RESEARCH ARTICLE

Quarterly Journal of the Royal Meteorological Society

Exceptionally poor and good medium-range forecasts of the large-scale circulation over Europe in ERA5 reforecasts

Seraphine Hauser[1,2,3] | Steven M. Cavallo[1*] | Linus Magnusson[4,5*] | Jonathan E. Martin[2*] | David B. Parsons[1*]

[1]School of Meteorology, University of Oklahoma, Norman, Oklahoma, USA

[2]Department of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, Madison, WI, USA

[3]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland

[4]European Centre for Medium-Range Weather Forecasts, Reading, UK

[5]Department of Meteorology, Stockholm University, Stockholm, Sweden

Correspondence
Seraphine Hauser, ETH Zurich, Department of Environmental Systems Science, Institute for Atmospheric and Climate Science (IAC), Universitätstrasse 16, 8092 Zürich, Switzerland
Email: seraphine.hauser@env.ethz.ch

Funding information
Work of SH was partially funded by the Office of Naval Research (Award Number N00014-18-1-2163) and the Swiss National Science Foundation (project 228019).

KEYWORDS
forecast busts, forecast dropout, predictability, large-scale dynamics, medium-range forecasts, ERA5 reforecasts, weather regimes, regime transition

Abbreviations: ACC, anomaly correlation coefficient; AR, Atlantic ridge; AT, Atlantic trough; CAPE, Convective available potential energy; ECMWF, European Centre for Medium-Range Weather Forecasts; EuBL, European blocking; GL; Greenland blocking; RMSE, root mean squared error; ScBL, Scandinavian blocking; ScTr, Scandinavian trough; Z500, geopotential height at 500 hPa; WRI, Weather regime index

[*]Equally contributing authors.

Abstract

Despite continuous improvements in weather forecasting, large-scale forecast busts—sudden drops in accuracy—still occur. In this study, we extend the concept of busts to define 'exceptionally poor forecasts' and introduce the notion of 'exceptionally good forecasts,' both derived using a methodology that accounts for seasonality in forecast skill. We apply this framework to 6-day forecasts over Europe in ERA5 reforecasts (1979–2023) to identify and compare their characteristics. The analysis explores potential links between these forecasts and large-scale weather regimes in the North Atlantic–European region, with particular attention to the occurrence and timing of regime transitions. We identify a declining trend in the annual rate of poor forecasts and an increasing trend in the rate of good forecasts, consistent with advances in the number and quality of observations. Poor forecasts occur more often in the warm season and good forecasts are found throughout the year, and their mean patterns contrast sharply: Rossby wave trains characterize poor forecasts, whereas blocking over northern Europe dominates good forecasts. Periods of poor forecast performance coincide with an above-average frequency of cyclonic regimes and persistent no-regime periods. Conversely, good forecasts show an above-average association with anticyclonic regimes, particularly with Scandinavian blocking. Whereas the share of cases with a regime transition is similar in both skill categories (60%), transitions occur significantly later during poor forecasts and earlier in good forecasts, providing a so-called 'window of opportunity' when initialized early in a regime life cycle. If regime transitions during poor forecasts occur early, the errors are not necessarily linked to wrong regime predictions, suggesting a fairly correct representation of the large-scale circulation while synoptic-scale systems may drive large errors restricted to Europe. In summary, our study contributes to understanding the large-scale circulation configurations and stages of regime evolution that favour exceptionally poor or good forecasts over Europe.

1 | INTRODUCTION

Over the last few decades, numerical weather prediction has steadily advanced in what Bauer et al. (2015) describes as a "quiet revolution", leading to more accurate weather forecasts – marked by a one-day-per-decade increase in forecast skill (Bauer et al., 2015, their Figure 1). This improvement in forecast accuracy is attributable to several factors, including an expanded observational network, improvements in the representation of model physics, and enhancements in data assimilation techniques (e.g., Magnusson and Källén, 2013; Bauer et al., 2015). Despite this overall progress, forecast skill remains variable and can fluctuate significantly on a day-to-day basis, due to the inherently chaotic nature of atmospheric flow (Lorenz, 1963). While subseasonal-to-seasonal forecast skill can benefit from broader sources of predictability—such as teleconnections and land–atmosphere coupling (e.g., Vitart, 2017)—medium-range forecasts (3–14 days lead time) are primarily governed by synoptic-scale dynamics, accuracy of initial conditions and quality of model physics. This implies that predictive skill can be lost through the non-linear amplification of small errors, model limitations, and insufficient or misused observational data (Palmer, 1999).

Occasionally, forecast performance deteriorates abruptly and substantially, resulting in dramatically incorrect predictions—a phenomenon referred to as a 'forecast bust' or 'dropout' (Rodwell et al., 2013). These events are defined by large deviations between predicted and observed atmospheric states, often occurring within well-observed regions and at lead times where forecast skill is typically high. While the term 'forecast bust' can be applied across scales, it is primarily used to describe large-scale events, such as those affecting the entire European region at medium-range forecast lead times (Rodwell et al., 2013). Forecast busts underscore the inherent limits to atmospheric predictability and are especially concerning due to their potential socioeconomic impacts (e.g., Magnusson, 2017). The fact that such busts often occur across multiple forecast systems simultaneously indicates that some events may be charac-

terized by intrinsic unpredictability rather than system-specific shortcomings (e.g., Rodwell et al., 2013; McLay and Satterfield, 2022). Several past bust events have resulted in substantial surface weather errors, revealing the influence of upper-tropospheric flow on surface conditions (e.g., Magnusson, 2017; Grams et al., 2018). In some instances, these busts led to large 2-meter temperature errors over European sub-regions, highlighting the need for a deeper understanding of the mechanisms driving forecast busts.

Characteristic initial-condition patterns associated with forecast busts over Europe include a trough over the Rockies and a high-pressure system over Canada, often accompanied by positive CAPE anomalies to the east (Rodwell et al., 2013). Subsequent research has linked such setups to mesoscale convective systems (MCSs) over North America that can modify the downstream upper-level flow and reduce European predictability several days later (e.g., Parsons et al., 2019; Lojko et al., 2022). Error-sensitivity analyses further point to key regions such as the tropical eastern Pacific, western/central Canada, and the western Atlantic as important for bust development (Magnusson, 2017). Consistent with this, feature-based diagnostics show that forecast errors are increase in the presence of weather systems, likely due to diabatic heating errors associated with latent heat release (Grams et al., 2018; Wandel et al., 2024; Yu et al., 2025). Recurving tropical cyclones in the North Atlantic during the autumn bust peak also emerge as important triggers, as their extratropical transitions can strongly perturb the midlatitude jet and downstream circulation (Lillo and Parsons, 2017; Keller et al., 2019; Brannan and Chagnon, 2020).

Medium-range forecast skill fluctuations partly stem from variations in the intrinsic predictability of the atmosphere, with certain flow regimes offering larger predictability than others (e.g., Ferranti et al., 2015; Matsueda and Palmer, 2018). Regime-dependent forecast skill horizon is approximately 3–5 days longer in winter compared to the other seasons (Büeler et al., 2021), with the two North Atlantic Oscillation (NAO) phases having the longest skill horizon, particularly in winter (e.g., Ferranti et al., 2018; Matsueda and Palmer, 2018). Anticyclonic regimes with blocking over Europe are generally less predictable, especially in spring and summer; an exception is Scandinavian blocking, which, is the most predictable regime during summer on the medium-range timescale (Büeler et al., 2021). Predictability is notably lower in situations in the absence of a regime, suggesting that transient, non-persistent flow patterns are particularly challenging to forecast (e.g. Osman et al., 2023). Transitions between regimes also pose significant difficulties, especially the onset of blocked regimes (e.g., Ferranti et al., 2015; Wandel et al., 2024). Forecast busts have been linked to the initiation and amplification of Rossby wave activity over the Atlantic, leading to large-scale circulation changes and, in some cases, missed onsets of blocked regimes over Europe (e.g., Lillo and Parsons, 2017; Magnusson, 2017; Grams et al., 2018; Hauser et al., 2023). However, the connection between busts and regime transitions remains insufficiently explored.

This study presents a substantially updated and extended characterization of 6-day forecast busts over Europe and provides new insights into the role of large-scale circulation patterns and regime transitions. We revise the original definition of busts to account for the seasonality of skill measures, enabling the detection of forecasts that perform anomalously poorly for their season; these are referred to in this study as 'exceptionally poor forecasts'. For the first time, these forecasts are compared to their counterpart—exceptionally good forecasts—allowing a detailed comparison of the characteristics of forecasts at the two extremes of the skill distribution. As all previous systematic studies (e.g., Rodwell et al., 2013; Lillo and Parsons, 2017) relied on ERA-Interim reforecasts (Dee et al., 2011), which have since become outdated, this study uses ERA5 reforecasts from ECMWF (Hersbach et al., 2020) for the period from 1979 to 2023. Despite being based on an older model cycle (41r2), compared to the current operational cycle (49r1), the value of the dataset lies in its 45-year consistency, which allows for a more robust analysis of large-scale atmospheric variability. Large-scale circulation changes are analysed within a weather regime framework based on the year-round North Atlantic–European classification of Grams et al. (2017), whose regimes are physically meaningful (Hochman et al., 2021) and widely used in dynamical and predictability studies across various time scales, and practi-

cal applications such as in the energy sector (e.g., Büeler et al., 2021; Osman et al., 2023; Teubler et al., 2023; Hauser et al., 2023; Mockert et al., 2023). Specifically, this study addresses the following research questions:

- How do exceptionally poor forecasts, identified using a seasonally adjusted method based on ERA5, compared with busts derived from ERA-Interim, and how do these poor forecasts differ from exceptionally good forecasts?
- Under which large-scale circulation regimes are exceptionally poor and good forecasts over Europe initialized, and which regimes do the models struggle to predict at forecast day 6?
- Do regime transitions occur within the 6-day period of the exceptional forecasts, and is there a systematic difference in this evolution between poor and good forecasts?

The paper is organized as follows: Section 2 introduces the datasets and methodology. Section 3 presents the results. The study concludes with a summary and final remarks in Section 4.

2 | DATA AND METHODS

2.1 | ERA5 reforecast and reanalysis datasets

The analyses in this study are based on reforecasts—also known as hindcasts—which are retrospectively generated forecasts produced using a fixed model version (e.g., Hamill et al., 2013). Specifically, we utilize deterministic 10-day control reforecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) produced with the ERA5 model (Hersbach et al., 2020), with forecast start dates ranging from 1 January 1979 00 UTC to 31 December 2023 12 UTC. While deterministic forecasts do not provide explicit representations of forecast uncertainty like ensemble systems, our dataset—consisting of twice-daily deterministic forecasts (00 UTC and 12 UTC) spanning 45 years (32,850 forecasts)—offers a uniquely dense and long-term record that enables robust statistical characterization of forecast skill variability. This extensive temporal coverage surpasses that of ensemble forecasts using a fixed model version and thus provides greater opportunity to sample a wide range of atmospheric conditions and rare events. As a reference, ECMWF's IFS sub-seasonal ensemble reforecasts (Vitart et al., 2017) span the past 20 years with forecasts initialized twice a week and 11 ensemble members, yielding around 22,280 forecasts in total. However, because ensemble members share the same initialization date, the forecasts are not independent and represent far fewer distinct large-scale circulation patterns than the raw total suggests.

