Windows of opportunity in subseasonal weather regime forecasting: A statistical-dynamical approach

Fabian Mockert[1,2*] | Christian M. Grams[1,3] | Sebastian Lerch[4,5,6] | Julian Quinting[1]

[1]Institute of Meteorology and Climate Research Troposphere Research (IMKTRO), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[2]now at: Institute of Statistics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[3]now at: Federal Office of Meteorology and Climatology, MeteoSwiss, Zurich-Airport, Switzerland

[4]Institute of Statistics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[5]Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

[6]now at: Department of Mathematics and Computer Science, University of Marburg, Marburg, Germany

Correspondence
Fabian Mockert, Institute of Statistics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
Email: fabian.mockert@kit.edu

Present address
Institute of Statistics, Karlsruhe Institute of Technology (KIT), Blücherstr.17, 76185 Karlsruhe, Germany

The Madden-Julian Oscillation (MJO) and stratospheric polar vortex (SPV) are prominent sources of subseasonal predictability in the Extratropics. With relevance for European weather it has been shown that the joint interaction of MJO and the SPV can modulate the preferred phase of the North Atlantic Oscillation (NAO) and the occurrence of weather regimes. However, improving extended-range numerical weather prediction (NWP) at three-week lead times remain under-explored. This study investigates how MJO and SPV phases affect Greenland Blocking (GL) activity and integrates atmospheric state information into a neural network to enhance week-three weather regime activity forecasts.

We define a weather regime activity metric using ECMWF reanalysis and reforecasts. In reanalyses we find increased GL activity following MJO phases 7, 8 and 1, as well as weak SPV phases, indicating climatological windows of opportunity in line with previous studies. However, ECMWF forecast skill improves only in MJO phases 8 and 1 and weak SPV phases, identifying somewhat different model windows of opportunities. Next we explore using these findings in post-processing tools. Climatological forecasts based on MJO/SPV-NAO relationships provide a purely statistical approach to extended-range GL activity forecasting, independent of NWP models. Notably, MJO conditioned climatological forecasts show clear signals when evaluated against observed GL activity.

Statistical-dynamical models, using neural networks that combine historical atmospheric state data with NWP-derived weather regime metrics improve weather regime activity forecasts across all regimes considered, achieving an absolute accuracy increase of 2.9% in forecasting the dominant weather regime compared to ECMWF. This is particularly beneficial to Blocking in the European domain, where NWP models often underperform. Atmospheric conditioned and neural network forecasts serve as valuable decision-support tools alongside NWP models, enhancing the reliability of subseasonal-to-seasonal (S2S) predictions.

KEYWORDS

weather regimes, windows of opportunity, Madden-Julian Oscillation, Stratospheric Polar Vortex, neural networks

1 | INTRODUCTION

The demand for accurate subseasonal to seasonal forecasting – extending beyond 10 days and up to 30 days (White et al., 2017) – has grown over the past decade, driven by socio-economic needs. Decision-makers in sectors such as public health, agriculture, and energy rely on accurate forecasts several weeks ahead, as briefly reviewed with the following examples: In the United Kingdom, winter cold spells increase hospital admissions due to respiratory diseases (Charlton-Perez et al., 2019; Elliot et al., 2008). Improved drought forecasting can save millions of dollars by mitigating crop shortages (Salient et al., 2023). Additionally, large-scale weather patterns can serve as indicators of peak winter energy demand, placing immense stress on the power systems (Bloomfield et al., 2020; Millin et al., 2024). Bloomfield et al. (2021) show that in the context of renewable energy forecasting, pattern-based methods yield higher forecast skill than grid-point based methods for extended-range lead times > 12 days.

Large-scale quasi-stationary, recurrent, and persistent weather patterns are commonly referred to as weather regimes (Michelangeli et al., 1995; Vautard, 1990). They represent the complex large-scale circulations through a finite set of states, facilitating the interpretation and categorisation of the prevailing or forecasted atmospheric flow. These regimes are associated with distinct impacts on surface variables, including 2-m temperature, 10-m wind and radiation (Grams et al., 2017; Beerli and Grams, 2019). Often the weather regimes are identified on the basis of the 500-hPa geopotential height field (Z500). For the European region, a year-round definition of seven weather regimes has been established (Grams et al., 2017) which can be formalised with a seven-dimensional normalised weather regime index. Two of the seven weather regimes, Zonal Regime and Greenland Blocking, closely resemble the positive and negative phase of the North Atlantic Oscillation (NAO), respectively. The forecast skill of Greenland Blocking and Zonal Regime is generally the best among the seven weather regimes, whereas the European and Scandinavian Blocking regimes are the worst predicted (Büeler et al., 2021; Osman et al., 2023). These three blocking regimes are especially important for the prediction of renewable power generation in winter as periods of low wind and solar power output occur mainly within these regimes. Greenland Blocking is further connected with cold temperature anomalies, leading to increased heating demand (Mockert et al., 2023). Improvements in the predictions for these three regimes, along with the other four regimes, in winter would greatly benefit the energy sector, particularly energy trading companies and transmission grid operators.

Previous studies suggest links between the atmospheric state at initialisation time and the occurrence of weather regimes weeks later (e.g., Cassou, 2008; Baldwin et al., 2001). Though much weaker, these connections are reflected in the skill of numerical weather prediction models in forecasting these regimes (e.g., Vitart, 2017). Two major atmospheric modes of variability influencing Greenland Blocking (or NAO–) are the Madden Julian Oscillation (MJO, (Madden and Julian, 1971)), and the stratospheric polar vortex (SPV). Cassou (2008) and Lin et al. (2009) showed a lagged connection between MJO phases at initialisation time and the preferred flow pattern in the North Atlantic-European region. The occurrence of NAO– is anomalously high two weeks after MJO phases 6–8, while NAO+ is anomalously more frequent following MJO phases 3 and 4. Similarly, although considerably weaker than in reanalysis, reforecasts by the European Centre for Medium-Range Weather Forecasts show positive NAO forecasts for lead times of 11–15 days following phase 3 and negative NAO forecasts following phase 7 (Vitart, 2017). Ferranti et al. (2018) describe an asymmetric connection of the MJO to forecasts of the NAO phases. Forecasts of NAO– tend to have more (less) skill when the MJO is active (inactive) at initialisation time, whereas little sensitivity of the forecast skill to the MJO activity at initial time is found for the NAO+ phase. Thus, despite the shown MJO teleconnection to Europe, numerical weather prediction still struggles exploiting this potential source of extended-range predictability. A potential reason, is the multi-scale interaction with other modes of subseasonal variability and the chaotic growth of error (Roberts et al., 2023).

Most importantly the relation between the MJO and European weather regimes is further influenced by the Quasi Biennial Oscillation (QBO), El Niño-Southern Oscillation (ENSO), and SPV. The QBO, the leading mode of the interannual variability in the tropical stratosphere, involves a reversal of the mean-zonal wind direction, alternating between easterly (EQBO) and westerly (WQBO) phases approximately every 28 months (Feng and Lin, 2019; Baldwin et al., 2001). The MJO-NAO link (tendency of

NAO+/NAO– following MJO phases 3/7 as described by Vitart (2017)) is more pronounced and persistent during the WQBO, whereas it is weaker and less statistically significant during the EQBO (Feng and Lin, 2019). ENSO modulates the MJO-NAO signal such that the NAO+ following MJO phases 1–5 is strongly enhanced during El Niño years and suppressed during La Niña years. Conversely, NAO– following MJO phases 7–8 is most pronounced during La Niña years and suppressed during El Niño years (Lee et al., 2019). Roberts et al. (2023) analyse the performance of ECMWF forecasts and conclude that the model underestimates the MJO-NAO connection and fails to reproduce the modulation of the MJO-NAO connection by ENSO. Finally, Kent et al. (2023) indicate that improvements in monthly NAO forecasts are hard to achieve, even with a perfect MJO forecast and MJO-NAO relationship, but they suggest that they are potentially achievable from currently unknown sources of skill.

Greenland Blockings (NAO–) occur with an enhanced probability of up to two months following weak SPV conditions due to persistent anomalies in the troposphere (Baldwin et al., 2001; Beerli and Grams, 2019). Further, Domeisen et al. (2020) state that sudden stratospheric warmings (SSWs) are also frequently followed by an Atlantic Trough in the weeks after. Greenland Blocking is most likely if a blocking situation over western Europe and the North Sea prevailed, while the Atlantic Trough is most likely when Greenland Blocking was present around the SSW onset.

The connection between the atmospheric state at initialisation time and the frequency of occurrence of weather regimes, as well as the enhanced forecast skill after forecast initialisation in certain atmospheric states, indicates that the phases of the MJO and SPV could potentially be considered as windows of opportunities for subseasonal predictions. Window of opportunities can generally be defined as specific atmospheric states in which either the frequency of occurrence of an event is increased, referred to as a climatological window of opportunity, or the forecast of an event is improved, referred to as a model window of opportunity (following the definition of Specq and Batté (2022) applying window of opportunities on the MJO and heavy tropical precipitation events).

This study explores windows of opportunities linking the MJO and SPV to weather regimes during the extended winter period (November to March). A key objective is to determine whether the established links between these atmospheric modes of variability and weather regimes, typically observed up to two weeks, extend into forecast week three. To explore this, we investigate whether a climatological forecast, informed by the atmospheric state at initialisation, can provide meaningful indications of weather regime occurrence beyond week two. Additionally, we assess whether statistical-dynamical models, particularly neural networks that integrate prior knowledge and numerical weather prediction outputs, enhance weather regime forecasts.

Before answering these questions, data and methods are introduced in Section 2. This study provides a detailed analysis of window of opportunities conditioned on the MJO and SPV for predicting the activity of Greenland Blocking on a lead time of three weeks (Section 3.1). That information is then further used to provide an atmospheric-based climatology conditioned on the state of the MJO or SPV as a tool alongside a numerical weather prediction model (Section 3.2). In a final step, all information from the numerical weather prediction and the atmospheric state at initialisation time are joined by a neural network to improve the forecasts of the activity of weather regimes (Section 3.3). Section 4 summarises and discusses the main findings.

2 | DATA AND METHODS

2.1 | ECMWF reforecast and reanalysis

This study utilises sub-seasonal to seasonal reforecast data of the ensemble prediction system of the European Centre for Medium-Range Weather Forecasts (ECMWF), provided through the Subseasonal-to-Seasonal (S2S) Prediction Project Database (Vitart et al., 2017). To increase the number of forecast initial dates, we include forecasts from two consecutive model cycles, Cy46R1 and Cy47R1 (ECMWF, 2018). These reforecasts are initialised twice a week (Mondays and Thursdays) from ERA5 reanalysis data (Hersbach et al., 2020) for the past 20 years and consist of 11 ensemble members. The forecasts cover a forecast lead time of 0–46 days at a native horizontal grid spacing of 16 km up to day 15 and 32 km from day 15 onwards. Forecast data were remapped from their native resolution to a regular latitude-longitude grid of 1×1°grid spacing. The two model cycles were operational from June 11, 2019, to May 11, 2021, with a cycle change from Cy46R1 to Cy47R1 on June 30, 2020. As a result, for the period between May 11 and June 30 reforecasts are only available from Cy46R1. Our reforecast dataset comprises a total of 4,000 initial dates, each with 11 ensemble members. For forecast evaluation, we treat ERA5 as a "perfect ensemble member" by aligning its data with each initialisation date and lead time (cf. Wandel et al., 2024). Additionally, ERA5 data is remapped from its native grid to match the reforecast grid spacing. Whenever possible (subject to predictor selection constraints), we enhance neural network training by incorporating not only the 4,000 reforecast dates but also daily ERA5 data from 1979–2021.

2.2 | Weather regimes

In this study, we identify the seven year-round North Atlantic-European weather regimes introduced by Grams et al. (2017) in ERA5 reanalysis as described in Hauser et al. (2023a,b) and in the reforecasts following the approach of Büeler et al. (2021) and Osman et al. (2023). These weather regimes (WR) represent the most common large-scale circulation patterns in the North Atlantic-European region (30–90°N, 80°W–40°E). In brief, we conduct an empirical orthogonal function (EOF) analysis of 6 hourly (1979–2019), 10-day low pass filtered (filter-width of 20 days, hence ±10 days), seasonally normalised geopotential height anomalies (Z500A, relative to 91-day running mean climatology) within the domain of the weather regimes. We then apply a k-means clustering algorithm on the first seven EOFs and set $k = 7$. These seven clusters represent the seven distinct weather regimes (for a visualisation of the weather patterns, see Figure S1 in Mockert et al. (2024)), with three cyclonic (Atlantic Trough (AT), Zonal Regime (ZO), and Scandinavian Trough (ScTr)) and four anticyclonic regime types (Atlantic Ridge (AR), European Blocking (EuBL), Scandinavian Blocking (ScBL), and Greenland Blocking (GL)). The ZO and GL regimes correspond to the positive and negative phases of the NAO, respectively.

