RESEARCH ARTICLE

Accounting for moist processes in a sub-grid orographic drag scheme

Samantha Smith

Met Office, Exeter, UK

Correspondence
S.A. Smith, Met Office, FitzRoy Road, Exeter EX1 3PB, UK.
Email: samantha.smith@metoffice.gov.uk

Funding information
Met Office Weather and Climate Science for Service Partnership India as part of the Newton Fund

Abstract
A method for estimating the effective moist stability within a sub-grid orographic drag scheme is described. It accounts for latent heat effects during low-level saturated sub-grid orographic ascent, with the lifting condensation level determining how much of the assumed sub-grid ascents are saturated. The scheme has been verified for two cases with moist south to southwesterly low-level flow impinging on mountain ranges of the USA and India, within both low-resolution global simulations and high-resolution limited-area model simulations with degraded orography. It produces significant improvements to the model wind fields in both regions. In particular, the amount of low-level orographic blocking predicted by the orographic drag scheme is reduced, particularly at low resolution, as are the associated large negative near-surface wind biases over the mountains. Precipitation is shifted slightly downstream. A suite of 24 five-day global NWP forecasts at N320 resolution was used to demonstrate that these results are robust, and that similar effects are seen over a number of mountain ranges. Over India, the convection scheme is very active in the global simulations, and has a larger effect on the winds at all levels than the orographic drag scheme. However the two schemes interact. Reducing the sub-grid low-level orographic blocking drag resulted in a reduction in the convergence occurring as the westerly flow decelerates towards the mountains. This decreased the amount of convection occurring over the sea upstream and along the coast, reducing the associated negative low-level wind bias there, as well as reducing the negative wind biases over land.

KEYWORDS
Drag parametrization, orographic flow blocking, model wind errors, latent heat effects on stability
1 | BACKGROUND

Mountains influence the large-scale atmospheric circulation and hydrological cycle. They block low-level flow, while flow over the unblocked mountain tops produces (a) mountain waves in stable air which decelerate the flow at high altitudes where they break, and (b) condensation which enhances precipitation. In addition, orography may initiate or strengthen convection as a result of additional ascent and/or horizontal convergence. All of these processes interact with each other over a wide range of scales. For example, a number of recent studies (e.g., Valenzuela and Kingsmill, 2017) have demonstrated that low-level blocked flow upstream of the Californian mountains turns to form a terrain-parallel jet. This acts as a virtual barrier, shifting ascent and precipitation upstream. Evaporation of any precipitation falling below cloud base may enhance low-level stability, reinforcing the blocked layer (Davolio et al., 2016).

Numerical models used for weather and climate simulations are unable to produce enough orographic drag over hills and mountains which are not adequately resolved. Despite their shortcomings, coarse-resolution forecasts (grid-spacings greater than 10 km) still play an important role in that they provide greater forecast lead times due to their lower demand on computer resources. They are thus used for global numerical weather prediction (NWP) and climate predictions, while ensembles are typically run at lower resolution than their deterministic counterparts. These models therefore rely on orographic drag (OD) schemes to represent the drag on winds due to unresolved orography. The large grid-spacings also mean that these models are unable to explicitly represent convection and, instead, rely on convection schemes to represent convective vertical fluxes of heat, water and momentum and associated surface precipitation.

Recent studies suggest that sub-grid OD schemes in many models, including the Met Office Unified Model (MetUM; Elvidge et al., 2019,van Niekerk et al., 2020), may produce too much low-level blocking at low resolutions. The importance of correctly representing low-level drag for orographic precipitation forecasts was identified by Colle et al., (1999), who found that the 10 km resolution Eta model had a tendency to produce excessive low-level blocking upwind of major barriers, resulting in precipitation falling upstream of its actual location. If orographic flow blocking is overdone, a positive feedback may take place whereby the excessive upstream precipitation evaporates as it falls through any underlying sub-saturated layer, leading to low-level diabatic cooling and increased stratification (Colle et al., 2005).

Moist processes occurring in saturated flow have a significant effect on orographic dynamics. Latent heat is released by condensation during saturated ascent, reducing the cooling rate of an ascending air parcel. This process enables air to rise over mountains more easily than in the absence of moisture due to the reduction in the effective stability, reducing the amount of low-level flow blocking (Miglietta and Buzzi, 2001). Latent heating will also modify the mountain wave field (Durran and Klemp, 1982a; 1983). Jiang and Doyle (2009) found that wave damping higher in the troposphere competes against the effects of increased ascent over the mountain at low levels when blocking is reduced.

Sub-grid OD parametrization schemes, and the theories on which they are based, were derived for dry orographic flows. They use the standard buoyancy frequency, which is based only on the vertical potential temperature (θ) gradient. Previous studies such as Jiang (2003) have suggested that a more accurate prediction of low-level flow blocking in saturated orographic flows can be obtained by replacing this standard N, which will henceforth be referred to as ND, with the saturated value NSAT given by Durran and Klemp (1982b). Previously sub-saturated air may reach saturation during its ascent over a mountain range and this should also be accounted for. Reeves et al., (2008) demonstrated that NSAT should be used in combination with the lifting condensation level (LCL) to determine how much of the total ascent over the hill would actually be saturated. Barcilon et al., (1979) suggested that NSAT can also be used to derive the wave solution for saturated flows, for which the mountain wave drag is substantially reduced.

The UM OD scheme has therefore been modified to account for these latent heat effects by using NSAT (Durran and Klemp, 1982b) instead of ND during low-level saturated ascent. An estimate of the LCL, combined with the maximum ascents and descents assumed by the OD scheme, gives information on the fraction of the sub-grid orographic vertical motions which are saturated, allowing for the estimate of an effective moist buoyancy frequency Nm on each model level. This study focuses on latent heat effects at low levels, where most atmospheric moisture is located: the Nm profile at each location is used to estimate the low-level average values used to predict the amount of low-level flow-blocking and the total wave drag in a column. Section 2 summarizes these modifications to the OD scheme, while Section 3 describes the model set-up. Section 4 describes the verification for two localized cases of low-level moist westerly to southwesterly flow impinging on the Rockies and the Western Ghats. Section 5 discusses tuning issues, while Section 6 demonstrates that the results of Section 4 are robust using a suite of 24 global simulations. A summary and discussion are provided in Section 7.
2 | THE MODIFIED OROGRAPHIC DRAG SCHEME

The standard OD scheme used in the UM is based on Lott and Miller (1997) (LM97 below) and is described in Vosper (2015). The sub-grid orography in a model grid-box is assumed to take the form of elliptical mountains of height $H_l = n $ $h_b$, where $h_b$ is the standard deviation of the surface height about the mean value and $n$ is a tuning parameter, currently set to 2.5. The scheme first calculates a vertical profile of the squared Brunt–Vaisala (or buoyancy) frequency $N^2$, a measure of the stability of the flow. This is an extremely important parameter which is used frequently by the OD scheme as described below. Two different low-level averages of $N$ are used, and these are listed in Table 1. The depth $z_b$ of any sub-grid orographic blocked layer is estimated using the low-level average Froude number $Fr_{av} = U_{av}/N_{av}H_l$, where $U$ is the horizontal wind speed resolved in the direction of the flow over the sub-grid mountain top. $U_{av}$ and $N_{av}$ are averaged over a layer of depth $z_{av}$, extending from the surface up to an altitude above the sub-grid mountain peak $H_l$ (or any neutral layer, if that is deeper) by a distance $U_{av}/N_{av}$ (limited to between 250 m and 3 km). This averaging, and the reasoning behind it, is described in Vosper et al., (2009). Once $z_b$ is obtained, the winds below $z_b$ are decelerated using the LM97 blocking drag equation. Air above this blocked layer is able to flow over the sub-grid mountain tops, launching gravity waves just above the sub-grid hill peak with an initial amplitude equal to the effective mountain height $h_{eff}$ (given by $H_l - z_b$). These waves are assumed to be vertically propagating hydrostatic waves, the vertical changes in $N$ being used to modify the gravity wave amplitude $A$ at each successive level above. Meanwhile, the value of $N$ at each level is used to determine the critical wave amplitude $\alpha_{crit} = Fr_{sat} U/N$ at which the waves will saturate, where $Fr_{sat}$ is the critical value of the local saturation Froude number. On any level where $\alpha > \alpha_{crit}$, gravity wave breaking is assumed to occur and $A$ is reset to the saturation amplitude $\alpha_{crit}$, with some of the drag deposited on the mean flow, spread vertically over a layer of depth $\lambda_{av} = 2\pi U/N$ (forced to remain between 1.5 and 19 km). The total wave stress $\tau_w$ deposited in a column during this process (LM97) depends on $h_{eff}$ and $N_{low}$, where $N_{low}$ is the average value of $N$ over the upper half of the sub-grid mountain.

