Received: 11 December 2023 | Revised: 29 July 2024 | Accepted: 20 August 2024

DOI: 10.1002/qj.4856

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Predictability of midlatitude Rossby-wave packets

Isabelle Prestel-Kupferer[1] | **Michael Riemer**[1] | **Sören Schmidt**[1] | **Franziska Teubler**[2]

[1]Johannes Gutenberg University of Mainz, Mainz, Germany

[2]Rheinland-Pfalz Kompetenzzentrum für Klimawandelfolgen, Trippstadt, Germany

Correspondence
Isabelle Prestel-Kupferer, Johannes Gutenberg University, Institut for Physics of the Atmosphere, Mainz, Germany.
Email: iprestel@uni-mainz.de

Funding information
German Research Foundation (DFG), Grant/Award Numbers: RI 1771/4-1, SFB/TRR 165

Abstract

Rossby-wave packets (RWPs) along the midlatitude jet are a fundamental ingredient of extratropical dynamics. RWPs have been linked to enhanced atmospheric predictability and the occurrence of extreme weather events. We here investigate the predictability of Northern Hemispheric RWPs as physical entities, defined by enhanced values of the Rossby-wave envelope field. Using a catalogue of RWPs identified and tracked in reanalysis data, we analyze RWP predictability in a 19-year period of NOAA GEFSv12 reforecasts. Our analysis adopts the so-called distance and amplitude score (DAS), a verification metric that avoids both the double-penalty issue of field-based verification of coherent features and the complications of an object-based approach. Applied to the envelope field, forecast errors defined by this metric asymptote towards saturation, but do not completely reach saturation within the 10 days lead time available to this study. The growth rate of the median DAS is highest initially and decreases with lead time. This is a nontrivial result, because the underlying envelope field largely de-emphasizes phase information, but still exhibits very similar error-growth characteristics to fields that contain the full phase information. Variations in RWP predictability are dominated by the stage of the RWP life cycle, with higher predictability found for the propagation stage than the onset and decay stages. In addition, RWP predictability exhibits a seasonal cycle, with higher predictability in winter than in summer. Controlling for seasonality and the stage of the life cycle, we find that (i) high-amplitude RWPs exhibit higher predictability than low-amplitude RWPs up to 6 days lead time and (ii) there is a general pattern of relatively high predictability over Eurasia. Finally, predictability of the propagating stage is higher if forecasts are initialized after RWP onset than if initialized before onset. In this sense, RWP onset acts as a predictability barrier to the subsequent propagation stage.

KEYWORDS

ensemble reforecasts, error growth, forecast verification, predictability barrier

1 | INTRODUCTION

Rossby-wave packets (RWPs) propagating along the strong gradients of potential vorticity (PV) associated with jet streams (Wirth *et al.*, 2018) are a fundamental ingredient of atmospheric dynamics in the extratropics. Midlatitude weather systems are often embedded in RWPs, and the significance of these wave packets for weather forecasting has long been recognized (Cressman, 1948; Hovmöller, 1949). More recently, special attention has been given to the role of RWPs as precursors to high-impact weather events (e.g., Chang, 2005; Fragkoulidis *et al.*, 2018; Martius *et al.*, 2008; Piaget *et al.*, 2015; Shapiro & Thorpe, 2004).

It may be expected that RWPs, as large-scale flow features obeying balanced dynamics, exhibit a large degree of predictability, which may then be inherited by smaller-scale weather features (e.g., Grazzini & Vitart, 2015; Wirth & Eichhorn, 2014). On the other hand, forecast errors grow and spread within RWPs, affecting predictability severely in the downstream region (e.g., Anwender *et al.*, 2008; Baumgart *et al.*, 2018; Clarke *et al.*, 2019; Lojko *et al.*, 2022; Rodwell *et al.*, 2013; Zheng *et al.*, 2013). An important question that therefore arises is: Under which conditions do RWPs exhibit high and low predictability, respectively? This study works towards answering this question by assessing and characterizing RWP predictability in 19 years of reforecast data.

We here consider RWPs as coherent, physical entities described by the wave envelope. The group velocity, and thus large-scale energy propagation, relates to the wave envelope. It is this large-scale feature of RWPs that can be expected to exhibit higher predictability, rather than the phases of the individual troughs and ridges (Lee & Held, 1993). In order to assess the predictability of RWPs as physical entities, forecast verification needs to be focused on these coherent entities. Grid-point-based verification metrics for coherent features suffer the double-penalty issue (Baldwin & Kain, 2006; Ebert & McBride, 2000). Object-based methods have been developed to overcome this issue (e.g., Davis *et al.*, 2006; Ebert & McBride, 2000; Wernli *et al.*, 2008), and an object-based perspective is often applied to RWPs (e.g., Glatt & Wirth, 2014; Souders *et al.*, 2014; Wolf & Wirth, 2017).

There are, however, complications in the application of object-based methods to forecast verification, including the handling of missing objects in the forecast and a degree of arbitrariness in matching forecast and analysis objects when forecasts are poor. The verification metric used in this study, the so-called distance and amplitude score (DAS), seeks to avoid both the double-penalty issue and the complications of object-based methods. The DAS is a field-based method here applied to an upper tropospheric, two-dimensional field of the Rossby-wave envelope. The DAS measures the similarity of a forecast to an analysis using an image-matching procedure. The DAS is a single number that measures forecast skill. Object-based verification yields important information for forecast users and model developers, such as hit rates, false-alarm rates, biases in propagation speed, and amplitude biases (e.g., for cyclones, Froude *et al.*, 2007; Froude, 2010). This information is not provided by the DAS and, ingeneral, object-based approaches and the DAS are not directly comparable. Because the focus of the current study is not on model evaluation but on more general questions about atmospheric predictability. We consider it a virtue of the DAS, for our particular purpose that forecast skill is measured by a single number.

Our focus on the wave envelope differs from previous predictability studies that evaluated more traditional metrics of forecast skill, for example, area-averaged root-mean-square errors or anomaly correlation coefficients, conditioned on the presence or lack of RWPs (e.g., Grazzini & Vitart, 2015; Quinting & Jones, 2016). Object-based approaches that consider RWPs as physical entities are mostly concerned with model evaluation (Pérez-Fernández & Barreiro, 2023; Quinting & Vitart, 2019). One exception is found in the work of Pérez-Fernández and Barreiro (2023), who investigate, inter alia, the role of the Madden–Julian Oscillation in the (intrinsic) predictability of long-lived RWPs in the Southern Hemisphere. The current study is, however, to our knowledge the first to provide a comprehensive examination of basic features of RWP's atmospheric predictability.

The basic premise of this study is that, on average over a large number of cases, particularly good and bad forecast skill corresponds statistically to high and low atmospheric predictability, respectively. Our study investigates forecasts for over 1000 RWPs. We claim that this number is large enough to meet the basic premise. Furthermore, we assume that the model bias is small compared with the forecast errors. If this assumption holds, predictability characteristics are dominated by atmospheric conditions rather than by model errors. One well-established bias of numerical weather prediction models is an underestimation of Rossby-wave amplitude with increasing lead time (e.g., Gray *et al.*, 2014; Martínez-Alvarado *et al.*, 2018). In addition, Quinting and Vitart (2019) report an underestimation of RWP decay over the eastern Atlantic, and thus an overestimation of propagation into eastern Europe, which is linked to an under-representation of blocking onset over Europe. This bias, however, decreases with increasing model resolution and is small at the resolution of the forecast model considered in this study. For the forecast model examined in this study, the ratio between bias and mean (absolute) forecast errors of the Rossby-wave envelope is largest during

summer, but with an average small value of approximately 5% (shown for 5-day lead time in Doensen *et al.*, 2024, their fig. 8). In addition, we do not find an indication that a potential (small) overestimation of RWP propagation into Europe would affect our results. We thus accept the assumption that the model bias is sufficiently small, but acknowledge that quantitative details of our predictability signal may be affected by the use of an imperfect model.

The next section describes the reforecast data, the DAS as verification metric, and our adaptations to use the DAS, which has been developed for and previously applied to the verification of convective-scale precipitation and clouds, for RWPs. Our results are presented in Section 3, comprising general aspects of error growth as seen by the DAS, predictability characteristics as a function of the stage of the RWP life cycle, season, and RWP characteristics, and finally the interpretation of the onset stage as a predictability barrier to the propagation stage. We conclude with a summary and discussion in Section 4 .

2 | DATA AND METHODS

2.1 | Data

The variables considered in this study are horizontal wind and geopotential height at 300 hPa. Our analysis focuses on the Northern Hemisphere. We use Global Ensemble Forecast System version 12 (GEFSv12) Phase 2 (2000–2018) reforecasts (Guan *et al.*, 2022) verified against the unperturbed (control) reanalysis (Hamill *et al.*, 2022) to assess the predictability of RWPs. GEFS reforecasts have been used in previous predictability studies (Miller & Gensini, 2023; Ye *et al.*, 2023). The reforecasts are run with an approximate native horizontal resolution of 25 km and comprise a five-member ensemble initialized at 0000 UTC each day. Each member is verified deterministically by the method described below. Predictability of an RWP is then assessed by the mean of these five verifications. Reforecasts are available with one-day lead time increments on a 0.5° horizontal grid. We investigate lead times of 1–10 days. For the verification of RWPs, data are degraded to a 1° horizontal grid.

European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5: (Hersbach *et al.*, 2020)) on a 1° horizontal grid is used to identify episodes of RWP occurrence and to describe RWP characteristics. ERA5 is used because of the higher available temporal resolution than GEFS, which considerably reduces ambiguities in the tracking of RWPs. Our tracking uses 3-hourly data. Note that the mere purpose of the ERA5 data is to identify and characterize RWPs. Forecast verification is done completely within the Global Forecast System (GFS) model system. There are minor differences between the specific values of the envelope field in the two reanalyses, with a root-mean-square error of approximately $4\,\mathrm{m\cdot s^{-1}}$. The values of the envelope field associated with RWPs are usually in the range of $25–35\,\mathrm{m\cdot s^{-1}}$. Associated large-scale patterns are consistent between the two reanalyses (not shown). We have confirmed visually for a number of cases that the region containing an RWP identified in ERA5 is associated with locally enhanced values of the envelope field in GFS reanalysis (illustrated for one case in Figure 1a). We have not found any indication that our results are affected by the minor differences between ERA5 and the GFS reanalysis.

