https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

Isolating flow-dependent uncertainty in ensemble reanalysis data and its relation to Euro-Atlantic weather regimes and warm conveyor belts

Henry Schoeller[1] and Stephan Pfahl[1]

[1]Institute of Meteorology, Freie Universität Berlin, Berlin, Germany

Correspondence: Henry Schoeller (henry.schoeller@fu-berlin.de)

Abstract. Data assimilation uncertainty varies greatly based on the quality and amount of ingested observations and the uncertainty in the background forecast. Ensemble data assimilation (EDA) schemes quantify this combined uncertainty. We investigate climatologically what governs this uncertainty on daily to synoptic time scales, but isolating the effects of changes to the observation system from the truly flow-dependent part of the uncertainty is not straightforward.

Drawing on the EDA produced for the ECMWF 5th Generation Reanalysis product (ERA5), we isolate the long- from the short-time-scale uncertainty components using grid-point-wise statistical models. We investigate the patterns and causes of the short time-scale component for a set of weather regimes and the occurrence of warm conveyor belts (WCBs) in the European North Atlantic sector. This provides coherent regions of higher and lower assimilation uncertainty in different weather regimes that can be explained by a combination of mathematical and physical arguments. Moisture presence is a key contributor for elevated assimilation uncertainty and our results indicate that this influence is mediated largely through state-dependent observation uncertainties in addition to model uncertainties usually considered to govern sub-daily error growth.

Consequently, we argue that what is commonly conceived as a model or weather system specific uncertainty (the “error-of-the-day") in ERA5 reanalysis is actually highly influenced by derived instrument uncertainties and assimilation challenges. Given ERA5’s wide use as well as the role of EDA in providing initial ensemble forecast spread, considering the dynamics of assimilation uncertainty should be standard practice.

1 Introduction

Knowledge about the state of the atmosphere at a given point in time is crucial both for forecasting as well as weather and climate research. Such knowledge is of limited value, however, if not accompanied by a measure of uncertainty, given the substantial differences in the quality of estimates across space and time. While such uncertainty is explicitly represented in forecasting contexts through initial ensemble perturbations, the uncertainty associated with reanalysis datasets deserves attention in its own right, given their paramount importance for climatological analyses of the atmosphere.

In reanalysis production, data assimilation combines observations with a numerical model such that the resulting state estimate is consistent with both (Kalnay et al., 2024, ch. 5). There are numerous approaches to data assimilation and, correspondingly, several ways to quantify the uncertainty of an assimilated state. In ensemble-based methods, an ensemble of realizations is computed, the spread of which serves as a measure of assimilation uncertainty. In variational frameworks, these uncertainty estimates determine the relative weighting of the background (first guess) and the observations according to their associated errors (Bannister, 2017). The ensemble of data assimilations (EDA) is commonly used to characterize uncertainty in the estimated atmospheric state.

Apart from their diagnostic purpose, assimilation uncertainties also serve as initial perturbations for ensemble forecasts (Kalnay et al., 2024, ch. 6). Their dual role in producing both initial perturbations for ensemble prediction and uncertainty estimates for reanalysis datasets provides the opportunity to study assimilation uncertainty on a larger data base than any single ensemble data set affords.

Ensemble forecasts are intended to provide a range of probable future developments given imperfect knowledge about the system and have successfully extended the forecast-skill horizon (Buizza and Leutbecher, 2015). The growth of the initial ensemble spread with time varies across space, time, variable and flow-state. This is because the error-growth characteristics of a model depend on its design, especially with respect to how small-scale processes are resolved or parametrized (Palmer et al., 2005; Palmer, 2019b; Matsunobu et al., 2024), but also because physical processes responsible for error growth are more prevalent in some atmospheric situations than others (Clarke et al., 2019; Puh et al., 2024). At short lead times, the construction of the initial perturbations itself additionally plays a role (Lang et al., 2012; Raynaud and Bouttier, 2016; Wang et al., 2023).

For forecasts up to a few days, dedicated experiments with grid-point-level initial perturbations have shown that convective processes efficiently propagate small-scale errors to larger scales (Zhang et al., 2007; Selz and Craig, 2015; Judt, 2018; Baumgart et al., 2019), reminiscent of the “butterfly effect” conceptualized by Lorenz (1969). At operational perturbation levels, however, error growth at short lead times is concentrated at larger scales and driven by adiabatic processes, with ensemble spread occasionally even decreasing at the smallest scales (Selz et al., 2022). Consistent with this, cyclogenesis case studies identify baroclinic rather than convective instability as the main driver of spread near the tropopause (Rodwell and Wernli, 2023), a conclusion supported by process-based analyses (Baumgart et al., 2018; Baumgart and Riemer, 2019). The practical relevance of the butterfly effect has thus been called into question, since relatively small errors at the large scales may dominate the spread before small-scale errors escalate (Durran and Gingrich, 2014; Lloveras et al., 2023).

Yet, convectively active regions are frequently highlighted in studies concerned with error growth that utilize operational ensemble forecasts initialized with assimilation uncertainty (Lillo and Parsons, 2017; Sánchez et al., 2020; Puh et al., 2024). Given that the upscale propagation pathway is of limited relevance at realistic perturbation levels, the elevated spread in these regions may instead reflect enhanced initial uncertainty due to weaker observational constraints (Rodwell et al., 2013), or be produced during the assimilation process itself.

Several lines of evidence support this interpretation. Globally, Žagar (2017) found that initial ensemble spread is concentrated in the unbalanced modes, especially in the tropics, and that these dominate the error growth for roughly the first day — a pattern more consistent with an initial-uncertainty origin than with an upscale convective cascade. Rodwell and Wernli (2023) analysed short-range background-forecast error growth in cyclogenesis and found that it was not spatially congruent with regions of strong precipitation, indicating that the spread enters the forecast through observations or through the singular-vector perturbations rather than through convective error production. In a similar case study, Cardinali et al. (2014) found that considerable EDA spread was concentrated in cloudy regions of developing cyclones in an assimilation run without stochastic physics. Consistently, the projection onto unbalanced modes mentioned above was already present also; the addition of model-error perturbations only marginally enhanced both.

Observational and assimilation difficulties in cloudy scenes provide a natural explanation. Satellite radiance uncertainties are larger in cloudy conditions (Geer and Bauer, 2011; Li et al., 2016; Geer et al., 2018), aircraft typically avoid regions of intense convection, and rapid small-scale variability combined with representation challenges in models further complicates the assimilation (Janjić et al., 2018). In particular, convectively active regions are characterized by shorter decorrelation length scales, implying that, even for the same observational coverage and quality, the amount of information extractable via assimilation shrinks (Cardinali and Prates, 2011; Lee and Huang, 2020; Li et al., 2023).

Crucially, although convection acts on small spatial scales, the spread it produces in the analysis is not confined to small wavenumbers. As Lloveras et al. (2022) emphasise, an uncertainty localised in physical space is necessarily broad in spectral space; convectively driven assimilation uncertainty therefore does not rely on an up-scale transfer via the butterfly effect to limit predictability.

Taken together, these findings suggest that forecast error growth cannot be understood independently of the uncertainty structure of the assimilated initial state. Just like model-intrinsic error growth, assimilation uncertainty is inherently flow dependent, and its spatial structure may strongly shape subsequent forecast uncertainty (Magnusson et al., 2019). Studying which flow states and processes dominate assimilation uncertainty is thus necessary both for interpreting predictability studies and for identifying where improvements in observations, data assimilation, and model formulation are most relevant for practical forecast skill.

These considerations carry over to reanalysis datasets, where the associated uncertainty is primarily informative about the quality and density of the observation system at a given point in time, though the flow-state dependent uncertainty dominant in the forecast setting plays a role as well. To the authors' knowledge, a systematic climatological characterization of this flow-dependence in reanalysis uncertainty has not yet been undertaken. Such a characterization would clarify how regime-conditioned climatologies derived from these datasets are affected by uncertainty in the conditioning fields themselves – a potential source of bias that has so far received little attention.

In this work, we characterize the uncertainty of the geopotential at 500 hPa in the ERA5 reanalysis (Hersbach et al., 2020) from a climatological perspective, with particular attention to its flow-dependent nature and the role of moist processes therein. ERA5 is widely considered one of the highest quality reanalysis datasets to date and is widely treated as a standard reference in atmospheric research. Derived from a 4D-Ensemble-Variational approach, it features the desired characteristics for our analysis, namely a state-dependent "error-of-the-day" inferred by the EDA approach mentioned above (Bonavita et al., 2016). How exactly EDA characterizes the uncertainty of a certain state can be studied in dedicated assimilation experiments (Cardinali, 2013; Cardinali et al., 2014), but only on very limited scale compared to a full reanalysis and characterizing diabatic influences in this way is not straight forward (Janisková and Cardinali, 2017).

A key difficulty in adopting a climatological perspective is that changes in the observing system dominate variability on long time scales. We address this issue by statistically modeling this influence independently at every grid point in our domain, which

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

offers additional insights into the regional disparities in how changes in the observing system have affected assimilation quality. Additionally accounting for seasonal effects, we obtain a data set comprising assimilation uncertainty variability predominantly on short time scales, whose drivers we then investigate by means of a set of second-stage statistical models.

The markedly flow-dependent nature of short-time-scale uncertainty becomes evident when stratifying the data into a set of weather regimes (WRs) defined over the North Atlantic–European region. In our analysis, we employ the set of seven such regimes introduced by Grams et al. (2017). These regimes are well suited for the present purpose, as they effectively distinguish between the dominant synoptic flow patterns on the time scales of interest, are defined year-round and have been widely used in studies of predictability (Spaeth et al., 2024; Osman et al., 2023; Büeler et al., 2021; Hauser et al., 2026), though we are not aware of any studies on their relation to observation and assimilation uncertainty.

The regime-specific patterns of assimilation uncertainty provide a basis for uncovering the physical and dynamical processes governing uncertainty at a local level. We do so by evaluating the extent to which the regime signal can be accounted for by variables more directly indicative of the underlying processes, such as those characterizing cloud cover, moisture content, hydrometeor amount, baroclinicity or jet stream intensity.

Alongside this local perspective, we carry out a more non-local, process-oriented analysis by examining the role of warm conveyor belts (WCBs). These features are of particular relevance in the context of uncertainty due to the strong influence of moist processes associated with them (Madonna et al., 2014), and can act as amplifiers of forecast uncertainty as small perturbations in their characteristics — such as ascent or moisture availability — efficiently project onto larger scales through convective and baroclinic instabilities (Pickl et al., 2023; Rodwell et al., 2018). This sensitivity to environmental perturbations is well documented (Schäfler and Harnisch, 2015; Oertel et al., 2025; Mathews and Czaja, 2024; Christ et al., 2025), as is the resulting link between WCBs and operational forecast uncertainty (Grams et al., 2018; Wandel et al., 2024). Much of this work attributes the uncertainty to model formulation, in particular the representation of microphysical and diabatic processes (Joos and Wernli, 2012; Joos and Forbes, 2016; Hieronymus et al., 2022; Oertel et al., 2023; Rivière et al., 2021; Mazoyer et al., 2023). The extent to which the amplification instead originates in elevated initial uncertainty has received little dedicated attention; Pickl et al. (2022) compared stochastic model physics with initial-condition perturbations and found the latter comparatively minor, though their target was WCB occurrence rather than forecast error growth. In addition, the interaction of WCBs with atmospheric blocking (Steinfeld and Pfahl, 2019; Wandel et al., 2024) is particularly relevant here, providing a direct link to several of the WRs considered.

The remainder of this paper is structured as follows. Sections 2 and 3 introduce the datasets utilized in this study. Section 3 also establishes the factors responsible for spread in the ERA5 EDA; readers familiar with data assimilation may skim through it. Section 4 presents the methodological framework to statistically separate long- from short-time-scale components of uncertainty and link them to dynamical covariates. Section 5 presents the results, while Sect. 6 summarizes the main findings and discusses their implications.

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

2 Data

We make use of the fifth-generation European Reanalysis product (ERA5) provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). Technical aspects and evaluation have been documented by Hersbach et al. (2020), while its extensions back to 1950 and 1940 have been described by Soci et al. (2024) and Bell et al. (2021) respectively. Note that we do not make use of the ERA5.1 product as we do not investigate stratospheric levels (Simmons et al., 2020). Since we want to investigate the flow-dependent uncertainty of the best estimate of the atmosphere at any time (the “analysis” or HRES product) we access both the ensemble spread of the data assimilation ensemble (EDA) used to estimate the assimilation uncertainty and the actual analysis. Given that the characteristics of these data sets are at the core of this study, we provide a brief introduction into EDA in Sect. 3.

For comparability, we use the same 0.5° spatial resolution in the analysis that is used for the ensemble. Since we are interested in the variability of the uncertainty on synoptic time scales, daily data at 12:00 UTC is sufficient. Our region of interest is the same as employed in the definition of the Euro-Atlantic year-round WRs (see below) and consists of grid points between 80° W and 40° E, and between 30° N and 90° N. This domain lends itself well for the analysis of weather-dependent reanalysis uncertainty as it has a long history of relatively dense atmospheric measurements (e.g. Soci et al. (2024)) and since its dynamics are highly variable and have been thoroughly researched (Woollings, 2010).

Weather classification is an ubiquitous tool in atmospheric sciences and a wide range of different approaches exist (Hannachi et al., 2017). To group spatially varying levels of uncertainty by weather situations we make use of the well-established year-round Euro-Atlantic WRs introduced by Grams et al. (2017) and updated by Hauser et al. (2024); Grams (2026). The regimes are defined as patterns of the geopotential height at 500 hPa derived from a `k-means` clustering in EOF space obtained after spatio-temporal filtering. While more detailed information about the similarity of any instantaneous geopotential height field with each of the WR prototypes is available (through projection on each of the patterns), we simply use the multinomial classification of each day into one of the eight categories (seven regimes and one transitional category), such that instead of seven continuous numerical time series, we simply use one categorical time series of eight categories. The data is available back to 1950.

To study the influence of WCBs on assimilation uncertainty, we make use of a Eulerian WCB identification dataset. It consists of two-dimensional fields indicating the occurrence of a WCB in inflow, ascent or outflow stage at each grid point. The identification process involves Lagrangian trajectory calculation, WCB classification according to an ascent criterion (at least 600 hPa in 48 h) and proximity to an extratropical cyclone and, finally, grouping into the three vertical categories: below 800 hPa (inflow), 800-400 hPa (ascent) and above 400 hPa (outflow). The occurrence is subsequently mapped to a regular $1^\circ \times 1^\circ$ latitude-longitude grid. We use data in 12-hourly intervals to account for lagged responses. The identification was originally introduced by Madonna et al. (2014) and later extended by Sprenger et al. (2017). The dataset we use contains data from 1979 to 2022, which were kindly provided by Quinting et al. (2022) (we use the original training data rather than the “learned” fields their transformer predicts).

3 Ensemble Data Assimilation

This study is concerned with the flow-state dependence of assimilation uncertainty and the various factors and processes influencing it. It is therefore instrumental to first briefly review how both the analysis and the associated uncertainty are computed. For a comprehensive introduction into data assimilation we refer to Park (2022) and to Park et al. (2022) for recent research directions. The data assimilation system used by the ECMWF (both for ERA5 and operational forecasting), an incremental hybrid ensemble 4D-Var system, is described by Hersbach et al. (2020) and in more detail by Bonavita et al. (2016), which this section largely follows.

Denote[1] by $\mathbf{d} = \mathbf{y}^{\mathrm{O}} - \mathbf{y}$ the misfit between the observations and the guess projected into observation space by $\mathbf{y} = H(\mathbf{x})$, where H is the observation operator and $\mathbf{x}$ is the guess (we neglect bias terms, because they make no difference in our analysis). In each of the 12-hourly time windows through which the production of the data set iterates, the best estimate – the "analysis" $\mathbf{x}^{\mathrm{a}}$ – is obtained as the minimizer of the following cost function:

$$J(\boldsymbol{\delta}\mathbf{x}) = \underbrace{\frac{1}{2}\boldsymbol{\delta}\mathbf{x}^T\mathbf{B}^{-1}\boldsymbol{\delta}\mathbf{x}}_{\text{background term}} + \underbrace{\frac{1}{2}\mathbf{d}^T\mathbf{R}^{-1}\mathbf{d}}_{\text{misfit term}}, \tag{1}$$

where $\boldsymbol{\delta}\mathbf{x} = \mathbf{x} - \mathbf{x}^{\mathrm{b}}$ is the increment from the initial guess (the background) obtained via numerical integration of the model. The solution of Eq. (1) – i.e. the minimizing $\mathbf{x}^{\mathrm{a}}$ – is also referred to as "HRES" in the context of ERA5. To achieve a balance between observations and model prediction that is scaled by their respective uncertainty, each of the terms contains weighting by the appropriate covariance matrices $\mathbf{B}$ and $\mathbf{R}$ (Courtier et al., 1994). $\mathbf{R}$ depends on the observation and is either prescribed (constant in time) or derived (state-dependent).

The background covariance matrix $\mathbf{B}$ follows a hybrid formulation (Isaksen et al., 2010; Bonavita et al., 2016) between a climatological and a state-dependent part according to

$$\mathbf{B} = (1-\alpha)\mathbf{B}_{\mathrm{cli}} + \alpha\mathbf{B}_{\mathrm{EDA}}, \tag{2}$$

with α increasing by wavenumber from .15 to .74. α was inadvertently kept constant during production of ERA5, which affects the ensemble spread for data between January 2001 to June 2005, and from January 2010 to October 2014 (see Sect. 5 also). $\mathbf{B}_{\mathrm{cli}}$ is supposed to encompass the climatological component of the background uncertainty dominated by the respective instrument generation and has been kept constant from January 1940 to November 1972 (H. Hersbach, 2025, personal communication), from November 1972 to December 1978, from January 1979 to December 1999 and from January 2000 onwards (Bell et al., 2021).

$\mathbf{B}_{\mathrm{EDA}}$ is calculated from the spread of a 10 member (for ERA5; 1 control and 9 perturbed) ensemble run of Eq. (1) at a lower spatial resolution (TL319 as opposed to TL639) which is referred to as "EDA" (ensemble data assimilation) in contrast to

---

[1] The notation referring to the assimilation methodology follows Ide et al. (1997) for consistency and is notably different to the statistical notation below.

HRES. The perturbations are added to the observations $\mathbf{y}^{\mathrm{O}}$ and are, crucially, scaled by the expected observation uncertainties. Each member also calculates its own independent background forecast through numerical integration initialized by its previous analysis and including stochastically perturbed physics (except for the control; Ecmwf (2016b); Leutbecher et al. (2017)). Perturbed physics are also applied in the nonlinear part of the construction of $\mathbf{x}$ during the minimization of Eq. (1) – the so-called “outer-loop". The EDA is a Monte Carlo estimate of the background-error posterior; because its sample size is small for computational reasons, it is supplemented by a climatological component to offset sampling error (Bonavita et al., 2011, 2016; Bannister, 2017). As such, it provides the “error-of-the-day" as an impactful improvement to the analysis quality compared to climatological background errors especially during extreme events (Bonavita et al., 2012), though its small sample size precludes trustworthy detailed structures (Isaksen et al., 2010; Bonavita et al., 2011).

Apart from providing the flow-dependent part of the background forecast uncertainty, the EDA serves also as a relative estimate of the uncertainty of the HRES[2]. It is for both of these reasons that the EDA dataset is at the core of this study. There are several aspects determining the spread of the EDA ensemble, which we discuss in the following.

3.1 Ensemble Spread

Though variational bias correction is applied in the EDA production (Dee, 2005), bias parameters are constant across the ensemble members (Hersbach et al., 2020). We therefore neglect bias terms again. With each ensemble member's deviation from the ensemble mean $\mathbf{x}_i^{\prime \mathrm{a}} = \mathbf{x}_i^{\mathrm{a}} - \overline{\mathbf{x}}^{\mathrm{a}}$, the linear approximation of the ensemble spread is (Dee and Da Silva, 1998; Higham, 2002):

$$\mathbb{E}\left[\mathbf{x}^{\prime \mathrm{a}} \mathbf{x}^{\prime \mathrm{a}^T}\right] = (\mathbf{I} - \mathbf{KH})\mathbf{B} = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1}, \text{ where} \tag{3}$$

$$\mathbf{H} = \left.\frac{\partial H}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^{\mathrm{b}}}, \text{ the linearized observation operator} \tag{4}$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\top}\left(\mathbf{H}\mathbf{B}\mathbf{H}^{\top} + \mathbf{R}\right)^{-1}, \text{ the Kalman gain.} \tag{5}$$

The causes for differences between ensemble members can be divided into three parts: firstly, the differences in the perturbations to the observations $\mathbf{y}^{\mathrm{O}}$ that are scaled by $\mathbf{R}$ and mapped into models space by $\mathbf{K}$, secondly, the differences in the background forecast $\mathbf{x}^{\mathrm{b}}$ that are scaled by $\mathbf{B}$ and, thirdly, the ensemble spread introduced by the physics perturbations by the full, non-linear outer-loop simulation across the analysis window that is affected by all terms in Eq. (3). Post-hoc statistical analysis by Cardinali (2013) revealed that observations have a share of roughly 20% of the influence on the assimilated state directly, while the remaining part is accounted for by the background forecasts (at least with the operational setup and the observations around that time). This proportion does not, however, translate cleanly to relative shares of observations and model

---

[2] Relative in the sense that the numerical values of the ensemble spread should not be taken at face value, since the EDA is underdispersive (Buizza et al., 2008; Soci et al., 2024), which is why they are augmented with singular vector perturbations in the operational ECMWF ensemble forecast (Leutbecher and Palmer, 2008; Palmer, 2019a; Lang et al., 2021a). Its small sample size entails substantial noise that would need to be filtered first (Raynaud et al., 2009; Bonavita et al., 2011).

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

uncertainties on EDA spread, since past (perturbed) observations exert an influence on spread as each EDA member calculates its background from its own previous analysis. For EDA spread specifically, Cardinali et al. (2014) found that disabling stochastic physics lead to a spread reduction by roughly a quarter in the Northern Hemisphere in the mid-troposphere (their figure 5). The contribution of both observation and model uncertainty to the EDA spread was shown to project onto balanced vs. unbalanced modes (according to the methodology by Žagar et al. (2011)) in a similar way.

3.1.1 Observation Perturbations

Since observations constrain the dynamical model, a stronger constraint should also induce a smaller spread of the ensemble around that constraint. Two distinct mechanisms link observational properties to EDA spread: the magnitude of the perturbations applied to each ingested observation, and the effective constraint exerted by the observation network as a whole. As uncertain observations also project onto a more uncertain background forecast, observation perturbations are supplied to estimate background error covariances (Isaksen et al., 2010).

The EDA observation perturbations are drawn from the observation-error model used in the Integrated Forecast System (IFS): each perturbation variance equals the per-observation error variance (the diagonal of $\mathbf{R}$) or the operator-computed, scene-dependent variance for radiances and similar instruments (Ecmwf, 2016a; Isaksen et al., 2010; Hersbach et al., 2020). This implies, for instance, that larger perturbations are applied for cloudy areas compared to clear-sky conditions for all-sky microwave radiances (Geer and Bauer, 2011; Geer et al., 2018). The high coverage and accuracy have made satellite instruments the most important source of observation nowadays (80% of the observational influence according to Cardinali (2013)). Except for TEMP and SYNOP observations, uncertainty in all humidity observations are modeled as a function of temperature, pressure and humidity itself, also implying state-dependence (Ecmwf, 2016a). The variances are computed from instrument noise plus representativeness error, modified by the observation operator and routinely tuned (Hollingsworth and Lönnberg, 1984; Desroziers et al., 2005). Sea surface temperature and sea ice concentration perturbations are drawn according to Hirahara et al. (2016), but are of minor importance given our focus on the mid-troposphere.

Spread increase due to larger observation perturbations are partially compensated by lower observational weight in the gain. We briefly discuss this issue in Appendix B, noting that this effect is predominantly of second order for operational cases.

Beyond the magnitude of the individual perturbations, EDA spread is also shaped by the effective observational constraint exerted by the network as a whole. Adding observations increases the dimension of $\mathbf{R}$ effectively adding positive definite terms to $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$, whereby the expected analysis ensemble variance decreases. This has a dominant effect on ensemble spread since the number of observations ingested in the production of ERA5 has grown by more than 3 orders of magnitude between 1940 and today (Soci et al., 2024).

Observational coverage also varies on shorter time scales, e.g. for socioeconomic or geopolitical reasons, but more importantly, severe weather may also degrade observational coverage: storms, systems of intense organized convection and regions of strong turbulence are typically avoided by aircraft (Moninger et al., 2003) and routinely undersampled given the important small scale structures defining them (Janjić et al., 2018; Emanuel, 2021; Ko et al., 2025).

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

Finally, through quality control, the effective amount of observations is reduced by aerosol or cloud presence in satellite instruments, in addition to the elevated observation uncertainty associated with these conditions. Inter-channel and spatial error correlations are also larger in cloudy conditions (Bormann et al., 2011), reducing the effective information content of each radiance and thereby the overall constraint.

3.1.2 Error growth and physics perturbations

The short-range background forecast for each assimilation window is obtained by numerical integration using the previous window's analysis three hours before the start of the window as an initial condition, so the total length of the background forecast run is 15 hours. Note that the background term in Eq. (1) is only evaluated at the assimilation window's outset, while the short range forecast is used as an initial guess in the misfit term throughout the full 12-hour width. Conceiving the ensemble of previous analyses provided by the different members as a probability distribution, the integration will cause the distribution to shift and deform in model space in a fashion governed by the dynamical model's short time-scale error-growth characteristics.

