# *Linking rapid forecast error growth to diabatic processes*

Article

Accepted Version

Sánchez, C. ORCID: https://orcid.org/0000-0002-5069-6849, Methven, J. ORCID: https://orcid.org/0000-0002-7636-6872, Gray, S. ORCID: https://orcid.org/0000-0001-8658-362X and Cullen, M. (2020) Linking rapid forecast error growth to diabatic processes. Quarterly Journal of the Royal Meteorological Society, 146 (732). pp. 3548-3569. ISSN 0035-9009 doi: 10.1002/qj.3861 Available at https://centaur.reading.ac.uk/91525/


Published version at: http://dx.doi.org/10.1002/qj.3861
To link to this article DOI: http://dx.doi.org/10.1002/qj.3861

Publisher: Wiley

**ORIGINAL ARTICLE**

**Journal Section**

# Linking rapid forecast error growth to diabatic processes


**Claudio Sánchez[1] | John Methven[2] | Suzanne Gray[2] | Mike Cullen[1]**

[1]Met Office, Exeter, United Kingdom

[2]Department of Meteorology, University of Reading, UK

**Correspondence**
Claudio Sanchez, Met Office, Fitzroy Rd, Exeter EX1 3PB, U.K
Email: claudio.sanchez@metoffice.gov.uk



**Funding information**
This work was supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra, and was carried out under the umbrella of the Met Office Academic Partnership. Authors would like to thanks the NAWDEX international community - in particular George Craig- for their useful input on the topic of predictability barriers at annual NAWDEX workshops. Two anonymous referees provided useful and helpful suggestions to improve the manuscript.



The predictability of high impact weather events over the North Atlantic is controlled by synoptic-scale systems and the mesoscale structures embedded within them. Despite forecast uncertainty being greatest at small scales at the initial time, forecast error projects strongly onto synoptic and larger scales within days. Different stages of error growth have previously been identified including: convective instability, baroclinic instability and the influence of divergent outflow on the tropopause position, and interactions between disturbances at tropopause level.

Evidence is presented for "predictability barriers" (PBs) identified with events on certain validation dates during the North Atlantic Waveguide and Downstream impact Experiment (NAWDEX) where ensemble spread grows more quickly than usual, but ensemble mean forecast error grows even faster. An advective mechanism for diabatic influence on the development of tropopause ridges is hypothesised to be linked to the PB events. A semi-geostrophic balance tool is used to attribute the response of the 3-D ageostrophic flow to geostrophic and diabatic forcing, enabling a novel diagnostic for Diabatically-Induced Ageostrophic Advection of potential vorticity (DIAA).

It is shown that predictability barriers are linked to events with strong diabatic influence on tropopause advection during the NAWDEX period. Error growth exceeds ensemble spread rate by approximately 4/3 during strong DIAA events, showing that predictive skill is considerably lower in these situations.

**KEYWORDS**
flow-dependent predictability, Rossby waves, ageostrophic advection, semi-geotriptic theory, NAWDEX

## 1 | INTRODUCTION

Although there has been relentless improvement in mid-latitude predictive skill over the last 40 years (Bauer et al., 2015), pushing back the limits of predictability to longer lead times, there are still occasions when predictive skill is much lower than usual, resulting in very low skill on lead times as short as five days (e.g., near zero anomaly correlation of geopotential height at 500 hPa over the North Atlantic region). Rodwell et al. (2013) showed that the worst 100 of these "forecast busts" in the global forecasts from the European Centre for Medium-range Weather Forecasts (ECMWF), evaluated over Central Europe at a lead time of six days, shared a common precursor Rossby wave pattern: a trough over the Rockies and warm, moist air extending polewards across the eastern USA beneath an upper-level ridge. They also noted that mesoscale convective systems were active over the USA in this situation and hypothesised that mis-representation of convection and diabatic processes in these systems may contribute to greater forecast error.

However, "forecast busts" are unlikely to be a result of model error alone and typically occur during particular flow configurations that are inherently less predictable. So, it is not obvious whether the forecast system fails or if it is a particularly challenging situation even for a perfect model running from an uncertain initial state. For example, Grazzini and Vitart (2015) have shown a link between predictability over Europe and transient Rossby wave packets, associated partly with uncertainty in the onset of large-scale blocking. Ferranti et al. (2015) have shown that, out of the transitions between large-scale North Atlantic patterns of variability, the transition to blocking has lowest predictability. Disturbances to the jet stream are central to this flow-dependent predictability. For example, Frame et al. (2011) showed that there is probabilistic predictive skill for jet latitude (over the eastern North Atlantic) beyond 15 days lead time for forecasts when the jet starts in the South. However, predictive skill is lost much earlier when the jet starts in the North which is when Rossby wave breaking and transitions in and out of blocking are more likely to occur (Woollings et al., 2008). In this paper we demonstrate a strong diabatic influence on flow-dependent predictability by applying a novel diagnostic to cases that occurred during the North Atlantic Waveguide and Downstream impact EXperiment (NAWDEX). This period (September-October 2016) featured a succession of active systems, such as recurving tropical cyclones (TCs) and growing extratropical cyclones with their ascending warm conveyor belt (WCB) air-streams, contributing to large-scale ridge building episodes and high impact weather events downstream in Europe. The link with NAWDEX is also important because many of the individual events have been investigated in detail in connection with the aircraft observations and modelling studies (Schäfler et al., 2018) and this paper puts those events into context with variations in predictability over the two month window. Moreover, most of the additional sonde observations during the NAWDEX period were assimilated to create the operational analyses and therefore the quality of the analysis of the atmospheric state is better during this period as a result of the campaign (Schindler et al., 2020).

Many authors have proposed different models and explanations for forecast error growth. One of the most employed methodologies is the examination of the amplification of differences between "twin forecasts", which differ only in their initial conditions (sometimes with the inclusion of stochastic parametrization of unresolved motions). For example, based on idealised simulations Zhang et al. (2007) divided the upscale error growth process into three stages: small errors emerge from the uncertain representation of convective-scale processes, then these errors propagate upscale towards the large-scales via "geostrophic adjustment", and ultimately growth occurs in the balanced flow driven by dry dynamic baroclinic instability. Recently, similar twin experiments have been examined using both global and limited-area models at "convection-permitting" resolution (2.8–4km grid spacing: Selz and Craig 2015; Judt 2018; Zhang et al. 2019). The dynamical mechanisms attributed to the various stages of error amplification are still debated (e.g., Zagar et al. 2017; Zhang et al. 2019). Baumgart et al. (2019) used the PV error analysis framework introduced by Baumgart et al. (2018) to examine mechanisms behind different stages of error growth in twin forecasts using the global ICON model with the stochastic Plant-Craig convection scheme. They identified four stages of upscale error growth at tropopause level:

1. Convective-scale growth [to 12 hours]
2. Influence of advection of the tropopause by divergent wind [0.5-2 days]
3. Error growth from nonlinear near-tropopause dynamics [2-14 days]
4. Planetary scale Rossby wave packets (phase-filtered wave envelope) [beyond 14 days]

Note that the convection (leading to stage 1 error growth) was parametrized in the Baumgart et al. (2019) simulations, but similar behaviour has been found in convection-permitting simulations with explicit representation of convective motion. Their method for attribution to different dynamical processes used piecewise PV inversion for lower-tropospheric and upper-level disturbances separately and then considered the advection of PV in one component by the winds attributed to PV in another component, following Teubler and Riemer (2016). The concept of interaction between Rossby waves at different levels through isentropic advection by the wind attributed to different wave components is central to an explanation of baroclinic instability (Heifetz et al., 2004). Surprisingly, when the methodology was applied to the operational ECMWF ensemble forecasts by Baumgart and Riemer (2019), they found the baroclinic interaction between low-level temperature anomalies and tropopause-level PV anomalies to be a weak contributor to the overall error growth in the ensemble, even though this interaction dominates the growth of the weather systems themselves. However, an important consideration here is that the divergent flow was found from Helmholtz decomposition of the full wind in the model and therefore contains the influence of ageostrophic motion associated with baroclinic wave dynamics, as well as the enhanced divergent outflow attributable to diabatic heating and unbalanced divergent motions (such as gravity waves). Baumgart et al. (2019) argue that stage 2 error growth is associated synoptic-scale ascent in baroclinic waves and the influence of latent heat release there, rather than the "geostrophic adjustment" paradigm of Zhang et al. (2007). Other examples of this error growth behaviour exist. For example, Grams et al. (2018) examined a case where the small-scale error in the structure of an upper-level PV cut-off altered the outflow of a WCB, leading to under-development of a tropopause ridge. Similarly, Martínez-Alvarado et al. (2016) attributed large errors in a ridge development episode in January 2011 to an underestimation of the WCB strength and outflow extent in a cyclone, coming from an under-active cyclogenesis (a baroclinic growth mechanism coupled with divergent outflow error).

Perhaps the most studied situations to date influencing flow-dependent predictability are when re-curving TCs undergo extra-tropical transition (Archambault et al., 2013; Keller et al., 2019). Ex-TCs moving polewards in a large-scale ridge typically have a high-altitude divergent outflow associated with the strong ascent. Grams and Archambault (2016) showed that the primary impact of this outflow is to advect the tropopause, expanding the ridge and the associated negative PV anomaly (stage 2 error growth). Depending on the phase of existing troughs approaching along the waveguide from upstream, the ex-TC can lead to a downstream amplification of a Rossby wave packet, or conversely to a more zonal jet state (Agustí-Panareda et al., 2005; Grams et al., 2011; Grams and Blumer, 2015). This error growth would be associated with stages 3 and 4.

NAWDEX was designed to observe jet stream structure in detail, using multiple observation platforms, as well as the weather systems giving rise to jet stream disturbances, downstream propagation of wave activity and forecast error. Schäfler et al. (2018) state the primary hypothesis of NAWDEX: "Diabatic processes have a major influence on the jet stream structure, downstream development of Rossby waves and eventually high impact weather". In terms of forecast error growth, this evolution corresponds to stages 2 and 3 above.

In this paper we test the "NAWDEX hypothesis" in three steps:

**A** Seek evidence for flow-dependent predictability using operational forecasts during the NAWDEX period;
**B** Quantify the diabatic influence on the balanced flow through the advection of potential vorticity mechanism; and
**C** Test whether or not the situations with lowest predictability are associated with strong diabatic influence.

We seek not only to identify diabatic influence on downstream weather, but also to quantify the effect on forecast error growth rate.

In step A, three sets of operational global forecasts are examined over a 35-day period encompassing the NAWDEX campaign in September and October 2016: the Met Office Global and Regional Ensemble Prediction System - Global version (MOGREPS-G), the Met Office high resolution global forecast (H-Res), and the European Centre for Medium-range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) high resolution forecast. The numerical weather prediction (NWP) systems employed are briefly described in Section 2. Evidence for flow-dependent predictability based on error growth and rate of ensemble spread are shown in Section 3.1.

In step B, the Semi-Geotriptic (SGT) balance tool created by (Cullen, 2018) is used to partition the 3-D ageostrophic flow in the H-Res forecasts into a "balanced-flow component" calculated from the geostrophic forcing of a generalisation to the omega-equation, a "diabatic component" attributed to ageostrophic flow response to diabatic heating and a remainder described as the "unbalanced component". These components are used in turn to calculate the influence of advection of PV at tropopause level in generating Rossby wave disturbances. Note that "geotriptic balance" is essentially like geostrophic balance, but with Ekman friction in the boundary layer which results in a three-way balance and turning of the wind vector. While this friction is essential to the global solutions found, in the mid-latitude tropopause region examined here the model is essentially semi-geostrophic (although there will be a non-local influence of the boundary in inversion). The SGT balance tool is described in Section 2.3. The quantification of diabatic influence on ageostrophic advection of PV in the NAWDEX cases is presented in Section 3.2.

Finally, step C relating the predictability barriers to diabatic influence is explored in Section 3.3. Conclusions are drawn in Section 4.

# 2 | METHODOLOGY

## 2.1 | Met Office global high resolution and ensemble prediction systems

The Met Office global forecasts are calculated using the Met Office Unified Model (MetUM). The configuration operational in autumn 2016 was the Global Atmosphere configuration (GA6.1, Walters et al., 2017). It included the dynamical core named "Even Newer Dynamics for the General Atmospheric Modelling of the Environment" (ENDGame, Wood et al., 2014) solving the deep atmosphere, non-hydrostatic compressible equations in spherical geometry using terrain-following coordinates. The GA6.1 physics components are a radiation scheme (Manners et al., 2012) with sub-grid cloud structure (Hill et al., 2011), a microphysics scheme (Wilson and Ballard, 1999) with improvements to the representation of rain (Abel and Boutle, 2012), a boundary layer scheme (Lock et al., 2000), an orographic drag scheme (Lott and Miller, 1997) with improvements detailed in Vosper (2015), a non-orographic scheme (Scaife et al., 2002), the prognostic cloud fraction and prognostic condensate (PC2) cloud scheme (Wilson et al., 2008), convection scheme (Gregory and Rowntree, 1990) and a land surface model (Best et al., 2011). For all forecasts, 70 terrain-following levels were used with a model top at 80 km.

