University of Reading

Impact of model upgrades on diabatic processes in extratropical cyclones and downstream forecast evolution

Article

Accepted Version

Maddison, J. W., Gray, S. L. ORCID: https://orcid.org/0000-0001-8658-362X, Martínez-Alvarado, O. ORCID: https://orcid.org/0000-0002-5285-0379 and Williams, K. D. (2020) Impact of model upgrades on diabatic processes in extratropical cyclones and downstream forecast evolution. Quarterly Journal of the Royal Meteorological Society, 146 (728). pp. 1322-1350. ISSN 0035-9009 doi: 10.1002/qj.3739 Available at https://centaur.reading.ac.uk/88544/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.
Published version at: http://dx.doi.org/10.1002/qj.3739
To link to this article DOI: http://dx.doi.org/10.1002/qj.3739

Publisher: Royal Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

University of Reading

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

ORIGINAL ARTICLE

Journal Section

Impact of model upgrades on diabatic processes in extratropical cyclones and downstream forecast evolution

Jacob W. Maddison[1]* | **Suzanne L. Gray**[1] | **Oscar Martínez-Alvarado**[2,1] | **Keith D. Williams**[3]

[1]Department of Meteorology, University of Reading, United Kingdom

[2]National Centre for Atmospheric Science-Atmospheric Physics, United Kingdom

[3]Met Office, United Kingdom

Correspondence
Department of Earth Physics and Astrophysics, Universidad Complutense de Madird, Madrid, Spain
Email: jacobmad@ucm.es

Funding information
NERC, Grant/Award Number: NE/M009610/1; UK BEIS/Defra, Grant/Award Number: GA01101

Models are continuously developed at numerical weather

* Corresponding Author.

prediction (NWP) centres to improve forecast skill, with new operational model configurations adopted every few years. The parameterisations of diabatic processes are probably the most frequently updated part of NWP models as they are crucial for accurate weather predictions and contain uncertainties in their formulation. The impact of model developments is assessed here in forecasts from the Met Office's weather forecast model initialised throughout the North Atlantic Waveguide and Downstream Impact Experiment field campaign period in autumn 2016. Planned model parameterisation developments are considered, together with an 'inexpensive coupled' forecast with daily updating of the sea surface temperature and sea-ice fraction. Forecasts produced from the coupled system have, on average, indistinguishable skill from the control forecasts, suggesting the benefits of coupled atmosphere-ocean NWP systems can be small. In contrast, a reduction in forecast error (~4%) is identified in forecasts produced using an upgraded convection scheme. Periods of low forecast skill during the study period are shown to be associated with the onset and decay of blocking events and increased diabatic heating of air masses reaching the upper troposphere. In forecasts of a specific block development case that was not accurately predicted in any of the experiments or in the operational ensemble forecast from the Met Office, the representation of diabatic heating in the warm conveyor belt of an upstream cyclone is shown to moderate the subsequent block development: forecasts in which the heating is stronger generally have a more-amplified blocking ridge and amplified heating contributions from all parameterisations as diagnosed using diabatic tracers. Hence, we demonstrate that plausible changes to the representation of several different diabatic processes in models can impact forecast block development via changes within upstream cyclones.

numerical weather prediction; parameterised diabatic processes; atmospheric blocking; warm conveyor belts

1 | INTRODUCTION

Numerical weather prediction (NWP) centres across the globe routinely produce forecasts of the weather for the coming days, weeks and even months, with the use of highly complex numerical models of the atmosphere. The models generally consist of a dynamical core that is used to numerically solve the atmosphere's equations of motion; a data assimilation system that ingests millions of observations into the model to produce an initial state of the atmosphere; a set of parameterisation schemes needed to represent processes occurring on scales smaller than those producible by the model; and boundary conditions for the model such as a description of roughness length, sea surface temperature (SST), soil moisture and vegetation cover. Each component of the NWP model is being constantly developed in an effort to improve the forecasts and new model configurations, consisting of many changes to its various parts, become operational every few years. The parameterisations within the model, representing processes occurring on

scales smaller than the grid scale, are one of the most frequently updated components of the model. They are known to have a large impact on forecast error (e.g. Bauer et al., 2015), and contain uncertain parameters that need to be tuned to correspond with changes elsewhere in the model. The purpose of this article is to quantify the relative impact of several operational improvements to parameterisations that affect diabatic processes on forecasts of extratropical circulation features.

Weather forecasts can have errors that arise from uncertainty in defining an initial state of the atmosphere and imperfections in NWP model formulation. These are commonly termed initial condition (or analysis) and model errors, respectively. Ensemble prediction systems (EPSs) have been introduced at many operational NWP centres in recent decades to account for both initial condition and model error (Buizza et al., 2005). To represent initial condition uncertainty, the probability distribution function (PDF) of possible initial conditions is sampled by adding dynamically-defined perturbations to the model's analysis (e.g. Molteni et al., 1996; Bowler et al., 2008); an ensemble of simulations is produced giving a probabilistic description of the weather. Model error is typically accounted for by the addition of stochasticity to parameterisation schemes (Buizza et al., 1999). Initial condition error is typically larger than model error in medium-range weather forecasts (Arpe et al., 1985; Rabier et al., 1996), though several studies have shown that model error cannot be ignored (Harrison et al., 1999; Buizza et al., 1999; Stensrud et al., 2000; Orrell et al., 2001), and both initial condition and model error can depend on geographical region and weather pattern. Forecast uncertainty in operational NWP models is generally highest for the transition to, and maintenance of, a blocked state (Ferranti et al., 2015; Matsueda and Palmer, 2018). Improvements to the models' parameterisation schemes (Jung et al., 2010; Dawson and Palmer, 2015; Joos and Forbes, 2016) and dynamical cores (Martínez-Alvarado et al., 2018), as well as increased resolution (Matsueda, 2009; Davini and D'Andrea, 2016; Schiemann et al., 2017), have been shown to improve forecasts of extratropical circulation features and atmospheric blocking in NWP and climate models. Boundary conditions, in particular SST, have also been shown to affect block development (Scaife et al., 2011; O'Reilly et al., 2016) and increasing the spatial resolution of SST can improve operational forecasts from the European Centre for Medium-range Weather Forecasts (ECMWF, Chelton, 2005).

Diabatic processes active in extratropical cyclones typically occur on scales smaller than the grid in global weather forecast models and hence must be parameterised. These diabatic processes, especially latent heat released by the condensation of water vapor, can contribute considerably to the intensification of extratropical cyclones (Davis and Emanuel, 1991; Stoelinga, 1996). Therefore their parameterisation is key for the accurate forecast of these systems. Latent heat release in the mid-troposphere produces a positive low-level potential vorticity (PV) anomaly below the region of maximum heating which can induce rapid cyclogenesis if it favorably interacts with an upper-level positive PV anomaly (Wernli et al., 2002; Ahmadi-Givi et al., 2004). A negative PV anomaly is produced above the region of maximum heating and can influence the development of upper-level ridges and modify the downstream flow pattern (Pomroy and Thorpe, 2000; Grams et al., 2011). One effect of parameterised diabatic processes in extratropical cyclones is to enhance the tropopause-level PV gradient (Chagnon et al., 2013) where a sharp increase from low PV values in the troposphere to high PV values in the stratosphere is observed. Warm conveyor belts (WCBs), the broad cloudy ascending airstreams flowing poleward in the warm sector of extratropical cyclones, are the regions of primary latent heat release in cyclones (Browning and Roberts, 1994) and can be crucial for the rapid intensification of explosive storms (Binder et al., 2016). The representation of WCBs in extratropical cyclones is also sensitive to the parameterisation of diabatic processes (Joos and Wernli, 2012; Joos and Forbes, 2016) which, in turn, can affect the upper-tropospheric PV structure and downstream flow development (Pomroy and Thorpe, 2000). This relationship suggests that the parameterisation of diabatic processes could also be important for the representation of the upper-tropospheric negative PV anomalies that define blocks. We investigate the impact of model physics uncertainty on the representation of upper-tropospheric ridge amplification and atmospheric blocking here.

Extratropical cyclones force the geopotential height rises in developing blocks through thermal and vorticity advection (Colucci, 1985; Nakamura and Wallace, 1993) and the repeated transfer of low-PV air polewards and upwards within cyclones into blocking ridges can act to maintain them against dissipation (Shutts, 1983; Yamazaki and Itoh, 2013; Luo et al., 2014). Diabatically-heated air masses can contribute considerably (>50%) to the total mass of blocked regions in the northern hemisphere (Pfahl et al., 2015) and diabatic heating is typically strong during the onset of blocking (Steinfeld and Pfahl, 2019). Atmospheric blocks are notoriously difficult to forecast in NWP models (e.g. Tibaldi and Molteni, 1990; Pelly and Hoskins, 2003a; Matsueda, 2009) and are the cause of some of the worst forecasts produced at operational NWP centres (Rodwell et al., 2013; Lillo and Parsons, 2017). Grams et al. (2018) showed

in a case study of a large forecast error originating from a missed blocking event that it was the poor forecast of an upstream cyclone's WCB that was the cause of the error. Forecasts of the most uncertain block onset cases in recent years were shown to be strongly influenced by the representation of upstream extratropical cyclones by Maddison et al. (2019). Furthermore, ridges in the tropopause are systematically misrepresented in operational NWP models (Gray et al., 2014) with the area and isentropic PV gradient of ridges decreasing with lead time. The sharpening of the isentropic PV gradient by parameterised diabatic processes can be too weak to maintain the strong PV-gradients in forecasts (Saffin et al., 2017). A case study that had an error representative of this systematic bias was caused by the poor forecast (too weak and too far south) of WCB outflow (Martínez-Alvarado et al., 2016).

Previous studies investigating the sensitivity of extratropical cyclones and atmospheric blocking to model physics or boundary conditions have often made large changes in parts of the model to obtain a large response in the representation of the feature of interest, such as removing moisture (Davis et al., 1993; Wernli et al., 2002; Coronel et al., 2015), changing parameterisation schemes (Carrera et al., 1999; Dearden et al., 2016) and reducing the physical tendencies from parameterisations (Martínez-Alvarado and Plant, 2014). Generally, cyclones are less intense and downstream upper-level ridges smaller when the diabatic effects are reduced. Smoothing the SST gradient near the Gulf Stream, or reducing the SST and maintaining the sharp front, can also reduce the intensity of cyclones (Sheldon et al., 2017) and reduce the frequency of blocking over Europe (O'Reilly et al., 2016). The approach taken here is different: we make small changes to various parameterisation schemes that constitute (or could constitute) an operational upgrade to the scheme. Joos and Forbes (2016) showed that an operational upgrade to the microphysical parameterisation in the ECMWF Integrated Forecasting System (IFS) had an impact on the simulation of a cyclone's WCB and downstream ridge building. Whilst changes to the forecast evolution were initially small, they were advected with and amplified by the flow resulting in marked differences in the upper-tropospheric PV pattern. An operational improvement to the radiation scheme in the IFS was also shown to systematically improve the skill of medium-range weather forecasts of geopotential at 200, 500 and 1000 hPa (Morcrette et al., 2008). A new operational version (cycle) of the ECMWF model, that differed primarily in its parameterisations of physical processes, was shown to improve the representation of blocking in seasonal forecasts (Branković and Molteni, 1996). In this article, planned operational upgrades to the Met Office Unified Model (MetUM) to various physical parameterisations are introduced separately to assess the impact of each parameterisation on the forecast evolution.

The research presented in this article addresses the following research questions:

1. Do model upgrades to physical parameterisations and boundary conditions have a systematic effect on forecasts of upper-tropospheric Rossby wave development and blocking?
2. How do the effects on the forecast evolution from model changes compare to those from initial condition uncertainty?
3. Does the previously found error in block forecasts associated with upstream cyclone representation and WCB structure (e.g. Matsueda, 2011; Grams et al., 2018; Maddison et al., 2019) originate from uncertainty in the representation of diabatic processes in extratropical cyclones?

These questions are motivated by relatively new results showing that the dynamics and predictability of upper-tropospheric Rossby waves and atmospheric blocking events are influenced strongly by diabatic processes within extratropical cyclones (Pfahl et al., 2015; Steinfeld and Pfahl, 2019), especially those active in their WCBs (Martínez-Alvarado et al., 2016; Grams and Archambault, 2016; Grams et al., 2018). In this article we follow forecast evolution changes that are triggered by the model upgrade's effect on parameterised diabatic heating from cyclone and WCB development to upper-level Rossby wave amplification and blocking.

The article is organised as follows. Details of the model simulations used in this study are given in section 2 and the methods used to analyse them are described in section 3. An overview of the results from forecasts initialised across the study period are presented in section 4. In section 5, a particular case study of block onset following extratropical cyclone intensification is described together with a verification of the operational forecast of the case and its representation in the study experiments. The sensitivity experiments are compared with the control simulation in section 6, with a focus on the role of diabatic processes in the sensitivity of upper-tropospheric flow to parameterised physical processes. We summarise the results and give conclusions in section 7.

2 | MET OFFICE UNIFIED MODEL EXPERIMENTS

In this section, details are given of the sensitivity experiments that are performed with the MetUM and analysed in this study. The MetUM is the numerical model developed and used by the Met Office for both weather and climate modelling. The model is continuously developed with new configurations being implemented operationally every few years. The experiments used in this study are mainly based around an operational upgrade in the MetUM from configuration Global Atmosphere (GA)6.1 to GA7.0, which is scheduled for November 2019. Experiments implementing the GA7 cloud, microphysics and boundary layer schemes into the GA6.1 configuration of the model are performed as well as a convection parameterisation experiment and a SST updating experiment. The control run is a MetUM GA 6.1 (Walters et al., 2017) run at N768 resolution with 70 model levels and an 80 km top. This is a rerun of the operational configuration of the MetUM that was used during the study period (autumn 2016). The experiments are summarised in Table 1 and a more detailed description of the MetUM and each of the sensitivity experiments is given in the Appendix. The experiments performed for this article are all single member runs and together create an ensemble of model perturbed forecasts. We compare the individual experiments to the operational ensemble forecasts with perturbed initial conditions.

TABLE 1 Summary of forecast experiments.

Name	Experiment description
SST-update	SST and sea-ice fractions are updated daily during the forecast evolution to the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA, Donlon et al., 2012)
Prog-ent	The convection scheme in the GA6.1 configuration is changed to the prognostic entrainment convection scheme that has been developed at the Met Office to include memory of recent convective activity
GA7Mp	The modifications to the microphysical parameterisation scheme for the GA7 configuration are implemented, including a new treatment of sub-grid cloud water content variability; a change to the warm rain microphysics; and a change to the turbulent production of liquid water in mixed-phase clouds
GA7Cl	The modifications to the large-scale cloud parameterisation scheme for the GA7 configuration are implemented, including a representation of convective core radiative effect; a new critical relative humidity based on turbulence; and the removal of complexity when dealing with ice cloud
GA7Bl	The modifications to the boundary layer parameterisation scheme for the GA7 configuration are implemented, including a dependence on coupling for boundary layer entrainment; a representation of clouds forming at boundary layer top; and a retuned cloud threshold used to diagnose a shear dominated boundary layer

The sensitivity experiments are separated into two groups: (1) the SST-update and Prog-ent experiments and (2) the experiments in which the different schemes are modified to their GA7 versions. The Prog-ent and SST-update experiments have been initialised every 12 hours (at 00 and 12 UTC) during the North Atlantic Wave and Downstream impact EXperiment (NAWDEX, Schäfler et al., 2018) field campaign period (20 September–16 October 2016) and integrated for 12 days, giving a total set of 54 12–day runs. NAWDEX was a field campaign investigating the diabatic influence on the jet stream and high impact weather over Europe. The NAWDEX campaign period included several extratropical cyclones and upper-level ridge building events which make it an ideal study period for this work. Whilst the SST-update experiment could never be implemented to produce an operational forecast, it is used to provide insight into the potential benefit of using a coupled model system for NWP. The GA7 physics experiments (GA7Mp, GA7Cl, GA7Bl) have been run for a single chosen forecast initiation date to determine their impact for a blocking case

study. All of the experiments are global model runs with output of certain diagnostics restricted to every 12 hours and in the Euro-Atlantic region (defined as 20.25°–80.55°N, 79.875°W–40.725°E, which covers the region of interest for the NAWDEX campaign period).

