RESEARCH ARTICLE

Processes governing the amplification of ensemble spread in a medium-range forecast with large forecast uncertainty

Marlene Baumgart | Michael Riemer

Institut für Physik der Atmosphäre, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany

Correspondence
Marlene Baumgart, Institut für Physik der Atmosphäre, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany.
Email: mbaumga@uni-mainz.de

Abstract

This study provides a process-based perspective on the amplification of forecast uncertainty and forecast errors in ensemble forecasts. A case from the North Atlantic Waveguide and Downstream Impact Experiment that exhibits large forecast uncertainty is analysed. Two aspects of the ensemble behaviour are considered: (a) the mean divergence of the ensemble members, indicating the general amplification of forecast uncertainty, and (b) the divergence of the best and worst members, indicating extremes in possible error-growth scenarios. To analyse the amplification of forecast uncertainty, a tendency equation for the ensemble variance of potential vorticity (PV) is derived and partitioned into the contributions from individual processes. The amplification of PV variance is, on average for the midlatitudes of the Northern Hemisphere, dominated by near-tropopause dynamics. Locally, however, other processes can dominate the variance amplification, for example, in the region where tropical storm *Karl* interacts with the Rossby-wave pattern during extratropical transition. In this region, the variance amplification is dominated by upper-tropospheric divergence and tropospheric–deep interaction and is thereby mostly related to (moist baroclinic) cyclone development. The differences between the error growth in the best and worst ensemble members can, to a large part, be attributed to differences in the representation of cut-off evolution around 3 days, which subsequently amplifies substantially in the highly nonlinear region of the Rossby-wave pattern until 5 days. In terms of the processes, the differences in error growth are dominated by differences in the error growth by near-tropopause dynamics. The approach presented provides flow-dependent insight into the dynamics of forecast uncertainty and forecast errors and helps to understand better the different contributions of specific weather systems to the medium-range amplification of ensemble spread.

KEYWORDS
atmospheric dynamics, ensemble forecasts, error growth, numerical weather prediction, potential vorticity, predictability, Rossby waves

1 | INTRODUCTION

1.1 | Flow-dependent forecast uncertainty

Ensemble forecasts have become an essential component of operational weather forecasts in the past 25 years. Weather prediction has thereby changed from a deterministic to a probabilistic approach, that means, from providing a single forecast to providing an estimate of the uncertainty associated with a forecast and the range of possible future scenarios (e.g., Buizza, 2018). On average, the forecast accuracy of ensemble forecasts has improved substantially over the past decades (for the European Centre for Medium-Range Weather Forecasts (ECMWF), see, for example, fig. 1a of Rodwell *et al.*, 2018). In addition to forecast accuracy, forecast reliability is a second important characteristic of the quality of an ensemble forecast. A forecast system is reliable if the forecast distribution matches the observed frequency of occurrence. For a reliable forecast system, the ensemble standard deviation should match the root-mean-square error of the ensemble mean when averaged over many cases (as measured by the so-called spread–error relationship, Leutbecher and Palmer, 2008). Considering the annual mean of the spread–error relationship for the Northern Hemisphere, large improvements have been achieved in the last decades, resulting in almost indistinguishable spread–error curves in 2014 (see, for example, fig. 1a of Rodwell *et al.*, 2018).

Due to the chaotic nature of atmospheric flow and the associated sensitive dependence on the initial conditions (Lorenz, 1963), atmospheric predictability exhibits a pronounced flow dependence. A central aim of ensemble forecasts is to provide a reliable estimate of this day-to-day variability of forecast uncertainty. On a day-to-day basis, and averaged over a local domain, however, the spread–error relationship holds only partially (see, for example, fig. 1b of Rodwell *et al.*, 2018). While the spread–error relationship can never hold perfectly on a day-to-day basis (Whitaker and Loughe, 1998), there is arguably still room for improvement (e.g., Rodwell *et al.*, 2018). In addition, a better understanding of the flow dependence of atmospheric predictability and associated local forecast reliability may help to improve the interpretation of ensemble forecasts.

The design of ensemble perturbations to generate appropriate ensemble spread is a nontrivial task and remains a major area of research (e.g., Bauer *et al.*, 2015; Berner *et al.*, 2017; Buizza, 2018). Ensemble forecast systems account for uncertainties in both the initial conditions and the model formulation (Palmer and Hagedorn, 2006). At ECMWF, for example, initial-condition uncertainty is accounted for by ensemble data assimilation (EDA) in combination with singular-vector perturbations, and model uncertainties by stochastically perturbed parametrization tendencies (ECMWF, 2018). There is, however, a wide range of approaches used at different operational weather prediction centres (e.g., Buizza, 2018). An improved understanding of the processes governing the amplification of ensemble spread is deemed important to design the best suitable ensemble perturbations.

Rodwell *et al.* (2018) presented an approach to studying flow-dependent deficiencies in short-term reliability using a budget equation of ensemble variance in observation space (Rodwell *et al.*, 2016). Deficiencies in short-term reliability could be identified for a composite of initial flow type that is associated with large forecast uncertainty, highlighting the importance of moist processes in mesoscale convective systems and in warm conveyor belts as sources for forecast uncertainty. In addition, the authors used a (spatio-temporally filtered) growth rate for the ensemble standard deviation of potential vorticity (PV) to highlight uncertainty growth that projects on the synoptic scale.

The current study also considers a PV framework. The focus here, in contrast to Rodwell *et al.* (2018), is not on identifying model deficiencies, but rather on diagnosing quantitatively the processes governing the flow-dependent amplification of initially small ensemble perturbations. Our PV framework builds on the PV perspective of midlatitude dynamics (Hoskins *et al.*, 1985) and provides a quantitative partitioning of the dynamics into individual processes, including the influence of near-tropopause Rossby-wave dynamics, baroclinic growth, and moist processes. The PV framework has recently been applied to quantify the governing processes of Rossby-wave packets (Piaget *et al.*, 2015; Teubler and Riemer, 2016) and a large-amplitude ridge (Schneidereit *et al.*, 2017). The PV framework has also been applied to a case study of error growth in an operational ECMWF forecast (Baumgart *et al.*, 2018) and to study upscale error growth in dedicated numerical experiments (Baumgart *et al.*, 2019). Applying this framework to an operational ensemble forecast using a tendency equation for the ensemble variance of PV is a key novelty of this study. The current study focuses on the medium-range amplification of forecast uncertainty. Notwithstanding the importance of short-range sources of forecast uncertainty, we emphasize that the amplification of forecast errors is highly nonlinear, that means, ensemble members with relatively small errors at short lead times may have relatively large errors at medium-range lead times, and vice versa (illustrated in Figure 1, see the discussion below).

1.2 | A case of large forecast uncertainty

We analyse one case from the North Atlantic Waveguide and Downstream Impact Experiment (NAWDEX; Schäfler *et al.*, 2018), namely the extratropical transition of tropical storm *Karl* to *Ex-Karl*. This case was associated with large forecast uncertainty (Schäfler *et al.*, 2018), and the medium-range ensemble forecast showed very different developments for the
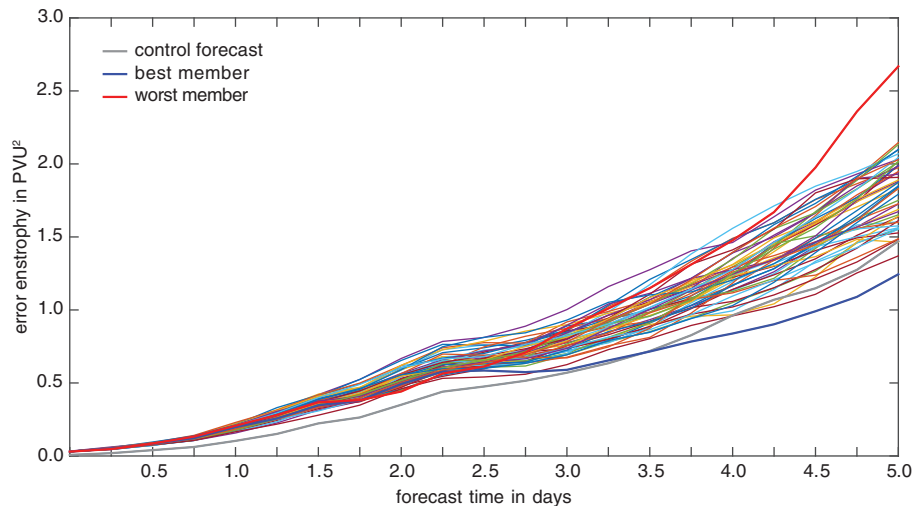
FIGURE 1 Error enstrophy of each ensemble member (spatially averaged over the midlatitudes of the Northern Hemisphere, 30°N–80°N) as a function of forecast time. Thick lines denote error enstrophy of the control forecast (grey), the best member (blue), and the worst member (red)

interaction between *Ex-Karl* and the waveguide, with only a few ensemble members capturing this interaction correctly (Kumpf *et al.*, 2018). Here, the same ensemble forecast as in Kumpf *et al.* (2018) is investigated. Besides the extratropical transition of *Karl* in the North Atlantic, the ensemble forecast indicates a second region of very large forecast uncertainty, which is associated with the interaction of a cut-off and a high-amplitude ridge over North America. A hemispheric perspective on the ensemble evolution is therefore provided first, before discussing important local differences from this hemispheric perspective in more detail.

To provide an overview of the ensemble evolution in our case, Figure 1 shows the error enstrophy of each ensemble member (spatially averaged over the midlatitudes of the Northern Hemisphere, 30°N–80°N) as a function of forecast time. Error enstrophy is here defined as the squared PV error, that is, $\frac{1}{2}P^{*2}$, with P^* denoting the PV error. Note that error enstrophy is thus directly related to a standard error metric: the root-mean-square error.

At the initial time, the error enstrophy of the control forecast is close to zero, whereas the perturbed members start with a notable error enstrophy that is of similar magnitude for each perturbed forecast.[1] The error enstrophy amplifies steadily for approximately 2 days in both the perturbed forecasts and the control forecast. The control forecast maintains the smallest error until 3 days. Around 2 days, nonlinearities in the error growth become apparent: some members with smaller errors exhibit larger error-growth rates than members with larger errors, that is, there are prominent intersections of the error curves of the individual members.

At the end of the period considered here (5 days), two ensemble members are distinct from the rest of the ensemble

members in terms of error enstrophy. One of the members (in the following referred to as the "worst member") exhibits a substantially larger error than the other ensemble members, whereas another member (the "best member") exhibits a substantially smaller error than the other ensemble members. Interestingly, the error enstrophy of the best and worst members is very similar in the first 2.5 days, with both members having a relatively small error. Afterwards, the worst member exhibits, however, much larger error growth than the best member and the two members diverge prominently until they become the apparent outliers of the ensemble at 5 days.

The comparison of the error growth of the individual ensemble members illustrates two main aspects of interest: (a) the mean divergence of the ensemble members, indicating the general amplification of forecast uncertainty, and (b) the divergence of the best and worst members, indicating extremes in possible error-growth scenarios. These two aspects are discussed in more detail in the following. Section 2 first describes the data and methods used to quantify the amplification of forecast uncertainty and forecast errors. We then discuss the amplification of forecast uncertainty from both a hemispheric and a localized perspective (section 3). Subsequently, the error growth of the best and worst members is compared in section 4. We conclude with a summary and discussion of the results in section 5.

2 | DATA AND METHODS TO QUANTIFY THE AMPLIFICATION OF FORECAST UNCERTAINTY AND FORECAST ERRORS

2.1 | Ensemble data

We use real-time data from the Atmospheric model Ensemble 15-day forecast (ENS) of ECMWF with a 3-hr temporal resolution in the first 3 days and 6-hr resolution afterwards. These

---

[1]Note that the control forecast is at the same resolution as the perturbed members.

data were archived manually on model levels in spectral space during the NAWDEX campaign. Operationally archived data on pressure levels are not sufficient for our diagnostic, as the vertical resolution is too low. For the PV inversion, we interpolate the manually archived data to a 1° × 1° grid and from model to pressure levels from 900–100 hPa, with a grid distance of 50 hPa. For the further analysis, all variables are interpolated to isentropic levels.