In the standard OD scheme, the square of the buoyancy frequency is calculated simply as

$$N^2 = N_d^2 = \left( \frac{g}{\theta} \right) \frac{d\theta}{dz}, \quad (1)$$

where $\theta$ is the potential temperature (K) and $z$ is the altitude (metres). This is referred to as the dry value $N_d$, as it is only strictly valid for dry flows with no moist processes. Modifications to this calculation of $N^2$ to account for the effects of latent heating are described in the following sections.

2.1 | Sub-saturated and saturated values of $N$

There are two alternative methods for calculating the buoyancy frequency $N$ which will be used in the new scheme. When air is saturated with respect to water, ascent results in condensation and latent heat release, reducing the effective stability to the saturated value $N_{sat}$ given by Durran and Klemp (1982b)

$$N_{sat}^2 = \frac{g}{T} \left( \frac{dT}{dz} + \frac{\Gamma_s}{(1 + L_v q T)} - \frac{g}{(1 + q_T)} \frac{dT}{dz} \right), \quad (2)$$

where $q_T$ is the total water mixing ratio (kg-kg$^{-1}$), $T$ is the temperature (K) and $z$ is altitude (km). $\Gamma_s$ is the saturated adiabatic lapse rate (K-m$^{-1}$), given by (from the AMS Glossary)

$$\Gamma_s = \frac{(R_d T^2) + (L_v q T)}{C_{pd}(R_d T^2) + (L_v q e)}, \quad (3)$$

where $g$ is 9.8 m-s$^{-2}$, the latent heat of vaporization $L_v$ is a weak function of temperature but is approximately 2.5×10$^6$ J-kg$^{-1}$, the specific heat capacity of dry air at constant pressure $c_{pd}$ is 1,005 J-kg$^{-1}$-K$^{-1}$ and the gas constant for dry air $R_d$ is 287 J-kg$^{-1}$-K$^{-1}$. The ratio of the gas constants for dry air and water vapour, $e$, is 0.622.

During sub-saturated adiabatic ascent, there are no water phase changes and therefore no latent heating. However, the buoyancy effect of water vapour is

### Table 1: The low-level averages of the buoyancy frequency $N$ used in the OD scheme

| Variable | Averaging layer | Used to calculate | Ascent height |
|----------|------------------|------------------|---------------|
| $N_{av}$ | Below $z_{av}$   | $Fr_{av} = U_{av}/N_{av}H_l$, giving $z_b$ | Sub-grid hill height $H_l$ |
| $N_{low}$ | $0.5H_l$ to $H_l$ | Total wave stress $\tau_w$ | Wave launch amplitude |

Note: The mean Froude number $Fr_{av}$ is calculated over an averaging depth $z_{av}$ given by Vosper et al., (2009). The effective mountain height above this blocked layer, $h_{eff} = H_l - z_b$, is the upper part of the mountain producing ascent and gravity waves.
accounted for by using the moist sub-saturated value of the Brunt–Vaisala frequency, given by

\[ N_{\text{unsat}}^2 = \frac{g}{\Gamma_{\text{unsat}}} \left( \frac{d \theta_v}{d z} \right), \quad (4) \]

where \( \theta_v \) is the virtual potential temperature given by

\[ \theta_v = \theta(1 + 0.61q - q_L - q_I) \quad (5) \]

\( q_L \) and \( q_I \) are the cloud liquid water and ice mixing ratios (kg·kg\(^{-1}\)) respectively. Note that this has much less effect than using \( N_{\text{sat}} \).

### 2.2 Estimation of the effective moist stability \( N_m \)

The two different buoyancy frequencies described in Section 2.1, \( N_{\text{unsat}} \) and \( N_{\text{sat}} \), are only valid for unsaturated and saturated flow respectively. This section explains how these are combined in each grid-box and on each model level to give an effective moist stability \( N_m \). This is simplest in the case of completely saturated sub-grid orographic ascent. The grid-box mean relative humidity \( RH \) is given by \( q/q_{\text{sat}} \), where \( q \) is the grid-box mean water vapour mixing ratio (kg·kg\(^{-1}\)) and \( q_{\text{sat}} \) is its saturation value. A grid-box is specified as saturated when \( RH \geq 1 \), and the effective stability \( N_m \) is simply set to the saturated value \( N_{\text{sat}} \) given by Equation (2).

For sub-saturated grid-boxes, estimating \( N_m \) is more complicated. From Smith et al., (2019), the ascent required for an air parcel of temperature \( T \) and dewpoint temperature \( T_{\text{dew}} \) (K) to reach saturation is

\[ \eta_c = \frac{(T - T_{\text{dew}})}{(\Gamma_d - \Gamma_{\text{dew}})}, \quad (6) \]

where \( \Gamma_d \) is the dry adiabatic lapse rate and \( \Gamma_{\text{dew}} \) is the dewpoint lapse rate (K·km\(^{-1}\)). The dewpoint temperature is estimated using

\[ T_d = \frac{B_1 \left\{ \ln(RH) + \frac{A_1 T}{B_1 T} \right\}}{A_1 - \ln(RH) - \frac{A_1 T}{B_1 T}}, \quad (7) \]

where \( RH \) is the fractional relative humidity with respect to liquid water. Both \( T \) and \( T_d \) are in °C, and the coefficients \( (A_1 = 17.625 \) and \( B_1 = 243.04^\circ \)C) are those recommended by Alduchov and Eskridge (1996). \( \Gamma_{\text{dew}} \) is derived from the Clausius–Clapeyron equation (McIlveen, 1991)

\[ \Gamma_{\text{dew}} = \frac{T_{\text{dew}}^2 g R \rho}{L_v R_d T}. \quad (8) \]

When compared against the assumed sub-grid orographic vertical displacements, this saturation level \( \eta_c \) can be used to determine how much of the assumed sub-grid vertical orographic displacements are saturated, allowing for the estimation of a moist effective buoyancy frequency \( N_m \) using a weighted average of \( N_{\text{sat}} \) and \( N_{\text{unsat}} \). The assumptions made about the sub-grid vertical orographic displacements are slightly different for the two low-level averages listed in Table 1. The first, \( N_{\text{av}} \), is used to calculate the low-level mean Froude number \( Fr_{\text{av}} \), a measure of whether the incoming flow will be able to ascend over the sub-grid hill of height \( H_t \) or not. While rising over the sub-grid hill, the air would become saturated only if \( \eta_c < H_t \), in which case the effective moist buoyancy frequency could be given by

\[ N_m^2 = \frac{(N_{\text{unsat}}^2 \eta_c) + (N_{\text{sat}}^2 (H_t - \eta_c))}{H_t}. \quad (9) \]

The schematic diagram in Figure 1 demonstrates this scenario. If \( \eta_c > H_t \), saturation would not be reached and \( N_m \) is simply set to \( N_{\text{unsat}} \).

The second low-level average stability, \( N_{\text{low}} \), is used to determine the total gravity wave saturation stress \( \tau_w \) in the column. Air below the predicted sub-grid blocked layer depth \( \Delta z \) flows around the hill, so only air above this level actually flows over the hill. The gravity wave launch amplitude at the sub-grid mountain top is therefore set to \( h_{\text{eff}} \), the effective height of the sub-grid hill rising above \( \Delta z \) and therefore actually forcing vertical motion. So the moist buoyancy frequency profile is recalculated using...
the same method as for $N_{\text{uns}}$, but replacing $H_{\text{t}}$ with $h_{\text{eff}}$ in Equation (9) and Figure 1.

In this way, the effect of latent heating on both the low-level flow blocking and on the total sub-grid wave stress has been accounted for. These are likely to be the dominant effects because (a) most atmospheric moisture is located at low levels, and (b) the difference between $\Gamma_{\text{a}}$ and $\Gamma_{\text{s}}$, which is responsible for most of the difference between $N_{\text{unsat}}$ and $N_{\text{sat}}$, is largest at low levels, reducing with height towards zero in the upper troposphere. Moisture and latent heating may have an effect on the vertical propagation of the gravity waves above the sub-grid mountain tops, modifying exactly where they break, but accounting for this is more complex and seems to have little effect on the winds in these cases, as discussed in the Appendix.

Supercooled water is assumed to form above the freezing level rather than ice. This is because condensation removes supersaturation almost immediately, while ice processes respond over longer time-scales. Ice processes are likely to be of secondary importance anyway, as latent heat modifications to $N$ are much reduced at lower temperatures.

For simplicity, the above calculations neglect random sub-grid horizontal moisture and temperature variations due to turbulence which may result in partial saturation, which is dealt with by the sub-grid cloud scheme. No attempt is made to incorporate these variations as they are totally unrelated to the sub-grid orography, so there is no way of knowing where any resulting sub-grid cloud is located relative to the sub-grid mountain peaks and valleys.