3 | METHODS

3.1 | Block identification

Blocking is identified in this study using a modified version of the 2D 500 hPa geopotential height ($Z500$) blocking index introduced by Scherrer et al. (2006), an extension of the 1D index of Tibaldi and Molteni (1990). The index is calculated using meridional gradients in $Z500$, termed GHGN and GHGS for northern and southern gradients respectively. The gradients are calculated at each longitude for latitudes (ϕ_0) between 35° and 75°N:

$$\text{GHGN} = \frac{Z(\phi_N) - Z(\phi_0)}{\phi_N - \phi_0}, \text{GHGS} = \frac{Z(\phi_0) - Z(\phi_S)}{\phi_0 - \phi_S}, \quad (1)$$

where $\phi_S = \phi_0 - 15°$ and $\phi_N = \phi_0 + 15°$. A latitude, longitude grid point is then defined as being blocked if GHGS > 0 and GHGN < -10 ($m/°$). A schematic showing an example $Z500$ field that satisfies these criteria is shown in Fig. 1 of Martínez-Alvarado et al. (2018). The index is extended here to identify blocks north of 75°N. For latitudes north of 75°N, GHGN is calculated using the remaining latitudes to ϕ_N at the pole and the threshold on GHGN to identify a block is scaled relative to the distance to the pole; GHGS is calculated in the normal way.

3.2 | Anticyclone tracking

Anticyclones are tracked using the objective feature tracking algorithm, TRACK (Hodges, 1994, 1995, 1999), which has been used extensively to track tropical (e.g. Hodges and Emerton, 2015) and extratropical cyclones (e.g. Hoskins and Hodges, 2002, 2005, 2019), polar lows (Zappa et al., 2014) and Tibetan plateau vortices (Curio et al., 2018) using relative vorticity to identify these features.

In this work, TRACK is used to identify anticyclones corresponding to positive $Z500$ anomalies with respect to the instantaneous zonal component. Small scales are removed by spectral filtering, lowering the original resolution of the data to T42 resolution. Once the maxima in the Z500 anomaly field are identified, a track is constructed by finding nearest neighbours in consecutive time steps (Hodges, 1994, 1999). Anticyclones are tracked in the forecasts initiated at 1200 UTC 27 September 2016 in the control simulation, model physics experiments, and the analysis. The location of the anticyclones in the analysis corresponds well with the blocking centre.

3.3 | Trajectory calculation

WCBs are identified in this article using a Lagrangian trajectory method. Air parcel trajectories are calculated with the Lagrangian Analysis Tool (LAGRANTO, Wernli and Davies, 1997; Sprenger and Wernli, 2015). Forward trajectories are started at every horizontal grid point over the North Atlantic (in the region of an extratropical cyclone) and vertically every 20 hPa from 1010 hPa to 790 hPa for various times throughout the cyclone evolution. The trajectories are calculated forward using the output from the MetUM experiments using the 3D wind field, temperature, specific humidity and surface pressure. Air parcel trajectories that ascend by more than a chosen ascent threshold in a given time period are defined as part of the WCB and used to compare WCB features among the experiments. The ascent thresholds are varied between 500 and 600 hPa in either 48 or 72 hour forecast periods to provide comparisons of WCB ascent rates in different forecast periods. Previous studies (e.g. Wernli and Davies, 1997; Madonna et al., 2014; Grams and Archambault, 2016) have used similar criteria to identify WCBs from ascending air trajectories, typically 600 hPa ascent in 48 hours for the most rapidly ascending trajectories within a WCB. We include additional ascent criteria here, namely 500 hPa in 48 hours and 600 hPa in 72 hours, for a broader comparison of WCBs in the forecast evolutions.

3.4 | Block forecast verification

Forecasts of atmospheric blocking are verified using a version of the Structure Amplitude Location (SAL) diagnostic introduced by Wernli et al. (2008) for verification of precipitation forecasts and adapted here for block forecasts. Block forecasts are compared against the analysis fields, and experiment forecasts against the control forecast, using amplitude (A) and location (L) values calculated from the output of the blocking index. The A and L components are calculated over all grid points (i,j) within the Euro-Atlantic domain in exactly the same way as used in Wernli et al. (2008), except with the grid point values of precipitation (their $R_{i,j}$) replaced by grid point values of an area-scaled blocking index, $B_{i,j}$. The blocking index $B_{i,j}$ is set equal to either $\cos(\phi)$, if the grid point at latitude ϕ is blocked, or zero, if it is not blocked. Note that we do not set a minimum threshold for the blocking index whereas a minimum precipitation amount threshold is applied in Wernli et al. (2008). Following Wernli et al. (2008) A is the normalized difference of the domain-averaged $B_{i,j}$ values between two sets of outputs: either output from an experiment simulation and the analysis or output from an experiment and the control run. Similarly, L is the sum of two components measuring (i) the normalised distance between the centres of mass of the two sets of $B_{i,j}$ values and (ii) the difference for the two runs between the weighted averaged distance between the centre of mass of the $B_{i,j}$ values in the individual blocked regions and the centre of mass of the total $B_{i,j}$ field. We do not compute the structure (S) component of the SAL diagnostic in Wernli et al. (2008). The S component takes into account the intensity of the field, looking for differences between the forecast and observations in both the intensity and coverage of precipitation. Because the blocking index values used here for block forecasts lie between zero and unity (a much smaller range than for precipitation values) the S component does not provide any additional useful information compared to that obtained from the A and L components.

3.5 | Potential temperature tracers

To study how the parameterisations of diabatic processes are affecting the flow and contributing to model error a set of θ-tracers has been incorporated in the MetUM. The method is similar to that described in Martínez-Alvarado and Plant (2014) and uses the temperature increments for several physical processes that are converted into potential temperature (θ) increments. The tracer method utilises the conservation property of θ by writing it in terms of a conserved component (θ_0) and a non-conserved component ($\Delta\theta$). θ_0 is advected with the flow, but not altered by the parameterisation schemes, and gives the isentropic origin of air masses at the start of the forecast. $\Delta\theta$ is advected as well but also accumulates the effect of the parameterisation schemes throughout the integration. θ at a grid point x and time t can be written as

$$\theta(x,t) = \theta_0(x,t) + \Delta\theta(x,t) + \epsilon_\theta(x,t), \tag{2}$$

where $\epsilon_\theta(x,t)$ is an error term originating from numerical diffusion modifying θ. The non-conserved term can be separated into terms for each parameterised process (i)

$$\Delta\theta(x,t) = \sum_i \Delta\theta_i(x,t), \tag{3}$$

where $\Delta\theta_i(x,t)$ represents the cumulative contribution to θ due to parameterisation i. The parameterisations contributing to θ that are considered in this study are:

- $\Delta\theta_{slow-phys}$: θ contribution from the slow physics scheme. This includes both the short and long wave radiation increments as well as the microphysics increments;
- $\Delta\theta_{BL}$: θ contribution from the boundary layer parameterisation;
- $\Delta\theta_{conv}$: θ contribution from the convection parameterisation;
- $\Delta\theta_{cloud-rebal}$: θ contribution from the cloud rebalancing scheme. This scheme calculates condensate and cloud fraction changes due to the change in temperature in a grid box that occurs due to condensation from adiabatically cooling ascending air (see Wilson et al. (2008b) for more details).

The θ-tracers are calculated online within the model simulation code from the transport equation and the error ϵ_θ is computed as

$$\epsilon_\theta = \theta - \theta_0 - \Delta\theta. \tag{4}$$

θ_0 and each $\Delta\theta_i$ are output as model diagnostics. The total non-conserved part of θ (neglecting the tracer error) can then be calculated as the difference between θ and θ_0 (or $\sum_i \Delta\theta_i$) and illustrates where diabatic processes have heated or cooled air masses during the model integration. More details about the θ tracer calculation can be found in Martínez-Alvarado and Plant (2014).

The θ-tracer diagnostic can be used to quantify how important diabatic heating is in the development of blocking in the forecast evolution and assess how strongly air masses that arrive in blocks have been heated in the preceding days. An example of the θ-tracer diagnostic is shown in Fig. 1. The θ, θ_0, $\Delta\theta$ and ϵ terms are shown at a forecast lead time of seven days and at model level 34 (8634 m) from a model run initiated at 1200 UTC on 27 September 2016. At this time there is a large-scale blocking ridge present over Northern Europe. Generally, we expect regions of diabatic heating to be located in ridges as air parcels arriving in them have ascended from lower levels and have experienced diabatic heating (Pfahl et al., 2015). The conserved part of θ, that has just been advected by the flow (Fig. 1(b)), shows the value of θ that each of the air masses had at the start of the forecast. Comparing this field to the full field highlights where diabatic or frictional processes have modified θ and the scales at which these processes are active. The non-conserved component of the θ field shows regions that have been heated or cooled during the integration and can be used to trace air masses. For example, the region of strong heating to the north of the UK depicts an air mass that has ascended and undergone latent heating, most likely in the WCB of a cyclone, before arriving at the model level shown. Air in the cut-off region over central Europe has experienced descent and cooling. The θ-tracer error term is shown in Fig. 1(d). The error is generally small across the entire domain and considerably smaller than the non-conserved θ shown in Fig. 1c (note that the colour bars in panels c and d are different). The error at even seven days into the forecast evolution is generally $<$10% of the non-conserved θ and hence we can neglect error in our analysis and consider the term $\theta - \theta_0$ as the change in θ due to diabatic processes (termed the total diabatic heating here).

4 | NAWDEX CAMPAIGN PERIOD

In this section, results are presented from the control forecast together with the SST-update and Prog-ent experiments for forecasts from the whole of the NAWDEX period. Forecasts from the operational ensemble are also included to compare the magnitude of the change in skill due to initial condition uncertainty and model uncertainty. The skill of the forecasts is assessed using the Root Mean Square Error (RMSE) of the mean sea level pressure (MSLP) and Z500. The variability in forecast skill is also related to weather patterns and diabatic heating, and the representation of blocking in forecasts during the NAWDEX period is assessed.

4.1 | Overview of forecast performance

The average RMSE of Z500 and MSLP over the Euro-Atlantic region (the domain shown in Fig. 1) is shown for forecasts during NAWDEX from the control forecast, SST-update and Prog-ent experiments and the operational ensemble mean forecast as a function of forecast lead time in Fig. 2. At early lead times, the average RMSEs in all the forecasts are almost identical for both MSLP and Z500. For forecasts longer than five days, the ensemble mean out-performs the control forecast and both model experiments, as expected (e.g. Toth and Kalnay, 1997). The RMSE is lower in the Prog-ent experiment than in the control forecasts for forecasts between five and ten days lead time for both fields. The RMSEs remain nearly indistinguishable between the control forecast and SST-update experiment until the longest lead times. It is clear from Figure 2 that running a forecast model with SST updating daily did not improve or degrade the forecast skill on average for this period. This suggests that running a coupled atmosphere-ocean model for medium-range forecasts may not be beneficial for this period and region, though having SST changing on only daily timescales may limit the potential impact. There was a slight improvement in skill with the Prog-ent experiment with a

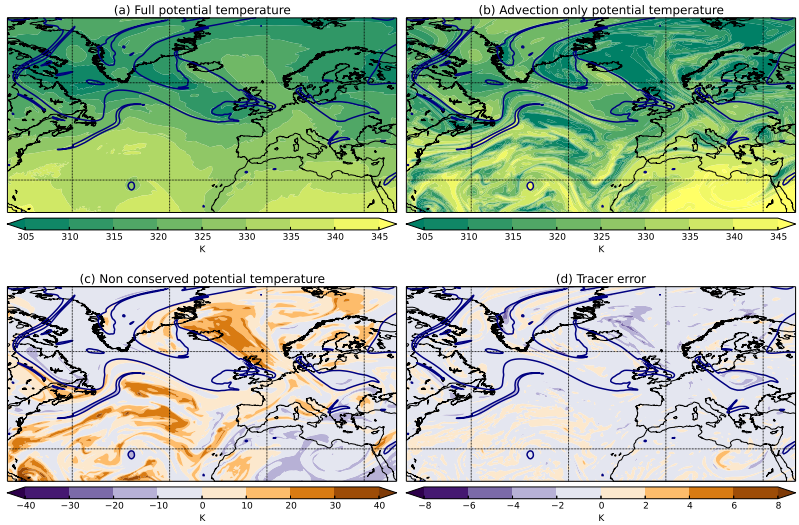

FIGURE 1 (a) Full field (θ), (b) advected initial field (θ_0), (c) the non-conserved θ ($\theta - \theta_0$) and (d) the tracer error ($\theta - (\theta_0 + \sum_i \Delta\theta_i)$) from the θ-tracer output on model level 34 (8634 m) in the forecast of 12 UTC 4 October 2016 at a lead time of 7 days. Note that the colour bar in (d) is one fifth of the scale in (c). The 2 PVU contour is plotted in each case to show Rossby-wave structure.

maximum reduction in RMSE of around 4% in both fields at nine days lead time. The reduction in RMSE is relevant in the operational forecast context, with similar reductions in RMSE found with the full GA7.0 implementation (Walters et al., 2019).

Forecast skill can also vary depending on the weather pattern (Ferranti et al., 2015; Matsueda and Palmer, 2018) so it can be insightful to consider how the RMSE values vary throughout the NAWDEX period. We now only consider RMSE of Z500 as this can be used to describe the general flow pattern, or weather regime, and relate this to periods of high and low forecast skill. The RMSE of Z500 for forecasts valid on dates during the NAWDEX period is shown in Fig. 3(a) together with a measure of blocking activity (the area blocked in the analysis). Both measures are calculated over the Euro-Atlantic domain defined above. The RMSE on the given valid date is averaged for forecasts of that date for all lead times between six and ten days inclusive.

Periods of low forecast skill (high RMSE) during the NAWDEX period are associated with the onset and decay of blocked periods. The average RMSE increases quickly before the onset of a blocked period around 12 UTC on 4 October 2016, suggesting that the onset of this block was poorly forecast in the control and model experiments. The RMSE of Z500 then decreases dramatically once the block is mature and remains low during the maintenance period of this block (from around 6–12 October 2016). The RMSEs in all forecasts increases again before the block decays (after 13 October 2016), highlighting the difficulty in forecasting the decay, as well as the onset, of blocked periods. The variation in RMSE throughout the NAWDEX period is similar in the control and SST-update experiment, consistent with their almost identical average Z500 RMSEs. There is a larger difference between the RMSE in the control forecast and the Prog-ent experiment: the RMSE in the Prog-ent experiment is generally lowest, except for forecasts for the period of block onset at the start of the NAWDEX period.

