term in the SSO4dx runs, see Figure 7). The resolution dependence of the two terms does not seem to depend on the choice made for the SSO fields and the parameters in the two schemes (dotted and solid lines are parallel).

The total orographic torque (RESOLVED+SSO+TOFD) shows a fairly good resolution independence from TCo319 to TCo1279 for both the control and the final orographic configuration. However, for the final orographic drag configuration, the total orographic torque increases somewhat at the lowest resolutions. This is likely related to the too strong resolution dependency of the blocking component of the Lott and Miller (1997) scheme pointed out by previous studies (Brown 2004, Sandu et al. 2019, van Niekerk et al. 2016). As explained in previous sections, this suggests that an optimization across this entire range of resolution (9 to 150 km) is not straightforward.

5.2.2 Impact on forecast skill

As can be seen from Figure 10, the final orographic drag configuration improves medium-range forecast skill in the NH for geopotential height and temperature from the lower troposphere to 50 hPa, at all

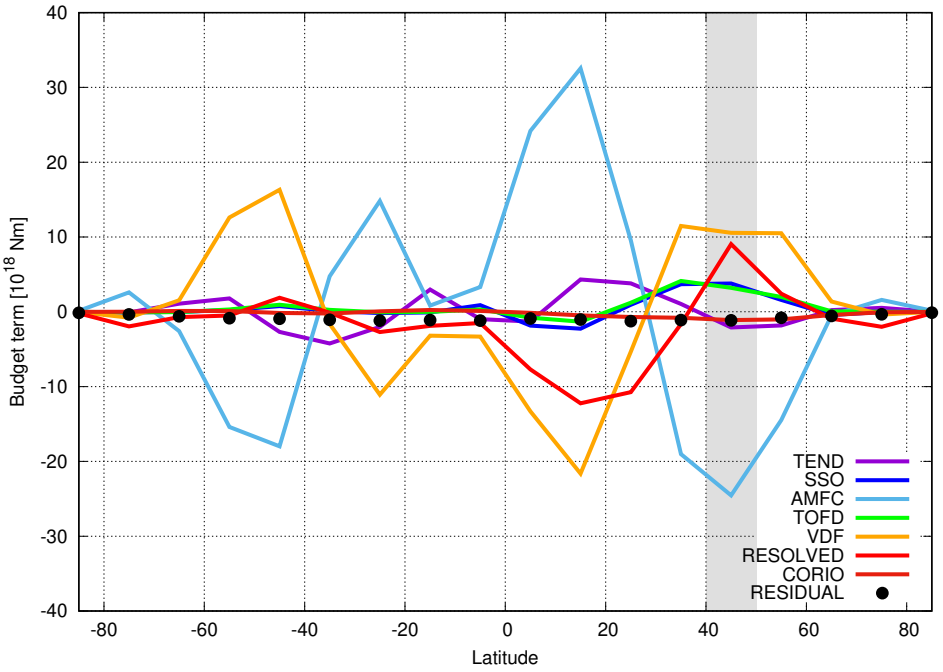


Figure 8: Individual terms [N m] in the zonally averaged vertically integrated relative angular momentum budget equation for the control forecast-only experiments at TCo639 for January 2015, as follows: total tendency (TEND; purple), SSO (blue), angular momentum flux convergence (AMFC; skyblue), TOFD (green), turbulent diffusion (VDF; orange), resolved orography (RESOLVED; red) and Coriolis forcing (CORIO; brown). The residual term is shown with black dots and the convection term is neglected due to nearly zero values. Each term is averaged over 10-degree latitudinal bands and over the first 24-hour of the ten-day forecasts, and its value is divided by 10^{18}. The 40N to 50N latitude band for which the resolution independence of the orography related terms is analysed in Figure 9 is indicated by the gray shading.

resolutions above TCo79. Similar results are found for wind. As the resolution is increased, more orographic features become resolved and less remain subgrid. One would therefore expect the impact of any changes to the representation of subgrid orographic drag to be largest at the lowest resolutions and to gradually decrease as the resolution increases. In the troposphere the skill improvement, as measured by the RMSE change (Figure 10), decreases indeed as expected from TCo319 to TCo1279, but at the lowest resolutions, it is either equal (TCo199) or smaller (TCo79) to the one obtained at TCo319. At TCo79, the skill is even degraded slightly at some levels. These results may be due to the fact that at such low resolutions the interaction between the TOFD and SSO schemes is not necessarily optimal as discussed by Kanehama et al. (2019). The lack of improvement at TCo79 is also due to the fact that the drag is overestimated in the troposphere at this resolution, as discussed above.

In the stratosphere, the impact is beneficial at all resolutions and is fairly resolution independent, apart for the highest resolution (TCo1279). Note however that the changes in RMSE shown in Figure 10 are relative differences with respect to the control experiment, and the impact of the SSO scheme in the control TCo1279 experiment is pretty small (top right in Figure 2). The 50 hPa temperature, geopotential and winds are strongly affected by the gravity wave drag component of the Lott and Miller (1997) scheme, and the weaker resolution sensitivity of the change in RMSE at this level is due to the weak resolution sensitivity of the scheme discussed above and in van Niekerk and Vosper (2021), Vosper et al. (2019).

