Check for updates

Received: 2 April 2024 | Revised: 25 June 2024 | Accepted: 9 August 2024

DOI: 10.1002/qj.4840

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Multivariate post-processing of probabilistic sub-seasonal weather regime forecasts

Fabian Mockert[1] | **Christian M. Grams**[1] | **Sebastian Lerch**[2,3] | **Marisol Osman**[1] | **Julian Quinting**[1]

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

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

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

Correspondence
Fabian Mockert, Institute of Meteorology and Climate Research Troposphere Research (IMKTRO), Karlsruhe Institute of Technology (KIT), PO Box 3640, 76021 Karlsruhe, Germany.
Email: fabian.mockert@kit.edu

Present address
Christian M. Grams, Federal Office of Meteorology and Climatology, MeteoSwiss, Zurich Airport, Switzerland; Marisol Osman, Facultad de Ciencias Exactas y Naturales, Departamento de Ciencias de la Atmósfera y los Océanos, Universidad de Buenos Aires, Buenos Aires, Argentina; Marisol Osman, CONICET–Universidad de Buenos Aires, Centro de Investigaciones del Mar y la Atmósfera (CIMA), Buenos Aires, Argentina; and Marisol Osman, CNRS–IRD–CONICET–UBA, Instituto Franco-Argentino para el Estudio del Clima y sus Impactos (IRL 3351 IFAECI), Buenos Aires, Argentina

Funding information
KIT Centre MathSEE; Axpo Solutions AG; Vector Stiftung; Helmholtz-Gemeinschaft, Grant/Award Number: VH-NG-1243; European Research Council, Grant/Award Number: 101077260

Abstract

Reliable forecasts of quasi-stationary, recurrent, and persistent large-scale atmospheric circulation patterns—so-called weather regimes—are crucial for various socio-economic sectors, including energy, health, and agriculture. Despite steady progress, probabilistic weather regime predictions still exhibit biases in the exact timing and amplitude of weather regimes. This study thus aims at advancing probabilistic weather regime predictions in the North Atlantic–European region through ensemble post-processing. Here, we focus on the representation of seven year-round weather regimes in sub-seasonal to seasonal reforecasts of the European Centre for Medium-Range Weather Forecasts (ECMWF). The manifestation of each of the seven regimes can be expressed by a continuous weather regime index, representing the projection of the instantaneous 500-hPa geopotential height anomalies ($Z500A$) onto the respective mean regime pattern. We apply a two-step ensemble post-processing involving first univariate ensemble model output statistics and second ensemble copula coupling, which restores the multivariate dependence structure. Compared with 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 1 day (from 13.5 to 14.5 days). Additionally, to our knowledge our study is the first to evaluate the multivariate aspects of forecast quality systematically for weather regime forecasts. Our method outperforms current practices in the multivariate aspect, as measured by the energy and variogram score. Still, our study shows that, even with advanced post-processing, weather regime prediction becomes difficult beyond 14 days, which likely points towards intrinsic limits of predictability for daily/instantaneous regime forecasts. 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.

KEYWORDS

ensemble copula coupling, ensemble model output statistics, forecasting, post-processing, weather regimes

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2024 The Author(s). *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society.

1 | INTRODUCTION

Weather regimes, defined as quasi-stationary, recurrent, and persistent large-scale circulation patterns (Michelangeli *et al.*, 1995; Vautard, 1990), provide valuable information for decision-making in the energy (Bloomfield *et al.*, 2021; Mockert *et al.*, 2023), health (Charlton-Perez *et al.*, 2019), and agricultural (Lavaysse *et al.*, 2018) sectors. These weather regimes are associated with distinct conditions on surface variables, including 2-m temperature, 10-m wind, and radiation, and thus prove beneficial for extended-range prediction. The representation of complex large-scale circulations through a finite set of states (e.g., weather regimes) facilitates the interpretation and categorisation of the prevailing large-scale circulation and its impact on surface weather. In the context of renewable energy forecasts for the European region, Bloomfield *et al.* (2021) conducted a comprehensive study comparing grid-point-based forecasting methods with pattern-based methods (including weather regimes). While grid-point forecasts exhibit superior skill up to 10 days lead time, pattern-based methods demonstrate better performance at extended-range lead times (12+ days).

In this article, we adopt the year-round definition of seven North Atlantic–European weather regimes proposed by Grams *et al.* (2017). These weather regimes, illustrated in Supplement S1 in the Supporting Information, represent large-scale circulation patterns within the 500-hPa geopotential height field ($Z500$). The regime definition is rooted in continuous information about the amplitude of the seven weather regimes via a normalised weather regime index (IWR). The seven-dimensional IWR vector thus provides additional information about the current regime characteristics beyond a mere categorisation, which proved to be useful in sub-seasonal prediction of weather regimes (cf. discussion in Grams *et al.*, 2020).

However, extended-range forecasts of $Z500$ have substantial biases (e.g., Büeler *et al.*, 2021; Ferranti *et al.*, 2018). One commonly used method to address these mean biases involves calibrating the $Z500$ field before computing weather regimes (cf. Büeler *et al.*, 2021). Instead of computing the $Z500$ anomalies relative to the reanalysis climatology, this method computes anomalies relative to the 90-day running-mean reforecast climatology at the respective lead time. While this correction mitigates the $Z500$ bias in the forecast field, it does not address all systematic errors, for example, the flow dependence of systematic errors. Additionally, this specific approach is impractical for operational on-the-fly reforecasts like those produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), due to the need for compromises regarding averaging windows when computing running-mean climatologies.

The presence of systematic errors and biases is a common challenge in ensemble weather forecasting (Lerch *et al.*, 2020; Vannitsem *et al.*, 2021). Addressing this challenge can involve statistical post-processing methods, where systematic forecasting errors are corrected by analysing the statistical error distribution of past forecasts. Most research efforts have been focused on univariate post-processing methods where different weather variables, locations, or lead times are treated separately. However, many practical applications require accurate representations of temporal, spatial, and inter-variable dependences. Preserving these multivariate dependence structures is crucial. In low-dimensional scenarios, this could be achieved by fitting a specific multivariate probability distribution (Schefzik & Möller, 2018). A more broadly applicable strategy involves a two-step process. In the first step, forecasting variables are post-processed individually in all dimensions. Then, in the second step, the multivariate dependence structure is restored by rearranging univariate sample values based on the rank order structure of a specific multivariate dependence template. From a mathematical perspective, this second step corresponds to applying a parametric or non-parametric copula. Commonly used multivariate approaches include ensemble copula coupling (Schefzik *et al.*, 2013), the Schaake Shuffle (Clark *et al.*, 2004), or a Gaussian copula approach (Möller *et al.*, 2013).

The combination of ensemble model output statistics (EMOS: Gneiting & Raftery, 2007) and ensemble copula coupling (ECC: Schefzik *et al.*, 2013) grounded in Sklar's theorem from multivariate statistics (Schefzik *et al.*, 2013), EMOS-ECC, has proven to be effective across various meteorological forecast variables. Schefzik *et al.* (2013) explored different EMOS-ECC configurations for sea-level pressure forecasts and consistently found its superior performance compared with alternative calibration techniques. Additionally, the configuration where equidistant quantiles are drawn from the forecast distribution (EMOS-ECC-Q) outperformed other EMOS-ECC configurations in the context of pressure forecasts. In a study by Scheuerer and Hamill (2015) on wind-speed forecasts, EMOS-ECC-Q consistently outperformed raw ensemble forecasts, EMOS-Q (without restoring multivariate dependences), and another EMOS-ECC configuration. Applying the EMOS-ECC approach from Schefzik *et al.* (2013) to temperature forecasts, Schefzik (2017) compared it with a member-by-member post-processing (MBMP) method. Schefzik's (2017) findings indicate that both methods exhibit good performance, with EMOS-ECC consistently outperforming MBMP in predictive skill. Further, various comparative studies of multivariate post-processing methods have found that the differences between different variants of ECC or observation-based approaches

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840 by Karlsruher Institut F., Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

such as the Schaake Shuffle tend to be small, and that EMOS-ECC-Q usually constitutes a competitive benchmark (e.g., Lakatos *et al.*, 2023; Lerch *et al.*, 2020; Perrone *et al.*, 2020; Wilks, 2015). Commonly, as also in the studies mentioned above, post-processing is applied directly to meteorological forecast variables, for example, temperature, precipitation, or wind-speed forecasts. The promising results from these recent studies motivate us to apply the EMOS-ECC-Q approach to our problem at hand, enabling comprehensive multivariate post-processing. Instead of directly post-processing a forecast variable (e.g., $Z500$) generated by the forecasting model, we post-process the weather regime index forecast, which is derived from the $Z500$ forecast.