We now compare the variability of forecast skill with the variability of diabatic heating in air masses reaching the

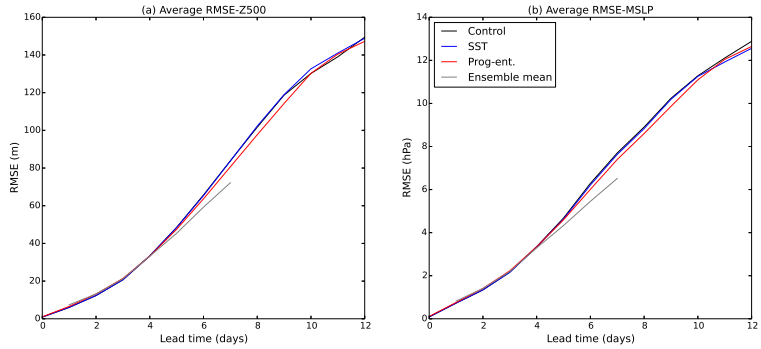

FIGURE 2 Average RMSE as a function of lead time for the control forecast, SST-update experiment, Prog-ent experiment and the mean of the operational ensemble for all of the forecasts initiated during the NAWDEX period. The RMSE of Z500 (a) and MSLP (b) are shown throughout the operational and experiment forecast integrations (7 and 12 days, respectively). Note that the black and blue lines are nearly indistinguishable until the longest lead times and the grey line of the operational ensemble extends only to 7 days lead time.

upper-troposphere using the θ-tracers. The fraction of the ridges that have undergone heating (FH) is calculated as

$$\mathrm{FH} = \frac{\sum_{i,j \in \hat{R}} cos\phi}{\sum_{i,j \in R} cos\phi}, \tag{5}$$

where R are ridges (points on the 315 K isentropic surface north of the model grid row with the latitude of 51.33°N and with a PV value less than 2 PVU), $\hat{R}$ represents the regions within R with total diabatic heating ($\theta - \theta_0$) greater than 2 K (the same value used in Pfahl et al. (2015)) and ϕ is the latitude of grid point i, j. Note that the results presented in this section are not sensitive to the arbitrary choice of latitude used to define ridges. The mean fraction of ridges heated in all forecasts between six and ten days lead time inclusive for forecasts valid on days during the NAWDEX period is shown in Fig. 3(b), as well as the area blocked in the analysis and the average area blocked in the control forecasts of the same lead times. The onset of the first blocked period in the control forecast is preceded by an increase in the fraction of the ridges heated. This is a result of air masses being heated and arriving at upper-levels as the ridge is amplifying before the block develops fully and is identified by the blocking index. The increase in the fraction of the ridges heated coincides with the increase in RMSE prior to block development. This supports the hypothesis that the uncertainty in block forecast (diagnosed as RMSE in Z500) is originating from lower levels and associated with diabatic heating and its representation by the parameterisation schemes. The uncertainty in the block forecast could also be due to uncertainty in the strength and location of the WCB due to initial condition uncertainty. The FH then remains relatively constant during the maintenance phase of the blocked period. There is an increase in the FH as this block decays, this increase is associated with heating in an amplifying ridge that develops into the blocked area towards the end of the NAWDEX period (not shown).

Low forecast skill during the NAWDEX period was related to an increase in diabatic heating (rather than a large net fraction of the ridge heated). This is partly a property of the θ-tracer calculation as it is the accumulated heating that is measured and a large heated fraction of a ridge does not necessarily mean that diabatic processes were active recently. However, an increase in the heated ridge fraction does imply recent diabatic heating which is uncertain in the model and potentially causing the low forecast skill. The increase in RMSE Z500 and total diabatic heating generally precedes block onset reflecting the model's known difficulty in transitioning to a blocked state (Matsueda, 2009; Martínez-Alvarado et al., 2018). The representation of blocking during the NAWDEX period in the control forecasts, model experiments and operational ensemble is now reviewed.

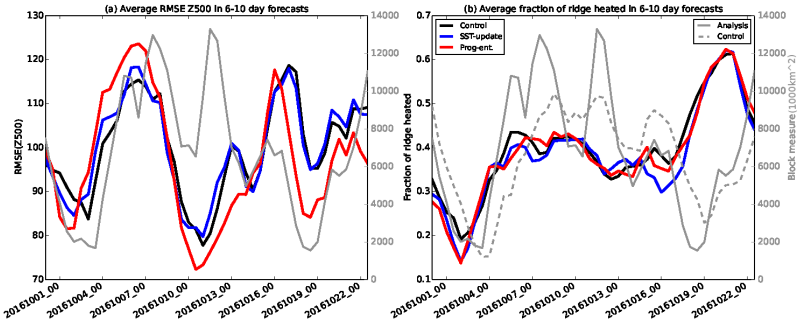

FIGURE 3 (a) RMSE of Z500 (m) in the Euro-Atlantic region for forecasts valid on days during the NAWDEX period, averaged for all forecast lead times between six and ten days lead time inclusive. The area blocked over the Euro-Atlantic domain during the NAWDEX period in the analysis is also shown (grey line). (b) The fraction of ridges that have been heated ($\theta - \theta_0 > 2$) for forecasts valid on days during NAWDEX period, averaged in forecasts of that date between six and ten days lead time. Again the area blocked in the analysis is given by the grey line and additionally the area blocked in the control forecast (averaged between six and ten days lead time) is given by the grey dashed line. Note that the right hand axis corresponds to the grey lines in both panels.

4.2 | Predictability of blocking during NAWDEX

Block representation in the forecasts is now compared to that in the analysis during the NAWDEX period. The average area difference in block area between the analysis and each of the forecasts and the operational ensemble forecast is shown together with the inter-quartile range in block area difference among each individual forecasts from the control and model experiments in Fig. 4(a). All the forecasts (both operational and experiments) exhibit the longstanding model deficiency of a reduction in block area with lead time (Tibaldi and Molteni, 1990; Matsueda, 2009; Martínez-Alvarado et al., 2018), though the reduction is not large for forecasts during this period especially compared to the spread in the individual forecasts. The difference in block area in all experiments and in the ensemble mean is small until after five days lead time, consistent with other studies (e.g Tibaldi and Molteni, 1990; Pelly and Hoskins, 2003a; Matsueda, 2009). As the lead time increases, the average difference then becomes increasingly negative (the blocks are smaller in the forecasts) with the error generally similar in the control forecast, SST-update and Prog-ent experiments and smaller in the average of the operational ensemble. There is a large spread in block area difference between the forecasts and analysis for specific forecast initiation dates (thin dashed lines).

Block forecasts are now verified using the modified version of the SAL technique (Wernli et al., 2008) described in section 3.4 to provide additional information about the error in both amplitude and location of the forecast blocks. A and L values are calculated for each of the operational ensemble members for forecasts of seven days lead time, and in the control simulation, SST-update and Prog-ent experiments at seven and ten days lead time, for every forecast initiated during the NAWDEX period. These lead times were chosen as they span the lead times at which block onset forecast typically deteriorates. The means (points) and the standard errors (lines) in A and L are plotted for values calculated against the analysis (Fig. 4(b)). Recall that a perfect forecast will have A and L values equal to zero. The mean A values are close to zero for all the simulations, reflecting the small average area difference shown in Fig. 4(a). The mean of the operational ensemble forecast has the lowest absolute A value, implying the smallest forecast error. The mean L values in the control, model experiments and operational ensemble lie between 0.2 and 0.3 which suggests some systematic error in the location of blocking events. The A and L values can also be used to contrast the SST-update and Prog-ent experiments against the control forecasts (Fig. 4(c)) to better quantify how much impact the experiments are having on the simulation of blocking. In general, the sensitivity experiments are more similar to the control forecast than all the forecasts are to the analysis (the points are closer to the origin). The Prog-ent experiment has more of an effect on block forecasts than the SST-update, consistent with the larger effect seen in RMSE MSLP

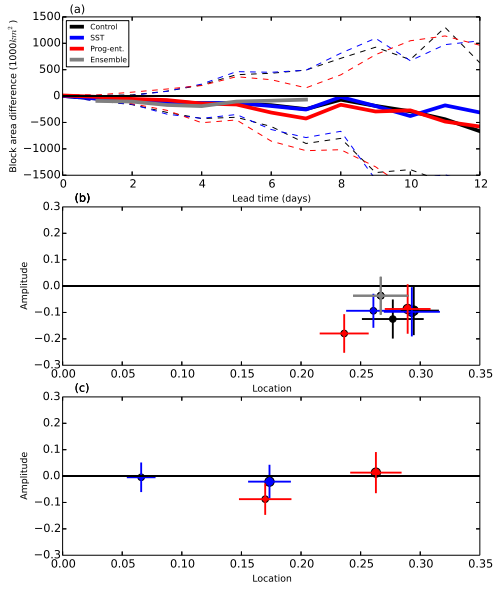

FIGURE 4 (a) The average difference in block area between the analysis and forecasts from the control, SST-update and Prog-ent experiments, and operational ensemble as a function of forecast lead time. The difference is shown in the mean (thick lines) and the interquartile range (thin dashed lines) of the forecasts for the control, model experiments and the operational ensemble mean. (b) *AL* diagram for forecasts of blocking during the NAWDEX period compared to the analysis in the control, model experiment and operational ensemble forecasts. Mean (points) and standard errors (lines) of *A* and *L* values are shown for forecasts of 7 (small dots) and 10 (big dots) days lead time. (c) As in (b) for forecasts from the model experiments compared to the control forecasts.

and Z500 values (Figs. 2 and 3). The difference in amplitude of the blocks in the simulations is much smaller than the difference in their locations. This is true for their means and standard errors, suggesting verifying block forecasts based only on their amplitudes might hide some of the forecast differences.

In this section, it has been shown that updating SST daily into the forecast integration to their observed values did not systematically improve or degrade the forecasts during NAWDEX. Running the forecast model with a new convection scheme had a larger impact and generated a slight reduction in average forecast error. Both the SST-update and Prog-ent experiments had different representations of blocking events than in the control simulation. In the next section, we will show in a case study that the SST-update and Prog-ent experiments, together with the GA7 parameterisation experiments, can nevertheless have a large impact on the evolution of individual forecasts and make use of the θ-tracers to explain the evolution changes observed.

5 | BLOCK CASE STUDY: FORECASTS VERSUS ANALYSIS

The case study is introduced first in this section after which the operational forecast and the full set of model experiments (GA7 parameterisation experiments as well as the SST-update and Prog-ent experiments) are compared against the analysis. The case study is focused on the block onset around 12 UTC on 4 October 2018 over Scandinavia following a cyclone known as the Stalactite cyclone (Schäfler et al., 2018). The forecast initiation time (1200 UTC on 27 September 2016, a week prior to block onset) is chosen to be far enough in advance of the block onset so that the

experiments have diverged considerably. This is a lead time at which models are known to poorly predict the onset of a blocked flow (e.g. Pelly and Hoskins, 2003a; Matsueda, 2011; Martínez-Alvarado et al., 2018).

5.1 | Synoptic overview

In Fig. 5, the synoptic situation is presented for the days preceding block onset. The analysed MSLP and Z500 fields as well as the 2-PVU contour on the 315 K isentropic surface (Tp315, representing the tropopause) and the location of the Stalactite cyclone are shown between 1–4 October 2016. On 1 October 2016, the Stalactite cyclone is identified in the central North Atlantic with an upper-tropospheric trough above it and to the west. An amplifying ridge is also visible in both the Tp315 and Z500 contours downstream of the cyclone, with a more zonal flow further downstream and upstream (Fig. 5(a)). On 2 October, the Stalactite cyclone has intensified and moved north-eastward toward Iceland and the ridge in the tropopause has amplified. An upper-level PV streamer is formed to the south of the Stalactite cyclone as the cyclone interacts with the upper-level trough and the streamer cyclonically wraps up around the cyclone centre. Z500 has risen over the UK associated with the incipient block and warm air advection in the Stalactite cyclone. By 3 October, the ridge has expanded dramatically and now extends over a large part of the North Atlantic, Z500 has risen further over UK and towards Scandinavia as the transition to a blocked state continues. The Stalactite cyclone has now reached maximum intensity and moved nearer to Iceland with the upper-level trough completely wrapped around the cyclone centre. At this time there is a sign of a secondary cyclone developing as a kink in the MSLP contours in the Gulf Stream region off the coast of North America. On 4 October, the analysed block onset date over Scandinavia, the block is clear in Z500 (almost a cut off high region) and Tp315 (a large ridge over Scandinavia). The Stalactite cyclone has dissipated by this time and a secondary cyclone has intensified leading to a secondary ridge ahead of it. This synoptic evolution provides a clear example of regime transition to a blocked state following extratropical cyclogenesis similar to those presented in previous studies (e.g. Colucci, 1985; McLeod and Mote, 2015; Grams et al., 2018). The flow evolution in this case is now assessed in the forecasts from the chosen initiation date.

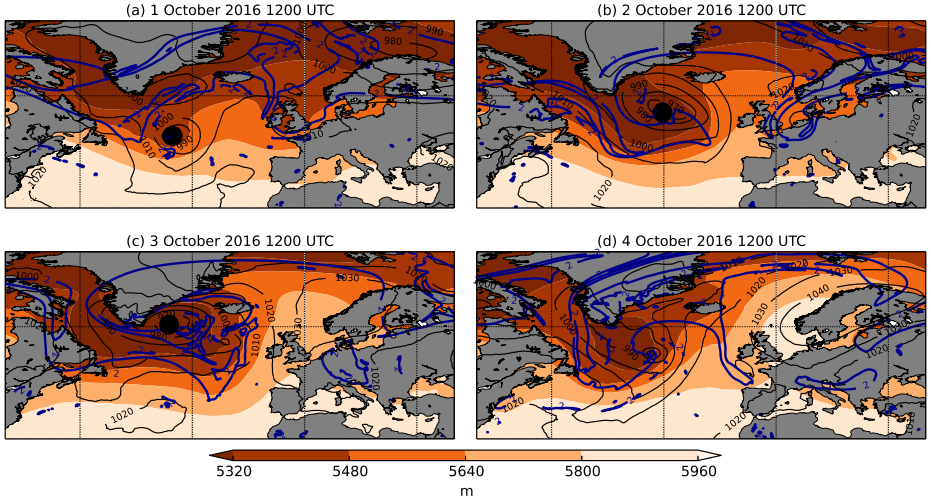

FIGURE 5 Synoptic overview for days preceding block initiation. MSLP (contours), Z500 (filled contours) and the Tp315 contour (thick blue contour) from the Met Office analyses are shown for (a) 1, (b) 2, (c) 3 and (d) 4 October 2016. The location of the Stalactite cyclone when present is shown by the black circle.

5.2 | Forecast verification

The track of the Stalactite cyclone and its minimum MSLP as seen in the analysis, operational ensemble and model physics experiments are shown in Fig. 6(a),(c). The tracks and intensities are shown between 12 UTC 30 September 2016 and 12 UTC 4 October 2016, which corresponds to lead times between three and six days for the forecasts initiated on 12 UTC 27 September 2016. The tracks and intensities (diagnosed by the associated positive Z500 anomalies) of the downstream block are also shown in Fig. 6 for the control forecast and model experiments.

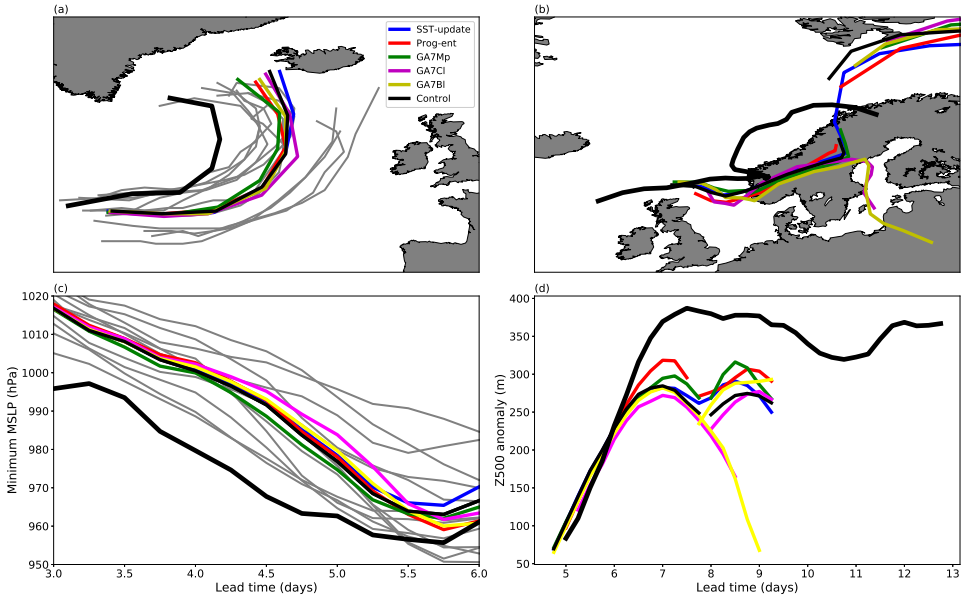

FIGURE 6 Stalactite cyclone and downstream block tracks and intensities for the forecast initiated at 1200 UTC 27 September 2016 from the control simulation (thin black), model physics experiments (colours), operational ensemble (grey, for the cyclone only) and the analysis (thick black): (a) cyclone tracks, (b) anticyclone tracks, (c) cyclone minimum MSLP, and (d) Z500 anomaly.