All ERA5 reforecasts within this period were produced using cycle 41r2 of the Integrated Forecasting System (IFS), which was operational from March to November 2016, with a global horizontal resolution of 36 km. The dataset spans the Northern Hemisphere and is available on a 1° × 1° latitude–longitude grid. The temporal resolution of the 10-day forecasts varies with lead time from 3-hourly intervals up to the 12-hour forecast, 6-hourly intervals up to the day 5 forecast, and 12-hourly intervals from day 5 to day 10 forecasts (i.e., up to t = 240 h).

In addition to the reforecasts, the ERA5 reanalysis dataset (Hersbach et al., 2020) is used for certain analyses and for verification for the period 1 January 1979 to 10 January 2024, since the final forecast in this dataset (31 December 2023, 12 UTC) extends through to 10 January 2024.

2.2 | Forecast skill measures

Rodwell et al. (2013) developed the first systematic dataset of forecast busts over Europe, based on ERA-Interim

reanalysis data. The Rodwell et al. (2013) definition of a bust was based on two criteria: the anomaly correlation coefficient (ACC) and the root mean square error (RMSE) for geopotential height at 500 hPa (Z500) over Europe. Both skill measures reflect the accuracy of large-scale circulation forecasts and are evaluated at forecast day 6 over a European domain (35 °–75 °N, 12.5 °W–42.5 °E). Due to the 1 ° spatial resolution of the dataset used here, we slightly adjusted their European box to 35°–75°N and 13°W–43°E. We tested several box sizes and found that variations of 1–2 ° do not significantly affect forecast skill.

In line with Rodwell et al. (2013), ACC is used as a metric for forecast performance, capturing the spatial correlation between the forecast and the analysis while accounting for the underlying climatology. Specifically, we use the centred ACC calculation (e.g., Wilks, 2020), defined as

$$ACC = \frac{\frac{1}{N}\sum_{n=1}^{N}(f'_n - \overline{f'})(a'_n - \overline{a'})}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}(f'_n - \overline{f'})^2 \frac{1}{N}\sum_{n=1}^{N}(a'_n - \overline{a'})^2}}, \tag{1}$$

where the index n runs over all N latitude–longitude grid points within the European domain. Forecast anomalies (f') and analysis anomalies (a') are calculated by subtracting a 30-day centred running-mean climatology—based on ERA5 reanalysis data from 1979 to 2023—from the respective absolute fields. Using a reanalysis-based climatology to derive forecast anomalies is justified at medium-range lead times: the model exhibits only a small seasonal bias over Europe, slightly negative in winter (5 gpm), increasing in spring, peaking in summer (+5–10 gpm), and decreasing again in autumn (cf. Figure S1 in the supplementary material). Bias correction has only a modest effect on RMSE (median reduction 1.5 %) and often slightly reduces ACC, indicating that forecast-to-forecast variability and pattern errors dominate skill at day 6, and that the model climate remains consistent with the reanalysis. The spatial mean of the anomalies over Europe ($\overline{f'}$, $\overline{a'}$) is removed to enforce zero-mean anomalies, as required for the centred ACC calculation.

The root mean square error (RMSE), the second criterion used by Rodwell et al. (2013) to define forecast busts, is given by

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(f_n - a_n)^2}, \tag{2}$$

with the absolute forecast (f) and analysis (a) Z500 fields over Europe. Note that the data points are weighted by the cosine of the latitude for the calculation of both metrics.

2.3 | Year-round weather regimes in the North Atlantic–European region

The Z500-based year-round weather regime classification in the North Atlantic-European region by Grams et al. (2017) is used. The original definition was based on ERA-Interim (Dee et al., 2011) but has since been applied to ERA5 reanalysis data (Hersbach et al., 2020) and used in several recent studies (e.g., Hauser et al., 2024; Lemburg and Fink, 2024). Weather regimes are detected using six-hourly low-pass filtered and normalized Z500 anomalies in the period 1979–2019 over the North Atlantic-European region (80°W–40°E, 30–90°N). After performing an empirical orthogonal function (EOF) analysis, k-means clustering is applied to the seven leading EOFs which explain 74.4% of the variability. This clustering analysis yields seven weather regimes with three cyclonic (Zonal regime, Scandinavian trough, Atlantic trough) and four anticyclonic regime types (Atlantic ridge, European blocking, Scandinavian blocking, Greenland blocking). The mean patterns of the Z500 regimes are available in the supplementary material (Figure S2).

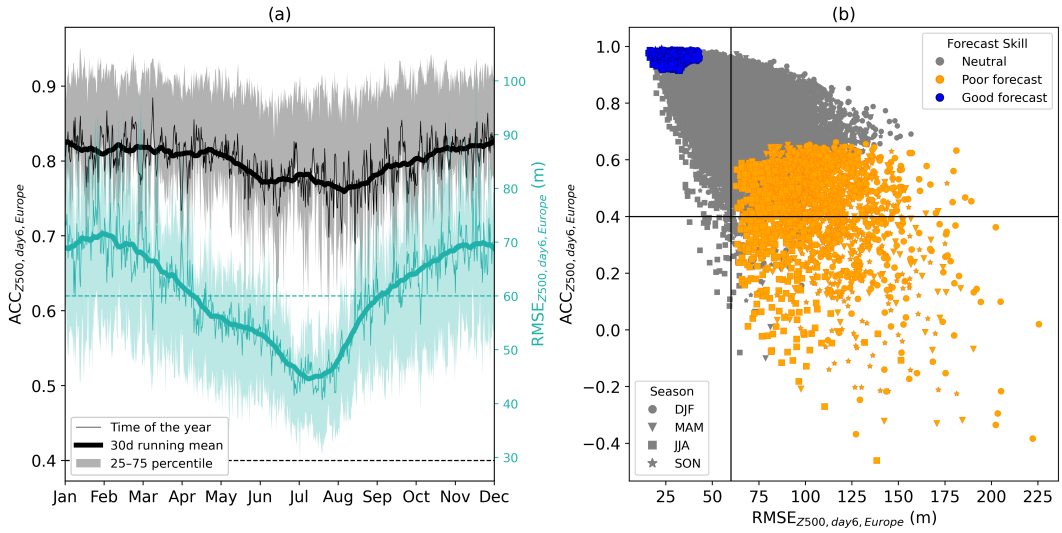

FIGURE 1 (a) Seasonal variations in skill measures. Evolution of the Z500-based ACC over Europe at forecast day 6(left y-axis, black) and Z500-based RMSE over Europe at forecast day 6 (in m, right y-axis, turquoise) depending on the time of the year. Thin lines represent means based on 12-hourly data, bold lines show the centred 30-day running means, and shading corresponds to the 25–75th percentile. (b) Scatter plot of ACC and RMSE of Z500 over Europe at forecast day 6 for all ERA5 reforecasts. Colours indicate the forecast category: orange for poor forecasts, blue for good forecasts, and grey for neutral cases. Marker types denote the season of each forecast. The straight vertical and horizontal lines in both panels represent the Rodwell et al. (2013) thresholds used to identify busts in ERA-Interim.

In this study, we use the weather regime perspective to characterise the large-scale circulation pattern for a given date by assigning one of the seven or no regime to it. For this purpose, we use the weather regime index (WRI) by Michel and Rivière (2011) and Grams et al. (2017). It measures the projection (dot product) of Z500 anomaly fields for a given time onto a fixed weather regime pattern (centroid of a cluster), and then standardizes it over time. Consequently, the WRI describes how strongly a specific large-scale pattern resembles a specific weather regime, expressed in standard deviations from an average projection. For physically meaningful regime periods, we follow the life cycle definition of Grams et al. (2017). A regime life cycle is detected when the following conditions are met: (1) The WRI needs to be equal or exceed 1.0 for consecutive time steps for a minimum duration of 5 days and (2) the regime must have the highest WRI out of all seven regimes for at least one time step within its life time. More details and further criteria for rare cases are documented in Hauser et al. (2024). Using regime life cycles rather than just the regime with the highest WRI for regime assignment leads to an additional category: the so-called no-regime category. This category includes periods when the large-scale flow pattern does not closely resemble any of the seven regimes or lacks sufficient persistence.

From the reanalysis perspective, the WRI is computed using low-pass–filtered 3-hourly Z500 anomaly fields spanning January 1979 to January 2024. The forecast perspective, in contrast, uses instantaneous 12-hourly fields to accommodate 10-day reforecasts and prevent data loss at the edges. Verification is performed by comparing instantaneous WRI projections from the forecasts with their reanalysis counterparts.

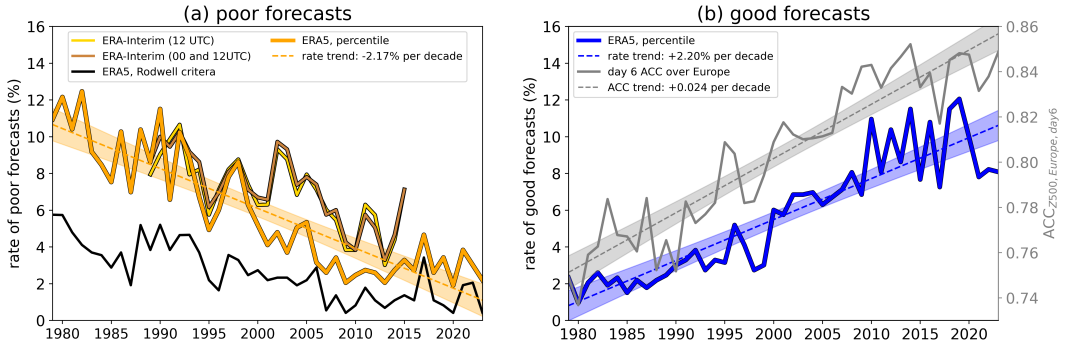

FIGURE 2 (a) Occurrence rates of poor forecasts (relative to all forecasts within a year, in %) for multiple database set-ups and thresholds: ERA-Interim based using the Rodwell et al. (2013) criteria (ACC < 40 % and RMSE > 60 m) for once daily initialized forecasts (gold, solid) and twice daily initialized (orange brown, solid), ERA5-based using the Rodwell et al. (2013) criteria (black, solid), the final selected percentile thresholds criteria (orange, solid) and the trend as determined by linear regression (orange, dashed). (b) Occurrence rates of good forecasts (left y-axis) using the percentile threshold (blue, solid) and its trend (blue, dashed). The annual mean ACC over Europe at day 6 (grey, solid) and the trend (grey, dashed) are displayed on the right y-axis. The shaded area around the linear regression lines (dashed) correspond to the 95 % confidence interval. The trends in the annual frequencies of poor and good forecasts, as well as in ACC, are statistically significant and remain robust after accounting for serial autocorrelation

3 | RESULTS

3.1 | Revision and extension of the forecast bust definition

The most established definition of forecast busts in the large-scale atmospheric circulation over Europe is that of Rodwell et al. (2013), who defined a bust as a Z500 forecast at day 6 with an anomaly correlation coefficient (ACC) below 0.4 and a root mean square error (RMSE) above 60 m within the European domain. This definition has been adopted in subsequent studies (e.g., Lillo and Parsons, 2017). Applied to ERA-Interim reforecasts over the period 1989–2015, 7.2% of all reforecasts were classified as forecast busts (Rodwell et al., 2013). Using the same criteria, we analysed ERA5 reforecasts and identified 2.6% of all reforecasts as busts for the period 1979–2023, and 2.4% for the period 1989–2015, for consistency with Rodwell et al. (2013). This indicates a significantly lower rate of forecast busts in ERA5, likely due to substantial improvements in the forecasting system (i.e. model development, data assimilation, and observation usage) with the transition from IFS cycle 31r2 (ERA-Interim) to cycle 41r2 (ERA5).