We analyse the ensemble forecast initialized on September 22, 2016 at 0000 UTC. This case is related to one of the observational highlights during the NAWDEX campaign, namely the extratropical transition of Hurricane *Karl*, which was related to a heavy precipitation event in Norway (Schäfler *et al.*, 2018; Kumpf *et al.*, 2019). Forecasting this event was rather difficult, as the evolution was very sensitive to uncertainties in the timing and location of the interaction between *Ex-Karl* and the midlatitude waveguide (Schäfler *et al.*, 2018). The ensemble forecast investigated here shows very different developments of individual members in the medium-range forecast of *Ex-Karl* (Kumpf *et al.*, 2018). A few members capture the interaction between *Ex-Karl* and the waveguide correctly, whereas the majority of members show a distinct error during the interaction (Kumpf *et al.*, 2018).

2.2 | Quantitative PV framework for the amplification of forecast uncertainty and forecast errors

To attribute the evolution of forecast uncertainty to individual processes, we extend our recently developed PV diagnostic for error growth (Baumgart *et al.*, 2018, 2019) to the evolution of ensemble variance. In isentropic coordinates and the primitive-equations framework, Ertel (1942) PV is defined as follows:

$$P = -g \frac{\partial \theta}{\partial p} (\zeta_\theta + f) , \qquad (1)$$

where g is the gravitational acceleration, θ the potential temperature, p the pressure, ζ_θ the vertical component of the isentropic relative vorticity, and f the Coriolis parameter. PV changes locally due to advection and nonconservative processes:

$$\frac{\partial P}{\partial t} = -\mathbf{v} \cdot \nabla_\theta P + \text{Nres}. \qquad (2)$$

Nonconservative tendencies from the parametrization schemes are not archived for ensemble forecasts, so the direct PV modification by nonconservative processes due to diabatic heating and nonconservative momentum change, which is given by

$$N = \dot{\theta} \frac{\partial P}{\partial \theta} + P \frac{\partial \dot{\theta}}{\partial \theta} - g \frac{\partial \theta}{\partial p} \left(\mathbf{k} \times \frac{\partial \mathbf{v}}{\partial \theta} \right) \cdot \nabla \dot{\theta} - g \frac{\partial \theta}{\partial p} \mathbf{k} \cdot (\nabla \times \dot{\mathbf{v}}) , \quad (3)$$

is here interpreted as part of the residual Nres.[2] The residual also includes the influence of other processes that cannot be quantified even if the tendencies from all parametrization schemes are available, such as numerical diffusion, analysis increments due to data assimilation, and numerical inaccuracies due to the discretization and interpolation of data. Near the tropopause, diabatic processes can play an important role for Rossby-wave dynamics (e.g., Chagnon *et al.*, 2013; Martnez-Alvarado *et al.*, 2014). Relative to the advective tendencies, however, the direct diabatic PV changes are only of second-order importance (Teubler and Riemer, 2016). The impact of latent heat release is arguably most prominently communicated to the tropopause region by the associated upper-tropospheric divergent outflow (e.g., Davis *et al.*, 1996; Teubler and Riemer, 2016) and furthermore by enhancing baroclinic coupling (e.g., Gutowski *et al.*, 1992). Both processes are included in the advective tendencies analysed here. In addition, a case study of error growth in an operational ECMWF forecast including nonconservative tendencies from the parametrization schemes (Baumgart *et al.*, 2018) clearly demonstrated that these nonconservative tendencies are negligible, even at very short lead times. For the operational ensemble forecast investigated here, neglecting direct diabatic tendencies is therefore not expected to affect the analysis.

The PV perspective has proven very helpful to gain insight into the dynamics of forecast errors (e.g., Snyder *et al.*, 2003; Davies and Didone, 2013; Baumgart *et al.*, 2018). Based on the difference between the PV tendency in the forecast and the analysis, Baumgart *et al.* (2018) derived a tendency equation for error (potential) enstrophy:

$$\frac{\partial}{\partial t} \frac{P^{*2}}{2} = -P^* \mathbf{v}^* \cdot \nabla_\theta P - \nabla_\theta \cdot \left(\frac{P^{*2}}{2} (\mathbf{v}^* + \mathbf{v}) \right)$$
$$+ \frac{P^{*2}}{2} \nabla_\theta \cdot (\mathbf{v}^* + \mathbf{v}) + \text{Nres} , \qquad (4)$$

where variables with index * denote error fields (forecast − analysis) and variables without an index denote analysis fields. This equation will be used in section 4 to quantify the differences between the error growth in best and worst members, respectively.

Deriving a similar equation for the amplification of forecast uncertainty in ensemble forecasts is a natural extension of our previous work. For that purpose, we use the ensemble variance in PV as our metric for forecast uncertainty:

$$V_P := \frac{1}{n-1} \sum_{i=1}^{n} (P_i - \overline{P})^2 , \qquad (5)$$

---

[2]For convenience, we will denote the residual in all tendency equations as Nres, although the mathematical forms of the terms in the respective tendency equations differ.

where n describes the number of ensemble members (50 perturbed forecasts and 1 control forecast in the ECMWF system) and $\overline{P}$ is the ensemble-mean PV.

The local change of PV variance can be written as

$$
\frac{\partial V_P}{\partial t} = \frac{2}{n-1} \left(\sum_{i=1}^{n} (P_i - \overline{P}) \frac{\partial}{\partial t}(P_i - \overline{P}) \right)
$$
$$
= \frac{2}{n-1} \left(\left(\sum_{i=1}^{n} P_i \frac{\partial P_i}{\partial t} \right) - n\overline{P}\frac{\partial \overline{P}}{\partial t} \right). \quad (6)
$$

For the second step in Equation 6, we made use of the fact that the terms

$$
\sum_{i=1}^{n} \overline{P}\frac{\partial \overline{P}}{\partial t}, \qquad \sum_{i=1}^{n} \overline{P}\frac{\partial P_i}{\partial t}, \qquad \text{and} \qquad \sum_{i=1}^{n} P_i \frac{\partial \overline{P}}{\partial t}
$$

can be written as

$$
n\overline{P}\frac{\partial \overline{P}}{\partial t}.
$$

To gain further insight into the variance amplification, we insert the PV tendency of the individual ensemble members ($\partial P_i/\partial t = -\mathbf{v_i} \cdot \nabla P_i + \text{Nres}_i$) and of the ensemble mean $\left(\partial \overline{P}/\partial t = -\frac{1}{n}\sum_{i=1}^{n}(\mathbf{v_i} \cdot \nabla P_i + \text{Nres}_i) \right)$ into Equation 6. By expressing P_i as the sum of the ensemble mean value and the perturbation thereof (i.e., $P_i = \overline{P} + \delta P_i$, with δ denoting the perturbation from the ensemble mean) and rearranging and combining terms, Equation 6 can be written as

$$
\frac{\partial V_P}{\partial t} = \frac{2}{n-1} \left[-\sum_{i=1}^{n} \delta P_i \mathbf{v_i} \cdot \nabla \overline{P} - \sum_{i=1}^{n} \mathbf{v_i} \cdot \nabla \frac{(\delta P_i)^2}{2} \right] + \text{Nres.} \quad (7)
$$

Using $\mathbf{v_i} = \overline{\mathbf{v}} + \delta\mathbf{v_i}$ for the first term on the right-hand side, together with noting that

$$
-\sum_{i=1}^{n} \delta P_i \overline{\mathbf{v}} \cdot \nabla \overline{P} = -\overline{\mathbf{v}} \cdot \nabla \overline{P} \sum_{i=1}^{n} \delta P_i = 0,
$$

and using the chain rule for the second term, finally leads to

$$
\frac{\partial V_P}{\partial t} = \frac{2}{n-1} \left[\underbrace{-\sum_{i=1}^{n} \delta P_i \delta\mathbf{v_i} \cdot \nabla \overline{P}}_{\text{nonlinear production term}} \underbrace{-\sum_{i=1}^{n} \nabla \cdot \left(\frac{(\delta P_i)^2}{2} \mathbf{v_i} \right)}_{\text{flux term}} \right.
$$
$$
\left. \underbrace{+ \sum_{i=1}^{n} \frac{(\delta P_i)^2}{2} \nabla \cdot \mathbf{v_i}}_{\text{area change term}} \right] + \text{Nres.} \quad (8)
$$

In analogy to the tendency equation for error enstrophy (Equation 4), the first term on the right-hand side of Equation 8 can be interpreted as a nonlinear production term. The flux term (second term in Equation 8) merely redistributes variance and does not contribute to the global variance amplification. The third term in Equation 8 is associated with an area change of PV variance due to the divergent flow. This term leads to variance amplification (decay) when the quasihorizontal flow is divergent (convergent).

By using $\mathbf{v_i} = \overline{\mathbf{v}} + \delta\mathbf{v_i}$, the flux and area-change terms can be combined to

$$
-\overline{\mathbf{v}} \cdot \nabla V_p - \frac{1}{n-1} \sum_{i=1}^{n} \delta\mathbf{v_i} \cdot \nabla(\delta P_i)^2.
$$

Rodwell *et al.* (2018) used a similar form of equation for ensemble spread, but using standard deviation instead of variance. They derived a "material" derivative for ensemble spread following the ensemble mean flow, which would correspond to the term $-\overline{\mathbf{v}} \cdot \nabla V_p$ in our diagnostic. For this study, we decided to use, instead, the local derivative of PV variance as indicated in Equation 8, in order to have an exact budget equation for PV variance. The use of Equation 8 has also the advantage that the second term can be interpreted as a boundary term when integrating spatially over a specific domain.

To gain further insight into the processes governing the variance amplification, we use the same partitioning of processes as in the previous works by Teubler and Riemer (2016) and Baumgart *et al.* (2018, 2019). This partitioning is based on the PV perspective of midlatitude dynamics (Hoskins *et al.*, 1985). From this perspective, the evolution of PV anomalies near the tropopause can be described by advective tendencies associated with (a) upper-level (near-tropopause) dynamics, (b) midtropospheric PV anomalies and potential temperature anomalies just above the boundary layer, and (c) upper-tropospheric divergent flow. The influence of upper-level (near-tropopause) PV anomalies on the upper-level (near-tropopause) evolution describes the influence of nonlinear Rossby-wave dynamics (Hoskins *et al.*, 1985) and will here be referred to as the contribution from near-tropopause dynamics (index nTP). The influence of lower-level anomalies on the upper-level evolution describes the influence of vertical interaction, including baroclinic instability (Eady, 1949; Hoskins *et al.*, 1985; Heifetz *et al.*, 2004), and will here be referred to as tropospheric–deep interaction (index TPd). Upper-tropospheric divergence (index div) can be associated with dry balanced dynamics and diabatic processes (see, for example, chapter 6.4 in Holton and Hakim, 2013) and is of particular importance during ridge building (Grams *et al.*, 2011; Teubler and Riemer, 2016). Pronounced upper-tropospheric divergence is often associated with latent heat release below (e.g., Davis *et al.*, 1993; Riemer *et al.*, 2014; Quinting and Jones, 2016) and is usually expected to be of larger importance to Rossby-wave dynamics than direct diabatic PV modification (e.g., Davis *et al.*, 1993; Riemer and Jones, 2010; Teubler and Riemer, 2016).

The technicalities of the flow partitioning are also the same as in Teubler and Riemer (2016) and Baumgart *et al.* (2018; 2019): We use a Helmholtz partitioning to separate the divergent flow from the nondivergent flow, following Lynch (1989). The nondivergent flow is further partitioned into those parts associated with upper- and lower-level PV anomalies, respectively, using piecewise PV inversion (PPVI) under non-linear balance (Charney, 1955), following Davis and Emanuel (1991) and Davis (1992). PPVI is performed on the Northern Hemisphere from 25°N–85°N and 850–150 hPa. Potential temperature anomalies at 875 and 125 hPa serve as vertical boundary conditions for the inversion. Anomalies are defined as deviations from a background state, which is here defined as the 30 day-temporal mean centred on September 23, 2016 at 0000 UTC in the analysis. A midtropospheric pressure level (600 hPa) is used as the separation level between upper- and lower-level anomalies. The flow partitioning yields an uncertainty, $\mathbf{v}_{\text{unc}}$, due to the harmonic flow component, the uncertainty in the horizontal boundary conditions, and nonlinearities of the piecewise PV inversion. This uncertainty is calculated as the difference between the wind field in the ECMWF data and the sum of the near-tropopause, tropospheric–deep, and divergent wind fields. It is, in general, small and does not affect the physical interpretation of the results.