All of the forecasts (control, model experiments and operational ensemble) have a clear eastward shift in the track of the cyclone compared to the analysis when it is present in the forecast between three and six days lead time (Fig. 6(a)). None of the model experiments result in a large change in the cyclone track forecast when compared to the control forecast in this case (they are spread about the control (thin black line)). All of the model experiments and the control forecast also forecast a less intense Stalactite cyclone than analysed (Fig. 6(c)), particularly at earlier lead times. The rate of MSLP drop and minimum MSLP reached is quite similar in the experiments to that analysed but starts from an incorrect value (at three days lead time). Again, the control forecast of minimum MSLP is in the middle of forecasts in the ensemble of experiments. There is larger spread among the operational ensemble than the model experiments (initial condition uncertainty develops quicker than model physics uncertainty in this case) in both the track and intensity of the cyclone, although the analysis still generally remains outside the ensemble spread in both track and intensity of the Stalactite cyclone. The cyclone in the analysis was further west than any ensemble member, and had a deeper MSLP than any member, at nearly all lead times. As none of the operational ensemble or model experiments were able to capture the development of the Stalactite cyclone it could be that there exists a deficiency somewhere in the model, e.g. in the dynamical core or set of physics parameterisations, which meant the model could never predict this evolution. Alternatively, the PDF of possible initial conditions may not have been sampled sufficiently and the ensemble may not have enough members.

Similar errors in the track and intensity are identified for the block that develops downstream of the Stalactite

cyclone. The tracks of the block centre are located further east in the forecasts (thin lines) than the analysis (thick line), the same error as in the track of the Stalactite cyclone. The intensity (maximum Z500 anomaly) of the block is less in the forecasts than the analysis, the same error as in the intensity of the Stalactite cyclone. Whilst this does not prove a causal relationship between the development of the Stalactite cyclone and downstream block it is highly suggestive of one and motivates the analysis in the subsequent sections of this article. The tracks of the blocking anticyclone are not continuous in all of the forecasts because the anomaly in Z500 did not always exceed the threshold for block detection, though these gaps are generally less than 12 hours long.

The misrepresentation of the track and intensity of the Stalactite cyclone and downstream block in the forecasts is consistent with errors in the forecast of the upper-tropospheric flow pattern. This can be summarised by looking at the Tp315 contour and the location of the Stalactite cyclone in the analysis and model experiments, as shown in Fig. 7, for the same dates as those in Fig. 5 to allow a direct comparison. On 1 October 2016, four days into the forecast evolution, the developing ridge is generally well represented in all of the experiments except for the upstream trough to the south of Greenland which is shifted too far east. This is consistent with the position of the Stalactite cyclone being too far east. Over the next 48 hours, the amplification both northward and westward of the ridge is underestimated in all of the experiments. By 3 October 2016, six days into the forecast evolution, the ridge in the analysis extends across Greenland which is not reproduced in any of the forecasts. The ridge position in the forecasts being further east than the analysis is also consistent with the eastward shift in the block tracks in the forecast. The position of the Stalactite cyclone remains too far east at this time. On 4 October, seven days into the forecast, the Stalactite cyclone has weakened and there is a large-scale blocking ridge present in all of the forecasts. The location and shape of the upper-level ridge in the control and model experiments does not match the analysis. The analysed blocking ridge has formed over Scandinavia and the secondary ridge has amplified over the North Atlantic. In the forecasts, the primary ridge is extending too far over Greenland (resembling the analysed ridge on the previous day). The developing ridge ahead of the secondary cyclone is completely missed in all the experiments. In this case, the forecasts do predict the onset of a blocking ridge following the Stalactite cyclone but are not able to predict the exact timing and location of the features accurately.

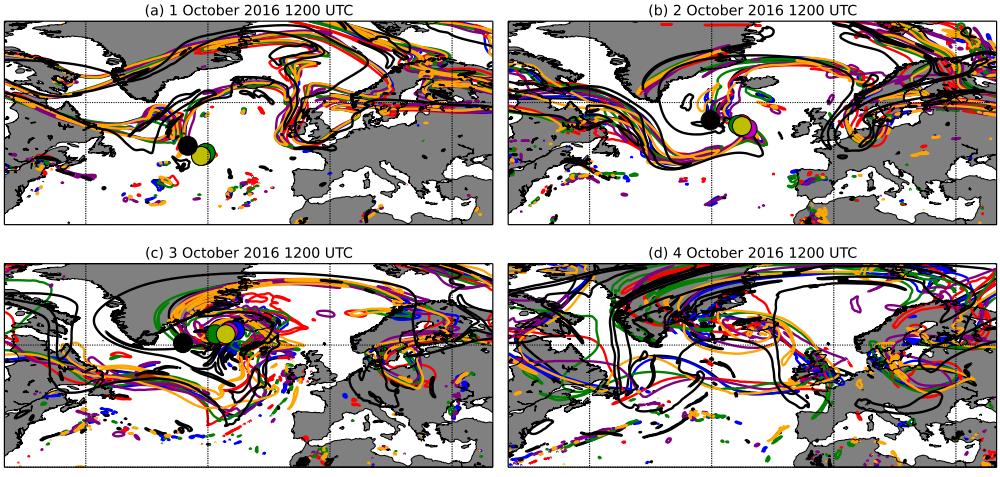

FIGURE 7 Tropopause location (Tp315 contour) in the analysis (thick black line) and model experiment forecasts (coloured lines) initialised at 1200 UTC 27 September 2016 and valid on the same date for (a) 1, (b) 2, (c) 3 and (d) 4 October 2016. The location of the Stalactite cyclone in the analysis is marked by the black circle and in the model experiments by the coloured circles (colours as defined in Fig. 6.

The poor forecast of the upper-level ridge and associated block can be analysed from a different perspective using

the blocking index. The area identified as blocked within the Euro-Atlantic region in the forecasts as a function of lead time, and on the corresponding date in the analysis, is shown in Fig. 8(a). The control forecast and model experiments do a reasonable job of capturing the analysed blocked area in the region. All forecasts exhibit an increase in block area around 160 hours lead time at roughly the same rate as that in the analysis. The spread in the operational ensemble (grey shading) is larger than the spread among the physics ensemble and the operational ensemble shows a tendency to underestimate the area blocked. The control forecast and SST-update experiment remain similar, with their forecast of block area nearly identical. There is greater spread among the parameterisation experiments, particularly at later lead times, with the analysis generally lying among the forecasts. The clear error in the forecast of the Stalactite cyclone and upper-tropospheric ridge highlighted previously is not evident in the forecast of block area over the Euro-Atlantic. A large-scale blocking ridge is clearly evident in all of the forecasts (Figs.7(c),(d)) and the blocking index reveals the block areas forecast are similar to those in the analysis. The error in the location of the block apparent in Fig. 7 is now quantified using the modified SAL diagnostic described in section 3.4.

A and L values calculated against the analysis are presented in Fig. 8(b). Values are calculated for forecasts from the operational ensemble, control and model experiment forecasts. A and L values are calculated at forecast lead times of seven (small dots) and eight days (big dots) lead time for the control and physics experiments and at seven days lead time for the operational ensemble forecasts. These lead times were chosen because they are the times of block development in the forecast (recall the operational ensemble only extends to seven days). In the control forecast and model experiments, A values are generally low at both forecast lead times implying that the forecasts did a reasonable job in predicting the total area blocked in the region. However, A values of less than -0.4 are found for some of the physics experiments at eight days lead time, representing an underestimation by more than a factor of 1.5 in the area blocked. L values are also non-zero, highlighting the error in the blocked location in this forecast. Whilst the GA7 parameterisation experiments resulted in a smaller area blocked than in the control forecast (increased error), they have smaller L values than the control forecast (particularly in the GA7Mp experiment) at both lead times and hence reduced error in the predicted location of the block. L values are typically larger at seven days lead time, consistent with the error in upper-level ridge location at this lead time (Fig 7(d)). A values in the operational ensemble members are typically more negative and span a greater range of values than those in the physics experiments. A and L values are also calculated in each of the model experiments using the control forecast as a reference (Fig. 8(c)). If two forecasts' A and L values are similar when compared to the analysis but different when compared to the control (e.g. the SST-update and Prog-ent experiments at eight days lead time) it suggests they were similar to the analysis in different ways. For example, this could occur if two forecasts predicted blocks equally far from the block in the analysis but in opposite directions. Not all of the model experiments are closer to the control forecast than to the analysis at this lead time.

This case study of block onset, following the development of the Stalactite cyclone, was very unpredictable for forecasts beyond four days lead time. This is in line with the findings from Schäfler et al. (2018), who showed that this period is associated with reduced forecast skill. The location and intensity of the Stalactite cyclone appear to be important for the block development in this case. This association has been confirmed using ensemble sensitivity analysis (using ECMWF operational ensemble forecasts in Maddison et al. (2019)). In that paper it was hypothesised that the sensitivity of block development to upstream cyclones originated from diabatic heating in WCBs. This hypothesis is investigated in the next section by comparing the model physics experiments to the control run. Although all these forecasts diverge from reality, insight can still be gained into how diabatic processes in the Stalactite cyclone affect the large-scale development, and in particular the block. An advantage of restricting our analysis to the model forecasts is that the θ diagnostic can be used to assess the diabatic influence on dynamical evolution; this diagnostic is not available for the analysis.

6 | BLOCK CASE STUDY: IMPACT OF MODEL UPGRADES

The block onset case study is now compared in each of the model experiments to the control (operational) forecast to quantify how much model uncertainty can affect the dynamical evolution. The θ tracers are used to determine the role that diabatic heating in the WCB of the Stalactite cyclone had on the upper-level Rossby wave pattern and block development. It was shown in Fig. 6 that there was some spread in the intensity ($\approx$ 10 hPa) and location ($\approx 5^\circ$) of the

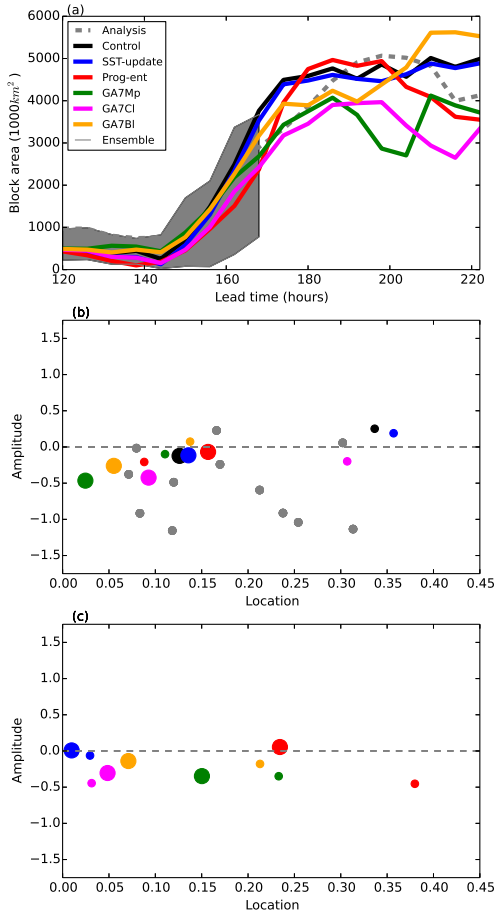

FIGURE 8 (a) Block area as a function of lead time in forecast initiated at 1200 UTC on 27 September 2016 in the control simulation and the model physics experiments and block area in the analysis on the corresponding day. (b) Diagram showing the *A* and *L* components of the modified SAL diagnostic for blocks in the operational ensemble at seven days lead time (grey dots) and each of the experiments at 7 (small coloured dots) and 8 (large) days lead time calculated against the analysis. (c) as in (b) but calculated against the control simulation and for the experiments only.

Stalactite cyclone among the model experiments. But how much did the model experiments change the WCB of the Stalactite cyclone? WCB trajectories have been computed in the control forecast and each of the model experiments in the forecast initialised at 12 UTC on 27 September 2016 and are shown as an example in Fig. 9 for the control forecast only. The trajectories shown start at three days into the forecast evolution, finish at six days into the forecast and have ascended more than 500 hPa in that period. The end points of the trajectories are at 12 UTC on 3 October 2016 (cf. Fig. 7.(c)) when the blocking ridge is amplifying in the control forecast.

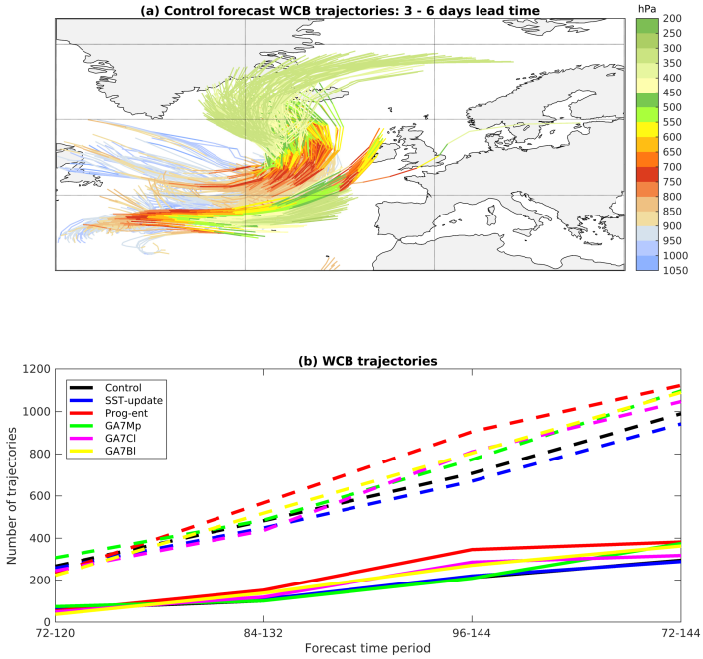

FIGURE 9 (a) WCB trajectories (having ascended more than 500 hPa) in the control forecast initiated on 1200 UTC 27 September 2016 between three and six days lead time. (b) The total number of identified WCB trajectories (see text) during different forecast periods in the control forecast and model experiments for ascent thresholds of 500 hPa (dashed lines) and 600 hPa (solid lines).

A large air mass ascends from the central northern Atlantic in the WCB of the Stalactite cyclone into the upper-level ridge, linking the Stalactite cyclone and the downstream block development. To quantify the intensity of the WCB in each of the experiments the number of trajectories classified as part of the WCB are compared for different time periods during the forecast run and for different ascent thresholds. The comparison is presented in Fig. 9(b). The Prog-ent experiment has the highest number of WCB trajectories for the majority of the different selection time periods and ascent thresholds; the control forecast and SST-update experiment often have the least. The Prog-ent and SST-update experiments were not outliers in the experiments when considering the track and intensity of the cyclone (Fig. 6). This suggests that, whilst neither the track nor the intensity of the cyclone changed considerably in the physics experiments, the WCB intensity differed because of the changes to the latent heating from the parameterised diabatic processes that were modified in the physics experiments. We expect the convection scheme to be active in the WCB of a deep extratropical cyclone so this may explain why we see the largest change in the Prog-ent experiment.