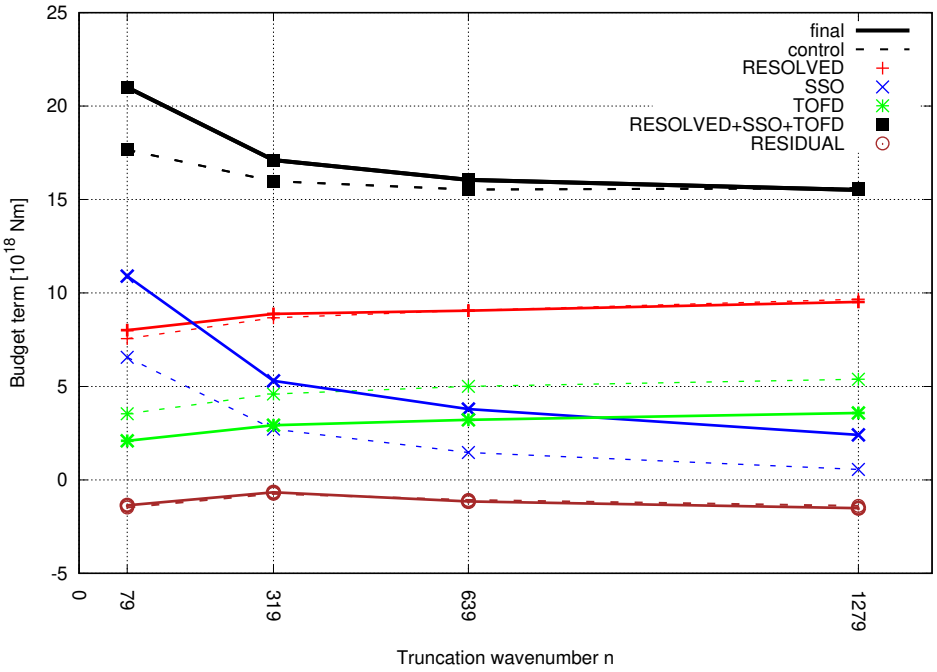


Figure 9: Orography related terms [N m] in the zonally averaged vertically integrated relative angular momentum budget equation at various resolutions ranging from TCo79 to TCo1279. Values are averaged over the 40 to 50N latitude band and over the first 24 hours of the forecasts performed for January 2015, and are divided by 10^{18}. Solid and dotted lines represent the forecasts performed using the control and final orographic drag configurations. Different colours show different terms, as shown in the legend.

5.3 Data Assimilation Experiments

The final orographic drag configuration significantly improves forecast skill in the NH in the data assimilation experiments in the short to medium-range. The improved skill in the short-range is well illustrated by the better fit of the background forecasts (12 hours forecasts used as first guess in the data assimilation cycles) to a range of observations (Figure 11). For example, improvements of up to 1 to 1.5% can be seen in the lower troposphere against radionsonde winds and aircraft temperature observations, respectively. A remarkable 3% improvement of the fit to surface pressure observations from Synop stations is also found. This large impact on surface pressure is not surprising, given that surface drag directly and rapidly impacts surface pressure (Sandu et al. 2016). The fit to radiosonde observations is slightly degraded in the lower stratosphere, but this degradation does not persist throughout the short and medium-range, apart from some degradation above 10 hPa north of 60N.

The final orographic drag configuration also results in statistically significant improvement in skill of the geopotential height in the NH, from the beginning of the forecasts up to 3 to 5 days ahead (depending on the height level) as illustrated in Figure 12. Similarly to what was seen for the forecast-only runs, the RMSE of the geopotential height is improved from the surface to about 50 hPa, with the biggest skill gain around 50 hPa. Similar results are found for winds and temperature (not shown). Some degradation is seen for all variables above 50 hPa, and particularly above 10 hPa, north of 60N, in certain periods. This does not seem to be related to sudden stratospheric events.

Since surface winds are one of the most important variables from a forecast user perspective, we also assessed the changes in surface winds induced by the revised representation of the orographic drag. Even if

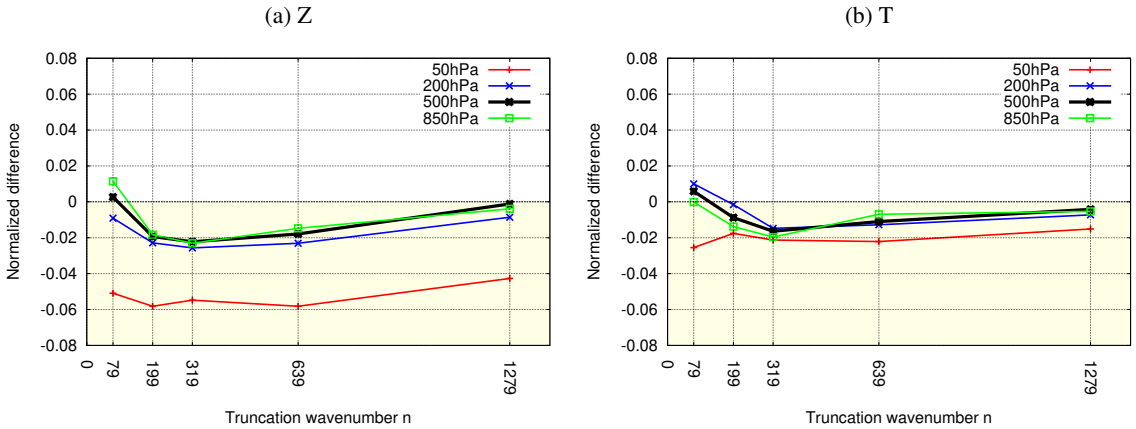


Figure 10: Normalized differences in root mean square error (RMSE) of geopotential height (left) and temperature (right) (%) between forecast-only experiments using the final and the control orographic drag configurations at various resolutions: TCo79, TCo199, TCo319, TCo639, TCo1279 (spectral truncation wavenumber indicated on the x-axis). The values are computed at a lead time of 120 hr over the NH. The colours indicate different levels: red, blue, black and green for 50 hPa, 200 hPa, 500 hPa and 850 hPa, respectively. Negative values (where the background is coloured in light yellow) indicate a decrease in RMSE, which indicates improved forecast skill in the experiments using the final orographic drag configuration.

wind verification in mountainous regions is not straightforward, current evidence based on user feedback and comparison with Synop observations suggests that the surface winds in the IFS are currently too weak over orography. This can also be seen for the forecasts of the control data assimilation experiment in Figure 13a. It appears that for this period, the mean 10m wind bias with respect to Synop observations is positive over flat land, and negative over the Alps and parts of Norway. When using the final configuration, the 10m wind speed is slightly increased over these regions, partly mitigating the existing bias.