The horizontal resolution used for the high resolution global forecasts was N768 ($\sim$18 km grid spacing in the mid-latitudes). The MOGREPS-G ensemble was run with a horizontal resolution of N400 ($\sim$35 km) and was comprised of 11 members plus a control. The MOGREPS-G configuration included the Ensemble Transform Kalman Filter (ETKF, Bishop et al., 2001) used to determine the initial condition perturbations to the control, plus the random parameter scheme (Bowler et al., 2008) and the Stochastic Kinetic Energy Backscatter scheme (SKEB, Tennant et al., 2011) designed to represent model uncertainty.

## 2.2 | ECMWF Integrated Forecasting System (IFS)

The high resolution IFS forecast is included to highlight forecast error characteristics in common with the high resolution MetUM forecasts in section 3.1. The version of the operational deterministic IFS during the NAWDEX campaign period was Cycle 41r2. It consists of a shallow atmosphere, hydrostatic, semi-Lagrangian, spectral dynamical core, with an O1280 octahedral reduced Gaussian grid-mesh ( $\sim$9 km horizontal grid spacing in the mid latitudes) and 137 hybrid-pressure terrain-following vertical levels (lid at 0.01 hPa). The introduction of the octahedral reduced Gaussian grid-mesh in that cycle led to a substantial increase in the effective grid-point resolution without the need to increase the spectral truncation of the model (Malardel et al., 2016).

The IFS radiation code is based on the Rapid Radiation Transfer Model (Iacono et al., 2008). Cloud radiation interactions are parametrized with the McICA (Monte Carlo Independent Column Approximation) method (McRad, Morcrette et al., 2008). The clouds and large-scale precipitation scheme is based on Tiedtke (1993), but with an enhanced representation of mixed-phase clouds and prognostic precipitation. Orographic blocking and the orographic gravity wave drag scheme are based on Lott and Miller (1997) and the non-orographic gravity wave drag parametrization in Orr et al. (2010). The moist convection scheme is based on the mass-flux approach of Tiedtke (1989), but including improved representation of tropical variability and Convective Available Potential Energy (CAPE) closure (Bechtold et al., 2008, 2014). The land-surface model is the Tiled ECMWF Scheme for Surface Exchanges over Land (TESSEL, Balsamo et al., 2011). Further details on the IFS physics can be found in (ECMWF, 2016b). The IFS 4D-Var data assimilation scheme is described in (ECMWF, 2016a).

## 2.3 | Semi-Geotriptic inversion tool

The extratropical atmosphere is close to geostrophic and hydrostatic balance on scales larger than the Rossby deformation radius. This observation has been used to derive a heirarchy of "balance models". The Quasi-Geostrophic (QG) model has been extensively used as a theoretical framework to understand observed mid-latitude dynamics. It can be obtained from the primitive equations by asymptotic expansion methods where the small parameter is the Rossby number (e.g., Vallis 2006). In such an approach, geostrophic balance is obtained at zeroth order and the QG equations are obtained at first order. It is not sufficient to consider Rossby number alone and the usual approach to QG is to assume that the Froude number and Rossby number are both small in a matched asymptotic expansion ($Ro \sim Fr \ll 1$). Two important results in the QG system are: i) vertical velocity can be deduced from the knowledge of the geostrophic flow alone using the "omega-equation" and ii) the system possesses a form of PV that is conserved following the geostrophic flow. Important limitations of the QG system are that momentum and temperature are also only advected by the geostrophic (horizontal) flow and that the background static stability is assumed to be uniform in the horizontal, which is far from the observed reality on levels crossing the tropopause.

Several higher order balance models have been derived. The semi-geostrophic (SG) model can be obtained from the primitive equations by assuming only hydrostatic balance and the geostrophic momentum approximation, where the full velocity advects momentum and temperature but the momentum itself is approximated by the geostrophic quantity. Hoskins (1975) showed that this approximation is valid if the rate of change in momentum following trajectories is small compared with the Coriolis acceleration. The full 3-D flow (geostrophic plus ageostrophic parts) advects all variables and no limitations on background static stability are imposed. However, the SG system inherits two important conceptual properties from the QG system:

**1** vertical velocity can be deduced given knowledge of the pressure field (and geostrophic motion) only through the SG omega equation (Hoskins and Draghici, 1977) and

**2** in the constant rotation case the system possesses a conserved PV which has the same form as Ertel PV but with an approximation to the relative vorticity vector.

This system can be derived using matched asymptotic analysis with perturbation parameter $\epsilon \sim Ro \sim Fr^2$ and is obtained at second order as $\epsilon$ tends to zero. This limit implies that the radius of curvature of trajectories is greater than the Rossby radius of deformation (Hoskins, 1975). At smaller scales, SG is no more accurate than QG except in the particular case of parallel flows, such straight fronts and jets, with no variation along the front. In these circumstances SG retains second order even if small scales develop in the cross-frontal direction. A similar second order balance approximation can be constructed for strictly axisymmetric flows, approximating winds around tropical cyclones. In the realistic intermediate curved flow situation, quantitatively the SG balanced flow may be no closer to the full wind than QG, although qualitatively the SG model is more similar to the full system in terms of allowing horizontal variation of static stability and advection by the horizontal ageostrophic wind.

The semi-geotriptic (SGT) model improves the semi-geostrophic model by the inclusion of Ekman friction. Where friction is active, the zeroth order solution becomes a three-way balance between friction, Coriolis and pressure gradient forces (Beare and Cullen, 2010). The SGT model then makes the "geotriptic momentum approximation" while retaining advection of all variables by the full flow and a new equation set is derived. Importantly, the SGT balance model enables deduction of the 3-D ageostrophic motion (including vertical velocity) from knowledge of the pressure field alone, even in the presence of a frictional atmospheric boundary layer. Furthermore, outside the boundary layer the PV is conserved (in the constant rotation case) following the full flow, just as in the SG system. The conserved PV is a function of pressure field alone through the geostrophic and hydrostatic balance assumptions. Thus the PV inversion procedure can be achieved by first solving an evolution equation for pressure, and then inferring the ageostrophic velocity. In the variable rotation case this PV is no longer conserved, but the evolution is still completely determined by the pressure field and the same inversion procedure can be followed. Moreover, the system will conserve the Ertel PV to order $\epsilon^2$. Both of these properties will be used to diagnose the diabatic influence on motion and development of tropopause disturbances.

The SGT inversion tool introduced by Cullen (2018) estimates the ageotriptic wind (meaning deviation from geotriptic balance) and pressure tendency from knowledge of the pressure field and corresponding geotriptic state. Although other balance approximations, such as Charney Nonlinear Balance (Davis et al., 1996) or Alternative Balance (Riemer et al., 2014), will be formally more accurate than SG in some situations, a key point is that the SGT diagnostic tool has been created to solve a global problem using data from the Met Office model (MetUM). No similar balance tool exists that can obtain global solutions. Furthermore, the tool also uses numerical discretisation consistent with the MetUM (in the horizontal and vertical) and therefore is ideally suited to examine the difference between the flow associated with balance and the full flow from the deep atmosphere, non-hydrostatic model. The diagnostic tool will extract the geotriptic part of the MetUM evolution without any differences associated with numerical discretisation or geometry. The balance component of the ageostrophic flow is obtained by inversion without the effects of lateral boundaries imposed on limited area inversion domains. The consistency also enables us to compare the effects of diabatic and frictional processes, as parametrized in the operational MetUM, on the full model solution (which produces the forecasts) and the balanced flow.

A brief description of the mathematical formalism of the SGT tool is outlined here; for the derivation and further details tha reader is referred to sections 2.1 and 2.2 of Cullen (2018). Geotriptic and hydrostatic balance are defined by:

$$\begin{aligned} c_{pd}\theta_v \nabla_h \pi - (f v_e, -f u_e) &= \frac{\partial}{\partial r}\Big(K_m \frac{\partial \mathbf{u}_e}{\partial r}\Big), \\ c_{pd}\theta_v \frac{\partial \pi}{\partial r} &= -g, \end{aligned} \tag{1}$$

where $c_{pd}$ is the specific heat of dry air at constant pressure, $\theta_v$ the virtual potential temperature, $\pi$ Exner pressure, $K_m$ a stability dependent diffusion coefficient in the boundary layer and $g$ the gravitational acceleration. Consistent with the deep atmosphere formulation of the MetUM, $r$ is used to denote the radial coordinate from the Earth's centre, $\nabla_h \pi$ denotes the local horizontal gradient of Exner pressure and the "ageotriptic wind vector" is written in component form as $(u - u_e, v - v_e, w)$ where $u$, $v$ and $w$ refer to the zonal, meridional and vertical velocity components. Outside the boundary layer $(u_e, v_e)$ denotes the geostrophic wind; within the boundary layer Ekman friction results in turning of this "geotriptic wind vector" with height.

Re-arrangement of the three components of the momentum equation and making the geotriptic momentum approximation leads to (2):

$$\mathbf{BQ'}\begin{bmatrix} u - u_e \\ v - v_e \\ w \end{bmatrix} + c_{pd}\theta_v \frac{\partial}{\partial t}(\nabla \pi) = \mathbf{BH'}, \tag{2}$$

where the first term on the left hand side is a vector with the ageotriptic wind and vertical motion as components, weighted by the matrices **B** and **Q′** described in (17) and (21) of Cullen (2018). The second term on the left hand side is the Eulerian time derivative of the pressure gradient. The "forcing term", represented by **H′** in (2) is given by:

$$\mathbf{H'} = \begin{bmatrix} -(\mathbf{u}_e\theta_v)\cdot\nabla\Big(\frac{v_e}{\theta_v}\Big) + S_v - \frac{v_e}{\theta_v}S_{\theta_v} \\ (\mathbf{u}_e\theta_v)\cdot\nabla\Big(\frac{u_e}{\theta_v}\Big) - S_u + \frac{u_e}{\theta_v}S_{\theta_v} \\ -\mathbf{u}_e\cdot\nabla\theta_v + S_{\theta_v} \end{bmatrix} \tag{3}$$

This matrix represents a sum of diabatic forcing (terms involving $S_{\theta_v}$), frictional forcing $(S_u, S_v)$ and advection terms that describe the "geostrophic forcing" of ageotriptic motion (i.e., ageostrophic motion outside the boundary layer). Hence, the linearity of the ageotriptic flow response deduced from the SGT model to the "forcing terms" (3) is demonstrated.

Cullen (2018) shows how eliminating the wind velocity from (2) using the continuity equation yields an elliptic equation for the pressure tendency $\partial\pi/\partial t$ which can be stepped forwards in time (eq. 33 in Cullen 2018). The updated pressure field can be used as a prognostic variable to derive the next time-step values for the geotriptic wind, $\theta_v$ and also the ageotriptic wind from (2). Therefore, the SGT system is a complete balance model that can be integrated in time.

The SGT model is used to calculate a one hour time-step in pressure, given the current MetUM state, and the pressure tendency is also used to find the ageotriptic velocity components by inverting (2). The numerical approximations (e.g., estimates of spatial derivatives) used are consistent with the formulation of the MetUM dynamical core. The MetUM fields are pre-processed by interpolation to coarser resolution (N80, approx. 120 km in the mid-latitudes) as appropriate for the SGT balance. The diabatic forcing term ($S_{\theta_v}$) is calculated from the latest hourly accumulations at the validity time of temperature increments from the parametrizations of shortwave and longwave radiation, large-scale precipitation, convection and boundary layer processes. The cloud increments are embedded in each parametrization increment as a result of the prognostic nature of the PC2 cloud scheme (Wilson et al., 2008). The effect of momentum forcing terms ($S_u$, $S_v$) is not included here, apart from the Ekman friction in the boundary layer which is included as part of the geotriptic balance.

Note that the procedure can be regarded as analogous to inverting the QG "omega equation" to obtain vertical motion except that here the 3-D components of the ageostrophic velocity are obtained, a frictional boundary layer is included in the balance and the full NWP model is used to provide the pressure field and all the forcing terms. In principal the pressure tendency term could be eliminated from (2) to obtain an elliptic equation for ageostrophic velocity (equivalent to the QG omega equation), but the approach to find pressure tendency first is preferred numerically because pressure is the smoothest variable (Cullen, 2018) and the tendency defines the balanced evolution.

Note that SGT balance does not represent situations where the flow satisfies conditions for symmetric instability ($f\mathrm{PV} < 0$, where $f$ is the Coriolis parameter). The diagnostic tool cannot reproduce the unstable motions of symmetric instability, since that is outside the scope of the model. However, slantwise ageostrophic circulations, arising from geostrophic or diabatic forcing, are part of the SGT diagnostic (as long as $f\mathrm{PV} \geq 0$). This will be a further reason for differences between the ageostrophic winds diagnosed from balance and the full flow, in addition to the inaccuracy of the balance approximation and other unbalanced fast motions.

In summary, the SGT model overcomes the following limitations, as identified by Davies (2015), of the application of the QG system to studies of extratropical dynamics:

**1** The SGT model is suitable where the tropopause slopes, resulting in large contrasts in static stability on horizontal surfaces intersecting the tropopause. The QG model assumes uniform static stability on each level;

**2** The SGT model includes advection by the ageostrophic velocity. This has profound consequences for the evolution of the balanced flow, for example in accelerating frontogenesis (Hoskins and Bretherton, 1972); and

**3** Ertel PV is conserved by SGT, to the order of accuracy of the SGT model, in adiabatic and frictionless conditions.

Furthermore, SGT balance (and estimated vertical motion) is expected to be quantitatively more accurate than QG where the jet stream and associated PV gradient is approximately straight (an upper-level front).