2.2 | Identification, tracking, and characterization of RWPs

The identification of RWPs in ERA5 reanalysis is based on the envelope of the 300-hPa horizontal wind following a background state, essentially as in Zimin *et al.* (2006). We define a slowly evolving background state by using a Lanczos filter (Duchon, 1979) with a cutoff period of 40 days and a window width of 80 days, which yields to results comparable to the 20-day moving average used by Zimin *et al.* (2006). For the further procedure, we follow Wolf and Wirth (2017) and use the component of the horizontal wind perpendicular to the streamlines of the background state. A semi-geostrophic coordinate transformation is applied to this wind signal (in the wave number range 4–17) to reduce the asymmetry in the spatial scale of troughs and ridges, and thus to reduce spurious fracture of RWPs. Finally, the envelope E is calculated by a Hilbert transform for wavelengths that correspond to a wave number range between 4 and 15 in the midlatitudes (at 50°N).

Distinct RWP objects are defined by thresholding the envelope field E. Wolf and Wirth (2017) introduced a time-varying threshold, which adopts to temporal variations of a background envelope field. This background field is defined as the hemispheric mean between 15° and 85°N, E_{15-85}. We here employ the version of the threshold that Wolf and Wirth (2017) recommend for tracking RWPs (their eq. A2). This (time-varying) threshold τ is defined by

$$\tau = \tau_0 + \frac{2\Delta}{\pi}\arctan\left(\frac{w_1 E_{15-85} - \tau_0}{w_2 \Delta}\right), \qquad (1)$$

with empirically determined, fixed parameters $\tau_0 = 20\,\mathrm{m\cdot s^{-1}}$, $\Delta = 10\,\mathrm{m\cdot s^{-1}}$, $w_1 = 3.0$, and $w_2 = 0.8$. With these parameters, the threshold may theoretically vary between $10\,\mathrm{m\cdot s^{-1}}$ and $30\,\mathrm{m\cdot s^{-1}}$. On average, values lie between

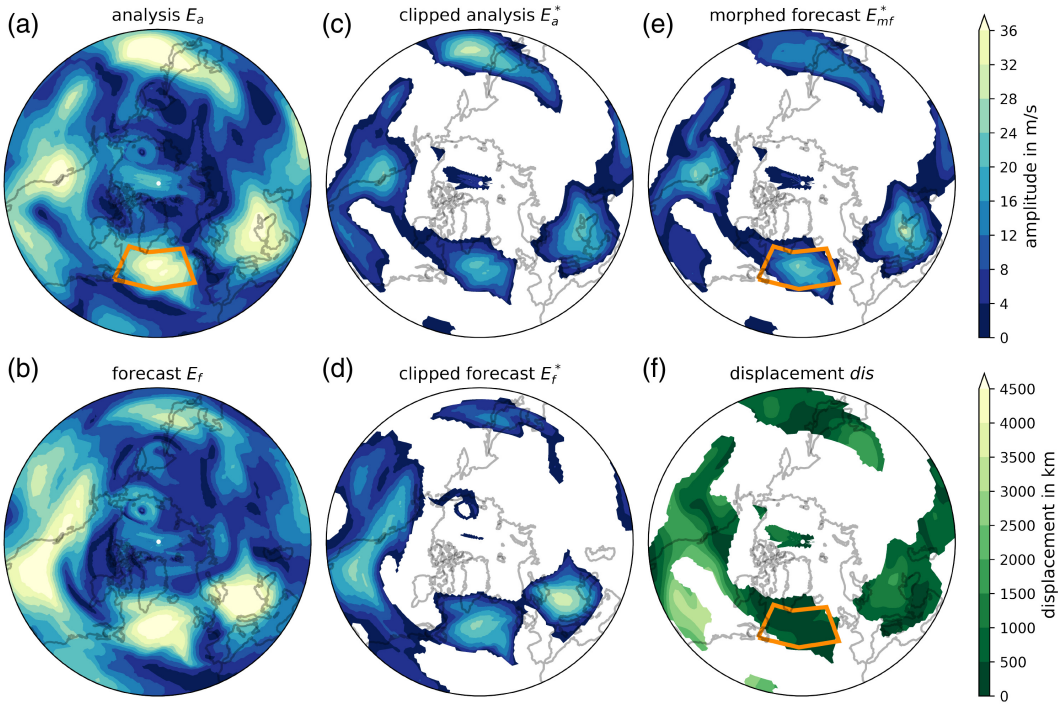

FIGURE 1 Illustration of the procedure for the DAS computation. The first column shows the field of the full envelope in (a) the analysis and (b) the corresponding forecast. The second column shows the envelope field after subtracting the (spatially constant and temporally varying) noise level E_{bg} and setting negative values to 0 (here denoted as a "clipped" field) for (c) the analysis and (d) the corresponding forecast. In the last column, (e) shows the result of matching the clipped forecast field to the clipped analysis field and (f) the resulting displacement field. The orange box in panels (a), (e), and (f) depicts the region of the associated RWP that is used to attribute a DAS value to that specific RWP.

21 $m \cdot s^{-1}$ (August) and 27 $m \cdot s^{-1}$ (November and April, not shown).

Tracking of RWP objects provides temporal coherence and is used to define the life cycle of RWPs. We apply the method employed in Hauser *et al.* (2022), which uses a criterion of spatial overlap between time steps and handles merging and splitting events. In the case of splitting, the object with the higher area-integrated amplitude is defined as the continuation of the original RWP. The object with the weaker amplitude defines the generation of a new RWP. In the case of merging, the RWP with higher area-integrated amplitude is defined to continue, whereas the object with weaker amplitude is defined to decay.

Defining RWP life cycles, each RWP is characterized by a specific life time, defined as the time between the first and the last detection of the associated RWP object. Furthermore, we define distinct stages of the RWP life cycle. We define the onset stage of RWPs as the time at which the RWP is detected for the first time, or is generated by a splitting event. We define the decay stage of RWPs as the time at which the RWP is detected for the last time, or decays in a merging event. We define the propagation stage of RWPs as the central time of its life time. In the case of a life time of even-numbered days, RWP characteristics are interpolated linearly to the central life time. For a robust distinction between life-cycle stages, we focus on RWPs that exhibit a life time of at least 5 days, such that the propagation stage is separated from onset and decay by at least 2 days. In addition, consistent with previous work, we omit the rare outliers of excessively long-lived RWPs and restrict our attention to RWPs with a life time of 14 days or less. At each time step during the life cycle, the RWP object is characterized further by its size, amplitude, and position. Size is defined as the area of the RWP object, that is, the area of the region in which the value of the envelope exceeds the threshold τ. Amplitude is defined as the value of the envelope averaged over the region of the RWP object *minus* a seasonally varying envelope threshold (E_{bg}, introduced below for the calculation of the DAS). Position is defined as the envelope-weighted center of mass of the RWP object.

The sample of RWPs that underlies this study is largely consistent with climatologies of RWPs presented in previous studies. Specifically, we have verified that seasonal and longitudinal variations of RWP onset and decay are consistent with Grazzini and Vitart (2015). Onset and decay constrain the propagation stage and it can be assumed that a good agreement between onset and decay characteristics implies a good agreement more generally. A visual indication of onset and decay regions can be found

later in Figures 6 and 7. In addition, the monthly ratio of long-lived versus short-lived RWPs is also consistent. Furthermore, differences between the onset longitude of long-lived and short-lived RWPs are consistent with results presented in Glatt and Wirth (2014). Differences between our sample and the climatology of Grazzini and Vitart (2015) occur in the monthly variation of total RWP numbers. Our sample contains, in relative terms, less winter than summer cases. The ratio between winter and summer cases in our sample is approximately 0.6, whereas we estimate that ratio to be 0.85 in Grazzini and Vitart (2015). This difference implies that year-round averages presented in this study may be biased towards predictability characteristics of summer cases. To estimate the magnitude of this potential bias, we compared results computed separately for the extended summer and winter season (May–September and November–March, respectively; not shown). The magnitude of the differences between seasons implies a potential bias with a magnitude that is similar to that of the standard error of the mean presented in the relevant figures and is not larger than the uncertainty range that arises due to setting free parameters for the computation of the DAS, as described in the following subsection. We are confident that none of our main results is affected by a potential bias towards summer cases in our sample.

2.3 | Verification of RWP forecasts: the DAS

The DAS provides a measure of the similarity of two fields, taking into consideration similarity of position, shape, and amplitude. We first describe general concepts of the DAS, before we discuss our specific DAS computation for RWPs. Figure 1 serves as illustration for the overall procedure.

The basis for the DAS is an image-matching procedure performed by so-called pyramidal matching, an algorithm originating from the field of computer vision. In a meteorological context, the algorithm is described and illustrated in detail by Zinner *et al.* (2008). Pyramidal matching aligns images by comparing coarse-grained representations of the images recursively, with a decreasing level of coarse-graining at each step. This hierarchical approach thus first generates a large-scale displacement vector field, which is then refined progressively for efficient and robust matching. For illustration, the result of matching a forecast field (Figure 1d) to the associated analysis field (Figure 1c) is depicted in Figure 1 e.

The magnitude of the displacement vector field (illustrated for our example in Figure 1f) provides the basis for the displacement score D. The amplitude differences between the matched fields (in our example, the differences between Figures 1c and 1e) provides the basis for the amplitude score A. Upon suitable normalization, these two scores are combined into a single DAS value:

$$DAS := \frac{D}{D_{\max}} + \frac{A}{A_{\mathrm{clim}}}, \tag{2}$$

with normalization factors $D_{\max}$ and A_{clim}. In the context of forecast verification, that is, matching a forecast with an analysis field, a DAS value of 0 identifies a perfect forecast. A specific upper limit of the DAS does not exist, but the idea of the normalization is that the maximum value of both individual terms is approximately 1, such that the upper limit of the DAS is approximately 2.

It is important to note that the individual metrics D and A do not correspond directly to a position and amplitude error, respectively, of an RWP. In addition to mere displacement of the envelope field, differences in the shape also contribute to D. Importantly, the image-matching algorithm may attempt to account for amplitude differences by shrinking or extending parts of the envelope field, thereby projecting amplitude difference onto the D metric (Keil & Craig, 2009, their figs. 3 and 4). A straightforward interpretation of the individual terms is thus not feasible. The DAS, however, as the combination of the A and the D metric, provides a robust metric for the similarity of fields.