The experiments modifying parameterisations of physical processes also had different developments of the upper-level Rossby wave pattern and block. The difference in PV on the 315 K isentropic surface (PV315) between each of the model experiments and control forecast (experiment minus control) is shown in Fig. 10, together with the Tp315 contour in both the control (solid black contour) and experiment (dashed black contour). The forecasts are shown at seven days lead time (cf. Fig 7(d)), which is when the block first formed over Scandinavia and was identified in the blocking index. There are clear differences in the amplitude and phase of the Rossby-wave pattern among the experiments. The main differences are in the northern extent of the ridge, the structure of the PV streamer forming on the upstream trough, and the phase and amplitude of the smaller ridge to the south of Greenland. The northern extent

of the ridge and cyclonic wrap up of the upstream trough are coupled to the development of the Stalactite cyclone and the divergent outflow in its WCB. The Prog-ent and GA7Cl experiments have a ridge that extends further to the north (extending out of the domain shown), consistent with the increased number of WCB trajectories compared to the control in these experiments, and different structure of the PV streamer. The GA7Bl experiment has a larger ridge extent in the north eastern corner of the domain. The PV difference between the experiments and control forecasts is much smaller than the difference compared to the analysis (not shown). We now use the θ-tracers to investigate if the difference in PV is originating from changes to the parameterised diabatic heating in the WCB of the cyclone.

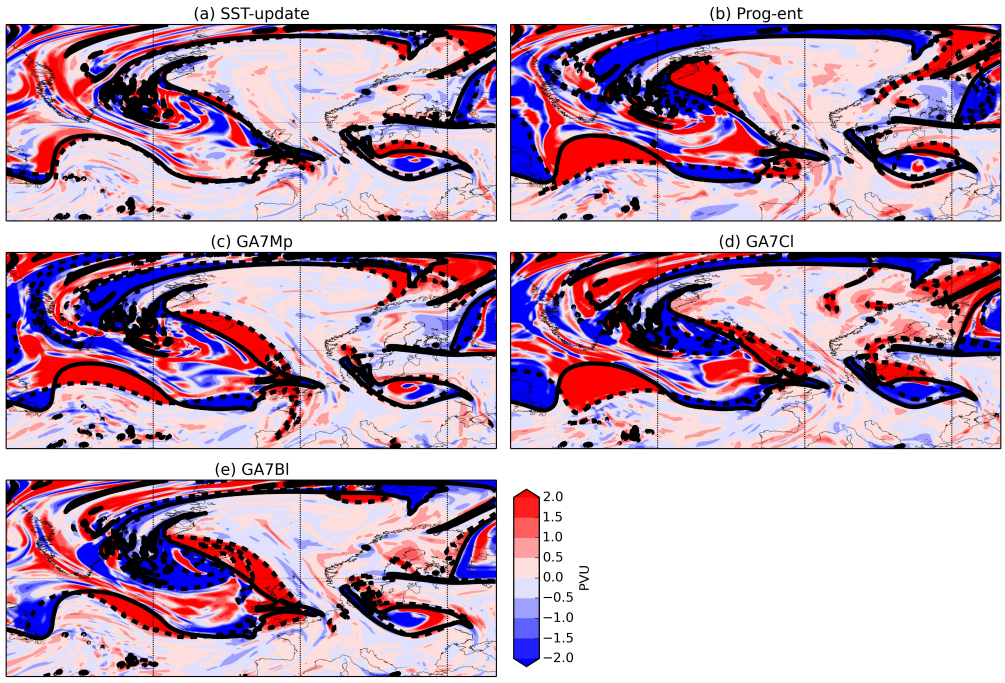

FIGURE 10 Difference (experiment - control) in PV315 in the (a) SST-update, (b) Prog-ent, (c) GA7Mp, (d) GA7Cl and (e) GA7Bl experiments. The Tp315 contour in the experiments (dashed contour) and control forecast (black contour) are plotted also. The difference is shown at seven days into the forecast initiated at 1200 UTC on 27 September 2016.

6.1 | Role of heating

The difference in total diabatic heating (as defined in section 3.5) on the 315 K surface between each experiment and the control simulation is plotted at seven days lead time in Fig. 11 so the results can be directly compared to the PV315 differences shown in Fig. 10. The heating represents the non-conserved part of θ and indicates where diabatic processes are active and air masses have undergone ascent or descent. There are large differences in the total diabatic heating of air parcels on the 315 K isentropic surface during the first seven days of forecast integration. The structure of the difference in PV315 between the control and each experiment clearly resembles that of that difference in total diabatic heating (which can exceed 20 K). Given that the total diabatic heating in the control forecast is generally less than 30 K (Fig. 1), these differences are considerable. There is generally more heating within the large-scale ridge for each of the experiments when compared to the control. In particular, the GA7Cl and Prog-ent experiments exhibit

more heating across the ridge interior and as well as in the region where the ridge is extending further north. This suggests that the ridge is more amplified because of the increased ascent of air masses resulting in the air masses reaching higher levels and the ridge becoming larger. The regions of largest difference in heating between the control simulation and the model physics ensemble are in the location of the WCB ascent and outflow as shown for the control experiment in Fig. 9(a). This suggests that changes to the heating in the WCB of the cyclone arising from the changes to the parameterised diabatic processes in our experiments are causing the different amplifications of the ridge. Note that here we are using large values of $\theta - \theta_0$ (as defined in (2)) as a proxy for WCB air masses here (i.e. they have experienced strong ascent and heating). For example, in the Prog-ent experiment, $\theta - \theta_0$ is increased meaning there is greater latent heat release in the WCB which results in its outflow reaching higher isentropic levels and the ridge being larger than in the control simulation. The region of negative difference in heating on the western flank of the ridge that is present in nearly all of the experiments is likely a result of a shift in the WCB ascent region in the experiments.

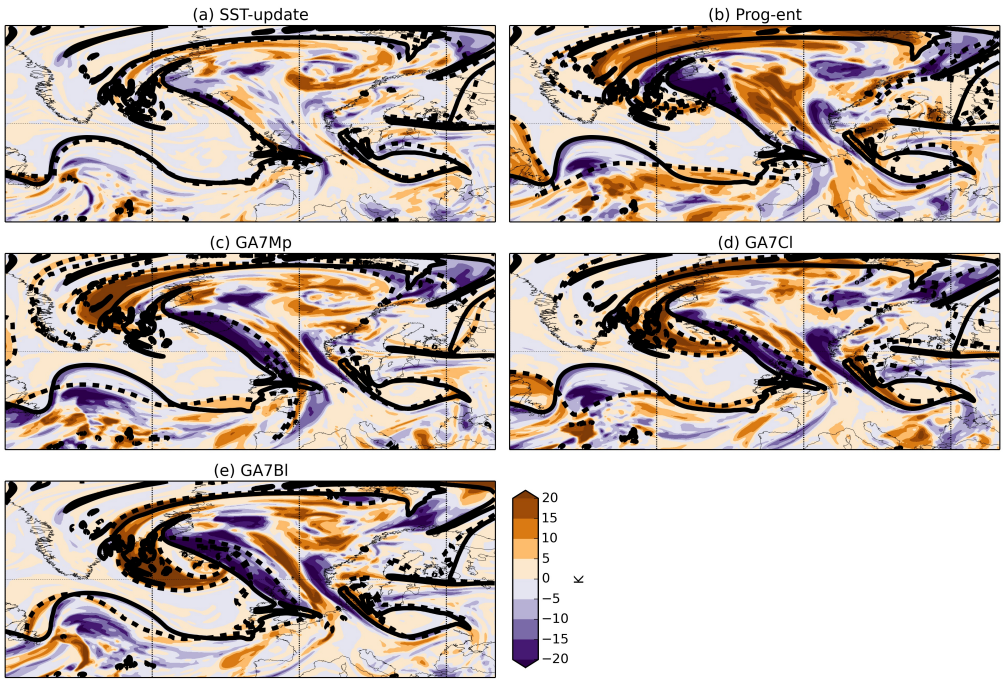

FIGURE 11 Difference (experiment - control) in total diabatic heating at 315 K in the (a) SST-update, (b) Prog-ent, (c) GA7Mp, (d) GA7Cl and (e) GA7Bl experiments. The Tp315 contour for the experiments (dashed line) and control forecast (solid) are plotted also. The fields plotted are from the forecast initiated at 1200 UTC on 27 September 2016 at a lead time of seven days.

The WCB intensities in the control simulation and experiments are consistent with the difference in heating in the ridges observed in Fig. 11: experiments with more WCB trajectories have more heating in the ridge when compared to the control and also a ridge that extends further north. To quantify this relationship the area of the ridges in the experiments was calculated as the area of regions north of 51.33°N with PVU values on the 315 K isentropic surface less than 2 PVU (not shown). The difference between each of the model experiments and the control forecast in the number of WCB trajectories, diabatic heating at 315 K, and ridge area appear to be highly correlated (correlations greater than 0.7), though the small sample size does not allow for a robust statement regarding these correlations. It is clear however that modifying the model parameterisations has an effect on WCB intensity and experiments with

stronger WCBs had increased total diabatic heating and larger ridges.

6.2 | Vertical structure of heating

Vertical cross sections of the total diabatic heating between 40-80°N in each of the simulations are shown in Fig. 12, averaged between 60°W and 40°E (the domain shown in Fig. 11), with the averaged 2 PVU contour overlain for tropopause identification. The region used to average the heating captures the WCB region and its outflow into the upper-level ridge. The blocking ridge is evident in all simulations to the north of 55°N as a slight increase in average tropopause height with latitude.

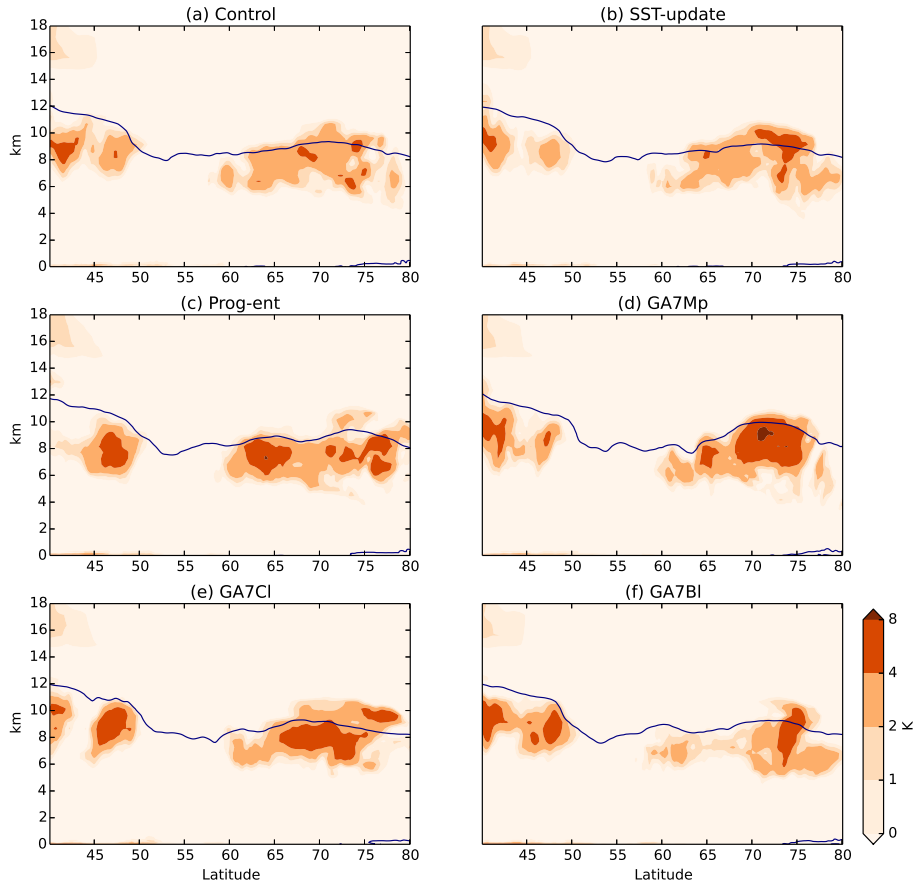

FIGURE 12 Cross section of total diabatic heating (colours) and the 2 PVU contour (lines) between 40 and 80° and averaged between 60°W and 40°E in the (a) control forecast and the (b) SST-update, (c) Prog-ent, (d) GA7Mp, (e) GA7Cl and (f) GA7Bl experiments at 1200 UTC 4 October 2016 (168 hours lead time). We only plot positive total diabatic heating for clarity though there are many regions with total diabatic cooling.

The general vertical structure of heating in the control and each model experiment is similar. There is a positive total diabatic heating throughout most of the upper-troposphere in the blocking ridge in all the simulations. The strongest regions of diabatic heating are generally near the tropopause (or above the average location of the tropopause) north of 60°N and extending towards the pole. This shows that air masses have been heated (in the WCBs of the cyclone) in the simulations prior to their arrival at upper levels in the blocking ridge. The average heating throughout this cross section is strongest in the Prog-ent, GA7Cl and GA7Mp experiments. This is consistent with these experiments having more heating throughout the ridge on the 315 K isentropic surface and more intense WCBs than in the control (Figs 11 and 9). We would expect the convection, cloud and microphysics parameterisation schemes to be active in the cloudy, ascending air of the WCB and changing these schemes to result in different diabatic heating rates. Figures 11 and 12 show only snapshots of the heating structure in the simulations for one lead time and for particular isentropic surfaces or longitude bands. To quantify how the total heating in the WCB during the forecast evolution is changing in the experiments, and to verify that this is driving the ridge amplification differences, an integrated measure of the diabatic heating is now calculated.

6.3 | Integrated heating

The θ-tracer technique accumulates the diabatic heating along Lagrangian-trajectories throughout the forecast integration. The $\theta - \theta_0$ output at t days lead time therefore represents the total heating that air parcels arriving at each grid point have undergone throughout the t days of the simulation. To determine how much diabatic heating has occurred during the WCB ascent we can integrate the heating term in the region of WCB outflow at the time it reaches the upper-troposphere. The θ-tracer diagnostic then gives us the heating the WCB trajectories have experienced since the start of the forecast. This calculation assumes the integration only covers the WCB period, i.e. there was not strong heating in the early part of the forecast, which is true for this case (the WCB started around two days into the forecast). Furthermore, the forecast evolution in the control forecast and model experiments is very similar for the first two days (prior to WCB development) meaning differences in total heating in the first seven days of forecast integration are due to differences in the diabatic heating associated with the WCB ascent. It has been seen using trajectory analysis that the WCB outflow is within the ridge and hence we can integrate the diabatic heating over the ridge at various isentropic levels to quantify the total heating of air parcels arriving in the upper-level ridge.

We calculate the integrated heating measure (IHM), which represents the total heating in grid boxes within the WCB outflow (grid boxes that have strong ascent) normalised by the ridge area at a certain lead time, as

$$IHM(t) = \frac{\sum_{i,j \in W \cap R} cos(\phi)(\theta - \theta_0)_{i,j}}{(\sum_{i,j \in R} cos(\phi)) \times t}, \tag{6}$$

where W is the WCB outflow region ($\theta - \theta_0 > 10\ K$), R is the ridge ($PVU < 2$ and north of 51.33°N within the domain shown in Figure 11) and ϕ is the latitude of grid point i, j. Note the conclusions presented in this section are robust to the choice of latitude used to define a ridge. This gives IHM units of K/day. The IHM is calculated on isentropic surfaces between 310 and 340 K to measure the strength of heating of the air parcels in the WCB outflow within the ridge reaching (passing through) the different levels.