The standard definition of busts by Rodwell et al. (2013) relies on fixed thresholds for the full year and might therefore neglect potential seasonality in forecast performance. Figure 1a shows the seasonal cycle of RMSE and ACC for Z500 at day 6 over Europe, based on ERA5 reforecasts. First, the RMSE of Z500 is biased by the mean seasonal cycle, with lower errors during summer and higher errors during winter (turquoise line). In contrast, the ACC measures pattern correlation and is less sensitive to the seasonal mean (black line). Nevertheless, it also displays a seasonal cycle, reflecting general seasonal predictability with lower skill in summer and higher skill in winter. The interquartile range (shading in Figure 1a) highlights the skewness in the distributions of RMSE and ACC at day 6 over Europe: ACC is strongly negatively skewed, while RMSE is strongly positively skewed. A RMSE threshold of 60 m captures extreme events well in summer but is less suitable in winter, where most forecasts exceed this threshold (dashed turquoise line). For ACC, the threshold of 0.4 lies well outside the interquartile range, making it more appropriate for identifying

exceptional events (grey dashed line and grey shading). Therefore, defining busts using fixed, year-round thresholds introduces a seasonal bias, causing bust frequency to reflect seasonal ACC variability rather than true performance declines.

To better account for this seasonality, we explore alternative approaches and investigate four objective methods that explicitly incorporate seasonal effects. The first method follows Yamagami and Matsueda (2021), who defined busts over the Arctic as cases whereACC at day 6 falls below the monthly 10th percentile and RMSE at day 6 exceeds the monthly 90th percentile, based on month-specific climatological distributions. The second approach involves standardizing ACC and RMSE to remove seasonal means and variances, using centred 30-day running mean climatologies and standard deviations (1979–2023). These standardized values are combined into a single composite index, defined as $\mathrm{CI} = -\mathrm{ACC}_{\mathrm{standardized}} + \mathrm{RMSE}_{\mathrm{standardized}}$. Busts are then defined as dates for which the CI exceeds the 90th percentile of all CI values. In the third approach, rather than combining the standardized ACC and RMSE values, we determine percentile thresholds for each variable independently. Busts are defined as dates on which $\mathrm{ACC}_{\mathrm{standardized}}$ falls below the 10th percentile and $\mathrm{RMSE}_{\mathrm{standardized}}$ exceeds the 90th percentile. The fourth method uses empirical percentile scores and actual ACC and RMSE values are compared to their distributions within a centred 30-day climatological window (1979–2023). A forecast is classified as a bust if the ACC percentile score is below 3% and the RMSE percentile score is above 97%. For the latter three approaches, the percentile thresholds were chosen to yield a number of busts comparable to those obtained with the established Yamagami and Matsueda (2021) method. Additionally, we adjusted the original, fixed thresholds of Rodwell et al. (2013) to achieve a similar total number of identified busts, resulting in a choice of 0.5 for ACC and 45 m for RMSE.

For this study, we chose to base our analysis on the third approach, defining busts from deseasonalized, standardized ACC and RMSE values using separate percentile thresholds for each metric. Specifically, the thresholds are set at the 10th percentile for ACC and the 90th percentile for RMSE. This method identifies events that are exceptional relative to the typical forecast performance for that time of year, requiring both unusually low ACC and unusually high RMSE. Compared to the other methods, the resulting pool of bust dates is robust: most bust dates (73–95%) identified by the alternative approaches are also captured by this final definition. Using this method, we identify 1,934 busts, corresponding to 5.9% of all reforecasts. Figure 1b presents a scatter plot of all absolute ACC and RMSE values at day 6 from the reforecasts, with bust cases highlighted in orange. The plot shows that nearly all busts defined by Rodwell et al. (2013) are still captured under the new definition. The few excluded cases likely fail to simultaneously meet both the low-ACC and high-RMSE values or are not exceptional when accounting for seasonal forecast performance. Notably, many events with ACC values above 0.4 are now classified as busts—an important refinement that reflects both the seasonality of forecast skill and the overall improvement of ERA5 over ERA-Interim.

Extending beyond previous studies that focused solely on busts (e.g., Rodwell et al., 2013; Lillo and Parsons, 2017; Yamagami and Matsueda, 2021), we also analyse the counterpart of busts, namely exceptionally good forecasts. Periods of above-average predictability (so-called 'windows of opportunity') have been well studied on sub-seasonal to seasonal timescales (e.g., Mariotti et al., 2020), but remain less explored for the medium range. We here define good forecasts as those with deseasonalized, standardized ACC values above the 90th percentile and RMSE values below the 10th percentile. These events are highlighted in blue in Figure 1b and occupy a small region in the scatter plot, reflecting the concentration of good forecast skill in the highly skewed skill measure distributions. To distinguish our approach from previous definitions and to enable comparison with exceptionally good forecasts, we hereafter refer to busts as exceptionally poor forecasts.

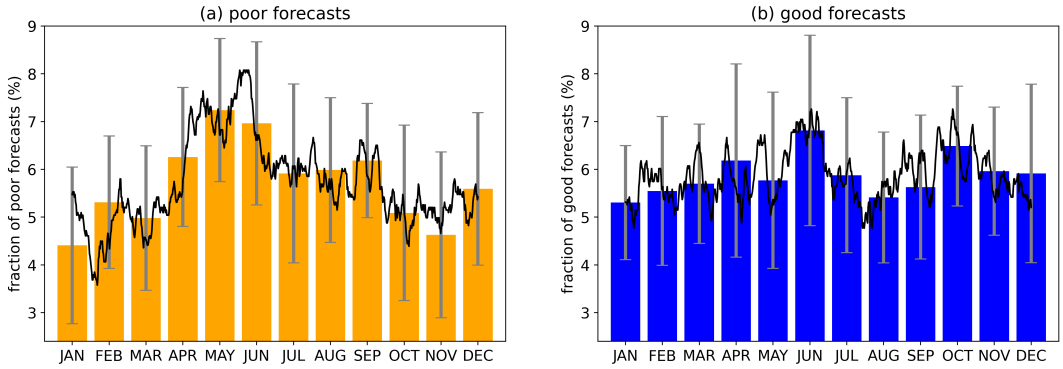

FIGURE 3 Seasonal distribution of the fraction of exceptionally poor and good forecasts (in %). The bars show the fraction of poor (panel a) and good (panel b) forecasts, respectively, aggregated over full months. The thin black line represents the smoothed fraction at higher temporal resolution (originally 12 h, smoothed using a 20-day moving window). Overlaid bars indicate the fraction of poor and good forecasts (in %) after removing consecutive forecasts. The 95% confidence intervals are shown in grey.

3.2 | Occurrence rate and seasonality

The frequency of poor forecasts over Europe at forecast day 6 has steadily declined with time (Figure 2a). Starting with a mean rate of around 11% in 1979 that declined to around 3% in 2023, a trend has been detected with a rate of −2.17% per decade over the analysis period. The negative trend likely corresponds closely with the increase in forecast accuracy, shown by the increase in annual mean ACC at day 6 (grey line in Figure 2b), which has risen by 0.024 per decade. The decreasing number of poor forecasts corresponds well with the results of Rodwell et al. (2013) and Lillo and Parsons (2017), who found a decreasing frequency for their shorter time periods using ERA-Interim reforecasts and the absolute thresholds for RMSE and ACC. The annual rate based on their bust definition for a 22-year period of twice daily ERA-Interim reforecasts is illustrated in Figure 2a for once-daily (yellow) and twice-daily (orange) initial times. Both curves reveal a very similar evolution of poor forecast rate, hence demonstrating that forecast skill is largely independent of the time of day of initialization. While the evolution of poor forecast rates before the year 2000 is very comparable between the different datasets and methods, the rate afterwards diverges between the datasets with a stronger decline in frequency for the ERA5-based reforecasts. One reason for this is the increased assimilation of satellite-derived data that began around that time, together with more advanced satellite data assimilation methods in ERA5 compared to ERA-Interim (cf. Hersbach et al., 2020).

In contrast to exceptionally poor forecasts, we found an increasing rate of good forecasts over Europe at day 6 in ERA5 reforecasts (Figure 2b, blue curve). Despite inter-annual variability, we identify a statistically significant increase in their rate, with a trend of +2.20% per decade. Again, around the year 2000, the sudden increase in forecast accuracy is reflected by a marked jump in in good forecast rate from 2% to 6%.

Using standardized skill measures that consider the forecast accuracy at the respective time of the year dampens the seasonality of poor and good forecasts compared to the year-round threshold approach by Rodwell et al. (2013). However, the frequency can still show seasonality, because the variability and tail behaviour of forecast errors is again seasonally dependent. Some seasons naturally produce more or fewer extreme anomalies, which affects the likelihood of both exceptionally good and poor forecasts despite standardized skill measures. The monthly occurrence of exceptional forecasts is illustrated in Figure 3 and reveals an uneven frequency of both, poor and good forecasts.

This unevenness should be interpreted with caution, as the 95% confidence intervals for monthly bust frequencies are broad and overlapping (Figure 3a,b,; grey), indicating high uncertainty in the estimates. Such wide intervals suggest that the apparent seasonal patterns may not be statistically significant. The rate of poor forecasts is largest from April to September with a major peak from May to June (Figure 3a). As discussed in previous studies, a peak of poor forecasts in late spring/early summer may be linked to MCSs over North America, where MCS interactions with Rossby wave dynamics can influence downstream Rossby Wave Packet development (e.g., Grazzini and Isaksen, 2002; Rodwell et al., 2013; Parsons et al., 2019). During the cold season, the rates are lower indicating less exceptionally poor forecasts during winter. The seasonal signals found here deviate from the seasonality in the ERA-Interim bust dataset, in which 24% of the annual busts were identified in the months of September and October alone. Lillo and Parsons (2017) emphasized the prevalence of poleward-recurving tropical storms across the central North Atlantic in their bust cases, hence linked their peak in busts to the North Atlantic hurricane season. This peak is absent in our dataset, likely because the mean forecast accuracy is removed for each time of the year, which reduces the influence of seasonally lower skill during the North Atlantic hurricane season.

For exceptionally good forecasts, the seasonality is less pronounced compared to poor forecast (Figure 3b). The highest rates are found in June and from October to December. The fact that both exceptionally poor and good forecast show a peak rate in June indicates an increased variability with a broader forecast error distribution during that time of the year (cf. Figure 1a), making it more likely to observe a considerable number of both exceptionally poor and good forecasts in the same month. For the extended winter months, sources of enhanced skill (although more importantly for extended-range forecasts) include the Madden–Julian Oscillation, the El Niño–Southern Oscillation, and the Stratospheric Polar Vortex (e.g., Mariotti et al., 2020).