The DAS was developed for application to convective-scale clouds and precipitation (Keil & Craig, 2009; Zinner *et al.*, 2008) and assumed planar geometry and isotropy regarding displacement. For our application to RWPs, we use the Rossby-wave envelope field in longitude and latitude, $E(\lambda, \phi)$, and account for the convergence of the meridians towards the pole when performing spatial integration and when computing the amplitude of the displacement vector. Furthermore, we introduce anisotropy regarding displacements, which is motivated by the preferred zonal extent and zonal propagation of RWPs. The image-matching algorithm operates on the full Northern Hemisphere and matches the forecast field to the analysis field within a given search distance. Anisotropy is introduced in this search distance, allowing the algorithm to perform larger displacements in the zonal than in the meridional direction, with a ratio that we set to 3. This ratio needs to be an integer due to algorithmic constraints and is motivated by the mean aspect ratio of RWPs during the propagation stage (the method described in Section 2.2 identifies a mean extent in longitude and latitude of 53° and 17°). An alternative, reasonable choice, a factor of 4 motivated by the ratio of the zonal and meridional scale of a hemisphere (in degrees), yields consistent results (not shown). Reasonable choices of the search distance are of the same scale as the analyzed signal and are also constrained by algorithmic requirements. Under these constraints, reasonable choices for RWPs are 64° and 16°, 32° and 8°, and 48° and 16° in the zonal and

meridional direction, respectively. We have tested all three options and found no qualitative and little quantitative sensitivity of our results. The presented results have been computed with zonal and meridional search distance of 48° and 16°, respectively.

To avoid excessive displacement of small-amplitude features ("noise") by the matching algorithm, we filter noise from the envelope field by subtracting a seasonally varying climatological background value E_{bg} and setting negative values to zero. The effect of this noise filter is illustrated in Figure 1 by comparing the fields in the first column (Figure 1a,b) with those in the middle column (Figure 1c,d). We define E_{bg} as the daily average (from 2000–2018) of the envelope field spatially averaged between 40°N and 60°N, smoothed by a 30-day running mean. Our noise level E_{bg} is thus a fixed value in space but (slowly) varies with the calendar day. This seasonal variation is very similar to that of the threshold τ to define RWP objects, but E_{bg} is about 10 $\mathrm{m{\cdot}s^{-1}}$ smaller than τ (not shown). We deliberately choose a value for E_{bg} that is substantially smaller than τ, because the DAS then diagnoses the underprediction of RWP amplitude more sensitively. Our results are not sensitive to the specific value of E_{bg} (tested within a range of ±20%).

The DAS for a given RWP is computed from the results of the image matching (Figure 1e,f) in a region defined by the extent of the RWP object, identified as described in Section 2.2. This region is defined as a rectangle (in λ and ϕ) encompassing the zonal and meridional extent of the RWP object (orange box in Figure 1). Our results are quantitatively sensitive to the definition of this region, in the sense that regions that are more focused on the large envelope values associated with RWPs yield larger differences in the predictability characteristics that we will present below. We did not find any qualitative sensitivities. To yield a sharper signal, a region that focuses on large values of the RWP envelope is thus desirable.

The displacement and the amplitude scores are defined following Keil and Craig (2009). The amplitude score A is defined as the root-mean-square difference of the noise-subtracted envelope field E^* of the analysis and the matched forecast, averaged over the rectangle encompassing the RWP:

$$A := \sqrt{\frac{1}{a}\int_{\mathcal{R}} (E^*_{\mathrm{a}}(\lambda,\phi) - E^*_{\mathrm{mf}}(\lambda,\phi))^2 \cos(\phi)\, d\lambda\, d\phi}, \quad (3)$$

where a is the area of the rectangle $\mathcal{R}$ encompassing the RWP, $E^*(\lambda,\phi) := E(\lambda,\phi) - E_{\mathrm{bg}}$ is the noise-subtracted envelope field, and the indices a and mf refer to the analysis and the matched forecast, respectively. The factor $\cos(\phi)$ accounts for the convergence of the meridians. The displacement score D is defined as the magnitude of the displacement vector, weighted by the noise-subtracted envelope of the analysis, and averaged over the rectangle encompassing the RWP:

$$D := \frac{\int_{\mathcal{R}} |\mathbf{d}(\lambda,\phi)|\, E^*_{\mathrm{a}}(\lambda,\phi)\cos(\phi)\, d\lambda\, d\phi}{\int_{\mathcal{R}} E^*_{\mathrm{a}}(\lambda,\phi)\cos(\phi)\, d\lambda\, d\phi}, \quad (4)$$

where $\mathbf{d}$ denotes the displacement vector field. This weighting by the underlying field deviates from Keil and Craig (2009) and is introduced here to reduce further the sensitivity of the results to the specific threshold used to define noise.

Finally, D and A are normalized to yield comparable values that can be combined to yield a single DAS value. The basic idea of the normalization is that the maximum displacement score should equal the amplitude score of forecasts that miss an observed feature with characteristic amplitude. The normalization is chosen such that a forecast that completely misses an observed feature with characteristic amplitude yields DAS = 1. To compare forecasts of a large number of RWPs, a climatological value A_{clim} is used for normalization, which we compute as the daily average over the investigated period 2000–2018 of the RWP envelope field that exceeds the noise level ($E(\lambda,\phi) - E_{\mathrm{bg}} > 0$), spatially averaged between 40°N and 60°N, and smoothed by a 30-day running mean, that is, the computation follows that of the noise level itself. This procedure yields an average normalized value of $A/A_{\mathrm{clim}} = 0.67$ at the longest considered lead time of 10 days (not shown). At long lead times, when forecasts have lost skill, errors can be expected to saturate on average at half of the value of the maximum error. Our normalized amplitude score at lead time 10 days is consistent with this expectation. For D_{max}, we choose a fixed value such that, on average, the normalized displacement score is equal to the normalized amplitude score at the longest considered lead time of 10 days. The idea of this approach is that both parts of the DAS should contribute equally at saturation, and that forecasts at 10 days lead time have lost on average enough skill that they can be considered to be close to saturation. The approach yields a value of $D_{\mathrm{max}} = 2590$ km, which corresponds to approximately 36° at 50°N. D_{max} is thus of the same scale, albeit 25% smaller at this latitude, than the search distance in the zonal direction, which is largely consistent with theoretical considerations under several simplifying assumptions (Keil & Craig, 2007).

3 | RESULTS

Our results are based on a sample of 1147 RWPs during the period 2000–2018. All characteristics of the results highlighted below, except for the spatial patterns discussed

in Section 3.4.1, have been tested for statistical significance by a *t*-test and have passed the 95% significance level.

3.1 | General characteristics of error growth in terms of the DAS

Figure 2 presents the average evolution of the DAS with lead time. DAS values exhibit exponential growth at short lead time (up to 72 h) when errors are small. Subsequently, growth rates decrease monotonically with lead time and, apparently, the DAS approaches a saturation value asymptotically. The evolution of the DAS with lead time thus exhibits similarity with logistic growth (e.g., Lorenz, 1982). At 10-day lead time, the longest lead time available to us, the median DAS value is 1.35 with an interquartile range of 1.17–1.56. While there is substantial slowdown of error growth at longer lead times, the DAS on average still increases between 9 and 10 days lead time, indicating that saturation of forecast errors in terms of the DAS has not yet been fully reached after 10 days.

Using the envelope field for verification largely de-emphasizes phase information of the underlying wave pattern. Still, the evolution of the envelope-based DAS growth rates with lead time—decreasing monotonically from an initial maximum value—is very similar to the results of many previous studies of error growth in a more general context considering error fields that contain the full phase information (e.g., Dalcher & Kalnay, 1987; Lorenz, 1982; Žagar *et al.*, 2017). This similarity is a non-trivial result, because it is reasonable to hypothesize that phase and amplitude errors develop differently with lead time. While this difference has not yet been demonstrated explicitly, a decomposition of (phase-information containing) forecast errors by Jankov *et al.* (2021) indicates that at early lead times, when errors grow most rapidly, phase errors dominate. Amplitude errors, according to their decomposition, become of comparable importance only later on. The results of this decomposition are consistent with the lead-time dependence of error-growth mechanisms. Error growth in current-day medium-range weather forecasts is dominated by the nonlinearities in the Rossby-wave dynamics (Baumgart *et al.*, 2018; Baumgart & Riemer, 2019; Selz *et al.*, 2022), which arguably project predominantly on location errors of individual troughs and ridges, that is, phase errors. As lead time increases, error-growth mechanisms associated with moist baroclinic development, which constitutes the main contribution to RWP amplification (Teubler & Riemer, 2021), tend to increase (the 'divergent' and 'tropospheric-deep' terms in Baumgart *et al.*, 2018; Baumgart & Riemer, 2019). In contrast, the contribution of the amplitude score A to the DAS is larger than that of the displacement score D at lead times up to day 6 (not shown). Bearing the caveats of interpreting A and D individually in mind, this result at least indicates that positional errors of the envelope field, analogous to phase errors of troughs and ridges in Jankov *et al.* (2021), do not dominate the initial error growth. The distinct qualitative similarity of error growth as diagnosed from the envelope-based DAS and from error metrics that incorporate phase errors fully is thus not necessarily expected.

3.2 | Dependence of predictability on the stage of the RWP life cycle

Rossby-wave packets undergo a distinct life cycle of onset, propagation, and decay. Conceptually, these stages can be associated with different dominating dynamical processes: moist baroclinic growth during onset, quasi-linear Rossby-wave propagation during propagation, and Rossby-wave breaking during decay (Wirth *et al.*, 2018). It is thus plausible that the stage of the life cycle at verification time may impact predictability. At this point in particular it should be kept in mind that model biases are ingeneral flow dependent and may affect the

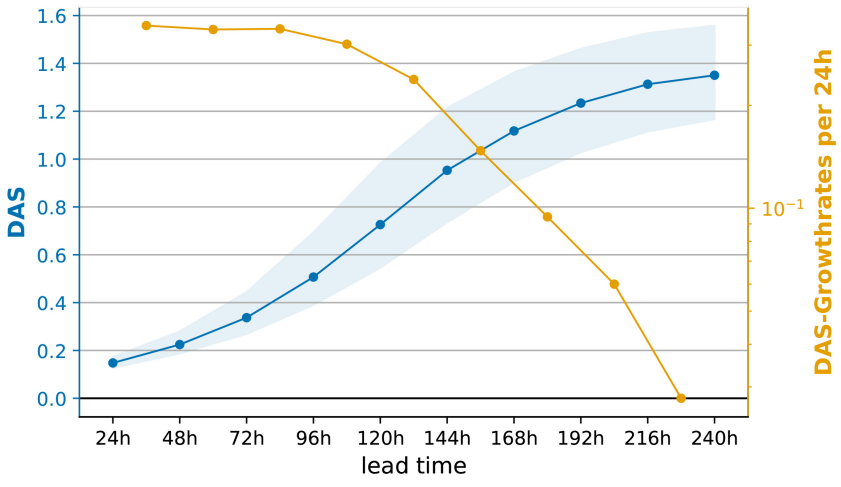

FIGURE 2 RWP predictability and error growth rates according to lead time. Shown are the median of the DAS distributions (blue) and an indication of the width of the distribution by the interquartile range (blue shading). The corresponding growth rates, defined as the finite time difference between two consecutive lead times of the median DAS and normalized by the respective DAS value, are shown in orange.