In our study, we develop a two-step post-processing method that adapts EMOS-ECC to forecasts of the continuous weather regime index IWR. In a first step, we address the univariate marginal distributions of each weather regime index independently. Subsequently, in the second step, we restore the multivariate dependence structure among weather regimes by applying a copula function. This function learns the dependence structure from the raw weather regime index forecasts. To our knowledge, our study is the first to develop multivariate post-processing methods for the weather regime index, and presents the first comprehensive evaluation of the multivariate probabilistic forecast skill for weather regime forecasts.

The structure of the article is outlined as follows. In Section 2, we introduce the reforecasts, the weather regime index, and the statistical post-processing methods. Additionally, we discuss the scoring rules used to evaluate the forecasts. Section 3 starts with a brief discussion of the mean biases in ECMWF Integrated Forecasting System (IFS) reforecasts and is then divided into three parts. In Section 3.2 we analyse the univariate skill. Then we shift our focus to assess multivariate skill (Section 3.3) and, lastly, we test the sensitivity of the EMOS-ECC approach to the frequency of reforecast initialisations and the historical period covered by the reforecasts (Section 3.4). In Section 4 we conclude and discuss our findings, and give an outlook on further research avenues.

2 | DATA AND METHODS

2.1 | ECMWF reforecast and reanalysis

For our study, we utilise sub-seasonal to seasonal reforecast data by ECMWF, provided through the Subseasonal-to-Seasonal (S2S) Prediction Project Database (Vitart *et al.*, 2017). To increase the number of forecast initial dates available for our analysis, we merge forecasts from two consecutive model cycles, Cy46R1 and Cy47R1 (Vitart & Mladek, 2023). These reforecasts are computed twice a week (Mondays and Thursdays) and consist of 11 ensemble members, covering a forecast lead time of 0–46 days with 91 vertical levels and a native horizontal grid spacing of 16 km up to day 15 and 32 km from day 15 onwards. Forecast data were remapped from their native resolution to a regular latitude–longitude grid with 1° grid spacing. The two model cycles were operational from June 11, 2019, to May 11, 2021, with a cycle change on June 30, 2020. As a result, for the period between May 11 and June 30, reforecasts are only available from Cy46R1, which may impact the training and evaluation of our methods within that specific time period. It is likely that in this period the post-processing performs worse, due to fewer training dates. Nonetheless, with the combination of Cy46R1 and Cy47R1, our dataset spans 21 years of reforecasts, from June 11, 1999–May 11, 2020, comprising a total of 4000 initial dates, each with 11 ensemble members. The reforecasts are initialised using the ECMWF Reanalysis v5 (ERA5) data (Hersbach *et al.*, 2020). For verification purposes, we treat the ERA5 dataset as a "perfect ensemble member" by matching ERA5 data to each initialisation date and lead time (cf. Wandel *et al.*, 2024). Further, the ERA5 data are remapped to the same grid spacing as the reforecasts.

2.2 | Weather regimes

In this study, we use the seven year-round North Atlantic–European 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, 2023b) 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 North Atlantic–European region (30–90°N, 80°W–40°E). In brief, we conduct an empirical orthogonal function (EOF) analysis of six-hourly (1979–2019), 10-day low-pass filtered (filter width of 20 days, hence ±10 days), seasonally normalised geopotential height anomalies ($Z500A$, relative to 91-day running-mean climatology) within the domain of the weather regimes. We then apply a k-means clustering algorithm on the first seven EOFs and set $k = 7$. These seven clusters represent the seven distinct weather regimes (Figure S1) originally introduced by Grams *et al.* (2017), with three cyclonic (Atlantic Trough (AT), Zonal Regime (ZO), and Scandinavian Trough (ScTr)) and four anticyclonic regime types (Atlantic Ridge (AR), European Blocking (EuBL), Scandinavian Blocking (ScBL), and Greenland Blocking (GL)).

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840 by Karlsruher Institut F., Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 briefly outline the steps to compute the weather regime index (see also Figure S2) for two sets of forecasts, raw and $Z500$ bias-corrected (notation adopted by Osman *et al.* (2023)), and refer to Büeler *et al.* (2021) and Osman *et al.* (2023) for a more detailed description of the computation of the IWR.

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 S2, top middle panel). Subsequently, we then apply a low-pass filter and normalisation to obtain standardised and filtered $Z500A$ (Φ_{m^*}), where m^* denotes all 11 ensemble members and the ERA5 perfect member.

Computing $Z500A$ directly 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 S2, 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–May 2015, excluding the subsequent test period of June 2015–May 2020. This ensures that model data from the test period are 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

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

Here, $P_{(\mathrm{wr},m^*)}(t,\tau)$ 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 $\Phi_{(\mathrm{wr})}(\lambda,\varphi)$ for the weather regimes $\mathrm{wr}\in\{AT, ZO, ScTr, AR, EuBL, ScBL, GL\}$. Here, λ and φ denote the longitudinal and latitudinal degrees, respectively.

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

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

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 1979–2019. For the $Z500$ bias-corrected forecasts, i iterates over all available dates and ensemble members in the model forecast data from June 1999–May 2015 (the reduced training set). Combining the weather regime index for each weather regime yields a seven-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 S2, bottom right panel).

2.3 | Statistical post-processing

We employ a two-step statistical post-processing method based on Sklar's theorem (Lerch *et al.*, 2020; Schefzik *et al.*, 2013; Sklar, 1959) for the seven-dimensional weather regime index forecast: first, univariate processing using ensemble model output statistics EMOS (Gneiting & Raftery, 2007), and, second, multivariate processing through ensemble copula coupling ECC (Schefzik *et al.*, 2013). We adapt the notation of Lerch *et al.* (2020) and Chen *et al.* (2024) with slight modifications for our specific setup.

According to Sklar's theorem, a multivariate cumulative distribution function (CDF) H can be decomposed into a copula function C representing the dependence structures and its marginal univariate CDFs $F_{AT}, \ldots, F_{GL}$ obtained through univariate post-processing. Specifically, for $x_{AT}, \ldots, x_{GL} \in \mathbb{R}$, we have $H(x_{AT}, \ldots, x_{GL}) = C(F_{AT}(x_{AT}), \ldots, F_{GL}(x_{GL}))$ (Lerch *et al.*, 2020), where the subscript iterates through the seven weather regimes $wr \in \{AT, ZO, ScTr, AR, EuBL, ScBL, GL\}$ and x represents the weather regime index.

The unprocessed 7D ensemble forecast of the weather regime index with $M = 11$ ensemble members is denoted

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840 by Karlsruher Institut F., Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

as $\mathbf{X}_1, \ldots, \mathbf{X}_M \in \mathbb{R}^7$, where $\mathbf{X}_m = \left(X_m^{(AT)}, \ldots, X_m^{(GL)}\right)$. Similarly, the observations of the weather regime index are denoted as $\mathbf{y} = \left(y^{(AT)}, \ldots, y^{(GL)}\right) \in \mathbb{R}^7$.

2.3.1 | Ensemble model output statistics

The first step of the two-step post-processing method is to apply EMOS univariately to fit a Gaussian predictive distribution $\mathcal{N}^{(\mathrm{wr})}$ with mean μ and variance σ^2,

$$y^{(\mathrm{wr})} | X_1^{(\mathrm{wr})}, \ldots, X_M^{(\mathrm{wr})} \sim \mathcal{N}(\mu, \sigma^2) = F_\theta^{(\mathrm{wr})}, \tag{3}$$

where the distribution parameters $\theta = (\mu, \sigma)$ are linked to the ensemble forecasts via $\boldsymbol{\theta} = g(X_1, \ldots, X_M)$.

