Improving subseasonal weather regime forecasts using statistical-dynamical approaches

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Fabian Dominique Mockert
aus Karlsruhe

Tag der mündlichen Prüfung:	01.08.2025
Referent:	Prof. Dr. Peter Knippertz
Korreferent:	Dr. Julian Quinting
Ext. Korreferent (Uni Marburg)	Prof. Dr. Sebastian Lerch

Abstract

As the share of renewable energy sources such as wind and solar continues to grow, so does the need for reliable subseasonal forecasts (10–30 days ahead) to balance supply and demand, plan infrastructure maintenance, and manage energy storage and trading. Forecasts of surface weather variables – like temperature, wind, and incoming solar radiation – provide useful information. However, their skill quickly deteriorates and vanishes after two weeks at the latest due to the chaotic nature of the atmosphere and imperfect models. To extend forecast skill into the subseasonal range, some degree of spatial and/or temporal aggregation of forecast information is essential. In this context, weather regimes – recurring patterns in the large-scale atmospheric circulation that persist for several days to weeks – offer a promising framework.

This thesis aims to develop more useful forecast information tailored to the needs of the energy sector. To this end, it newly establishes a link between weather regimes and "Dunkelflauten" in Germany – energy-related extreme events characterised by simultaneous low wind and solar power generation. To better predict these regimes at subseasonal lead times, we apply cost-effective, data-driven post-processing techniques to numerical weather prediction (NWP) model outputs.

Dunkelflauten pose a particular challenge for energy supply, especially during winter, when energy demand is high due to low temperatures. Anticipating the occurrence of cold Dunkelflauten, which present a worst-case stress scenario for the energy system, is therefore of high importance. Using ECMWF's ERA5 reanalysis data from 1979 to 2018, we objectively identify Dunkelflauten in Germany. As weather regimes shape regional surface weather conditions across Europe, we examine the statistical occurrence of Dunkelflauten with respect to a year-round seven weather regime classification over the North Atlantic-European region, as introduced by Grams et al. (2017). Dunkelflauten predominantly occur during winter and are primarily associated with three anticyclonic weather regimes, so-called blocking patterns. Among these, the Greenland Blocking (GL) regime stands out due to its frequent association with extremely cold conditions. Moreover, Dunkelflauten typically occur in the well established phase of longer-lived weather regimes. Identifying GL as the key regime for cold Dunkelflauten and its persistence suggest that improved regime forecasting could enhance the anticipation of such energy-critical events.

At lead times beyond 12 days, regime-based forecasts outperform grid-point-based surface variable forecasts for predicting country-aggregated energy metrics like demand or wind power generation. However, challenges remain when forecasting weather regimes on the subseasonal time scale. Next to the intrinsic predictability limit due to the chaotic nature of the atmosphere, systematic errors in NWP models, such

as biases and dispersion errors, further limit the probabilistic forecast skill. This thesis addresses the systematic errors by applying a two-step ensemble post-processing technique to ECMWF reforecasts spanning from 1999 to 2020. First, univariate ensemble model output statistics are applied to forecasts for each weather regime individually. Second, ensemble copula coupling restores the multivariate dependence structure between weather regimes. Compared to current forecast calibration practices, which rely on correcting the Z500 field by the lead-time-dependent mean bias, our approach extends the forecast skill horizon for daily/instantaneous regime forecasts moderately by one day (from 13.5 to 14.5 days). The proposed method can easily be applied to operational weather regime forecasts, offering a neat alternative for cost- and time-efficient post-processing of real-time weather regime forecasts. To our knowledge, our study is the first to evaluate the multivariate aspects of forecast quality systematically for weather regime forecasts.

Despite the improvements through statistical post-processing, average daily regime forecast skill remains limited to about two weeks. To extend the probabilistic forecast skill beyond this period, we introduce a novel, temporally aggregated weather regime metric – the so-called weather regime activity – which captures the probability of strong projections into individual regimes over a one-week period. From literature it is known that the Madden-Julian Oscillation (MJO) and the stratospheric polar vortex (SPV) are prominent sources of predictability in the Extratropics. They modulate the occurrence of weather regimes, though their effects into forecast week 3 remain under-explored. We show that for GL activity in forecast week 3, phases of the MJO with convection over the central Pacific, Western Hemisphere, and central Maritime Continent, along with periods of weak SPV conditions at initialisation time identify as windows of opportunity – favourable atmospheric conditions that enhance forecast skill or increase the likelihood of impactful events. Utilising this knowledge, statistical-dynamical models using neural networks are trained on historical atmospheric states and NWP-derived regime metrics. These models serve as valuable decision-support tools alongside NWP forecasts, improving probabilistic forecasts of weekly regime activity in week 3 by up to 10%, as measured by the mean squared error. The largest improvements are found for blocking regimes over the European domain, where NWP models tend to underperform.

By bridging physical understanding, statistical correction, and data-driven modelling, this thesis outlines pathways to enhance weather regime forecasts at subseasonal lead times and demonstrates their value as decision-support tool for anticipating regime configurations prone to Dunkelflauten. These advancements can greatly benefit the energy sector, where long-term foresight is of growing importance.

Zusammenfassung

Mit dem wachsenden Anteil erneuerbarer Energiequellen wie Wind- und Solarenergie steigt auch der Bedarf an verlässlichen Vorhersagen auf subsaisonalen Zeitskalen (10–30 Tage im Voraus), um Angebot und Nachfrage auszugleichen, Wartungsarbeiten an der Infrastruktur zu planen sowie Energiespeicherung und -handel zu managen. Vorhersagen von bodennahen Wettervariablen – wie Temperatur, Wind und eingehende Solarstrahlung – liefern nützliche Informationen. Ihre Vorhersagegüte nimmt jedoch aufgrund der chaotischen Natur der Atmosphäre und den fehlerbehafteten numerischen Modellen rasch ab und verschwindet spätestens nach zwei Wochen. Um die Vorhersagegüte in den subsaisonalen Bereich zu verlängern, ist eine gewisse räumliche und/oder zeitliche Aggregation der Vorhersageinformationen essenziell. In diesem Zusammenhang bieten Wetterregime – wiederkehrende Muster in der großskaligen atmosphärischen Zirkulation, die über mehrere Tage bis Wochen andauern – einen vielversprechenden Ansatz.

Diese Dissertation hat zum Ziel, nützlichere Vorhersageinformationen für den Energiesektor bereitzustellen. Zu diesem Zweck wird ein neuer Zusammenhang zwischen Wetterregimen und sogenannten „Dunkelflauten“ – energierelevante Extremereignissen, die durch gleichzeitig geringe Wind- und Solarstromerzeugung gekennzeichnet sind – in Deutschland hergestellt. Um diese Regime bei subsaisonalen Vorhersagezeiten besser erfassen zu können, werden kosteneffiziente, datengetriebene Nachbearbeitungsmethoden (post-processing) auf die Vorhersagen von numerischen Wettervorhersagemodellen (NWP) angewendet.

Dunkelflauten stellen insbesondere im Winter eine große Herausforderung für die Energieversorgung dar, da in dieser Jahreszeit der Energiebedarf aufgrund niedriger Temperaturen besonders hoch ist. Die frühzeitige Erkennung kalter Dunkelflauten, die ein Stressszenario für das Energiesystem darstellen, ist daher von großer Bedeutung. Auf Basis der ERA5-Reanalysedaten des europäischen Zentrums für mittelfristige Wettervorhersage (ECMWF) für den Zeitraum 1979 bis 2018 werden Dunkelflauten in Deutschland objektiv identifiziert. Da Wetterregime das regionale Wettergeschehen in Europa wesentlich prägen, wird das statistische Auftreten von Dunkelflauten im Zusammenhang mit einer ganzjährigen Sieben-Regime-Klassifikation für den nordatlantisch-europäischen Raum, wie sie von Grams et al. (2017) eingeführt wurde, untersucht. Dunkelflauten treten überwiegend im Winter auf und sind hauptsächlich mit drei antizyklonalen Wetterregimen, sogenannten Blocking-Mustern, assoziiert. Unter diesen sticht das Greenland Blocking (GL) Regime durch seine häufige Verbindung mit extrem kalten Bedingungen hervor. Zudem treten Dunkelflauten typischerweise in der etablierten Phase langlebiger Wetterregime auf. Die Identifikation von GL als Schlüsselregime für kalte Dunkelflauten sowie die Langlebigkeit dieser Regime legt

nahe, dass eine verbesserte Regimevorhersage die Früherkennung solcher energiekritischen Ereignisse deutlich verbessern kann.

Bei Vorhersagezeiten von mehr als 12 Tagen übertreffen Regime-basierte Ansätze die Verwendung bodennaher Variablen auf Gitterpunkt-Basis in Bezug auf Vorhersagen landesweit aggregierter Energiekennzahlen wie Stromnachfrage oder Windkraftproduktion. Dennoch bestehen weiterhin Herausforderungen bei der Vorhersage von Wetterregimen auf subsaisonaler Zeitskala. Neben den inhärenten Vorhersagegrenzen durch die chaotische Atmosphäre wirken sich systematische Fehler in NWP-Modellen, etwa Vorhersagefehler (Bias) und Streuungsfehler, zusätzlich negativ auf die probabilistische Vorhersagegüte aus. Diese Arbeit adressiert systematische Fehler durch eine zweistufige Nachbearbeitungsmethode, angewendet auf Re-Vorhersagen des ECMWF für den Zeitraum von 1999 bis 2020. Zunächst wird eine univariate Methode (die Ensemble Model Output Statistics (EMOS)) auf die Vorhersagen der einzelnen Regime angewendet. Anschließend wird mithilfe von der Methode Ensemble Copula Coupling (ECC) die multivariate Abhängigkeitsstruktur zwischen den Wetterregimen wiederhergestellt. Im Vergleich zu herkömmlichen Kalibrierungsverfahren, bei denen das Z500-Feld lediglich um den vorhersagezeitabhängigen mittleren Bias korrigiert wird, verlängert unser Ansatz den Horizont der täglichen Regimevorhersage moderat um einen Tag (von 13,5 auf 14,5 Tage). Die vorgeschlagene Methode lässt sich leicht auf operationelle Regimevorhersagen anwenden und stellt eine praktikable Alternative für eine kosteneffiziente und zeitsparende Nachbearbeitung dar. Nach unserem Kenntnisstand ist unsere Studie die Erste, welche die multivariaten Aspekte der Vorhersagegüte systematisch für Wetterregime analysiert.

Trotz der Verbesserungen durch statistische Nachbearbeitung bleibt die mittlere tägliche Regimevorhersagegüte auf etwa zwei Wochen begrenzt. Um die probabilistische Vorhersagegüte über diesen Zeitraum hinaus zu erweitern, wird ein neuartiges, zeitlich aggregiertes Regimemaß eingeführt – die sogenannte Regimeaktivität – welche die Wahrscheinlichkeit starker Projektionen auf einzelne Regime innerhalb eines Wochenintervalls beschreibt. Aus der Literatur ist bekannt, dass die Madden-Julian-Oszillation (MJO) und der stratosphärische Polarwirbel (SPV) bedeutende Quellen von Vorhersagbarkeit in den mittleren Breiten darstellen. Beide beeinflussen das Auftreten von Wetterregimen, allerdings sind ihre Effekte in der dritten Vorhersagewoche bislang kaum untersucht. Für die GL-Aktivität in Woche 3 zeigen wir, dass bestimmte Phasen der MJO mit Konvektion über dem zentralen Pazifik, der westlichen Hemisphäre und dem zentralen maritimen Kontinent sowie Perioden mit schwachem SPV zum Initialisierungszeitpunkt der Vorhersage sogenannte „windows of opportunity“ darstellen – günstige atmosphärische Bedingungen, die die Vorhersagegüte erhöhen oder die Eintrittswahrscheinlichkeit wirkungsvoller Ereignisse steigern. Aufbauend auf diesem Wissen werden statistisch-dynamische Modelle auf Basis neuronaler Netze trainiert, welche historische atmosphärische Zustände sowie NWP-basierte Regimemaße nutzen. Diese Modelle dienen als wertvolle Entscheidungsunterstützung neben den NWP-Vorhersagen und verbessern die probabilistische Vorhersage der wöchentlichen Regimeaktivität in Woche 3 um bis zu 10% (gemessen am mittleren quadratischen Fehler). Die größten Verbesserungen treten bei Blocking-Regimen über dem europäischen Raum auf, in dem NWP-Modelle tendenziell unterdurchschnittlich performen.

Durch die Verbindung von physikalischem Verständnis, statistischer Korrektur und datengetriebener Modellierung zeigt diese Arbeit Wege zur Verbesserung der Wetterregimevorhersage auf Zeitskalen größer 10 Tage auf und demonstriert deren Mehrwert als Entscheidungsunterstützung bei der Antizipation von Regimekonfigurationen, die für Dunkelflauten prädisponiert sind. Diese Fortschritte können dem Energiesektor erheblich zugutekommen, in dem langfristige Planbarkeit zunehmend an Bedeutung gewinnt.

Preface

The PhD candidate confirms that the research presented in this thesis contains significant scientific contributions by himself. This thesis reuses material from the following publications:

> Mockert, F., C.M. Grams, T. Brown & F. Neumann (2023). Meteorological conditions during periods of low wind speed and insolation in Germany: The role of weather regimes. Meteorological Applications, 30(4), e2141.
> Available from: `https://doi.org/10.1002/met.2141`

> Mockert, F., C.M. Grams, S. Lerch, M. Osman & J. Quinting (2024) Multivariate post-processing of probabilistic sub-seasonal weather regime forecasts. Quarterly Journal of the Royal Meteorological Society, 1–17.
> Available from: `https://doi.org/10.1002/qj.4840`

> Mockert, F., C.M. Grams, S. Lerch & J. Quinting (2025): Windows of opportunity in subseasonal weather regime forecasting: A statistical-dynamical approach. Submitted to Quarterly Journal of the Royal Meteorological Society.
> Preprint available from: `https://doi.org/10.48550/arXiv.2505.02680`

The research leading to the results has been accomplished within the MathSEE funding program for PhD students at the Karlsruhe Institute of Technology.

The candidate confirms that appropriate credit has been given within the thesis, where reference has been made to the work of others. This copy has been supplied on the understanding that this is copyright material and that no quotation from the thesis may be published without proper acknowledgement.
©2025, Karlsruhe Institute of Technology and Fabian Mockert

Contents

Acronyms

AAO	Antarctic Oscillation
AI	Artificial Intelligence
AO	Arctic Oscillation
AR	Atlantic Ridge
AT	Atlantic Trough
BR	Base Rate
CRP(S)S	Continuous Ranked Probability (Skill) Score
DJF	Winter Period, December–February
DOY	Day of Year
ECC(-Q)	Ensemble Copula Coupling (Quantile)
ECMWF	European Centre for Medium-Range Weather Forecasts
EMOS(-G/Q)	Ensemble Model Output Statistics (Gaussian/Quantile)
ENSO	El Niño-Southern Oscillation
EOF	Empirical Orthogonal Function
EQBO	Easterly Quasi-Biennial Oscillation
ERA	ECMWF Reanalysis
E(S)S	Energy (Skill) Score
EuBL	European Blocking
FAR	False Alarm Rate
FN	False Negative
FP	False Positive
FR	Forecast Rate
GL	Greenland Blocking
HR	Hit Rate
IWR	Weather Regime Index
MJO	Madden-Julian Oscillation
MSE(S)	Mean Squared Error (Skill Score)
NAE	North Atlantic-European
NAO	North Atlantic Oscillation
NDJFM	Extended Winter Period, November–March

NN	Neural Network
NOAA	(United States) National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
OLR	Outgoing Longwave Radiation
OMI	OLR-based MJO Index
ONI	Ocean Niño Index
PNA	Pacific-North American Pattern
PSS	Peirce Skill Score
QBO	Quasi-Biennial Oscillation
RMM	Real-time Multivariate MJO Index
RQ	Research Question
ScBL	Scandinavian Blocking
ScTr	Scandinavian Trough
SPV	Stratospheric Polar Vortex
SST(ano)	Sea Surface Temperature (Anomaly)
SSW	Sudden Stratospheric Warming
S2S	Subseasonal-to-Seasonal
TN	True Negative
TP	True Positive
V(S)S	Variogram (Skill) Score
WQBO	Western Quasi-Biennial Oscillation
WR	Weather Regime
$WRact_{mean}$	Weekly Mean Weather Regime Activity
$WRact_{agg}$	Aggregated Daily Weather Regime Activity
WR_{cat}	Categorical Weather Regime Index
ZO	Zonal Regime
Z500(A)	500-hPa Geopotential Height (Anomaly)

1. Introduction

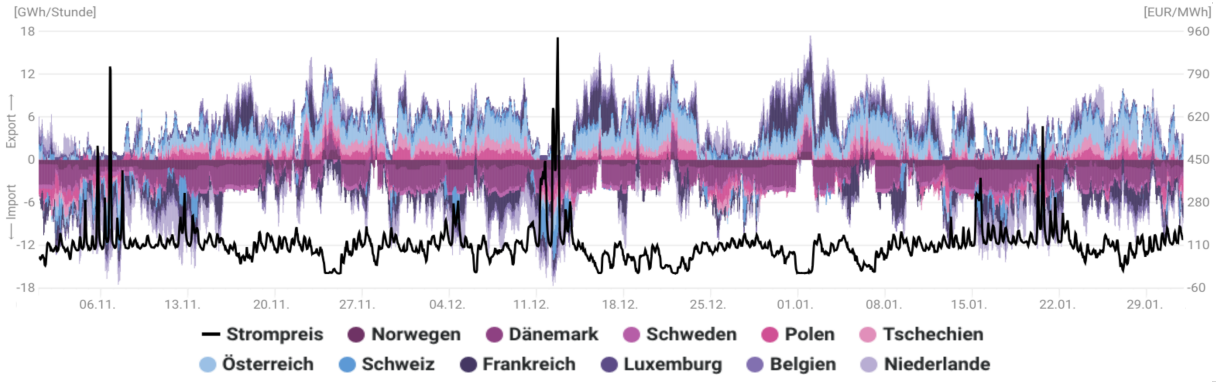
> The price per kilowatt hour [of electricity], usually between five and 20 cents, even rose well above one euro in daily trading. Companies that need a lot of electricity and buy it daily on the exchange cut back production and sent staff home.

This excerpt from the German news broadcast Tagesschau (2025) refers to an extreme weather event in Germany from 11 to 13 December 2024 (Figure 1.1a). Low wind and solar power generation led to temporarily high electricity costs for companies trading electricity directly on the market. A similar example occurred in December 2022 (Figure 1.1b), where reduced power supply from renewable energy sources (8 to 13 December) coincided with temperatures dropping to –12 °C in Karlsruhe and on average below –10 °C at night across southern Germany. Events like these, known as Dunkelflauten (the German word for "dark wind lulls"), are a matter of public discussion in Germany, as they are often believed to cause German wide blackouts. Besides exaggerated claims[1] and panic about potential blackouts (e.g., Kumar, 2025; Rodriguez, 2023), understanding these periods of low electricity generation is crucial for the energy transition towards a safe, sustainable, and economically successful future.

Three major paths to mitigate the impact of Dunkelflauten are energy storage technologies, diversification of energy sources, and an international electricity network (EnBW, 2025; TransnetBW, 2025). Daily variations in solar power can be balanced by storage or through flexibility in energy demand, such as bidirectional charging of battery-powered electric vehicles (Brown et al., 2018). Seasonal variability can be partially balanced by co-deploying wind and solar technologies (Heide et al., 2010; Pozo-Vázquez et al., 2004; Santos-Alamillos et al., 2015). Power-to-gas units, which transform surplus electrical energy into gas (e.g., hydrogen), and long-term thermal energy storage can further support balancing large-scale and seasonal variations in electricity supply and demand. However, the currently available storage capacity in Germany (20 GWh battery storage (Energy-Charts, 2025)) remains insufficient for a full replacement of fossil energy sources. As the share of renewable energy continues to grow (from 49.9% in 2015 to 68.4% in 2025 (Energy-Charts, 2025)), the need to understand and predict weather-dependent fluctuations in renewable energy generation on multi-day to weekly timescales is becoming increasingly critical (Bloomfield et al., 2016; Wiel et al., 2019a).

[1] Several science communication channels critically examine common blackout narratives surrounding Dunkelflauten, suggesting that fears are often exaggerated and not supported by energy system data. See, e.g., Doktor Whatson: "Wie wir den Blackout verhindern" `https://youtu.be/O1ELU21wiJA`, "So wird Deutschland klimaneutral" `https://youtu.be/mbI5WJ_KvOY`; Quarks: "Wie unser Stromsystem jetzt komplett auf den Kopf gestellt wird" `https://youtu.be/oSxIEx7pNho`; Illwerke vkw: "50 Hz - Wie wir einen Blackout vermeiden" `https://youtu.be/dVJYG44Wswg`.

(a) Winter 2024/2025 energy price and imports/exports.



(b) Winter 2022/2023 energy price and energy mix.

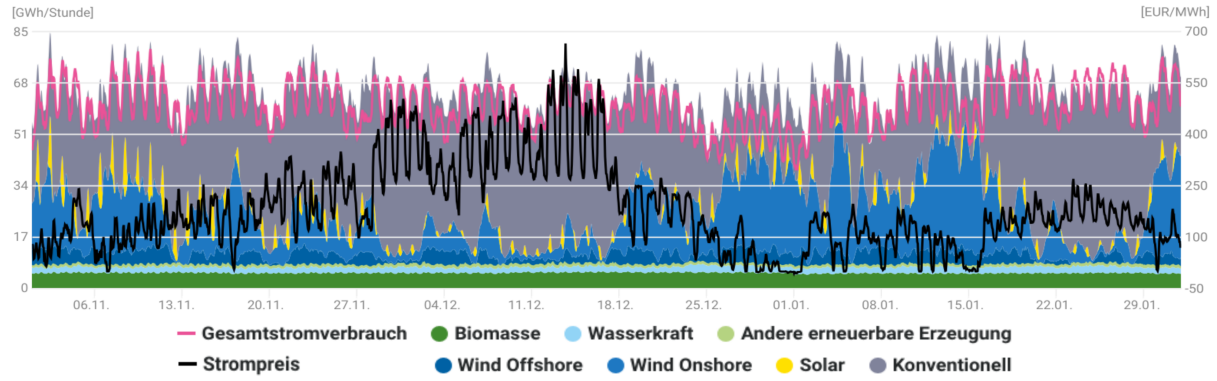


Figure 1.1.: Hourly electricity market data for Germany during two winter periods (November–January), each featuring a Dunkelflaute. (a) Winter 2024/2025: electricity prices and net imports/exports by country. A pronounced Dunkelflaute from 11 to 13 December 2024 coincides with peak prices reaching up to 950EUR/MWh and a strong reliance on imports. (b) Winter 2022/2023: electricity price, electricity demand and the share of renewable and conventional generation. During the Dunkelflaute from 8 to 13 December 2022, the contribution of renewable energy sources to the energy mix is especially low, coinciding with long lasting high electricity prices. Adapted from Agora Energiewende (`www.agora-energiewende.de/daten-tools/agorameter`), with minor visual adjustments.

The new frontier of weather forecasting is the timescale of 14 days up to six weeks, also referred to as subseasonal to seasonal (S2S) range. Forecasts on this time-range are becoming increasingly important, not only for the energy sector but also for public health and agriculture, as they fill the gap between weather forecasts and monthly or seasonal outlooks (White et al., 2017, 2022). On the S2S range, forecasts aim for a probabilistic description of the general characteristics of weather, using an ensemble of forecasts with varying initial conditions and model physics to quantify forecast uncertainty. The general characteristics of the large-scale atmospheric circulation are best described by pattern-based methods, such as weather regimes. They are defined as quasi-stationary, persistent, and recurrent large-scale flow patterns in the mid-latitudes (Michel and Rivière, 2011; Vautard, 1990). A weather regime definition tailored for the North Atlantic-European (NAE) region has been introduced by Grams et al. (2017), representing the large-scale atmospheric state by a 7-dimensional vector. These weather regimes modulate surface weather in continent-sized regions on multi-day to weekly timescales (Büeler et al.,

2021) and can explain substantial multi-day variability in the European energy sector (Wiel et al., 2019b; Zubiate et al., 2017).

While the chaotic nature of the atmosphere imposes a fundamental limit on the intrinsic predictability, the range of skilful forecasts is further constrained by systematic model errors. Despite major advances in forecasting systems over recent decades (Bauer et al., 2015), these limitations remain significant, particularly at subseasonal lead times (e.g., Büeler et al., 2021; Ferranti et al., 2018). Current calibration practices for weather regime forecasts rely on correcting the 500 hPa geopotential height field by the lead-time-dependent mean bias. In recent years, statistical post-processing techniques (Vannitsem et al., 2021) have been successfully applied to a wide range of weather variables, but not yet in the context of probabilistic weather regime forecasts. We apply a two-step statistical post-processing approach rooted in Sklar's theorem from multivariate statistics (Schefzik et al., 2013; Scheuerer and Hamill, 2015; Schefzik, 2017) to post-process the ensemble forecast of the 7-dimensional weather regime index. Ensemble model output statistics are applied to each regime individually to calibrate the marginal distributions (e.g., correct bias and spread). Ensemble copula coupling then restores the original dependence structure across regimes by reordering samples from these calibrated marginals to match the rank structure of the raw ensemble – effectively using its empirical copula.

These statistical methods learn relations from past forecasts and observations, correcting the forecasts independently of the prevailing atmospheric state. However, studies have shown that the performance of numerical weather prediction (NWP) models is also influenced by the initial atmospheric state (e.g., Vitart et al., 2010; Roberts et al., 2023). Hybrid approaches that combine information from the slowly varying components of the climate system, e.g. the stratospheric polar vortex (SPV) or tropical weather phenomena like the Madden-Julian Oscillation (MJO), with probabilistic weather regime forecasts from NWP systems provide a promising avenue to enhance predictability and better exploit available information.

The overarching goal of this thesis is to develop more useful forecast information tailored to the needs of the energy sector.

To achieve this, the research builds on the concept of weather regimes. The first part of the work investigates the link between Dunkelflauten and weather regimes, with a particular focus on identifying whether specific regimes are associated with cold Dunkelflauten which present a worst-case stress scenario for the energy system. For the first time, we apply probabilistic, multivariate post-processing to weather regime forecasts. While previous research has concentrated on teleconnections affecting the occurrence of weather regimes on the medium-range forecast horizon (up to two weeks), this work shifts the focus to the subseasonal range, with a particular focus on week 3. A new weather regime metric is introduced, designed to provide weekly aggregated, probabilistic information on strong regime projections. Using this metric, a hybrid approach that integrates a-priori information about the atmospheric state with NWP model outputs is developed.

The thesis is structured as follows: Chapter 2 provides the necessary background, introducing key concepts including weather regimes, atmospheric modes of variability, and ensemble forecasting. It also covers why statistical post-processing and machine learning approaches are used to improve forecast skill, and how weather regimes are relevant for the energy sector. Chapter 3 outlines the research questions addressed in this work. Chapter 4 describes the data and methods used throughout the thesis. In Chapter 5, the relationship between Dunkelflauten in Germany and weather regimes is established and analysed. Weather regime forecasts are statistically post-processed in Chapter 6. Chapter 7 analyses windows of opportunity in the context of weather regime forecasts and explores hybrid modelling approaches that combine a-priori knowledge of the atmospheric state with NWP model outputs. Chapter 8 integrates all research components in a comprehensive case study, discussing case-specific findings in relation to the overall results. Finally, Chapter 9 summarises the main findings and highlights their practical relevance.

2. Theoretical background

This chapter provides the theoretical foundation for the topics addressed in this thesis. It begins with an introduction to mid-latitude dynamics (Section 2.1) and the concept of weather regimes (Section 2.2), followed by a discussion of their relevance to the energy sector (Section 2.3). Next, the chapter explores large-scale atmospheric modes of variability (Section 2.4), the fundamentals of ensemble forecasting (Section 2.5), and the importance of the subseasonal to seasonal timescale, with particular attention to the influence of teleconnections on forecast skill (Section 2.6).
Finally, it introduces statistical post-processing techniques and highlights the growing role of machine learning in atmospheric science (Section 2.7).

2.1. Introduction to mid-latitude dynamics

The mid-latitudes refer to the region between 30° and 60° latitude in both hemispheres, situated between the (sub-)tropics and the polar regions. This zone is home to the majority of the world's population and economic activity. The atmosphere in this region is characterised by pronounced temperature variability, strong horizontal temperature gradients, distinct seasonal changes, and rapidly evolving weather systems. Temperature gradients influence both the horizontal and vertical structure of the wind. Horizontally, they shape the pressure gradient, which drives the geostrophic wind. Vertically, through the thermal wind relationship, stronger horizontal temperature contrasts result in greater vertical wind shear – that is, a larger change in wind speed and direction with height. In the upper troposphere, this shear intensifies the westerlies into narrow bands of high-velocity winds known as jet streams. These jets form where horizontal temperature gradients are strongest and include two primary types: the subtropical jet, located near 30° latitude at the poleward edge of the Hadley cell, and the polar front jet, typically near 60° latitude, marking the boundary between cold polar and warm mid-latitude air. The latter is often referred to as the mid-latitude jet, due to its variable position between 30° and 60° latitude.
The strength and position of the jet streams vary seasonally. During winter, the equator-to-pole temperature contrast – and hence the pressure gradient – is stronger, leading to a more intense and zonally oriented polar front jet which can shift further equatorward. In summer, the poleward region warms and the reduced temperature gradient weakens the jet and causes it to shift poleward, exhibiting more meridional features and it is often fragmented.
In addition to seasonal changes, topography and land-sea contrasts can influence the atmospheric flow by inducing stationary waves. These manifest as persistent undulations in the geopotential height field

and are observed as equatorward or poleward displacements of the westerlies. These displacements are also known as Rossby waves. The jet streams act as waveguides for Rossby waves, channelling the development and propagation of synoptic-scale weather systems. The meandering of the jet stream gives rise to alternating ridges and troughs that steer cyclones and anticyclones along their paths.
The characteristics of Rossby wave propagation depend on jet stream strength. A strong jet supports faster-moving Rossby waves with smaller meridional amplitude, efficiently transporting energy downstream. A weaker jet allows for slower, larger-amplitude waves that can amplify, become quasi-stationary, or even break. These dynamics can lead to atmospheric blocking – quasi-stationary high-pressure systems that interrupt the typical west-to-east progression of weather systems (e.g., Rex, 1950). Blocking events are associated with persistent and anomalous weather conditions. Depending on their location and season, they can lead to cold air outbreaks, heatwaves, droughts, or prolonged periods of precipitation (see review by Kautz et al., 2022). For example, a blocking high over Northern Europe may cause hot and dry conditions locally, while a downstream trough can result in sustained rainfall and flooding.

2.2. Weather regimes

The impact of atmospheric blocking on weather in the Atlantic-European region strongly depends on the block's location. One approach to categorising such variability is through the concept of weather regimes.
Weather regimes are based on the idea that the large-scale atmospheric circulation can in practice be represented by a finite number of possible atmospheric states that manifest themselves in quasi-stationary, persistent, and recurrent large-scale flow patterns (Grams et al., 2020). These regimes represent the low-frequency variability of the weather systems affecting continent-sized regions over periods typically longer than 10 days.
Because weather regimes modulate surface weather variables – specifically temperature, wind speed, precipitation, and cloud cover – they directly affect the energy sector due to fluctuations in renewable energy yields by wind and solar (e.g., Grams et al., 2017; Beerli and Grams, 2019; Wiel et al., 2019a; Gerighausen et al., 2024).
The concept of weather regimes has its origins in the mid-20th century. Early work by Baur et al. (1944), later refined by Hess and Brezowsky (1969), introduced the Grosswetterlagen – a set of synoptic-scale flow types over the East Atlantic-European region. Although still used operationally and maintained by the German Weather Service, the Grosswetterlagen classification is based on subjective analysis of daily synoptic patterns and differs methodologically from modern, statistically derived weather regimes. In particular, it does not require criteria such as persistency or quasi-stationarity, which are fundamental to the definition of weather regimes.
A more dynamical perspective emerged with Rex (1950), who introduced the phenomenon of atmospheric blocking based on events over the North Pacific and North Atlantic. Motivated by the fact that

the variability between the blocking events was small, he introduced a European weather regime associated with atmospheric blocking over the eastern North Atlantic (Rex, 1951).
Later, theoretical and modelling studies offered further insights. Charney and Devore (1979) used a barotropic model with topography to demonstrate bimodal flow behaviour – zonal versus blocked – which may be produced by the same forcing. This idea was extended by Legras and Ghil (1985), who found evidence for multiple equilibria when allowing more degrees of freedom. Hoskins and Karoly (1979) suggested that planetary waves and orographic forcing are key drivers of atmospheric blocking.

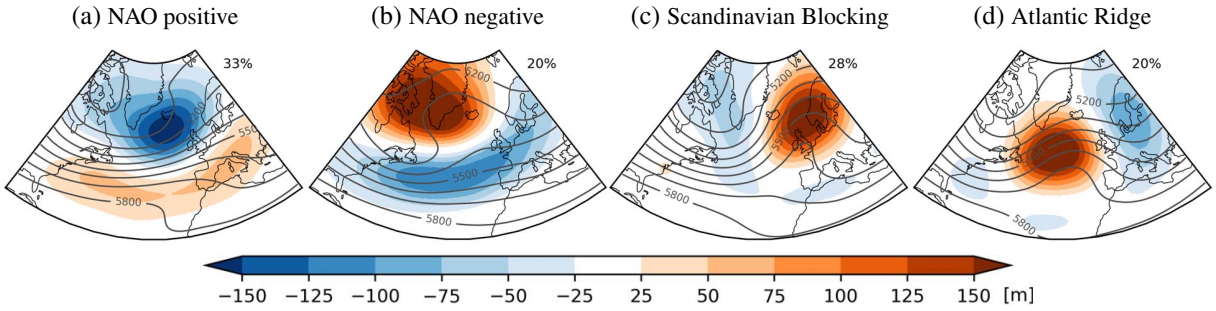


Figure 2.1.: 500 hPa geopotential height (contours in *m*) and anomalies to climatology (shading in *m*) for the four winter weather regimes from Cassou (2008), based on ERA5 data (1979–2018, DJF, 90°W–30°E, 20°N–80°N). Values indicate the percentage of total days categorised into each regime. Reprinted from Wiel et al. (2019b).

Statistical-dynamical approaches emerged in the late 1980s. Vautard and Legras (1988) introduced a regime classification that incorporated feedbacks from transient eddies, which themselves depend nonlinearly on the large-scale flow, splitting the regimes into two main families: zonal and blocking types. In a subsequent study, Vautard (1990) applied this method to 37 winters and identified nine significant patterns, which were later reduced to four dominant regimes. These include two regimes resembling the positive and negative phases of the North Atlantic Oscillation (NAO) – one zonal and one with blocking over Greenland (Figure 2.1a,b), respectively – as well as two regimes dominated by anticyclonic anomalies: Scandinavian Blocking (Figure 2.1c) and the Atlantic Ridge (Figure 2.1d) – with a blocking over the European/Scandinavian and eastern Atlantic region, respectively.
Since then, many weather regime definitions have been proposed, differing by method, season, and time period (e.g., Hannachi et al., 2017). Despite methodological differences – like cluster analysis (e.g., Michelangeli et al., 1995; Cassou et al., 2004) or principal component analysis (e.g., Vautard, 1990) of geopotential height anomalies or mean sea level pressure – all approaches aim to identify regimes with three key properties:

Recurrence: The most frequent patterns occurring during a climatological period, indicating the highest occurrence frequency of certain flow configurations.

Quasi-stationarity: The large-scale atmospheric motion remains relatively unchanged over time, characterised by a vanishing time derivative, which allows the flow pattern to be treated as stationary on synoptic timescales.

Persistence: the tendency of the large-scale flow pattern to reside for an extended period – typically from a week to a month (Grams et al., 2020) – requiring circulation anomalies associated with the pattern to be maintained over that duration.

The commonly used four-regime framework by Vautard (1990) is primarily defined for winter only. Cassou et al. (2005) introduced a summer counterpart with four regimes. The "Blocking" regime (NAO+) features 500 hPa geopotential height anomalies over Scandinavia, while the "NAO–" regime shows a dipole of anomalies between Greenland and Northern Europe. Two additional regimes – the Atlantic Low and Atlantic Ridge – are characterised by a deep anomalous trough over the North Atlantic or a strong positive anomaly off Western Europe, respectively.

To find a definition of weather regimes applicable across all seasons, Grams et al. (2017) introduced a year-round regime classification consisting of seven regimes that represent the North Atlantic-European large-scale flow throughout the year. Since this seven-regime framework provides the basis for the analysis in this thesis, we describe it in more detail in the following.

The seven year-round weather regimes represent recurring atmospheric circulation patterns that combine characteristics of the classical four-regime structures from both summer and winter (e.g., Cassou et al., 2005, 2004). These regimes consolidate to seven rather than eight categories because the Greenland Blocking regime – with a pronounced NAO– pattern – exhibits a similar structure across seasons.

The regimes are primarily characterised by anomalies in the 500 hPa geopotential height field (Z500) relative to the climatological mean. The climatology typically features a trough over North America, weak ridging over the eastern North Atlantic, and a dominant westerly upper-level flow across the North Atlantic into Europe. Based on the prevailing Z500 anomalies, the seven regimes can be grouped into two broad categories:

Cyclonic regimes: Atlantic Trough, Zonal Regime, and Scandinavian Trough (Figure 2.2a–c)

Anticyclonic (blocked) regimes: Atlantic Ridge, European Blocking, Scandinavian Blocking, and Greenland Blocking (Figure 2.2d–g)

Each regime is described below in terms of its Z500 characteristics (Figure 2.2), seasonal occurrence, and its influence on the surface variables 2 m temperature (Figure 2.3a–g for November–March) and 100 m wind speed (Figure 2.3h–n for November–March).

The Atlantic Trough (AT) is characterised by a prominent negative Z500 anomaly west of Ireland and a strong westerly upper-level flow across the North Atlantic into Europe, especially between 40°–50°N. Surface temperatures show positive anomalies across Europe year-round, with winter warming reaching up to +3 K. Wind speeds are enhanced across the eastern Atlantic and parts of Western Europe, especially in winter. AT occurs throughout the year and is the most frequent cyclonic regime during summer.

The Zonal Regime (ZO) features a broad negative Z500 anomaly centred over Iceland and southern Greenland, accompanied by weakly anticyclonic conditions due to an enhanced ridge over western continental Europe. The resulting upper-level flow remains zonal across the eastern North Atlantic, shifting

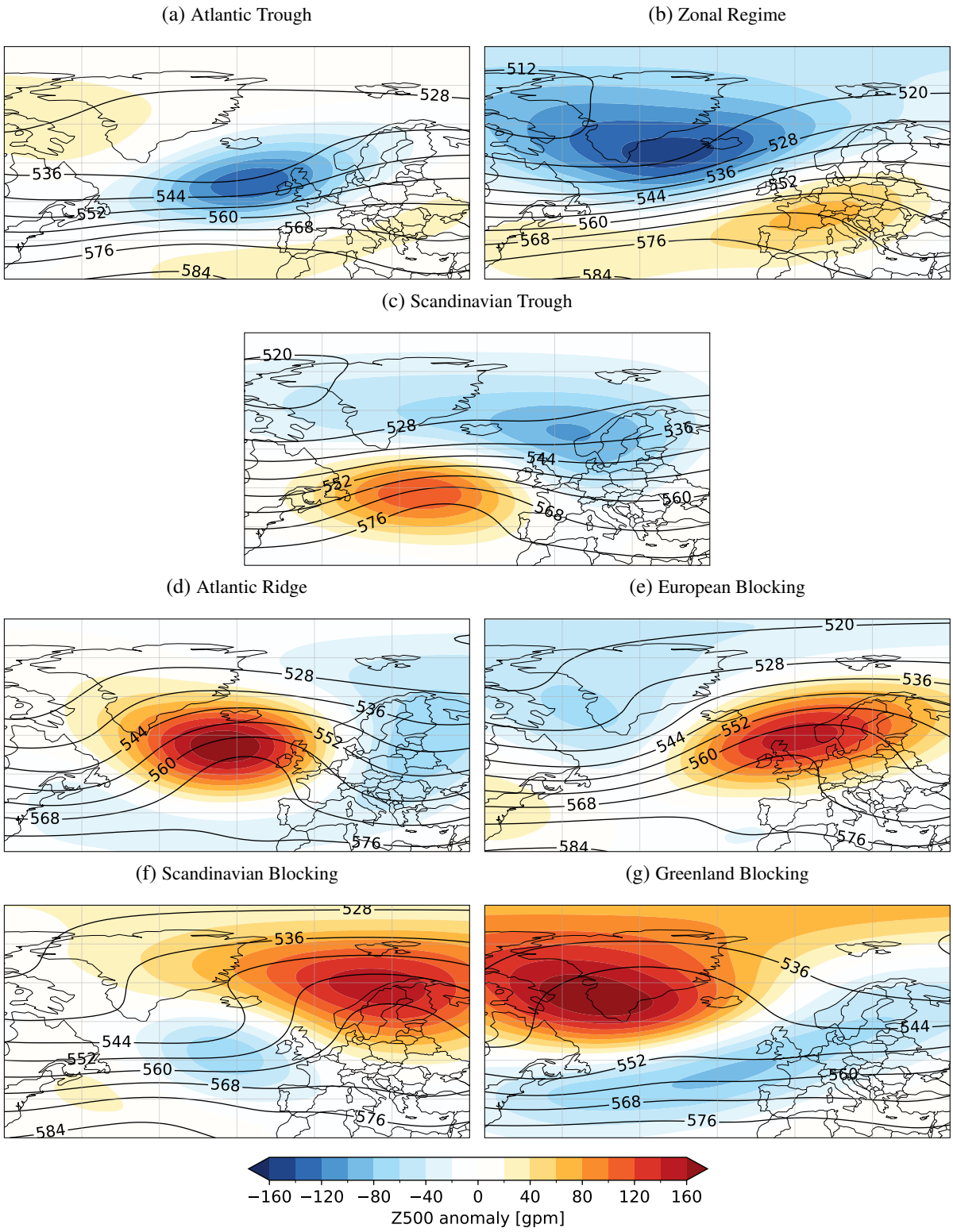


Figure 2.2.: 500 hPa geopotential height field (contour in gpdm) with their anomalies to climatology (shading in gpm) for the seven year-round weather regimes, splitted into (a–c) cyclonic regimes and (d–g) anticyclonic or blocked regimes. Applied to data from 1979–2019. The region shown (80°W–40°E, 30°N–90°N) equals the seven year-round weather regime region.

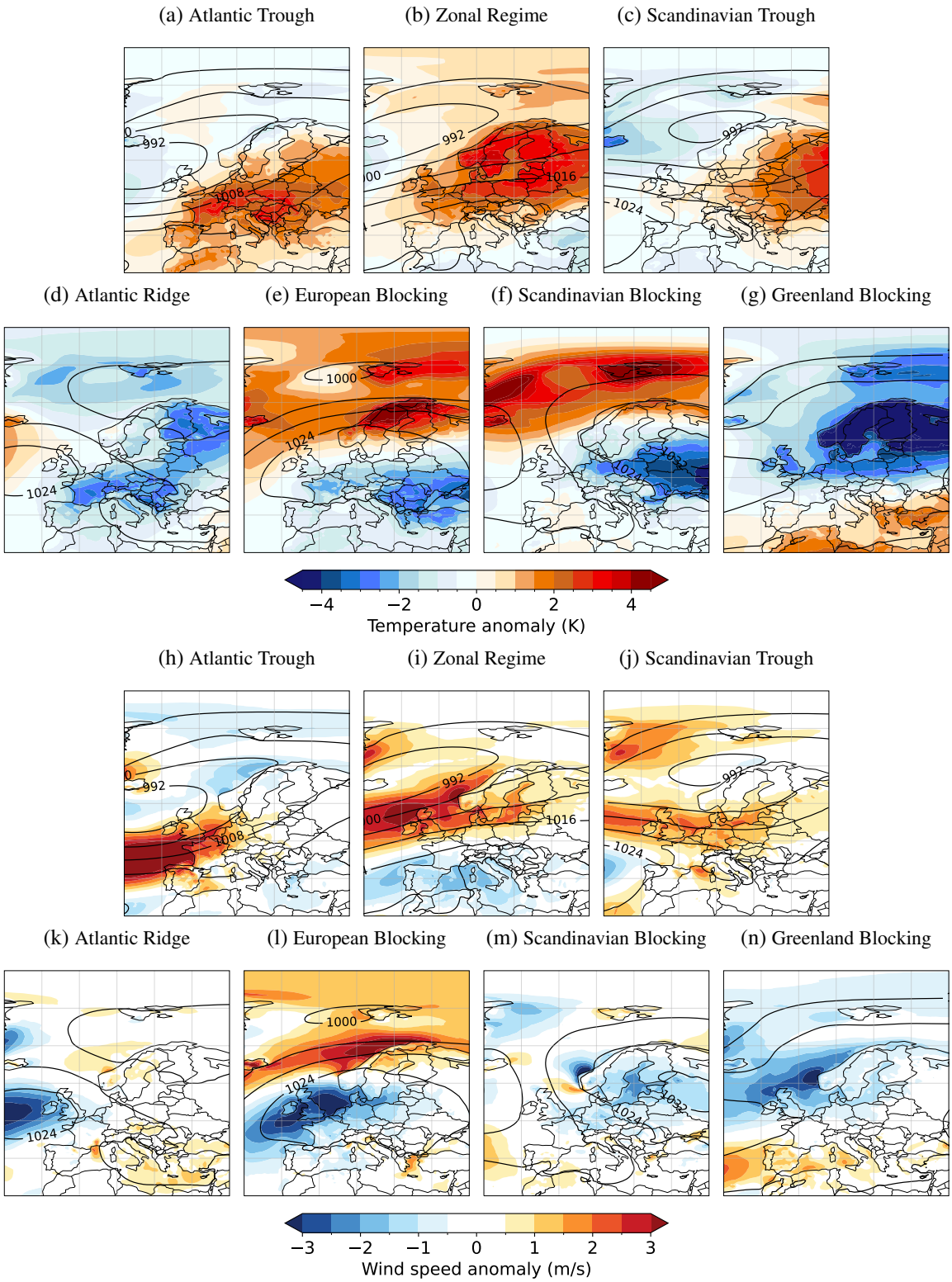


Figure 2.3.: Composites of atmospheric field variables for weather regimes. (a–g) 2 m temperature anomalies (shading in K) and (h–n) 100 m wind speed anomalies (shading in $\frac{m}{s}$) with respect to a 31-day running mean climatology. The contours indicate the composite of the mean sea level pressure (hPa). Applied to data from 1979–2022 and within the extended winter period (November–March). The region shown (20°W–40°E, 30°N–90°N) equals the European part of the seven year-round weather regime region.

to south-westerly over the North Sea and Scandinavia. Surface temperature anomalies are positive across Central and Northern Europe, peaking at +4 K in winter. Wind anomalies are strongest in the North Atlantic south of Iceland. ZO is most common in winter, when it dominates the regime distribution.

The Scandinavian Trough (ScTr) is defined by a negative Z500 anomaly over Northern Europe, shifted eastward compared to ZO. A weak ridge is present over the North Atlantic (40°–50°N). Westerly upper-level winds prevail further north, compared to AT. Positive wind anomalies are most notable around the British Isles and the North Sea, especially during winter. Near-surface temperature anomalies are positive in northeastern Europe and western Russia, extending into Central Europe during winter, while negative anomalies dominate over Central Europe during the remaining seasons. ScTr primarily occurs in winter and transitional seasons.

The Atlantic Ridge (AR) is marked by a strong positive Z500 anomaly south of Iceland, resulting in a blocking ridge west of Ireland and a downstream trough influencing much of Europe. The upper-level westerly flow is blocked and deflected northward over Northern and Western Europe. Surface temperatures are notably colder across continental Europe, particularly pronounced in winter and spring. Wind anomalies are minor, with weak negative values west of Ireland, whereas continental Europe does not experience any anomalies. AR is most frequent in autumn and winter rather than in summer.

The European Blocking (EuBL) displays a positive Z500 anomaly centred over the North Sea and surrounding Western and Central Europe. Flanking troughs appear up- and downstream over the Labrador Sea and southeastern Europe. The prevailing upper-level flow is deflected into southwesterlies from Newfoundland over Iceland to northern Scandinavia. Warm anomalies dominate Northern Europe, extending into Central Europe in spring and summer, whereas cold anomalies prevail in Central Europe in winter and autumn. Wind speed is reduced over the North Sea and west of Great Britain, while enhanced in the Norwegian Sea, strongest in winter. EuBL frequency peaks in summer.

The Scandinavian Blocking (ScBL) features a positive Z500 anomaly over northern Scandinavia, flanked by weaker negative anomalies over the eastern North Atlantic and Western Europe. The upper-level flow splits, with a branch deflected poleward around the anticyclone and westerly flow over Iberia and the Mediterranean. Surface temperatures are warm across Northern and Central Europe except during winter, when cold anomalies from the east push into Central Europe. Wind anomalies are weak, with reductions in Northern Europe and slight positive anomalies north-west of Portugal. ScBL is the most frequent regime during summer.

The Greenland Blocking (GL) is dominated by a strong positive Z500 anomaly over Greenland and a zonally elongated negative anomaly stretching from the eastern North Atlantic into Northern Europe. The prevailing upper-level westerly flow is deflected southward (around 40°N), reaching into the Mediterranean. Surface temperatures show pronounced negative anomalies from the North reaching into Central Europe, particularly during winter (up to –4.5 K in Northern Europe), while southern Europe experiences mild warming. Wind speeds increase west of Portugal during winter and autumn and decrease over the

Norwegian and North Seas, especially in winter. GL is most frequent in spring and is the dominant springtime regime.