In Section 3.2, we use the SGT inversion tool to diagnose the influence of diabatic processes on ageostrophic motion, especially divergent outflow in the upper troposphere. We exploit the PV conservation property of the SG model (item 3) to partition advection of Ertel PV by the diabatically-forced ageostrophic wind from the full advection and to isolate the role of this component in the growth of tropopause ridges. Since the study focuses at tropopause level, where the SGT model has no friction and $(u_e, v_e)$ is the geostrophic velocity, we will refer to the ageotriptic wind diagnosed from (2) as the "ageostrophic wind" which is a more familiar concept in the literature. However, due to the non-local nature of the inversion equation, it is not the same as would be obtained by the SG model where friction is neglected at all locations.

# 3 | RESULTS

The results are divided into three sections, relating to the three steps in testing the NAWDEX hypothesis that diabatic processes have a major influence on predictability of weather systems developing across the North Atlantic (see Introduction). In Section 3.1 evidence is sought for flow-dependent predictability using operational forecasts spanning the NAWDEX period. Section 3.2 uses the new SGT tool to quantify the diabatic influence on Rossby waves arising through the response to heating in ageostrophic motion and the enhanced advection of PV. Finally, Section 3.3 tests whether or not the situations with lowest predictability are associated with stronger than average diabatic influence on tropopause advection.

## 3.1 | Flow-dependent predictability and identification of predictability barriers

One of the most employed metrics for forecast error is the Root Mean Square Error (RMSE) measuring the difference between a forecast and verifying analysis at a given time. Typically, grid point values are compared and a spatial mean is used because it does not require pre-calculation of a climatology and so is easier to compute for a fixed time framework than other measures such as the Anomaly Correlation Coefficient (ACC). We compute the RMSE of the geopotential height at 500 hPa ($Z500$) field over the North Atlantic (60–0°W and 30–70°N), hereafter denoted as $E$, to highlight errors in the location and amplitude of the large-scale balanced flow over this region.

$E$ is computed every six hours in each forecast for the Met Office H-Res and ECMWF IFS forecasts (see Section 2) against their own analyses and for the MOGREPS-G ensemble mean and control against the Met Office H-Res analysis. Forecasts are initialized every 12 hours for the IFS and H-Res systems and every 6 hours for MOGREPS-G. The chosen period is 10 September to 15 October 2016, starting 6 days earlier than the NAWDEX campaign period in order to capture the major cyclone development on 10–15 September which is likely to have influenced the subsequent extratropical transition of TC Ian leading into NAWDEX IOP1.

| | |
|---|---|
| $\partial_f E$ | Rate of increase of $Z500$ RMSE with respect to $f$ |
| $\partial_f \sigma$ | Rate of increase of $Z500$ ensemble spread with $f$ |
| $f$ | Forecast lead time (hours since the start of the forecast) |
| $s$ | Forecast initialization time |
| $t$ | Validation time ($s + f$ in forecasts) |

**TABLE 1** Definition of the metrics and time variables employed in section 3.1.

The rate of increase of $E$ with forecast lead time ($f$) represents the forecast error growth and is denoted as $\partial_f E$ (see Table 1 for a description of the different metrics and time variables employed in this section). The times when error grows faster become clear when $\partial_f E$ is shown over all forecast lead times, $f$, and forecast initialization times, $s$, for the Met Office H-Res forecasts (Figure 1a). If error growth were not flow-dependent, the figure would be expected to have horizontal bands with higher values for increasing $f$ and with very small variability in the $s$-direction. However, there are distinct diagonal stripes where the rate of error increase is large (for $f > 60$ hr) which are parallel to the green diagonal isolines representing constant validation times, $t$ (e.g. a 120 hour forecast started on a particular day will have the same validation time as a 72 hour forecast started 48 hours later). The behaviour is not isolated to the Met Office forecast system. Similarly large rates of forecast error increase have been found around the same times for the IFS Ensemble Mean ACC (George Craig, pers. comm.) and which can also be deduced from the T+120 ACC shown in Fig. 5b

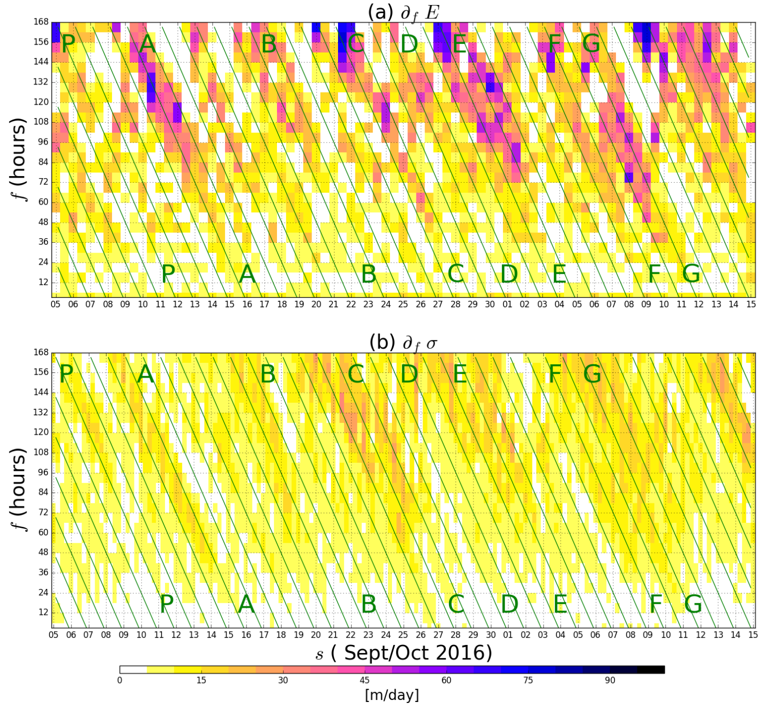


**FIGURE 1** Rate of increase (m day$^{-1}$) with respect to forecast lead time, $f$, of (a) H-Res forecast RMSE in $Z500$ ($\partial_f E$) and (b) MOGREPS-G ensemble spread in $Z500$ ($\partial_f \sigma$) in forecasts started on initialization times, $s$, spanning the NAWDEX campaign. Green diagonal lines represent constant validation times, $t$ (corresponding to 00Z at each forecast start date on the x-axis). Green letters indicate the Predictability Barrier events described in Table 2.

of Schäfler et al. (2018).

The rate of increase of $Z500$ ensemble spread in MOGREPS-G (the standard deviation of ensemble members from the ensemble mean) with forecast lead time, $f$, (hereafter $\partial_f \sigma$) is also greater for approximately the same dates as the large error growth rates. The dependence of the rate of ensemble spread on the dates shows that there is *marked flow-dependent predictability*. However, the rate of forecast error growth ($\partial_f E$) is considerably faster than ensemble growth rate in low predictability situations (compare Fig. 1a with b). We define such events as *predictability barriers (PB)* because forecast skill is considerably lower for a particular date, at all medium-range lead times. The PB events coincide with cases intensively observed during the NAWDEX campaign (Table 2).

To illustrate this flow-dependence further, the error $E$ has been plotted against forecast lead time, $f$, in Figure 2 for four different initialization times, $s$, related to the PB Case **A**. The error, $E$, of the H-Res and IFS forecasts and the MOGREPS-G ensemble mean and control are included, as well as the MOGREPS-G ensemble spread, $\sigma$. The verification time when $\partial_f E$ for PB event A is greatest in Fig. 1a, 00Z 16 September, is marked by an upward arrow in Fig. 2. The arrow coincides with the largest error increase in the four forecasts and all the systems included. There are small differences across systems: the largest increases of $E$ in the IFS (green) and MOGREPS-G control (solid blue) forecasts start a bit earlier than in the Met Office H-Res forecast (red). Only for the forecast initialized at 00Z 14 September (Fig. 2.c) is

| NAME | Date | NAWDEX IOP | Event |
|---|---|---|---|
| P | 11-12 Sept | Pre-NAWDEX | Growth of a large-scale mid-Atlantic cyclone |
| A | 15-16 Sept | IOP1 | Tropical cyclone Ian transitioning into an extratropical cyclone |
| B | 23 Sept | IOP3 | Development of cyclone Vladiana with a strong warm conveyor belt |
| C | 28 Sept | IOP5 | Extratropical transition of TC Karl and the following cyclone Walpurga and atmospheric river |
| D | 1 Oct | IOP6 | Growth of the "Stalactite cyclone" |
| E | 4-5 Oct | IOP7 | Growth of frontal cyclone after Stalactite and onset of blocking over Western Europe |
| F | 9 Oct | IOP9 | Development of upper-level cut-off low and cyclone Sanchez |
| G | 11 Oct | IOP10 | Poleward expansion of ridge Thor |

**TABLE 2** List of Predictability Barrier events: letters shown in Figure 1, their associated dates, the corresponding Intense Observation Periods (IOP) during the NAWDEX campaign and a brief description of key features.

there is a clear improvement of the ensemble mean (dotted blue) over the control. The ensemble spread (dashed blue) increases faster during PB event A, but it does not grow as fast as the error afterwards.

Averaging $\partial_f E$ across all forecast lead times $f$ for each validation time $t$, i.e. along the green diagonal lines in Figure 1, produces the timeseries shown in Figure 3. Again, PB events are clearly distinguishable in all forecasting systems though with some differences, such as the comparatively earlier emergence of the error in the IFS forecasts for PB A, and the lower error from the IFS forecasts for PB D (the Stalactite cyclone). The average rate of error increase is larger during October when the the flow transitions into a Scandinavian blocking regime (see blue line in Figure 5.a of Schäfler et al. 2018). Correlation indices between the timeseries, shown in Table 3, indicate statistically significant correlations between all of them. Thus, all systems indicate rapid error growth over the North Atlantic domain on the same specific dates. The rate of MOGREPS-G ensemble spread $|\partial_f \sigma|_t$ is clearly weaker than rate of growth in error $|\partial_f E|_t$, even for the ensemble mean, and is less well correlated with the error growth variability.

| | IFS | EM | SPREAD |
|---|---|---|---|
| H-Res | 0.725 | 0.849 | 0.488 |
| IFS | · | 0.608 | 0.567 |
| EM | · | · | 0.467 |

**TABLE 3** Pearson correlation coefficients amongst the different forecast error and ensemble spread time series shown in Figure 3. All coefficients are statistically significant at the 99% confidence level.

For a reliable forecast system, ensemble spread is expected to match the ensemble mean error, for all lead times, on averaging over many forecast dates (Leutbecher and Palmer, 2008). There are three plausible reasons why the forecast error growth rate for all forecasts might be larger than the rate of ensemble spread during the PB events:

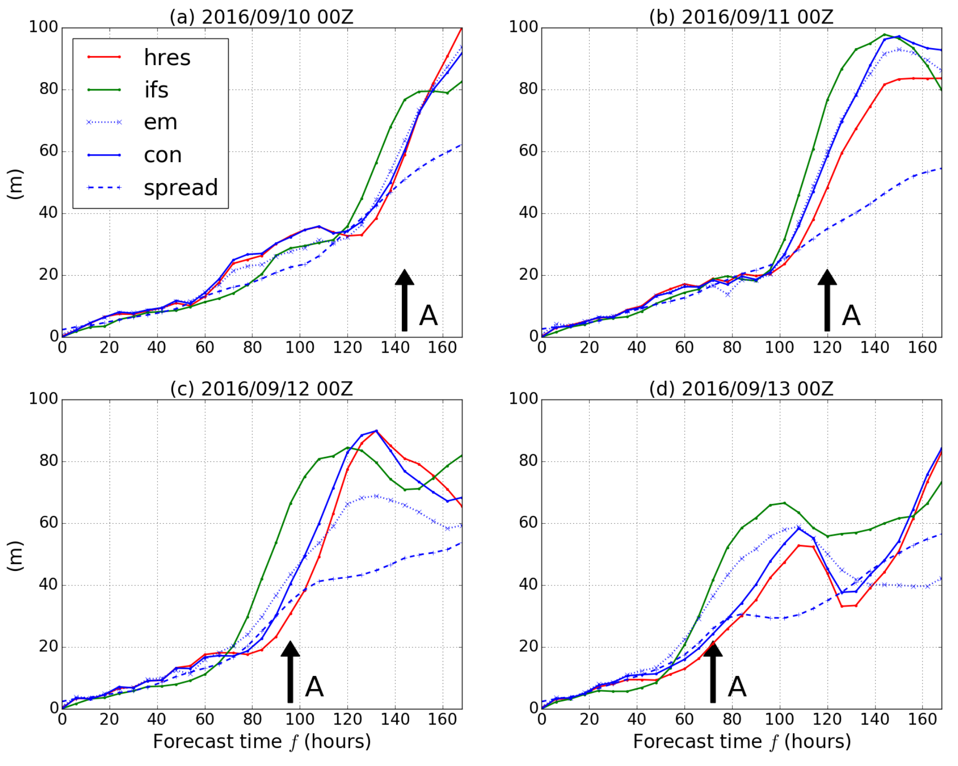


**FIGURE 2** Case study for the predictability barrier event A occurring at 00Z 16 September 2016 (see Table 2). Timeseries of forecast error ($Z500$ RMSE) for H-Res (red), IFS (green), MOGREPS-G ensemble mean (EM, slim solid blue) and control (thick solid blue) across forecast lead time. Ensemble spread in $Z500$ is also shown (dashed blue). Different forecasts initialized at $s$ of (a) 00Z 10 (b) 00Z 11 (c) 00Z 12 and (d) 00Z 13 September 2016. Upright arrows mark the PB event A.