The projection of instantaneous anomalies onto the mean regime patterns, whether in reanalysis or reforecast, is determined by a seven-dimensional weather regime index (IWR) following the approach of Michel and Rivière (2011). As this study relies on the dataset computed in Mockert et al. (2024), we refer the interested reader to their publication for a detailed explanation and only briefly outline the computation of the IWR. Here, we specifically use the Z500 bias-corrected IWR reforecasts and reanalysis from Mockert et al. (2024):

1. Compute Z500 anomalies for reforecasts/ERA5 relative to a 91-day running mean model/ERA5 climatology (1999-2015 and 1979-2019, respectively).
2. Apply low-pass filtering and normalisation to obtain filtered and standardised Z500 anomalies.
3. Project filtered and standardised Z500 anomalies onto seven cluster mean Z500 anomalies of distinct weather regimes
4. Compute weather regime index by normalising the projection against its climatological mean and standard deviation

2.3 | Weather regime activity

This study focuses on the predictability of weather regimes at three weeks forecast lead time. Due to the reduced skill of daily weather regime forecasts at this lead time (Büeler et al., 2021) and the practical needs of decision-makers, we consider weekly aggregated forecast information based on daily data. To achieve this, we introduce the concept of weather regime activity, which quantifies the presence of a specific weather regime over a given period. The weather regime activity is defined in two complementary ways. The *weekly mean weather regime activity* ($WRact_{mean}$), suited for deterministic decision-making, considers a weather regime active in ERA5 or an ensemble member ($WRact_{mean} = 1$) in the respective week, if the weekly mean of daily IWR exceeds a predefined activity threshold of $IWR_{min} = 1$, otherwise $WRact_{mean} = 0$. Applied to the ensembles forecast across 11 ensemble members, $WRact_{mean}$ takes values between 0 and 1 in increments of 1/11, representing a probability in the ensemble. In contrast, the *aggregated daily weather regime activity* ($WRact_{agg}$) provides a more detailed view by measuring the fraction of days within a given week where the daily IWR exceeds IWR_{min}. Applied to ERA5 or an ensemble member, this results in values between 0 and 1, with discrete steps of 1/7, representing the fraction of days per week with active regime conditions. When applied to ensemble forecasts across 11 ensemble members, $WRact_{agg}$ yields values between 0 and 1 in steps of 1/77, allowing for a probabilistic perspective, representing the fraction of days per week with active regime conditions in the ensemble.
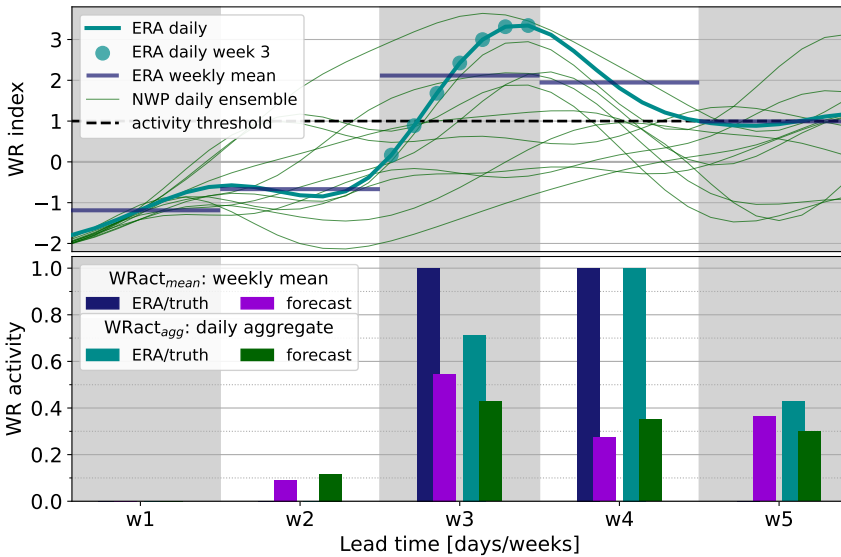

FIGURE 1 Visualisation of $WRact_{mean}$ and $WRact_{agg}$ by using the weather regime index. The upper panel shows the ensemble forecast (green lines) and the perfect member (ERA5, turquoise line) of the daily IWR. Further, the dark blue horizontal lines indicate the mean weekly IWR in ERA5 and the turquoise dots the daily IWR in ERA5 for forecast week 3. The latter is used for the computation of the $WRact_{agg}$ by comparing them against the activity threshold (black dashed line). In the lower panel, for each forecast week, the $WRact_{mean}$ (dark blue bars) and $WRact_{agg}$ (turquoise bars) in ERA5 and in forecasts (purple and green bars respectively) are shown. By definition $WRact_{mean}$ (dark blue bars) for ERA5 takes values of either 0.0 or 1.0.

Figure 1 illustrates the computation process. Seven consecutive daily IWR values from ERA5 reanalysis, shown as turquoise circles in the upper panel, are averaged to compute the weekly mean IWR (blue circles). This value (here 2.1 in week 3) is then

compared to the activity threshold IWR$_{min}$ (black dashed line, here 1.0), determining whether the regime is considered active (1.0) or inactive (0.0). The corresponding $WRact_{mean}$ is represented as blue bar in the lower panel. The $WRact_{agg}$ follows a similar approach, but instead assesses each daily IWR value individually against IWR$_{min}$. The fraction of days exceeding IWR$_{min}$ defines the $WRact_{agg}$, visualised by the turquoise bar in the lower panel.

The same procedure applies to ensemble forecasts, where the IWR forecast for each ensemble member, initialised at the beginning of the shown period, is processed in the same manner (green lines in the upper panel of Figure 1). The resulting $WRact_{mean}$ for the ensemble is displayed as purple bars in the lower panel, while the $WRact_{agg}$ is shown in green. This methodology enables a comprehensive assessment of weather regime activity across both, reanalysis and forecasts. Since this process is applied individually to each weather regime, multiple regimes can be considered active simultaneously, reflecting the inherent variability in large-scale atmospheric patterns.

To further simplify the interpretation of weather regime activity, we introduce the maximum weather regime activity ($WRact_{max}$), which assigns a single dominant regime to each day independent of an IWR threshold. This is achieved by identifying the weather regime with the highest $WRact_{agg}$ within a given period. In case multiple weather regimes share the highest $WRact_{agg}$ at a time, the regime persisting the longest is chosen. By designating the most active regime, $WRact_{max}$ provides a practical and intuitive summary, ensuring that each atmospheric scenario is linked to the prevailing large-scale circulation pattern. With this definition, we aim to offer a simplified yet informative representation of regime dominance, making it accessible to users beyond the meteorological sector, such as decision-makers who rely on large-scale weather variability information.

2.4 | Contingency table and metrics

We define windows of opportunity and validate the $WRact_{max}$ on the basis of contingency tables. We utilise both, traditional and probabilistic contingency tables (following the methodology in Gold et al. (2020)), as well as verification metrics which are derived from these.

Forecast	Weather regime active/inactive	
	Yes	No
Yes	Hit / True Positive (TP)	False alarm / False Positive (FP)
No	Miss /False Negative (FN)	Correct rejections / True Negative (TN)

(a) Traditional contingency table.

Forecast	Weather regime active/inactive	
	Yes	No
Yes	$E_{11} = \mathbf{p} \cdot \mathbf{k}$	$E_{12} = \mathbf{p} \cdot (1 - \mathbf{k})$
No	$E_{21} = (1 - \mathbf{p}) \cdot \mathbf{k}$	$E_{22} = (1 - \mathbf{p}) \cdot (1 - \mathbf{k})$

(b) Probabilistic contingency table.

TABLE 1 General classification terms of (a) the traditional 2×2 contingency table and (b) probabilistic equations for the occurrence of weather regime activities. $\mathbf{p}$ represents a vector of probabilities for a weather regime activity to occur and $\mathbf{k}$ is a vector of binary indicators whether the weather regime activities have occurred at the given dates (see Section 2.4). The operator "·" denotes the inner product of two vectors and "1" the unit vector. The tables are adopted by Gold et al. (2020) and modified.

A traditional contingency table (Table 1a) represents the relationship between forecasted and observed categorical events (e.g., whether a weather regime is active or not). The contingency table classifies outcomes into four categories: Hit (or True Positive, TP), where weather regime activity is forecasted and the weather regime activity occurs; Miss (or False Negative, FN), where no weather regime activity is forecasted but a weather regime activity occurs; False Alarm (or False Positive, FP), where weather regime activity is forecasted but no weather regime activity occurs; and Correct Rejection (or True Negative, TN), where

no weather regime activity is forecasted and no weather regime activity occurs. These categories form the basis for several verification metrics, which are detailed in Table 2. The key metrics include Base Rate, representing the overall occurrence of a weather regime activity; Forecast Rate, which is the proportion of forecasts predicting a weather regime activity; Hit Rate (also known as Recall or Accuracy), which is the proportion of actual weather regime activities correctly predicted; False Alarm Rate, the proportion of incorrect predictions of weather regime activities; Peirce Skill Score, which is the difference between the hit rate and the false alarm rate; Precision, representing the proportion of true positive forecasts out of all positive predictions; and F1-Score, the harmonic mean of precision and Hit Rate, balancing both metrics.

Metric	Formula
Base Rate	$\text{BR} = \frac{TP+FN}{TP+FN+FP+TN}$
Forecast Rate	$\text{FR} = \frac{TP+FP}{TP+FN+FP+TN}$
Hit Rate (Recall, Accuracy)	$\text{HR} = \frac{TP}{TP+FN}$
False Alarm Rate	$\text{FAR} = \frac{FP}{FP+TN}$
Peirce Skill Score	$\text{PSS} = \text{HR} - \text{FAR}$
Precision	$\text{Prec} = \frac{TP}{TP+FP}$
F1-Score	$\text{F1} = 2 \times \frac{\text{Prec} \times \text{HR}}{\text{Prec} + \text{HR}}$

TABLE 2 Collection of verification metrics originating from the contingency table in Table 1 used to identify windows of opportunity and to validate the skill of a categorical $WRact_{max}$ forecast.

An alternative to the traditional contingency table is the probabilistic contingency table (Table 1b), which incorporates probabilistic forecasts (Gold et al., 2020). In this framework, the vector $\mathbf{p} = (p_1, p_2, ..., p_n)$ represents the forecast probabilities for an event (weather regime active) occurring, where n is the total number of dates in the sample. Each p_i is the mean of the binary predictions from all ensemble members for date i, with M being the number of ensemble members, in this study $M = 11$. Specifically, each ensemble member m provides a binary prediction $p_{i,m}$ (which is either 0 or 1), and the overall probability for date i is calculated as

$$p_i = \frac{1}{M} \sum_{m=1}^{M} p_{i,m}. \tag{1}$$

Similarly, the vector $\mathbf{k} = (k_1, k_2, ..., k_n)$ represents the actual binary outcomes, where $k_i = 1$ if the weather regime activity occurs and $k_i = 0$ if it does not occur. The scalar product of the forecast probabilities p_i and the binary outcomes k_i generates the probabilistic outcomes, which are analogous to the traditional contingency table.

In addition to the verification metrics derived from the contingency tables, we use the Mean Squared Error (MSE) to assess the quality of the forecasts (verifying $WRact_{agg}$). The MSE quantifies the average squared differences between the predicted values and the observed values across all data points:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{2}$$

where y_i is the forecasted value and $\hat{y}_i$ is the observed value for the i-th data point. To compare the MSE of a forecast to that of

a reference forecast, we compute the MSE score (MSES):

$$MSES = 1 - \frac{MSE_{forecast}}{MSE_{reference}}.$$

(3)

Here, the MSES represents the normalised performance of the forecast relative to the reference, which in this case is the numerical weather prediction (NWP) forecast. A higher MSES indicates better forecast skill compared to the reference forecast.

2.5 | Windows of opportunity

In this study, we adopt the window of opportunity definition from Specq and Batté (2022). The atmospheric state at initialisation time can be characterised by several indicators, such as the phase of the Madden-Julian Oscillation or the strength of the stratospheric polar vortex.

We categorise the data according to the prevalent phase. Following the window of opportunity definition by Specq and Batté (2022), verification metrics (base rate (BR), hit rate (HR), Peirce skill score (PSS)) are then calculated for each subsample, based on the definitions in Table 1 and Table 2. To assess whether the differences in the weather regime activity after active and inactive phases are statistically significant, we use confidence intervals. Specifically, we calculate a 90% confidence interval for each verification metric during active phases using bootstrapping. This is done by resampling the dates of each phase 1,000 times with replacement. If the corresponding metric for the inactive phase falls outside this confidence interval, we consider the difference to be statistically significant.

Once the verification metrics (BR, HR, PSS) are calculated, we categorise windows of opportunities into two main types: climatological and model windows of opportunities, with the model windows of opportunities further divided into three subcategories. A climatological window of opportunity is based on the historical relationship between the atmospheric state at initialisation time and the likelihood of a weather regime to occur three weeks later as derived from ERA5 reanalysis. A certain atmospheric state is considered to represent a climatological window of opportunity if the frequency of weather regime occurrence – referred to as the base rate – is higher after an active phase than after an inactive phase. By identifying these climatological windows of opportunities, a forecaster can estimate the likelihood of a weather regime activity occurring simply by relying on past observational data corresponding to a given atmospheric condition at initialisation.

In contrast, a model window of opportunity takes into account both the base rate and the ability of the forecasting system (here the ECMWF reforecasts) to detect and benefit from the atmospheric signal. For a model window of opportunity to be useful, it might not be enough for the system to simply identify a higher base rate and hit rate. The system should also demonstrate a reliable forecast quality, which is where the Peirce skill score comes into play. The Peirce skill score assesses the difference between the hit rate and false alarm rate, allowing us to distinguish between forecasts that are genuinely improved and those that may have increased hit rates at the cost of more false alarms. Within the model window of opportunity category, there are three subtypes:

Type 1: Characterised by an increase in both the base rate and the hit rate, but without an improvement in the Peirce skill score due to an excessive false alarm rate. In this case, while the forecast may indicate a higher likelihood of the event, it comes with a higher number of false alarms, reducing overall forecast reliability.

Type 2: Shows improvements in the base rate, hit rate, and the Peirce skill score. This subtype indicates a higher-quality forecast where the hit rate improves while minimising false alarms, resulting in a more accurate and reliable forecast.

Type 3: Exhibit a positive anomaly in both the hit rate and the Peirce skill score compared to the neutral phase, even if this improvement cannot be explained by a climatological signal, due to no increase in the base rate. These phases are still considered valuable because they reflect an improvement in forecast performance over what would be expected without the

active phase.

2.6 | Atmospheric-conditioned climatology

The existence of windows of opportunities under certain atmospheric conditions naturally motivates the computation of a climatology conditioned on the atmospheric state at initialisation time, which we refer to as an atmospheric-conditioned climatology. Unlike an unconditioned climatology, which averages over all historical periods, the atmospheric-conditioned climatology is computed conditioned on each atmospheric state (e.g., a specific MJO phase) individually. It provides the climatological mean of the $WRact_{agg}$ at a specific lag. Since atmospheric variables used for conditioning are sensitive to seasonality, all climatologies are computed using a 91-day sliding window. For each atmospheric state, we select all historical dates within the sliding window that meet the state condition at initialisation. From these dates, we compute the atmospheric-conditioned climatology of $WRact_{agg}$. These climatologies offer insight into how frequently weather regime activity occurs under different atmospheric conditions (see Section 3.2).

Building on the atmospheric-conditioned climatology, we construct an atmospheric-conditioned climatological forecast by selecting the corresponding atmospheric-conditioned climatology for each initialisation date based on the prevailing atmospheric condition. Although this simple, condition-based selection is straightforward and effective, it can lead to abrupt changes in forecasted $WRact_{agg}$ when transitioning between atmospheric states. To address this issue, we introduce a sliding forecast, which smooths these transitions. This approach differs depending on whether the conditioning variable is two-dimensional (e.g., the MJO) or one-dimensional (e.g., the SPV index). Below, we detail the method for the MJO and highlight modifications required for one-dimensional variables.