2.3. Weather regimes and the energy sector

The increasing share of variable renewable energy sources in the power system, combined with rising electricity demand driven by electrification (Edenhofer et al., 2011; Bloomfield et al., 2021b), has made the energy system increasingly sensitive to meteorological variability (Bloomfield et al., 2016; Wiel et al., 2019a). As a result, understanding and forecasting weather patterns has become essential for managing energy supply and demand.

Large-scale weather regimes have proven particularly valuable for analysing energy-relevant metrics, such as wind power generation across Europe. For instance, blocked regimes – such as the Scandinavian Blocking and Greenland Blocking – are typically associated with reduced wind power generation and increased energy demand. These conditions raise the risk of positive residual load, by factors of approximately 1.5 and 2.0, respectively (Wiel et al., 2019b). This general finding is supported by several studies; for example, Charlton-Perez et al. (2019) showed that GL is often linked to prolonged cold spells across Europe, while Bloomfield et al. (2020b) found that high-pressure systems extending westward from Russia to Scandinavia are associated with peak European power demand, owing to cold temperatures and moderate wind speeds. Grams et al. (2017) demonstrated that average wind power output across Europe and across different regimes ranges from 22 GW (AT) to 44 GW (EuBL) (numbers for the power system of 2015), a spread expected to triple by 2030 under the planning strategies at the time. Brayshaw et al. (2011) showed that United Kingdom mean wind power output varies by up to 10% depending on the state of the NAO, when days are categorised into high, medium, and low NAO index values.

However, characterising impacts based solely on inter-regime differences may overlook important nuances. Intra-regime variability can be equally critical for operational decision-making (Gerighausen et al., 2024; Wiel et al., 2019b). Gerighausen et al. (2024) explored the GL regime and highlighted how secondary regime projections influence cold extremes. Strong projections into GL combined with positive projections into the AR were linked to severe cold events, whereas similar GL strengths paired with the AT did not show extreme cold temperatures.

The transition to renewables also affects the baseload structure of the power system. As renewable capacity grows, the average output from baseload plants (e.g., nuclear) diminishes, while inter-annual variability in system load increases, introducing additional challenges (Bloomfield et al., 2016). Although peak demand events are generally expected to decrease in frequency, the reduction is spatially heterogeneous. For instance, peak demand reductions are projected to be more pronounced in Spain than in Central Europe (Bloomfield et al., 2020b). Spatial deployment strategies for new wind capacity also play a key role; deploying wind farms in the Balkans instead of the North Sea, for example, could balance inter-annual variability (Grams et al., 2017).

On the forecasting side, pattern-based methods (such as weather regimes) have shown considerable promise at the S2S scale. Bloomfield et al. (2021a) compared traditional grid-point-based forecasts (e.g., using 10 m wind to estimate national wind power output) with pattern-based approaches that first classify the synoptic-scale circulation before estimating surface impacts. While grid-point forecasts exhibit superior skill at short lead times (days 0–10), pattern-based forecasts often outperform at subseasonal lead times (day 12+). A hybrid approach, where grid-point forecasts are conditioned on the confidence of regime forecasts (e.g., when $\geq$50% of ensemble members agree on a specific regime), has been shown to enhance skill in week 1 (days 5–11) and week 2 (days 12–18)(Bloomfield et al., 2021a). Cionni et al. (2022) emphasise that including additional relevant physical processes in the dynamical system, e.g. by including statistical models of teleconnections like the NAO, could further improve S2S forecasting performance.

Improved S2S forecasts have clear operational benefits. Weather-related risks are key drivers of electricity pricing, production planning, and demand patterns (White et al., 2017). Incorporating reliable S2S information into energy sector operations can reduce such risks by informing decisions on generation dispatch and market operations. Grid operators could optimise pricing schemes and generator scheduling – especially for slow-start sources like nuclear – based on forecasts of key supply-side drivers (e.g., wind speed for wind power, precipitation and temperature for hydropower operations) and demand-side indicators (primarily temperature for heating or cooling). Such enhancements are critical to maintaining the system in balance, especially under growing renewable shares and variable loads (White et al., 2017).

2.4. Atmospheric modes of variability

Though weather regimes represent preferred quasi-stationary states of the large-scale atmospheric circulation, their frequency and persistence are not solely determined by internal atmospheric dynamics. Instead, they are modulated by various large-scale modes of variability and teleconnection patterns (Figure 2.4).

Ångström (1935) originally described teleconnection patterns as: “The weather at a given place is not an isolated phenomenon but is intimately connected with the weather at adjacent places.” This foundational concept of teleconnections highlights how temporal variations in an atmospheric variable at one location can be significantly correlated with variations in the same or a different variable at a distant location. Furthermore, the strength and structure of these correlations can be influenced by the atmospheric state at a third location.

Evidence of humans recognising such long-distance linkages dates back to as early as approximately 1000 AD, when Viking settlers are believed to have observed opposing pressure patterns between Greenland and Europe – a phenomena known today as the North Atlantic Oscillation (Feldstein and Franzke, 2017; Zavadoff and Arcodia, 2022). Today, the impacts of teleconnection patterns are widely recog-

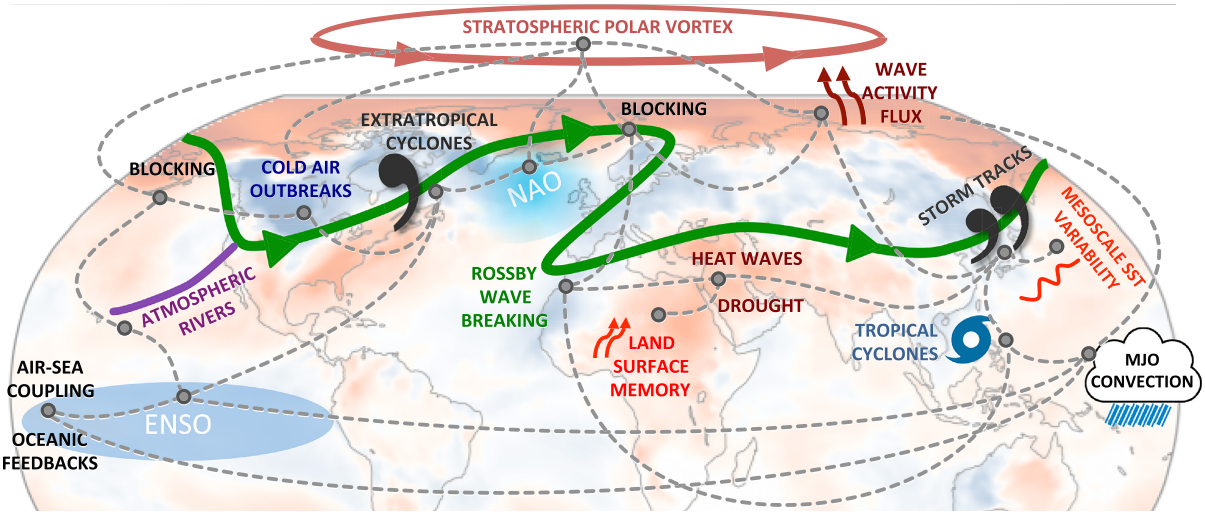


Figure 2.4.: A schematic overview of many of the atmospheric phenomena and the teleconnections necessary to make accurate forecasts in the subseasonal to seasonal time scale. Adopted from Lang et al. (2020)

nised, although developing physically robust explanations for individual linkages remains a continuing challenge.

Numerous connections have been established between atmospheric modes of variability and NAE weather regimes – whether considering the classical four winter regimes or the expanded set of seven year-round regimes. The following provides a brief overview of the main large-scale modes of variability and their associated teleconnections with these regimes.

2.4.1. Madden-Julian Oscillation

The Madden-Julian Oscillation is the dominant mode of intraseasonal (30–90 days) variability in the tropical atmosphere, first described by Madden and Julian (1971, 1972). It represents the atmospheric circulation and moist convection in the tropics (Figure 2.5), characterised by large-scale convective anomalies that propagate eastward across the equatorial Indian Ocean and the western to central Pacific at an average speed of approximately 5 $\frac{m}{s}$ (Zhang, 2005). The MJO continually interacts with the underlying ocean and exerts significant influence on weather and climate systems around the globe (Matthews, 2008).

At its core, the MJO features a dipole structure consisting of an eastward-propagating centre of enhanced deep convection and precipitation, flanked by regions of suppressed convection. These convective anomalies are linked to vertically extensive zonal overturning circulations. In the lower troposphere (around 850 hPa), the convective centre is typically accompanied by anomalous strong westerly winds to the west and easterly winds to the east. In the upper troposphere (around 200 hPa), the zonal wind anomalies reverse direction (e.g., Zhang, 2005; NOAA, 2025c).

Although the term “oscillation” implies regularity, the MJO is not strictly periodic and by no means a regular pattern. Its return period varies between 30 and 100 days and exhibits notable variations in propagation speed, seasonal strength, and interannual variability (Zhang, 2005). The propagation and ampli-

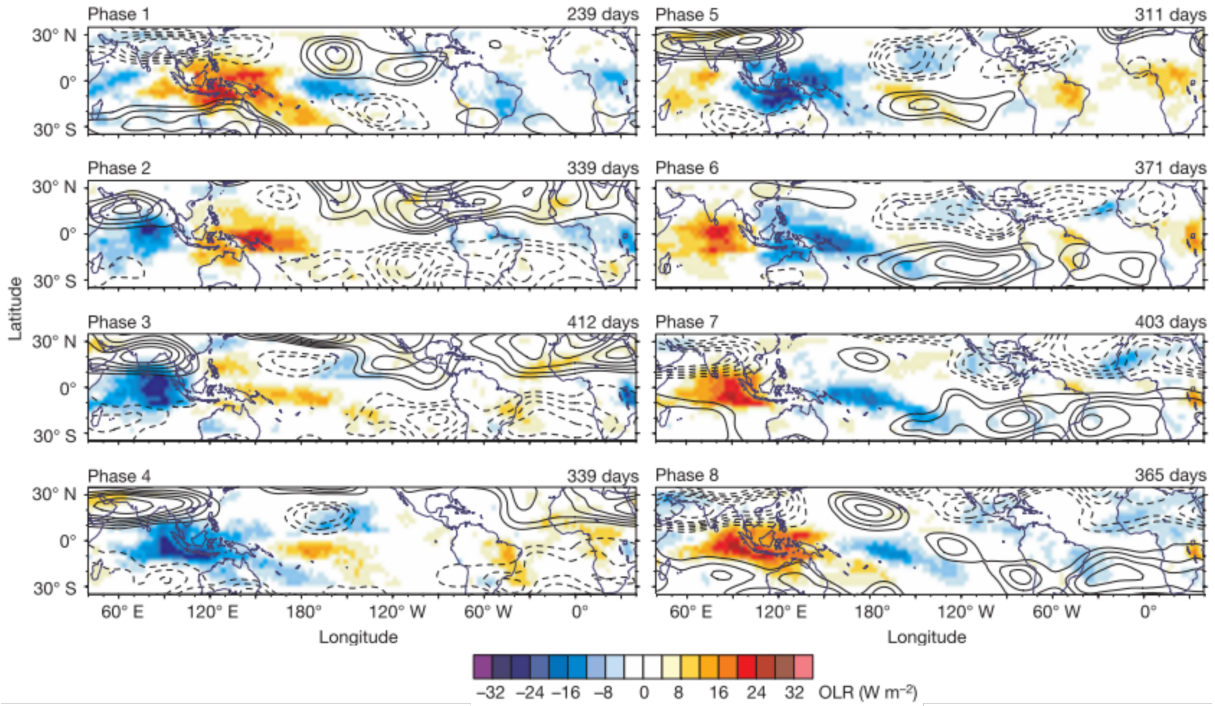


Figure 2.5.: Dynamical and thermodynamical signatures of the eight MJO phases. Wintertime composites of outgoing longwave radiation (colour shading, intervals of 4 $\frac{W}{m^2}$) and 300 hPa stream function anomalies (contours) for the eight MJO phases. Only strong MJO events are included; the number of days per phase is indicated. Blue shading indicates enhanced convection and wetter conditions; red shading indicates suppressed convection and drier conditions. Solid contours in the Northern Hemisphere and dashed contours in the Southern Hemisphere represent anomalous anticyclonic circulation. Figure adapted from Cassou (2008) and modified.

tude of the MJO are commonly tracked using the real-time multivariate MJO (RMM) index (Wheeler and Hendon, 2004), which captures its evolution through a characteristic eight-phase lifecycle (Figure 2.5). The MJO also influences the mid-latitude circulation via the excitation and propagation of Rossby waves, which can modulate the atmospheric flow over the North Atlantic-European region (Cassou, 2008; Lin et al., 2009). When the MJO is active over the eastern Indian Ocean and Maritime Continent (phases 2–5), it tends to force Rossby waves that act as precursors to NAO+ phases in the following two weeks. Conversely, activity over the central Pacific (phases 6–8) is associated with precursors to NAO– phases (e.g., Cassou, 2008; Roberts et al., 2023). These Rossby waves do not directly trigger NAO phases but instead modify the background atmospheric flow, creating a more favourable environment for the development of particular regimes. The full manifestation of an NAO phase often requires additional interaction with transient eddies (e.g., cyclones or anti-cyclones in the troposphere) – particularly through wave breaking processes (Cassou, 2008).

Beyond the NAO, the MJO also modulates blocking frequencies over the North Atlantic and Europe. Henderson et al. (2016) found that blocking over the North Atlantic decreases in the two weeks following MJO phases 1–5 and increases after phases 6–8. Similarly, blocking over Europe decreases after phases 3–5 and increases following phases 6–8.

2.4.2. Quasi-Biennial Oscillation

The Quasi-Biennial Oscillation (QBO) is the leading mode of interannual variability in the tropical stratosphere and is characterised by a regular reversal of the mean zonal wind direction between easterly (EQBO) and westerly (WQBO) phases, with a typical period of approximately 28 months (Feng and Lin, 2019; Baldwin et al., 2001). This alternating wind pattern occurs primarily in the equatorial lower stratosphere and arises from the interaction between vertically propagating waves and the mean flow.
The QBO modulates tropical convection and has been shown to significantly influence the amplitude of the Madden-Julian Oscillation during boreal winter. Specifically, MJO activity tends to be enhanced during EQBO phases (i.e., when lower stratospheric winds are anomalously easterly), while it is suppressed during WQBO phases (anomalously westerly winds) (Yoo and Son, 2016; Zhang and Zhang, 2018).
While the QBO exerts a relatively direct influence on the MJO activity, it also plays a modulatory role in the teleconnection between the MJO and the NAO. According to Feng and Lin (2019), the tendency for NAO+/NAO– patterns to follow MJO phases 3 and 7, respectively, is more pronounced and persistent during WQBO phases, but is weaker and less statistically significant during EQBO phases.

2.4.3. El Niño-Southern Oscillation

The El Niño-Southern Oscillation (ENSO) is the most prominent mode of interannual climate variability on Earth. It originates in the tropical Pacific Ocean and exerts widespread influence on global atmospheric circulation and weather patterns. ENSO consists of alternating phases: El Niño, characterised by warmer-than-normal sea surface temperatures (SSTs) in the central and eastern equatorial Pacific, and La Niña, associated with cooler-than-normal SSTs in the same region. El Niño events typically persist for 9–12 months, whereas La Niña can last from 1 to 3 years. A full ENSO cycle recurs every 2–7 years and El Niño and La Niña tend to peak during boreal winter (Dutton, 2021).
The atmospheric component of ENSO is referred to as the Southern Oscillation, which is commonly measured by the sea-level pressure difference between Darwin (Australia) and Tahiti. This pressure difference is typically negative during El Niño and positive during La Niña, reflecting changes in the strength and orientation of the Walker circulation – a large-scale tropical convective circulation driven by SST gradients. La Niña events are associated with a strengthening of the Walker circulation, while El Niño weakens or even reverses it, leading to drastic shifts in the Walker circulation due to deep atmospheric convection (Dutton, 2021).
ENSO also modulates extratropical atmospheric circulation, including impacts on the NAE region. In early winter (December), El Niño events can produce a precipitation dipole over the tropical Indian Ocean, characterised by enhanced rainfall in the west and suppressed rainfall in the east. This dipole affects the subtropical South Asian jet, promoting a circulation pattern that projects onto NAO+. In late winter (February), however, the North Atlantic-European response shifts, with ENSO projecting onto

NAO–, likely due to the Pacific Ocean exerting a stronger influence than the Indian Ocean at that time (Abid et al., 2021).
Moreover, ENSO modulates the MJO-NAO teleconnection. During El Niño years, the likelihood of NAO+ following MJO phases 1–5 is enhanced, whereas this signal is suppressed during La Niña years. Conversely, NAO– following MJO phases 7–8 is more likely during La Niña and dampened during El Niño (Lee et al., 2019; Henderson and Maloney, 2018).
ENSO also interacts with the QBO-MJO relationship. Sun et al. (2019) extend the relations described from Yoo and Son (2016); Zhang and Zhang (2018) by showing that the QBO's modulation of MJO activity is further conditioned by the ENSO phase. Specifically, MJO activity is most suppressed during La Niña combined with WQBO, and enhanced during La Niña combined with EQBO. In contrast, during El Niño years, MJO activity is generally enhanced, with minimal additional modulation by the QBO.

2.4.4. Stratospheric polar vortex

The stratospheric polar vortex is a planetary-scale cyclonic circulation that forms in the winter hemisphere's stratosphere due to the pronounced meridional temperature gradient between the cold polar regions and the warmer tropics (Baldwin and Dunkerton, 2001; Lee, 2021). This thermal contrast, via the thermal wind balance, generates a band of strong westerly winds – the polar night jet – which encircles the winter pole, peaking near 60° latitude and extending from the upper troposphere (around 100 hPa) into the mesosphere (above 1 hPa). Together, the cold stratospheric air and the polar night jet define the structure of the SPV.
The SPV typically begins to form in late August, reaches its maximum strength in mid-winter (around January), and breaks down by April. During boreal summer, the stratosphere is dominated by easterly winds, and the vortex is absent. Common diagnostics for SPV strength include the zonal mean zonal wind at 60°N at 10 hPa or 100 hPa, the area enclosed by low geopotential heights surrounded by westerlies, and regions of high potential vorticity (Waugh et al., 2017). The SPV plays a central role in stratosphere-troposphere coupling, especially in modulating the position and strength of the tropospheric jet stream. A strong vortex tends to support a poleward-shifted and intensified jet, limiting cold air to the Arctic. Conversely, a weak vortex is associated with a weaker, more meandering jet stream that facilitates cold air intrusions into mid-latitudes (Figure 2.6). These conditions are closely related to the phases of the Arctic Oscillation (AO) and the North Atlantic Oscillation (Kidston et al., 2015).
A key process influencing the SPV is the upward propagation of planetary-scale Rossby waves from the troposphere. These waves, especially those with low zonal wavenumbers (1–3), can propagate vertically into the stratosphere under westerly wind conditions. Upon reaching the stratosphere, they can decelerate the polar night jet via wave breaking and momentum deposition. The Northern Hemisphere, due to greater topographic forcing and land-sea contrasts, produces stronger wave activity than the Southern Hemisphere, making the Arctic vortex weaker and more variable than its Antarctic counterpart (Lee, 2021; Waugh et al., 2017).

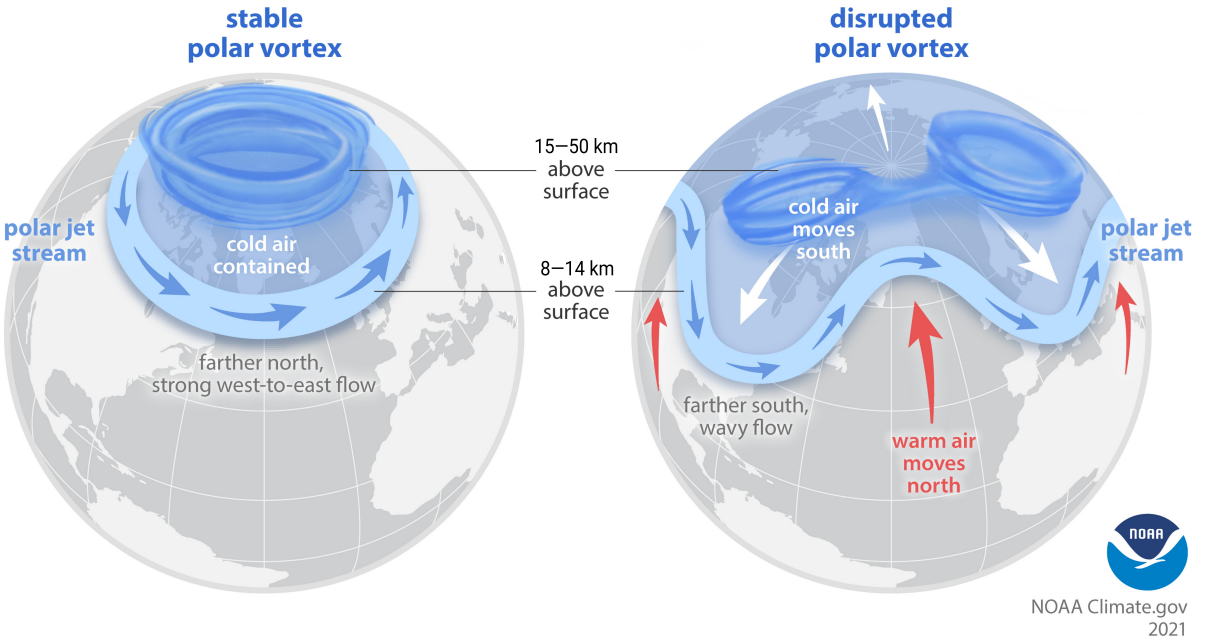


Figure 2.6.: Schematic of a stable vs. disrupted Arctic polar vortex and associated polar jet stream patterns. The left panel illustrates a stable SPV with a strong, zonally oriented polar jet stream that confines cold air masses to the Arctic region. In contrast, the right panel shows a disrupted SPV, characterised by a wavier polar jet stream that allows cold Arctic air to move southward and warm air to intrude northward. Adopted from Lindsey (2021).

An extreme manifestation of vortex weakening is known as a sudden stratospheric warming (SSW). These events are marked by a rapid increase in polar stratospheric temperatures – often exceeding 50 K in a matter of days – and a dramatic weakening or even reversal of the zonal winds (Charlton and Polvani, 2007). Major SSWs, defined by a complete reversal in wind direction, occur roughly every second winter in the Arctic. They are typically classified into (1) displacement events, where the vortex is shifted off the pole (often linked to amplified wave-1 activity) or (2) split events, where the vortex splits into two smaller vortices (associated with amplified wave-2 activity). These disturbances can persist for several weeks before the vortex reforms (Charlton and Polvani, 2007).

The SPV state modulates NAE weather regimes, particularly through its influence on the NAO. According to Charlton-Perez et al. (2018), NAO– is especially sensitive to stratospheric variability and occurs on approximately 33% of days following a weak SPV, compared to only 5% of days following a strong SPV. In contrast, NAO+ and the Atlantic Ridge pattern exhibit weaker sensitivity to the SPV. Beyond the occurrence frequency, the SPV also alters the transition probabilities and persistence of regimes: weak vortex conditions favour transitions into NAO–, while strong vortex conditions favour NAO+.

These findings are supported by Beerli and Grams (2019), who investigated the relationship between the SPV and the seven year-round weather regimes. They found that regimes related to NAO+ (e.g., Zonal Regime, Scandinavian Trough) are more frequent during strong vortex states, while regimes related to NAO– (e.g., Greenland Blocking, and to a lesser extent Scandinavian Blocking) are favoured during and up to two months after a weak vortex, reflecting the persistence of tropospheric anomalies.

Further, Domeisen et al. (2020) show that Atlantic Trough patterns often follow SSWs, particularly when Greenland Blocking was present near the onset of the SSW. Conversely, Greenland Blocking becomes more likely when a blocking situation over Western Europe and the North Sea precedes the SSW.

2.5. Forecasting models

2.5.1. From deterministic to probabilistic forecasts

Bjerknes et al. (1904) introduced the concept that predicting the atmospheric state could be treated as an initial value problem in mathematical physics. They proposed that future weather conditions could be derived from current atmospheric observations by solving a set of partial differential equations. These prognostic equations include the Navier-Stokes equations, the continuity equation, the first law of thermodynamics, and the ideal gas law. Together, they describe the temporal evolution of wind, pressure, density, and temperature in the atmosphere. Due to their nonlinearity and complexity, these equations are not analytically solvable and must instead be approximated numerically using discretisation in space and time.

This discrete numerical framework cannot resolve all atmospheric processes explicitly. Sub-grid-scale processes – those occurring at spatial scales smaller than the grid resolution or at shorter timescales than the numerical time step – are represented through parameterisations that introduce source terms for mass, momentum, and energy (Bauer et al., 2015).

The skill of numerical weather prediction (NWP) is constrained by both model imperfections – including limitations in physical parameterisations – and uncertainties in initial conditions. The extent to which future atmospheric states can be predicted by an NWP model defines its practical predictability, which is shaped by observational capabilities, data assimilation quality, modelling fidelity, and computational resources. However, even with a hypothetical perfect model and error-free observations, the intrinsically chaotic nature of the atmosphere imposes a fundamental limit on predictability – referred to as intrinsic predictability (Lorenz, 1968; Melhauser and Zhang, 2012; Bauer et al., 2015).

Lorenz (1968) conceptualised atmospheric variability as a combination of periodic and non-periodic components. Periodic components, such as diurnal and seasonal cycles, are in principle infinitely predictable. In contrast, the chaotic, non-periodic components exhibit finite predictability: small errors in the initial state grow non-linearly over time, eventually rendering forecasts unreliable. Crucially, even substantial reductions in initial errors cannot entirely suppress this effect unless those errors are eliminated entirely. The intrinsic predictability limit is scale-dependent and can vary with the flow regime (Lorenz, 1995).

According to Zhang et al. (2019), the current practical predictability limit for mid-latitude instantaneous weather using deterministic models is approximately 10 days. Their findings indicate that reducing initial condition errors by an order of magnitude may extend this limit by approximately five additional days, implying that operational NWP systems still perform well below the theoretical predictability ceiling.

In response to these challenges, forecasting practice has undergone a major shift. Buizza and Leutbecher (2015) documented the transition over the past 25 years from a deterministic to a probabilistic approach. Instead of relying a single deterministic forecast run, modern systems now employ ensemble forecasting, whereby multiple forecasts are generated from perturbed initial conditions and/or model configurations. This ensemble approach samples the uncertainty space around the initial conditions and enables the estimation of the full probability distribution of possible future atmospheric states.
Improvements in ensemble methods – alongside advances in model physics, data assimilation, and computational capabilities – have extended the forecast skill horizon beyond the previously assumed two-week limit (Lorenz, 1968). These probabilistic approaches build the foundation of modern operational forecasting systems such as those at the European Centre for Medium-Range Weather Forecasts (ECMWF).

2.5.2. Ensemble forecasts

The representation of the atmospheric state to initialise a NWP model inevitably differs from the true atmospheric state. These small discrepancies – referred to as initial condition errors – amplify with increasing lead time due to the chaotic nature of the atmosphere. Consequently, forecasts deviate progressively from the actual atmospheric evolution. Ensemble forecasting offers a practical approach to account for these uncertainties in the initial conditions, as well as for the inherent limitations of deterministic modelling and the analytic intractability of fully detailed stochastic dynamic equations (Wilks, 2011b).
In an ensemble system, a finite number of plausible initial conditions are sampled from a probability distribution that characterises the uncertainty in the initial conditions (illustrated in Figure 2.7). Each member of the ensemble (represented by a circle) constitutes a valid realisation of the atmospheric state (represented by the ellipse). The NWP model is then integrated forward independently for each member (indicated by arrows), producing a range of forecast outcomes.
At the time of initialisation, ensemble members are tightly clustered, indicating the small magnitude of initial uncertainty. As the forecast progresses, the spread among members increases, reflecting the growing uncertainty in the system. This ensemble spread provides an approximation of how the initial probability distribution evolves under the system's dynamics, thereby offering insight into the forecast uncertainty. The rate and structure of this error growth are sensitive to the prevailing flow regime (Leutbecher and Palmer, 2007).
Forecast errors arise not only from initial condition errors but also from model deficiencies, such as limited resolution, imperfect physical parameterisations, and inaccuracies in boundary conditions. These deficiencies are accounted for by introducing stochastic perturbations or using alternative parameterisation schemes throughout the forecast integration (ECMWF, 2018).
As shown in Figure 2.7, forecast trajectories may diverge substantially over time. At longer lead times, the ensemble may form clusters of similar outcomes, revealing distinct scenarios. The increasing dis-

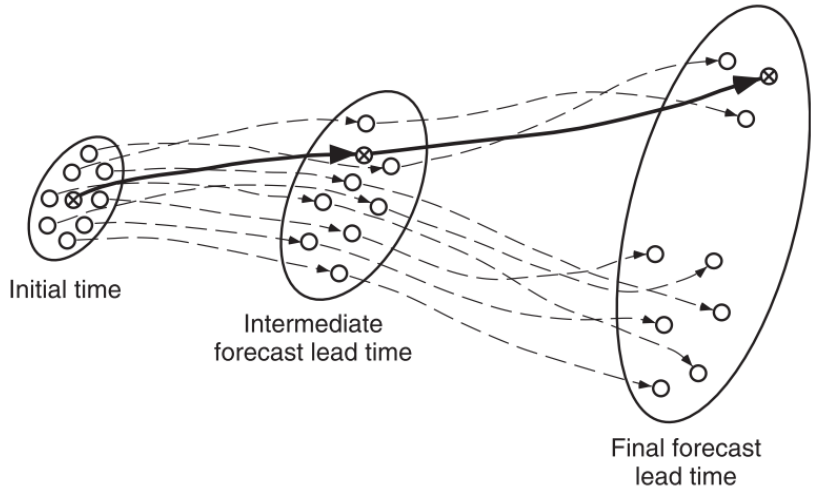


Figure 2.7.: A schematic illustration of ensemble forecasting in an idealised two-dimensional phase space. Each ensemble member is represented as a circle, with the best estimate of the initial condition marked by an "X." Ellipses indicate the probabilistic distribution at different times. Arrows represent the forecast trajectories of the ensemble members. Adopted from Wilks (2011b), Figure 7.24.

persion of ensemble members – illustrated by the widening ellipses – serves as a proxy for forecast uncertainty from both initial condition errors and model deficiencies. Such probabilistic information is absent from a single deterministic forecast, which reflects only one realisation (solid arrow) of a much broader range of potential outcomes.

Each ensemble member requires a full rerun of the dynamical model, making ensemble forecasting computationally expensive. Operational centres must balance trade-offs between ensemble size, model resolution, and lead time. They also apply different strategies to perturb initial conditions and represent model uncertainty (see Chapter 4.2 for further details).

2.6. Subseasonal to seasonal timescale and forecast skill

2.6.1. Definition and importance of subseasonal to seasonal time scale

Traditionally, weather and climate prediction have been treated as distinct disciplines. Weather forecasting typically targets daily atmospheric variability with lead times up to approximately 10 days, while climate prediction addresses long-term, time-aggregated statistics of weather conditions over periods ranging from several weeks to decades (Figure 2.8). Despite this historical divide, both rely on similar numerical approaches (Robertson et al., 2020). In recent decades, the recognition that atmospheric processes occur across a continuum of spatial and temporal scales has driven a convergence between weather and climate prediction. This has led to the emergence of a new forecasting range: the subseasonal to seasonal timescale, spanning roughly two weeks to a season ahead. This interval was once considered a "predictability desert" due to the lack of skilful operational forecasts between medium-range (3–10 days)

and seasonal (monthly to three-month average) predictions (e.g., Franzke and O'Kane, 2017; Robertson et al., 2020).

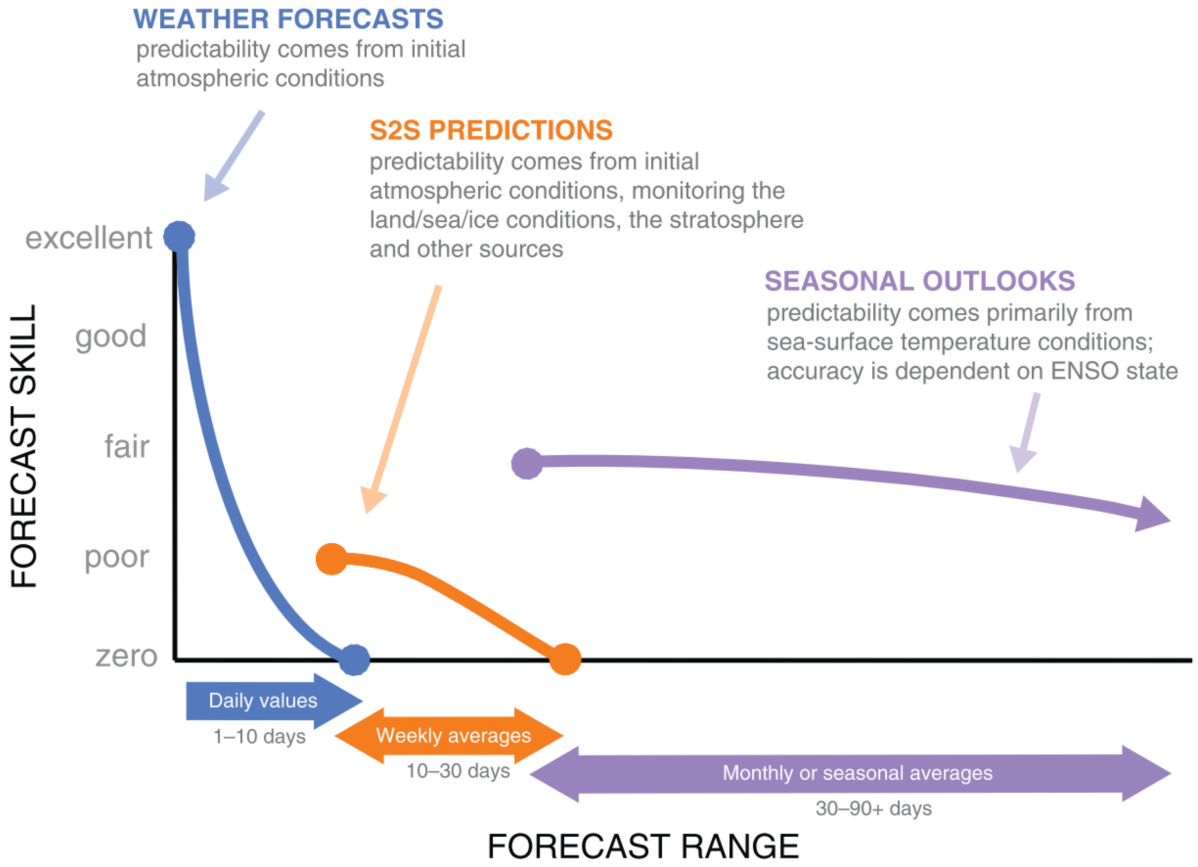


Figure 2.8.: Qualitative estimate of forecast skill across different forecast ranges, from short-term weather forecasts to long-range seasonal predictions. Relative skill levels are based on varying forecast averaging periods. Potential sources of predictability are indicated for each timescale. Adopted from White et al. (2017), based on earlier work by Tony Barnston and Elisabeth Gawthrop (both International Research Institute for Climate and Society).

The S2S time scale represents a unique forecasting challenge. It is generally too long for initial atmospheric conditions to retain much memory, as they do in short-term (1–3 days) and medium-range weather forecasts, yet too short for the more slowly evolving boundary conditions (such as sea surface temperatures, soil moisture, snow cover, or sea ice) that underpin seasonal predictions to exert their full influence (Robertson et al., 2020).

Until the 1980s and 1990s, there was little evidence that forecasts within the S2S range could be skilful. However, substantial advances over the past few decades have transformed this view and interest in S2S forecasting grows. Vitart and Robertson (2019) attribute these developments to four key factors: (1) identification of sources of S2S predictability associated with atmospheric, oceanic, and land-surface processes; (2) improved skill in medium-range forecasts; (3) the development of seamless prediction systems; and (4) increased demand from end users.

Several sources of predictability on the S2S timescale have now been identified (Robertson et al., 2020). These include:

- The Madden-Julian Oscillation
- Soil moisture, which retains memory on timescales of weeks and influences the atmosphere via evapotranspiration and surface energy fluxes
- Snow cover, whose radiative and thermal properties can modulate surface temperature and circulation
- Stratosphere-troposphere interactions, such as signals from the stratospheric polar vortex that can propagate downward and influence tropospheric circulation for up to two months
- Ocean conditions, particularly SST anomalies, which affect atmospheric convection and circulation patterns through changes in air-sea heat fluxes

A second crucial element of S2S prediction, next to the existence of predictable phenomena on the S2S time scales like the MJO, is temporal aggregation. While forecasting individual weather events beyond one to two weeks remains inherently limited by atmospheric chaos, aggregated weather statistics (e.g., weekly means) may still be predictable (Robertson et al., 2020).

Ensemble forecasts are central to S2S prediction. One of their key diagnostic outputs is the ensemble spread, which serves as an estimate of forecast uncertainty (Scherrer et al., 2004). A well-calibrated ensemble forecast exhibits a robust spread-error relationship: the ensemble spread (i.e., variance across members) should statistically reflect the magnitude of forecast errors relative to observations. In other words, larger ensemble spread typically corresponds to lower forecast accuracy, and vice versa (Hopson, 2014).

Forecast skill in NWP systems has steadily improved over the past four decades, with roughly one additional day of skill gained per decade (Figure 2.9). This trend has been described as a “quiet revolution” in weather prediction (Bauer et al., 2015). Key contributors to these advancements include international initiatives such as the Subseasonal to Seasonal Prediction Project (Vitart et al., 2017), launched in 2014 under the joint coordination of the World Weather Research Programme (WWRP) and the World Climate Research Programme (WCRP).

Despite these gains, significant challenges remain. Scientific gaps include limited understanding of land-surface processes, ensemble generation techniques, stratospheric influences, and the mechanisms of systematic errors in the models and error growth. Operational gaps include the need for more accessible and faster data access, multi-model calibrated products, and more extensive reforecast datasets with longer time spans and larger ensemble sizes for verification purposes. For end-users, barriers include low forecast skill for decision-making, lack of post-processing to ensure reliable probabilistic forecast outputs, and difficulties in interpreting forecast probabilities (Robertson et al., 2020).

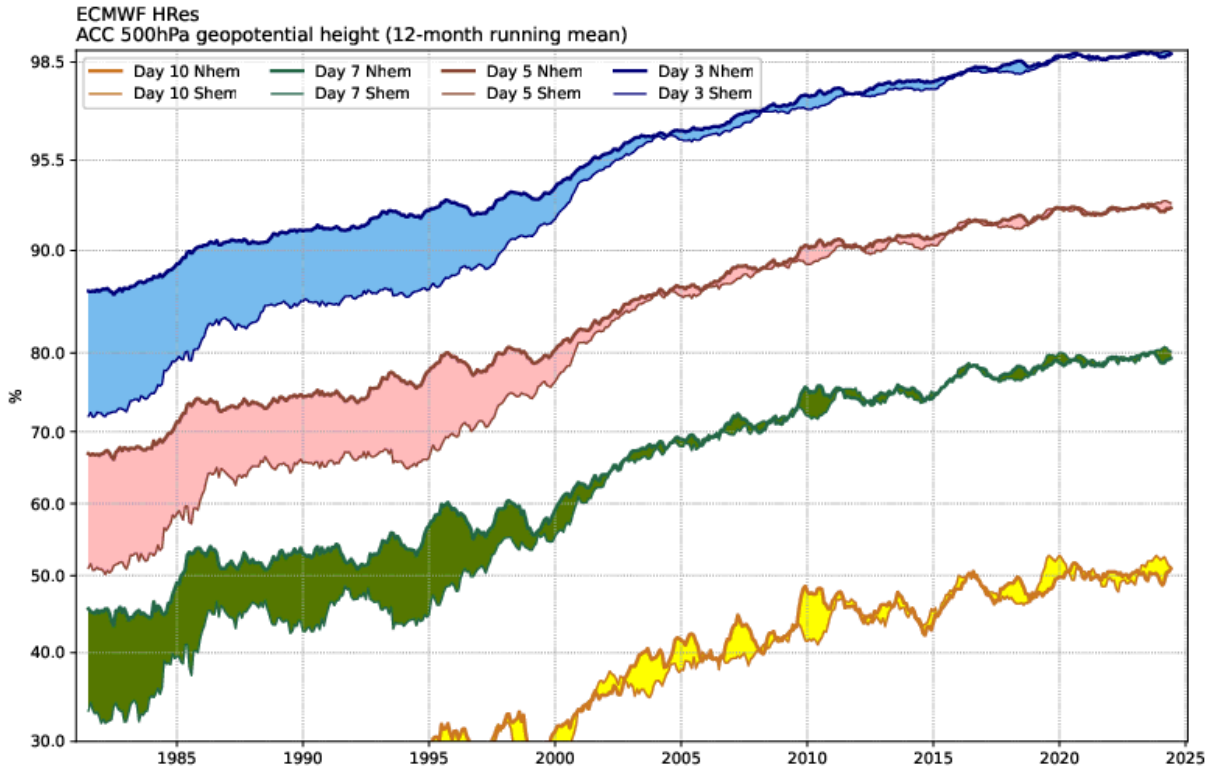


Figure 2.9.: Forecast skill of the ECMWF high-resolution model in predicting 500 hPa geopotential height fields at various lead times from 1980 to 2025. Shown are anomaly correlation coefficients (ACC, in %) for forecasts at days 3 (blue), 5 (red), 7 (green), and 10 (yellow) for the Northern Hemisphere (thick lines) and Southern Hemisphere (thin lines). Higher ACC values indicate greater similarity to verifying analyses. Adopted from ECMWF (`https://charts.ecmwf.int/`).

Despite these hurdles, the societal value of skilful S2S forecasts is substantial. 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); Enhanced drought forecasting can help mitigate agricultural losses (crop shortages) and economic damage (Salient et al., 2023); In the energy sector, long-range forecasts of anomalous weather conditions can improve preparedness for peak demand periods (Bloomfield et al., 2020b; Millin et al., 2024).

2.6.2. Flow dependent predictability of weather regimes

A number of studies have evaluated the ability of forecasting models to represent atmospheric modes of variability and their associated teleconnection patterns (introduced in Section 2.4) affecting the NAE region. In general, the extension of the forecast skill horizon over recent decades can be attributed to significant improvements in the NWP models. These include the transition from deterministic to probabilistic forecasting, improved representation of physical processes, and the implementation of advanced data assimilation techniques (Buizza and Leutbecher, 2015).

The ECMWF model has demonstrated skill in forecasting the amplitude of the MJO up to about 20 days (as of 2010, Vitart et al. (2010)). By 2017 and reaffirmed in 2022, various forecasting models, including ECMWF, exhibited skill in predicting the MJO at lead times of 2 to 4 weeks (Vitart, 2017; Stan et al., 2022). The enhanced skill in representing the MJO and teleconnections to Europe in recent years is likely attributable to improvements in model physics and for the teleconnections due to a better simulation of the MJO itself (Vitart, 2013). However, forecast skill for the MJO amplitude deteriorates due to the model's tendency to simulate the eastward propagation of the MJO too slowly. It particularly struggles in representing the passage across the Maritime Continent accurately (Vitart et al., 2010; Vitart, 2017). Moreover, the QBO modulates the MJO forecast skill. During winter, forecasts initiated under EQBO conditions exhibit up to 10 days longer skilful forecasts compared to those under WQBO (Wang et al., 2019; Lim et al., 2019). These results are insensitive to the initial MJO amplitude but rather due to the greater persistence of active MJO events during EQBO winters (Lim et al., 2019).

Forecasting models tend to overpredict the occurrence of NAO– and Atlantic Ridge, while underpredicting the frequency of NAO+. This bias is primarily driven by overly frequent transitions into NAO– from other regimes, rather than an overestimation of NAO– persistence (based on a four-regime wintertime classification) (Matsueda and Palmer, 2018). Forecast skill is generally highest for initialisations within NAO–, particularly during prolonged NAO– episodes (Matsueda and Palmer, 2018). This is confirmed by (Büeler et al., 2021; Osman et al., 2023), as they also demonstrate high forecast skill for Greenland Blocking (which is closely linked to NAO–) throughout the year, with a notable peak during winter. In contrast, the European Blocking regime remains challenging to predict, particularly when forecasts are initialised from NAO+ or Atlantic Ridge conditions (Matsueda and Palmer, 2018). Both Büeler et al. (2021) and Osman et al. (2023) report lower skill for European and Scandinavian Blocking, with lead times typically 3–5 days shorter than for other regimes.

While forecasting models adequately simulate the general MJO-induced precipitation patterns in the tropics, they tend to underestimate the amplitude of associated teleconnections to Europe when compared to reanalysis data (Vitart et al., 2010; Vitart, 2017). This discrepancy is partly due to the slow propagation of the MJO in the forecasts. Additionally, sea surface temperature fronts in the Gulf Stream region influence the eddy-driven jet. Systematic biases in the representation of these SST features in coupled ocean-atmosphere S2S models may adversely affect the simulation of MJO teleconnections in the NAE sector.

The ECMWF model captures several MJO-induced modulations of European weather regimes. These include an increased likelihood of NAO+ following MJO phase 3 (enhanced convection over the eastern Indian Ocean), reduced NAO+ activity after phase 6 (suppressed convection over the same region), and increased NAO– occurrence following MJO phase 7 (enhanced convection over the western Pacific) (Vitart et al., 2010; Vitart, 2017). In the absence of an active MJO, there is almost no forecast skill for probabilistic forecasts over Europe, whereas the presence of an active MJO substantially enhances forecast reliability, particularly at lead times between 19 and 25 days (Vitart et al., 2010).

Ferranti et al. (2018) highlight an asymmetric connection of the MJO on forecasts of the NAO phases. The skill of NAO– forecasts, which are often associated with cold spells in Europe, tends to be higher when the MJO is active at forecast initialisation and lower when the MJO is inactive at forecast initialisation. In contrast, NAO+ forecasts exhibit only weak sensitivity to the initial MJO activity.
In contrast to its influence on MJO forecast skill, the QBO modulates MJO teleconnections to Europe in the opposite manner. Forecast skill for 500 hPa geopotential height fields increases following WQBO conditions, especially during active MJO phases, although some improvement is also evident when the MJO is inactive (Mayer and Barnes, 2020). Most models simulate enhanced NAO+ activity following MJO phase 3 under WQBO conditions. However, they often fail to reproduce the corresponding enhancement of NAO– following MJO phase 7 under WQBO (Feng and Lin, 2021). During EQBO periods, some models – including ECMWF – simulate NAO– following MJO phase 7, but generally do not capture enhanced NAO+ after MJO phase 3 (Feng and Lin, 2021).
Roberts et al. (2023) report that the ECMWF model fails to capture the modulation of MJO-NAO teleconnections by ENSO. They identify two contributing factors: first, an insufficient propagation of Rossby wave activity from the Pacific to the Atlantic following MJO phase 3 during El Niño conditions, when the direct tropospheric teleconnection pathway is dominant; and second, an underrepresentation of the stratospheric polar vortex response following MJO phase 8 during La Niña conditions, when the indirect stratospheric pathway plays a central role.
On average, the forecast skill horizon for weather regimes in winter – not focusing on a specific weather regime but rather the mean over all – is extended during strong SPV states and reduced during weak SPV conditions (Büeler et al., 2021).
Despite robust evidence of teleconnections between tropical variability and European weather regimes, current numerical weather prediction systems still struggle to fully exploit these potential sources of subseasonal predictability. A likely explanation lies in the complex, multi-scale interactions among modes of subseasonal variability and the chaotic growth of forecast error (Roberts et al., 2023). In addition to model limitations, it is important to recognise that the intrinsic predictability of the atmosphere varies depending on the prevailing large-scale atmospheric state (e.g. Lorenz, 1995; Ferranti et al., 2015). Certain configurations, such as a strong MJO amplitude and specific phases or a well-established SPV anomaly, can enhance the potential for skilful forecasts. Thus, understanding when and under which conditions the atmosphere is more predictable is key to advancing subseasonal forecasting.

2.7. Post-processing and machine learning

Ensemble forecasts ideally represent uncertainty by the spread/dispersion of their members. However, in practice, ensemble members do not fully sample the uncertainty in initial conditions (Wilks, 2011b), and additional sources of error – such as inaccurate boundary conditions and structural model deficiencies – further degrade forecast quality (Vannitsem et al., 2021). These issues often result in systematic biases

and misrepresented dispersion, making statistical post-processing an essential step to enhance the reliability of ensemble predictions, meaning the predicted probabilities should accurately reflect the observed frequencies (Gneiting, 2014; Vannitsem et al., 2021). Post-processing methods are routinely used by operational weather forecasting centres worldwide (Wilks, 2011b). Early approaches, such as Perfect Prog and Model Output Statistics, relied on linear regression techniques to link observations to model outputs (Vannitsem et al., 2021). Today, a broader set of methods – including logistic regressions, neural networks, and other machine learning algorithms – are employed to improve forecast skill.

Post-processing typically addresses three core goals: (1) the forecast should be as close as possible to the truth, (2) the forecast should be climatologically reliable, and (3) forecasts and observations should be statistically indistinguishable (Vannitsem et al., 2021). Most methods fall into two broad categories: parametric approaches, which assume a specific probability distribution and model its parameters based on NWP-derived predictors, and non-parametric approaches, which avoid distributional assumptions and estimate forecast distributions directly from data (Vannitsem et al., 2021). While parametric methods are interpretable and easier to implement, they can be inflexible. Non-parametric methods – such as quantile regression (e.g. Taillardat et al., 2016), analogue techniques (e.g. Hamill and Whitaker, 2006), or machine learning techniques (e.g. Rasp and Lerch, 2018) – offer greater flexibility at the cost of increased computational complexity.

