University of Reading

Atmospheric blocking and upper-level Rossby wave forecast skill dependence on model configuration

Article

Accepted Version

Martinez-Alvarado, O. ORCID: https://orcid.org/0000-0002-5285-0379, Maddison, J., Gray, S. ORCID: https://orcid.org/0000-0001-8658-362X and Williams, K. (2018) Atmospheric blocking and upper-level Rossby wave forecast skill dependence on model configuration. Quarterly Journal of the Royal Meteorological Society, 144 (716). pp. 2165-2181. ISSN 1477-870X doi: 10.1002/qj.3326 Available at https://centaur.reading.ac.uk/76959/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1002/qj.3326

Publisher: Royal Meteorological Society

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

University of Reading

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Atmospheric blocking and upper-level Rossby wave forecast skill dependence on model configuration

O. Martínez-Alvarado[a,b], J. W. Maddison[b], S. L. Gray[b] and K. D. Williams[c]

May 3, 2018

[a] National Centre for Atmospheric Science–Atmospheric Physics, United Kingdom
[b] Department of Meteorology, University of Reading, United Kingdom
[c] Met Office, United Kingdom

Abstract

Weather models differ in their ability to forecast, at medium range, atmospheric blocking and the associated structure of upper-level Rossby waves. Here, we evaluate the effect of a model's dynamical core on such forecasts. Operational forecasts from the ensemble prediction systems (EPSs) of the European Centre for Medium-range Weather Forecasts (ECMWF), the Met Office (MO) and the Korean Meteorological Administration (KMA) are used. Northern hemisphere model output is analysed from winters before and after a major upgrade to the dynamical core of the MO-EPS. The KMA-EPS acts as a control as it uses the same model as the MO-EPS, but used the older dynamical core throughout. The confounding factor of resolution differences between the MO-EPS and the KMA-EPS is assessed using a MO forecast model hindcast experiment with the more recent dynamical core, but the operational resolution of the KMA-EPS. The introduction of the new dynamical core in the MO-EPS has led to increased forecast blocking frequency, at lead times of five and seven days, counteracting the typically-observed reduction in blocking frequency with lead time. Hit rates of blocking activity, onset and decay are also increased in the main blocking regions (without a corresponding increase in false positive rate). The previously-found reduction of upper-level ridge area and tropopause sharpness (measured by isentropic potential vorticity gradient) with lead time is also reduced with the new dynamical core. This dynamical core improvement (associated with a reduction in implicit damping) is thus demonstrated to be at least as effective as operational resolution improvements in improving forecasts of upper-level Rossby waves and associated blocking.

Keywords

Operational ensemble prediction systems; predictive skill; atmospheric blocking; Rossby waves

1 Introduction

Atmospheric blocks are nearly stationary large-scale weather patterns that effectively redirect (or block) mobile cyclones. They are often associated with a large-amplitude, synoptic-scale, quasi-stationary anticyclone in the extratropics. This phenomenon has a strong influence on mid-latitude weather as it can lead to high-impact weather events, locally and downstream, due to its scale and persistence. Despite its importance, atmospheric blocking remains difficult to represent in weather and climate models. For example, Schiemann *et al.* (2017) showed that even though the representation of Euro-Atlantic blocking tends to improve with resolution, models still exhibit large biases, tending to underestimate winter northern European blocking even at a relatively fine 25-km grid spacing. The purpose of this study is to evaluate the impact of a model's dynamical core on its representation of blocking, specifically a change of dynamical core leading to a reduction in implicit damping.

Medium-range forecasts have demonstrated skill in predicting aspects of blocking for more than a decade. For example, Pelly and Hoskins (2003) found skill in the European Centre for

Medium-Range Weather Forecasts (ECMWF) Ensemble Prediction System (EPS) predictions of (instantaneous) blocking and blocking episodes (with persistence of at least four days) out to 10 days over the Euro-Atlantic sector. They also showed that control forecasts remain skilful for block onset relative to climatology until day five of the forecast in the ECMWF-EPS and concluded that onsets are harder to predict than the decay of blocking. After classifying ECMWF-EPS Euro-Atlantic sector forecasts by weather regimes, Ferranti *et al.* (2015) found that blocking leads to the least accurate forecasts, with an underestimation of blocking persistence and large ensemble spread in forecasts initiating blocking as well as difficulties in the prediction of the transition to blocking (in agreement with Pelly and Hoskins (2003) and Tibaldi and Molteni (1990)). Matsueda (2009) found that blocking frequency tended to be underestimated by ensemble forecasts from several operational centres beyond a lead time of five days in winters (December–February: DJF) between 2006/07 and 2009/10. Using single-member hindcasts from the NCEP Climate Forecast System version 2, Jia *et al.* (2014) found skilful forecasts of wintertime blocking activity in the Northern Hemisphere (NH) at lead times up to nine and seven days over the Euro-Atlantic and Pacific sectors, respectively, but less skill in blocking onset and decay.

Upper-level Rossby waves, manifest in the strong potential vorticity (PV) gradient region along the extratropical (dynamical) tropopause, are associated with mid-latitude tropospheric cyclones and anticyclones. Therefore, Rossby waves greatly influence the generation of mid-latitude weather. Through analysis of seven winters from 2006/07–2012/13 in three EPSs, Gray *et al.* (2014) found systematic forecast errors in the structure of Rossby waves in terms of a reduction in Rossby-wave amplitude and tropopause sharpness with lead time. In agreement with those results, Giannakaki and Martius (2016) found systematic errors in the area and strength of Rossby waveguides, defined as long and narrow bands of strong isentropic PV gradient, in ECMWF forecasts compared to reanalyses. There is a causal relationship between diabatic processes and ridge development (Davis *et al.*, 1993). Latent heat release is known to have an effect on ridge building by advection of low-PV air into the ridge, which enhances the divergent flow at upper levels (Riemer and Jones, 2010) and 'tropopause uplifting' (Bosart and Lackmann, 1995). Furthermore, it has been shown that the reduction in Rossby-wave amplitude is linked to diabatic processes through errors in forecasts of warm conveyor belts (Martínez-Alvarado *et al.*, 2016). Harvey *et al.* (2016) has also shown that the reduction of isentropic PV gradient can be linked to slower eastward propagation of Rossby waves and to a reduction in Rossby-wave amplitude. The correspondence between errors in the structure of Rossby waves (Gray *et al.*, 2014) and those in the structure of Rossby waveguides (Giannakaki and Martius, 2016), and the relationship of the former with warm conveyor belts (Martínez-Alvarado *et al.*, 2016) suggest that these errors have a direct impact on the synoptic variability of models. However, their impact on the representation of stationary waves is not known.

Atmospheric blocking and the structure of upper-level Rossby waves are not independent of each other. As well as an anomaly in geopotential height, atmospheric blocking can be conceptualised as a negative anomaly in the PV field (Schwierz *et al.*, 2004; Röthlisberger *et al.*,

2016). Negative PV anomalies can, in turn, be viewed as ridges in Rossby-wave structure. The dynamical association between Rossby-wave ridges and the anticyclones that define blocking imply that changes in the numerical representation of Rossby-wave ridges should lead to changes in the representation of atmospheric blocking. Consistent with this relationship, de Vries *et al.* (2013) found that future changes of seasonal atmospheric blocking activity can be explained by changes in the strength and variance of the mean upper-level zonal circulation.

The Met Office (MO) introduced a new dynamical core into its operational weather forecast model (the Unified Model: MetUM) in 2014. The new dynamical core is characterised by a reduction in implicit damping with respect to its predecessor. It has been previously shown that the reduced implicit damping has increased extratropical atmospheric variability in the model as measured, for example, by eddy kinetic energy (Walters *et al.*, 2014, 2017). Consistent with the increase in eddy kinetic energy, the new dynamical core has removed a detected loss in extratropical cyclone intensity with forecast lead time in models based on the previous core (Walters *et al.*, 2017). Mid-latitude cyclones and jets are more intense with the new dynamical core and, occasionally, too intense in comparison with corresponding analyses (Mittermaier *et al.*, 2016; Walters *et al.*, 2017).

The MetUM dynamical core upgrade and the availability of The International Grand Global Ensemble (TIGGE, Bougeault *et al.*, 2010), which is an archive of operational forecasts from several forecast centres including the Met Office from 2006 to date, provides an opportunity to study the effects of the improved representation of extratropical variability and cyclone intensity on large-scale circulation features, using ensemble forecasts from other operational centres as control cases. In this study we focus on atmospheric blocking and the structure of upper-level Rossby waves. Thus, this article aims at answering the following questions: does a reduction in implicit damping in a model's dynamical core change the representation of atmospheric blocking and upper-level Rossby waves?; and, if so, are these changes consistent with the known dynamical link between the two features?

The rest of the article is organised as follows. A description of the new dynamical core features in the MetUM is given in Section 2. The data and methodology are presented in Section 3. The results are presented as two linked studies: (i) an analysis of atmospheric blocking in EPSs (presented in Section 4) and (ii) an analysis of upper-level Rossby waves in EPSs (presented in Section 5). Section 6 summarises and concludes this work.

2 The MetUM dynamical cores

The *NewDynamics* dynamical core (Davies *et al.*, 2005) of the Met Office Unified Model (MetUM) was upgraded to the *ENDGame* (Even Newer Dynamics for General atmospheric modelling of the environment) dynamical core (Wood *et al.*, 2014) in July 2014. Both the NewDynamics and ENDGame dynamical cores use a finite-difference discretisation of the non-hydrostatic deep-atmosphere dynamical equations with semi-implicit, semi-Lagrangian integration schemes (Walters *et al.*, 2017). Moreover, both cores use Arakawa C-grid staggering in the

horizontal (Arakawa and Lamb, 1977) and are terrain-following with a hybrid-height Charney–Phillips (Charney and Phillips, 1953) vertical staggering. The following are the differences between the two cores:

- ENDGame introduces a nested iterative approach for each atmospheric time step reducing the need for off-centring (time weights) used in the calculation of the departure point in the semi-Lagrangian scheme (see e.g. Shutts and Vosper, 2011).
- The special treatment of potential temperature and the continuity equation in NewDynamics are abandoned for a full semi-Lagrangian discretisation of all prognostic variables in ENDGame.
- The horizontal staggering of variables has been modified in ENDGame to avoid solving the Helmholtz equation at the poles.
- The explicit horizontal and targeted diffusion used in NewDynamics are no longer required in ENDGame.

Together with improvement to physical parametrizations and increased resolution, the introduction of ENDGame constitutes the Global Atmosphere 6 (GA6) configuration of the MetUM. Full details on the GA6 configuration and the differences between ENDGame and NewDynamics can be found in Walters *et al.* (2017). As discussed in that paper, these changes in the MetUM have improved the accuracy, scalability and numerical stability of the model. The improvement in extratropical circulation with the GA6 configuration seen in Walters *et al.* (2017) was attributed to the reduced implicit damping with the ENDGame dynamical core rather than a result of physical parametrization improvements.

3 Data and methodology

The operational forecasts used and a bespoke MetUM simulation are presented in Section 3.1, and the diagnostics used to assess atmospheric blocking and the structure of Rossby waves are presented in Section 3.2.