The Madden–Julian Oscillation is the dominant mode of sub-seasonal atmospheric variability in the tropics (Madden and Julian, 1971). Unlike the commonly used real-time multivariate MJO index (RMM) by Wheeler and Hendon (2004), we adopt the Outgoing Longwave Radiation (OLR) MJO Index (OMI) (NOAA, 2025). The OMI is derived solely from OLR anomalies, without incorporating zonal winds at 850 hPa and 200 hPa. This results in a smoother temporal evolution of the MJO signal compared to RMM.

The sliding atmospheric-conditioned climatological forecast for the MJO is generated as follows (illustrated in Figure 2): First, all historical OMI1 and OMI2 pairs (the first two principal components of the OLR anomaly empirical orthogonal function) within a 91-day sliding climatological window (grey dots) centred on the day of year of the current initialisation date (red diamond) are collected. Rather than selecting only the single MJO phase for the current date, we include all historical dates whose OMI1/OMI2 pairs fall within a 45°arc (blue dots) around the current OMI pair in the phase space. Dates with an OMI amplitude less than 1 (inner circle), representing the MJO inactive phase, are excluded. Additionally, to ensure independence from recent events all events in a windows of ± 45 days (orange dots) are removed from consideration. Finally, the mean $WRact_{agg}$ is computed from the remaining historical dates and used as the MJO-conditioned climatological forecast for the current initialisation.

For one-dimensional atmospheric variables, such as the SPV index, the sliding forecast approach is adjusted accordingly. Instead of using a 45°arc in a two-dimensional phase space, we select the 5% of historical values closest to the current SPV value from within the 91-day sliding climatological window, excluding dates from the current seasonal cycle. The mean $WRact_{agg}$ from these selected historical dates is then used to produce the SPV-conditioned climatological forecast.

2.7 | Neural networks

To combine information from (atmospheric-conditioned) climatological forecasts, NWP forecasts, and recent $WRact_{agg}$, we employ a statistical-dynamical approach represented with fully connected dense neural networks (NN). With these networks,
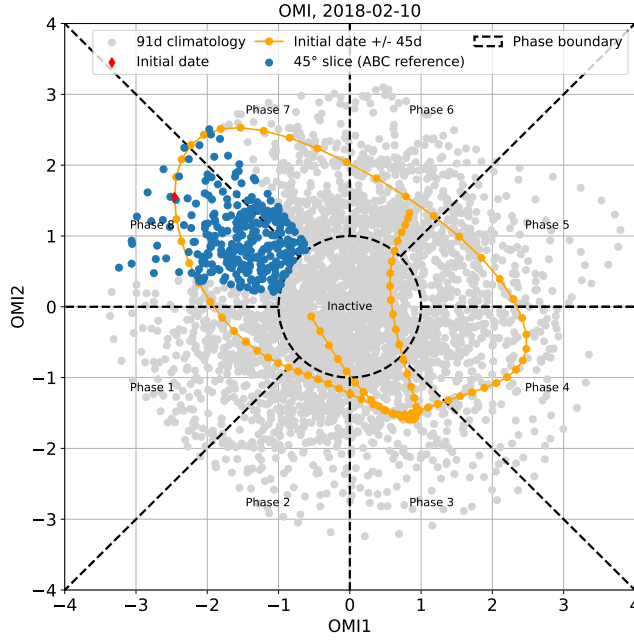
FIGURE 2 Phase diagram of the outgoing long-wave radiation Madden-Julian Oscillation index. The first and second principal components derived from empirical orthogonal function analysis of filtered OLR data are given on the x- and y-axis. Further the phases of the MJO are indicated with dashed lines and labelling. For visualisation how the atmospheric-conditioned climatological forecast works, one initial date, February 10, 2018, is marked with a red diamond and all values in the 45°arc used for the MJO-conditioned climatological forecast of that initialisation date (blue dots). Values in the ±45 days window around the initial date (orange dots) are excluded from the data used for the computation. With grey dots, all other data points of the 91 day running climatology are indicated. For a visualisation of the OMI phase space similar to the real-time multivariate MJO index (RMM1/2), OMI1 is multiplied by -1 and OMI1 and OMI2 are exchanged with each other.

we aim to identify complex relationships between predictors and enhance the predictability of $WRact_{agg}$, particularly for strong weather regime events in forecast week 3. For readers unfamiliar with machine learning techniques and terminology, we recommend the machine learning tutorials by Chase et al. (2022, 2023). A fully connected dense neural network setup is selected after extensive experimentation with different architectures, depths, loss functions, and output configurations. The final models, one model per lead time and weather regime, consist of two hidden layers with 64 and 16 nodes, each using a ReLU activation function. A dropout layer with a rate of 0.2 is inserted between the hidden layers to reduce overfitting. The output layer has a single node with a sigmoid activation function, producing a probability-like output representing the $WRact_{agg}$ for a specific weather regime.

The network is trained for up to 50 epochs, but early stopping is applied to prevent overfitting. Early stopping monitors the validation loss with a patience of 10 epochs, halting training if the validation loss does not improve. Additionally, a learning rate scheduler reduces the learning rate, which starts at 0.001, by a factor of 0.5 after 5 epochs without improvement, with a minimum learning rate of 10^{-6}. During training, the model uses a batch size of 32 and a validation split of 20%, without shuffling the data to preserve temporal dependencies. To robustly assess model performance, we apply 4-fold cross-validation. The 1720 data points from the extended winter period (NDJFM) are split into four consecutive subsets of 430 points each. Each fold trains on 1290 data points (data includes the validation split of 20%) and tests on the remaining 430. To ensure results are not dependent

on a single random initialisation parameter setting, ten fully independent models with identical configurations though different random initialisation parameters are trained simultaneously, and their ensemble mean is used as the final prediction. The model predicts the $WRact_{agg}$ for a single weather regime and one specific lead time, such as GL activity for forecast week 3.

To improve model performance while keeping model complexity low, we apply a stepwise feature selection procedure, restricting ourselves to the most informative predictors and keeping the signal-to-noise ratio low. The stepwise selection begins by training the model using each predictor individually and recording the mean squared error from cross-validation. In subsequent steps, the best-performing predictor from the previous step is retained, and additional predictors are tested one by one. The set of predictors with the lowest cross-validated MSE is selected. This procedure is performed separately for each lead time and weather regime, tailoring the predictor set to the specific forecast horizon and regime dynamics.

To ease interpretability, we group all available predictors into four distinct subcategories. Climatology predictors include a 91-day mean $WRact_{agg}$ climatology, day-of-year (DOY), and atmospheric-conditioned climatological forecasts based on large-scale drivers such as the MJO and SPV. Atmospheric state predictors describe the current state of the atmosphere, including indices representing large-scale circulation patterns such as the MJO and SPV, as well as sea surface temperature anomalies (SSTano) in the North Atlantic. NWP predictors capture forecast-based $WRact_{agg}$ indicators, including trends and variability across different forecast weeks. Recent weather regime predictors represent recent observed $WRact_{agg}$ and IWR from reanalysis data. A comprehensive description of these predictors is provided in Tables S1 and S2.

Predictor pools	NN_{all}	NN_{NWP+WR}	NN_{noNWP}
Climatology	X		X
Atmospheric State	X		X
NWP	X	X	
Recent WR	X	X	X

TABLE 3 Overview of predictor pools used as input predictors for the three neural network configurations: NN_{all} (all predictors), NN_{NWP+WR} (NWP and weather regime predictors), and NN_{noNWP} (excluding NWP predictors). An "X" indicates inclusion of the respective predictor pool.

Using these predictor subcategories, we train three separate neural networks with distinct predictor pools to explore the contributions of different predictor types (Table 3). The climatological neural network NN_{noNWP} uses no data of the NWP model. As it does not rely on forecast data, this model is trained on a larger dataset, consisting of daily data from 1979 to 2020, though it is evaluated on the same test set as the other models for fair comparison. The NWP and weather regime neural network NN_{NWP+WR} uses only NWP and weather regime based predictors, such as NWP weather regime forecast evolutions and recent past $WRact_{agg}$ and IWR from ERA. This model is trained exclusively on the reforecast period. Finally, the all-predictors neural network NN_{all} combines all available predictors, integrating (atmospheric-conditioned) climatology, NWP, and recent past weather regime information. This model provides insight into the added value of integrating all sources of predictability.

3 | RESULTS

We approach our investigation of windows of opportunity in subseasonal weather regime forecasting with an analysis of the modulation of occurrence and forecast quality for Greenland Blocking by different states of the MJO and SPV (Section 3.1). Next, we utilise climatological knowledge of the atmospheric state to produce an atmospheric-conditioned climatological forecast for the occurrence of Greenland Blocking (Section 3.2). Finally, we integrate information from numerical weather prediction regarding $WRact_{agg}$ with knowledge about the atmospheric state at initialisation, employing a fully connected dense neural network to enhance activity forecasts and expand our analysis on other weather regimes (Section 3.3).

The research presented here focuses exclusively on the extended winter period (November to March) and a forecast lead time of three weeks.

3.1 | Windows of opportunity

In our setting, a window of opportunity indicates whether $WRact_{mean}$ is more likely to occur or to be forecasted three weeks after the initialisation date, conditioned on the state of the MJO or SPV phase at forecast initialisation time. To identify these opportunities, we analyse the base rate of $WRact_{mean}$ following a given phase, along with the hit rate and Peirce skill score (see Section 2.5).
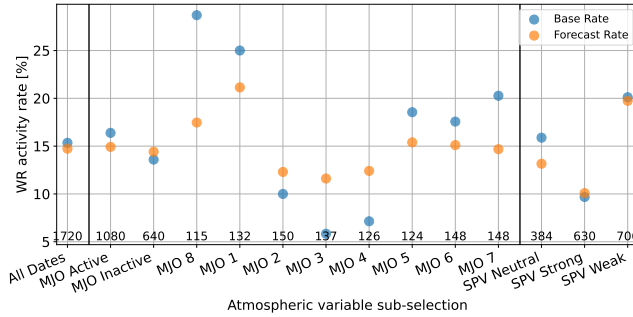

FIGURE 3 Base rate (blue dots) and forecast rate (orange dots) for weekly mean Greenland Blocking activity ($WRact_{mean,GL}$) in the extended winter period (November–March). These rates are split into subsets based on phases of the MJO or SPV with a lead time of three weeks. The size of the subset is indicated at the bottom of the respective phase-column.

As a preliminary step, we compare the base rate and forecast rate of weekly Greenland Blocking activity three weeks after specific states of the MJO and SPV (Figure 3). For an analysis across all forecasting weeks see Figure S1 in the Supplementary Material. On average, throughout the analysis period, the base rate and the forecast rate of $WRact_{mean,GL}$ are approximately 15%, with the forecast rate slightly lower than the base rate, indicating a slight negative bias in $WRact_{mean,GL}$ and thus in GL frequency as found earlier (e.g., Osman et al., 2023). An analysis conditioned on the MJO phases reveals that during the inactive MJO phase, the base rate and forecast rate align closely, both at approximately 14%. The base rate is generally higher for MJO phases 5–8 and 1, while it is lower for phases 2–4. Therefore, the phases 5–8 and 1 can be considered as climatological windows of opportunities (phases 7, 8 and 1 identify as significant for GL, according to the bootstrap test, see Section 2.5). Notably, forecast rates are always closer to the climatological base rate and the inactive phase than the respective base rates. This discrepancy may be attributed to the three-week forecast lead time, as extended-range forecasts tend to regress toward climatological values. A similar behaviour between the base rate and forecast rate can be observed for strong and weak SPV
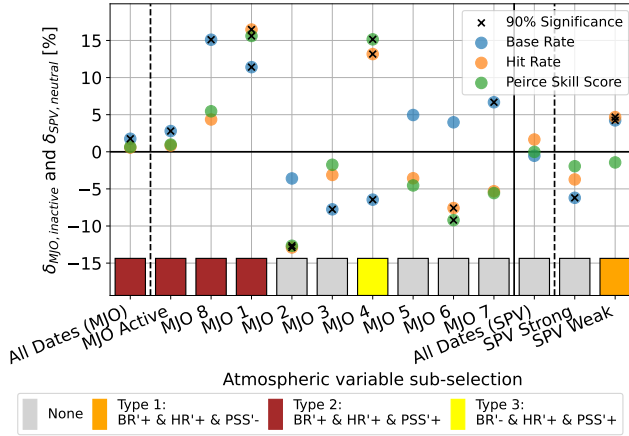
FIGURE 4 Indication of windows of opportunity for Greenland Blocking activity in week 3 by the base rate, hit rate and Peirce skill score anomalies and colour-coded indication. Similar to Figure 3 the metrics are separated by the state of the MJO and the SPV. Here, the base rate, hit rate and Peirce skill score (blue, orange, and green dots, respectively) are anomalies (δ in %) respective to the neutral phases of the respective atmospheric variables. The combination of positive and negative anomalies (denoted with a prime) for the base rate (BR'), hit rate (HR') and Peirce skill score (PSS') are responsible for the model window of opportunity type indicated by the coloured boxes at the bottom with its legend below. Further crosses in the dots for anomalies indicate whether the rates in the specific categories are significantly different to their neutral phases with a 90% confidence interval.

states, although the differences are negligible. The difference between the base rate and forecast rate for the neutral SPV phase is larger, which could be an indicator that the model is able to represent the occurrence of $WRact_{mean,GL}$ better when an anomalous SPV state (either strong or weak) is present at initialisation time. The positive base rate of $WRact_{mean,GL}$ after phases of weak SPV, with a significant increase compared to the neutral SPV phase, makes the weak SPV state a climatological window of opportunity. One can also consider the strong SPV state as a climatological window of opportunity but for a reduced occurrence of Greenland Blocking, which is also captured by the forecast rate.