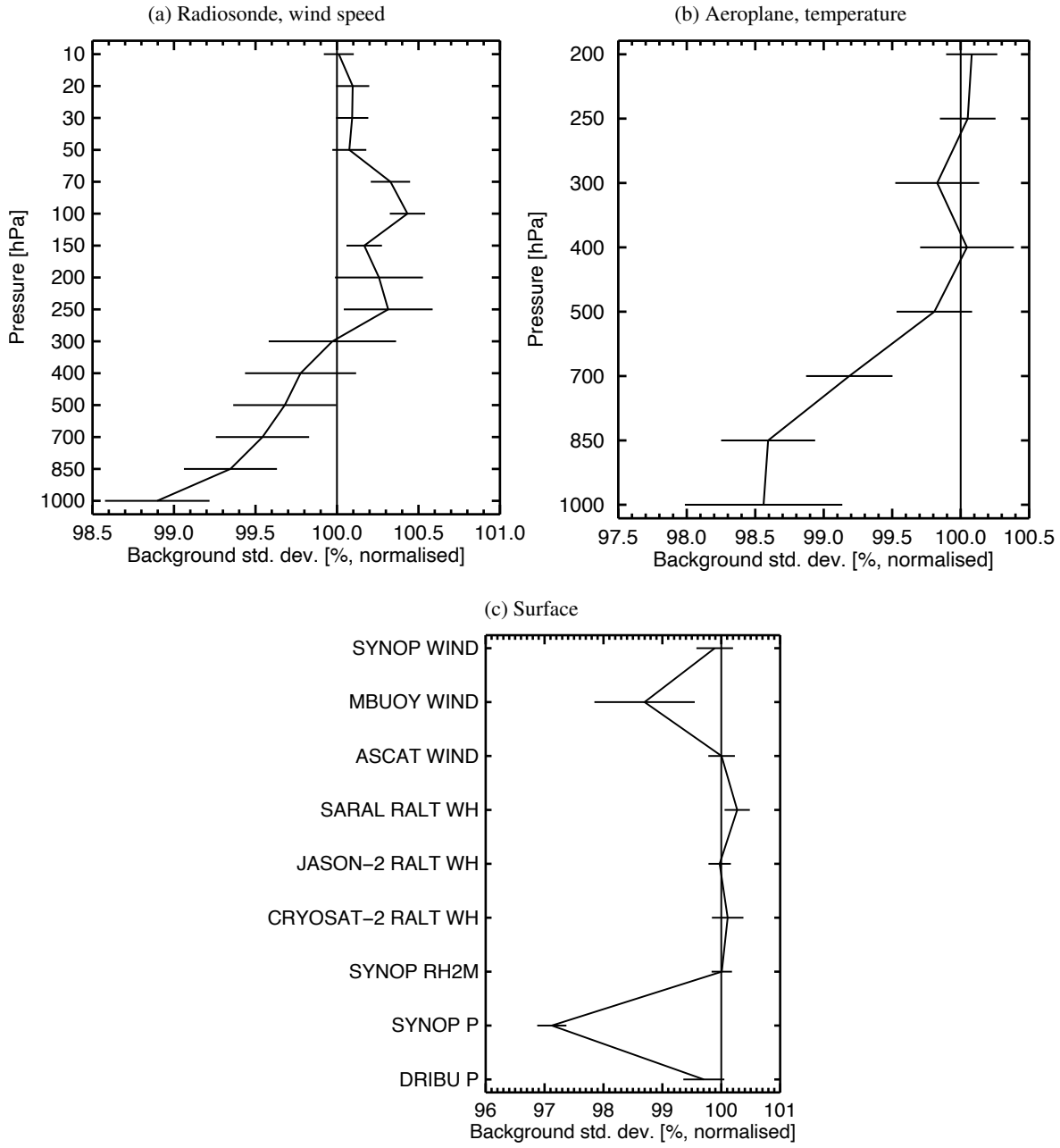


Figure 11: Normalized differences in standard deviation of observations minus background forecasts, between the data assimilation experiments using the final and control orographic drag configurations for December 2018 - January 2019. The different panels show different observations types: (a) wind speed from radiosondes, (b) temperature from aircraft and (c) surface observations of winds from synop, buoys, scatterometer (ASCAT), altimeter (JASON, CRYOSAT), of humidity from Synop and of pressure from Synop and drifting buoys. Error bars indicate 95% statistically significant intervals. Values smaller than 100% indicate a reduction of the standard deviation in the final configuration with respect to the control (the 100% line), and hence an improved fit to observations.

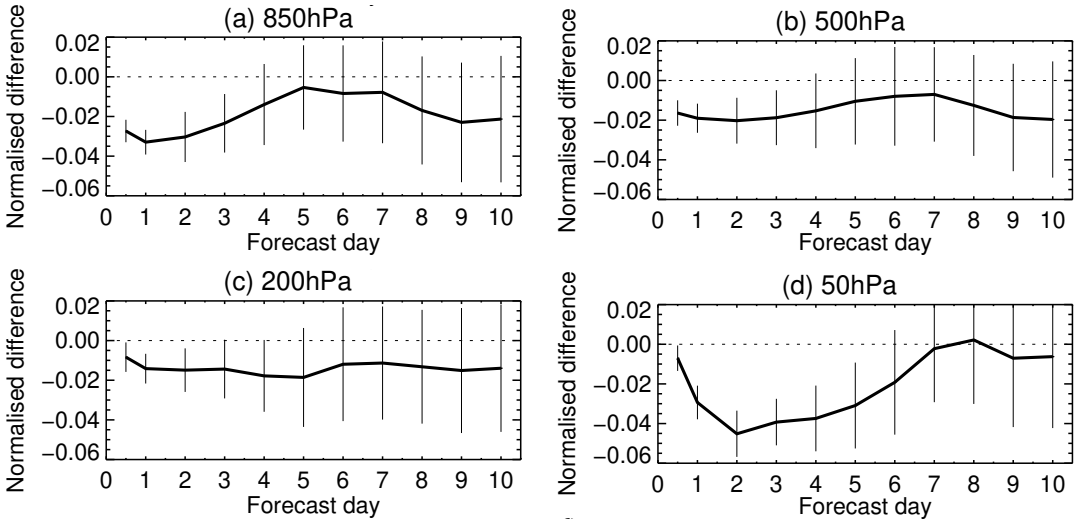


Figure 12: Normalized differences in root mean square error (RMSE) of geopotential height (%) between the data assimilation experiments using the final and the control orographic drag configurations for December 2018 - January 2019, for the NH and as a function of lead-time. Negative values indicate a reduction in the RMSE when using the final configuration compared to the control experiment, and hence indicate an improvement in skill. When error bars do not cross the zero line, the performance of the final experiment is significantly worse/better (95% interval) than the control.