3.1 Model forecast data

3.1.1 Operational forecasts

The present study focuses on four winters, DJF 2012/13–2015/16, motivated by the introduction of ENDGame into the operational version of the Global configuration of the Met Office Global Ensemble Prediction System (MOGREPS-G) (Bowler *et al.*, 2008, 2009; Tennant *et al.*, 2011; Williams *et al.*, 2015) in July 2014. This choice spans two winters before and two winters after the introduction of ENDGame; these periods are hereafter referred to as the NewDynamics and ENDGame eras, respectively

Table 1: Configurations of the ECMWF-EPS (EC), MOGREPS (MO) and the KMA-EPS (KMA) during the four winter seasons under analysis where 4D Var is four-dimensional variational data assimilation, SV is singular vectors, EDA is ensemble of data assimilations and ETKF is ensemble transform Kalman filter.

Season	Horizontal resolution (zonal grid spacing at 50°N)			#vertical levels (top of model)			Initial perturbation method (data assimilation method)			#ensemble members			Length of fcts available (days)		
	EC	MO	KMA	EC	MO	KMA	EC	MO	KMA	EC	MO	KMA	EC	MO	KMA
2012/13	T639 (N320) (40 km)	N216 (60 km)	N320 (40 km)	62 ($\approx 35\ km$)	70 (80 km)	70 (80 km)	SV and EDA (4D Var)	ETKF (Hybrid 4D Var)	ETKF (4D Var)	50	23	23	15	15	10
2013/14				91 ($\approx 75\ km$)	85 (85 km)				ETKF (Hybrid 4D Var)						
2014/15		N400 (32 km)			70 (80 km)						11			7	
2015/16															

Daily 1200 UTC data from three operational EPSs, namely MOGREPS (Bowler *et al.*, 2008, 2009; Tennant *et al.*, 2011; Williams *et al.*, 2015), the ECMWF EPS (Molteni *et al.*, 1996; Buizza *et al.*, 1999), and the KMA-EPS, are used for both studies (blocking and upper-level Rossby waves). The KMA-EPS is included because its underlying numerical model is the MetUM with NewDynamics throughout the period of analysis; this offers an opportunity to compare the two dynamical cores in operational setups. A comparison between the configuration of the three EPSs in terms of horizontal and vertical resolution and the generation of initial ensemble perturbations is presented in Table 1.

The data has been obtained from the TIGGE archive (Park *et al.*, 2008). Geopotential height at 500 hPa (Z_{500}) interpolated onto a regular 2.5° grid is used in the study of atmospheric blocking. PV on the 320-K isentropic surface interpolated onto a regular 1° grid is used in the study of Rossby-wave structure; this particular isentropic surface is often used for the study of Rossby-wave structure and is chosen here because it is the only isentropic surface for which PV is available in the TIGGE archive. Tests using a regular 1° grid (instead of the regular 2.5° grid) for the study of atmospheric blocking showed that the conclusions are not sensitive to the grid resolution within this range (not shown) and the coarser resolution data were used for computational speed.

3.1.2 MetUM experiment

Differences in the representation of atmospheric blocking and upper-level Rossby waves are likely to be affected by differences in resolution as well as differences in the dynamical core. To determine the impact of this confounding factor on the comparison of forecasts from MOGREPS and the KMA-EPS, a single-member MetUM run (hereafter ENDGame-RERUN) was performed for the winter 2013/14 with ENDGame and the associated physical parametrization package at a horizontal resolution of N320 (the same resolution as used by the KMA-EPS) and 70 vertical levels (with model top at 80 km). Forecasts were initiated from the Met Office analyses of the day. The output of this run is compared to the operational control-member forecasts from MOGREPS and the KMA-EPS which were both produced using the NewDynamics dynamical core; the resolution of the ENDGame-RERUN is the same as that used by the KMA-EPS and higher than that used by MOGREPS, allowing a relatively clean diagnosis of the relative roles of resolution and dynamical core differences in the atmospheric evolution.

3.2 Diagnostic methods

3.2.1 Atmospheric blocking

The blocking index proposed by d'Andrea *et al.* (1998) is used in this study, following Matsueda (2009). This index is one-dimensional, instantaneous (no persistence criteria) and based on the meridional gradients of Z_{500}. Whilst this is a simple index and has several limitations, it is known to be capable of identifying basic blocking features (Barriopedro *et al.*, 2010) and is sufficient for a forecast/reanalysis comparison such as this study. The gradient to the south

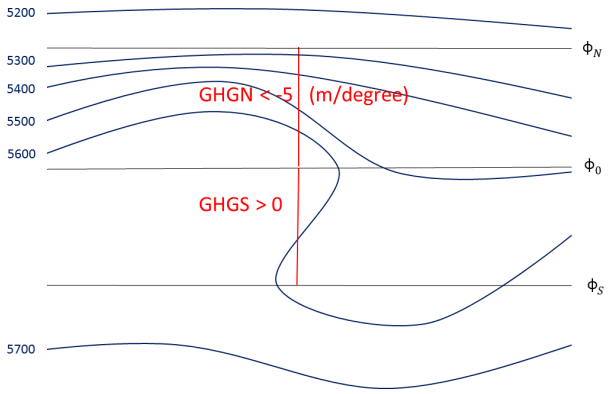

Figure 1: The typical configuration of Z_{500} contours (m) during a blocking event. The red line represents a blocked longitude. The quantities ϕ_n, ϕ_0 and ϕ_s as well as GHGN and GHGS are defined in the text.

(GHGS) and north (GHGN) of a central latitude are defined as follows:

$$\begin{aligned} GHGS &= \frac{Z_{500}(\phi_0) - Z_{500}(\phi_s)}{\phi_0 - \phi_s}, \\ GHGN &= \frac{Z_{500}(\phi_n) - Z_{500}(\phi_0)}{\phi_n - \phi_0}, \end{aligned} \tag{1}$$

where $\phi_n = 77.5°\text{N} \pm \Delta$, $\phi_0 = 60°\text{N} \pm \Delta$, $\phi_s = 40°\text{N} \pm \Delta$, and $\Delta = 0°, 2.5°, 5°$. These Δ values were chosen because the data are interpolated onto a 2.5° grid following Matsueda (2009). The same Δ is used for ϕ_n, ϕ_0 and ϕ_s.

A specific longitude is defined as blocked if (for at least one value of the same Δ) both

$$GHGS > 0, \tag{2}$$

and

$$GHGN < -5 \text{ m } (°)^{-1}. \tag{3}$$

Figure 1 illustrates an example Z_{500} field that satisfies these conditions. The first condition (2) ensures easterly flow to the south of a central blocked latitude, ϕ_0, while the second (3) ensures strong westerly flow to the north. Several studies (e.g. Lejenäs *et al.*, 1983) provide evidence that these conditions are suitable for identifying the characteristics of a blocked situation.

The regions that are most prone to blocking are defined similarly to Matsueda (2009) as the Euro-Atlantic (EA) sector (27.5°W–40°E) and the Pacific (PA) sector (120°E–140°W). To take into account the longitudinal extent of blocking, a sector is then defined as blocked if three or more adjacent longitudinal grid boxes within the sector are blocked on a specific day (also following Matsueda (2009)). The *onset* date of a blocked sector is defined as the date when the sector transitions from a non-blocked to a blocked state and the *decay* date is the date when the sector becomes non-blocked after being blocked previously.

Blocking frequency, defined as the fractional number of blocked days in a winter, was computed from the ensemble-mean forecasts from each EPS for the four winters considered and

Table 2: Contingency table for the hit rate analysis.

		Analysis	
		yes	no
Forecast	yes	A	B
	no	C	D

compared to that calculated using the ECMWF interim reanalysis (ERA-I) reanalyses (Dee *et al.*, 2011) as a reference. Lead times of five, seven and nine days were used for comparison with Matsueda (2009), who found that blocking frequency is well forecast up to a lead time of five days and is under-predicted at longer lead times, taking the Japanese 25-year Reanalysis (Onogi *et al.*, 2007) as the reference.

To assess the ability of forecasts to predict the timing of blocking, hit rate analyses for blocking activity, block onset and block decay were performed for ensemble forecasts at lead times from three to seven days also using ERA-I as the reference. The hit rate, H, and false positive rate, F, for a given event are defined, using the contingency table in Table 2, as follows (e.g. Wilks, 2011; Jia *et al.*, 2014):

$$H = \Pr\{\text{event predicted}|\text{event observed}\} = \frac{A}{A+C}, \tag{4}$$

$$F = \Pr\{\text{event not observed}|\text{event predicted}\} = \frac{B}{A+B}. \tag{5}$$

A good forecast will have a high hit rate and low false positive rate. Hit rates and false positive rates were also calculated for a randomly generated set of each event (e.g. block onsets) to determine if the operational forecasts performed better than a random forecast. The random sequence of events was constructed by randomly choosing whether an event occurred or not on a day in winter given the probability that it occurred in ERA-I in that period. 10000 random sequences of events were constructed and the hit rates and false positive rates were calculated and then averaged to give a hit rate and false positive rate for random forecasts of each event.

Blocking activity is defined as unity for a day when blocking is present and zero otherwise. *Blocking onset* is defined as occurring in the ERA-I data if blocking is present on the day being considered, but absent on the previous day. For the ensemble forecasts, onset occurs on a particular day and for a particular lead time if blocking is present on that day for that lead time forecast, but absent in that same forecast on the previous day. *Blocking decay* is defined analogously as occurring on the first day that a block is absent after being present on the previous day. No persistence criteria for blocking events is set to keep the sample size of events as large as possible and for direct comparability with Matsueda (2009). Blocking frequency, hit rates and false positive rates were compared for events occurring in the NewDynamics and ENDGame eras.

Ensemble spread (inter-quartile range) is included for the analysis of blocking forecasts. The forecast of blocking frequency, activity, onset and decay are calculated separately for each ensemble member to calculate the ensemble spread. This assumes that the forecast from a

given ensemble member for two different initialisation days are somehow related, which is not necessarily true. To test the representativity of the spread calculated in this way we also calculated the spread by choosing multiple random paths through the different ensemble members for each day to follow possible evolutions the system could have taken. The spread calculated with 10000 possible random ensemble member sequences is very similar to the spread calculated using consistent ensemble members, supporting the approach taken. For blocking frequency the spread is calculated at the peak of blocking frequency seen in ERA-I within the PA sector. The spread is similar in the EA sector and much smaller where there is infrequent blocking (not shown). For the hit rate analyses, ensemble spread is only plotted for hit rates for clarity of presentation. However, it is of similar magnitude for the false positive rates (not shown).

3.2.2 Rossby-wave structure

Following Gray *et al.* (2014), forecasts of Rossby-wave structure are evaluated via two parameters: Rossby-wave ridge area and isentropic PV gradient at the tropopause. To define these two parameters, we need to first define the concepts of equivalent latitude, the tropopause, Rossby-wave pattern and Rossby-wave ridges on the 320-K isentropic surface used in this work.

Equivalent latitude, ϕ_e, for a given PV contour is the latitudinal circle of a zonally-symmetric background state that contains the same mass and circulation as that contour in the full flow. Computing equivalent latitudes for all PV contours yields a background state, known as the modified Lagrangian mean (e.g. Nakamura, 1995), given by the latitudinal location of the resulting zonally-symmetric PV contours. Further details on the calculation and interpretation of equivalent latitudes can be found in Methven and Berrisford (2015).

The *tropopause* on the 320-K isentropic surface is defined here as the 2.24 PVU (PV units where 1 PVU = 10^{-6} K m^2 kg^{-1} s^{-1}) contour. The value 2.24 PVU has been chosen because it corresponds to the average location of the strongest PV gradient in the background state (Gray *et al.*, 2014).

A *Rossby-wave pattern* on the 320-K isentropic surface at a given time is defined here as the (unique) tropopause contour that spans every longitude. Contours that span only a limited range of longitudes correspond to cut-off lows or highs depending on their location (south or north) with respect to the Rossby-wave pattern.