There is a wealth of literature studying the processes responsible for error growth in atmospheric models at different lead times (Krishnamurthy, 2019). Error growth is flow-dependent on short (Puh et al., 2024; Rodwell and Wernli, 2023), medium (Baumgart and Riemer, 2019; Sánchez et al., 2020) and long (Spaeth et al., 2024; Osman et al., 2023) time scales, and it also depends on the characteristics of the initial perturbations themselves (Raynaud and Bouttier, 2016; Wang et al., 2023). In the present setting, the initial perturbations are the differences between the previous window's analyses across ensemble members, which themselves are the product of the preceding EDA cycle. The resulting spread in the background forecast therefore reflects a mixture of error growth during the 15-hour integration and uncertainty inherited from the previous analyses, two contributions that are difficult to separate.

As outlined in the introduction, the dominant role of moist convection in upscale error propagation (Selz and Craig, 2015; Selz, 2019; Judt, 2018; Bechtold et al., 2014; Zhang et al., 2007) – a foundational result for short-range predictability – has been established primarily for grid-point-level initial perturbations. At operational perturbation magnitudes such as those of the EDA, error growth is instead concentrated at larger scales and driven by adiabatic, baroclinic processes, in particular near tropopause (Selz et al., 2022; Rodwell and Wernli, 2023; Baumgart et al., 2019). Convective regions nonetheless tend to show large EDA spread (Žagar, 2017), which is likely inherited from elevated initial uncertainty as well as produced de novo by convective error growth during the short background integration itself (Cardinali et al., 2014). While disentangling these contributions within any single analysis cycle is not possible, the long time span of ERA5 with a fixed IFS configuration does provide leverage: in its early decades, with few observations to constrain the analysis, the assimilation system runs quite freely, so flow-dependent spread patterns then more prominently reflect model-internal mechanisms compared to more constrained assimilation nowadays.

A long history of parametrization development (Christensen and Zanna, 2022) has shown that ensemble forecasts suffer from incorrect dispersion and biases when model uncertainty in unresolved processes is not accounted for. This motivated the development of stochastic physics schemes (Palmer, 2019b), of which two are used in the production of ERA5: the stochas-

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

tically perturbed parametrization tendencies (SPPT) scheme (Buizza et al., 1999; Lock et al., 2019; Leutbecher et al., 2017) represents uncertainty in physical – especially diabatic – tendencies via multiplicative perturbations, while the stochastic kinetic energy backscatter (SKEB) scheme (Shutts, 2005; Berner et al., 2009) accounts via additive perturbations for the missing upscale energy transfer caused by spectral truncation.

Both schemes have been found to alleviate biases (Berner et al., 2012, 2017; Lang et al., 2021b) and to increase EDA spread throughout the atmosphere, most effectively above the planetary boundary layer (Palmer et al., 2009; Isaksen et al., 2010), with SPPT typically dominating. Correctly representing model uncertainty in assimilation – judging by proper EDA spread – avoids overconfidence in the background forecast and, by extension, incorrect deprioritization of observations (cf. Eq. (1)). Crucially, the stochastic perturbations are sampled independently for each ensemble member, so that even members supplied with identical observation perturbations and identical initial conditions would diverge through the iterative cycling. Note that ECMWF has since deactivated SKEB (Ecmwf, 2018) and replaced SPPT with the Stochastically Perturbed Parametrizations (SPP) scheme (Ollinaho et al., 2017; Lang et al., 2021b; Ecmwf, 2024), but these changes post-date ERA5.

The interplay between stochastic physics and the model's representation of large-scale flow regimes – particularly atmospheric blocking and the associated WCBs – has received considerable attention (Christensen et al., 2015; Dawson and Palmer, 2015; Filippucci et al., 2024; Davini et al., 2021; Pickl et al., 2022; Wandel et al., 2024). While these influences typically unfold on longer time scales than those relevant for the EDA, they motivate the regime- and WCB-conditioned analyses we present below. In the absence of any stochastic perturbations, members initialized identically and supplied with identical observations would still diverge slightly through the amplification of numerical errors by turbulent cascades (Leith, 1971; Leutbecher and Palmer, 2008; Lang et al., 2021a), an effect much smaller in magnitude than any of the perturbations discussed above.

4 Methods

4.1 Time Series

To isolate the influence of variations in observational coverage and quality on long time scales and later study the dynamics of the short time scale variability in assimilation uncertainty, we employ statistical models for time series data. Such models are well-established (e.g. Deistler and Scherrer (2022)) and appropriate for the problem at hand given the substantial autocorrelation. In the spirit of Occam's razor, we aim to use statistical models that are as simple as possible and as complex as necessary (Box, 1979). As will be proven in Sect. 5 the time series data exhibits heteroskedasticity and autocorrelation as well as breakpoints (see Sect. 4.2) that divide the time series into distinct segments. We have therefore chosen to adopt a generalized least squares (GLS) framework in the following, most general, form.

For a time series divided into N segments by $N-1$ breakpoints, let $y_{n,t}$ denote the response variable in segment $n \in \{1,\ldots,N\}$ at timestep $t \in \{1,\ldots,T\}$. We fit

$$y_{n,t} = \beta_n^0 + \sum_{m=1}^{M} \beta_n^m x_t^m + \varepsilon_{n,t}, \tag{6}$$

subject to

$$\mathrm{Cor}\left(\varepsilon_{n_1,t_1},\varepsilon_{n_2,t_2}\right)=\rho^{|t_1-t_2|},\ \text{if } n_1=n_2 \tag{7}$$

$$\mathrm{Var}\left(\varepsilon_{n,t}\right)=\sigma_n^2. \tag{8}$$

This amounts to a marginal GLS model with an arbitrary set of covariates x_t^m, segment specific regression coefficients β_n^m, AR(1) structure in the residuals with an autocorrelation coefficient ρ pooled across segments, but with segment specific error variances σ_n^2. We pool ρ because the segments will be defined from area-averaged data rather than the individual grid-point time series; estimating a separate autocorrelation per segment per grid point would be noisy and potentially unstable and would, moreover, assume the weather system's intrinsic uncertainty evolution differs by segment.

The model formulation 7 implies we assume the autocorrelation is present in the error structure. One can drop the AR(1) structure in the errors by assuming the autocorrelation is caused by the response variable by introducing lagged variables as covariates (e.g. $y_{n,t-1}$) or by assuming one of the covariates is the cause for the autocorrelation. Apart its differences in physical logic, this distinction has consequences for the fitting procedure. Under 7, GLS estimation prewhitens both response and design. Any covariate whose own autocorrelation is comparable to ρ has most of its variance removed by this transformation, and the estimate of ρ then competes with such a covariate for the same low-frequency variance in y, producing weakly identified or singular fits. For one set of models in this study (Sect. 5.3.3) this is the case; for these we drop 7 and fit 6 by ordinary least squares with heteroskedasticity- and autocorrelation-consistent (HAC) standard errors (Sect. 4.3), at the cost of some efficiency relative to a correctly specified GLS.

Segment-specific variances are retained to reflect genuine differences in variability between segments, which is more important for parameter and uncertainty estimates as well as diagnostics than locally correct autocorrelation coefficients (Hamilton, 1994). It is also logical to assume the variance of the data depends on the instrument suite and, therefore, the segment rather than being an intrinsic property and, thus, constant across segments.

We note in passing that we could have also split the time series into N segments and fit N separated models, which would have been more granular, more complex, but probably less likely to exhibit convergence issues. As we rarely experienced convergence issues anyway, we have chosen the above approach since it enables joint estimation of parameters within a single likelihood, ensuring coherent inference and streamlined implementation. All models were fitted by maximum likelihood for GLS using the `nlme` package in `R` (Pinheiro and Bates, 2000) version 3.1-162 (Pinheiro et al., 2025).

For the seasonal cycle, we employ K pairs of harmonic covariates,

$$x_t^{\sin,k}=\sin\left(2k\pi\frac{\mathrm{doy}_t}{365}\right), \tag{9}$$

$$x_t^{\cos,k}=\cos\left(2k\pi\frac{\mathrm{doy}_t}{365}\right), \tag{10}$$

for $k = 1, \ldots, K$, where $\mathrm{doy}_t \in \{1, \ldots, 366\}$ is the day of year (this is exchanged for the month of the year for monthly data). Together these contribute $2K$ terms to eq. 6; we refer to their fitted sum as the estimated seasonal cycle

$$\hat{s}_{n,t} = \sum_{k=1}^{K} \left(\hat{\beta}_n^{\sin,k} x_t^{\sin,k} + \hat{\beta}_n^{\cos,k} x_t^{\cos,k} \right). \tag{11}$$

More harmonics add more degrees of freedom and therefore more detail. For reasons that will become clear upon presentation of the results (cf. Sect. 5.2) we use three harmonics associated with periods of 12, 6 and 4 months ($K = 3$).

This way of parameterizing a seasonal dependence is almost always preferred over directly estimating phases and amplitudes since the coefficients enter the model linearly, the associated least-squares problem is convex and has a unique global solution and standard inference applies (Bloomfield, 2000).

Given fitted parameters $\hat{\beta}_n^m$, we define the fitted values $\hat{y}_{n,t}$ and the empirical residuals $\hat{\varepsilon}_{n,t}$ by

$$\hat{y}_{n,t} = \hat{\beta}_n^0 + \sum_{m=1}^{M} \hat{\beta}_n^m x_t^m, \qquad \hat{\varepsilon}_{n,t} = y_{n,t} - \hat{y}_{n,t}. \tag{12}$$

To assess the effect of categorical covariates, we use estimated marginal means (EMMs): model-implied means for each category after averaging over the remaining model terms. Consider the covariates x_t^m and partition them as $x_t = (x_t^{(c)}, x_t^{(-c)})$, where $x_t^{(c)}$ is categorical taking values in a set of levels $\mathcal{L} = \{1, \ldots, L\}$ and $x_t^{(-c)}$ collects the remaining covariates; then the EMM at level $l \in \mathcal{L}$ is defined as

$$\hat{\mu}_n^c(l) = \mathbb{E}_{x^{(-c)}} \left[\hat{\beta}_n^0 + \sum_{m=1}^{M} \hat{\beta}_n^m x_t^m \,\middle|\, x_t^{(c)} = l \right], \tag{13}$$

where $\mathbb{E}_{x^{(-c)}}[\cdot]$ denotes expectation with respect to the distribution of the remaining covariates $x_t^{(-c)}$ that is replaced with the sample average over observed values of $x_t^{(-c)}$ in practice. We will drop the subscript n when referring to the total time series EMMs as $\hat{\mu}^c(l)$. EMMs answer the question of what the expected response is for each category after adjustment for the other terms in the model and may be viewed as adjusted composites: unlike raw composite means, they account for additional covariates and inherit the assumptions of the fitted regression, including the functional form and covariance structure (Searle et al., 1980). We use the `emmeans` package in R (Lenth and Piaskowski, 2026).

To measure the variability between the categories l, the standard deviation between EMMs $\hat{\mu}^c(l)$ across the $|\mathcal{L}|$ different levels can be calculated as

$$\sigma_c = \sqrt{\frac{1}{|\mathcal{L}|} \sum_{l \in \mathcal{L}} \left(\hat{\mu}^c(l) - \bar{\hat{\mu}}^c \right)^2}, \tag{14}$$

with $\bar{\hat{\mu}}^c = |\mathcal{L}|^{-1} \sum_{l \in \mathcal{L}} \hat{\mu}^c(l)$. This measure can be compared between different model variants that feature the categorical covariate to assess the influence of other covariates.

4.2 Breakpoints

Breakpoints (or change points, structural breaks) in time series are points at which the underlying data-generating process undergoes an abrupt change in some or all of its statistical properties. For a comprehensive introduction to the theory and applications of detecting and modeling such breaks, see Chen and Gupta (2012) and the review by Truong et al. (2020).

The topic finds application in this paper, since the uncertainty of atmospheric measurements and hence the uncertainty of reanalysis data exhibits such breakpoints, for instance, due to the introduction of satellites. In the present setting, we want to objectively identify an unknown number of breakpoints in a time series given assumptions about the nature of the underlying statistical process – these assumptions are given above in Sect. 4.1. We do this by finding the number and location of breakpoints that minimizes the Bayesian Information Criterion (BIC), which is a measure for model fit balanced by the number of parameters used to prevent overfitting and is consistent in finding the true model under regularity as $T \to \infty$ (Schwarz, 1978).

The minimization uses a dynamic programming algorithm to find the global optimum across all possible segmentations given either a maximum number of breakpoints or a minimum segment length. Algorithmic details are given by Zeileis et al. (2003) and implemented in the `strucchange` package version 1.5-4 in `R` (Zeileis et al., 2002). The theory on this goes back to Bai (1994) with extension to simultaneous estimation of multiple change points by Bai (1997); Bai and Perron (2003).

4.3 Model Evaluation and Inference

Given the complexity of the dataset and the statistical models, distinguishing signals from noise warrants careful treatment. The aim here is always to balance rigor and interpretability with computational effort. These details are not essential for interpreting the main results.

Parameter estimates $\hat{\beta}_n^m$ are attached with uncertainties out of the box by the `nlme` package's fitting algorithm and significance is assessed by a two-sided t-test (Pinheiro and Bates, 2000). For the estimated seasonal cycle $\hat{s}_{n,t}$ (cf. Eq. 11), we visualize parameter uncertainty jointly in simultaneous confidence intervals for the whole year within which the true seasonal cycle lies with probability $1-\alpha$. These bands are computed using the Scheffé method for linear combinations of regression coefficients, based on the estimated covariance matrix of the harmonic parameters. This approach accounts for the joint uncertainty of all seasonal coefficients and ensures coverage of the entire seasonal curve rather than individual time points (Scheffé, 1953; Dickhaus, 2014).

Because each estimated marginal mean is a linear combination of the fitted regression coefficients, its uncertainty follows directly from the estimated coefficient covariance, yielding Wald confidence intervals and two-sided significance tests (Searle, 1987). We compute these with the `emmeans` package in R (Lenth and Piaskowski, 2026).

When a model is fitted by ordinary least squares rather than GLS, we replace the coefficient covariance with a heteroskedasticity- and autocorrelation-consistent (HAC, or "sandwich") estimator (White, 1980; Newey and West, 1987, 1994) with automatic bandwidth selection, computed via the `sandwich` package (Zeileis, 2006; Zeileis et al., 2020). Nothing else in the EMM calculation changes.

We often want to know how an EMM shifts between two model variants A and B fitted to the same response – for instance, the same model with and without a particular block of covariates. For each level we simply take the difference of the two EMMs. Because both coefficient sets are estimated from the same data, the covariance between them is available analytically (from the design matrices and the working covariance, or its HAC counterpart), so the uncertainty of the difference follows by the same linear propagation as for a single EMM (Goldstein and Healy, 1995).

The quantity σ_c introduced in Eq. 14 summarises how strongly the EMMs vary across levels. It is zero exactly when all EMMs coincide, so asking whether σ_c differs from zero is the same as asking whether the levels differ at all; we assess this with a joint Wald test.

Comparing σ_c between two model variants is less direct, because σ_c is a nonlinear function of the EMMs. We therefore proceed by simulation. Using the joint covariance of the two coefficient sets (again obtained analytically, as above), we draw a large sample (here 10,000 realizations) of the EMMs from their joint normal distribution, recompute σ_c for each variant on every draw, and examine the resulting distribution of the difference. Its central 95% gives a confidence interval, and the proportion of draws falling on the far side of zero serves as a two-sided significance test (Davison and Hinkley, 1997).

Since we perform a large number of identical statistical tests at different grid points, it is reasonable to consider the proportion of false positives across all significance tests rather than the probability of Type 1 errors (false rejections of the null hypothesis) individually for every test (Wilks, 2019, ch. 5). Among the many ways of doing this (Dickhaus, 2014; Benjamini, 2010), we apply the classical false discovery rate (FDR) correction based on Benjamini and Hochberg (1995), focusing on a balance of simplicity, power and conservativeness. The method implies we are always at least as conservative in rejecting the null compared to individual testing. We note that the method is robust under positive dependence between tests (like spatial correlation between neighboring grid points; Benjamini and Yekutieli (2001)). The algorithm is implemented in the core package `stats` in `R` (R Core Team, 2022).

Comparison between model variants can be naturally drawn from respective BIC values and likelihood ratio tests. Models are also investigated based on their goodness-of-fit by reporting the classical adjusted coefficient of determination R^2_{adj} together with its decomposition. R^2_{adj} measures the fraction of response-level variance described by the model after accounting for model dimensionality (Hittner, 2020). Then the total empirical variation in response space and its part not accounted for by the model is given by

$$\mathrm{TSS} = \sum_{t=1}^{T} (y_{n,t} - \bar{y})^2, \qquad \mathrm{RSS} = \sum_{t=1}^{T} \hat{\varepsilon}^2_{n,t}, \tag{15}$$

with $\bar{y} = T^{-1}\sum_{t=1}^{T} y_{n,t}$, the empirical mean. With the number of parameters p estimated in the model the associated mean-square (variance) components,

$$V_{\text{res}} = \frac{\mathrm{RSS}}{T-p}, \qquad V_{\text{total}} = \frac{\mathrm{TSS}}{T-1}, \tag{16}$$

represent the data-based counterparts to the error variance for the whole time series σ^2, which is an average of the segment-wise assumed error variances σ^2_n with weights according to the relative share of observations. The adjusted coefficient of determination is given by

$$R^2_{\text{adj}} = 1 - \frac{V_{\text{res}}}{V_{\text{total}}}, \tag{17}$$

which measures the fraction of empirical response-level variability in $y_{n,t}$ accounted for by the deterministic component of the model, while the covariance structure 7 governs the assumed second-order properties of the underlying error process $(\varepsilon_{n,t})$. We decide to provide these scale-consistent and degrees-of-freedom adjusted quantities rather than their unadjusted counterparts for a directly interpretable assessment of model fit.

Segments with high variance dominate the above response-level metrics. Consistent with our assumption of heteroskedasticity, we thus also compute its generalized-least-squares analogue,

$$R^{2,\text{GLS}}_{\text{adj}} = 1 - \frac{\text{GRSS}/(T-p)}{\text{GTSS}/(T-1)}, \tag{18}$$

where GRSS and GTSS are the residual and total sums of squares evaluated in whitened space, i.e. after transforming the data by the inverse square-root of the estimated error-covariance matrix, so that the transformed errors are homoskedastic and uncorrelated.

Uncertainty quantification in the estimation of breakpoints can be obtained from the limit distribution of the break point estimator derived under general assumptions by Bai (1997); Bai and Perron (2003). The framework notably also allows for autocorrelation and heteroskedasticity in the errors under mild assumptions. Neither of the two generalizations modify the point estimates for the breakpoints, but do affect the estimator's asymptotic variance and, hence, the confidence intervals around the breakpoints. In practice, we employ the same HAC “sandwich" estimator from above (White, 1980) for the covariance matrix introduced by Newey and West (1987, 1994) and implemented in the `sandwich` package in R (Zeileis, 2006; Zeileis et al., 2020).

5 Results

Our variable of interest throughout will be the logarithmic ensemble spread (standard deviation) of geopotential on 500 hPa (zg500) between the EDA members for the whole time period, which we denote by $\log(\sigma_{\text{EDA}})$. We apply a logarithm transform because the data varies on an exponential scale as expected from considerations in Sect. 3. The logarithm of the ensemble spread also allows us to assume Gaussian statistics, which would be an incorrect assumption for the standard deviation directly due to its non-negativity. The response variable is then in uncommon units of $\log(m^2/s^2)$, but that presents no loss, since the ensemble spread should not be taken at face value, anyways. We use zg500 as the classic variable for synoptic variability. Moreover, it is reasonable to assume zg500 is an integrative variable in the sense that it does not depend exclusively on a specific kind of measurement, but rather is influenced by satellite instruments, air-bourne and ground-based measurements alike. This is desirable because it prevents local small scale influences that would occur e.g., upon analyzing surface variables.

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

5.1 Breakpoints

We begin by applying the breakpoint identification algorithm introduced in Sect. 4.2 to the monthly and latitude-weighted spatially averaged time series of $\log(\sigma_{\text{EDA}})$. Our goal is to find common break points in time to be applied to the statistical models at each grid point. The assumption here is that abrupt changes in the observation or assimilation system ($\mathbf{B}_{\text{cli}}$) affect the whole region roughly at the same point in time, though possibly with a varying magnitude, so breakpoint identification at each grid point individually would be unnecessarily complex. We average monthly to be consistent with the spatial coarse-graining.

The statistical model assumptions underlying the breakpoint detection are according to Sect. 4.1 with covariates being the seasonal cycle represented by three harmonics with periods 12, 6 and 4 months and the year ($x_t^{\text{year}} \in \{1940, \ldots, 2024\}$). Using standard notation for statistical models (Pinheiro and Bates, 2000), this reads

$$y \sim x^{\text{year}} + \sum_{k=1}^{3} \left(x^{\sin,k} + x^{\cos,k} \right). \tag{19}$$

A minimal BIC value is achieved for a five breakpoint segmentation shown in fig. 1 applying a minimum segment length of 10% of the time series which equates to a length of 101 months or eight years and five months. We provide minimal RSS and BIC values for different numbers of breakpoints in the Appendix D, Fig. D1. Fig. 1 also presents the according fit to the “observed" time series along with 95% confidence intervals for both the fit and the breakpoints[3]. As mentioned in Sect. 4.3, the confidence intervals for the breakpoints are estimated with a HAC estimator, while the model fit implicitly accounts for autocorrelation in the error structure. From a meteorological stand point we think of the autocorrelation as a feature of the atmosphere itself instead of assuming it is caused by the instrument and assimilation suite.

As expected, the ensemble spread has dropped tremendously by more than three orders of magnitude across the time span under investigation. The drop is particularly strong in the 1940s and 1950s, which is expected in light of the sparsity and inhomogeneity of the observation system prior to the international geophysical year (IGY) 1957/1958 (Odishaw, 1959). The data also follows the fit more regularly after this period, especially with respect to the visible seasonal cycle. In the following we give a brief account of the possible explanations for the different breakpoints and the behaviour of the time series in the segments between them.

1. **1948–09.** The first segment is characterized by considerable irregularities alongside an order-of-magnitude reduction in spread, likely resulting from increased coverage and quality of radiosonde, aircraft and surface-based measurements, especially following the end of World War 2. The steep decline just before the first breakpoint in September 1948 is perhaps attributable to the establishment of “weather ships” by the International Civil Aviation Organization (ICAO), founded in 1947 (Downes, 1977).

2. **1957–08.** The second segment exhibits a slightly flatter decline and smaller noise as coverage and quality of observations increased. This is especially true for the years leading up to the IGY in 1957/1958, which involved expansion, standard-

---

[3] To avoid confusion between “observations" in the assimilation sense and “observations" in the statistical sense, we will refer to the ensemble spread at individual points in time as “data points" or “assimilated values".

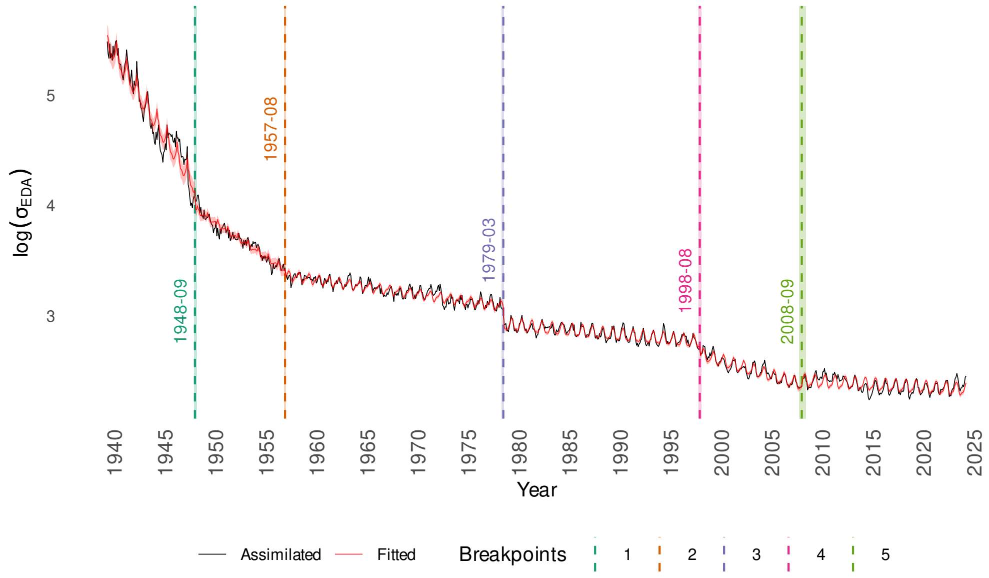

Figure 1. Monthly and spatially averaged logarithmic EDA ensemble spread of geopotential at 500 hPa assimilated (black line) and fitted with 95% confidence intervals (red line and shading). Identified change points annotated with the respective months (vertical dashed lines). Shading around change points indicate 95% confidence intervals.

ization and international coordination of geophysical measurements (Joselyn and Ismail-Zadeh, 2019) and to which the breakpoints location is probably related.