Overall, Figure 3 suggests that while the forecasted frequency of weather regime activity aligns with observed occurrences, discrepancies remain in capturing the observed $WRact_{mean,GL}$ three weeks after a specific atmospheric state. Large differences exist particularly when separating by MJO phases at initialisation, with the mean frequency of $WRact_{mean,GL}$ ranging from 6% following MJO phase 3 to 28% following MJO phase 8. The absolute values of the base rates and forecast rates should be treated with care. The sample size for individual MJO phases, where an NWP reforecast is initialised, is small (1720 reforecasts considered across 21 winter seasons, sample size per phase is indicated at bottom of Figure 3). MJO phases 7,8 and 1 as well as the weak SPV can be considered as climatological window of opportunity with significantly different base rates compared to their neutral phases, marking them as promising phases for further analysis to identify whether these climatological signals also translate into model windows of opportunities.

Following the approach described in Section 2.5, we now analyse the model windows of opportunities by computing the base rate, hit rate, and Peirce skill score for each state of the MJO and SPV. We then compute rate and skill score anomalies (denoted with a prime) relative to the inactive MJO and neutral SPV, respectively (Figure 4): $R'_{state} = R_{state} - R_{ref}$, where $R \in BR, HR, PSS$ and ref indicates either the inactive or neutral phase.

The HR' and PSS' values closely align for almost all MJO and SPV states, suggesting that the false alarm rates (FAR) in these sub-selections are similar to those of the neutral phases (not shown). The only exception is the weak SPV state during which

PSS' is smaller than HR', indicating that the FAR is greater than for the neutral phase. As shown in Figure 3, the BR' values are positive for MJO phases 5–8 and 1. Furthermore, for phases 8 and 1, both HR' and PSS' are positive, indicating that these phases act as a type 2 model window of opportunity, where BR', HR', and PSS' are all positive. Note that only for MJO phase 1 all three anomalies are significantly different to the neutral phase on a 90% confidence interval. MJO phase 4 is classified as a type 3 model window of opportunity for $WRact_{mean,GL}$ with a three-week lead time. In this case, HR' and PSS' is distinctly positive (significant difference to neutral phase), despite the BR' being negative. This suggests that, although $WRact_{mean,GL}$ is significantly less frequent following phase 4, the forecast accurately predicts these fewer occurrences, with an even lower FAR than in the neutral phase. The signals associated with SPV phases are less distinct, likely due to the broad classification into only three categories (strong, neutral and weak SPV) which encompasses a wide range of individual cases. Further, with the SPV index being a scalar, no information about the further temporal evolution of the index is provided. Three weeks after a weak SPV, $WRact_{mean,GL}$ is increased (positive BR'), and the forecast also predicts more frequent activity, leading to a higher HR. However, the FAR also increases, which reduces PSS'. In fact, PSS' is negative, indicating the HR/FAR ratio is worse than for the neutral SPV phase, categorising this as a type 1 model window of opportunity. In contrast, strong SPV phases are followed by less frequent $WRact_{mean,GL}$, along with a decrease in HR, PSS, and FAR, ultimately resulting in no identifiable window of opportunity type. The significant positive and negative BR' for the weak and strong SPV state, respectively, indicate that there is a strong link between the SPV and the occurrence of a weather regime activity. The negative PSS' (and therefore the higher FAR) following the weak SPV state indicate that the model has trouble predicting GL activities correctly following an extreme state of the SPV. The model shows a GL activity response too often following a weak SPV state, even when it does not realise.

In summary, the frequency of $WRact_{mean,GL}$ (base rate) varies across different MJO and SPV states. It is particularly high (significantly different to the neutral phase) for MJO phases 8 and 1, as well as after weak SPV conditions. The magnitude of the forecast rate is generally too low. MJO phases 8 and 1 emerge as type 2 model windows of opportunities, meaning that not only is the base rate increased for $WRact_{mean,GL}$ three weeks after these phases, but both the hit rate and Peirce skill score are also improved compared to the neutral phase. All three values are significantly different to the neutral phase for MJO phase 1. Forecasts initialised during a weak SPV phase show an improved hit rate, but the false alarm rate also increases compared to the neutral phase, classifying this as a type 1 model window of opportunity with significance. Finally, MJO phase 4 is a model window of opportunity for correct prediction of less GL activity.

3.2 | Atmospheric-conditioned climatological weather regime activity forecasts

The previous section and existing literature (Baldwin et al., 2001; Beerli and Grams, 2019) suggest that the $WRact_{agg,GL}$ three weeks after forecast initialisation is influenced by the prevailing states of the MJO and SPV. To better understand this relationship, an atmospheric-conditioned climatology is introduced with a 91 day sliding window over extended winter. This climatology expresses the likelihood of $WRact_{agg}$ three weeks after a given MJO or SPV phase, demonstrating that phase-specific climatological frequencies of GL occurrence differ significantly from the overall climatology and that this has important intra-seasonal variability (Figure 5). During the extended winter period, the 91-day running mean of the $WRact_{agg,GL}$ fluctuates between 15 and 18% (black lines). When categorised by MJO phases (Figure 5a), notable deviations from this baseline emerge. The most prominent signal appears for MJO phase 8, where $WRact_{agg,GL}$ increases from 11% in November to 30% in February (brown line). Similarly, MJO phase 1 (red line) exhibits an increase from 15% in November to 22% by late December, maintaining an elevated activity level of approximately 20% until March. For SPV states (Figure 5b), the separation is even more pronounced. The neutral phase (grey line) largely follows the overall climatological cycle, while conditions following a weak SPV (blue line) show increased $WRact_{agg,GL}$, with two maxima reaching 24% in January and March. Conversely, when a strong SPV (red line) is present, $WRact_{agg,GL}$ decreases, reaching a minimum of 7% in January.

To use the atmospheric-conditioned climatology as a forecast tool, a slight modification is introduced to the computation

(a) MJO-conditioned climatology.
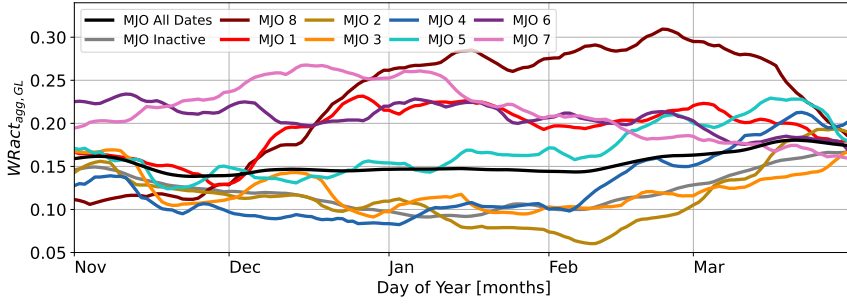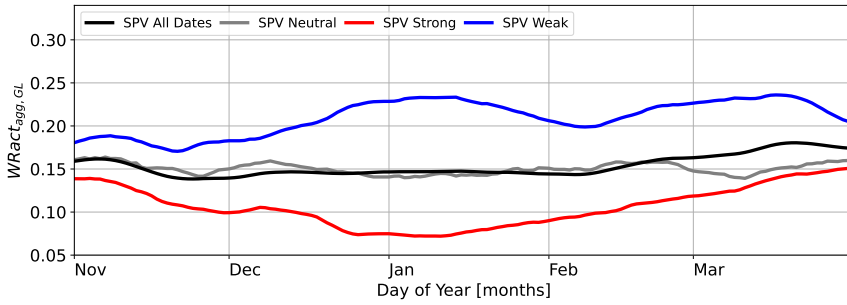

(b) SPV-conditioned climatology.

FIGURE 5   Atmospheric-conditioned climatology for $WRact_{agg,GL}$ with a time delay of three weeks after specific states of the (a) MJO and (b) SPV for the extended winter period (November–March). The black line in both sub-figures indicate the 91-day running climatology without any atmospheric-conditioned sub-selection.

method. Instead of strictly categorising the atmospheric-conditioned climatology by discrete phases, a sliding window approach is applied. This adjustment ensures a smoother transition between categorical phases, preventing abrupt shifts in activity forecast (see Section 2.6 for further details).

As a proof of concept, we select periods of active GL and analyse atmospheric-conditioned climatological and NWP forecasts leading up to these GL activities. While this approach may seem selective, as it focuses only on active GL periods, it enables a targeted examination of two key forecasting questions:

1. Do single forecast runs of the NWP model and atmospheric-conditioned climatological forecasts indicate the onset of $WRact_{agg,GL}$ in week 3 and the subsequent evolution of $WRact_{agg,GL}$?
2. Do consecutively initialised week 3 forecasts show an increasing signal leading up to the $WRact_{agg,GL}$?

To address these questions, two mean composite plots are computed (Figure 6). For both analyses, the actual GL activity onset is defined as the first time when the ERA5 $WRact_{mean,GL}$ reaches one. For the first question, individual forecasts initialised three weeks prior to GL activity onset (or as close as possible, given the availability of NWP reforecasts) are analysed with lead times ranging from 0–36 days (Figure 6a). For the second question, only those forecasts at a fixed lead time of three weeks (18 days) are considered. Thus, instead of a single forecast per event, consecutive forecasts leading up to the GL activity are analysed to track the signal of GL activity over time (Figure 6b). The mean composite of forecasts in Figure 6 is computed centring all

(a) Forecasts (0–36 days lead time) initialised three weeks (18 days) prior to GL onsets. (b) Forecasts at a fixed lead time of 18 days (3 weeks) at valid times ±18 days around GL onsets.
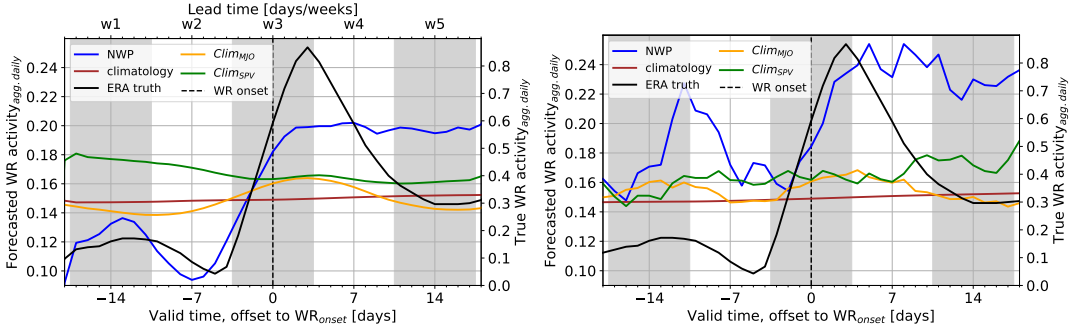

FIGURE 6 Mean composite of different forecasts at valid times centred around observed Greenland Blocking onsets (black dashed vertical line). Due to values on different scales, forecast values for the $WRact_{agg}$ are indicated on the left y-axis, the mean of the ERA5 $WRact_{agg}$ are indicated on the right y-axis. In (a), the mean composite is generated by forecasts (0–36 day lead time) initialised three weeks (18 days) prior to GL onset, whereas in (b), the forecasts have a fixed lead time of three weeks (18 days) for each valid time ±18 days around the GL onset (x-axis). For a fair comparison, between the atmospheric-conditioned climatological and NWP forecasts, we restrict the data for the computation of the mean composite to the extended winter period from 1999–2020, the period in which NWP forecasts are available. A Greenland Blocking onset is defined as the time step where the $WRact_{mean}$ for Greenland Blocking changes from 0 to 1.

events around the GL onset. For the individual events see Figures S2 and S3 in the Supplementary Material.

The true (ERA5) $WRact_{agg,GL}$ (black line, right y-axis) is identical in both sub-figures of Figure 6. As expected, there is a clear increase in $WRact_{agg,GL}$ around the onset (vertical black dashed line), peaking three days after the onset. Additionally, a secondary local maximum appears around 12 days prior to the GL onset, indicating that in some of the scenarios, two GL events occur in close succession (e.g., GL event indices 129, 137, 181 in Figure S2). This secondary maximum may also indicate that the state of WR activity prior to, at, and after initialisation carries valuable predictive information. Due to the centring of the data around the activity onset, the mean composite shows actual $WRact_{agg,GL}$ reaching up to 85%, whereas forecasts only reach up to 25% of $WRact_{agg,GL}$ in the mean composite. Therefore, the $WRact_{agg,GL}$ forecast values are displayed on a separate y-axis for better visualisation (left y-axis).

The unconditional climatology (brown line) serves as a reference, showing no distinct signal apart from the seasonal base rate of approximately 15% $WRact_{agg,GL}$. Comparing the MJO- and SPV-conditioned climatological forecasts (orange and green lines, respectively) to the NWP model (blue line), the NWP forecast exhibits the strongest $WRact_{agg,GL}$ signal. This result is not surprising, as the NWP model incorporates full dynamical atmospheric information, while the atmospheric-conditioned climatological forecasts rely solely on the MJO or SPV phase at initialisation. Thus, rather than comparing absolute values, the temporal changes of the forecast signals relative to the unconditional climatology are the primary focus.

For forecasts initialised three weeks prior to GL onsets (Figure 6a), the NWP forecast shows a strong increase in $WRact_{agg,GL}$ beginning approximately seven days before onset, mirroring the observed activity trend. An additional increase appears at earlier lead times, specifically 11–17 days before onset, aligning with the secondary maximum observed in the ERA5 $WRact_{agg,GL}$. At longer lead times (7–18 days after onset), the NWP forecast suggests a prolonged high $WRact_{agg,GL}$, even when the observed activity declines. This behaviour is likely an artifact caused by forecast timing errors, where some forecasts predict $WRact_{agg,GL}$ too late (e.g., GL indices 142, 166, 202 in Figure S2). As a result, in the mean composite, $WRact_{agg,GL}$ appears misleadingly prolonged in the forecast.

The MJO-conditioned climatological forecast impressively captures the key characteristics of $WRact_{agg,GL}$ evolution. At earlier lead times (up to five days before onset), $WRact_{agg,GL}$ is lower than the unconditional climatology. However, from 4 days before to 10 days after onset, forecasted $WRact_{agg,GL}$ exceeds the unconditioned climatology, indicating an increased likelihood of $WRact_{agg,GL}$. This result reinforces previous findings that MJO phases modulate $WRact_{agg,GL}$ two to four weeks later, validating the teleconnection mechanism.