Rossby-wave ridges are outlined by all the points along the tropopause contour with a latitude $\phi > \phi_e$; similarly, Rossby-wave troughs are outlined by all the points along the tropopause contour with a latitude $\phi < \phi_e$. Rossby-wave ridge area is defined as the area enclosed between the Rossby-wave ridge outline and ϕ_e.

The same set of equivalent latitudes for the tropopause contour used in Gray *et al.* (2014) has also been used in this work. This set was computed from the ERA-I reanalyses for the fifteenth day of each month from November to March 2009/10. These values were then linearly interpolated to daily values for each day. The calculated ridge areas will be dependent on the prescribed ϕ_e. However, the conclusions of the present investigation are independent of the

precise values chosen for ϕ_e because the ridge areas are classified and compared according to validation times: in a perfect forecast the ridge area would be the same as that in the analysis for the same validation time at all lead times as long as the same ϕ_e is used for both forecast and analysis.

The isentropic PV gradient at the tropopause, calculated as the magnitude of the 2D vector $\nabla_\theta \mathrm{PV}$ written in spherical coordinates for an isentropic surface and evaluated at the tropopause, is computed by bi-linearly interpolating the magnitude of the PV gradient onto a set of equally-spaced points along the tropopause contour (this is a different calculation method to that used by Gray *et al.* (2014)). The uniform spatial separation between tropopause points has been arbitrarily set to 50 km; however, the results are not particularly sensitive to this choice (not shown). The number of points changes from day to day with the length of the tropopause contour. The PV gradient is calculated for a given validation time as the median of the PV gradients at the tropopause points. The median has been chosen as the statistic that best represents the centre of the resulting skewed distribution (not shown).

The methodology used to characterise Rossby-wave structure in Gray *et al.* (2014) is here extended in two ways. First, rather than only presenting hemispheric results, a sector analysis is introduced by defining the Greenland–Euro-Atlantic (GEA, 90°W–40°E) and Pacific–North America (PNAm, 120°E–90°W) sectors, akin to those in the atmospheric blocking analysis (see Section 3.2.1) but spanning larger longitude ranges under the assumption that a blocked region would be surrounded by a ridge outline. Second, rather than exclusively studying the control members in each EPS, whole ensembles are investigated by computing Rossby-wave ridge area and isentropic PV gradient at the tropopause, hemispherically and sector-by-sector, for each member in each ensemble.

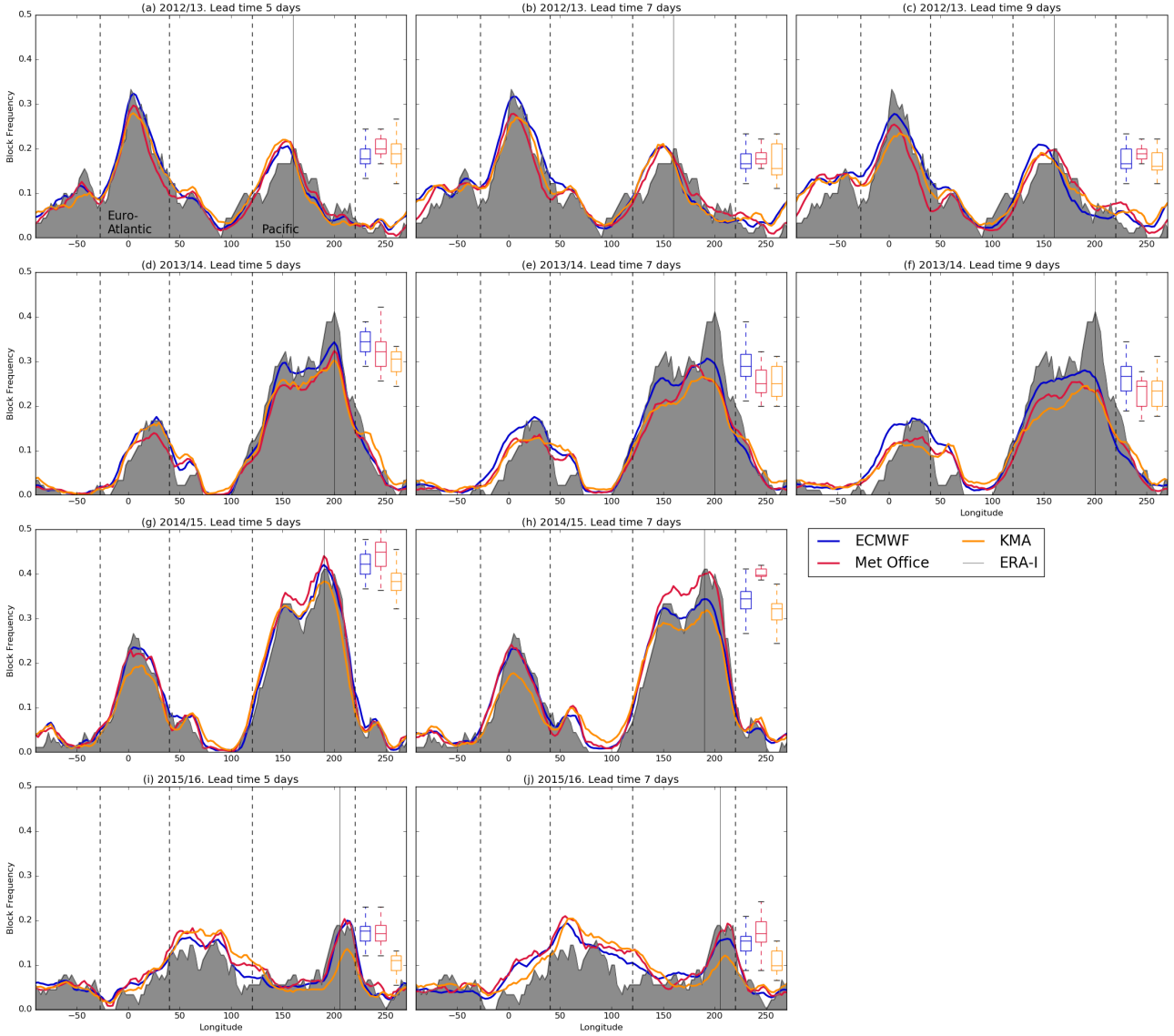

Figure 2: Blocking frequency in the NH for ERA-I (grey shading) and for each EPS (colours) for winters (a–c) 2012/13, (d–f) 2013/14, (g–h) 2014/15 and (i–j) 2015/16. The left, middle and right columns represent forecasts of lead times five, seven, and nine days respectively. Data for lead times beyond seven days is not available from the TIGGE archive for MOGREPS after the winter of 2013/14. Vertical dashed lines represent the limits of the EA and PA sectors, as labelled in (a). Box and whisker diagrams indicating ensemble spread for each EPS are included at the right side of each panel for forecasts of blocking frequency at the longitude in the PA sector with the highest blocking frequency in ERA-I. The longitudes where the ensemble spreads are calculated are indicated by the vertical solid lines.

4 Atmospheric blocking forecast skill

4.1 Blocking frequency

4.1.1 ERA-I

The ERA-I data reveals a large inter-annual variability in the pattern of NH blocking frequency (grey shadings in Fig. 2). Atmospheric blocking over the EA sector is more frequent than that over the PA sector during winter 2012/13. This pattern is reversed during the next two winters, especially during 2013/14. Among the four winters considered, 2015/16 stands out as a winter with suppressed atmospheric blocking in both the EA and PA sectors in comparison with the three previous years.

4.1.2 Operational EPS forecasts

Blocking forecasts are first considered separately for each winter in the study period before synthesizing the results. During 2012/13, every EPS performed reasonably well at predicting blocking frequency over the EA and PA sectors, even at nine days lead time (Fig. 2(a–c)). However, the EA maximum was underestimated and the PA maximum was slightly overestimated by every EPS beyond seven days lead time.

In five-day forecasts for 2013/14, the ECMWF-EPS and the KMA-EPS accurately predicted blocking frequency over the EA sector (Fig. 2d). At that lead time, the ECMWF-EPS also produced an accurate prediction of the secondary blocking frequency peak over the Pacific (around 150°E), but underestimated the primary peak (around 200°E). The KMA-EPS underestimated blocking frequency over the whole PA sector. During that same winter, five-day MOGREPS forecasts slightly underestimated blocking frequency over both sectors. Nevertheless, every EPS was able to produce the single- and double-peaked patterns observed over the EA and PA sectors, respectively. However, as lead time increased, the quality of the forecasts decreased. Even at seven days lead time (Fig. 2e), a wider sector of blocking activity was forecast over the EA sector than was observed and the forecast maxima in blocking frequency over the Pacific was underestimated by every EPS; this underestimation is worse at nine days lead time.

During 2014/15, every EPS reproduced the hemispheric structure of blocking frequency, mainly given by a single blocking frequency peak over the EA sector and a double peak over the PA sector, at five days lead time (Fig. 2g). However, while the ECMWF-EPS accurately predicted blocking frequency over the two sectors, the KMA-EPS underestimated it over the EA sector and MOGREPS overestimated it over the PA sector. In contrast with the previous winter, both the ECMWF-EPS and MOGREPS performed well at seven days lead time even though the overestimation of blocking frequencies by MOGREPS was enhanced and those predicted by the ECMWF-EPS over the PA sector had started to decay by this lead time (Fig. 2h). Blocking frequencies forecast by the KMA-EPS, on the other hand, had noticeably reduced over both the EA and PA sectors by seven days lead time.

The suppressed blocking frequency during DJF 2015/16 was slightly over-predicted in the

region around 75°E (blocking frequency peaked to the east of the EA region in this winter) by every EPS and well predicted by the ECMWF-EPS and MOGREPS, but slightly under-predicted by the KMA-EPS, over the PA sector at five days lead time (Fig. 2). The same forecast error pattern was produced at seven days lead time, but the errors were enhanced with respect to those at five days lead time (Fig. 2j).

The spread in the ECMWF-EPS, shown in the box and whisker plots in Figure 2, is generally consistent for each winter and lead time considered. The MOGREPS and KMA-EPS have similar ensemble distributions in winters 2012/13–2013/14, consistent with the EPSs having similar ensemble mean forecasts. In winters 2014/15–2015/16 the forecast of blocking frequency is clearly increased in the whole MOGREPS ensemble compared to the KMA-EPS at both five and seven days lead time.

We conclude that the ECMWF-EPS performance was consistent across the four winters despite changes in model configuration (Table 1). The performance of the KMA-EPS was similar to that of MOGREPS at all lead times in the NewDynamics era (i.e. before the introduction of ENDGame). This similarity was no longer present in the ENDGame era. Another aspect that highlights the contrasting performance between MOGREPS and the KMA-EPS in the two eras is a change in the tendency of the forecast frequency of frequently blocked regions to decrease with forecast lead time. This blocking frequency reduction has been a long-standing issue: it was already identified by Tibaldi and Molteni (1990), who showed a reduction in amplitude of the main peaks in blocking frequency with lead time in operational ECMWF winter forecasts between 1980 and 1987. During the NewDynamics era, there is generally a decay in forecast frequency with lead time in all three EPSs (Figs. 2(a–c) and 2(d–f)). This tendency is absent and even opposite in MOGREPS during the ENDGame era, whereas it is maintained in the other two EPSs (Figs. 2(g–h) and 2(i–j)).