- Intrinsic errors in the model formulation (e.g. from physics parametrizations) that stochastic perturbation schemes used by the ensemble system do not represent;
- The ensemble design of initial conditions may be sub-optimal and result in an ensemble that is under-dispersive
- The observed events have inherently low predictability because they are unlikely trajectories given the initial conditions. In this situation, even if we had a perfect ensemble forecast system, the observed trajectory would be on the edge of the ensemble and the ensemble mean would be expected to have a larger error than usual relative to the observations.

In summary, the forecast error grows rapidly over the same events when simulated by two state-of-the-art NWP models, MetUM and IFS, and also in ensemble mean error. The ensemble spread in MOGREPS-G does not grow as fast as the ensemble mean error. The remainder of the paper investigates the hypothesis that diabatic processes have an influence on tropopause ridges (Section 3.2) and that the rapid error growth in PB events is associated with such

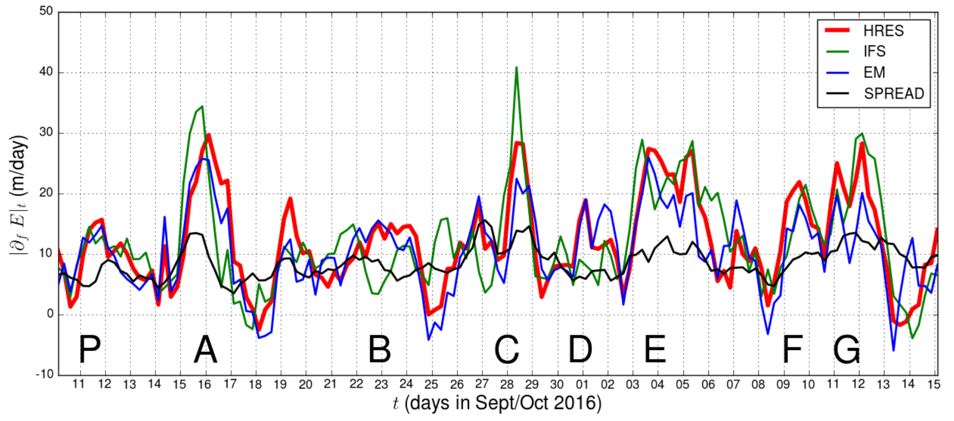


**FIGURE 3** $\partial_f E$ averaged across forecast lead times as a function of validation time for different NWP systems (same colours as in Fig. 2). Letters are placed at the times when the PB events denoted in Table 2 occur.

diabatically-induced perturbations (Section 3.3).

## 3.2 | Diagnosing the influence of diabatic processes on ageostrophic outflow and tropopause advection

The SGT inversion tool described in 2.3 enables the calculation of the ageostrophic flow associated with SGT balance dynamics and partitioning of that flow into a response to geostrophic forcing and forcing by diabatic processes. The focus is the extent to which the ageostrophic flow contributes to advection of PV at tropopause level, thereby altering PV anomalies and balanced flow during the NAWDEX campaign period. Therefore, the SGT inversion is applied six-hourly in each MetUM H-Res forecast out to lead time $f = 120$ hours.

As an example of application of the SGT diagnostic, Figure 4 shows the balanced vertical velocity, $w$, the ageostrophic wind in vectors, $\mathbf{v}_{ag}$, and the dynamical tropopause (PV isoline of 2 PVU) averaged over upper-tropospheric MetUM levels (8.6 to 10.7 km). The MetUM fields, regridded to the same large-scale grid as the SGT fields, are shown in Figure 4a, where $\mathbf{v}_{ag}$ is the difference between the full velocity $(u, v)$ from the MetUM and the geostrophic wind $(u_g, v_g)$. There are three nodes of vertical ascent associated with: (i) ex-TC Ian on the west side of the ridge (48°N, 38°W) with the cyclone PV tower at its centre; (ii) an extratropical cyclone to the Northeast (56°N, 35°W); and (iii) the southern tip of the downstream trough (50°N, 8°W). There is also strong subsidence at the base of the upstream trough (46°N, 42–56°W) associated with tropopause fold deepening.

In the longer forecast of the ex-TC Ian case (PB **A**), the downstream cyclone is weaker and the downstream ridge grows more slowly than in shorter forecasts (e.g., Fig. 5). In long forecasts the ex-TC follows a track much further south relative to the analysis and short forecasts, producing large $\partial_f E$ (shown in Figures 1a and 3) which will result in poorer forecasts of the location and intensity of high weather impact events as the ex-TC and its downstream trough reach northern Europe. The ex-TC Ian case is thus another example where a recurving TC may have had an impact on predictability downstream (Keller et al., 2019).

The vertical motion, $w$, and ageotriptic wind obtained from SGT balance, inverting (2) with the full forcing on the right hand side, is shown in Figure 4b. The structure of the balance resembles quite well the MetUM forecast fields

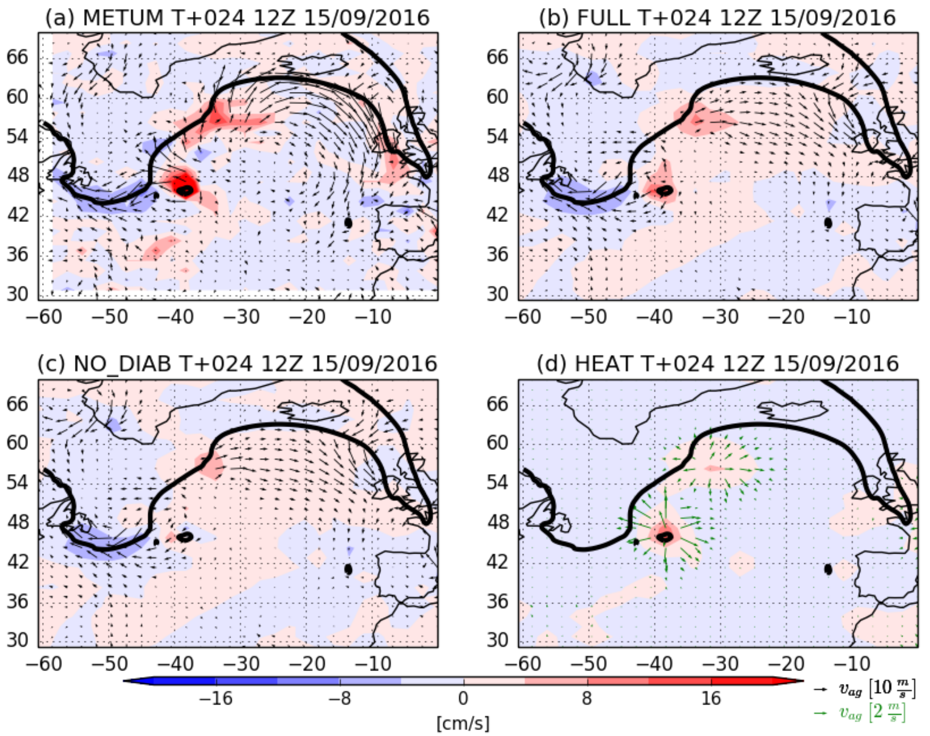


**FIGURE 4** Vertical velocity (cm/s) from (a) MetUM regridded to N80 and (b–d) inverted from the SGT model with (b) all forcing, (c) geostrophic forcing only and (d) diabatic forcing only. In all panels, $t$ is 12Z 16 September 2016 and $f = 24$ hours. Vectors denote horizontal ageotriptic wind: note green ones in (d) have a different scale whose key is in the lower right corner. Thick line is the 2 PVU contour. Fields have been vertically averaged over MetUM levels 35 and 39 (8.6 to 10.7 km). Area shown matches domain for computation of $E$ and $\sigma$.

shown in Figure 4a, but with weaker ageotriptic winds. This is to be expected, given that the balance approximation applied is most accurate on scales larger than the Rossby radius, while the MetUM represents smaller-scale dynamics, including unbalanced motions, that can have larger magnitudes. The decomposition of the FULL SGT fields into the separate contributions from geostrophic and diabatic forcing (as described in section 2.3) produces a $1-2\%$ residual error (over-estimate relative to FULL) in the vertical velocity in the WCB outflow region with marginal values elsewhere (not shown).

The ageostropic winds obtained as a response to geostrophic forcing only are shown in Figure 4c. This solution shows a mainly rotational component to the ageostrophic wind with weaker flow normal to the tropopause on the western flank of the secondary ridge developing downstream of the extratropical cyclone than the solution with both sources (Figure 4b) . The descent at the southern end of the upstream trough is almost entirely a result of geostrophic forcing. In contrast, the ageotriptic wind induced in response to diabatic forcing (Fig. 4d) is purely divergent and produces a component normal to the tropopause on the western side of the ridges. The divergent nature of the flow above strong heating motivated the TC-extratropical flow interaction metric employed in the TC recurving studies of

Archambault et al. (2013, 2015) and Grams and Archambault (2016). Their interaction metric is based on the advection of PV by the irrotational wind (obtained by Helmholtz decomposition of global model winds) and links the strength of TC outflow to the anchoring and amplification of a downstream ridge. However, their metric is not directly attributable to heating because the irrotational wind also includes the balanced response to geostrophic forcing. In particular, within a baroclinic wave the primary ascending air stream is the WCB. Even in dry simulations, the ascent in this region ahead of the upper-level trough results in horizontal divergence in the upper troposphere in the WCB outflow which could advect the tropopause and expand the ridge (Schemm et al., 2013). In the moist atmosphere, latent heat release in the WCB creates a positive feedback on ascent and divergent outflow and is therefore expected to enhance ridge building. Unlike the TC-extratropical flow interaction metric, the new SGT tool enables attribution of ridge building to diabatic processes.

Here, the SGT balance tool is used to calculate the Ageostrophic Advection of PV (AA, $-\mathbf{v}_{ag}.\nabla q$), the dot product of the horizontal ageostrophic wind and the gradient of Ertel PV. The balanced ageostrophic velocity is interpolated to the native H-Res model grid to calculate the dot product with the model's PV field. The justification is that the balanced winds must vary on large-scales to be consistent with the balance approximation, while the PV field has sharp gradients, especially along the tropopause, that require higher resolution to represent. The AA metric can be partitioned according to two contributions to ageostrophic wind: the advection of PV by the diabatically-induced ageostrophic wind (DIAA) and the advection of PV by the ageostrophic wind in response to geostrophic forcing (SGAA). Due to the linearity of (2) with respect to the ageostrophic wind, the contributions to AA are additive because the responses to the different forcings are also additive.

| | |
|---|---|
| DIAA | Diabatic forcing of ageostrophic advection of PV |
| SGAA | Geostrophic forcing of ageostrophic advection of PV |
| S-EG | Strong Error Growth: times matching the upper tercile of $\lvert\partial_f E\rvert_t$ timeseries (points above the dashed black line in in Figure 8) |
| W-EG | Weak Error Growth: times matching the lower tercile of $\lvert\partial_f E\rvert_t$ timeseries (points below the dotted black line in in Figure 8) |
| S-DI | Strong Diabatic Influence: times matching the most negative tercile of $\lvert < DIAA > \rvert_t$ timeseries (points below the dashed green line in Figure 8) |
| W-DI | Weak Diabatic Influence: times matching the most positive tercile of $\lvert < DIAA > \rvert_t$ timeseries (points above the dotted green line in Figure 8) |

**TABLE 4** Definition of the measures of diabatic influence and their time series employed in section 3.2. See Table 1 for variables from previous sections.

The SGAA and DIAA diagnostics for the PB Case **A** are shown in Figure 5.a,c for $f = 24$ hours and Figure 5.b,d for $f = 72$ hour forecasts. The sign of the SGAA and DIAA diagnostics is chosen to indicate the sign of the local tendency of PV arising from advection by the ageostrophic flow. Where the ageostrophic wind vectors are directed from troposphere to stratosphere then AA is negative and it implies ridge expansion as low PV values are being advected over the observer. For example, ridge building occurs on the western side of ex-TC Ian at (45°W, 50°N). On the other hand, when the ageostrophic winds are directed from stratosphere to troposphere then AA is positive.