In contrast, the SPV-conditioned climatological forecast produces a less distinct signal. Forecasted $WRact_{agg,GL}$ remains higher than the unconditional climatology across all lead times, yet the highest values appear at early lead times, followed by a steady decline, with only a minor incline 0–7 days after the weather regime onset. One possible explanation is that the one-dimensional SPV index lacks the complexity of the two-dimensional MJO index (amplitude and phase angle). Previous studies (e.g., Domeisen et al., 2020) suggest that sudden stratospheric warming events (low SPV index values) can trigger concurrent GL activity, but not all GL activities are linked to SSWs. Since the SPV-conditioned climatological forecast does not account for changes in the index over time, it merges scenarios where the atmosphere is both approaching and departing from an SSW event, leading to a diffuse climatological signal. Incorporating temporal changes in the SPV index could improve the clarity of the forecast signal and better represent the dependence between the SPV index at initialisation and the $WRact_{agg,GL}$ with a lag in time. To account for this, we introduced a 2D index that incorporates the SPV change over the past seven days. However, integrating this into the atmospheric-conditioned climatological forecast did not yield noticeable improvements (therefore not shown).

When analysing consecutive forecasts with a fixed 3-week lead time (Figure 6b), similar signals emerge, though with distinct differences. The NWP forecast still predicts GL activity onsets, indicated with an activity increase starting for valid times two days prior to the onset. In contrast to the single forecast runs, the $WRact_{agg,GL}$ forecasts prior to the weather regime onset are already higher than climatology. The delayed signal increase compared to ERA5 (two days rather than five days prior to the onset) and the later maximum of the $WRact_{agg,GL}$ (five days rather than three days after the onset) suggest a systematic delay in predicted onset timing. Additionally, the previously active GL activity is still visible in the forecasts by the NWP (7–14 days prior to the GL onset). This could also indicate that the NWP model favours the persistence of a GL activity rather than the transition into another weather regime. The atmospheric-conditioned climatological forecasts continue to provide meaningful signals. The MJO-conditioned climatological forecast exhibits a clear increase relative to the unconditional climatology, beginning three days before the onset and persisting for up to ten days after the onset. The SPV-conditioned climatological forecast, while more diffuse, shows a gradual increase across the forecasted periods with all values being above climatological forecast values, supporting the notion that the SPV phase influences GL activity, albeit less distinctly than the MJO phase.

These results demonstrate that atmospheric-conditioned climatologies offer a valuable framework for forecasting $WRact_{agg,GL}$. The MJO-conditioned climatological forecast exhibits a robust increase in $WRact_{agg,GL}$, aligning well with observed $WRact_{agg,GL}$ and confirming the role of the MJO in modulating $WRact_{agg,GL}$ two to four weeks later. The SPV-conditioned climatological forecast also indicates a weak modulation of GL activity, though it is more diffuse and less clear to interpret. While the NWP model predicts $WRact_{agg,GL}$ well, it exhibits a systematic delay of predicting the onset. Overall, these findings support the use of atmospheric-conditioned climatological forecasts as a complementary tool to dynamical NWP models. However, further refinements, particularly incorporating temporal changes in the SPV index, could enhance predictive accuracy. It is important to emphasise that these findings provide only a partial interpretation of the forecast signals, as the analysis is limited to mean composites centred around actual ERA5 $WRact_{agg,GL}$ onsets. The forecasts within this subset exhibit inherent variability, and the composites do not account for scenarios where, for example, the NWP model predicts $WRact_{agg,GL}$ that does not verify.

3.3 | Statistical-dynamical approach

Given the promising results observed in $WRact_{agg,GL}$ forecasts conditioned on the MJO and SPV, as well as the performance of the NWP model, we explore whether these forecasts –along with additional atmospheric variables– can be combined to enhance the NWP model's skill in predicting $WRact_{agg}$ three weeks in advance. To achieve this, we employ statistical-dynamical models using fully connected neural networks (as described in Section 2.7) to predict $WRact_{agg}$ at week 3.

To systematically assess the impact of different information sources, we construct three neural networks based on distinct predictor pools (see Table S1 and S2 for a complete list of predictors). The "NWP and weather regime" neural network (NN_{NWP+WR}), combines information from the NWP model with recent $WRact_{agg}$. The "no NWP" neural network (NN_{noNWP}), incorporates all available predictors except those derived from the NWP model. The all-predictors neural network (NN_{all}) utilises the full set of available predictors, including climatological forecasts, atmospheric indicators at initialisation, NWP-based weather regime forecasts and temporal changes, and $WRact_{agg}$ history.

Each neural network undergoes a stepwise feature selection process (see Section 2.7) to identify the most relevant predictors. Once the predictors yielding the highest skill are determined, models are retrained using k-fold cross-validation, ensuring a robust evaluation across multiple forecast instances. While this approach introduces minor information leakage – since the predictor selection is performed on the mean over all folds before final model training on the same folds – the effect is negligible and outweighed by the benefit of a larger testing dataset.

Our analysis begins with the predictor selection for the NN_{all} in predicting $WRact_{agg,GL}$ for Greenland Blocking, followed by a comparison of the predictor order across all weather regimes. The most influential predictor for $WRact_{agg,GL}$ in NN_{all} is the NWP model's week-three mean IWR forecast, followed closely by its $WRact_{agg}$ forecast of the NWP model for the same week, which is inherently derived from the IWR forecast (Figure S4). The second most important predictor is the NAO index at initialisation. As GL strongly correlates with the NAO–, this likely highlights the persistence of GL activity. The next two predictors include the GL climatological occurrence and the day of the year, both are indicators that there is a seasonal cycle of the GL occurrence within the extended winter period on which the neural network can rely on. Other key predictors include the atmospheric-conditioned climatological forecasts and actual indices of the SPV and QBO, two major atmospheric modes of variability influencing the GL activity (as shown in the Introduction (Section 1) and in the previous results for SPV (Section 3.1 and 3.2). Interestingly, MJO-related predictors are absent in NN_{all} for $WRact_{agg,GL}$, despite strong signals in prior analyses.

Extending the predictor selection analysis to all seven weather regimes (Figure 7) reveals that, except for GL, the NWP forecast of the $WRact_{agg}$ (target variable) remains the dominant predictor. NWP-derived predictors designed to mimic a human forecaster's analysis – such as the temporal change of the IWR across consecutive forecasts or its evolution in week two – play a minor role in the neural networks. The week-two trend is the last predictor selected for AT, EuBL, and ScBL, while temporal changes from previous forecasts with the same valid time are excluded entirely. Similar to GL, the ZO network (representing the NAO+ phase, the counterpart to GL/NAO–) includes climatology and DOY. However, here, the MJO phase, the QBO-conditioned climatological forecast, and the SPV index take precedence over climatology and DOY. An unexpected but recurrent predictor is the Antarctic Oscillation (AAO). It ranks second for EuBL, third for ScBL, eighth for AT, fourth and fifth (atmospheric-conditioned climatological forecast) for AT and GL, highlighting its potential link to $WRact_{agg}$. The AAO has been associated with its Northern Hemisphere counterpart, the Arctic Oscillation (AO), which is closely linked to the NAO. Tachibana et al. (2018) demonstrate that the AAO and AO exhibit correlations, particularly in October and February, while Song et al. (2009) find that negative AAO phases correspond to anomalously high 300 hPa geopotential heights over the North Atlantic-European region, with a lag of 25–40 days. This is consistent with the AAO being a good predictor for EuBL and ScBL in our neural networks.

The highly variable predictor selection across neural networks for different weather regimes highlights the diverse sources of predictability. Key predictors stem from the Arctic (AO-conditioned climatological forecast for AR and ScBL), the stratosphere
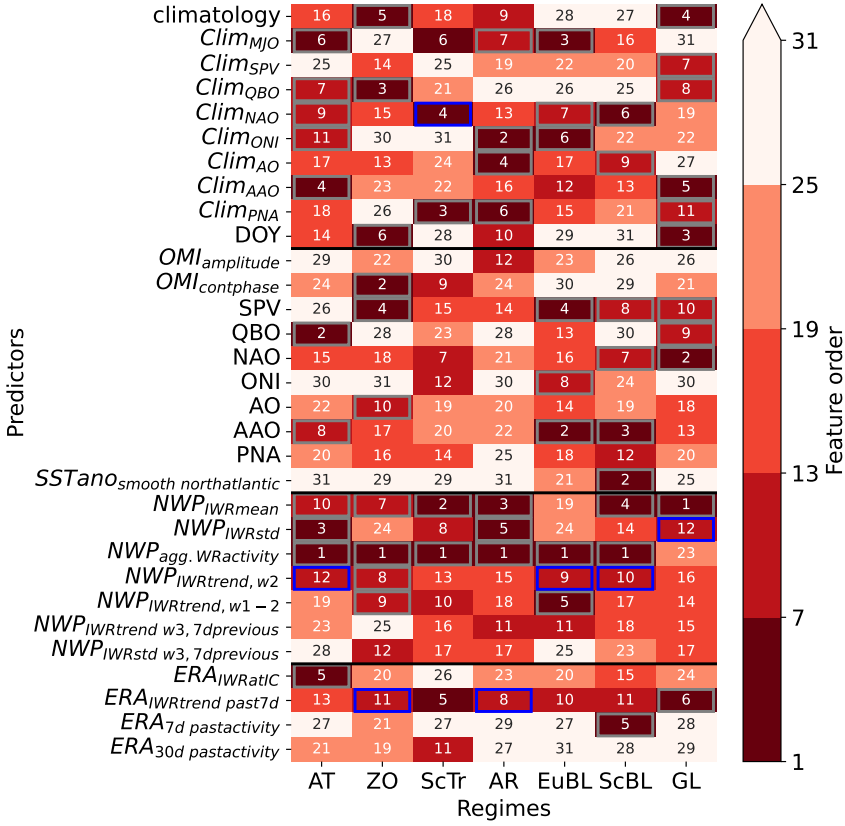
FIGURE 7 Summary of the predictor order for all-predictor neural networks (NN_{all}). The predictor order is shown for all seven weather regimes. Blue boxes indicate the last predictor which is considered due to the best MSE for the predictor combination of all predictors with a lower predictor order number (indicated with grey boxes). A detailed explanation of the predictors can be found in Table S1 and S2.

(SPV-related predictors for ZO, EuBL, ScBL, and GL; QBO-related predictors for AT, ZO, and GL), and even the Southern Hemisphere (AAO-related predictors for AT, EuBL, ScBL, and GL; El Niño-related predictors for AT, AR, EuBL). This diversity underscores the complexity of weather regime forecasting in the North Atlantic-European region, where no single factor – aside from NWP predictors – dominates $WRact_{agg}$ across all regimes.

Following the predictor selection, we evaluate the neural networks' performance in forecasting the $WRact_{agg}$ (Figure 8). Comparing the MSE skill score ($MSES_{NWP}$) across all forecasts in the extended winter season provides key insights into the statistical-dynamical model performance relative to both the NWP model and climatology. The climatological forecast, based on a standard 91-day running mean of $WRact_{agg}$, serves as a baseline for assessing NWP skill. Consistent with previous research (e.g., Osman et al., 2023; Büeler et al., 2021), the NWP model struggles to outperform climatology for ScBL and, particularly, EuBL at week 3, while performing well for ZO and GL. All three neural network setups generally outperform the NWP model for EuBL and ScBL, with NN_{all} showing the most pronounced improvements of 9.7%. Notably, NN_{noNWP} surpasses the NWP model for EuBL and ScBL, demonstrating that even without NWP-derived inputs, it achieves lower mean squared errors in $WRact_{agg}$ forecasts. Meanwhile, NN_{NWP+WR}, provides only minor improvements compared to the NWP
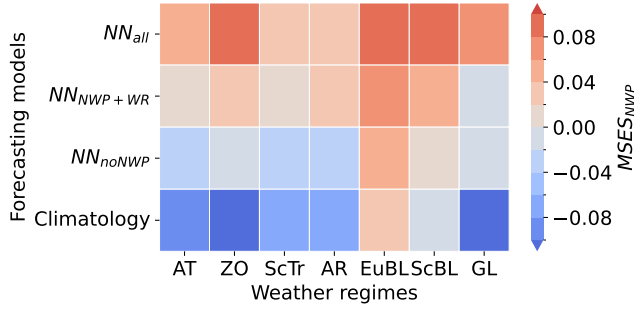
FIGURE 8 Performance of the different neural network and climatological forecasts in comparison to the NWP forecast for $WRact_{agg}$. As performance measure, the MSE is calculated for each model and weather regime and put into perspective against the MSE of the NWP model (MSE skill score).

model, suggesting that it mainly corrects biases rather than introducing new sources of skill. The most striking result comes from NN_{all}, which consistently outperforms the NWP model across all weather regimes, achieving MSE reductions of 3.0–9.7%, with the largest gains observed for ScBL activity. Furthermore, NN_{all} is the only neural network that enhances the already high forecast skill of $WRact_{agg,GL}$. These improvements go beyond simple error reduction; as illustrated in the $WRact_{agg}$ forecasts (Figure S5), some neural networks better capture key $WRact_{agg}$ patterns than the NWP model alone. Notably, NN_{noNWP} successfully captures the prolonged GL activity during the winters of 2009/2010 and 2010/2011. Additionally, for forecasts with the largest x% differences in predicted activity between NN_{all} and NWP, NN_{all} consistently achieves lower MSE, regardless of the magnitude of the difference (not shown here).

In the final stage of our analysis, we evaluate whether the improvements observed for individual $WRact_{agg}$ forecasts translate into better predictions of the $WRact_{max}$ three weeks after forecast initialisation. We determine the weather regime with the maximum predicted activity for both the NN_{all} and the NWP model, then compare their performance using verification metrics derived from the contingency table (Table 1a and Table 2), including accuracy (hit rate), false alarm rate, precision, and F1-score (Figure 9). Across all metrics, a rather uniform image emerges, the NN_{all} performs equal or better for all weather regimes except the ScTr. The hit rate improves for all weather regimes except ScTr, with the overall accuracy increasing from 28.7% for the NWP model to 31.6% for the NN_{all}, representing a relative improvement of 10%. False alarm rates decrease for ZO, ScTr, EuBL and ScBL, stay the same for GL, and increase for AT and AR. The precision and F1-score improve for all weather regimes except the F1-score for ScTr, which is due to the decrease in the hit rate.