It should be noted that, for a sub-saturated grid-box, $\Gamma_{\text{s}}$ (Equation (3)) is determined using the cloud-base temperature $T_{\text{cloud}}$, which is the grid-box temperature $T$ corrected for dry adiabatic ascent to the LCL $\eta_{\text{c}}$. This is necessary because $\Gamma_{\text{s}}$ is valid only once the air becomes saturated.

A comprehensive set of physical parametrizations are used to represent sub-grid-scale processes, as described in Walters et al., (2017). Which parametrizations are switched on, and the values of any tuning parameters, are defined by the science configuration. A global atmosphere (GA) configuration is used when the MetUM is run as a global model, while a regional atmosphere and land (RAL) science configuration is used when the MetUM is run as a kilometre-scale limited-area model (LAM). The GA configuration used for all global simulations develops over time to incorporate new or modified sub-grid parametrization schemes. While it would be ideal also to have a single defined RAL configuration that performs effectively in all LAM regions, this has not yet been possible. Instead, there are two sub-releases, one for midlatitudes (RAL1-M) and one for tropical regions (RAL1-T). The physics of these set-ups, the reasons for their existence and their performance is documented in Bush et al., (2020). As explained there, one of the major reasons for needing two configurations is that convection is sometimes very under-resolved in the UK in kilometre-scale simulations, particularly in cases of small, shallow showers. This can manifest itself as small showers initiating too late or not at all. In order to cope with this, RAL1-M has relatively weak turbulent mixing, and has stochastic temperature and moisture perturbations, to encourage the model fields to be less uniform and to help convection initiate. If the model is run with these in the Tropics, the model initiates convection too early and convective cells tend to be too small.

The main differences between the GA and RAL configurations are the representation of processes which are explicitly resolved in kilometre-scale simulations but need to be parametrized in the global simulations. The GA configuration uses the OD scheme described in Section 2 to represent flow deceleration due to low-level blocking and the breaking of mountains waves aloft produced by sub-grid orography. In the RAL configurations, the tuning parameters in the OD scheme are set such that the scheme does very little, effectively turning it off. The global simulations also rely on a convection parametrization scheme. The kilometre-scale simulations are able to represent deep convection, although small-scale convection is not truly resolved. So, in the absence of a scale-aware convection scheme (in development), the convection parametrization scheme is switched off in both RAL configurations. The sub-grid cloud scheme also differs between science configurations. The GA and RAL1-T configurations use the prognostic cloud fraction and prognostic condensate (PC2) scheme, which calculates the various condensate sources and sinks, and then advects the updated cloud fields. Meanwhile, the RAL1-M configuration uses the diagnostic Smith cloud scheme. Further details of these schemes,
and relevant references, are given by Walters et al., (2017) and Bush et al., (2020).

In addition to using different science configurations, the global and LAM simulations also use different vertical model levels and time steps. The 70 model levels go up to 80 km in the global simulations but only 40 km in the LAM simulations, giving the LAM simulations a higher vertical resolution. The model time step decreases with decreasing horizontal grid spacing from 20 min in the N96 simulation to 7.5 min in the N768 simulation, and 1 min in the 1.5 km resolution LAM simulations.

4 | CASE-STUDY SIMULATIONS

Section 2 described changes made to the OD scheme in order to account for the effect of latent heating on the stability profile. This section describes the effects of these changes on model wind fields in the MetUM using 36 hr long NWP-type case-study simulations. All simulations were initialised using global analyses from the operational archive. These were created by the data assimilation system when the operational forecast was produced, so was based on a model simulation using the original dry OD scheme. Data assimilation is not used in the current simulations.

Two case-studies were chosen because they were situations with moist westerly to southwesterly airflow impinging on a major mountain range. These are:

(a) the Kerala Floods case in southern India during summer monsoonal westerly flow on 15 August 2018, and
(b) the Atmospheric River (AR) episode accompanying the passage over California (USA) of a frontal system on 29 and 30 November 2012, as described in Eiserloh and Chiao (2015).

Vertically integrated liquid water condensate is shown in Figure 2, along with horizontal wind vectors on the model level which has an altitude of 180 m over the sea. Note that, unless otherwise stated, all times are specified as Coordinated Universal Time (UTC). Regions with significant liquid water condensate will be saturated at low levels, resulting in latent heat effects during sub-grid orographic vertical motions. Sub-saturated air may still become saturated during ascent over the sub-grid mountains, resulting in reduced $N_m$ as described above. However, the impact of the OD modifications will reduce for lower RH values.

During the Kerala Floods case over India, low-level winds were westerly and changed little during the simulation. Most precipitation was convective in nature, and Figure 2b shows that, while convective showers form upstream over the sea, they are enhanced or even triggered more extensively over the Western Ghats. The N96 global simulation shows none of this fine-scale detail in the vertically integrated liquid water condensate.

During the USA case, precipitation produced by a frontal system was enhanced over the very-large-scale orography due to moist southwesterly low-level flow (Smith et al., 2019). The low-level wind field was extremely complex, and at some locations it changed significantly with time due to the southward movement of the frontal system and associated low-level moisture from the position in Figure 2a. The wind fields from the N96 global model were much simpler because the coarse model grid is unable to properly represent the winds associated with either the frontal structure or small-scale orographic motions. However, the location of the frontal system in the N96 model compares well with that in the 1.5 km resolution model.

Verification is performed against a LAM simulation with a 1.5 km horizontal grid spacing and a 1 min time step, driven using lateral boundary conditions derived from a global N768 (17 km resolution) simulation. The first 12 hr of each LAM simulation was discarded to allow for model spin-up. The standard midlatitude RAL configuration, RAL1-M, is used over the USA, while the RAL configuration recommended for the Tropics, RAL1-T, is used over India (Section 3). The 1.5 km resolution orography (Figure 3) captures most of the important orographic features, so that the OD is resolved explicitly by the model. Verifying against a 1.5 km LAM simulation instead of observations allows for verification of the entire model wind field, rather than being able to verify only at specific observation locations and times. When verifying against observations, any non-orographic errors in model evolution make verification in terms of orographic processes difficult. Verification against the LAM, however, reduces this problem. This makes it much easier to pinpoint the errors of interest, i.e. those related to the representation of the orography at low resolution, both resolved and sub-grid.

The modified OD scheme will be tested in global simulations over a range of resolutions, with grid-spacings of 17 km (N768), 42 km (N320) and 140 km (N96). The N768 model (which means that N has a value of 768, where the model grid has 2N longitudes and 1.5N latitudes), was used for operational NWP forecasts until fairly recently. While the N96 model is used only for climate simulations, NWP simulations allow for the analysis of the direct effect over the mountains themselves, without complications due to changes to the general circulation and hydrological cycle as seen for the seeder–feeder scheme in Smith et al., (2019). It provides a more stringent test of the change: the extremely large grid spacing means that the OD scheme is likely to produce a larger proportion of the OD. These global simulations are used to produce NWP and climate
Vertically integrated condensed liquid water (kg m\(^{-2}\)), along with horizontal wind vectors on the 180 m model level, from the 1.5 km resolution LAMf simulations: (a) at 1200 UTC on 29 November 2012 over the USA (127.5–117.5\(^{\circ}\)W, 34–45\(^{\circ}\)N), and (b) at 1200 UTC on 15 August 2018 over India (71–78\(^{\circ}\)E, 8–16\(^{\circ}\)N).

Forecasts, including seasonal and decadal predictions, and therefore need to be as accurate as possible. However, problems such as the inability of the lower-resolution global simulations to properly represent complex background non-orographic flows make verification in terms of orographic effects challenging. Therefore, for a clean investigation of its performance, the modified OD scheme is first tested within a set of LAM simulations in which the orography is degraded to that used in the global simulations mentioned previously.

4.1 High-resolution simulations varying the representation of orography

This section describes tests of the modified OD scheme within a set of 1.5 km resolution LAM simulations in which the orography is degraded to that used in the above-mentioned global simulations. This is done by interpolating the global model orography fields onto the high-resolution LAM grid. These fields include the resolved orography, and the sub-grid variables used by the OD scheme as described by Elvidge et al., (2019). The OD scheme is switched on, with the same tuning parameter values as in the GA configuration, in order to account for the drag due to the smaller-scale orographic features which are now missing. These simulations will be referred to as LAMn96, for example, for the LAM simulation using orography from the N96 global model. For each case-study and orographic resolution, a control (CTRL) simulation is performed using the standard OD scheme. The simulation is then repeated using the modified OD scheme (the NMOIST simulation).

The original LAM simulation using the full 1.5 km resolution orography, which will be referred to as the LAMf simulation, will be used to verify the simulations with degraded orography. As the model grid, time step,
science configuration (RAL1-M for the USA simulations, RAL1-T for the India simulations) and ancillary fields (other than orography) for simulations with degraded orography remain consistent with the LAMf simulation, errors compared to the LAMf simulation should be due purely to differences in the resolved orography and the representation of the sub-grid orography by the OD scheme.