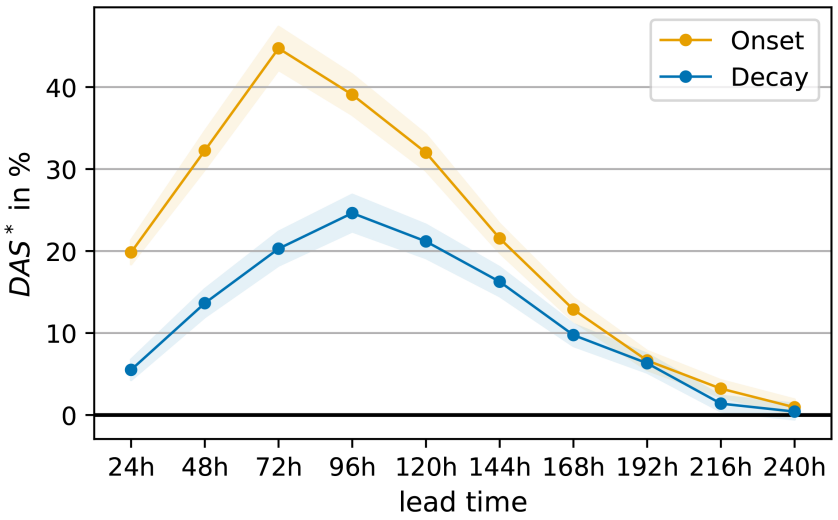

FIGURE 3 RWP predictability for different life-cycle stages. Shown are relative differences of the DAS for the onset $DAS^* = (DAS_{\text{Onset}} - DAS_{\text{Prop}})/DAS_{\text{Prop}}$ (orange) and decay $DAS^* = (DAS_{\text{Decay}} - DAS_{\text{Prop}})/DAS_{\text{Prop}}$ (blue) compared with the propagation stage. The corresponding standard errors are represented by the shading.

different stages of the RWP life cycle to different degree. This flow dependence may affect the quantitative details of our results.

Figure 3 depicts the DAS of the onset and decay stages, respectively, relative to that of the propagation stage of the same RWP. Evidently, the onset stage is the least predictable stage, with difference in the DAS relative to the propagation stage maximizing at 45% at 72-h lead time. The decay stage is also less predictable than the propagation stage, with relative differences maximizing at 25% at 96-h lead time. The results thereby indicate that quasi-linear Rossby wave propagation exhibits higher predictability than both moist-baroclinic growth the highly nonlinear process of Rossby wave breaking, as may have been expected. Rossby wave breaking at the end of an RWP's life cycle is preceded by the relatively well predictable propagation stage. In contrast, moist-baroclinic growth at the beginning of the life cycle can in general not be expected to exhibit precursor dynamics with suchrelatively high predictability, which may partly explain the lower predictability of onset compared to decay. Our results, however, do not allow a comparison of the predictability of Rossby wave breaking vs. moist-baroclinic growth per se, i.e., independent of the predictability of the preceding processes. The relative differences between life-cycle stages maximize for short medium-range lead times and decrease gradually for longer lead times. Approaching error saturation at longer lead times evidently diminishes differences between stages. In this sense, forecasts for all stages of the life cycle apparently have equally little skill at 216- and 240-h lead time.

3.3 | Seasonal variations in RWP predictability

General verification scores of operational centers (e.g., Haiden *et al.*, 2022) typically exhibit a seasonal cycle, indicating lower predictability during summer than winter. In addition, characteristics of RWPs (Glatt & Wirth, 2014; Souders *et al.*, 2014) and the general amplitude of the Rossby-wave envelope (as defined here in Section 2) vary with season. Here, we first examine the seasonality of RWP predictability. The potential importance of RWP characteristics will be discussed below.

Figure 4 presents the relative differences with lead time between the DAS for extended summer (May–September) and extended winter (November–March), individually for all three life-cycle stages. The DAS during summer is on average higher than that during winter for all stages, indicating in general less predictability of RWPs during summer than during winter. Similarly to the differences between the life-cycle stages, relative differences tend to be largest in the early medium range. For the decay stage, relative differences are largest and around 15% up to day 5, diminishing for longer lead times. For this stage, relative differences of the DAS do not grow substantially during early lead times, in contrast to the differences between life-cycle stages in Figure 3 . For the onset stage, seasonal differences are smallest, with a maximum of 10% at 4 days lead time. Maximum values for the propagation stage are between 10% and 15% from day 4–6.

Averaging over forecasts verifying at all three stages of the RWP life cycle, the relative difference of the DAS from the yearly average is depicted in Figure 5 as a function of month. This monthly perspective confirms a distinct seasonal cycle, with maximum amplitude for early medium-range lead times and decreasing amplitude for longer lead times. The largest positive relative difference (around 10%), and thus lowest predictability, occurs during July and August. Highest predictability (around −5% relative differences) for the early medium range is found rather consistently from October–April, with the exception of January. For longer lead times, the largest signal is found during late summer (August and September),

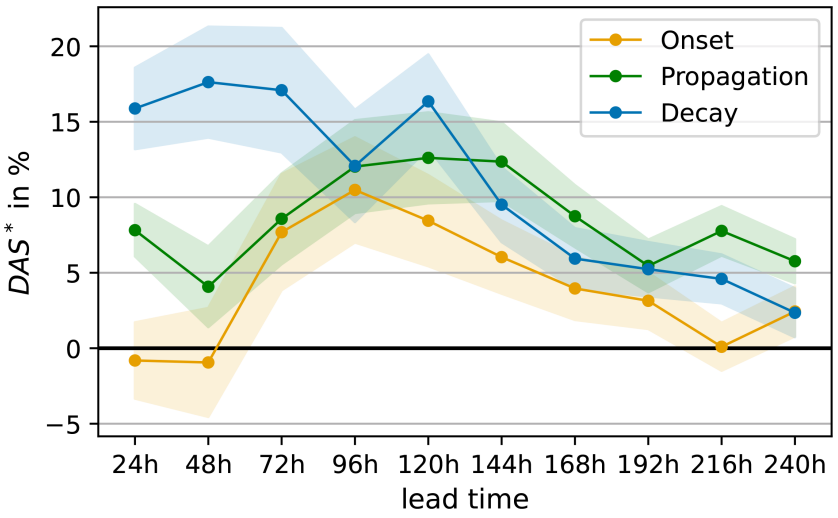

FIGURE 4 Difference of RWP predictability between extended summer and winter. The relative differences $DAS^* = (DAS_{\text{ext. summer}} - DAS_{\text{ext. winter}})/DAS_{\text{ext. winter}}$ are shown separately for each life-cycle stage: onset (orange), propagation (green), and decay (blue). The shading shows the corresponding standard error.

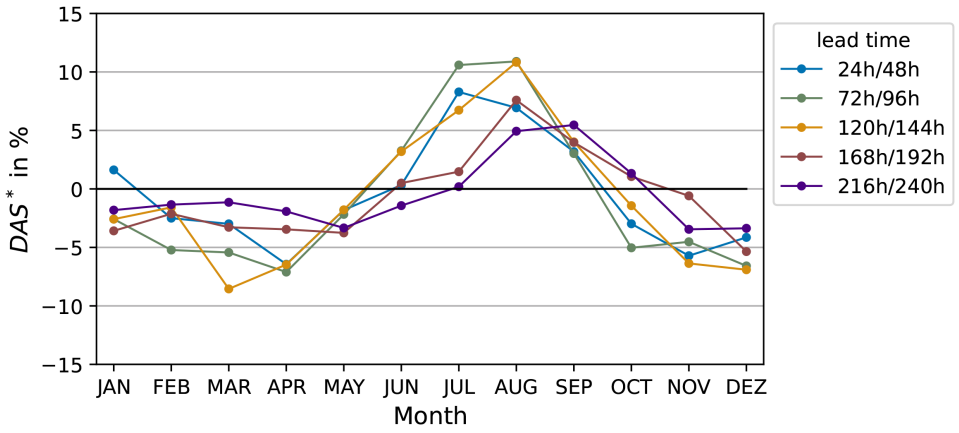

FIGURE 5 Seasonality of RWP predictability. Shown are relative values compared with its climatology $DAS^* = (DAS - DAS_{\text{clim}})/DAS_{\text{clim}}$ for different lead-time groups (colors).

shifting to early fall. Based on the relative differences found in Figures 4 and 5, medium-range predictability depends less strongly on season than on the stage of the RWP life cycle (cf. Figure 3).

3.4 | RWP characteristics and predictability

This subsection investigates the association of predictability with the RWP characteristics position, amplitude, size, and life time, defined as described in Section 2.2. We consider the three life-cycle stages individually and account for seasonality by stratifying data for each month separately before accumulating year-round. When stratifying data, here we consider only RWPs that occur in a meridional band representative of the main occurrence region of the centroids of RWPs. The rationale for this restriction is that stratifying data focuses on the tails of the distribution of RWP characteristics. Preliminary analysis indicated that the tails of the distributions tend to be associated with unusually high or low meridional position. The focus on a representative meridional band is thus introduced to avoid an over-representation of these unusual positions. We present results for 30°–50°N, which comprises more than 90% of all cases. Results are qualitatively similar for the meridional band of 30°–60°N, which comprises approximately 99% of all cases.

3.4.1 | Geographical characteristics

To investigate geographical characteristics of RWP predictability, we need to split the sample of RWP forecasts into different geographical regions, which reduces sample size drastically compared with the analyses presented above. With this reduced sample size, investigating geographical differences in the mean or median DAS is affected by the statistical effect that small samples exhibit the largest variations in the estimates of their mean values, and thus the most prominent extrema (not shown). We consider these extrema to be statistical artefacts, which in addition tend to emphasize regions with small RWP occurrence.

To avoid this issue, we define a metric of predictability as an alternative to the mean DAS and consider the occurrence frequency of "good" and "bad" forecasts, which we define as forecasts with DAS values in the upper and lower terciles of the hemispheric DAS distribution, respectively. We define regions in which good forecasts occur more often than bad forecasts as regions with relatively high RWP predictability, and vice versa.

To provide a spatial map of RWP predictability, we attribute each forecast to a region around the centroid of the associated RWP object. To partly account for the anisotropic nature of RWPs, we define this region as an ellipse with half axes of 10° and 5° in the zonal and meridional direction, respectively. To de-emphasize regions with small RWP occurrence, we examine absolute (not relative) differences in the number of good and bad forecasts. To increase the sample size, we aggregate forecasts with lead times of 4–6 days and 8–10 days, respectively. We show and discuss results only in regions (i) that comprise at least 20 forecasts at each grid point (10 for the extended seasons) and (ii) in which average relative differences (the difference in the number of good and bad forecasts divided by the total number of forecasts) exceed at least 15%.