In conclusion, neural networks can extract additional predictive information beyond what is provided by the NWP model alone. While the NWP-based $WRact_{agg}$ forecast (or IWR forecast for GL) remains the most influential predictor, added value comes from atmospheric indicators such as the SPV, QBO, MJO, and AAO indices. The NN_{all} not only outperforms the NWP model for individual weather regime forecasts (based on the MSE) but also achieves a 10% relative improvement in predicting the dominant weather regime at week 3. Furthermore, the NN_{noNWP}, which excludes NWP-derived inputs, performs equally well or better for blocked regimes (AR, EuBL, ScBL, GL) highlighting the importance of non-NWP sources of predictability. These findings underscore the potential of hybrid forecasting approaches that integrate dynamical and statistical models to enhance sub-seasonal $WRact_{agg}$ predictions.
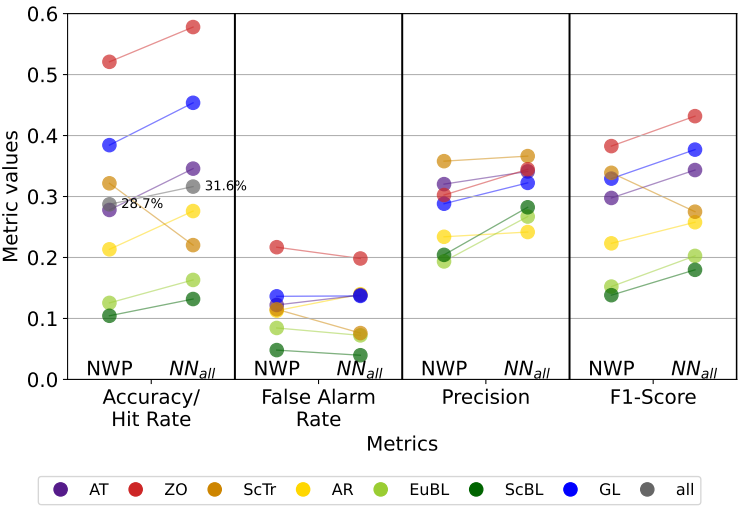
FIGURE 9 Verification metrics comparing the skill of the NWP model and the all-predictors neural network in predicting the $WRact_{max}$ at a lead time of three weeks. Four metrics, the hit rate, the false alarm rate, the precision, and the F1-score are visualised (x-axis). Further together with the hit rate, the accuracy for the full time series is given and indicated by percentages in text. For each metric, on the left the NWP model and on the right the all-predictors neural network (NN_{all}) is represented, joint by a thin line which serves as a guiding line for the reader to directly see whether the respective metric increases or decreases for the neural network in comparison to the NWP model.

4 | CONCLUSIONS AND DISCUSSIONS

Various studies documented the connection between subseasonal sources of predictability (e.g., MJO or SPV) and the large-scale circulation in the North Atlantic-European region (e.g., Cassou, 2008; Lee et al., 2019, 2020; Domeisen et al., 2020; Roberts et al., 2023). However, only few studies so far could demonstrate its practical applicability to improve extended-range NWP forecasts (e.g., Scaife et al., 2022).

In this study, we investigate the influence of the state of the MJO and SPV at initialisation on the occurrence and forecast of seven year-round North-Atlantic European weather regimes, with a focus on the activity of Greenland Blocking, three weeks later. Furthermore, we leverage information describing the atmospheric state prior to and at initialisation to enhance NWP week 3 weather regime activity forecasts using neural networks.

We compute a weather regime activity metric ($WRact_{mean}$ and $WRact_{agg}$), representing the fraction of daily weather regime index values in a given week exceeding a given threshold, based on reanalysis and reforecasts. Our findings show that Greenland Blocking $WRact_{mean,GL}$ is significantly enhanced following MJO phases 7, 8 and 1, as well as weak SPV state, suggesting a climatological window of opportunity (base rate anomaly to neutral state is positive). However, ECMWF reforecast hit rates show improvement over the respective neutral phase only in MJO phases 8 (though not significant) and 1 and in the weak SPV phase. A positive Peirce skill score in MJO phases 8 and 1 classifies them as a model window of opportunity type 2 (according to Specq and Batté (2022), with a positive anomaly in base rate, hit rate and Peirce skill score with respect to neutral state), while the weak SPV state corresponds to a model window of opportunity type 1 (base rate and hit rate anomalies to neutral state are positive and Peirce skill score anomaly is negative). Greenland Blocking is rare following the MJO phase 4, though the NWP model is performing well in correctly predicting these rare Greenland Blocking, classifying MJO phase 4 as a window of opportunity type 3 (hit rate and Peirce skill score anomalies are positive and the base rate anomaly is negative). We demonstrate the potential skill for predicting Greenland blocking activity in week 3 purely derived from reanalysis data by constructing an atmospheric-conditioned climatological forecast that utilises historical information of the MJO state (amplitude and phase) or SPV state to estimate the mean historical $WRact_{agg,GL}$ for the prevailing situation. Incorporating this information – alongside MJO, SPV, and other atmospheric variability indices – with NWP-derived $WRact_{agg}$ metrics in a statistical-dynamical model using neural networks improves $WRact_{agg}$ forecasts across all seven weather regimes. When aggregating individual $WRact_{agg}$ forecasts to determine the dominant regime ($WRact_{max}$), the best-performing neural network achieves accuracy increases for all regimes individually, except the Scandinavian Trough. The overall accuracy increase of the neural network is 2.9%, representing a relative improvement of 10% compared to the NWP model.

Our findings on windows of opportunities align with previous studies on the MJO-NAO (Cassou, 2008; Vitart, 2017; Ferranti et al., 2018) and SPV-NAO (Büeler et al., 2020; Baldwin et al., 2001; Beerli and Grams, 2019) connection. In particular, we find an increased climatological occurrence of Greenland Blocking following weak SPV states, a relationship that is well captured by the ECMWF forecast model. This supports the results of Spaeth et al. (2024), who show that ensemble forecasts exhibit greater confidence in predicting the dominant regime – typically Greenland Blocking – after a weak SPV state compared to neutral or strong states. However, our results also reveal a downside: the NWP model tends to over-predict Greenland Blocking in these situations, leading to - as we can show - frequent false alarms. Vice-versa the NWP model shows poor performance in predicting the relatively rare Greenland Blocking occurrence during strong SPV states. These limitations are likely related to the findings of Domeisen et al. (2020), who show that the tropospheric response to a sudden stratospheric warming (i.e., a weak SPV state) can bifurcate depending on the tropospheric conditions at the time of the warming. It is plausible that the model lacks the ability to distinguish between these different pathways of stratospheric forcing.

To our knowledge, this study is the first to demonstrate how the MJO/SPV-NAO relationship can be harnessed as a decision-support tool for extended-range GL forecasts, based on purely statistical information and independent of NWP models. Additionally, our neural network approach shows particular value for predicting Blocking over Europe (EuBL and

ScBL). Consistent with previous findings that NWP models struggle with these regimes – often performing worse than simple climatological forecasts (e.g., Büeler et al., 2021; Osman et al., 2023; Wandel et al., 2024) – we find that two of our neural networks, one excluding and one including NWP information, both outperform the raw NWP forecasts for these blocked regimes. These results highlight that predictability sources should not be approached with a "one-size-fits-all" strategy but rather tailored to specific weather regimes. This perspective may have uncovered previously unknown sources of skill, such as the Antarctic Oscillation index's potential relevance for European and Scandinavian Blocking.

Further research should focus on refining predictor selection for individual weather regimes to establish robust connections. One approach could be incorporating time-series analysis of indices in more complex machine learning models such as transformers to capture temporal dependencies. Additionally, conducting case studies on specific forecasts and their associated atmospheric states may provide deeper insights into the mechanisms driving predictability. Another important step is extending the analysis to longer lead times within the extended range to evaluate how different predictors contribute to forecast skill over increasing time horizons. By enhancing our understanding of regime-specific sources of predictability, we can further improve extended-range forecasting and develop more reliable decision-support tools for subseasonal-to-seasonal prediction.

To conclude, our study demonstrates the potential of statistical–dynamical forecasting with neural networks for the subseasonal prediction of North Atlantic-European weather regimes, providing valuable improvements to NWP performance – particularly in situations where traditional NWP models struggle.

CONFLICT OF INTEREST

All authors declare that they have no conflicts of interest.

CODE AND DATA AVAILABILITY

Code is available on the GitHub repository `https://github.com/fmockert/windowsofopportunity_WR`. The ERA5 data can be obtained from the Climate Data Store `https://cds.climate.copernicus.eu/#!/home`. Weather regime data are available from CMG upon request. The original S2S database is hosted at ECMWF as an extension of the TIGGE database.

ACKNOWLEDGEMENTS

REFERENCES

M. P. Baldwin, L. J. Gray, T. J. Dunkerton, K. Hamilton, P. H. Haynes, W. J. Randel, J. R. Holton, M. J. Alexander, I. Hirota, T. Horinouchi, D. B. A. Jones, J. S. Kinnersley, C. Marquardt, K. Sato, and M. Takahashi. The quasi-biennial oscillation. 2001. doi: `https://doi.org/10.1029/1999RG000073`.

R. Beerli and C. M. Grams. Stratospheric modulation of the large-scale circulation in the atlantic–european region and its implications for surface weather events. *Quarterly Journal of the Royal Meteorological Society*, 145:3732–3750, 10 2019. ISSN 1477870X. doi: 10.1002/qj.3653.

H. C. Bloomfield, C. C. Suitters, and D. R. Drew. Meteorological drivers of european power system stress. *Journal of Renewable Energy*, 2020:1–12, 8 2020. ISSN 2314-4386. doi: 10.1155/2020/5481010.

H. C. Bloomfield, D. J. Brayshaw, P. L. Gonzalez, and A. Charlton-Perez. Pattern-based conditioning enhances sub-seasonal prediction skill of european national energy variables. *Meteorological Applications*, 28, 7 2021. ISSN 14698080. doi: 10.1002/met.2018.

D. Büeler, R. Beerli, H. Wernli, and C. M. Grams. Stratospheric influence on ecmwf sub-seasonal forecast skill for energy-industry-relevant surface weather in european countries. *Quarterly Journal of the Royal Meteorological Society*, 146:3675–3694, 10 2020. ISSN 1477870X. doi: 10.1002/qj.3866.

D. Büeler, L. Ferranti, L. Magnusson, J. F. Quinting, and C. M. Grams. Year-round sub-seasonal forecast skill for atlantic–european weather regimes. *Quarterly Journal of the Royal Meteorological Society*, 147:4283–4309, 10 2021. ISSN 1477870X. doi: 10.1002/qj.4178.

C. Cassou. Intraseasonal interaction between the madden-julian oscillation and the north atlantic oscillation. *Nature*, 455:523–527, 9 2008. ISSN 14764687. doi: 10.1038/nature07286.

A. J. Charlton-Perez, R. W. Aldridge, C. M. Grams, and R. Lee. Winter pressures on the uk health system dominated by the greenland blocking weather regime. *Weather and Climate Extremes*, 25, 9 2019. ISSN 22120947. doi: 10.1016/j.wace.2019.100218.

R. J. Chase, D. R. Harrison, A. Burke, G. M. Lackmann, and A. McGovern. A machine learning tutorial for operational meteorology. part i: Traditional machine learning. *Weather and Forecasting*, 37:1509–1529, 8 2022. ISSN 15200434. doi: 10.1175/WAF-D-22-0070.1.

R. J. Chase, D. R. Harrison, G. M. Lackmann, and A. McGovern. A machine learning tutorial for operational meteorology. part ii: Neural networks and deep learning. *Weather and Forecasting*, 38:1271–1293, 8 2023. ISSN 15200434. doi: 10.1175/WAF-D-22-0187.1.

D. I. V. Domeisen, C. M. Grams, and L. Papritz. The role of north atlantic–european weather regimes in the surface impact of sudden stratospheric warming events. *Weather and Climate Dynamics*, 1:373–388, 8 2020. doi: 10.5194/wcd-1-373-2020.

ECMWF. Forecast ensemble (ens) - rationale and construction, 2018. URL `https://confluence.ecmwf.int/display/FUG/Section+5+Forecast+Ensemble+%28ENS%29+-+Rationale+and+Construction`.

A. J. Elliot, K. W. Cross, and D. M. Fleming. Acute respiratory infections and winter pressures on hospital admissions in england and wales 1990-2005. *Journal of Public Health*, 30:91–98, 3 2008. ISSN 17413842. doi: 10.1093/pubmed/fdn003.

P. N. Feng and H. Lin. Modulation of the mjo-related teleconnections by the qbo. *Journal of Geophysical Research: Atmospheres*, 124: 12022–12033, 11 2019. ISSN 21698996. doi: 10.1029/2019JD030878.

L. Ferranti, L. Magnusson, F. Vitart, and D. S. Richardson. 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, 7 2018. ISSN 1477870X. doi: 10.1002/qj.3341.

S. Gold, E. White, W. Roeder, M. McAleenan, C. S. Kabban, and D. Ahner. Probabilistic contingency tables: An improvement to verify probability forecasts. *Weather and Forecasting*, 35:609–621, 4 2020. ISSN 15200434. doi: 10.1175/WAF-D-19-0116.1.

C. M. Grams, R. Beerli, S. Pfenninger, I. Staffell, and H. Wernli. Balancing europe's wind-power output through spatial deployment informed by weather regimes. *Nature Climate Change*, 7:557–562, 8 2017. ISSN 17586798. doi: 10.1038/NCLIMATE3338.

S. Hauser, F. Teubler, M. Riemer, P. Knippertz, and C. M. Grams. Towards a holistic understanding of blocked regime dynamics through a combination of complementary diagnostic perspectives. *Weather and Climate Dynamics*, 4:399–425, 5 2023a. ISSN 26984016. doi: https://doi.org/10.5194/wcd-4-399-2023.

S. Hauser, F. Teubler, M. Riemer, P. Knippertz, and C. M. Grams. Life cycle dynamics of greenland blocking from a potential vorticity perspective. *EGUsphere Preprint repository*, 2023b. doi: https://doi.org/10.5194/egusphere-2023-2945. URL `https://doi.org/10.5194/egusphere-2023-2945`.

H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. D. Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. d. Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J. N. Thépaut. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146:1999–2049, 7 2020. ISSN 1477870X. doi: 10.1002/qj.3803.

C. Kent, A. A. Scaife, and N. Dunstone. What potential for improving sub-seasonal predictions of the winter nao? *Atmospheric Science Letters*, 24, 4 2023. ISSN 1530261X. doi: 10.1002/asl.1146.

J. C. K. Lee, R. W. Lee, S. J. Woolnough, and L. J. Boxall. The links between the madden-julian oscillation and european weather regimes. *Theoretical and Applied Climatology*, 141:567–586, 7 2020. ISSN 14344483. doi: 10.1007/s00704-020-03223-2.