The parameters

$$(\mu, \sigma^2) = (a_0 + a_1 \overline{X}, b_0 + b_1 S^2) = g\left(X_1^{(\mathrm{wr})}, \ldots, X_M^{(\mathrm{wr})}\right) \tag{4}$$

are determined by minimising the continuous ranked probability score (CRPS) via optimisation of a_0, a_1, b_0, b_1 on a training period from June 1999–May 2015 (Figure 1a,b). The resulting model for each weather regime index is referred to as EMOS-G. Initially, we conducted the above-mentioned univariate post-processing step using different setups for the training period to discern variations in performance. These setups involved training one EMOS method on the full dataset, splitting it into two seasons (winter half-year (October–March) and summer half-year (April–September)), four seasons (winter (December–February), spring (March–May), summer (June–August), and autumn (September–November)), or training EMOS for each calendar day using running windows of 9, 31, and 91 days. The performance of all methods was comparable, with a slight tendency towards better performance for the four-season approach and a running window of 31 days. We therefore focus our analysis on the 31-day running-window setup, as it is amongst the best-performing setups and compatible with on-the-fly generated reforecast data.

For each weather regime index, we reduce the continuous Gaussian forecast distributions to an ensemble with the same number of members ($M = 11$) as the unprocessed forecast $\mathbf{X}_1, \ldots, \mathbf{X}_M$ (Figure 1b). This process involves drawing equidistant quantiles at levels $1/(M+1), \ldots, M/(M+1)$ from the forecast distributions $F_\theta^{(\mathrm{wr})}$ and the resulting forecasts are referred to as EMOS-Q (Figure 1c).

2.3.2 | Ensemble copula coupling

In the second step, we utilise ECC to restore the multivariate dependence structure of the weather regime index.

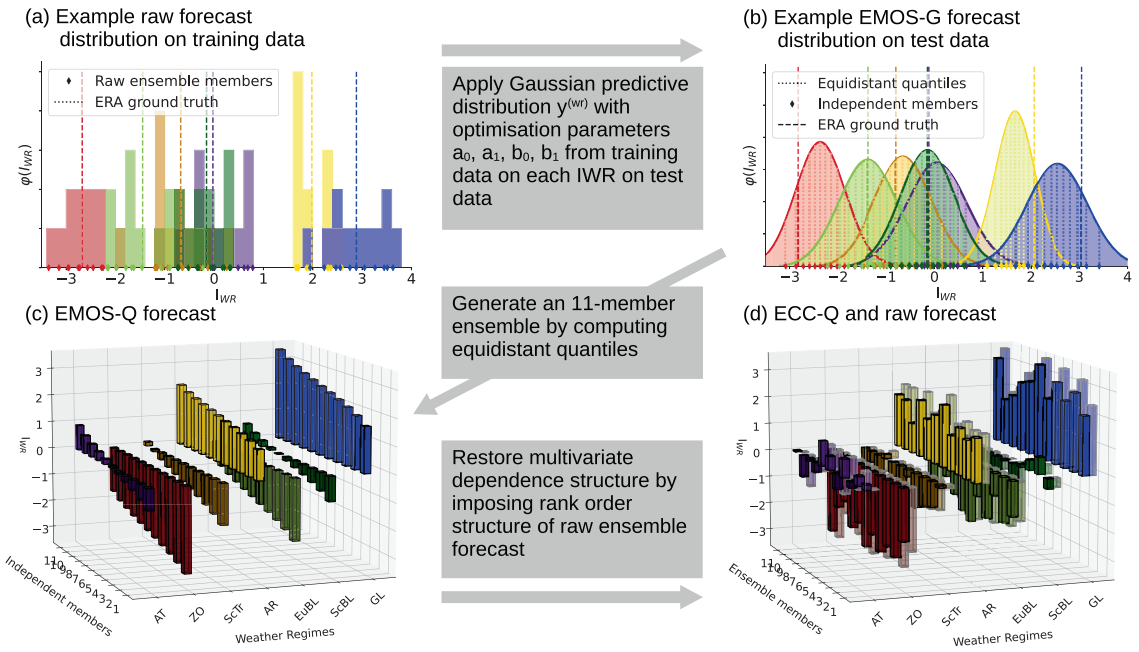

FIGURE 1 Post-processing workflow for the seven-dimensional weather regime index using EMOS-ECC. (a) Initially, the optimisation parameters are fitted by minimising the continuous ranked probability score on the raw ensemble forecasts in the training data. The obtained optimising parameters from the training data are then applied to 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 dependence structure of the raw forecast (transparent bars in d). [Colour figure can be viewed at wileyonlinelibrary.com]

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840 by Karlsruher Institut F., Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

We achieve this by retrieving the rank order structure of the unprocessed ensemble member forecasts (Figure 1d, pale bars) and sorting the post-processed ensemble members accordingly (Figure 1c,d). To formalise this process, for each weather regime index we define $\sigma_{(\mathrm{wr})}(m) = \mathrm{rank}\left(X_m^{(\mathrm{wr})}\right)$ as a permutation, and aim to find $\tilde{X}_m^{(\mathrm{wr})} = \hat{x}_{\sigma_{(\mathrm{wr})}(m)}^{(\mathrm{wr})}$, where $\hat{x}_1^{(\mathrm{wr})}, \dots, \hat{x}_M^{(\mathrm{wr})}$ form a sample under the assumption that $\hat{x}_1^{(\mathrm{wr})} \leq \cdots \leq \hat{x}_M^{(\mathrm{wr})}$ to simplify the notation. Here,

$$\hat{x}_1^{(\mathrm{wr})} := \left(F_\theta^{(\mathrm{wr})}\right)^{-1}\left(\frac{1}{M+1}\right), \dots, \hat{x}_M^{(\mathrm{wr})} := \left(F_\theta^{(\mathrm{wr})}\right)^{-1}\left(\frac{M}{M+1}\right)$$

represent the quantile-based EMOS-Q predictions of the weather regime indices. This non-parametric, empirical copula approach is referred to as ECC-Q. Unlike in the Introduction, for the sake of simplicity we will refer to the execution of the two-step process as ECC and not EMOS-ECC, since we did not test any other univariate methods. Next to ECC, we also tested the Schaake Shuffle approach, which ranks ensemble members based on past observations rather than the actual forecast.

2.4 | Skill metrics and their skill scores

In this study, we aim to compare the univariate and multivariate skill of the post-processed ensemble forecasts using EMOS-G and ECC-Q with the skill of the current practice of processing the $Z500$ field prior to computing the weather regime index. To conduct this comparison, we introduce the continuous ranked probability skill score (CRPSS) for univariate evaluation and the energy skill score (ESS) and variogram skill score (VSS) for assessing multivariate aspects of forecast quality. While these scores are discussed extensively in Gneiting and Raftery (2007) and Scheuerer and Hamill (2015), we will provide a concise summary of the underlying metrics and skill scores in this work and direct interested readers to the literature mentioned for more detailed information.

The CRPS is a metric used for evaluating univariate probabilistic forecasts and generalises the absolute error to which it reduces when the forecast is deterministic. It is defined as

$$\mathrm{CRPS}(F, y) = \int_{-\infty}^{\infty} \left(F(x) - \mathbb{1}\{x \geq y\}\right)^2 dx, \tag{5}$$

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

To assess the skill of multivariate probabilistic forecasts, we employ the energy score (ES) and the variogram score (VS) of order p (VS^p). The ES is a generalisation of the CRPS and the variogram score originates in the concept of variograms (also referred as structure functions) from geostatistics. The ES is calculated as

$$ES(F, \mathbf{y}) = \frac{1}{M}\sum_{i=1}^{M}\|\mathbf{X}_i - \mathbf{y}\| - \frac{1}{2M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\|\mathbf{X}_i - \mathbf{X}_j\|, \tag{6}$$

where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^D$, $\mathbf{X}_i, \mathbf{X}_j$ are samples from the multivariate forecast distribution, and $\mathbf{y}$ represents the multivariate observation, respectively. The VS^p is given by

$$VS^p(F, \mathrm{y}) = \sum_{i=1}^{D}\sum_{j=1}^{D} w_{i,j}\left(\left|y^{(i)} - y^{(j)}\right|^p - \frac{1}{M}\sum_{k=1}^{M}\left|X_k^{(i)} - X_k^{(j)}\right|^p\right)^2, \tag{7}$$

with $w_{i,j}$ being a non-negative weight for pairs of component combinations and p representing the order of the VS. In accordance with Scheuerer and Hamill (2015), we use an unweighted version of the VS with $w_{i,j} = 1$ and set $p = 0.5$. The values of both the energy and the variogram score can be interpreted similarly to the CRPS, where lower values represent better forecast skill.