We first focus on medium-range forecasts (4–6 days lead time), because the results above indicate that both relative and absolute differences in predictability can be expected to be relatively large in this lead-time range (Figure 6). For the onset stage, relatively low RWP predictability is predominantly found in the main onset region over the North Pacific. By contrast, the North Atlantic onset region off the east coast of North America exhibits relatively high RWP predictability. A further coherent region of relatively high predictability extends from the eastern Atlantic over Europe. For the decay stage, the predominant signal is a dipole pattern with relatively high predictability over Eurasia and relatively low predictability extending from the North Pacific over North America into the central North Atlantic.

For the propagation stage, predictability characteristics during extended summer and extended winter seasons partly oppose each other. The year-round perspective presented for the onset and decay stages is thus of limited use to describe these characteristics. We thus discuss both seasons separately. During winter, the main occurrence region mostly exhibits relatively low predictability (from the easternmost North Pacific, over North America, and into the central North Atlantic). By contrast, the region upstream (from the central to the eastern North Pacific) exhibits relatively high predictability. Predictability is also relatively high over Eurasia. During summer, the pattern is more complex. The clearest signal occurs in the North Pacific, predominantly with relatively low predictability. Over the continents, relatively high predictability dominates, in particular over Eurasia, which is consistent with the results of Chang (1999) who showed that RWP propagation is most coherent over Eurasia in both winter and summer. Over the main occurrence region in the North Atlantic region, we consider our results to be inconclusive.

As discussed in the Introduction, RWPs as large-scale atmospheric features are often expected to exhibit predictability at longer lead times than synoptic-scale features. It is therefore also of interest to investigate relative RWP predictability at longer lead times (here, 8–10 days). The results for this longer lead-time range are largely consistent with those of the shorter 4–6 day lead time range (Figure 7, cf. Figure 6). We thus focus our discussion on the differences between the lead-time ranges. For the onset stage, there is now relatively high predictability polewards and just downstream of the main onset region in the North Pacific, which still exhibits relatively low predictability. Relatively high predictability is also found just upstream and downstream of the onset region off the east coast of North America. Finally, the signal of relatively high predictability over Eurasia at longer lead times is found more clearly over central Asia than over Europe. For the decay stage, the signal of relatively low predictability over the eastern North Pacific has vanished and the pattern in the North Atlantic–European region now exhibits a meridional dipole with relatively low predictability Equatorwards and relatively high predictability polewards of the main decay region. For the propagation stage, during winter, a clear signal over central Asia and the central North Pacific is missing and during summer a localized region of relatively high predictability is found amidst the general region of relatively low predictability over the North Pacific.

In summary, geographical characteristics of RWP predictability depend on the stage of the life cycle, partly on season (as illustrated for the propagation stage), and to a much lesser extent on the lead-time range considered here. For the same stage of the life cycle, characteristics vary between the ocean basins of the North Pacific and the North Atlantic, and in particular between the ocean basins and the large continental region of Eurasia. The rather consistent signal of relatively high predictability over Eurasia indicates that RWPs over land masses, "continental" RWPs for which moist processes arguably play a small role, are more predictable than their "maritime" counterparts, which are arguably impacted by moist processes to a much larger extent. Substantiating this indication, as well as achieving understanding of other geographical differences, however, requires further research that is beyond the scope of the current study.

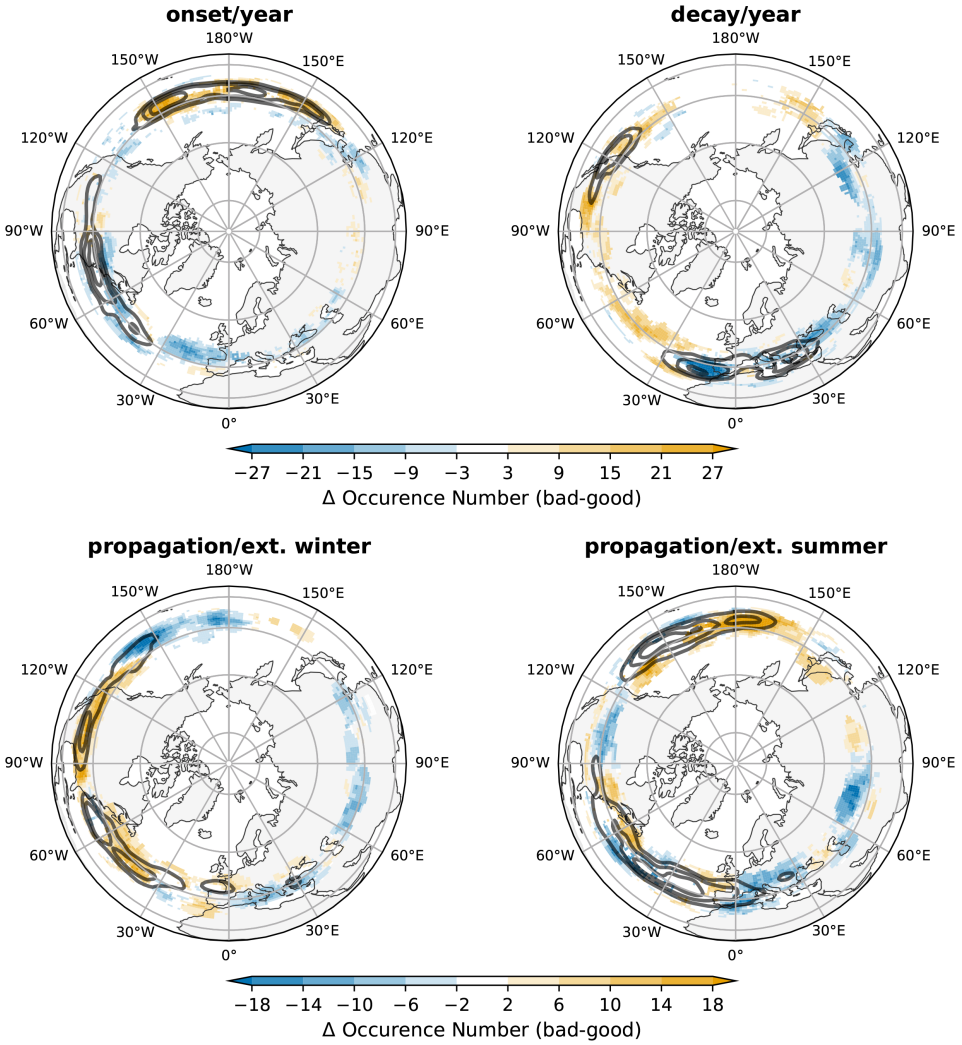

FIGURE 6 Difference in the occurrence number of forecasts exhibiting DAS values in the upper and lower terciles ("bad − good" forecasts) for the lead-time range of 4–6 days. Blue (orange) denotes relatively high (low) predictability. Contours show climatological occurrence numbers, values drawn for onset and decay: 80, 100, 120. For propagation ext. winter: 30, 40, 50 and for ext. summer: 40, 60, 80.

3.4.2 | Amplitude and size

We consider RWP amplitude and size during the propagation stage only. Amplitude and size are not useful metrics to characterize RWPs during the onset and decay stage, because these stages are defined as the first and last time step, respectively, when an RWP is detected. In the limit of infinitesimal temporal resolution of the available data, amplitude and size would thus go to zero. We find very similar predictability characteristics, but with a 5%–10% stronger signal for amplitude than for size (not shown). We attribute this similarity to the correlation between amplitude and size (with $r = 0.59$, not shown). For the sake of brevity, here we focus the presentation on our results for RWP amplitude.

Similarly to good and bad forecasts, we define "weak" and "strong" RWPs in terms of upper and lower percentiles of the amplitude distribution. We investigated terciles, quartiles, and quintiles (Figure 8). Up to six days lead time, strong RWPs exhibit higher predictability than weak RWPs, with the largest differences again found during the short medium range: at three days lead time, relative differences are above 25% for the quintiles and quartiles and around 20% for the terciles of the RWP amplitude distributions. In general, for lead times up to day 6 the differences increase systematically for the more extreme tails of the distribution, that is, from terciles to quintiles. Differences in RWP predictability due to amplitude can thus be considered somewhat less pronounced than differences due to the stage of the life cycle, but more pronounced than differences due to the seasonal cycle (cf. Figures 3 and 4, respectively). For lead times of 7 days and longer, the signal changes sign, that is, weak RWPs are then more predictable than strong RWPs. Absolute relative differences increase up to around 10% at 240-h lead time, with hardly any differences between the different percentiles considered.

We argue that this change of sign does not denote a predictability characteristic of the atmosphere, but rather relates to the choice of specific parameters used in the identification of RWP packets. When forecasts lose skill at longer lead time, the predicted amplitude of RWPs tends towards a climatological average. The alleged higher predictability of weak RWPs indicates that our samples of

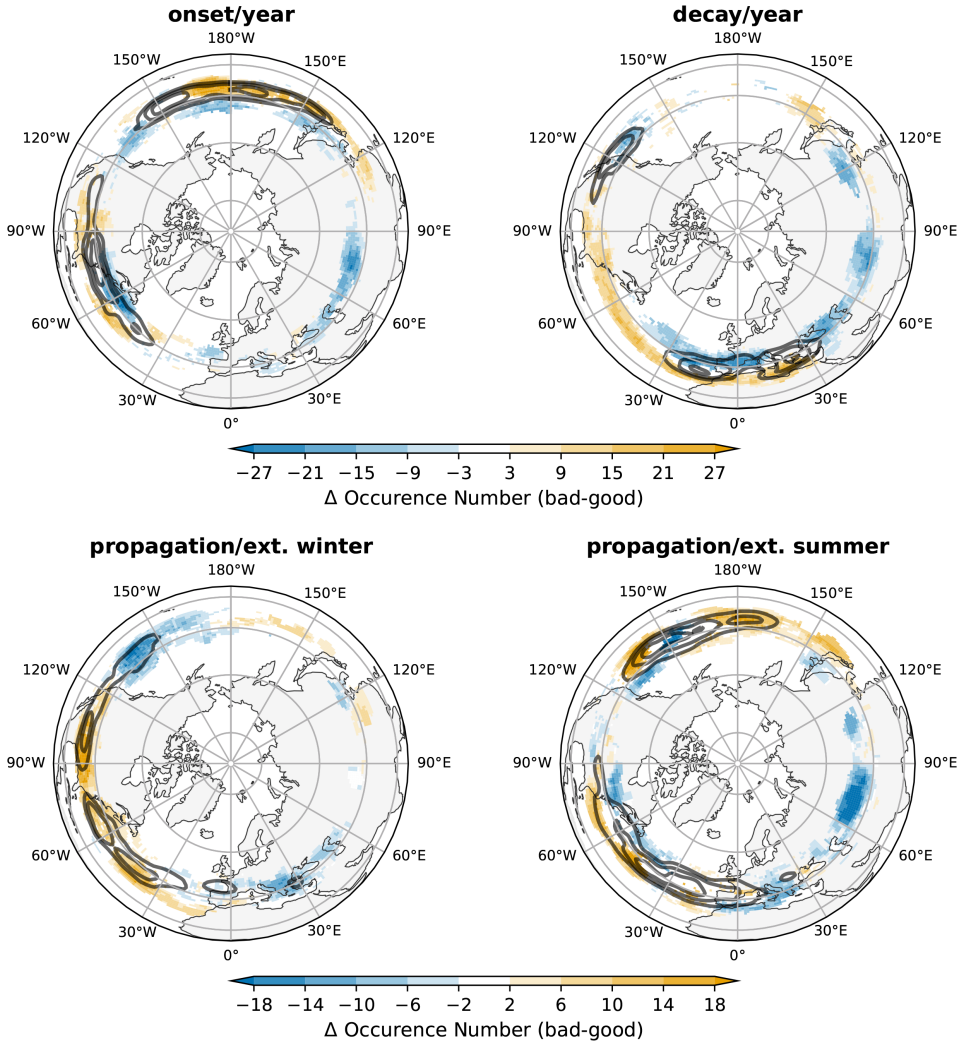

FIGURE 7 Same as Figure 6, but for the lead-time range of 8–10 days.