4.1.3 ENDGame-RERUN

The blocking frequency produced by the ENDGame-RERUN, together with that from the control members from the three operational EPSs considered and ERA-I, is presented in Fig. 3. Blocking frequency for winter 2013/14 is shown for forecasts at five, seven, and nine days lead time. The control members from MOGREPS and the KMA-EPS show similar features to the ensemble mean forecasts at five days lead time (cf. Fig. 3a and Fig. 2d). For instance, the maxima in blocking frequency over the Pacific is under-predicted by both control members. However, unlike the MOGREPS ensemble mean, which underestimated the peak in EA-sector blocking frequency, the MOGREPS control member reproduced this feature. The control member from the ECMWF-EPS generally performs better than the ensemble mean, particularly at a lead time of nine days.

At lead times of five and seven days, the control members from the ECMWF-EPS and the ENDGame-RERUN perform better than those from the KMA-EPS and MOGREPS (Fig. 3(a,b)). Furthermore, the ENDGame-RERUN forecast the highest peak in blocking frequency over the PA sector more accurately than any other EPS at these lead times; this is consistent with

the accurate forecasts from MOGREPS with ENDGame run operationally during 2014/15 and 2015/16 (Fig. 2(g–j)). At nine days lead time, the ENDGame-RERUN under-predicts blocking in the PA sector (as does the control member from the ECMWF-EPS), although it performs better than both MOGREPS and the KMA-EPS; in contrast, the ENDGame-RERUN over-predicts blocking over the EA sector and Eastern Europe, although it does not capture the peak blocking frequencies.

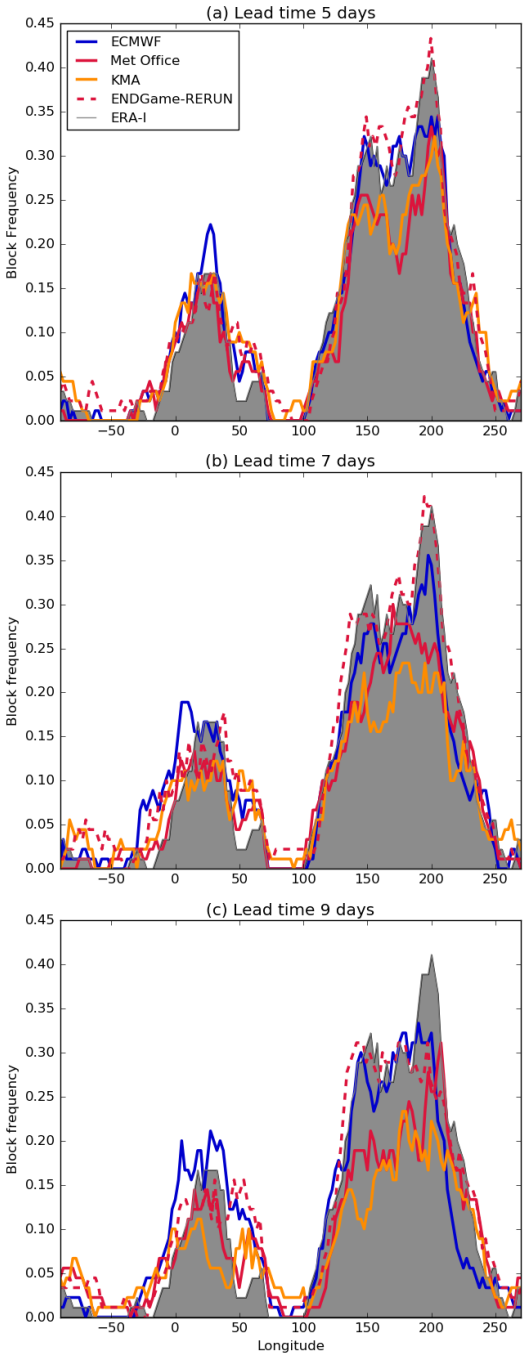

Figure 3: Blocking frequency during winter 2013/14 as diagnosed from ERA-I (grey shading). Coloured lines represent the frequency predicted by the control members from the operational forecasts from the ECMWF-EPS, MOGREPS and the KMA-EPS. The dashed line represents the frequency forecast by the control member of the ENDGame-RERUN. Forecasts of lead times five, seven and nine days are shown in panels (a), (b) and (c), respectively.

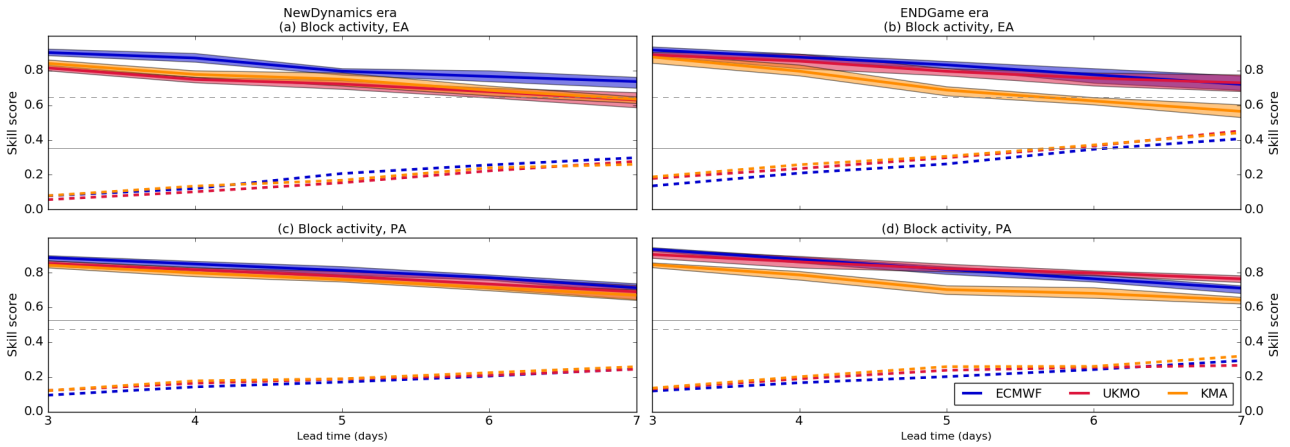

Figure 4: Hit rates (solid lines) and false positives rates (dashed lines) for blocking activity in winters (a, c) 2012/13–2013/14 and (b, d) 2014/15–2015/16 in (a, b) the EA and (c, d) PA sectors. Grey lines represent hit rates/false positive rates for a randomly-generated set of events (see text). The shading on the hit rate curves represents the ensemble spread.

4.2 Hit rate analysis

In this section, the model representation of the temporal behaviour of blocking in the EA and PA sectors is assessed in terms of blocking activity (Section 4.2.1) and blocking onset and decay (Section 4.2.2). For ease of presentation, results for the two winters during which MOGREPS used the NewDynamics dynamical core (2012/13 and 2013/14) are combined, as are the results for the two winters during which MOGREPS used the ENDGame dynamical core (2014/15 and 2015/16). It is better to compare the performance of different EPSs within each of these periods than compare how each EPS performs for the two periods because hit rates and false positive rates are sensitive to the observed blocking frequency.

4.2.1 Blocking activity

Hit rates and false positive rates as functions of lead time for each EPS are presented in Fig. 4, which also includes the hypothetical hit rates and false positive rates for a randomly-generated period of blocking activity with the same probability of blocking as in ERA-I (grey lines in Fig. 4). Blocking activity hit rates and false positive rates for all the EPSs remain above the hit rates and below the false positive rates, respectively, for the randomly-generated blocking activity, which implies that the ensemble forecasts have more skill at forecasting blocking activity than a random forecast for lead times up to (at least) seven days.

We first discuss hit rates and false positive rates over the EA sector (Fig. 4a,b), followed by those for the PA sector (Fig. 4c,d). Over the EA sector, MOGREPS and the KMA-EPS have similar hit rates and false positive rates in the NewDynamics era, which is consistent with the blocking frequency being similar for these two EPSs during those winters (cf. Fig. 2). The spread in MOGREPS and KMA-EPS is also similar in the NewDynamics era. The hit rates of both EPSs are well below those of the ECMWF-EPS, which indicates better forecasting skill in the ECMWF-EPS (using ERA-I as the reference). The false positive rates over the EA sector are similar across the EPSs. Moving to the ENDGame era (Fig. 4b) , there is a clear change in

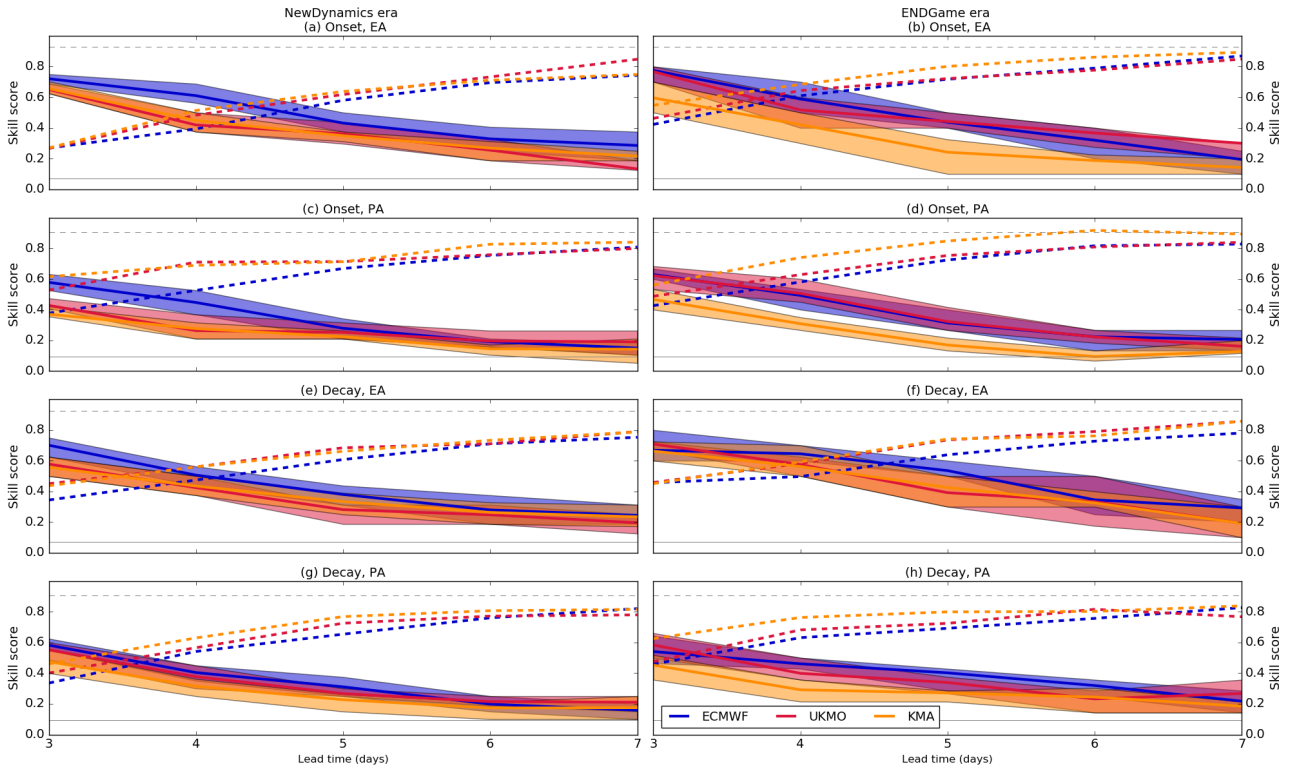

Figure 5: Hit rates (solid lines) and false positives rates (dashed lines) for (a, b, c, d) block onset and (e, f, g, h) block decay in (a, c, e, g) winters 2012/13–2013/14 and (b, d, f, h) 2014/15–2015/16 in (a, b, e, f) the EA and (c , d, g, h) PA sectors. Grey lines represent hit rates/false positive rates for a randomly-generated set of events (see text). The shading on the hit rate curves represents the ensemble spread.