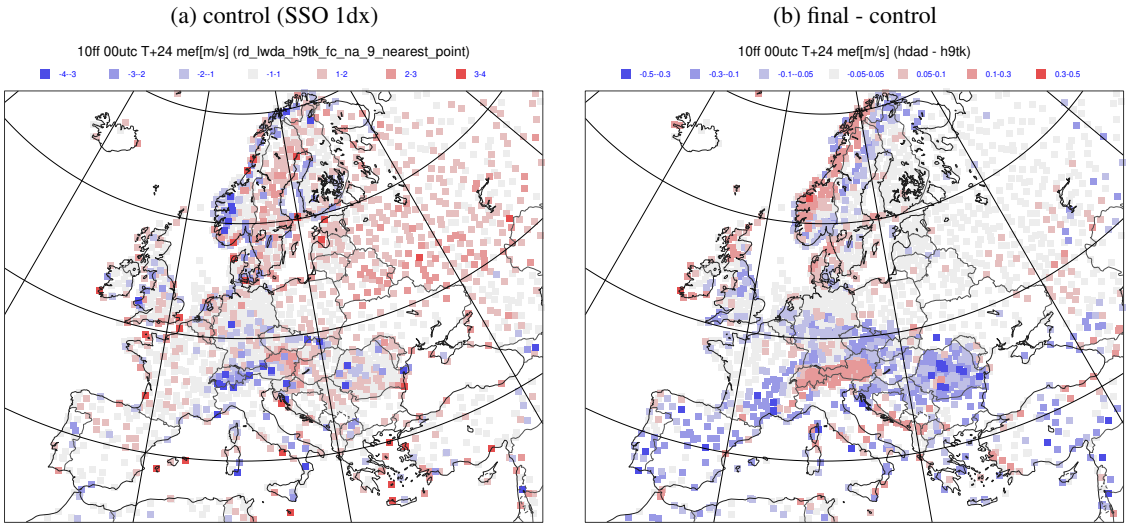


Figure 13: (a) Mean 10m wind speed bias [m/s] against Synop observations for the control data assimilation experiment. (b) Mean difference in 10m wind speed [m/s] between the final and control experiment at the nearest model grid-point to the Synop observations. The forecasts values are at T+24 hr, and the averages are performed over December 2018 - January 2019.

5.4 Medium-range ensemble experiments

Figure 14 illustrates that significant improvements in skill are obtained in the NH, as well as more locally over Europe and East Asia, when using the final instead of the control orographic drag configuration in the medium-range ensemble experiments. The improvements are significant up to the medium-range, i.e. that is to a lead-time of 4 to 7 days ahead, depending on the variable, region and level. The largest improvements are in geopotential height, and they maximise in the lower to mid-stratosphere, similarly to what was seen in the forecast-only experiments (Figure 6b). The changes in skill are consistent when verified against operational analysis and observations, except for the first 24 hours when a deterioration can be seen for certain variables/levels/regions when verifying against own analysis. Given that the ensemble is initialized both in the control and final orographic configuration experiment from the operational analysis, the verification at the very short-range (day 1 to 2) favours the control run which has an orographic drag representation that is consistent with the one used to produce the operational analysis. Verification against observations is therefore more reliable at these short-ranges. In the Tropics and the Southern Hemisphere, changes in skill are relatively small compared to the Northern Hemisphere, which is expected because the effects of mountains on the large scale circulation are known to be the largest in the Northern Hemisphere winter.

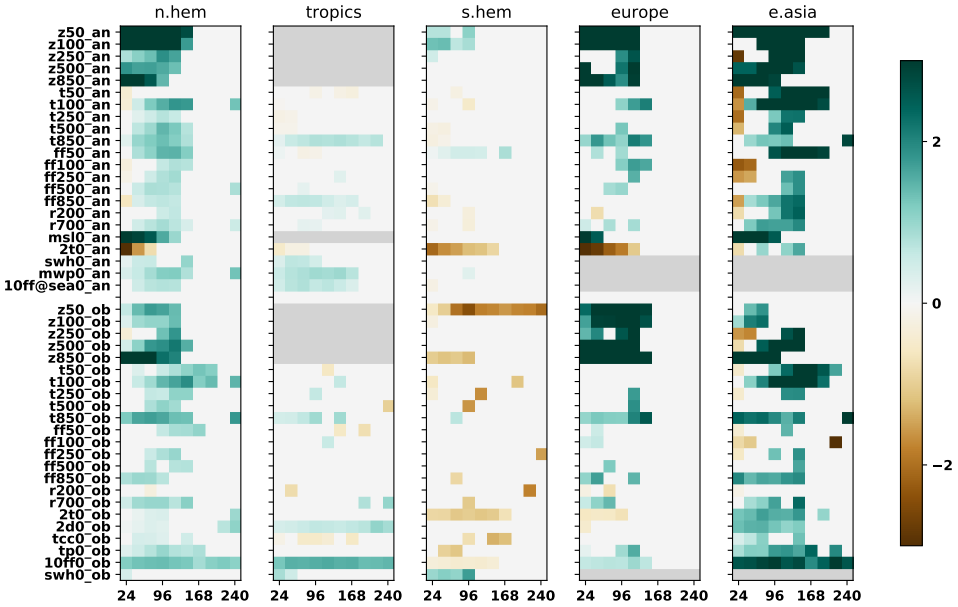


Figure 14: The summary of the differences in the fair continuous ranked probability score (fCRPS) for the 8-member medium-range ensemble experiments between the final and the control orographic drag configurations for the winter 2018/2019. Verified regions are the Northern Hemisphere (n.hem), Tropics (tropics), Southern Hemisphere (s.hem), Europe (europe) and East Asia (e.asia) up to T+240 hr. Only statistically significant values are plotted and the decrease (increase) in fCRPS in the experiment with the final orographic drag configuration is coloured in green (brown). The suffixes "_an" and "_ob" on the y-axis indicate verification against operational analysis and observations, respectively.