In recent years, machine learning (ML) has emerged as a powerful tool in post-processing, particularly for handling large, high-dimensional NWP datasets. Deep learning models are capable of capturing complex spatial and physical relationships in input data. While training such models is computationally intensive, their operational use is relatively lightweight compared to running full NWP ensembles (Haupt et al., 2021). One key challenge is interpretability – ML models are often seen as "black boxes." However, techniques like variable importance measures and Shapley values offer insights into model behaviour and individual predictions (McGovern et al., 2019; Vannitsem et al., 2021). ML also allows for customised post-processing by optimising task-specific loss functions, enabling forecasts to be tailored to the needs of end-users and specific decision-making contexts (Haupt et al., 2021).

3. Research questions

As discussed in Chapter 1, the growing share of renewable energy, particularly in Germany, makes energy generation increasingly weather-dependent, leading to greater volatility (Bloomfield et al., 2016; Wiel et al., 2019a). This variability poses significant challenges for grid stability and energy markets, particularly during Dunkelflauten. Consequently, there is growing interest in the energy sector in predicting both the weather and its impacts on the energy system, across time scales ranging from hours to days, weeks, and seasons.

While advancing NWP models is a long-term goal, it remains computationally expensive and time consuming. In the meantime, statistical post-processing offers a cost-effective alternative to enhance probabilistic forecast skill by correcting model biases and improving ensemble spread. However, this approach critically relies on the ability of the NWP forecast to provide at least some predictive information, which Büeler et al. (2021) showed that it tends to be lost on average at roughly two weeks for daily weather regime forecasts. One way to address this limitation is by using statistical knowledge and the atmospheric state at forecast initialisation to identify windows of opportunity – periods of enhanced (subseasonal) forecast skill.

The overarching aim of this thesis is to develop more useful forecast information tailored to the needs of the energy sector. This chapter outlines the five research questions (RQ), which are also summarised schematically in Figure 3.1.

Dunkelflauten pose a major risk to energy security, especially in winter when demand is high. Weather regimes offer a promising framework for analysing Dunkelflauten because they capture large-scale atmospheric patterns that correspond well with the spatially integrated nature of energy systems (e.g., national electricity supply and demand). Drücke et al. (2021) identified a high-pressure pattern over Germany as the most frequently associated atmospheric condition during Dunkelflauten in Germany. However, this does not necessarily imply that this pattern is responsible for the most critical energy system impacts. On a broader European scale, Wiel et al. (2019b) showed that anticyclonic weather regimes are linked to high energy shortfall events. Despite these insights, the specific connection between NAE weather regimes and the occurrence of Dunkelflauten in Germany remains insufficiently explored. Therefore, the main research question in Chapter 5 is:

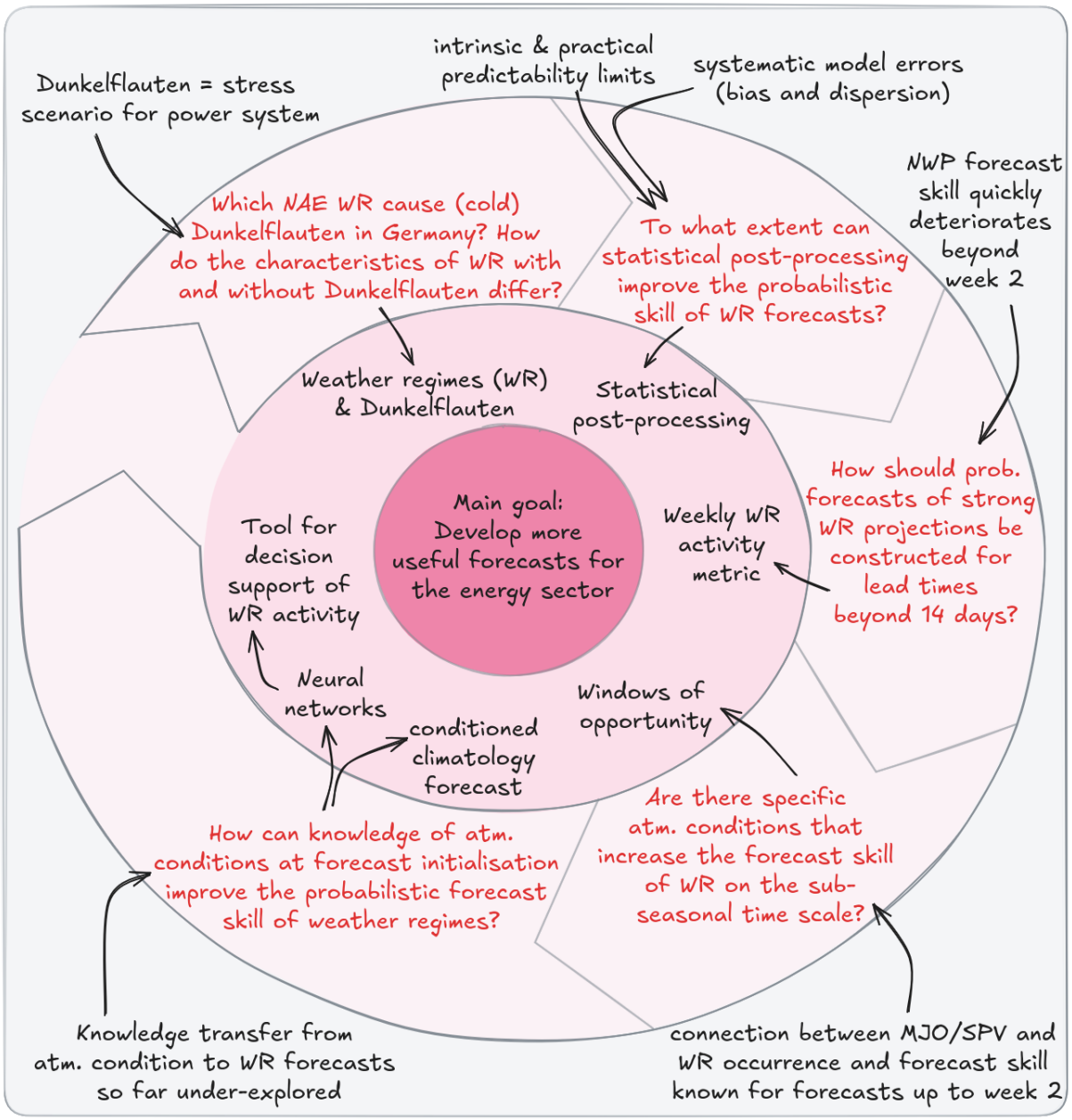


Figure 3.1.: Schematic overview of the red thread of this PhD thesis. The overarching goal – to develop more useful forecasts for the energy sector – is placed at the core. Moving clockwise, the diagram illustrates the progression from research questions (in red, outermost ring), motivated by knowledge from literature (in black, outer box), to the corresponding topics of research (in black, inner ring).

RQ-1 Which North Atlantic-European weather regimes cause Dunkelflauten in Germany, especially cold Dunkelflauten? How do the characteristics of these regimes differ from those not associated with Dunkelflauten?

While previous studies identified general atmospheric patterns linked to Dunkelflauten, this is the first to analyse their timing within the weather regime life cycle and how these life cycles differ between regimes with and without Dunkelflauten. This provides new insights into predicting if and when a Dunkelflaute is likely to occur.

To exploit the link between weather regimes and Dunkelflauten, forecasts of weather regimes must be both skilful and statistically reliable – meaning their predicted probabilities should accurately reflect the observed frequencies. However, raw ensemble forecasts in the subseasonal range suffer from systematic biases and dispersion errors. While statistical post-processing methods are well-established for surface variables such as sea-level pressure (Schefzik et al., 2013), wind speed (Scheuerer and Hamill, 2015), and temperature (Schefzik, 2017), their application to regime forecasting is pending, hence the main research question in Chapter 6 is:

RQ-2 To what extent can statistical post-processing improve the probabilistic skill of ECMWF reforecast-based weather regime forecasts, particularly when accounting for systematic biases, dispersion errors and multivariate dependencies?

In this chapter, we apply the EMOS-ECC approach to statistically post-process probabilistic forecasts of the weather regime index. For the first time, we evaluate these probabilistic forecasts not only in a univariate framework using the continuous ranked probability score, but also in a multivariate setting, based on the seven dimensional weather regime index vector as a whole, using the energy and variogram score.

Probabilistic forecast skill for daily regime index forecasts typically vanishes beyond two weeks (Büeler et al., 2021; Osman et al., 2023), limiting their practical value on subseasonal time scales. This motivates the development of a more robust metric for capturing longer-term regime tendencies:

RQ-3 How should probabilistic forecasts of strong weather regime projections be best constructed for lead times beyond 14 days?

We introduce a new metric, described in Chapter 4, which is used for the remaining research questions. This metric combines the detailed but difficult to interpret seven dimensional weather regime index and a more easily interpretable but strongly simplified categorical regime classification.

In the medium-range, several studies have shown that weather regime forecast skill is modulated by the atmospheric state at initialisation (e.g., Vitart et al., 2010; Feng and Lin, 2021; Roberts et al., 2023). However, similar analyses for the subseasonal range remain scarce. This leads to the first research question in Chapter 7:

RQ-4 Are there specific atmospheric conditions that increase the forecast skill of weather regimes on the subseasonal time scale?

In this chapter, we focus on Greenland Blocking forecasts in week 3 and examine how forecast skill of the ECMWF NWP model and regime occurrence are modulated by the Madden-Julian Oscillation and stratospheric polar vortex. The findings are summarised in a categorisation of windows of opportunity.

While identifying windows of opportunity provides insight into potential forecast skill, end users often lack the tools or expertise to translate this information into actionable decisions (Robertson et al., 2020). Communicating both regime forecasts and skill expectations remains a challenge. To address this gap, we investigate in Chapter 7 the following question:

RQ-5 How can knowledge of atmospheric conditions at forecast initialisation improve the probabilistic forecast skill of weather regimes?

We develop atmospheric-conditioned climatology forecasts for the newly introduced weather regime metric and combine them with NWP outputs using a statistical-dynamical approach by training neural networks. These models predict the likelihood of strong weather regime projections in week 3, making the forecast more user-friendly and informative. The goal is to support decision-making by integrating statistical knowledge and dynamical predictions into a single, interpretable product.

Before addressing the research questions in Chapters 5–7, Chapter 4 provides an overview of the data sources and methodologies used throughout the thesis.

4. Data & methods

This chapter provides an overview of the data sources and methodological approaches employed throughout this thesis. A range of meteorological and energy-related datasets form the foundation for the analyses. These include ECMWF reanalysis data (Section 4.1), which serve as a proxy for the atmospheric "truth", and the ECMWF reforecast dataset (Section 4.2). Electricity demand data (Section 4.3) for selected regions is incorporated to explore the link between meteorological variability and electricity consumption patterns.

A central element of the thesis is the use of weather regimes (Section 4.4), which are employed to describe the large-scale atmospheric circulation using both continuous and categorical indices. Next to the weather regime framework, several large-scale atmospheric modes of variability are introduced (Section 4.5), such as the Madden-Julian Oscillation, the Stratospheric Polar Vortex, the Antarctic Oscillation, and others. These modes are described by established climate indices and play a critical role in characterising the background state of the atmosphere.

To verify forecast performance, the forecasts are rigorously evaluated using a combination of discrete and continuous error metrics, verification rank histograms, and statistical significance testing (Section 4.6). This enables a comprehensive assessment of forecast quality and provides a solid basis for evaluating the added value of post-processing techniques and neural networks which are applied to the raw ensemble forecasts (Section 4.7). These include Ensemble Model Output Statistics, Ensemble Copula Coupling, and neural networks trained to combine NWP outputs with information about the atmospheric state at initialisation time.

4.1. Reanalysis

Reanalysis data from ECMWF, specifically the ECMWF Reanalysis v5 (ERA5), is used as our observational reference (Hersbach et al., 2020). ERA5 serves as the basis for computing weather regimes (using 500 hPa geopotential height), generating composites of atmospheric field variables during Dunkelflauten (including 100 m wind speed, surface net solar radiation, surface net solar radiation for clear skies, 2 m temperature, and sea level pressure), and calculating capacity factors (based on 100 m wind speed, surface roughness, direct, diffuse, and top-of-atmosphere solar radiation, and albedo).

In Chapter 5, we use ERA5 data interpolated from its native reduced-Gaussian grid to a regular latitude-longitude grid (0.5° spacing) with a three-hour temporal resolution for weather regime calculations (see Section 4.4) and meteorological composites. For computing capacity factors and Dunkelflauten, we use

the original ERA5 grid spacing of 0.25° with hourly data, processed via the "atlite" library (see Sections 5.1.1 and 5.1.2).
In Chapters 6 and 7, we remap the ERA5 dataset to match the grid spacing of the reforecast dataset (introduced in Section 4.2), which is a 1° horizontal grid-spacing and daily temporal resolution. For verification, we treat ERA5 as a "perfect ensemble member" by aligning it with each forecast initialisation date and lead time (cf. Wandel et al., 2024).

4.2. ECMWF reforecast

In Chapters 6 and 7, we use subseasonal to seasonal reforecast data from ECMWF, obtained from the Subseasonal-to-Seasonal Prediction Project Database (Vitart et al., 2017). To increase the number of forecast initialisation dates available for analysis, we merge reforecasts from two consecutive ECMWF model cycles: Cy46R1 and Cy47R1 (Vitart and Mladek, 2023). These reforecasts are computed twice a week (Mondays and Thursdays), with 11 ensemble members, covering lead times from 0 to 46 days. The model runs on 91 vertical levels with a native horizontal resolution of 16 km (days 0–15) and 32 km (day 15 onward). All forecast data are remapped to a regular 1° latitude-longitude grid.
The two model cycles were operational between 11 June 2019, and 11 May 2021, with the cycle transition occurring on 30 June 2020. During the period between 11 May and 30 June, only Cy46R1 reforecasts are available, which could impact the training and evaluation of our methods due to fewer training dates. As a result, post-processing performance may be lower during this period.
Nonetheless, by combining Cy46R1 and Cy47R1, we obtain a 21-year dataset of reforecasts, spanning 11 June 1999 to 11 May 2020, with a total of 4000 initialisation dates, each containing 11 ensemble members. These reforecasts are initialised using ERA5 reanalysis data.
The ECMWF ensemble forecast includes one control forecast and 10 perturbed members that start from slightly modified initial conditions. These perturbations are generated using singular vectors, which identify patterns in the atmosphere where small errors grow fastest over a 48-hour period, based on total energy (Leutbecher and Palmer, 2007). Singular vectors are found by iteratively evolving a small perturbation forward with a tangent linear model, measuring how much it grows, then using an adjoint model to trace this growth back and adjust the initial guess. This process is repeated to find the five fastest-growing, mutually independent (orthogonal) singular vectors. Each is added and subtracted from the control to create 10 perturbed members. To represent model uncertainty, ECMWF also uses Stochastically Perturbed Parametrisation Tendencies, which adds spatio-temporally correlated random noise to the model's physical tendencies, mimicking the effects of unresolved sub-grid-scale processes (e.g., convection or cloud processes) and improving ensemble spread and realism (Palmer et al., 2009).

4.3. Electricity demand

To investigate weather-induced variability in electricity demand, we also incorporate daily mean electricity demand data from Bloomfield et al. (2020a) into our analysis. The model builds on 2 m temperature data from the ERA5 reanalysis and uses a multiple-linear regression model for 28 European countries from 1979 to 2018. The model focuses on weather-dependent parameters, therefore neglecting human behavioural factors (day-of-week and long-term socio-economic trends). The demand time series we use can be interpreted as the demand which would have been expected on each weather day in 1979–2018, with no day-of-week effects and the prevailing socio-economic conditions of 2017. This data reflects the temperature dependence of the electricity demand in 2017. We use the daily demand time series to compute the average daily demand anomalies during Dunkelflauten, compared to daily climatology. Due to electrification in the space heating sector in Germany (BDEW, 2022), it can be expected that the electricity demand will become more temperature-dependent in future scenarios, so results with the historical demand model are likely to underestimate the significance of the weather-dependent demand. Data is limited to daily, weather-dependent demand, excluding the diurnal cycle of demand (Bloomfield et al., 2020a).

4.4. Weather regimes

In this thesis, we use the seven year-round NAE weather regimes introduced by Grams et al. (2017) based on ERA-Interim reanalysis but adapted here for the newer ERA5 reanalysis as described in Hauser et al. (2023a,b) and applied to IFS reforecasts following the approach of Büeler et al. (2021) and Osman et al. (2023). These weather regimes represent the most common large-scale circulation patterns in the NAE region (30–90°N, 80°W–40°E).

4.4.1. Weather regime patterns

To construct the climatological mean patterns of the weather regimes, we compute 6 hourly 500 hPa geopotential height anomalies (Z500A) for the period 1979–2019. Anomalies are calculated relative to a 91-day running mean climatology centred on each calendar date ($\pm$45 days), providing a smooth representation of the seasonal cycle. To reduce short-term variability and highlight persistent, large-scale patterns, we apply a 10-day low-pass filter to the anomaly fields.

A key step in the regime definition is the seasonal normalisation of the Z500A, which accounts for the inherently stronger variability in winter compared to summer. This is achieved by dividing the filtered anomaly fields at each grid point by a calendar-day-dependent normalisation factor. The factor is computed as the spatial average (across the NAE domain) of the 31-day running standard deviation of the Z500A for the corresponding calendar day. This normalisation ensures that anomalies from different

seasons are comparable in magnitude, allowing for a consistent year-round classification of weather regimes.

We then apply an empirical orthogonal function (EOF) analysis to the filtered and seasonally normalised Z500A to identify the leading modes of variability. The first seven EOFs are retained, and a k-means clustering (with $k = 7$) is applied. The resulting seven clusters represent the seven distinct weather regimes (Figure 2.2) originally introduced by Grams et al. (2017), with three cyclonic (Atlantic Trough, Zonal Regime, and Scandinavian Trough) and four anticyclonic regime types (Atlantic Ridge, European Blocking, Scandinavian Blocking, and Greenland Blocking).

4.4.2. Continuous weather regime index

To describe the projection of instantaneous anomalies onto mean regime patterns, whether in reanalysis or (re-)forecast, we introduce a seven-dimensional weather regime index (IWR). We outline the steps to compute the weather regime index (see also Figure B.1) for two sets of forecasts, raw and Z500 bias-corrected (notation adopted by Osman et al. (2023)).

Similar to the definition of the weather regime patterns, we compute the Z500A for raw forecasts relative to the 91-day running mean ERA5 climatology (1979–2019, Figure B.1, top middle panel). Subsequently, we then apply a low-pass filter and seasonal normalisation to obtain standardised and filtered Z500A (Φ_{m^*}), where m^* denotes all 11 ensemble members and the ERA5 perfect member.

Directly computing Z500A by subtracting ERA5 climatology from model (re)forecasts does not account for systematic model biases. To address this, systematic Z500 forecast biases are eliminated in the Z500 bias-corrected forecasts (abbreviated as Z500 cor., Figure B.1, top right panel). Here, the underlying Z500 calendar day climatology derived from ERA5 reanalysis is replaced with a model climatology (similar to Büeler et al., 2021). To ensure a fair comparison later on, the model climatology covers June 1999 to May 2015, excluding the subsequent test period of June 2015 to May 2020. This ensures that model data from the test period in the post-processing results (Chapter 6) is not included. For the perfect member, the ERA5 climatology of this reduced period is used rather than model climatology. Due to the different reference periods in climatology (1979–2019 vs. 1999–2015), the ERA5 perfect members of the raw and Z500 bias-corrected forecasts are not identical (though differences are minimal).

After computing the standardised and filtered Z500A (Φ_{m^*}) for each forecast, we project these onto the seven cluster mean Z500A following the method of Michel and Rivière (2011) via

$$P_{(wr,m^*)}(t,\tau) = \frac{1}{\sum_{\lambda,\varphi\in(region)}\cos(\varphi)} \sum_{\lambda,\varphi\in(region)} \left(\Phi_{m^*}(\lambda,\varphi,t,\tau)\cdot\Phi_{(wr)}(\lambda,\varphi)\cdot\cos(\varphi)\right). \tag{4.1}$$

Here, $P_{(wr,m^*)}(t,\tau)$ represents a scalar measure for the spatial correlation of the instantaneous anomaly field $\Phi_{m^*}(\lambda,\varphi,t,\tau)$ at lead time day τ initialised at date t, and ensemble member m^*, with the cluster mean anomaly field $\Phi_{(wr)}(\lambda,\varphi)$ for the weather regimes $wr \in \{AT, ZO, ScTr, AR, EuBL, ScBL, GL\}$. Here, λ and φ denote the longitudinal and latitudinal degrees, respectively.

The weather regime index $I_{(wr,m^*)}(t,\tau)$ is then computed for each weather regime, ensemble member, initialisation date and lead time based on anomalies of the projections $P_{(wr,m^*)}(t,\tau)$. These anomalies are relative to the climatological mean projection $\overline{P_{(wr)}} = \frac{1}{N}\sum_{i=1}^{N} P_{(wr)}(i)$ and the estimated climatological standard deviation of the projection,

$$I_{(wr,m^*)}(t,\tau) = \frac{P_{(wr,m^*)}(t,\tau) - \overline{P_{(wr)}}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_{(wr)}(i) - \overline{P_{(wr)}}\right)^2}}. \tag{4.2}$$

The variable i in the climatological mean projection and estimated standard deviation has different meanings for the two different forecast sets. For the raw forecasts, i iterates over the ERA5 data from 1979 to 2019. For the Z500 bias-corrected forecasts, i iterates over all available dates and ensemble members in the model forecast data from June 1999 to May 2015 (the reduced training set). Combining the weather regime index for each weather regime yields a 7-dimensional weather regime index vector $IWR_{m^*}(t,\tau)$ for each initialisation date, lead time, and ensemble member.

For verifying IWR forecasts, respective ERA5 perfect members are considered. When computing skill scores (e.g., the continuous ranked probability skill score), we use a climatological reference forecast, computed by the perfect member of the raw IWR forecast (Figure B.1, bottom right panel).

4.4.3. Categorical weather regime index

To facilitate the analysis of large-scale atmospheric patterns in the reanalysis data, we derive a categorical weather regime index (WR_{cat}) from the continuous seven-dimensional IWR. A weather regime is considered active when its IWR exceeds 1.0 for at least five consecutive days and includes the maximum IWR value among all regimes during this period. The onset of a regime is defined as the first time step where IWR>1.0 (cf. Grams et al., 2017; Büeler et al., 2021). Time steps that do not meet these criteria are classified as "no regime," representing periods of transition or weak regime projections. This classification provides a discrete representation of regime life cycles, simplifying the interpretation of large-scale atmospheric variability in the reanalysis data.

4.4.4. Weather regime activity

In subseasonal forecasting (beyond 10 days), achieving skill at a daily resolution becomes increasingly challenging. At the same time, decision-makers such as energy traders and transmission system operators often do not require daily details but are more interested in the probability of a weather regime occurring within a given forecast week. To address this, we introduce a metric called 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 $\text{IWR}_{min} = 1$, otherwise $WRact_{mean} = 0$. Applied to the ensemble 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.

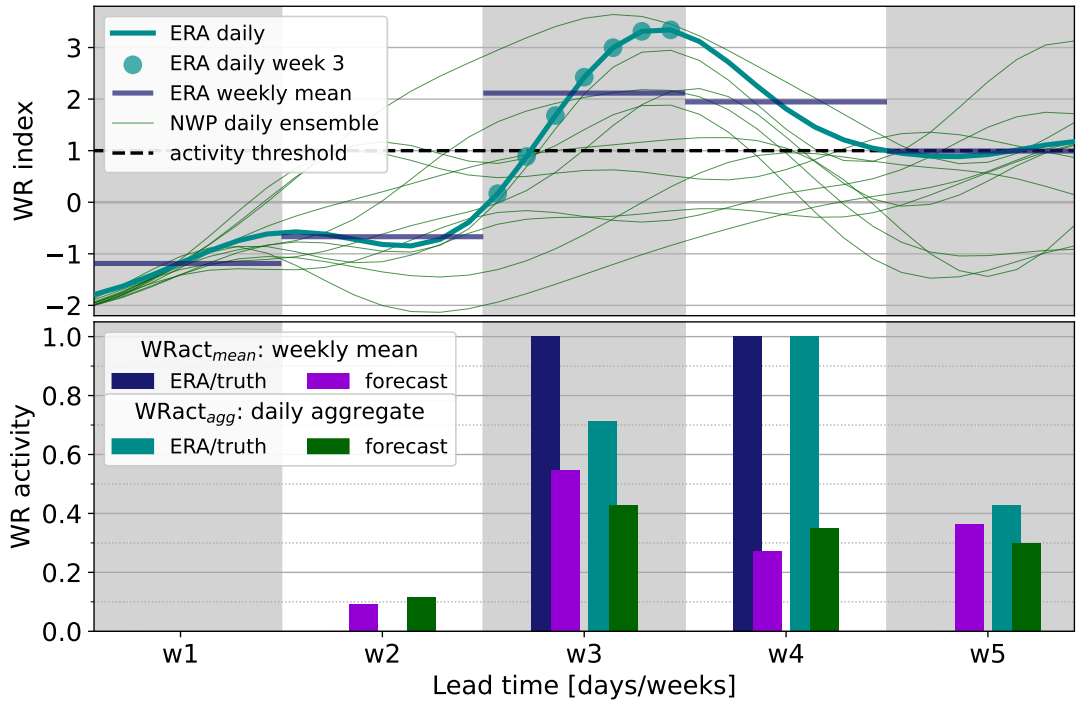


Figure 4.1.: Visualisation of $WRact_{mean}$ and $WRact_{agg}$ by using the weather regime index. The upper panel shows the ensemble forecast (green lines) and the perfect member (ERA5, turquoise line) of the daily IWR. Further, the dark blue horizontal lines indicate the mean weekly IWR in ERA5 and 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 4.1 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 lines). 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 4.1). 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.

4.5. Atmospheric modes of variability

While the weather regimes described above provide a useful framework for categorising large-scale atmospheric circulation patterns in the NAE region, they are influenced by broader atmospheric modes of variability. These modes capture large-scale oscillations and teleconnections operating across a range of spatial and temporal scales, and they serve as valuable predictors for the occurrence and persistence of weather regimes.
This section briefly introduces the calculation methods for these atmospheric modes, with an overview of the associated data sources presented in Table C.1. The primary application of these modes is in Chapter 7, where they are used to identify windows of opportunity, create climatological forecasts based on atmospheric conditions, and serve as predictors for neural networks.

Madden-Julian Oscillation The Madden-Julian Oscillation (MJO) is a key mode of sub-seasonal atmospheric variability that modulates tropical precipitation patterns (NOAA, 2025c). In contrast to the commonly used Real-time Multivariate MJO (RMM) index (Wheeler and Hendon, 2004), this thesis employs the Outgoing Longwave Radiation (OLR) MJO Index (OMI) (NOAA, 2025a), which is derived solely from OLR anomalies and does not include zonal winds at 850 hPa or 200 hPa. As a result, the OMI provides a smoother representation of MJO evolution.

The OMI is based on a pair of EOFs computed from near-equatorial (15°N–15°S) satellite-derived OLR data. By projecting forecast or analysis data onto these EOFs – after removing the annual cycle and inter-annual variability – two principal components, OMI1 and OMI2, are obtained. These components define the MJO's position in a two-dimensional phase space, where each of eight segments corresponds to a distinct phase of MJO activity. Typical regions of enhanced convection are associated with consecutive phase pairs: phases 2–3 correspond to the western and eastern Indian Ocean, phases 4–5 to the Maritime Continent (west and east of Oceania), phases 6–7 to the western and central Pacific, and phases 8–1 to the western Hemisphere and Africa. The amplitude ($\sqrt{OMI1^2 + OMI2^2}$) quantifies MJO strength, with values greater than one indicating an active MJO.

Stratospheric Polar Vortex The strength of the stratospheric polar vortex is assessed via the zonal-mean zonal wind at 100 hPa and 60°N. To account for seasonality, the SPV is categorised as weak, neutral, or strong based on daily climatological percentiles (33rd and 66th) derived from a 91-day moving window over the period 1979–2019. Days below the 33rd percentile are classified as weak SPV, those above the 66th percentile as strong SPV, and intermediate values as neutral.

Quasi-Biennial Oscillation The Quasi-Biennial Oscillation index used here is defined as the 3-month running mean of the zonal-mean zonal wind anomaly at 50 hPa over the tropics (10°S–10°N), following the approach of Yoo and Son (2016). Based on its standardised anomaly, the index is classified into easterly, westerly, or neutral phases using thresholds of ± 0.5 standard deviations (approximately $\pm 5 \frac{\mathrm{m}}{\mathrm{s}}$) (Lim et al., 2019; Mayer and Barnes, 2020).

El Niño Southern Oscillation Among the various El Niño Southern Oscillation indices, this work uses the Oceanic Niño Index (ONI), the operational definition adopted by NOAA (Trenberth, 2025). The ONI is based on SST anomalies relative to the 1971–2000 climatology, averaged over 5°N–5°S and 170°W–120°W. A 3-month running mean is applied, and El Niño (La Niña) conditions are defined when anomalies exceed +0.5°C (–0.5°C) for five consecutive months. The ONI, originally available as 3-month means, is linearly interpolated to daily values.

North Atlantic Oscillation The North Atlantic Oscillation is characterised by a dipole in sea level pressure between two quasi-permanent pressure systems, the Azores High and the Icelandic Low. The index used here is a two-point pressure difference, calculated as the standardised sea level pressure between Ponta Delgada (Azores) and Reykjavik (Iceland), following Wanner et al. (2001).

Arctic Oscillation The Arctic Oscillation is derived as the leading EOF of the monthly 1000 hPa geopotential height field north of 20°N. The AO index is computed by projecting daily anomalies of 1000 hPa geopotential height onto this loading pattern.

Antarctic Oscillation Analogous to the AO, the Antarctic Oscillation (AAO) is defined via the leading EOF of the monthly 700 hPa geopotential height field south of 20°S. Daily anomalies are projected onto the resulting loading pattern to compute the AAO index.

Pacific-North American Pattern The Pacific-North American pattern (PNA) index is derived by projecting daily 500 hPa geopotential height anomalies over the Northern Hemisphere (0–90°N) onto the second leading mode of a rotated EOF analysis of monthly mean 500 hPa geopotential height fields from 1950–2000 (NOAA, 2025b).

Sea surface temperature anomaly As an additional predictor for the neural networks, daily mean sea surface temperatures anomalies over the North Atlantic are used. These are computed by latitudinally weighting daily mean SST values within 80°W–10°W and 0°N–80°N. A 31-day running mean climatology (1991–2020) is subtracted to obtain anomalies.

4.6. Forecast verification

Forecast verification assesses the quality of forecasts by quantifying their relationship with corresponding observations (Wilks, 2011a). It plays a crucial role in comparing forecasts across different geographical regions, forecasting models, and lead times. Verification also helps evaluate a forecast's performance against baseline forecasts – typically climatological or persistence forecasts – to identify strengths and weaknesses within a forecasting system.
Since forecasts and observations vary in nature – ranging from deterministic to probabilistic forecasts, discrete to continuous predictands, and univariate to multivariate verification – a variety of verification metrics exist to evaluate different forecast types. This section introduces the verification metrics and methods applied in this thesis.

4.6.1. Discrete error metrics from contingency tables

A common example of discrete error metrics is the forecasting of precipitation days. The observed outcome is binary: either it rained (event observed) or it did not (event not observed). When forecasts are deterministic and discrete, the four possible forecast-observation pairs (for two possible outcomes) can be represented in a traditional contingency table (Table 4.1a). Various metrics can be derived from this table to evaluate forecast performance. The outcomes in a traditional contingency table are: Hit (True Positive, TP), the event is forecasted and occurs; Miss (False Negative, FN), the event occurs but was not forecasted; False Alarm (False Positive, FP), the event is forecasted but does not occur; and Correct Rejection (True Negative, TN), the event is not forecasted and does not occur.
Key metrics derived from contingency tables (Table 4.2) include: Base Rate, representing the overall occurrence of the event; Forecast Rate, the proportion of forecasts predicting the event; Hit Rate (also

Forecast	Event observed / not observed	
	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	Event observed / not observed	
	Yes	No
Yes	$E_{11} = \mathbf{p} \cdot \mathbf{k}$	$E_{12} = \mathbf{p} \cdot (\mathbf{1} - \mathbf{k})$
No	$E_{21} = (\mathbf{1} - \mathbf{p}) \cdot \mathbf{k}$	$E_{22} = (\mathbf{1} - \mathbf{p}) \cdot (\mathbf{1} - \mathbf{k})$

(b) Probabilistic contingency table.

Table 4.1.: General classification terms of (a) the traditional 2×2 contingency table and (b) probabilistic equations for the occurrence of an event. **p** represents a vector of probabilities for an event to occur and **k** is a vector of binary indicators whether the events have occurred at the given dates. 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.

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 4.2.: Verification metrics derived from the contingency table (Table 4.1) to evaluate discrete forecasts.

known as Recall or Accuracy), the proportion of actual events correctly predicted; False Alarm Rate, the proportion of incorrect predictions of the event; Peirce Skill Score, the difference between the hit rate and the false alarm rate; Precision, the proportion of true positive forecasts out of all positive predictions; and F1-Score, the harmonic mean of precision and recall, balancing both metrics.

An alternative to the traditional contingency table is the probabilistic contingency table (Table 4.1b), which incorporates probabilistic forecasts (Gold et al., 2020). Here, $(p_1, p_2, ..., p_n)$ represents the forecast probabilities of an event (e.g. precipitation day) occurring over n dates. Each p_i is the mean of the binary predictions from all ensemble members for date i, calculated as:

$$p_i = \frac{\sum_{m=1}^{M}}{M} p_{i,m}, \tag{4.3}$$

where M is the number of ensemble members (in this study $M = 11$). Each ensemble member $p_{i,m}$ provides a binary prediction (0 or 1), and the actual event outcomes $(k_1, k_2, ..., k_n)$ are represented by binary values, where $k_i = 1$ if the event occurs and $k_i = 0$ otherwise.

By combining forecast probabilities and observed binary outcomes, probabilistic contingency tables provide a framework analogous to traditional contingency tables, allowing for a more nuanced evaluation of probabilistic forecasts.

The metrics from this section are utilised in Chapter 7 to evaluate forecasts, considering both windows of opportunity and neural network predictions.

4.6.2. Continuous error metrics

Extending the example of precipitation forecasting, a related question is not just whether it will rain, but how much precipitation will occur. In this case, the predictand is a continuous variable. While precipitation is limited to non-negative values – introducing certain constraints – other continuous variables, such as temperature or the weather regime index, do not share this limitation.

For deterministic forecasts, the absolute error (AE) measures the absolute difference between the forecasted (f) and observed (o) value for a given time (i): $\mathrm{AE}_i = |o_i - f_i|$. The mean absolute error (MAE) measures the average absolute error over all forecasts: $\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{AE}_i$, where N is the number of forecasts. A related metric, the mean squared error (MSE) penalises larger errors more heavily by squaring the differences: $\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{AE}_i^2$. This metric is particularly sensitive to outliers, as large deviations contribute disproportionately to the total error.

For the evaluation of probabilistic forecasts – such as those derived from ensemble predictions – for continuous predictands, we introduce the continuous ranked probability score (CRPS) for univariate evaluation and the energy and variogram score (ES and VS, respectively) for assessing multivariate aspects of forecast skill. While these metrics are extensively discussed in Gneiting and Raftery (2007) and Scheuerer and Hamill (2015), we provide a concise summary here and refer interested readers to the mentioned literature for further details.

The CRPS measures the accuracy of a univariate probabilistic forecast and generalises the absolute error, to which it reduces in the case of a deterministic forecast. It is defined as:

$$CRPS(F, y) = \int_{-\infty}^{\infty} \left(F(x) - \mathbb{1}\{x \geq y\}\right)^2 \mathrm{d}x, \tag{4.4}$$

where $\mathbb{1}$ is the indicator function, F represents the predictive distribution, and y is the observed value. Lower CRPS values represent higher forecast skill, with a score of 0 representing a perfect forecast.

For multivariate probabilistic forecasts, we employ the Energy Score (ES) and Variogram Score (VS) of order p. The ES generalises the CRPS to multiple dimensions, while the VS originates from the concept of variograms (also known as structure functions) in geostatistics.

The ES is given by:

$$ES(F,\mathbf{y}) = \frac{1}{M}\sum_{i=1}^{M}\|\mathbf{X}_i-\mathbf{y}\| - \frac{1}{2M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\left\|\mathbf{X}_i-\mathbf{X}_j\right\|, \tag{4.5}$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^D$, $\mathbf{X}_i, \mathbf{X}_j$ are samples from the multivariate forecast distribution, and $\mathbf{y}$ is the observed multivariate outcome.

The VS^p is defined as:

$$VS^p(F,\mathbf{y}) = \sum_{i=1}^{D}\sum_{j=1}^{D} w_{i,j}\left(\left|y^{(i)}-y^{(j)}\right|^p - \frac{1}{M}\sum_{k=1}^{M}\left|X_k^{(i)}-X_k^{(j)}\right|^p\right)^2, \tag{4.6}$$

where $w_{i,j}$ is a non-negative weight for each component pair, and p representing the order of the VS. In accordance with Scheuerer and Hamill (2015), we use an unweighted version ($w_{i,j}=1$ with $p=0.5$). Both, the ES and VS follow the same interpretation as the CRPS, where lower values indicate better forecast performance.

To compare a forecast's skill against a reference forecast (e.g., climatology), we introduce skill scores (SS_f), which are computed as:

$$SS_f = \frac{\overline{S_{ref}}-\overline{S_f}}{\overline{S_{ref}}-\overline{S_{opt}}} = 1-\frac{\overline{S_f}}{\overline{S_{ref}}}, \tag{4.7}$$

where $\overline{S_f}$ is the mean score of a given forecast method f, and $\overline{S_{ref}}$ is the mean score of the reference forecast. Since the metrics used here (MSE, CRPS, ES, VS) all have a perfect score (S_{opt}) of 0, the equation simplifies to a skill score of 1 for a perfect forecast, 0 for a forecast equal to the reference, and negative values for forecasts worse than the reference.

The MSE serves as both the loss function and evaluation metric for neural networks in Chapter 7, while CRPS, ES, and VS are used to assess post-processed forecasts in Chapter 6.

4.6.3. Verification rank histograms

Verification rank histograms are widely used to assess the calibration of ensemble forecasts for a scalar predictand, which in this case is the IWR (Wilks, 2011a). To construct a rank histogram, we analyse an ensemble forecast consisting of $M = 11$ members across N forecast instances. For each instance, the observation is ranked among the $M+1$ values (the 11 ensemble members plus the observation itself) after sorting them in ascending order. The frequency of each rank is then recorded, forming a histogram that represents the distribution of observation ranks across all forecasts. A uniform rank distribution indicates a well-calibrated forecast, meaning the ensemble properly captures the observation uncertainty. In contrast, a U-shaped histogram suggests that the forecast is underdispersed, meaning the ensemble spread is too narrow, frequently placing the observation outside the ensemble range. Conversely, a bell-shaped histogram indicates overdispersion, where the ensemble spread is too large. A skewed histogram reveals systematic biases in the forecast. To provide a comprehensive visualisation, rank histograms for all lead times are presented in a joint 2D histogram. The x-axis represents the observation rank, the

y-axis shows lead times, and colour-coded cells indicate deviations from the expected uniform frequency of 1/12 per rank, given that there are $M+1=12$ possible ranks. The exact deviations are also displayed as numbers within the cells, offering a clear representation of forecast calibration across different lead times.

The verification rank histograms evaluate the raw, Z500 bias-corrected, and post-processed IWR forecasts in Chapter 6.

4.6.4. Diebold-Mariano test of equal performance

To assess whether the difference in predictive performance between two forecasts is statistically significant, we apply the Diebold-Mariano test of equal performance (Diebold and Mariano, 1995). This test evaluates whether the mean scores of two competing forecasts, denoted as F and G, differ in a statistically meaningful way. The test statistic is given by:

$$t_n = \sqrt{n}\frac{\overline{S_n^F} - \overline{S_n^G}}{\hat{\sigma}_n}, \tag{4.8}$$

where $\overline{S_n^F} = \frac{1}{n}\sum_{i=1}^{n} S(F_i, \mathbf{y}_i)$ and $\overline{S_n^G} = \frac{1}{n}\sum_{i=1}^{n} S(G_i, \mathbf{y}_i)$ represent the mean scores of forecasts F and G over N samples. The standard deviation of the score differentials is estimated as:
$\hat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(S(F_i, \mathbf{y}_i) - S(G_i, \mathbf{y}_i)\right)^2}$.

Following Gneiting and Katzfuss (2014) and assuming independence between score differentials, the test statistic t_n asymptotically follows a standard normal distribution. A negative value of t_n suggests that forecast F outperforms G based on the chosen score S. We apply a significance level of $\alpha = 0.05$, meaning that differences exceeding this threshold are considered statistically significant.

The Diebold-Mariano tests assess significant differences among the raw, Z500 bias-corrected, post-processed, and climatological IWR forecasts in Chapter 6.

4.7. Statistical post-processing techniques and neural networks

We employ a two-step statistical post-processing method (Chapter 6) based on Sklar's theorem (Sklar, 1959; Schefzik et al., 2013; Lerch et al., 2020) to improve the probabilistic forecast skill of the seven dimensional weather regime index forecast. The first step involves univariate processing using ensemble model output statistics (EMOS (Gneiting and Raftery, 2007)), followed by multivariate post-processing through ensemble copula coupling (ECC (Schefzik et al., 2013)). We adapt the notation from Lerch et al. (2020) and Chen et al. (2024) with modifications tailored to our specific setup.

Sklar's theorem states that a multivariate cumulative distribution function H can be decomposed into a copula function C – which captures the dependence structure – and its marginal univariate cumulative distribution functions $F_{AT}, ..., F_{GL}$, obtained via univariate post-processing. Specifically, for weather regime indices $x_{AT}, ..., x_{GL} \in \mathbb{R}$, the decomposition is given by: $H(x_{AT}, ..., x_{GL}) = C(F_{AT}(x_{AT}), ..., F_{GL}(x_{GL}))$

(Lerch et al., 2020). x represents the weather regime index and the subscript iterates through the 7 weather regimes $wr \in \{AT, ZO, ScTr, AR, EuBL, ScBL, GL\}$.
The raw seven-dimensional ensemble forecast consists of $M = 11$ ensemble members, denoted as $\mathbf{X}_1, ..., \mathbf{X}_M \in \mathbb{R}^7$, where each member is given by $\mathbf{X}_m = \left(X_m^{(AT)}, ..., X_m^{(GL)}\right)$. Similarly, the observations are represented as $\mathbf{y} = \left(y^{(AT)}, ..., y^{(GL)}\right) \in \mathbb{R}^7$.

4.7.1. Ensemble model output statistics

The first step of the two-step post-processing pipeline applies univariate EMOS to fit a Gaussian predictive distribution $\mathcal{N}^{(wr)}$ with mean μ and variance σ^2:

$$y^{(wr)} | X_1^{(wr)}, ..., X_M^{(wr)} \sim \mathcal{N}(\mu, \sigma^2) = F_\theta^{(wr)}, \tag{4.9}$$

where the distribution parameters $\theta = (\mu, \sigma)$ are linked to the ensemble forecasts via:

$$(\mu, \sigma^2) = (a_0 + a_1 \overline{X}, b_0 + b_1 S^2) = g\left(X_1^{(wr)}, ..., X_M^{(wr)}\right). \tag{4.10}$$

These parameters are determined by minimising the CRPS (introduced in Section 4.6) via optimisation of

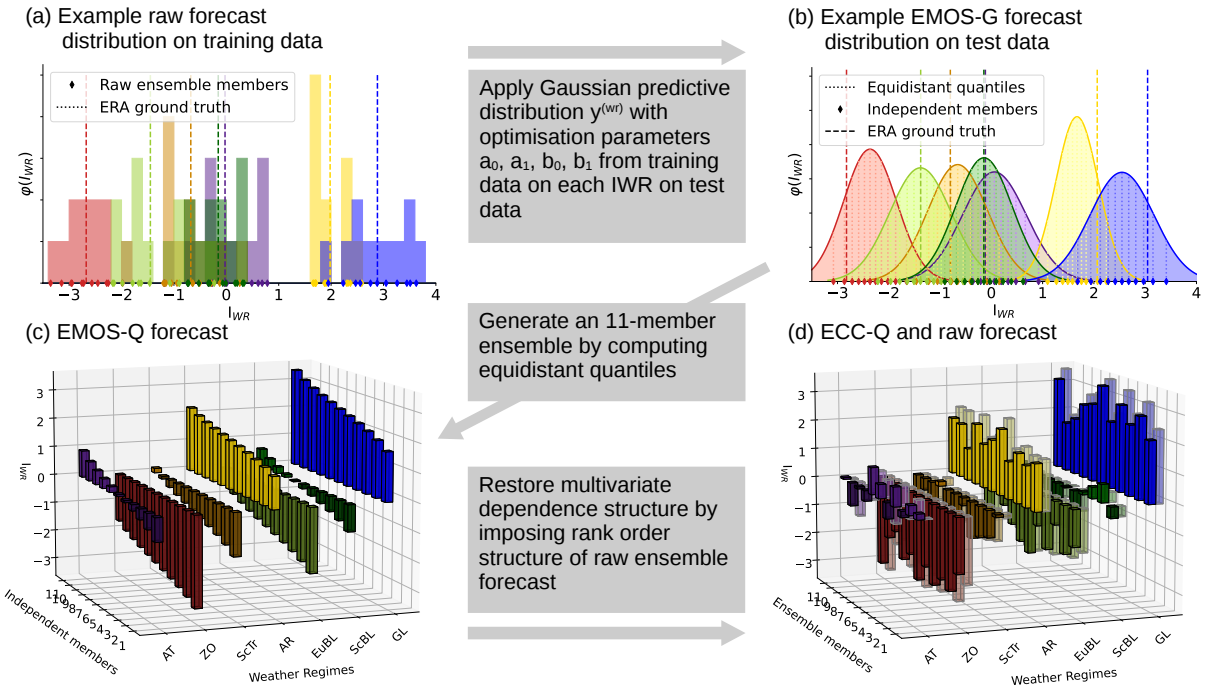


Figure 4.2.: Post-processing workflow for the seven dimensional weather regime index using EMOS-ECC. (a) Initially, the optimisation parameters are fitted by minimising the continuous ranked probability score on the raw ensemble forecasts in the training data. The obtained optimising parameters from the training data are then applied on ensemble forecasts in the test data to transform them into Gaussian predictive distributions (a to b). (b) The resulting EMOS-G forecast is used to sample equidistant quantiles, generating (c) the post-processed ensemble forecast EMOS-Q, which is organised by the rank. Finally, (d) ECC-Q is employed to restore the multivariate dependency structure of the raw forecast (transparent bars in (d)).

a_0, a_1, b_0, b_1 over a training period from June 1999 to May 2015 (Figure 4.2a, b). The resulting model for each weather regime index is referred to as EMOS-G. To evaluate the sensitivity of EMOS performance to different training setups, we tested several configurations. One approach trained a single EMOS model on the full dataset, while another divided the data into two seasonal models, one for the winter half-year (October–March) and one for the summer half-year (April–September). A third configuration used four seasonal models, distinguishing between winter (December–February), spring (March–May), summer (June–August), and autumn (September–November). Additionally, we explored daily models trained using running windows of 9, 31, and 91 days. The differences in performance across these setups were small, with the four-season model and the 31-day running window showing slightly improved results. As the 31-day window is also compatible with on-the-fly generated reforecast data, we focus our analysis on this configuration.

After univariate post-processing, the continuous EMOS-G Gaussian distributions need to be converted into an ensemble with the same number of members ($M = 11$) as the raw forecast. This is done by extracting equidistant quantiles at levels $\frac{1}{M+1}, ..., \frac{M}{M+1}$ from the forecast distributions $F_\theta^{(wr)}$ (Figure 4.2b). The resulting post-processed ensemble forecasts are referred to as EMOS-Q (Figure 4.2c).

4.7.2. Ensemble copula coupling

In the second step, we utilise ECC to restore the multivariate dependency structure of the weather regime index. This is done by retrieving the rank order structure of the unprocessed ensemble member forecasts (Figure 4.2d, pale bars) and sorting the post-processed ensemble members accordingly (Figure 4.2c to 4.2d). To formalise this process, for each weather regime index, we define $\sigma_{(wr)}(m) = \text{rank}\left(X_m^{(wr)}\right)$ as a permutation, and aim to find $\tilde{X}_m^{(wr)} = \hat{x}_{\sigma_{(wr)}(m)}^{(wr)}$, where $\hat{x}_1^{(wr)}, ..., \hat{x}_M^{(wr)}$ form a sample under the assumption that $\hat{x}_1^{(wr)} \leq ... \leq \hat{x}_M^{(wr)}$ to simplify the notation. Here, $\hat{x}_1^{(wr)} := \left(F_\theta^{(wr)}\right)^{-1}\left(\frac{1}{M+1}\right), ..., \hat{x}_M^{(wr)} := \left(F_\theta^{(wr)}\right)^{-1}\left(\frac{M}{M+1}\right)$ represent the quantile-based EMOS-Q predictions of the weather regime indices. This non-parametric, empirical copula approach is referred to as ECC-Q.

For clarity, and for consistency, we will refer to the execution of the two-step process as ECC, rather than EMOS-ECC, since no other univariate methods were tested. Besides ECC, we also tested the Schaake Shuffle approach, which ranks ensemble members based on past observations rather than the actual forecast.