3.3 | Consecutive exceptional forecasts

Forecast busts can occur in successive episodes rather than as isolated events (e.g., Rodwell et al., 2013, their Figure 1). It is therefore of interest to examine whether exceptionally poor and good forecasts tend to occur as isolated or consecutive events, whether any changes in this behaviour have occurred over the 45-year period and whether this depends on the time of year.

We find that many exceptionally poor and good forecasts occur not isolated, but with a preceding or following poor and good forecast, respectively. Considering the full period (1979–2023), 58 % of all poor forecasts and 47 % of all good forecasts occur in a series of at least two consecutive poor or good forecasts. These numbers rapidly rise when we consider a period of +/- 3 days around each exceptional forecast; then 79 % of poor and 73 % of good forecasts occur in a sequence. Figure 4a shows the annual share of good and poor forecasts that exhibit a same-skill preceding or following forecast. The annual share is subject to a high inter-annual variability (thin lines in Figure 4a), but the 5-year means (thick lines with markers) show a more clear picture on the trends within the dataset. Consecutive poor forecasts dominate in earlier years (≈1979–2007), but we find a tendency of decreasing consecutive poor forecasts after this period (bold orange line in Figure 4a). This decrease corresponds to a shift from clustered toward more sporadic and isolated poor forecasts and coincides with gradual improvements in observations that may have reduced the likelihood of forecasts persisting in a poor state across consecutive runs. The frequency of consecutive poor forecasts strongly fluctuates towards the end of the reforecast period (≈2015–2023), suggesting that these fluctuations may reflect enhanced inter-annual climate variability. Good forecasts, in contrast, increasingly occur as consecutive events, indicating a more stable forecast performance in the medium range, that is, the system is able to maintain high forecast skill over multiple days (bold blue line in Figure 4a). This suggests the presence of windows of opportunity.

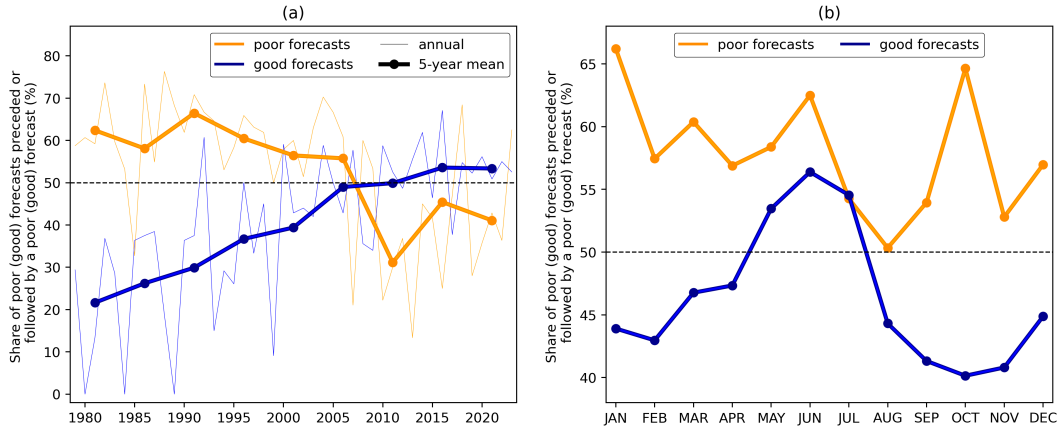

FIGURE 4 (a) Evolution of the frequency of consecutive, exceptional forecasts. Share of poor (good) forecasts preceded or followed by a poor (good) forecast (in %) as a function of year. Orange lines point to poor forecasts, blue lines to good forecasts. Thin lines represent the annual share, thick lines with bullets indicate the 5-year mean. (b) Seasonality of consecutive, exceptional forecasts. Share of poor (good) forecasts preceded or followed by a poor (good) forecast (in %) as a function of the month of the year.

Seasonal differences are evident when comparing the occurrence of consecutive poor and good forecasts over the course of the year (Figure 4b). There is high variability with several peaks for consecutive poor forecasts, with preferred occurrences in January, June and October (orange line). A minimum is found from July to September. This seasonality pattern of only consecutive poor forecasts shows close similarities with the seasonality of all ERA-Interim based busts of Rodwell et al. (2013); Lillo and Parsons (2017). Although our novel identification of poor forecasts considers the seasonality in skill measures and therefore significantly reduces seasonality on poor forecast frequency, the seasonality re-emerges when considering consecutive busts. This suggests that while our normalization accounts for seasonal variations in baseline forecast skill, it does not remove the underlying dynamical processes that govern consecutive poor forecasts. As a result, the probability of extended periods of poor forecast performance still peaks in the same months (January, June, and October) identified in the ERA-Interim study. In contrast to poor forecasts, consecutive good forecasts show a clear peak from spring to summer (blue line in Figure 4b), suggesting that medium-range windows of opportunity are more likely during the warm season.

3.4 | Spatial mean patterns

We examine year-round patterns of exceptionally poor forecasts at initial time and day 6 in the verifying analysis, comparing them with the ERA-Interim busts of Rodwell et al. (2013). Consecutive events are retained in the following analyses, consistent with previous studies. Figure 5 (top row) shows Northern Hemisphere composites of Z500 and CAPE, highlighting large-scale patterns extending well upstream and downstream of Europe. At day 0, poor forecasts feature a pronounced Rossby wave train spanning the western Northern Hemisphere, from the dateline to Europe (Figure 5a), in agreement with the ERA-Interim composite (cf. Rodwell et al., 2013, their Figure 4a). The 'Rockies trough' is again a robust feature, while positive Z500 anomalies dominate Western and Southern Europe and negative anomalies appear over Iceland and Scandinavia. However, three differences stand out: (1) the 'Canada high' is absent, (2) the eastern US ridge is stronger and more robust, and (3) the Scandinavian trough extends northwestward into

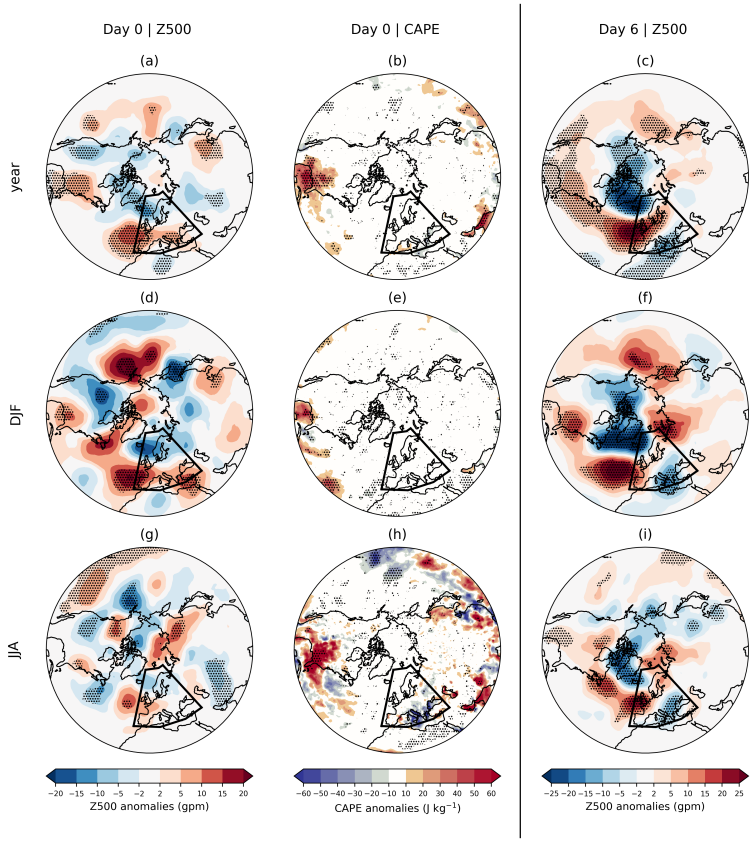

FIGURE 5 Mean patterns of exceptionally poor forecasts of Z500 at day 0 (left column), CAPE at day 0 (middle column), and Z500 at day 6 (right column) based on the verifying analysis. The rows show the composites year-round (upper row), Northern hemispheric winter (DJF, middle row), and Northern hemispheric summer (JJA; upper row). The black box marks the area used to determine the ACC and RMSE at day 6 (35°–75°N, 13°W–43°E). Dots represent grid points where the 95% confidence interval of the block-bootstrap distribution excludes zero, indicating robust mean anomalies against temporal sampling variability. Note that to remove long-term trends, we de-trend our fields by subtracting a linear least-squares fit at each grid point and for each calendar day, smoothed with a 20-day running mean (±10 days around the centred date).

Canada. CAPE anomalies (Figure 5b) reveal enhanced instability from the Gulf of Mexico to the southeastern US, consistent with the 'North American CAPE region' of Rodwell et al. (2013, their Figure 4b), suggesting MCSs may contribute to downstream forecast degradation (e.g., Grazzini and Isaksen, 2002). By day 6, the verifying analysis shows ridging from the eastern North Atlantic into Europe and negative Z500 anomalies over the Mediterranean, Iceland, and Greenland (Figure 5c). The upstream Pacific Rossby wave evident at day 0 disappears, and the main ridge is centred over the eastern North Atlantic rather than over the North Sea as in Rodwell et al. (2013, their Figure 3). Such discrepancies likely reflect the differences in bust/poor forecast definitions, reanalyses (ERA-Interim vs. ERA5), and analysis periods (1989–2010 vs. 1979–2023).

The year-round mean patterns of good forecasts differ substantially from those of poor forecasts (Figure 6, upper row). At initial time, the large-scale circulation over Europe is characterized by strong positive Z500 anomalies extending from eastern coast of North America to northeastern Europe, accompanied by a trough over the western Mediterranean (Figure 6a); both being robust features. Unlike in poor forecasts, no clear Rossby wave pattern is apparent upstream of Europe. The elevated CAPE over North America observed in poor forecasts is absent (Figure 6b), with some regions even exhibiting negative anomalies. These results suggest that enhanced convective instability over North America can act as a source of downstream forecast error, whereas its absence is associated with improved forecast skill over Europe. Negative CAPE anomalies prevail in the western Mediterranean and positive anomalies in the eastern Mediterranean, which indicates an inverse signal compared to poor forecasts. By day 6, the verifying analysis shows a circulation pattern very similar to that at initial time over Europe (Figure 6c), suggesting only minor structural evolution over the 6-day period. Strong positive Z500 anomalies continue to dominate high-latitude regions in Europe but are shifted towards Greenland and centred over Iceland, while negative anomalies appear more compact and prevail over western Europe.

While year-round composites highlight robust differences between good and poor forecasts, they can mask important seasonal variations in circulation and convective signals. Figures 5 and 6 show winter (DJF) and summer (JJA) composites; spring and autumn are in the supplementary material (Figures S3, S4). Seasonal composites reveal marked contrasts, particularly at initialization, reflecting high intra-composite variability. Poor forecasts show a positive Z500 anomaly over the western North Atlantic in winter but negative anomalies in summer, consistent with a stronger summer Rossby wave train (Figure 5d,g). The North American CAPE signal of poor forecasts appears in summer but is absent in winter, linking warm-season convection to poor European forecasts (Figure 5e,h). For good forecasts, winter shows a clear wave pattern spanning the Northern Hemisphere at day 0, which breaks down by day 6, while summer patterns are more stable, with persistent ridging over Scandinavia (Figure 6d,f,i). Summer composites at initialization exhibit weaker signals outside the North Atlantic–European domain, reflecting high internal variability (Figure 6g).