Model surface heights in the LAMn96 and LAMf simulations are compared for USA and India in Figure 3. The resolved orography in the N96 simulations is extremely smooth with absolutely no fine-scale detail, and much of the orography is sub-grid: sub-grid surface heights are comparable to the resolved surface heights, making the OD scheme very important in correctly forecasting the wind fields at this low resolution. The N768 model orography (not shown) has more fine-scale detail. For example, the Central Valley in California and the Western Ghats of India are at least partially represented. Note that the mountains of the USA are of much larger scale than those of India.

The values of $N_{\text{low}}$ and $h_{\text{eff}}$ predicted by the OD scheme determine how the scheme behaves in each grid box, so these are shown over the USA in Figure 4 for the LAMn96 CTRL and NMOIST simulations. In the CTRL simulations (Figure 4a,c), $N_{\text{low}}$ is calculated from the dry buoyancy frequency $N_d$, and generally has values of the order of 0.01 s$^{-1}$. The low-level flow is strongly blocked by the large sub-grid mountains of the USA, giving effective mountain heights $h_{\text{eff}}$ which are generally smaller than $z_0$. Plots from the NMOIST simulations are shown in Figure 4b,d. Accounting for the effect of latent heating on $N$ in the OD scheme reduces the value of $N_{\text{av}}$ used to calculate the Froude number $F_{\text{av}}$ and therefore $z_0$, and the low-level blocking drag, are reduced. The corresponding increase to $h_{\text{eff}}$ launches larger-amplitude sub-grid gravity waves and acts to increase the total wave drag. However, this is offset by a reduction to $N_{\text{low}}$, on which the total wave drag also depends. If $N_{\text{low}} \leq 0$, then both $z_0$ and $h_{\text{eff}}$ are set to zero and absolutely no sub-grid OD is applied in that grid box (this was already done in the standard OD scheme). Over the USA, this occurs in 50 to 60% of the land points in the LAMn96 NMOIST simulation, while it rarely occurred in the CTRL simulation (<2% of land points).

The mountains of India are lower and narrower than the mountains of the USA, giving smaller resolved and sub-grid mountain heights in the LAMn96 simulations. As there is also a moist unstable boundary layer in tropical regions, low-level orographic blocking is much weaker over India. There is very little sub-grid blocking in the LAMn96 NMOIST simulation, with zero blocking in 94% of grid-boxes. This moist unstable boundary layer is not captured in the CTRL simulation using the dry buoyancy frequency, so sub-grid blocking still occurs: values of $z_0$ up to 800 m are seen and zero blocking occurs in only 37% of grid-boxes.

The OD scheme directly modifies the model wind fields by applying drag at levels where either low-level blocking or gravity wave saturation is occurring. The effect of the modified OD scheme on the model wind fields will therefore now be examined. Wind fields from each simulation were first regridded from the rotated LAM grid onto a regular lat–long grid of comparable resolution. Mean wind profiles are first compared, in order to investigate the mean effects of the change and how this varies with altitude. As discussed previously, the complexity of the low-level wind fields shown in Figure 2 makes verification of the wind profiles at a single place or time extremely difficult. Horizontally averaging the profiles over all land points in the region gets around this problem, showing the mean effect over the mountains. Wind speeds are shown for the southwesterly winds over the USA, while the easterly wind component $u$ is shown over India. The (unaveraged) high-resolution wind fields will be compared shortly.

Figure 5 shows these profiles for the LAMn96 CTRL and NMOIST simulations over the USA, along with the profiles from the LAMf for comparison. Figure 5b shows the wind errors compared to the LAMf for the LAMn96 and LAMn768 simulations. (Profiles from the LAMn320 simulations lie between these, but are more similar to the LAMn768 profiles.) A frontal system associated with the jet stream is passing over the USA, bringing strong, moist southwesterly flow. Low-level winds become more southerly towards the surface due at least in part to low-level orographic blocking. This appears to be overdone in the LAMn96 CTRL simulation, with low-level winds that are too slow and too southerly compared to the LAMf winds. The dominant errors in the CTRL simulations are these negative biases in the near-surface winds, reaching up to $-9$ m s$^{-1}$ in the lowest kilometre of the atmosphere in the LAMn96 CTRL simulation. While there is also a negative near-surface wind bias in the LAMn320 and LAMn768 CTRL simulations, this error is smaller at higher resolution, in agreement with van Niekerk et al., (2020). These low-level wind errors are greatly reduced in all of the NMOIST simulations, particularly at larger grid-spacings, and the errors vary less with resolution. These improvements to the low-level winds in the NMOIST simulations are brought about by a reduction in the blocked layer depth $z_0$ and reduced decelerations due to sub-grid blocking, as shown in Figure 6.

While the largest change to the winds is this reduction in excess sub-grid low-level blocking at low resolution, accounting for the effects of latent heating on $N$ also modifies the drag due to breaking gravity waves aloft. In the CTRL simulations, drag due to gravity wave breaking in the troposphere (Figure 6) is kept to a minimum over the
USA by the increasing wind speed with altitude. Once wind speeds start to decrease above the jet maximum, gravity waves are able to reach saturation and break within the stratosphere. The associated drag applied at these levels is spread vertically over a layer of depth $\lambda_z$, as described in Section 2. This behaviour is modified in the NMOIST simulations in two ways. Firstly, reduced blocking allows a deeper layer of air to ascend the sub-grid hills (increased $h_{\text{eff}}$), launching larger-amplitude sub-grid gravity waves which are able to break at lower levels. This increased tropospheric drag is largest in the LAMn96 simulation, where a positive wind bias in the CTRL simulation is reduced in the NMOIST simulation. Secondly, smaller values of $N_{\text{low}}$ act to reduce the total wave stress ($\tau_w$) in the column, often reducing it to zero as discussed above. This, in combination with a larger fraction of the drag being applied in the troposphere, reduces the drag in the stratosphere. Stratospheric wind speeds are therefore increased in the NMOIST simulations. A small positive wind speed error develops between altitudes of 14 and 19 km, an increase in a trend which was already present in the CTRL simulations. Overall there is less vertical variation in the wind errors in the NMOIST simulations than in the CTRL simulations and the wind profiles are improved. There is also less variation in wind error between different model resolutions.

Figure 7 shows the LAMn96 and LAMf $u$ profiles over India, along with the $u$ errors in the LAMn96 and LAMn768 simulations with degraded orography. (As for the USA, the LAMn320 error profiles lie between these, but are more similar to the LAMn768 profile.) The low-level flow is westerly within the summer monsoon jet. These westerly winds decrease with altitude above about 1.5 km, until the wind direction reverses to become easterly in the...
FIGURE 6 Mean sub-grid $u$ accelerations produced by (a, b) gravity wave saturation and (c, d) low-level blocking over (a, c) the USA and (b, d) India at the final output times. Black (grey) lines show profiles from the LAMn96 (LAMn768) simulations, and solid (dashed) lines show profiles for the CTRL (NMOIST) simulations.

upper troposphere. Most of the sub-grid wave saturation drag (Figure 6b) is applied to the tropospheric winds due to this critical level. As for the USA case, the largest wind errors in the CTRL simulations (Figure 7b) are a negative wind speed bias in the lowest 2 km of the atmosphere due to excessive sub-grid low-level blocking, which is greatly reduced in the NMOIST simulations. In fact it is replaced by a small positive wind bias. However, the remaining errors in the NMOIST wind profiles are generally within $\pm 1 \text{ m s}^{-1}$.

The negative low-level wind biases in the CTRL simulations, which are reduced in the NMOIST simulations, are concentrated over the mountains themselves (Figures 8 and 9). Small-scale errors remain in the NMOIST wind fields due to flow over unresolved peaks and valleys, and due to unresolved gap flows and wakes in southern India. This is to be expected, as any parametrization scheme can only represent the mean effect of the unresolved orography. It cannot add small-scale detail.

4.2 Global case-study simulations

The previous section demonstrated that the modified OD scheme performs well in a suite of LAM simulations in which the orography was degraded. This provided a clean test of the scheme, as the only differences between simulations were the resolved orography and the performance of the OD scheme. However, it is the global simulations which actually use the OD scheme. The GA6.1 Global Atmosphere science configuration described in Walters et al., (2017) is used, which has been the operational NWP configuration since July 2014. For each case-study and orographic resolution, a CTRL (NMOIST) simulation...
**Figure 7** (a) Mean $u$ profiles over the mountains of India at 0000 UTC on 16 August 2018 from the LAMf (grey) and LAMn96 (black) simulations. (b) Errors in $u$ compared to the LAMf for the LAMn96 (black) and LAMn768 (grey) simulations. In both panels, solid (dashed) lines show the CTRL (NMOIST) simulations.

**Figure 8** Horizontal wind speed errors (m·s$^{-1}$) compared to the LAMf on the model level which is 180 m above the sea at 1200 UTC on 29 November 2012 over the USA in the (a) CTRL and (b) NMOIST LAMn96 simulations. Black lines indicate coastlines and low-resolution surface height at 500 m, 1 km and 1.5 km.