The usual QG (adiabatic) theory for Rossby wave propagation considers advection of PV by the geostrophic flow. It is expected to exhibit a wave pattern in the PV advection term that must integrate to zero hemispherically, even in

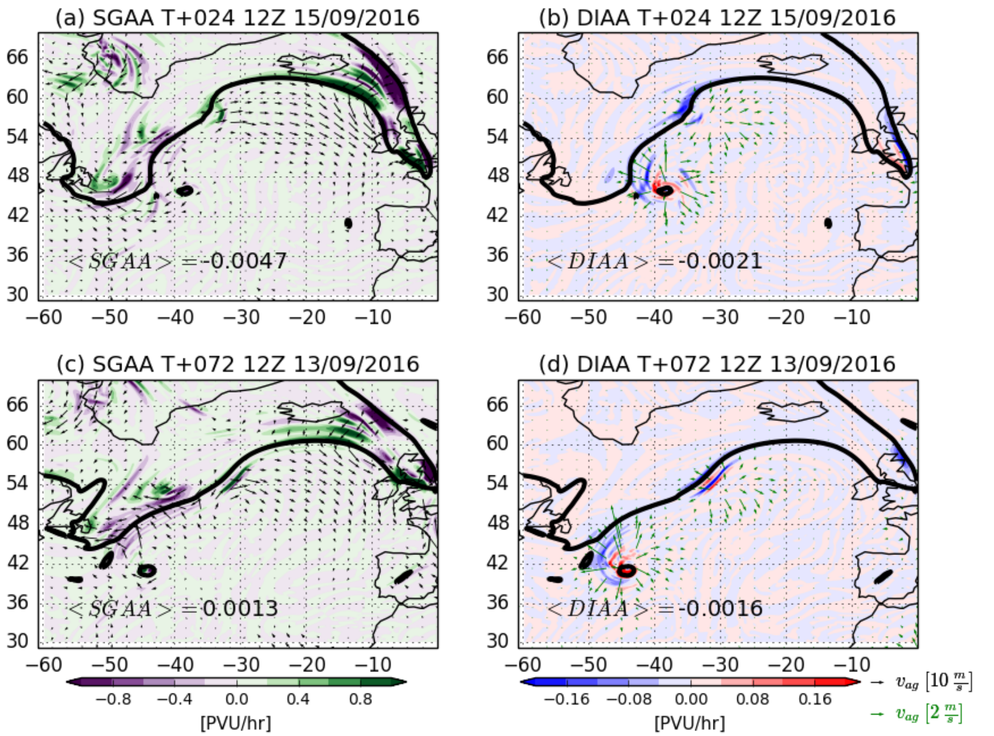


**FIGURE 5** Ageostrophic advection of PV diagnostics on PB Case $A$ for validation time 12Z 16 September. (a) and (c) show SGAA coloured and $\mathbf{v}_{ag}$ from geostrophic forcing as vectors, whereas (b) and (d) show DIAA coloured and $\mathbf{v}_{ag}$ from diabatic forcing as vectors. Solid thick black line shows PV at 2 PVU. For (a) and (b) $f = 24$, and for (c) and (d) $f = 72$. All fields shown have been averaged between model levels L35 and L39 (8.6–10.7 km). The domain-averaged values of the $AA$ diagnostics are included in each panel. Note different colourbars represent the same variable but the (a) and (c) one is 5 times bigger than the one for (b) and (d).

the presence of baroclinic interaction (Heifetz et al., 2004). In the case of SGAA, it is expected that the ageostrophic component of dry balanced baroclinic wave motions would also contribute both to positive and negative PV advection. For example, in Fig. 5 SGAA shows a positive-negative dipole across the downstream trough. The sign of AA indicates that it is contributing to westward (upstream) propagation of this trough. A similar behaviour can be seen on the wider upstream trough. In the longer forecast ($f = 72$ hr) there is stronger negative SGAA to the north of ex-TC Ian (Fig. 5c) which would tend to retard the movement of the large-scale trough towards the east (see Figure 3a of Keller et al. 2019).

The ageostrophic advection induced by diabatic forcing, DIAA, has negative values at almost all locations along the waveguide, with greatest magnitude on the upstream side of the ridge next to the ex-TC Ian and the extratropical cyclone (Figure 5b). The domain-average value (denoted as $< DIAA >$) over 60–0°W and 30–70°N (as for $\partial_f E$) is also negative, indicating an overall ridge building effect from the ageostrophic circulation attributable to diabatic forcing. Longer lead time forecasts of the same event show smaller negative values of DIAA on the western flank of the ridge and smaller negative domain-averaged values too (Fig. 5d). Therefore, we anticipate that DIAA averaged over a baroclinic wave is always net negative because:

- There is asymmetry between diabatic heating and cooling in baroclinic waves where heating rates are typically much faster than cooling rates; and
- Furthermore, large-scale ascent and latent heating occur in the warm conveyor belt running polewards, along the surface cold front, which must lie ahead of the upper-level trough in a baroclinic wave. The WCB outflow and associated divergent component of ageostrophic motion is therefore positioned such that the vectors always point from troposphere to stratosphere.

The synoptic situation and quantification of diabatic influence on PV advection is now presented for some of the other PB events during NAWDEX to reinforce the points above. Figure 6 shows DIAA in the same format as in Figure 5, but for different PB cases at $f = 24$ and $f = 72$ hours; the domain-averaged value <DIAA> is also written in the southwest corner of each panel. For cyclone Vladiana (PB Case **B** discussed in Oertel et al. 2019), the ridge-building activity driven by DIAA is clear on the west side of the ridge (south of Iceland). The $f = 72$ hours forecast shows a smaller ridge and weaker <DIAA> in comparison to the $f = 24$ hours forecast (less than one third in the domain average; Fig. 6a and b). The Stalactite cyclone wrap up (PB Case **D**) produces a ridge formation south of Greenland (at 48°N, 30°E), a phenomenon well captured at $f = 24$ hours, but weaker at $f = 72$ hours (compare Fig. 6c and d). The frontal cyclone that follows the Stalactite cyclone generates a prominent ridge that later develops into a blocking regime over Europe (PB Case **E**) and is a case discussed in detail in Maddison et al. (2019, 2020). DIAA helps with ridge building on the westward side and its contribution is weaker in the longer forecast (Figure 6e,f). The early formation of ridge Thor (PB Case **G**) is also clear in the $f = 24$ hours forecast over NewFoundland (Fig. 6g). It is weaker in the $f = 72$ hour forecast where the cyclone is too far east (Fig. 6h). Additionally, Fig. 6g also shows negative PV advection downstream of the cut-off Sanchez (west of Iberian peninsula), that later brought high impact weather to the western Mediterranean (see Schäfler et al. 2018).

## 3.3 | Relating diabatic influence to predictability barriers

Since the diabatic influence on ageostrophic advection of PV (DIAA) is expected to be negative-definite in the average over the domain (60–0°W and 30–70°N as used to calculate $\partial_f E$), its magnitude is proposed as a measure of the diabatic influence on ridge building and Rossby wave development. The measure is now related to forecast error growth during the NAWDEX campaign period by plotting them together as functions of forecast start times, $s$, and lead times, $f$, in Figure 7. It is clear that $< DIAA >$ is strongly flow-dependent (aligned along the diagonals that represent constant validation times) just as error growth is. Gaps with the weakest values of $< DIAA >$ (during validation time periods 19–22 September and 29–30 September) exhibit weak error growth $< \partial_f E >$. Conversely, there are large magnitudes simultaneously in $< DIAA >$ and $< \partial_f E >$ at validation times associated with the PB Cases **A** and **E**. There are some very strong $< DIAA >$ events where the error growth is only slightly enhanced (**P** and **B**). Finally, there are events where enhanced error growth rate slightly lags DIAA (**C** and **G**).

Other authors have discussed cases with two stages of development related to low predictability (e.g., Martínez-Alvarado et al. 2016; Grams et al. 2018). For example, in the extratropical transition of TCs and their interaction with the jet stream, authors have often referred to the development of a "pre-vortex" in the extratropics ahead of the poleward advancing TC. Latent heat release typically has an important role in the pre-vortex growth and disturbance to the tropopause which becomes important as the upper-tropospheric outflow of the advancing TC enters the larger-scale ridge (see Keller et al. 2019). This situation results in high sensitivity to initial conditions. NAWDEX PB **A** resembles this situation (Fig. 5) with the small-scale extratropical cyclone ahead of ex-TC Ian. Martínez-Alvarado et al. (2016) discussed a “pre-conditioning” event, represented by a weak cyclogenesis on a cold front prior to a major ridge building event

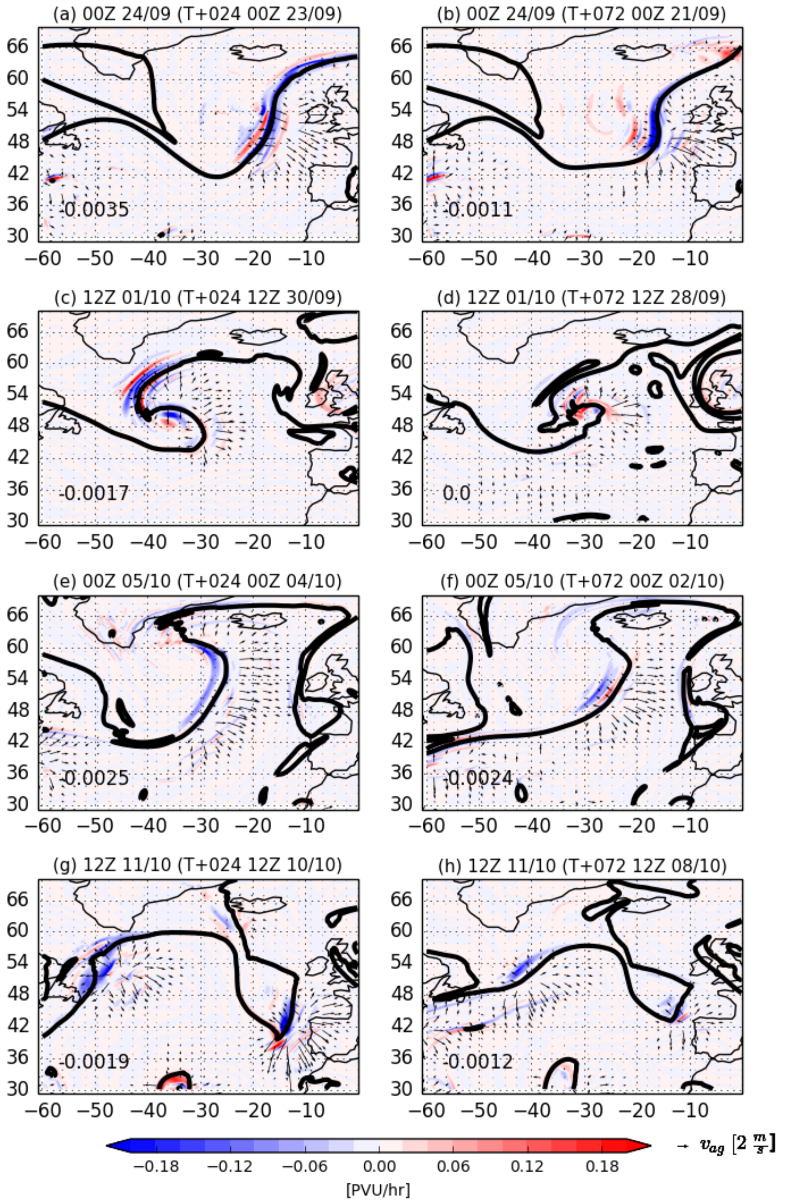


**FIGURE 6** DIAA, as in Figure 5b,d shown at validation times corresponding to difference PBs, as represented by 24 and 72 hour forecasts: (a,b) PB **B** at 00Z 24 September, (c,d) PB **D** 12Z 1 October, (e,f) PB **E** 00Z 5 October, and (g,h) PB **G** 12Z 11 October. $f = 24$ hours in the left column (a,c,e,g) and $f = 72$ hours in the right column (b,d,f,h).

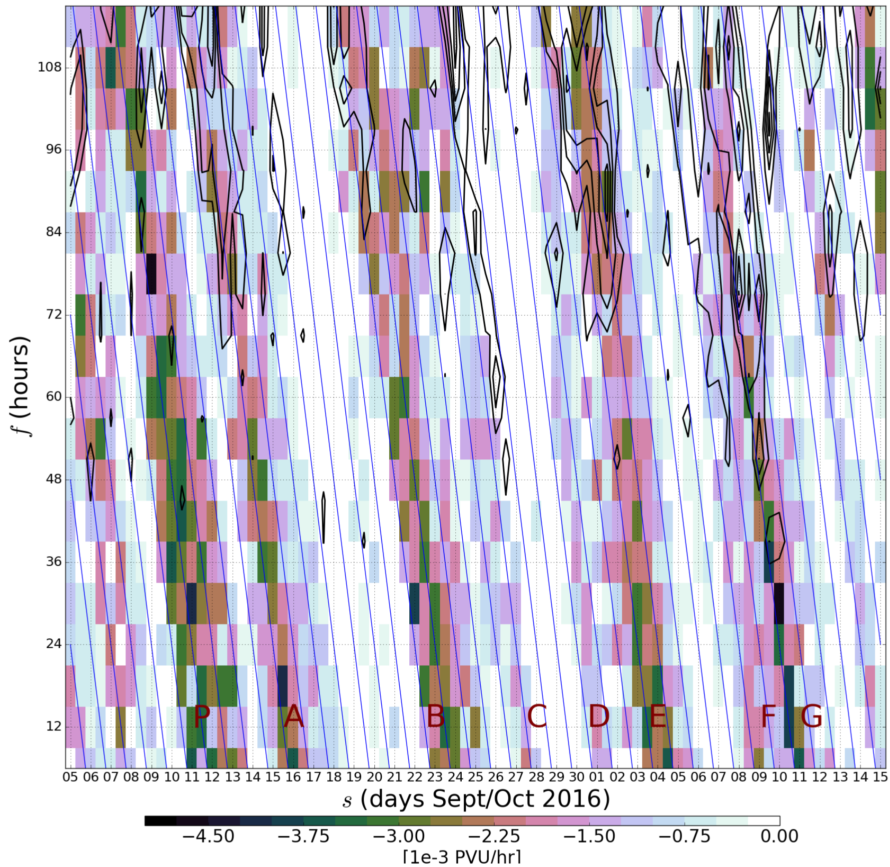


**FIGURE 7** $< DIAA >$ coloured and $< \partial_f E >$ contoured. Similar layout as for Figure 1, but extending up to 120 hours on the $f$ axis. Contours of $< \partial_f E >$ start from 20 $m/day$ with interval 10 $m/day$.

associated with a WCB in a larger cyclone. Grams et al. (2018) discussed the misrepresentation of a PV cut-off as a precursor to a sensitive situation. Figure 7 gives some support to the two-stage hypothesis, with four pairs of PB events where the error growth is much faster in the second of the two events, even though both have enhanced $< DIAA >$: **P–A**, **B–C**, **D–E**, **F–G**.