The IHM is shown in Fig. 13 at a forecast lead time of seven days. The total diabatic heating, and level at which the heating is maximum, is different in each of the experiments. The Prog-ent and GA7Cl experiments generally have the strongest heating in the WCB air parcels when looking at the different θ levels; the GA7Bl experiment and the control forecast generally have the least. The control simulation has the least integrated heating among the experiments on the 315 K surface, consistent with Fig. 11. The θ surface for which the integrated heating is strongest depends on experiment, though is between 315 and 330 K, with the experiments with larger total diabatic heating in the WCB having the maximum heating at higher levels, illustrating the link between WCB diabatic heating and ascent. The large range of θ surfaces at which diabatic heating in the WCBs is maximum reflects the uncertainty in WCB dynamics and the level their outflows reach. The control simulation and GA7Bl experiment have no outflow reaching 330 K; the SST-update, GA7Cl and GA7Mp have no outflow reaching 335 K. The Prog-ent experiment has outflow reaching the highest isentropic level among the experiments. Averaging the IHM for the different WCB outflow levels yields an

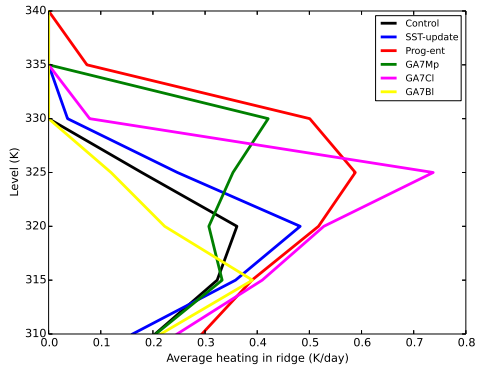

FIGURE 13 IHM as a function of isentropic level in the control forecast and model experiments. The total accumulated heating is averaged over the seven days of forecast evolution.

estimation of the total heating of WCB air parcels reaching the upper troposphere in the ridge. Averaged values were smallest for the GA7Bl experiment and control forecast (0.135 K/day and 0.152 K/day, respectively) and largest in the Prog-ent and GA7Cl experiments (0.338 K/day and 0.286 K/day, respectively). The vertically-averaged heating rates are also shown in the first column of Fig. 14.

The IHM calculated provides consistent results with the analysis previously presented and the hypothesis introduced in the introduction of this article. The Prog-ent and GA7Cl experiments have larger ridges, forced by increased heating across the ridge at 315 K, stronger heating meridionally throughout the troposphere across the WCB region and stronger heating throughout the WCB outflow reaching most isentropic levels. The results demonstrate that the diabatic heating in the WCB trajectories is driving the amplification of the upper-level ridge. The GA7Bl experiment for example had a large number of WCB trajectories, but the diabatic heating was not strong, its WCB did not reach high altitudes and its ridge was not amplified. The Prog-ent and GA7Cl experiments had increased diabatic heating in their WCB developments, the outflow from their WCBs reached higher isentropic levels and their ridges were larger. This demonstrates that by changing the parameterisations of convection and cloud, the diabatic heating in the WCB ascent was increased which resulted in its outflow reaching higher levels and a more amplified upper-level ridge developing in the forecasts. The picture is not as clear when looking at the block metrics: whilst the Prog-ent experiment had a similar area blocked to the control forecast (slightly larger at certain lead times) the GA7Cl experiment had a smaller area blocked. The block index only considers geopotential height at 500 hPa and may be less sensitive to changing ridge amplitudes at higher levels.

6.4 | Dominant processes

The total diabatic heating for air parcels arriving in the block as part of the WCB outflow can be separated into the diabatic heating from each parameterisation. By doing this, it is possible to quantify how much diabatic heating is added by each process and which parameterisation scheme is contributing most to the total diabatic heating. The IHM can be computed for each parameterisation by replacing the $(\theta - \theta_0)$ term in (6) with the $\Delta\theta_i$ terms listed in section 3.5.

The IHM from each parameterisation considered is shown averaged for isentropic surfaces between 310 and 340 K and at seven days lead time in Fig. 14. The total diabatic heating $(\theta - \theta_0)$ and θ-tracer error $(\theta - (\theta_0 + \sum_i \Delta\theta_i))$ are also shown. In the control forecast and each of the experiments, the convection parameterisation contributes most to the total diabatic heating along the WCB trajectories, followed by the cloud rebalancing scheme. The accumulated tendencies for each of the parameterisation schemes are larger than the tracer error in all experiments, despite the long forecast lead time. The amount of heating from the different parameterised processes in each experiment scales

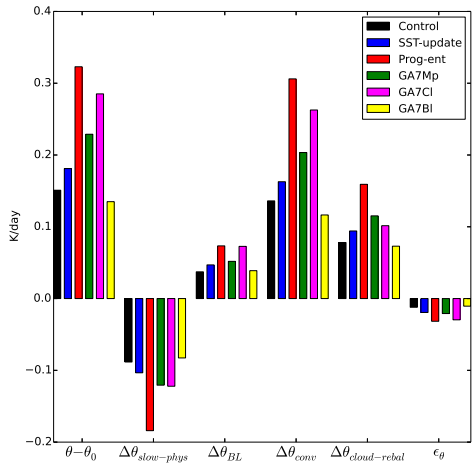

FIGURE 14 Contributions to the total IHM from the different physical parameterisations and for the different experiments at seven days lead time.

approximately linearly with total heating. For example, the Prog-ent experiment has the highest magnitude of total heating rate as well as heating from each parameterisation scheme, rather than a disproportionately large increase in heating coming from the convection scheme. This implies that the changes to the individual parameterisation schemes in the experiments are changing the evolution of the cyclone and its WCB and not just the heating increments from the modified scheme.

7 | CONCLUSIONS

We have quantified the effects of operational changes to parameterisation schemes that impact diabatic processes, and those of running a forecast with updating SST, on midlatitude forecast evolution and compared this effect to the forecast evolution change from different initial conditions. We have also used the forecast evolution changes identified when modifying the parameterisations to show that upper-level ridge amplification depends on parameterised diabatic heating rates in the WCB of an extratropical cyclone. The forecast experiments were performed for the time period of the NAWDEX field campaign (Schäfler et al., 2018) because this included a variety of interesting weather events including a case study of block onset downstream of extratropical cyclogenesis that was the main focus of this study. The development of atmospheric blocks has been related to upstream cyclones since some of the earliest studies of blocking (e.g. Berggren et al., 1949; Colucci, 1985; Lupo and Smith, 1995). More recently, forecasts of atmospheric blocking and upper-level ridges have been shown to depend on the forecast of upstream cyclones (Maddison et al., 2019), the representation of their WCBs (Grams et al., 2018), and their parameterised diabatic processes (Joos and Forbes, 2016; Martínez-Alvarado et al., 2016). Here it is demonstrated that this sensitivity can arise from diabatic heating from parameterised processes.

In forecasts initialised every 12 hours during the NAWDEX campaign period, updating the SST at daily intervals during the forecast evolution did not have a systematic effect on the forecast skill. SST is important for the evolution of the overlying atmosphere because of the influence surface fluxes have on the stability of the marine boundary layer and near surface winds (Park et al., 2006; Donlon et al., 2012). Coupling the atmospheric model to an ocean model is known to improve prediction on subseasonal to seasonal timescales (Palmer and Anderson, 1994; Goddard et al., 2001; de Andrade et al., 2019). The small impact of evolving SST shown here suggests that running a coupled model

may be less important in medium-range forecasts, where initial condition uncertainty usually dominates (Rabier et al., 1996; Kalnay, 2003), though giving the forecasts analysed SST and the relatively small sample of forecasts used here are caveats. Forecasts from the model with an improved convection scheme (one including memory about recent convective activity) led to an average reduction in forecast error. Periods of low forecast skill during the NAWDEX campaign period were associated with the onset and decay of atmospheric blocking events. The poor prediction of the transition to and from a blocked state has been a persistent problem in NWP (e.g. Tibaldi and Molteni, 1990; Pelly and Hoskins, 2003b; Ferranti et al., 2015; Lillo and Parsons, 2017). Periods of low forecast skill also tended to be associated with an increase in the amount of diabatic heating in upper-level ridges, consistent with previous studies showing that parameterised diabatic processes can be important for forecast skill (Rodwell et al., 2013; Martínez-Alvarado et al., 2016). The forecast error for atmospheric blocking events during the NAWDEX period was characterised by an underestimation of block area. This was the case in forecasts from both the model experiments performed for this article and the operational ensemble forecasts. The location of the predicted blocking events was also assessed using a modified version of the *SAL* technique introduced in Wernli et al. (2008). The location of blocks in the analysis were systematically misforecast with the mean values far from zero even when taking into account the standard error.

The impact of the model upgrades on forecast evolution was considered in detail for a particular forecast initiation date preceding the intensification of an extratropical cyclone and the downstream development of a block. The cyclone development, between three and six days lead time, and downstream block development, after six days lead time, were missed in the control forecast, model experiments and operational ensemble. Forecasts of this event were identified as those having some of the least skill during the NAWDEX period in Schäfler et al. (2018). The cyclone in all of the forecasts was located too far east throughout its lifecycle. The downstream block error was forecast dependent, either misplaced or too small (or both) in the control and model experiment forecasts and in the operational ensemble forecasts. The development of blocking in this case was very unpredictable: forecast runs with different initial conditions, updated SST, and improved parameterisations of diabatic processes were not able to capture the evolution of the flow that really happened. In the forecast of the cyclone and downstream block, there was more spread among the operational ensemble than the model experiments, particularly at early lead times. Initial condition uncertainty normally is larger for medium-range forecasts than model uncertainty (Lorenz, 1963; Rabier et al., 1996), though the small model experiment ensemble size and the method of modifying the model here were not intended to represent fully model uncertainty. The model experiments did however diverge from the control forecast considerably by six or seven days lead time.

A set of θ-tracers implemented in the MetUM was used to attribute the different upper-level forecast evolutions to the different diabatic heating rates from parameterised processes. The implemented new versions of the parameterisation schemes in the MetUM resulted in differences in the phase and amplitude of the upper-level blocking ridge. The different amplitudes of the upper-level ridge in the experiments resulted from differences in the total diabatic heating of air parcels arriving in the ridge in the WCB of an extratropical cyclone. Forecasts that produced larger-amplitude ridges had stronger diabatic heating throughout the WCB development. Changing the parameterisation of convection was shown to have the biggest increase in number of WCB trajectories, the diabatic heating of air masses in the upper-level ridge and downstream block forecast. The model experiments changed the cyclone location and intensity by approximately 5° (zonally) and 10 hPa, respectively, but the diabatic heating changed by more than 20 K. We hypothesise that the small changes in the cyclone location and intensity resulted from the changes to the parameterised diabatic processes which then resulted in large changes in the total diabatic heating in the ascending WCB. WCB ascent is sensitive to diabatic heating (Joos and Wernli, 2012) and errors have shown to amplify rapidly on the arrival of the WCB at upper levels (Davies and Didone, 2013; Martínez-Alvarado et al., 2016; Grams et al., 2018). WCB ascent has been shown to include sporadic periods of intense, convective like ascent (e.g. Browning, 1971; Oertel et al., 2019) and may explain the strongest dependence of the WCB to the convection and cloud parameterisations found here. For this case study we have demonstrated that modifying parameterised diabatic processes changed the properties of the WCB, the diabatic heating of air parcels in the WCB ascent, and the upper-level ridge amplification and block forecast. Hence, model error attributable to the parameterisation of diabatic processes contributes to the documented role of extratropical cyclones and WCBs in the forecast uncertainty of upper-level ridges and blocking events (Grams et al., 2018; Maddison et al., 2019) and the large forecast errors associated with some blocking events (Rodwell et al., 2013; Ferranti et al., 2015). While this relationship was demonstrated by Joos and Forbes (2016) by implementing a new microphysical parameterisation in the IFS model and assessing the heating rate from each mi-

crophysical process with consistent results to those found here, this is the first time (to the authors' knowledge) that this relationship has been quantified and compared for several parameterisation modifications and using all contributions to the total diabatic heating (such that the heating budget is closed). This study thus provides strong evidence that improving the representation of diabatic processes could reduce the frequency of poor forecasts associated with blocking events. Future work should investigate whether the relationship between parameterised diabatic heating in WCBs and upper-level ridges in the case study included here is systematic across case studies and contributing to poor forecast skill.

acknowledgements

We thank Dr Claudio Sanchez for the implementation of the θ tracers into the MetUM and assistance running the forecast experiments and Dr Kevin Hodges for running the anticyclone tracking code. J.W.M.'s contribution was funded through a NERC Industrial CASE studentship in collaboration with the Met Office (NE/M009610/1). O.M-A.'s contribution was funded by the United Kingdom's Natural Environment Research Council (NERC) as part of the National Centre for Atmospheric Sciences. K.D.W. was supported by the joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101). The source code for the MetUM is available to use. To apply for a license for the MetUM go to http://www.metoffice.gov.uk/research/collaboration/um-collaboration. For more information on the exact model versions and branches applied please contact the authors. Data from the simulations is archived at the Met Office and available for research use through the Centre for Environmental Data Analysis JASMIN platform (http://www.jasmin.ac.uk/); please contact the authors for details.

A | APPENDIX

The forecast experiments performed for this article are described in detail in this appendix.

A.1 | The MetUM

The MetUM GA6.1 uses the ENDGame dynamical core to solve the non-hydrostatic, fully compressible deep-atmosphere equations of motion with a semi-implicit semi-Lagrangian formulation (Wood et al., 2014). The prognostic fields are the three-dimensional wind components, virtual dry potential temperature, Exner pressure and dry density. Moist prognostic fields (e.g. mass mixing ratio of water vapour and prognostic cloud fields) are advected as free tracers. Prognostic fields are discretised horizontally onto a regular latitude-longitude grid with Arakawa C-grid staggering (Arakawa and Lamb, 1977). A Charney-Phillips staggering is used in the vertical (Charney and Phillips, 1953) with terrain-following hybrid height coordinates. ENDGame uses a nested iterative approach at each atmospheric time step with processes split into an outer loop and an inner loop (see Wood et al. (2014) for details). Processes that are parameterised in the MetUM include solar and terrestrial radiation, large-scale precipitation, large-scale cloud, sub-grid orographic drag, non-orographic gravity wave drag, the boundary layer, convection, and atmospheric aerosols and chemistry. The MetUM is coupled to the Joint UK Land Environment Simulator (JULES) land-surface model (Clark et al., 2011). For a full description of the MetUM GA6.1 the reader is referred to Walters et al. (2017).

A.2 | Sensitivity experiments

A.2.1 | Prognostic entrainment

In the Prog-ent experiment, the convection scheme in GA6.1 is changed to a scheme that includes prognostic entrainment rates, designed to give the convection scheme memory about recent convective activity. The change to the Prog-ent scheme also includes the change to the 6A convection scheme that is part of the GA7 convection scheme (Walters et al., 2019). The numerics of the convection scheme were revised for the 6A scheme. A convection parameterisation represents sub-grid scale cumulus clouds within a grid box and their associated transport of heat, moisture and momentum. The MetUM GA6.1 uses a mass flux convection scheme that is an extension of Gregory and Rown-

tree (1990). The scheme consists of three steps: (i) initial convective diagnosis of whether convection is possible given boundary layer properties; (ii) a call to either the shallow or deep convection schemes at all points diagnosed as shallow or deep from step (i); and (iii) a call to the mid-level convection scheme at all grid points.

In the GA6.1 scheme fully developed, deep convective clouds can appear within a single timestep without any gradual development, which is not usually realistic. This can occur because the entrainment rate given to a cloud diagnosed as a deep convective cloud is one that is appropriate for deep convection that is fully developed, as there is no option in the scheme for using higher entrainment rates appropriate for developing cumulus. In the real world, regions with a small amount of convective activity would be likely to have relatively small convective clouds with relatively high entrainment rates (if convection was present) and regions with lots of recent convective activity the opposite. A scheme has been developed at the Met Office that accounts for this difference by adding an additional 3D model prognostic ($\bar{P}$) that is a measure of recent convective activity (described in Willett and Whitall (2017)). It is defined as

$$\frac{D\bar{P}}{Dt} = \frac{1}{\tau}(\tilde{p}_{surf}^{conv} - \bar{P}), \tag{7}$$

where

$$\tilde{p}_{surf}^{conv}(x, y, z) = C(x, y, z)\max[p_{min}, p_{surf}^{conv}(x, y)] \tag{8}$$

is a 3D extension of the 2D surface convective precipitation rate (p_{surf}^{conv}). This equation gives $\bar{P}$ the same units as precipitation rate. $C(x, y, z)$ is defined as equal to unity at a given point if convection is active and zero if not, p_{min} is an arbitrarily small precipitation rate (set to 10^{-5} kg m^{-2} s^{-1}) allowing non-precipitating clouds to contribute to $\bar{P}$, and τ is the e-folding time (set to three hours) that defines the memory timescale. $\bar{P}$ is related to the entrainment rate such that grid points that have had relatively little amounts of recent convection have higher entrainment rates and vice versa. To do this, the standard entrainment rate is scaled by the 3D factor

$$F = C_{grad}\log_{10}\left(\bar{P}\frac{q_s^{ref}}{q_s^{LCL}}\right) + C_{int}, \tag{9}$$

where $C_{grad} = -1.1$ is a constant controlling the strength of the coupling between the entrainment and $\bar{P}$, $C_{int} = -2.9$ is an intercept, q_s^{LCL} is the specific humidity of an undilute parcel-ascent to the lifting condensation level (LCL) calculated in the convective diagnostics, which is normalised by a reference value $q_s^{ref} = 20$ g kg^{-1}. Estimates of the range of tropical precipitation rates and associated entrainment rates (from observations and model data) were used to derive the values of C_{grad} and C_{int}. The scaling factor F is limited to the range 0.5–2.5 which requires the scaling to have a logarithmic dependence on $\bar{P}$ because precipitation rates can vary over several orders of magnitude. The dependence on q_s^{LCL} reflects the strong control temperature has on precipitation rates. The Prog-ent scheme has been tested for several cases over tropical regions for a variety of tropical processes (Willett and Whitall, 2017), and its performance assessed in terms of mean climate and average NWP scores across the globe. The impact that the Prog-ent scheme has in simulations of extratropical cyclones is investigated here.