**Figure 9** As Figure 8, but at 0000 UTC on 15 August 2018 over India. Surface height contours of 200 and 400 m are shown.
Wind errors compared to the LAMf for the global N96 (black) and N768 (grey) simulations over (a) the USA and (b) India. In both panels, solid (dashed) lines show profiles from the CTRL (NMOIST) simulations.

Mean wind profiles over land at the final simulation time over (a) the USA and (b) India from the N96 global CTRL simulation (solid), a simulation with the OD scheme switched off (dotted) and a simulation with the convection scheme switched off (dashed).

is performed using the standard (modified) OD scheme, as in the previous section, and errors are calculated by comparison against the LAMf simulation. To obtain wind error fields, the LAMf wind fields were regridded from the rotated grid onto a regular lat–long grid of comparable resolution, while the global wind fields were horizontally interpolated onto the same high-resolution grid for direct comparison. When an altitude is given for a wind field, this is the altitude of the model level over the sea.

Over the mountains, the comparative performance of the standard and modified OD schemes in the global simulations is similar to that described for the LAM simulations in the previous section. This is demonstrated by the error profiles in Figure 10. Large negative wind errors occur at low levels over the mountains due to the standard OD scheme producing too much sub-grid low-level flow blocking. This error is greatly reduced in the NMOIST simulations in both regions. Meanwhile, stratospheric winds over the USA are increased in the NMOIST simulations due to a reduction in wave drag, indicating the need for retuning of the OD scheme. Other errors in the global simulations, which were not present in the LAM simulations, are unlikely to be directly due to the representation of orography. Therefore they should not necessarily be reduced in the NMOIST simulations. However, the large errors above an altitude of 6 km in the N96 CTRL simulation over India are slightly reduced. This could be due to wind changes modifying the behaviour of the convection scheme, which itself modifies the winds. Figure 11 shows that, while the OD scheme has the largest effect on mid-latitude winds over the USA mountains, the convection scheme dominates in tropical regions such as India.
Over the sea, additional low-level wind errors can be seen in the global simulations which were not present in the LAM simulations. For the USA case, positive and negative localised errors are due to the poor representation of the frontal system on the low-resolution model grid. For the Indian case, large negative low-level wind errors just off the coast (Figure 12) appear to be related to the convection scheme. The convection scheme modifies the model wind fields through vertical momentum transport. The 180 m wind speeds are increased by about 10 m s⁻¹ along the coast if the convection scheme is switched off. This negative coastal wind error is reduced in the NMOIST simulation, suggesting an interaction between the sub-grid OD and convection schemes. Speeding up the wind over the orography in the NMOIST simulation probably results in less convergence just off the coast, reducing sub-grid convection and its associated momentum fluxes: convective precipitation is indeed reduced in that region.

Global simulations are used to produce precipitation forecasts, which may be improved in some circumstances by this modification to the OD scheme, in particular by the reduction in low-level sub-grid flow blocking. Figure 13 shows the changes to the mean surface precipitation produced in the N320 NMOIST simulations compared to the N320 CTRL simulations. For each case, the precipitation from the LAMf has been averaged onto the global grid, and the precipitation from each simulation has been averaged over the 24 hr or so of interest. The modified OD scheme has a similar effect on the rain patterns in both regions, shifting some of the surface rain slightly downstream of that in the CTRL simulation so that it is closer to the resolved mountain peak. In the Indian case, the dominant change is the reduction in precipitation along the coast and over the sea upstream due to reduced convergence, as discussed above. The precipitation pattern is improved in the NMOIST simulations, as shown by the
5 | TUNING OF THE OD SCHEME

Retuning of the OD scheme will be required when this modification is included in any future global atmospheric model configuration, as the present settings were chosen for the original dry version of the scheme. The aim of the tuning will be to give the best possible global performance primarily in midlatitude winds (Walters et al., 2017). As discussed by Hordin et al., (2017), tuning is an important step in optimizing model performance. The values of the various tuning parameters on which sub-grid parametrizations depend are often poorly constrained by observations. There may be preferred values based on theory or numerical experiments, for example, but these will not necessarily give the best global model performance. Optimal values vary with location at a given model resolution. For example, Vosper (2015) and Vosper et al., (2016) demonstrated that, while the OD scheme can be tuned to accurately represent the OD compared to high-resolution simulations, the optimal tuning varies with location. The hand-over between resolved and sub-grid drag was only well-behaved when the scheme was tuned optimally for each region. The Unified Model is used for simulations on a wide range of temporal and spatial scales, from short-duration NWP to long-duration low-resolution climate simulations. The seamless modelling approach adopted by the Met Office requires that the various tuning parameters be set to the same value in the global model at all horizontal resolutions. This is to provide robust GA configurations that our collaborators can use with confidence. However, optimal tuning parameter values for one resolution have not been found to be optimal at all other resolutions, so it has previously been necessary to prioritize performance at NWP resolutions. As a result, the current OD scheme produces too much low-level blocking and too little wave drag at climate resolutions (van Nierkerk et al., 2020), a problem that is shared by a number of models. Recent attempts to re-tune the current scheme to reduce this problem have proved fruitless. Williams et al., (2020) found that reducing \( n \) (which decreases the sub-grid hill height \( H_t \)) improved the performance of the climate simulation, but NWP forecast skill was degraded. This problem of too much low-level blocking at lower resolutions was responsible for the large negative low-level wind speed biases shown, for example, by the solid lines in Figure 5b. The modification to the OD scheme described in this paper greatly reduces this problem, so that the wind profiles from the different resolution simulations are much more comparable. This raises the possibility that inclusion of this modification may enable successful re-tuning of the OD scheme across a range of horizontal resolutions.

6 | A SUITE OF GLOBAL NWP FORECAST SIMULATIONS

This section describes results from a suite of 5-day global NWP forecast simulations at N320 resolution. These were initialised using Met Office operational global NWP analyses at 0000 UTC on a number of random dates, with the proviso that they be spread at least two weeks apart in order to minimize the synoptic correlation between cases. There were 12 simulations during the summers (JJA) of 2013 and 2014, plus 12 winter/spring simulations between 1 November 2013 and 1 April 2014. Fields from all simulations were averaged together at various lead times (out to 5 days) for each season. The effect of the modified OD scheme is investigated by examining the differences in the model fields from the NMOIST and CTRL simulations (NMOIST – CTRL).

These simulations used two more recent GA configurations. The first is GA7 (Walters et al., 2019), as this is the basis of the UK’s contribution to the Sixth Coupled Model Intercomparison Project (CMIP6; Eyring et al., 2016). The second is GA8, which is GA7 with some additional science changes. The science change most relevant to this study is the addition of the drag package described in Williams et al., (2020), which included a reduction in the amount of smoothing of the mean orography. This increases the resolved orography while reducing sub-grid orography, which means that more of the OD is explicitly resolved, and the effect of the modification to the OD scheme will be slightly reduced. The drag package included a number of other changes, such as the use of newer source orography datasets, improved roughness lengths and changes to the orographic form drag. (Note that the attempt to reduce the OD tuning parameter \( n \) described in that paper was not part of the final drag package for reasons discussed in the section on tuning.) Tests using these two different GA configurations yielded similar qualitative results, so only results using GA8 will be shown.

The previous section pointed out the obvious need to re-tune the OD scheme now that it has been modified to account for latent heating in its calculation of \( N \). The aim of any future OD tuning would be to give the best possible global performance for winds in the midlatitudes, where the OD scheme dominates, while avoiding any degradation of the verification against a range of other NWP metrics. This tuning will be performed on global NWP simulations incorporating a number of science changes (in addition to this one) to be included in the next GA configuration, some of which may also modify the winds. So
for now, changes to the mean wind fields produced by the modified OD scheme in this NWP suite will be investigated to see whether the effects identified for the localized case studies are robust.

The standard OD scheme used in the CTRL simulations predicts blocking over most major mountain ranges in both seasons, as shown in Figure 14a,b. Figure 14c,d demonstrate that the sub-grid blocked-layer depth $z_b$ is reduced over many mountainous regions of the globe in both seasons in the NMOIST simulations compared to the CTRL simulations. Although Figure 14 shows $z_b$ changes at T+24, changes are qualitatively similar at later lead times. Figure 15 shows the mean 850 mb zonal winds at T+24 in each season, along with the differences between the NMOIST and CTRL simulations. As a result of the reduced amount of sub-grid blocking (and wave) drag, mean low-level zonal winds are increased over mountains, particularly those lying across the moist westerly flow associated with a midlatitude storm track. Winds over the Southern Andes are faster in both seasons. There are regions of faster winds over the Rockies and the mountains of Norway which are located further north during summer as the storm tracks are displaced northward. Similarly, effects are seen more strongly over southern Europe and the Middle East during winter. Winds are strongly accelerated over the southern Himalayas during winter, although this is not seen in the 850 mb winds due to the high altitude of these mountains. (It is clearly seen in the 700 mb winds, which also show strong increases over the Rockies and southern Andes). The 850 mb westerly flow associated
with the Asian summer monsoon is faster over land in the NMOIST simulations.