To bring out the relationship between $< DIAA >$ and $\partial_f E$, their averages across different forecasts for the same validation times (denoted by $||_t$) are shown in Figure 8. The results are contrasted with those using the geostrophic forcing of ageostrophic advection term $< SGAA >$. Both AA variables (response to diabatic and geostrophic forcing) have been averaged over short lead times $f = 6$ to $f = 60$ hours. Longer lead times are excluded because synoptic-scale features are likely to have different structures and positions moving into the medium range and the aim is to capture the diabatic influence on the observed event. In contrast, the error growth $\partial_f E$ has been averaged across all forecast lead times because the differences between forecast trajectories on synoptic scales only emerge after two days lead time and continue to grow. Similar results are found if averaged only beyond $f = 60$ hours (not shown). As anticipated, there is a strong anti-correlation between $< DIAA >$ and error growth rate as a function of validation date. However, in some of the labelled events, the validation time with strongest error growth lags the strongest diabatic influence by 12-24 hours. This lag is most marked for PB **C** and **G** (as can be seen in Fig. 7). The relationships between time series are quantified using Pearson correlation coefficients shown in Table 5. The negative correlation of the $|$<DIAA>$|_t$ and $|\partial_f E|_t$ time series is significant at the 99% level, whereas the $|$<SGAA>$|_t$ timeseries varies between negative and positive values and has no significant correlation with either the $\partial_f E$ or $< DIAA >$ time series.

The time-average $|$<DIAA>$|_t$ is systematically negative, estimated at $-0.97 \times 10^{-3}$ PVU/hr with a confidence interval of $(-0.89, -1.05) \times 10^{-3}$ PVU/hr, computed using a bootstrapping technique at 99% significance. In contrast, the time-average of $|$<SGAA>$|_t$ is smaller than the fluctuations from positive to negative. As explained earlier, $< SGAA >$ is expected to integrate approximately to zero over such a large domain, while $< DIAA >$ is negative due to the typical position of large-scale heating immediately ahead of upper-level troughs. The systematically-negative DIAA values imply that diabatic processes have, on average, an influence in ridge building across the NAWDEX period and, as they are correlated to the $|\partial_f E|_t$ timeseries, that they are also associated to the development of $Z500$ errors over the North Atlantic region.

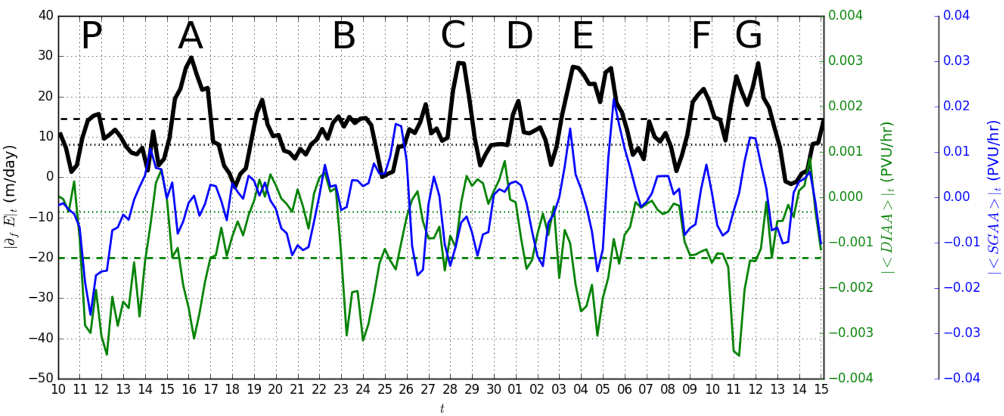


**FIGURE 8** Time series for diagnostics averaged over forecasts with varying lead times ($f$) for fixed validation dates, $t$: $\partial_f E$ (black), $< DIAA >$ (green), and $< SGAA >$ (blue). Only forecasts with $f < 60$ hour are included in the averages for <DIAA> and <SGAA>. Dashed (dotted) thick lines denote the absolute upper (lower) terciles for $|\partial_f E|_t$ (black) and $|$<DIAA>$|_t$ (green).

| Time series A | series B | correlation | P-coefficient |
|---|---|---|---|
| $\lvert\partial_f E\rvert_t$ | $\lvert$<DIAA>$\rvert_t$ | $-0.395$ | $< 0.01$ |
| $\lvert\partial_f E\rvert_t$ | $\lvert$<SGAA>$\rvert_t$ | 0.028 | 0.74 |
| $\lvert$<DIAA>$\rvert_t$ | $\lvert$<SGAA>$\rvert_t$ | 0.121 | 0.15 |

**TABLE 5** Pearson correlation coefficients between $|\partial_f E|_t$, $|$<DIAA>$|_t$ and $|$<SGAA>$|_t$ timeseries shown in Figure 8 and their P-values testing the significance of the correlation.

To quantify the conditional relationship between the $|\partial_f E|_t$ and $|$<DIAA>$|_t$ variables further, dates with $|\partial_f E|_t$ in the upper tercile are classified as "Strong Error Growth" (S-EG) and dates in the lower tercile as "Weak Error Growth" (W-EG). It is seen in Fig. 7 that $< DIAA >$ declines with lead time on the validation dates of certain PB events (e.g. **A**, **B**, **E** or **G**), with stronger values for $f < 60$ hours. Therefore, the conditional dependence of $< DIAA >$ on PBs is considered by calculating three statistics versus lead time: i) the average of $< DIAA >$ across all forecasts, ii) the average across forecasts at validation times associated with S-EG, and iii) the average across times associated with W-EG.

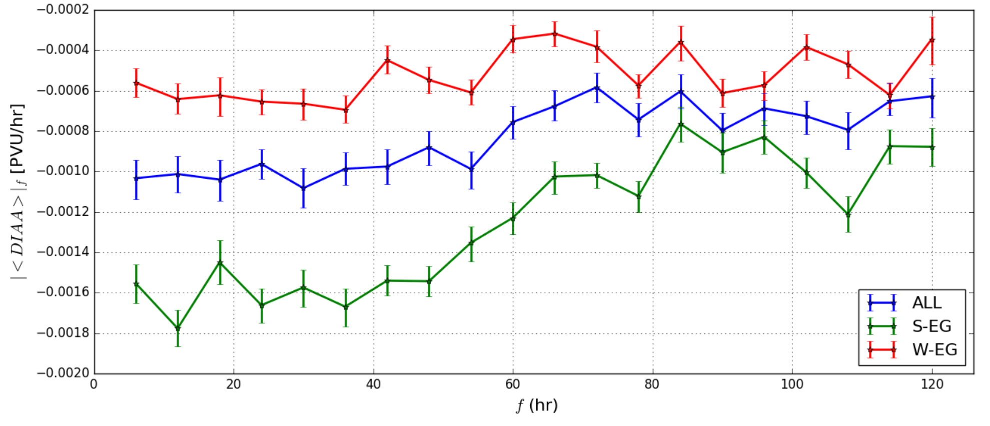


**FIGURE 9** $< DIAA >$ against forecast lead time, $f$, averaged across all validation times (blue), averaged on validation times coinciding with strong error growth activity (S-EG; green) and averaged on validation times coinciding with weak error growth activity (W-EG; red). See text for definitions. Confidence intervals (at 99% significance) are obtained from bootstrapping 1000 samples of $< DIAA >$.

On dates corresponding with weak error growth (W-EG; red line in Figure 9) the diabatic influence $< DIAA >$ is always negative but has no significant dependence on lead time. During S-EG events the diabatic influence $< DIAA >$ is almost three times stronger on average for lead times of less than two days. Intriguingly, the average $< DIAA >$ for S-EG events declines markedly between two and four days lead time. This results in part from the conditional sampling effect mentioned above: if events at analysis time are selected with strong $< DIAA >$ then as the average over forecasts approaches climatology at long lead times, $< DIAA >$ must decrease. However, the unconditional time-average of $< DIAA >$ (blue line) also declines on average over all forecasts by about 1/3 of its magnitude between lead times of two to four days. This reduction suggests that there may be a systematic model error where the diabatic influence on ageostrophic advection of the tropopause is weaker in longer-range forecasts.

The $< DIAA >$ diagnostic has shown that diabatic influence on the tropopause is almost three times stronger on average during PB events, than when PB events are not occurring. However, to make a connection with flow-dependent predictability the statistics of the MOGREPS-G ensemble are considered conditional on the strength of diabatic influence. As before, dates are partitioned into three categories based on error growth rate: S-EG, W-EG and the remaining "neutral" third. A similar approach is used to define dates with strong diabatic influence (S-DI) as those in the top tercile in terms of $< DIAA >$ magnitude, weak diabatic infuence (W-DI) as those in the bottom tercile and the remainder unclassified. The MOGREPS-G ensemble mean error, $E$, and ensemble spread, $\sigma$, are partitioned and then averaged over four different time subsets: validation times coinciding with S-EG, W-EG, S-DI and W-DI. The relationship between the two diagnostics for each of these four subsets is shown in Figure 10. Note that ensemble spread has been re-scaled upwards to account for the small ensemble size, following the argument in Section 2.2 of Leutbecher and Palmer (2008). When the rate of error growth is smaller than average (W-EG) the re-scaled ensemble spread matches ensemble mean error at all lead times, except at the shortest lead time when the spread exceeds the error. This $1:1$ relationship between spread and error is a necessary condition for a reliable ensemble in a probabilistic forecasting sense. However, for the set of validation times corresponding to S-EG events ensemble mean error grows considerably faster than ensemble spread for lead times beyond two days. By definition, error grows faster in the S-EG events, but the important finding is that ensemble spread does not keep pace. It is also clear from Fig. 1 that ensemble spread is on average weaker than error growth for lead times beyond two days (without any conditional sampling).

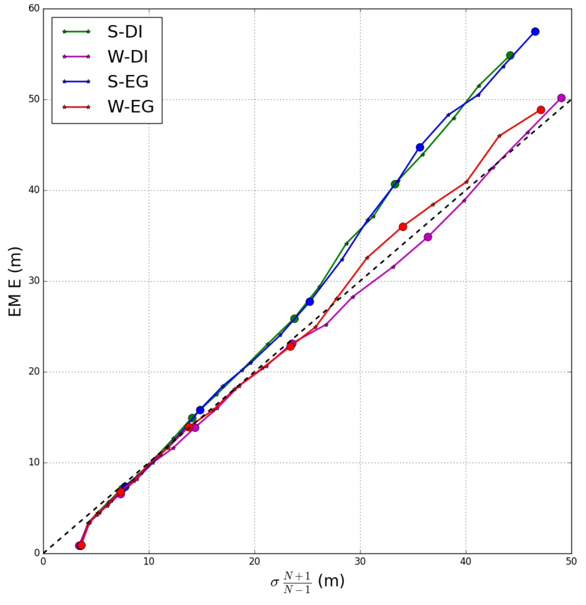


**FIGURE 10** Ensemble mean $Z500$ RMSE, $EME$, versus scaled $Z500$ ensemble spread, $\sigma$, for events with strong $< DIAA >$ (S-DI, green), negligible $< DIAA >$ (W-DI, magenta), strong error growth (S-EG, blue) and weak error growth (W-EG, red). Dots indicate forecast lead time (dots every 24 hours from $f = 0$). The classifications are based on the upper (lower) terciles shown in Figure 8. Dashed black line is the $1:1$ line.

If instead the statistics conditional on diabatic influence are considered, it is found that the spread matches ensemble mean error when there is weak diabatic influence (W-DI), but error grows much faster than spread for the S-DI events. This is anticipated due to the strong negative correlation between $< DIAA >$ and $\partial_f E$. For S-DI events the average rate of error growth is approximately a factor of $4/3$ larger than rate of ensemble spread at times when $E$ grows fastest. Interestingly, the divergence between the S-DI and W-DI ensemble statistics in Fig. 10 begins after $f = 48$ hours (the third dot near the 15 m error and spread intersection), which coincides with the start of decay in the magnitude of $|\text{<DIAA>}|_f$ shown in Figure 9 even when averaged over all forecasts. It suggests that the ensemble is under-dispersive only in situations when there is strong diabatic influence on the tropopause in the first two days. The faster rate of error growth during S-DI events, and decline of diabatic influence in forecasts, together suggest that mis-representation of diabatic processes in forecast models may be contributing to error, but is not captured in the design of the ensemble. Rodwell et al. (2018) also found, using the ECMWF ensemble of data assimilations and forecasts, that initial ensemble spread and growth rate early in forecasts are too small in the vicinity of WCBs, relative to analysis error and forecast error growth rate. The evidence therefore supports the NAWDEX Hypothesis (see Introduction): situations with lowest predictability, and also lowest predictive skill with current forecast systems, are associated with events when diabatic processes are highly active, such as WCBs or recurving tropical cyclones, with pronounced influence on upper-tropospheric divergence in response to diabatic heating.