A.2.2 | Evolving SST

The SST-update experiment is designed to mimic a coupled NWP model that has evolving SST. The SST of the atmospheric model is updated at daily intervals in the forecast evolution. The sea-ice fraction is also updated each day as these fields both contribute to the surface fluxes we are trying to evolve. The SST are updated daily at 12 UTC using the OSTIA analysis (Donlon et al., 2012) produced by the Met Office for the corresponding date in the forecast evolution. The OSTIA analysis is a global SST and sea-ice fraction field at a resolution of $1/20^{\circ}$ produced using several

different sources of both satellite data and in-situ measurements of SST and sea-ice fraction. The product is described fully in Donlon et al. (2012). Operational forecasts produced at the Met Office are initialised with the OSTIA analysis of that date and the fields are not updated during the forecast integrations.

A.3 | GA7 parameterisations

The parameterisations that are modified to their GA7 versions in the experiments are now described: the microphysics, large-scale cloud and boundary layer parameterisation schemes. These processes are likely to have an impact on cyclone and WCB development and hence will likely affect PV modification near the tropopause and have an impact on the downstream forecast flow pattern. The convection scheme is also likely to have an impact here, but we will use the Prog-ent convection scheme described above to elucidate the effects that a new convection scheme can have. Details of each of the schemes being modified are first given, followed by the specific GA7 changes that are made to the scheme from the GA6.1 control run (a full description of the MetUM GA6.0/6.1 configuration can be found in Walters et al. (2017) and of GA7.0/7.1 in Walters et al. (2019)). The GA ticket numbers (#GA) used in the GA7.0/7.1 development and documentation are included here for reference.

A.3.1 | Microphysics

The microphysics (also known as the large-scale precipitation) scheme is responsible for the formation and evolution of precipitation due to grid scale processes, including phase changes between vapour, liquid water and ice and their vertical advection. The GA6.1 (control) scheme is an extensively modified version of the Wilson and Ballard (1999) scheme. The individual changes made to the microphysics scheme that constitute the GA7 upgrade are summarised below.

- **New sub-grid scale cloud water content variability treatment (#GA15)**: the standard deviation of cloud water content in a grid box (divided by its mean value) is now used to represent the sub-grid scale water content variability effect on radiation in the microphysics scheme. This has been changed from a globally constant value to a more realistic one dependent on cloud fraction, vertical layer thickness and whether the cloud is convective or not.
- **Change to warm rain microphysics (#GA52)**: within the microphysics scheme the warm rain microphysics part has been nearly totally rewritten for GA7. The autoconversion and accretion parameterisations are changed and the evaporation and sedimentation code have been improved. The new scheme also explicitly represents how sub-grid variability affects microphysical process rates. Sub-grid rain fraction is now consistently set by either the fraction of autoconverting cloud or melting snow when rain is created.
- **Turbulent production of liquid water in mixed-phase clouds (#GA120)**: the parameterisation of liquid water production in mixed-phase clouds is changed to improve the production and maintenance of super-cooled liquid and mixed-phase clouds, which is a known problem across many models (Illingworth et al., 2007). A probability distribution of supersaturation is calculated and then the liquid-cloud properties are calculated as moments of this distribution. A lack of super-cooled liquid in cold clouds has been shown to contribute to model biases in, for example, surface radiative fluxes (Curry et al., 2000) and liquid water path (Klein et al., 2009).

A.3.2 | Large-scale cloud

Clouds form on scales smaller than the grid scale before the grid box average humidity reaches saturation. The large-scale cloud scheme is required to determine how much of the grid box is covered by cloud and how much condensed water is contained in those clouds. Within a grid box, the cloud scheme calculates the amount and phase of condensation at each time step and calculates or updates the cloud fractions. The cloud scheme can create latent heat release when it converts water vapour into liquid or ice. The calculated cloud cover and ice water contents are passed to the radiation and microphysics schemes to calculate the radiative impact of the clouds and determine if any precipitation has formed. The prognostic cloud and prognostic condensate (PC2) scheme is used (Wilson et al., 2008b,a) in both

GA6 and GA7. Modifications to the GA6.1 scheme for the upgrade to GA7 are made as follows:

- **Representation of the radiative impact of convective cores (#GA44)**: for some convective cloud types, e.g. shallow fair-weather cumulus, detrainment into the environment may be small but the radiative impact of the convective core considerable and needs to be represented. The impacts of these clouds are included using a convective cloud model that includes the radiative effects of the convective cores. Previously it was assumed a convective plume rises and mixes with the environment and it is only once condensate has detrained from the plume that it will have a radiative effect.
- **New critical relative humidity value that is based on turbulence (#GA89)**: the critical relative humidity used to determine the initiation of cloud in grid boxes that are cloud free, but with increasing (in time) relative humidity, is changed from a constant global value for each model level to a value based on turbulence. This new value is also used for the removal of cloud from grid boxes that are full of cloud when relative humidity is decreasing. The value is calculated using the resolved vertical gradients in temperature and humidity as well as the sub-grid mixing length, eddy diffusivity and turbulent kinetic energy calculated from the boundary layer scheme.
- **Removal of unnecessary complexity when dealing with ice cloud (#GA98)**: the value of the cirrus spreading rate has been reduced from $1.0 \times 10^{-3} s^{-1}$ to $1.0 \times 10^{-5} s^{-1}$. The cirrus spreading rate acts to increase the frozen cloud fraction and was introduced into an earlier configuration of the MetUM to counteract an unrealistic reduction in mean ice cloud fraction.

A.3.3 | Boundary layer

The atmospheric boundary layer scheme parameterises vertical turbulent transports of heat, moisture and horizontal momentum. These turbulent motions are not resolved in global NWP models but are important for producing realistic vertical structure in wind and thermodynamic profiles. The scheme primarily handles the lowest layers in the MetUM, but does include a free-tropospheric component that can extend to model levels approaching the tropopause. The boundary layer scheme is based on Lock et al. (2000), with the modifications of Lock (2001) and Brown et al. (2008). Upgrades from the GA6 boundary layer parameterisation for GA7 are described below.

- **A boundary layer entrainment dependence on decoupling (#GA13)**: surface driven turbulence entrainment at cloud top is restricted when the cloud layer is decoupled from the surface so the parameter value diagnosing decoupling is increased in GA7 and a linear weighting is introduced to better match large eddy simulations. The boundary layer scheme includes the parameterisation of turbulent entrainment at the top of cloudy boundary layers. There are sources from both the cloud top (radiative and evaporative cooling) and the surface (positive buoyancy fluxes and wind shear).
- **Forced convective clouds and resolved mixing across the top of the boundary layer (#GA83)**: a new parameterisation of the profile of cloud fraction for clouds forming in the capping inversion is implemented that varies linearly in height between the cloud base and cloud top. Forced convective clouds can form in undulations of the top of convective boundary layers when the capping inversion thickness is parameterised. These are too shallow to reach the level of free convection and become cumulus clouds. Previously, capping inversions at the top of convective boundary layers were assumed to be thin enough to be unresolved by the model so the entrainment flux across the boundary layer could be applied at one vertical level. As the model resolution increases this is no longer the case.
- **Retuned cloud threshold for diagnosing a shear dominated boundary layer (#GA162)**: the fraction of the cloud layer through which the Richardson number is calculated to diagnose a shear-dominated boundary layer is changed from 0.3 to 0.4. This change was not scientifically formulated but instead is a tuning change implemented in GA7.

references

Ahmadi-Givi, F., Graig, G. and Plant, R. (2004) The dynamics of a midlatitude cyclone with very strong latent-heat release. *Quart. J. Roy. Meteor. Soc.*, **130**, 295–323.

de Andrade, F. M., Coelho, C. A. and Cavalcanti, I. F. (2019) Global precipitation hindcast quality assessment of the Subseasonal to Seasonal (S2S) prediction project models. *Climate Dyn.*, **52**, 5451–5475.

Arakawa, A. and Lamb, V. (1977) Computational design of the basic dynamical processes of the UCLA General Circulation Model. *Methods Comput. Phys.*, **17**, 173–265.

Arpe, K., Hollingsworth, A., Tracton, M., Lorenc, A., Uppala, S. and Kållberg, P. (1985) The response of numerical weather prediction systems to FGGE level IIb data. Part II: Forecast verifications and implications for predictability. *Quart. J. Roy. Meteor. Soc.*, **111**, 67–101.

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

Berggren, R., Bolin, B. and Rossby, C.-G. (1949) An aerological study of zonal motion, its perturbations and break-down. *Tellus*, **1**, 14–37.

Binder, H., Boettcher, M., Joos, H. and Wernli, H. (2016) The role of warm conveyor belts for the intensification of extratropical cyclones in Northern Hemisphere winter. *J. Atmos. Sci.*, **73**, 3997–4020.

Bowler, N. E., Arribas, A., Mylne, K. R., Robertson, K. B. and Beare, S. E. (2008) The MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **134**, 703–722.

Branković, Č. and Molteni, F. (1996) Sensitivity of the ECMWF model northern winter climate to model formulation. *Climate Dyn.*, **13**, 75–101.

Brown, A., Beare, R., Edwards, J., Lock, A., Keogh, S., Milton, S. and Walters, D. (2008) Upgrades to the boundary-layer scheme in the Met Office numerical weather prediction model. *Boundary Layer Meteorol.*, **128**, 117–132.

Browning, K. (1971) Radar measurements of air motion near fronts. *Weather*, **26**, 320–340.

Browning, K. and Roberts, N. (1994) Structure of a frontal cyclone. *Quart. J. Roy. Meteor. Soc.*, **120**, 1535–1557.

Buizza, R., Houtekamer, P., Pellerin, G., Toth, Z., Zhu, Y. and Wei, M. (2005) A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Weather Rev.*, **133**, 1076–1097.

Buizza, R., Milleer, M. and Palmer, T. (1999) Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908.

Carrera, M. L., Gyakum, J. R. and Zhang, D.-L. (1999) A numerical case study of secondary marine cyclogenesis sensitivity to initial error and varying physical processes. *Mon. Weather Rev.*, **127**, 641–660.

Chagnon, J., Gray, S. and Methven, J. (2013) Diabatic processes modifying potential vorticity in a North Atlantic cyclone. *Quart. J. Roy. Meteor. Soc.*, **139**, 1270–1282.

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

Chelton, D. B. (2005) The Impact of SST Specification on ECMWF Surface Wind Stress Fields in the Eastern Tropical Pacific. *J. Climate*, **18**, 530–550.

Clark, D., Mercado, L., Sitch, S., Jones, C., Gedney, N., Best, M., Pryor, M., Rooney, G., Essery, R., Blyth, E. et al. (2011) The Joint UK Land Environment Simulator (JULES), model description–Part 2: carbon fluxes and vegetation dynamics. *Geosci. Model Dev.*, **4**, 701–722.

Colucci, S. J. (1985) Explosive cyclogenesis and large-scale circulation changes: Implications for atmospheric blocking. *J. Atmos. Sci.*, **42**, 2701–2717.

Coronel, B., Ricard, D., Rivière, G. and Arbogast, P. (2015) Role of Moist Processes in the Tracks of Idealized Midlatitude Surface Cyclones. *J. Atmos. Sci.*, **72**, 2979–2996.

Curio, J., Chen, Y., Schiemann, R., Turner, A. G., Wong, K. C., Hodges, K. and Li, Y. (2018) Comparison of a manual and an automated tracking method for Tibetan Plateau vortices. *Adv. Atmos. Sci.*, **35**, 965–980.

Curry, J., Hobbs, P., King, M., Randall, D., Minnis, P., Isaac, G., Pinto, J., Uttal, T., Bucholtz, A., Cripe, D. et al. (2000) FIRE Arctic clouds experiment. *Bull. Amer. Meteor. Soc.*, **81**, 5–30.

Davies, H. C. and Didone, M. (2013) Diagnosis and dynamics of forecast error growth. *Mon. Weather Rev.*, **141**, 2483–2501.

Davini, P. and D'Andrea, F. (2016) Northern Hemisphere atmospheric blocking representation in global climate models: Twenty years of improvements? *J. Climate*, **29**, 8823–8840.

Davis, C. A. and Emanuel, K. A. (1991) Potential vorticity diagnostics of cyclogenesis. *Mon. Weather Rev.*, **119**, 1929–1953.

Davis, C. A., Stoelinga, M. T. and Kuo, Y.-H. (1993) The integrated effect of condensation in numerical simulations of extratropical cyclogenesis. *Mon. Weather Rev.*, **121**, 2309–2330.

Dawson, A. and Palmer, T. (2015) Simulating weather regimes: Impact of model resolution and stochastic parameterization. *Climate Dyn.*, **44**, 2177–2193.

Dearden, C., Vaughan, G., Tsai, T. and Chen, J.-P. (2016) Exploring the diabatic role of ice microphysical processes in two North Atlantic summer cyclones. *Mon. Weather Rev.*, **144**, 1249–1272.

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E. and Wimmer, W. (2012) The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sens. Environ.*, **116**, 140–158.

Ferranti, L., Corti, S. and Janousek, M. (2015) Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. *Quart. J. Roy. Meteor. Soc.*, **141**, 916–924.

Goddard, L., Mason, S. J., Zebiak, S. E., Ropelewski, C. F., Basher, R. and Cane, M. A. (2001) Current approaches to seasonal to interannual climate predictions. *Int. J. Climatol.*, **21**, 1111–1152.

Grams, C. M. and Archambault, H. M. (2016) The key role of diabatic outflow in amplifying the midlatitude flow: A representative case study of weather systems surrounding western North Pacific extratropical transition. *Mon. Weather Rev.*, **144**, 3847–3869.

Grams, C. M., Magnusson, L. and Madonna, E. (2018) An atmospheric dynamics 'perspective on the amplification and propagation of forecast error in numerical weather prediction models: a case study. *Quart. J. Roy. Meteor. Soc.*, **144**, 2577–2591.

Grams, C. M., Wernli, H., Böttcher, M., Čampa, J., Corsmeier, U., Jones, S. C., Keller, J. H., Lenz, C.-J. and Wiegand, L. (2011) The key role of diabatic processes in modifying the upper-tropospheric wave guide: a North Atlantic case-study. *Quart. J. Roy. Meteor. Soc.*, **137**, 2174–2193.

Gray, S. L., Dunning, C., Methven, J., Masato, G. and Chagnon, J. M. (2014) Systematic model forecast error in Rossby wave structure. *Geophys. Res. Lett.*, **41**, 2979–2987.

Gregory, D. and Rowntree, P. (1990) A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Mon. Weather Rev.*, **118**, 1483–1506.