In summary, our flow partitioning yields

$$\mathbf{v} = \mathbf{v}_{\text{nTP}} + \mathbf{v}_{\text{TPd}} + \mathbf{v}_{\text{div}} + \mathbf{v}_{\text{unc}}. \tag{9}$$

The flow partitioning is performed separately for each individual ensemble member and then inserted into Equation 8, yielding

$$
\frac{\partial V_P}{\partial t} = \frac{2}{n-1}
$$
$$
\underbrace{\left[-\sum_{i=1}^{n} \delta P_i (\delta \mathbf{v}_{i,\,\text{nTP}} + \delta \mathbf{v}_{i,\,\text{TPd}} + \delta \mathbf{v}_{i,\,\text{div}} + \delta \mathbf{v}_{i,\,\text{unc}}) \cdot \nabla \overline{P} \right.}_{\text{nonlinear production term}}
$$
$$
\underbrace{- \sum_{i=1}^{n} \nabla \cdot \left(\mathbf{v}_i \frac{(\delta P_i)^2}{2} \right)}_{\text{flux term}} + \underbrace{\left. \sum_{i=1}^{n} \frac{(\delta P_i)^2}{2} \nabla \cdot \mathbf{v}_{i,\,\text{div}} \right]}_{\text{area change term}} + \text{Nres}. \tag{10}
$$

Note that the divergent wind contributes to both the non-linear production term and the area change term. These two contributions will be considered together as the divergent contribution in our discussion below. For a quantitative view on the relative importance of the individual processes, we spatially integrate Equation 10:

$$
\int_A \frac{\partial V_P}{\partial t}\, dA = \frac{2}{n-1} \left[-\int_A \left(\sum_{i=1}^{n} \delta P_i (\delta \mathbf{v}_{i,\,\text{nTP}} + \delta \mathbf{v}_{i,\,\text{TPd}} \right. \right.
$$
$$
\left. \left. + \delta \mathbf{v}_{i,\,\text{div}}) \cdot \nabla \overline{P} \right) dA
$$

$$
\left. + \int_A \left(\sum_{i=1}^{n} \frac{(\delta P_i)^2}{2} \nabla \cdot \mathbf{v}_{i,\,\text{div}} \right) dA \right]
$$
$$
+ \text{bnd.} + \text{unc.} + \text{Nres}, \tag{11}
$$

where $dA = a^2 \cos\phi\, d\lambda\, d\phi$ is the area element in spherical coordinates with Earth radius a, longitude λ, and latitude ϕ. The uncertainty term,

$$
\text{unc.} = -\frac{2}{n-1} \int_A \left(\sum_{i=1}^{n} \delta P_i \delta \mathbf{v}_{i,\,\text{unc}} \cdot \nabla \overline{P} \right) dA,
$$

describes the uncertainty of the flow partitioning, while the boundary term bnd. describes the contribution from the flux term,

$$
\text{bnd.} = \frac{2}{n-1} \int_A \left(-\sum_{i=1}^{n} \nabla \cdot \left(\mathbf{v}_i \frac{(\delta P_i)^2}{2} \right) \right) dA
$$
$$
= \frac{2}{n-1} \oint_S \left(-\sum_{i=1}^{n} \frac{(\delta P_i)^2}{2} \mathbf{v}_i \cdot \mathbf{n} \right) dS,
$$

where dS denotes the boundary of the area and $\mathbf{n}$ the normal vector of the boundary pointing outward. The variance change observed between consecutive time steps acts as an indication of the representativeness of our diagnostic and is calculated by centred differences:[3]

$$
\text{obs.} = \int_A \frac{\partial V_P}{\partial t}\, dA \approx \frac{\int_A V_p(t + \Delta t)\, dA - \int_A V_p(t - \Delta t)\, dA}{2\Delta t}, \tag{12}
$$

with Δt being 3 hr in the first three forecast days and 6 hr afterwards. Equation 11 thus provides a novel diagnostic to quantify the relative importance of near-tropopause dynamics, tropospheric–deep interaction, and upper-tropospheric divergence in the amplification of forecast uncertainty.

3 | QUANTITATIVE VIEW OF THE VARIANCE AMPLIFICATION

3.1 | Synoptic overview and variance evolution

Before discussing the variance amplification in more detail, we provide a synoptic overview of our case, together with a description of the variance evolution (Figure 2).

The synoptic evolution of our case is characterized by a large-amplitude Rossby-wave pattern spanning from (counterclockwise) 180–60°E (as seen by the blue and black contour denoting the 2-PVU surface of the analysis and the ensemble mean, respectively, in Figure 2). Most interesting

---

[3]For the first and last time steps, we use instead forward and backward differences, respectively.
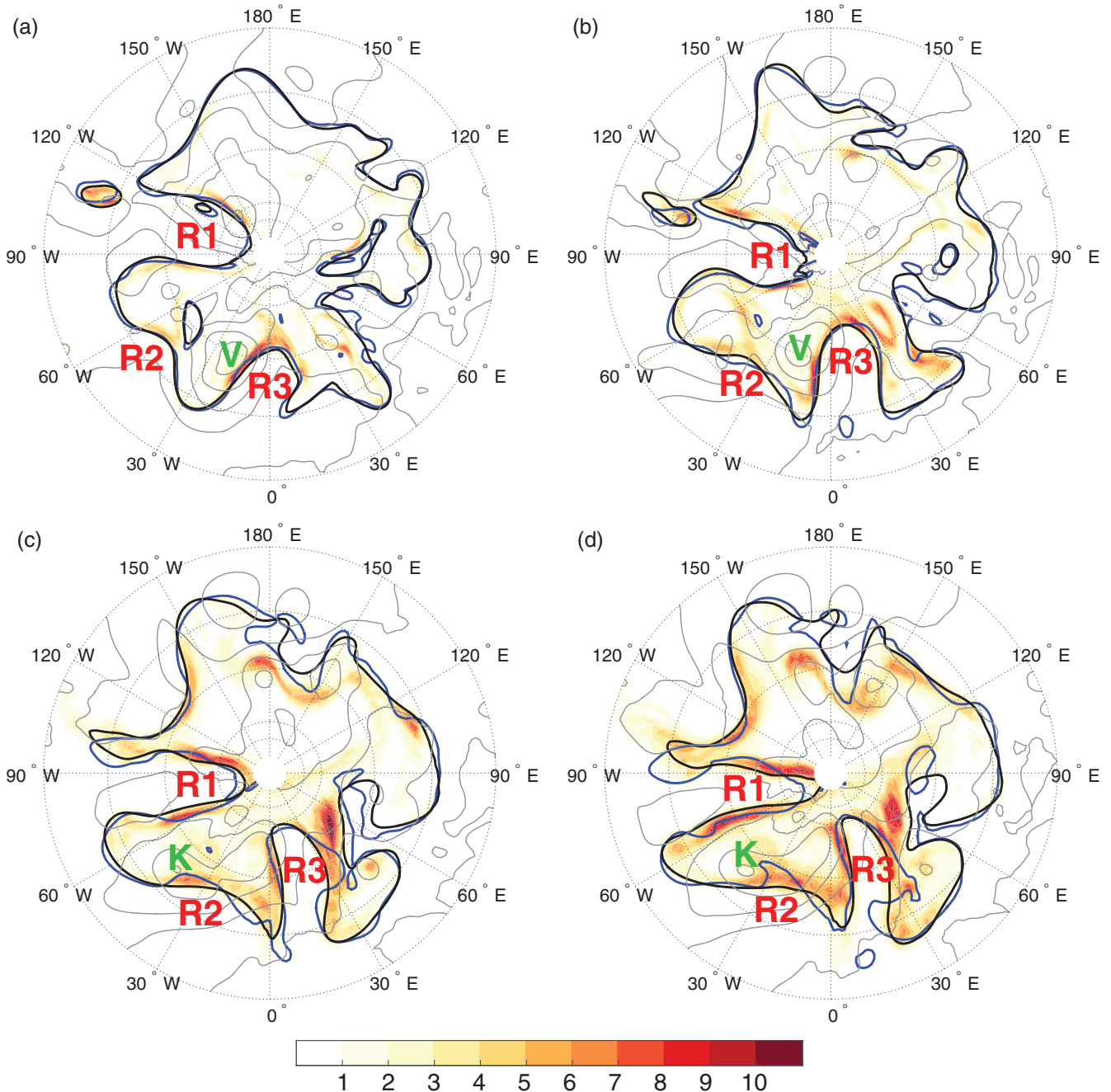
FIGURE 2 PV variance (coloured shading in PVU^2) at 325 K at (a) 2 days, (b) 3 days, (c) 4 days, and (d) 4.5 days. The black and blue contour denote the 2-PVU contour (smoothed over a box of 5×5 grid points using a mean filter) of the ensemble mean and the analysis, respectively. Grey contours denote the ensemble mean of mean sea-level pressure every 10 hPa (smoothed over a box of 3×3 grid points using a mean filter). Labels refer to individual ridges (prefix R) and the cyclones *Vladiana* (label V) and *Ex-Karl* (label K)

for the Rossby-wave evolution is the evolution of several ridges (labelled R1–R3 in Figure 2). At 2 days, a large-amplitude ridge exists around 150–75°W (label R1). Its upstream trough interacts with a cut-off around 110°W (2–3 days, Figure 2a,b) and completely reabsorbs this cut-off shortly after 3 days (Figure 2b). Ridge R1 is characterized by a large extension in the meridional direction and a contraction in the zonal direction and thereby exhibits highly nonlinear evolution (Figure 2c,d). Another ridge exists around

20°W–10°E (label R3), which is characterized by ridge building between 2 and 3 days (Figure 2a,b). This ridge building was associated with the development of cyclone *Vladiana* (labelled V in Figure 2), which was another observational highlight during the NAWDEX period, due to the occurrence of pronounced warm-conveyor-belt ascent (Schäfler *et al.*, 2018; Oertel *et al.*, 2019). In the following, ridge R3 is characterized by a large amplitude and a similar nonlinear evolution to ridge R1, albeit its spatial extent is smaller than that of

ridge R1 (Figure 2c,d). In between these larger-amplitude ridges, a smaller-amplitude ridge exists (label R2, Figure 2). Between 4 and 5 days, this ridge is influenced by the interaction with *Ex-Karl* around 40°W (labelled K in Figure 2), which leads to cyclonic wave-breaking of the ridge (as seen by the PV wrap-up of the analysis ridge around 4.5 days; blue contour in Figure 2d).

The ensemble variance of PV, which we use as our metric for forecast uncertainty, is not distributed homogeneously over the hemisphere (coloured shading in Figure 2), manifesting the well-known flow dependence of forecast uncertainty (e.g., Palmer and Hagedorn, 2006; Rodwell *et al.*, 2018). PV variance is maximized, in general, along the dynamical tropopause and exhibits a larger amplitude in the region of the Rossby-wave pattern (counterclockwise from 180°E–60°E) than in the more zonally-oriented region of the tropopause. In the time range investigated here, variance amplification occurs in both amplitude and scale. Several local maxima of variance amplification can be identified: (a) the cut-off evolution and its reabsorption by the waveguide around 2–3 days, (b) ridge-building events in association with cyclone development, for example, the ridge building of ridge R3 around 2–3 days, (c) highly nonlinear regimes of the wave pattern, for example, the large-amplitude ridge R1, and (d) the interaction between *Ex-Karl* and ridge R2 around 4–5 days.

3.2 | Individual contributions to variance amplification: spatial illustration

The previous subsection revealed several local maxima of variance amplification. To illustrate spatially the mechanisms that govern this variance amplification, we partition the variance tendency into the contributions of individual processes as detailed in section 2. Before looking at these individual contributions, the representativeness of our diagnostic for the actual variance amplification is assessed by comparing the advective variance tendency (right-hand side of Equation 8) with the observed variance tendency (left-hand side of Equation 8 approximated by centred differences) at two forecast lead times (2 and 4.5 days, Figure 3).

The observed variance tendency is characterized by dipole patterns along the tropopause that are associated with an eastward displacement of PV variance. This displacement is consistent with the eastward phase propagation of the Rossby-wave anomalies, in which the maxima of PV variance are located. In most of the dipole patterns, the positive part is larger than the negative part, leading to an overall amplification of PV variance. The main patterns of the observed variance tendency are captured well by the advective tendency, in terms of both the variance displacement and the overall variance amplification. The magnitude of the observed variance tendency, however, is smaller than that of the advective tendency, in particular at 4.5 days. One reason

for the smaller magnitude might be that the observed variance change has to be approximated by centred differences of 3-hr data for lead times smaller than 3 days and 6-hr data afterwards, which yields a smoothing and thereby a reduction of maxima and minima of the observed tendency. There thus exists a physically meaningful explanation for the differences between the observed and advective tendencies.