5.5 Seasonal prediction experiments

The results of the seasonal experiments are analysed in terms of the ensemble mean biases with respect to the ERA5 reanalysis, and the frequency of occurrence of atmospheric blocking events, for the December-January-February period (that is for a lead time between 1 and 4 months ahead).

The control experiment shows a too strong and northwardly shifted mid-latitude tropospheric jet in the NH (Figure 15a). As shown in Figure 15c the jet bias extends from the Pacific to the Atlantic. The geopotential height at 500 hPa and the mean sea surface pressure are too weak over the high latitudes, and too strong at lower latitudes in the NH (Figure 15e and g). These biases are consistent with too little surface drag (Sandu et al. 2016, Williams et al. 2020). Not surprisingly, when the orographic, and consequently the surface drag, is increased in the NH in the final configuration these biases are reduced. The NH mid-latitude jet position is more accurate and the wind speed biases are reduced, particularly over the Pacific and Atlantic (Figure 15b and d). The magnitude of the biases in geopotential height at 500 hPa and in mean sea level pressure are also reduced (Figure 15f and h), albeit not completely alleviated, their structure remains similar to that of the control experiment. Tests performed in a later IFS cycle (47r1) and in a coupled configuration, for which the biases in the control experiment are smaller, showed that the effects of the orographic drag changes are similar but of smaller amplitude (not shown).

The representation atmospheric blocking events on seasonal to climate timescales remains notoriously difficult (Davini and D'Andrea 2020, Davini et al. 2021b). Recent evidence suggests that the deficiencies in the representation of blocking events are related in part with an underestimation of the atmospheric variability and with errors in mean state (Davini et al. 2021b). Several recent studies also suggested that a better representation of orographic drag can result in an improved representation of blocking frequency and duration (Davini et al. 2021a, Pithan et al. 2016, Williams et al. 2020). This suggests that the changes to the orographic drag representation proposed here could lead to an improved representation of the atmospheric blocking events in the seasonal integrations.

The frequency of atmospheric blocking events in the seasonal integrations with the control and final orographic drag configurations, as measured by the Tibaldi-Molteni index (Tibaldi and Molteni 1990), is shown in figure 16. The climatological blocking frequency shows three peaks, around 0° E (North Atlantic blocking), 60° E (Ural blocking) and 180° E (Pacific blocking). The control experiment tends to predict more eastward North Atlantic blocking and less frequent Pacific blocking compared to ERA5, and shows a relatively good agreement for the Ural blocking. The underestimation of the blocking frequency is a well documented deficiency of the operational seasonal prediction systems at ECMWF, including the latest SEAS5 system (Davini et al. 2021b, Stockdale et al. 2018). Figure 16 suggests that the experiment with the final orographic drag configuration leads to a better agreement with ERA5 in terms of both the position of the North Atlantic peak, and the magnitude of the blocking frequency over the North Atlantic and the Pacific. Although this result is encouraging, it requires to be corroborated through a more in depth analysis based on experimentation with more ensemble members.

6 Conclusions

In this study, we revisited the representation of orographic drag in the ECMWF IFS. This revision relied on two recently proposed approaches: the framework proposed in the WGNE/GASS COORDE project, to constrain parametrized orographic low-level blocking and gravity wave drag by using km-scale simulations (van Niekerk et al. 2020), and a method to estimate the effective orographic resolution of a numerical model. Here, the COORDE framework was first used to compare the parametrized drag in

IFS, across a range of resolutions from 9 to 150 km, with the resolved drag from km-scale simulations performed with the MetUM over the Middle East region. This comparison corroborated the results of van Niekerk et al. (2018; 2020), Vosper et al. (2019), namely that the parametrized orographic drag in the IFS is too weak in the stratosphere at all resolutions, and the tropospheric drag is overestimated at the coarsest resolutions.

To alleviate these deficiencies, at least in part, the subgrid scale orography fields were reconstructed to represent scales up to the effective orographic resolution rather than the grid-length, building on the ideas of Vosper et al. (2016). The SSO fields were generated to represent scales smaller than 4-dx rather than 1-dx, following the findings of Kanehama et al. (2019), that for the cubic discretization used for all operational ECMWF forecasts, the effective orographic resolution is approximately 4 times the size of the grid box. This change results in an increased standard deviation and slope of the subgrid orography, and thereby stronger orographic drag, particularly due to low-level flow blocking. This change to the SSO fields was combined with the removal of some unjustified choices made in the SSO scheme over the years and with a rebalancing of the contributions of the TOFD and SSO schemes to low-level drag. These changes bring the two schemes closer to their original design, eliminating some of the arbitrary choices made during repeated tuning exercises over the years. In summary, the revised orographic drag representation includes: the 4-dx SSO climate fields, a cleaner SSO code, a stronger flow blocking and a weaker form drag.

The final orographic drag configuration results in an increased surface stress in the NH mid-latitudes during winter. It also leads to a better agreement of the parametrized drag with the resolved drag from the km-scale simulations over the Middle East region, apart for the coarsest resolution evaluated here (150 km) which is not used for operational applications at ECMWF. At this coarsest resolution, the drag exerted through the troposphere is too strong. The total orographic torque is fairly resolution independent across the range of resolutions used for operational forecasts at ECMWF (TCo319 to TCo1279), although it becomes less so at the coarsest resolutions with the final drag configuration. As also discussed in Kanehama et al. (2019) this undesired behaviour at resolutions larger than 100 km may be due to an suboptimal interaction of the TOFD and SSO schemes at these resolutions.