# 4 | CONCLUSIONS

The role of diabatic processes in forecast busts, where forecast errors grow much more rapidly than usual, has been previously postulated based on case studies and composites of the worst performing forecasts. For example, Rodwell et al. (2013) used such a composite approach to identify that a common precursor Rossby wave pattern was associated with forecast busts over Central Europe five to six days into the forecast. Subsequently, Rodwell et al. (2018) has shown that the analysis uncertainty (from ensemble data assimilation) in such situations is greatest in the upper troposphere in a ridge extending polewards across eastern USA, in the region where diabatic processes are very active in mesoscale convective systems. Studies on the extratropical transition of tropical cyclones have also shown how predictability downstream can be lower than normal as the TC approaches the jet stream. The role of diabatic processes in ascending air masses, and particularly their influence on divergent outflow and expansion of the ridge by advecting the tropopause further away, has been identified as a key process introducing uncertainty into the forecasts (Grams and Archambault, 2016). More generally, Baumgart et al. (2019) have identified four characteristic stages of error growth in the mid-latitudes, the second stage involving advection of the tropopause by divergent wind. However, the hypothesis, posed by NAWDEX (Schäfler et al., 2018), that diabatic processes affect large-scale predictability downstream through enhancing divergent outflow in ridges, advecting the tropopause and generating Rossby wave perturbations, has not been tested systematically over many forecasts.

This study set out to establish the influence of diabatic processes on mid-latitude predictability in three steps, outlined in the Introduction. The conclusions from each of these steps are summarised here.

**A. Evidence for flow-dependent predictability using operational forecasts during the NAWDEX period**

- Flow-dependent predictability was identified with a strong dependence of the rate of ensemble spread and forecast error growth on particular dates, irrespective of lead time, all characterised by expanding large-scale tropopause ridges early in the forecasts.
- The same events, termed "predictability barriers", are identified with strong error growth (within the upper tercile) in both the Met Office and ECMWF forecast systems.
- On average during PB events, ensemble mean forecast error grows faster than ensemble spread by a factor of $4/3$ beyond two-days lead time. Ensemble spread matches ensemble mean forecast error on days without PB events.

**B. Quantification of the diabatic influence on the balanced flow through the ageostrophic advection of PV mechanism**

- A new semi-geotriptic (SGT) balance tool (Cullen, 2018) was used to estimate the balanced ageostrophic flow at tropopause level and to attribute it as a response to geostrophic or diabatic forcing. The model is appropriate in the situation of the large horizontal static stability contrast where horizontal surfaces intersect the tropopause (outside the regime of validity for the QG model).

- In SGT dynamics the Ertel PV is approximately conserved, and the advecting velocity is well approximated by the geostrophic and ageostrophic winds deduced as functions of pressure. This property was used to introduce two new diagnostics: PV advection by the ageostrophic velocity attributed to geostrophic forcing (SGAA) or attributed to diabatic forcing (DIAA). These diagnostics are considered at tropopause level due to the influence of PV advection there on Rossby wave propagation and downstream error growth. Other ways in which diabatic processes influence the balanced flow have not been investigated.
- It was discovered that the domain average of DIAA is always negative because the horizontal ageostrophic wind vectors forced by heating point from the troposphere towards the stratosphere, resulting in advection of the tropopause and ridge expansion. Therefore, DIAA can be used as a measure of the magnitude of "indirect diabatic influence on the tropopause" as distinct from the direct diabatic impact on PV at the tropopause which has been demonstrated to be small (Chagnon et al., 2013; Chagnon and Gray, 2015).

**C. Situations with lowest predictability are associated with strong diabatic influence on tropopause advection**

- DIAA and error growth rate are strongly correlated, although in some PB events (two out of eight) the rapid error growth occurs at validation times 12–24 hours after the maximum DIAA.
- Although the average magnitude of DIAA is much less than SGAA, only DIAA is correlated with the predictability barriers. SGAA fluctuates about zero and is uncorrelated with PBs.
- DIAA is almost three times larger in strong PB events, compared with dates with weak error growth.
- The PB events identified during NAWDEX were associated with strong DIAA on the western flanks of developing tropopause ridges (see Figs. 5 and 6).

Diabatic influence on the ageostrophic advection of the tropopause (DIAA) declines between two and four days lead time when averaged over all Met Office high resolution global forecasts. A lead time dependence should not be expected for a perfect forecast system averaged over many events. For strong DI events, the error growth rate is found to exceed the rate of ensemble spread for lead times beyond two days, while spread matches error on average over weaker DI events. Taken together this indicates that part of the excess error growth (relative to ensemble spread) may result from model error in the representation of diabatic processes. A number of distinct reasons may explain this finding:

**I** Under-representation of the magnitude of heating arising from physics parametrizations;
**II** Misrepresentation of the position of heating relative to large-scale shear and position of the tropopause, which would change the influence of heating on PV through non-advective fluxes (Harvey et al., 2020) and ageostrophic advection;
**III** Misrepresentation of the response of the balanced flow to heating;
**IV** Coupled moist dynamics that is not described by balance; and
**V** Design of the ensemble initial perturbations may not adequately reflect sensitivity to moisture, latent heat release and other physical processes.

In conclusion, results shown here provide evidence consistent with the NAWDEX hypothesis: predictability barriers exist that are associated with diabatic influence on ageostrophic wind in upper-tropospheric outflow, and the tendency for this wind to build ridges by advecting the tropopause on the western flank of the ridges. The latent heating occurs within the warm conveyor belts of cyclones in situations where meridional tropopause displacements are growing rapidly. Therefore the sensitivity is connected to the second stage of error growth identified by Baumgart et al. (2019).

This study is unable to answer why DIAA declines on average for forecast lead times beyond two days, and whether this decline can be attributed to the model representation of physical processes. Further research is needed to understand the model processes active during PB events and the ways in which they might contribute to model error. It is also found that forecast error grows much faster than ensemble spread in flow configurations with strong DIAA, in a forecasting system that is otherwise well calibrated. However, the reasons why this occurs have not been determined. Specific model experiments would be required to establish whether the diabatic processes cause the error growth and whether such errors could be reduced by improving the model. There are some tools that have been useful to diagnose specific sources of model error, such as diabatic increments over backward Lagrangian trajectories in WCBs (Joos and Wernli, 2012; Joos and Forbes, 2016), passive tracers accumulating PV, moisture and potential temperature tendencies from physical parametrizations (Chagnon et al., 2013; Martínez-Alvarado and Plant, 2014; Martínez-Alvarado et al., 2014; Saffin et al., 2016), or The initial uncertainty growth rate diagnostic of Rodwell et al. (2018). The combination of those tools with NAWDEX observations could shed some light into the representation of diabatic processes pointing the way towards model improvements. Alternatively, ensemble design might be improved to capture a greater rate of ensemble trajectory separation in situations where diabatic influence is strong, for example through flow-dependent stochastic parametrization more closely linked to the physical processes in play than used operationally at present (e.g. Clarke et al. 2019).

## REFERENCES

Abel, S. J. and Boutle, I. A. (2012) An improved representation of the raindrop size distribution for single-moment microphysics schemes. *Quarterly Journal of the Royal Meteorological Society*, **138**, 2151–2162. URL: `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.1949`.

Agustí-Panareda, A., Gray, S. L., Craig, G. C. and Thorncroft, C. (2005) The extratropical transition of tropical cyclone lili (1996) and its crucial contribution to a moderate extratropical development. *Monthly Weather Review*, **133**, 1562–1573. URL: `https://doi.org/10.1175/MWR2935.1`.

Archambault, H. M., Bosart, L. F., Keyser, D. and Cordeira, J. M. (2013) A climatological analysis of the extratropical flow response to recurving western north pacific tropical cyclones. *Monthly Weather Review*, **141**, 2325–2346. URL: `https://doi.org/10.1175/MWR-D-12-00257.1`.

Archambault, H. M., Keyser, D., Bosart, L. F., Davis, C. A. and Cordeira, J. M. (2015) A composite perspective of the extratropical flow response to recurving western north pacific tropical cyclones. *Monthly Weather Review*, **143**, 1122–1141. URL: `https://doi.org/10.1175/MWR-D-14-00270.1`.

Balsamo, G., Boussetta, S., Dutra, E., Beljaars, A., Viterbo, P. and van den Hurk, B. (2011) Evolution of land-surface processes in the ifs. *ECMWF Newsletter*, 17–22. URL: `https://www.ecmwf.int/node/17444`.

Bauer, P., Thorpe, A. and Brunet, G. (2015) The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55.

Baumgart, M., Ghinassi, P., Wirth, V., Selz, T., Craig, G. and Riemer, M. (2019) Quantitative view on the processes governing the upscale error growth up to the planetary scale using a stochastic convection scheme. *Monthly Weather Review*, **147**, 1713–1731.

Baumgart, M. and Riemer, M. (2019) Processes governing the amplification of ensemble spread in a medium-range forecast with large forecast uncertainty. *Quarterly Journal of the Royal Meteorological Society*, Early online, doi:10.1002/qj.3617.

Baumgart, M., Riemer, M., Wirth, V., Teubler, F. and Lang, S. (2018) Potential vorticity dynamics of forecast errors: A quantitative case study. *Monthly Weather Review*, **146**, 1405–1425.

Beare, R. J. and Cullen, M. J. P. (2010) A semi-geostrophic model incorporating well-mixed boundary layers. *Quarterly Journal of the Royal Meteorological Society*, **136**, 906–917. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.612.

Bechtold, P., Köhler, M., Jung, T., Doblas-Reyes, F., Leutbecher, M., Rodwell, M. J., Vitart, F. and Balsamo, G. (2008) Advances in simulating atmospheric variability with the ecmwf model: From synoptic to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*, **134**, 1337–1351. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.289.

Bechtold, P., Semane, N., Lopez, P., Chaboureau, J.-P., Beljaars, A. and Bormann, N. (2014) Representing equilibrium and nonequilibrium convection in large-scale models. *Journal of the Atmospheric Sciences*, **71**, 734–753. URL: https://doi.org/10.1175/JAS-D-13-0163.1.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B. and Harding, R. J. (2011) The joint uk land environment simulator (jules), model description - part 1: Energy and water fluxes. *Geoscientific Model Development*, **4**, 677–699. URL: https://www.geosci-model-dev.net/4/677/2011/.

Bishop, C. H., Etherton, B. J. and Majumdar, S. J. (2001) Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Monthly Weather Review*, **129**, 420–436. URL: https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

Bowler, N. E., Arribas, A., Mylne, K. R., Robertson, K. B. and Beare, S. E. (2008) The mogreps short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, **134**, 703–722. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.234.

Chagnon, J. M. and Gray, S. L. (2015) A diabatically-generated potential vorticity structure near the extratropical tropopause in three simulated extratropical cyclones. *Mon. Weather Rev.*, **143**, 2337–2347.

Chagnon, J. M., Gray, S. L. and Methven, J. (2013) Diabatic processes modifying potential vorticity in a north atlantic cyclone. *Quarterly Journal of the Royal Meteorological Society*, **139**, 1270–1282. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2037.

Clarke, S. J., Gray, S. L. and Roberts, N. M. (2019) Downstream influence of mesoscale convective systems. part 2: Influence on ensemble forecast skill and spread. *Quarterly Journal of the Royal Meteorological Society*, **145**, 2953–2972. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3613.

Cullen, M. (2018) The use of semigeostrophic theory to diagnose the behaviour of an atmospheric gcm. *Fluids*, **3**. URL: https://www.mdpi.com/2311-5521/3/4/72.

Davies, H. C. (2015) The quasigeostrophic omega equation: Reappraisal, refinements, and relevance. *Monthly Weather Review*, **143**, 3–25. URL: https://doi.org/10.1175/MWR-D-14-00098.1.

Davis, C. A., Grell, E. D. and Shapiro, M. A. (1996) The balanced dynamical nature of a rapidly intensifying oceanic cyclone. *Monthly Weather Review*, **124**, 3–26. URL: https://doi.org/10.1175/1520-0493(1996)124<0003:TBDNOA>2.0.CO;2.

ECMWF (2016a) *Part II: Data Assimilation*. No. 2 in IFS Documentation. ECMWF. URL: https://www.ecmwf.int/node/16666.

— (2016b) *Part IV: Physical Processes*. No. 4 in IFS Documentation. ECMWF. URL: https://www.ecmwf.int/node/16648.

Ferranti, L., Corti, S. and Janousek, M. (2015) Flow-dependent verification of the ecmwf ensemble over the euro-atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, **141**, 916–924. URL: https://doi.org/10.1002/qj.2411.

Frame, T., Ambaum, M., Gray, S. and Methven, J. (2011) Ensemble prediction of transitions of the North Atlantic eddy driven jet. *Quarterly Journal of the Royal Meteorological Society*, **137**, 1288–1297.

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. Wea. Rev.*, **144**, 3847–3869.