To remove those dipole patterns that are only associated with a displacement of variance and not with a net amplification of variance, we exclude the flux term of Equation 8 from our investigation, as it is only associated with a redistribution of PV variance. The main pattern of this net advective tendency (Figures 4a and 5a) is no longer characterized by dipole patterns, as it was the case for the full advective tendency (Figure 3b,d). Regions of variance amplification are thus much easier to identify. For the partitioning into the contributions from individual processes (Equation 10), we will thus discuss only the net advective tendency, and we will drop the prefix "net" for brevity.

At 2 days (Figure 4), a large amplitude of the advective tendency is, in general, found within the Rossby-wave pattern around 180°E–60°E (counterclockwise), with a particularly large variance amplification on the western flanks of ridges R1 and R3 around 130°W and 20°W, respectively, and in the trough around 40°E (Figure 4a). The individual contributions to this tendency are shown in Figure 4b–d. The advective tendency is mostly dominated by the near-tropopause tendency (Figure 4b). This tendency is particularly large where the Rossby-wave pattern exhibits a large amplitude, such as, for example, in ridge R1 around 120°W or in the trough around 40°E. One region in which not only the near-tropopause tendency makes a dominating contribution to the variance amplification is ridge R3 around 20°W. This ridge is characterized by ridge building in association with a cyclone development (Figure 2a,b) and its variance amplification is governed largely by both the near-tropopause and divergent tendency (Figure 4b,d). Compared with the near-tropopause and divergent tendency, the tropospheric–deep tendency is much smaller in amplitude (Figure 4c, note the different scale of the associated colour bar).

At 4.5 days (Figure 5), the variance tendency has become larger in scale than at 2 days and now exhibits a large amplitude almost everywhere along the dynamical tropopause. The large-amplitude ridge R1 (120–60°W) is still associated with large variance amplification (Figure 5a). Another region that exhibits large variance amplification is the smaller-amplitude ridge R2 where *Ex-Karl* interacts with the Rossby-wave pattern around 40°W. The processes that govern the variance amplification in the two ridges are distinct: While the near-tropopause tendency dominates the variance amplification in the large-amplitude ridge R1, it cannot explain the variance amplification in the
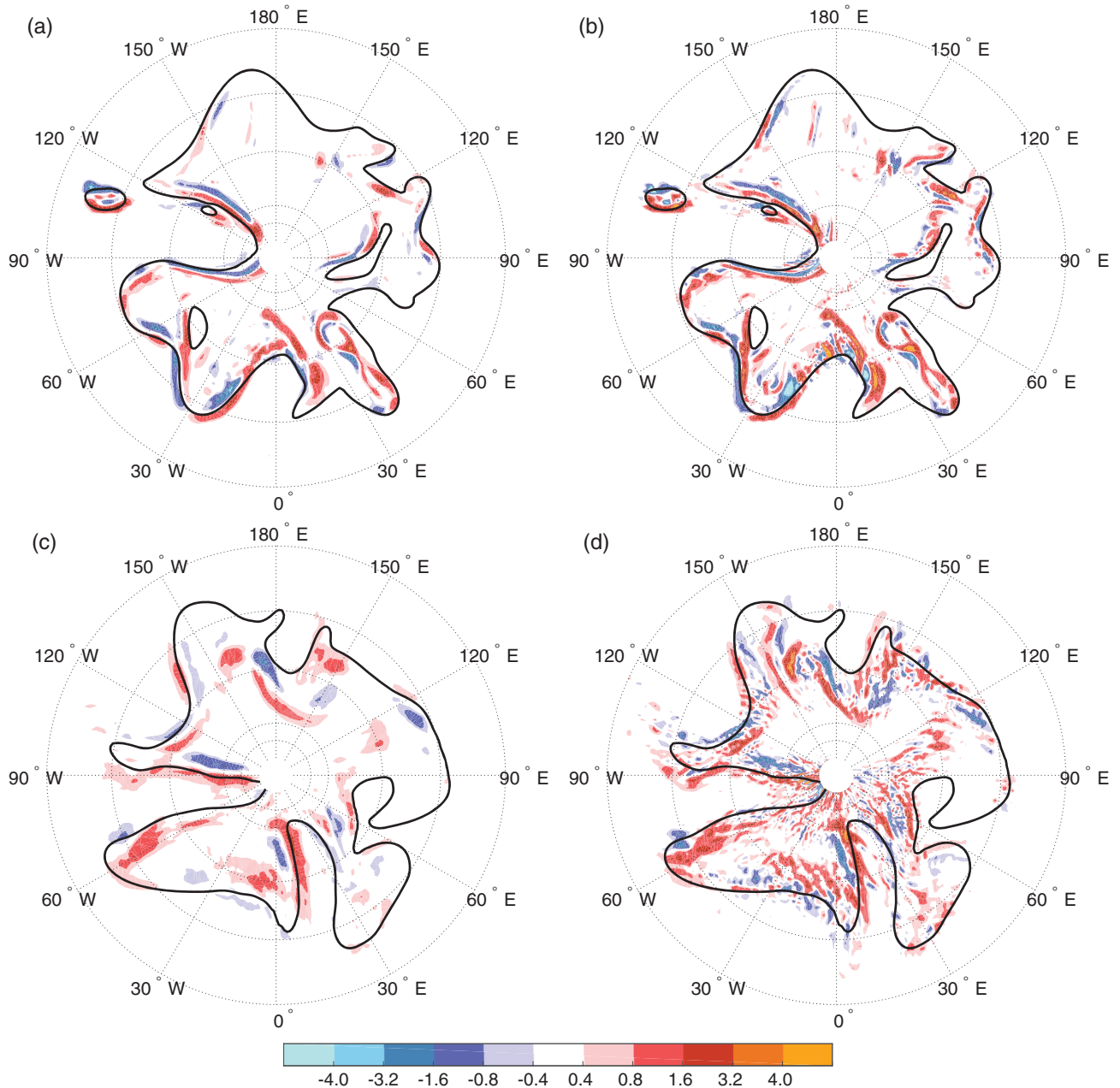
FIGURE 3 Comparison between the observed variance tendency (left-hand side of Equation 8, approximated by centred differences in (a) and (c)) and the advective variance tendency (right-hand side of Equation 8 in (b) and (d)) at 325 K at (a,b) 2 days and (c,d) 4.5 days. The coloured shading shows the respective variance tendency in 10^{-4} PVU^2/s and the black contour denotes the 2-PVU contour (smoothed over a box of 5×5 grid points using a mean filter) of the ensemble mean

smaller-amplitude ridge R2 (Figure 5b). In this ridge, the divergent and tropospheric–deep tendency make dominating contributions to the variance amplification (Figure 5c,d), which suggests that the variance amplification in this ridge is related mostly to the (moist baroclinic) cyclone development of *Ex-Karl*. A locally dominating contribution from the tropospheric–deep and divergent tendency is also found in the smaller-amplitude ridge around 130°W. The main pattern of the advective tendency, however, is dominated by the near-tropopause tendency, as at 2 days. Interestingly, the variance tendency in ridge R3 (around 10–40°W), which was

dominated by the divergent tendency at 2 days (Figure 4), is now dominated by the near-tropopause tendency, indicating a multi-stage behaviour of variance amplification, as was observed for upscale error growth (Baumgart *et al.*, 2019).

From the previous discussion, it is evident that the variance tendency is dominated in most regions by the near-tropopause tendency, but that there are also localized regions in which other processes dominate the variance tendency. In the following two subsections, we will thus quantify the variance amplification both from a hemispheric-averaged perspective for the midlatitudes of the Northern Hemisphere and from a
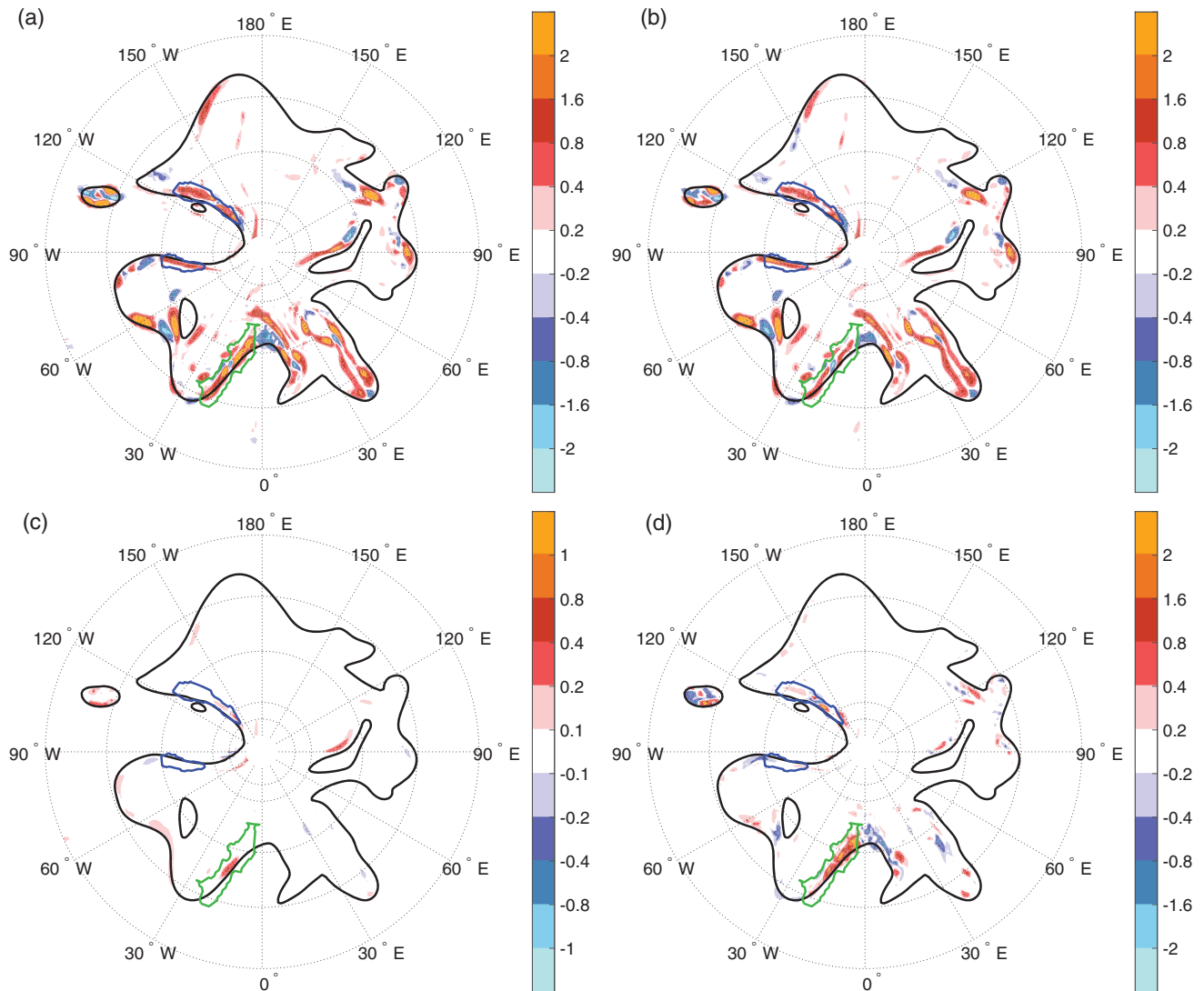
FIGURE 4 Variance tendency (coloured shading in 10^{-4} PVU^2 s^{-1}) at 325 K at 2 days for (a) full wind, (b) near-tropopause wind, (c) tropospheric–deep wind, and (d) divergent wind. The black contour denotes the 2-PVU contour (smoothed over a box of 5×5 grid points using a mean filter) of the ensemble mean. The blue and green contours indicate the regions used for the localized perspective on the variance amplification (Figure 7a,b)

local perspective for contrasting regions exhibiting (a) large near-tropopause variance amplification and (b) large divergent and/or tropospheric–deep variance amplification.