3.5 | Initial and forecast verification regime signatures from a reanalysis perspective

Given the seasonal variability of exceptionally poor and good forecasts, we examine large-scale flow at initialization (day 0) and verification (day 6) using the seven North Atlantic–European weather regimes of Grams et al. (2017). Figure 7 presents the year-round weather regime frequencies and relative weather regime frequency anomalies for poor and good forecasts, as represented in the reanalysis, with red colours corresponding to blocked regimes and blue colours to cyclonic regimes. Figures showing seasonal patterns are provided in the supplementary material (Figure S6), and significant seasonal deviations from the year-round picture are highlighted where relevant.

Both poor and good forecasts were initialized and verified during all seven regimes, as well as during periods without a dominant regime, referred to as the no regime category (Figure 7a). This aligns with the findings of Yamagami and Matsueda (2021), who showed that forecast busts over the Arctic can be initialized under all Arctic weather

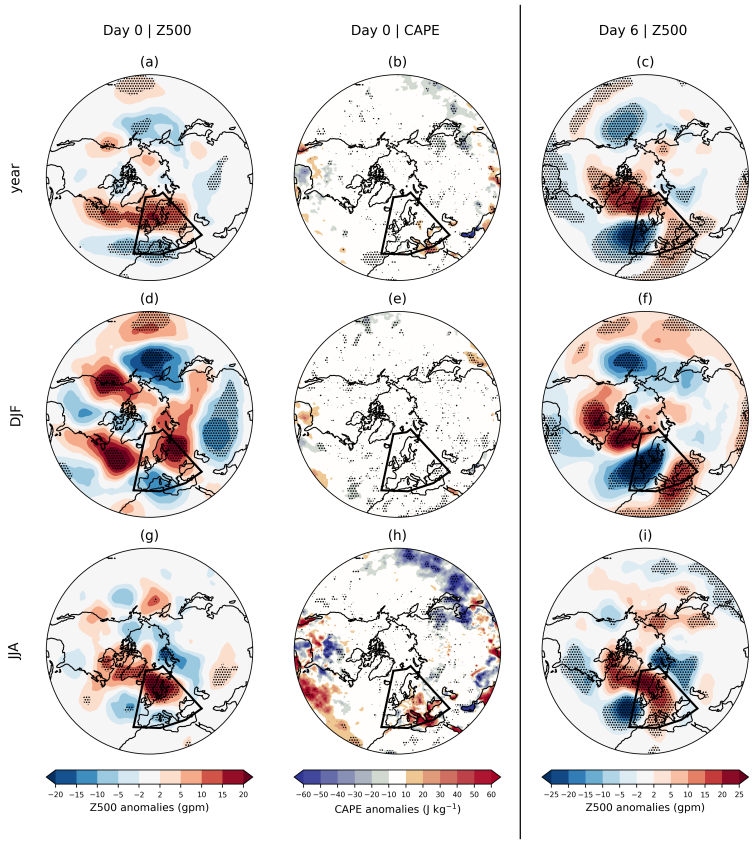

FIGURE 6 Same as Figure 5, but for exceptionally good forecasts.

regime patterns, rather than being limited to specific regimes. However, an anomaly-based perspective provides more distinct insights by accounting for the underlying climatological regime frequencies. Poor forecasts are unusually often initialized during the no regime and the two cyclonic regimes ZO and ScTr (Figure 7b, left, day 0). The increases are statistically significant for the no regime and ZO. Positive ZO frequency anomalies are evident in all seasons but are strongest in summer and autumn (Figure S6a). The increased frequency of the no regime aligns with Büeler et al. (2021); Osman et al. (2023), who attribute generally low skill to the no regime category of Grams et al. (2017), as the atmosphere is in a highly transient state at that time. Additionally, these authors note decreased forecast skill for ZO in summer, which is consistent with our results. Decreased relative frequency anomalies were found for AT and three out of four blocked regimes (AR, ScBL, GL) at day 0 for poor forecasts. At forecast validation time (Figure 7b, left, day 6), days are again more frequently assigned to the no regime category and the two cyclonic regimes (ZO and ScTr), with significant increases for the no regime and ScTr. Although not significant, a positive anomaly is detected for the blocked regime EuBL, most evident in spring (Figure S6b). This is consistent with a poorly forecasted event linked to the onset of EuBL in spring 2016, which was analysed in detail as one of the most significant forecast busts over Europe (Magnusson, 2017; Grams et al., 2018; Hauser et al., 2023). Large negative and significant frequency anomalies were found for AT, GL, and ScBL at validation time for poor forecasts (day 6), indicating the rare co-occurrence of these

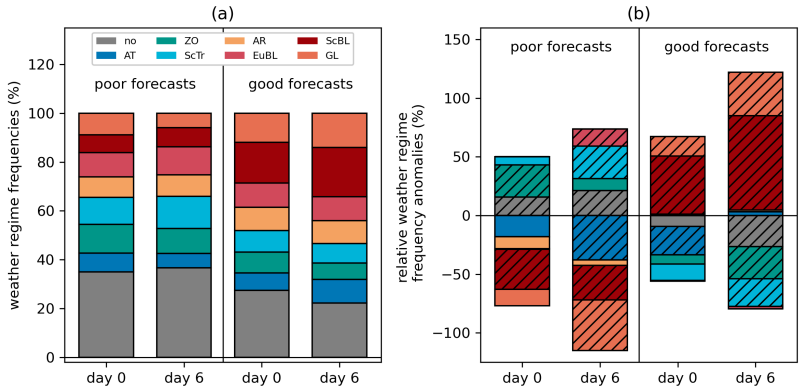

FIGURE 7 (a) Year-round weather regime frequencies (stacked, in %) for poor (left) and good (right) forecasts at initial time (day 0) and verification time (day 6), using the no re regime category (no), the three cyclonic regimes Atlantic trough (AT), Zonal regime (ZO), Scandinavian trough (ScTr), and the four anticyclonic regimes Atlantic ridge (AR), European blocking (EuBL), Scandinavian blocking (ScBL), and Greenland blocking (GL). (b) Year-round weather regime frequency anomalies relative to climatological frequency (in %). The climatological frequency of weather regimes is based on the period 1979–2023 and displayed in Figure S5 in the supplementary material. Masked bars point to statistically significant anomalies based on a bootstrapping, taking into account consecutive dates (N = 2000, < 1% or > 99%).

regimes with poor forecasts over Europe.

The signals of increased regime frequencies differ markedly for good forecasts. At day 0, the frequency of the two blocked regimes ScBL and GL is significantly increased, particularly for ScBL (Figure 7b, right, day 0). This suggests that good forecasts are preferentially initialized during active regime patterns, with a strong preference for certain blocked regimes. The increased GL frequency is dominated by summer events, while the dominant positive ScBL anomaly is evident and significant across all seasons but is largest and significant in winter and summer (Figure S6c). These results are consistent with the generally high skill of ScBL forecasts in summer and autumn (Büeler et al., 2021; Osman et al., 2023). The frequency of the remaining two blocked regimes (EuBL and AR) corresponds to their climatological occurrence. All cyclonic regimes and the no regime exhibit negative frequency anomalies for good forecasts at initial time, supporting the idea that good forecasts are less frequently initialized during cyclonic and highly transient circulation patterns. At forecast validation time (Figure 7b, right, day 6), the positive anomalies of GL and ScBL increase compared to day 0. Again, the positive ScBL anomaly is evident and this time the increased frequency of ScBL is significant across all seasons, particularly in summer when ScBL is climatologically the most frequent of the seven regimes (Figure S6d). Negative anomalies also grow for cyclonic regimes and the no regime, with significant reductions in frequency for ZO, ScTr, and the no regime. This further highlights that the model performs exceptionally well in predicting the two blocked regimes, a finding that may seem contradictory given the intrinsically low predictability and often sudden onset of blocking (e.g., Nakamura and Huang, 2018). We examine this apparent contradiction further below.

3.6 | Regime evolution during the forecast period for cases of poor and good forecasts

This and all remaining subsections of Section 3 primarily analyse how weather regimes behave during periods of exceptionally good and poor forecasts, as our goal is to explore variability *within* regimes rather than to attribute exceptional forecast skill to specific regime types. We first classified all periods based on the type of flow change to investigate the large-scale pattern evolution over 6-day periods from a reanalysis perspective: persistent regime (or persistent no-regime), regime onset (transition from no-regime to one of the seven regimes), regime decay (transition from one of the seven regimes to no-regime), and regime-to-regime transition (transition between two regimes). Across all forecasts between 1979 and 2023, 37% of the 6-day periods showed no regime change (persistent regime; Figure 8a, grey bar). The second most frequent case was regime-to-regime transitions, accounting for 28% of periods. Regime onsets and regime decays occurred at similar frequencies, with 17% and 18%, respectively.

Coloured bars in Figure 8a illustrate how the distribution of these categories differs between exceptionally poor and good forecasts compared to the full dataset. Persistence of the large-scale circulation is more common in good forecasts and slightly less common in poor forecasts, but the differences are not statistically significant (z-test for proportions). The share of regime onsets is comparable between poor and good forecasts and slightly higher than the climatological proportion, suggesting that regime onsets are somewhat more likely to be associated with exceptional forecasts. A clear difference emerges for regime decays: they occur significantly more often in poor forecasts than in good forecasts. The opposite holds for regime-to-regime transitions, which are significantly more frequent in good forecasts. This indicates that forecasts tend to perform better in situations with transitions between regimes than in cases of regime decay into a no-regime state.

The categorical analysis provides a first overview of transition types, but more detailed insights emerge when we directly compare the active regime on day 0 with that on day 6 in the reanalysis, focusing on how the percentage of cases with poor forecasts differs from that with good forecasts (Figure 8b). First, the largest contrasts are found for the persistence of no regime activity and the ScBL regime. No-regime persistence within the 6-day period is much more common in poor forecasts, with the odds being only half as high in good forecasts. In contrast, ScBL persistence within the 6-day period is much more common in good forecasts with odds nearly four times higher in good forecasts. Second, beyond persistence, several other transitions highlight systematic differences between poor and good forecasts. Onsets or transitions into ScBL, such as from the no regime, AT and EuBL, are consistently more common in good forecasts. In particular, the transition from EuBL to ScBL is strongly favoured in good forecasts, showing that this transition is nearly ten times more likely in good forecasts than in poor ones. This suggests that good forecasts tend to depict the transition into ScBL more frequently and with greater persistence. A similar strong signal appears for the AR-GL transition, which is more than four times as likely to occur in good compared to poor forecasts. Overall, higher frequency of regime-to-regime transitions in good forecasts (cf. Figure 8a) mainly reflects transitions between blocked regimes (central square in Figure 8b). And third, several regime decays show a strong association with poor forecasts, namely AT, EuBL, and ZO decays, all of which exhibit odds ratios (ORs) below 0.5, indicating that these transitions occur at least twice as often in poor forecasts as in good ones. To conclude, these results highlight the added value of analysing regime types individually, as overall signals in broad regime evolution categories can be dominated by specific regime transitions.