3. **1979–03.** The third segment extends up to the massive drop in ensemble spread in 1979, caused by the introduction of satellite instruments. A slightly smaller drop is appreciable around 1972 when the first satellite instrument (NOAA-2) started service (Bell et al., 2021). Notably, from 1972 to 1978 a new $\mathbf{B}_{\text{cli}}$ was also used (cf. Sect. 3).

4. **1998–08.** Changes across this breakpoint roughly coincide with the advent of new satellite instruments, particularly AMSU (Duncan et al., 2022) and ATOVS (English et al., 2000) microwave sounders, atmospheric motion vectors (Bormann et al., 2014), and later radio occultation instruments (Anthes et al., 2008). The absence of an abrupt jump but subsequent stronger decrease reflects the continuing establishment of satellite instruments (Hersbach et al., 2020). A new $\mathbf{B}_{\text{cli}}$ was used starting 2000. A possible additional influence is the establishment of the WMO Aircraft Meteorological Data Relay (AMDAR) Panel in 1998 (World Meteorological Organization, 2003), though no sharp changes are visible in upper-air wind forecast quality around that time (Petersen, 2016).

5. **2008–09.** A peculiar periodic increase in uncertainty is discernable until roughly 2015, which we assume results from an erroneous configuration of α during ERA5 production (cf. Sect. 3). This is notably unrelated to the ERA5.1 production changes, which affected only the stratosphere and data from 2000 to 2006 (Simmons et al., 2020). The comparatively larger confidence interval reflects the relatively small change in statistical properties across this breakpoint. Nevertheless,

we retain it to adhere to objective (BIC-based) segmentation, noting that the resulting statistical model is far from overfit (48 parameters on 1,020 degrees of freedom)–much less so on the daily timescale.

5.1.1 Seasonality

We consider the seasonal cycle in the assimilation uncertainty a subject of relevance, so we discuss briefly its precise nature and possible explanation in the following before filtering it out for the ensuing analyses. The pronounced seasonality in the time series becomes visibly clearer with increasing time. This is also evident from the statistical model fitted to the data and presented in fig. 1. The 36 seasonal parameters can be grouped into 18 pairs of the same segment n and wavenumber k. Of the 18 pairs there are four pairs for which neither of the two parameters are significant; those for wavenumber 3 in segments 1, 3 and 5 and the pair for wavenumber 1 in segment 2. There are multiple conclusions to be drawn from this.

Firstly, it is fair to say that there is a significant seasonality associated with the time series across all segments, even though not all coefficients are significant (we keep all terms in the ensuing models for consistency and comparability). This makes it quite certain that the seasonality is a property of the system itself (at least on a spatially averaged level) rather than a property introduced by some instrument, nevertheless we filter it out because we want to assess model uncertainty independent of season in the spirit of the year-round WRs of Grams et al. (2017). The seasonality itself is presumably caused largely by the increased zg500 gradients prevalent during the boreal cold season, which would effectually distort uncertainty differences between regimes once considering their impact year-round – especially so because the regimes defined occur in very different frequency across seasons and, truly, months (cf. supplementary figure S2 in Grams et al. (2017)).

Secondly, judging from the fact that wavenumber three terms exhibit the lowest significance, the ensemble spread tends to vary rather on a (semi-)annual timescale, consistent with the above assumption. This can be recognized in fig. 2, too, where the estimates of the seasonal cycles in terms of $\hat{s}_{n,t}$ (cf. Eq. 11) are displayed for every segment as well as for the whole time series. The estimate for the whole time series is obtained from a separate model with breakpoints consistently estimated without assuming the seasonal cycle as a covariate. The corresponding fit is provided in the Supplementary Fig. S2.

The confidence intervals get smaller with time, which is expected from fig. 1. The best estimates converge to the estimate obtained from fitting a combined seasonal cycle for the whole time series, which acquires its maximum in January and its minimum in May. The curves prove that the seasonal dependence cannot satisfyingly be modelled by a pure sine wave, which necessitates using multiple harmonics. It is a well-known fact that synoptic activity is generally strongest in winter due to the meridional temperature gradient (e.g. (Holton and Hakim, 2013, ch. 6,7), Wettstein and Wallace (2010)), so a peak somewhere within DJF was expected. The gradual increase during summer towards autumn is probably related to synoptic activity in the form of storms and organized deep convection considering the potential causes for ensemble spread (cf. Sect. 3.1), but what exactly governs the spread is hard to pinpoint from this perspective. It will be subject of later sections.

Although the per-segment seasonal cycles converge towards the whole-period estimate, suggesting one could simplify to a common cycle, the variations in amplitude and phase between segments are significant—more so on daily time scales for individual grid points. Crucially, these varying amplitudes are tied to the heteroskedastic nature of the time series; it would therefore be inconsistent to hold the seasonal coefficients constant across breakpoints while allowing residual variance and all

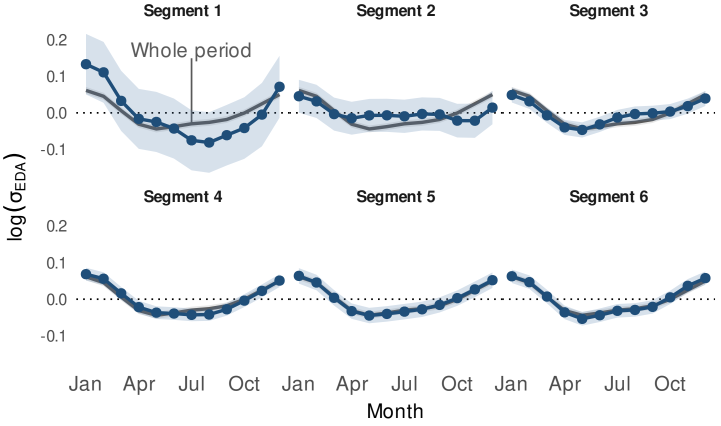

Figure 2. Statistical model estimates of the seasonal cycle $\hat{s}_{n,t}$ for every segment individually (blue curves with circles) and for the whole time series combined (gray curve in the background). The seasonal cycle for the whole time series is estimated based on a separate model using break points obtained assuming no seasonal cycle. 95% confidence intervals are for the whole seasonal cycle combined rather than for individual months (cf. Sect. 4.3).

other covariates to change. As a side note, the number and location of breakpoints (Sect. 5.1) depend only mildly on whether a seasonal cycle is assumed; estimates without one are given in the Supplementary Fig. S1.

5.2 Model Evaluation

With the breakpoints identified, we now estimate separate time series models for each grid point and using data on daily time 545 scales. For a first glimpse into the results, we report the adjusted coefficient of determination R^2_{adj} across the investigated region together with its decomposition as introduced in Sect. 4.3. Since the number of data points T and parameters p is identical across grid points ($T = 31{,}047$ and $p = 48$), dividing by the respective degrees of freedom amounts to a constant rescaling that does not alter spatial patterns, while yielding quantities that are directly interpretable as variance components. Figure 3 provides R^2_{adj} along with its decomposition.

550 The models generally cover a substantial amount of response-level variation across the domain with values of R^2_{adj} ranging from roughly .56 to .88 – a level of spatial diversity expected for our rather simple modelling assumptions. The highest values are achieved in the polar region (upward triangle), where the total variance is also largest as can be seen from the right panel in figure 3. The high total variance is of due to the historically sparse observational coverage of the region and the entailing wide range of values across the time span considered. This argument is backed up by the fact that the total reduction in uncertainty 555 has a similar spatial pattern (cf. Fig. D2 in App. D). Along our assumptions on the structure of the dataset, we argue that the high values of R^2_{adj} in the high latitudes imply that most of the variation here is due to the development of the observation and assimilation system, though in absolute terms, an appreciable amount of variation is still present in the residuals as can be seen in the center of Fig. 3.

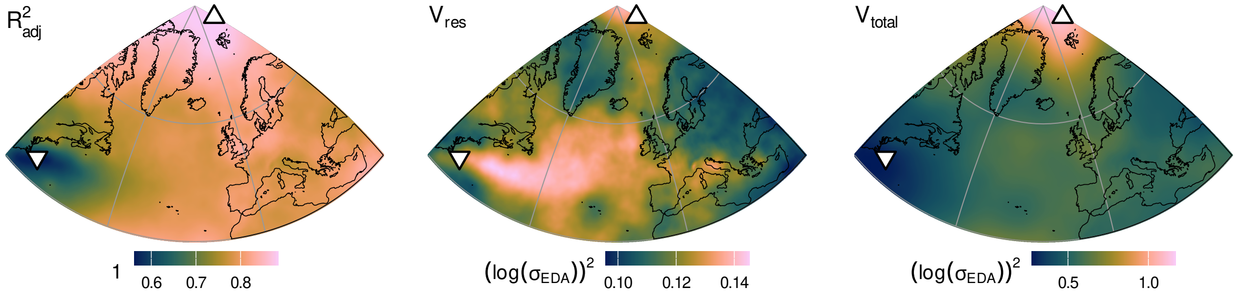

Figure 3. Grid-point-wise model evaluation. Left: Adjusted coefficient of determination in response variable space R^2_{adj}. Center: Response-level variance not accounted for by each model per degree of freedom V_{res}. Right: Total response-level variance per degree of freedom V_{total}. Indicated with an upward (downward) triangle is the location of the maximum (minimum) R^2_{adj} in the domain.

On the other end of the spectrum, the lowest values of R^2_{adj} are evident over the Atlantic off the coast of New England (downward triangle). The decomposition makes it clear that this comes about both due to a below average total variance and an above average residual variance. The reasons for either are less obvious, but one part of the explanation is surely the pronounced land-sea contrast in the residual variance—a feature eminently visible around the Azores.

The presence of moisture over the oceans and the location of the mid-latitude storm track as a major source of synoptic activity surely plays a role in the large amount of residual variation over the oceans. We consider it safe to say, however, that the presence or absence of land-based measurements play a considerable role in this effect, too (for a proof, again, consider the Azores region). Given that the residuals of the models will be used in ensuing analyses, the spatial pattern in the model's residual variance opens an interesting question. Should the time series of residuals as proxies for short time scale assimilation uncertainty be standardized by the grid-point wise total variance in time? After all, it has just been argued that the land-sea contrast is overwhelmingly caused by the observation system rather than the system's intrinsic dynamics. We decide not to, since spatial variability in the total level of model uncertainty is expected and, truly, a central feature of the atmosphere. What's more, the WR specific uncertainties at the heart of this study are shaped and governed by the spatial configuration of the Euro-Atlantic region, and assuming unit variance in time across the region would severely contradict that assumption. Finally, the scale of the spatial heterogeneity is considerably smaller than the level of heterogeneity in the scale of the total response-level variation as can be judged from the color scales in the center and right panel in figure 3.

The patterns discussed above remain clearly visible when accounting for the difference in (temporal) error variances between the different segments. We provide the appropriate measure in App. D in fig. D3. Its spatial structure is qualitatively identical to that of R^2_{adj}, but the values are systematically lower. This has a simple explanation: in response space the early high-variance periods contribute most of the total sum of squares, so R^2_{adj} is dominated by the (good) fit there. Whitening rescales each segment's residuals by its variance, giving all segments comparable weight; the resulting drop therefore shows that the model

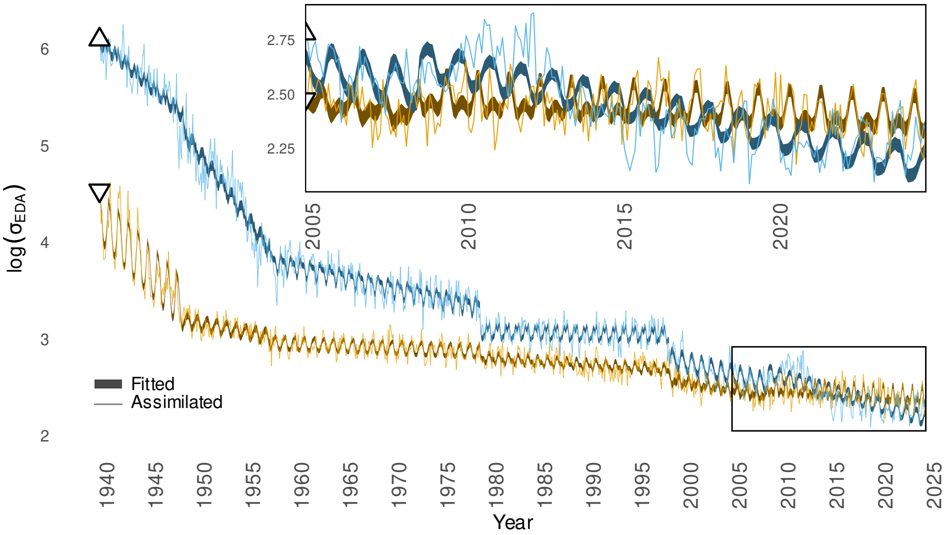

Figure 4. Monthly averaged logarithmic EDA ensemble spread of geopotential at 500 hPa assimilated (thin lines) and 95% confidence intervals for the mean estimates (shadings). The blue (orange) line and shading corresponds to the grid point indicated by an upward (downward) triangle in fig. 3. The inset details the data for the time slice from 2005 to 2024.

accounts for a larger fraction of the variance in the high-variance (earlier) segments than in the low-variance (later) ones. This is expected, since the elevated early variance largely reflects instrument and assimilation uncertainty, whose strong secular decline is well captured by the model (cf. Fig. 1). A decomposition analogous to Fig. 3 is meaningless, since whitening equalizes the residual variance across segments by construction.

5.2.1 Two example grid points

To shed some light on the diversity of the different models and on the behaviour of the time series on daily time scales, we present in detail the assimilated and fitted values at the grid points with the highest and lowest values of R^2_{adj}, indicated by the respective triangles in fig. 3. For a comparison with the spatially aggregated time series in figure 1, we first provide the monthly averages at the two grid points along with the confidence intervals for the estimated means appropriately propagated to the monthly level in figure 4.

The data at both locations exhibits the same general shape, validating our assumption of common covariance and breakpoints retrospectively (at least for these two grid points). Both time series have a seasonal cycle, an interannual dependence and show changes in statistical properties roughly at the breakpoints used, albeit all in differing magnitude. For both locations, most of the coefficients are statistically significant at the (FDR-corrected) 95% level except for the higher order seasonal cycle components. Note that significance is assessed against the null of zero coefficients, not against the hypothesis that coefficients change across breakpoints; we do not perform the latter model selection at every grid point, for simplicity and comparability.

Despite the drastically larger range of variability for the Arctic grid point ($V_{\text{total}} = 1.11\,(\log(\sigma_{\text{EDA}}))^2$ vs. to $.29\,(\log(\sigma_{\text{EDA}}))^2$), the residual variance is comparable: the Arctic uncertainty is approximately 1.5 orders of magnitude higher for the 1940s, but becomes virtually indistinguishable – even slightly lower – by recent years (inset, Fig. 4).

With a mature and mostly satellite-based observation system, the differences in the uncertainties at the two grid points become a marginal fraction of their absolute uncertainty. Nevertheless, the two time series do have distinct statistical properties as indicated by the different (statistically significant) model estimates for the mean. In particular, the data from the Arctic location declines on the time scale of years to decades whereas the data from the Gulf Stream location remains rather constant. In Sect. 5.1 we argued that the reason for the change in statistical properties identified by the breakpoint identification algorithm in 2008 is due to a misspecification of α during the production of ERA5. Consistent with this, we hypothesize that the Arctic grid point is more affected by this (the onset of nearby radar rain-rate measurements around this time appears not to affect the response; cf. Fig. 2 in Soci et al. (2024)).

The decline in the uncertainty at the Arctic grid point does make it seem like mean uncertainty levels are lower there compared to the Gulf Stream region for the assimilation system nowadays, which ties well with our understanding of synoptic activity as zg500 gradients are usually smaller in the Arctic compared to the mid-latitudes and the Gulf Stream region in particular. The influence of parametrized processes and satellite instrument peculiarities may also have a role in this. In the final two segments the two series show opposing seasonal cycles (Fig. D4, App. D): as in the spatial average, uncertainty peaks in winter at the Arctic point, but in summer at the Gulf Stream point.

In Sect. 5.1 we argued the seasonal cycle in the spatially averaged time series was due to larger zg500 gradients in winter compared to summer. For the two grid points presented, this explanation would certainly not hold for the Gulf Stream grid point for the last two segments. Generally, the seasonal cycles estimated by the models for the individual segments and grid points vary substantially, and daily, grid point wise data is way noisier and confounded by local idiosyncratic noise structures, so interpretation has to be taken with a grain of salt. Provided our hypothesis that the observation and assimilation suite's influence on the seasonality diminishes with later segments, one could argue that the uncertainty in the higher latitudes is influenced more by the variations in zg500 gradients whereas deep convection and other moist processes dominate intra-annual variability around the Gulf Stream grid point.

5.2.2 Spectral Analysis

The example grid points were discussed on monthly aggregated scale for simplicity and visual clarity, but variability in the residuals is large on daily time scales as well. In the supplement, we provide an analogue of figure 4 but on daily time scale for the most recent year 2024, which demonstrates that variability on daily timescales is considerable. A more quantitative view is provided in figure 5, where we present the power spectral density (PSD) of the original dataset together with the PSD of the residual time series for both grid points. The estimation of the spectra followed standard procedure with detrending, demeaning (meaningless for the residuals), application of a cosine taper of 5% and a moving-average smoother in frequency space with a span of 7 neighboring values (a “Daniell smoother"; cf. (Brockwell and Davis, 1991, ch. 10)). We use the standard implementation in the `stats` package (R Core Team, 2022).

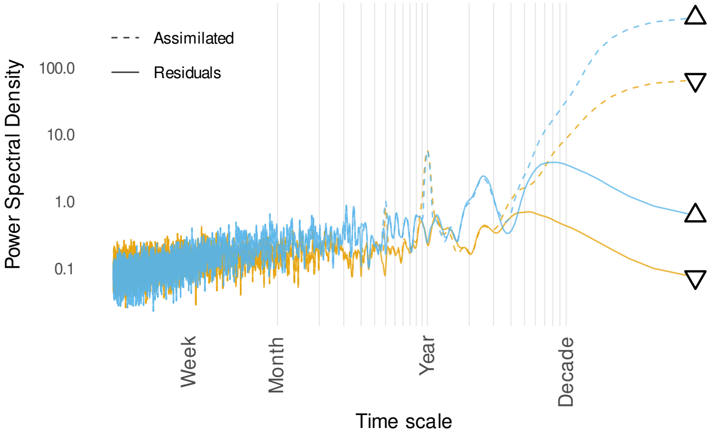

Figure 5. Power spectral density for the daily times series of raw values (dashed lines) as well as residuals (solid lines) at the two grid points indicated in figure 3. Colour and marker style identify the respective grid points. Shown on double logarithmic axes as a function of period (days); vertical reference lines indicate characteristic calendar scales. Spectra were estimated using a periodogram after linear detrending and demeaning, with a cosine taper of 5% applied to the time series and two subsequent Daniell smoothings with span 7 in the frequency domain (Brockwell and Davis, 1991, ch. 10). Spectra are truncated to periods $\geq$ 2 days.

As expected, the PSD peaks at the low-frequency end for both grid points. The models absorb a vast amount of this into the mean estimate consistent with their conception (cf. figure 4). The models also clearly erase local maxima at annual and semi-annual frequencies as well as power at a frequency of four months, though the original data features no prominent peak there. More generally, the models only really start absorbing spectral power at annual and longer scales. The spectra exhibit further peaks at roughly 2.5 years in both raw and residual time series matching the frequency of the Quasi-Biennial Oscillation (QBO; Baldwin et al. (2001)), but we refrain from going into further detail about this. The spectra also unveil that the higher total variability for the Arctic grid point is caused dominantly by time scales longer than a month while the power at smaller time scales is comparable between the two, the Gulf stream grid point possibly even dominating on sub-weekly time scales, though we do not rigorously test this hypothesis.

We extend the analysis above to all other grid points and provide the appropriate spectra in figure 6. In contrast to the above we apply no additional smoothing in the frequency domain since the averaging over the distribution of grid points already has a regularizing effect. Just like for the two example grid points, the average raw spectrum also unveils prominent peaks at half- and annual time scales with another minor peak at four months, retrospectively justifying the implementation of three harmonics in the modelling of the seasonality. More specifically, the two peaks at sub-annual periods indicate the seasonal cycles at a majority of grid points deviate from pure sine waves. Absorbing variability of the time series at higher frequencies risks erasing signals actually attributable to variability in WR occurrence and synoptic activity, but synoptic time scales are considerably shorter than four months and the seasonal cycles are estimated across a number of years, so we are confident that the applied model specification is appropriate for the goals of our analysis.

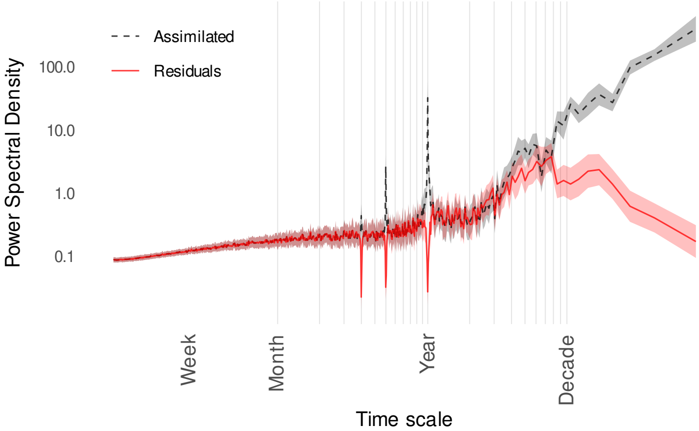

Figure 6. Mean (lines) and interquartile range (shading) of power spectral densities across grid points for the daily times series of raw values (black dashed lines) as well as residuals (red solid lines). Mean and quartiles are latitude-weighted. Shown on doubly logarithmic axes as a function of period (days); vertical reference lines indicate characteristic calendar scales. Spectra were estimated using a periodogram after linear detrending and demeaning, with a cosine taper of 5% applied to the time series (Brockwell and Davis, 1991, ch. 10). Spectra are truncated to periods ≥ 2 days.

Apart from the aforementioned peaks, the spectra exhibit clearly a non-constant (red noise) spectrum on intra-annual scales. This proves the existence of autocorrelation implied also in our model assumptions. In contrast to the spectra for the two example grid points, this average spectrum has no broad local maximum between two and three years, but, if anything, a local maximum around five years. This might be related to the the break point structure in the original dataset, but we refrain from in-depth interpretation.

5.3 Assimilation uncertainty on synoptic time scales

Having obtained a reasonable understanding of the estimated models and the underlying data, we go on by investigating the models' residual component climatologically. We interpret the residuals of our first-stage models as proxies for assimilation uncertainty on synoptic time scales, which is the relevant quantity with regard to ensemble forecasting. Since we assume the segment-specific heteroskedasticity is an artifact of the observation and assimilation suite, we weight the residuals by the respective estimated error variance in each segment, without changing the overall variance across the whole time series. Given the empirical residuals $\hat{\varepsilon}_{n,t}$ in segment n and at time step t according to Eq. 12 we call the variance-balanced residuals $\hat{\varepsilon}^*_{n,t}$:

$$\hat{\varepsilon}^*_{n,t} = \left(\frac{1}{T}\sum_{n=1}^{N} T_n \hat{\sigma}^2_n\right)\frac{\hat{\varepsilon}_{n,t}}{\hat{\sigma}^2_n}, \tag{20}$$

where T_n is the number of time steps in segment n and $\hat{\sigma}^2_n$ the empirical error variance in each segment. This means, we can basically “forget" about the different break points from now on apart from the characterization of the autocorrelation according

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

to Eq. 7. That being said, we checked most of the results presented in the following for consistency between segments and have found no noteworthy discrepancies (not shown).

We will use $\hat{\varepsilon}^*$ as a response variable in a set of second stage models, which aligns with our assumptions, but is not without alternative. In particular, fitting a separate second-stage model comes at the cost of not having fully joint likelihood-based estimation, so that uncertainty from the first stage is not propagated completely and the resulting estimates are somewhat less efficient than those from an integrated model. Its main advantages, however, are a substantially simpler and more stable estimation problem, straightforward implementation, and a clearer interpretation of the inferred effects as acting on the residual uncertainty component rather than on the original response. This makes the second-stage framework suitable for the present exploratory analysis, whose aim is not to exhaustively close the uncertainty budget, but to identify which variables and associated physical processes are most strongly connected with the residual uncertainty patterns.

As a first step, we examine whether the residual uncertainty is flow-state dependent by estimating the effects of the WRs introduced in Sect. 2. Figure 7 shows the estimated marginal mean effect (cf. Eq. 13) of each of the WRs, from a model with the WR time series as the sole covariate:

$$\hat{\varepsilon}^* \sim x^{\mathrm{WR}} \tag{21}$$

For comparison, two closely related diagnostics based on alternative assumptions are provided in App. D. Figure D5 are composite mean maps of $\hat{\varepsilon}^*$ according to Eq. A2, which represent the same flow-state dependence in a more descriptive, less model-based and therefore assumption-free form. Significance is assessed using the block permutation test described in App. A. Figure D6 presents regime effects derived directly from the full first-stage model on the original response scale (i.e. we added x^{WR} to the model in 19) and therefore answers a related, but not identical, question, namely how WRs project onto the total modeled quantity before isolating the residual uncertainty component. Importantly, all three representations yield very similar spatial patterns, demonstrating that the inferred flow-state dependence is robust across all three choices—descriptive versus model-based, residual versus response scale, and ordinary versus block-permutation significance. This supports our choice of second-stage models over full omnibus models, which in any case suffered convergence failures at some grid points—requiring relaxed assumptions and convergence criteria—whose causes we do not pursue here.