MOGREPS performance. MOGREPS hit rates are higher than those of the KMA-EPS (such that the ensemble spreads no longer overlap from five days lead time) and more comparable to those of the ECMWF-EPS for all lead times. False positive rates are similar across both eras for every EPS which, for MOGREPS in particular, implies that the over-estimation of blocking frequency over the EA sector during 2015/16 (Figs. 2(i,j) is not worse than it would have been with the NewDynamics dynamical core.

The patterns of hit rates and false positive rates in the PA sector are similar to those in the EA sector (Fig. 4c,d). MOGREPS and the KMA-EPS have similar hit rates, which are generally below those of the ECMWF-EPS during the NewDynamics era. During the ENDGame era, MOGREPS has higher hit rates than the KMA-EPS at all lead times and, at lead times of five, six and seven days, MOGREPS exhibits the highest hit rates among the three EPSs considered. In terms of false positive rates, the three EPSs exhibit similar performance during the two eras.

4.2.2 Blocking onset and decay

Hit rates and false positive rates are shown in Fig. 5 for onset and decay of blocking occurring in the EA (Figs. 5(a,b,e,f)) and PA (Figs. 5(c,d,g,h)) sectors. Figures 5(a–h) also include the hit rates and false positive rates for a randomly-generated set of onset/decay events with the same corresponding probabilities as in ERA-I. Overall, the timescales for accurate prediction

of block onset and decay (considered as hit rates above 0.5) by the EPSs are comparable to those found in Pelly and Hoskins (2003) and Jia *et al.* (2014) and are an improvement on those found in earlier studies (e.g. Tibaldi and Molteni, 1990).

Hit rates are much lower (and false positive rates higher) for the onset and decay of blocking than for blocking activity. The ensemble spread is also larger for forecasts of onset and decay. In the EA sector, there is a general downward trend with lead time in hit rate for the onset of blocking in the every EPS in both the NewDynamics and ENDGame eras. MOGREPS and the KMA-EPS have similar hit rates for blocking onsets in the NewDynamics era and their ensemble spreads are similar. In contrast, in the ENDGame era, MOGREPS has hit rates greater than the KMA-EPS at all lead times and especially at longer lead times; the ensemble spreads for the forecasts from MOGREPS the KMA-EPS do not overlap at five and six days lead time and only just meet at 7 days lead time. This suggests an improvement in the forecast of block onsets with the ENDGame dynamical core in the EA sector, although analysis of more winters would be needed to confirm this.

For block onsets in the PA sector, hit rates for MOGREPS and the KMA-EPS are again similar in the NewDynamics era and there is an increase in hit rate for MOGREPS in the ENDGame era when compared to the KMA-EPS with separated ensemble spreads at four to six days lead time. This increased hit rate is not associated with an increased false positive rate: false positive rates for MOGREPS in the ENDGame era are lower than for the KMA-EPS and similar to the ECMWF-EPS. Hit rates for the ECMWF-EPS for block onset in the PA sector are similar in both eras. Hit rates for block onset are generally lower in the PA sector than in the EA sector at all lead times and for all EPSs. A possible explanation for this could be the different mechanisms driving the formation of blocking in each sector: block formation in the European region is most dependent on low-frequency dynamics, whereas forcing by transient eddies is crucial for block formation in the Pacific (Nakamura *et al.*, 1997). All EPSs perform better than for a randomly-generated list of onset dates.

Hit rates and false positive rates for the decay of blocking show similar patterns to those for block onset: a general downward trend in hit rate and increase in false positive rate with lead time. The values for each EPS are comparable for onset and decay, implying that EPSs do not clearly forecast either the onset or decay of blocking best. Other studies (e.g. Pelly and Hoskins, 2003) have found that models tend to predict the decay of blocking events more accurately. In the EA sector, the hit rate and false positive rate for MOGREPS and the KMA-EPS are similar for all lead times and the ECMWF-EPS performs best at short lead times of four and five days in both the NewDynamics and ENDGame eras. In the PA sector, the three EPSs perform similarly in the NewDynamics era. MOGREPS performs better than the KMA-EPS, and more similarly to the ECMWF-EPS, in the ENDGame era at lead times up to five days although the ensemble spreads are not separated: hit rates are generally higher and false positive rates lower than for the KMA-EPS. Apart from this short lead time difference in the PA sector, the hit rates, false positive rates and ensemble spreads for each EPS are consistent in both sectors and both eras. Hence, there has not been a clear impact on the forecast of block decay in

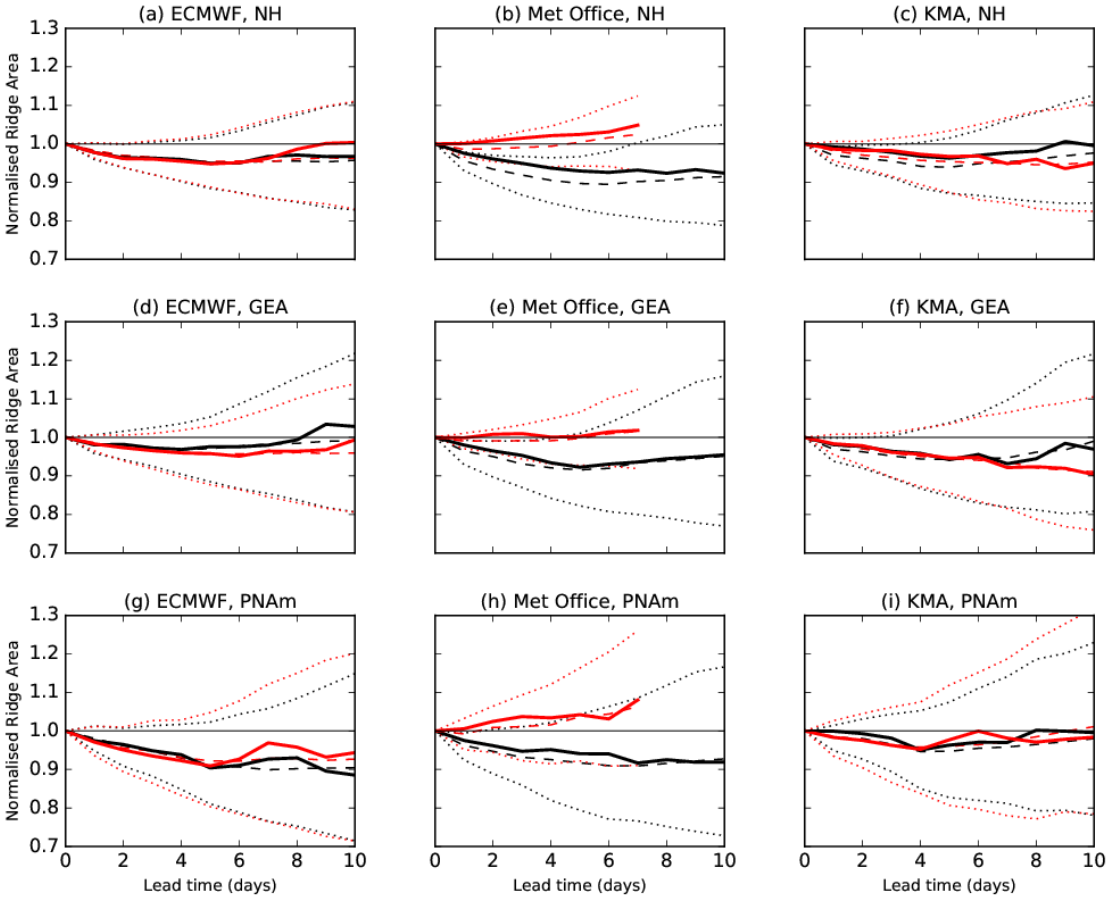

Figure 6: Ridge area normalised by its value at analysis time as a function of forecast lead time over (a–c) the NH, (d–f) the GEA sector and (g–i) the PNAm sector for ECMWF (left), Met Office (middle) and KMA (right), showing the control member's median (solid) and the ensemble's median (dashed) and first and third quartiles (dotted) over winters 2012/13–2013/14 (black) and 2014/15–2015/16 (red).

MOGREPS due to the introduction of the ENDGame dynamical core.

The improvements in the representation of atmospheric blocking in MOGREPS due to the new dynamical core are hypothesised to be related to the improvement in the representation of upper-level Rossby waves. The changes in the representation of Rossby-waves in MOGREPS are presented in the next section.

5 Rossby-wave structure forecast skill

5.1 Rossby-wave ridge area

The 320-K Rossby-wave ridge area as a function of forecast lead time is shown in Fig. 6 for the NH (Figs. 6a–c), and the GEA (Figs. 6d–f) and PNAm (Figs. 6g–i) sectors (as defined in Section 3.2.2). The results in each panel are grouped according to the NewDynamics and ENDGame era winters. The ensemble results are presented in terms of the first, second (median) and third quartiles of all ensemble members across the winter days for each forecast lead time. The control member median over winter is also shown for comparison. The definition

of equivalent latitude prevents us from comparing absolute values of ridge areas between two different years (see Section 3.2.2). Thus, the results are presented as ridge area normalised by its value at analysis time (T+0).

During the NewDynamics era over the NH, the three EPSs exhibit a decrease in ridge area with lead time (as described by Gray *et al.* (2014) for earlier years in the TIGGE archive) in the medians of both the control member and the ensemble. The decrease in hemispheric ridge area is less evident in the control member of the KMA-EPS, but it is still noticeable in its ensemble median (Fig. 6c). Considering only the ensemble data, the ECMWF-EPS and the KMA-EPS display very similar characteristics with a maximum decrease in the median in both EPSs of about 5% with respect to T+0 at ten days lead time, while the first and third quartiles are located at about 80% and 110% of the T+0 value, respectively. In contrast, MOGREPS exhibits the strongest ridge area decrease with a maximum decrease in the median of about 10% with respect to T+0 at ten days lead time; even the third quartile for this EPS only reaches 95% of the T+0 ridge area value at a lead time of five days. Considering that MOGREPS and the KMA-EPS use the same dynamical core during this era, the difference in behaviour between these two EPSs can be attributed primarily to resolution (see Table 1).

During the ENDGame era over the NH, the ECMWF-EPS exhibits strikingly similar statistical behaviour to the NewDynamics era (Fig. 6a), both in the control member and in the rest of the ensemble. Similarly, the KMA-EPS exhibits similar behaviour over the NH for the two eras, at least up to six days lead time (Fig. 6c) (though the match is not as close as that for the ECMWF-EPS). In clear contrast with the ECMWF-EPS and the KMA-EPS, MOGREPS exhibits large differences in performance over the NH in the two eras (Fig. 6b). During the ENDGame era, the ridge area value is maintained above 90% of its value at T+0 even by the first quartile; the median of the control member displays an increase in normalised ridge area as lead time increases. At a lead times exceeding five days, forecast NH ridge area is slightly more likely to be larger than (rather than less than) its T+0 value (Fig. 6b).