3.3 | Hemispheric variance amplification

For a quantitative perspective on the hemispheric-averaged variance amplification, Figure 6 shows the spatially averaged variance amplification for the midlatitudes of the Northern Hemisphere (30–80°N) and its partitioning into individual processes (Equation 11) as a function of lead time. The observed variance tendency (as approximated by centred differences, Equation 12) is mostly increasing with lead time until around 4.5 days and then mostly decreasing until it is almost zero after around 8 days. A particularly large growth of the observed tendency is found in the first 1.5 days

and between 3.5 and 4.5 days. Between these two periods (day 1.5–3.5), the observed tendency is almost constant. The initially large variance amplification might be related to the singular-vector approach at ECMWF, which maximizes the growth of ensemble perturbations in the first 2 days, whereas the later nonhomogeneous growth of the observed tendency points to a flow dependence of the variance amplification with a period of lower predictability and thus larger variance amplification between 3.5 and 4.5 days.

Our diagnostic yields a residual, which increases in magnitude in the first 2 days and is then almost constant at a value of around -3×10^{-6} PVU^2/s. This residual is related to those processes that cannot be measured with the available data (see section 2). Based on the results of Baumgart *et al.* (2018), dissipation can be expected to make a dominant negative contribution. Dissipation yields a smoothing of small-scale PV
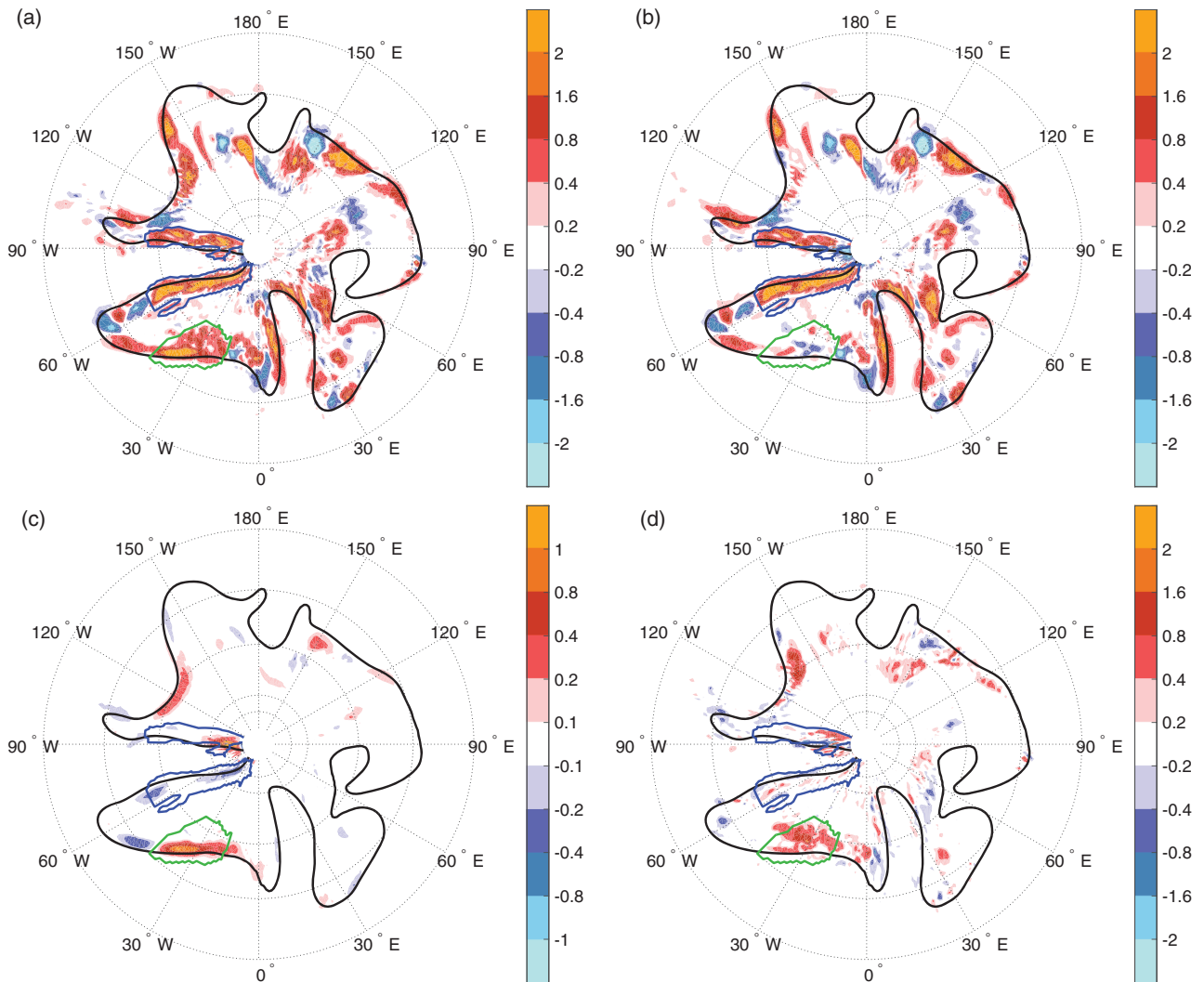
FIGURE 5 Variance tendency (coloured shading in 10^{-4} PVU^2/s) at 325 K at 4.5 days for (a) full wind, (b) near-tropopause wind, (c) tropospheric–deep wind, and (d) divergent wind. The black contour denotes the 2-PVU contour (smoothed over a box of 5×5 grid points using a mean filter) of the ensemble mean. The blue and green contours indicate the regions used for the localized perspective on the variance amplification (Figure 7c,d)

features in association with the downscale cascade of enstrophy. In this way, dissipation also leads to a smoothing of the PV variance associated with small-scale PV features and thereby provides a variance sink that explains the negative residual observed in Figure 6. This important role of dissipation was also discussed in further previous studies (e.g., Zhang *et al.*, 2007; Saffin *et al.*, 2016; 2017; Baumgart *et al.*, 2019).

The near-tropopause tendency clearly dominates the hemispheric variance amplification. This tendency makes by far the largest contribution to the observed variance tendency in the first 7 days and its general time evolution corresponds well with that of the observed variance tendency. In the first two days, the near-tropopause tendency increases almost linearly with lead time, suggesting that the singular-vector approach, which maximizes the perturbation growth in the first 2 days, projects most strongly on to the near-tropopause tendency.

The divergent tendency is almost as large as the near-tropopause tendency at the first time step, but much smaller afterwards. This small importance suggests that the divergent variance amplification observed in the spatial maps (Figures 4d and 5d) is only important in a localized sense, not in a hemispheric sense. Interestingly, the divergent tendency shows a small peak around 4–4.5 days, which corresponds to the time when *Ex-Karl* interacts with the waveguide.

The tropospheric–deep tendency is small in the first 4 days. Afterwards, it increases slowly and becomes larger than the near-tropopause tendency after 7 days, when the near-tropopause tendency is small. This slow increase of the tropospheric–deep tendency suggests that the tropospheric–deep variance amplification depends on the horizontal scale of the potential-temperature uncertainty at the lower boundary (875 hPa), which has to become large enough in scale to penetrate vertically up to the tropopause
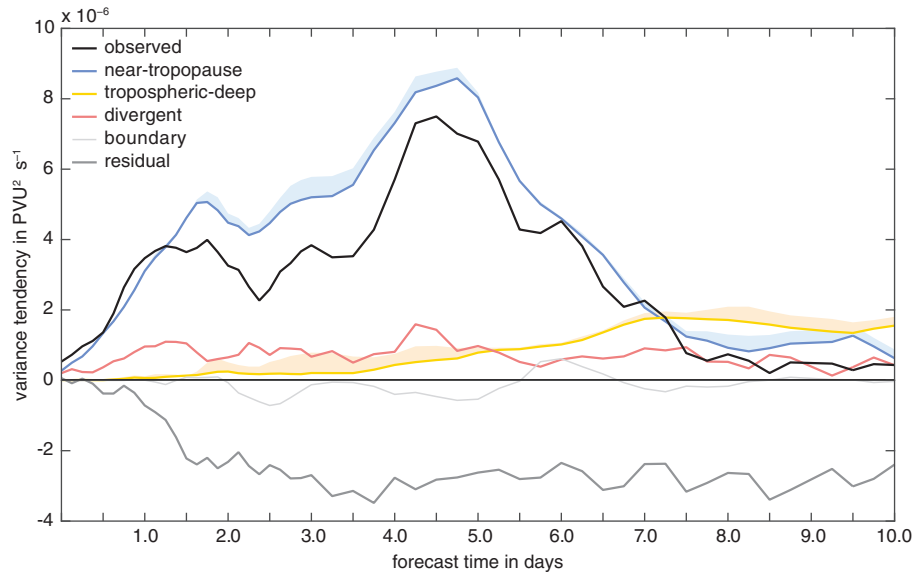
FIGURE 6 Individual contributions (coloured lines) to the variance tendency (spatially averaged over the midlatitudes of the Northern Hemisphere, 30°N–80°N) on the 325-K surface as indicated by Equation 11 as a function of forecast time. The observed tendency is approximated by centred differences in time of the spatially averaged variance (obs. in Equation 12). The shading next to the near-tropopause and tropospheric–deep tendencies indicates the uncertainty of the result due to the uncertainty of PPVI (unc. in Equation 11)

region. Such a scale dependence would be consistent with the scale dependence found for error growth in the Eady (1949) model (Baumgart *et al.*, 2018).

3.4 | Variance amplification in localized regions

The hemispheric variance amplification investigated in the previous subsection was dominated by near-tropopause dynamics (Figure 6). In the spatial maps (Figures 4 and 5), we found, however, that in localized regions the variance amplification can be dominated by other processes. In the following, we will thus consider the two time steps used for the spatial maps (2 and 4.5 days) and quantify the variance amplification for two localized regions at each time step that exhibit different characteristics in terms of the processes governing the variance amplification. The localized regions of variance amplification are chosen by a threshold of the smoothed advective tendency ($> 0.2 \times 10^{-4}$ PVU^2/s, smoothing is performed over a box of 5×5 grid points using a mean filter). These integration regions are indicated by the blue and green contours in Figures 4 and 5.

At 2 days, there are two localized regions of variance amplification that exhibit different characteristics in terms of the processes, the large-amplitude ridge R1 around 150–75°W and the smaller-amplitude ridge R3 around 20–0°W (Figure 4). From a quantitative perspective, the variance amplification in the large-amplitude ridge is dominated by the near-tropopause tendency, which contributes about 90% to the advective variance amplification (Figure 7a). The variance amplification in the smaller-amplitude ridge,

instead, is dominated by upper-tropospheric divergence, which contributes about two thirds to the advective variance amplification (Figure 7b).

At 4.5 days, there are again two localized regions of variance amplification that exhibit different characteristics in terms of the processes, the large-amplitude ridge R1 around 120–60°W, which was also investigated at 2 days, and the smaller-amplitude ridge R2 around 45–15°W (Figure 5). In the large-amplitude ridge, the largest contribution to the variance amplification is again given by the near-tropopause tendency (almost 100% of the full advective amplification, Figure 7c). In the small-amplitude ridge, instead, the near-tropopause tendency makes only a very small contribution to the variance amplification (about 1% of the full advective amplification, Figure 7d). In this ridge, the variance amplification is largely governed by the divergent tendency (almost 50% of the full advective amplification) and the tropospheric–deep tendency (about 25% of the full advective amplification).[4]

This comparison of localized variance amplification shows that there can be large flow-dependent differences in the processes that govern variance amplification. The large importance of near-tropopause dynamics in the large-amplitude ridge R1 at 2 and 4.5 days suggests that