R. W. Lee, S. J. Woolnough, A. J. Charlton-Perez, and F. Vitart. Enso modulation of mjo teleconnections to the north atlantic and europe. *Geophysical Research Letters*, 46:13535–13545, 11 2019. ISSN 19448007. doi: 10.1029/2019GL084683.

H. Lin, G. Brunet, and J. Derome. An observed connection between the north atlantic oscillation and the madden-julian oscillation. *Journal of Climate*, 22:364–380, 1 2009. ISSN 08948755. doi: 10.1175/2008JCLI2515.1.

R. A. Madden and P. R. Julian. Detection of a 40–50 day oscillation in the zonal wind in the tropical pacific. *Journal of the Atmoshperic Sciences*, 28:702–708, 3 1971. doi: https://doi.org/10.1175/1520-0469(1971)028%3C0702:DOADOI%3E2.0.CO;2.

C. Michel and G. Rivière. The link between rossby wave breakings and weather regime transitions. *Journal of the Atmospheric Sciences*, 68:1730–1748, 8 2011. ISSN 00224928. doi: 10.1175/2011JAS3635.1.

P. Michelangeli, R. Vautard, and B. Legras. Weather regimes: Recurrence and quasi stationarity. *Journal of Atmospheric Sciences*, 52: 1237–1256, 1995. doi: https://doi.org/10.1175/1520-0469(1995)052%3C1237:WRRAQS%3E2.0.CO;2.

O. T. Millin, J. C. Furtado, and C. Malloy. The impact of north american winter weather regimes on electricity load in the central united states. *npj Climate and Atmospheric Science*, 7, 12 2024. ISSN 23973722. doi: 10.1038/s41612-024-00803-1.

F. Mockert, C. M. Grams, T. Brown, and F. Neumann. Meteorological conditions during periods of low wind speed and insolation in germany: The role of weather regimes. *Meteorological Applications*, 30, 7 2023. ISSN 1350-4827. doi: 10.1002/met.2141. URL https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2141.

F. Mockert, C. M. Grams, S. Lerch, M. Osman, and J. Quinting. Multivariate post-processing of probabilistic sub-seasonal weather regime forecasts. *Quarterly Journal of the Royal Meteorological Society*, 9 2024. ISSN 0035-9009. doi: 10.1002/qj.4840. URL https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840.

NOAA. Daily mjo index time series, 2 2025. URL https://psl.noaa.gov/mjo/mjoindex/.

M. Osman, R. Beerli, D. Büeler, and C. M. Grams. Multi-model assessment of sub-seasonal predictive skill for year-round atlantic-european weather regimes. *Quarterly Journal of the Royal Meteorological Society*, 7 2023. ISSN 0035-9009. doi: https://doi.org/10.1002/qj.4512.

C. D. Roberts, M. A. Balmaseda, L. Ferranti, and F. Vitart. Euro-atlantic weather regimes and their modulation by tropospheric and stratospheric teleconnection pathways in ecmwf reforecasts. *Monthly Weather Review*, 151:2779–2799, 2023. doi: 10.1175/MWR-D-. URL https://doi.org/10.1175/MWR-D-.

Salient, M. Deo, and A. Atlas. On the money: How salient's novel s2s forecasts can save the agriculture industry millions of dollars, 5 2023. URL https://www.salientpredictions.com/blog/on-the-money-how-salients-novel-s2s-forecasts-can-save-the-agriculture-industry-millions-of-dollars.

A. A. Scaife, L. Hermanson, A. v. Niekerk, M. Andrews, M. P. Baldwin, S. Belcher, P. Bett, R. E. Comer, N. J. Dunstone, R. Geen, S. C. Hardiman, S. Ineson, J. Knight, Y. Nie, H. L. Ren, and D. Smith. Long-range predictability of extratropical climate and the length of day. *Nature Geoscience*, 10 2022. ISSN 17520908. doi: 10.1038/s41561-022-01037-7.

J. Song, W. Zhou, C. Li, and L. Qi. Signature of the antarctic oscillation in the northern hemisphere. *Meteorology and Atmospheric Physics*, 105:55–67, 2009. ISSN 01777971. doi: 10.1007/s00703-009-0036-5.

J. Spaeth, P. Rupp, M. Osman, C. M. Grams, and T. Birner. Flow-dependence of ensemble spread of subseasonal forecasts explored via north atlantic-european weather regimes. *Geophysical Research Letters*, 51, 7 2024. ISSN 0094-8276. doi: 10.1029/2024GL109733. URL https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024GL109733.

D. Specq and L. Batté. Do subseasonal forecasts take advantage of madden–julian oscillation windows of opportunity? *Atmospheric Science Letters*, 23, 4 2022. ISSN 1530261X. doi: 10.1002/asl.1078.

Y. Tachibana, Y. Inoue, K. K. Komatsu, T. Nakamura, M. Honda, K. Ogata, and K. Yamazaki. Interhemispheric synchronization between the ao and the aao. *Geophysical Research Letters*, 45:13,477–13,484, 12 2018. ISSN 19448007. doi: 10.1029/2018GL081002.

R. Vautard. Multiple weather regimes over the north atlantic: Analysis of precursors and successors. *Monthly Weather Review*, 118: 2056–2081, 1990. doi: https://doi.org/10.1175/1520-0493(1990)118%3C2056:MWROTN%3E2.0.CO;2.

F. Vitart. Madden—julian oscillation prediction and teleconnections in the s2s database. *Quarterly Journal of the Royal Meteorological Society*, 143:2210–2220, 7 2017. ISSN 1477870X. doi: 10.1002/qj.3079.

F. Vitart, C. Ardilouze, A. Bonet, A. Brookshaw, M. Chen, C. Codorean, M. Déqué, L. Ferranti, E. Fucile, M. Fuentes, H. Hendon, J. Hodgson, H. S. Kang, A. Kumar, H. Lin, G. Liu, X. Liu, P. Malguzzi, I. Mallas, M. Manoussakis, D. Mastrangelo, C. MacLachlan, P. McLean, A. Minami, R. Mladek, T. Nakazawa, S. Najm, Y. Nie, M. Rixen, A. W. Robertson, P. Ruti, C. Sun, Y. Takaya, M. Tolstykh, F. Venuti, D. Waliser, S. Woolnough, T. Wu, D. J. Won, H. Xiao, R. Zaripov, and L. Zhang. The subseasonal to seasonal (s2s) prediction project database. *Bulletin of the American Meteorological Society*, 98:163–173, 1 2017. ISSN 00030007. doi: 10.1175/BAMS-D-16-0017.1.

J. Wandel, D. Büeler, P. Knippertz, J. F. Quinting, and C. M. Grams. Why moist dynamic processes matter for the sub-seasonal prediction of atmospheric blocking over europe. *Journal of Geophysical Research: Atmospheres*, 129, 4 2024. ISSN 21698996. doi: 10.1029/2023JD039791.

M. C. Wheeler and H. H. Hendon. An all-season real-time multivariate mjo index: Development of an index for monitoring and prediction. 2004. doi: https://doi.org/10.1175/1520-0493(2004)132%3C1917:AARMMI%3E2.0.CO;2.

C. J. White, H. Carlsen, A. W. Robertson, R. J. Klein, J. K. Lazo, A. Kumar, F. Vitart, E. C. d. Perez, A. J. Ray, V. Murray, S. Bharwani, D. MacLeod, R. James, L. Fleming, A. P. Morse, B. Eggen, R. Graham, E. Kjellström, E. Becker, K. V. Pegion, N. J. Holbrook, D. McEvoy, M. Depledge, S. Perkins-Kirkpatrick, T. J. Brown, R. Street, L. Jones, T. A. Remenyi, I. Hodgson-Johnston, C. Buontempo, R. Lamb, H. Meinke, B. Arheimer, and S. E. Zebiak. Potential applications of subseasonal-to-seasonal (s2s) predictions. *Meteorological Applications*, 24:315–325, 7 2017. ISSN 14698080. doi: https://doi.org/10.1002/met.1654.

S | SUPPLEMENTARY MATERIAL

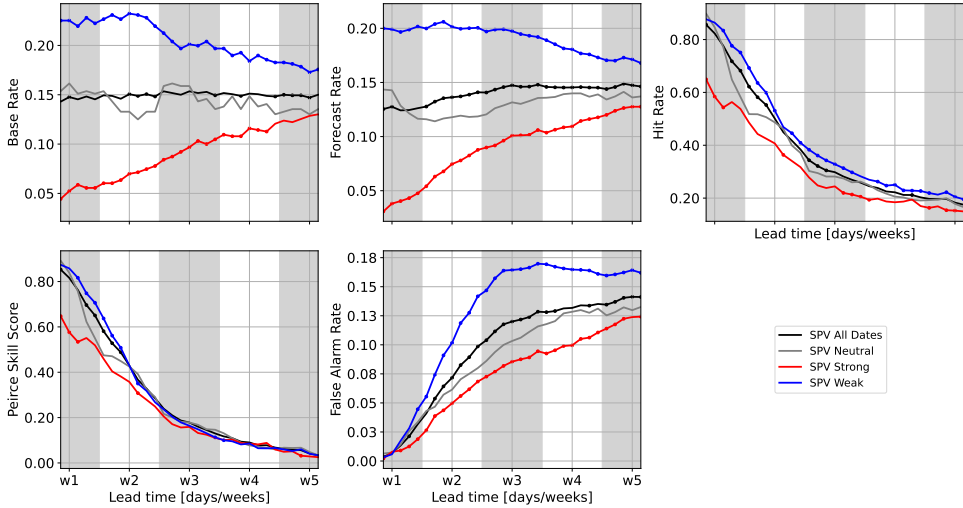(a) MJO-conditioned rates.
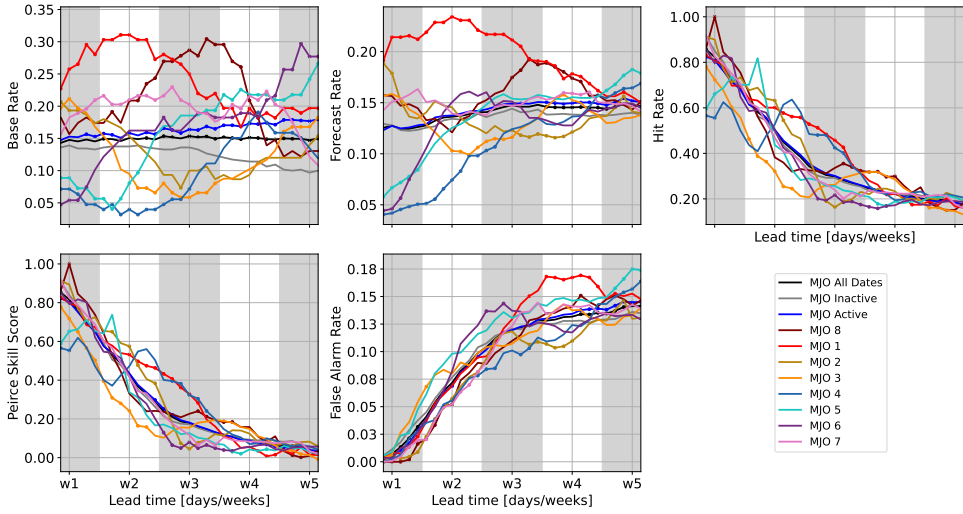
(b) SPV-conditioned rates.