The skill scores (SS_f) for the mentioned skill metrics ($S \in \{CRPS, ES, VS^p\}$) are calculated as

$$SS_\mathrm{f} = \frac{\overline{S_\mathrm{ref}} - \overline{S_\mathrm{f}}}{\overline{S_\mathrm{ref}} - \overline{S_\mathrm{opt}}} = 1 - \frac{\overline{S_\mathrm{f}}}{\overline{S_\mathrm{ref}}}, \tag{8}$$

where $\overline{S_\mathrm{f}}$ is the mean score of a forecasting method f. For the skill metrics considered here, a perfect score $\overline{S_\mathrm{opt}}$ equals 0. As a reference forecast ($\overline{S_\mathrm{ref}}$), we generally use a 31-day ensemble climatology, where the ensemble members are represented by reanalysis data for the S2S reforecast dates throughout the training period of June 1999–May 2015 of the raw weather regime index (Figure S2, bottom right panel). A value of 1 indicates perfect skill, a value of 0 equal skill, and negative values worse skill than the reference forecast.

2.5 | Diebold–Mariano test of equal performance

To assess the statistical significance of the differences in predictive performance between the post-processed forecast, raw forecast, $Z500$ bias-corrected forecast, and climatological reference forecast, we employ the Diebold–Mariano test of equal performance (Diebold & Mariano, 1995), the test statistic t_n of which is given by

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840 by Karlsruher Institut F., Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$t_n = \sqrt{n}\frac{\overline{S_n^F} - \overline{S_n^G}}{\hat{\sigma}_n}, \tag{9}$$

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, respectively. Following Gneiting and Katzfuss (2014) and assuming independence between the score differentials, we estimate the standard deviation by $\hat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(S(F_i, \mathbf{y}_i) - S(G_i, \mathbf{y}_i)\right)^2}$. Under standard regularity assumptions, t_n asymptotically follows a standard Gaussian distribution. Negative values of t_n indicate that F outperforms G with respect to the considered score S. We use a level of $\alpha = 0.05$ to assess the significance of the performance. Values falling outside this level (indicated with grey shading in Figures 6 and 8) are considered statistically significant.

2.6 | Verification rank histograms

Verification rank histograms are essential tools for assessing the calibration of a collection of ensemble forecasts for a scalar predictand, which here is the IWR (Wilks, 2011). To construct a rank histogram, we analyse n ensemble forecasts, each with $M = 11$ ensemble members. For every ensemble forecast, we determine the rank of the observation within the $M + 1$ values, hence we sort the IWR in ensemble forecasts and observation (reanalysis) in ascending order and determine the rank of the observation value. These ranks are then tabulated, resulting in a histogram that represents the distribution of observation ranks across all ensemble forecasts. A calibrated and reliable forecast would be represented by a uniform distribution of ranks, while deviations may indicate biases or under/overdispersion in the ensemble forecasts. To provide a comprehensive view, we visualise the verification rank histograms of each lead time in one joint two-dimensional histogram. The rank is on the x-axis, the lead time on the y-axis, and the frequency distribution is indicated by coloured boxes and numbers inside the boxes, indicating the deviations from a perfect frequency distribution of 1/12 for each rank due to 11 + 1 members.

3 | RESULTS

In this section, we first analyse $Z500$ forecast biases and how they are connected to biases in the weather regime index forecasts. Then we present a comprehensive evaluation of the post-processed forecasts in comparison with both the raw and $Z500$ bias-corrected forecasts of the weather regime index (IWR). Our analysis is divided into two main aspects: the assessment of univariate skill, focusing on EMOS-G (Section 3.2), and the evaluation of multivariate skill, centred around ECC-Q (Section 3.3). In addition, we investigate the method's sensitivity to variations in training data availability (Section 3.4). It is important to note that we are assessing the ensemble forecast's capability to predict IWR on given days, which is a challenging forecasting question, in particular at extended-range lead times.

3.1 | $Z500$ forecast biases

Biases in the $Z500$ forecast have a direct impact on the seven-dimensional weather regime index, as these project directly into the IWR forecasts. The analysis of $Z500$ biases as a function of forecast lead time (Figure 2) unveils the reasons for systematic biases in the IWR (Figure S3). In the weather regime region (denoted by the grey dashed box in Figure 2), forecast biases grow from values near 0 gpm (geopotential meter) 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 2) and extend from Canada into the high-latitude North Atlantic. This positive bias anomaly projects most prominently into the Atlantic Ridge weather regime (yellow in Figure S3). The $Z500$ biases are seasonally dependent, which also manifests in the IWR. In winter (second row in Figure 2), $Z500$ biases exhibit a dipole structure projecting into the Atlantic Trough and Greenland Blocking (purple and blue in Figure S3). In spring (third row in Figure 2), the positive $Z500$A in the North Atlantic grows, projecting into the Atlantic Ridge and Greenland Blocking (yellow and blue in Figure S3). During summer (fourth row in Figure 2), 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 S3). In autumn (fifth row in Figure 2), 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 S3). This seasonally differentiated analysis of bias growth in the $Z500$ field and its projection into weather regimes emphasises the need for bias correction that accounts for seasonality.

3.2 | Univariate post-processing

Here, we utilise EMOS for univariate post-processing. EMOS is trained on a reforecast dataset with 3101 forecasts spanning June 1999–May 2015, and subsequently tested

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840 by Karlsruher Institut F., Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