The revised orographic drag representation was evaluated in a range of configurations used for various applications at ECMWF. The evaluation spanned different resolutions, from TCo319 to TCo1279, and timescales, from hours to a season ahead. Medium-range forecast skill is improved in the NH winter throughout the troposphere, and particularly in the stratosphere, both in deterministic and ensemble forecasts. The largest improvements were found in geopotential height, though winds and temperature are also improved. The fit of short-range forecasts to observations in the data assimilation window is also improved, particularly against aircraft, radiosondes and surface pressure observations. The improvements due to the revised orographic drag configuration are expected to be smaller at the higher resolutions used for operational deterministic (TCo1279) and ensemble forecasts (TCo639) than in the TCo399 experiments discussed here (because more orographic scales are explicitly resolved and less scales are parametrized). In the experiments performed here, mean aspects of the NH winter circulation (mid-latitude jet strength and position, surface pressure, geopotential height), as well as atmospheric blocking are also improved at seasonal timescale. Given the large variability during NH winter, robust statistics are needed to assess circulation changes here. These results, therefore, require corroboration using experiments using more ensemble members. Nonetheless, the optimization of orographic drag parametrizations using the COORDE framework was thus proven to be a promising approach for improving the representation of orographic drag in a numerical weather prediction system and for improving forecast skill.

Acknowledgements

Author TK thanks Chihiro Matsukawa at JMA for pointing out the incorrect formulation of aspect ratio r of elliptic mountains. This study was carried out in a collaboration between ECMWF and JMA while TK was visiting ECMWF.

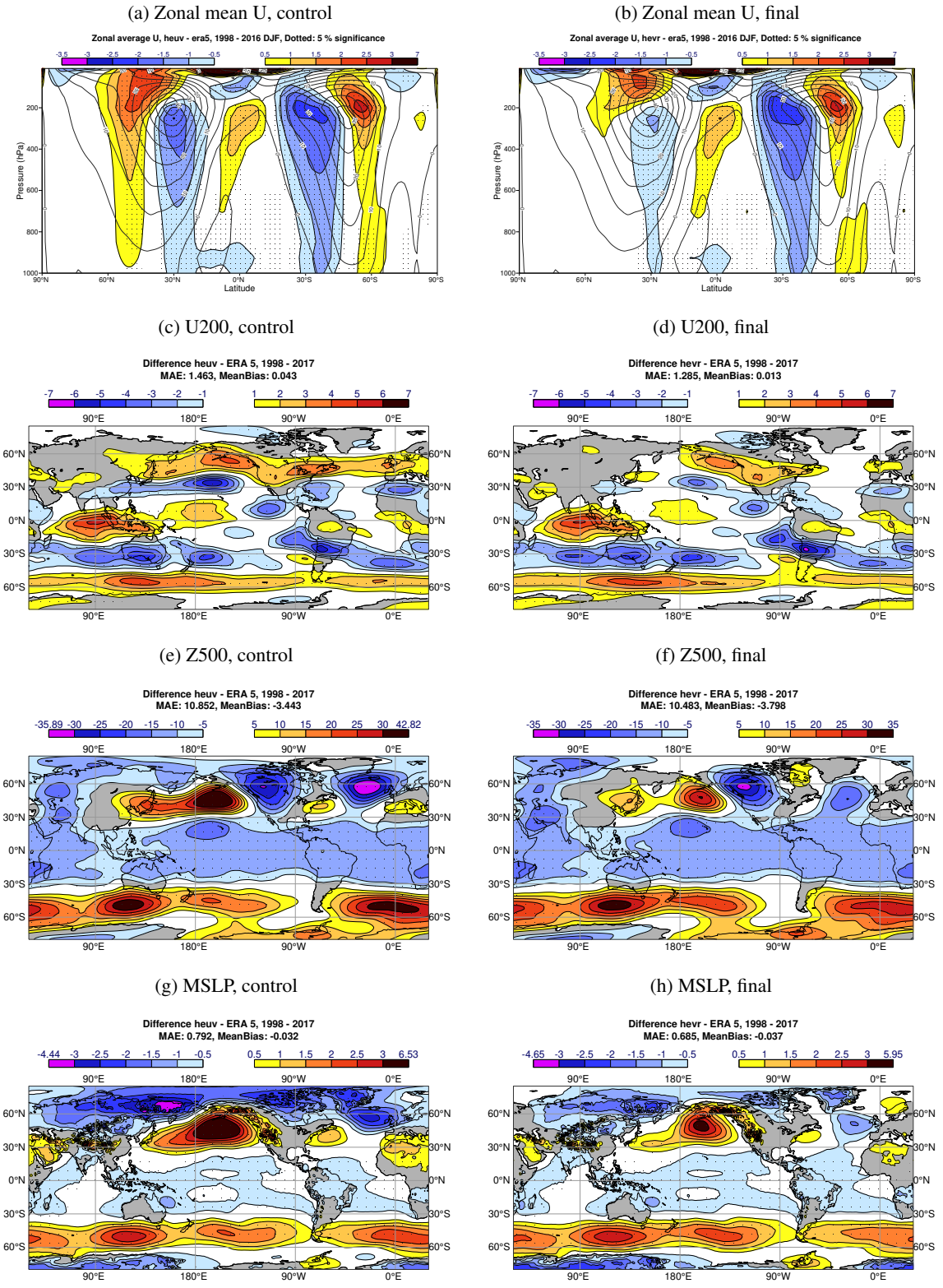


Figure 15: Ensemble mean bias with respect to ERA5 of the seasonal integrations performed with the control (left) and final orographic drag (right) configurations, averaged over DJF 1998 to 2016. From top to bottom: (a,b) zonal mean zonal wind [m/s], (c,d) zonal wind at 200 hPa (U200) [m/s], (e,f) geopotential height at 500hPa (Z500) [m] and (g,h) mean sea level pressure (MLSP) [hPa]. Dots indicate regions where the differences between the ensemble mean and ERA5 are statistically significant (using a 95% threshold).