Figure 16 shows the mean surface precipitation in each season at T+24, along with the changes produced in the NMOIST simulations. The precipitation fields are noisier than the wind fields, and averaging over many cases will reduce the effects as precipitation will not always accompany drag changes. However, it is still possible to identify regions near the windward slopes of significant mountains where there has been a downstream shift in precipitation. There is a definite downstream precipitation shift over the southwestern Himalayas, while shifts in precipitation over Europe are more complex, due to the three-dimensional nature of the mountains. For coastal midlatitude mountains such as the Rockies and the southern Andes, there is less precipitation over the sea immediately upstream of the orography, which is sometimes accompanied by an increase in precipitation over the windward mountain slopes. During the Asian summer monsoon, there is a decrease in precipitation over the seas immediately upstream of the Western Ghats (India) and the Arakan Yoma and Bilauktang (Myanmar).

At these early lead times, direct wind changes over the mountains were clearly seen, because far-field changes to the flow were still small and conditions upstream of the mountains were largely unchanged. (This was also true for the localized case-study simulations shown previously). As these simulations progress, however, the wind difference fields become much more complex, and changes are no longer confined to mountainous regions after 5 days (not shown). Changes to conditions upstream of the mountains complicate the analysis. In spite of this, the T+120 mean wind profiles from the NMOIST simulations show better agreement with radiosonde observations at all altitudes, in both seasons of both hemispheres. This is shown for the winter Northern Hemisphere (from 20° to 90°N) in Figure 17. Comparison against analyses also show improved tropospheric winds, but suggest that stratospheric winds are already too fast in the CTRL simulation. This is possibly due to insufficient sub-grid stratospheric wave drag, a point which was noted by van Niekerk et al., (2020) for the MetUM and a number of other models at low resolution. Faster winds in the NMOIST simulations make this positive stratospheric wind bias slightly worse, but this may be improved by the aforementioned retuning of the OD scheme. Similar results in terms of wind profiles were also obtained in more extensive tests using data assimilation.

Williams et al., (2020) described winter mean sea level pressure ($P_{msl}$) biases associated with excess amounts of low-level drag and low-level midlatitude winds that are too slow: a positive $P_{msl}$ bias at high latitudes and a negative bias at low latitudes. These biases are reduced in the NMOIST simulations, both in the N320 NWP case-study suite and in a 20-year N96 resolution climate simulation. Detailed results of the N96 climate simulations are not shown, as it is difficult to see the direct effects over orography due to long-term changes in the general circulation and hydrological cycle. However, RMS errors in tropospheric winds are generally reduced, as are surface precipitation errors. In addition, a dry precipitation bias over the Maritime Continent, and a negative screen temperature bias over northern land masses, are reduced.

In summary, the results from this suite of 24 NWP forecasts demonstrate that the low-level wind accelerations (and associated downstream shift in precipitation) produced in the earlier case-study simulations are a common occurrence over the Rockies and the Western Ghats, and that similar changes are observed over a number of other mountain ranges. Midlatitude winds are generally improved in the troposphere, which reduces the associated $P_{msl}$ biases, but stratospheric wave drag needs to be increased by a re-tuning of the scheme.

## 7 SUMMARY AND DISCUSSION

This paper describes a method for estimating the effects of moisture and latent heating on the $N$ profiles calculated by the orographic drag (OD) scheme. This is done by using $N_{sat}$ (Durran and Klemp, 1982b) during saturated sub-grid orographic vertical displacements. An estimate of the saturation level $\eta_c$, combined with the assumptions made about the sub-grid orographic ascent, gives information on the fraction of the sub-grid orographic vertical motions which are saturated. This allows for the calculation of an effective moist stability $N_{m}$ profile, which is then used in place of the dry value $N_d$ for low-level processes. The assumptions about the sub-grid ascent are (i) to estimate the depth of the sub-grid low-level blocking $z_{bl}$, air is assumed to ascend to the top of the sub-grid mountain $H_t$ in order to determine the mean Froude number $Fr_{av}$, and (ii) mountain waves are produced by ascent over the upper part of the sub-grid mountain rising above the blocked layer, referred to as the effective mountain height $h_{eff}$.

Verification of the scheme was performed against 1.5 km-resolution LAM simulations. Two case-studies were chosen for which moist westerly to southwesterly airflow at low levels impinged on significant orography: (a) the passage of a frontal system over California, USA with orographic precipitation enhancement over the mountains, and (b) summer monsoon flow over India with enhanced convective precipitation over the Western Ghats. The effect on the model winds was tested within a set of LAM simulations in which (a) the resolved and sub-grid
model orography was degraded to that used in global model at various resolutions, and (b) the OD scheme was switched on as in the global model. All other model settings and ancillary files were kept the same. The modified OD scheme performs well in both regions. In particular, it reduces the excessive sub-grid orographic blocking that is applied by the standard OD scheme, improving the low-level wind fields compared to the LAMf simulation using the full 1.5 km resolution orography.

The modified OD scheme was also shown to perform well in the global simulations, again reducing the low-level wind errors associated with excessive sub-grid blocking over the mountains at low resolution. This had the knock-on effect of improving the surface precipitation pattern over and upstream of the mountains, shifting some of the precipitation downstream towards the mountain peaks. For the Indian global simulations, it was found that the convection scheme has a larger effect on the winds at all levels than the OD scheme, but that the two schemes interact. Reducing the sub-grid low-level orographic blocking drag resulted in a reduction in the convergence occurring as the westerly flow decelerates towards the mountains, reducing the amount of convection occurring along the coast and over the sea upstream, and the associated negative low-level wind bias there.

A suite of 5-day global NWP forecasts at N320 resolution, consisting of 12 summer and 12 winter/spring simulations, was used to demonstrate that these results are robust. Faster low-level winds were produced over mountains lying in the path of the midlatitude storm tracks in both seasons, while the low-level westerlies associated with the summer monsoon were increased over mountainous regions of Asia. Associated with these wind accelerations are regions with a local downstream displacement of precipitation, with precipitation reduced over the sea immediately upstream of coastal mountains.

The representation of moisture and latent heat effects on the buoyancy frequency, $N$, is a physical improvement, representing a process that has previously been neglected by OD schemes. Some version of this change should therefore be included in coarse-resolution simulations. Re-tuning of the OD scheme will be required when this modification is included in any future global atmospheric model configuration, as the present settings were chosen for the original dry version of the scheme. In particular, evidence has been presented that stratospheric wave drag...

**FIGURE 15** Mean 850 mb zonal windspeeds (m s$^{-1}$) during the (a) twelve winter and (b) twelve summer NWP case-studies at N320 resolution after 24 hr of simulation. (c, d) show the corresponding changes produced by the NMOIST simulation.
should be increased, although this problem already exists to a smaller extent in the current OD scheme. Tuning has previously been found to be highly problematic because optimal values vary with location for a given model resolution due to differences in the dominant mountain length-scale (Vosper, 2015; Vosper et al., 2016). In addition,
optimal OD tuning parameter values for one resolution have so far not been found to be optimal at all other resolutions. So it has previously been necessary to prioritize performance at NWP resolutions, with the result that there is too much low-level blocking and too little wave drag at low resolutions, particularly those used for climate simulations. Accounting for latent heat effects in the OD scheme has greatly reduced this problem. Low-level winds from different resolution simulations are now much more comparable, suggesting that it will now be easier to obtain optimal tuning of the OD scheme over a wide range of resolutions.

This modification means that the OD scheme now uses a more accurate stability profile for estimating blocking and total wave drag, with low-level stability now reduced due to moist effects. In particular, the modified OD scheme is able to identify moist unstable or neutral layers, demonstrated by the larger number of grid-boxes in the NMOIST simulations with \( N_{\text{low}} \leq 0 \) (Figure 4). Unstable or neutral layers are common in the boundary layer, particularly in the daytime or in the Tropics. Examples of \( N \) profiles over the mountains during these two cases are shown in Figure 18. There is a moist unstable boundary layer present in the \( N_{\text{sat}} \) profile over India which grows during the day. In contrast, \( N_d \) remains positive at nearly all altitudes for this profile. The \( N \) profiles over the USA are more difficult to assess, with large vertical variations in both \( N_d \) and \( N_{\text{sat}} \), with \( N_{\text{sat}} \) sometimes becoming negative but \( N_d \) remaining positive. This large vertical variation means that values of \( N_{\text{low}} \) will depend on the exact averaging layer used.