The sector analysis reveals a longitudinal variation in the systematic forecast errors of Rossby-wave ridge area. However, this longitudinal variation depends on both the EPS and era. There are differences between the GEA and the PNAm sectors in the ECMWF-EPS during the NewDynamics era. In the ECMWF-EPS, the ensemble median over the GEA sector remains above 95% of the T+0 ridge area value (Fig. 6d), while in the PNAm sector the ensemble median slowly decreases to reach 90% of the T+0 ridge area value over that sector (Fig. 6g). In MOGREPS, the ensemble median and inter-quartile range during the NewDynamics era over the GEA (Fig. 6e) and the PNAm (Fig. 6h) sectors are similar to each other throughout the ten days of forecast lead time considered with the decrease in the PNAm sector being slightly larger than in the GEA sector. In the KMA-EPS, the two sectors also behave similarly to each other during the NewDynamics era up to six days lead time (see Figs. 6(f,i)).

The statistical behaviour of sector ridge-area forecasts during the ENDGame era is similar to that of the NewDynamics era in both the ECMWF-EPS and the KMA-EPS. However, there is a small displacement towards lower values in the GEA sector and towards higher values in the

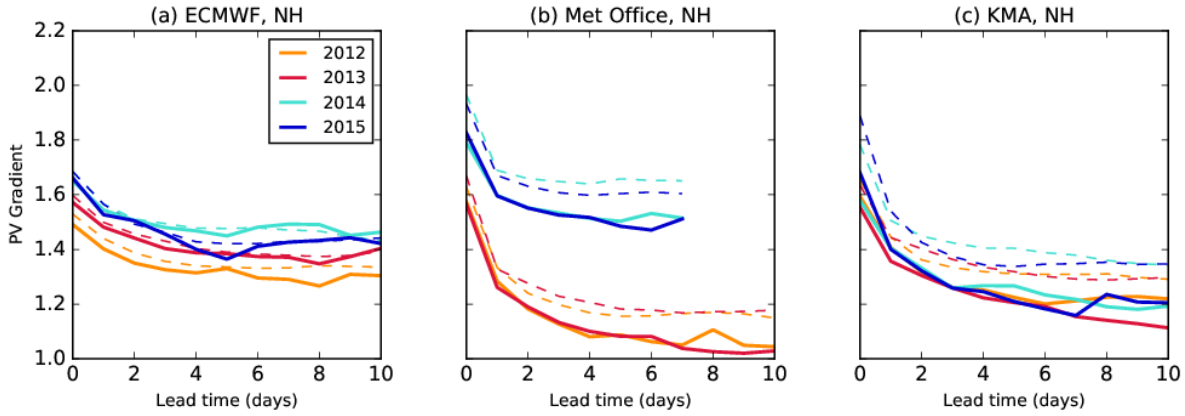

Figure 7: NH isentropic PV gradient, in PVU $(100\ \mathrm{km})^{-1}$, at the tropopause as a function of forecast lead time for winter 2012/13 (yellow), 2013/14 (red), 2014/15 (turquoise) and 2015/16 (blue) showing the control median (solid) and the ensemble median (dashed) for (a) ECMWF, (b) Met Office and (c) KMA.

PNAm sector in the ENDGame era relative to the NewDynamics era in both EPSs. Considering that there are no significant changes in the configuration of these two EPSs between eras, the statistical differences between eras hint at flow dependence of the development of systematic Rossby-wave forecast error. Consistent with the findings of the hemispheric analysis (cf. Fig. 6b and related discussion), sector ridge-area forecasts in MOGREPS during the ENDGame era are very different from those during the NewDynamics era. During the ENDGame era, the median of the normalised ridge area over the GEA sector stays around 1, indicating the virtual absence of systematic forecast error (implying the forecast error is purely random) over this sector throughout the seven-day lead time interval considered (Fig. 6e). However, this result should not be interpreted in isolation from the NH and PNAm results. Over the PNAm sector, the median of the normalised ridge area increases during the seven lead-time days so that there is a greater likelihood of an overestimation of ridge area over this sector (Fig. 6h). The sector analysis also reveals larger ensemble spread in the sectors than in the hemisphere as a whole, implying that the narrower hemispheric distribution arises as a result of compensations between sectors. This effect is larger for longer lead times, which explains the apparent recovery of ridge area at longer lead times: the forecasts may be displaying total hemispheric ridge area values close to those at T+0. However, the recovery of ridge area may be taking place in localised hemispheric sectors. This effect can be found in both eras in the three EPSs considered.

5.2 Isentropic PV gradient at the tropopause

The ensemble representation of the isentropic PV gradient at the NH tropopause for the four individual winters is shown in Fig. 7. Unlike ridge area, whose non-normalised values depend on the prescribed equivalent latitude, non-normalised values of isentropic PV gradient do not depend on any arbitrary reference and therefore can be compared directly. The ensemble statistics are represented by the median of all ensemble members over all winter days. The control member median over winter days is also shown for comparison. The tropopause PV

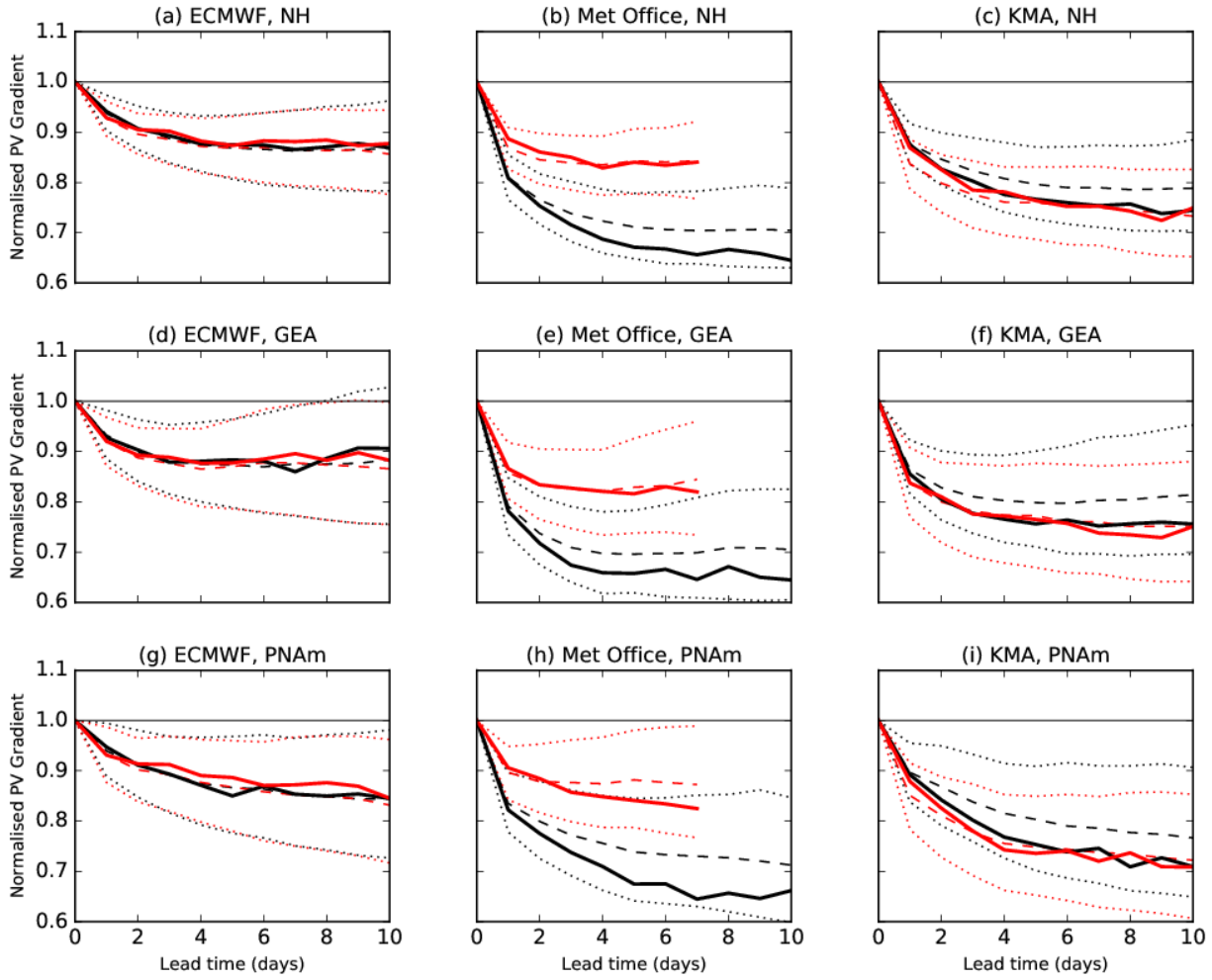

Figure 8: As in Fig. 6, but for PV gradient at the tropopause normalised by its value at analysis time as a function of forecast lead time.

gradient exhibits a decrease with forecast lead time in every year and all the three EPSs considered, similar to that described by Gray *et al.* (2014) for earlier years in the TIGGE archive. However, there is inter-annual variability in the values. Moreover, there is a lack of agreement in the values of the PV gradient at the tropopause among EPSs even at T+0. For any given year, MOGREPS tends to yield the largest values. There is also a systematic difference between the control member and the rest of the ensemble members in every EPS: the median of the ensemble corresponds to sharper PV gradients than the corresponding control member. This effect is present even at T+0, which suggests that it occurs as the ensemble perturbations are generated, and is most noticeable in MOGREPS. In this EPS (and in the KMA-EPS) the initial condition perturbations are produced using the local ensemble transform Kalman filter (Bowler *et al.*, 2009) and are, therefore, a linear combination of the forecast perturbations. Since the forecast perturbations are affected by the stochastic physics perturbations, these physics perturbations influence the initial condition perturbations (Bowler *et al.*, 2008). In MOGREPS, the stochastic physics perturbations consist of the 'random parameters' and the stochastic kinetic energy backscatter schemes (Bowler *et al.*, 2008, 2009). The latter introduces vorticity perturbations into the forecast to account for kinetic energy dissipated by other model components such as

numerical advection errors, horizontal diffusion and parametrization schemes (Bowler *et al.*, 2009). These vorticity perturbations, introduced in regions where gradients are already large such as the tropopause, are a likely source of the stronger PV gradients in the ensemble members (compared to the control member) at T+0.

Tropopause PV gradient normalised by its T+0 value as a function of time is shown in Fig. 8. As for the Rossby-wave ridge area, results are presented for the NH (Fig. 8a–c) and for the GEA (Fig. 8d–f) and the PNAm (Fig. 8g–h) sectors and grouped according the New-Dynamics and ENDGame eras. During the NewDynamics era over the NH, the ECMWF-EPS (Fig. 8a) displayed the smallest decrease in tropopause PV gradient relative to its T+0 value while MOGREPS (Fig. 8b) displayed the largest decrease, as indicated by both the control members' medians and the ensembles' inter-quartile ranges. The reduction of PV gradient in the ECMWF-EPS is very similar during the two eras (Fig. 8a). The behaviour of the KMA-EPS control member median is very similar during both eras too; however, the ensemble exhibited a slight reduction during the ENDGame era in comparison to the NewDynamics era throughout the ten lead-time days considered, with a difference in medians of about 5% of the T+0 value at ten days lead time. Given the lack of changes in the configuration of the KMA-EPS between eras, the differences in response between eras could be attributed to differences in atmospheric flows. However, given that similar differences in response are not evident in the ECMWF-EPS, this flow-dependent sensitivity might be model dependent.