3.7 | Reanalysis perspective on regime persistence and transition timing

While poor and good forecasts over Europe show a similar share of cases without regime transitions, their comparison reveals statistically significant differences (Figure 8b). These motivate a more detailed analysis of the persistence of

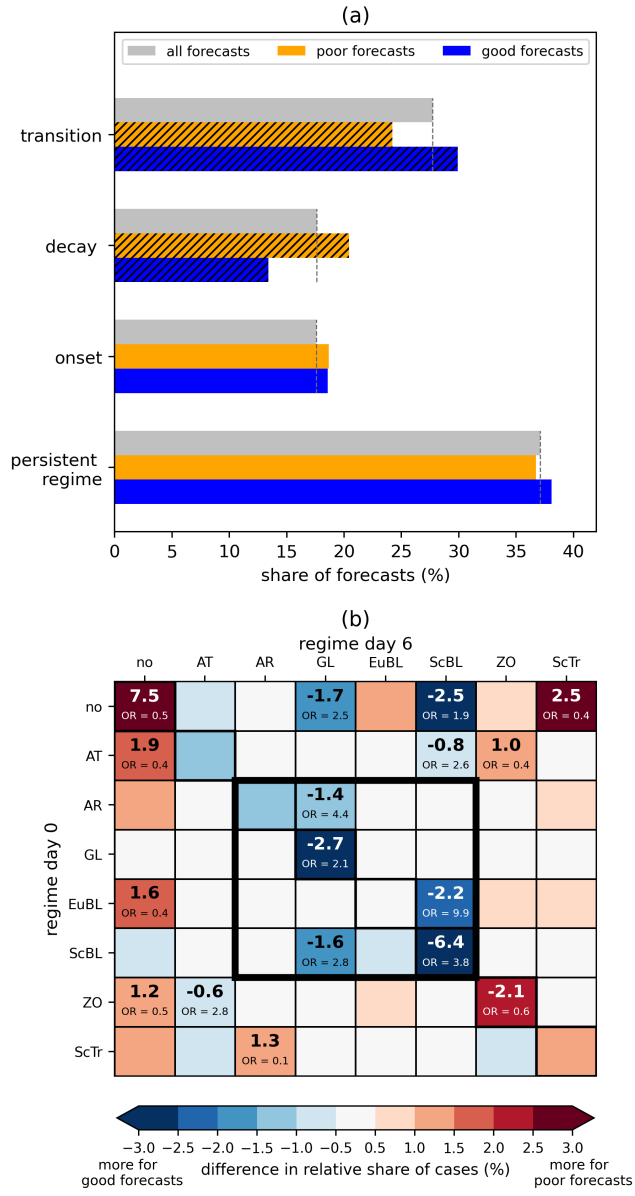

FIGURE 8 (a) Fraction of forecasts (in %) for the four different categories of regime evolution within the 6-day period as represented in the reanalysis: (i) persistent regime or persistent no-regime, (ii) onset of a regime out of the no regime, (iii) decay of a regime into the no regime, and (iv) transition from one into another regime (both not the no regime). Grey bars correspond to all forecasts (1979–2023), orange bars to poor forecasts and blue bars to good forecasts. Hatched bars show categories where the difference in proportions between good and poor forecasts is statistically significant, based on a z-test for proportions ($p<0.05$). (b) Differences in the relative share of regime-to-regime combinations between poor and good forecasts (in %), based on reanalysis data from day 0 to day 6. Cases with regime persistence (i.e., the same regime at day 0 and day 6) appear along the diagonal. The box marked by thick lines in the centre of the matrix indicates persistent blocked regimes (in the diagonal) or blocked-to-blocked regime transitions. Cells of the 8×8 transition matrix were tested for differences between poor and good forecasts using z-tests or permutation tests (for rare transitions), with FDR correction (α = 0.05). For transitions that are statistically significant and meet a practical effect threshold ($\geq$ 2 percentage points difference or odds ratio (OR) $\geq$ 2 2 or $\leq$ 2 0.5), the exact relative difference (in bold) and the OR are displayed.

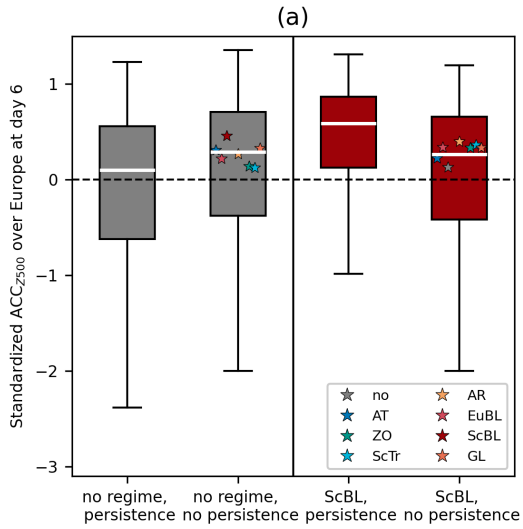

FIGURE 9 Forecast skill for all forecasts from 1979 to 2023, grouped by the initial regime and 6-day persistence in the reanalysis: 'persistence' indicates the regime remains unchanged after 6 days, while 'no persistence' indicates a transition to a different regime within 6 days in the reanalysis. White lines correspond to the median. Forecast skill is measured by the Z500-based, standardized ACC over Europe at forecast day 6. Star markers within the no-persistence category show the median skill for each specific transition into a different regime.

no-regime episodes in poor forecasts and of ScBL in good forecasts, again from a reanalysis perspective. Around 16 % of poor forecasts are linked to persistent no-regime periods, whereas if poor forecasts were randomly distributed across all periods, only about 12 % would be expected. This represents a modest but meaningful enrichment, with poor forecasts being roughly 30 % more likely than chance to coincide with persistent no-regime episodes. Most poor forecasts during persistent no-regime periods (≈65 %) occur consecutively, thereby resulting in extended intervals of low predictability. Both consecutive and isolated events occur year-round, but peak in spring (not shown). The modestly higher fraction of poor forecasts (16 %) compared to good forecasts (9 %) occurring during persistent no-regime periods raises the question of how forecast skill over Europe depends on initialization in a no-regime state and on the persistence of such conditions. Using all 6-day forecasts from 1979 to 2023, we find that forecast skill over Europe, measured by the standardized ACC, is higher during periods with a regime onset within the 6-day window than during persistent 6-day no-regime periods (Figure 9a, left). The exact type of regime onset also plays a role, with the highest forecast skill over Europe occurring for transitions into ScBL, and the lowest skill for transitions into cyclonic regimes such as ZO and ScTr (stars in Figure 9a, left). The difference between the two forecast skill distributions is statistically significant (Mann–Whitney U test $p<0.05$, effect size>0.1) and suggests that the persistence of large-scale flow regimes may influence forecast skill. Forecasts initialized during persistent no-regime periods show lower skill because the atmosphere remains disorganized, lacking slowly varying structures that constrain future evolution. In contrast, when a transition toward a regime is underway, the flow organizes into a more stable large-scale pattern, reducing the number of possible evolutions.

Of all good forecasts over Europe, nearly 17 % are initialized during ScBL periods. Among these, the majority, 54 % (corresponding to 8.6 % of all good forecasts), are associated with persistent ScBL periods. For comparison, if good forecasts were randomly distributed across all periods, only 4.1 % would be expected to occur during persistent ScBL.

This shows that persistent ScBL is roughly twice as frequent in good forecasts as would be expected by chance. Most good forecasts occur as consecutive good forecasts and hence provide a window of opportunity for enhanced skill. The seasonality of isolated events is rather flat, while consecutive events exhibit a clear peak occurrence between May and August, when the climatological frequency of ScBL reaches its maximum for the year (cf. Figure S5). Again, we compare the forecast skill over Europe depending on the persistence of ScBL (Figure 9a, right). Skill over Europe is higher when forecasts start in and cover persistent ScBL periods, whereas those with regime transitions or decay (into a no-regime state) exhibit lower skill. This indicates that forecasts initialized early in the life cycle of ScBL benefit from the regime's persistence, and this time the differences between the two distributions are statistically significant. Notably, RMSE-based distributions are significantly different for ScBL, in addition to ACC, highlighting that both accuracy and error magnitude benefit from persistent ScBL (Figure S8). This pattern is unique to ScBL. For the other regimes, statistically significant differences in ACC are found only for AT (Figure S7), and differences in RMSE are not significant for any regime other than ScBL (Figure S8). When the non-persistent ScBL cases are separated by transition type, forecast skill is lowest during ScBL decay. This suggests that the model has more difficulty representing the decay of the blocking pattern than its transition into a well-defined regime. Predictability requires persistence within a dynamically grounded attractor, and ScBL provides a clear example of this: In summer, ScBL has been shown to be the most physically grounded regime (Hochman et al., 2021, their Figures S4f and S5), combining strong dynamical constraints with persistence, which supports high forecast skill over Europe.

From the reanalysis perspective, both exceptionally poor and good forecasts feature a similar share of 6-day periods with regime transitions (Figure 8a), indicating that the occurrence of a transition alone is not a sufficient predictor of forecast skill. Factors such as the nature of the involved regimes, sensitivity to initial errors, and the timing of the regime transition likely determine forecast skill. Here, we focus on the latter and investigate the timing of regime transitions within the 6-day forecasts, as represented in the reanalysis, using three transition categories (onset, decay, and regime-to-regime) for poor and good forecasts (Figure 10). For exceptionally good 6-day forecasts, regime transitions tend to occur early in the period (days 1–2; solid blue line), suggesting that the large-scale flow is already evolving predictably at initialization. In contrast, transitions in poor forecasts occur later, between days 4 and 6 (solid orange line), and this difference compared with good forecasts is statistically significant. Such a pattern aligns with the general decline of predictability with lead time, with regime transitions being particularly sensitive due to the combined effects of initial uncertainty growth and transient dynamics. A more nuanced picture emerges when considering the transition categories separately. In the onset category, corresponding to transitions from no-regime to regime (dashed lines in Figure 10), good forecasts peak within the first day after initialization, while poor forecasts show two peaks, one very early and the other on day 5. This suggests that poor forecasts can still capture early regime onsets when the initial conditions already contain the processes leading to the onset, although forecast skill may deteriorate later due to error growth. For the decay category, corresponding to transitions from regime to no-regime (dotted lines in Figure 10), good forecasts show peaks in regime decay at both very early and late lead times, with the later peak being dominant. Poor forecasts show a tendency toward later decays, with a peak around day 5, although the timing distributions are not significantly different. This indicates differences in the timing of decays in the reanalysis, without implying whether the forecasts themselves captured the transition. Finally, regime-to-regime transitions show no major differences in timing between poor and good forecasts, and their distributions are not significantly different (dash-dotted lines).