4.7.3. Neural networks

An alternative to purely statistical and parametric approaches like ECC for improving weather regime forecasts is to integrate information from multiple sources, e.g. atmospheric-conditioned climatological forecasts, NWP model output, and recent values of aggregated weather regime activity, through non-parametric approaches like neural networks (NNs). We adopt a statistical-dynamical approach using fully connected NNs. These networks are designed to capture complex, non-linear relationships

among the predictors, with the goal of improving the forecast skill of $WRact_{agg}$, particularly for strong weather regime events at three-weeks lead time. 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 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 C.1 and C.2.

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 4.3.: 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 4.3). 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.

5. The role of weather regimes for Dunkelflauten detection in Germany

Dunkelflauten are extreme weather events with significant implications for the energy sector as only little solar and wind power is available during these periods. They pose the greatest challenge to the power system when they coincide with cold temperatures, which drive up electricity demand and, consequently, the residual load. Understanding the frequency, seasonality, and atmospheric drivers of Dunkelflauten in a specific country – in this case, Germany – is crucial for energy trading companies and transmission grid operators to anticipate and mitigate potential impacts.

To address these questions, we first define Dunkelflauten using capacity factors for wind and solar power output (Section 5.1). In Section 5.2, we provide a general overview of Dunkelflauten in Germany and their relationship with weather regimes. We then examine the meteorological origins of these events by analysing composites of key atmospheric variables (Section 5.3). In Section 5.4, we explore whether weather regime life cycles associated with Dunkelflauten differ from typical regime life cycles. Finally, in Section 5.5, we incorporate the electricity demand model of Bloomfield et al. (2020b) to assess how the meteorological conditions identified in Section 5.3 influence electricity demand during Dunkelflauten events.

This chapter is based on and slightly adapted from Mockert et al. (2023), with the working title "Meteorological conditions during periods of low wind speed and insolation in Germany: The role of weather regimes."

5.1. Definition and computation of Dunkelflauten

Dunkelflauten – periods of persistently low wind and solar power output – are identified using the capacity factor, which represents the ratio between actual renewable energy generation (solar, onshore, and offshore wind power) and the total installed capacity at a given time. Before the actual analysis of Dunkelflauten in Germany (starting in Section 5.2), we first introduce the concept of the capacity factor and outline the methodology used to compute Dunkelflauten. This foundational understanding is essential for interpreting the results of our study.

5.1.1. Capacity factors

The capacity factors are computed with the Python library "atlite", which converts weather data (e.g. wind speed and solar influx) into energy system data (e.g. capacity factors) (Hofmann et al., 2021).

To generate the capacity factor, atlite uses the ERA5 reanalysis dataset at 0.25° grid spacing and one hourly temporal resolution. For the wind capacity factor calculation, atlite uses the 100 m wind speed and the surface roughness as input data. For calculating the solar capacity factor, the direct, diffuse and top-of-the-atmosphere influx and the albedo are needed as input data.

Atlite assumes the geographic distribution of installed wind and solar capacities to be proportional to the technology's capacity factors (i.e. more wind farms in windy areas, see the Supplementary Material A.1). We use a solar panel model based on Huld et al. (2010) with a solar azimuth of 180° (South) and a slope of 35°. As a reference onshore wind turbine, we use the Vestas V112 3 MW model, and for offshore the 5 MW NREL Reference Turbine.

Based on a linear regression between simulated and historical capacity factor time series, the capacity factors are corrected to match the observed time series (OPSD, 2020) more closely (Supplementary Material A.2). The total power output by renewable energy sources is expressed by combining the three different capacity factor time series to one combined capacity factor by applying a weighted mean with currently installed capacities per technology as weights. For Germany in 2018, these are: 44% solar (45.9 GW), 50% onshore wind (53.0 GW), 6% offshore wind (6.4 GW) (IRENA, 2019).

The optimised approach of the geographic distribution of installed wind and solar capacities will likely lead to a too high estimate of the capacity factors, but it is less dependent on political decision-making. The effect of the too-high capacity factor is damped by scaling the capacity factors with the historical capacity factor time series of Germany given by OPSD (2020).

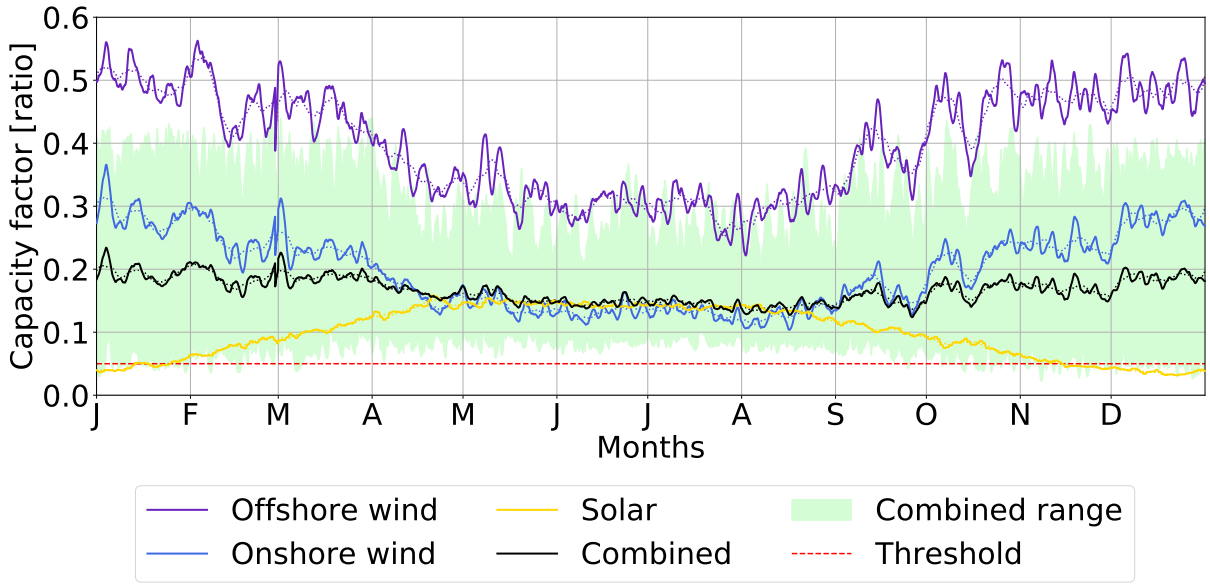


Figure 5.1.: Average seasonal cycle of the capacity factors in Germany for 1979–2018 (based on hourly time steps, 48 h running mean (solid lines) and weekly running mean (dotted lines)). The combination of onshore wind (blue), offshore wind (purple) and solar (yellow) capacity factors, weighted on the installed capacity, is represented by the combined capacity factor (black). The range (minimum, maximum) of the 48 h running mean combined capacity factor throughout the 40 year (1979–2018) period is visualised (mint green shading). The red horizontal line marks a threshold of 0.06.

For Germany, the individual capacity factors of wind and solar have a pronounced seasonal cycle (Figure 5.1). On average, the on- and offshore wind capacity factors are lower in summer than in winter, and the solar capacity factor is higher in summer than in winter. By combining the capacity factors of wind and solar, weighted by the installed capacity, the mean seasonal variation is balanced.
From here on, we refer to the combined capacity factor when writing capacity factor unless otherwise stated.

5.1.2. Dunkelflauten

We are interested in periods when only little or no energy is available by wind and solar power for at least 2 days. In this study, we define Dunkelflauten as periods with 48 h running mean capacity factors below a threshold of 0.06. We have defined the 48 h threshold according to the study of Kaspar et al. (2019) and since energy storage is more problematic for periods with low renewable feed-in longer than two days (Schmidt et al., 2019). The frequency and duration of Dunkelflauten events can be controlled by modifying the threshold value. Here, we set the threshold to 0.06 to obtain a similar number of Dunkelflauten per year as shown by Kaspar et al. (2019) (3–4 Dunkelflauten per year with a threshold of 0.1). Our results regarding the central time of occurrence and meteorological characteristics of Dunkelflauten hold under modification of the threshold value (e.g. 0.05 or 0.10), although we detect fewer and shorter events for smaller thresholds and more and longer events for higher thresholds.
All time steps contributing to a running mean below the threshold are considered part of the Dunkelflaute. Therefore, the minimum duration of a Dunkelflaute is 48 h.
We categorise Dunkelflauten according to the dominant weather regime in this period. In case several weather regimes contribute the same number of hours to a Dunkelflaute, the Dunkelflaute is associated with the weather regime that occurs at the onset of the Dunkelflaute since this is the weather regime that triggers the Dunkelflaute in the first place. With this definition, there are eight categories of Dunkelflauten: Dunkelflauten related to one of the seven weather regimes or to the “no regime.”
To capture the sequence of challenging renewable supply conditions, we consider years centred on winter, starting in July and ending in June of the following year. For simplicity, we refer to these periods as winter-centred years.

5.2. Characteristic of Dunkelflauten in Germany

By our definition, a Dunkelflaute is a rare event indicating a period of low combined wind and solar power output. In the 40 year period from 1979–2018, we detected 169 Dunkelflauten which cumulates to an average of four Dunkelflauten per year. These Dunkelflauten have a pronounced seasonal cycle (Figure 5.2) with the bulk of Dunkelflauten in autumn (September–November) and winter (December–February) and no Dunkelflaute in summer (July–August). The seasonality of Dunkelflauten is explainable by the seasonal cycle of the capacity factors (Figure 5.1).

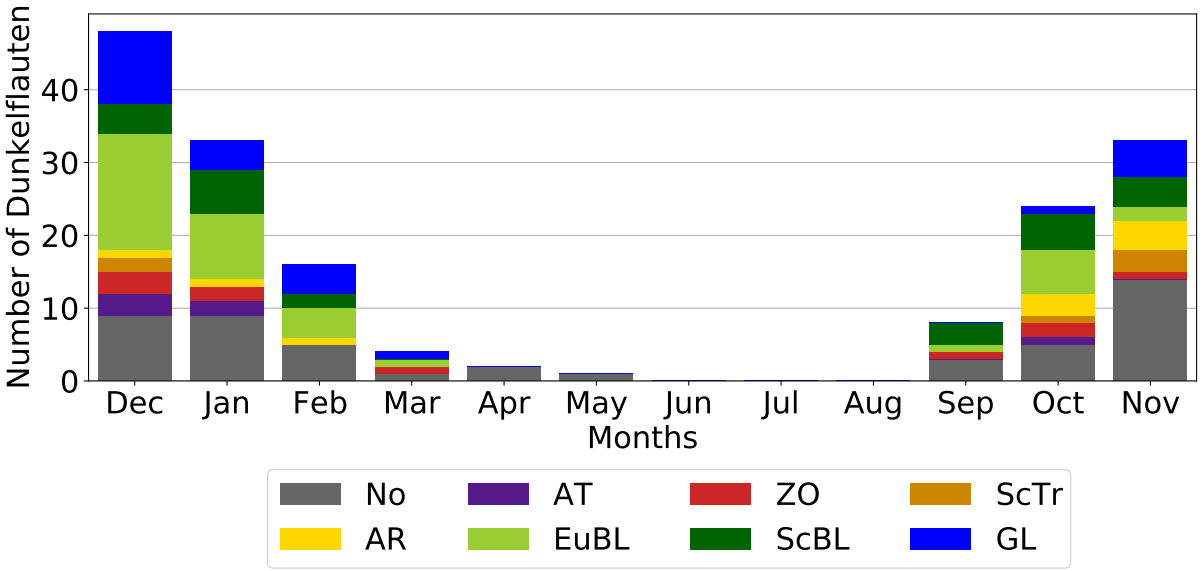


Figure 5.2.: Monthly distribution of Dunkelflauten with the dominant weather regimes of Dunkelflauten indicated by the colour of the bars. The amount of Dunkelflauten is measured on an absolute scale for the 40 year period from 1979–2018.

The distribution width of the combined capacity factor, using the 48 hour running mean in the 40 year period, is the smallest in summer, and the minimal value of the summer distribution (0.064) is above the Dunkelflauten threshold. In the winter half-year, the combined capacity factor distribution varies on a larger range, occasionally reaching values below the Dunkelflauten threshold. Values below the threshold are due to the combination of low 48 hour running mean solar capacity factors (average of 0.053 in winter time) and occasional drops in the wind capacity factors below the threshold. In winter, solar power cannot always compensate for the occasional lack of wind power. In summer, the solar capacity factor (average of 0.139) is well above the Dunkelflauten threshold and can thus compensate for periods of low wind capacity factors.

The frequency of Dunkelflauten does not only vary throughout the year but also interannually. On average, 4.3 Dunkelflauten occur per year (Supplementary Material A.3). The most Dunkelflauten per winter-centred year have occurred in 1995/1996 and 1996/1997 with 9 and 10 Dunkelflauten respectively. The median duration of a Dunkelflaute is 3.2 days (77.5 hours) with a maximum duration of 8 days (Supplementary Material A.4). Only 15 Dunkelflauten events (9%) exceed a duration of 5 days, which is the minimum duration of weather regimes. No Dunkelflaute reaches the mean duration of weather regimes, which is 11.3 days. The shortest possible duration of a Dunkelflaute is 2 days by definition. The duration of Dunkelflauten has a similar seasonal cycle as the amount of Dunkelflauten, with the longest Dunkelflauten in December. Supplementary Material A.5 provides a detailed visualisation of all Dunkelflauten over the years.

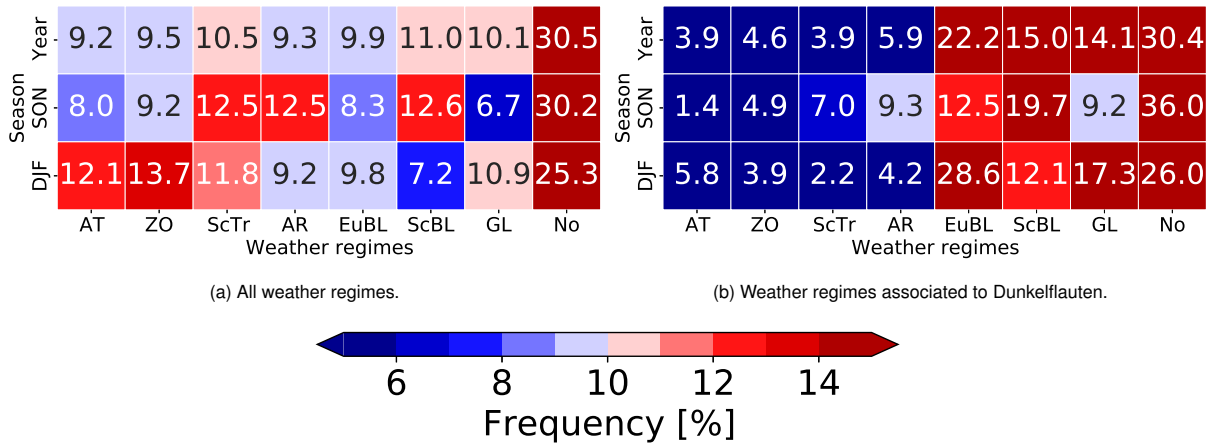


Figure 5.3.: Weather regime frequencies for (a) all weather regimes and (b) all weather regimes associated with a Dunkelflaute. Shown for all seasons of the year combined, only autumn (SON) and only winter (DJF). Data for the years 1979 to 2018.

With the help of weather regimes for the Northern Atlantic-European region, we can describe and forecast large-scale circulation and wind and solar irradiation patterns. The frequency distribution of weather regimes during Dunkelflauten differs from its climatological distribution (Figure 5.3).

The year-round climatological frequency distribution of weather regimes is well balanced throughout cyclonic (Atlantic Trough, Zonal Regime, Scandinavian Trough) and blocked weather regimes (Atlantic Ridge, European Blocking, Scandinavian Blocking, Greenland Blocking) with frequencies between 9.2% and 11.0% (Figure 5.3a, top row). However, for Dunkelflauten periods the distribution of the weather regimes is not in balance (Figure 5.3b, top row). Dunkelflauten occur predominantly in three of the four blocked weather regimes, European, Scandinavian and Greenland Blocking, among which Dunkelflauten occur most frequently during European Blocking (22.2%). As Dunkelflauten mainly occur in autumn and winter (Figure 5.2), we compare the frequency distributions for these seasons separately (Figure 5.3a and 5.3b, second and third row). The preference of Dunkelflauten occurrence during the three blocked weather regimes over the cyclonic weather regimes remains but differences in the preferred frequency of blocked regimes occur. In autumn, Scandinavian Blocking is by 7% more frequent during Dunkelflauten compared to all autumn days (12.6% vs 19.7%, Figure 5.3). In winter, the frequency of European and Greenland Blocking is increased by 19 and 6% (an increase from 9.8% up to 28.6% and from 10.9% up to 17.3%, respectively, Figure 5.3) during Dunkelflauten compared to the overall frequency in winter. Thus, Dunkelflauten preferentially occur during the blocked European, Scandinavian and Greenland Blocking regimes.

5.3. Meteorological parameters during Dunkelflauten

To further shed light on the meteorological conditions during Dunkelflauten and to estimate the severity of Dunkelflauten during different regimes in terms of surface weather, we analyse anomaly composites

(respective to the 30 day running mean climatology) of different atmospheric field variables (100 m wind speed, solar irradiation, 2 m temperature, sea level pressure) in the North-Atlantic-European region. The solar irradiation is given as the ratio of the daily sums of the surface net solar radiation (SSR in ERA5 reanalysis) and the surface net solar radiation for clear skies (SSRC in ERA5 reanalysis). This quantity reflects the fraction of the maximum possible daily insolation (cf. Grams et al., 2017).

The composites of all Dunkelflauten events (Figure 5.4a–c) represent the mean atmospheric conditions. Low wind speeds are associated with weak surface pressure gradients. Weak pressure gradients can occur in the centre of a high-pressure system or the saddle point between multiple high and low-pressure systems. On average, weak pressure gradients in the centre and northwestern edge of a surface anticyclone over Europe cause low wind speeds during Dunkelflauten in Germany (Figure 5.4a). Insolation is only marginally altered in comparison to the 30 day running mean climatology (Figure 5.4b). This suggests that Dunkelflauten are primarily due to a lack of wind power. On average, 2 m temperature anomalies of up to 2 K occur in Central Europe (Figure 5.4c).

Analysing Dunkelflauten separated by weather regimes demonstrates that the conditions for Dunkelflauten can be achieved by different atmospheric conditions (Figure 5.4d–l). Thus, the mean conditions for all events do not reflect variability in meteorological conditions during Dunkelflauten events imposed by different weather regimes.

The three dominant Dunkelflauten types, classified by weather regimes as European, Scandinavian and Greenland Blocking (shown in Figure 5.3), all indicate low wind speeds in the North Sea region due to weak pressure gradients, but the pressure system patterns differ between the Dunkelflauten types (Figure 5.4d,g,j). For the three dominant Dunkelflauten types, there are only weak modulations of irradiation (Figure 5.4e,h,k). Strong differences in the temperature anomalies amongst these Dunkelflauten types exist (Figure 5.4f,i,l).

During European Blocking Dunkelflauten low wind speeds in Northern Germany and the North Sea region occur in the centre of a high-pressure system centred over Germany (Figure 5.4d).

In contrast, low wind speeds for Scandinavian and Greenland Blocking Dunkelflauten occur in a region of weak pressure gradients due to a quadripole of two high and two low-pressure systems centred over Germany. For Scandinavian Blocking Dunkelflauten (Figure 5.4g), the low pressure dominates near Iceland and in the Mediterranean, high pressure prevails in the vicinity of the Azores and Scandinavia. The quadripole is reversed for Greenland Blocking Dunkelflauten (Figure 5.4j). Low-pressure centres are located over the Atlantic north of the Azores and Scandinavia, high pressure extends over the Icelandic region and Southeastern Europe.

Solar irradiation is only marginally altered for the different Dunkelflauten types. Minor positive (negative) anomalies occur in the south (north) of Germany during European Blocking Dunkelflauten (Figure 5.4e). For Scandinavian Blocking Dunkelflauten, only a negative solar irradiation anomaly is detected in northern Germany (Figure 5.4h), and no solar irradiation anomalies in Germany are present during Greenland Blocking Dunkelflauten (Figure 5.4k).

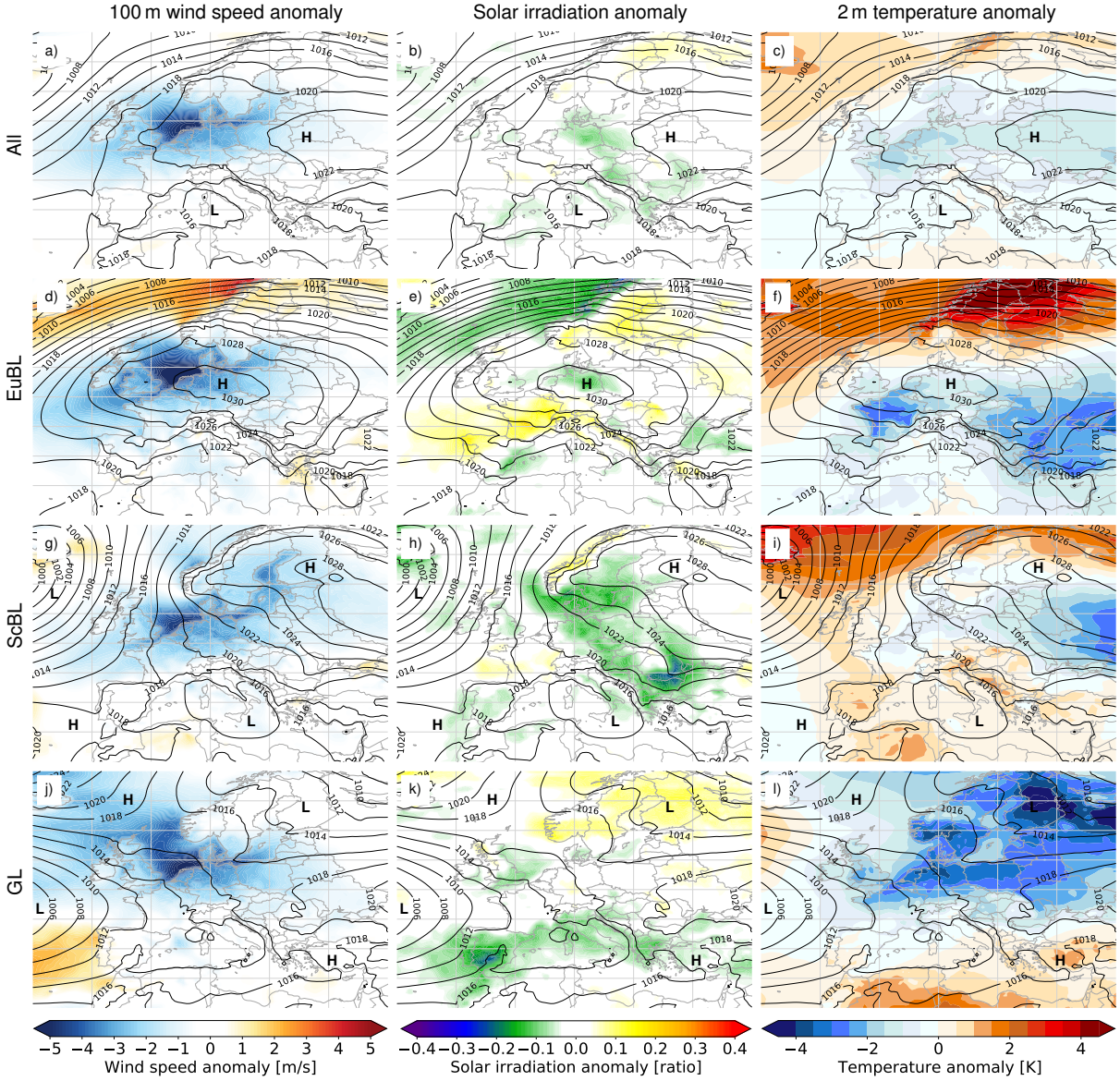


Figure 5.4.: Composites of atmospheric field variables for different types of Dunkelflauten. The rows show (a–c) all Dunkelflauten events, (d–f) European Blocking Dunkelflauten, (g–i) Scandinavian Blocking Dunkelflauten and (j–l) Greenland Blocking Dunkelflauten. The first column shows 100 m wind speed anomalies (with respect to a 30 day running mean climatology) as shading, the second column the anomaly in the daily fraction of maximum solar insolation (with respect to the 30 day running mean climatology) and the third column 2 m temperature anomalies (with respect to the 30 day running mean climatology). Each composite shows the absolute sea level pressure with a 2 hPa contour interval.

For European Blocking Dunkelflauten, the 2 m temperature, in particular in Germany but also in Western and Eastern Europe in general, is up to 2 K below the 30 day running mean climatology (Figure 5.4f). For Scandinavian Blocking Dunkelflauten, the south of Germany experiences marginally warmer and the north of Germany marginally colder 2 m temperatures in comparison to climatology (Figure 5.4i). During Greenland Blocking Dunkelflauten, Germany is under the influence of substantial negative 2 m temperature anomalies, which are up to 4 K colder compared to the 30 day running mean climatology

(Figure 5.4l). The negative 2 m temperature anomaly is present in Western, Central, Eastern, and Northern Europe, as well as in Russia.
Analysing the temperature anomalies in more detail for Greenland Blocking Dunkelflauten shows that the negative anomalies in Germany and Northern Europe are already present 6 days prior to the onset of the Dunkelflauten (Supplementary Material A.6). Cold polar air is advected to Europe prior to the onset of the Dunkelflaute. Subsequently, the cold air mass becomes stationary in Germany and even endures the Dunkelflaute itself. The long-lasting negative temperature anomalies amplify the potential stress that an energy system faces during and after Dunkelflauten related to Greenland Blocking.
In summary, exploring the meteorological conditions during Dunkelflauten related to different weather regimes unveiled that not all Dunkelflauten are caused by high pressure over Germany and go along with cold conditions. For Dunkelflauten related to Greenland Blocking, Germany is located in a saddle point between weather systems. Importantly, Greenland Blocking Dunkelflauten are cold Dunkelflauten, causing unusually cold conditions for a prolonged period, and, therefore, likely stress the energy system.

5.4. Characteristics of weather regimes associated to Dunkelflauten

The strong link of Dunkelflauten to specific weather regimes raises the question of whether regime life cycles associated with Dunkelflauten differ from the general characteristics of regime life cycles. Therefore, we now explore life cycle characteristics such as the regime duration and the relation of Dunkelflauten occurrence and regime onset and decay.
First, we investigate the duration of weather regimes with Dunkelflauten. To account for the small sample size of weather regimes, such as 52 Greenland Blocking life cycles in the months from November to March over the years 1979–2018, a direct comparison of average life cycle durations is not suitable. Instead, we employ a bootstrapping method following the approach by Wernli and Papritz (2018) and explained in detail in Section 2.5 of Büeler et al. (2020): For each weather regime category, we create two bootstrapped distributions, one for the Dunkelflauten and one as a climatological reference. Each distribution consists of 1000 random samples with the same number of elements as the number of Dunkelflauten for the respective weather regime category. The elements for the Dunkelflauten samples are drawn from all Dunkelflauten events with replacements. The elements for a sample of the climatological reference distribution are drawn with replacements from all regime life cycles in the 40-year period. To retain the seasonal cycle of Dunkelflauten in the random climatological samples, we draw for each Dunkelflaute event a corresponding random element in a ± 45 day window around the same day of the year but for a random year. For each of the 1000 random samples, we then calculate the mean life cycle duration to obtain a distribution of the mean life cycle duration for Dunkelflauten and in the climatological reference.
Following Wernli and Papritz (2018) and Büeler et al. (2020), we test if the life cycle duration for Dunkelflauten is statistically significantly longer on a 20% level if the 20% percentile of the Dunkelflauten

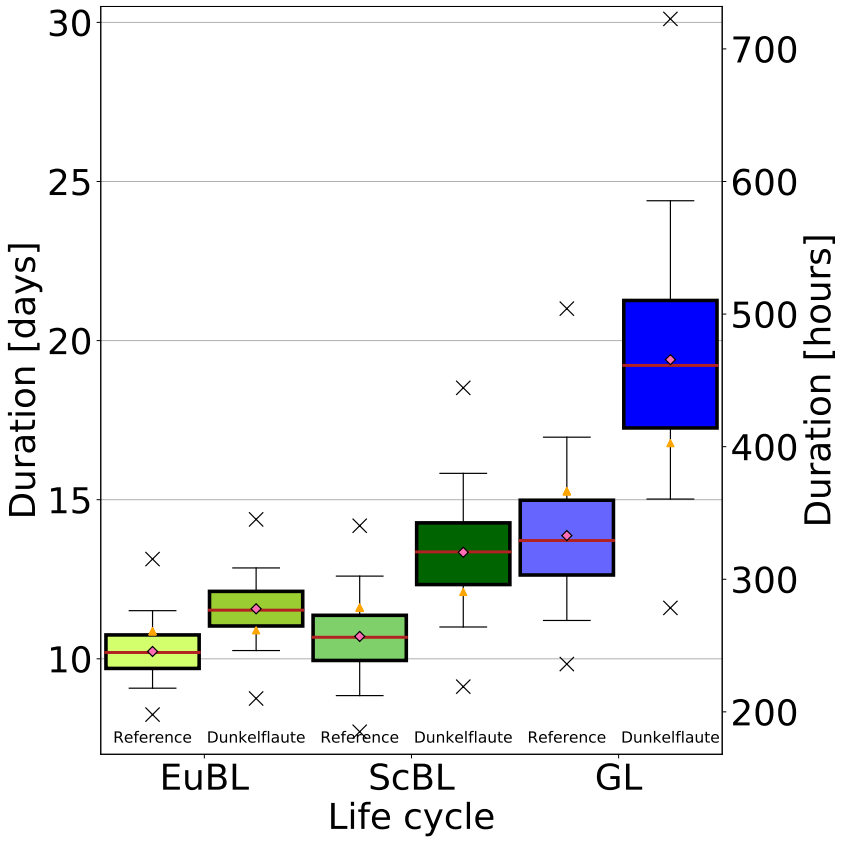


Figure 5.5.: Bootstrapped distribution of mean regime life cycle duration associated with Dunkelflauten: Climatological distribution (pale coloured bar plots left) and life cycles associated with Dunkelflauten (dark coloured bar plots right) for European Blocking (EuBL, light green), Scandinavian Blocking (ScBL, dark green) and Greenland Blocking (blue). Pink diamonds represent the mean, red horizontal lines the median, boxes the 25–75% percentile, whiskers the 5–95% percentiles and outliers the minimum and maximum values. Orange triangles indicate the 80% and 20% percentiles for the climatological and Dunkelflauten distribution, respectively.

distribution is greater than the 80% percentile of the reference distribution (indicated as orange triangles in Figure 5.5). The rather high significance level of 20% is chosen to account for the relatively high number of life cycles with a Dunkelflaute compared to those without a Dunkelflaute in winter time. For instance, of the 52 Greenland Blocking life cycles in NDJFM, 18 feature a Dunkelflaute, roughly a share of 1/3. Drawing the climatological reference samples only from those life cycles not featuring a Dunkelflaute would yield similar results as reported below, but on a 10% significance level (not shown). The reference samples' mean life cycle duration distribution reveals differences between the European, Scandinavian and Greenland Blocking weather regimes (Figure 5.5, left columns). Whereas European and Scandinavian Blocking weather regime life cycles are active for 10.2 and 10.7 days respectively, Greenland Blocking life cycles have a mean lifespan of 13.8 days.

Comparing the Dunkelflauten distribution (Figure 5.5, right columns) to the reference distribution (Figure 5.5, left columns) for each weather regime separately shows significantly longer life cycle durations for the Dunkelflauten samples. Life cycles associated with European and Scandinavian Block-

ing Dunkelflauten have a mean duration of 11.6 and 13.3 days respectively (an increase in duration of 14 and 24% compared to the reference samples). For life cycles associated with Greenland Blocking Dunkelflauten, the difference is even more pronounced with an average life cycle duration of 19.4 days (an increase of 41%). All three life cycle duration differences are significant on the 20% level.

The results of this bootstrapping method indicate that Dunkelflauten occur in longer-lived life cycles compared to life cycles in the same seasonal period. Considering the average life cycle duration of 11–

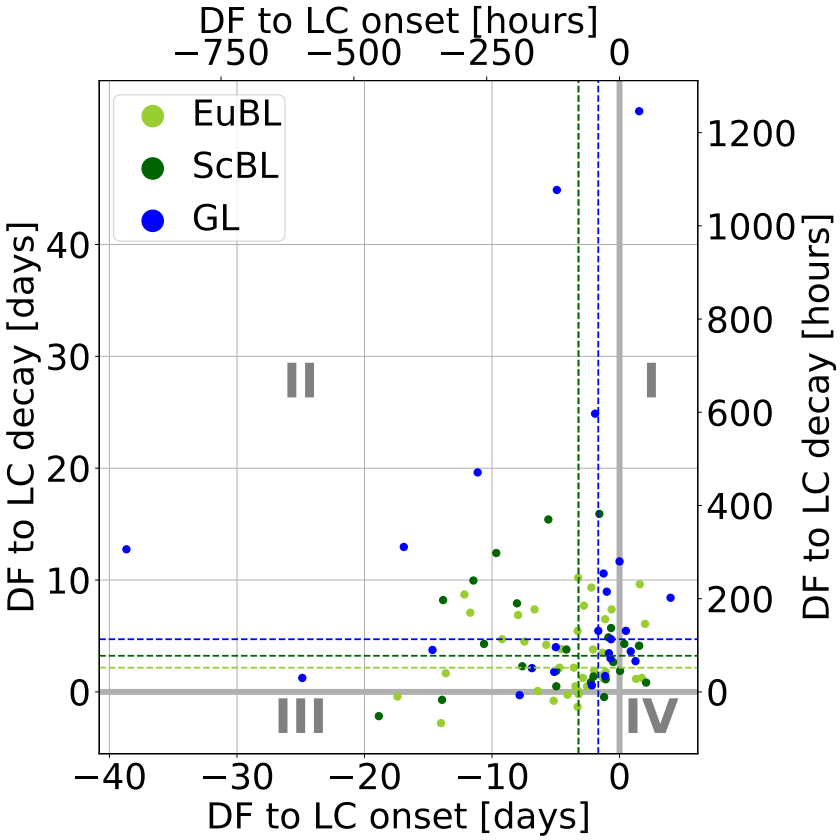


Figure 5.6.: Onset and decay of Dunkelflauten (DF) in relation to the onset and decay of the associated regime life cycle (LC) for European, Scandinavian and Greenland Blocking Dunkelflauten. The median time differences between the onsets and decays for each Dunkelflauten category are represented by the dashed vertical and horizontal lines, respectively. The four quadrants distinguish the following categories: I. A Dunkelflaute sets in before the life cycle and decays during the active life cycle; II. A Dunkelflaute sets in during the associated life cycle and decays during the active life cycle; III. A Dunkelflaute sets in during an active life cycle and decays after the decay of the life cycle; IV. A Dunkelflaute sets in prior to the onset of the life cycle and lasts longer than the life cycle; therefore decays after the decay of the life cycle.

20 days and the Dunkelflauten duration of 3–9 days, the subsequent question is whether Dunkelflauten occur at specific times of the associated life cycle.

Dunkelflauten are mostly fully embedded in the regime life cycle associated with the Dunkelflaute (Figure 5.6 quadrant II, indicating that the Dunkelflauten onset is after the life cycle onset and the Dunkelflauten decay is prior to the life cycle decay). In numbers, this translates to an average (median) onset of Dunkelflauten 3.3/3.2/1.7 days after the onset of the life cycle and an average decay

2.2/3.2/4.7 days prior to the decay of the life cycle for European/Scandinavian/Greenland Blocking, respectively. Thus, Dunkelflauten occur mainly in the well-developed and stable phases of regime life cycles.

5.5. Electricity demand during Dunkelflauten

With negative 2 m temperature anomalies during European and Greenland Blocking Dunkelflauten (shown in Section 5.3), we expect a higher heating demand in Germany due to colder than normal temperatures in wintertime. The combination of a Dunkelflauten event with low renewable power output and an increased electricity demand due to cold temperatures will likely put stress on the energy system.

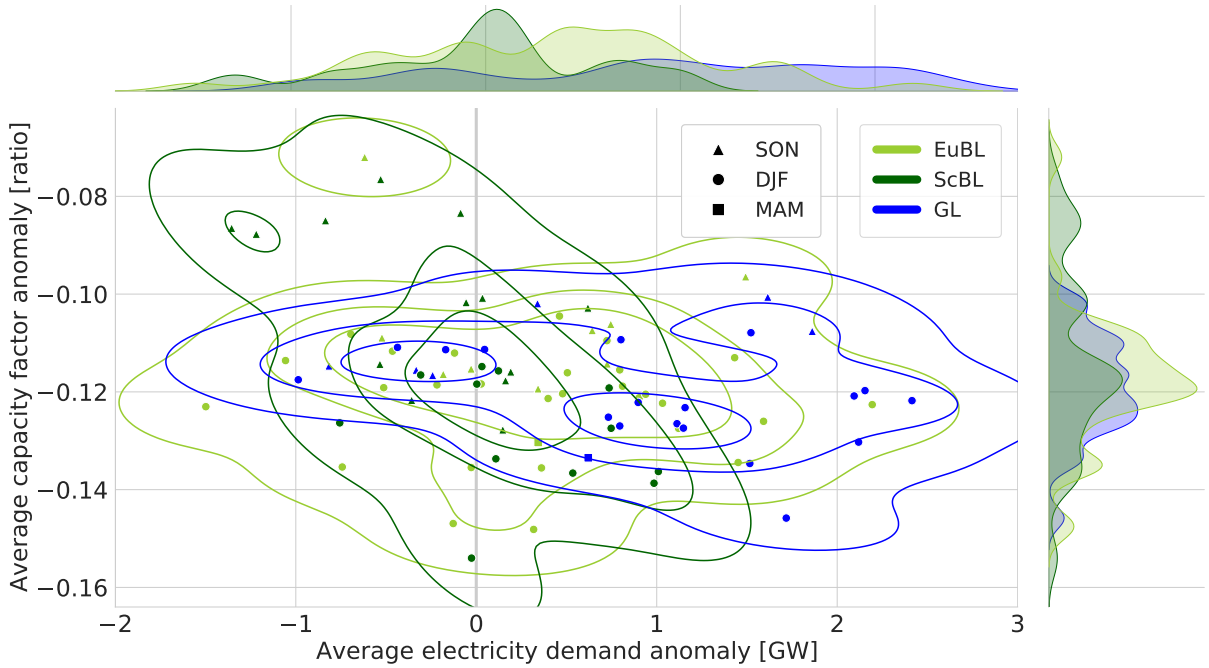


Figure 5.7.: Average capacity factor and electricity demand anomalies during different Dunkelflauten types in Germany. The electricity demand is based on the prevailing socioeconomic conditions of 2017. All European, Scandinavian, and Greenland Blocking Dunkelflauten are represented with their regime colour. The shape of the marker indicates the season of each Dunkelflaute. The 2D density distribution of each Dunkelflauten type is given as shading in the main figure and as 1D distribution in the two marginal figures.

The capacity factors during the three dominant Dunkelflauten types are reduced on average from 0.18 to 0.06 (Figure 5.7, right marginal figure), which is a relative reduction of 66% compared to winter climatology.

Whereas there is no anomalously high average electricity demand for Scandinavian Blocking Dunkelflauten (Figure 5.7, top marginal figure), the average electricity demand during Dunkelflauten is marginally increased for European Blocking Dunkelflauten (average increase of 0.5 GW) and more strongly increased for Greenland Blocking Dunkelflauten (average increase of 1.0 GW, but values also reaching up to 3.2 GW). Absolute values of the electricity demand in the period from November to March

range from 44.5 GW up to 76.5 GW. Therefore, during Dunkelflauten, the electricity demand may rise by up to 7%, using the socioeconomic conditions of 2017. Although we did not directly investigate the dependence of electricity demand and 2 m temperature, the latter is likely a main meteorological driver of demand (Bloomfield et al., 2020b) (Figure 5.7). Also, as shown in Section 5.3, marked temperature anomalies occur in Germany during Dunkelflauten. In particular, persistent low temperatures accompany the low power output during the cold Greenland Blocking Dunkelflaute (Figure 5.4l and Supplementary Material A.6). Thus we conclude that Greenland Blocking Dunkelflauten are likely the most challenging periods for the operation of an energy system with high shares of renewables.
An even larger increase in the electricity demand can be expected in the future due to the transition to a fully renewable energy system, so this problem amplifies.

5.6. Conclusions and discussion

The present study links large-scale atmospheric circulation patterns, called weather regimes, and periods with low wind and solar power in Germany called Dunkelflauten. Following Kaspar et al. (2019), Dunkelflauten are defined as periods with a low combined capacity factor of solar photovoltaic, onshore wind and offshore wind in Germany, lasting for at least two days, using the atlite energy system model framework and meteorological data from ERA5 reanalysis. We then link each Dunkelflauten event from 1979–2018 to one of the seven year-round weather regimes or the "no regime" (Grams et al., 2017).
Dunkelflauten are mainly found in Germany during autumn and winter. This seasonal effect is due to the low mean solar capacity factor in winter. Thus Dunkelflauten are primarily caused by low wind speed conditions.
Differentiating Dunkelflauten by the prevalent large-scale atmospheric flow, using weather regimes, helps to identify different atmospheric patterns leading to Dunkelflauten in Germany. Three blocking weather regimes are most frequently associated with Dunkelflauten: European Blocking with 22% of all Dunkelflauten, Scandinavian Blocking with 15%, and Greenland Blocking with 14%. A high-pressure system over Germany and Central Europe is the most frequent Dunkelflauten pattern and is associated with the European Blocking weather regime.
The meteorological conditions during European, Scandinavian, and Greenland Blocking Dunkelflauten differ: A high-pressure system extends over Central Europe during European Blocking Dunkelflauten, and Germany is located in the centre of the high pressure with prevalent low pressure gradients and consequently low winds. In contrast, during Scandinavian Blocking and Greenland Blocking Dunkelflauten, Central Europe is located in a saddle point of a quadripole of two high and two low-pressure systems, likewise causing low-pressure gradients and low winds in Germany. Thus, it is essential to analyse Dunkelflauten not as one composite but separately between different weather regime patterns. Greenland Blocking Dunkelflauten are considered to be cold Dunkelflauten. These Dunkelflauten periods are up to 4 K colder than the 30 day running climatology, increasing the electricity demand from electrical

heating. Increased electricity demand in periods of lower-than-normal power output by renewable energy sources puts stress on the energy system.

Weather regime life cycles associated with Dunkelflauten also differ in their general characteristics: Life cycles associated with Dunkelflauten are longer-lived than usual, especially for Greenland Blocking weather regimes, with life cycles on average 5 days longer than the NDJFM climatology (19 days compared to 14 days). The life cycles of weather regimes begin well before the onset of the Dunkelflauten and end thereafter. Thus Dunkelflauten occur when weather regime life cycles are well-established. This is useful from a forecast perspective, as knowledge about upcoming blocking weather regimes can help prepare the energy system for an imminent supply shortage.

Our results confirm findings from previous studies and extend the knowledge of Dunkelflauten. Kaspar et al. (2019) and Drücke et al. (2021) also find Dunkelflauten in Germany, mainly in autumn and winter. They categorise Dunkelflauten by Grosswetterlagen (Grosswetterlagen classify the circulation in Europe with 29 different weather types focusing more on the regional conditions than the continental-scale large-scale weather regimes used here) and identify the "Grosswetterlage GWL9" as the most frequent pattern for Dunkelflauten in Germany. GWL9 has comparable characteristics to the European Blocking weather regime. Wiel et al. (2019b) identify the NAO– weather regime, derived from the four weather regime classification by Vautard (1990); Michelangeli et al. (1995); Cassou (2008), and strongly correlating with the Greenland Blocking weather regime, to be the scenario with reduced energy production and increased energy demand for Europe. The increased electricity demand in Greenland Blocking Dunkelflauten, using the electricity demand model of Bloomfield et al. (2020b), is likely to intensify in the future due to Germany's transition to electrical heating. In the past 10 years, the contribution of electrical heating to the heating structure of new buildings has doubled, from 24% in 2012 up to 50% in 2022 (BDEW, 2022). Otero et al. (2022) show the link of energy compound events in Germany, simultaneous episodes of low renewable energy production (wind plus solar power) and high electricity demand, and weather regimes. Energy compound events in Germany are more frequent in European and Greenland Blocking weather regimes. These results are consistent with the cold Dunkelflauten in our research.

We extend the knowledge of Dunkelflauten in Germany by analysing weather regime life cycles and meteorological conditions associated with different Dunkelflauten. Dunkelflauten, especially the cold Greenland Blocking Dunkelflauten, are positioned in the well-established phase of longer-lived weather regime life cycles. Furthermore, Dunkelflauten not only occur when high pressure prevails but also in conditions when Germany is in the middle between pronounced weather systems elsewhere.

Although we find robust results and confirm previous findings, our choices for defining Dunkelflauten (minimum duration, capacity factor threshold) and weather regimes (low-pass filter, minimum duration, life cycle definition) might affect the results. Therefore we stress that the results are valid for the specific definitions used here. In particular, other regime definitions might identify different weather regimes relevant for Dunkelflauten. Also, our regime definition allows for parallel regime life cycles. However, here we attributed Dunkelflauten based on the highest number of time steps with maximum regime

projections during the Dunkelflaute which does not consider the potential simultaneous projection into another regime. Future avenues of research could explore factors that may contribute to the occurrence of Dunkelflauten besides the already identified long-duration of the weather regime and investigate whether Dunkelflauten occur in parallel regime life cycles.

The link of Dunkelflauten to weather regimes provides a forecast opportunity on the subseasonal to seasonal range. Bloomfield et al. (2021a) showed that pattern-based methods, such as weather regimes, outperform grid-point forecasts for lead times larger than 12 days for the European national power forecasts. Weather regime forecasts can fill the gap between short-range weather prediction and long-range seasonal outlooks for the energy sector (White et al., 2017). With the results of our study, we expect to more accurately forecast Dunkelflauten using weather regimes, as the likelihood of a potential Dunkelflaute could be identified not only by the forecasted weather regime but also by the forecasted duration of a weather regime life cycle.

Based on our results in combination with the results by Büeler et al. (2021), showing the modulation of surface weather by weather regimes and the promising skill of weather regimes on the subseasonal to seasonal range, we see weather regime forecasts as an essential tool for energy system operators to prepare for multi-day supply shortages.

This chapter has demonstrated a strong link between Dunkelflauten and weather regimes. However, forecasting these regimes remains challenging, as systematic errors in numerical models tend to grow with lead time, limiting forecast skill. To address this, the next chapter evaluates the skill of weather regime forecasts and employs statistical post-processing techniques to investigate whether it can improve forecast skill.

6. Statistical post-processing of weather regime forecasts

In the previous chapter, we established a strong link between Dunkelflauten and weather regimes. Accurate weather regime forecasts could provide significant benefits for decision-makers in the energy sector. However, due to the chaotic nature of the atmosphere and errors in NWP models, the forecast skill of ECMWF's model declines with lead time, limiting its probabilistic reliability for subseasonal predictions. We begin by analysing biases in Z500 forecasts and their impact on weather regime index forecasts (Section 6.1). We then conduct a comprehensive evaluation of statistical post-processing techniques, comparing the skill of raw forecasts, Z500 bias-corrected forecasts, and post-processed IWR forecasts. Our analysis focuses on two key aspects: the assessment of univariate forecast skill, using EMOS-G (Section 6.2), where each component of the 7-dimensional weather regime index is evaluated individually; and the evaluation of multivariate forecast skill, applying ECC-Q (Section 6.3), which considers the joint forecast performance across all seven dimensions simultaneously. Additionally, we examine the sensitivity of these methods to variations in training data availability (Section 6.4). It is important to note that our evaluation specifically assesses the ensemble forecasts ability to predict IWR on given days – an especially challenging task at subseasonal lead times.
This chapter is based on and slightly adapted from Mockert et al. (2024), with the working title: "Multivariate post-processing of probabilistic sub-seasonal weather regime forecasts."