As for the Rossby-wave ridge area, MOGREPS exhibited clear differences in the forecasts of tropopause PV gradient in the two eras. The ensemble median changed from 70% of the T+0 value at seven-days lead time during the first era to just below 85% of the T+0 value at the same lead time during the second era, making the MOGREPS response more comparable to that of the ECMWF-EPS. Nevertheless, the ECMWF-EPS exhibited the smallest decrease in normalised PV gradient throughout the seven days of comparable lead time. The PV gradient in MOGREPS at T+0 is sharper than in other analyses (Fig. 7). Thus, even though the drop with lead time in MOGREPS in the ENDGame era is comparable to that in the ECMWF-EPS (Fig. 8a–c), the gradient in MOGREPS at day five is comparable with that of the ECMWF-EPS at T+0 (Fig. 7). It is not possible to compare the analysed PV gradient with observations; indeed, the current lack of observations with which to verify and constrain tropopause PV gradients was one of the motivations for the the recent North Atlantic Waveguide and Downstream Impact Experiment (NAWDEX).

The sector-by-sector analysis shows that even though there is a longitudinal dependence of systematic errors in tropopause PV gradient in the three EPSs considered, the statistical behaviour of each EPS over each sector is consistent with that over the NH. The PV gradient forecasts of the ECMWF-EPS are very consistent across the two eras for both sectors (Figs. 8(d,g)). As observed when considering the NH, the PV gradient forecasts from the KMA-EPS exhibit differences across the two eras for both sectors (Figs. 8(f,i)). However, these differences are much smaller than those exhibited by MOGREPS, for which there is clearly a reduction in the decrease of tropopause PV gradient with lead time during the ENDGame era

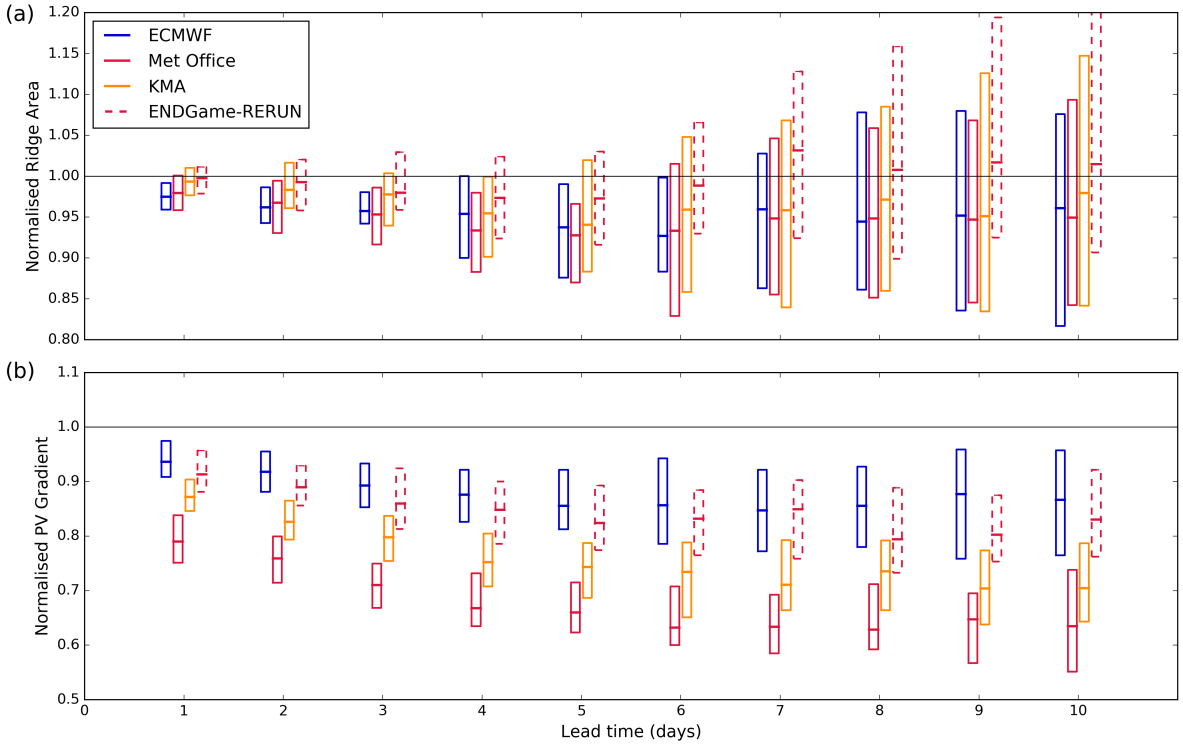

Figure 9: (a) Ridge area and (b) isentropic PV gradient at the tropopause, normalised by their values at T+0, as functions of forecast lead time over the NH during winter 2013/14. The boxes indicate the first, second and third quartiles of the control members of the indicated EPSs and the ENDGame-RERUN. Each group of boxes correspond to the central lead time labelled on the horizontal axis.

relative to the NewDynamics era over both sectors (Figs. 8(e,h)).

5.3 ENDGame-RERUN

The comparison between MOGREPS and the KMA-EPS leaves open the possibility that the changes seen in the MOGREPS forecasts on the introduction of the ENDGame dynamical core are just due to the increased resolution. To assess this possibility, the results for the ENDGame-RERUN are compared with the EPS control members for winter 2013/14 in Fig. 9.

The control members of the ECMWF-EPS, MOGREPS and the KMA-EPS show the same general features during 2013/14 as those discussed previously for the NewDynamics era, both in terms of ridge area (Fig. 9a) and tropopause isentropic PV gradient (Fig. 9b). Regarding ridge area, the three operational EPSs behaved in a very similar way during this particular winter, even though the KMA-EPS conserves ridge area marginally better than MOGREPS and the ECMWF-EPS (Fig. 9a). However, it should be noted that the apparent similarity between MOGREPS and ECMWF-EPS for this winter is not present in winter 2012/13 (not shown). The performance of the ENDGame-RERUN at maintaining ridge area is similar to that of the KMA-EPS control member for the first three days, after which time the median ridge area is greater for the ENDGame-RERUN. The median stays within 97% of the T+0 ridge-area value for up to five-days lead time (Fig. 9a) and then increases so that it is about 1% larger than its

value at T+0 at ten-days lead time although the spreads for all of the models overlap at all times. This slight increase in normalised ridge area, relative to the T+0 value, by 10-days lead time is consistent with the behaviour found for MOGREPS during the ENDGame era (Fig. 6b). Regarding tropopause PV gradient, the ENDGame-RERUN values are comparable, but slightly weaker, than those of the ECMWF-EPS control member and clearly stronger than those of the control members of both MOGREPS and the KMA-EPS (the spreads of the MOGREPS control member and the ENDGame-RERUN do not overlap, and the upper-limit (third quartile) of the spread of KMA control member is below the median of the ENDGame-RERUN).

Direct comparisons between the control members of MOGREPS and the KMA-EPS and between the latter and the ENDGame-RERUN show that the changes in the representation of the ridge area and tropopause isentropic PV gradient are partly due to the changes in the dynamical core and partly due to the resolution increase. Thus, improving the dynamical core can be as important as increasing resolution for the improvement of the representation of upper-level Rossby wave structure as well as blocking, as discussed in Section 4.

6 Summary and conclusion

We have compared the performance of the ECMWF-EPS, MOGREPS and the KMA-EPS at forecasting two inter-related large-scale aspects of the mid-latitude circulation, atmospheric blocking and upper-level Rossby-wave structure, to assess the impact of changing a model's dynamical core. Forecasts of blocking have been evaluated in terms of blocking frequency and sector hit rate analysis of blocking activity, blocking onset and blocking decay. Forecasts of upper-level Rossby-wave structure have been evaluated in terms of non-conservation of ridge area and isentropic PV gradient at the tropopause with lead time. The study has focused on the winters 2012/13–2015/16. This period was chosen because it includes two years before and two years after the change of dynamical core from NewDynamics to ENDGame in MOGREPS in July 2014. During this same period, the ECMWF-EPS and the KMA-EPS maintained a relatively stable configuration (see Table 1) providing an opportunity to examine the impact of the improvements in the dynamical core by using the performance of the ECMWF-EPS and the KMA-EPS as references. The KMA-EPS, in particular, has effectively provided a direct control experiment (apart from resolution differences) as it is based on the same underlying model as MOGREPS, but used the NewDynamics dynamical core throughout the period of analysis. The confounding factor of resolution has been addressed using a single-member hindcast, comparable to a control member run, in which the MetUM was run with ENDGame for winter 2013/14 at N320 resolution, the same resolution as that used by the KMA-EPS.

The long-standing issue that forecast frequency of frequently blocked regions exhibits a tendency to decrease with lead time (e.g. Tibaldi and Molteni, 1990) has been identified in the present study in both the ECMWF-EPS and the KMA-EPS during the four winters included in the analysis. The effect was also identified in MOGREPS during the NewDynamics era, during which the performance of MOGREPS was similar to that of the KMA-EPS at all

lead times. However, it has been shown that the introduction of ENDGame into MOGREPS has led to forecast frequency increases with lead time relative to the KMA-EPS and ERA-I in several longitude bands and these changes are robust across the ensemble. The impact of ENDGame was confirmed by the single-member hindcast experiment, which showed that control forecasts with the ENDGame dynamical core performed better than control forecasts with the NewDynamics core (even at the same resolution) for 2013/14.

Consistency was also found in the hit rates and false positive rates for blocking activity in the ECMWF-EPS and the KMA-EPS throughout the four winters considered. The KMA-EPS and MOGREPS exhibited similar hit rates, generally below those of the ECMWF-EPS, during the NewDynamics era. During the ENDGame era, MOGREPS exhibited higher hit rates than the KMA-EPS and similar to the ECMWF-EPS, while maintaining a similar performance in terms of false positive rates throughout the four winters (as did the other two EPSs). Hit rates for onset and decay of blocking were lower than that for blocking activity and the ensemble spread was larger, which highlights that the models struggle to forecast the onset and decay of blocking accurately. The EPSs were found to be more skilful at forecasting onsets in the EA than PA region, while decays were more consistent across the two sectors during the four winters considered. For MOGREPS in the ENDGame era, hit rates for blocking onset were clearly above, and false positives below, those for the KMA-EPS for most lead times; during the NewDynamics era, the performance of the two EPSs was more comparable.

The tendency of the frequency of frequently blocked regions to decrease with lead time is consistent with a reduction in Rossby-wave ridge amplitude with lead time (first identified by Gray *et al.* (2014)); this tendency has also been found in the present study. The ECMWF-EPS and the KMA-EPS exhibited a decrease in Rossby-wave ridge area with lead time that was consistent across the four winters considered. In contrast, the reduced reduction (or increase) in blocking frequency with lead time in several locations in MOGREPS in the ENDGame era (but not in the NewDynamics era) was associated with a clear improvement in conserving, and even increasing, Rossby-wave ridge area with lead time.

Finally, there is still a tendency for isentropic PV gradient at the tropopause to rapidly decrease with lead time, as previously identified by Gray *et al.* (2014). However, the introduction of ENDGame has improved the performance of MOGREPS in this respect. There may be a link between this result and the assessment of atmospheric blocking through the mechanism outlined by Harvey *et al.* (2016): smoothing of the PV gradient was found to lead not only to slower Rossby waves, but also to a decrease in their amplitude.

In summary, the ENDGame dynamical core has led to noticeable changes in forecasts of blocking frequency as well as in blocking activity and the onset of blocking; consistent changes were not seen in the decay of blocking. These results are consistent with those for upper-level Rossby-wave structure, as expected from the relationship between this and atmospheric blocking.