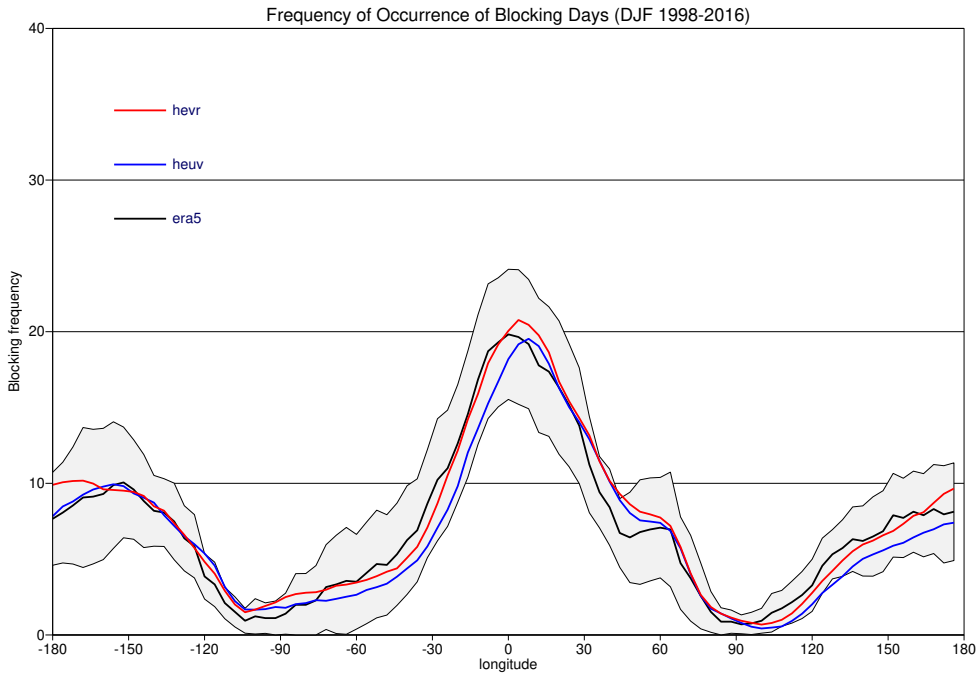


Figure 16: Frequency of occurrence of blocking days in the Northern Hemisphere, measured with the Tibaldi-Molteni index for ERA5 (black) and for the seasonal integrations performed with the control (blue) and final orographic drag configuration (red). The blocking frequency is averaged over December-January-February 1998–2016. The grey shading indicates 95% confidence intervals for ERA5 derived from a t-test.

References

Abdalla, S., L. Isaksen, P. A. E. M. Janssen, and N. Wedi 2013. Effective spectral resolution of ECMWF atmospheric forecast models. *ECMWF Newsletter*, 137:19–22.

Baines, P. and T. Palmer 1990. Rationale for a new physically-based parametrization of subgrid-scale orographic effects. *ECMWF Technical Memorandum*, (169).

Bauer, P., A. Thorpe, and G. Brunet 2015. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47.

Beljaars, A. 2020. Towards optimal parameters for the prediction of near surface temperature and dew-point. *ECMWF Technical Memorandum*, (866).

Beljaars, A., 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.

Brown, A. 2004. Resolution dependence of orographic torques. *Quarterly Journal of the Royal Meteorological Society*, 130(603):3029–3046.

Davies, L. A. and A. R. Brown 2001. Assessment of which scales of orography can be credibly resolved in a numerical model. *Quarterly Journal of the Royal Meteorological Society*, 127(574):1225–1237.

Davini, P. and F. D'Andrea 2020. From CMIP3 to CMIP6: Northern Hemisphere atmospheric blocking simulation in present and future climate. *Journal of Climate*, 33(23):10021–10038.

Davini, P., F. Fabiano, and I. Sandu 2021a. Orographic resolution driving the improvements in Northern Hemisphere winter mid-latitudes associated with horizontal resolution increases. *Journal of Advances in Modeling Earth Systems*, submitted.

Davini, P., A. Weisheimer, M. Balmaseda, S. J. Johnson, F. Molteni, C. D. Roberts, R. Senan, and T. N. Stockdale 2021b. The representation of winter Northern Hemisphere atmospheric blocking in ECMWF seasonal prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 147(735):1344–1363.

ECMWF 1997. Introduction and working group repots. In *Proceedings of a workshop on orography*.

Elvidge, A. D., I. Sandu, N. Wedi, S. B. Vosper, A. Zadra, S. Boussetta, F. Bouyssel, A. van Niekerk, M. A. Tolstykh, and M. Ujiie 2019. Significant uncertainty in the representation of orography in numerical weather prediction and implications for atmospheric drag and circulation. *Journal of Advances in Modeling Earth Systems*, 11(8):2567–2585.

Ferro, C. A. T. 2014. Fair scores for ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 140(683):1917–1923.

Hersbach, H., B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, et al. 2020. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049.

IFS documentation 2019. CY46r1. Technical report, ECMWF, available at: https://www.ecmwf.int/en/elibrary/19308-ifs-documentation-cy46r1-part-iv-physical-processes.

Iwasaki, T., S. Yamada, and K. Tada 1989. A parameterization scheme of orographic gravity wave drag with two different vertical partionings part I: Impacts on Medium-range forecasts. *Journal of the Meteorological Society of Japan. Ser. II*, 67(1):11–27.

Johnson, S. J., T. N. Stockdale, L. Ferranti, M. A. Balmaseda, F. Molteni, L. Magnusson, S. Tietsche, D. Decremer, A. Weisheimer, G. Balsamo, S. P. E. Keeley, K. Mogensen, H. Zuo, and B. M. Monge-Sanz 2019. SEAS5: the new ECMWF seasonal forecast system. *Geoscientific Model Development*, 12(3):1087–1117.

Kanehama, T., I. Sandu, A. Beljaars, A. van Niekerk, and F. Lott 2019. Which orographic scales matter most for medium-range forecast skill in the Northern Hemisphere winter? *Journal of Advances in Modeling Earth Systems*, 11(12):3893–3910.

Lander, J. and B. J. Hoskins 1997. Believable scales and parameterizations in a spectral transform model. *Monthly Weather Review*, 125(2):292–303.

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

Malardel, S., N. Wedi, W. Deconinck, M. Diamantakis, C. Kühnlein, G. Mozdzynski, M. Hamrud, and P. Smolarkiewicz 2016. A new grid for the IFS. *ECMWF Newsletter*, 146:23–28.