Taken together, our results indicate that forecast skill over Europe is higher when the large-scale flow is already organized at initialization, either through persistent regimes such as ScBL or through early regime onsets, whereas persistent no-regime periods and late transitions are generally associated with lower skill. This emphasizes that both the persistence of dynamically grounded regimes and the timing of regime transitions appear to influence, with ScBL

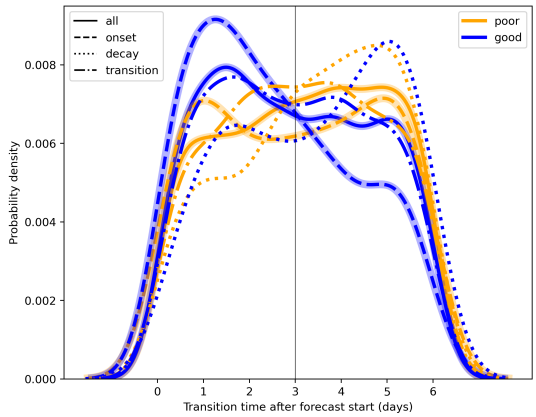

FIGURE 10 Probability density functions of weather regime transition times from reanalysis. Kernel density estimates are shown for good (blue) and poor (orange) forecasts over Europe, distinguishing: (1) all transitions (solid), (2) regime onset (dashed), (3) regime decay (dotted), and (4) regime-to-regime transitions (dash-dotted). The x-axis shows transition time after forecast start (0–6 days), and the y-axis shows probability density. Shading highlights statistically significant differences between poor and good forecasts within each category (Kolmogorov–Smirnov, Mann–Whitney U, and permutation tests). KDE smoothing slightly extends curves beyond 0–6 days, though all data lie within this range.

representing a particularly favourable case.

3.8 | Link between exceptional forecasts over Europe and regime predictions over the North Atlantic–European region

The previous section provided insights into regime development during periods of exceptionally poor and exceptionally good forecasts in the reanalysis. This raises the question of how large-scale regime developments are represented in the reforecasts. Here, we aim to systematically investigate whether exceptionally poor and good forecasts over Europe are associated with incorrect or correct regime assignments over the North Atlantic–European region at forecast day 6. For each forecast and reanalysis, we identify the dominant regime at each lead time by assigning the regime with maximum WRI above 1.0 (or no regime if none exceed 1.0) and consider the forecast correct when the active regime matches that of the reanalysis.

We found significant differences in the number of correct regime forecasts between poor and good forecasts (Figure 11). Less than 45% of all poor forecasts have a correct regime assignment at day 6, indicating that most poor forecasts over Europe do not capture the large-scale regime over the broader North Atlantic–European domain (orange line). In contrast, more than 80% of good forecasts over Europe show a correct regime prediction at day 6 (blue line). The evolution of this correct regime share is very similar in the first two days. This does not necessarily mean that the magnitude of forecast errors is similar during these two days but rather suggests that errors need to grow in scale first to modulate the large-scale circulation in such a way, that it is recognizable in the WRI. After day 2, the differences between poor and good forecasts become statistically significant and also differ from the evolution based on all forecasts from 1979 to 2023.

The co-occurrence of poor forecasts over Europe with correct regime forecasts, as well as good forecasts with incorrect regime forecasts raises the question of how this can occur given that Europe lies fully within the regime do-

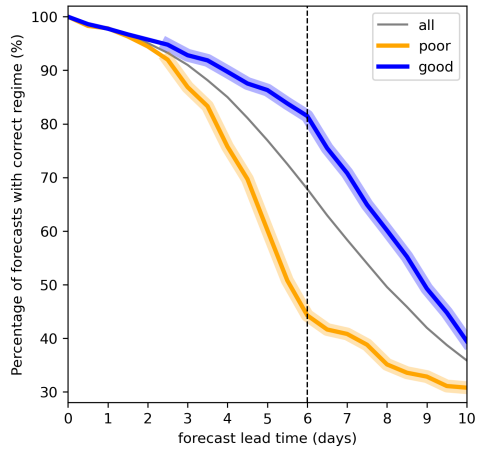

FIGURE 11 Percentage of ERA5 forecasts that show a correct dominant regime prediction (in %) as a function of forecast lead time (days) for all forecasts (grey line), poor forecasts over Europe (orange), and good forecasts over Europe (blue). The vertical black line marks the day we use for validation of forecast skill over Europe (day 6). Shading around the lines indicates lead times where the difference in percentage between good and poor forecasts is statistically significant, as determined by per-lead permutation tests corrected using the Benjamini–Hochberg false discovery rate (FDR).

main. Figure 12 shows differences in absolute Z500 errors at day 6 between correct and incorrect regime forecasts for poor and good forecasts separately. For poor forecasts over Europe, larger errors over the North Atlantic extend into Europe, with the region of significantly larger errors spreading from the Central North Atlantic to Scandinavia (Figure 12a). This reveals slight differences in errors within Europe, suggesting that larger errors over northern Europe are necessary for a simultaneous incorrect regime forecast over the North Atlantic–European region. Eastern and central Europe show slightly higher absolute errors when the regime forecast is correct, indicating that the largest errors tend to occur in the southern and eastern parts of Europe near the edges of the regime domain. When errors are largest in regions where the seven regimes have their main centres, such as the North Atlantic and northwestern Europe (cf. Figure S2), they are more likely to be associated with incorrect regime forecasts. For good forecasts, the 20% that coincide with an incorrect regime exhibit significantly larger errors over the North Atlantic (red shading), particularly southern Greenland and the area south of Greenland and Iceland, while differences over Europe are negligible (Figure 12b). This indicates that incorrect regime forecasts can co-occur with good forecasts over Europe if large errors are confined to the North Atlantic, which could later propagate eastward. Overall, not all poor forecasts over Europe lead to incorrect regimes, since some regimes have their centres of action over the North Atlantic, while good forecasts over Europe can still coincide with incorrect regimes if significant errors over the North Atlantic are present. This demonstrates that forecast errors over Europe and regime errors are not linearly related.

Lastly, we revisit the timing of regime transitions for poor and good forecasts, now splitting each skill group into forecasts with a correct regime prediction at day 6 and those without. Figure 13 shows the distribution of transition timing between day 0 and day 6 for cases with a regime transition. For poor forecasts, transitions occur earlier when the regime is correctly predicted, while incorrect regime forecasts mostly happen later in the forecast period. The two distributions differ significantly and are robust given balanced sample sizes (55 % incorrect vs. 45 % correct). This indicates that for poor forecasts, predicting the correct regime at day 6 strongly depends on early transitions, which

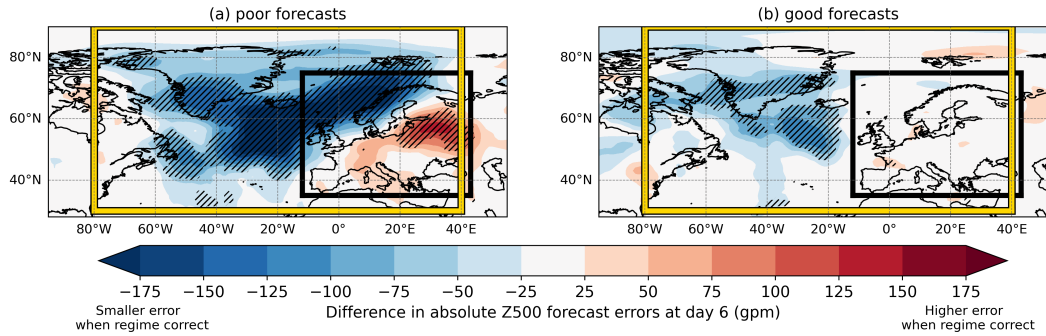

FIGURE 12 Difference in absolute Z500 forecast errors (gpm) at day 6 between correct and incorrect regime forecast for (a) poor forecasts and (b) good forecasts. Negative values indicate smaller errors for correct regimes or larger errors for incorrect regimes, and vice versa for positive values. The black box illustrates the box to calculate skill measures over Europe, the yellow box shows the domain used to define the North Atlantic–European regimes. The black box marks the area for calculating skill over Europe, and the yellow box shows the North Atlantic–European regime domain. Diagonal hatching (black) marks grid points where the difference in absolute Z500 forecast errors between correct and incorrect regime forecasts is statistically significant, based on a two-sided Mann–Whitney U test at each grid point with p-values corrected for multiple comparisons using the FDR method (α = 0.05).

provide a stronger signal, whereas later transitions reduce consistency with the evolving flow. For good forecasts, transition timing is less decisive. Correct regime forecasts tend to have early transitions (day 1–2), but incorrect forecasts occur at both early and late times, and the two distributions are not significantly different, partly due to sample size imbalance (80 % correct vs. 20 % incorrect). Overall, transition timing influences whether the regime is predicted correctly or incorrectly for poor forecasts, but it is less important for good forecasts.

4 | SUMMARY AND CONCLUSIONS

Despite considerable progress in medium-range weather forecasting over recent decades, models can still occasionally fail to accurately predict atmospheric conditions, resulting in so-called forecast busts. Research into these events has expanded in the last two decades, however, many questions remain unresolved, especially regarding the large-scale pattern evolution and the role of weather regime transitions. Moreover, previous systematic studies do not include the most recent years and rely on now outdated model versions, such as ERA-Interim, highlighting the need for a renewed investigation using more modern forecast systems. In this study, we revisited the original definition of forecast busts—poor forecasts of Z500 over Europe at day 6—proposed by Rodwell et al. (2013), introduced a revised definition that incorporates the seasonal variability of the two skill measures (ACC and RMSE), and based it on the objectively identified anomalous behaviour of both metrics. Using this updated definition, we systematically investigated forecast busts over Europe at day 6 based on a 45-year dataset of ERA5 reforecasts from ECMWF, which includes 32,850 forecasts. For the first time, this study extends the original notion of busts to the updated terminology of 'exceptionally poor forecasts' and introduces 'exceptionally good forecasts', providing a systematic and consistent characterization of both over Europe. Year-round North Atlantic–European weather regimes by Grams et al. (2017) were used to analyse the evolution of the large-scale circulation in the 6-day forecast periods. Using this regime perspective, we gained insights into the occurrence and role of large-scale circulation changes for exceptionally poor

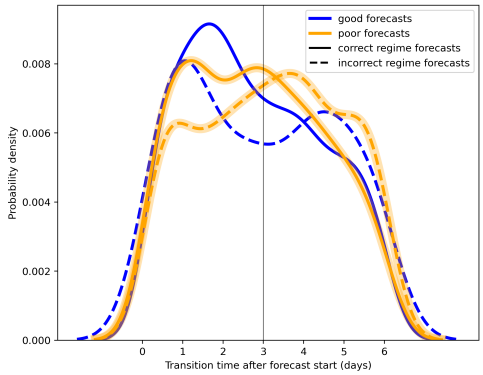

FIGURE 13 Probability density functions of weather regime transition times. Kernel density estimates are shown for poor (orange) and good (blue) forecasts over Europe, with correct regimes in solid lines and incorrect regimes in dashed lines at day 6. The x-axis shows transition time after forecast start (0–6 days), and the y-axis shows probability density. Shading around the orange lines marks statistically significant differences between correct and incorrect forecasts (Kolmogorov–Smirnov, Mann–Whitney U, and permutation tests). KDE smoothing slightly extends curves beyond 0–6 days, though all data points lie within this range.

and good forecasts over Europe.