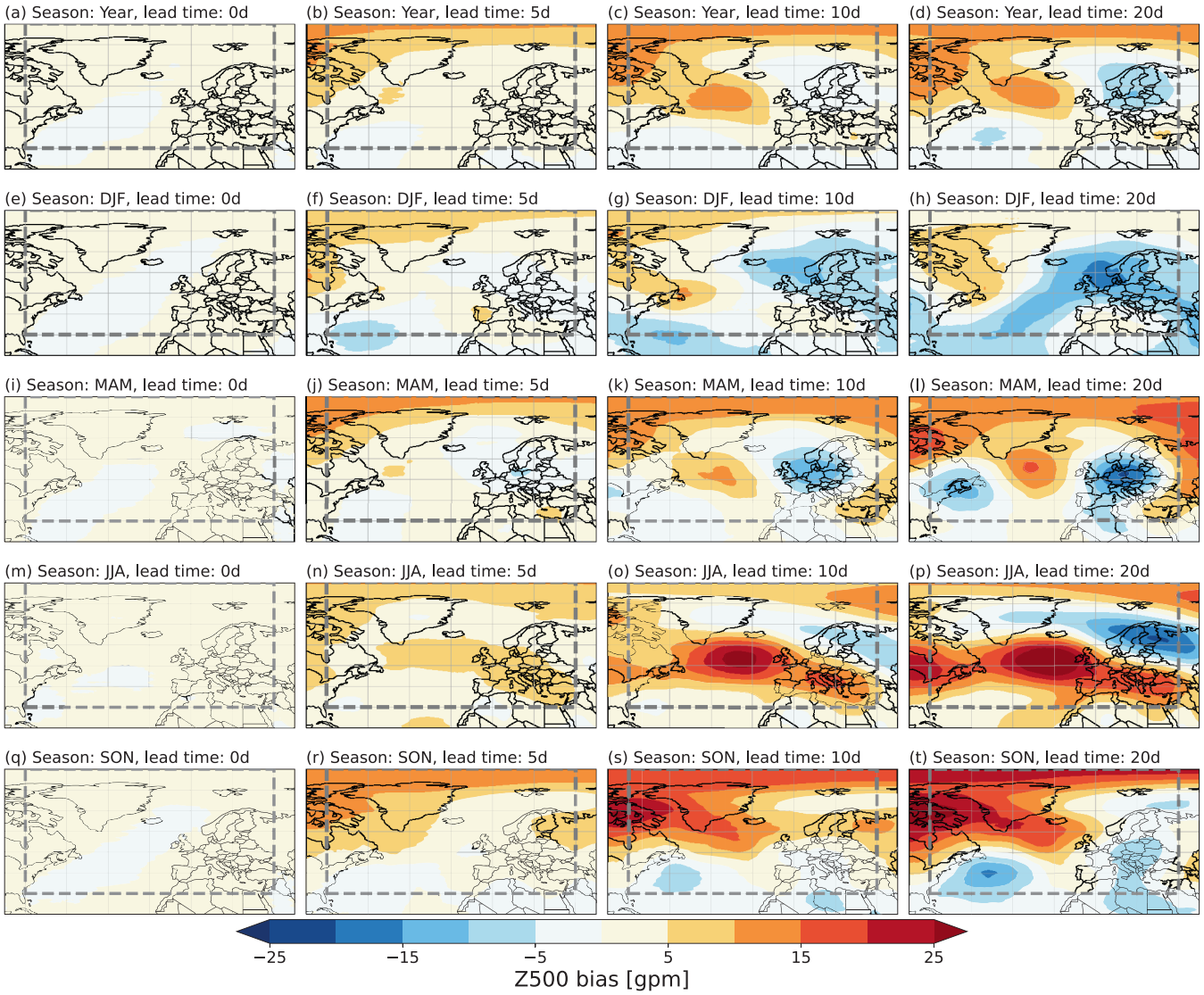

FIGURE 2 Displayed are the mean 500-hPa geopotential height anomaly biases of the ensemble mean forecasts. The $Z500A$ bias is computed by subtracting the ERA perfect member from the ensemble mean. The mean bias fields are presented for the entire year and four seasons (winter: DJF, spring: MAM, summer: JJA, autumn: SON) across rows, and at lead times of 0, 5, 10, and 20 days across columns. The region corresponding to North Atlantic–European weather regimes is indicated by the dashed grey box. Note that 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. [Colour figure can be viewed at wileyonlinelibrary.com]

on 899 forecasts from the reforecast dataset covering June 2015–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, for example, Jordan *et al.* (2019).

In the previous section, we identified biases in the $Z500$ field forecasts (Figure 2). 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 S4 and S5 in the Supplement.

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840 by Karlsruher Institut F., Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

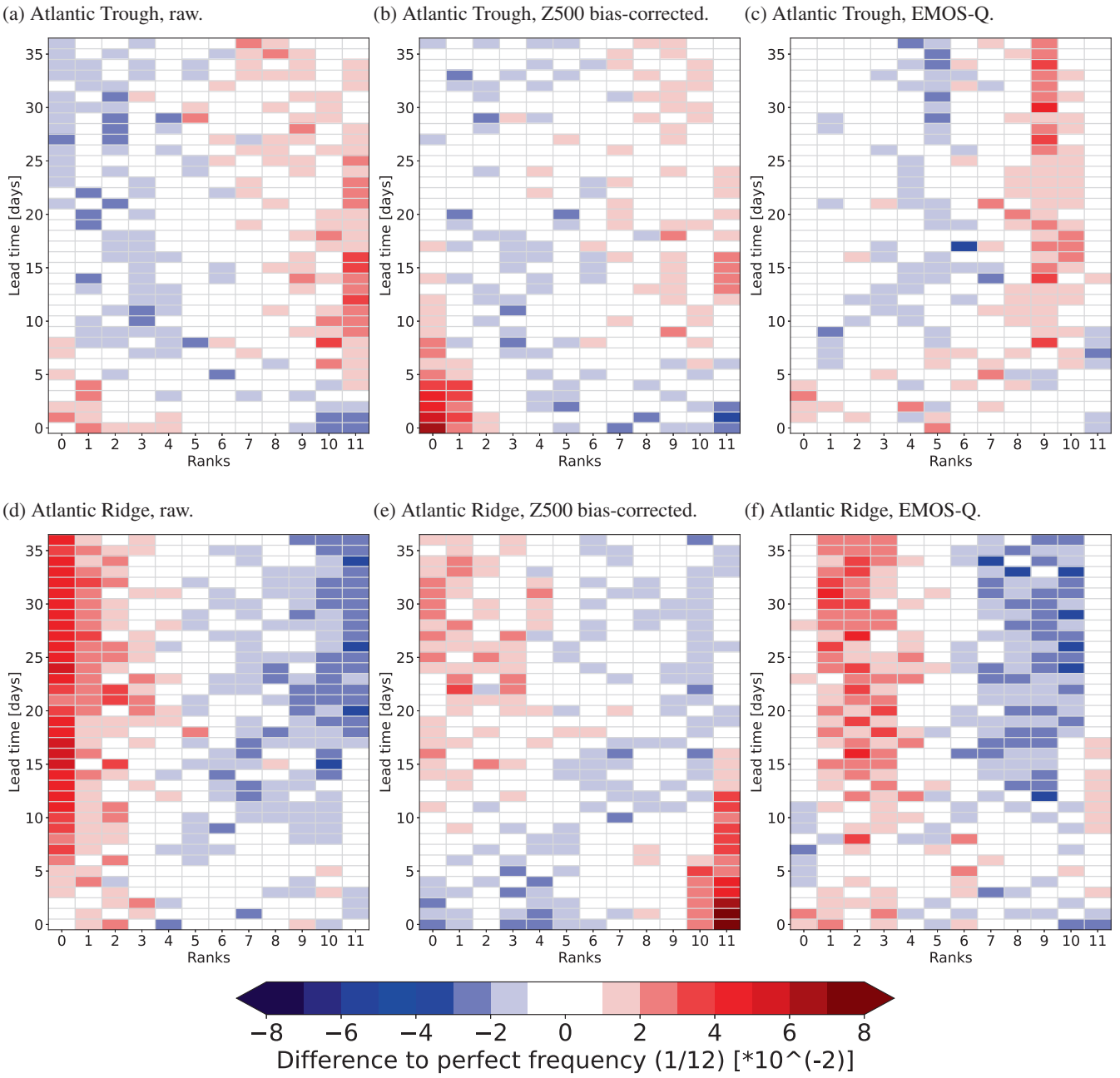

FIGURE 3 The verification rank histograms for (a–c) the Atlantic Trough and (d–f) the Atlantic Ridge are visualised. Additional rank histograms for all cyclonic and anticyclonic weather regimes can be found in supplementary Figures S4 and S5. The figures illustrate the rank histograms for (a,d) raw forecasts, (b,e) *Z*500 bias-corrected forecasts, and (c,f) EMOS-Q post-processed forecasts 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. [Colour figure can be viewed at wileyonlinelibrary.com]

Due to the observed bias in the *Z*500 forecast, the raw IWR forecast for Atlantic Trough tends to be underforecast for lead times exceeding 10 days (Figure 3a, noticeable in the red colours at higher ranks), while it is overforecast at shorter lead times. Similarly, the Atlantic Ridge is consistently overforecast across all lead times (Figure 3d, evident in the red colour at lower ranks). Although the bias correction of the *Z*500 field improves the reliability for these two weather regimes at extended lead times (Figure 3b,e), it deteriorates the reliability at shorter lead times. Forecasts up to a 10-day lead time tend to overforecast the Atlantic Trough regime, while the Atlantic Ridge is strongly underforecast at lead times up to 12 days after the *Z*500 bias correction. Verification rank histograms

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840 by Karlsruher Institut F., Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