---

[4]Note that there is also a large uncertainty of the piecewise PV inversion (about 25% of the full advective amplification), which could be attributed to both the near-tropopause and the tropospheric–deep tendency. The general observation that the near-tropopause tendency does not make the largest contribution to the variance amplification, however, holds even when the whole uncertainty of the diagnostic is attributed to the near-tropopause tendency.
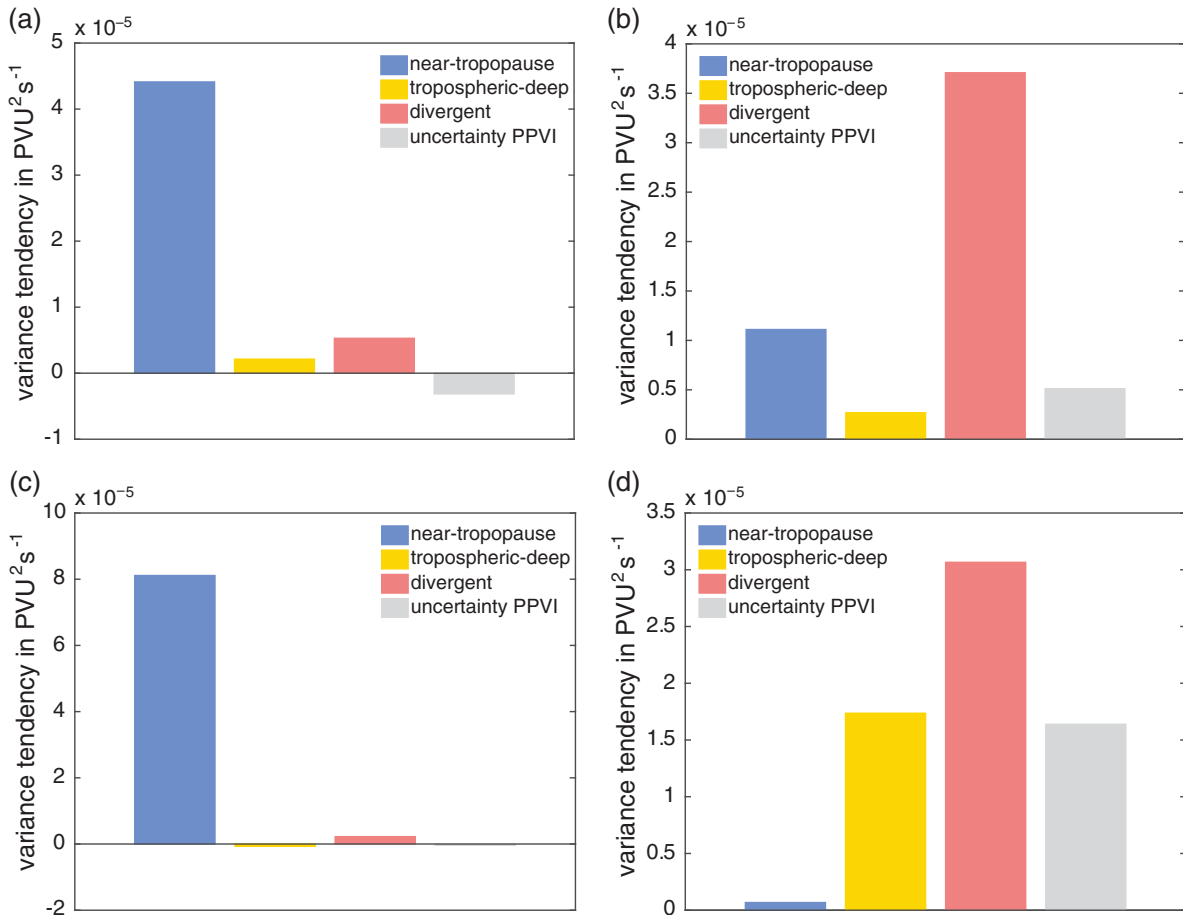
FIGURE 7 Individual contributions (coloured bars) to the variance tendency on the 325-K surface. The variance tendency is averaged over localized regions of variance amplification: (a) in the flanks of large-amplitude ridge R1 around 150–75°W at day 2 (integration area indicated by blue contours in Figure 4), (b) on the western flank of the smaller-amplitude ridge R3 around 20°W–0° at day 2 (green contour in Figure 4), (c) in the flanks of large-amplitude ridge R1 around 120–60°W at day 4.5 (blue contours in Figure 5), and (d) in the smaller-amplitude ridge R2 around 45–15°W at day 4.5 (green contour in Figure 5)

this contribution is particularly large when the flow is highly nonlinear. The large importance of the divergent and tropospheric–deep tendency in the smaller-amplitude ridges R2 and R3, in which ridge building occurs in association with a cyclone development, suggests that the variance amplification during the earlier phase of ridge building is mainly related to uncertainties in the development of the respective cyclones, in terms of both the associated warm conveyor belt and divergent outflow (large divergent contribution at 2 and 4.5 days) and the associated baroclinic growth (large tropospheric–deep contribution at 4.5 days).

4 | COMPARISON BETWEEN ERROR GROWTH IN THE BEST AND WORST MEMBERS

As noted in the Introduction, two members are distinct from the rest of the ensemble members at 5 days in terms of error enstrophy being exceptionally small and large, respectively

(Figure 1). In addition to our analysis of variance amplification (section 3), which provides a mean picture of the divergence of ensemble members, it is also interesting to analyse the divergence of these two individual members, which are associated with the largest differences in error enstrophy of the ensemble. This section thus investigates in more detail the mechanisms leading to the pronounced differences in the error evolution of the two members.

The difference between the hemispheric error enstrophy of the worst and best members is shown in Figure 8 (black line) as a function of forecast time. In the first 2.5 days, this difference is small, but it increases prominently afterwards, until the two members become the best and worst members of the ensemble at 5 days (Figure 1). To gain insight into the origin of this large difference in the hemispheric error enstrophy of the two members, the hemisphere is split into four quadrants and the difference in error enstrophy is calculated separately for each quadrant (coloured lines in Figure 8). From this partitioning, it is evident that all four quadrants contribute in a similar way to the hemispheric error-enstrophy difference
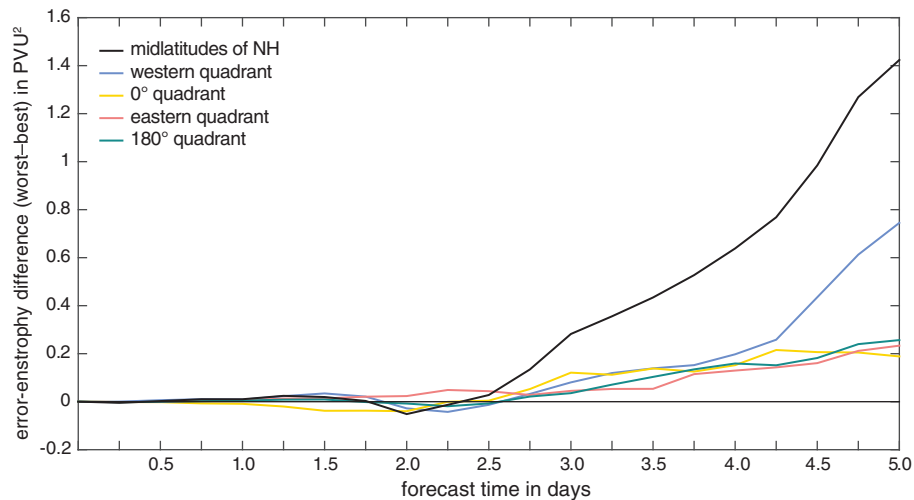
FIGURE 8 Difference between error enstrophy of the worst and best members at 325 K (spatially averaged over the midlatitudes of the Northern Hemisphere, 30–80°N, black line), partitioned into the contributions from individual segments of the hemisphere, with each segment spanning a longitude range of 90° (coloured lines)

between 2.5 and 4 days. Between 4 and 5 days, however, the western quadrant contributes much more strongly to the hemispheric error-enstrophy difference than the other three quadrants. Finally, this quadrant contributes more than 50% to the hemispheric error-enstrophy difference at 5 days, whereas the other three quadrants contribute individually only 13–18%.

Comparing spatial maps of the PV error of the best and worst members (Figure 9) reveals that the large error-enstrophy difference in the western quadrant is mostly related to the highly nonlinear region of the wave pattern (120–60°W). At 4.5 days (Figure 9c,d), this region is captured rather well by the best member, while the worst member exhibits large errors that are related to both the phase and the shape of the wave pattern. This difference in the error pattern can be traced back in time (a manual error tracking is applied as in Magnusson, 2017) to around 3 days when the cut-off interacts with the Rossby-wave pattern (around 110°W, Figure 9a,b). The worst member exhibits a distinct phase error in the trough that interacts with the cut-off, whereas the error of the best member is small in this region. This difference between the error pattern of the two members then amplifies largely in the highly nonlinear region of the wave pattern, leading to the pronounced differences in the error pattern around 4–5 days (Figure 9c,d). The important role of a misrepresented cut-off evolution leading to large error growth in the further evolution was also discussed by Grams *et al.* (2018) for a forecast-bust event in March 2016.

For a quantitative and process-based perspective on the error-growth differences between the best and worst members, Figure 10 compares the spatially averaged error-enstrophy tendency (Equation 4) in the midlatitudes of the Northern Hemisphere (30–80°N) for both members, together with its partitioning into the contribution from individual processes. In the first 2.5 days, the error tendency

of the best and worst members is rather similar in terms of the observed error-tendency magnitude, consistent with the fact that the error enstrophy of the two members first diverges after 2.5 days (Figures 1 and 8). During this time, the error growth of the two members is also similar in terms of the processes governing the error growth, with the near-tropopause tendency dominating the error growth in both members.

Around 2.5 days, the best member shows a distinct drop in the observed error tendency, with a magnitude around zero between 2.5–2.75 days. The error tendency in the worst member does not show such a distinct drop, but instead increases largely with forecast time, starting from a tendency of about 0.3×10^{-5} PVU^2/s around day 2.5 and ending at a tendency of about 1.3×10^{-5} PVU^2/s between 4 and 5 days. A particularly large error growth is found between 4 and 5 days, which corresponds to the time when the error in the worst member becomes much larger than that of all other ensemble members (Figure 1). Between 4 and 5 days, the observed error tendency of the best member is only about 0.5×10^{-5} PVU^2/s and thereby less than half that of the worst member.

In terms of the processes, the difference between the error growth in the best and worst members can mostly be attributed to the near-tropopause tendency. While the near-tropopause tendency between 2.5 and 5 days is about 0.4×10^{-5} PVU^2/s for the best member, it is about 1.4×10^{-5} PVU^2/s for the worst member. The magnitude of the divergent tendency, instead, is very similar and both best and worst members show a modest maximum around 4.5 days when *Ex-Karl* interacts with the waveguide. Between 2.5 and 5 days, the tropospheric–deep tendency is positive for the best member, whereas it is negative for the worst member. This tendency, however, is much smaller than the near-tropopause tendency and thus not of large importance to the error-growth differences between the two members. Baumgart *et al.*
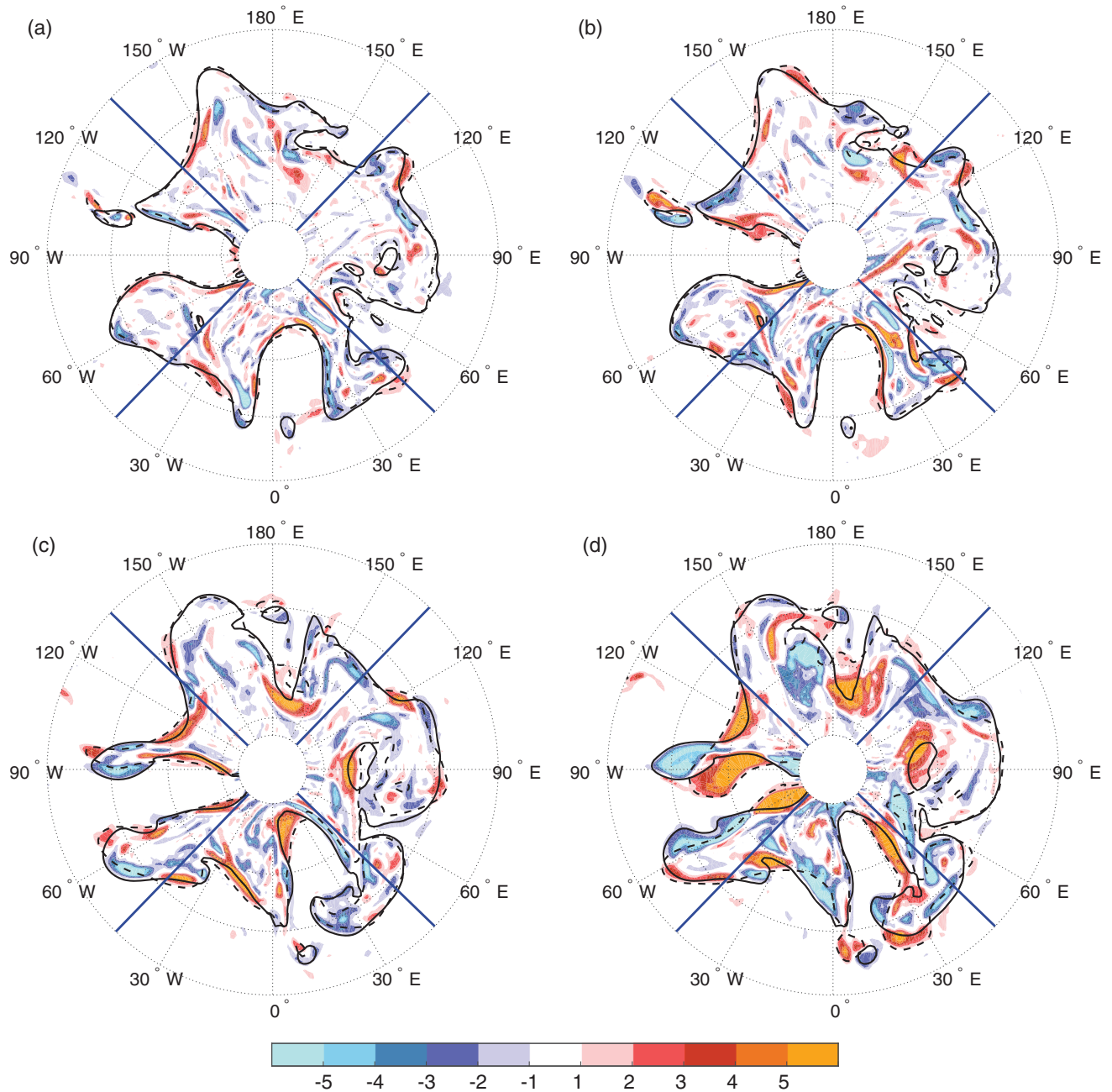
FIGURE 9 Comparison between the PV error (coloured shading in PVU) in the best member (a,c: member 9) and worst member (b,d: member 18) at 325 K at (a,b) 3 days and (c,d) 4.5 days. The solid (dashed) black contours denote the 2-PVU contour (smoothed over a box of 5×5 grid points using a mean filter) of the analysis (respective ensemble member). The blue lines indicate the regions used for the partitioning of error-enstrophy difference between the worst and best members in Figure 8