The flow states themselves are also shown as green contours and are derived as composite means of the target variable (zg500) in ER5 HRES for the whole period for which the regime classification is available (from 1950 to 2024). We remark that the patterns are not the cluster centroids that result from the `k-means` clustering in the original derivation of the regimes and are also not calculated from seasonality adjusted fields, but resemble the archetypal regimes very closely (Grams, 2026).

Figure 7 reveals distinct patterns for each of the WRs. At first glance, the patterns resemble the composites of the 10-day low-pass filtered geopotential height anomalies on 500 hPa associated with the archetypal regimes (Grams, 2026), but with opposing sign. This is somewhat surprising because the data we use ($\hat{\varepsilon}^*$) encodes not the flow patterns themselves, let alone their low-pass filtered anomalies, but rather the assimilation uncertainty on sub-annual timescales. It therefore seems like a non-trivial correlation exists between the deviation of the flow from the mean state (in the HRES analysis) and the uncertainty

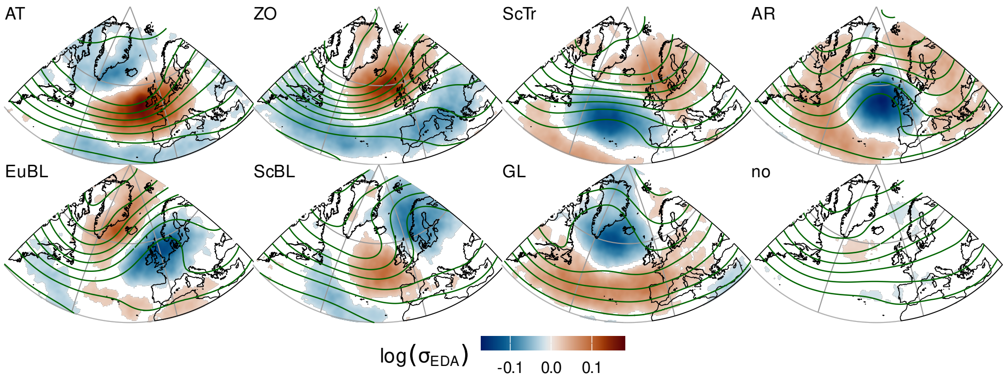

Figure 7. Effect of the occurrence of each WR on the synoptic time scale assimilation uncertainty. Estimated marginal means $\hat{\mu}^{\mathrm{WR}}$ for each grid point are obtained from second-stage models of the kind 21 (shading). Shown for grid points with FDR-adjusted p-values below 5%. Green contours display the composite mean z500 maps for each of the WRs based on the ERA5 HRES product. Contour lines are drawn every 80 gpm.

surrounding that deviation. The signed response adds further complexity to this connection. We test this hypothesis in the App. C and provide interpretations for the observed behaviour in that context.

From a mathematical standpoint, it is obvious that the local level of uncertainty in any field correlates strongly with the gradient and higher order derivatives of that field, since a small displacement of that field will entail a larger error compared to a field with a shallower gradient. This effect might contribute to some of the patterns, in particular for locations with highly zonal configuration and dense isohypses and for locations with anticyclonic patterns and weak gradients. Since there are locations contradicting that mechanism, there have to be other factors involved as well, though.

Anomalously low EMMs in the central Atlantic for the Scandinavian trough regime (ScTr) for example coincide with quite dense isohypses. A possible explanation may be the locally anticyclonic pattern apparently linked to lower EMMs in other WRs as well. Since anticyclones typically tend to suppress lifting and therefore cloud processes, the negatively anomalous patterns visible in ridges and blocked regions may also be caused by moist processes via stochastic physics or derived observation uncertainties on daily time scales. These hypotheses will be tested in the following section. We will use the WRs as a handle to understand the drivers of our derived response variable $\hat{\varepsilon}^*$. After screening possible explaining variables in an exploratory analysis, we will calculate quantitatively, which variables attenuate the WR signal, how strongly and where.

As a side note, the chosen statistical setup neither guarantees a zero mean for the shown WR patterns spatially (because the models are fitted for each grid point independently) nor across WRs (because their frequencies differ). Of interest might be the question, whether regimes have an (cosine latitude weighted) spatially averaged effect on overall uncertainty. Fitting an according model reveals rather weak signals with the “AR" regime exhibiting a significantly elevated EMM and the “ZO", “ScBL" and “no" regimes significantly reduced EMMs. Results for all regimes are provided in fig. D7 in App D.

5.3.1 Temporally varying correlations

As elaborated in Sect. 3.1, the spread in the EDA ensemble is caused by the scaled observation perturbations directly, other observational constraints and error growth in the background forecast including perturbations in the stochastic physics schemes. None of these aspects can be cleanly isolated from the signal visible in the residuals, especially because all three act principally via moist processes.

Firstly, the observation uncertainty ingested in the perturbations varies not only on the long time scales isolated by means of our first-stage models, but is also state-dependent. As discussed in Sect. 3.1.1, atmospheric moisture and the presence of hydrometeors in particular exert a state-dependent effect on the observation perturbations. Since zg500 is an integrative field and radiances and humidity measurements are important observation types for constraining the assimilation system, it is quite plausible, that an influence is detectable. Perturbations for humidity observations are not applied to SYNOP and TEMP measurements, but are among the measurements with the longest history (Soci et al., 2024). Together with the scene-dependent uncertainties from satellite instruments, this implies that the influence of observation perturbations on our response variable has to be varying in time. This is also true for aircraft based measurements since, their influence, too has increased over the period we consider.

For a preliminary screening of variables for later use in second stage models, we therefore investigate this presumed relation on interannual time scales. More specifically, we calculate the simple Pearson correlation coefficient r (Wilks, 2019, ch. 3) between time series of our response variable $\hat{\varepsilon}^*$ and several candidate variables on yearly subsets and show the resulting median and interquartile ranges (across grid points) in fig. 8. Calculation on yearly subsets implies sensitivities to interannual and decadal variability, so a certain degree of noise is certainly anticipated.

All variables chosen are either directly obtained or calculated from the ERA5 HRES product. The top row shows variables associated with moist processes exclusively while the bottom row consists of variables with possible associations for other reasons. The mid- and high-level cloud cover are chosen because their association with synoptic conditions is well-established and more direct compared to low-level clouds. Global circulation models such as the IFS still generally struggle with the representation of clouds – especially at low (Konsta et al., 2022; Yoo et al., 2013; Forbes and Ahlgrimm, 2014) and high (Kärcher, 2017; Bland et al., 2024) altitudes – so interpretations of these variables have to be taken with a grain of salt particularly in the earlier, observation sparse periods. This also goes for the total column hydrometeor mass (“Hydrometeors") which is calculated by adding the total column cloud liquid, ice, rain and snow water. We include this variable because it most closely resembles the variable determining the scene-dependent uncertainties derived for microwave radiances (Geer and Bauer, 2011). Finally, we also include the average relative humidity between 500 and 850 hPa to capture the variable moisture availability in regions most relevant for large scale ascent and convection. As mentioned before, associated observations are also equipped with derived (state-dependent) uncertainties rather than prescribed (fixed) ones.

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

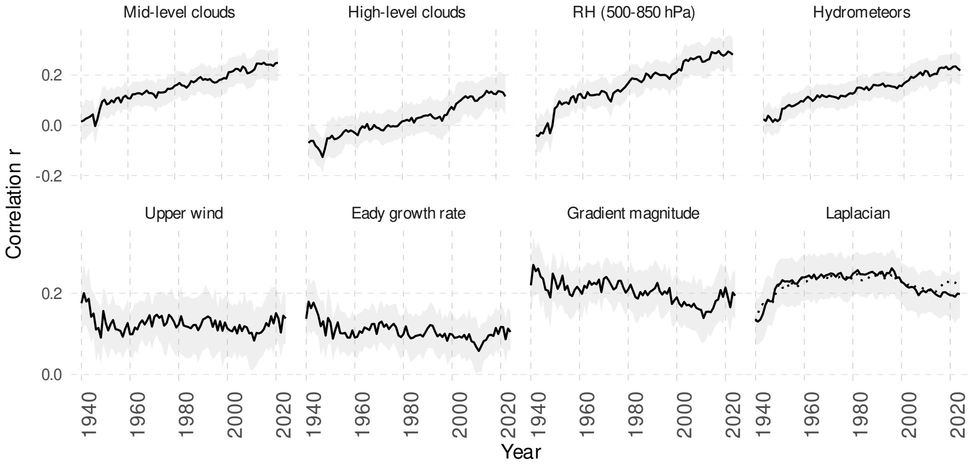

Figure 8. Pearson correlation coefficient r between the variance-balanced residuals $\hat{\varepsilon}^*$ and the variable indicated atop the facets based on annual time slices. See text for full definition of the variables. Line indicates median whereas shading indicates interquartile range across grid points. Statistical moments are latitude-weighted. The dotted line in the last panel indicates the median annual correlation for the absolute Laplacian.

All time series related to moist processes exhibit roughly the same trends indicating a strong collinearity. This trend points to an increase in the correlation with time, though the only sudden jump discernable is around 1950, so no single specific satellite instrument and associated uncertainties-of-the-day exert a major influence here. This argument is supported by the fact that a differentiation by land and sea grid points paints an essentially equivalent picture (not shown), whereas microwave instruments treat land and sea grid points drastically different (Geer and Bauer, 2011).

The first ten years of these time series are also interesting in so far as ERA5 can be conceived almost as a free IFS run due to its weak constrain by observations (especially with respect to tropospheric moisture) in this period. It is remarkable that virtually no correlation exists between these moist variables and the uncertainty as that contradicts the hypothesis of stochastic physics exerting a major influence. It is, however compatible with the hypothesis that short time-scale error growth relevant for the background forecast surely dominant in this period is not dominated by convective processes (cf. 3.1).

At any rate, the top row of fig. 8 proves there is a predominantly positive correlation between the selected moist variables and the residuals, except for the high level-cloud cover. The positive trend with time indicates that this is an effect modulated by the observational coverage, since the IFS model and most of the assimilation set-up are consistent across the period. This, again, points to the assimilation uncertainty being dominated by varying observation uncertainties rather than stochastic physics. The relative humidity achieves the highest values of r which is why we carry it over to the ensuing second stage models as a covariate, omitting other moist variables because of high collinearity. We performed a sensitivity analysis on the choice of the vertical levels for the relative humidity and found the layer 500-850 hPa to achieve the highest correlation. Alternative choices are presented in App. D in fig. D8.

The lower row in fig. 8, by contrast, presents a number of variables representative of essentially dry processes and mathematical effects. We use the wind magnitude ($|\mathbf{u}_\mathrm{h}| = \sqrt{u^2+v^2}$) at 300 hPa (“Upper wind") as a measure of the upper-level jet stream forcing on the mid-tropospheric geopotential field under investigation due to its considerable influence on Rossby wave propagation (Wirth et al., 2018). Near-tropopause dynamics have been identified as highly relevant for error growth on short time-scales (Baumgart et al., 2018, 2019). We also study the Eady growth rate as a measure of baroclinic development (Eady, 1949; Pierrehumbert and Swanson, 1995), which we calculate according to Hoskins and Valdes (1990), between 500 and 850 hPa from instantaneous values (Simmonds and Lim, 2009). This is motivated by baroclinic instability being identified as a driver of error growth as well (Selz et al., 2022; Rodwell and Wernli, 2023).

Both variables unveil similar correlation values throughout the investigated period with r remaining somewhere between .1 and .2 most of the time. The similar trends hint at some collinearity between the two variables as well as the geopotential gradient itself (see below). It stands to reason that these variables are not independent from each other, still we infer from their comparatively low correlation with the residuals that jet-stream forcing and baroclinic instability play a secondary role in shaping the assimilation uncertainty.

The two last variables are derived from the underlying target field itself, namely the geopotential at 500 hPa. Gradient magnitude ($|\nabla\phi|$) and Laplacian ($\nabla^2\phi$) are calculated in a standard way in spherical coordinates (Holton and Hakim, 2013, ch. C.3) using a second-order accurate differencing scheme provided by the `xarray` package (Hoyer and Hamman, 2017). We calculate both variables on the finer .25° resolution and coarsen by spatial averaging to .5° to control small-scale noise and aliasing introduced by the Laplacian in particular (Ferziger et al., 2020, ch. 3).

As mentioned before, derivatives of a field affect positional errors $\delta\mathbf{x}$, by virtue of

$$\delta\phi = \nabla\phi\cdot\delta\mathbf{x} + \tfrac{1}{2}\,\delta\mathbf{x}^\mathsf{T}\,\mathsf{H}_\phi\,\delta\mathbf{x} + \ldots, \tag{22}$$

where H_ϕ is the Hessian of ϕ and $\operatorname{tr}\mathsf{H}_\phi = \nabla^2\phi$ is the Laplacian, which captures its isotropic part. Note that here $\delta\mathbf{x}$ denotes a spatial displacement rather than the analysis increment of Sect. 3. Both variables therefore give insight into the degree to which an observed error is caused by the horizontal structure of the field ϕ, but in complementary ways. Imagine, for example a strong low pressure center, where the gradient would vanish, but the Laplacian assume a local extremum. In such a case, a displacement will result in an above average error if the (absolute) Laplacian is large just like in a high gradient case.

The example above hints at a delicate issue with both of the variables, though. The horizontal structure of the geopotential in terms of its gradient and Laplacian is of course not some arbitrary mathematical aspect, but intricately linked to the prevalence of dynamical processes, in turn chiefly associated with error growth. From (quasi-)geostrophic theory one has (Holton and Hakim, 2013, ch. 6)

$$|\mathbf{u}_{\mathrm{g,h}}| = \frac{1}{f_0}|\nabla\phi|; \quad \zeta_\mathrm{g} = \frac{1}{f_0}\nabla^2\phi, \tag{23}$$

so the gradient magnitude is directly linked to the geostrophic wind magnitude and the Laplacian to the (quasi-)geostrophic relative vorticity. The Laplacian also appears in the quasi-geostrophic vorticity and omega equations, so errors in this term project onto the dynamical tendencies that drive the short-range evolution. Maxima of the gradient magnitude are typically associated with high wind regions such as mid-tropospheric jet streaks or severe storm systems while its minima locate high and low centers. The idea that high gradient regions are associated with high uncertainties, possibly in accord with dynamical implications is not new; for example, part of what makes WCBs sources of uncertainty is due to the their linkage with cold fronts. These are defined by their exceptionally high temperature gradient and have been identified as relevant objects for forecast uncertainty, too (Wernli and Gray, 2024). This holds in a similar manner for other synoptic features as well, in accord with increased baroclinicity (Yu et al., 2025) and the two aforementioned variables presumed to be indicative for regions of high error growth – namely the Eady growth rate and the upper-level wind speed – are not independent of the geopotential gradient on 500 hPa, but rather closely connected.

The Laplacian, on the other hand, measures the vorticity of the flow field, dynamically identifying (anti-)cyclonic regions like troughs and ridges. It frequently attains its maxima in low or high pressure centers, though regions of sharp curvature in the jet stream feature large Laplacian values as well. In an effort to disentangle effects specific to cyclonic and anticyclonic features, we have also performed the correlation analysis with the absolute of the Laplacian, which is indicated as a dotted line in the last panel. If elevated assimilation uncertainty due to a greater Laplacian was – even partially – caused by processes specific to either cyclonic or anticyclonic flow – like differences in cloud cover –, its absolute value would exhibit higher or lower correlations. Yet, it reveals barely any differences, so we reject that hypothesis and assume the evident connection is less via quasi-geostrophic dynamics, since cyclonic and anticyclonic flow features behave quite distinct dynamically. We will still keep the gradient's and Laplacian's dual role as mathematical and dynamic variables in mind for the following interpretations.

The same field geometry that signals dynamical activity also degrades the assimilation locally. Where ϕ varies rapidly, observation representativeness errors are inflated and the effective observational constraint is reduced; consistently, background-error correlation length scales shorten in dynamically active regions, so each observation constrains only a local neighbourhood and the residual uncertainty per unit area is correspondingly larger (Lee and Huang, 2020; Li et al., 2023). Sharp gradients and large Laplacians therefore tend to coincide with elevated assimilation uncertainty through two distinct routes—enhanced dynamical error growth and weakened observational constraint—which our covariates cannot disentangle but which both point in the same direction.

Both the gradient magnitude and the (absolute) Laplacian exhibit positive correlation throughout the time series. Even though some variations are apparent, the correlation does not follow a long-term trend comparable to the moist variables. The curve for the gradient bears resemblance to the curves of the upper wind speed magnitude and the Eady growth rate, which makes sense, since both of those relate to geopotential gradients, albeit on different vertical levels. Given its likely collinearity and a correlation of roughly .1 higher, we decide to carry over the gradient for our second stage models. Fitting a complementary set of models including both Eady growth rate and upper level wind magnitude confirmed this notion, but is not discussed for brevity. The curve for the Laplacian in fig. 8 seems particularly well behaved and we decide to use that variable in our second stage models.

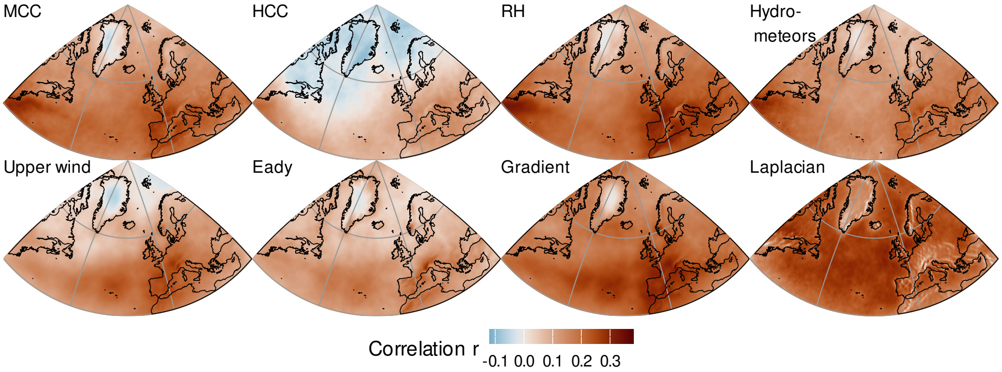

Figure 9. Pearson correlation coefficient r between the variance-balanced residuals $\hat{\varepsilon}^*$ and the variable indicated atop the facets. See text for full definition of the variables. Correlation calculated for the time period 1950-2024.

5.3.2 Spatially varying correlations

We also provide maps of correlation coefficients calculated at each grid point to give some insight into the spatial diversity of the statistical relations. They are presented in fig. 9. The patterns are quite diverse indeed, with variations fairly pronounced even within the group of moist variables. Shared across all variables except for the Laplacian seems to be a negative anomaly over Greenland, the reasons for which remain elusive to the authors. On a broader scale, polar regions exhibit lower correlation values than subtropical regions – a property perhaps also related to differences in observational coverage. Why exactly a higher cloud cover in upper tropospheric layers is associated with lower assimilation uncertainty is outside the scope of this study – our focus is also more on the mid-latitudes than on the polar regions.

Strong correlation is visible for the moist variables particularly in the Gulf Stream region and over Western Europe. Both of these regions are marked by fairly high variance overall (cf. fig. 3), which contributes to painting a picture of assimilation uncertainty dominated by moist processes on short time scales here (note that r is insensitive to the variance scaling 20). Compared to the other moist variables, the field of correlation coefficients for the relative humidity is smoother and generally less variable, which we consider a further argument in favor of choosing that variable for the ensuing analysis. Analogous maps for the relative humidity averaged between different levels are attached in App. D in fig. D9.

To test an influence due to the introduction of satellite instruments, we prepared maps analogous to the ones in fig. 9, but separately for the period before and after 1979 (not shown). The signals stay broadly consistent with changes attributable to changes in the mean apparent in fig. 8. Consistently, the quartiles visible as shading in fig. 8 move roughly in accord with the medians.

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

The dry dynamical variables (“Upper wind", and “Eady") feature a fairly similar pattern again, but the surface topography’s influence is more pronounced for the Eady growth rate, which is visible around mountain ranges. This comes as no surprise given it is anchored in lower levels of the troposphere. Both variables correlate most strongly in the central Atlantic, which complements well the pattern for the relative humidity that exhibits higher values over the Gulf Stream region and over Europe. The correlation for the geopotential gradient is somewhat similar to the upper wind magnitude and the Eady growth rate, but a little more constant across the whole region. Both of these aspects highlight the dynamical nature of the gradient apart from its mathematical association with uncertainty. Its strong correlation and evident mediation by the aforementioned two variables are further arguments in favor of its selection for the ensuing models.

The Laplacian, finally, acquires quite high values across a wide region, albeit visibly exhibiting small scale features especially at grid points close to prominent topography, probably related to computational and numerical issues. We still decide to include it in our second stage models due to its high correlation and idiosyncratic spatial pattern. The analogous map for the absolute of the Laplacian is provided in the appendix in fig. D9 and predominantly mirrors the basic structure, which supports our assumption that the relation between the residual uncertainty and the Laplacian is not dominantly caused by processes differing between cyclonic and anticyclonic circulation. It exhibits lower correlation values and is marred by small scale noise even more severely, so we stick with the Laplacian for the following analysis.

5.3.3 Explaining the weather regime patterns

After exploring possible variables as covariates for the assimilation uncertainty on synoptic time scales, we apply again the familiar model-based framework presented in Eq. 6 for meaningful handling of multiple covariates and robust inference. We select as covariates the relative humidity averaged between 500-850hPa as a proxy for the moisture influence as well as the geopotential’s gradient magnitude and Laplacian as proxies for the barotropic and baroclinic influences. In separate sets of models, the other variables analyzed hitherto (cf. fig.8) proved to be quite collinear, adding little explanatory power, but a lot of computational complexity, which is why we opted for this more concise set-up.

To infer the change in WR specific structure caused by the addition of the covariates, we also include the time series of WRs and compare results to those using models with only the WRs as a covariate (cf. eq. 21). This nested formulation offers inferential model comparison as will be explained below. Compared to 21 the advanced model is defined by:

$$\hat{\varepsilon}^* \sim x^{\mathrm{WR}} + x^{\mathrm{RH}} + x^{|\nabla\phi|} + x^{\nabla^2\phi} \tag{24}$$

Compared with the other models estimated thus far, we drop the AR(1) specification on the residuals, instead accounting for it post-hoc through HAC standard errors (cf. Sect. 4.3). This is because the geopotential gradient and Laplacian are strongly autocorrelated, too, and we assume part of the autocorrelation in $\hat{\varepsilon}^*$ is rooted in these variables. This decision does not primarily affect the point estimates, but rather the inference. In particular, both LRT and BIC rely on iid Gaussian residuals, which is now not accounted for anymore. Instead we employ Wald tests as specified in Sect. 4.3 to assess significant changes between model variants.

Consistent with our aim to understand the spatial uncertainty patterns shown in fig. 7, we want to see how the strength of the effect exerted by the WR time series changes upon addition of the above covariates. We argue that the WRs covariate is a flow-state indicator rather than directly linked to physical processes, so its effect has to be mediated by some other physical effects. If those effects are accounted for by further covariates, the signal strength will change. The covariates we add are local, have a finer structure and are more directly linked to physical processes, which is why they don't "compete" with the WR time series for signal strength, but absorb it to the extent they share variance with it. Estimated marginal means are the correct tool for this kind of analysis, since they evaluate the category means based on a marginal sample rather than a conditional sample – the conditional means (composites) do not change between the model variants.

We first assess the attenuation in the overall variability covered by the WRs. To that end, the standard deviation of the EMMs $\hat{\mu}^{\mathrm{WR}}(l)$ across the $|\mathcal{L}| = 8$ different levels of the WR variable at each grid point for the simple model 21, which we call $\sigma_{\mathrm{WR}}^{(\mathrm{simple})}$ in accord with Eq. 14, is provided in fig. 10. Calculating the standard deviation between the patterns for each WR means we disregard their differing occurrence frequencies. Critically, the variability measured on the left of fig. 10 is not equivalent to the total variability in the residuals at each grid point – that measure is presented in the center of fig. 3.

As suggested from the patterns in fig. 7, the largest variability in the WRs' effect on the residual uncertainty is found in the eastern Northern Atlantic. This is in contrast to the overall residual variance (cf. fig. 3), where the largest variability was found in the south western Atlantic, indicating the reason(s) for that maximum are not well represented by the WR clustering. That is not surprising, since the WRs are inherently designed for differentiating large scale structures, whereas it is sensible to assume, as we have seen, that the assimilation uncertainty is governed by small scale processes and observation specifics.

Additionally, the WR specific zg500 patterns themselves exhibit a fairly large variability in the North Atlantic compared to the Gulf Stream region, as shown in fig. D10 in App. D. This is because the WRs partition the state space not along the axis of the largest variability in assimilation uncertainty of zg500, but along the axis of largest variability in zg500 itself. Compared to the variability in WR patterns, high values of σ_{WR} do not extend so much into the polar regions. Given the above considerations, we argue the spatial pattern of $\sigma_{\mathrm{WR}}^{(\mathrm{simple})}$ can be understood as a combination of the variability in the response itself (V_{res}) driven mainly by strom track dynamics and variability in the WR fields themselves.