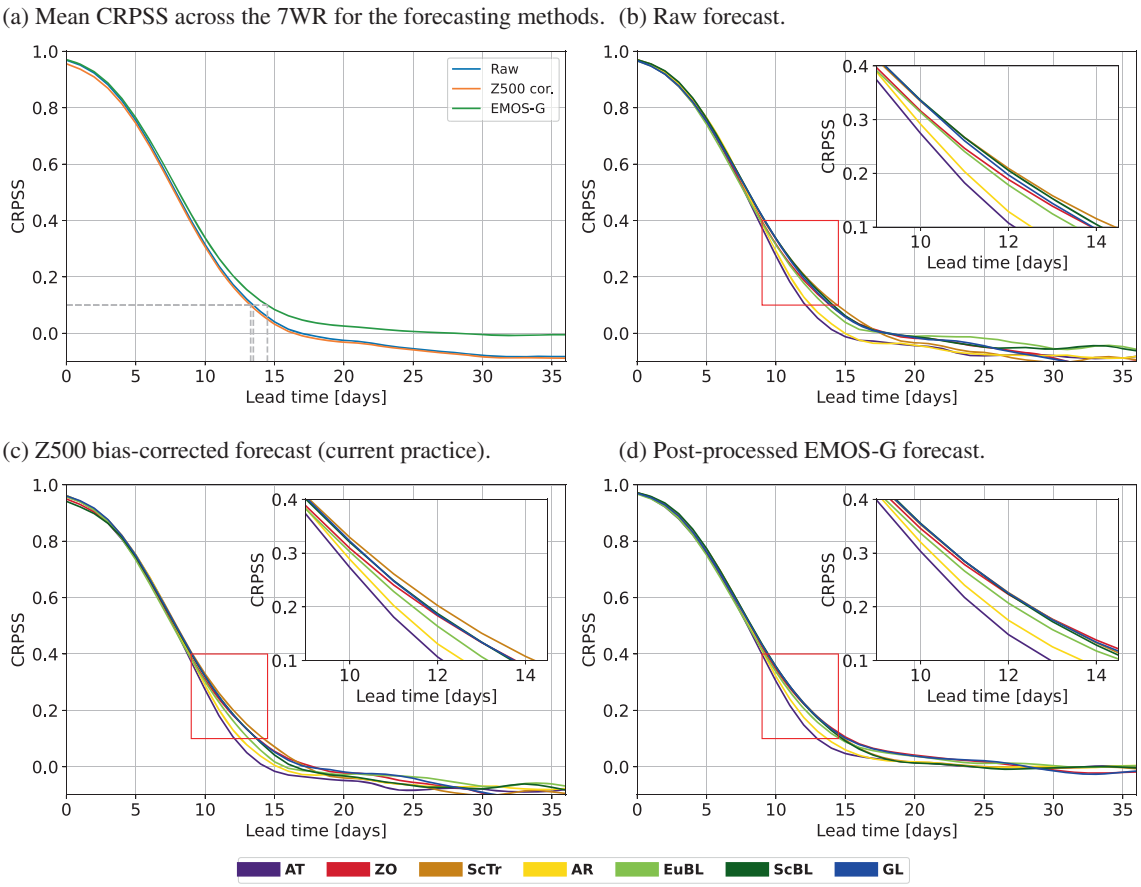

FIGURE 4 CRPSS as a function of lead time. In (a), the aggregated mean CRPSS across the weather regimes is shown for the three forecasting methods, raw (blue), $Z500$ bias-corrected (orange), and EMOS-G (green). In (b)–(d) the CRPSS is shown separately for each weather regime, indicated by the regime colours, for (b) the raw forecasts, (c) the $Z500$ bias-corrected forecasts, and (d) the EMOS-G forecasts. The red box indicates the area where we zoom in to visualise the differences between those regimes better. The CRPSS is calculated against a 31-day rolling climatological forecast. [Colour figure can be viewed at wileyonlinelibrary.com]

show that forecasts post-processed with EMOS exhibit improved reliability (Figure 3c,f 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 the 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 S4). 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 S5). In summary, the $Z500$ bias-corrected forecast, especially at 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 suggest that EMOS post-processing generates more reliable and consistent forecasts for the weather regime index compared with $Z500$ bias-corrected forecasts. To investigate the corrections via EMOS in more detail, we analyse the estimated EMOS coefficients introduced in Equation (4) for different weather regimes (Figure S6). Across all weather regimes (Figure 4a–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

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840 by Karlsruher Institut F., Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 CRPSS with climatology as the reference forecast. As an initial comparison among the forecasting methods, we examine the mean skill score over the seven weather regimes (Figure 4a).

The mean skill scores of the three forecasting methods (Figure 4a) mainly differ for extended-range 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 forecasts slightly outperforming the $Z500$ bias-corrected forecasts. At extended lead times, the $Z500$ bias-corrected forecasts exhibit slightly lower CRPSS values than the raw forecasts, though the difference is minimal.

In Figure 4a, we observed that the mean skill scores of the raw and EMOS-G forecasts 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 (Figure 4b–d) to reveal commonalities and differences between the individual forecasting methods. All three forecasting methods have in common that the differences 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 at lead times between 10 and 15 days (the lead time range until which all forecasting methods still exhibit CRPSS values larger than 0 for each weather regime). The lowest skill is observed for the 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 the 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 with 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, the largest differences in terms of forecast skill for the different forecasting methods and regimes are observed for lead times of 10–14 days (insets in Figure 4b–d).

To asses the skill score improvements of the EMOS-G forecasts in comparison with the raw forecast in more detail, we calculate the lead-time gain across a range of forecast skill horizon thresholds (Figure 5). 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 with 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 5).

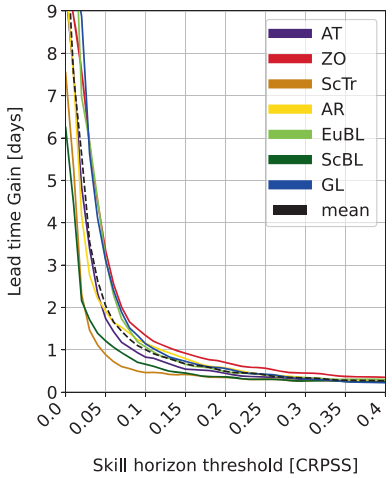

FIGURE 5 Forecast skill horizon gain as a function of different CRPSS thresholds from raw forecasts to EMOS-G forecasts. Lead-time gain separated for weather regimes. [Colour figure can be viewed at wileyonlinelibrary.com]

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 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 4d. 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 mean forecast skill horizons of 13.5 and 14.5 days for the raw and EMOS-G forecasts, respectively, indicating a lead-time gain of 1 day.

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840 by Karlsruher Institut F., Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

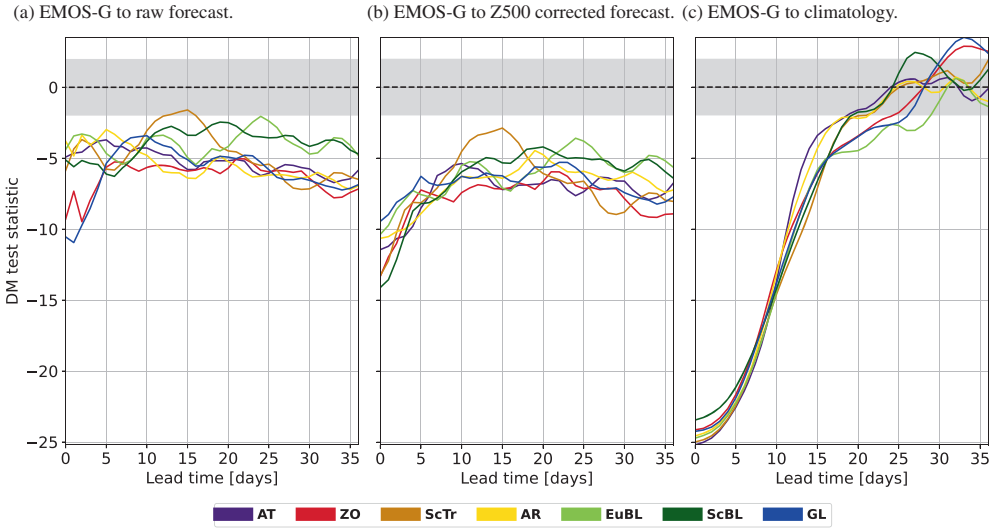

FIGURE 6 Testing the significance of the univariate CRPS improvements of the EMOS-G method with respect to (a) the raw forecast, (b) the $Z500$ bias-corrected forecast, and (c) the climatological forecast using a Diebold–Mariano test of equal performance with an α-level of 0.05 (indicated by the grey shading). [Colour figure can be viewed at wileyonlinelibrary.com]

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 6a) and $Z500$ bias-corrected (Figure 6b) forecasts, as well as the climatological reference forecast (Figure 6c).