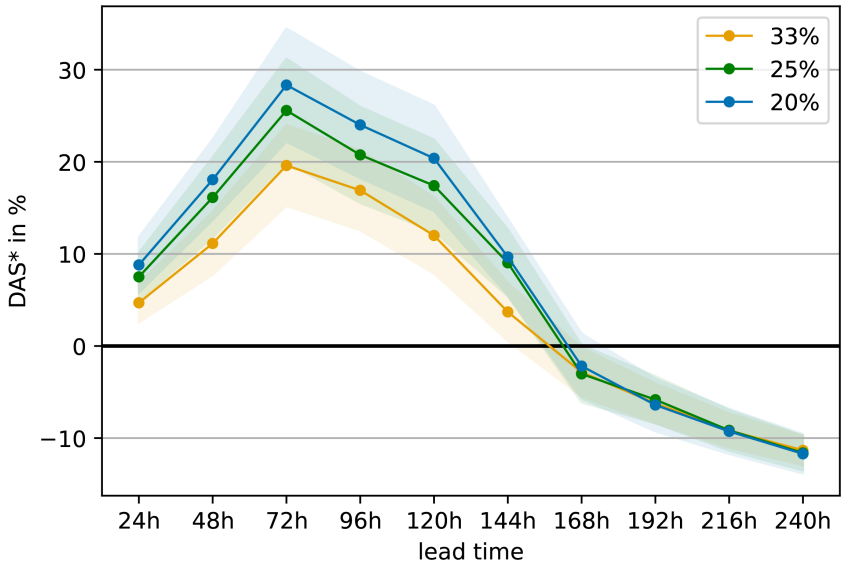

FIGURE 8 RWP predictability for weak and strong amplitude RWPs. Shown are the relative differences $DAS^* = (DAS_{\text{weak}} - DAS_{\text{strong}})/DAS_{\text{strong}}$ for upper and lower terciles (orange), quartiles (green), and quintiles (blue), respectively. Shading indicates the standard error.

weak RWPs are on average closer to that climatological average than our samples of strong RWPs. We thus interpret the higher predictability of weak RWPs at long lead times as an artificial signal. This putative artificial signal appears excessively when using the root-mean-square error (RMSE) instead of the DAS as error metric. It is

well known in the verification community that the RMSE tends to indicate large errors when the underlying signal is large. Repeating the same analysis but with the RMSE (not shown), weak RWPs already appear to be more predictable by 10% at the shortest lead times, increasing to 20%–25% at day 10, with a systematic increase of higher predictability for weaker RWPs. Similarly, the RMSE identifies the decay stage as the most predictable stage at all lead times (not shown), in stark contrast to the result obtained by the DAS (Figure 3). It seems counterintuitive to us that the weak signals of decay and low-amplitude RWPs would exhibit higher predictability than the stronger signals of the propagation stage and high-amplitude RWPs, respectively, for skillful forecasts. The DAS is evidently much less affected by such a putative artificial signal and is thus suitable to identify differences in predictability characteristics of RWPs.

3.4.3 | Life time and initialization relative to onset

Previous work (Grazzini & Vitart, 2015; Wirth & Eichhorn, 2014) suggested that long-lived RWPs imply enhanced predictability. We here investigate the role of RWP life time on predictability by defining two RWP categories: "short-lived" RWPs with life times of 5–8 days and "long-lived" RWPs with life times of 9–14 days, a definition consistent with previous work (e.g., Grazzini & Vitart, 2015; Wirth *et al.*, 2018). We investigate the relative differences in the DAS between these two groups, similar to our analysis of relative differences between weak and strong RWPs in Section 3.4.2. Between our samples of long-lived and short-lived RWPs, however, the differences are so small that they are not statistically significant (not shown).

We consider the lack of a statistically significant result to be counterintuitive. The onset stage clearly exhibits the lowest predictability (Figure 3), which suggests that RWP forecasts that are initialized before onset have lower predictability than forecasts initialized after onset. In this sense, RWP onset can be expected to act as a predictability barrier for the subsequent evolution (cf. Sánchez *et al.*, 2020). This hypothesized predictability barrier implies differences between the predictability of short-lived and long-lived RWPs. To see this, consider the following example. A four-day forecast verifying at the propagation stage of an RWP with a life time of 5 days is initialized before onset, whereas a four-day forecast verifying at the propagation stage of an RWP with a life time of 10 days is initialized after onset. From this perspective alone, we would expect a better forecast for the long-lived RWP than for the short-lived RWP. Possibly, the stratification into two life time groups above does not reveal these predictability differences due to compounding factors such as amplitude or geographical location. Therefore, we now test the hypothesized predictability barrier more explicitly.

Forecasts verifying at propagation stage are stratified with respect to their initialization time relative to onset. For example, we compare the DAS of forecasts initialized at onset time with forecasts that are initialized, say, three days before onset. The complication of this approach is that forecasts with different lead times are compared, because the verification time is fixed at the propagation stage of a given RWP, but initialization time varies. To account for this difference, we normalize each forecast with the average DAS of the respective lead time. We then examine the relative difference of the normalized DAS at different initialization times with the normalized DAS of forecasts initialized at onset.

The results of this analysis are shown in Figure 9. For both short- and long-lived RWPs, predictability is highest if onset is represented by the initial conditions, that is, for forecasts initialized at onset. Differences increase with increasingly earlier initialization time before onset, and reach around 20% at the earliest initialization times considered (disregarding the last data point for both short- and long-lived RWPs as a potential outlier due to small sample size). This analysis provides a much robuster result than stratifying by life time alone and verifies that RWP onset acts as a partial predictability barrier for the propagation stage of RWPs.

4 | SUMMARY AND DISCUSSION

This study investigates the predictability of RWPs as physical entities. To this end, we first compile a catalogue of RWPs from ERA5 reanalysis following previous work by Wolf and Wirth (2017) to identify RWPs. From this catalogue, we select RWPs with life times between 5 and 14 days for further investigation. Predictability of these RWPs is assessed using five-member GEFS-v12 ensemble reforecasts, with each member verified deterministically against the corresponding GFS reanalysis, and averaging over all five members. The reforecasts are available to us for the period 2000–2018 and a total of over 1000 RWPs in the Northern Hemisphere are analyzed. Our study is not concerned with evaluating the skill of the model at hand. Rather, the basic premise here is that, for a large number of cases, particularly good and bad forecast skill correspond statistically to high and low atmospheric predictability, respectively. We are thus concerned with the intrinsic predictability of the atmosphere and not with the practical predictability of a specific forecast system. It is important to note, however, that our results

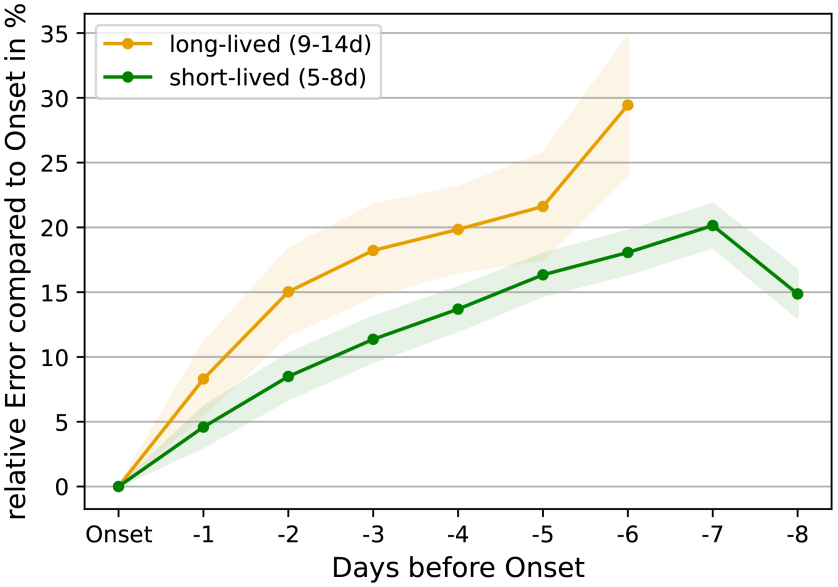

FIGURE 9 Relative differences in DAS for short- (green) and long-lived (orange) RWPs compared with forecasts initialized at onset.

may be affected by the use of data from an imperfect model.

We use the distance and amplitude score (DAS) as a skill metric. Originally developed and applied for convective-scale clouds and precipitation (Keil & Craig, 2007, 2009; Zinner *et al.*, 2008), we adopt the DAS to planetary-scale RWPs. The DAS is applied to the Rossby-wave envelope field but avoids the double-penalty issue of many grid-based skill metrics. At the same time, the DAS avoids the complications of an object-based approach, for example, missing objects or arbitrary matching of objects for poor forecasts. The DAS asymptotes towards a saturation value for long lead times, but does not completely reach saturation within the 10-day lead time considered in this study. The growth rate of the median DAS is largest initially and decreases gradually with lead time, consistent with logistic growth and other models for error growth. This consistency is not necessarily expected, because other studies using error decomposition (Jankov *et al.*, 2021) and illustrating the lead-time dependence of error-growth mechanisms (Baumgart *et al.*, 2018; Baumgart & Riemer, 2019) indicate that the early rapid error growth diagnosed by more common error metrics is associated with phase errors, but phase information is largely de-emphasized by the DAS applied to the envelope field.