Also presented is the relative difference in WR related EMM variability between the simple and the full model with added covariates 24:

$$\Delta\sigma_{\mathrm{WR}} := \frac{\sigma_{\mathrm{WR}}^{(\mathrm{simple})} - \sigma_{\mathrm{WR}}^{(\mathrm{full})}}{\sigma_{\mathrm{WR}}^{(\mathrm{simple})}}. \qquad (25)$$

The construction as a relative measure means values can be large but still insignificant (when they are inflated by a small denominator), because in both absolute and relative terms, the same test ($\sigma_{\mathrm{WR}}^{(\mathrm{simple})} = \sigma_{\mathrm{WR}}^{(\mathrm{full})}$) is applied. The right panel in figure 10 proves that the second stage model reduces the variability in the WR associated uncertainty patterns for most of the region considered. For grid points with high pattern variability $\sigma_{\mathrm{WR}}^{(\mathrm{simple})}$, around half of the signal are attenuated due to the inclusion of the covariates. Even larger shares of variance are drawn from the WR covariate over much of the European mainland and in the subtropical eastern North Atlantic. The reasons will be explored later with the help of the attenuation's decomposition across the three covariates, but we deduct from the above that a meaningful part of the assimilation uncertainty

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

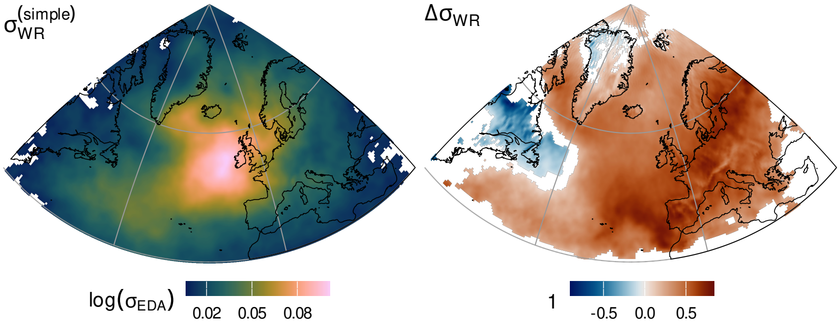

Figure 10. Weather regime pattern variability as measured by the standard deviation of the EMMs across the 8 weather regime categories. Left: standard deviation for the models with the weather regimes as the sole covariate. Right: Difference in standard deviation to a model with additional covariates introduced in the text relative to the standard deviation across the regimes for the simple model (shown on the left). Positive values indicate pattern attenuation. Values shown for grid points with FDR-adjusted p-values below 5%.

on synoptic time scales can be explained by the physical and mathematical processes associated with the three covariates utilized. We formulate this as a causal connection on purpose because the WRs themselves are not indicative of some process in particular but rather spatially constant flow state descriptors. The differences in the EMMs for each WR individually are also given in fig. D11 in App. D. In agreement with the aforementioned, the WR specific differences are largest in regions where the raw uncertainty patterns are also pronounced.

Interestingly, there are also regions where the additional covariates actually increase the standard deviation of the WRs uncertainty patterns, most notably over northeastern North America (this is also consistent across second stage models including other covariates from fig. 8; not shown). In this region, the uncertainty patterns do not exhibit a strong variability in either the simple or the full model and neither do the WR patterns in terms of zg500 show considerable variability (see. fig. D10). Though seemingly paradoxical, an increase in the between group differences through inclusion of further covariates implies these covariates mask true differences as long as they are not accounted for explicitly in the model. In other words, the WR specific assimilation uncertainty on synoptic time scales has to be driven by processes not related to either of the variables included. We do not explore this aspect further, especially since this region is arguably of secondary relevance. Finally, we note that the polar region, again, stands out by noisy signals also for this measure, which we think is, again, related to this analysis framework being tailored more to the mid-latitudes.

The WR pattern attenuation achieved by the second stage models may also be approximately decomposed into the contributions by each of the added covariates with a “leave-one-out" strategy: Instead of comparing $\sigma_{\mathrm{WR}}^{(\mathrm{simple})}$ with $\sigma_{\mathrm{WR}}^{(\mathrm{full})}$, we leave out one of the three additional covariates each and compute that models respective $\sigma_{\mathrm{WR}}^{(-\mathrm{pred})}$, which is the standard deviation of the EMMs across WRs in the model with all covariates except x^{pred}. This isolates the unique contribution of each covariate after accounting for the others, which an analogous “add-one" strategy would not. Results from “add-one" models mentioned above

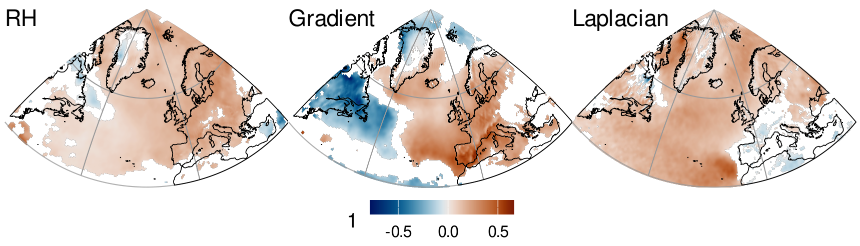

Figure 11. Relative weather regime pattern variability attenuation as measured by $\Delta\sigma_{\mathrm{WR}}$, but with one covariate removed from the full model 24. The removed covariate is indicated atop the facets. Values shown for grid points with FDR-adjusted p-values below 5%

would look virtually identical to fig. 9. The “leave-one-out" procedure provides an inexact decomposition in the sense that the three components do not exactly sum to the effect of the full model due to compensating effects between the covariates. That being said, the sum of the three components nonetheless approximates the full-model effect well, consistent with the covariates being largely mutually independent in their contribution to the regime signal. The results are presented in fig. 11.

A large part of the variability in the uncertainty that is projected onto the WRs clustering is due to the uncertainties’ correlation with the geopotential gradient itself. Though we cannot determine how much of this reflects the gradient’s role in QG quantities such as the geostrophic wind and the Eady growth rate, we consider it likely that this fact is predominantly caused by the mathematical and assimilation considerations laid out before. This is supported by the fact that attenuation by the gradient is strongest in regions where the WR patterns themselves show large variability (cf. fig. D10). The large and significant patch of negative values across northeastern North America and along its coast imply a similar conclusion. This is because the WR patterns exhibit minuscule variability in the zg500 gradient in that region whereas the region is generally quite variable in that respect. In other words, the variability in zg500 gradient within WR categories compared to the between category contrast, in effect masks contrasts between categories in the assimilation uncertainty. We stress that the gradient neither dominates the overall uncertainty variability nor is negatively correlated with it anywhere (see fig. 9).

The Laplacian’s effect – though generally a little smaller in magnitude – is significantly positive for most of the domain. An exception to this are large parts of Europe and Northern Africa, which we consider caused by topographic influences (cf. Fig. 9) and the other two covariates attaining larger effects here. The Laplacian’s secondary role is in agreement with it being a second-order derivative. We do maintain, however, that it accounts for a considerable part of the signal and in a regionally constant manner that speaks in favor of a general effect. We consider it also a further argument against the Laplacian’s effect being mediated through its role in QG dynamics, since QG factors certainly have varying effect magnitudes across the domain.

The inclusion of lower tropospheric humidity content, finally, also accounts for a non-negigible amount of signal in the WRs’ association with assimilation uncertainty. Its signal is also rather constant across space, indicating an effect independent of local climates. An exception is visible, again, for the northeastern North American region, which we attribute, again, to the WR classification's focus on regions downstream.

Given the particularly high correlation of this variable with the short time-scale assimilation uncertainty (cf. 9) in the Gulf Stream region, but a rather sparse and fragmented signal in this WR attenuation framework, it seems like there might be an important linkage missed. Indeed, for the region, the relative humidity attenuates the WR signal quite strongly – an effect that is masked by large internal variability as measured by our hypothesis tests. To further explore this intriguing aspect, we therefore turn to a more object oriented perspective by considering the role of WCBs.

5.3.4 Warm conveyor belts

As outlined in the introduction, WCBs are well established as amplifiers of forecast uncertainty, with much of that uncertainty commonly attributed to model formulation rather than to the initial state. Our setting differs in this respect: as seen in the previous section, the target variable—the variance-balanced residuals of the first-stage models $\hat{\varepsilon}^*$—is severely influenced by specifics of the observation system. We therefore do not consider model uncertainty in isolation, as is often done in studies of the impact of WCBs on predictability. Since observation-related uncertainties are very much present and relevant in the assimilation process for both reanalysis production and operational forecasts, we regard the relation between WCBs and our target variable as equally relevant, and we aim to contribute to the body of literature reviewed above by studying how WCBs relate to assimilation uncertainty specifically.

WCBs are often studied in a Lagrangian manner, but such an approach is cumbersome for our framework given our Eulerian setting. As outlined in Sect. 2, we utilize a WCB identification dataset that has been cast into a gridded format of four dimensions: latitude, longitude, time and lifecycle stage (inflow, ascent and outflow). Similar datasets have been used in the study of WCB-associated uncertainty, with particular focus on atmospheric blockings (Wandel et al., 2024, 2021; Pickl et al., 2023). As a first measure, we investigate the local in time and space association between WCBs and the residual uncertainty without stratifying according to WRs by simply fitting

$$\hat{\varepsilon}^* \sim x^{\mathrm{inf}} + x^{\mathrm{asc}} + x^{\mathrm{out}}. \tag{26}$$

The three covariates used are binary indicators of WCB occurrence in the respective stage – the simplest form of categorical covariates – so EMMs are, again, the appropriate measure to establish their influence. The results are shown in fig. 12. We note that the signals appear noisier in space compared to the EMMs of the WRs because the WCB data is present on a 1° resolution and from 1979 onwards only.

As expected from the positive influence of moisture variables generally, the presence of WCBs relates to anomalously high uncertainty at most grid points. The signal exhibits a meridional gradient attaining its highest values in the subtropics – even turning (significantly) negative for the outflow at higher latitudes. The signal is generally strongest for the ascent stage, which is expected given that we study the uncertainty in geopotential on 500 hPa, which coincides with the vertical layers considered for the ascent stage. The patch of negatively anomalous uncertainty associated with WCB outflow in high latitudes might be

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

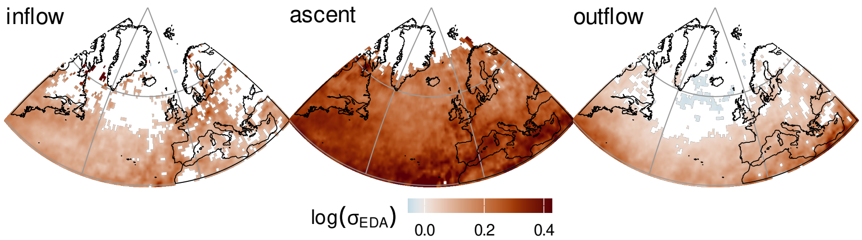

Figure 12. EMMs for each warm conveyor belt lifecycle stage. Values shown for grid points with FDR-adjusted p-values below 5%. For some polar grid points, warm conveyor belt occurrence frequency is not sufficient for model estimation.

caused by the ridge building effect such outflow tends to have (Steinfeld and Pfahl, 2019; Saffin et al., 2021). The divergent outflow aloft decreases the geopotential gradient such that WCBs effect might be mediated by the geopotential field's structure rather than the humidity. Using WCB centered forecast error composites, Pickl et al. (2023) uncovered positive anomalies in zg500 also during outflow stage, but those composites are averaged over forecast lead time and the whole region impeding a proper comparison.

We assume the stronger effect in the subtropics comes about due to more severe convection on the one hand and a larger fraction of variance not dominated by structures in the geopotential field itself on the other hand. Indeed, moist variables (cloud cover in particular) reach higher Pearson correlation coefficients in the subtropics compared to the Polar regions, too. Drawing attention to the differences in the range of values covered by the colour bar of fig. 12 compared to the colour bar in fig. 7, we maintain that the WCBs effect on uncertainty can at times be quite marked. Some of the differences in magnitude may also stem from the lower frequencies and spatial locality of the WCBs compared to the WRs, which makes them being more strongly related to dynamically active weather situations locally.

In contrast to the covariates analyzed in the last section, spatial and temporal lag presumably play a role in the way in which WCBs exert an influence on uncertainty on synoptic scales. This goes especially for the inflow and outflow stages as these processes occur at different vertical levels compared to our target variable zg500. For instance, Wandel et al. (2024) found that correct WCB representation days and weeks ahead influences the correct prediction of EuBL occurrences. More systematically, Pickl et al. (2023) showed that WCB occurrence at any stage is linked to (mostly elevated) anomalies in forecast-error growth, though, again, their analysis is probably not dominated by initial condition influences. Still, as mentioned before, assimilation uncertainty is also partially influenced by model error growth through the background forecast ensemble forecasts, suggesting factors determining uncertainty may carry over.

Crucially, such lagged influences need not be local in space: a WCB ascending in one region may shape the uncertainty downstream days later, an effect that a grid-point-wise analysis like that of the previous section cannot capture. A second

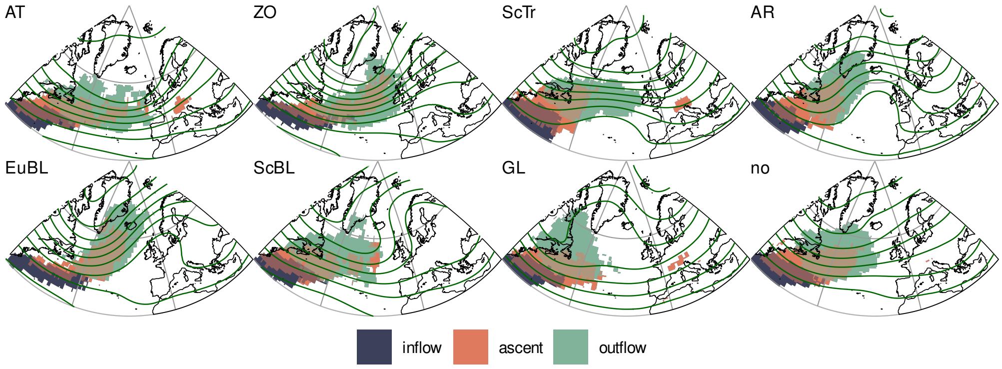

Figure 13. Climatological inflow, ascent and outflow warm conveyor belt masks for each weather regime for lag 12h (in- and outflow) and lag 0h for the ascent stage. The grid points in the indicated regions account for the top third of the climatological occurrence frequency mass. Green contours display the composite mean z500 maps for each of the weather regimes based on the ERA5 HRES product. Contour lines are drawn every 80 gpm.

obstacle is one of sample size. We are interested in the effects WCBs have across different WRs, but an analysis akin to the one performed in the last section would require dividing the already limited WCB dataset among at least eight WR classes. Our calculations accordingly indicate unstable optimizations and estimates blurred by large parameter uncertainty. For both reasons—the inherently non-local character of WCB influence and the prohibitive fragmentation of the sample—we adopt a more non-local approach.

As a first step, we compile climatologies of WCB occurrence at every lifecycle stage and at lags 0h, 12h and 24h (WCB occured 0, 12, 24 hours before) for each WR separately. We then sort all grid point in the region considered according to their individual occurrence frequency and select for every combination of lag, lifecycle stage and WR those grid points that taken together account for a third of the cosine latitude weighted occurrence frequency mass. The procedure yields nine binary masks per WR, of which we select one per lifecycle stage by identifying the highest values of the Pearson correlation coefficient as in Sect. 5.3.2. We provide maps of the correlation coefficient at every lag and lifecycle stage in App.D in fig. D12. Note that correlation peaks for lag 12h for the inflow and outflow stages, while it peaks for lag 0h for the ascent stage – a result very much in line with the fact that the ascent stage includes the pressure level of our target variable (500hPa) and suggesting disturbances from lower and higher levels take time to propagate to the mid-troposphere. The same lag for each lifecycle stage is selected across all WRs for consistency. Climatological masks are depicted for every WR in fig. 13.

In line with our understanding of WCBs, the climatologically most prominent regions for occurrence differ more between the WRs the higher the air streams ascend. The gulf stream is clearly identified as a common source region only differing by how far and wide the region extends out into the open Atlantic. The ascent and outflow stages on the other hand exhibit more

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

distinct patterns across the different WRs. By and large, they reflect how extratropical cyclones are guided differently by the differing background flow. We remind the reader that the masks shown in fig. 13 account for “only" a third of the climatological frequency mass, so WCBs of course also occur at other locations. Nevertheless, we argue that the masks do capture the cores of the climatological air stream pathways associated with each WR.

Recall that our goal with these masks is to be able to aggregate the WCB occurrences to a level that is sufficient for meaningful estimation of the WR specific influence of WCBs on assimilation uncertainty. To that end, we use these masks to conflate individual occurrences to a single, spatially integrated variable we consider to represent WCB activity. More specifically, for each time step in our data set, we count the WCB occurrences within each of the three masks belonging to the currently active WR for each of the respective lifecycle stages. This results in three one-dimensional variables $x^{\mathrm{inf,clim}}, x^{\mathrm{asc,clim}}, x^{\mathrm{out,clim}}$. For example, if at some time step at which the WR “AT" was active, there was not a single ascending WCB detected at any of the grid points in the climatological mask, then $x^{\mathrm{asc,clim}} = 0$ for that time step. If on the other hand WCBs at inflow stage were detected 12 hours before at every second of the grid points that are part of the climatological mask (and these grid points were distributed equally across latitudes), then $x^{\mathrm{inf,clim}} = .5$ for that time step.

The spatial aggregation makes the estimation problem more robust, since the three covariates are spatially constant and problems due to rare occurrences are avoided. We proceed by fitting

$$\hat{\varepsilon}^{*} \sim x^{\mathrm{WR}} * \left(x^{\mathrm{inf,clim}} + x^{\mathrm{asc,clim}} + x^{\mathrm{out,clim}}\right), \tag{27}$$

where the operator “$*$" indicates an interaction including a main effect for the WCB covariates (Pinheiro and Bates, 2000). In other words, we are estimating separate slopes for each of the WCB variables within each WR category and include a WR independent offset. The set-up implies that the results truly reflect WCB influence and are not confounded by WR specific patterns in fig. 7. Though this formulation allows for an attenuation analysis like performed above, that is not our goal, since it is unlikely that WCBs influence accounts for a dominant part of the WR specific influence and because such influence would likely be mediated through the variables analyzed in the last section. Experiments carried out confirm this conjecture (not shown).

The effect of the three WCB covariates individually is, as expected, small, so we instead show their combined effect when all three are simultaneously one standard deviation above their regime-specific average. This is admittedly a crude premise, but it conveys what elevated WCB activity entails for each weather regime. Concretely, for each regime l we define this $+1\,\mathrm{SD}$ effect as

$$\hat{\delta}(l) = \hat{\mu}\big(l \,\big|\, x^{(c)} = \mathrm{sd}_l\big) - \hat{\mu}\big(l \,\big|\, x^{(c)} = 0\big), \tag{28}$$

i.e. the difference between the estimated mean uncertainty $\hat{\mu}$ with the WCB covariates $x^{(c)}$ raised to their regime-specific standard deviation sd_l and that at the regime average ($x^{(c)} = 0$, since the covariates are centered within regime). Because $\hat{\delta}(l)$ is a linear contrast of the fitted GLS coefficients, significance is again assessed with Wald z-tests. We do not propagate the sampling uncertainty of sd_l itself, which renders the test mildly liberal in finite samples but is negligible relative to the uncertainty in the slope estimates.

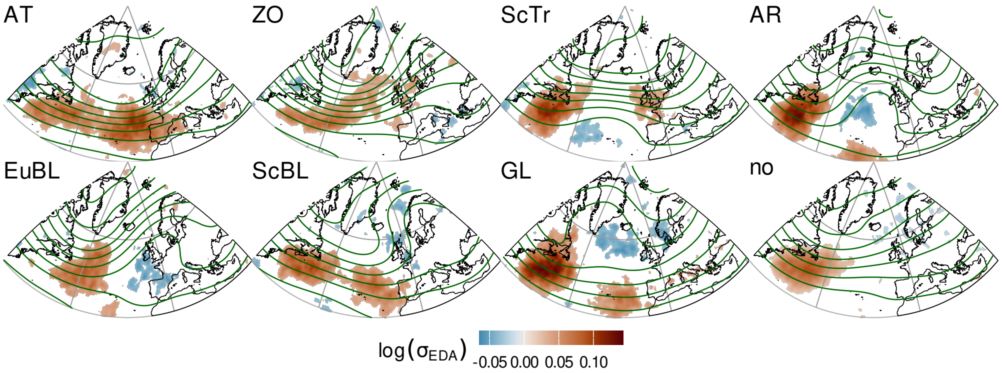

Figure 14. Effects of warm conveyor belt activity being one standard deviation above the mean for $x^{\mathrm{inf,clim}}$, $x^{\mathrm{asc,clim}}$ and $x^{\mathrm{out,clim}}$ simultaneously on assimilation uncertainty. Values shown for grid points with FDR-adjusted p-values below 5%. Green contours display the composite mean z500 maps for each of the weather regimes based on the ERA5 HRES product. Contour lines are drawn every 80 gpm.

Even though positively anomalous uncertainties are visible around the Gulf Stream region throughout all WRs, the wider ramifications are a lot more distinct. Interestingly, significant signals are not confined to the grid points included in the climatological WCB masks defined in fig. 13, though they are clearly largest there. In addition, negative anomalies appear for several regimes – all located within or close to anticyclonic patterns. This is further evidence in support of the hypothesis that upper-tropospheric outflow by WCBs decreases assimilation uncertainty by making the zg500 field smoother. Apparently, atmospheric states associated with above average WCB activity in regions where that activity predominantly occurs exhibit significantly (and mostly positively) anomalous assimilation uncertainty in a wide range of grid points across the whole region. In the present non-local perspective, an inverse contingency is also possible – namely that a higher uncertainty in the vicinity of anticyclonic patterns (caused by whatever mechanism) is a condition of the large scale flow that in turn facilitates WCB activity. We cannot rule out this process chain, but find it rather unlikely given our lag-structure and that WCB activity is typically dependent on upstream preconditions (Schäfler and Harnisch, 2015; Oertel et al., 2025).

Notwithstanding the minor magnitude and regionally confined signals, we maintain the results discussed above serve as significant evidence in favor of a solid (and mostly increasing) effect of WCBs on assimilation uncertainty on short time scales. In agreement with common conceptions of error growth in atmospheric blockings in particular, we find that WCBs induce uncertainty not within a blocking itself, but rather upstream, which is then propagated along the mid-latitude waveguide, where it grows and perturbs the ridge-building especially at upper levels. Consistent with our insights into the important role of moisture in governing such assimilation uncertainty, it is therefore reassuring to recognize such upstream uncertainty generation uncovered in a range of studies is represented – at least to some degree – in the initial uncertainty estimate governing the forecast uncertainty in ensemble forecasts. We do, however, also highlight that these spatially distinct and WR dependent patterns of

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

assimilation uncertainty need to be explicitly considered in such ensemble forecasting studies to avoid confounding statements about error growth with uncertainty actually caused by observation uncertainty.

6 Conclusions

This study investigated synoptic-scale assimilation uncertainty in the ERA5 reanalysis dataset by isolating it from signals introduced by changes in the observation system and seasonality using a statistical modeling framework applied to ensemble data assimilation (EDA) output. By explicitly accounting for structural breaks associated with abrupt changes in the observing system and filtering long-term variability, we constructed a temporally consistent estimate of flow-dependent uncertainty spanning multiple decades. While being agnostic to the choice of atmospheric variable, here this methodology's application to geopotential on 500hPa in the Atlantic-European sector provided exemplary insights into assimilation uncertainty for midlatitude synoptic dynamics.

The results show that a substantial fraction of ensemble spread variability is linked to changes in the observational network. The identified breakpoints and statistical fits for the long term variability can be explained rather exhaustively by known historical changes to the observational and assimilation suite. After removing these long-term effects, the remaining variability exhibits clear flow dependence and organizes coherently across Euro-Atlantic WRs. We take this stratification along WRs as a reference to further explore the processes and mechanisms governing the uncertainty on short time scales.

We find no single cause, but rather multiple variables significantly and independently related to the above described uncertainty. A combination of mathematical mechanisms and moist dynamical processes, we argue, is able to explain the largest part of the WR dependent uncertainty patterns. In particular, our results show that assimilation uncertainty of a field correlates strongly with the local structure of this field as quatified by its covariation with the gradient magnitude $|\nabla\phi|$ and its Laplacian $\nabla^2\phi$. Both these variables are deeply rooted in quasigeostrophic theory and it is likely that part of the association is via physical rather than mathematical mechanisms. Consistently, we find smaller, but consistent influences by baroclinic instability (Eady growth rate) and jet stream forcing.

Moist processes account for a considerable and independent part of the uncertainty and we find the strongest association with the integrated relative humidity in the lower troposphere (850-500 hPa). That covariation varies in time indicating a mediation by observation rather than model uncertainties. Contrary to expectations, our analysis does not support the notion that stochastic parametrization schemes dominate the assimilation uncertainty. Rather, the presence of moisture exerts a state-dependent (positive) effect on the assimilation uncertainty through elevated observation uncertainties due to both increased observation perturbations and decreased observational coverage – in particular in regions with prevalent convective activity.