We conclude with the formulation of a feature chain, linking the physical basis of the changes in the dynamical core to the changes in the large-scale circulation. At the grid-point

level, ENDGame has improved the accuracy of the MetUM's large-scale dynamics, leading to a reduction of the model's implicit damping and, as a consequence, to more kinetic energy at mid-latitudes (Walters *et al.*, 2014, 2017). More mid-latitude kinetic energy has upscaled, leading to synoptic effects such as stronger extratropical cyclones (Walters *et al.*, 2017). Stronger extratropical cyclones have led to an improved tropopause structure and so improved Rossby-wave structure and development as shown by our results in terms of improvements in the conservation of Rossby-wave ridge area and tropopause sharpness (diagnosed by isentropic PV gradient at the tropopause) with lead time. In turn, the cumulative effect of a better representation of upper-level Rossby waves has led to an improved representation of atmospheric blocking. We hypothesise that a better representation of mid-latitude weather systems will also lead to improvements to the representation of large-scale modes of variability (i.e. NAO, PNA). This hypothesis is supported by Williams *et al.* (2015), who showed improvement in correlation and variability in the NAO, and Dunstone *et al.* (2016), who showed that the skill at predicting the NAO in DePreSys3-GC2 (whose atmospheric component is based on ENDGame) is similar to that using GloSea5.

Finally, our study has revealed that changes in a model's dynamical core can be at least as effective as realistic increases in operational model resolution in improving forecasts of upper-level Rossby wave structure and associated atmospheric blocking. This finding is important because the computational cost of dynamical core changes is likely to be substantially less than that associated with typical resolution increases.

Acknowledgements

We thank Prof John Methven for providing the equivalent latitudes used for the calculation of Rossby-wave ridge area and Drs Warren Tennant and Neill Bowler for useful discussion. We also thank two anonymous reviewers, whose useful comments have helped to improve the original manuscript. O.M-A.'s contribution was funded by the United Kingdom's Natural Environment Research Council (NERC) as part of the National Centre for Atmospheric Sciences. J.W.M.'s contribution was funded through a NERC Industrial CASE studentship in collaboration with the Met Office (NE/M009610/1). K.D.W. was supported by the joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101). The TIGGE data for the three EPSs was retrieved via the ECMWF data server. The source code for the UM is available to use. To apply for a license for the UM go to http://www.metoffice.gov.uk/research/collaboration/um-collaboration. For more information on the exact model versions and branches applied please contact the authors. Data from the simulations is archived at the Met Office and available for research use through the Centre for Environmental Data Analysis JASMIN platform (http://www.jasmin.ac.uk/); please contact the authors for details.

References

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

Barriopedro D, García-Herrera R, Trigo RM. 2010. Application of blocking diagnosis methods to general circulation models. Part I: a novel detection scheme. *Clim. Dyn.* **35**: 1373–1391.

Bosart LF, Lackmann GM. 1995. Postlandfall tropical cyclone reintensification in a weakly baroclinic environment: A case study of hurrica ne david (september 1979). *Monthly Weather Review* **123**: 3268–3291, doi:10.1175/1520-0493(1995)123⟨3268:PTCRIA⟩2.0.CO;2.

Bougeault P, Toth Z, Bishop C, Brown B, Burridge D, Chen DH, Ebert B, Fuentes M, Hamill TM, Mylne K, Nicolau J, Paccagnella T, Park YY, Parsons D, Raoult B, Schuster D, Dias PS, Swinbank R, Takeuchi Y, Tennant W, Wilson L, Worley S. 2010. The THORPEX interactive grand global ensemble. *Bull. Amer. Meteor. Soc.* **91**(8): 1059–1072, doi: 10.1175/2010BAMS2853.1.

Bowler NE, Arribas A, Beare SE, Mylne KR, Shutts GJ. 2009. The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.* **135**: 767–776.

Bowler NE, Arribas A, Mylne KR, Robertson KB, Beare SE. 2008. The MOGREPS short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.* **134**: 703–722.

Buizza R, Milleer M, Palmer T. 1999. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* **125**: 2887–2908.

Charney JG, Phillips NA. 1953. Numerical integration of the quasi-geostrophic equations for barotropic and simple baroclinic flows. *J. Meteor.* **10**(2): 71–99.

d'Andrea F, Tibaldi S, Blackburn M, Boer G, Déqué M, Dix M, Dugas B, Ferranti L, Iwasaki T, Kitoh A, *et al.* 1998. Northern hemisphere atmospheric blocking as simulated by 15 atmospheric general circulation models in the period 1979-1988. *Clim. Dyn.* **14**: 385–407.

Davies T, Cullen MJP, Malcolm AJ, Mawson MH, Staniforth A, White AA, Wood N. 2005. A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Q. J. R. Meteorol. Soc.* **131**: 1759–1782.

Davis CA, Stoelinga MT, Kuo YH. 1993. The integrated effect of condensation in numerical simulations of extratropical cyclogenesis. *Mon. Wea. Rev.* **121**: 2309–2330.

de Vries H, Woollings T, Anstey J, Haarsma RJ, Hazeleger W. 2013. Atmospheric blocking and its relation to jet changes in a future climate. *Clim. Dyn.* **41**(9-10): 2643–2654.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hlm EV, Isaksen L, Kllberg P, Khler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thpaut JN, Vitart F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**: 553–597.

Dunstone N, Smith D, Scaife A, Hermanson L, Eade R, Robinson N, Andrews M, Knight J. 2016. Skilful predictions of the winter North Atlantic Oscillation one year ahead. *Nat. Geosci.* **9**: 809–814, doi:10.1038/ngeo2824.

Ferranti L, Corti S, Janousek M. 2015. Flow-dependent verification of the ecmwf ensemble over the euro-atlantic sector. *Q. J. R. Meteorol. Soc.* **141**: 916–924, doi:10.1002/qj.2411.

Giannakaki P, Martius O. 2016. An object-based forecast verification tool for synoptic-scale Rossby waveguides. *Weather Forecast.* **31**: 937–946, doi:10.1175/WAF-D-15-0147.1.

Gray SL, Dunning C, Methven J, Masato G, Chagnon J. 2014. Systematic model forecast error in Rossby wave structure. *Geophys. Res. Lett.* **41**, doi:10.1002/2014GL059282.

Harvey BJ, Methven J, Ambaum MHP. 2016. Rossby wave propagation on potential vorticity fronts with finite width. *J. Fluid Mech.* **794**: 775–797, doi:10.1017/jfm.2016.180.

Jia X, Yang S, Song W, He B. 2014. Prediction of wintertime northern hemisphere blocking by the NCEP climate forecast system. *Journal of Meteorological Research* **28**: 76–90.

Lejenäs H, *et al.* 1983. Characteristics of Northern Hemisphere blocking as determined from a long time series of observational data. *Tellus A* **35**(5): 350–362.

Martínez-Alvarado O, Madonna E, Gray S, Joos H. 2016. A route to systematic error in forecasts of Rossby waves. *Q. J. R. Meteorol. Soc.* **142**: 196–210.

Matsueda M. 2009. Blocking predictability in operational medium-range ensemble forecasts. *SOLA* **5**: 113–116.

Methven J, Berrisford P. 2015. The slowly evolving background state of the atmosphere. *Q. J. R. Meteorol. Soc.* **141**: 2237–2258, doi:10.1002/qj.2518.

Mittermaier M, North R, Semple A, Bullock R. 2016. Feature-based diagnostic evaluation of global NWP forecasts. *Mon. Wea. Rev.* **144**: 3871–3893.

Molteni F, Buizza R, Palmer TN, Petroliagis T. 1996. The ECMWF ensemble prediction system: Methodology and validation. *Q. J. R. Meteorol. Soc.* **122**: 73–119.

Nakamura H, Nakamura M, Anderson JL. 1997. The role of high-and low-frequency dynamics in blocking formation. *Mon. Wea. Rev.* **125**(9): 2074–2093.

Nakamura N. 1995. Modified lagrangian-mean diagnostics of the stratospheric polar vortices. Part I. Formulation and analysis of GFDL SKYHI GCM. *J. Atmos. Sci.* **52**: 2096–2108.

Onogi K, Tsutsui J, Koide H, Sakamoto M, Kobayashi S, Hatsushika H, Matsumoto T, Yamazaki N, Kamahori H, Takahashi K, Kadokura S, Wada K, Kato K, Oyama R, Ose T, Mannoji N, Taira R. 2007. The JRA-25 Reanalysis. *J. Meteorol. Soc. Jpn. Ser. 2* **85**: 369–432, doi:10.2151/jmsj.85.369.

Park YY, Buizza R, Leutbecher M. 2008. TIGGE: preliminary results on comparing and combining ensembles. *Q. J. R. Meteorol. Soc.* **134**: 2029–2050, doi:10.1002/qj.334.

Pelly JL, Hoskins BJ. 2003. How well does the ECMWF ensemble prediction system predict blocking? *Q. J. R. Meteorol. Soc.* **129**(590): 1683–1702.

Riemer M, Jones SC. 2010. The downstream impact of tropical cyclones on a developing baroclinic wave in idealized scenarios of extratropical transition. *Q. J. R. Meteorol. Soc* **136**: 617–637.

Röthlisberger M, Pfahl S, Martius O. 2016. Regional-scale jet waviness modulates the occurrence of midlatitude weather extremes. *Geophys. Res. Lett.* **43**, doi:10.1002/2016GL070944.

Schiemann R, Demory ME, Shaffrey LC, Strachan J, Vidale PL, Mizielinski MS, Roberts MJ, Matsueda M, Wehner MF, Jung T. 2017. The resolution sensitivity of Northern Hemisphere blocking in four 25-km atmospheric global circulation models. *J. Clim.* **30**: 337–358, doi: 10.1175/JCLI-D-16-0100.1.

Schwierz C, Croci-Maspoli M, Davies HC. 2004. Perspicacious indicators of atmospheric blocking. *Geophys. Res. Lett.* **31**, doi:10.1029/2003GL019341.

Shutts GJ, Vosper SB. 2011. Stratospheric gravity waves revealed in NWP model forecasts. *Q. J. R. Meteorol. Soc.* **137**: 303–317, doi:10.1002/qj.763.

Tennant WJ, Shutts GJ, Arribas A, Thompson SA. 2011. Using a stochastic kinetic energy backscatter scheme to improve MOGREPS probabilistic forecast skill. *Mon. Wea. Rev.* **139**: 1190–1206.

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

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 PL, Wilkinson J, Willett M, Woolnough S, 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, doi:10.5194/gmd-10-1487-2017.

Walters D, Wood N, Vosper S, Milton S. 2014. ENDGame: A new dynamical core for seamless atmospheric prediction. Technical report, Met Office. [Available online at: http://www.metoffice.gov.uk/media/pdf/s/h/ENDGameGOVSci_v2.0.pdf].

Wilks DS. 2011. *Statistical methods in the atmospheric sciences*. Academic Press, Elsevier: Amsterdam, 3rd edn.

Williams KD, Harris CM, Bodas-Salcedo A, Camp J, Comer RE, Copsey D, Fereday D, Graham T, Hill R, Hinton T, Hyder P, Ineson S, Masato G, Milton SF, Roberts MJ, Rowell DP, Sanchez C, Shelly A, Sinha B, Walters DN, West A, Woollings T, Xavier PK. 2015. The Met Office Global Coupled model 2.0 (GC2) configuration. *Geosci. Model Dev.* **8**: 1509–1524, doi:10.5194/gmd-8-1509-2015.

Wood N, Staniforth A, White A, Allen T, Diamantakis M, Gross M, Melvin T, Smith C, Vosper S, Zerroukat M, Thuburn J. 2014. An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Q. J. R. Meteorol. Soc.* **140**: 1505–1520, doi:10.1002/qj.2235.