(2018) showed that the near-tropopause error tendency is mostly related to differences in the nonlinear Rossby-wave dynamics. The large importance of near-tropopause tendency to the error-growth differences between the best and worst members is thus consistent with the observation that the largest differences between the best and worst members occur in the western quadrant of the hemisphere, where the Rossby-wave pattern is highly nonlinear (Figures 8 and 9).

5 | SUMMARY AND DISCUSSION

This study provides a quantitative framework to investigate the processes governing the amplification of forecast uncertainty and forecast errors in ensemble forecasts. A tendency equation for the ensemble variance of PV is derived and partitioned into the contributions from individual processes. The framework is applied to a case from the North Atlantic Waveguide and Downstream Impact Experiment (NAWDEX), namely the interaction of tropical storm *Karl*
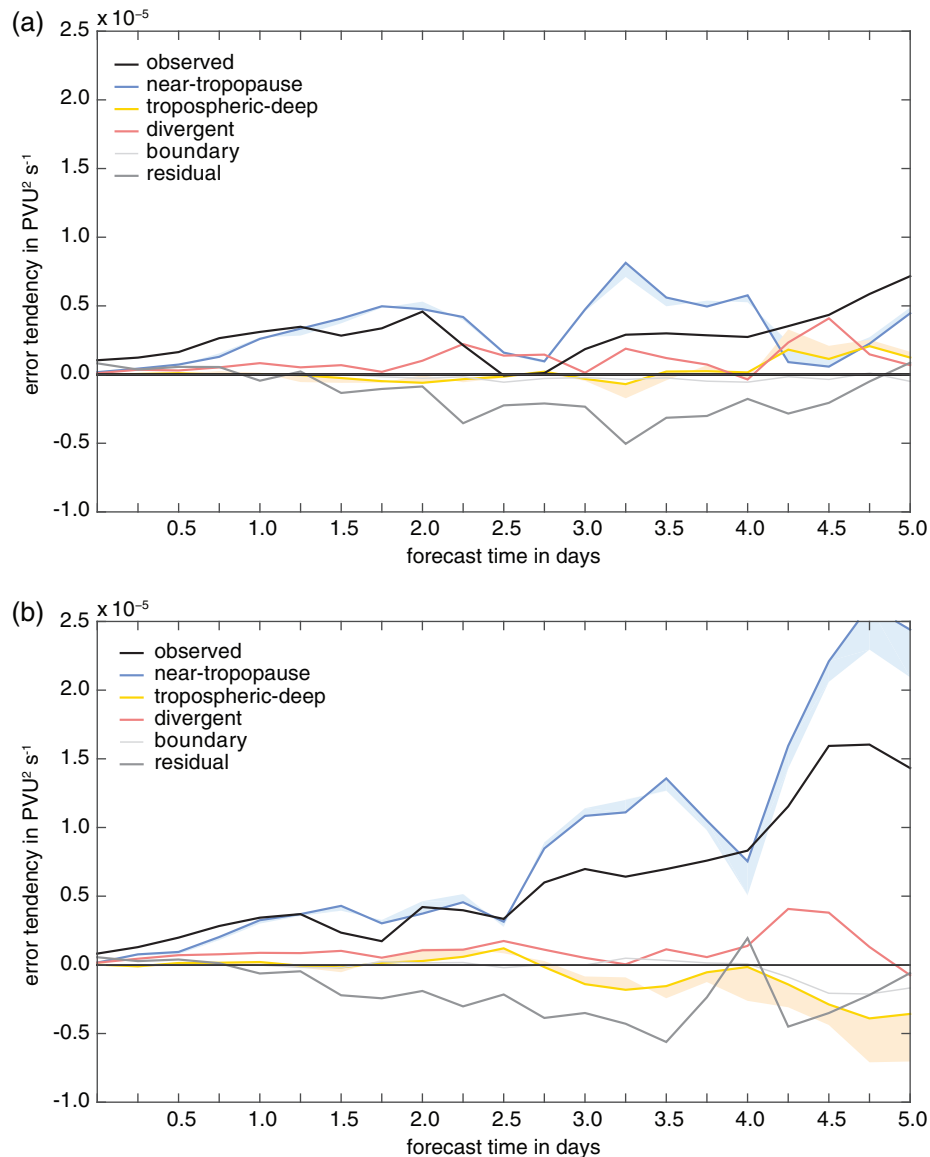
FIGURE 10 Individual contributions (coloured lines) to the error-enstrophy tendency (spatially averaged over the midlatitudes of the Northern Hemisphere, 30–80°N) on the 325-K surface as indicated by Equation 4 as a function of forecast time, for (a) the best member (member 9), and (b) the worst member (member 18). The observed tendency is approximated by centred differences in time of the spatially averaged error enstrophy. The shading next to the near-tropopause and tropospheric–deep tendencies indicates the uncertainty of the result due to the uncertainty of PPVI

and the Rossby-wave pattern during extratropical transition. In the medium range, this interaction was associated with large forecast uncertainty (Schäfler *et al.*, 2018) and only a few members captured the evolution correctly (Kumpf *et al.*, 2018). Here, the same ensemble forecast as in Kumpf *et al.* (2018) is investigated. However, we not only focus on the region of *Ex-Karl*, but also provide a hemispheric perspective on the ensemble evolution, before local differences, including the evolution of *Ex-Karl*, are highlighted. Two aspects of the ensemble behaviour are our main interests: (a) the mean divergence of the ensemble members, indicating the general amplification of forecast uncertainty, and (b) the divergence of the best and worst members, indicating extremes in possible error-growth scenarios.

The synoptic evolution of our case is characterized by a large-amplitude Rossby-wave pattern spanning from (counterclockwise) 180–60°E. This Rossby-wave pattern is, in general, associated with larger PV variance than the other part of the hemisphere, in which the tropopause is more zonally oriented. Several local maxima of PV variance exist, including a cut-off, a large-amplitude ridge, a ridge-building event, and the interaction between *Ex-Karl* and the strong midlatitude PV gradient. The variance amplification is, on average for the midlatitudes of the Northern Hemisphere, dominated by near-tropopause dynamics. This contribution is particularly large in highly nonlinear regions of the wave pattern, such as, for example, in the large-amplitude ridge around 120–60°W. Locally, however, the variance amplification can also be

dominated by other processes. One prominent example is the region in which *Ex-Karl* interacts with the Rossby-wave pattern around 4.5 days lead time. In this region, the variance amplification is dominated by upper-tropospheric divergence and tropospheric–deep interaction and is thereby mostly related to uncertainties in (moist baroclinic) cyclone development.

The differences between the error growth in the best and worst members can, to a large part, be attributed to the highly nonlinear evolution of the large-amplitude ridge around 120–60°W. Around 4–5 days, this region is captured well by the best member, whereas the worst member exhibits large errors associated with both the phase and the shape of the wave pattern. This different error pattern can be traced back in time until 3 days, when a cut-off interacts with the upstream trough. During this interaction, the worst member is characterized by a distinct phase error, which largely amplifies in the highly nonlinear region of the wave pattern until 5 days. In terms of the processes, the differences in error growth between the two members are dominated by differences in the error growth due to near-tropopause dynamics, which manifests the large importance of nonlinear Rossby-wave dynamics.

The large importance of near-tropopause dynamics to the variance amplification and the error-growth differences between the best and worst members is consistent with the large importance of near-tropopause dynamics found in a recent case study of error growth in a deterministic ECMWF operational forecast (Baumgart *et al.*, 2018) and in numerical error-growth experiments (Baumgart *et al.*, 2019). In the current study, however, we also identify distinct local differences in the variance amplification, such as, for example, in the region of *Ex-Karl*. In addition, the large (hemispherically averaged) error-growth differences between the best and worst ensemble members around 5 days could be linked to a specific local event, namely the cut-off interaction around 2–3 days. The approach presented is thus able to identify flow features that are of large importance for the amplification of ensemble spread. Because the interaction of both *Ex-Karl* and the cut-off with the strong midlatitude PV gradient can be interpreted as vortex–wave interactions, we speculate that such interactions are particularly sensitive to perturbations and are therefore prominent "amplifiers" of uncertainty in medium-range forecasts.

To extend our case study, future work could apply the approach presented in a more systematic way, for example, to the trough/CAPE-type flow situations identifed by Rodwell *et al.* (2013; 2018) as associated with increased forecast uncertainty in the medium range. Rodwell *et al.* (2018) demonstrated that there are deficiencies in the short-range reliability associated with this flow situation, by investigating a budget equation of ensemble variance in observation space. Our complementary approach is able to identify the processes governing the flow-dependent amplification of ensemble spread in the medium range. As illustrated in Figure 1, the amplification of forecast errors and forecast uncertainty can be highly nonlinear, in the sense that ensemble members with relatively small errors at short lead times may have relatively large errors at medium-range lead times, and vice versa. Notwithstanding the importance of short-range sources of forecast uncertainty, a better understanding of the flow-dependent nature of the medium-range amplification of forecast errors and forecast uncertainty (as also highlighted in previous studies, for example, Ferranti *et al.*, 2014) is also of importance for the design and interpretation of ensemble forecasts.

Even though the focus of this study is on the medium range, it is of interest to consider the evolution of ensemble spread in the short range briefly as well. During the first 1–2 days, the near-tropopause tendency dominates, on average, the amplification of PV variance. This result is in contrast to results from upscale-error-growth simulations, in which the ensemble members differ only in terms of the stochastic seed of the convection scheme (Selz, 2019). In these simulations, error growth in the early phase of the simulations is dominated by moist processes (Baumgart *et al.*, 2019). This difference indicates (i) that the ensemble perturbations representing initial-condition uncertainty in the ECMWF operational system in 2016 project most strongly on the near-tropopause tendency and (ii) that this initial-condition uncertainty dominates, on average, over the (direct) generation of ensemble spread by processes with low intrinsic predictability, most notably moist processes. The latter notion is consistent with the lack of ensemble spread in low-predictability scenarios identified by Rodwell *et al.* (2018). Extending the focus of our analysis to the spread evolution at short lead times in low-predictability scenarios will require nonconservative tendency data that are not available in the archives of operational centres. Results from such an analysis, however, may yield important insight into the deficiencies of flow-dependent reliability and may be of practical relevance for the design of stochastic parametrization schemes

REFERENCES

Bauer, P., Thorpe, A. and Brunet, G. (2015) The quiet revolution of numerical weather prediction. *Nature*, 525, 47–55. https://doi.org/10.1038/nature14956.

Baumgart, M., Ghinassi, P., Wirth, V., Selz, T., Craig, G.C. and Riemer, M. (2019) Quantitative view on the processes governing the upscale error growth up to the planetary scale using a stochastic convection scheme. *Monthly Weather Review*, 147(5), 1713–1731. https://doi.org/10.1175/MWR-D-18-0292.1.