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 6a). The least significant results at extended lead times when comparing EMOS-G forecasts with raw forecasts are observed for European and Scandinavian Blocking, which is in line with the respective verification rank histograms in supplementary Figure S5d,f,g,i, respectively. Compared with climatology (Figure 6c), EMOS-G demonstrates significant performance improvements across all weather regimes out to 19 days forecast lead time, with even longer significant improvements for European Blocking, reaching out to 28 days.

In conclusion, applying EMOS-G post-processing to the raw forecasts leads to significant skill improvements across all weather regimes and lead times compared with 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.

3.3 | Multivariate post-processing

Similar to the evaluation of univariate post-processing skill (Section 3.2), 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 ES, which is a multivariate extension of the CRPS. We also evaluate our results by using the variogram score (VS) 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 7a and VSS in Figure 7b) 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 value), 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 of the multivariate post-processing step. The forecast skill improvement of ECC-Q in comparison with EMOS-Q is notable as early as 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 7a,b). At 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 at extended lead times

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840 by Karlsruher Institut F., Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

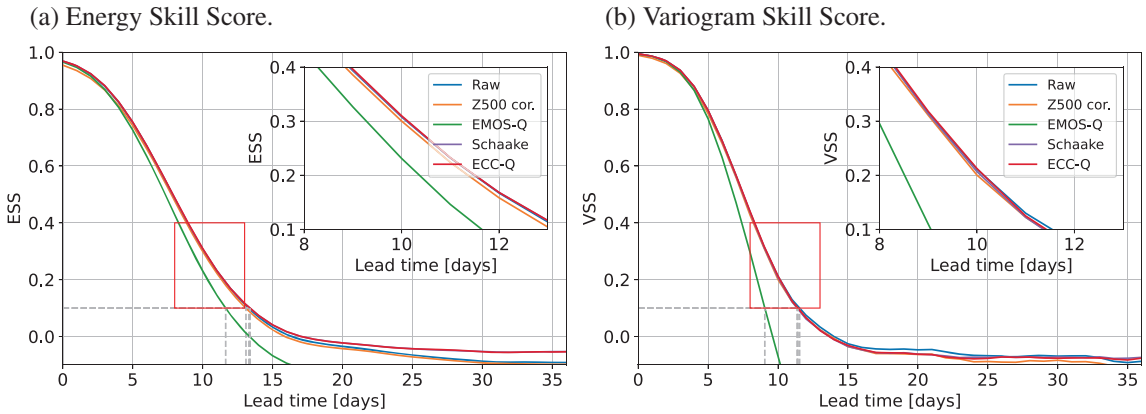

FIGURE 7 Multivariate skill scores for raw forecast (blue), $Z500$ bias-corrected forecast (orange), EMOS-Q (green), Schaake Shuffle (purple), and ECC-Q (red), using (a) the energy skill score and (b) the variogram skill score. [Colour figure can be viewed at wileyonlinelibrary.com]

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 ESS and VSS (Figure 7, 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 8), gives a clearer insight into the actual differences of the skill scores. Using the energy score, ECC-Q does perform better than the raw and $Z500$ bias-corrected forecasts at 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 lead time. Comparing with 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 a lead time of 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-processing method of $Z500$ calibration and shows it continuously outperforms the pre-processing method using the energy skill score and for most lead times using the variogram skill score. Restoring the multivariate dependence 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 with univariate post-processing (EMOS-Q), ensuring a more accurate representation of the true relationship between weather

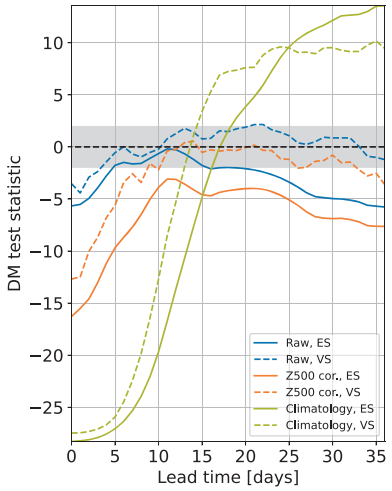

FIGURE 8 Diebold–Mariano tests based on the multivariate scores (energy score (solid) and variogram score (dashed)) for equal predictive performance of ECC-Q and the raw forecast (blue), the $Z500$ bias-corrected forecast (orange), and the climatological reference forecast (olive green). [Colour figure can be viewed at wileyonlinelibrary.com]

regime indices of each ensemble member. However, it is important to acknowledge that the multivariate forecast skill in the extended range for all forecasting methods, evaluated over the entire testing period, is inferior to climatology.

3.4 | Sensitivity of post-processed weather regime forecasts to training data availability

The main setup of our analysis involves combining two ECMWF model cycles (Cy46R1 and Cy47R1) to train

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840 by Karlsruher Institut F., Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

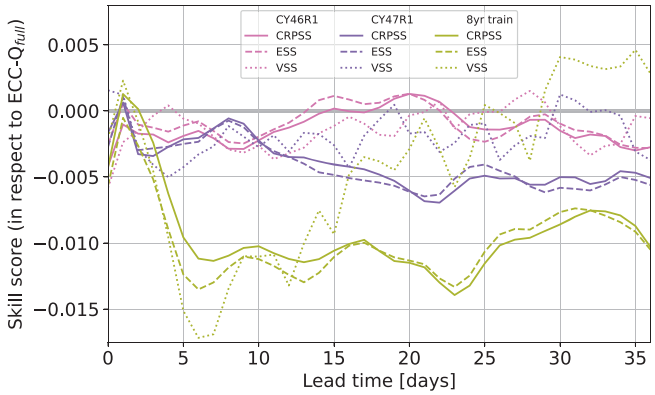

FIGURE 9 Comparison of performance when reducing training data to the cycles CY46R1 (pink) and CY47R1 (purple) separately, as well as using only the eight most recent years of both cycles (green) together. We compare the performance of these method setups using skill scores (CRPSS: solid lines, ESS: 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. [Colour figure can be viewed at wileyonlinelibrary.com]

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 relies solely 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 in the training period by keeping both forecasting cycles but reducing the number of training years to the eight most recent years of the entire training period (June 2007–May 2015). The testing period is identical to that in the previous sections, using the 899 forecasts from June 2015–May 2020, combining data from both forecast model cycles. To ensure a fair comparison, we exclude data from April 26–July 15 of each year for Cy47R1, as this cycle did not run for 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 compare the performance of the ECC-Q setups with a reduced training period directly with the ECC-Q setup with the full training set (Figure 9) 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 with the combination of both forecast cycles (skill score values around 0). ECC-Q trained only on Cy47R1 (purple) exhibits slightly lower skill scores than ECC-Q trained on Cy46R1, which is potentially due to the unavailability of training data for April 26–July 15. When training on an eight-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, e.g., 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 applied directly to operational ECMWF extended-range forecasts without considerable losses in skill compared with 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 seven-dimensional weather regime index.

4 | 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 seven-dimensional weather regime index (IWR) that identifies four anticyclonic and three cyclonic regimes. The IWR represents the projection of 500-hPa geopotential height anomalies ($Z500A$) 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

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840 by Karlsruher Institut F., Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 traced directly 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 at all lead times. When evaluating EMOS in a 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 $\mathrm{CRPSS_{clim}}$ exceeds a specific threshold, indicates that EMOS outperforms the $Z500$ bias-corrected and raw forecasts for a range of skill horizon thresholds ($0 \leq \mathrm{CRPSS_{clim}} \leq 0.4$). 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 with the raw forecast and 1.2 days compared with the $Z500$ bias-corrected forecasts.