WCBs have been identified as key drivers of uncertainty in the mid-latitudes. Our study extends the existing body of research by proving this not only holds for forecast uncertainty, but also assimilation uncertainty. The extent to which larger assimilation uncertainties already preposition larger ensemble forecast spreads in case studies of WCBs is an interesting and relevant question warranting further investigation. In addition to spatio-temporally local effects, it appears the area-integrated level of WCB activity also has a bearing on assimilation uncertainty with, again, WR specific spatial patterns.

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

On a more general level, the results presented herein point to short time scale assimilation uncertainties being driven by observation specific aspects more than physics of short-forecast error growth. While regions and times associated with elevated uncertainties may relate to both – namely in convectively active regions – it has been shown, that this is not always the case. We therefore argue considering flow-dependent assimilation uncertainty is vital, yet frequently overlooked in studies concerned with error growth or ensemble forecast uncertainty, especially from a WR perspective.

A central contribution of this work is the novel combination of multiple datasets and perspectives: long-term EDA-based uncertainty estimates, statistical models accounting for non-stationarity, WR classifications, and WCB occurrence. This integrated approach enables a more comprehensive characterization of atmospheric uncertainty than any single component could provide in isolation.

The datasets produced in this study are made available for further research along with the accompanying code. Given its temporal extent and the explicit separation of observational and flow-dependent components, we expect that it contains valuable information beyond the aspects explored here. It offers opportunities for future work on predictability, model evaluation, regime dynamics, and the role of moist processes in atmospheric uncertainty.

Several limitations should be noted. First, the EDA spread represents random uncertainty and does not capture systematic model biases or correlated errors, which may also be flow-dependent. Second, the interpretation of results relies on statistical inference rather than controlled numerical experiments, limiting the ability to establish causality. Third, the analysis is based solely on ERA5, although many findings are likely transferable to similar ensemble-based reanalysis systems. Finally, the focus on the Euro-Atlantic region restricts the generality of the conclusions, and extending the framework to other regions would be a natural next step.

Overall, the findings highlight the importance of considering observation-driven, flow-dependent uncertainty – especially in moist regimes – in reanalysis-based studies. Explicitly accounting for such effects can improve the interpretation of atmospheric variability and supports more informed use of reanalysis and ensemble forecast data in both weather and climate research.

Code and data availability. The code necessary to produce the results are provided in a ZENODO archive based on a GitHub repository (Schoeller, 2026). It also contains the variance balanced residuals $\hat{\varepsilon}^*$. ERA5 reanalysis data is available from the ECMWF's climate data store (CDS; cds.climate.copernicus.eu).

Appendix A: Composite means

We provide classical composites for comparison to EMMs. Define $\{I_{l,t}\}_{l=1}^{L}$, the binary indicators for the L different levels

$$I_{l,t} = \begin{cases} 1, & \text{if } x_t^{(c)} = l, \\ 0, & \text{otherwise,} \end{cases} \tag{A1}$$

with $T_l = \sum_{t=1}^{T} I_{l,t}$ the total number of observations in category l and, naturally, $T = \sum_{l=1}^{L} T_l$. Then the unadjusted composite mean is

$$\overline{y}^{(l)} = \frac{1}{T_l} \sum_{t=1}^{T} I_{l,t}\, y_{n,t}, \tag{A2}$$

which is the empirical average of $y_{n,t}$ over all times assigned to category l. Unlike EMMs, composite means do not adjust for other covariates and do not inherit model assumptions, making them useful as a simple baseline for comparison.

For the composite means $\overline{y}^{(l)}$ calculated according to equation A2, we infer statistical significance under the null hypothesis of no association between the levels $\{l\}_{l=1}^{L}$ and the observed variable $y_{n,t}$. This can be done under minimal assumptions by employing a non-parametric permutation (Monte Carlo) test on the time series of categories to account for its block structure (Good, 2005), which is an advantage compared to EMMs. Partition the full label sequence $(x_1^{(c)}, x_2^{(c)}, \ldots, x_T^{(c)})$ into contiguous blocks $(\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_B)$ in which the category $x_t^{(c)}$ remains constant:

$$\mathcal{B}_b = (x_{s_b}^{(c)}, x_{s_b+1}^{(c)}, \ldots, x_{e_b}^{(c)}); \quad b = 1, \ldots, B, \tag{A3}$$

$$x_{s_b}^{(c)} = x_{s_b+1}^{(c)} = \ldots = x_{e_b}^{(c)}, \tag{A4}$$

with $s_1 = 1$, $e_B = T$, and $s_{b+1} = e_b + 1$. The block lengths are

$$\tau_b = e_b - s_b + 1, \quad \sum_{b=1}^{B} \tau_b = T. \tag{A5}$$

To construct P surrogate sequences under the null hypothesis of no category effect, we sample P random permutations $\{\pi^{(p)}\}_{p=1}^{P}$ of the block indices $\{1, \ldots, B\}$, subject to the constraint that adjacent blocks in each permutation do not contain the same category (preserving block lengths for each level of $x_t^{(c)}$), then concatenate

$$\mathcal{B}_{\pi^{(p)}(1)}, \mathcal{B}_{\pi^{(p)}(2)}, \ldots, \mathcal{B}_{\pi^{(p)}(B)} \tag{A6}$$

to obtain the surrogate sequence $(x_1^{(c)}, x_2^{(c)}, \ldots, x_T^{(c)})^{(p)}$. With the corresponding indicators $I_{l,t}^{(p)}$ defined as in A2, the permuted composite mean for each surrogate is

$$\overline{y}^{(l)(p)} = \frac{1}{T_l} \sum_{t=1}^{T} I_{l,t}^{(p)}\, y_t. \tag{A7}$$

The empirical distribution $\{\overline{y}^{(l)(p)}\}_{p=1}^{P}$ approximates the null distribution of the composite mean for category $l \in \mathcal{L}$, against which the observed value $\overline{y}^{(l)}$ is compared to assess statistical significance.

Appendix B: Observation Uncertainty and the Perturbation–Weight Trade-off

Consider a scalar analogue of Eq. (3) with a single observation, $H = 1$, background-error variance σ_b^2 and observation-error variance σ_o^2. The gain reduces to

$$K = \frac{\sigma_\mathrm{b}^2}{\sigma_\mathrm{b}^2 + \sigma_\mathrm{o}^2}. \tag{B1}$$

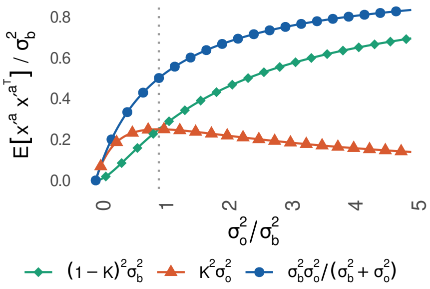

Figure B1. Impact of varying the observation uncertainty on the ensemble spread and its decomposition according to Eq. (B2).

A perturbed member assimilates $y^{\mathrm{O}} + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \sigma_{\mathrm{o}}^2)$, so its analysis deviation from the ensemble mean is $x'^{\mathrm{a}} = (1-K)\,x'^{\mathrm{b}} + K\epsilon$. With independent background and observation perturbations, the analysis spread decomposes as

$$\mathbb{E}[x'^{\mathrm{a}} x'^{\mathrm{a}T}] = \underbrace{(1-K)^2\sigma_{\mathrm{b}}^2}_{\text{filtered background}} + \underbrace{K^2\sigma_{\mathrm{o}}^2}_{\text{observation perturbations}} = \frac{\sigma_{\mathrm{b}}^2\sigma_{\mathrm{o}}^2}{\sigma_{\mathrm{b}}^2+\sigma_{\mathrm{o}}^2}. \tag{B2}$$

The two mechanisms of Sect. 3.1 appear as the two terms. Raising σ_{o}^2 enlarges each perturbation but lowers K, so the observation term $K^2\sigma_{\mathrm{o}}^2 = \sigma_{\mathrm{b}}^4\sigma_{\mathrm{o}}^2/(\sigma_{\mathrm{b}}^2+\sigma_{\mathrm{o}}^2)^2$ is non-monotonic: it vanishes for both $\sigma_{\mathrm{o}}^2 \to 0$ and $\sigma_{\mathrm{o}}^2 \to \infty$ and peaks at $\sigma_{\mathrm{o}}^2 = \sigma_{\mathrm{b}}^2$. The direct injection of spread by the observation perturbations is thus partially offset by the reduced weight in the gain. This behaviour is illustrated in fig. B1

The total spread in Eq. (B2), however, increases monotonically with σ_{o}^2 and is bounded above by σ_{b}^2: as observations are downweighted, the analysis simply retains more of the background spread. The compensation therefore does not reverse the response of the spread to observation uncertainty but only bends it sublinearly towards the background-spread ceiling – a second-order modulation rather than a first-order cancellation. The same limit holds in the matrix case, where $(\mathbf{I}-\mathbf{KH})\mathbf{B} \to \mathbf{B}$ as $\mathbf{R} \to \infty$. In principle, this allows for the possibility of large observation uncertainty being masked by low background uncertainty. Imagine, for instance, all background forecasts agreeing on a convective cell not appearing, but such a cell increasing observation uncertainty drastically. Such degenerate cases are rather hypothetical, in particular, because for the operational range of σ_{o}^2, in which strongly cloud- or aerosol-affected observations are already removed by quality control (Sect. 3.1.1), the assimilated observations sit away from the saturated tail, so the trade-off leaves the leading-order dependence on observation uncertainty intact.

Appendix C: Relation between anomalies and uncertainty

As mentioned in Sect. 5.3, the patterns of anomalous assimilation uncertainty in zg500 for each WR (cf. fig. 7) look similar to the normalised, low-pass filtered anomalies in the geopotential field itself for each of the WRs (cf. Supplementary fig. S1 in Grams et al. (2017) or fig. 5 in Grams (2026)), but with a negative sign. One might be tempted to try to interpret the apparent

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

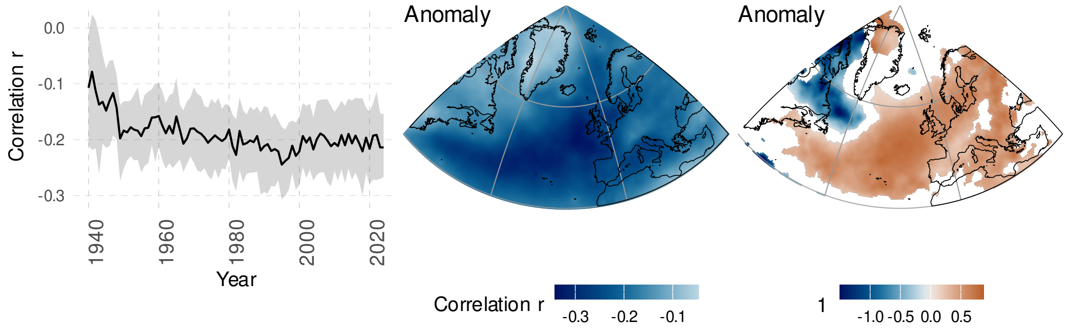

Figure C1. Correlation and attenuation analysis for normalised, low-pass filtered zg500 anomalies as mentioned in the text. Left: Pearson correlation coefficient r between the variance-balanced residuals $\hat{\varepsilon}^*$ and the zg500 anomalies ased on annual time slices. Line indicates median whereas shading indicates interquartile range of values across grid points. Statistical moments are latitude-weighted. Center: Pearson correlation coefficient at each grid point calculated for the time period 1950-2024. Right: Relative attenuation of WR patterns by the inclusion of the zg500 anomalies. Same quantity as on the right of fig. 10. Values below or equal to -1.5 are shown in the same colour (colourbar has been clipped). Values shown for grid points with FDR-adjusted p-values below 5%

relation between the deviation from a mean state and the uncertainty from a dynamical systems perspective, but that lense would arguably speak rather in favor of a correlation with positive sign.

To better understand the relation between the two variables, we perform the same analysis steps as for the other variables investigated in our study (cf. Grams (2026) for a thorough definition of the seasonally adjusted anomalies). We present in fig. C1 the time series and maps of the Pearson correlation coefficient r akin to figs. 8 and 9. Also shown is the relative attenuation of the WR patterns in uncertainty due to inclusion of the variable as a covariate $\Delta\sigma_{\mathrm{WR}}$ – the same quantity as in fig. 10.

The figures reveals a correlation between the values of roughly -.2 throughout the period except for the first decade where correlation is weaker. The map proves that, similar to the other dry variables the strongest signal is spatially over the Atlantic and Western Europe, which comes as no surprise given they are essentially derived from (mostly) the same variable. The negative correlation is in agreement with our understanding of the assimilation uncertainty, because positive anomalies impose anticyclonic flow patterns when added to the background field (which is essentially equivalent to the flow field of the “no" category). We would, thus, argue that the negative correlation between the uncertainty and the anomalies is mediated through the horizontal structure of the field (and its resulting association with moist variables) rather than the deviation from some mean state. The weaker correlation magnitude gives us additional confidence in performing our analysis using the covariates selected in Sect. 5.3.3.

The attenuation achieved upon inclusion of the variable as a covariate is quite remarkable (note the colourbar scaling). The effect is significantly positive roughly in the same regions where the correlation is also largest, whereas we find stronger

WR specific uncertainty emerge for the Northeastern Northern American region again. This implies the center of maximal variability has shifted to the North West compared to the raw model. As explained above, we consider this effect indirect – at least it is unclear why mathematically positive anomalies in a field should entail negative anomalies in the assimilation uncertainty of that field in a direct sense.

1210 Appendix D: Additional Figures

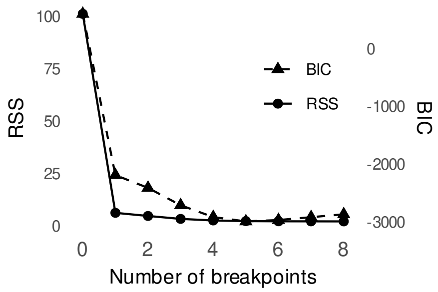

Figure D1. BIC and RSS for various segmentations. A minimum of the BIC corresponds to the preferred segmentation. Data and model assumption as described in Sect. 5.1. The resulting model and break points are shown in fig. 1.

Author contributions. HS conceptualized the study, developed the methodology, carried out the analysis, and wrote the manuscript. SP provided supervision and contributed to the conceptual development and manuscript revision.

Competing interests. The authors declare no competing interest. SP is a member of the editorial board of *Weather and Climate Dynamics*.

Acknowledgements. This research has been funded by Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1114 “Scaling Cascades 1215 in Complex Systems", Project Number 235221301, Project A08 “Characterization and Prediction of Quasi-Stationary Atmospheric States". Furthermore, the authors would like to thank Christian Grams for providing the WR classification, Julian Quinting for providing the WCB dataset and Hans Hersbach, Mark Rodwell and Sarah-Jane Lock from the ECMWF as well as Henning Rust for helpful comments. The authors would also like to thank Shreyas Deshpande and Bennett et al. (2020) for computational support. The authors used chat bots hosted by FU Berlin (ki.fu-berlin.de/chatbot) to belp with formulations.

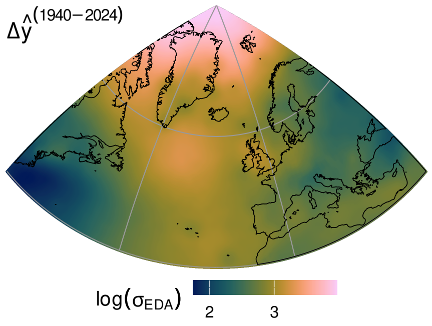

Figure D2. Difference in the model-estimated annual mean between the years 2024 and 1940 respectively as a proxy for the total reduction in observation uncertainty across the time span considered.

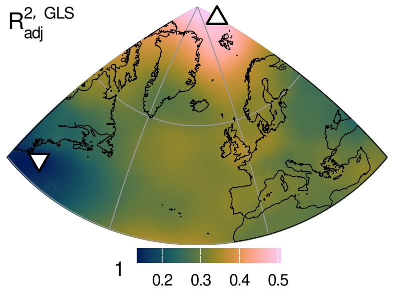

Figure D3. Same as figure 3 but for the precision weighted diagnostic introduced in equation 18.

References

Anthes, R. A., Bernhardt, P. A., Chen, Y., Cucurull, L., Dymond, K. F., Ector, D., Healy, S., Ho, S.-P., Hunt, D. C., Kuo, Y.-H., Liu, H., Manning, K., McCormick, C., Meehan, T. K., Randel, W. J., Rocken, C., Schreiner, W. S., Sokolovskiy, S. V., Syndergaard, S., Thompson, D. C., Trenberth, K. E., Wee, T.-K., Yen, N. L., and Zeng, Z.: The COSMIC/FORMOSAT-3 Mission: Early Results, Bull. Amer. Meteor. Soc., 89, 313–334, https://doi.org/10.1175/BAMS-89-3-313, 2008.

Bai, J.: Least Squares Estimation of a Shift in Linear Processes, J. Time Ser. Anal., 15, 453–472, https://doi.org/10.1111/j.1467-9892.1994.tb00204.x, 1994.

Bai, J.: Estimating Multiple Breaks One at a Time, Econ Theory, 13, 315–352, https://doi.org/10.1017/S0266466600005831, 1997.

Bai, J. and Perron, P.: Computation and Analysis of Multiple Structural Change Models, J of Applied Econometrics, 18, 1–22, https://doi.org/10.1002/jae.659, 2003.

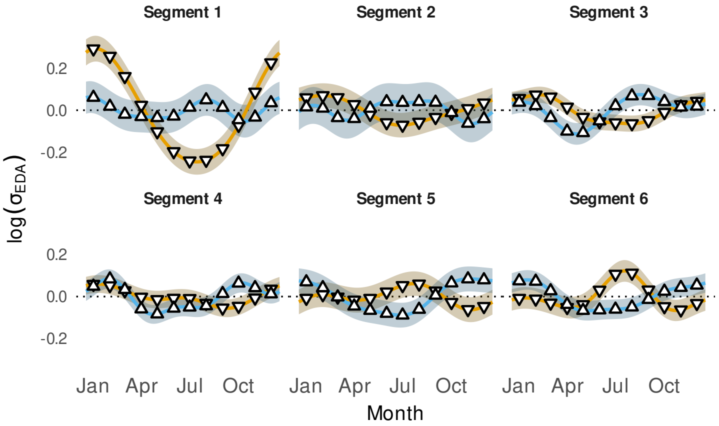

Figure D4. Statistical model estimates of the seasonal cycle $\hat{s}_{n,t}$ for every segment individually based on the models on daily scale shown in fig. 4. Blue (orange) line and shading and upward (downward) triangle indicate the Arctic (Gulf Stream) grid point shown in fig. 3. 95% confidence intervals are for the whole seasonal cycle combined rather than for individual months (cf. Sect. 4.3).

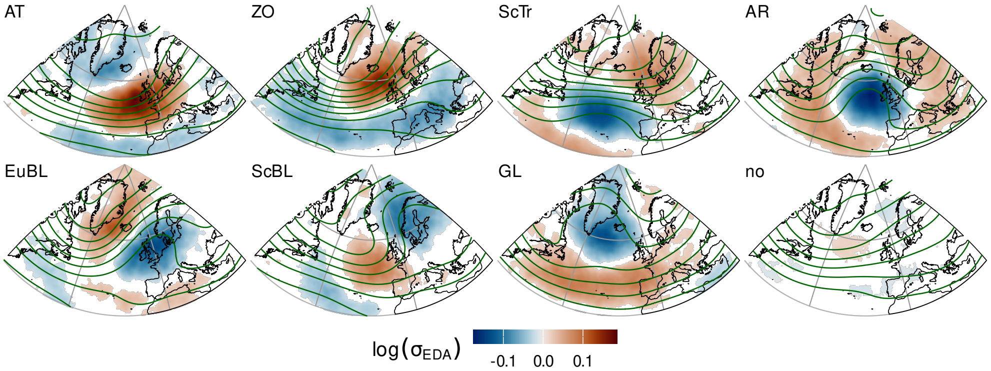

Figure D5. Composite mean maps of the variance-balanced residuals $\hat{\varepsilon}^*$. Values shown for grid points with FDR-adjusted p-values below 5% derived from the block-permutation test described in Sect. 4.3 with 10,000 draws. Green contours display the composite mean z500 maps for each of the weather regimes based on the ERA5 HRES product.

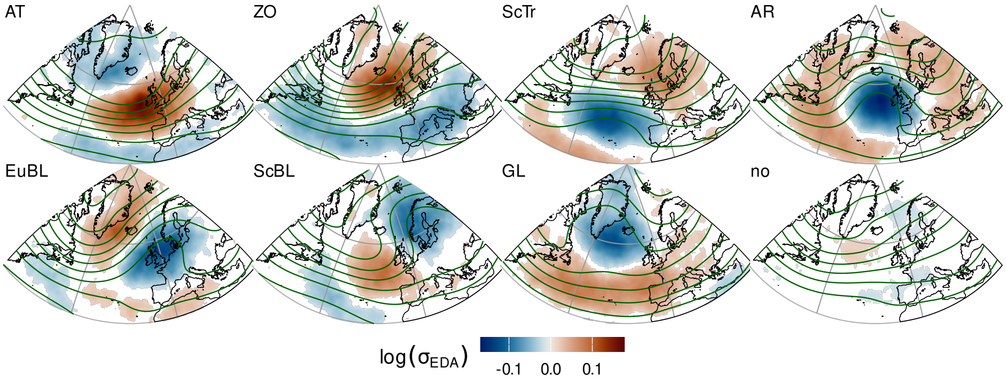

Figure D6. Same as fig. 7, but marginal means estimated from the full models from Sect. 5.2 extended by the weather regime time series as a covariate.

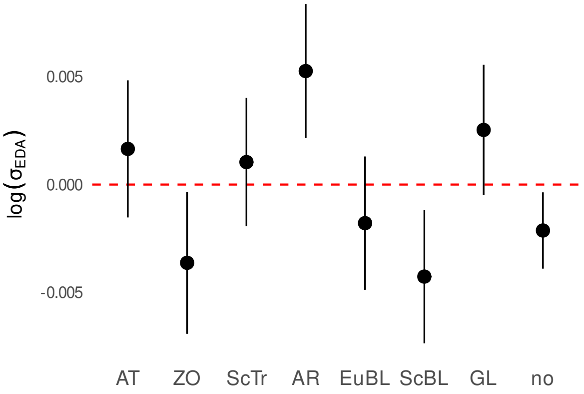

Figure D7. Estimated marginal mean effects of each of the weather regimes on the (latitude-weighted) spatially averaged variance-balanced residuals $\hat{\varepsilon}^*$. Points indicate estimate and lines the 95% confidence intervals. Confidence intervals excluding 0 imply statistical significance.

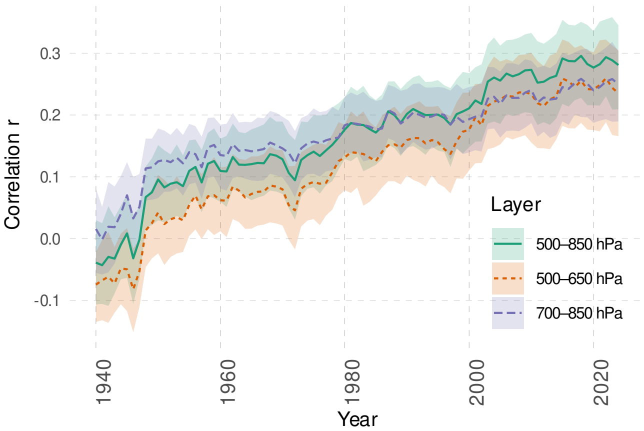

Figure D8. Pearson correlation coefficient r between the variance-balanced residuals $\hat{\varepsilon}^*$ and the vertically averaged relative humidity within the layers indicated in the legend based on annual time slices. Line indicates median whereas shading indicates interquartile range of values across grid points. All statistical moments are latitude-weighting.

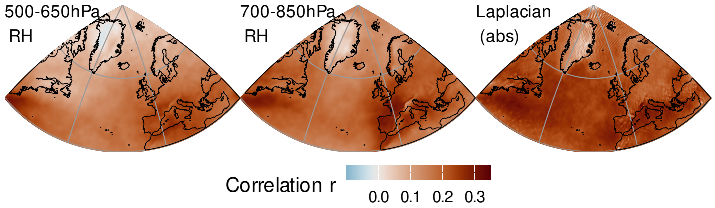

Figure D9. Pearson correlation coefficient r between the variance-balanced residuals $\hat{\varepsilon}^*$ and the variable indicated atop the facets. See text for full definition of the variables. Correlation is calculated for the time period 1950-2024.

Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnersley, J. S., Marquardt, C., Sato, K., and Takahashi, M.: The Quasi-Biennial Oscillation, Rev. Geophys., 39, 179–229, https://doi.org/10.1029/1999RG000073, 2001.

Bannister, R. N.: A Review of Operational Methods of Variational and Ensemble-variational Data Assimilation, Q. J. R. Meteorolog. Soc., 143, 607–633, https://doi.org/10.1002/qj.2982, 2017.

Baumgart, M. and Riemer, M.: Processes Governing the Amplification of Ensemble Spread in a Medium-range Forecast with Large Forecast Uncertainty, Quart J Royal Meteoro Soc, 145, 3252–3270, https://doi.org/10.1002/qj.3617, 2019.