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

Berner, J., Achatz, U., Batté, L., Bengtsson, L., Adl, C., Christensen, H.M., Colangeli, M., Coleman, D.R.B., Crommelin, D., Dolaptchiev, S.I., Franzke, C.L.E., Friederichs, P., Imkeller, P., Järvinen, H., Juricke, S., Kitsios, V., Lott, F., Lucarini, V., Mahajan, S., Palmer, T.N., Penland, C., Sakradzija, M., von Storch, J.S., Weisheimer, A., Weniger, M., Williams, P.D. and Yano, J.I. (2017) Stochastic parametrization: toward a new view of weather and climate models. *Bulletin of the American Meteorological Society*, 98(3), 565–588. https://doi.org/10.1175/BAMS-D-15-00268.1.

Buizza, R. (2018) Introduction to the special issue on "25 years of ensemble forecasting". *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/10.1002/qj.3370.

Chagnon, J.M., Gray, S.L. and Methven, J. (2013) Diabatic processes modifying potential vorticity in a North Atlantic cyclone. *Quarterly Journal of the Royal Meteorological Society*, 139(674), 1270–1282. https://doi.org/10.1002/qj.2037.

Charney, J. (1955) The use of the primitive equations of motion in numerical prediction. *Tellus*, 7(1), 22–26. https://doi.org/10.1111/j.2153-3490.1955.tb01138.x.

Davies, H.C. and Didone, M. (2013) Diagnosis and dynamics of forecast error growth. *Monthly Weather Review*, 141(7), 2483–2501. https://doi.org/10.1175/MWR-D-12-00242.1.

Davis, C.A. (1992) Piecewise potential vorticity inversion. *Journal of the Atmospheric Sciences*, 49(16), 1397–1411. https://doi.org/10.1175/1520-0469(1992)049<1397:PPVI>2.0.CO;2.

Davis, C.A. and Emanuel, K.A. (1991) Potential vorticity diagnostics of cyclogenesis. *Monthly Weather Review*, 119(8), 1929–1953. https://doi.org/10.1175/1520-0493(1991)119<1929:PVDOC>2.0.CO;2.

Davis, C.A., Grell, E.D. and Shapiro, M.A. (1996) The balanced dynamical nature of a rapidly intensifying oceanic cyclone. *Monthly Weather Review*, 124(1), 3–26. https://doi.org/10.1175/1520-0493(1996)124<0003:TBDNOA>2.0.CO;2.

Davis, C.A., Stoelinga, M.T. and Kuo, Y.H. (1993) The integrated effect of condensation in numerical simulations of extratropical cyclogenesis. *Monthly Weather Review*, 121(8), 2309–2330. https://doi.org/10.1175/1520-0493(1993)121<2309:TIEOCI>2.0.CO;2.

Eady, E.T. (1949) Long waves and cyclone waves. *Tellus*, 1(3), 33–52. https://doi.org/10.1111/j.2153-3490.1949.tb01265.x.

ECMWF (2018) Part V: Ensemble Prediction System. ECMWF. IFS Documentation No. 5. Available at: https://www.ecmwf.int/node/18715.

Ertel, H. (1942) Ein neuer hydrodynamischer Erhaltungssatz. *Meteorologische Zeitschrift*, 30(36), 543–544. https://doi.org/10.1007/BF01475602.

Ferranti, L., Corti, S. and Janousek, M. (2014) Flow-dependent verification of the ECMWF ensemble over the Euro–Atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, 141(688), 916–924. https://doi.org/10.1002/qj.2411.

Grams, C.M., Magnusson, L. and Madonna, E. (2018) An atmospheric dynamics perspective on the amplification and propagation of forecast error in numerical weather prediction models: a case study. *Quarterly Journal of the Royal Meteorological Society*, 144(717), 2577–2591. https://doi.org/10.1002/qj.3353.

Grams, C.M., Wernli, H., Bottcher, M., Campa, J., Corsmeier, U., Jones, S.C., Keller, J.H., Lenz, C.J. and Wiegand, L. (2011) The key role of diabatic processes in modifying the upper-tropospheric wave guide: a North Atlantic case-study. *Quarterly Journal of the Royal Meteorological Society*, 137(661), 2174–2193. https://doi.org/10.1002/qj.891.

Gutowski, W.J., Branscome, L.E. and Stewart, D.A. (1992) Life cycles of moist baroclinic eddies. *Journal of the Atmospheric Sciences*, 49(4), 306–319. https://doi.org/10.1175/1520-0469(1992)049<0306:LCOMBE>2.0.CO;2.

Heifetz, E., Bishop, C.H., Hoskins, B.J. and Methven, J. (2004) The counter-propagating Rossby-wave perspective on baroclinic instability. I: mathematical basis. *Quarterly Journal of the Royal Meteorological Society*, 130(596), 211–231. https://doi.org/10.1256/qj.02.184.

Holton, J.R. and Hakim, G.J. (2013) *An Introduction to Dynamic Meteorology*. Science Academic Press. Available at: https://books.google.com/books?id=-ePQ6x6VbjgC.

Hoskins, B.J., McIntyre, M.E. and Robertson, A.W. (1985) On the use and significance of isentropic potential vorticity maps. *Quarterly Journal of the Royal Meteorological Society*, 111(470), 877–946. https://doi.org/10.1256/smsqj.47001.

Kumpf, A., Rautenhaus, M., Riemer, M. and Westermann, R. (2019) Visual analysis of the temporal evolution of ensemble forecast sensitivities. *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 98–108. https://doi.org/10.1109/TVCG.2018.2864901.

Kumpf, A., Tost, B., Baumgart, M., Riemer, M., Westermann, R. and Rautenhaus, M. (2018) Visualizing confidence in cluster-based ensemble weather forecast analyses. *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 109–119. https://doi.org/10.1109/TVCG.2017.2745178.

Leutbecher, M. and Palmer, T.N. (2008) Ensemble forecasting. *Journal of Computational Physics*, 227(7), 3515–3539. https://doi.org/10.1016/j.jcp.2007.02.0140.

Lorenz, E.N. (1963) Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2), 130–141. https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Lynch, P. (1989) Partitioning the wind in a limited domain. *Monthly Weather Review*, 117(7), 1492–1500. https://doi.org/10.1175/1520-0493(1989)117<1492:PTWIAL>2.0.CO;2.

Magnusson, L. (2017) Diagnostic methods for understanding the origin of forecast errors. *Quarterly Journal of the Royal Meteorological Society*, 143(706), 2129–2142. https://doi.org/10.1002/qj.3072.

Martnez-Alvarado, O., Joos, H., Chagnon, J., Boettcher, M., Gray, S.L., Plant, R.S., Methven, J. and Wernli, H. (2014) The dichotomous

structure of the warm conveyor belt. *Quarterly Journal of the Royal Meteorological Society*, 140(683), 1809–1824. https://doi.org/10.1002/qj.2276.

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

Palmer, T. and Hagedorn, R. (2006) *Predictability of Weather and Climate*. Cambridge University Press. Available at: https://books.google.de/books?id=c-rY28QQCj8C.

Piaget, N., Froidevaux, P., Giannakaki, P., Gierth, F., Martius, O., Riemer, M., Wolf, G. and Grams, C.M. (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. https://doi.org/10.1002/qj.2496.

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

Riemer, M., Baumgart, M. and Eiermann, S. (2014) Cyclogenesis downstream of extratropical transition analyzed by Q-vector partitioning based on flow geometry. *Journal of the Atmospheric Sciences*, 71(11), 4204–4220. https://doi.org/10.1175/JAS-D-14-0023.1.

Riemer, M. and Jones, S.C. (2010) The impact of tropical cyclones on a developing baroclinic wave in idealized scenarios of extratropical transition. *Quarterly Journal of the Royal Meteorological Society*, 136(648), 617–637. https://doi.org/10.1002/qj.605.

Rodwell, M.J., Lang, S.T.K., Ingleby, N.B., Bormann, N., Hólm, E., Rabier, F., Richardson, D.S. and Yamaguchi, M. (2016) Reliability in ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 142(694), 443–454. https://doi.org/10.1002/qj.2663.

Rodwell, M.J., Magnusson, L., Bauer, P., Bechtold, P., Bonavita, M., Cardinali, C., Diamantakis, M., Earnshaw, P., Garcia-Mendez, A., Isaksen, L., Kaellen, E., Klocke, D., Lopez, P., McNally, T., Persson, A., Prates, F. and Wedi, N. (2013) Characteristics of occasional poor medium-range weather forecasts for Europe. *Bulletin of the American Meteorological Society*, 94(9), 1393–1405. https://doi.org/10.1175/BAMS-D-12-00099.1.

Rodwell, M.J., Richardson, D.S., Parsons, D.B. and Wernli, H. (2018) Flow-dependent reliability: a path to more skillful ensemble forecasts. *Bulletin of the American Meteorological Society*, 99(5), 1015–1026. https://doi.org/10.1175/BAMS-D-17-0027.1.

Saffin, L., Gray, S.L., Methven, J. and Williams, K.D. (2017) Processes maintaining tropopause sharpness in numerical models. *Journal of Geophysical Research*, 122(18), 9611–9627. https://doi.org/10.1002/2017JD026879.

Saffin, L., Methven, J. and Gray, S.L. (2016) The non-conservation of potential vorticity by a dynamical core compared with the effects of parametrized physical processes. *Quarterly Journal of the Royal Meteorological Society*, 142(696), 1265–1275. https://doi.org/10.1002/qj.2729.

Schäfler, A., Craig, G., Wernli, H., Arbogast, P., Doyle, J.D., McTaggart-Cowan, R., Methven, J., Rivière, G., Ament, F., Boettcher, M., Bramberger, M., Cazenave, Q., Cotton, R., Crewell, S., Delanoë, J., Dörnbrack, A., Ehrlich, A., Ewald, F., Fix, A., Grams, C.M., Gray, S.L., Grob, H., Groß, S., Hagen, M., Harvey, B., Hirsch, L., Jacob, M., Kölling, T., Konow, H., Lemmerz, C., Lux, O., Magnusson, L., Mayer, B., Mech, M., Moore, R., Pelon, J., Quinting, J., Rahm, S., Rapp, M., Rautenhaus, M., Reitebuch, O., Reynolds, C.A., Sodemann, H., Spengler, T., Vaughan, G., Wendisch, M., Wirth, M., Witschas, B., Wolf, K. and Zinner, T. (2018) The North Atlantic waveguide and downstream impact experiment. *Bulletin of the American Meteorological Society*, 99(8), 1607–1637. https://doi.org/10.1175/BAMS-D-17-0003.1.

Schneidereit, A., Peters, D.H.W., Grams, C.M., Quinting, J.F., Keller, J.H., Wolf, G., Teubler, F., Riemer, M. and Martius, O. (2017) Enhanced tropospheric wave forcing of two anticyclones in the prephase of the January 2009 major stratospheric sudden warming event. *Monthly Weather Review*, 145(5), 1797–1815. https://doi.org/10.1175/MWR-D-16-0242.1.

Selz, T. (2019) Estimating the intrinsic limit of predictability using a stochastic convection scheme. *Journal of the Atmospheric Sciences*, 76(3), 757–765. https://doi.org/10.1175/JAS-D-17-0373.1.

Snyder, C., Hamill, T.M. and Trier, S.B. (2003) Linear evolution of error covariances in a quasigeostrophic model. *Monthly Weather Review*, 131(1), 189–205. https://doi.org/10.1175/1520-0493(2003)131<0189:LEOECI>2.0.CO;2.

Teubler, F. and Riemer, M. (2016) Dynamics of Rossby-wave packets in a quantitative potential vorticity–potential temperature framework. *Journal of the Atmospheric Sciences*, 73(3), 1063–1081. https://doi.org/10.1175/JAS-D-15-0162.1.

Whitaker, J.S. and Loughe, A.F. (1998) The relationship between ensemble spread and ensemble mean skill. *Monthly Weather Review*, 126(12), 3292–3302. https://doi.org/10.1175/1520-0493(1998)126<3292:TRBESA>2.0.CO;2.

Zhang, F., Bei, N., Rotunno, R., Snyder, C. and Epifanio, C.C. (2007) Mesoscale predictability of moist baroclinic waves: convection-permitting experiments and multistage error growth dynamics. *Journal of the Atmospheric Sciences*, 64(10), 3579–3594. https://doi.org/10.1175/JAS4028.1.