6.1. Z500 forecast biases

Biases in the Z500 forecast have a direct impact on the seven dimensional weather regime index as these directly project into the IWR forecasts. The analysis of Z500 biases as a function of forecast lead time (Figure 6.1) unveils the reasons for systematic biases in the IWR (Figure B.2). In the weather regime region (denoted by the grey dashed box in Figure 6.1), forecast biases grow from values near 0 gpm at 0 day lead time to around 40 gpm at lead times beyond 20 days. Throughout the year, positive biases dominate in the northern part of the weather regime region (first row in Figure 6.1), and extend from Canada into the high-latitude North Atlantic. This positive bias anomaly most prominently projects into the Atlantic Ridge weather regime (yellow in Figure B.2). The Z500 biases are seasonally dependent which also manifest in the IWR. In winter (second row in Figure 6.1), Z500 biases exhibit a dipole structure projecting into the Atlantic Trough and Greenland Blocking (purple and blue in Figure B.2). In spring (third row in Figure 6.1), the positive Z500A in the North Atlantic grows, projecting into the Atlantic Ridge and Greenland Blocking (yellow and blue in Figure B.2). During summer (fourth row in

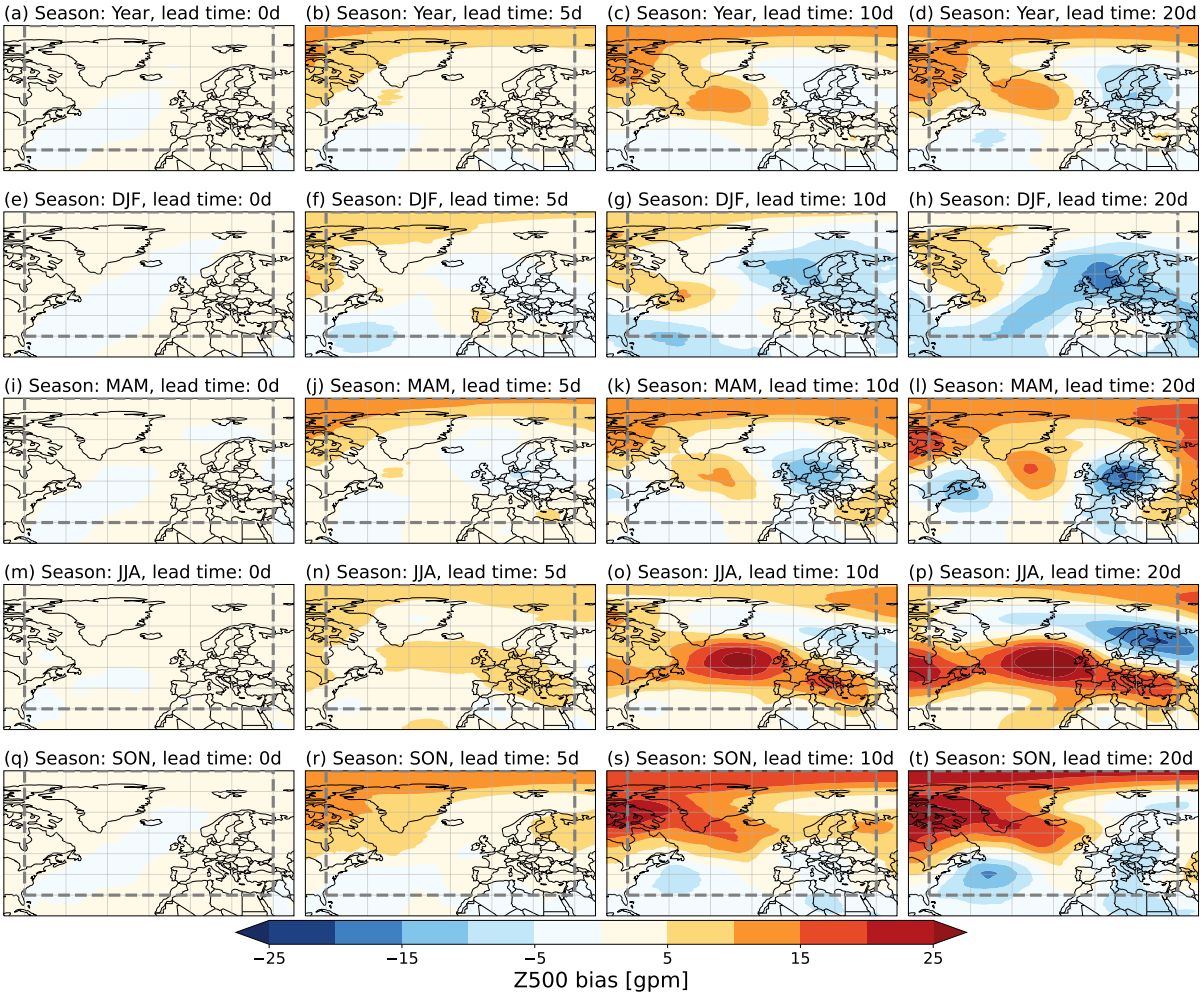


Figure 6.1.: Displayed are the mean 500 hPa geopotential height anomaly biases of the ensemble mean forecasts. The Z500A bias is computed by subtracting the ERA perfect member from the ensemble mean. The mean bias fields are presented for the entire year and four seasons (winter: DJF, spring: MAM, summer: JJA, autumn: SON) across rows, and at lead times of 0, 5, 10, and 20 days across columns. The region corresponding to NAE weather regimes is indicated by the dashed grey box. Note the bias field at lead time 0 days is not exactly 0 gpm as we present biases of the 10-day low-pass filtered Z500. Consequently, the Z500 forecast at 0 day lead time is influenced by forecasts up to 10 days ahead.

Figure 6.1), the positive anomaly in the North Atlantic intensifies, accompanied by a negative anomaly in northern Europe, projecting into the Scandinavian Trough and Atlantic Ridge (orange and yellow in Figure B.2). In autumn (fifth row in Figure 6.1), the positive anomaly resides mainly in northern America and the high-latitude North Atlantic, resulting in projections into the Greenland Blocking and Atlantic Ridge (blue and yellow in Figure B.2). This seasonally differentiated analysis of bias growth in the Z500 field and its projection into weather regimes emphasises the need for bias correction which accounts for seasonality.

6.2. Univariate post-processing

Here, we utilise EMOS for univariate post-processing. EMOS is trained on a reforecast dataset with 3101 forecasts spanning June 1999 to May 2015, and subsequently tested on 899 forecasts from the reforecast dataset covering June 2015 to May 2020. As we intend to apply the post-processing to operational forecasts, we train the EMOS for each calendar day using a running window approach with 31 days and raw forecasts. The univariate evaluation of the EMOS predictions is based on the analytical closed-form solution of the CRPS for the Gaussian forecast distribution (EMOS-G), whereas the verification rank histograms and multivariate score computations are based on the quantile-based forecasts (EMOS-Q and ECC-Q). For computational details and implications of these choices, see, e.g., Jordan et al. (2019).

In the previous section, we identified biases in the Z500 field forecasts (Figure 6.1). The (positive) bias is particularly prominent in the high latitudes of the North Atlantic, persisting across seasons and lead times, notably projecting into the Atlantic Ridge regime which is characterised by a positive Z500A in a similar region. Therefore, we focus on analysing the verification rank histograms for the Atlantic Ridge and its counterpart, the Atlantic Trough, aiming to assess forecast reliability and calibration. For a comprehensive overview of rank histograms for all weather regimes, please refer to Figures B.3 and B.4 in the Supplement.

Due to the observed bias in the Z500 forecast, the raw IWR forecast for Atlantic Trough tends to be underforecast for lead times exceeding 10 days (Figure 6.2a, noticeable in red colours at higher ranks), while overforecast on shorter lead times. Similarly, the Atlantic Ridge is consistently overforecast across all lead times (Figure 6.2d, evident in red colour at lower ranks). Although the bias correction of the Z500 field improves the reliability for these two weather regimes on subseasonal lead times (Figure 6.2b and Figure 6.2e), it deteriorates the reliability on shorter lead times. Forecasts up to 10 day lead time tend to overforecast the Atlantic Trough regime, while the Atlantic Ridge is strongly underforecast on lead times up to 12 days after the Z500 bias correction. Verification rank histograms show that forecasts post-processed with EMOS exhibit improved reliability (Figure 6.2c and Figure 6.2f for the Atlantic Trough and Atlantic Ridge, respectively). While forecast biases persist to some extent, their magnitude decreases, and the largest deviations shift from the outer ranks towards central ranks. This general trend is also observed across the other weather regimes. EMOS consistently enhances the raw IWR forecast, particularly for the Zonal Regime and Scandinavian Trough (see Supplementary Figure B.3). Forecasts are also better calibrated for Greenland Blocking. However, forecasts of European and Scandinavian Blocking exhibit similar or even larger miscalibration, with observation too frequently falling into the highest ranks (see Figure B.4).

In summary, the Z500 bias-corrected forecast, especially on shorter lead times, degrades the ensemble's calibration. The IWR is either underforecast (Scandinavian Trough), overforecast (Greenland Blocking), or overconfident (Zonal Regime, European and Scandinavian Blocking), as evident in the underdispersive distribution in the verification rank histograms. Overall, these findings suggests that EMOS post-

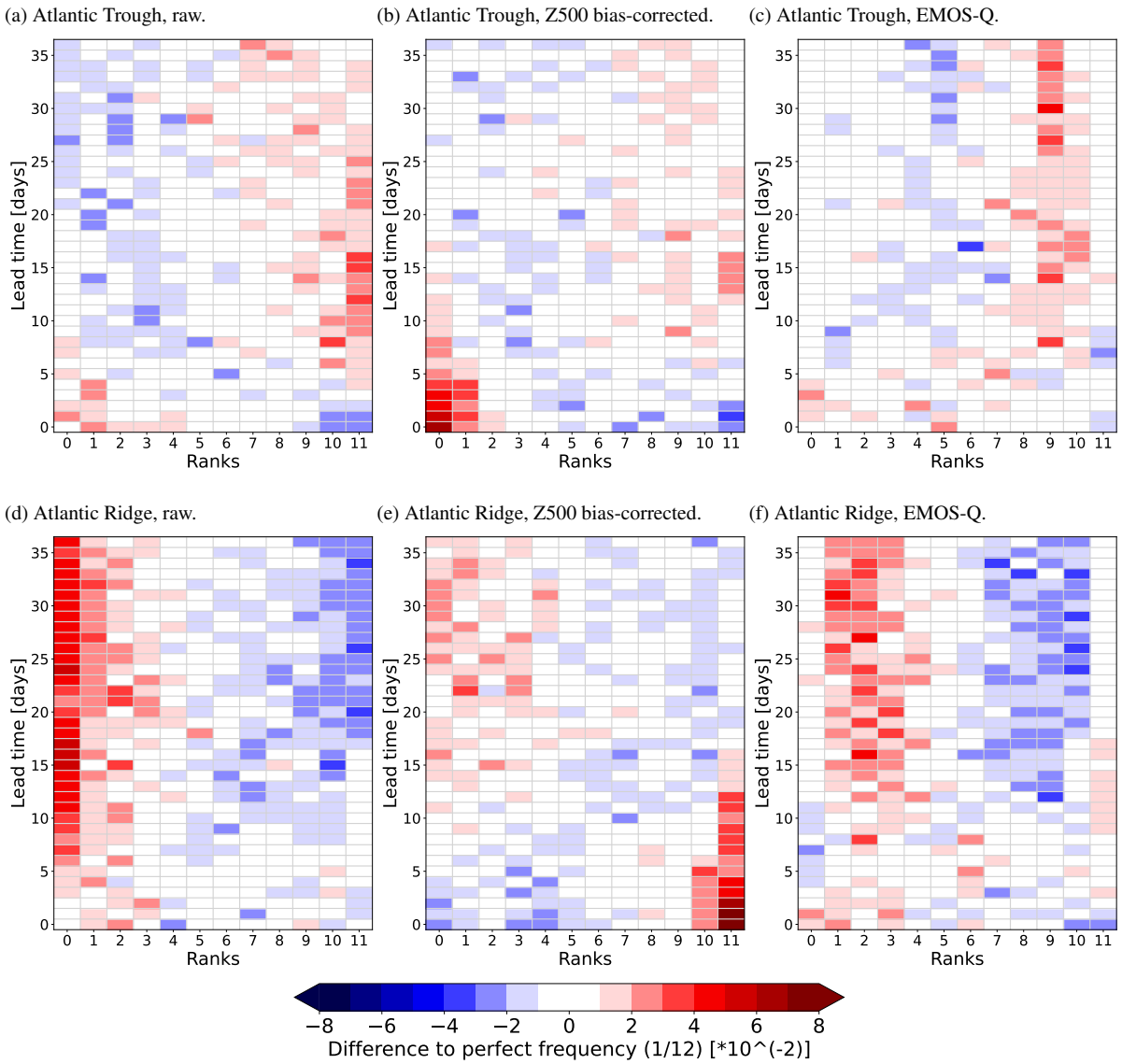


Figure 6.2.: The verification rank histograms for the Atlantic Trough (a–c) and the Atlantic Ridge (d–f) are visualised. Additional rank histograms for all cyclonic and anticyclonic weather regimes can be found in supplementary Figures B.3 and B.4. The figures illustrate the rank histograms for the raw forecasts (a, d), Z500 bias-corrected forecasts (b, e), and EMOS-Q post-processed forecasts (c, f) for both weather regimes. The rank is illustrated on the x-axis, the lead time on the y-axis and the frequency of occurrence, specific for each lead time, is indicated by the values and colours of the boxes. To facilitate the readability, the frequency is shown as anomaly to a perfect distribution of 1/12.

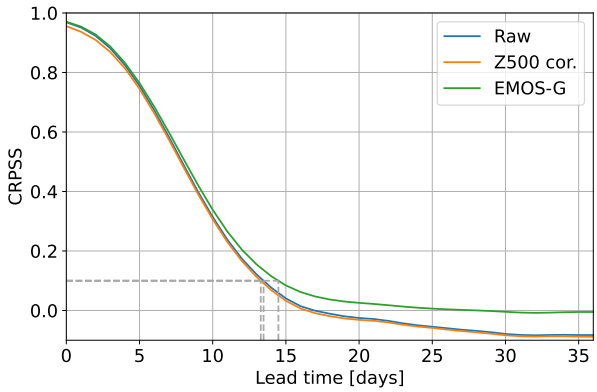
processing generates more reliable and consistent forecasts for the weather regime index compared to Z500 bias-corrected forecasts.

To investigate the corrections via EMOS in more detail, we analyse the estimated EMOS coefficients introduced in Equation 4.10 for different weather regimes (Figure B.5). Across all weather regimes (Figure 4.10a–g), the coefficients associated with the ensemble mean (a_1) and ensemble variance (b_1) converge to zero as lead time increases. The intercept coefficient of the location parameter (a_0) remains

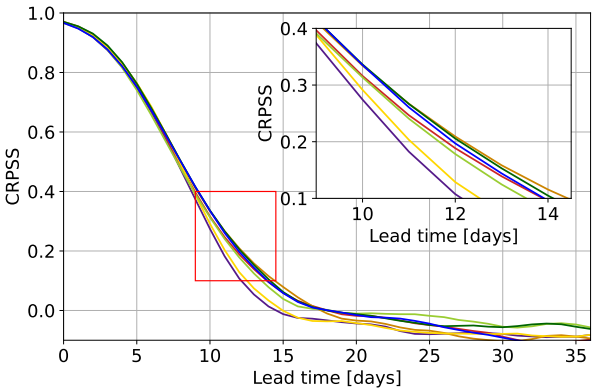
nearly constant at zero, while the intercept coefficient of the scale parameter (b_0) converges to one. This indicates that the EMOS model has learned to rely less on the flow-dependent information from the raw ensemble predictions and reverts to climatological forecasts as lead time increases. In particular for the first 10 days, there is no notable bias correction of the ensemble mean prediction with coefficient values of a_1 close to one and a_0 close to zero. Thus, the main effect of EMOS post-processing appears to be the adjustment of the ensemble spread.

We now dive deeper into the comparison of the univariate skill of the various forecasting methods (raw, Z500 bias-corrected, and EMOS-G). For this purpose, we assess the continuous ranked probability skill score (CRPSS) with climatology as the reference forecast. As an initial comparison among the forecasting methods, we examine the mean skill score over the 7 weather regimes (Figure 6.3a).

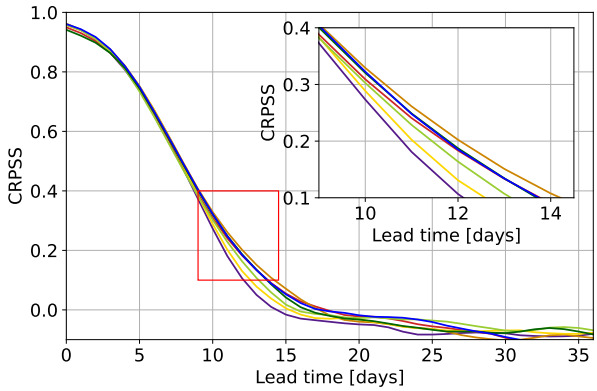
(a) Mean CRPSS across the 7WR for the forecasting methods.



(b) Raw forecast.



(c) Z500 bias-corrected forecast (current practice).



(d) Post-processed EMOS-G forecast.

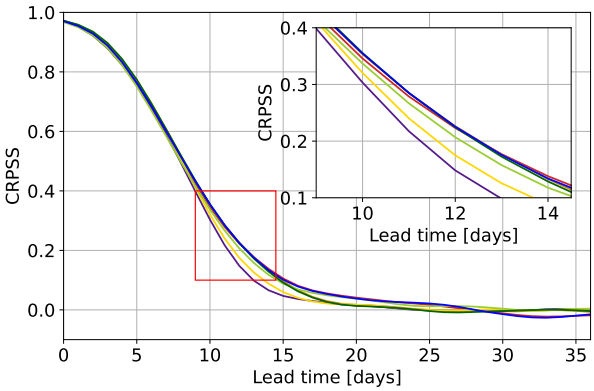


Figure 6.3.: CRPSS as a function of lead time. In (a), the aggregated mean CRPSS across the weather regimes is shown for the three forecasting methods, raw (blue), Z500 bias-corrected (orange) and EMOS-G (green). In (b)–(d) the CRPSS is shown separately for each weather regime, indicated by the regime colours, for (b) the raw forecasts, (c) the Z500 bias-corrected forecasts and (d) the EMOS-G forecasts. The red box indicates the area in which we zoom in to better visualise the differences between those regimes. The CRPSS is calculated against a 31-day rolling climatological forecast.

The mean skill scores of the three forecasting methods (Figure 6.3a) mainly differ for subseasonal lead times with the CRPSS for EMOS-G approaching 0 and only surpassing it minimally at lead times beyond 30 days. The CRPSS for the raw and Z500 bias-corrected forecasts is greater than 0 until day 17 and approaches a skill score of –0.09. For early lead times, all forecasting methods obtain high scores, with the raw and EMOS-G forecast slightly outperforming the Z500 bias-corrected forecasts. On the extended lead times, the Z500 bias-corrected forecasts exhibit slightly lower CRPSS values than the raw forecasts, though the difference is minimal.

In Figure 6.3a, we observed that the mean skill scores of the raw and EMOS-G forecast remain similar up to day 10 and afterwards EMOS-G forecasts show noticeably higher skill than the raw and Z500 bias-corrected forecasts. We now analyse the skill scores across the weather regimes (Figures 6.3b–6.3d) to reveal commonalities and differences between the individual forecasting methods.

All three forecasting methods have in common that the difference in the CRPSS of the individual weather regimes are close to indistinguishable up to a lead time of 7 days. A further commonality is the order of skill for the different weather regimes on lead times between 10 and 15 days (the lead time range until where all forecasting methods still exhibit CRPSS values larger than 0 for each weather regime). The lowest skill is observed for Atlantic Trough and Atlantic Ridge, which coincides with our choice in the analysis of verification rank histograms due to the region of largest Z500 bias in the North Atlantic that project on the anomalies associated with Atlantic Trough and Atlantic Ridge. The third lowest forecast skill is found for European Blocking, which is known to be particularly challenging to predict compared to other regimes, looking at the categorical weather regime definition (see Büeler et al. (2021)). For the remaining four regimes, a common order of skill cannot be discerned. However, it is noteworthy that for the EMOS-G forecasting method, the CRPSS across all four regimes (Zonal regime, Scandinavian Trough, Scandinavian, and Greenland Blocking) is remarkably similar.

Overall, largest differences in terms of forecast skill for the different forecasting methods and regimes is observed for lead times of 10–14 days (insets in Figures 6.3b–6.3d).

To asses the skill score improvements of the EMOS-G forecasts in comparison to the raw forecast in more detail, we calculate the lead time gain across a range of forecast skill horizon thresholds (Figure 6.4). The forecast skill horizon is defined as the lead time at which the CRPSS of a forecast falls below a certain threshold. Typical thresholds for the forecast skill horizon are 0, which indicates that the CRPS of the forecasting method achieves the same score as the climatological reference forecast or 0.1 (see Büeler et al. (2021) in the context of weather regimes), which indicates an improvement of the skill score compared to climatology of 10%. There is no set rule as to which threshold should be analysed as this is subject to the forecast question. Therefore, we provide a visualisation of a range of thresholds ranging from 0.0 up to 0.4 at intervals of 0.01, demonstrating the robustness of the results across a range of thresholds (Figure 6.4). EMOS-G consistently outperforms the raw forecast, as all lead time gain values are positive for each weather regime. The mean lead time gain ranges from 0.3 days for a CRPSS threshold of 0.4 up to 11.2 days for a threshold of 0.0. The most significant improvements

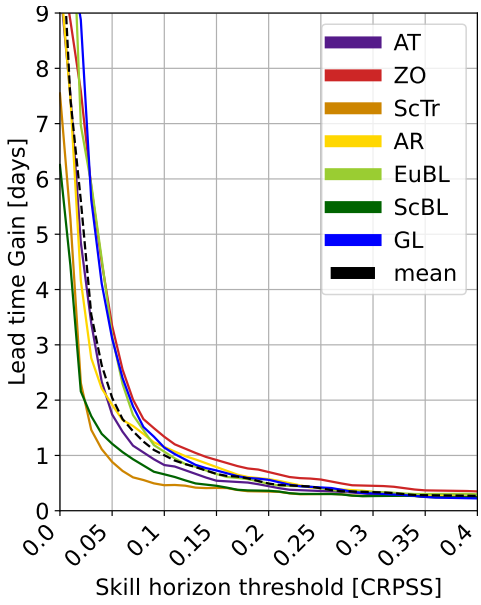


Figure 6.4.: Forecast skill horizon gain as a function of different CRPSS thresholds from raw forecasts to EMOS-G forecasts. Lead time gain separated for weather regimes.

occur in forecasts of the Zonal Regime, Greenland Blocking, and European Blocking, while the smallest improvements are observed for forecasts of the Scandinavian Trough and Scandinavian Blocking. This is in line with the close proximity of the CRPSS curves for these four regimes (excluding European Blocking) in Figure 6.3d. The apparent similarity in terms of forecast skill is due to the substantial increase of forecast skill for the Zonal Regime and Greenland Blocking after being post-processed with EMOS-G. Fixing the CRPSS threshold to 0.1, similar to Büeler et al. (2021), we find a mean forecast skill horizon of 13.5 and 14.5 days for the raw and EMOS-G forecast, respectively, indicating a lead time gain of 1 day.

To complete the univariate analysis, we evaluate the significance of skill improvements using a Diebold-Mariano test for equal performance on the CRPS. We compare the EMOS-G method with the raw (Figure 6.5a) and Z500 bias-corrected (Figure 6.5b) forecasts, as well as with the climatological reference forecast (Figure 6.5c). EMOS-G leads to significantly higher skill than the raw and Z500 bias-corrected forecasts. These results are statistically significant for all weather regimes and all lead times at a level of 0.05, except for the Scandinavian Trough at lead times between 12 and 16 days (Figure 6.5a). The least significant results on the extended lead time when comparing EMOS-G forecasts to raw forecasts are observed for the European and Scandinavian Blocking, which is in line with the respective verification rank histograms in the supplementary Figures B.4d and B.4f and Figures B.4g and B.4i, respectively. Compared to climatology (Figure 6.5c), EMOS-G demonstrates significant performance improvements across all weather regimes out to 19 days forecast lead time, with even longer significant improvements for the European Blocking, reaching out to 28 days.

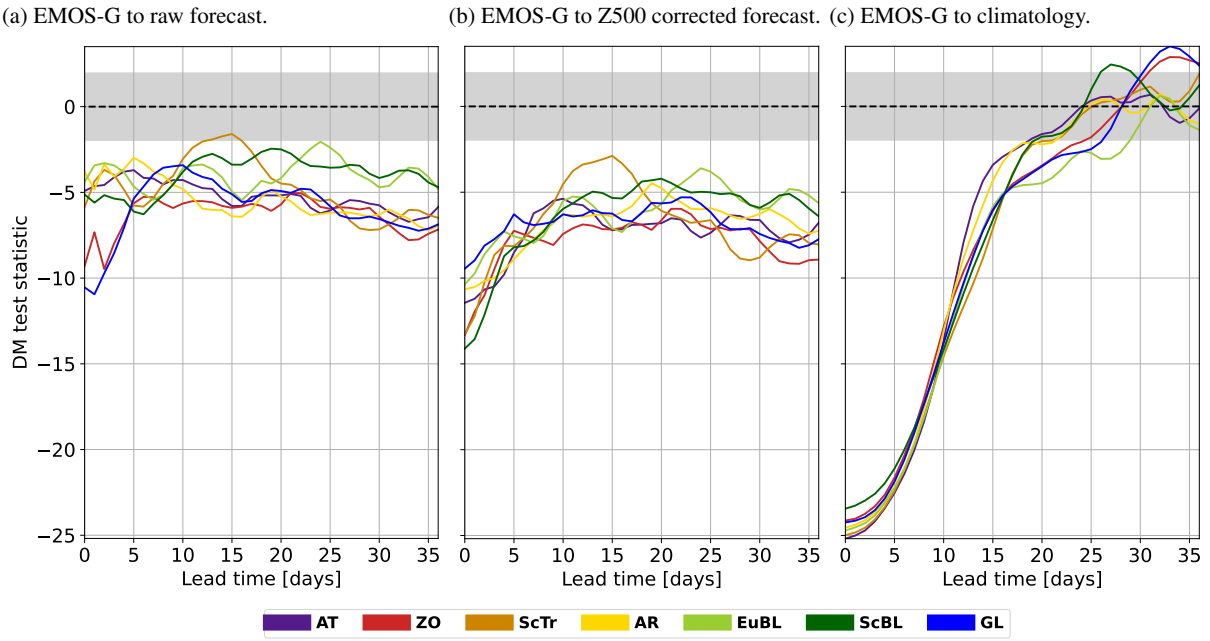


Figure 6.5.: Testing the significance of the univariate CRPS improvements of the EMOS-G method with respect to (a) the raw forecast, (b) the Z500 bias-corrected forecast and (c) the climatological forecast using a Diebold-Mariano test of equal performance with an α-level of 0.05 (indicated by the grey shading).

In conclusion, applying EMOS-G post-processing to the raw forecasts leads to significant skill improvements across all weather regimes and lead times compared to the Z500 bias-corrected method. The forecast skill horizon, measured by a 10% CRPSS improvement relative to climatology, extends to an average of 15.5 days for all weather regimes, surpassing the current practice of Z500 calibration by 1.2 days and the raw forecasts by 1 day.

6.3. Multivariate post-processing

Similar to the evaluation of univariate post-processing skill (Section 6.2), we compare skill scores with the raw and Z500 bias-corrected forecasts and assess the significance of the skill differences using a Diebold-Mariano test. To evaluate the multivariate skill of the forecasts, we employ the energy score, which is a multivariate extension of the CRPS. We also evaluate our results by using the variogram score as an alternative metric which has been argued to be more discriminative with respect to the correlation structure.

When comparing the multivariate skill scores (ESS in Figure 6.6a and VSS in Figure 6.6b) of the univariate post-processing of EMOS-Q (green lines) with the additional multivariate post-processing of EMOS-Q plus ECC-Q (red lines), the necessity of the multivariate step becomes clear. EMOS-Q ensemble members are sorted in ascending order (ensemble member 1 has the lowest values of each weather regime index and ensemble member 11 the highest values), while for ECC-Q the IWR values are sorted based on the rank order of the raw forecast ensembles. This comparison demonstrates the direct effect

(a) Energy Skill Score. (b) Variogram Skill Score.

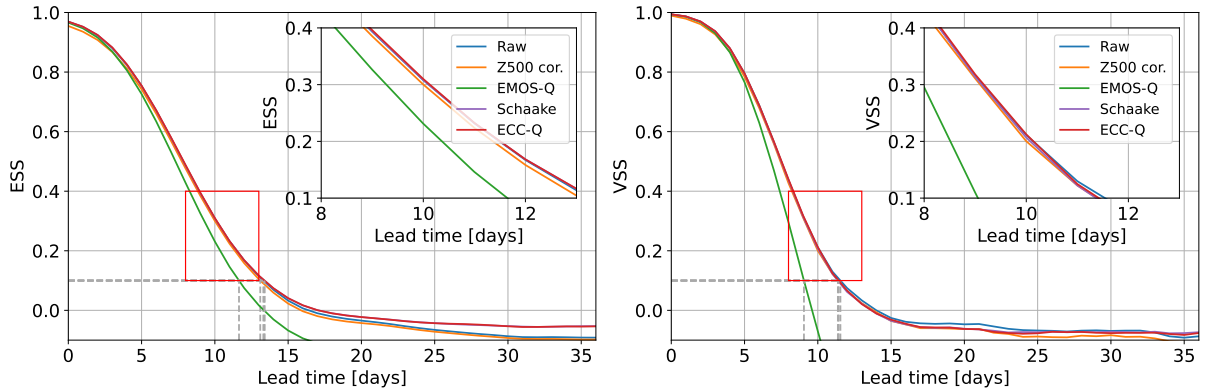


Figure 6.6.: Multivariate skill scores for raw forecast (blue), Z500 bias-corrected forecast (orange), EMOS-Q (green), Schaake Shuffle (purple) and ECC-Q (red). Using (a) the energy skill score and (b) the variogram skill score.

of the multivariate post-processing step. The forecast skill improvement of ECC-Q in comparison to EMOS-Q is notable as early as of 5 days lead time. The more relevant comparison of multivariate skill scores is between ECC-Q and raw and Z500 bias-corrected forecasts. The raw forecasts (blue), Z500 bias-corrected forecasts (orange) and ECC-Q post-processed forecasts (red) exhibit comparable skill for lead times up to 12 days for ESS and VSS (Figure 6.6a and 6.6b). On extended lead times, the energy skill score for the raw and Z500 bias-corrected forecasts is comparable and ECC-Q exhibits higher skill. However, the score of all three forecasting methods is below 0 after lead times of 16–18 days and hence less skilful than a climatological forecast. For the variogram skill score, the results are similar for the raw and Z500 bias-corrected forecasts but both exhibit less skill than climatology after 14 days of forecast lead time. The superiority of ECC-Q on the extended lead times does not prevail for the variogram skill score and its performance is similar to the other forecasting methods. In preliminary tests with the Schaake Shuffle, we observed no significant differences in multivariate performance with respect to ECC, based on the ESS and VSS (Figure 6.6, purple line covered by red line).

Analysing the significance of the ECC-Q skill scores against the other forecasting methods, by using a Diebold-Mariano test (Figure 6.7), gives a clearer insight of the actual differences of the skill scores. Using the energy score, ECC-Q does perform better than the raw and Z500 bias-corrected forecasts on all lead times. These results are significant for all lead times against the Z500 calibration and significant for all lead times, except day 5–15, against the raw forecasts. Comparing the energy score of ECC-Q against climatology it becomes apparent that ECC-Q has significant better scores until lead time 16 days. When using the variogram score, ECC-Q performs significantly better up to 8 days of lead time. Comparing against the raw forecasts, the variogram score of ECC-Q is significantly better until 3 days. Against climatology, the score of ECC-Q is significantly better until lead time 12 days.

In conclusion, the multivariate comparison of the results aligns with the findings from the univariate comparison. Post-processing using EMOS-G and ECC-Q demonstrates its competitiveness with the pre-

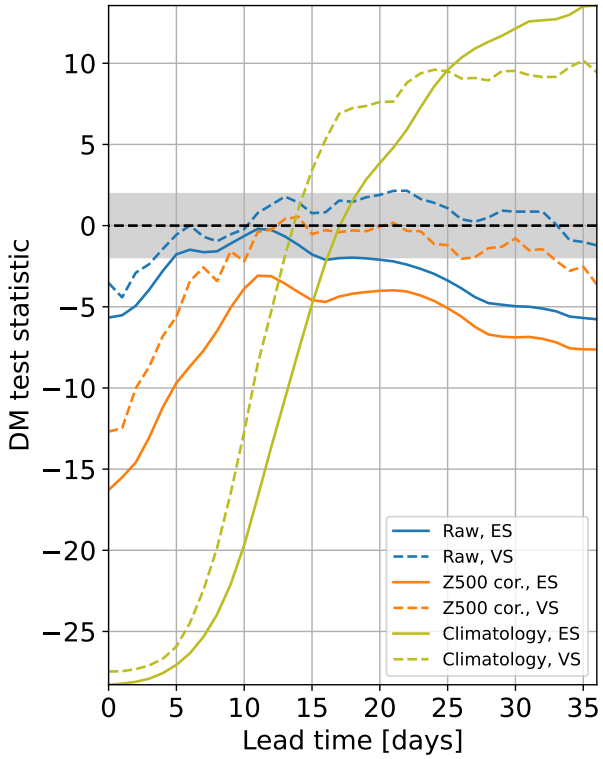


Figure 6.7.: Diebold-Mariano tests based on the multivariate scores (energy score (solid) and variogram score (dashed)) for equal predictive performance of ECC-Q and the raw forecast (blue), the Z500 bias-corrected forecast (orange), and the climatological reference forecast (olive green).

processing method of Z500 calibration and continuously outperforms it using the energy skill score and for most lead times using the variogram skill score. Restoring the multivariate dependency structure has been shown to be a crucial aspect of the post-processing method, as it leads to a substantial improvement in multivariate performance compared to the univariate post-processing (EMOS-Q), ensuring a more accurate representation of the true relationship between weather regime indices of each ensemble member. However, it is important to acknowledge that the multivariate forecast skill in the subseasonal range for all forecasting methods, evaluated over the entire testing period, is inferior to climatology.

6.4. Sensitivity of post-processed weather regime forecasts on training data availability

The main setup of our analysis involves combining two ECMWF model cycles (Cy46R1 and Cy47R1) to train and evaluate on a large dataset. While this approach allows us to test the potential performance limit of our post-processing method, it does not fully represent the data amount available in an operational forecasting scenario, where training solely relies on reforecast data from the operational model cycle. To address the issue of available reforecast data for training, we test the sensitivity of EMOS+ECC trained on two variants of a reduced set of reforecasts.

First, we test for a lower initialisation frequency of forecasts by splitting the training data into the respective forecast cycles, Cy46R1 and Cy47R1. Second, we test for the importance of interannual variability

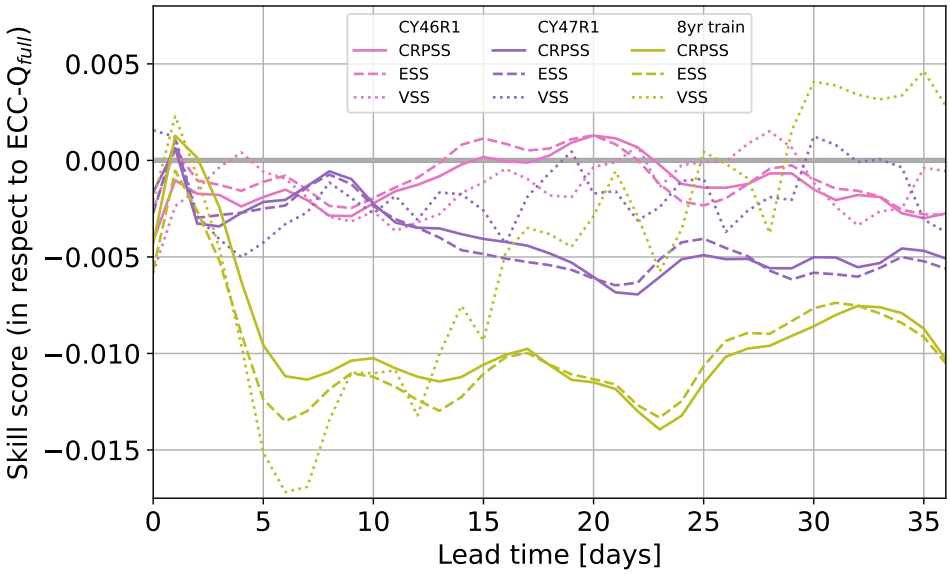


Figure 6.8.: Comparison of performance when reducing training data to the cycles CY46R1 (pink) and CY47R1 (purple), separately, as well as using only the 8 most recent years of both cycles (green) together. We compare the performance of these method setups using skill scores (CRPSS: solid lines, ESS: dashed lines, VSS: dotted lines) with respect to the ECC-Q method trained on both forecast cycles and the full training period of 16 years. Verification is performed on the joined forecasts of both cycles, as in the previous sections.

in the training period by keeping both forecasting cycles but reducing the amount of training years to the 8 most recent years of the entire training period (June 2007 to May 2015). The testing period is identical to the previous sections, using the 899 forecasts from June 2015 to May 2020, combining data from both forecast model cycles. To ensure a fair comparison, we exclude data from April 26 to July 15 of each year for Cy47R1, as this cycle did not run a full year operationally, hence training EMOS with a 31 day rolling window might not be possible at all or only on a minimal set of forecasts. We now directly compare the performance of the ECC-Q setups with a reduced training period against the ECC-Q setup with the full training set (Figure 6.8) for the skill scores: CRPSS (solid lines), ESS (dashed lines), and VSS (dotted lines).

ECC-Q trained on Cy46R1 (pink), which ran for more than a full year, exhibits skill scores nearly identical when compared to the combination of both forecast cycles (skill score values around 0). ECC-Q trained only on Cy47R1 (purple) exhibits slightly lower skill scores than the ECC-Q trained on Cy46R1, which potentially is due to the unavailability of training data for 26 April to 15 July. When training on an 8-year long training period (green) the performance is in general worse than ECC-Q trained only on one forecast cycle and consequently training on the full training data. This indicates that accounting for the interannual variability in the training period is more important than accounting for the initialisation frequency of the forecasts.

In a simplified operational scenario for end-users, it may be beneficial to define a categorical weather regime index by assigning the weather regime with the highest weather regime index (if it is above 1.0) to the corresponding initial date plus lead time (see Chapter 4.4 or Büeler et al. (2021) for a more detailed

definition of the categorical weather regime index). With this approach, our results hold when analysing the forecasts using a categorical forecast skill score, namely the Brier skill score (not shown here).
In conclusion, the post-processing approach we evaluated in our study could be directly applied to operational ECMWF forecasts without considerable losses in skill compared to the results shown above. Training the ECC-Q across many years of forecasts is of greater importance than training the ECC on forecasts with higher initialisation frequency. Our findings remain robust when using a simplified categorical weather regime index instead of the 7-dimensional weather regime index.

6.5. Conclusions and discussion

The present study explores the potential of a statistical post-processing technique that combines ensemble model output statistics and ensemble copula coupling to enhance the forecast skill of multivariate probabilistic weather regime forecasts. Following the approach of Grams et al. (2017), we employ a year-round 7-dimensional weather regime index that identifies four anticyclonic and three cyclonic regimes. The IWR represents the projection of 500-hPa geopotential height anomalies onto the mean anomaly patterns of the seven distinct weather regimes.
Our approach involves the computation and post-processing of weather regime indices, based on the Z500 field obtained from ECMWF's sub-seasonal reforecast ensemble data, utilising model cycles Cy46R1 and Cy47R1. The outcomes of this process are validated against ERA5 reanalyses. To enhance the accuracy of the raw multivariate probabilistic weather regime forecasts, a combined approach of EMOS and ECC is employed as part of the post-processing procedure.
Biases in the raw IWR forecasts can be directly traced back to biases in the Z500A fields. EMOS can effectively correct a portion of these biases and systematic forecasting errors, and thus improves univariate forecasting skill scores on all lead times. When evaluating EMOS in an univariate context against the current practice of using Z500 bias-corrected fields (where Z500A are computed against a model climatology rather than the ERA5 climatology), significant improvements are observed across all lead times and for all weather regimes. The forecast skill horizon, which is defined as the lead time until which the $\mathrm{CRPSS}_{\mathrm{clim}}$ exceeds a specific threshold, indicates that EMOS outperforms the Z500 bias-corrected and the raw forecast for a range of skill horizon thresholds ($0 \leq \mathrm{CRPSS}_{\mathrm{clim}} \leq 0.4$). When the threshold is set at 0.1, the mean forecast skill horizon across the seven weather regimes in the EMOS process is 14.5 days. This represents an improvement of 1 day compared to the raw forecast and 1.2 days compared to the Z500 bias-corrected forecasts.
The multivariate dependency structure of the IWRs is lost when post-processing ensemble forecasts in an univariate manner with EMOS. Hence, restoring this structure through ECC is crucial in multivariate post-processing. The effectiveness of ECC is evident when comparing multivariate skill scores of the univariate EMOS and multivariate ECC forecasts. The enhancements of ECC compared to EMOS become evident starting from a lead time of 5 days when considering multivariate skill scores (ESS and VSS).

Consistent with the univariate comparison against the Z500 bias-corrected forecasts, the multivariate comparison also demonstrates the superiority of the ECC process. Specifically, ECC significantly outperforms the Z500 bias-corrected forecasts in terms of the energy score across all lead times and for the variogram score up to a lead time of 8 days.

The EMOS-ECC process exhibits little sensitivity to the initialisation frequency of reforecasts in the training period but displays a more pronounced sensitivity to the interannual variability in the training data. Nonetheless, the skill achieved by EMOS+ECC when trained on a modified training data set surpasses the current practice of calibrating the Z500 forecast field prior to assigning the weather regime index.

In summary, the statistical post-processing approach of combining EMOS+ECC consistently outperforms the Z500 bias-corrected forecasts. Not only is this approach computationally efficient, but it also provides a compelling alternative due to its ease of implementation and ability to deliver comparable or superior forecasting skill. Additionally, it is versatile, being applicable to both on-the-fly and fixed reforecast configurations. Our findings remain robust when using a limited training data set. Capturing the interannual variability in the training data set is of greater importance than including more frequent initial times. Further, our findings also remain robust using a simplified categorical weather regime index instead of the 7-dimensional weather regime index.

In line with previous studies, such as Schefzik et al. (2013) on pressure, Scheuerer and Hamill (2015) on wind speed, and Schefzik (2017) on temperature, the EMOS-ECC-Q approach is also on the 7-dimensional weather regime index comparable or superior to other methods. Similar to the findings by Schefzik (2017), we observe that the individual EMOS-Q ensemble lacks representation of dependence structures (Figure 6.6 EMOS-Q vs. ECC-Q), leading to weaknesses in multivariate scores like the energy and variogram score. However, we address this limitation by combining the univariate EMOS-Q with the multivariate ECC-Q post-processing step, effectively restoring the dependency structure and enhancing predictive skill.

Although the EMOS-ECC approach outperforms the Z500 bias-corrected approach, it is important to note that on average skilful forecasts of the daily weather regime index are primarily limited to the medium-range, typically up to 15 days. This limitation is dependent on factors like the season, specific weather regime, and the state of the atmosphere. Similar findings were reported by Büeler et al. (2021) using the categorical weather regime index definition based on Z500 bias-corrected weather regimes and the Brier skill score. The limited skill horizons across all methods largely stem from the intrinsic predictability limit of the atmosphere and model deficiencies. All methods solely rely on Z500 ensemble forecast. The marginal improvements in forecast skill of post-processed probabilistic weather regime indices prompt the question of whether these improvements propagate into downstream applications (e.g., energy or hydrological forecasts) by utilising post-processed forecasts rather than raw forecasts. Further research is needed to address this question.

In practical applications of the EMOS-ECC post-processing, we believe that it is crucial to consider both the multivariate outcomes from ECC and the Gaussian distributions from the univariate EMOS step, weighing these outcomes based on the specific application. This approach ensures a comprehensive and accurate interpretation of the forecast. Thanks to the Gaussian distributions in the EMOS post-processing step, this method can be adapted to forecasting models with varying numbers of ensemble members in reforecasts (used as training data) and operational forecasts (e.g., 11 vs. 101 ensemble members in the ECMWF forecast cycle CY48R1).

During certain atmospheric conditions – so-called windows of opportunity – such as a weak stratospheric polar vortex or specific phases of the Madden-Julian Oscillation, the forecast skill horizon for weather regimes may extend (e.g., Büeler et al., 2021). Building on studies investigating predictive skill in mid-latitudes and teleconnection patterns (Ferranti et al., 2018; Lee et al., 2019; Mayer and Barnes, 2020), we propose that incorporating additional predictors representing key atmospheric modes of variability into post-processing methods could enhance weather regime forecasts at extended lead times. Neural networks offer a promising approach for effectively integrating these features (Rasp and Lerch, 2018; Vannitsem et al., 2021).

To explore this, we systematically define and analyse windows of opportunity for forecasting Greenland Blocking three weeks in advance. Furthermore, we employ neural networks to enhance the predictive skill of weather regimes at this extended range.

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

In the previous chapter, we applied statistical post-processing techniques to improve the skill of weather regime forecasts, extending the forecast skill horizon by approximately one day to 14.5 days. However, average daily forecast skill beyond two weeks remains limited and forecasts only marginally outperform climatology. Due to the atmosphere's chaotic nature, temporal aggregation may be necessary to achieve skilful forecasts at subseasonal lead times. In this chapter, we use a weekly aggregated metric – the weather regime activity, introduced in Chapter 4.4.4 – to capture the probability of weather regimes within a forecast week.

Previous research (e.g., Büeler et al. (2021)) suggests that forecast skill is not uniform across all forecasts but depends on the prevailing atmospheric conditions, such as the strength of the stratospheric polar vortex or the phases of the Madden-Julian Oscillation. We therefore focus on windows of opportunity: periods when the statistical likelihood or forecast skill of a specific regime is enhanced under certain atmospheric conditions.

Prior to the analysis in this chapter, we systematically introduce the concept of windows of opportunity (Section 7.1) and the computational concepts of an atmospheric-conditioned climatology and atmospheric-conditioned climatological forecast (Section 7.2). We then systematically identify MJO and SPV phases associated with improved predictability of Greenland Blocking at a three-week lead time (Section 7.3). Greenland Blocking is of particular interest, as it is associated with cold Dunkelflauten events in Germany during winter (see Chapter 5). Next, we incorporate climatological knowledge of the atmospheric state to develop an atmospheric-based climatological forecast (Section 7.4). Finally, we integrate this atmospheric state information with numerical weather prediction outputs using a fully connected neural network to enhance weather regime forecasts (Section 7.5).

This chapter focuses exclusively on the extended winter period (November to March, NDJFM). Sections 7.3 and 7.4 analyse Greenland Blocking at a three-week lead time, while Section 7.5 expands the scope to all seven weather regimes while maintaining the same forecast horizon.

This chapter is based on and slightly adapted from Mockert et al. (2025), with the working title: "Windows of opportunity in subseasonal weather regime forecasting: A statistical-dynamical approach."

7.1. Definition of windows of opportunity

Windows of opportunity integrate knowledge of the atmospheric state at the time of initialisation with the likelihood of a particular event – specifically, WR activity – occurring after a set time period. 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. In the case of the MJO, for example, there are eight active phases and one neutral (inactive) phase. A more detailed explanation of how MJO and SPV phases are computed can be found in Section 7.2.

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 4.1 and Table 4.2. 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.

7.2. 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 7.4).

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, 2025a). 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.

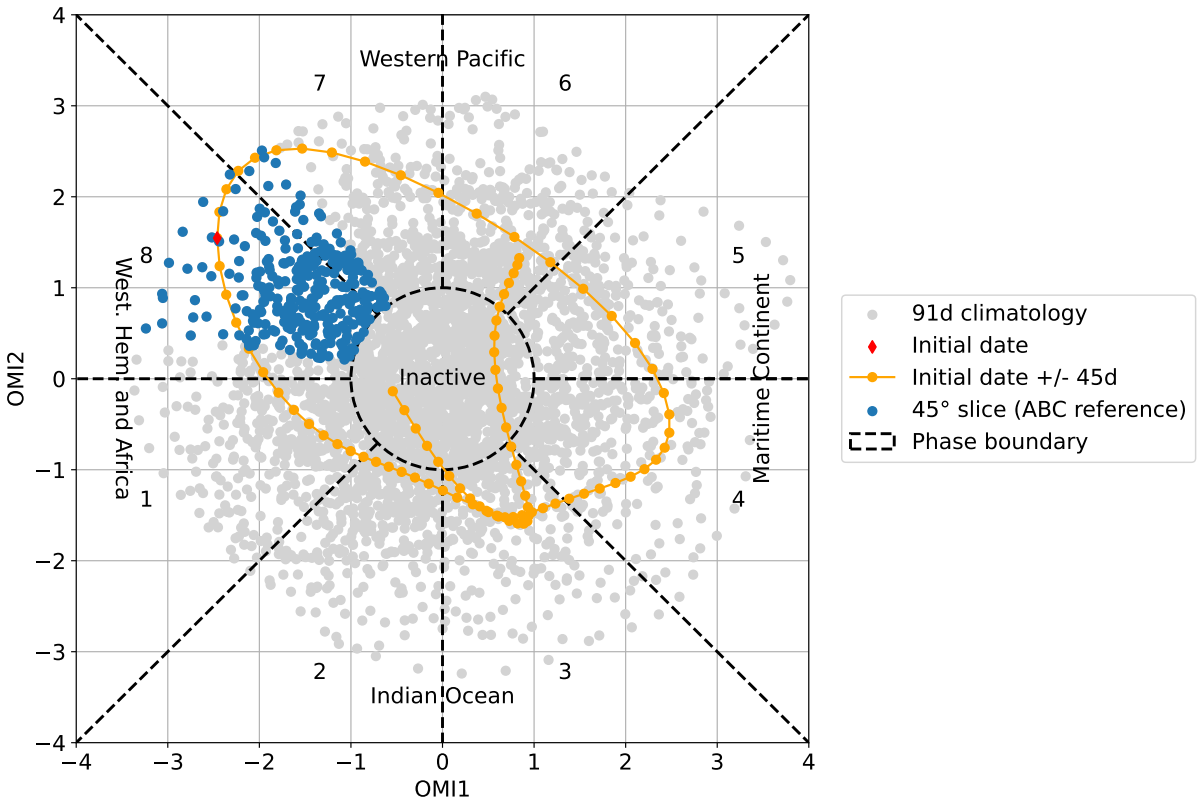


Figure 7.1.: Phase diagram of the outgoing long-wave radiation Madden-Julian Oscillation index. The first and second principal components derived from empirical orthogonal function analysis of filtered OLR data are given on the x- and y-axis. Further the phases of the MJO are indicated with dashed lines and labelling. For visualisation how the atmospheric-conditioned climatological forecast works, one initial date, 10 February 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.