Baumgart, M., Riemer, M., Wirth, V., Teubler, F., and Lang, S. T. K.: Potential Vorticity Dynamics of Forecast Errors: A Quantitative Case Study, Mon. Weather Rev., 146, 1405–1425, https://doi.org/10.1175/MWR-D-17-0196.1, 2018.

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

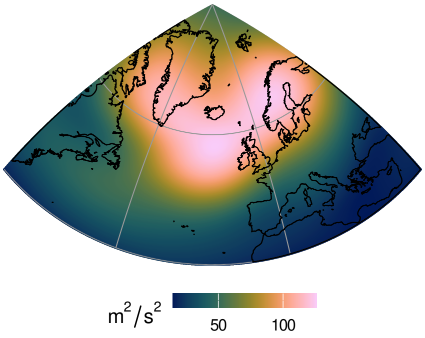

Figure D10. Standard deviation of zg500 across each of the eight weather regimes. The patterns are shown as dark green contours e.g. in fig. 7.

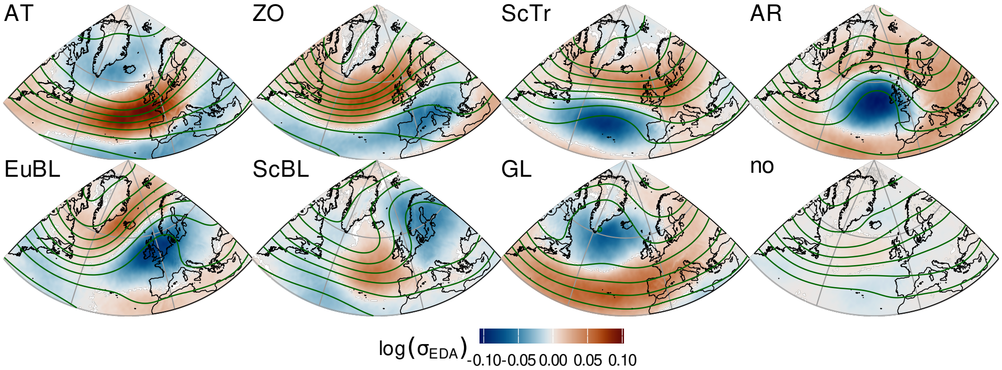

Figure D11. Difference in weather regime specific estimated marginal means between the simple model 21 and the full model 24. Positive (negative) values indicate higher (lower) values in the simple model. Values shown for FDR-adjusted p-values below 5%. Green contours display the composite mean z500 maps for each of the weather regimes based on the ERA5 HRES product. Contour lines are drawn every 80 gpm.

https://doi.org/10.5194/egusphere-2026-3137
Preprint. Discussion started: 16 June 2026

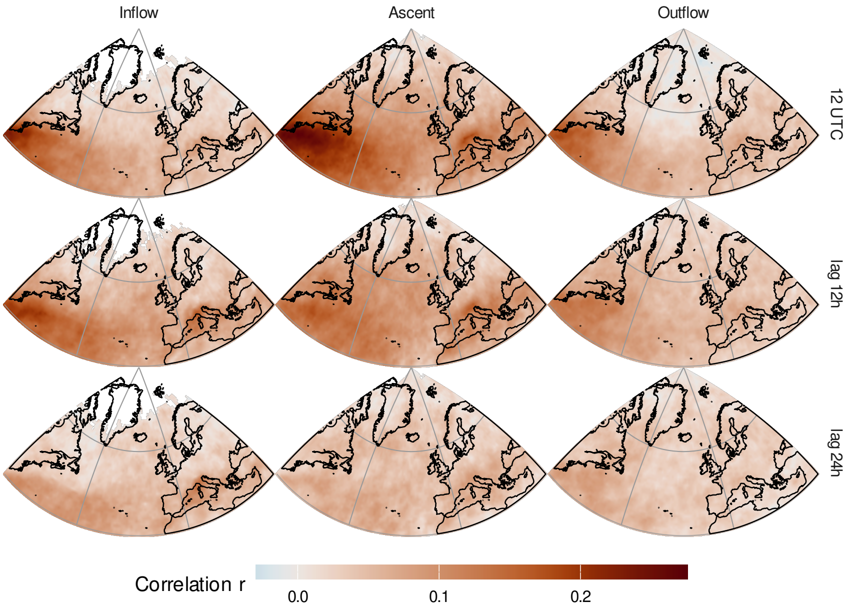

Figure D12. Pearson correlation coefficient r between the variance-balanced residuals $\hat{\varepsilon}^*$ and instantaneous occurrence of a warm conveyor belt in the lifecycle stage indicated atop the columns and at the temporal lag indicated aside each row. Correlation is calculated for the time period 1979-2024.

Baumgart, M., Ghinassi, P., Wirth, V., Selz, T., Craig, G. C., and Riemer, M.: Quantitative View on the Processes Governing the Upscale Error Growth up to the Planetary Scale Using a Stochastic Convection Scheme, Mon. Weather Rev., 147, 1713–1731, https://doi.org/10.1175/MWR-D-18-0292.1, 2019.

Bechtold, P., Sandu, I., Klocke, D., Semane, N., Ahlgrimm, M., Beljaars, A., Forbes, R. M., and Rodwell, M. J.: The Role of Shallow Convection in Ecmwf's Integrated Forecasting System, https://doi.org/10.21957/heba1qwem, 2014.

Bell, W., Hersbach, H., Simmons, A., Berrisford, P., Dahlgren, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Soci, C., Villaume, S., Bidlot, J.-R., Haimberger, L., Woollen, J., Buontempo, C., and Thépaut, J.-N.: The ERA5 Global Reanalysis: Preliminary Extension to 1950, Q. J. R. Meteorolog. Soc., 147, 4186–4227, https://doi.org/10.1002/qj.4174, 2021.

Benjamini, Y.: Discovering the False Discovery Rate, J. R. Stat. Soc. B, 72, 405–416, https://doi.org/10.1111/j.1467-9868.2010.00746.x, 2010.

Benjamini, Y. and Hochberg, Y.: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society. Series B (Methodological), 57, 289–300, 1995.

Benjamini, Y. and Yekutieli, D.: The Control of the False Discovery Rate in Multiple Testing under Dependency, Ann. Stat., 29, 1165–1188, 2001.

Bennett, L., Melchers, B., and Proppe, B.: Curta: A General-purpose High-Performance Computer at ZEDAT, Freie Universität Berlin, https://doi.org/10.17169/REFUBIUM-26754, 2020.

Berner, J., Shutts, G. J., Leutbecher, M., and Palmer, T. N.: A Spectral Stochastic Kinetic Energy Backscatter Scheme and Its Impact on Flow-Dependent Predictability in the ECMWF Ensemble Prediction System, J. Atmos. Sci., 66, 603–626, https://doi.org/10.1175/2008JAS2677.1, 2009.

Berner, J., Jung, T., and Palmer, T. N.: Systematic Model Error: The Impact of Increased Horizontal Resolution versus Improved Stochastic and Deterministic Parameterizations, Journal of Climate, 25, 4946–4962, https://doi.org/10.1175/JCLI-D-11-00297.1, 2012.

Berner, J., Achatz, U., Batté, L., Bengtsson, L., Cámara, A. D. L., 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.: Stochastic Parameterization: Toward a New View of Weather and Climate Models, Bull. Am. Meteorol. Soc., 98, 565–588, https://doi.org/10.1175/BAMS-D-15-00268.1, 2017.

Bland, J., Forbes, R. M., Gray, S. L., and Methven, J.: Processes Controlling Extratropical Near-Tropopause Humidity and Temperature in the ECMWF Global Weather Forecast Model, Q. J. R. Meteorolog. Soc., 150, 5356–5372, https://doi.org/10.1002/qj.4873, 2024.

Bloomfield, P.: Fitting Sinusoids, in: Fourier Analysis of Time Series, chap. 2, pp. 9–24, John Wiley & Sons, Ltd, ISBN 978-0-471-72223-6, https://doi.org/10.1002/0471722235.ch2, 2000.

Bonavita, M., Raynaud, L., and Isaksen, L.: Estimating Background-error Variances with the ECMWF Ensemble of Data Assimilations System: Some Effects of Ensemble Size and Day-to-day Variability, Q. J. R. Meteorolog. Soc., 137, 423–434, https://doi.org/10.1002/qj.756, 2011.

Bonavita, M., Isaksen, L., and Hólm, E.: On the Use of EDA Background Error Variances in the ECMWF 4D-var, https://doi.org/10.21957/3msfrh5zm, 2012.

Bonavita, M., Hólm, E., Isaksen, L., and Fisher, M.: The Evolution of the ECMWF Hybrid Data Assimilation System, Q. J. R. Meteorolog. Soc., 142, 287–303, https://doi.org/10.1002/qj.2652, 2016.

Bormann, N., Geer, A. J., and Bauer, P.: Estimates of Observation-Error Characteristics in Clear and Cloudy Regions for Microwave Imager Radiances from Numerical Weather Prediction, Q. J. R. Meteorolog. Soc., 137, 2014–2023, https://doi.org/10.1002/qj.833, 2011.

Bormann, N., Hernandez-Carrascal, A., Borde, R., Lutz, H.-J., Otkin, J. A., and Wanzong, S.: Atmospheric Motion Vectors from Model Simulations. Part I: Methods and Characterization as Single-Level Estimates of Wind, J. Appl. Meteorol. Climatol., 53, 47–64, https://doi.org/10.1175/JAMC-D-12-0336.1, 2014.

Box, G.: Robustness in the Strategy of Scientific Model Building, in: Robustness in Statistics, edited by Launer, R. L. and Wilkinson, G. N., pp. 201–236, Academic Press, ISBN 978-0-12-438150-6, https://doi.org/10.1016/B978-0-12-438150-6.50018-2, 1979.

Brockwell, P. J. and Davis, R. A.: Time Series: Theory and Methods, Springer Series in Statistics, Springer New York, New York, NY, ISBN 978-1-4419-0319-8 978-1-4419-0320-4, https://doi.org/10.1007/978-1-4419-0320-4, 1991.

Büeler, D., Ferranti, L., Magnusson, L., Quinting, J. F., and Grams, C. M.: Year-round Sub-seasonal Forecast Skill for Atlantic–European Weather Regimes, Q. J. R. Meteorolog. Soc., 147, 4283–4309, https://doi.org/10.1002/qj.4178, 2021.

Buizza, R. and Leutbecher, M.: The Forecast Skill Horizon, Q. J. R. Meteorolog. Soc., 141, 3366–3382, https://doi.org/10.1002/qj.2619, 2015.

Buizza, R., Milleer, M., and Palmer, T. N.: Stochastic Representation of Model Uncertainties in the ECMWF Ensemble Prediction System, Q. J. R. Meteorolog. Soc., 125, 2887–2908, https://doi.org/10.1002/qj.49712556006, 1999.

Buizza, R., Leutbecher, M., and Isaksen, L.: Potential Use of an Ensemble of Analyses in the ECMWF Ensemble Prediction System, Q. J. R. Meteorolog. Soc., 134, 2051–2066, https://doi.org/10.1002/qj.346, 2008.

Cardinali, C.: Observation Influence Diagnostic of a Data Assimilation System, in: Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II), edited by Park, S. K. and Xu, L., pp. 89–110, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-642-35087-0 978-3-642-35088-7, https://doi.org/10.1007/978-3-642-35088-7_4, 2013.

Cardinali, C. and Prates, F.: Performance Measurement with Advanced Diagnostic Tools of All-Sky Microwave Imager Radiances in 4D-var, Q. J. R. Meteorolog. Soc., 137, 2038–2046, https://doi.org/10.1002/qj.865, 2011.

Cardinali, C., Žagar, N., Radnoti, G., and Buizza, R.: Representing Model Error in Ensemble Data Assimilation, Nonlinear Processes Geophys., 21, 971–985, https://doi.org/10.5194/npg-21-971-2014, 2014.

Chen, J. and Gupta, A. K.: Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance, Birkhäuser Boston, Boston, ISBN 978-0-8176-4800-8 978-0-8176-4801-5, https://doi.org/10.1007/978-0-8176-4801-5, 2012.

Christ, S., Wenta, M., Grams, C. M., and Oertel, A.: From Sea to Sky: Understanding the Sea Surface Temperature Impact on an Atmospheric Blocking Event Using Sensitivity Experiments with the ICOsahedral Nonhydrostatic (ICON) Model, Weather Clim. Dynam., 6, 17–42, https://doi.org/10.5194/wcd-6-17-2025, 2025.

Christensen, H. M. and Zanna, L.: Parametrization in Weather and Climate Models, Oxford University Press, ISBN 978-0-19-022862-0, https://doi.org/10.1093/acrefore/9780190228620.013.826, 2022.

Christensen, H. M., Moroz, I. M., and Palmer, T. N.: Simulating Weather Regimes: Impact of Stochastic and Perturbed Parameter Schemes in a Simple Atmospheric Model, Clim Dyn, 44, 2195–2214, https://doi.org/10.1007/s00382-014-2239-9, 2015.

Clarke, S. J., Gray, S. L., and Roberts, N. M.: Downstream Influence of Mesoscale Convective Systems. Part 2: Influence on Ensemble Forecast Skill and Spread, Q. J. R. Meteorolog. Soc., 145, 2953–2972, https://doi.org/10.1002/qj.3613, 2019.

Courtier, P., Thépaut, J.-N., and Hollingsworth, A.: A Strategy for Operational Implementation of 4D-var, Using an Incremental Approach, Q. J. R. Meteorolog. Soc., 120, 1367–1387, https://doi.org/10.1002/qj.49712051912, 1994.

Davini, P., Weisheimer, A., Balmaseda, M. A., Johnson, S. J., Molteni, F., Roberts, C. D., Senan, R., and Stockdale, T. N.: The Representation of Winter Northern Hemisphere Atmospheric Blocking in ECMWF Seasonal Prediction Systems, Quart J Royal Meteoro Soc, 147, 1344–1363, https://doi.org/10.1002/qj.3974, 2021.

Davison, A. C. and Hinkley, D. V.: Bootstrap Methods and Their Application, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 1997.

Dawson, A. and Palmer, T. N.: Simulating Weather Regimes: Impact of Model Resolution and Stochastic Parameterization, Clim Dyn, 44, 2177–2193, https://doi.org/10.1007/s00382-014-2238-x, 2015.

Dee, D. P.: Bias and Data Assimilation, Q. J. R. Meteorolog. Soc., 131, 3323–3343, https://doi.org/10.1256/qj.05.137, 2005.

Dee, D. P. and Da Silva, A. M.: Data Assimilation in the Presence of Forecast Bias, Q. J. R. Meteorolog. Soc., 124, 269–295, https://doi.org/10.1002/qj.49712454512, 1998.

Deistler, M. and Scherrer, W.: Time Series Models, vol. 224 of *Lecture Notes in Statistics*, Springer International Publishing, Cham, ISBN 978-3-031-13212-4 978-3-031-13213-1, https://doi.org/10.1007/978-3-031-13213-1, 2022.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of Observation, Background and Analysis-error Statistics in Observation Space, Q. J. R. Meteorolog. Soc., 131, 3385–3396, https://doi.org/10.1256/qj.05.108, 2005.

Dickhaus, T.: Simultaneous Statistical Inference: With Applications in the Life Sciences, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-642-45181-2 978-3-642-45182-9, https://doi.org/10.1007/978-3-642-45182-9, 2014.

Downes, C. R.: History of the British Ocean Weather Ships, Mar. Obs., 47, 179–186, 1977.

Duncan, D. I., Bormann, N., Geer, A. J., and Weston, P.: Assimilation of AMSU-a in All-Sky Conditions, Mon. Weather Rev., 150, 1023–1041, https://doi.org/10.1175/MWR-D-21-0273.1, 2022.

Durran, D. R. and Gingrich, M.: Atmospheric Predictability: Why Butterflies Are Not of Practical Importance, Journal of the Atmospheric Sciences, 71, 2476–2488, https://doi.org/10.1175/JAS-D-14-0007.1, 2014.

Eady, E. T.: Long Waves and Cyclone Waves, Tellus, 1, 33–52, https://doi.org/10.1111/j.2153-3490.1949.tb01265.x, 1949.

Ecmwf: IFS Documentation CY41R2 - Part I: Observations, in: IFS Documentation CY41R2, 1, ECMWF, https://doi.org/10.21957/9phrx9eu, 2016a.

Ecmwf: IFS Documentation CY41R2 - Part V: Ensemble Prediction System, in: IFS Documentation CY41R2, 5, ECMWF, https://doi.org/10.21957/4btqaug2x, 2016b.

Ecmwf: IFS Documentation CY45R1 - Part V : Ensemble Prediction System, in: IFS Documentation CY45R1, 5, ECMWF, https://doi.org/10.21957/fwl3l91x, 2018.

Ecmwf: IFS Documentation CY49R1 - Part V: Ensemble Prediction System, in: IFS Documentation CY49R1, chap. 5, ECMWF, https://doi.org/10.21957/956d60ad81, 2024.

Emanuel, K.: Atlantic Tropical Cyclones Downscaled from Climate Reanalyses Show Increasing Activity over Past 150 Years, Nat. Commun., 12, 7027, https://doi.org/10.1038/s41467-021-27364-8, 2021.

English, S. J., Renshaw, R. J., Dibben, P. C., Smith, A. J., Rayer, P. J., Poulsen, C., Saunders, F. W., and Eyre, J. R.: A Comparison of the Impact of TOVS and ATOVS Satellite Sounding Data on the Accuracy of Numerical Weather Forecasts, Q. J. R. Meteorolog. Soc., 126, 2911–2931, https://doi.org/10.1002/qj.49712656915, 2000.

Ferziger, J. H., Perić, M., and Street, R. L.: Computational Methods for Fluid Dynamics, Springer International Publishing, Cham, ISBN 978-3-319-99691-2 978-3-319-99693-6, https://doi.org/10.1007/978-3-319-99693-6, 2020.

Filippucci, M., Bordoni, S., and Davini, P.: Impact of Stochastic Physics on the Representation of Atmospheric Blocking in EC-Earth3, Weather Clim. Dynam., 5, 1207–1222, https://doi.org/10.5194/wcd-5-1207-2024, 2024.

Forbes, R. M. and Ahlgrimm, M.: On the Representation of High-Latitude Boundary Layer Mixed-Phase Cloud in the ECMWF Global Model, Mon. Weather Rev., 142, 3425–3445, https://doi.org/10.1175/MWR-D-13-00325.1, 2014.

Geer, A. J. and Bauer, P.: Observation Errors in All-sky Data Assimilation, Q. J. R. Meteorolog. Soc., 137, 2024–2037, https://doi.org/10.1002/qj.830, 2011.

Geer, A. J., Lonitz, K., Weston, P., Kazumori, M., Okamoto, K., Zhu, Y., Liu, E. H., Collard, A., Bell, W., Migliorini, S., Chambon, P., Fourrié, N., Kim, M.-J., Köpken-Watts, C., and Schraff, C.: All-Sky Satellite Data Assimilation at Operational Weather Forecasting Centres, Q. J. R. Meteorolog. Soc., 144, 1191–1217, https://doi.org/10.1002/qj.3202, 2018.

Goldstein, H. and Healy, M. J. R.: The Graphical Presentation of a Collection of Means, J. R. Stat. Soc., Ser. (Stat. Soc.), 158, 175, https://doi.org/10.2307/2983411, 1995.

Good: Permutation, Parametric and Bootstrap Tests of Hypotheses, Springer Series in Statistics, Springer-Verlag, New York, ISBN 978-0-387-20279-2, https://doi.org/10.1007/b138696, 2005.

Grams, C. M.: A Life Cycle Definition of Year-Round Weather Regimes in the North Atlantic European Region, https://doi.org/10.5194/egusphere-2025-6385, 2026.

Grams, C. M., Beerli, R., Pfenninger, S., Staffell, I., and Wernli, H.: Balancing Europe's Wind-Power Output through Spatial Deployment Informed by Weather Regimes, Nature Clim Change, 7, 557–562, https://doi.org/10.1038/nclimate3338, 2017.

Grams, C. M., Magnusson, L., and Madonna, E.: An Atmospheric Dynamics Perspective on the Amplification and Propagation of Forecast Error in Numerical Weather Prediction Models: A Case Study, Q. J. R. Meteorolog. Soc., 144, 2577–2591, https://doi.org/10.1002/qj.3353, 2018.

Hamilton, J. D.: Time Series Analysis, Princeton University Press, ISBN 978-0-691-21863-2, https://doi.org/10.1515/9780691218632, 1994.

Hannachi, A., Straus, D. M., Franzke, C. L. E., Corti, S., and Woollings, T. J.: Low-frequency Nonlinearity and Regime Behavior in the Northern Hemisphere Extratropical Atmosphere, Reviews of Geophysics, 55, 199–234, https://doi.org/10.1002/2015RG000509, 2017.

Hauser, S., Teubler, F., Riemer, M., Knippertz, P., and Grams, C. M.: Life Cycle Dynamics of Greenland Blocking from a Potential Vorticity Perspective, Weather Clim. Dynam., 5, 633–658, https://doi.org/10.5194/wcd-5-633-2024, 2024.

Hauser, S., Cavallo, S. M., Magnusson, L., Martin, J. E., and Parsons, D. B.: Exceptionally Poor and Good Medium-Range Forecasts of the Large-Scale Circulation over Europe in ERA5 Reforecasts, Q. J. R. Meteorolog. Soc., p. e70117, https://doi.org/10.1002/qj.70117, 2026.

Hersbach, H., Bell, W., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J.-R., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D. P., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R. M., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 Global Reanalysis, Q.J.R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hieronymus, M., Baumgartner, M., Miltenberger, A. K., and Brinkmann, A.: Algorithmic Differentiation for Sensitivity Analysis in Cloud Microphysics, JAMES, 14, e2021MS002 849, https://doi.org/10.1029/2021MS002849, 2022.

Higham, N. J.: 14. Matrix Inversion, in: Accuracy and Stability of Numerical Algorithms, pp. 259–285, Society for Industrial and Applied Mathematics, Philadelphia, 2nd edition edn., https://doi.org/10.1137/1.9780898718027.ch14, 2002.

Hirahara, S., Alonso-Balmaseda, M., de Boisseson, E., and Hersbach, H.: Sea Surface Temperature and Sea Ice Concentration for ERA5, 2016.

Hittner, J. B.: Ezekiel's Classic Estimator of the Population Squared Multiple Correlation Coefficient: Monte Carlo-Based Extensions and Refinements, J. Gen. Psychol., 147, 213–227, https://doi.org/10.1080/00221309.2019.1679080, 2020.

Hollingsworth, A. and Lönnberg, P.: The Statistical Structure of Short Range Forecast Errors as Determined from Radiosonde Data Part I: The Wind Field, vol. II, pp. 7–70, ECMWF, Shinfield Park, Reading, 1984.

Holton, J. R. and Hakim, G.: An Introduction to Dynamic Meteorology, Elsevier, Amsterdam, fifth edition edn., ISBN 978-0-12-384867-3, 2013.

Hoskins, B. J. and Valdes, P. J.: On the Existence of Storm-Tracks, J. Atmos. Sci., 47, 1854–1864, https://doi.org/10.1175/1520-0469(1990)047<1854:OTEOST>2.0.CO;2, 1990.

Hoyer, S. and Hamman, J.: Xarray: N-D Labeled Arrays and Datasets in Python, J. Open Res. Software, 5, 10, https://doi.org/10.5334/jors.148, 2017.

Ide, K., Courtier, P., Ghil, M., and Lorenc, A. C.: Unified Notation for Data Assimilation : Operational, Sequential and Variational, J. Meteorol. Soc. Jpn., II, 75, 181–189, https://doi.org/10.2151/jmsj1965.75.1B_181, 1997.

Isaksen, L., Bonavita, M., Buizza, R., Fisher, M., Haseler, J., Leutbecher, M., and Raynaud, L.: Ensemble of Data Assimilations at ECMWF, https://doi.org/10.21957/obke4k60, 2010.

Janisková, M. and Cardinali, C.: On the Impact of the Diabatic Component in the Forecast Sensitivity Observation Impact Diagnostics, in: Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. III), edited by Park, S. K. and Xu, L., pp. 483–511, Springer International Publishing, Cham, ISBN 978-3-319-43414-8 978-3-319-43415-5, https://doi.org/10.1007/978-3-319-43415-5_22, 2017.

Janjić, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., Losa, S. N., Nichols, N. K., Potthast, R., Waller, J. A., and Weston, P.: On the Representation Error in Data Assimilation, Q. J. R. Meteorolog. Soc., 144, 1257–1278, https://doi.org/10.1002/qj.3130, 2018.

Joos, H. and Forbes, R. M.: Impact of Different IFS Microphysics on a Warm Conveyor Belt and the Downstream Flow Evolution, Q. J. R. Meteorolog. Soc., 142, 2727–2739, https://doi.org/10.1002/qj.2863, 2016.

Joos, H. and Wernli, H.: Influence of Microphysical Processes on the Potential Vorticity Development in a Warm Conveyor Belt: A Case-Study with the Limited-Area Model COSMO, Q. J. R. Meteorolog. Soc., 138, 407–418, https://doi.org/10.1002/qj.934, 2012.

Joselyn, J. A. and Ismail-Zadeh, A.: IUGG Evolves (1940–2000), Hist. Geo- Space Sci., 10, 45–72, https://doi.org/10.5194/hgss-10-45-2019, 2019.