Harrison, M., Palmer, T., Richardson, D. and Buizza, R. (1999) Analysis and model dependencies in medium-range ensembles: Two transplant case-studies. *Quart. J. Roy. Meteor. Soc.*, **125**, 2487–2515.

Hodges, K. I. (1994) A general method for tracking analysis and its application to meteorological data. *Mon. Weather Rev.*, **122**, 2573–2586.

— (1995) Feature tracking on the unit sphere. *Mon. Weather Rev.*, **123**, 3458–3465.

— (1999) Adaptive constraints for feature tracking. *Mon. Weather Rev.*, **127**, 1362–1373.

928 Hodges, K. I. and Emerton, R. (2015) The prediction of Northern Hemisphere tropical cyclone extended life cycles by the
929 ECMWF ensemble and deterministic prediction systems. Part I: Tropical cyclone stage. *Mon. Weather Rev.*, **143**, 5091–
930 5114.

931 Hoskins, B. and Hodges, K. (2019) The annual cycle of Northern Hemisphere storm tracks. Part I: Seasons. *J. Climate*, **32**,
932 1743–1760.

933 Hoskins, B. J. and Hodges, K. I. (2002) New perspectives on the Northern Hemisphere winter storm tracks. *J. Atmos. Sci.*, **59**,
934 1041–1061.

935 — (2005) A new perspective on Southern Hemisphere storm tracks. *J. Climate*, **18**, 4108–4129.

936 Illingworth, A., Hogan, R., O'connor, E., Bouniol, D., Brooks, M., Delanoë, J., Donovan, D., Eastment, J., Gaussiat, N., God-
937 dard, J. et al. (2007) Cloudnet: Continuous evaluation of cloud profiles in seven operational models using ground-based
938 observations. *Bull. Amer. Meteor. Soc.*, **88**, 883–898.

939 Joos, H. and Forbes, R. M. (2016) Impact of different IFS microphysics on a warm conveyor belt and the downstream flow
940 evolution. *Quart. J. Roy. Meteor. Soc.*, **142**, 2727–2739.

941 Joos, H. and Wernli, H. (2012) Influence of microphysical processes on the potential vorticity development in a warm conveyor
942 belt: A case-study with the limited-area model COSMO. *Quart. J. Roy. Meteor. Soc.*, **138**, 407–418.

943 Jung, T., Balsamo, G., Bechtold, P., Beljaars, A., Köhler, M., Miller, M., Morcrette, J.-J., Orr, A., Rodwell, M. and Tompkins, A.
944 (2010) The ECMWF model climate: Recent progress through improved physical parametrizations. *Quart. J. Roy. Meteor.*
945 *Soc.*, **136**, 1145–1160.

946 Kalnay, E. (2003) *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press.

947 Klein, S. A., McCoy, R. B., Morrison, H., Ackerman, A. S., Avramov, A., Boer, G. d., Chen, M., Cole, J. N., Del Genio, A. D., Falk, M.
948 et al. (2009) Intercomparison of model simulations of mixed-phase clouds observed during the ARM Mixed-Phase Arctic
949 Cloud Experiment. I: Single-layer cloud. *Quart. J. Roy. Meteor. Soc.*, **135**, 979–1002.

950 Lillo, S. P. and Parsons, D. B. (2017) Investigating the dynamics of error growth in ECMWF medium-range forecast busts.
951 *Quart. J. Roy. Meteor. Soc.*, **143**, 1211–1226.

952 Lock, A. (2001) The numerical representation of entrainment in parameterizations of boundary layer turbulent mixing. *Mon.*
953 *Weather Rev.*, **129**, 1148–1163.

954 Lock, A., Brown, A., Bush, M., Martin, G. and Smith, R. (2000) A new boundary layer mixing scheme. Part I: Scheme description
955 and single-column model tests. *Mon. Weather Rev.*, **128**, 3187–3199.

956 Lorenz, E. N. (1963) Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

957 Luo, D., Cha, J., Zhong, L. and Dai, A. (2014) A nonlinear multiscale interaction model for atmospheric blocking: The eddy-
958 blocking matching mechanism. *Quart. J. Roy. Meteor. Soc.*, **140**, 1785–1808.

959 Lupo, A. R. and Smith, P. J. (1995) Climatological features of blocking anticyclones in the Northern Hemisphere. *Tellus A*, **47**,
960 439–456.

961 Maddison, J. W., Gray, S. L., Martínez-Alvarado, O. and Williams, K. D. (2019) Upstream Cyclone Influence on the Predictability
962 of Block Onsets over the Euro-Atlantic Region. *Mon. Weather Rev.*, **147**, 1277–1296.

963 Madonna, E., Wernli, H., Joos, H. and Martius, O. (2014) Warm conveyor belts in the ERA-Interim dataset (1979–2010). Part
964 I: Climatology and potential vorticity evolution. *J. Climate*, **27**, 3–26.

965 Martínez-Alvarado, O., Maddison, J. W., Gray, S. L. and Williams, K. D. (2018) Atmospheric blocking and upper-level Rossby-
966 wave forecast skill dependence on model configuration. *Quart. J. Roy. Meteor. Soc.*, **144**, 2165–2181.

967 Martínez-Alvarado, O., Madonna, E., Gray, S. and Joos, H. (2016) A route to systematic error in forecasts of Rossby waves.
968 *Quart. J. Roy. Meteor. Soc.*, **142**, 196–210.

969 Martínez-Alvarado, O. and Plant, R. (2014) Parametrized diabatic processes in numerical simulations of an extratropical cy-
970 clone. *Quart. J. Roy. Meteor. Soc.*, **140**, 1742–1755.

971 Matsueda, M. (2009) Blocking predictability in operational medium-range ensemble forecasts. *Sola*, **5**, 113–116.

972 — (2011) Predictability of Euro-Russian blocking in summer of 2010. *Geophys. Res. Lett.*, **38**.

973 Matsueda, M. and Palmer, T. (2018) Estimates of flow-dependent predictability of wintertime Euro-Atlantic weather regimes
974 in medium-range forecasts. *Quart. J. Roy. Meteor. Soc.*, **144**, 1012–1027.

975 McLeod, J. T. and Mote, T. L. (2015) Assessing the role of precursor cyclones on the formation of extreme Greenland blocking
976 episodes and their impact on summer melting across the Greenland ice sheet. *Journal of Geophysical Research: Atmospheres*,
977 **120**, 12357–12377.

978 Molteni, F., Buizza, R., Palmer, T. N. and Petroliagis, T. (1996) The ECMWF ensemble prediction system: Methodology and
979 validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.

980 Morcrette, J.-J., Barker, H. W., Cole, J. N. S., Iacono, M. J. and Pincus, R. (2008) Impact of a New Radiation Package, McRad,
981 in the ECMWF Integrated Forecasting System. *Mon. Weather Rev.*, **136**, 4773–4798.

982 Nakamura, H. and Wallace, J. M. (1993) Synoptic behavior of baroclinic eddies during the blocking onset. *Mon. Weather Rev.*,
983 **121**, 1892–1903.

984 Oertel, A., Boettcher, M., Joos, H., Sprenger, M., Konow, H., Hagen, M. and Wernli, H. (2019) Convective activity in an
985 extratropical cyclone and its warm conveyor belt – a case-study combining observations and a convection-permitting
986 model simulation. *Quart. J. Roy. Meteor. Soc.*, **145**, 1406–1426.

987 O'Reilly, C. H., Minobe, S. and Kuwano-Yoshida, A. (2016) The influence of the Gulf Stream on wintertime European blocking.
988 *Climate Dyn.*, **47**, 1545–1567.

989 Orrell, D., Smith, L., Barkmeijer, J. and Palmer, T. (2001) Model error in weather forecasting. *Nonlinear Processes Geophys.*, **8**,
990 357–371.

991 Palmer, T. N. and Anderson, D. L. T. (1994) The prospects for seasonal forecasting —A review paper. *Quart. J. Roy. Meteor. Soc.*,
992 **120**, 755–793.

993 Park, K., Cornillon, P. and Codiga, D. L. (2006) Modification of surface winds near ocean fronts: Effects of Gulf Stream rings
994 on scatterometer (QuikSCAT, NSCAT) wind observations. *Journal of Geophysical Research: Oceans*, **111**.

995 Pelly, J. L. and Hoskins, B. J. (2003a) How well does the ECMWF Ensemble Prediction System predict blocking? *Quart. J. Roy.*
996 *Meteor. Soc.*, **129**, 1683–1702.

997 — (2003b) A new perspective on blocking. *J. Atmos. Sci.*, **60**, 743–755.

998 Pfahl, S., Schwierz, C., Croci-Maspoli, M., Grams, C. and Wernli, H. (2015) Importance of latent heat release in ascending air
999 streams for atmospheric blocking. *Nat. Geosci.*, **8**, 610–614.

1000 Pomroy, H. R. and Thorpe, A. J. (2000) The evolution and dynamical role of reduced upper-tropospheric potential vorticity in
1001 intensive observing period one of FASTEX. *Mon. Weather Rev.*, **128**, 1817–1834.

1002 Rabier, F., Klinker, E., Courtier, P. and Hollingsworth, A. (1996) Sensitivity of forecast errors to initial conditions. *Quart. J. Roy.*
1003 *Meteor. Soc.*, **122**, 121–150.

1004 Rodwell, M. J., Magnusson, L., Bauer, P., Bechtold, P., Bonavita, M., Cardinali, C., Diamantakis, M., Earnshaw, P., Garcia-Mendez,
1005 A., Isaksen, L. et al. (2013) Characteristics of occasional poor medium-range weather forecasts for Europe. *Bull. Amer.*
1006 *Meteor. Soc.*, **94**, 1393–1405.

1007 Saffin, L., Gray, S., Methven, J. and Williams, K. (2017) Processes maintaining tropopause sharpness in numerical models.
1008 *Journal of Geophysical Research: Atmospheres*, **122**, 9611–9627.

Scaife, A. A., Copsey, D., Gordon, C., Harris, C., Hinton, T., Keeley, S., O'Neill, A., Roberts, M. and Williams, K. (2011) Improved Atlantic winter blocking in a climate model. *Geophys. Res. Lett.*, **38**.

Schäfler, A., Craig, G., Wernli, H., Arbogast, P., Doyle, J. D., McTaggart-Cowan, R., Methven, J., Rivière, G., Ament, F., Boettcher, M. et al. (2018) The North Atlantic waveguide and downstream impact experiment. *Bulletin of the American Meteorological Society*, **99**, 1607–1637.

Scherrer, S. C., Croci-Maspoli, M., Schwierz, C. and Appenzeller, C. (2006) Two-dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns in the Euro-Atlantic region. *Int. J. Climatol.*, **26**, 233–250.

Schiemann, R., Demory, M.-E., Shaffrey, L. C., Strachan, J., Vidale, P. L., Mizielinski, M. S., Roberts, M. J., Matsueda, M., Wehner, M. F. and Jung, T. (2017) The resolution sensitivity of Northern Hemisphere blocking in four 25-km atmospheric global circulation models. *J. Clim.*, **30**, 337–358.

Sheldon, L., Czaja, A., Vannière, B., Morcrette, C., Sohet, B., Casado, M. and Smith, D. (2017) A 'warm path' for Gulf Stream–troposphere interactions. *Tellus A: Dynamic Meteorology and Oceanography*, **69**, 1299397.

Shutts, G. (1983) The propagation of eddies in diffluent jetstreams: Eddy vorticity forcing of 'blocking' flow fields. *Quart. J. Roy. Meteor. Soc.*, **109**, 737–761.

Sprenger, M. and Wernli, H. (2015) The LAGRANTO Lagrangian analysis tool–version 2.0. *Geosci. Model Dev.*, **8**, 2569–2586.

Steinfeld, D. and Pfahl, S. (2019) The role of latent heating in atmospheric blocking dynamics: a global climatology. *Climate Dynamics*, **53**, 6159–6180.

Stensrud, D. J., Bao, J.-W. and Warner, T. T. (2000) Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Weather Rev.*, **128**, 2077–2107.

Stoelinga, M. T. (1996) A potential vorticity-based study of the role of diabatic heating and friction in a numerically simulated baroclinic cyclone. *Mon. Weather Rev.*, **124**, 849–874.

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

Toth, Z. and Kalnay, E. (1997) Ensemble forecasting at NCEP and the breeding method. *Mon. Weather Rev.*, **125**, 3297–3319.

Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J., Morcrette, C., Mulcahy, J., Sanchez, C., Smith, C., Stratton, R., Tennant, W., Tomassini, L., Van Weverberg, K., Vosper, S., Willett, M., Browse, J., Bushell, A., Carslaw, K., Dalvi, M., Essery, R., Gedney, N., Hardiman, S., Johnson, B., Johnson, C., Jones, A., Jones, C., Mann, G., Milton, S., Rumbold, H., Sellar, A., Ujiie, M., Whitall, M., Williams, K. and Zerroukat, M. (2019) The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations. *Geoscientific Model Development*, **12**, 1909–1963.

Walters, D., Boutle, I., Brooks, M., Melvin, T., Stratton, R., Vosper, S., Wells, H., Williams, K., Wood, N., Allen, T., Bushell, A., Copsey, D., Earnshaw, P., Edwards, J., Gross, M., Hardiman, S., Harris, C., Heming, J., Klingaman, N., Levine, R., Manners, J., Martin, G., Milton, S., Mittermaier, M., Morcrette, C., Riddick, T., Roberts, M., Sanchez, C., Selwood, P., Stirling, A., Smith, C., Suri, D., Tennant, W., Vidale, P. L., Wilkinson, J., Willett, M., Woolnough, S. and Xavier, P. (2017) The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations. *Geoscientific Model Development*, **10**, 1487–1520.

Wernli, B. H. and Davies, H. C. (1997) A Lagrangian-based analysis of extratropical cyclones. I: The method and some applications. *Quart. J. Roy. Meteor. Soc.*, **123**, 467–489.

Wernli, H., Dirren, S., Liniger, M. A. and Zillig, M. (2002) Dynamical aspects of the life cycle of the winter storm 'Lothar' (24–26 December 1999). *Quart. J. Roy. Meteor. Soc.*, **128**, 405–429.

Wernli, H., Paulat, M., Hagen, M. and Frei, C. (2008) SAL— A novel quality measure for the verification of quantitative precipitation forecasts. *Mon. Weather Rev.*, **136**, 4470–4487.

Willett, M. R. and Whitall, M. A. (2017) A Simple Prognostic based Convective Entrainment Rate for the Unified Model: Description and Tests. *Tech. rep.*, Met Office.

Wilson, D. R. and Ballard, S. P. (1999) A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Quart. J. Roy. Meteor. Soc.*, **125**, 1607–1636.

Wilson, D. R., Bushell, A., Kerr-Munslow, A. M., Price, J. D., Morcrette, C. J., Bodas-Salcedo, A. et al. (2008a) PC2: A prognostic cloud fraction and condensation scheme. II: Climate model simulations. *Quart. J. Roy. Meteor. Soc.*, **134**, 2109–2125.

Wilson, D. R., Bushell, A. C., Kerr-Munslow, A. M., Price, J. D. and Morcrette, C. J. (2008b) PC2: A prognostic cloud fraction and condensation scheme. I: Scheme description. *Quart. J. Roy. Meteor. Soc.*, **134**, 2093–2107.

Wood, N., Staniforth, A., White, A., Allen, T., Diamantakis, M., Gross, M., Melvin, T., Smith, C., Vosper, S., Zerroukat, M. et al. (2014) An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quart. J. Roy. Meteor. Soc.*, **140**, 1505–1520.

Yamazaki, A. and Itoh, H. (2013) Vortex–vortex interactions for the maintenance of blocking. Part I: The selective absorption mechanism and a case study. *J. Atmos. Sci.*, **70**, 725–742.

Zappa, G., Shaffrey, L. and Hodges, K. (2014) Can polar lows be objectively identified and tracked in the ECMWF operational analysis and the ERA-Interim reanalysis? *Mon. Weather Rev.*, **142**, 2596–2608.