The sliding atmospheric-conditioned climatological forecast for the MJO is generated as follows (illustrated in Figure 7.1): 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.

7.3. 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 7.1).

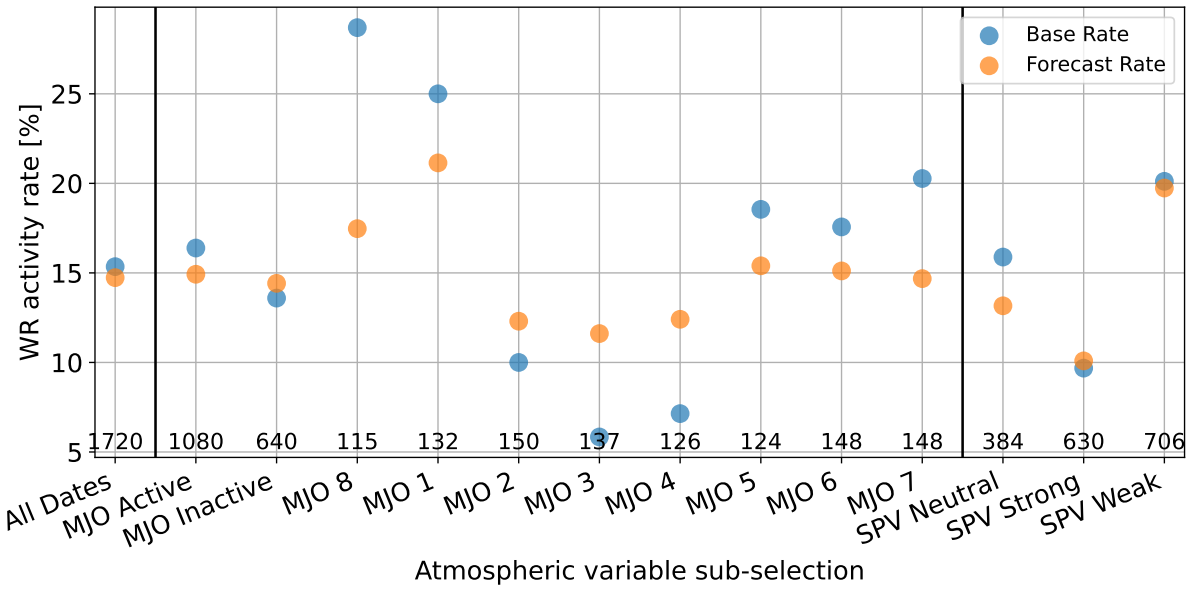


Figure 7.2.: Base rate (blue dots) and forecast rate (orange dots) for weekly mean Greenland Blocking activity ($WRact_{mean,GL}$) in the extended winter period (November–March). These rates are split into subsets based on phases of the MJO or SPV with a lead time of three weeks. The size of the subset 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 7.2). For an analysis across all forecasting weeks see Figure C.1 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 7.1). 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 subseasonal 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 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 7.2 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 7.2). 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 7.1, 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 7.3): $R'_{state} = R_{state} - R_{ref}$, 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 7.2, 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

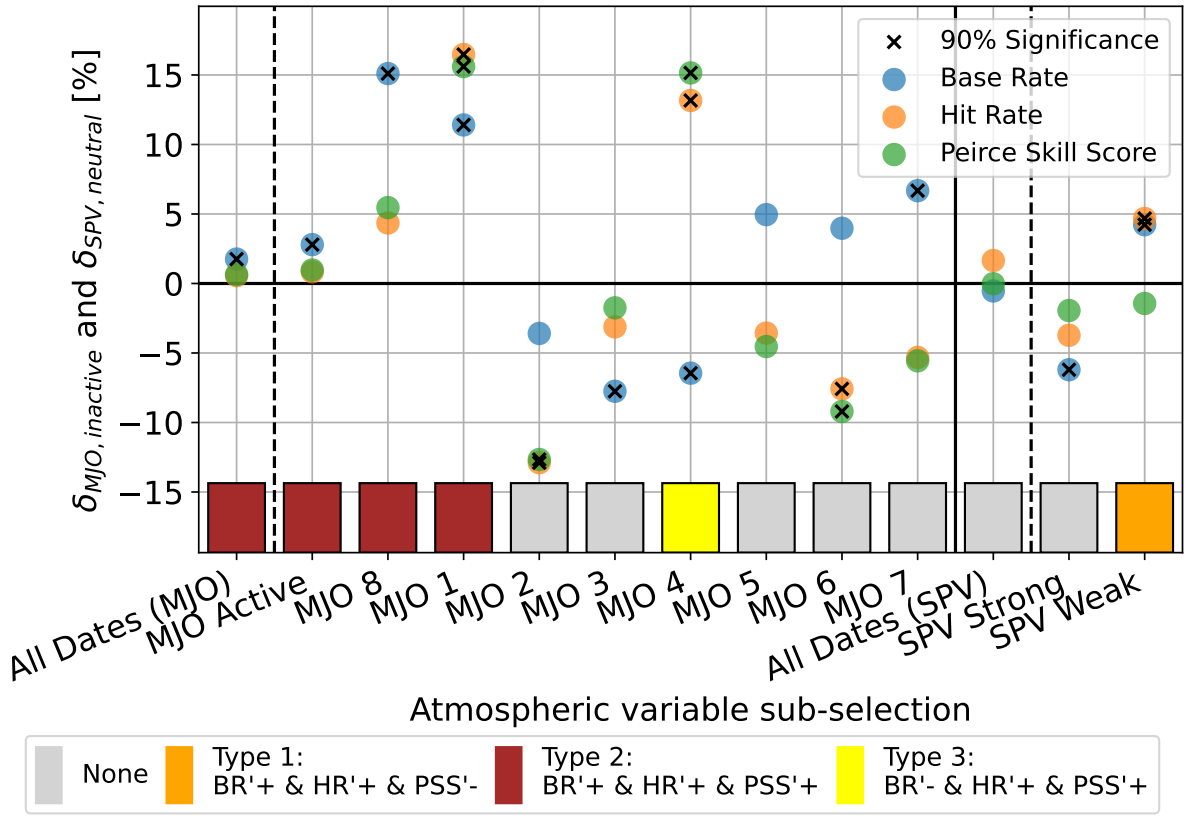


Figure 7.3.: Indication of windows of opportunity for Greenland Blocking activity in week 3 by the base rate, hit rate and Peirce skill score anomalies and colour-coded indication. Similar to Figure 7.2 the metrics are separated by the state of the MJO and the SPV. Here, the base rate, hit rate and Peirce 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.

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.

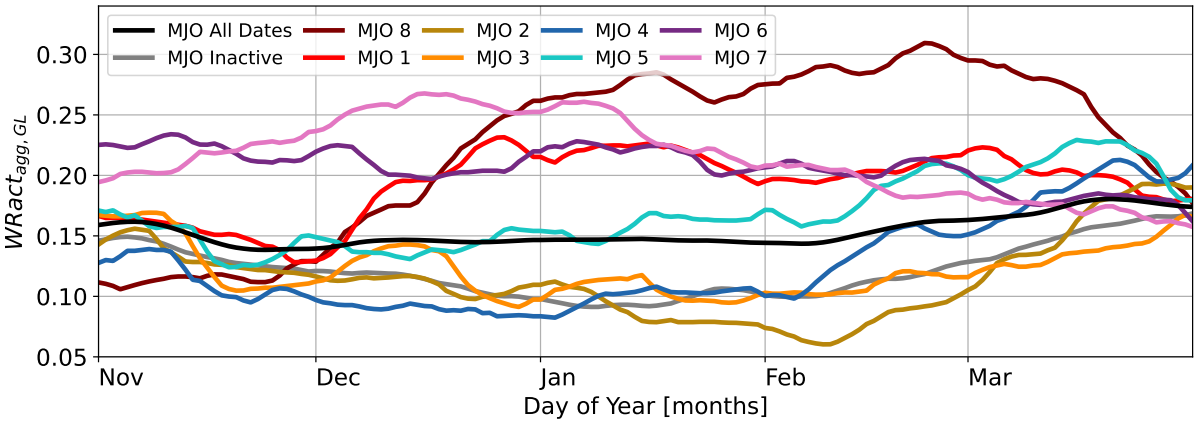
In summary, the frequency of $WRact_{mean,GL}$ (base rate) varies across different MJO and SPV states. It is particularly high (significantly different to the neutral phase) for MJO phases 8 and 1, as well as after weak SPV conditions. The magnitude of the forecast rate is generally too low. MJO phases 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.

7.4. 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 7.4). 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 7.4a), 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 7.4b), 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 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 7.2 for further details).

(a) MJO-conditioned climatology.



(b) SPV-conditioned climatology.

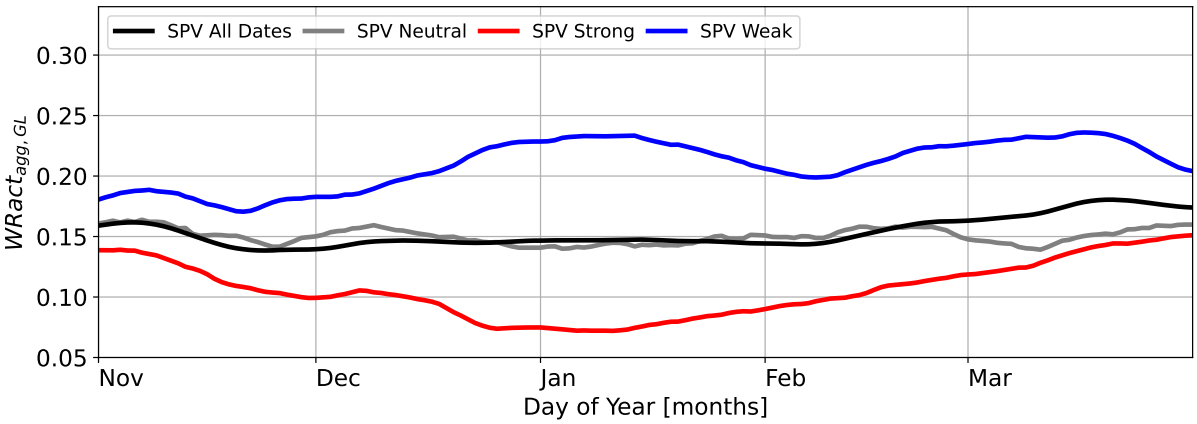


Figure 7.4.: Atmospheric-conditioned climatology for $WRact_{agg,GL}$ with a time delay of three weeks after specific states of the (a) MJO and (b) SPV for the extended winter period (November–March). The black line in both sub-figures indicate the 91-day running climatology without any atmospheric-conditioned sub-selection.

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 7.5).
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 7.5a). For the second question, only those forecasts at a fixed lead time of

(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.

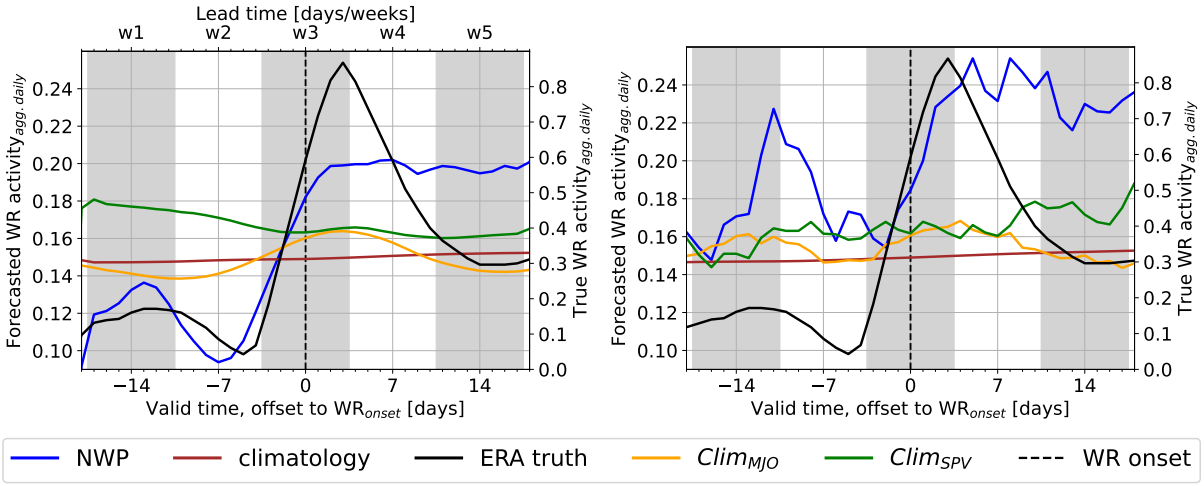


Figure 7.5.: Mean composite of different forecasts at valid times centred around observed Greenland Blocking onsets(black dashed vertical line). Due to values on different scales, forecast values for the $WRact_{agg}$ are indicated on the left y-axis, the mean of the ERA5 $WRact_{agg}$ are indicated 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.

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 7.5b). The mean composite of forecasts in Figure 7.5 is computed centring all events around the GL onset. For the individual events see Figures C.2 and C.3 in the Supplementary Material.

The true (ERA5) $WRact_{agg,GL}$ (black line, right y-axis) is identical in both sub-figures of Figure 7.5. 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 C.2). 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 7.5a), 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 C.2). 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 7.5b), 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.

7.5. 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 4.7.3) 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 C.1 and C.2 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 4.7.3) to identify the most relevant predictors. Once the predictors yielding the highest skill are determined, models are re-trained 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 C.4). The second most important predictor is the NAO index at initialisation. As GL strongly correlates with 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 Section 2.4 and in the previous results for SPV, Section 7.3 and 7.4). 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 7.6) 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 NAE 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.

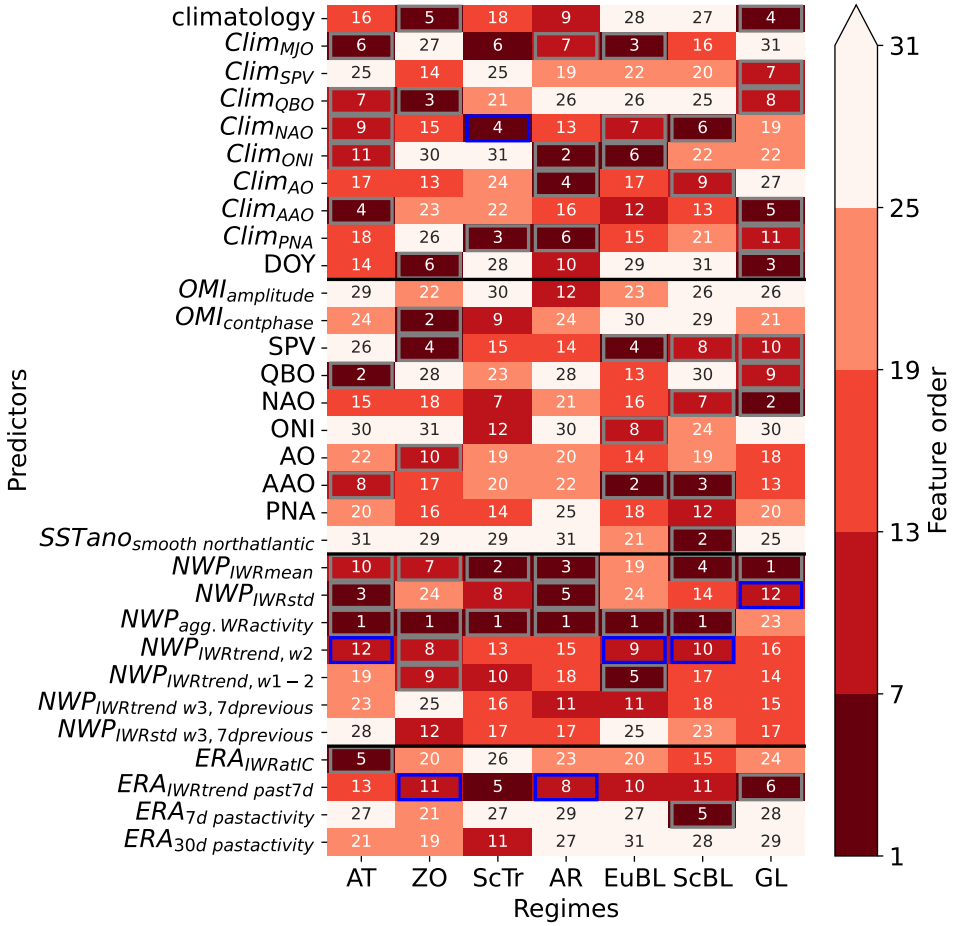


Figure 7.6.: Summary of the predictor order for all-predictor neural networks (NN_{all}). The predictor order is shown for all seven weather regimes. Blue boxes indicate the last predictor which is considered due to the best MSE for the predictor combination of all predictors with a lower predictor order number (indicated with grey boxes). A detailed explanation of the predictors can be found in Table C.1 and C.2.

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 (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 NAE 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 7.7). 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,

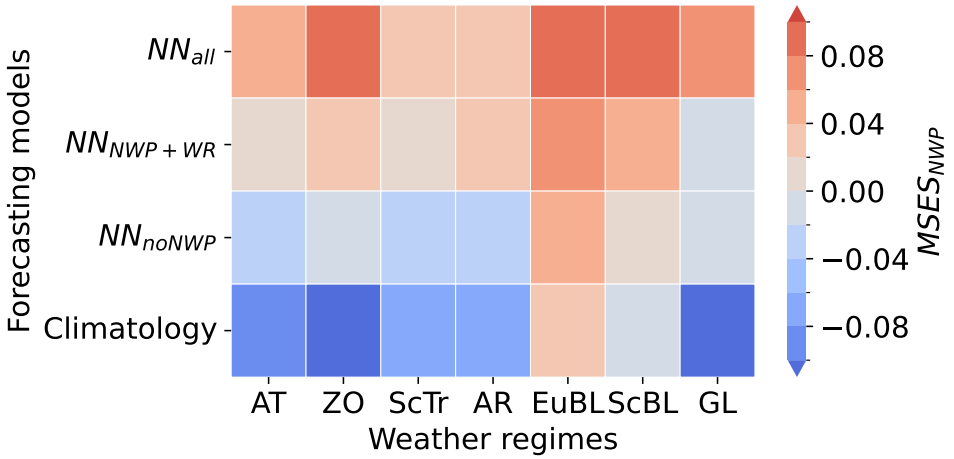


Figure 7.7.: Performance of the different neural network and climatological forecasts in comparison to the NWP forecast for $WRact_{agg}$. As performance measure, the MSE is calculated for each model and weather regime and put into perspective against the MSE of the NWP model (MSE skill score).

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 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 C.5), 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 4.1a and Table 4.2), including accuracy (hit rate), false alarm rate, precision, and F1-score (Figure 7.8). 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

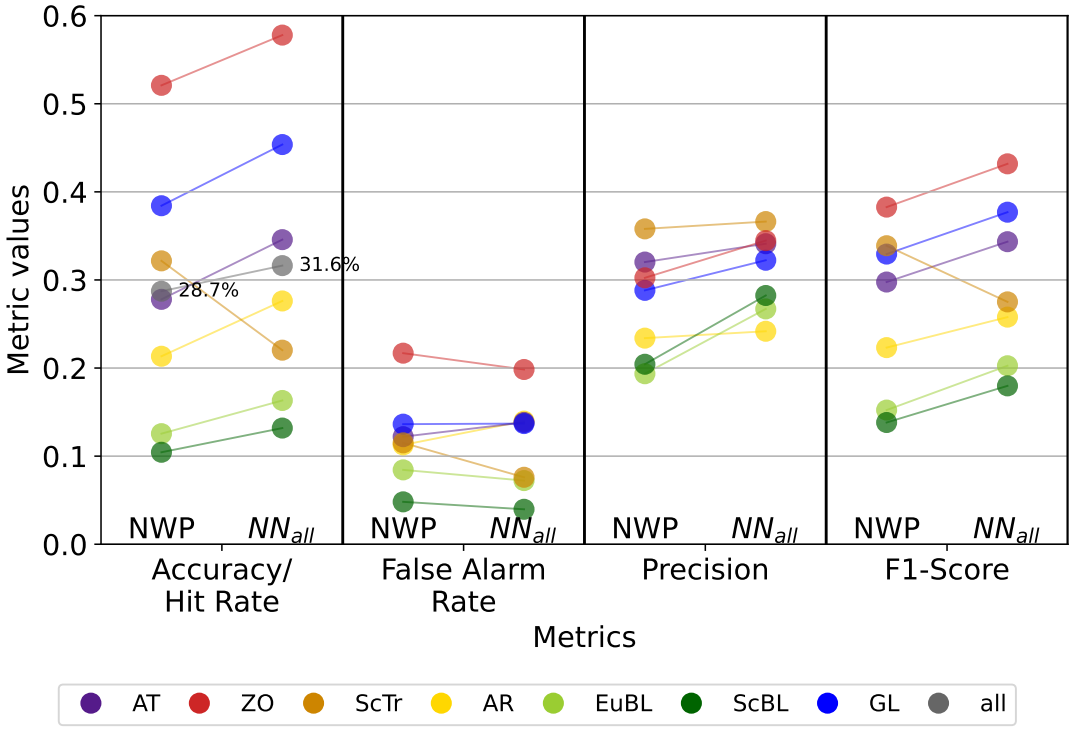


Figure 7.8.: Verification metrics comparing the skill of the NWP model and the all-predictors neural network in predicting the $WRact_{max}$ at a lead time of three weeks. Four metrics, the hit rate, the false alarm rate, the precision, and the F1-score are visualised (x-axis). Further together with 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.

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.

7.6. Conclusions and discussion

Various studies documented the connection between subseasonal sources of predictability (e.g., MJO or SPV) and the large-scale circulation in the NAE 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 subseasonal 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 percentage points, 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 subseasonal 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 subseasonal 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 subseasonal 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 NAE weather regimes, providing valuable improvements to NWP performance – particularly in situations where traditional NWP models struggle.

8. Case study: Integrating all research components

The previous three chapters (Chapters 5–7) addressed individual aspects of why weather regime forecasting is a crucial part for the energy sector when forecasting on the subseasonal time scale, as well as how to improve the forecasts and make them useful for the energy sector. In this chapter, we integrate all research aspects into one case study to illustrate the practical relevance and coherence of the findings. It serves as a concrete example of how the developed methods can be applied in an operational setting and provides a lens through which to critically assess both the capabilities and the current limitations of current weather regime forecasting techniques.

In the case study, the winter of 2009/2010 is analysed, with a special focus on an unusually-persistent Greenland Blocking episode lasting 27 days with two local maxima in the weather regime index – making it the fourth longest Greenland Blocking event in the period from 1979 to 2022. The Greenland Blocking episode (from 14 December 2009 to 10 January 2010) is preceded by a brief Scandinavian Blocking episode from 7 to 13 December 2009.

8.1. Cold Dunkelflauten

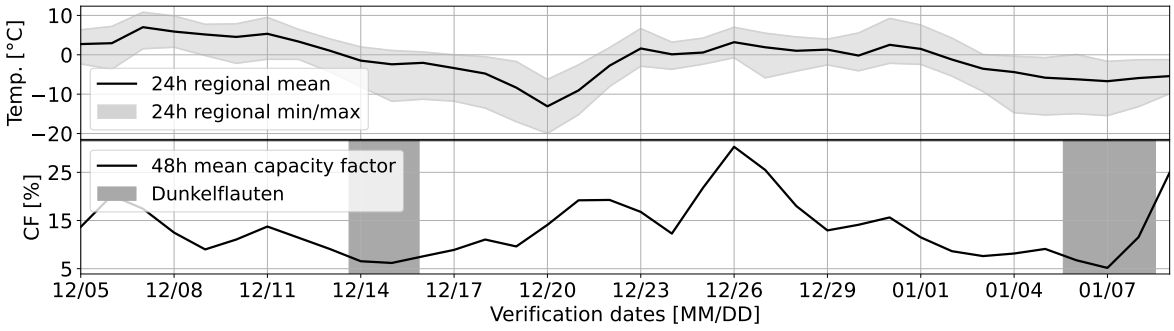


Figure 8.1.: Temperature and capacity factor during a Greenland Blocking event (December 2009/January 2010). The top panel shows the 24-hour running mean of 2 m temperature over Germany. The bottom panel presents the 48-hour running mean of the wind and solar capacity factor for Germany combined. Shaded intervals in the bottom panel highlight Dunkelflauten periods.

Two Dunkelflauten occurred during the Greenland Blocking episode (grey boxes in bottom panel in Figure 8.1), supporting the broader finding that Dunkelflauten tend to occur during longer-lasting weather regimes. However, the timing of the two Dunkelflauten events relative to the weather regime life cycle diverges from the mean picture. The first occurs shortly after the onset of Greenland Blocking, while

the second takes place just before its decay, suggesting that the timing of the Dunkelflaute relative to the regime life cycle varies somewhat from case to case.

The temperature evolution (top panel in Figure 8.1) during this case is characteristic for Greenland Blocking Dunkelflauten. A pronounced temperature drop precedes the first event, with daily spatial-mean averages falling from 7°C to –13°C and grid-point-wise minimum temperatures in Germany reaching –20°C by 20 December. After a milder period, from 23 December to 3 January, a second, less severe cooling period accompanies the second Dunkelflaute, with temperatures declining from 2°C to around –7°C. These observations align well with the general characterisation of Greenland Blocking Dunkelflauten as cold events.

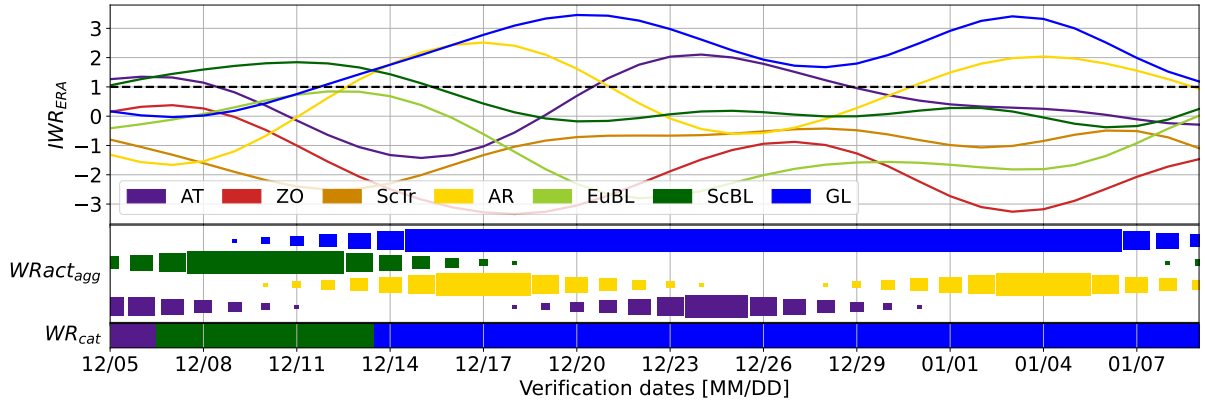


Figure 8.2.: Weather regime metrics during a Greenland Blocking event (December 2009/January 2010). The top panel shows the ERA5-based weather regime index. The middle panel presents the corresponding weather regime activity. The bottom panel indicates the categorical regime assignments.

A more detailed examination of the IWR (Figure 8.2) for this case highlights an important nuance: while the categorical regime classification captures the broad circulation, secondary regime contributions can significantly modulate conditions within these phases. This is consistent with findings by Gerighausen et al. (2024), who show that the coldest days within Greenland Blocking periods are typically associated with secondary contributions from the Atlantic Ridge, whereas warmer days correspond to influences from the Atlantic Trough. This emphasises the value of complementing categorical regime classification with a continuous or mixed-regime perspective when assessing weather impacts on energy systems.

8.2. Potential and limitations of statistical post-processing

Forecasts initialised on 9 December (top panel in Figure 8.3) – four days before the Greenland Blocking onset – perform well for predicting the onset of the Greenland Blocking. Both raw and post-processed forecasts closely track the IWR evolution, with the ensemble mean representing the amplitude of the first maximum rather close, though forecasting the maximum too early. Here, post-processing strengthens the

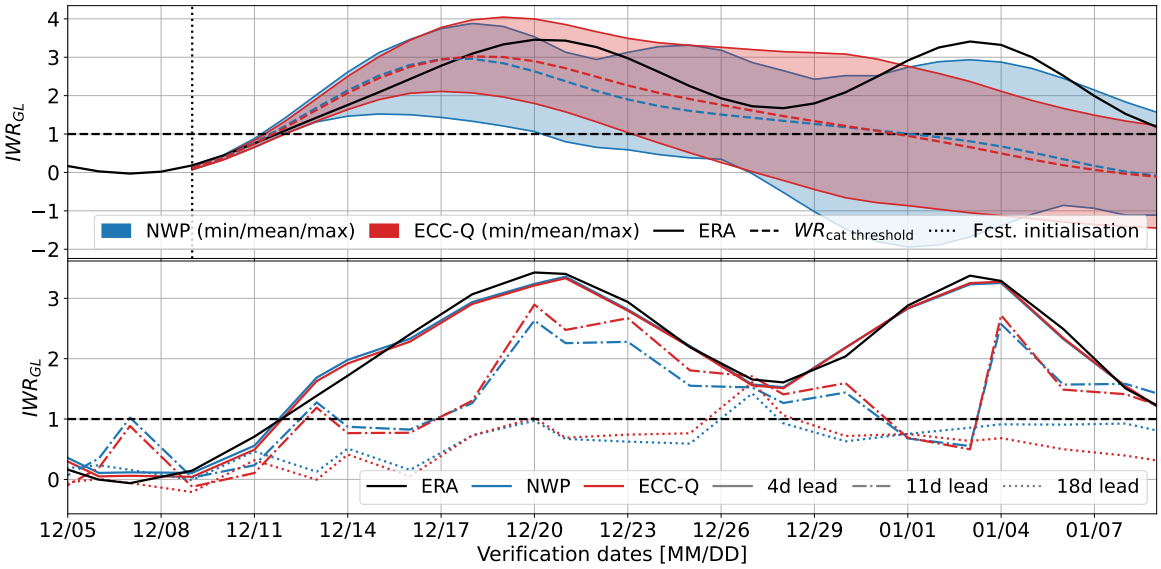


Figure 8.3.: Raw and statistically post-processed IWR forecasts during a Greenland Blocking event (December 2009/January 2010). In the top panel, Greenland Blocking forecasts are initialised on 9 December 2009, with lead times up to 36 days for the raw NWP and ECC-processed forecasts, along with the verifying ERA5 reanalysis. The bottom panel shows IWR forecasts at fixed lead times (4, 11, and 18 days), for both raw and ECC-processed models.

IWR and shifts the forecast toward higher IWR values during the maximum phase, slightly improving persistence and intensity.

The benefit is particularly evident when comparing forecasts at different lead times (4, 11, and 18 days) in the bottom panel of Figure 8.3. At four days lead time (solid lines), the raw forecast aligns well with the observed IWR, leaving little room for improvement. At eleven days lead time (dash-dotted lines), the raw forecast indicates a signal for the first Greenland Blocking maximum and the post-processing enhances the prediction of the first maximum.

Forecasts with an 18 day lead time (dotted lines), both raw and post-processed forecasts, remain in the ensemble mean close to climatology, with values only for a few days exceeding an IWR of 1, unable to reproduce the high IWR values. This suggests that while a signal of the upcoming regime can be present even beyond the average forecast skill horizon of 13.5 days, confidence in strong IWR projection is limited at these lead times. Post-processing adjusts the ensemble mean, slightly correcting low IWR values downward and high values upward, but the corrections are too modest to recover the full intensity of the Greenland Blocking signal.

Together, these findings highlight both the potential and the limitations of statistical post-processing. While the average skill increases, especially in the medium-range, the ability to forecast specific events – such as strong and persistent blocking – on a daily basis at long (subseasonal) lead times is constrained by the inherent limits of the raw ensemble and the conservative nature of post-processing corrections.

8.3. Windows of opportunity as affirmation for the end-user

Prior to the onset of the blocking episode, the MJO progressed into phase 8, and a sudden stratospheric warming event triggered a transition into a weak SPV state just at the onset of the blocking episode. The MJO phase 8 and the weak SPV state were both identified as climatological and model windows of opportunity (Chapter 7.3). The combination of these two atmospheric modes might be a reason why the specific NWP forecast initialised on 9 December 2009 (four days before the onset) performs well. It not only shows IWR skill up to 20 days (top panel in Figure 8.3), but also predicts Greenland Blocking activity with a probability exceeding 50% up to 23 days ahead (not shown). This highlights how leveraging knowledge of large-scale atmospheric modes can enhance the trust of an end-user in subseasonal forecasts. Although this forecast, initialised four days before the Greenland Blocking onset, shows good performance, the general skill of the 18 day/week 3 NWP forecasts is worse, as it does not confidently predict the onset and maintenance of the Greenland Blocking (dotted line in Figure 8.3, bottom panel).

8.4. Statistical-dynamical approaches as decision supporting tool

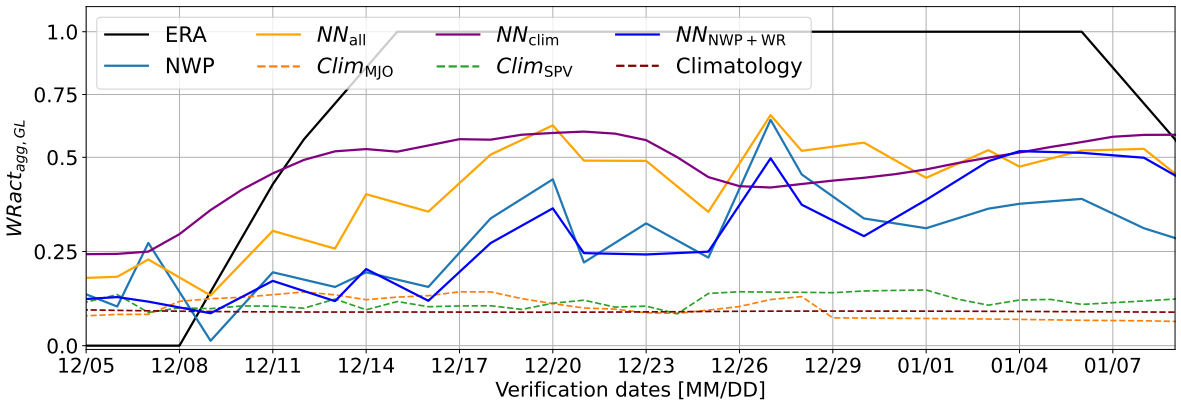


Figure 8.4.: Aggregated weather regime activity forecasts during a Greenland Blocking event during (December 2009/January 2010). Weather regime activity forecasts at three-week lead for the NWP model, the atmospheric-conditioned climatological forecasts, and neural networks. The figure highlights the added value of incorporating climatological and atmospheric predictors beyond the raw NWP outputs.

Greenland Blocking activity forecasts by the NWP model with a lead time of three weeks (light blue line in Figure 8.4) only show modest Greenland Blocking activity for the first maximum with values of up to 44%. Though, values over 40% are only achieved for one of the forecast initialisation, with previous and latter forecast initialisations predicting values well below 40% (34% and 22%, respectively). The MJO- and SPV-conditioned forecasts (dashed orange and green lines) show a stronger-than-climatology signal during the first maximum and the SPV-conditioned forecast also shows an elevated weather regime activity across the second maximum (Figure 8.4), suggesting a higher-than-usual likelihood of Greenland Blocking.

The neural network trained only on NWP data and weather regime activity before forecast initialisation (dark blue line in Figure 8.4) does not improve the forecast in week 3. However, the model trained exclusively on climatological and atmospheric predictors – with no access to the NWP forecast – successfully identifies increased Greenland Blocking activity throughout the regime's lifespan (purple line), with values reaching 60% likelihood for an activity. This suggests that, particularly at longer lead times, atmospheric precursors still hold valuable predictive power even in the absence of guidance by the NWP model. The full neural network model, combining all inputs (solid orange line), produces an improved forecast compared to the NWP alone, with activities up to 63% rather than 44% for the NWP in the first activity maximum. This underlines the benefits of a hybrid approach that goes beyond simple averaging of individual forecasts.

The hybrid neural network – informed by both atmospheric state and dynamical forecasts – provides a more consistent and confident signal of Greenland Blocking activity, showcasing its utility for subseasonal forecasts.

9. Conclusions, discussion, and outlook

This thesis investigated the role of weather regimes in controlling renewable energy production in Germany and explored both statistical post-processing and statistical-dynamical approaches to enhance the forecast skill of weather regimes.

As the energy transition progresses and the share of weather-dependent renewable sources continues to grow across Europe – particularly in Germany – weather forecasts to subseasonal outlooks become increasingly vital. They are essential not only for daily grid operation and energy trading but also for medium- to long-term planning of energy storage capacities. In this context, the ability to anticipate energy-relevant extremes, such as Dunkelflauten, gains strategic importance. Understanding the atmospheric precursors to such events, as well as the potential to detect them within the forecast, is crucial to improving the forecasts themselves.

This final chapter reflects on the key results and places them in a broader context by revisiting the five central research questions posed in Chapter 3. The chapter concludes with an outlook on future research directions and emerging operational needs.

Conclusions

The first part of this thesis explored the link between large-scale weather regimes and various types of Dunkelflauten events and revealed consistent atmospheric patterns associated with these critical low-energy generation events.

While today's ensemble forecasting systems, such as those operated by the ECMWF, offer valuable probabilistic outlooks of atmospheric conditions on the subseasonal time scale, their skill diminishes with increasing lead time. This is partly due to systematic biases and dispersion errors, especially in relation to regime transitions and strong regime projections.

To address these shortcomings, the second part of this thesis applied statistical post-processing methods – namely Ensemble Model Output Statistics and Ensemble Copula Coupling – correcting systematic biases and dispersion errors in ensemble forecasts. These methods successfully enhanced the skill of weather regime forecasts.

To further improve forecasting skill on the subseasonal timescale, the third part of the thesis developed data-driven statistical models to identify atmospheric windows of opportunity that indicate an increased likelihood of the occurrence of specific weather regimes. In addition, the development of hybrid models – where neural networks combine statistical information with dynamic forecasts – led to tangible im-

provements in forecasting strong positive weather regime projections. The identification of windows of opportunity and the resulting improvements in the forecast underscore the potential of combining physical understanding with modern data science to better anticipate energy-relevant weather extremes. The main research questions posed in Chapter 3 were addressed in Chapters 5–7 and the main findings are summarised and discussed in the following.

RQ-1 Which North Atlantic-European weather regimes cause Dunkelflauten in Germany, especially cold Dunkelflauten? How do the characteristics of these regimes differ from those not associated with Dunkelflauten?

Dunkelflauten in Germany predominantly occurred during the winter half of the year, with a peak in December. This timing coincided with minimal solar power output due to the short daylight hours and the low angle of the winter sun. Consequently, wind power became the decisive factor for such events – particularly when wind speeds drop below typical values for an extended period, especially over the North Sea, which is a critical region for wind energy generation.

These reductions in wind speed were most frequently linked to the presence of a high-pressure system centred over Northern Germany and the Baltic Sea. This synoptic pattern aligns closely with the European Blocking weather regime, which emerged as the dominant circulation pattern associated with Dunkelflauten. In addition to European Blocking, Scandinavian Blocking and Greenland Blocking led to a saddle point in the surface pressure field over Germany, where the pressure gradient equals zero. The position of Germany between a quadrupole of high and low pressure systems resulted in similarly weak wind conditions.

While European Blocking accounted for the majority of Dunkelflauten, Greenland Blocking led to a combination of low wind conditions and below-average temperatures (–4 K), which in turn amplified electricity demand due to increased heating needs. This effect is expected to become even more pronounced in the future as heating systems transition from gas to electric heat pumps, shifting more of the heating burden onto the power grid. A comparison of regime phases revealed that Dunkelflauten were more likely to occur during long-lasting weather regimes, typically during the well-established regime phase rather than during transitions.

RQ-2 To what extent can statistical post-processing improve the probabilistic skill of ECMWF reforecast-based weather regime forecasts, particularly when accounting for systematic biases, dispersion errors and multivariate dependencies?

Statistical post-processing methods improved the forecast skill of weather regime forecasts over the NAE region. This thesis applied a two-step approach comprising an univariate correction via Ensemble Model Output Statistics, followed by a multivariate reconstruction of the dependence structure among the weather regimes using Ensemble Copula Coupling. The EMOS step, which independently

corrected the forecast probabilities of the IWR for each regime and lead time, extended the forecast skill horizon by approximately one day, as measured by the continuous ranked probability skill score with a 0.1 threshold. In absolute terms, the average skill horizon increased from 13.5 days for the raw ECMWF ensemble forecast to 14.5 days for the post-processed forecast. These improvements were evident across all regimes and lead times.

After the multivariate dependence structure, necessary for a multivariate comparison of the forecasts across the weather regimes, was restored by employing ECC, its multivariate added value, measured by the energy and variogram score, compared to the raw forecast was more moderate than the univariate added value of EMOS, measured by the CRPSS. Nonetheless, when the forecast skill was assessed using the energy score, post-processed forecasts outperformed raw forecasts across all lead times, although not always with statistically significant margins. In contrast, the analysis of the variogram score revealed that improvements diminished beyond ten days lead time, suggesting that the structure in the ensemble became harder to recover at subseasonal ranges. A sensitivity analysis further showed that EMOS-ECC performed more robustly when trained on a longer and more temporally distributed dataset, rather than on a shorter period with more frequent forecast initialisations. This finding highlights the importance of a properly chosen training set in order to maximise the performance of post-processing methods.

RQ-3 How should probabilistic forecasts of strong weather regime projections be best constructed for lead times beyond 14 days?

The inability of NWP models to produce on average meaningfully better probabilistic daily IWR forecasts beyond two weeks (already indicated by Büeler et al. (2021); Osman et al. (2023) and shown in Chapter 6) led to developing a new metric aimed at better capturing the dynamical characteristics of weather regimes in a forecast context.

To bridge the gap between the high-dimensional IWR and the drastically simplified categorical regime index – while specifically targeting conditions associated with strong positive weather regime projections – a new diagnostic metric, the weather regime activity, was introduced.

This metric balances interpretability and informativeness by quantifying the probability of strongly positive IWR values for each weather regime, independently, over specific weekly forecast intervals. It is implemented in two complementary forms: the weekly mean weather regime activity, defined categorically (active if the weekly mean IWR exceeds 1, inactive otherwise), and the daily aggregated weather regime activity, defined probabilistically as the fraction of days within a week exceeding the IWR threshold of 1. Crucially, because the activity is computed separately for each regime, simultaneous activities of multiple weather regimes are possible – capturing overlapping regime signals that are lost in the categorical representation (this is visible in the case study in Figure 8.2, comparing the bot-

tom and middle panel). For interpretive consistency with standard approaches, the dominant weather regime activity can also be identified as the regime with the highest activity value.
This new metric served to address the fourth and fifth research questions.

RQ-4 Are there specific atmospheric conditions that increase the forecast skill of weather regimes on the subseasonal time scale?

Focusing on the Greenland Blocking regime in forecast week 3, we employed a window of opportunity approach following the methodology of Specq and Batté (2022). Each ensemble member was treated separately as a deterministic forecast, with the weekly mean weather regime index thresholded at 1 to identify regime activity. Specifically, we assessed how different phases of the Madden-Julian Oscillation and the strength of the stratospheric polar vortex modulated the likelihood and forecast skill of Greenland Blocking in forecast week 3.
We distinguished between climatological and model windows of opportunity. Climatological windows of opportunity indicate an increased base rate of weather regime activity following specific atmospheric conditions. Model windows of opportunity additionally consider the model's ability to forecast that activity. These are categorised into three types: Type 1 presents an increased base rate and hit rate, but no improvement in the Peirce skill score (quantifying the balance between hits and false alarms); Type 2 presents improvements in all three metrics compared to the neutral state; Type 3 presents improvements in hit rate and Peirce skill score, but no enhancements in the base rate.
The analysis revealed that Greenland Blocking activity occurred more frequently three weeks after MJO phases 5–8 and phase 1, as well as after weak SPV conditions – representing climatological windows of opportunity. When we examined forecasts initialised during active MJO phases, we found clear signs of improved predictability. For instance, MJO phases 8 and 1 qualified as model windows of opportunity type 2: both the base rate of Greenland Blocking and the forecast skill improved relative to forecasts initialised in the inactive phase. In contrast, MJO phase 4 represented a type 3 model window of opportunity: although Greenland Blocking was less likely to occur (lower base rate), forecasts predicting its activity initialised during this phase tended to be more often correct, showing improved skill despite lower climatological likelihood.
Forecasts initialised under weak SPV conditions exhibited an improved hit rate but this came at the cost of a higher false alarm rate, suggesting that while the regime may be more predictable, the signal-to-noise ratio remains a challenge, qualifying as a type 1 model window of opportunity.
Overall, these findings support the idea that specific atmospheric precursors can serve as strategic entry points for improved subseasonal forecasts of weather regimes.

RQ-5 How can knowledge of atmospheric conditions at forecast initialisation improve the probabilistic forecast skill of weather regimes?

To visualise and include knowledge about the atmospheric state at forecast initialisation, this thesis introduced a climatological forecasting approach conditioned on large-scale atmospheric indices. Specifically, the method constructs forecasts of weather regime activity by averaging historical outcomes in situations where the atmospheric state – such as the MJO or SPV index – closely resembles the current one. When compared against an unconditioned climatological forecast, this enables the identification of signals linked to particular atmospheric modes. These conditioned forecasts can then complement the NWP forecast: if both the NWP and atmospheric-conditioned climatological forecasts indicate a high likelihood of regime activity, confidence in the prediction increases.

Building on this idea, neural networks were used to merge information from multiple sources into a single, more robust forecast. Three configurations were tested. The first model integrated all available inputs: atmospheric precursors, conditioned climatological forecasts, recent regime activity, and the NWP forecast. The second omitted the NWP component, relying only on climatological and atmospheric-state-based predictors including the recent regime activity. The third used the NWP and recent regime activity but excluded atmospheric indices and climatological forecasts. Across all weather regimes, the full model – combining all sources – outperformed the raw NWP forecast, demonstrating the value of this hybrid approach. As one can expect, the neural network learned most from the NWP forecast, but interestingly, atmospheric precursors also contributed valuable signals.

Notably, while based on previous studies, one might expect the MJO to be the dominant atmospheric predictor. However, the neural network rarely learned from MJO information. Instead, a surprising result was the frequent appearance of the Antarctic Oscillation index (also known as the Southern Annular Mode) as an influential predictor. Although this connection has received little attention in the context of NAE weather regimes, previous studies offer possible explanations: Tachibana et al. (2018) identified correlations between the AAO and its Northern Hemisphere counterpart, the Arctic Oscillation, while Song et al. (2009) found that negative AAO phases precede anomalously high 300 hPa geopotential heights over the NAE region by 25–40 days. These findings align with the AAO emerging as a useful predictor for blocking regimes such as EuBL and ScBL in our models. This could suggest a previously unexplored teleconnection that warrants further investigation. If this connection is validated (for example through modelling studies), it may represent a new avenue for improving subseasonal weather regime forecasts.

Discussion and Outlook

The Greenland Blocking case study from the winter of 2009/2010 (Chapter 8) has served as a powerful example to illustrate how the research themes of this thesis are interconnected. It provided valuable

insight into where current forecasting systems perform well, where they fall short, and how combining different modelling strategies can help overcome those limitations.

Previous studies (e.g., Grams et al., 2017; Wiel et al., 2019b) have emphasised the importance of regime-dependent variability for renewable energy supply, particularly in the context of wintertime blocking. This thesis builds on this foundation by highlighting the dominant role of Greenland Blocking for cold Dunkelflauten in Germany, thereby clarifying the relevance of regime associations for the energy sector. While previous work on Dunkelflauten has often treated weather regimes as static categories (e.g., Drücke et al., 2021; Wiel et al., 2019b), this work demonstrates that the evolution and life cycle of regimes contain predictive information that should not be ignored in subseasonal risk assessment. In terms of methodology, this work advances the post-processing of ensemble forecasts tailored for weather regimes through the application and validation of multivariate ensemble calibration. In contrast to previous approaches that focus on correcting mean Z500 biases (cf. Büeler et al., 2021), the EMOS-ECC framework used here directly accounts for systematic errors in IWR forecasts, providing a computationally efficient and operationally feasible alternative for calibrating regime forecasts. A key contribution of this work is the development of a new metric – weather regime activity – that serves as a middle ground between the complex seven-dimensional weather regime index and the oversimplified categorical regime classification. This metric is particularly useful for detecting strong positive regime projections in weekly forecasts. Finally, by identifying and integrating subseasonal teleconnection signals – particularly the influence of the MJO and SPV on weather regime activity – into neural network-based models, this work extends previous medium-range studies of regime-teleconnection links into the subseasonal timescale (e.g., Cassou, 2008; Vitart et al., 2010). The resulting framework also offers a practical, data-driven tool to support decision-making based on weather regime forecasts.

At the same time, this thesis opened new research questions. Neural networks, as used here, rely strongly on the input data and training size. Predictor variables were initially selected based on meteorological expertise and further refined through testing to reduce noise. The networks themselves were deliberately kept simple, using architectures limited to scalar input variables. However, recent work by Bommer et al. (2025) has shown that more complex architectures, such as long short-term memory (LSTM) networks that take time series of predictors related to the MJO and SPV, can outperform traditional NWP forecasts for certain weather regimes on the subseasonal to seasonal range, even without using the NWP forecasts themselves. These results suggest that more complex models may better capture phenomena like sudden stratospheric warmings and clarify the expected – but not fully realised – relationship between MJO phases and regime predictability in our simpler models.