Judt, F.: Insights into Atmospheric Predictability through Global Convection-Permitting Model Simulations, J. Atmos. Sci., 75, 1477–1497, https://doi.org/10.1175/JAS-D-17-0343.1, 2018.

Kalnay, E., Mote, S., and Da, C.: Earth System Modeling, Data Assimilation and Predictability: Atmosphere, Oceans, Land and Human Systems, Cambridge University Press, 2 edn., ISBN 978-0-511-92060-8 978-1-107-00900-4 978-1-107-40146-4, https://doi.org/10.1017/9780511920608, 2024.

Kärcher, B.: Cirrus Clouds and Their Response to Anthropogenic Activities, Curr. Clim. Change Rep., 3, 45–57, https://doi.org/10.1007/s40641-017-0060-3, 2017.

Ko, H.-C., Chun, H.-Y., and Bechtold, P.: Evaluation and Improvement of the ECMWF Aviation Turbulence Forecasts, J. Geophys. Res.: Atmos., 130, e2024JD043 158, https://doi.org/10.1029/2024JD043158, 2025.

Konsta, D., Dufresne, J.-L., Chepfer, H., Vial, J., Koshiro, T., Kawai, H., Bodas-Salcedo, A., Roehrig, R., Watanabe, M., and Ogura, T.: Low-Level Marine Tropical Clouds in Six CMIP6 Models Are Too Few, Too Bright but Also Too Compact and Too Homogeneous, Geophys. Res. Lett., 49, e2021GL097 593, https://doi.org/10.1029/2021GL097593, 2022.

Krishnamurthy, V.: Predictability of Weather and Climate, Earth Space Sci., 6, 1043–1056, https://doi.org/10.1029/2019EA000586, 2019.

Lang, S. T. K., Leutbecher, M., and Jones, S. C.: Impact of Perturbation Methods in the ECMWF Ensemble Prediction System on Tropical Cyclone Forecasts, Q. J. R. Meteorolog. Soc., 138, 2031–2046, https://doi.org/10.1002/qj.1942, 2012.

Lang, S. T. K., Dawson, A., Diamantakis, M., Dueben, P., Hatfield, S., Leutbecher, M., Palmer, T. N., Prates, F., Roberts, C. D., Sandu, I., and Wedi, N.: More Accuracy with Less Precision, Q. J. R. Meteorolog. Soc., 147, 4358–4370, https://doi.org/10.1002/qj.4181, 2021a.

Lang, S. T. K., Lock, S.-J., Leutbecher, M., Bechtold, P., and Forbes, R. M.: Revision of the Stochastically Perturbed Parametrisations Model Uncertainty Scheme in the Integrated Forecasting System, Q. J. R. Meteorolog. Soc., 147, 1364–1381, https://doi.org/10.1002/qj.3978, 2021b.

Lee, J. C. K. and Huang, X.-Y.: Background Error Statistics in the Tropics: Structures and Impact in a Convective-Scale Numerical Weather Prediction System, Q. J. R. Meteorolog. Soc., 146, 2154–2173, https://doi.org/10.1002/qj.3785, 2020.

Leith, C. E.: Atmospheric Predictability and Two-Dimensional Turbulence, J. Atmos. Sci., 28, 145–161, https://doi.org/10.1175/1520-0469(1971)028<0145:APATDT>2.0.CO;2, 1971.

Lenth, R. V. and Piaskowski, J.: Emmeans: Estimated Marginal Means, Aka Least-Squares Means, 2026.

Leutbecher, M. and Palmer, T. N.: Ensemble Forecasting, J. Comput. Phys., 227, 3515–3539, https://doi.org/10.1016/j.jcp.2007.02.014, 2008.

Leutbecher, M., Lock, S.-J., Ollinaho, P., Lang, S. T. K., Balsamo, G., Bechtold, P., Bonavita, M., Christensen, H. M., Diamantakis, M., Dutra, E., English, S. J., Fisher, M., Forbes, R. M., Goddard, J., Haiden, T., Hogan, R. J., Juricke, S., Lawrence, H., MacLeod, D., Magnusson, L., Malardel, S., Massart, S., Sandu, I., Smolarkiewicz, P. K., Subramanian, A., Vitart, F., Wedi, N., and Weisheimer, A.: Stochastic Representations of Model Uncertainties at ECMWF: State of the Art and Future Vision, Quart J Royal Meteoro Soc, 143, 2315–2339, https://doi.org/10.1002/qj.3094, 2017.

Li, J., Wang, P., Han, H., Li, J., and Zheng, J.: On the Assimilation of Satellite Sounder Data in Cloudy Skies in Numerical Weather Prediction Models, J. Meteorolog. Res., 30, 169–182, https://doi.org/10.1007/s13351-016-5114-2, 2016.

Li, L., Žagar, N., Raeder, K., and Anderson, J. L.: Comparison of Temperature and Wind Observations in the Tropics in a Perfect-Model, Global EnKF Data Assimilation System, Q. J. R. Meteorolog. Soc., 149, 2367–2385, https://doi.org/10.1002/qj.4511, 2023.

Lillo, S. P. and Parsons, D. B.: Investigating the Dynamics of Error Growth in ECMWF Medium-Range Forecast Busts, Q. J. R. Meteorolog. Soc., 143, 1211–1226, https://doi.org/10.1002/qj.2938, 2017.

Lloveras, D. J., Tierney, L. H., and Durran, D. R.: Mesoscale Predictability in Moist Midlatitude Cyclones Is Not Sensitive to the Slope of the Background Kinetic Energy Spectrum, J. Atmos. Sci., 79, 119–139, https://doi.org/10.1175/JAS-D-21-0147.1, 2022.

Lloveras, D. J., Durran, D. R., and Doyle, J. D.: The Two- to Four-Day Predictability of Midlatitude Cyclones: Don't Sweat the Small Stuff, J. Atmos. Sci., 80, 2613–2633, https://doi.org/10.1175/JAS-D-22-0232.1, 2023.

Lock, S.-J., Lang, S. T. K., Leutbecher, M., Hogan, R. J., and Vitart, F.: Treatment of Model Uncertainty from Radiation by the Stochastically Perturbed Parametrization Tendencies (SPPT) Scheme and Associated Revisions in the ECMWF Ensembles, Q. J. R. Meteorolog. Soc., 145, 75–89, https://doi.org/10.1002/qj.3570, 2019.

Lorenz, E. N.: The Predictability of a Flow Which Possesses Many Scales of Motion, Tellus A: Dyn. Meteorol. Oceanogr., 21, 289, https://doi.org/10.3402/tellusa.v21i3.10086, 1969.

Madonna, E., Wernli, H., Joos, H., and Martius, O.: Warm Conveyor Belts in the ERA-Interim Dataset (1979–2010). Part I: Climatology and Potential Vorticity Evolution, Journal of Climate, 27, 3–26, https://doi.org/10.1175/JCLI-D-12-00720.1, 2014.

Magnusson, L., Chen, J.-H., Lin, S.-J., Zhou, L., and Chen, J.-H.: Dependence on Initial Conditions versus Model Formulations for Medium-Range Forecast Error Variations, Q. J. R. Meteorolog. Soc., 145, 2085–2100, https://doi.org/10.1002/qj.3545, 2019.

Mathews, J. and Czaja, A.: Oceanic Maintenance of Atmospheric Blocking in Wintertime in the North Atlantic, Clim. Dyn., https://doi.org/10.1007/s00382-024-07196-0, 2024.

Matsunobu, T., Puh, M., and Keil, C.: Flow- and Scale-Dependent Spatial Predictability of Convective Precipitation Combining Different Model Uncertainty Representations, Q. J. R. Meteorolog. Soc., 150, 2364–2381, https://doi.org/10.1002/qj.4713, 2024.

Mazoyer, M., Ricard, D., Rivière, G., Delanoë, J., Riette, S., Augros, C., Borderies, M., and Vié, B.: Impact of Mixed-Phase Cloud Parameterization on Warm Conveyor Belts and Upper-Tropospheric Dynamics, Mon. Weather Rev., 151, 1073–1091, https://doi.org/10.1175/MWR-D-22-0045.1, 2023.

Moninger, W. R., Mamrosh, R. D., and Pauley, P. M.: Automated Meteorological Reports from Commercial Aircraft, Bull. Am. Meteorol. Soc., 84, 203–216, https://doi.org/10.1175/BAMS-84-2-203, 2003.

Newey, W. K. and West, K. D.: A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, Econom. : J. Econom. Soc., 55, 703–708, 1987.

Newey, W. K. and West, K. D.: Automatic Lag Selection in Covariance Matrix Estimation, Rev. Econom. Stud., 61, 631–653, https://doi.org/10.2307/2297912, 1994.

Odishaw, H.: International Geophysical Year, Science, 129, 14–25, https://doi.org/10.1126/science.129.3340.14, 1959.

Oertel, A., Miltenberger, A. K., Grams, C. M., and Hoose, C.: Interaction of Microphysics and Dynamics in a Warm Conveyor Belt Simulated with the ICOsahedral Nonhydrostatic (ICON) Model, Atmos. Chem. Phys., 23, 8553–8581, https://doi.org/10.5194/acp-23-8553-2023, 2023.

Oertel, A., Miltenberger, A. K., Grams, C. M., and Hoose, C.: Sensitivities of Warm Conveyor Belt Ascent, Associated Precipitation Characteristics and Large-scale Flow Pattern: Insights from a Perturbed Parameter Ensemble, Quart J Royal Meteoro Soc, 151, e4986, https://doi.org/10.1002/qj.4986, 2025.

Ollinaho, P., Lock, S.-J., Leutbecher, M., Bechtold, P., Beljaars, A., Bozzo, A., Forbes, R. M., Haiden, T., Hogan, R. J., and Sandu, I.: Towards Process-level Representation of Model Uncertainties: Stochastically Perturbed Parametrizations in the ECMWF Ensemble, Q. J. R. Meteorolog. Soc., 143, 408–422, https://doi.org/10.1002/qj.2931, 2017.

Osman, M., Beerli, R., Büeler, D., and Grams, C. M.: Multi-model Assessment of Sub-seasonal Predictive Skill for Year-round Atlantic–European Weather Regimes, Q. J. R. Meteorolog. Soc., 149, 2386–2408, https://doi.org/10.1002/qj.4512, 2023.

Palmer, T. N.: The ECMWF Ensemble Prediction System: Looking Back (More than) 25 Years and Projecting Forward 25 Years, Q. J. R. Meteorolog. Soc., 145, 12–24, https://doi.org/10.1002/qj.3383, 2019a.

Palmer, T. N.: Stochastic Weather and Climate Models, Nat Rev Phys, 1, 463–471, https://doi.org/10.1038/s42254-019-0062-2, 2019b.

Palmer, T. N., Shutts, G. J., Hagedorn, R., Doblas-Reyes, F. J., Jung, T., and Leutbecher, M.: Representing Model Uncertainty in Weather and Climate Prediction, Annu. Rev. Earth Planet. Sci., 33, 163–193, https://doi.org/10.1146/annurev.earth.33.092203.122552, 2005.

Palmer, T. N., Buizza, R., Doblas-Reyes, F. J., Jung, T., Leutbecher, M., Shutts, G. J., Steinheimer, M., and Weisheimer, A.: Stochastic Parametrization and Model Uncertainty, https://doi.org/10.21957/ps8gbwbdv, 2009.

Park, S. K.: Principles of Data Assimilation, Cambridge University Press, Cambridge, 1 edn., ISBN 978-1-108-83176-5, 2022.

Park, S. K., Xu, L., Park, S. K., and Xu, L.: Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. IV), Springer International Publishing AG, Cham, 1st ed. 2022. edn., ISBN 978-3-030-77721-0, 2022.

Petersen, R. A.: On the Impact and Benefits of AMDAR Observations in Operational Forecasting—Part I: A Review of the Impact of Automated Aircraft Wind and Temperature Reports, Bulletin of the American Meteorological Society, 97, 585–602, https://doi.org/10.1175/BAMS-D-14-00055.1, 2016.

Pickl, M., Lang, S. T. K., Leutbecher, M., and Grams, C. M.: The Effect of Stochastically Perturbed Parametrisation Tendencies (SPPT) on Rapidly Ascending Air Streams, Quart J Royal Meteoro Soc, 148, 1242–1261, https://doi.org/10.1002/qj.4257, 2022.

Pickl, M., Quinting, J. F., and Grams, C. M.: Warm Conveyor Belts as Amplifiers of Forecast Uncertainty, Quart J Royal Meteoro Soc, 149, 3064–3085, https://doi.org/10.1002/qj.4546, 2023.

Pierrehumbert, R. T. and Swanson, K. L.: Baroclinic Instability, Annu. Rev. Fluid Mech., 27, 419–467, https://doi.org/10.1146/annurev.fl.27.010195.002223, 1995.

Pinheiro, J. C. and Bates, D. M.: Mixed-Effects Models in S and s-PLUS, Springer, New York, https://doi.org/10.1007/b98882, 2000.

Pinheiro, J. C., Bates, D. M., and R Core Team: Nlme: Linear and Nonlinear Mixed Effects Models, https://doi.org/10.32614/CRAN.package.nlme, 2025.

Puh, M., Tempest, K. I., Keil, C., and Craig, G. C.: Flow Dependence of Forecast Uncertainty in a Large Convection-permitting Ensemble, Q. J. R. Meteorolog. Soc., 150, 5113–5126, https://doi.org/10.1002/qj.4860, 2024.

Quinting, J. F., Grams, C. M., Oertel, A., and Pickl, M.: EuLerian Identification of Ascending AirStreams (ELIAS 2.0) in Numerical Weather Prediction and Climate Models – Part 2: Model Application to Different Datasets, Geosci. Model Dev., 15, 731–744, https://doi.org/10.5194/gmd-15-731-2022, 2022.

R Core Team: R: A Language and Environment for Statistical Computing, Vienna, Austria, 2022.

Raynaud, L. and Bouttier, F.: Comparison of Initial Perturbation Methods for Ensemble Prediction at Convective Scale, Q. J. R. Meteorolog. Soc., 142, 854–866, https://doi.org/10.1002/qj.2686, 2016.

Raynaud, L., Berre, L., and Desroziers, G.: Objective Filtering of Ensemble-based Background-error Variances, Q. J. R. Meteorolog. Soc., 135, 1177–1199, https://doi.org/10.1002/qj.438, 2009.

Rivière, G., Wimmer, M., Arbogast, P., Piriou, J.-M., Delanoë, J., Labadie, C., Cazenave, Q., and Pelon, J.: The Impact of Deep Convection Representation in a Global Atmospheric Model on the Warm Conveyor Belt and Jet Stream during NAWDEX IOP6, Weather Clim. Dyn., 2, 1011–1031, https://doi.org/10.5194/wcd-2-1011-2021, 2021.

Rodwell, M. J. and Wernli, H.: Uncertainty Growth and Forecast Reliability during Extratropical Cyclogenesis, Weather Clim. Dyn., 4, 591–615, https://doi.org/10.5194/wcd-4-591-2023, 2023.

Rodwell, M. J., Magnusson, L., Bauer, P., Bechtold, P., Bonavita, M., Cardinali, C., Diamantakis, M., Earnshaw, P., Garcia-Mendez, A., Isaksen, L., Källén, E., Klocke, D., Lopez, P., McNally, A. P., Persson, A., Prates, F., and Wedi, N.: Characteristics of Occasional Poor Medium-Range Weather Forecasts for Europe, Bull. Am. Meteorol. Soc., 94, 1393–1405, https://doi.org/10.1175/BAMS-D-12-00099.1, 2013.

Rodwell, M. J., Forbes, R. M., and Wernli, H.: Why Warm Conveyor Belts Matter in NWP, https://doi.org/10.21957/MR20VG, 2018.

Saffin, L., Methven, J., Bland, J., Harvey, B., and Sanchez, C.: Circulation Conservation in the Outflow of Warm Conveyor Belts and Consequences for Rossby Wave Evolution, Q. J. R. Meteorolog. Soc., 147, 3587–3610, https://doi.org/10.1002/qj.4143, 2021.

Sánchez, C., Methven, J., Gray, S. L., and Cullen, M.: Linking Rapid Forecast Error Growth to Diabatic Processes, Q. J. R. Meteorolog. Soc., 146, 3548–3569, https://doi.org/10.1002/qj.3861, 2020.

Schäfler, A. and Harnisch, F.: Impact of the Inflow Moisture on the Evolution of a Warm Conveyor Belt, Q. J. R. Meteorolog. Soc., 141, 299–310, https://doi.org/10.1002/qj.2360, 2015.

Scheffé, H.: A Method for Judging All Contrasts in the Analysis of Variance, Biometrika, 40, 87–110, https://doi.org/10.1093/biomet/40.1-2.87, 1953.

Schoeller, H.: Isolating Flow-Dependent Uncertainty in Ensemble Reanalysis Data and Its Relation to Euro-Atlantic Weather Regimes and Warm Conveyor Belts, Zenodo, https://doi.org/10.5281/ZENODO.20474275, 2026.

Schwarz, G.: Estimating the Dimension of a Model, Ann. Stat., 6, https://doi.org/10.1214/aos/1176344136, 1978.

Searle, S. R.: Linear Models for Unbalanced Data, Wiley Series in Probability and Mathematical Statistics, Wiley, New York, ISBN 978-0-471-84096-1, 1987.

Searle, S. R., Speed, F. M., and Milliken, G. A.: Population Marginal Means in the Linear Model: An Alternative to Least Squares Means, Am. Stat., 34, 216–221, https://doi.org/10.1080/00031305.1980.10483031, 1980.

Selz, T.: Estimating the Intrinsic Limit of Predictability Using a Stochastic Convection Scheme, Journal of the Atmospheric Sciences, 76, 757–765, https://doi.org/10.1175/JAS-D-17-0373.1, 2019.

Selz, T. and Craig, G. C.: Upscale Error Growth in a High-Resolution Simulation of a Summertime Weather Event over Europe*, Mon. Weather Rev., 143, 813–827, https://doi.org/10.1175/MWR-D-14-00140.1, 2015.

Selz, T., Riemer, M., and Craig, G. C.: The Transition from Practical to Intrinsic Predictability of Midlatitude Weather, Journal of the Atmospheric Sciences, 79, 2013–2030, https://doi.org/10.1175/JAS-D-21-0271.1, 2022.

Shutts, G. J.: A Kinetic Energy Backscatter Algorithm for Use in Ensemble Prediction Systems, Q. J. R. Meteorolog. Soc., 131, 3079–3102, https://doi.org/10.1256/qj.04.106, 2005.

Simmonds, I. and Lim, E.-P.: Biases in the Calculation of Southern Hemisphere Mean Baroclinic Eddy Growth Rate, Geophys. Res. Lett., 36, https://doi.org/10.1029/2008gl036320, 2009.

Simmons, A., Soci, C., Nicolas, J., Bell, W., Berrisford, P., Dragani, R., Flemming, J., Haimberger, L., Healy, S., Hersbach, H., Horányi, A., Inness, A., Sabater, J. M., Radu, R., and Schepers, D.: Global Stratospheric Temperature Bias and Other Stratospheric Aspects of ERA5 and ERA5.1, https://doi.org/10.21957/rcxqfmg0, 2020.

Soci, C., Hersbach, H., Simmons, A., Poli, P., Bell, W., Berrisford, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Villaume, S., Haimberger, L., Woollen, J., Buontempo, C., and Thépaut, J.-N.: The ERA5 Global Reanalysis from 1940 to 2022, Quart J Royal Meteoro Soc, 150, 4014–4048, https://doi.org/10.1002/qj.4803, 2024.

Spaeth, J., Rupp, P., Osman, M., Grams, C. M., and Birner, T.: Flow-Dependence of Ensemble Spread of Subseasonal Forecasts Explored via North Atlantic-European Weather Regimes, Geophysical Research Letters, 51, e2024GL109 733, https://doi.org/10.1029/2024GL109733, 2024.

Sprenger, M., Fragkoulidis, G., Binder, H., Croci-Maspoli, M., Graf, P., Grams, C. M., Knippertz, P., Madonna, E., Schemm, S., Škerlak, B., and Wernli, H.: Global Climatologies of Eulerian and Lagrangian Flow Features Based on ERA-interim, Bull. Am. Meteorol. Soc., 98, 1739–1748, https://doi.org/10.1175/BAMS-D-15-00299.1, 2017.

Steinfeld, D. and Pfahl, S.: The Role of Latent Heating in Atmospheric Blocking Dynamics: A Global Climatology, Clim Dyn, 53, 6159–6180, https://doi.org/10.1007/s00382-019-04919-6, 2019.

Truong, C., Oudre, L., and Vayatis, N.: Selective Review of Offline Change Point Detection Methods, Signal Process., 167, 107 299, https://doi.org/10.1016/j.sigpro.2019.107299, 2020.

Wandel, J., Quinting, J. F., and Grams, C. M.: Toward a Systematic Evaluation of Warm Conveyor Belts in Numerical Weather Prediction and Climate Models. Part II: Verification of Operational Reforecasts, Journal of the Atmospheric Sciences, 78, 3965–3982, https://doi.org/10.1175/JAS-D-20-0385.1, 2021.

Wandel, J., Büeler, D., Knippertz, P., Quinting, J. F., and Grams, C. M.: Why Moist Dynamic Processes Matter for the Sub-Seasonal Prediction of Atmospheric Blocking Over Europe, JGR Atmospheres, 129, e2023JD039 791, https://doi.org/10.1029/2023JD039791, 2024.

Wang, J., Chen, J., Zhang, H., Ma, R., and Chen, F.: Impacts of Multiscale Components of Initial Perturbations on Error Growth Characteristics and Ensemble Forecasting Skill, J. Appl. Meteorol. Climatol., 62, 1677–1692, https://doi.org/10.1175/JAMC-D-23-0108.1, 2023.

Wernli, H. and Gray, S. L.: The Importance of Diabatic Processes for the Dynamics of Synoptic-Scale Extratropical Weather Systems – a Review, Weather Clim. Dynam., 5, 1299–1408, https://doi.org/10.5194/wcd-5-1299-2024, 2024.

Wettstein, J. J. and Wallace, J. M.: Observed Patterns of Month-to-Month Storm-Track Variability and Their Relationship to the Background Flow*, Journal of the Atmospheric Sciences, 67, 1420–1437, https://doi.org/10.1175/2009JAS3194.1, 2010.

White, H.: A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, Econom. : J. Econom. Soc., 48, 817–838, 1980.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Elsevier, ISBN 978-0-12-815823-4, https://doi.org/10.1016/C2017-0-03921-6, 2019.

Wirth, V., Riemer, M., Chang, E. K. M., and Martius, O.: Rossby Wave Packets on the Midlatitude Waveguide—a Review, Mon. Weather Rev., 146, 1965–2001, https://doi.org/10.1175/MWR-D-16-0483.1, 2018.

Woollings, T. J.: Dynamical Influences on European Climate: An Uncertain Future, Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci., 368, 3733–3756, https://doi.org/10.1098/rsta.2010.0040, 2010.

World Meteorological Organization: Aircraft Meteorological Data Relay (AMDAR) Reference Manual, WMO Technical Document WMO-No. 958, World Meteorological Organization, Geneva, Switzerland, 2003.

Yoo, H., Li, Z., Hou, Y.-T., Lord, S., Weng, F., and Barker, H. W.: Diagnosis and Testing of Low-Level Cloud Parameterizations for the NCEP/GFS Model Using Satellite and Ground-Based Measurements, Clim. Dyn., 41, 1595–1613, https://doi.org/10.1007/s00382-013-1884-8, 2013.

Yu, Q., Spensberger, C., Magnusson, L., and Spengler, T.: Forecast Errors Attributed to Synoptic Features, Meteorol. Appl., 32, e70 093, https://doi.org/10.1002/met.70093, 2025.

Žagar, N.: A Global Perspective of the Limits of Prediction Skill of NWP Models, Tellus A: Dyn. Meteorol. Oceanogr., 69, 1317 573, https://doi.org/10.1080/16000870.2017.1317573, 2017.

Žagar, N., Tribbia, J., Anderson, J. L., and Raeder, K.: Balance of the Background-Error Variances in the Ensemble Assimilation System DART/CAM, Mon. Weather Rev., 139, 2061–2079, https://doi.org/10.1175/2011MWR3477.1, 2011.

Zeileis, A.: Object-Oriented Computation of Sandwich Estimators, J. Stat. Softw., 16, 1–16, https://doi.org/10.18637/jss.v016.i09, 2006.

Zeileis, A., Leisch, F., Hornik, K., and Kleiber, C.: **Strucchange** : An *R* Package for Testing for Structural Change in Linear Regression Models, J. Stat. Softw., 7, https://doi.org/10.18637/jss.v007.i02, 2002.

Zeileis, A., Kleiber, C., Krämer, W., and Hornik, K.: Testing and Dating of Structural Changes in Practice, Comput. Stat. Data Anal., 44, 109–123, https://doi.org/10.1016/S0167-9473(03)00030-6, 2003.

Zeileis, A., Köll, S., and Graham, N.: Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R, J. Stat. Softw., 95, https://doi.org/10.18637/JSS.V095.I01, 2020.

Zhang, F., Bei, N., Rotunno, R., Snyder, C., and Epifanio, C. C.: Mesoscale Predictability of Moist Baroclinic Waves: Convection-Permitting Experiments and Multistage Error Growth Dynamics, J. Atmos. Sci., 64, 3579–3594, https://doi.org/10.1175/JAS4028.1, 2007.