Lee and Held (1993) have hypothesized that the large-scale energy propagation implied by the evolution of the RWP envelope exhibits higher predictability than the underlying wave pattern that contains the information about the phase of individual troughs and ridges. To our knowledge, this hypothesis has not yet been tested. We do not present an explicit test either, but it is instructive to consider the growth of the DAS with lead time in the context of this hypothesis. As discussed above, the growth rates of the DAS do not indicate qualitative differences in how error growth affects the envelope and the underlying wave pattern. In addition, the DAS getting close to saturation within the 10-day time frame considered does not indicate a longer predictability horizon for the envelope than for the underlying waves in the respective wave-number range (here 4–15; see, e.g., fig. 3 in Žagar *et al.* (2017) for wave number range 4–14). Our results thereby do not provide supporting evidence for the hypothesis, but rather cast doubt on its validity. More explicit tests, however, are needed before the hypothesis may be refuted.

Relative differences in RWP predictability tend to maximize during the short medium range. In this summary, we present quantitative results as rounded values averaged over 3–5 days lead time. The different stages of the RWP life cycle exhibit the largest relative differences in predictability. The onset stage and the propagation stage are the least and most predictable stages, respectively. Forecasts verifying during the onset stage exhibit more than 35% higher DAS values than during the propagation stage. The decay stage exhibits more than 20% higher DAS values than the propagation stage. During winter, RWPs exhibit higher predictability than during summer, consistent with general verification scores of operational centers (e.g., Haiden *et al.*, 2022). This seasonal difference is more pronounced for the decay and propagation stages, with more than 10%–15% difference of the DAS values between extended summer and extended winter, and less pronounced for the onset stage (5%–10% relative difference). Investigating all months individually confirms that RWP predictability exhibits a robust seasonal cycle.

The characteristics of RWPs impact their predictability further. RWP characteristics correlate with season and the stage of the life cycle. Our analysis thus controls for these two factors by (i) considering life-cycle stages individually and (ii) stratifying data for each month individually

before aggregating year-round. RWP amplitude exerts a prominent impact on predictability during the propagation stage, with strong RWPs being more predictable than weak RWPs up to 6 days lead time. Comparing the upper and lower quintiles and terciles of the amplitude distribution, relative DAS differences are 25% and 15%, respectively. The differences in predictability due to RWP amplitude are thus even somewhat larger than those due to seasonality. RWPs during onset and decay are typically weaker than during the propagation stage. The predictability differences of the different life-cycle stages may thus be related to those found for different RWP amplitudes during the propagation stage. The impact of RWP size on predictability is overall very similar to that of amplitude, but somewhat weaker.

In addition, geographic location impacts RWP predictability. Due to limitations in sample size, we did not attempt a separation of amplitude and geographic effects. They thus act as confounding factors in our analysis. There is no indication, however, that the geographic variations of RWP predictability merely reflect the climatological amplitude distribution (not shown). The geographic analysis thus detects signals that are at least partly independent of amplitude. The geographical characteristics depend on the stage of the life cycle and partly on the season. For the same stage of the life cycle, characteristics may vary between the ocean basins of the North Pacific and the North Atlantic, and in particular between the ocean basins and the large continental region of Eurasia. Over Eurasia, RWPs rather consistently exhibit relatively high predictability, consistent with themost coherent propagation of RWPs in that region (Chang, 1999).

Previous work suggests that long-lived RWPs are associated with enhanced predictability (e.g., Grazzini & Vitart, 2015; Wirth & Eichhorn, 2014). Here, we do not find robust differences between short-lived and long-lived RWPs. Because the onset stage is the least predictable stage of the RWP life cycle, conceptually one would expect that the propagation stage exhibits higher predictability once RWP onset is represented in the initial conditions of the forecast, that is, the onset stage is expected to act as a predictability barrier for RWP propagation. Moist baroclinic growth and its associated disturbance of the tropopause, which relates to RWP onset, have been documented as predictability barriers in specific episodes over the North Atlantic (Sánchez *et al.*, 2020). Here we find that RWPs that are initialized before onset indeed exhibit lower predictability, with up to 20% higher DAS values. Our results thus confirm that the onset stage acts as a predictability barrier for the propagation stage.

This study focuses on phenomenological characteristics of RWP predictability. An important topic for future studies is to understand these phenomenological findings better in terms of differences in the RWP dynamics. To gain deeper further understanding, future work will also need to explore the dynamics of upscale error growth that eventually affect the RWP scale. This larger scale has hitherto not been the focus of studies on upscale error growth. There is indication that error-growth mechanisms on this scale assume characteristics different from those on smaller, synoptic scales (Baumgart *et al.*, 2019). Understanding upscale error growth up to the RWP scale may thus require us to move beyond long-standing paradigms of error-growth dynamics (Craig *et al.*, 2021).

AUTHOR CONTRIBUTIONS

Isabelle Prestel-Kupferer: data curation; formal analysis; methodology; visualization; writing – original draft. **Michael Riemer:** conceptualization; funding acquisition; project administration; supervision; writing – original draft. **Sören Schmidt:** data curation; formal analysis; methodology; writing – original draft. **Franziska Teubler:** conceptualization; supervision; writing – original draft.

ACKNOWLEDGEMENTS

The research leading to these results has been performed within the subprojects A1 and A8(N) of the Transregional Collaborative Research Center "Waves to Weather" (https://www.wavestoweather.de/). I. Prestel-Kupferer acknowledges further support by the German Research Foundation (DFG) grant no. RI 1771/4-1.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Isabelle Prestel-Kupferer https://orcid.org/0009-0001-5098-8219

REFERENCES

Anwender, D., Harr, P.A. & Jones, S.C. (2008) Predictability associated with the downstream impacts of the extratropical transition of tropical cyclones: Case studies. *Monthly Weather Review*, 136(9), 3226–3247. Available from: https://doi.org/10.1175/2008MWR2249.1

Baldwin, M.E. & Kain, J.S. (2006) Sensitivity of several performance measures to displacement error, bias, and event frequency. *Weather and Forecasting*, 21(4), 636–648. Available from: https://doi.org/10.1175/WAF933.1

Baumgart, M., Ghinassi, P., Wirth, V., Selz, T., Craig, G.C. & 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(5), 1713–1731. Available from: https://doi.org/10.1175/MWR-D-18-0292.1

Baumgart, M. & 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*, 145(724), 3252–3270. Available from: https://doi.org/10.1002/qj.3617

Baumgart, M., Riemer, M., Wirth, V., Teubler, F. & Lang, S.T.K. (2018) Potential vorticity dynamics of forecast errors: A quantitative case study. *Monthly Weather Review*, 146(5), 1405–1425. Available from: https://doi.org/10.1175/MWR-D-17-0196.1

Chang, E.K.M. (1999) Characteristics of wave packets in the upper Troposphere. Part II: Seasonal and hemispheric variations. *Journal of the Atmospheric Sciences*, 56, 1729–1747, https://doi.org/10.1175/1520-0469(1999)056<1729:COWPIT>2.0.CO;2

Chang, E.K.M. (2005) The impact of wave packets propagating across Asia on pacific cyclone development. *Monthly Weather Review*, 133(7), 1998–2015. Available from: https://doi.org/10.1175/MWR2953.1

Clarke, S.J., Gray, S.L. & Roberts, N.M. (2019) Downstream influence of mesoscale convective systems. Part 1: Influence on forecast evolution. *Quarterly Journal of the Royal Meteorological Society*, 145(724), 2933–2952. Available from: https://doi.org/10.1002/qj.3593

Craig, G.C., Fink, A.H., Hoose, C., Janjić, T., Knippertz, P., Laurian, A. et al. (2021) Waves to weather: Exploring the limits of predictability of weather. *Bulletin of the American Meteorological Society*, 102(11), E2151–E2164. Available from: https://doi.org/10.1175/BAMS-D-20-0035.1

Cressman, G.P. (1948) On the forecasting of long waves in the upper westerlies. *Journal of the Atmospheric Sciences*, 5(2), 44–57. Available from: https://doi.org/10.1175/1520-0469(1948)005<0044:OTFOLW>2.0.CO;2

Dalcher, A. & Kalnay, E. (1987) Error growth and predictability in operational ECMWF forecasts. *Tellus A*, 39(5), 474–491. Available from: https://doi.org/10.1111/j.1600-0870.1987.tb00322.x

Davis, C., Brown, B. & Bullock, R. (2006) Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Monthly Weather Review*, 134(7), 1772–1784. Available from: https://doi.org/10.1175/MWR3145.1

Doensen, O., Fragkoulidis, G., Magnusson, L., Riemer, M. & Wirth, V. (2024) Medium-range predictability of temperature extremes and biases in Rossby wave amplitude. *Quarterly Journal of the Royal Meteorological Society*, 150(765), 5390–5402. https://doi.org/10.1002/qj.4875

Duchon, C.E. (1979) Lanczos filtering in one and two dimensions. *Journal of Applied Meteorology and Climatology*, 18(8), 1016–1022. Available from: https://doi.org/10.1175/1520-0450(1979)018<1016:LFIOAT>2.0.CO;2, https://journals.ametsoc.org/view/journals/apme/18/8/1520-0450_1979_018_1016_lfioat_2_0_co_2.xml

Ebert, E.E. & McBride, J.L. (2000) Verification of precipitation in weather systems: Determination of systematic errors. *Journal of Hydrology*, 239(1-4), 179–202. Available from: https://doi.org/10.1016/S0022-1694(00)00343-7

Fragkoulidis, G., Wirth, V., Bossmann, P. & Fink, A.H. (2018) Linking Northern Hemisphere temperature extremes to Rossby wave packets. *Quarterly Journal of the Royal Meteorological Society*, 144(711), 553–566. Available from: https://doi.org/10.1002/qj.3228

Froude, L.S.R. (2010) TIGGE: Comparison of the prediction of northern hemisphere extratropical cyclones by different ensemble prediction systems. *Weather and Forecasting*, 25(3), 819–836. Available from: https://doi.org/10.1175/2010WAF2222326.1

Froude, L.S.R., Bengtsson, L. & Hodges, K.I. (2007) The prediction of extratropical storm tracks by the ECMWF and NCEP ensemble prediction systems. *Monthly Weather Review*, 135(7), 2545–2567. Available from: https://doi.org/10.1175/MWR3422.1

Glatt, I. & Wirth, V. (2014) Identifying Rossby wave trains and quantifying their properties. *Quarterly Journal of the Royal Meteorological Society*, 140(679), 384–396. Available from: https://doi.org/10.1002/qj.2139