FIGURE S1   Rates for weekly mean Greenland Blocking activity ($WRact_{mean,GL}$) in the extended winter period (November–March). These rates are split into subsets based on phases of the (a) MJO or (b) SPV. In contrast to Figure 3 and Figure 4, here all lead times are shown (x-axis). Dots on the individual lines indicate whether the rates in the specific categories and at the specific lead time are significantly different to their inactive/neutral phases – for MJO and SPV, respectively – with a 90% confidence interval.
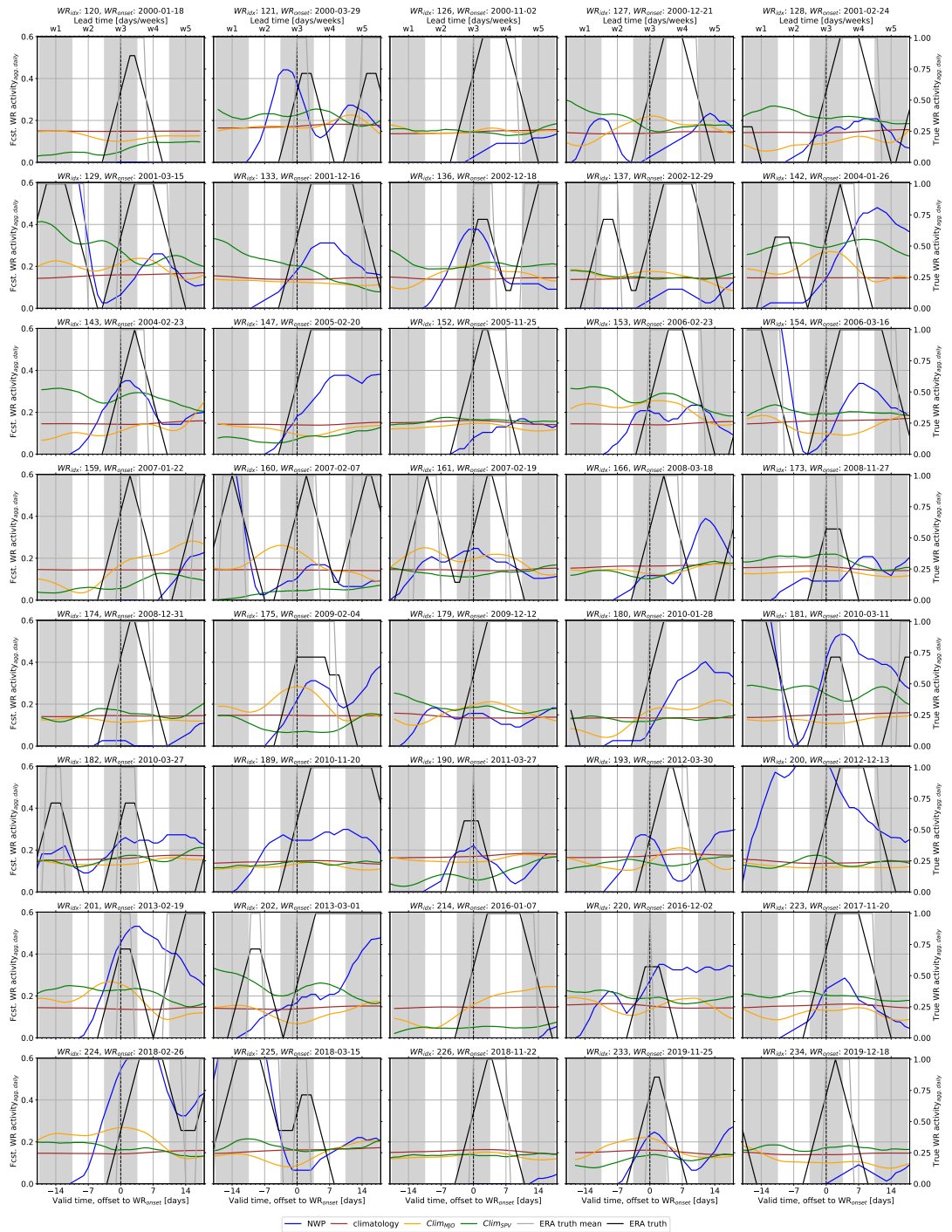
FIGURE S2 Forecasts of the aggregated daily Greenland Blocking activity used for the mean composite plot in Figure 6a for all Greenland Blocking activities. The forecasts are initialised three weeks ahead of the Greenland Blocking activity onset.
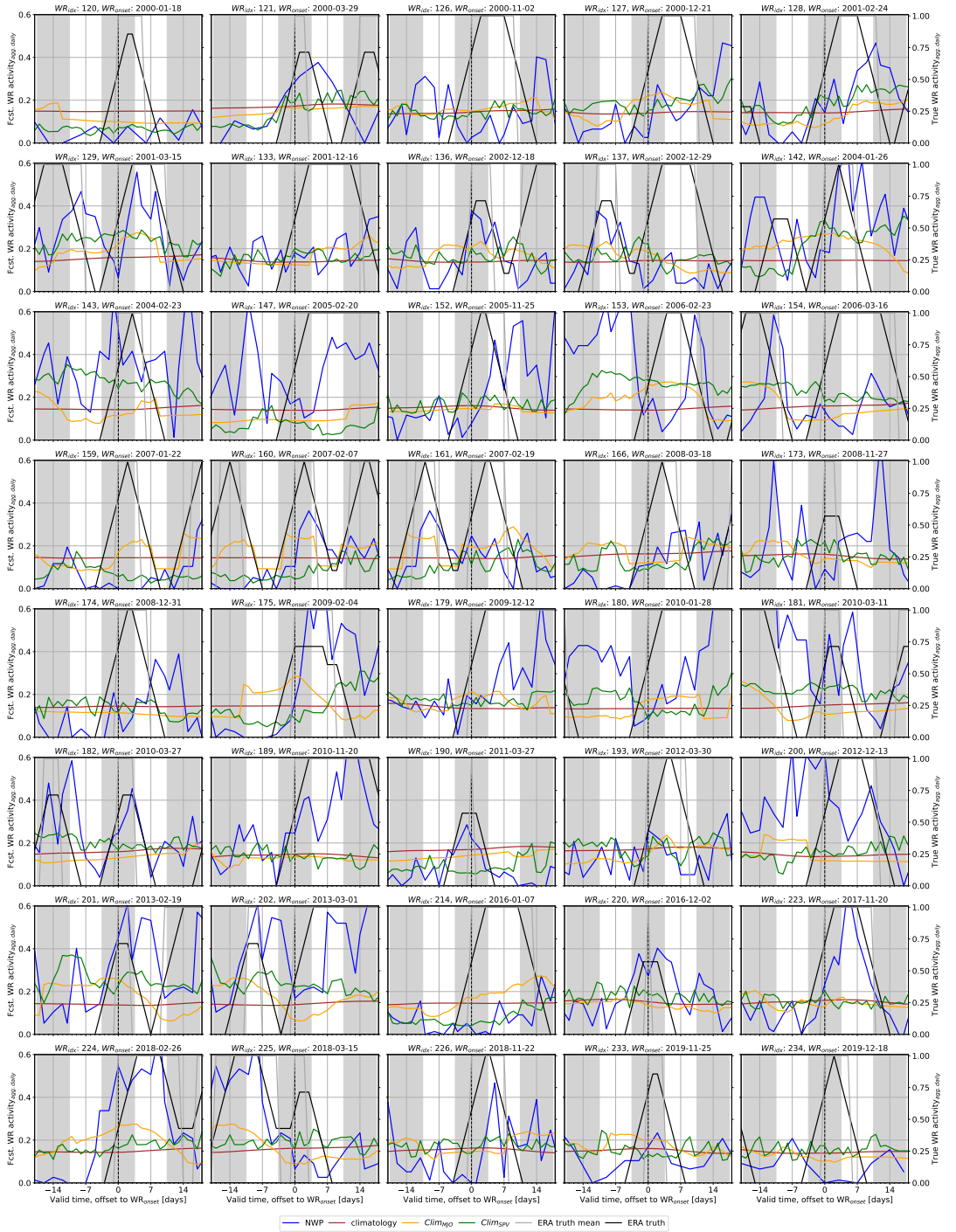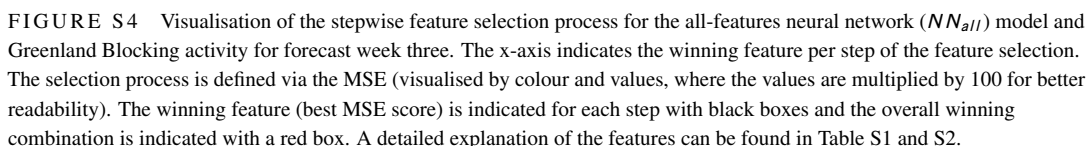
FIGURE S3 Forecasts of the aggregated daily Greenland Blocking activity used for the mean composite plot in Figure 6b for all Greenland Blocking activities. Consecutive forecasts with each a lead time of three weeks are used.
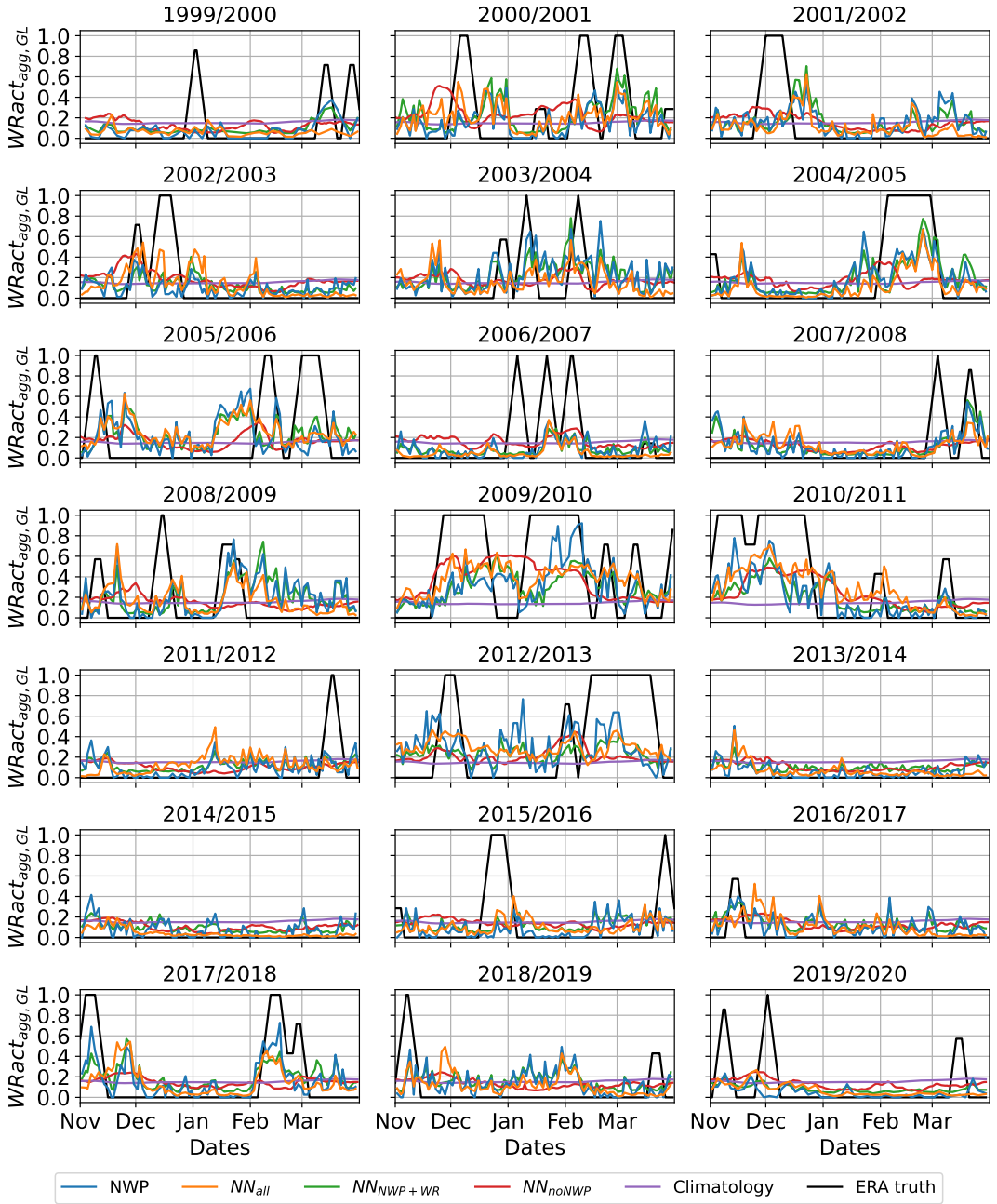
FIGURE S4 Visualisation of the stepwise feature selection process for the all-features neural network (NN_{all}) model and Greenland Blocking activity for forecast week three. The x-axis indicates the winning feature per step of the feature selection. The selection process is defined via the MSE (visualised by colour and values, where the values are multiplied by 100 for better readability). The winning feature (best MSE score) is indicated for each step with black boxes and the overall winning combination is indicated with a red box. A detailed explanation of the features can be found in Table S1 and S2.

FIGURE S5 Forecasts of the aggregated daily GL activity for each extended winter period from 1999 to 2020. The ERA5 (actual) weather regime activity is indicated in black, the neural networks all-features, NWP and ERA, and climatology in blue, orange and green, as well as the NWP model in red and a climatological forecast in purple.

Category	Feature	Description
Climatology	Climatology	Standard 91-days climatology
	Clim$_{MJO}$	MJO-conditioned climatological forecast using OMI1/2
	Clim$_{SPV}$	SPV-conditioned climatological forecast using 100 hPa zonal mean wind at 60°N
	Clim$_{QBO}$	QBO-conditioned climatological forecast using Quasi-Biennial Oscillation index
	Clim$_{NAO}$	NAO-conditioned climatological forecast using North Atlantic Oscillation index
	Clim$_{ONI}$	ONI-conditioned climatological forecast using Ocean Nino index
	Clim$_{AO}$	AO-conditioned climatological forecast using Arctic Oscillation index
	Clim$_{AAO}$	AAO-conditioned climatological forecast using Antarctic Oscillation index
	Clim$_{PNA}$	PNA-conditioned climatological forecast using Pacific North-American index
	DOY	Day of year represented by sinus curve (maxima at July 15th and minima at January 15th)
Atmospheric State	OMI$_{amplitude}$	MJO amplitude in OMI-phase space. Source: `https://psl.noaa.gov/mjo/mjoindex/omi.era5.1x.webpage.4023.txt`
	OMI$_{contphase}$	MJO angle in OMI-phase space. Source:
	QBO	QBO index. Source: `https://www.geo.fu-berlin.de/met/ag/strat/produkte/qbo/qbo.dat`
	ONI	ONI index. Source: `https://psl.noaa.gov/data/correlation/oni.data`
	NAO	NAO index. Source: `https://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/norm.nao.monthly.b5001.current.ascii.table`
	AO	AO index. Source: `https://ftp.cpc.ncep.noaa.gov/cwlinks/norm.daily.ao.index.b500101.current.ascii`
	PNA	PNA index. Source: `https://ftp.cpc.ncep.noaa.gov/cwlinks/norm.daily.pna.index.b500101.current.ascii`
	AAO	AAO index. Source: `https://ftp.cpc.ncep.noaa.gov/cwlinks/norm.daily.aao.index.b790101.current.ascii`
	SPV	SPV index computed by 60°N zonal mean wind at 100 hPa
	SSTano$_{smooth\ northatlantic}$	30-day running mean anomalies of seas surface temperature compared to 1991–2020 in the North Atlantic (0 to 80°N, 80 to 10°W)

TABLE S1 Overview of features available for neural network selection, categorised into Climatology and Atmospheric State. Features from the NWP and recent weather regime activity categories are listed separately in Table S2.

Category	Feature	Description
NWP	$NWP_{IWRmean}$	Mean weather regime index forecast for week 3
	NWP_{IWRstd}	Standard deviation of ensemble weather regime index forecast for week 3
	$NWP_{agg.\ WRactivity}$	Aggregated daily weather regime activity forecast for week 3
	$NWP_{IWRtrend,w2}$	Trend of daily IWR forecast across week 2 (linear regression)
	$NWP_{IWRtrend,w1-2}$	Trend of daily IWR forecast across week 1-2 (linear regression)
	$NWP_{IWRtrend,w3,7dprevious}$	Trend of the weather regime index across the forecasts available 7d prior to initialisation time and a valid time similar to the 3 week forecast from initialisation time
	$NWP_{IWRstd,w3,7dprevious}$	Standard deviation of the weather regime index across the forecasts available 7d prior to initialisation time and a valid time similar to the 3 week forecast from initialisation time
Recent WR	$ERA_{IWRatIC}$	IWR at initialisation time
	$ERA_{IWRtrend,past7d}$	IWR trend (linear regression) of past 7 days
	$ERA_{7d,pastactivity}$	Mean aggregated daily weather regime activity of past 7 days
	$ERA_{30d,pastactivity}$	Mean aggregated daily weather regime activity of past 30 days

T A B L E S 2   Overview of features available for neural network selection, categorised into NWP and recent weather regime activity. Features from the Climatology and Atmospheric State categories are listed separately in Table S1.