Grams, C. M. and Blumer, S. R. (2015) European high-impact weather caused by the downstream response to the extratropical transition of north atlantic hurricane katia (2011). *Geophysical Research Letters*, **42**, 8738–8748. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015GL066253`.

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. *Quarterly Journal of the Royal Meteorological Society*, **144**, 2577–2591. URL: `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3353`.

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. *Q. J. Roy. Meteor. Soc.*, **137**, 2174–2193.

Grazzini, F. and Vitart, F. (2015) Atmospheric predictability and rossby wave packets. *Quarterly Journal of the Royal Meteorological Society*, **141**, 2793–2802. URL: `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2564`.

Gregory, D. and Rowntree, P. R. (1990) A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Monthly Weather Review*, **118**, 1483–1506. URL: https://doi.org/10.1175/1520-0493(1990)118<1483:AMFCSW>2.0.CO;2.

Harvey, B., Methven, J., Sanchez, C. and Schäfler (2020) Diabatic generation of negative potential vorticity and its impact on the North Atlantic jet stream. *Quarterly Journal of the Royal Meteorological Society*, accepted.

Heifetz, E., Bishop, C., Hoskins, B. and Methven, J. (2004) The CRW perspective on baroclinic instability. Part I: Mathematical basis. *Quarterly Journal of the Royal Meteorological Society*, **130**, 211–231.

Hill, P. G., Manners, J. and Petch, J. C. (2011) Reducing noise associated with the monte carlo independent column approximation for weather forecasting models. *Quarterly Journal of the Royal Meteorological Society*, **137**, 219–228. URL: `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.732`.

Hoskins, B. and Bretherton, F. (1972) Atmospheric frontogenesis models: Mathematical formulation and solution. *Journal of the Atmospheric Sciences*, **29**, 11–37.

Hoskins, B. and Draghici, I. (1977) The forcing of ageostrophic motion according to the semi-geostrophic equations and in an isentropic coordinate model. *Journal of the Atmospheric Sciences*, **34**, 1859–1867.

Hoskins, B. J. (1975) The geostrophic momentum approximation and the semi-geostrophic equations. *Journal of the Atmospheric Sciences*, **32**, 233–242. URL: https://doi.org/10.1175/1520-0469(1975)032<0233:TGMAAT>2.0.CO;2.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A. and Collins, W. D. (2008) Radiative forcing by long-lived greenhouse gases: Calculations with the aer radiative transfer models. *Journal of Geophysical Research: Atmospheres*, **113**. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JD009944`.

Joos, H. and Forbes, R. M. (2016) Impact of different ifs microphysics on a warm conveyor belt and the downstream flow evolution. *Quarterly Journal of the Royal Meteorological Society*, **142**, 2727–2739. URL: `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2863`.

Joos, H. and Wernli, H. (2012) Influence of microphysical processes on the potential vorticity development in a warm conveyor belt: a case-study with the limited-area model cosmo. *Quarterly Journal of the Royal Meteorological Society*, **138**, 407–418. URL: `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.934`.

Judt, F. (2018) Insights into atmospheric predictability through global convection-permitting model simulations. *Journal of the Atmospheric Sciences*, **75**, 1477–1497.

Keller, J. H., Grams, C. M., Riemer, M., Archambault, H. M., Bosart, L., Doyle, J. D., Evans, J. L., Galarneau, T. J., Griffin, K., Harr, P. A., Kitabatake, N., McTaggart-Cowan, R., Pantillon, F., Quinting, J. F., Reynolds, C. A., Ritchie, E. A., Torn, R. D. and Zhang, F. (2019) The extratropical transition of tropical cyclones. part ii: Interaction with the midlatitude flow, downstream impacts, and implications for predictability. *Monthly Weather Review*, **147**, 1077–1106.

Leutbecher, M. and Palmer, T. (2008) Ensemble forecasting. *Journal of Computational Physics*, **227**, 3515 – 3539. URL: http://www.sciencedirect.com/science/article/pii/S0021999107000812. Predicting weather, climate and extreme events.

Lock, A. P., Brown, A. R., Bush, M. R., Martin, G. M. and Smith, R. N. B. (2000) A new boundary layer mixing scheme. part i: Scheme description and single-column model tests. *Monthly Weather Review*, **128**, 3187–3199. URL: https://doi.org/10.1175/1520-0493(2000)128<3187:ANBLMS>2.0.CO;2.

Lott, F. and Miller, M. J. (1997) A new subgrid-scale orographic drag parametrization: Its formulation and testing. *Quarterly Journal of the Royal Meteorological Society*, **123**, 101–127. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712353704.

Maddison, J. W., Gray, S. L., Martínez-Alvarado, O. and Williams, K. D. (2019) Upstream cyclone influence on the predictability of block onsets over the euro-atlantic region. *Monthly Weather Review*, **147**, 1277–1296. URL: https://doi.org/10.1175/MWR-D-18-0226.1.

Maddison, J. W., Gray, S. L., Martinez-Alvarado, O. 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*. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3739.

Malardel, S., Wedi, N., Deconinck, W., Diamantakis, M., Kuehnlein, C., Mozdzynski, G., Hamrud, M. and Smolarkiewicz, P. (2016) A new grid for the ifs. *ECMWF Newsletter*, 23–28. URL: https://www.ecmwf.int/node/17262.

Manners, J., Vosper, S. B. and Roberts, N. (2012) Radiative transfer over resolved topographic features for high-resolution weather prediction. *Quarterly Journal of the Royal Meteorological Society*, **138**, 720–733. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.956.

Martínez-Alvarado, O., Joos, H., Chagnon, J., Boettcher, M., Gray, S. L., Plant, R. S., Methven, J. and Wernli, H. (2014) The dichotomous structure of the warm conveyor belt. *Quarterly Journal of the Royal Meteorological Society*, **140**, 1809–1824. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2276.

Martínez-Alvarado, O., Madonna, E., Gray, S. L. and Joos, H. (2016) A route to systematic error in forecasts of rossby waves. *Quarterly Journal of the Royal Meteorological Society*, **142**, 196–210. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2645.

Martínez-Alvarado, O. and Plant, R. S. (2014) Parametrized diabatic processes in numerical simulations of an extratropical cyclone. *Quarterly Journal of the Royal Meteorological Society*, **140**, 1742–1755. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2254.

Morcrette, J.-J., Barker, H. W., Cole, J. N. S., Iacono, M. J. and Pincus, R. (2008) Impact of a new radiation package, mcrad, in the ecmwf integrated forecasting system. *Monthly Weather Review*, **136**, 4773–4798. URL: https://doi.org/10.1175/2008MWR2363.1.

Oertel, A., Boettcher, M., Joos, H., Sprenger, M., Konow, H., Hagen, M. and Wernli, H. (2019) Convective activity in an extratropical cyclone and its warm conveyor belt-a case-study combining observations and a convection-permitting model simulation. *Quarterly Journal of the Royal Meteorological Society*, **145**, 1406–1426. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3500.

Orr, A., Bechtold, P., Scinocca, J., Ern, M. and Janiskova, M. (2010) Improved middle atmosphere climate and forecasts in the ecmwf model through a nonorographic gravity wave drag parameterization. *Journal of Climate*, **23**, 5905–5926. URL: https://doi.org/10.1175/2010JCLI3490.1.

Riemer, M., Baumgart, M. and Eiermann, S. (2014) Cyclogenesis downstream of extratropical transition analyzed by q-vector partitioning based on flow geometry. *Journal of the Atmospheric Sciences*, **71**, 4204–4220. URL: https://doi.org/10.1175/JAS-D-14-0023.1.

Rodwell, M., Richardson, D., Parsons, D. and Wernli, H. (2018) Flow-dependent reliability: A path to more skillful ensemble forecasts. *Bulletin of the American Meteorological Society*, **99**, 1015–1026.

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

Saffin, L., Methven, J. and Gray, S. L. (2016) The non-conservation of potential vorticity by a dynamical core compared with the effects of parametrized physical processes. *Quarterly Journal of the Royal Meteorological Society*, **142**, 1265–1275. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2729.

Scaife, A. A., Butchart, N., Warner, C. D. and Swinbank, R. (2002) Impact of a spectral gravity wave parameterization on the stratosphere in the met office unified model. *Journal of the Atmospheric Sciences*, **59**, 1473–1489. URL: https://doi.org/10.1175/1520-0469(2002)059<1473:IOASGW>2.0.CO;2.

Schäfler, A., Craig, G., Wernli, H., Arbogast, P., Doyle, J. D., McTaggart-Cowan, R., Methven, J., Rivière, G., Ament, F., Boettcher, M., Bramberger, M., Cazenave, Q., Cotton, R., Crewell, S., Delanoë, J., Dörnbrack, A., Ehrlich, A., Ewald, F., Fix, A., Grams, C. M., Gray, S. L., Grob, H., Groß, S., Hagen, M., Harvey, B., Hirsch, L., Jacob, M., Kölling, T., Konow, H., Lemmerz, C., Lux, O., Magnusson, L., Mayer, B., Mech, M., Moore, R., Pelon, J., Quinting, J., Rahm, S., Rapp, M., Rautenhaus, M., Reitebuch, O., Reynolds, C. A., Sodemann, H., Spengler, T., Vaughan, G., Wendisch, M., Wirth, M., Witschas, B., Wolf, K. and Zinner, T. (2018) The North Atlantic Waveguide and Downstream Impact Experiment. *Bulletin of the American Meteorological Society*, **99**, 1607–1637.

Schemm, S., Wernli, H. and Papritz, L. (2013) Warm conveyor belts in idealized moist baroclinic wave simulations. *Journal of the Atmospheric Sciences*, **70**, 627–652. URL: https://doi.org/10.1175/JAS-D-12-0147.1.

Schindler, M., Weissmann, M., Schäfler, A. and Radnoti, G. (2020) The impact of dropsonde and extra radiosonde observations during nawdex in autumn 2016. *Monthly Weather Review*, **148**, 809–824. URL: https://doi.org/10.1175/MWR-D-19-0126.1.

Selz, T. and Craig, G. (2015) Upscale error growth in a high-resolution simulation of a summertime weather event over Europe. *Monthly Weather Review*, **143**, 813–827.

Tennant, W. J., Shutts, G. J., Arribas, A. and Thompson, S. A. (2011) Using a stochastic kinetic energy backscatter scheme to improve mogreps probabilistic forecast skill. *Monthly Weather Review*, **139**, 1190–1206. URL: https://doi.org/10.1175/2010MWR3430.1.

Teubler, F. and Riemer, M. (2016) Dynamics of rossby wave packets in a quantitative potential vorticity-potential temperature framework. *Journal of the Atmospheric Sciences*, **73**, 1063–1081. URL: https://doi.org/10.1175/JAS-D-15-0162.1.

Tiedtke, M. (1989) A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, **117**, 1779–1800. URL: https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2.

— (1993) Representation of clouds in large-scale models. *Monthly Weather Review*, **121**, 3040–3061. URL: https://doi.org/10.1175/1520-0493(1993)121<3040:ROCILS>2.0.CO;2.

Vallis, G. (2006) *Atmospheric and oceanic fluid dynamics*. Cambridge University Press. 745pp.

Vosper, S. B. (2015) Mountain waves and wakes generated by south georgia: implications for drag parametrization. *Quarterly Journal of the Royal Meteorological Society*, **141**, 2813–2827. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2566.

Walters, D., Brooks, M., Boutle, I., 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. *Geosci. Model Dev.*, **10**, 1487–1520.

Wilson, D. R. and Ballard, S. P. (1999) A microphysically based precipitation scheme for the uk meteorological office unified model. *Quarterly Journal of the Royal Meteorological Society*, **125**, 1607–1636. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712555707.

Wilson, D. R., Bushell, A. C., Kerr-Munslow, A. M., Price, J. D. and Morcrette, C. J. (2008) Pc2: A prognostic cloud fraction and condensation scheme. i: Scheme description. *Quarterly Journal of the Royal Meteorological Society*, **134**, 2093–2107. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.333.

Wood, N., Staniforth, A., White, A., Allen, T., Diamantakis, M., Gross, M., Melvin, T., Smith, C., Vosper, S., Zerroukat, M. and Thuburn, J. (2014) An inherently mass-conserving semi-implicit semi-lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quarterly Journal of the Royal Meteorological Society*, **140**, 1505–1520. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2235.

Woollings, T., Hoskins, B., Blackburn, M. and Berrisford, P. (2008) A new rossby wave-breaking interpretation of the north atlantic oscillation. *Journal of the Atmospheric Sciences*, **65**, 609–626. URL: https://doi.org/10.1175/2007JAS2347.1.

Zagar, N., Horvat, M., Zaplotnik, Z. and Magnusson, L. (2017) Scale-dependent estimates of the growth of forecast uncertainties in a global prediction system. *Tellus A*, **69**, 1287492.

Zhang, F., Bei, N., Rotunno, R., Snyder, C. and Epifanio, C. C. (2007) Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *Journal of the Atmospheric Sciences*, **64**, 3579–3594. URL: https://doi.org/10.1175/JAS4028.1.

Zhang, F., Sun, Y., Magnusson, L., Buizza, R., Lin, S., Chen, J. and Emanuel, K. (2019) What is the predictability limit of midlatitude weather? *Journal of the Atmospheric Sciences*, **76**, 1077–1091.