Gray, S.L., Dunning, C.M., Methven, J., Masato, G. & Chagnon, J.M. (2014) Systematic model forecast error in Rossby wave structure. *Geophysical Research Letters*, 41(8), 2979–2987. Available from: https://doi.org/10.1002/2014GL059282

Grazzini, F. & Vitart, F. (2015) Atmospheric predictability and Rossby wave packets. *Quarterly Journal of the Royal Meteorological Society*, 141(692), 2793–2802. Available from: https://doi.org/10.1002/qj.2564

Guan, H., Zhu, Y., Sinsky, E., Fu, B., Li, W., Zhou, X. et al. (2022) GEFSv12 reforecast dataset for supporting subseasonal and hydrometeorological applications. *Monthly Weather Review*, 150(3), 647–665. Available from: https://doi.org/10.1175/MWR-D-21-0245.1, https://journals.ametsoc.org/view/journals/mwre/150/3/MWR-D-21-0245.1.xml

Haiden, T., Janousek, M., Vitart, F., Ben-Bouallegue, Z., Ferranti, L., Prates, F. et al. (2022) Evaluation of ECMWF forecasts, including the 2021 upgrade. https://www.ecmwf.int/node/20469

Hamill, T.M., Whitaker, J.S., Shlyaeva, A., Bates, G., Fredrick, S., Pegion, P. et al. (2022) The reanalysis for the global ensemble forecast system, version 12. *Monthly Weather Review*, 150(1), 59–79. Available from: https://doi.org/10.1175/MWR-D-21-0023.1, https://journals.ametsoc.org/view/journals/mwre/150/1/MWR-D-21-0023.1.xml

Hauser, S., Teubler, F., Riemer, M., Knippertz, P. & Grams, C.M. (2022) Towards a diagnostic framework unifying different perspectives on blocking dynamics: insight into a major blocking in the North Atlantic-European region. *Weather and Climate Dynamics Discussions*, 4, 1–36. Available from: https://doi.org/10.5194/wcd-2022-44, https://wcd.copernicus.org/preprints/wcd-2022-44/

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. Available from: https://doi.org/10.1002/qj.3803

Hovmöller, E. (1949) The trough-and-ridge diagram. *Tellus*, 1(2), 62–66. Available from: https://doi.org/10.1111/j.2153-3490.1949.tb01260.x

Jankov, I., Gregory, S., Ravela, S., Toth, Z. & Peña, M. (2021) Partition of forecast error into positional and structural components. *Advances in Atmospheric Sciences*, 38, 1012–1019. Available from: https://doi.org/10.1007/s00376-021-0251-7

Keil, C. & Craig, G.C. (2007) A displacement-based error measure applied in a regional ensemble forecasting system. *Monthly Weather Review*, 135, 3248–3259. Available from: https://doi.org/10.1175/MWR3457.1

Keil, C. & Craig, G.C. (2009) A displacement and amplitude score employing an optical flow technique. *Weather and Forecasting*, 24, 1297–1308. Available from: https://doi.org/10.1175/2009WAF2222247.1

Lee, S. & Held, I.M. (1993) Baroclinic wave packets in models and observations. *Journal of Atmospheric Sciences*, 50(10),

1413–1428. Available from: https://doi.org/10.1175/1520-0469(1993)050<1413:BWPIMA>2.0.CO;2

Lojko, A., Payne, A. & Jablonowski, C. (2022) The remote role of North-American mesoscale convective systems on the forecast of a Rossby wave packet: A multi-model ensemble case-study. *Journal of Geophysical Research: Atmospheres*, 127(24), e2022JD037171. Available from: https://doi.org/10.1029/2022JD037171

Lorenz, E.N. (1982) Atmospheric predictability experiments with a large numerical model. *Tellus*, 34(6), 505–513. Available from: https://doi.org/10.3402/tellusa.v34i6.10836

Martínez-Alvarado, O., Maddison, J.W., Gray, S.L. & Williams, K.D. (2018) Atmospheric blocking and upper-level Rossby-wave forecast skill dependence on model configuration. *Quarterly Journal of the Royal Meteorological Society*, 144(716), 2165–2181. Available from: https://doi.org/10.1002/qj.3326

Martius, O., Schwierz, C. & Davies, H.C. (2008) Far-upstream precursors of heavy precipitation events on the Alpine south-side. *Quarterly Journal of the Royal Meteorological Society*, 134(631), 417–428. Available from: https://doi.org/10.1002/qj.229

Miller, D.E. & Gensini, V.A. (2023) GEFSV12 high- and low-skill day-10 tornado forecasts. *Weather and Forecasting*, 38(7), 1195–1207. Available from: https://doi.org/10.1175/WAF-D-22-0122.1, https://journals.ametsoc.org/view/journals/wefo/38/7/WAF-D-22-0122.1.xml

Pérez-Fernández, I. & Barreiro, M. (2023) How well do forecast models represent observed long-lived Rossby wave packets during southern hemisphere summer? *Atmospheric Science Letters*, 24, e1175. Available from: https://doi.org/10.1002/asl.1175

Piaget, N., Froidevaux, P., Giannakaki, P., Gierth, F., Martius, O., Riemer, M. et al. (2015) Dynamics of a local alpine flooding event in October 2011: Moisture source and large-scale circulation. *Quarterly Journal of the Royal Meteorological Society*, 141(690), 1922–1937. Available from: https://doi.org/10.1002/qj.2496

Quinting, J.F. & Jones, S.C. (2016) On the impact of tropical cyclones on Rossby wave packets: A climatological perspective. *Monthly Weather Review*, 144(5), 2021–2048. Available from: https://doi.org/10.1175/MWR-D-14-00298.1

Quinting, J.F. & Vitart, F. (2019) Representation of synoptic-scale Rossby wave packets and blocking in the S2S prediction project database. *Geophysical Research Letters*, 46(2), 1070–1078. Available from: https://doi.org/10.1029/2018GL081381

Rodwell, M.J., Magnusson, L., Bauer, P., Bechtold, P., Bonavita, M., Cardinali, C. et al. (2013) Characteristics of occasional poor medium-range weather forecasts for Europe. *Bulletin of the American Meteorological Society*, 94(9), 1393–1405. Available from: https://doi.org/10.1175/BAMS-D-12-00099.1

Sánchez, C., Methven, J., Gray, S. & Cullen, M. (2020) Linking rapid forecast error growth to diabetic processes. *Quarterly Journal of the Royal Meteorological Society*, 146(732), 3548–3569. Available from: https://doi.org/10.1002/qj.3861

Selz, T., Riemer, M. & Craig, G.C. (2022) The transition from practical to intrinsic predictability of midlatitude weather. *Journal of the Atmospheric Sciences*, 79(8), 2013–2030. Available from: https://doi.org/10.1175/JAS-D-21-0271.1

Shapiro, M.A. & Thorpe, A.J. (2004) Thorpex international science plan. *WMO/TD*, 1246, 57.

Souders, M.B., Colle, B.A. & Chang, E.K.M. (2014) The climatology and characteristics of Rossby wave packets using a feature-based tracking technique. *Monthly Weather Review*, 142(10), 3528–3548. Available from: https://doi.org/10.1175/MWR-D-13-00371.1

Teubler, F. & Riemer, M. (2021) Potential-vorticity dynamics of troughs and ridges within Rossby wave packets during a 40-year reanalysis period. *Weather and Climate Dynamics*, 2(3), 535–559. Available from: https://doi.org/10.5194/wcd-2-535-2021, https://wcd.copernicus.org/articles/2/535/2021/

Wernli, H., Paulat, M., Hagen, M. & Frei, C. (2008) SAL—a novel quality measure for the verification of quantitative precipitation forecasts. *Monthly Weather Review*, 136(11), 4470–4487. Available from: https://doi.org/10.1175/2008MWR2415.1

Wirth, V. & Eichhorn, J. (2014) Long-lived Rossby wave trains as precursors to strong winter cyclones over Europe. *Quarterly Journal of the Royal Meteorological Society*, 140(680), 729–737. Available from: https://doi.org/10.1002/qj.2191

Wirth, V., Riemer, M., Chang, E.K.M. & Martius, O. (2018) Rossby wave packets on the midlatitude Waveguide—A review. *Monthly Weather Review*, 146(7), 1965–2001. Available from: https://doi.org/10.1175/MWR-D-16-0483.1, https://journals.ametsoc.org/view/journals/mwre/146/7/mwr-d-16-0483.1.xml

Wolf, G. & Wirth, V. (2017) Diagnosing the horizontal propagation of Rossby wave packets along the midlatitude waveguide. *Monthly Weather Review*, 145(8), 3247–3264. Available from: https://doi.org/10.1175/MWR-D-16-0355.1, https://journals.ametsoc.org/view/journals/mwre/145/8/mwr-d-16-0355.1.xml

Ye, J., Wang, Z., Yang, F., Harris, L., Jensen, T., Miller, D.E. et al. (2023) Evaluation and process-oriented diagnosis of the GEFSV12 reforecasts. *Journal of Climate*, 36(12), 4255–4274. Available from: https://doi.org/10.1175/JCLI-D-22-0772.1

Žagar, N., Horvat, M., Zaplotnik, Ž. & Magnusson, L. (2017) Scale-dependent estimates of the growth of forecast uncertainties in a global prediction system. *Tellus A: Dynamic Meteorology and Oceanography*, 69(1), 1287492. Available from: https://doi.org/10.1080/16000870.2017.1287492

Zheng, M., Chang, E.K.M. & Colle, B.A. (2013) Ensemble sensitivity tools for assessing extratropical cyclone intensity and track predictability. *Weather and Forecasting*, 28(5), 1133–1156. Available from: https://doi.org/10.1175/WAF-D-12-00132.1, https://journals.ametsoc.org/view/journals/wefo/28/5/waf-d-12-00132_1.xml

Zimin, A.V., Szunyogh, I., Hunt, B.R. & Ott, E. (2006) Extracting envelopes of nonzonally propagating Rossby wave packets. *Monthly Weather Review*, 134(4), 1329–1333. Available from: https://doi.org/10.1175/MWR3122.1, https://journals.ametsoc.org/view/journals/mwre/134/4/mwr3122.1.xml

Zinner, T., Mannstein, H. & Tafferner, A. (2008) CB-TRAM: Tracking and monitoring severe convection from onset over rapid development to mature phase using multi-channel Meteosat-8 SEVIRI data. *Meteorology and Atmospheric Physics*, 101, 191–210. Available from: https://doi.org/10.1007/s00703-008-0290-y

How to cite this article: Prestel-Kupferer, I., Riemer, M., Schmidt, S. & Teubler, F. (2024) Predictability of midlatitude Rossby-wave packets. *Quarterly Journal of the Royal Meteorological Society*, 150(765), 5057–5073. Available from: https://doi.org/10.1002/qj.4856