The moist stability described in this paper should be closer to reality than the dry value currently used when sub-grid orographic ascent results in saturation. It should be noted, however, that it was necessary to make some simplifying assumptions, which will be discussed here. Firstly, ice processes were neglected. Supercooled water is likely to be important over the smaller-scale orographic surface variations represented by the sub-grid OD scheme. This is due to the rapid reaction of the condensation process to supersaturation, while ice formation generally requires much longer time-scales. Even if freezing does occur, it occurs only at low temperatures where the reduction in \( N \) due to latent heating is small. A number of previous studies support the importance of supercooled water over orography. In simulations of flow over a large-scale mountain using a cloud-resolving model, Kirshbaum and Smith (2008) demonstrated that adding a series of small-scale peaks produced locally strong updraughts containing supercooled water, while Morales et al., (2018) demonstrated that accretion and riming seem to be the most important mechanisms producing orographic precipitation.

Another assumption is that, once an air parcel rises above the LCL on the windward side of a sub-grid hill, it remains saturated until it descends back below the same LCL during lee-side descent. The removal of water by precipitation enhancement is neglected. The amount of blocking is predicted by considering whether the incoming flow will be able to ascend the windward slope of the sub-grid hill. This calculation would not be modified by precipitation unless it results in sub-saturation before the air reaches the hill top, which is less likely for the smaller-scale hills generally represented by the sub-grid OD scheme. However, drying by precipitation may raise the level at which the sub-grid orographic water is completely evaporated during lee-side descent (Miglietta and Rotunno, 2005), which may be important when calculating the total wave drag in a column. It would be quite difficult to represent this drying effect due to precipitation within the OD scheme as precipitation formation and fallout depends on the advection time across individual sub-grid hills.

There may be some assumptions already made by the OD scheme which will need to be reconsidered when using this modification to the low-level stability. The most obvious of these is that a neutral or unstable atmosphere is unable to support either gravity wave propagation or blocking. The OD scheme currently deals with \( N_{\text{low}} \leq 0 \) by not applying any drag at all. This assumption may be an oversimplification of a complex problem, which will be more of an issue now that the OD scheme can identify moist unstable and neutral layers. Some studies have suggested that, while the drag may be significantly reduced in the presence of a convective or neutral BL, it is not eradicated completely. For example, Jiang and Doyle (2008) found that for an extremely small (10 m high) hill, a daytime convective BL greatly weakens the waves and associated drag (reduced by 90%), but does not remove it completely. Meanwhile, Peng and Thompson (2003) found that a neutral BL only reduced the wave amplitudes, with the top of the boundary layer forcing waves instead of the mountain itself. Therefore, further studies should investigate how best to treat such neutral and unstable layers within OD schemes. Changes to the averaging depth for \( N_{\text{low}} \) could also be considered as a simpler way to include the effect of the stable layer aloft. This is already done for \( N_{\text{av}} \) (Vosper et al., 2009), the average used to estimate the Froude number and therefore the blocked-layer depth, although that study was only conducted for a neutral boundary layer. The exact choice of averaging layer for \( N_{\text{low}} \) will now have a much larger effect on the resulting drag in the presence of both stable and unstable layers.

The presence of convection complicates matters significantly. Convection modifies the wind, moisture and
FIGURE 18 Profiles of $N_2^2$ (solid) and $N_{sat}^2$ (dashed) (s$^{-2}$) calculated from profiles extracted from the N96 CTRL simulations over (a, b) the USA and (c, d) India at (a, c) 0000 UTC and (b, d) 1200 UTC. Local times are 1600 hr and 0400 hr on 30 November 2012 for the USA and 0530 hr and 1730 hr on 15 August 2018 for India.

Temperature profiles via vertical turbulent mixing. Evaporatively cooled downdraughts may produce cold air outflows below cloud base (Miglietta and Rotunno, 2009), while the evaporation of precipitation falling below cloud base will result in latent cooling. All of these processes modify the thermodynamic profile, which determines the drag due to underlying orography. These processes therefore need to be correctly represented at low resolution by the convection scheme in order to provide accurate thermodynamic profiles for use in the OD scheme. Even a perfect OD scheme would be unable to correctly represent the drag if the thermodynamic profiles are inaccurate. Modifications to the model wind fields by the OD scheme, in turn, affect when and where convection occurs in the model. This complex interaction between the OD and convection schemes could be an interesting avenue for future research.

The above is also true when convection occurs upstream of the orography. Satellite observations of Asian summer monsoon precipitation show that convective monsoon rainfall is anchored on the windward side of narrow mountain ranges such as the Western Ghats (Xie et al., 2006), with rainfall maxima generally located 50 to 200 km offshore. For midlatitude mountainous regions such as the European Alps, a pre-existing blocked layer may act as a virtual barrier to incoming flow, producing ascent upstream of the actual mountains. If this upstream ascent produces convection, that convection modifies the thermodynamic profile incident on the mountains downstream. For example, latent cooling due to the evaporation of precipitation, and cold air outflows below cloud base, may reinforce the pre-existing blocked layer, which otherwise may have been swept away (Davolio et al., 2016).

ACKNOWLEDGEMENTS
This work and contributors were supported by the Met Office Weather and Climate Science for Service.
Partnership (WCSSP) India as part of the Newton Fund. Thanks to Annelize van Neikerk and Simon Vosper for useful feedback on the manuscript.

ORCID
Samantha Smith https://orcid.org/0000-0001-7228-5162

REFERENCES

Alduchov, O.A. and Eskridge, R.E. (1996) Improved Magnus form approximation of saturation vapor pressure. Journal of Applied Meteorology, 35, 601–609

Barcilon, A., Jusem, J.C. and Drazin, P.G. (1979) On the two-dimensional, hydrostatic flow of a stream of moist air over a mountain ridge. Geophysical and Astrophysical Fluid Dynamics, 13, 125–140

Bush, M., Allen, T., Bain, C., Boutle, I., Edwards, J., Finnenkoetter, A., Franklin, C., Hanley, K., Lean, H., Lock, A.P., Manners, J., Mittermaier, M., Morcette, C.J., North, R., Petch, J., Short, C., Vosper, S.B., Walters, D., Webster, S., Weeks, M., Wilkinson, J., Wood, N. and Zerroukat, M. (2020) The first Met Office Unified Model–JULES Regional Atmosphere and Land configuration, RAL1. Geoscientific Model Development, 13, 1999–2029

Colle, B.A., Westrick, K.J. and Mass, C.F. (1999) Evaluation of MM5 andEta-10 precipitation forecasts over the Pacific Northwest during the cool season. Weather and Forecasting, 14, 137–154

Colle, B.A., Wolfe, J.B., Steenburgh, W.J., Kingsmill, D.E., Cox, J.A.W. and Shafer, J.C. (2005) High-resolution simulations and microphysical validation of an orographic precipitation event over the Wasatch Mountains during IPEX IOP3. Monthly Weather Review, 133, 2947–2971

Davolio, S., Volonté, A., Manzato, A., Pucillo, A., Cicogna, A. and Ferrario, M. (2016) Mechanisms producing different precipitation patterns over northeastern Italy: insights from HyMeX-SOP1 and previous events. Quarterly Journal of the Royal Meteorological Society, 142, 188–205

Durran, D.R. (1990). Mountain waves and downslope winds. In W. Blumen (Ed.), Atmospheric Processes over Complex Terrain, pp. 59–81. Berlin: Springer.

Durran, D.R. and Klemp, J.B. (1982a) The effects of moisture on trapped mountain lee waves. Journal of the Atmospheric Sciences, 39, 2490–2506

Durran, D.R. and Klemp, J.B. (1982b) On the effects of moisture on the Brunt–Väisälä frequency. Journal of the Atmospheric Sciences, 39, 2152–2158

Durran, D.R. and Klemp, J.B. (1983) A compressible model for the simulation of moist mountain waves. Monthly Weather Review, 111, 2341–2361

Eiserloh, A.J. and Chiao, S. (2015) Modeling studies of landfalling atmospheric rivers and orographic precipitation over northern California. Meteorology and Atmospheric Physics, 127, 1–16

Elvidge, A.D., Sandu, I., Wedi, N., Vosper, S.B., Zdra, A., Boussetta, S., Bouyssel, F., van Niekerk, A., Tolstykh, M.A. and Ujije, M. (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, 2567–2585

Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J. and Taylor, K.E. (2016) Overview of the Coupled Model Intercomparison Project phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development, 9, 1937–1958

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., Rauser, F., Rio, C., Tomassini, L., Watanabe, M. and Williamson, D. (2017) The art and science of climate model tuning. Bulletin of the American Meteorological Society, 98, 589–602

Jiang, Q. (2003) Moist dynamics and orographic precipitation. Tellus A: Dynamic Meteorology and Oceanography, 55, 301–316

Jiang, Q. and Doyle, J.D. (2009) The impact of moisture on mountain waves during T-REX. Monthly Weather Review, 137, 3888–3906

Jiang, Q. and Doyle, J.D. (2008) On the diurnal variation of mountain waves. Journal of the Atmospheric Sciences, 65, 1360–1377

Kirshbaum, D.J. and Smith, R.B. (2008) Temperature and moist-stability effects on midlatitude orographic precipitation. Quarterly Journal of the Royal Meteorological Society, 134, 1183–1199

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

McIlveen, J.F.R. (1991) Fundamentals of Weather and Climate. Abingdon: Routledge.