A particularly intriguing and unexpected result from this thesis is the apparent relevance of the Antarctic Oscillation for forecasting NAE weather regimes. This potential connection has not yet received much attention in literature. If this relationship holds under further analysis and a physical process understanding can be developed, it may open new pathways for enhancing prediction skill in the NAE sector.

Preliminary results (not shown in this thesis) suggest that the methods developed here – including EMOS-ECC post-processing and neural networks for regime activity forecasting – can be successfully transferred to North American weather regimes. We would expect the MJO to have a clearer influence on forecast skill for North American weather regimes than for NAE regimes, as the physical distance to the convective envelope is generally shorter and research shows a more direct impact of the MJO on North American weather (e.g., Zheng et al., 2018; Green and Furtado, 2019; Jenney et al., 2019). Exploring this application in more detail, and comparing results to those obtained for the NAE region, would be a natural next step and provide valuable insights into the generalisability of the approaches.

A further compelling future direction is to more directly link weather regime forecasts to energy-relevant events, such as Dunkelflauten. Is it feasible to derive a new metric that can reliably predict these rare but critical events? Or might it be more effective to post-process established energy-related variables, guided by insights from weather regime forecasts and atmospheric indicators? Addressing these questions would bring meteorological research even closer to practical applications in the energy sector.

In recent years, a new avenue has emerged in weather forecasting, challenging the longstanding dominance of NWP models (e.g., Bauer, 2024). Since 2022, artificial intelligence (AI) has evolved beyond its roles in post-processing, pattern recognition, and simple prediction tasks. It now stands as a serious contender in global weather forecasting (Jones, 2025). Several major technology companies and research institutions have introduced AI-based forecasting systems: Google DeepMind with GraphCast (Lam et al., 2023), NVIDIA with FourCastNet (Kurth et al., 2023), Huawei with Pangu-Weather (Bi et al., 2023), and ECMWF with AIFS (Lang et al., 2024). While these models still rely on components of NWP – particularly within the data assimilation stage – the recent development of Aardvark Weather marks a significant step forward. As the first fully end-to-end data-driven forecasting system, it attempts to tackle the entire NWP pipeline using AI alone (Allen et al., 2025).

Following the "quiet revolution" in NWP described by Bauer et al. (2015), these developments can certainly be seen as a "second revolution" in weather forecasting, as suggested by Charlton-Perez in Jones (2025). AI-based systems have already demonstrated forecast skill comparable to, and in some instances exceeding, that of ECMWF's operational models, while also offering improvements in computational efficiency by several orders of magnitude (Jones, 2025).

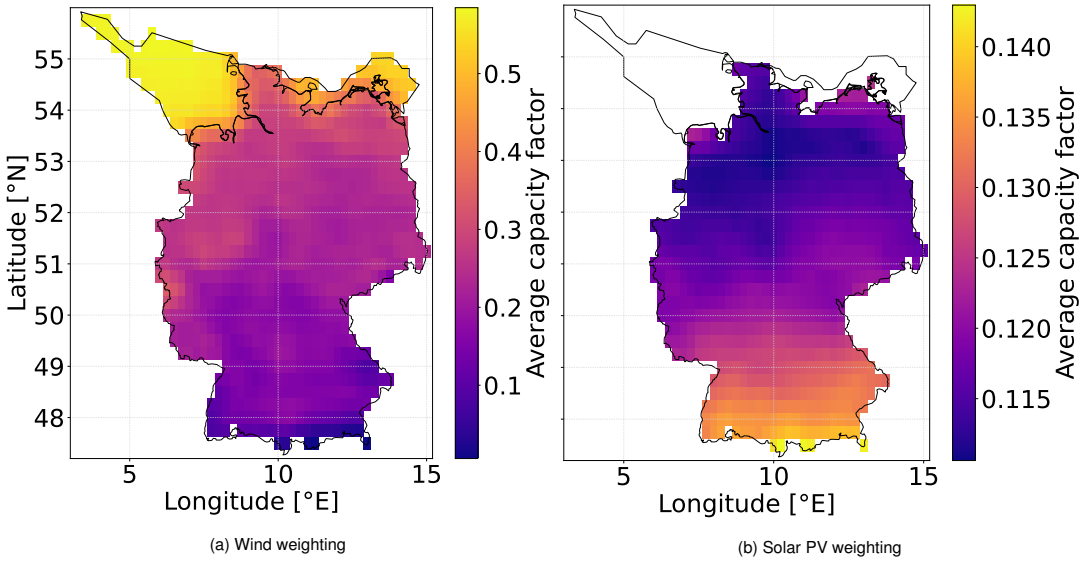
These advances open promising new avenues for research in the subseasonal to seasonal forecasting range. Key questions include: How well do AI models capture weather regimes? How do they compare to traditional NWP models in representing teleconnections, such as the MJO-NAO link? Can the analysis of AI forecasts reveal previously unidentified teleconnection patterns? Are the AI models capable of accurately forecasting strong regime projections, particularly considering that one of AI's known weaknesses lies in predicting extreme events that fall outside the distribution of the training data?

Final remarks

Today, weather forecasting may be undergoing a major technological evolution, one whose full trajectory and future developments remain uncertain. In 2020, Tim Palmer – a renowned physicist and meteorologist – stated, "For these reasons, I do not believe AI will ever compete seriously with NWP in the all-important day 7–10 medium range" (Palmer, 2020). Yet less than five years later, rapid advances in AI-based forecasting systems have begun to challenge that view – underscoring just how quickly scientific paradigms can evolve.

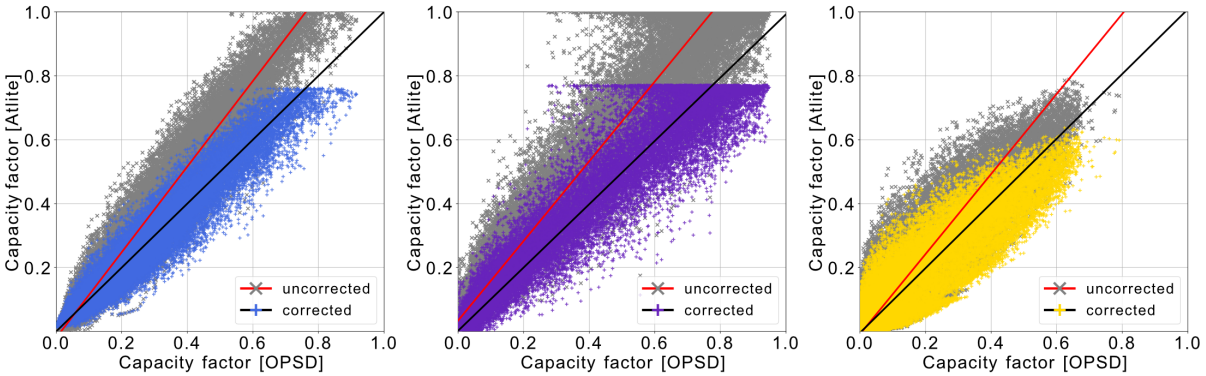
In this thesis, I have demonstrated how statistical post-processing can extend the skill of weather regime forecasts by approximately one day on average, enhancing predictability deeper into the subseasonal to seasonal range. Moreover, identifying so-called "windows of opportunity" complements the confidence information provided by ensemble spread by highlighting periods when specific weather events are not only more likely to occur but also more reliably forecasted (through climatological and model windows of opportunity, respectively), based on the prevailing atmospheric conditions. A statistical-dynamical approach that incorporates atmospheric information alongside the NWP model output has been shown to improve week 3 forecasts by up to 10%. Nevertheless, the landscape of weather forecasting is entering a dynamic and transformative phase – one marked by accelerating innovation and the convergence of statistical, dynamical, and AI-driven methods. This evolving landscape holds substantial promise for redefining what is considered possible in subseasonal to seasonal forecasting.

A. Appendix for Chapter 5



(a) Wind weighting (b) Solar PV weighting

Figure A.1.: Regional weighting of the (a) wind and (b) solar PV capacity factors in Germany and Germany's Exclusive Economic Zone, proportional to the potential power output at each location. The weighting corresponds to the average capacity factors.



(a) Onshore wind (b) Offshore wind (c) Solar PV

Figure A.2.: Scatter plot (crosses and pluses) of capacity factors given by OPSD (2020) (x-axis) and atlite (y-axis) for the (a) onshore wind, (b) offshore wind, and (c) solar PV. Compared are the uncorrected capacity factors generated by atlite (red lines) and the corrected capacity factors of atlite by using linear regression and the data of OPSD (2020) (orange lines). A perfect correlation is given by the black diagonal line. The availability of historical capacity factor time series by OPSD (2020) for Germany are limited to onshore wind: 2010–2018, offshore wind: 2016–2018, solar PV: 2012–2018, which are used as the basis for the regression.

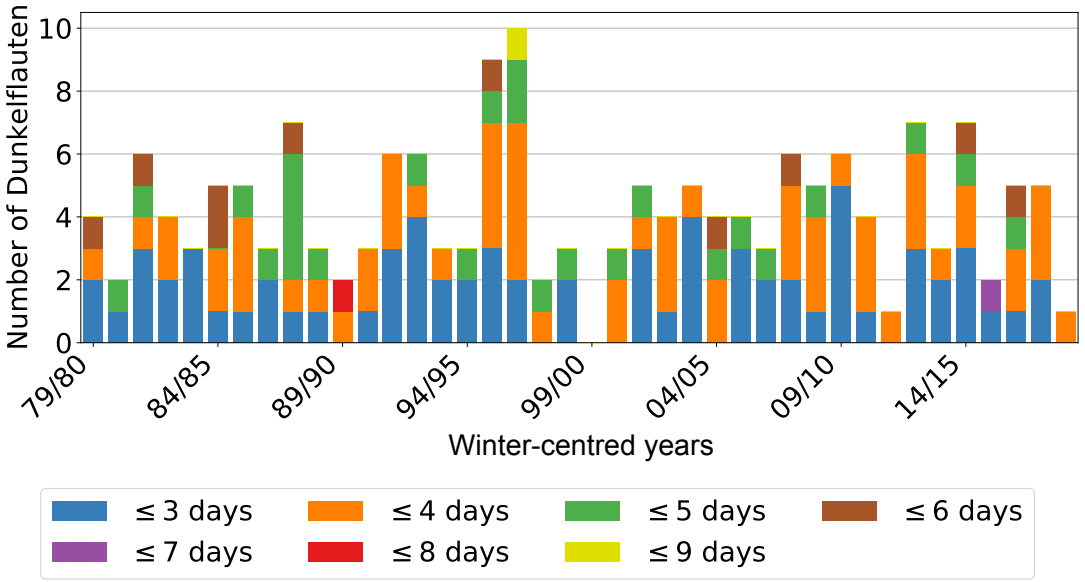


Figure A.3.: Amount and duration of Dunkelflauten over the 40 year period from 1979–2018. The years are represented as winter-centred years, starting in July of the first year and ending in June of the second year (x-axis).

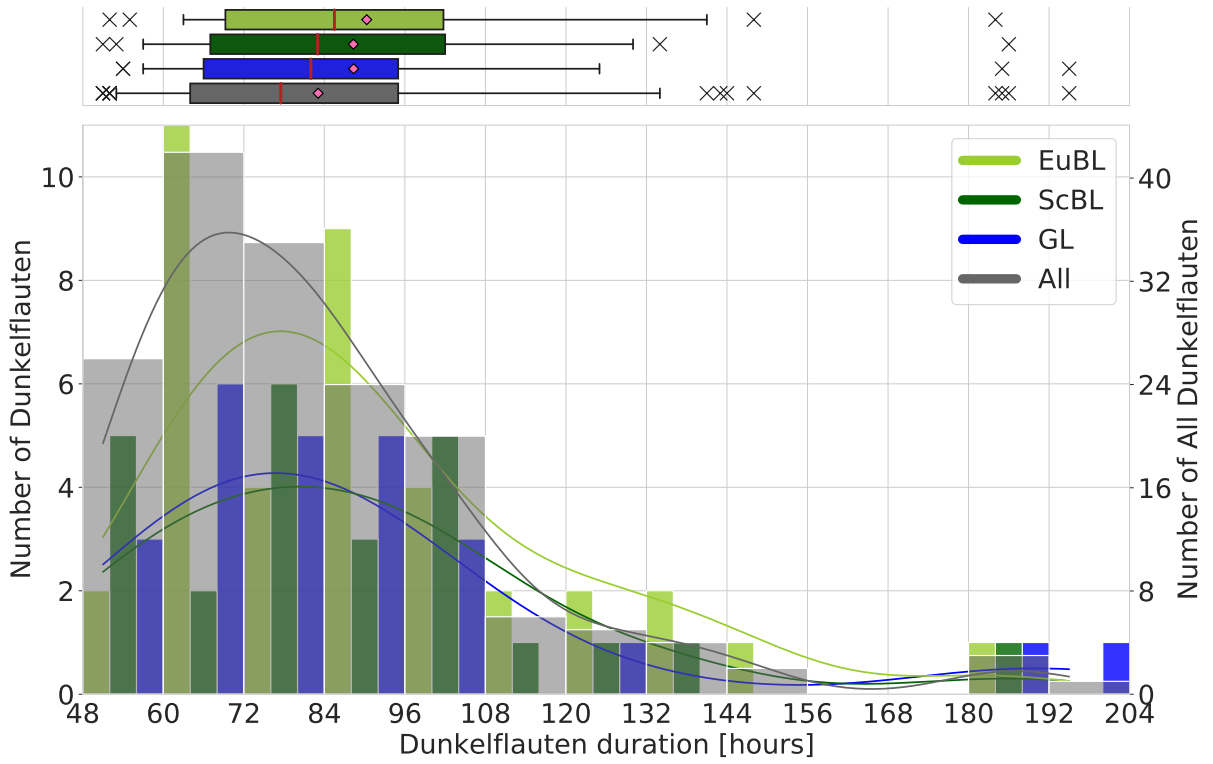


Figure A.4.: Distribution of the duration of Dunkelflauten for the European, Scandinavian and Greenland Blocking weather regimes (left y-axis) and for all Dunkelflauten (right y-axis). The histograms have a bin width of 12 h, and the curves are kernel density estimations. Box and Whisker plots of the distributions of the duration of Dunkelflauten for the three blocked weather regimes and all Dunkelflauten are given in the marginal plot. Pink diamonds represent the mean, red vertical lines the median, boxes the 25–75% percentile, whiskers the 5–95% percentiles and crosses the outliers.

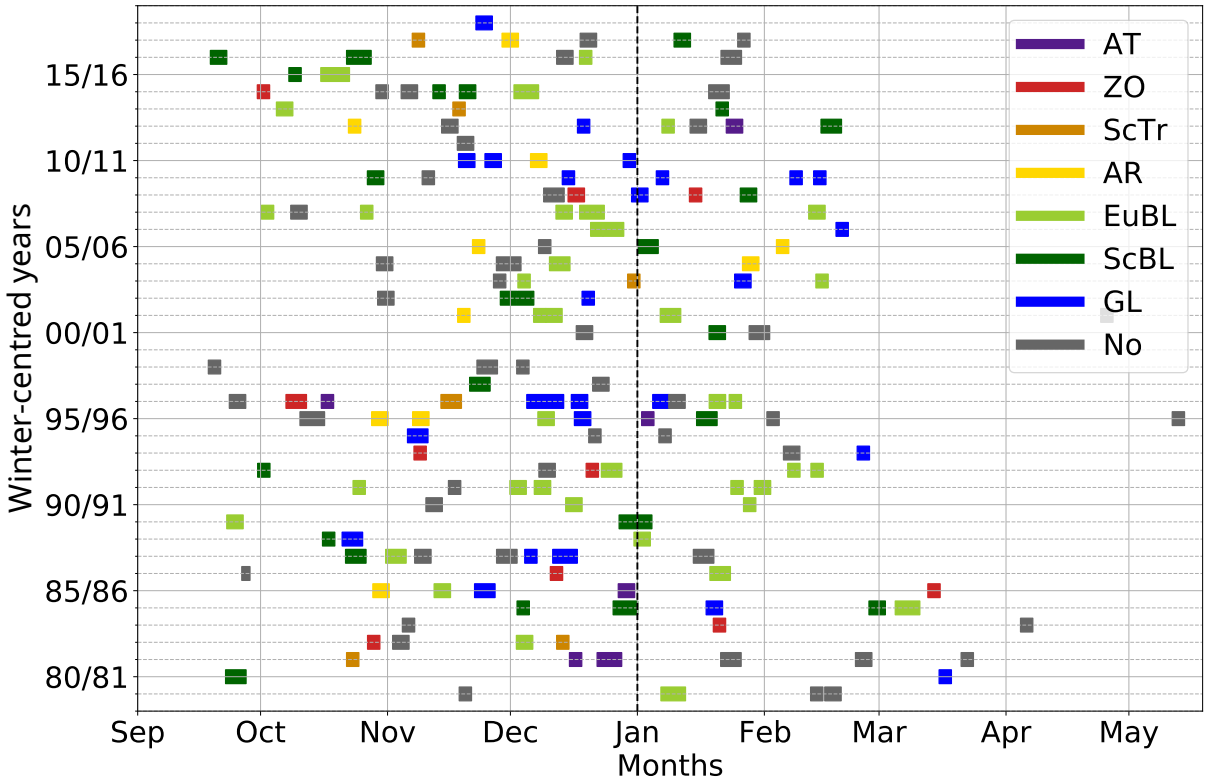


Figure A.5.: Calendar plot of all Dunkelflauten in the 40 year period from 1979–2018. The colours indicate the main weather regimes occurring in the Dunkelflauten phase. The years are represented as winter-centred years (y-axis) and months are shown from September to May (x-axis).

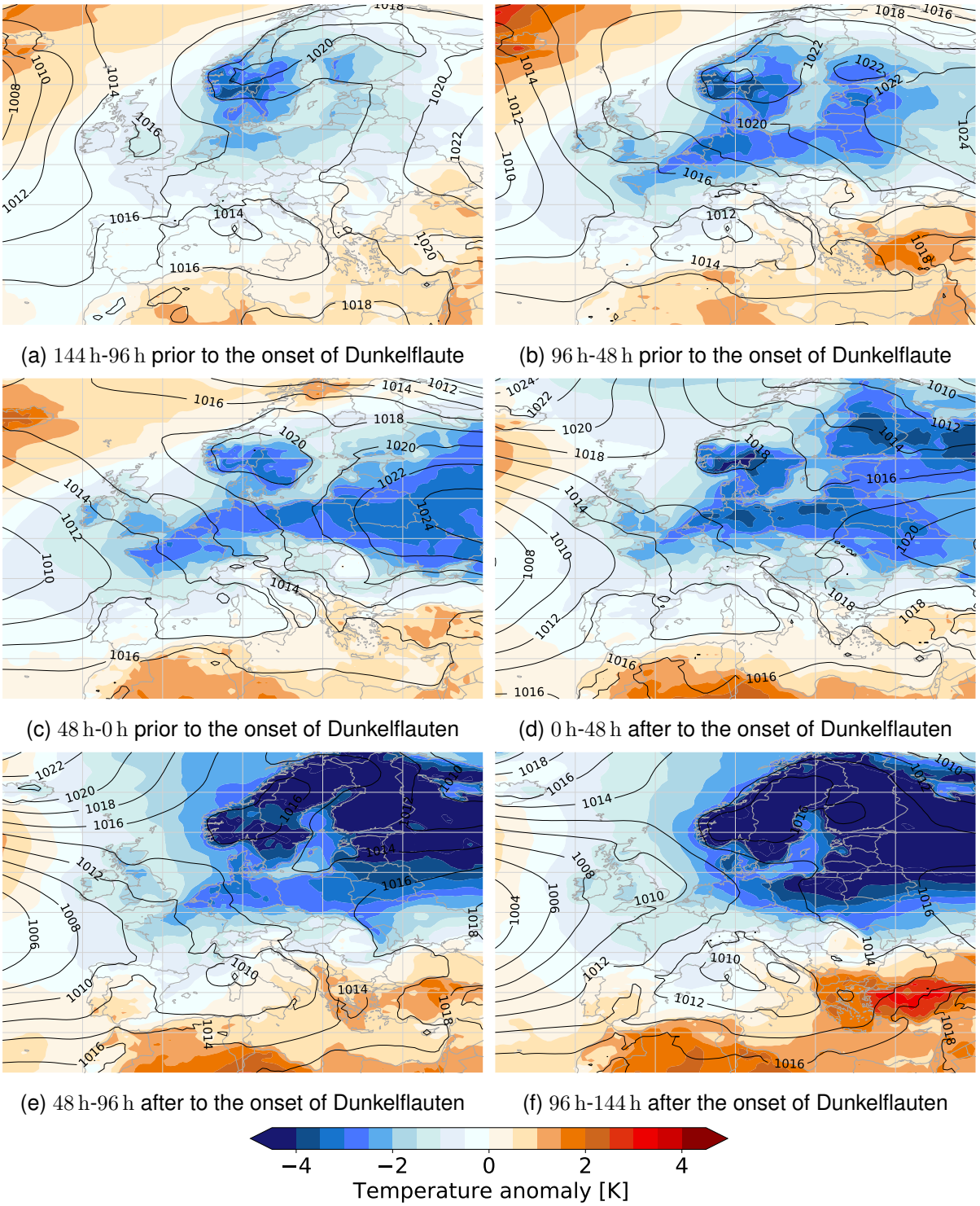


Figure A.6.: Mean temporal evolution of sea level pressure (contouring every 2 hPa and 2 m temperature anomalies (shading every 0.5 K) in the phase building up to Greenland Blocking Dunkelflauten (a–c, starting 6 days prior to the onset of Dunkelflauten) and the phase during and after Greenland Blocking Dunkelflauten (d–f, ending 6 days after the onset of Dunkelflauten).

B. Appendix for Chapter 6

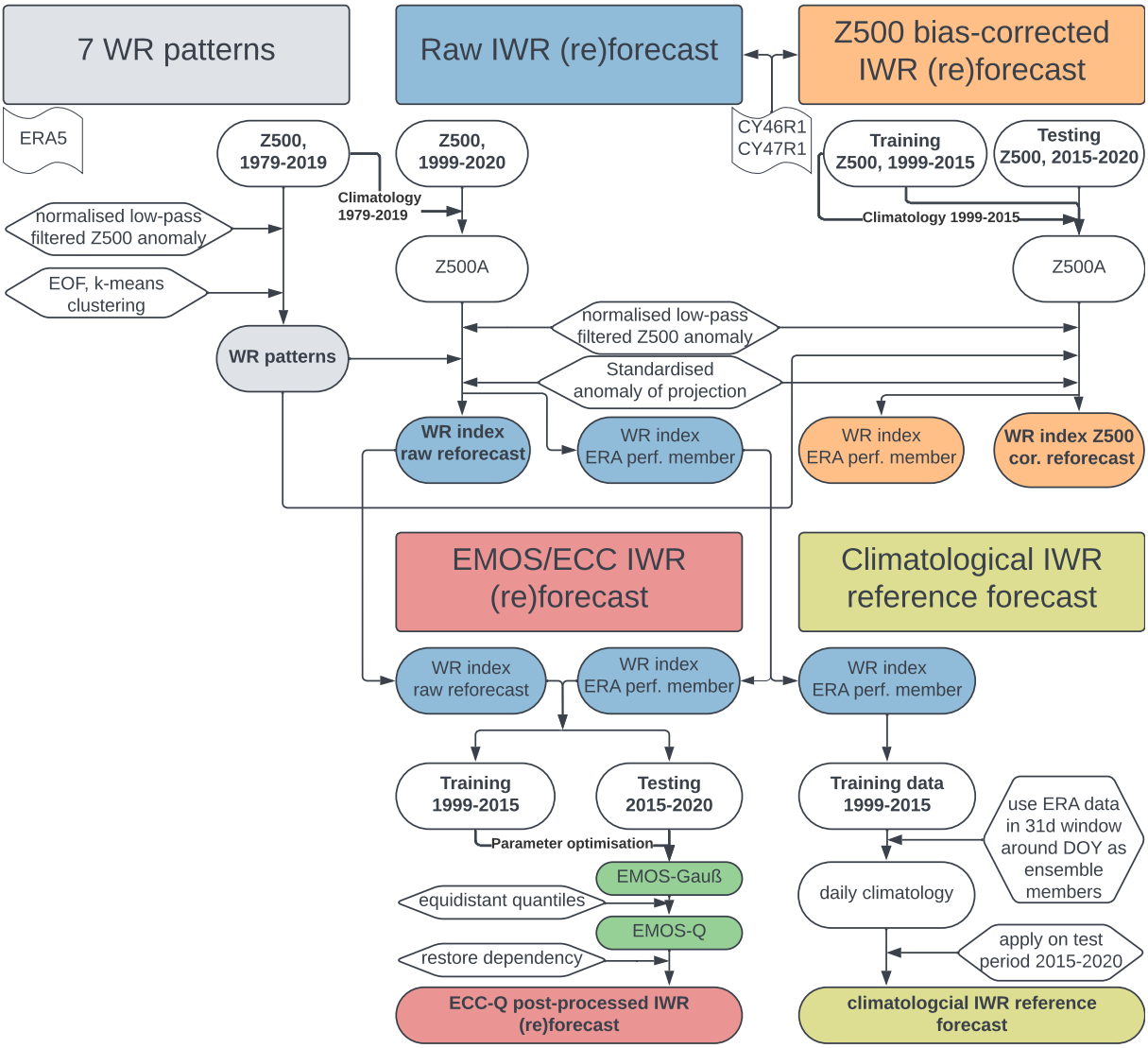


Figure B.1.: Work flow of computing the weather regime patterns (top left, grey) and the weather regime index (IWR) of the raw (top middle, blue), Z500 bias-corrected (top right, orange), post-processed EMOS/ECC forecasts (bottom middle, green/red) and climatological reference forecast (bottom right, olive). The step from Z500A to WR index for the raw and Z500 bias-corrected forecasts involves the Equations 4.1 and 4.2 from the data and method section.

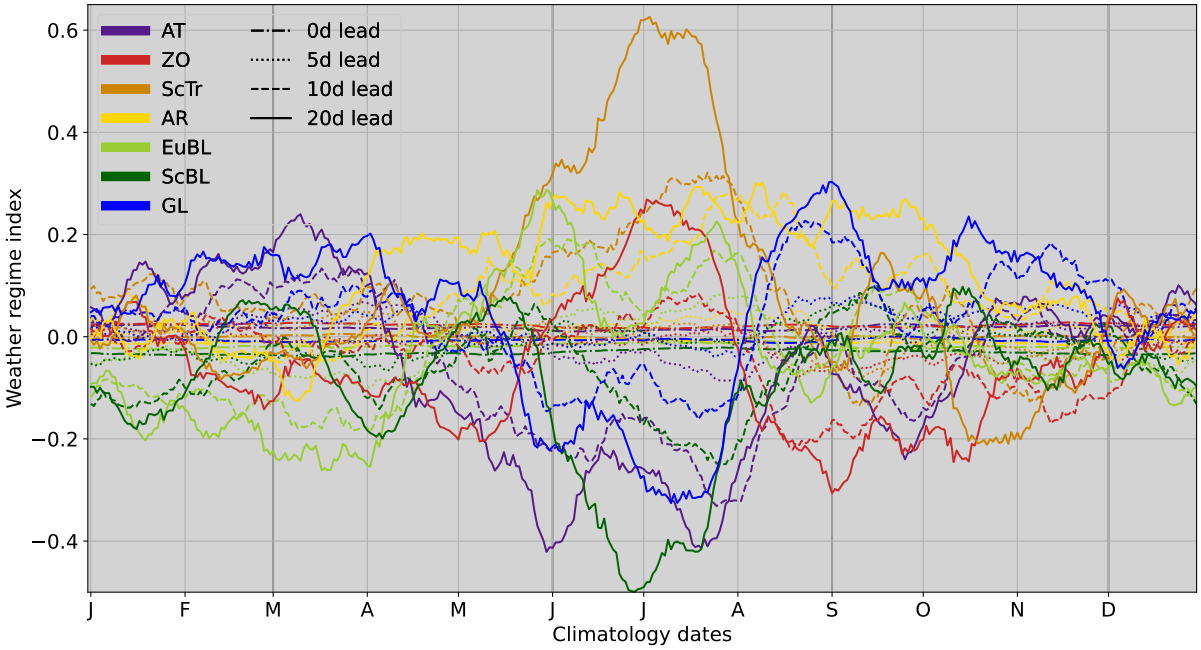


Figure B.2.: Climatological projection of Z500 biases (see Figure 6.1) onto the weather regime index for 0, 5, 10, and 20 day lead times. The Z500 biases from Figure 6.1 are mapped onto the Z500A patterns of the weather regimes (Figure 2.2) to derive the weather regime index shown here. For example, at a 20-day lead time in summer (JJA), the Z500 bias field (Figure 6.1p) demonstrates strong positive values in the northern North Atlantic and negative values in the Scandinavian region. This aligns well with the Z500A pattern of the Scandinavian Trough (Figure 2.2), hence the Z500 bias pattern is projecting into Scandinavian Trough (orange line).

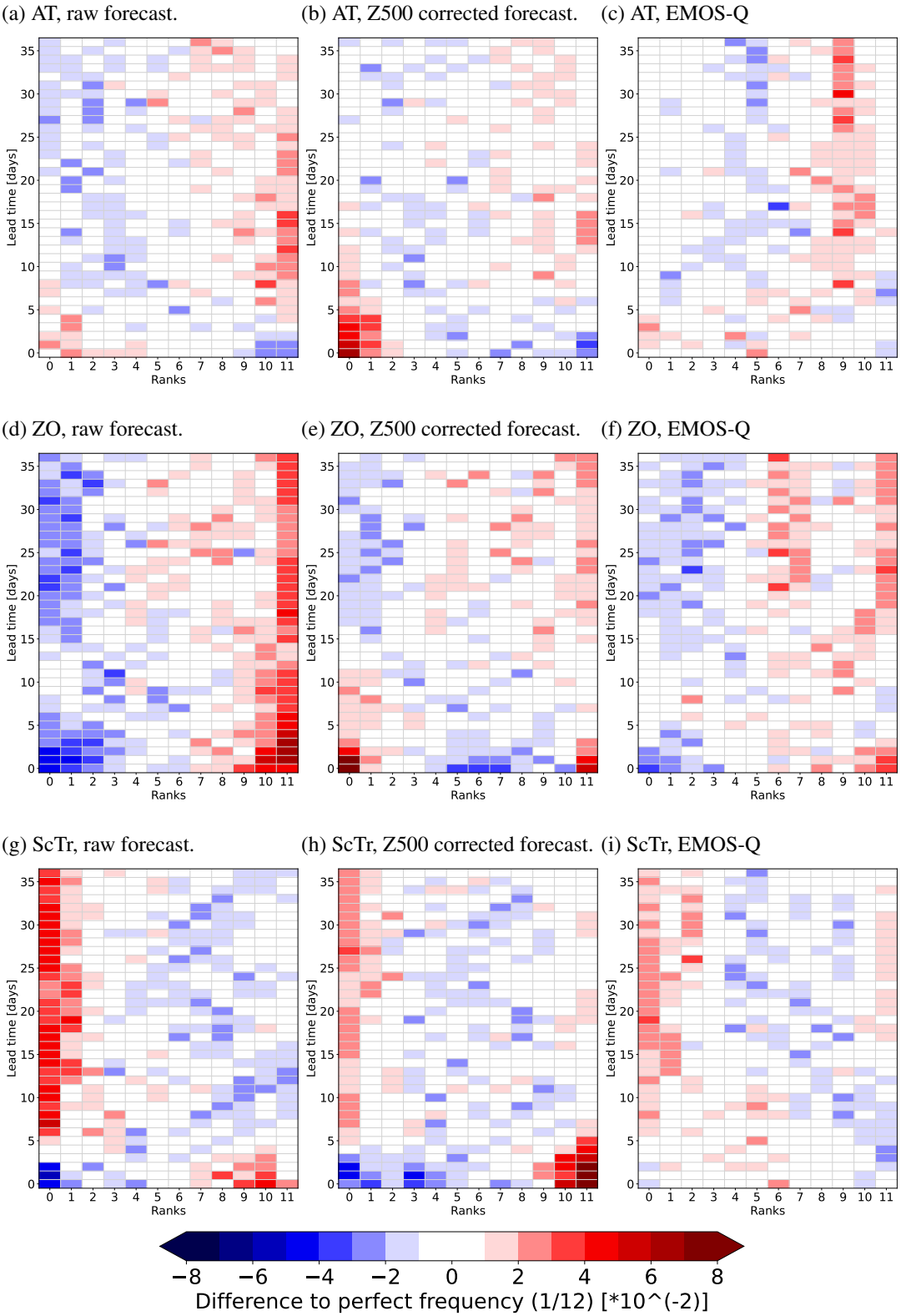


Figure B.3.: Rank histograms for all cyclonic weather regimes: (a–c) Atlantic Trough, (d–f) Zonal Regime, (g–i) Scandinavian Trough. With the raw forecast in the left column, Z500 bias-corrected forecasts in the middle column and EMOS-Q in the right column. The rank histograms for Atlantic Trough are also shown in the main manuscript in Figure 6.2a-c.

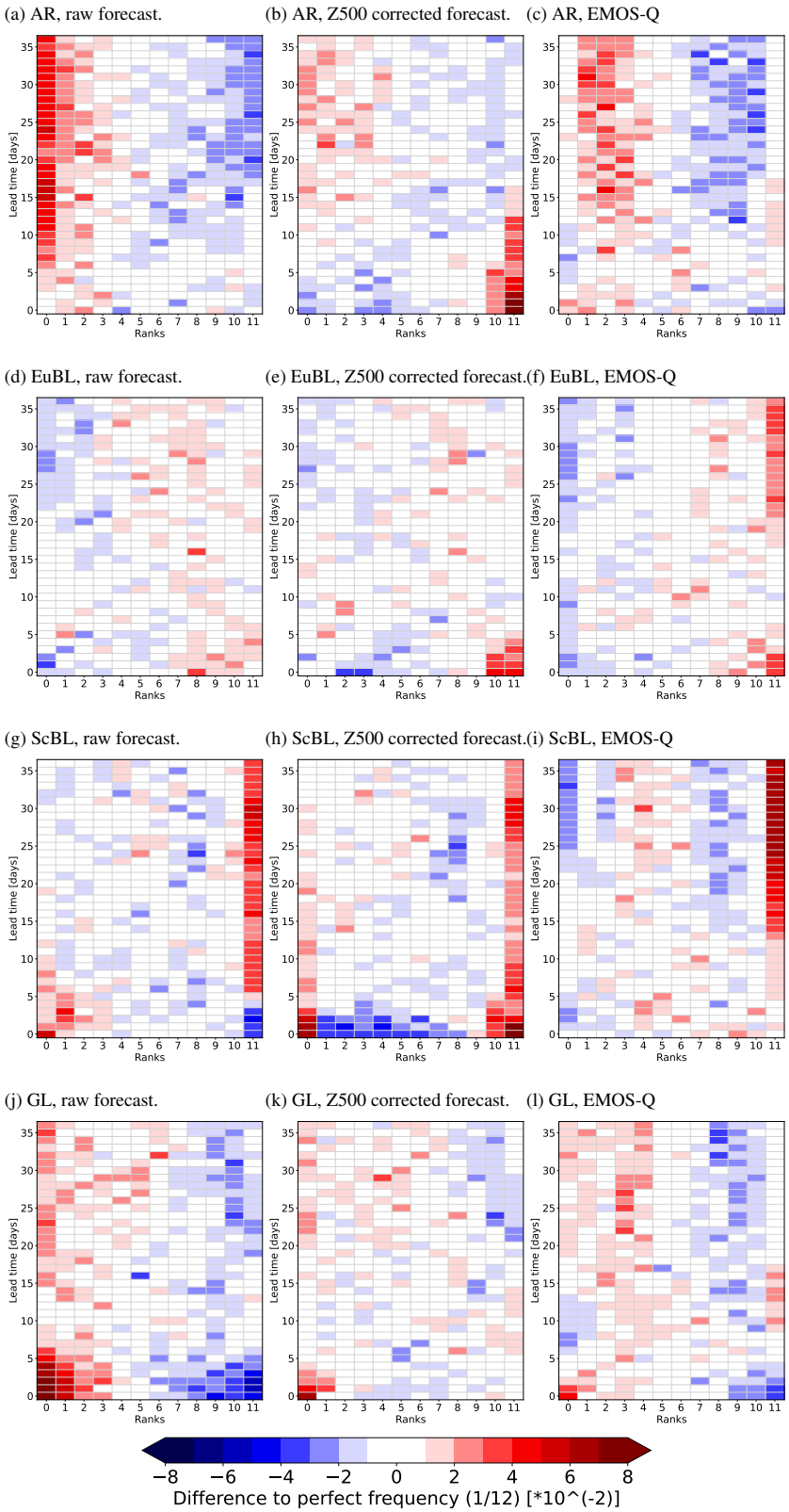


Figure B.4.: Rank histograms for all anticyclonic weather regimes: (a–c) Atlantic Ridge, (d–f) European Blocking, (g–i) Scandinavian Blocking, (j–l) Greenland Blocking. With the raw forecast in the left column, Z500 bias-corrected forecasts in the middle column and EMOS-Q in the right column. The rank histograms for Atlantic Ridge are also shown in the main manuscript in Figure 6.2d–f.

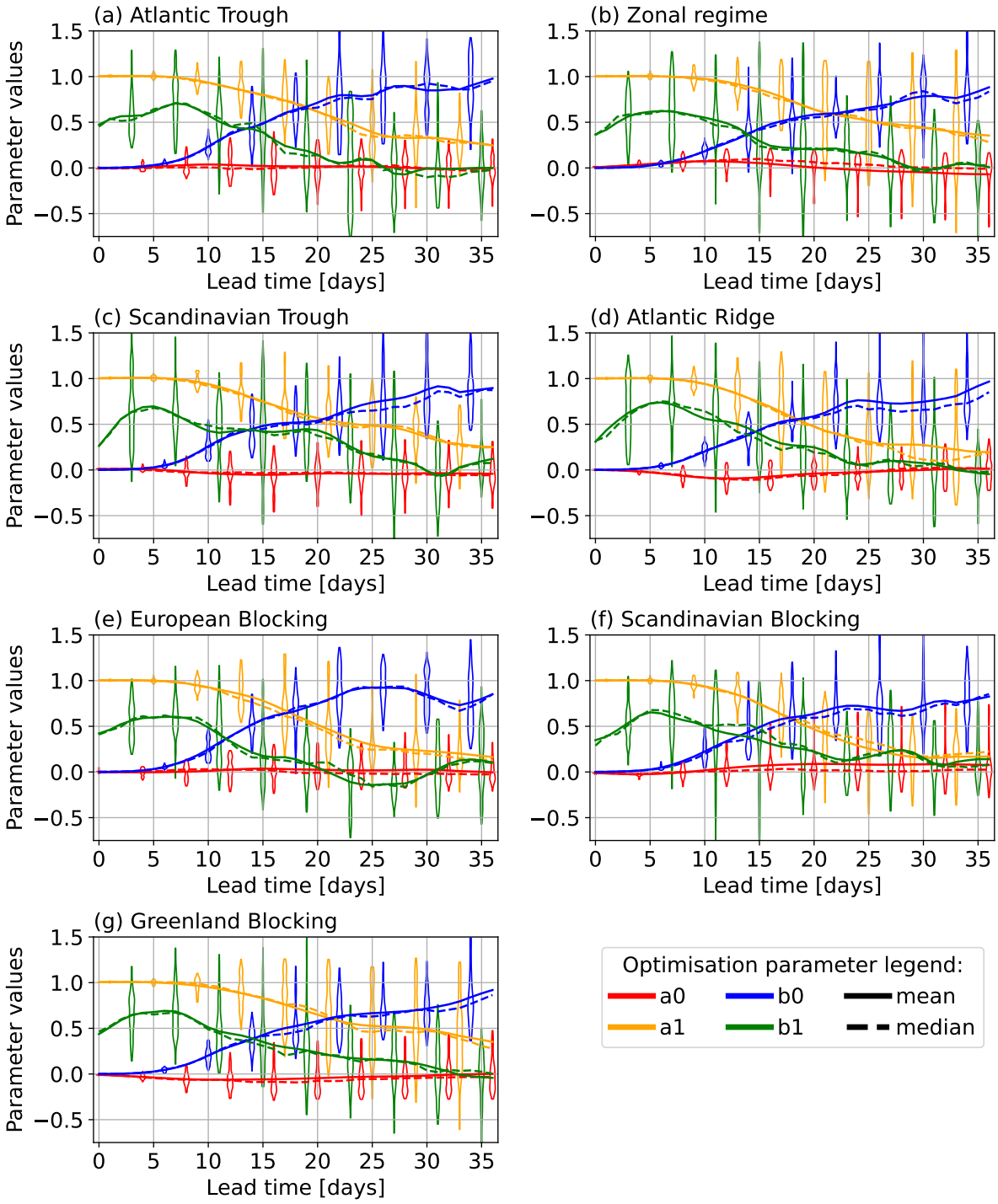
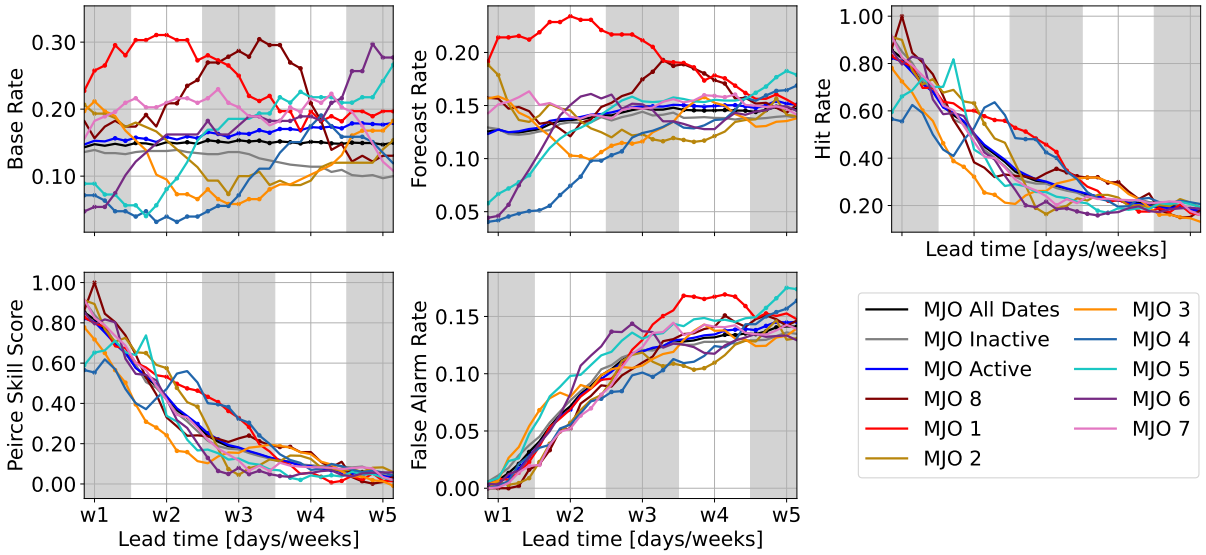


Figure B.5.: Optimised parameters a_0 (red), a_1 (orange), b_0 (blue), and b_1 (green) for the EMOS (Ensemble Model Output Statistics) procedure, which adjust the mean (μ) and variance (σ^2) of ensemble forecast distributions across lead times from 0 to 36 days. The parameters are introduced in Equation 4.10, where a_0 and b_0 are the intercept coefficients for the location and scale parameters, respectively, and a_1 and b_1 are the coefficients associated with the ensemble mean and ensemble variance, respectively. Each subplot corresponds to a specific weather regime: (a) Atlantic Trough, (b) Zonal regime, (c) Scandinavian Trough, (d) Atlantic Ridge, (e) European Blocking, (f) Scandinavian Blocking, and (g) Greenland Blocking. Solid and dashed lines represent the mean and median values of the parameters over the testing period from June 2015 to May 2020. Violin plots indicate the estimated density distribution of the coefficients, shown every fourth lead time only to improve visual clarity.

C. Appendix for Chapter 7

(a) MJO-conditioned rates.



(b) SPV-conditioned rates.

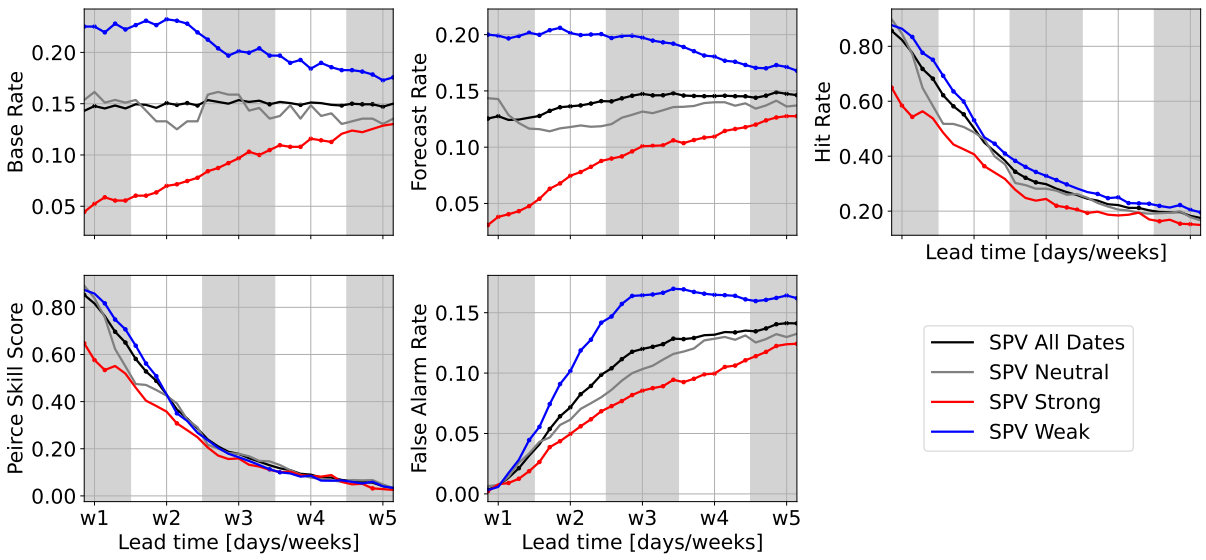


Figure C.1.: Rates for weekly mean Greenland Blocking activity ($WRact_{mean,GL}$) in the extended winter period (November–March). These rates are split into subsets based on phases of the (a) MJO or (b) SPV. In contrast to Figure 7.2 and Figure 7.3, here all lead times are shown (x-axis). Dots on the 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.

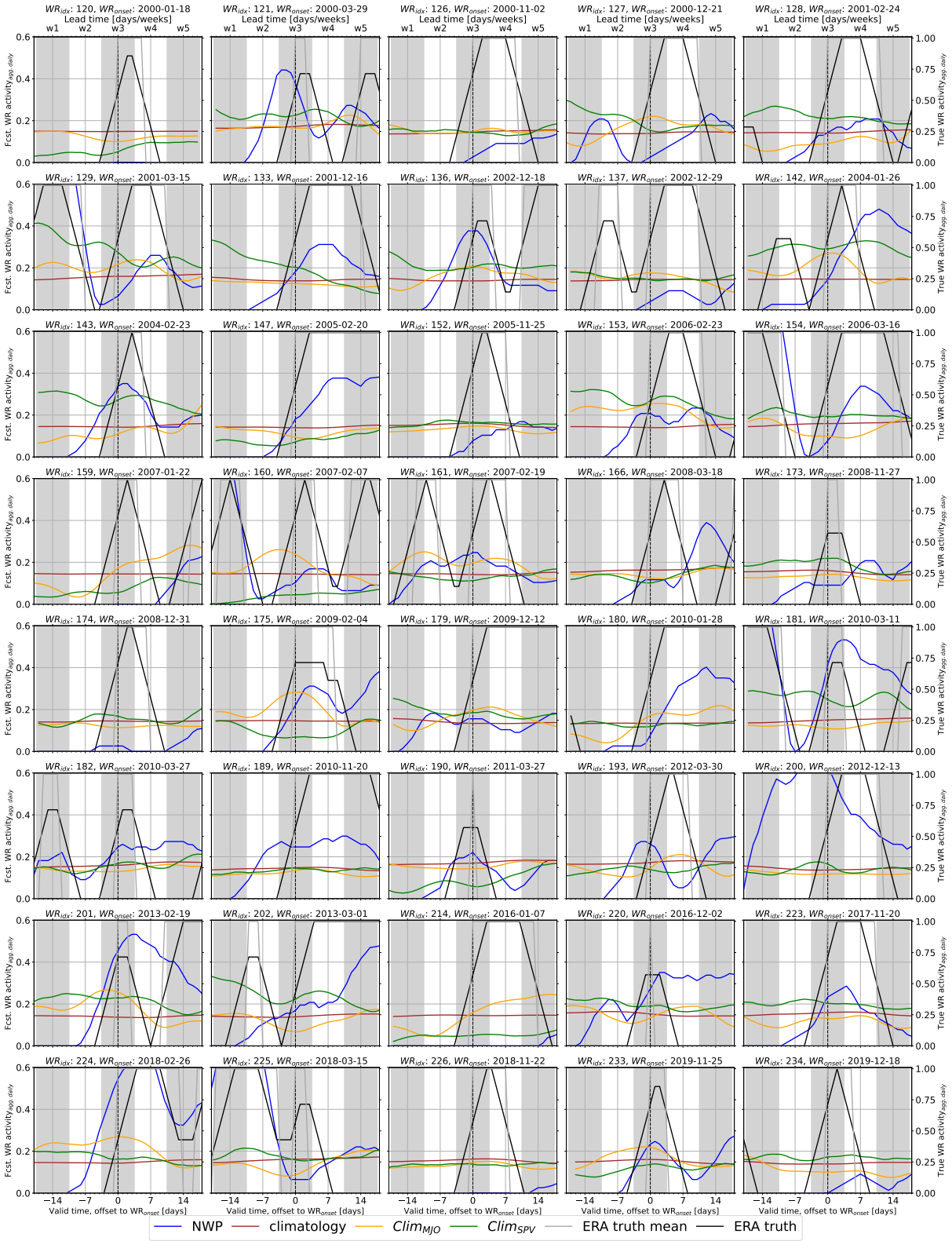


Figure C.2.: Forecasts of the aggregated daily Greenland Blocking activity used for the mean composite plot in Figure 7.5a for all Greenland Blocking activities. The forecasts are initialised three weeks ahead of the Greenland Blocking activity onset.