McFarlane, N. 1987. The effect of orographically excited gravity wave drag on the general circulation of the lower Stratosphere and Troposphere. *J. Atmos. Sci.*, 44(14):1775–1800.

Palmer, T., G. Shutts, and R. Swinbank 1986. Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parametrization. *Quarterly Journal of the Royal Meteorological Society*, 112(474):1001–1039.

Phillips, D. 1984. Analytical surface pressure and drag for linear hydrostatic flow over three-dimensional elliptical mountains. *Journal of the atmospheric sciences*, 41(6):1073–1084.

Pithan, F., T. G. Shepherd, G. Zappa, and I. Sandu 2016. Climate model biases in jet streams, blocking and storm tracks resulting from missing orographic drag. *Geophysical Research Letters*, 43(13):7231–7240.

Sandu, I., P. Bechtold, A. Beljaars, A. Bozzo, F. Pithan, T. G. Shepherd, and A. Zadra 2016. Impacts of parameterized orographic drag on the Northern Hemisphere winter circulation. *Journal of Advances in Modeling Earth Systems*, 8(1):196–211.

Sandu, I., A. Beljaars, and G. Balsamo 2014. Improving the representation of stable boundary layers. *ECMWF Newsletter*, 138.

Sandu, I., A. Beljaars, P. Bechtold, T. Mauritsen, and G. Balsamo 2013. Why is it so difficult to represent stably stratified conditions in numerical weather prediction (nwp) models? *Journal of Advances in Modeling Earth Systems*, 5(2):117–133.

Sandu, I., A. van Niekerk, T. G. Shepherd, S. B. Vosper, A. Zadra, J. Bacmeister, A. Beljaars, A. R. Brown, A. Dörnbrack, N. McFarlane, et al. 2019. Impacts of orography on large-scale atmospheric circulation. *npj Climate and Atmospheric Science*, 2(1):10.

Scinocca, J. and N. McFarlane 2000. The parametrization of drag induced by stratified flow over anisotropic orography. *Quarterly Journal of the Royal Meteorological Society*, 126(568):2353–2393.

Skamarock, W. C. 2004. Evaluating mesoscale nwp models using kinetic energy spectra. *Monthly Weather Review*, 132(12):3019–3032.

Slingo, A. and D. Pearson 1987. A comparison of the impact of an envelope orography and of a parametrization of orographic gravity-wave drag on model simulations. *Quarterly Journal of the Royal Meteorological Society*, 113(477):847–870.

Stockdale, T., M. Alonso-Balmaseda, S. Johnson, L. Ferranti, F. Molteni, L. Magnusson, S. Tietsche, F. Vitart, D. Decremer, A. Weisheimer, C. D. Roberts, G. Balsamo, S. Keeley, K. Mogensen, H. Zuo, M. Mayer, and B. Monge-Sanz 2018. SEAS5 and the future evolution of the long-range forecast system. *ECMWF Technical Memorandum*, (835).

Tibaldi, S. and F. Molteni 1990. On the operational predictability of blocking. *Tellus A: Dynamic Meteorology and Oceanography*, 42(3):343–365.

van Niekerk, A., I. Sandu, and S. B. Vosper 2018. The circulation response to resolved versus parametrized orographic drag over complex mountain terrains. *Journal of Advances in Modeling Earth Systems*, 10(10):2527–2547.

van Niekerk, A., I. Sandu, A. Zadra, E. Bazile, T. Kanehama, M. Khler, M.-S. Koo, H.-J. Choi, Y. Kuroki, M. D. Toy, S. B. Vosper, and V. Yudin 2020. COnstraining ORographic Drag Effects (COORDE): A model comparison of resolved and parametrized orographic drag. *Journal of Advances in Modeling Earth Systems*, 12(11).

van Niekerk, A., T. G. Shepherd, S. B. Vosper, and S. Webster 2016. Sensitivity of resolved and parametrized surface drag to changes in resolution and parametrization. *Quarterly Journal of the Royal Meteorological Society*, 142(699):2300–2313.

van Niekerk, A. and S. B. Vosper 2021. Towards a more scale-aware orographic gravity wave drag parametrization: description and initial testing. *Quarterly Journal of the Royal Meteorological Society*, accepted.

Vosper, S., A. Brown, and S. Webster 2016. Orographic drag on islands in the NWP mountain grey zone. *Quarterly Journal of the Royal Meteorological Society*, 142(701):3128–3137.

Vosper, S. B. 2015. Mountain waves and wakes generated by south georgia: Implications for drag parametrization. *Quarterly Journal of the Royal Meteorological Society*, 141(692):2813–2827.

Vosper, S. B., A. van Niekerk, A. Elvidge, I. Sandu, and A. Beljaars 2019. What can we learn about orographic drag parametrisation from high-resolution models? A case study over the Rocky Mountains. *Quarterly Journal of the Royal Meteorological Society*, 146(727):979–995.

Wallace, J., S. Tibaldi, and A. Simmons 1983. Reduction of systematic forecast errors in the ECMWF model through the introduction of an envelope orography. *Quarterly Journal of the Royal Meteorological Society*, 109(462):683–717.

White, R. H., J. M. Wallace, and D. S. Battisti 2021. Revisiting the role of mountains in the Northern Hemisphere winter atmospheric circulation. *Journal of the Atmospheric Sciences*, 78(7):2221 – 2235.

Williams, K. D., A. van Niekerk, M. J. Best, A. P. Lock, J. K. Brooke, M. J. Carvalho, S. H. Derbyshire, T. D. Dunstan, H. S. Rumbold, I. Sandu, and D. M. Sexton 2020. Addressing the causes of large-scale circulation error in the Met Office Unified Model. *Quarterly Journal of the Royal Meteorological Society*, 146.

Wood, N. and P. Mason 1993. The pressure force induced by neutral, turbulent flow over hills. *Quarterly Journal of the Royal Meteorological Society*, 119:1233–1267.

Zadra, A. 2013. Wgne drag project: An inter-model comparison of surface stresses. *Environment Canada Tech. Rep*, 1:36.