The main results of this study can be summarized as follows:

- Skill measures (ACC and RMSE) for Z500 exhibit seasonality, so exceptional forecasts are identified relative to typical skill at each time of the year, thereby capturing extreme deviations in amplitude and phase error.
- Over the ERA5 reforecast period, exceptionally poor forecasts became less frequent (−2.17 %/decade) and exceptionally good forecasts more frequent (+2.20 %/decade), likely reflecting improved observational data.
- Despite applying for seasonality in skill measures for the detection of exceptional forecasts, seasonal effects persist, with poor forecasts remaining more common during the warm season, while good forecasts occur nearly evenly throughout the year.
- Consecutive sequences of the same forecast category are common (58 % poor, 47 % good) but have evolved over time, with fewer successive poor forecasts and more successive good forecasts in recent years.
- The mean picture of the large-scale circulation pattern differs sharply with Rossby wave train signals from the Pacific to Europe and ridging over the eastern North Atlantic for poor forecasts and blocking over northern Europe and a lack of upstream wave activity for good forecasts.
- From a weather-regime perspective, poor forecasts are most often linked to cyclonic or no-regime states, whereas good forecasts correspond to anticyclonic, blocked regimes, particularly Scandinavian Blocking.
- Regime transitions occur in approximately 60 % of cases for both skill categories; however, transition type and timing matter, with good forecasts associated with earlier regime transitions and more frequent regime-to-regime transitions, whereas late transitions, which often involve a decay of a regime, are linked to poor forecasts.
- Day-6 regime accuracy ranges from roughly 55 % in poor forecasts to over 80 % in good forecasts over Europe, yet the decisive factors are the large-scale error distribution and the timing of regime transitions,

with early transitions typically captured, whereas late ones are often missed.

Referring back to the research questions addressed in the introduction, our results show that the characteristics of poor and good forecasts differ more strongly from each other than poor forecasts do across different analyses. The similarity between ERA-Interim-based (Rodwell et al., 2013; Lillo and Parsons, 2017) and ERA5-based poor forecasts indicates that, despite improvements in model quality and adjustments in the definition, the same types of events continue to challenge the forecasting system, suggesting that these errors are not model-specific. Although exceptionally poor and good forecasts occur during all regimes, their distributions differ systematically: poor forecasts are more likely than good forecasts to occur during no-regime periods (35 % vs. 27 %) and cyclonic regimes (31 % vs. 25 %), whereas good forecasts are more likely than poor forecasts to occur during anticyclonic regimes (48 % vs. 35 %). Overall, 16 % of poor forecasts are associated with persistent no-regime periods compared with 9 % of good forecasts, while good forecasts are roughly twice as likely as poor forecasts to be linked to persistent ScBL periods (9 % vs. 4 %), highlighting the importance of regime type and persistence. While the mere occurrence of a regime transition within the 6-day period in the reanalysis is not indicative of forecast skill, both the type of transition and, in particular, its timing within the 6-day period appear to have a significant influence on forecast performance. Overall, these findings suggest that improving forecasts requires not only continued investment in observational systems but also careful representation of regime dynamics and transition timing in numerical models.

As with any study, several limitations should be acknowledged. Exceptionally poor and good forecasts show substantial case-to-case variability, which we partly addressed by stratifying by season and regime transitions, though alternative classifications could provide additional insights. Our analysis is based on a single model configuration without ensembles, limiting generalizability and preventing assessment of forecast skill versus ensemble spread, as in Ferranti et al. (2015). Inter-model and ensemble-based comparisons could clarify which features are model-specific. While using year-round weather regimes and skill-based definitions improved interpretation, the identification of regimes and transitions depends on the chosen classification method, which potentially affects quantitative results. This study also does not explicitly quantify all dynamical or diabatic processes, such as upstream convection, Rossby wave breaking, or wave packet characteristics, which may influence forecast skill. Future research could leverage ensemble forecasts, investigate the societal impacts of extreme forecasts, track systematic errors, and expand analyses to additional regions and models to better understand medium-range forecast extremes and their underlying mechanisms.

Acknowledgements

The authors acknowledge funding by the Office Of Naval Research (ONR) and the Swiss National Science Foundation (SNSF). We would like to thank Christian M. Grams for his valuable contributions and insightful discussions, which significantly enriched the quality of this work in the review process. We are grateful to Mark Rodwell for helpful discussions on forecast busts and for providing the ERA-Interim forecast-bust dates. Finally, we sincerely thank the anonymous reviewers for their constructive and thoughtful comments and the editor for helpful editorial suggestions that improved the clarity and robustness of this manuscript.

Conflict of interest

The authors declare no conflict of interest.

Data availability statement

The ERA5 reanalyses used in this study are freely available online. The weather regime data can be downloaded from Zenodo at https://doi.org/10.5281/zenodo.17080146. The ERA5 reforecasts are made available by ECMWF under license to authorized users. Processed data and code employed in the analyses presented can be provided by the authors upon reasonable request.

References

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

Brannan, A. L. and Chagnon, J. M. (2020) A climatology of the extratropical flow response to recurving atlantic tropical cyclones. *Monthly Weather Review*, **148**, 541 – 558.

Büeler, D., Ferranti, L., Magnusson, L., Quinting, J. F. and Grams, C. M. (2021) Year-round sub-seasonal forecast skill for atlantic–european weather regimes. *Quarterly Journal of the Royal Meteorological Society*, **147**, 4283–4309.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F. (2011) The era-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137**, 553–597.

Ferranti, L., Corti, S. and Janousek, M. (2015) Flow-dependent verification of the ecmwf ensemble over the euro-atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, **141**, 916–924.

Ferranti, L., Magnusson, L., Vitart, F. and Richardson, D. S. (2018) How far in advance can we predict changes in large-scale flow leading to severe cold conditions over europe? *Quarterly Journal of the Royal Meteorological Society*, **144**, 1788–1802.

Grams, C. M., Beerli, R., Pfenninger, S., Staffell, I. and Wernli, H. (2017) Balancing europe's wind power output through spatial deployment informed by weather regimes. *Nature Clim Change*, **7**, 557–562.

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

Grazzini, F. and Isaksen, L. (2002) North american increments. *Technical Report OD/MOD/23*, ECMWF Operational Department, Reading, UK.

Hamill, T. M., Bates, G. T., Whitaker, J. S., Murray, D. R., Fiorino, M., Galarneau, T. J., Zhu, Y. and Lapenta, W. (2013) Noaa's second-generation global medium-range ensemble reforecast dataset. *Bulletin of the American Meteorological Society*, **94**, 1553 – 1565.

Hauser, S., Teubler, F., Riemer, M., Knippertz, P. and Grams, C. M. (2023) Towards a holistic understanding of blocked regime dynamics through a combination of complementary diagnostic perspectives. *Weather and Climate Dynamics*, **4**, 399–425.

— (2024) Life cycle dynamics of greenland blocking from a potential vorticity perspective. *Weather and Climate Dynamics*, **5**, 633–658.

Hersbach, H., Bell, B., 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., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J.-N. (2020) The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**, 1999–2049.

Hochman, A., Messori, G., Quinting, J. F., Pinto, J. G. and Grams, C. M. (2021) Do Atlantic-European Weather Regimes Physically Exist? *Geophysical Research Letters*, **48**, e2021GL095574.

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

Lemburg, A. and Fink, A. H. (2024) Investigating the medium-range predictability of european heatwave onsets in relation to weather regimes using ensemble reforecasts. *Quarterly Journal of the Royal Meteorological Society*, **150**, 3957–3988.

Lillo, S. P. and Parsons, D. B. (2017) Investigating the dynamics of error growth in ecmwf medium-range forecast busts. *Quarterly Journal of the Royal Meteorological Society*, **143**, 1211–1226.

Lojko, A., Payne, A. and Jablonowski, C. (2022) The remote role of north-american mesoscale convective systems on the forecast of a rossby wave packet: A multi-model ensemble case-study. *Journal of Geophysical Research: Atmospheres*, **127**, e2022JD037171.

Lorenz, E. N. (1963) Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130 – 141.

Magnusson, L. (2017) Diagnostic methods for understanding the origin of forecast errors. *Quarterly Journal of the Royal Meteorological Society*, **143**, 2129–2142.

Magnusson, L. and Källén, E. (2013) Factors influencing skill improvements in the ecmwf forecasting system. *Monthly Weather Review*, **141**, 3142 – 3153.

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., Dirmeyer, P. A., Ferranti, L., Johnson, N. C., Jones, J., Kirtman, B. P., Lang, A. L., Molod, A., Newman, M., Robertson, A. W., Schubert, S., Waliser, D. E. and Albers, J. (2020) Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bulletin of the American Meteorological Society*, **101**, E608 – E625.

Matsueda, M. and Palmer, T. N. (2018) Estimates of flow-dependent predictability of wintertime euro-atlantic weather regimes in medium-range forecasts. *Quarterly Journal of the Royal Meteorological Society*, **144**, 1012–1027.

McLay, J. G. and Satterfield, E. (2022) Forecast dropouts in the navgem model: Characterization with respect to other models, large-scale indices, and ensemble forecasts. *Weather and Forecasting*, **37**, 2049 – 2067.

Michel, C. and Rivière, G. (2011) The link between rossby wave breakings and weather regime transitions. *Journal of the Atmospheric Sciences*, **68**, 1730–1748.

Mockert, F., Grams, C. M., Brown, T. and Neumann, F. (2023) Meteorological conditions during periods of low wind speed and insolation in germany: The role of weather regimes. *Meteorological Applications*, **30**, e2141.

Nakamura, N. and Huang, C. S. Y. (2018) Atmospheric blocking as a traffic jam in the jet stream. *Science*, **361**, 42–47.

Osman, M., Beerli, R., Büeler, D. and Grams, C. M. (2023) Multi-model assessment of sub-seasonal predictive skill for year-round atlantic–european weather regimes. *Quarterly Journal of the Royal Meteorological Society*, **149**, 2386–2408.

Palmer, T. N. (1999) Predicting uncertainty in forecasts of weather and climate. *Tech. Rep. 294*, ECMWF Technical Memoranda.

Parsons, D. B., Lillo, S. P., Rattray, C. P., Bechtold, P., Rodwell, M. J. and Bruce, C. M. (2019) The role of continental mesoscale convective systems in forecast busts within global weather prediction systems. *Atmosphere*, **10**.

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, T., Persson, A., Prates, F. and Wedi, N. (2013) Characteristics of occasional poor medium-range weather forecasts for europe. *Bulletin of the American Meteorological Society*, **94**, 1393 – 1405.

Teubler, F., Riemer, M., Polster, C., Grams, C. M., Hauser, S. and Wirth, V. (2023) Similarity and variability of blocked weather-regime dynamics in the atlantic–european region. *Weather and Climate Dynamics*, **4**, 265–285.

Vitart, F. (2017) Madden—julian oscillation prediction and teleconnections in the s2s database. *Quarterly Journal of the Royal Meteorological Society*, **143**, 2210–2220.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C. and et al. (2017) The subseasonal to seasonal (s2s) prediction project database [dataset]. `https://apps.ecmwf.int/datasets/data/s2s/`. Accessed: 2025-07-23.

Wandel, J., Büeler, D., Knippertz, P., Quinting, J. F. and Grams, C. M. (2024) Why moist dynamic processes matter for the sub-seasonal prediction of atmospheric blocking over europe. *Journal of Geophysical Research: Atmospheres*, **129**, e2023JD039791.

Wilks, D. S. (2020) *Statistical methods in the Atmospheric Sciences*. Elsevier, 4 edn.

Yamagami, A. and Matsueda, M. (2021) Statistical characteristics of arctic forecast busts and their relationship to arctic weather patterns in summer. *Atmospheric Science Letters*, **22**, e1038.

Yu, Q., Spensberger, C., Magnusson, L. and Spengler, T. (2025) Forecast Errors Attributed to Synoptic Features. *Meteorological Applications*, **32**, e70093.