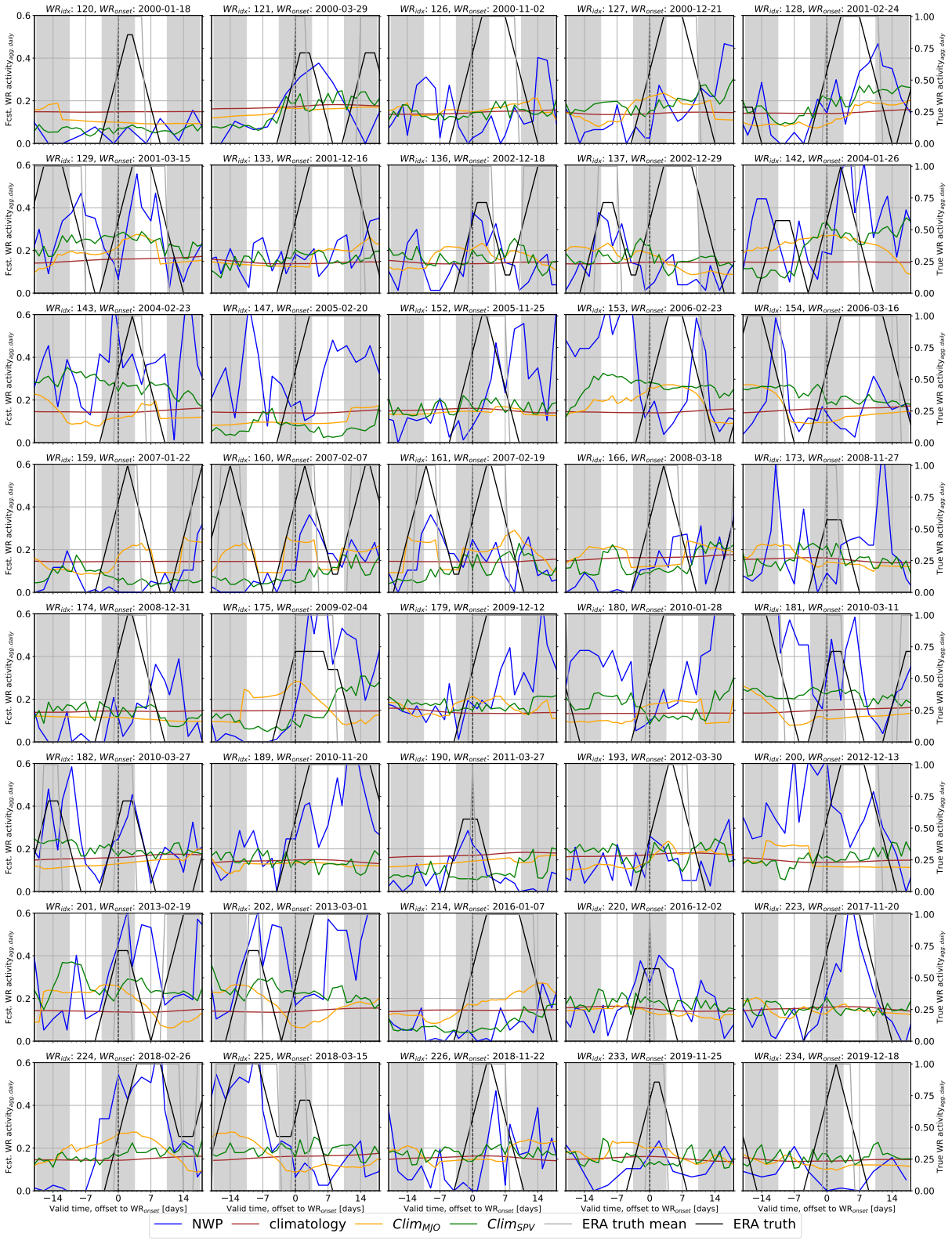


Figure C.3.: Forecasts of the aggregated daily Greenland Blocking activity used for the mean composite plot in Figure 7.5b for all Greenland Blocking activities. Consecutive forecasts with each a lead time of three weeks are used.

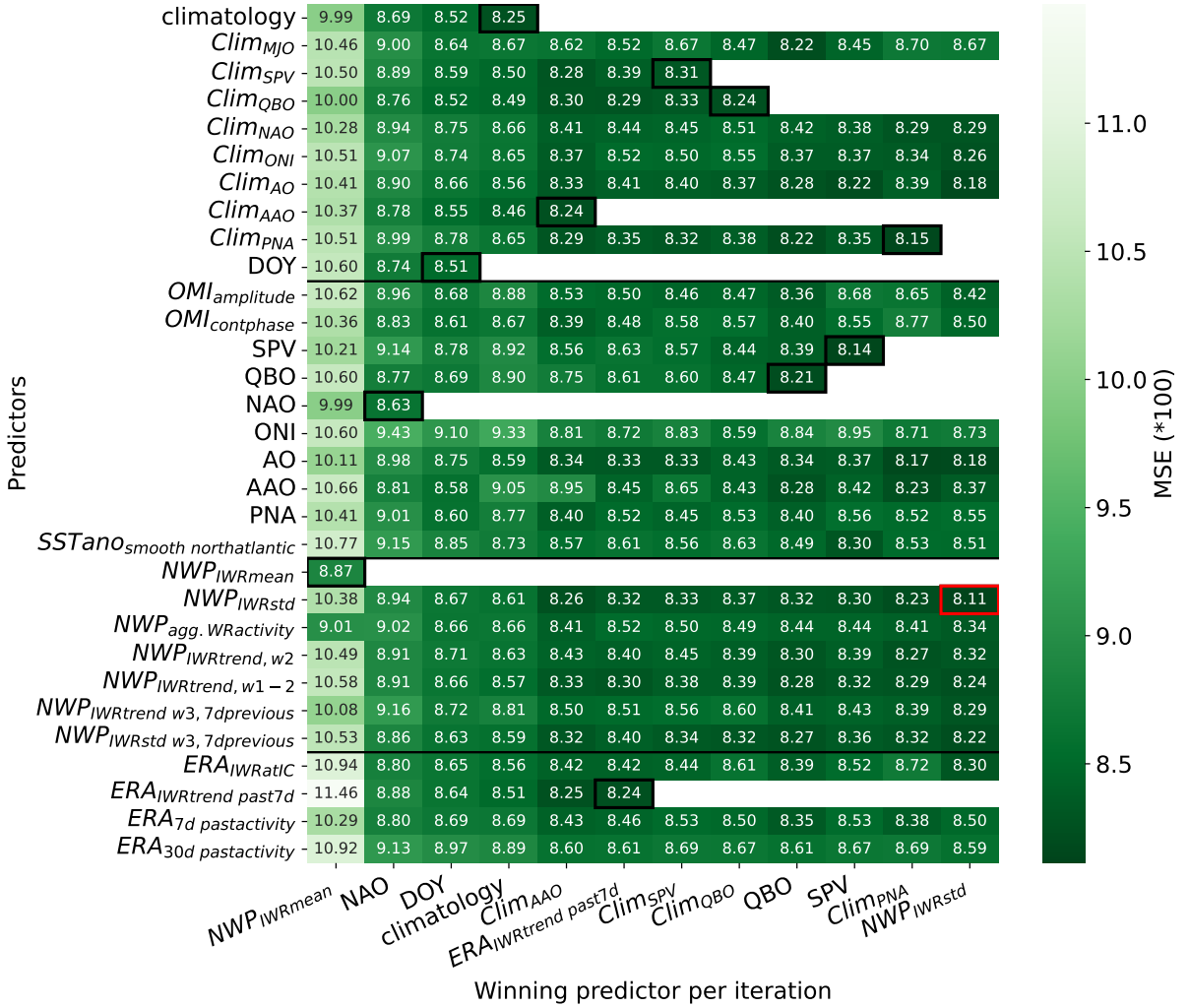


Figure C.4.: Visualisation of the stepwise feature selection process for the all-features neural network (NN_{all}) model and Greenland Blocking activity for forecast week three. The x-axis indicates the winning feature per step of the feature selection. The selection process is defined via the MSE (visualised 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 C.1 and C.2.

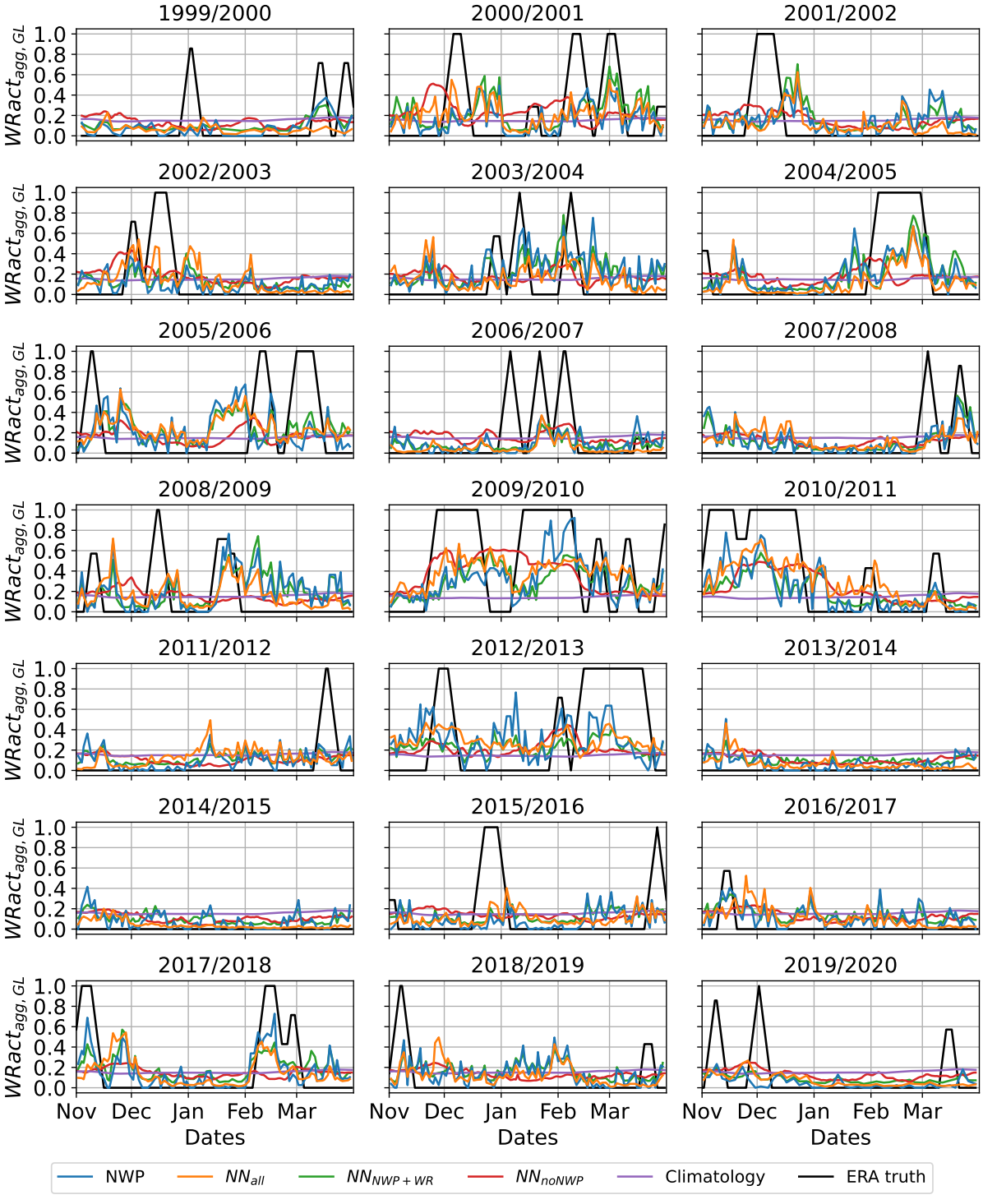


Figure C.5.: Forecasts of the aggregated daily GL activity for each extended winter period from 1999 to 2020. The ERA5 (actual) weather regime activity is indicated in black, the neural networks all-features, NWP and ERA, and climatology in blue, orange and green, as well as the NWP model in red and 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	$\text{OMI}_{\text{amplitude}}$	MJO amplitude in OMI-phase space. Source: `https://psl.noaa.gov/mjo/mjoindex/omi.era5.1x.webpage.4023.txt`
	$\text{OMI}_{\text{contphase}}$	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
	$\text{SSTano}_{\text{smooth northatlantic}}$	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 C.1.: 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 C.2.

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)
	$NWP_{IWRtrend,w1\text{-}2}$	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

Table C.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 C.1.

D. Bibliography

Abid, M. A., F. Kucharski, F. Molteni, I.-S. Kang, A. M. Tompkins, and M. Almazroui, 2021: Separating the Indian and Pacific Ocean Impacts on the Euro-Atlantic Response to ENSO and Its Transition from Early to Late Winter. *Journal of Climate*, **34**, 1531–1548.

Allen, A., S. Markou, W. Tebbutt, J. Requeima, W. P. Bruinsma, T. R. Andersson, M. Herzog, N. D. Lane, M. Chantry, J. S. Hosking, and R. E. Turner, 2025: End-to-end data-driven weather prediction. *Nature*.

Baldwin, M. P. and T. J. Dunkerton, 2001: Stratospheric Harbingers of Anomalous Weather Regimes. Tech. rep.

Baldwin, M. P., 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, 2001: The Quasi-Biennial Oscillation. *Review of Geophysics*, **39**, 179–229.

Bauer, P., 2024: What if? Numerical weather prediction at the crossroads. *Journal of the European Meteorological Society*, **1**, 100 002.

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

Baur, F., P. Hess, and H. Nagel, 1944: Kalender der Grosswetterlagen Europas vom 1.1.1881 bis 31.12.1943.

BDEW, 2022: Entwicklung der Beheizungsstruktur im Wohnungsneubau. URL `https://www.bdew.de/service/daten-und-grafiken/entwicklung-beheizungsstruktur-baugenehmigungen/`.

Beerli, R. and C. M. Grams, 2019: 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.

Bi, K., L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, 2023: Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, **619**, 533–538.

Bjerknes, V., E. Volken, and S. Brönnimann, 1904: The problem of weather prediction, considered from the viewpoints of mechanics and physics. *Meteorologische Zeitschrift*.

Bloomfield, H., D. Brayshaw, and A. Charlton-Perez, 2020a: ERA5 derived time series of European country-aggregate electricity demand, wind power generation and solar power generation. URL `https://doi.org/10.17864/1947.272`.

Bloomfield, H. C., D. J. Brayshaw, P. L. Gonzalez, and A. Charlton-Perez, 2021a: Pattern-based conditioning enhances sub-seasonal prediction skill of European national energy variables. *Meteorological Applications*, **28**.

Bloomfield, H. C., D. J. Brayshaw, L. C. Shaffrey, P. J. Coker, and H. E. Thornton, 2016: Quantifying the increasing sensitivity of power systems to climate variability. *Environmental Research Letters*, **11**.

Bloomfield, H. C., D. J. Brayshaw, A. Troccoli, C. M. Goodess, M. D. Felice, L. Dubus, P. E. Bett, and Y. M. Saint-Drenan, 2021b: Quantifying the sensitivity of european power systems to energy scenarios and climate change projections. *Renewable Energy*, **164**, 1062–1075.

Bloomfield, H. C., C. C. Suitters, and D. R. Drew, 2020b: Meteorological Drivers of European Power System Stress. *Journal of Renewable Energy*, **2020**, 1–12.

Bommer, P. L., M. Kretschmer, F. R. Spuler, K. Bykov, and M. M. C. Höhne, 2025: Deep Learning Meets Teleconnections: Improving S2S Predictions for European Winter Weather. *arXiv*.

Brayshaw, D. J., A. Troccoli, R. Fordham, and J. Methven, 2011: The impact of large scale atmospheric circulation patterns on wind power generation and its potential predictability: A case study over the UK. *Renewable Energy*, **36**, 2087–2096.

Brown, T., D. Schlachtberger, A. Kies, S. Schramm, and M. Greiner, 2018: Synergies of sector coupling and transmission reinforcement in a cost-optimised, highly renewable European energy system. *Energy*, **160**, 720–739.

Buizza, R. and M. Leutbecher, 2015: The forecast skill horizon. *Quarterly Journal of the Royal Meteorological Society*, **141**, 3366–3382.

Büeler, D., R. Beerli, H. Wernli, and C. M. Grams, 2020: Stratospheric influence on ECMWF subseasonal forecast skill for energy-industry-relevant surface weather in European countries. *Quarterly Journal of the Royal Meteorological Society*, **146**, 3675–3694.

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

Cassou, C., 2008: Intraseasonal interaction between the Madden-Julian Oscillation and the North Atlantic Oscillation. *Nature*, **455**, 523–527.

Cassou, C., J. W. Hurrell, and C. Deser, 2004: North Atlantic Winter Climate Regimes: Spatial Asymmetry, Stationarity with Time, and Oceanic Forcing. *Journal of Climate*, **17**, 1055–1068.

Cassou, C., L. Terray, and A. S. Phillips, 2005: LETTER Tropical Atlantic Influence on European Heat Waves. *Journal of Climate*, **18**, 2805–2811.

Charlton, A. J. and L. M. Polvani, 2007: A New Look at Stratospheric Sudden Warmings. Part I: Climatology and Modeling Benchmarks. Tech. rep.

Charlton-Perez, A. J., R. W. Aldridge, C. M. Grams, and R. Lee, 2019: Winter pressures on the UK health system dominated by the Greenland Blocking weather regime. *Weather and Climate Extremes*, **25**.

Charlton-Perez, A. J., L. Ferranti, and R. W. Lee, 2018: The influence of the stratospheric state on North Atlantic weather regimes. *Quarterly Journal of the Royal Meteorological Society*, **144**, 1140–1151.

Charney, J. G. and J. G. Devore, 1979: Multiple Flow Equilibria in the Atmosphere and Blocking. *Journal of the Atmospheric Sciences*, **36**, 1205–1216.

Chase, R. J., D. R. Harrison, A. Burke, G. M. Lackmann, and A. McGovern, 2022: A Machine Learning Tutorial for Operational Meteorology. Part I: Traditional Machine Learning. *Weather and Forecasting*, **37**, 1509–1529.

Chase, R. J., D. R. Harrison, G. M. Lackmann, and A. McGovern, 2023: A Machine Learning Tutorial for Operational Meteorology. Part II: Neural Networks and Deep Learning. *Weather and Forecasting*, **38**, 1271–1293.

Chen, J., T. Janke, F. Steinke, and S. Lerch, 2024: Generative machine learning methods for multivariate ensemble postprocessing. *Annals of Applied Statistics*, **18**, 159–183.

Cionni, I., L. Lledó, V. Torralba, and A. Dell'Aquila, 2022: Seasonal predictions of energy-relevant climate variables through Euro-Atlantic Teleconnections. *Climate Services*, **26**.

Diebold, F. X. and R. S. Mariano, 1995: Comparing predictive accuracy. *Journal of Business and Economic Statistics*, **13**.

Domeisen, D. I. V., C. M. Grams, and L. Papritz, 2020: The role of North Atlantic–European weather regimes in the surface impact of sudden stratospheric warming events. *Weather and Climate Dynamics*, **1**, 373–388.

Drücke, J., M. Borsche, P. James, F. Kaspar, U. Pfeifroth, B. Ahrens, and J. Trentmann, 2021: Climatological analysis of solar and wind energy in Germany using the Grosswetterlagen classification. *Renewable Energy*, **164**, 1254–1266.

Dutton, J., 2021: What is the El Niño/Southern Oscillation (ENSO)? URL https://www.worldclimateservice.com/2021/08/25/what-is-el-nino/?utm_source=chatgpt.com.

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

Edenhofer, O., R. Pichs-Madruga, Y. Sokona, K. Seyboth, D. Arvizu, T. Bruckner, J. M. Christensen, J.-M. Devarnay, A. Faaij, M. Fischedick, B. Goldstein, G. Hansen, J. Huckerby, A. Jäger-Waldau, S. Kadner, D. Kammen, V. Krey, A. Kumar, A. Lewis, and T. Zwickel, 2011: Summary for Policy Makers. Tech. rep., IPCC Special Report on Renewable Energy Sources and Climate Change Mitigation, URL https://orbit.dtu.dk/en/publications/summary-for-policy-makers.

Elliot, A. J., K. W. Cross, and D. M. Fleming, 2008: Acute respiratory infections and winter pressures on hospital admissions in England and Wales 1990-2005. *Journal of Public Health*, **30**, 91–98.

EnBW, 2025: Drei Gründe, warum der Strom bei einer Dunkelflaute nicht ausgeht. URL https://www.enbw.com/unternehmen/themen/netze/dunkelflaute.html.

Energy-Charts, 2025: Net installed electricity generation capacity in Germany. URL https://www.energy-charts.info/charts/installed_power/chart.htm?c=DE&expansion=installed_power&legendItems=4x1vnq.

Feldstein, S. B. and C. L. Franzke, 2017: *Nonlinear and stochastic climate dynamics: Atmospheric teleconnection patterns*. Cambridge University Press, 54-104 pp.

Feng, P. N. and H. Lin, 2019: Modulation of the MJO-Related Teleconnections by the QBO. *Journal of Geophysical Research: Atmospheres*, **124**, 12 022–12 033.

———, 2021: Modulation of the MJO-Related Teleconnection by the QBO in Subseasonal-to-Seasonal Prediction Models. *Atmosphere - Ocean*, **59**, 165–177.

Ferranti, L., S. Corti, and M. Janousek, 2015: Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, **141**, 916–924.

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

Franzke, C. L. E. and T. O'Kane, 2017: *Nonlinear and stochastic climate dynamics*. Cambridge University Press.

Gerighausen, J., J. Dorrington, M. Osman, and C. M. Grams, 2024: Quantifying intra-regime weather variability for energy applications. *arXiv*.

Gneiting, T., 2014: Calibration of Medium-Range Weather Forecasts. *ECMWF Technical Memorandum*, **719**.

Gneiting, T. and M. Katzfuss, 2014: Probabilistic forecasting. *Annual Review of Statistics and Its Application*, **1**, 125–151.

Gneiting, T. and A. E. Raftery, 2007: Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359–378.

Gold, S., E. White, W. Roeder, M. McAleenan, C. S. Kabban, and D. Ahner, 2020: Probabilistic contingency tables: An improvement to verify probability forecasts. *Weather and Forecasting*, **35**, 609–621.

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

Grams, C. M., L. Ferranti, and L. Magnusson, 2020: How to make use of weather regimes in extended-range predictions for Europe. *ECMWF Newsletter*, **165**.

Green, M. R. and J. C. Furtado, 2019: Evaluating the Joint Influence of the Madden-Julian Oscillation and the Stratospheric Polar Vortex on Weather Patterns in the Northern Hemisphere. *Journal of Geophysical Research: Atmospheres*, **124**, 11 693–11 709.

Hamill, T. M. and J. S. Whitaker, 2006: Probabilistic Quantitative Precipitation Forecasts Based on Reforecast Analogs: Theory and Application. *Monthly Weather Review*, **134**, 3209–3229.

Hannachi, A., D. M. Straus, C. L. Franzke, S. Corti, and T. Woollings, 2017: Low-frequency nonlinearity and regime behavior in the Northern Hemisphere extratropical atmosphere.

Haupt, S. E., W. Chapman, S. V. Adams, C. Kirkwood, J. S. Hosking, N. H. Robinson, S. Lerch, and A. C. Subramanian, 2021: Towards implementing artificial intelligence post-processing in weather and climate: Proposed actions from the Oxford 2019 workshop.

Hauser, S., F. Teubler, M. Riemer, P. Knippertz, and C. M. Grams, 2023a: Life cycle dynamics of Greenland blocking from a potential vorticity perspective. *EGUsphere Preprint repository*.

———, 2023b: Towards a holistic understanding of blocked regime dynamics through a combination of complementary diagnostic perspectives. *Weather and Climate Dynamics*, **4**, 399–425.

Heide, D., L. V. Bremen, M. Greiner, C. Hoffmann, M. Speckmann, and S. Bofinger, 2010: Seasonal optimal mix of wind and solar power in a future, highly renewable Europe. *Renewable Energy*, **35**, 2483–2489.

Henderson, S. A. and E. D. Maloney, 2018: The impact of the Madden-Julian oscillation on high-latitude winter blocking during El Niño-Southern Oscillation events. *Journal of Climate*, **31**, 5293–5318.

Henderson, S. A., E. D. Maloney, and E. A. Barnes, 2016: The influence of the Madden-Julian oscillation on Northern Hemisphere winter blocking. *Journal of Climate*, **29**, 4597–4616.

Hersbach, H., 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, 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**, 1999–2049.

Hess, P. and H. Brezowsky, 1969: *Katalog der Grosswetterlagen Europas*, Vol. 15. 113th ed., Dt. Wetterdienst.

Hofmann, F., J. Hampp, F. Neumann, T. Brown, and J. Hörsch, 2021: atlite: A Lightweight Python Package for Calculating Renewable Power Potentials and Time Series. *Journal of Open Source Software*, **6**, 3294.

Hopson, T. M., 2014: Assessing the ensemble spread-error relationship. *Monthly Weather Review*, **142**, 1125–1142.

Hoskins, B. J. and D. J. Karoly, 1979: The Steady Linear Response of a Spherical Atmosphere to Thermal and Orographic Forcing. *Journal of the Atmospheric Sciences*, **38**, 1179–1196.

Huld, T., R. Gottschalg, H. G. Beyer, and M. Topič, 2010: Mapping the performance of PV modules, effects of module type and data averaging. *Solar Energy*, **84**, 324–338.

IRENA, 2019: *Renewable Capacity Statistics 2019*. URL `https://www.irena.org/publications/2019/Mar/Renewable-Capacity-Statistics-2019`.

Jenney, A. M., K. M. Nardi, E. A. Barnes, and D. A. Randall, 2019: The Seasonality and Regionality of MJO Impacts on North American Temperature. *Geophysical Research Letters*, **46**, 9193–9202.

Jones, N., 2025: A.I. Is Quietly Powering a Revolution in Weather Prediction. Tech. rep., URL `https://e360.yale.edu/features/artificial-intelligence-weather-forecasting`.

Jordan, A., F. Krüger, and S. Lerch, 2019: Evaluating Probabilistic Forecasts with scoringRules. *Journal of Statistical Software*, **90**.

Kaspar, F., M. Borsche, U. Pfeifroth, J. Trentmann, J. Drücke, and P. Becker, 2019: A climatological assessment of balancing effects and shortfall risks of photovoltaics and wind energy in Germany and Europe. *Advances in Science and Research*, **16**, 119–128.

Kautz, L. A., O. Martius, S. Pfahl, J. G. Pinto, A. M. Ramos, P. M. Sousa, and T. Woollings, 2022: Atmospheric blocking and weather extremes over the Euro-Atlantic sector - A review. *Weather and Climate Dynamics*, **3**, 305–336.

Kidston, J., A. A. Scaife, S. C. Hardiman, D. M. Mitchell, N. Butchart, M. P. Baldwin, and L. J. Gray, 2015: Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nature Geoscience*, **8**, 433–440.

Kumar, C., 2025: Blaming the blackout on net zero won't keep the lights on. URL `https://neweconomics.org/2025/04/blaming-the-blackout-on-net-zero-wont-keep-the-lights-on`.

Kurth, T., S. Subramanian, P. Harrington, J. Pathak, M. Mardani, D. Hall, A. Miele, K. Kashinath, and A. Anandkumar, 2023: FourCastNet: Accelerating Global High-Resolution Weather Forecasting Using Adaptive Fourier Neural Operators. *Proceedings of the Platform for Advanced Scientific Computing Conference, PASC 2023*, Association for Computing Machinery, Inc.

Lam, R., A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, A. Merose, S. Hoyer, G. Holland, O. Vinyals, J. Stott, A. Pritzel, S. Mohamed, and P. Battaglia, 2023: Learning skillful medium-range global weather forecasting. *Science*, **382**, 1416–1421.

Lang, A. L., K. Pegion, and E. A. Barnes, 2020: Introduction to Special Collection: "Bridging Weather and Climate: Subseasonal-to-Seasonal (S2S) Prediction". *Journal of Geophysical Research: Atmospheres*, **125**.

Lang, S., M. Alexe, M. Chantry, J. Dramsch, F. Pinault, B. Raoult, M. C. A. Clare, C. Lessig, M. Maier-Gerber, L. Magnusson, Z. B. Bouallègue, A. P. Nemesio, P. D. Dueben, A. Brown, F. Pappenberger, and F. Rabier, 2024: AIFS – ECMWF's data-driven forecasting system. *arXiv*.

Lee, J. C. K., R. W. Lee, S. J. Woolnough, and L. J. Boxall, 2020: The links between the Madden-Julian Oscillation and European weather regimes. *Theoretical and Applied Climatology*, **141**, 567–586.

Lee, R. W., S. J. Woolnough, A. J. Charlton-Perez, and F. Vitart, 2019: ENSO Modulation of MJO Teleconnections to the North Atlantic and Europe. *Geophysical Research Letters*, **46**, 13 535–13 545.

Lee, S. H., 2021: The stratospheric polar vortex and sudden stratospheric warmings. *Weather*, **76**, 12–13.

Legras, B. and M. Ghil, 1985: Persistent Anomalies, Blocking and Variations in Atmospheric Predictability. *Journal of the Atmospheric Sciences*, **42**, 433–471.

Lerch, S., S. Baran, A. Möller, J. Groß, R. Schefzik, S. Hemri, and M. Graeter, 2020: Simulation-based comparison of multivariate ensemble post-processing methods. *Nonlinear Processes in Geophysics*, **27**, 349–371.

Leutbecher, M. and T. N. Palmer, 2007: 514 Ensemble Forecasting. Tech. rep., URL https://www.ecmwf.int/sites/default/files/elibrary/2007/10729-ensemble-forecasting.pdf.

Lim, Y., S. W. Son, A. G. Marshall, H. H. Hendon, and K. H. Seo, 2019: Influence of the QBO on MJO prediction skill in the subseasonal-to-seasonal prediction models. *Climate Dynamics*, **53**, 1681–1695.

Lin, H., G. Brunet, and J. Derome, 2009: An observed connection between the North Atlantic oscillation and the Madden-Julian oscillation. *Journal of Climate*, **22**, 364–380.

Lindsey, R., 2021: Understanding the Arctic polar vortex. URL https://www.climate.gov/news-features/understanding-climate/understanding-arctic-polar-vortex.

Lorenz, E. N., 1968: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307.

———, 1995: Predictability: a problem partly solved. ECMWF, URL https://www.ecmwf.int/en/elibrary/75462-predictability-problem-partly-solved.

Madden, R. A. and P. R. Julian, 1971: Detection of a 40–50 Day Oscillation in the Zonal Wind in the Tropical Pacific. *Journal of the Atmoshperic Sciences*, **28**, 702–708.

———, 1972: Description of Global-Scale Circulation Cells in the Tropics with a 40–50 Day Period. *Journal of the Atmospheric Sciences*, 1109–1123.

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

Matthews, A. J., 2008: Primary and successive events in the Madden-Julian Oscillation. *Quarterly Journal of the Royal Meteorological Society*, **134**, 439–453.

Mayer, K. J. and E. A. Barnes, 2020: Subseasonal midlatitude prediction skill following Quasi-Biennial Oscillation and Madden–Julian Oscillation activity. *Weather and Climate Dynamics*, **1**, 247–259.

McGovern, A., R. Lagerquist, D. J. Gagne, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, 2019: Making the black box more transparent: Understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society*, **100**, 2175–2199.

Melhauser, C. and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *Journal of the Atmospheric Sciences*, **69**, 3350–3371.

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

Michelangeli, P.-A., R. Vautard, and B. Legras, 1995: Weather Regimes: Recurrence and Quasi Stationarity. *Journal of Atmospheric Sciences*, **52**, 1237–1256.

Millin, O. T., J. C. Furtado, and C. Malloy, 2024: The impact of North American winter weather regimes on electricity load in the central United States. *npj Climate and Atmospheric Science*, **7**.

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

Mockert, F., C. M. Grams, S. Lerch, M. Osman, and J. Quinting, 2024: Multivariate post-processing of probabilistic sub-seasonal weather regime forecasts. *Quarterly Journal of the Royal Meteorological Society*.

Mockert, F., C. M. Grams, S. Lerch, and J. Quinting, 2025: Windows of opportunity in subseasonal weather regime forecasting: A statistical-dynamical approach. *arXiv*.

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

———, 2025b: Pacific / North American Pattern. URL `https://www.cpc.ncep.noaa.gov/products/precip/CWlink/ENSO/verf/new.pna.shtml`.

———, 2025c: What is an MJO and how does it affect the weather? URL `psl.noaa.gov/mjo/MJOprimer/`.

OPSD, 2020: Time series renewable power plants. Tech. rep.

Osman, M., R. Beerli, D. Büeler, and C. M. Grams, 2023: Multi-model Assessment of Sub-seasonal Predictive Skill for Year-round Atlantic-European Weather Regimes . *Quarterly Journal of the Royal Meteorological Society*.

Otero, N., O. Martius, S. Allen, H. Bloomfield, and B. Schaefli, 2022: Characterizing renewable energy compound events across Europe using a logistic regression-based approach. *Meteorological Applications*, **29**.

Palmer, T., 2020: A Vision for Numerical Weather Prediction in 2030. *arXiv*.

Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic Parametrization and Model Uncertainty. Tech. rep.

Pozo-Vázquez, D., J. Tovar-Pescador, S. R. Gámiz-Fortis, M. J. Esteban-Parra, and Y. Castro-Díez, 2004: NAO and solar radiation variability in the European North Atlantic region. *Geophysical Research Letters*, **31**.

Rasp, S. and S. Lerch, 2018: Neural Networks for Postprocessing Ensemble Weather Forecasts. *Monthly Weather Review*.

Rex, D. F., 1950: Blocking Action in the Middle Troposphere and its Effect upon Regional Climate. *Tellus*, **2**, 275–301.

———, 1951: The Effect of Atlantic Blocking Action upon European Climate. *Tellus*, **3**, 100–112.

Roberts, C. D., M. A. Balmaseda, L. Ferranti, and F. Vitart, 2023: Euro-Atlantic Weather Regimes and Their Modulation by Tropospheric and Stratospheric Teleconnection Pathways in ECMWF Reforecasts. *Monthly Weather Review*, **151**, 2779–2799.

Robertson, A. W., F. Vitart, and S. J. Camargo, 2020: Subseasonal to Seasonal Prediction of Weather to Climate with Application to Tropical Cyclones. *Journal of Geophysical Research: Atmospheres*, **125**.

Rodriguez, L., 2023: Why are renewables being blamed for energy blackouts? URL `https://ratedpower.com/blog/energy-blackouts/`.

Salient, M. Deo, and A. Atlas, 2023: On the Money: How Salient's Novel S2S Forecasts Can Save the Agriculture Industry Millions of Dollars. URL `https://www.salientpredictions.com/blog/on-the-money-how-salients-novel-s2s-forecasts-can-save-the-agriculture-industry-millions-of-dollars`.

Santos-Alamillos, F. J., D. Pozo-Vázquez, J. A. Ruiz-Arias, L. V. Bremen, and J. Tovar-Pescador, 2015: Combining wind farms with concentrating solar plants to provide stable renewable power. *Renewable Energy*, **76**, 539–550.

Scaife, A. A., 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, 2022: Long-range predictability of extratropical climate and the length of day. *Nature Geoscience*.

Schefzik, R., 2017: Ensemble calibration with preserved correlations: unifying and comparing ensemble copula coupling and member-by-member postprocessing. *Quarterly Journal of the Royal Meteorological Society*, **143**, 999–1008.

Schefzik, R., T. L. Thorarinsdottir, and T. Gneiting, 2013: Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, **28**, 616–640.

Scherrer, S. C., C. Appenzeller, P. Eckert, and D. Cattani, 2004: Analysis of the Spread-Skill Relations Using the ECMWF Ensemble Prediction System over Europe. Tech. rep.

Scheuerer, M. and T. M. Hamill, 2015: Variogram-Based Proper Scoring Rules for Probabilistic Forecasts of Multivariate Quantities*. *Monthly Weather Review*, **143**, 1321–1334.

Schmidt, O., S. Melchior, A. Hawkes, and I. Staffell, 2019: Projecting the Future Levelized Cost of Electricity Storage Technologies. *Joule*, **3**, 81–100.

Sklar, A., 1959: Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, 229–231.

Song, J., W. Zhou, C. Li, and L. Qi, 2009: Signature of the Antarctic oscillation in the northern hemisphere. *Meteorology and Atmospheric Physics*, **105**, 55–67.

Spaeth, J., P. Rupp, M. Osman, C. M. Grams, and T. Birner, 2024: Flow-Dependence of Ensemble Spread of Subseasonal Forecasts Explored via North Atlantic-European Weather Regimes. *Geophysical Research Letters*, **51**.

Specq, D. and L. Batté, 2022: Do subseasonal forecasts take advantage of Madden–Julian oscillation windows of opportunity? *Atmospheric Science Letters*, **23**.

Stan, C., C. Zheng, E. K.-M. Chang, D. I. Domeisen, C. I. Garfinkel, A. M. Jenney, H. Kim, Y.-K. Lim, H. Lin, A. Robertson, C. Schwartz, F. Vitart, J. Wang, and P. Yadav, 2022: Advances in the prediction of MJO-Teleconnections in the S2S forecast systems. *Bulletin of the American Meteorological Society*.

Sun, L., H. Wang, and F. Liu, 2019: Combined effect of the QBO and ENSO on the MJO. *Atmospheric and Oceanic Science Letters*, **12**, 170–176.

Tachibana, Y., Y. Inoue, K. K. Komatsu, T. Nakamura, M. Honda, K. Ogata, and K. Yamazaki, 2018: Interhemispheric Synchronization Between the AO and the AAO. *Geophysical Research Letters*, **45**, 13,477–13,484.

Tagesschau, 2025: Warum Kraftwerke trotz Dunkelfalute nicht anspringen. URL `https://www.tagesschau.de/wirtschaft/energie/dunkelflauten-strom-preise-deutschland-102.html`.

Taillardat, M., O. Mestre, M. Zamo, and P. Naveau, 2016: Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics. *Monthly Weather Review*, **144**, 2375–2393.

TransnetBW, 2025: Adequacy 2050: Security of supply in the power system. Tech. rep., TransnetBW GmbH, URL `https://www.transnetbw.de/de/adequacy-2050`.

Trenberth, K., 2025: The Climate Data Guide: Nino SST Indices (Nino 1+2, 3, 3.4, 4; ONI and TNI). URL `https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni`.

Vannitsem, S., J. B. Bremnes, J. Demaeyer, G. R. Evans, J. Flowerdew, S. Hemri, S. Lerch, N. Roberts, S. Theis, A. Atencia, Z. B. Bouallègue, J. Bhend, M. Dabernig, L. D. Cruz, L. Hieta, O. Mestre,

L. Moret, I. O. Plenković, M. Schmeits, M. Taillardat, J. V. D. Bergh, B. V. Schaeybroeck, K. Whan, and J. Ylhaisi, 2021: Statistical Postprocessing for Weather Forecasts-Review, Challenges and Avenues in a Big Data World. Tech. rep.

Vautard, R., 1990: Multiple Weather Regimes over the North Atlantic: Analysis of Precursors and Successors. *Monthly Weather Review*, **118**, 2056–2081.

Vautard, R. and B. Legras, 1988: On the Source of Midlatitude Low-Frequency Variability. Part II: Nonlinear Equilibration of Weather Regimes. *Journal of the Atmsopheric Sciences*, **45**, 2845–2867.

Vitart, F., 2013: Evolution of ECMWF sub-seasonal forecast skill scores over the past 10 years. Tech. rep., URL `https://www.ecmwf.int/sites/default/files/elibrary/2013/12932-evolution-ecmwf-sub-seasonal-forecast-skill-scores-over-past-10-years.pdf`.

———, 2017: Madden—Julian Oscillation prediction and teleconnections in the S2S database. *Quarterly Journal of the Royal Meteorological Society*, **143**, 2210–2220.

Vitart, F., 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, 2017: The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, **98**, 163–173.

Vitart, F. and R. Mladek, 2023: ECMWF Model. URL `https://confluence.ecmwf.int/display/S2S/ECMWF+Model`.

Vitart, F., F. Molteni, and T. Jung, 2010: Prediction of the Madden-Julian Oscillation and its impact on the European weather in the ECMWF monthly forecasts. URL `https://www.ecmwf.int/sites/default/files/elibrary/2012/12942-prediction-madden-julian-oscillation-and-its-impact-european-weather-ecmwf-monthly-forecasts.pdf`.

Vitart, F. and A. W. Robertson, 2019: *Sub-Seasonal to Seasonal Prediction: Chapter 1 - Introduction: Why Sub-seasonal to Seasonal Prediction (S2S)?* Elsevier, 3-15 pp., URL `https://www.sciencedirect.com/science/article/pii/B9780128117149000012`.

Wandel, J., D. Büeler, P. Knippertz, J. F. Quinting, and C. M. Grams, 2024: Why Moist Dynamic Processes Matter for the Sub-Seasonal Prediction of Atmospheric Blocking Over Europe. *Journal of Geophysical Research: Atmospheres*, **129**.

Wang, S., M. K. Tippett, A. H. Sobel, Z. K. Martin, and F. Vitart, 2019: Impact of the QBO on Prediction and Predictability of the MJO Convection. *Journal of Geophysical Research: Atmospheres*, **124**, 11 766–11 782.

Wanner, H., S. Brönnimann, C. Casty, D. Gyalistras, J. Luterbacher, C. Schmutz, D. B. Stephenson, and E. Xoplaki, 2001: North Atlantic Oscillation - Concepts and Studies. *Surveys in Geophysics*, **22**, 321–382.

Waugh, D. W., A. H. Sobel, and L. M. Polvani, 2017: What is the polar vortex and how does it influence weather?

Wernli, H. and L. Papritz, 2018: Role of polar anticyclones and mid-latitude cyclones for Arctic summertime sea-ice melting. *Nature Geoscience*, **11**, 108–113.

Wheeler, M. C. and H. H. Hendon, 2004: An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction. *Monthly Weather Review*, **132**, 1917–1932.

White, C. J., 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, 2017: Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological Applications*, **24**, 315–325.

White, C. J., D. I. Domeisen, N. Acharya, E. A. Adefisan, M. L. Anderson, S. Aura, A. A. Balogun, D. Bertram, S. Bluhm, D. J. Brayshaw, J. Browell, D. Büeler, A. Charlton-Perez, X. Chourio, I. Christel, C. A. Coelho, M. J. Deflorio, L. D. Monache, F. D. Giuseppe, A. M. García-Solórzano, P. B. Gibson, L. Goddard, C. G. Romero, R. J. Graham, R. M. Graham, C. M. Grams, A. Halford, W. T. Huang, K. Jensen, M. Kilavi, K. A. Lawal, R. W. Lee, D. Macleod, A. Manrique-Suñén, E. S. Martins, C. J. Maxwell, W. J. Merryfield, Ángel G. Muñoz, E. Olaniyan, G. Otieno, J. A. Oyedepo, L. Palma, I. G. Pechlivanidis, D. Pons, F. M. Ralph, D. S. Reis, T. A. Remenyi, J. S. Risbey, D. J. Robertson, A. W. Robertson, S. Smith, A. Soret, T. Sun, M. C. Todd, C. R. Tozer, F. C. Vasconcelos, I. Vigo, D. E. Waliser, F. Wetterhall, and R. G. Wilson, 2022: Advances in the Application and Utility of Subseasonal-to-Seasonal Predictions. *Bulletin of the American Meteorological Society*, **103**, E1448–E1472.

Wiel, K. V. D., L. P. Stoop, B. R. V. Zuijlen, R. Blackport, M. A. V. D. Broek, and F. M. Selten, 2019a: Meteorological conditions leading to extreme low variable renewable energy production and extreme high energy shortfall. *Renewable and Sustainable Energy Reviews*, **111**, 261–275.

———, 2019b: The influence of weather regimes on European renewable energy production and demand. *Environmental Research Letters*, **14**.

Wilks, D. S., 2011a: *Statistical Methods in the Atmospheric Sciences: Forecast Verification*, Vol. 100. 301-394 pp.

———, 2011b: *Statistical Methods in the Atmospheric Sciences: Statistical Forecasting*, Vol. 100. 215-300 pp.

Yoo, C. and S. W. Son, 2016: Modulation of the boreal wintertime Madden-Julian oscillation by the stratospheric quasi-biennial oscillation. *Geophysical Research Letters*, **43**, 1392–1398.

Zavadoff, B. and M. Arcodia, 2022: What are teleconnections? Connecting Earth's climate patterns via global information superhighways. URL `https://www.climate.gov/news-features/blogs/enso/what-are-teleconnections-connecting-earths-climate-patterns-global`.

Zhang, C., 2005: Madden-Julian Oscillation. *Reviews of Geophysics*, **43**, 1–36.

Zhang, C. and B. Zhang, 2018: QBO-MJO Connection. *Journal of Geophysical Research: Atmospheres*, **123**, 2957–2967.

Zhang, F., Y. Q. Sun, L. Magnusson, R. Buizza, S. J. Lin, J. H. Chen, and K. Emanuel, 2019: What is the predictability limit of midlatitude weather? *Journal of the Atmospheric Sciences*, **76**, 1077–1091.

Zheng, C., E. K. Chang, H. Kim, M. Zhang, and W. Wang, 2018: Impacts of the Madden-Julian Oscillation on Storm-Track Activity, Surface Air Temperature, and Precipitation over North America. *Journal of Climate*, **31**, 6113–6134.

Zubiate, L., F. McDermott, C. Sweeney, and M. O'Malley, 2017: Spatial variability in winter NAO–wind speed relationships in western Europe linked to concomitant states of the East Atlantic and Scandinavian patterns. *Quarterly Journal of the Royal Meteorological Society*, **143**, 552–562.

Ångström, A., 1935: Teleconnections of Climatic Changes in Present Time. *Geografiska Annaler*, **17**, 242–258.

Acknowledgments

The past three and a half years have certainly been a scientific challenge – from dry spells in the search for the right methods and metrics to receiving awards for my research and teaching. I would like to take this opportunity to sincerely thank everyone who stood by me over the years, supported and encouraged me during tough phases, and celebrated successes with me.

First and foremost, I would like to thank Christian Grams who, although not officially part of my reviewer team in the end, has been by my academic side from the very first moment I decided to pursue a career in meteorology. From supervising me as a student assistant and offering me Bachelor's and Master's theses, to welcoming me into his team for my PhD position, he has supported me at every stage. His extensive meteorological expertise still fascinates me, and his rigorous but constructive feedback on my paper drafts made the official review process a surprisingly pleasant one. Although he moved to MeteoSwiss during my time as a doctoral student, he remained available when I needed his input.

A big thank you also goes to Julian Quinting, who not only accepted me into his working group after Christian Grams left KIT but also agreed to be my official second reviewer. No matter the time of day – even early in the morning – Julian's door was always open, and he was always ready to listen to my doubts and took them away from me. The best ideas probably came up together during our bike rides home.

I would also like to thank Sebastian Lerch. Through the "KIT Graduate School on Computational and Data Science," Sebastian was part of my unofficial supervision team from the beginning, and after his move to the University of Marburg for his professorship, I am happy to now have him as an official external reviewer. Sebastian introduced me to the world of artificial intelligence during my Master's and sparked my enthusiasm for the subject. I look forward to continuing to work with him at the interface of meteorology, renewable energies, and artificial intelligence.

My sincere thanks also go to Peter Knippertz, who agreed to take on the official role of first reviewer. His detailed feedback on the thesis itself, along with his critical input over the years, helped me enormously in gaining an external perspective on my work.

I am also grateful to two people who played a key role in my academic journey: Tom Brown and Fabian Neumann (now at TU Berlin). Together with Christian Grams, they offered me my first student assistant position and introduced me to the world of energy systems.

Thanks also go to Roswitha Marioth, Rosalba Gräbner, Alexandra Beideck, and Gabi Klinck for their administrative and IT support, as well as to Angela Hühnerfuß for organizing the graduate school and to Uwe Ehret for being part of my graduate school supervision team.

I would also like to thank all my colleagues from my three working groups: the “Large-scale Dynamics and Predictability” group led by Christian Grams, the “Artificial Intelligence for Probabilistic Weather Forecasting” group led by Sebastian Lerch, and the “Meteorological Data Science” group led by Julian Quinting. A special thank you goes to Marisol Osman, who helped me a lot with the creation of my dataset, guided me through “Frankenstein’s monster,” and was always available for scientific discussions. A final academic thank you also goes to my loyal virtual “colleague”, who tried to answer my scientific questions day and night, helped me with programming, and corrected my grammar while writing the thesis.

Last but not least, I would like to thank my family and friends. Thank you for your unconditional love and support throughout my academic journey. Thank you to Lea, who has been with me throughout my PhD – through all the ups and downs – endlessly listening to my current challenges, but most importantly, actively encouraging me to switch off, stop thinking about research, and enjoy life in nature. None of this would have been possible without you.