Miglietta, M.M. and Buzzi, A. (2001) A numerical study of moist stratified flows over isolated topography. Tellus A, 53, 481–499

Miglietta, M.M. and Rotunno, R. (2005) Simulations of moist nearly neutral flow over a ridge. Journal of the Atmospheric Sciences, 62, 1410–1427

Miglietta, M.M. and Rotunno, R. (2009) Numerical simulations of conditionally unstable flows over a mountain ridge. Journal of the Atmospheric Sciences, 66, 1865–1885

Morales, A., Morrison, H. and Posselt, D.J. (2018) Orographic precipitation response to microphysical parameter perturbations for idealized moist nearly neutral flow. Journal of the Atmospheric Sciences, 75, 1933–1953

Peng, M.S. and Thompson, W.T. (2003) Some aspects of the effect of surface friction on flows over mountains. Quarterly Journal of the Royal Meteorological Society, 129, 2527–2557

Reeves, H.D., Lin, Y.-L. and Rotunno, R. (2008) Dynamic forcing and mesoscale variability of heavy precipitation events over the Sierra Nevada Mountains. Monthly Weather Review, 136, 62–77

Smith, S.A., Field, P.R., Vosper, S.B., Shipway, B.J. and Hill, A.A. (2016) A parametrization of subgrid orographic rain enhancement via the seeder–feeder effect. Quarterly Journal of the Royal Meteorological Society, 142, 132–142

Smith, S.A., Field, P.R., Vosper, S.B. and Derbyshire, S.H. (2019) Verification of a seeder–feeder orographic precipitation enhancement scheme accounting for low-level blocking. Quarterly Journal of the Royal Meteorological Society, 145, 2909–2932

Valenzuela, R.A. and Kingsmill, D.E. (2017) Terrain-trapped airflows and orographic rainfall along the coast of northern California. Part I: Kinematic characterization using a wind profiling radar. Monthly Weather Review, 145, 2993–3008

Vosper, S.B. (2015) Mountain waves and wakes generated by South Georgia: implications for drag parametrization. Quarterly Journal of the Royal Meteorological Society, 141, 2813–2827

Vosper, S.B., Wells, H. and Brown, A.R. (2009) Accounting for non-uniform static stability in orographic drag parametrization.
APPENDIX A. ESTIMATE OF $N_m$ FOR GRAVITY WAVE VERTICAL PROPAGATION

Section 2 described how to estimate an effective moist buoyancy frequency $N_m$ for calculating the low-level averages in Table 1, accounting for the effect of latent heating on both the low-level blocking and on the total sub-grid wave stress in the column. Moisture and latent heating may also have an effect on the vertical propagation of the gravity waves above the sub-grid mountain tops, and therefore the levels at which they saturate. Estimating an effective moist buoyancy frequency $N_m$ for these gravity wave propagation and saturation calculations requires comparing the sub-grid orographic vertical displacements against the saturation level $\eta$ as in Section 2. However, both ascents ($\eta_{\text{up}}$) and descents ($\eta_{\text{down}}$) will occur, so the total saturated and sub-saturated displacements from the model level at altitude $z$ are

$$\eta_{\text{sat}} = \eta_{\text{up}} - \eta_c, \quad (A1)$$

$$\eta_{\text{unsat}} = \eta_{\text{down}} + \eta_c, \quad (A2)$$

where $\eta_c$ is positive for a sub-saturated grid-box, or zero for a grid-box that is saturated with no cloud water. For a saturated grid-box containing resolved cloud water, $\eta_c$ is negative, as it is the descent required to evaporate all of the resolved model cloud liquid water $q_L$ as described in Smith et al. (2016)

$$\eta_c = -q_L \frac{dq_L}{dz}, \quad (A3)$$

where $dq_L/dz$ is the adiabatic rate of change of $q_L$ per metre of vertical displacement

$$\frac{dq_L}{dz} = -\frac{dq_{\text{sat}}}{dz} = \frac{(\epsilon + q_{\text{sat}})q_{\text{sat}}L_v}{R_d T^2} \Gamma_s - \frac{q_{\text{sat}}P_g}{(P - e_{\text{sat}})R_d T}. \quad (A4)$$

$P$ is the air pressure and $e_{\text{sat}}$ is the saturation vapour pressure, which must both be in the same units (e.g. Pa). If the grid-box is saturated, or if saturation is expected to occur during the assumed sub-grid orographic ascent, then the effective moist stability is given by

$$N_m^2 = \frac{(N_{\text{unsat}}^2 \eta_{\text{unsat}}) + (N_{\text{sat}}^2 \eta_{\text{sat}})}{(\eta_{\text{up}} + \eta_{\text{down}})} = 0, \quad (A5)$$

giving $N_m = N_{\text{unsat}}$ if all of the orographic displacements are unsaturated and $N_m = N_{\text{sat}}$ if all of the orographic displacements are saturated.

The vertical displacements produced by an actual gravity wave field are complex and may consist of many horizontal scales, with shorter-scale (non-hydrostatic) waves propagating horizontally, as well as vertically. The

How to cite this article: Smith S. Accounting for moist processes in a sub-grid orographic drag scheme. Q.J.R. Meteorol. Soc. 2021;147:2124–2146. https://doi.org/10.1002/qj.4013
simplest assumption that could be made is

\[ \eta_{\text{up}}(z) = \eta_{\text{down}}(z) = 0.5 A(z). \]  \hspace{1cm} \text{(A6)}

where \( A(z) \) is the gravity wave amplitude predicted by the OD scheme at altitude \( z \). Simulations using this assumption will be referred to as MoistProp1 (MP1).

Other assumptions could be also made. Figure 4.3b in Durran (1990) shows the pattern of vertical displacements produced by a hydrostatic wave. Immediately above a hill, the wave phase oscillates with altitude from pure ascent at altitudes given by \( h + n \lambda_z \), to pure descent at altitudes given by \( h + (n + 0.5) \lambda_z \). The integer \( n \) corresponds to the number of full wavelengths already encountered at lower levels, starting from 0 within the first wavelength above the hill. These variations in the amount of ascent and descent with altitude are assumed here, for simplicity, to be linear over each half a wavelength. Therefore over the lower half of a wave, \( \eta_{\text{up}}(z) \) decreases linearly from the wave amplitude to zero, while \( \eta_{\text{down}}(z) \) increases linearly from zero to the wave amplitude

\[ \eta_{\text{up}}(z) = A(z) \left( 1 - \frac{2 \Delta z_w}{\lambda_z} \right), \]  \hspace{1cm} \text{(A7)}

\[ \eta_{\text{down}}(z) = A(z) \frac{2 \Delta z_w}{\lambda_z}, \]  \hspace{1cm} \text{(A8)}

where \( \Delta z_w \) is the altitude through the current half-wavelength, varying from 0 to 0.5. The opposite is true for the upper half of a wave

\[ \eta_{\text{up}}(z) = A(z) \frac{2 \Delta z_w}{\lambda_z}, \]  \hspace{1cm} \text{(A9)}

so that \( \eta_{\text{up}}(z) \) increases and \( \eta_{\text{down}}(z) \) decreases with altitude. The vertical wavelength, \( \lambda_z = 2 \pi U / N \), is averaged vertically from the gravity wave launch altitude to the level being considered. Simulations using this assumption will be referred to as MoistProp2 (MP2).

As both \( A \) and \( \lambda_z \) depend on \( N \), it is necessary to perform these calculations iteratively until \( N_m^2 \) has converged.

As this calculation of \( N_m \) could therefore be more expensive, it is only performed when \( \Gamma_d \) and \( \Gamma_s \) differ by more than 0.3% of \( \Gamma_s \).

The LAMn96 USA simulations were rerun using \( N_m \) estimated using these two different assumptions. Figure A1 compares the wind profiles extracted from these simulations against the CTRL and NMOIST simulations described in the main body of the paper. The differences between the winds from the NMOIST, MoistProp1 and MoistProp2 simulations are so small that they look almost identical, and there are only extremely small differences in the wind errors. In contrast, the winds from the CTRL simulation look significantly different. Similar results were obtained over India and in the LAMn768 simulations.

This suggests that, while it is important to account for low-level latent heat effects when estimating both the sub-grid orographic blocking and the total wave stress in a column \( \tau_w \), it is possible that latent heat effects on the wave vertical propagation and saturation calculations may be safely neglected. (This part of the scheme only determines how \( \tau_w \) is distributed in the vertical.) Further studies are needed to confirm this, perhaps using longer simulations or idealized studies.