The multivariate dependence structure of the IWRs is lost when post-processing ensemble forecasts in a 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 with EMOS become evident starting from a lead time of 5 days when considering multivariate skill scores (ESS and VSS). Consistent with the univariate comparison with 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 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 seven-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 with the seven-dimensional weather regime index is also 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 7 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 dependence 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 limited primarily 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 stem largely from the intrinsic predictability limit of the atmosphere and model deficiencies. All methods rely solely on the $Z500$ 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 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).

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840 by Karlsruher Institut F., Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

During specific atmospheric situations ("windows of forecast opportunity"), such as a strong stratospheric polar vortex or specific phases of the Madden–Julian Oscillation (Madden & Julian, 1971), the forecast skill horizon for weather regimes may extend, as demonstrated by Büeler *et al.* (2021). Building on results from studies exploring predictive skill in the midlatitudes and teleconnection patterns (Ferranti *et al.*, 2018; Lee *et al.*, 2019; Mayer & Barnes, 2020), we believe that enhancing weather regime forecasts in the extended range can be achieved by incorporating additional information/predictors representing the state of relevant atmospheric modes into our post-processing method. Neural networks are likely to be the most suitable method for the effective implementation of these features (Rasp & Lerch, 2018; Vannitsem *et al.*, 2021). By introducing neural networks into the post-processing framework, we anticipate not only an improvement in extended-range predictive skill but also the ability to investigate the characteristics of windows of forecasting opportunity through the application of explainable artificial intelligence (explainable AI) methods. In addition, it is likely that advancing forecast skill to sub-seasonal lead times can be achieved by focusing on skill during windows of forecast opportunity, rather than average skill over all available forecasts. Hence, it is crucial to identify and explore a priori knowledge linked to improved flow-dependent predictability. We hypothesise that evaluating forecasts based on windows of forecast opportunity will be of imminent importance for downstream applications, providing additional information to the user on whether to trust the forecast or not.

We are currently engaged in optimising neural networks for the effective post-processing of extended-range weather regime forecasts, aiming not only to enhance predictive skill but also to explore—a priori—forecasting windows of opportunity using explainable AI.

ACKNOWLEDGEMENTS

F. Mockert has received funding from the KIT Center for Mathematics in Sciences, Engineering, and Economics under the seed funding programme. The contribution of J. Quinting was partly funded by the European Union (ERC, ASPIRE, 101077260). The work of J. Quinting and C. M. Grams was funded by the Helmholtz Association as part of the Young Investigator Group "Sub-seasonal Predictability: Understanding the Role of Diabatic Outflow" (SPREADOUT, Grant VH-NG-1243). S. Lerch gratefully acknowledges support by the Vector Stiftung through the Young Investigator Group "Artificial Intelligence for Probabilistic Weather Forecasting." The contribution of M. Osman was supported by Axpo Solutions AG.

CONFLICT OF INTEREST STATEMENT

All authors declare that they have no conflicts of interest.

DATA AVAILABILITY STATEMENT

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

ORCID

Fabian Mockert https://orcid.org/0000-0002-3222-6667
Christian M. Grams https://orcid.org/0000-0003-3466-9389
Sebastian Lerch https://orcid.org/0000-0002-3467-4375
Marisol Osman https://orcid.org/0000-0002-6275-1454
Julian Quinting https://orcid.org/0000-0002-8409-2541

REFERENCES

Bloomfield, H.C., Brayshaw, D.J., Gonzalez, P.L. & Charlton-Perez, A. (2021) Pattern-based conditioning enhances sub-seasonal prediction skill of European national energy variables. *Meteorological Applications*, 28, e2018.

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

Charlton-Perez, A.J., Aldridge, R.W., Grams, C.M. & Lee, R. (2019) Winter pressures on the UK health system dominated by the Greenland blocking weather regime. *Weather and Climate Extremes*, 25, 100218.

Chen, J., Janke, T., Steinke, F. & Lerch, S. (2024) Generative machine learning methods for multivariate ensemble post-processing. *Annals of Applied Statistics*, 18, 159–183.

Clark, M., Hay, L., Rajagopalan, B. & Wilby, R. (2004) The Schaake shuffle: a method for reconstructing space-time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology*, 5, 243–262.

Diebold, F.X. & Mariano, R.S. (1995) Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 253–263.

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

Gneiting, T. & Katzfuss, M. (2014) Probabilistic forecasting. *Annual Review of Statistics and its Application*, 1, 125–151.

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

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

Grams, C.M., Ferranti, L. & Magnusson, L. (2020) How to make use of weather regimes in extended-range predictions

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840 by Karlsruher Institut F., Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

for Europe. *ECMWF Newsletter*, 165. www.ecmwf.int/en/about/media-centre/media-resourcesfrom

Hauser, S., Teubler, F., Riemer, M., Knippertz, P. & Grams, C.M. (2023a) Life cycle dynamics of Greenland blocking from a potential vorticity perspective. EGUsphere Preprint Repository. https://doi.org/10.5194/egusphere-2023-2945

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

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049.

Jordan, A., Krüger, F. & Lerch, S. (2019) Evaluating probabilistic forecasts with scoringRules. *Journal of Statistical Software*, 90, 1–37.

Lakatos, M., Lerch, S., Hemri, S. & Baran, S. (2023) Comparison of multivariate post-processing methods using global ECMWF ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 149, 856–877.

Lavaysse, C., Vogt, J., Toreti, A., Carrera, M.L. & Pappenberger, F. (2018) On the use of weather regimes to forecast meteorological drought over Europe. *Natural Hazards and Earth System Sciences*, 18, 3297–3309.

Lee, R.W., Woolnough, S.J., Charlton-Perez, A.J. & Vitart, F. (2019) ENSO modulation of MJO teleconnections to the North Atlantic and Europe. *Geophysical Research Letters*, 46, 13535–13545.

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

Madden, R.A. & Julian, P.R. (1971) Detection of a 40-50 day oscillation in the zonal wind in the tropical Pacific. *Journal of the Atmospheric Sciences*, 28, 702–708.

Mayer, K.J. & Barnes, E.A. (2020) Subseasonal midlatitude prediction skill following quasi-biennial oscillation and Madden–Julian oscillation activity. *Weather and Climate Dynamics*, 1, 247–259.

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

Michelangeli, P.-A., Vautard, R. & Legras, B. (1995) Weather regimes: recurrence and quasi stationarity. *Journal of Atmospheric Sciences*, 52, 1237–1256.

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

Möller, A., Lenkoski, A. & Thorarinsdottir, T.L. (2013) Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Quarterly Journal of the Royal Meteorological Society*, 139, 982–991.

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

Perrone, E., Schicker, I. & Lang, M.N. (2020) A case study of empirical copula methods for the statistical correction of forecasts of the ALADIN-LAEF system. *Meteorologische Zeitschrift*, 29, 277–288. Available from: https://doi.org/10.1127/metz/2020/1034

Rasp, S. & Lerch, S. (2018) Neural networks for postprocessing ensemble weather forecasts. *Monthly Weather Review*, 146, 3885–3900.

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. & Möller, A. (2018) Ensemble postprocessing methods incorporating dependence structures. In: *Statistical postprocessing of ensemble forecasts*. Amsterdam, The Netherlands: Elsevier, pp. 91–125. Available from: https://www.sciencedirect.com/science/article/pii/B9780128123720000042

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

Scheuerer, M. & Hamill, T.M. (2015) Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review*, 143, 1321–1334. Available from: https://doi.org/10.1175/MWR-D-14-00269.1

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.

Vannitsem, S., Bremnes, J.B., Demaeyer, J., Evans, G.R., Flowerdew, J., Hemri, S. et al. (2021) Statistical postprocessing for weather forecasts: review, challenges, and avenues in a big data world. *Bulletin of the American Meteorological Society*, 102, E681–E699.

Vautard, R. (1990) Multiple weather regimes over the North Atlantic: analysis of precursors and successors. *Monthly Weather Review*, 118, 2056–2081.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C. et al. (2017) The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, 98, 163–173.

Vitart, F. & Mladek, R. (2023) ECMWF model. https://confluence.ecmwf.int/display/S2S/ECMWF+Model

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

Wilks, D.S. (2011) *Forecast verification*, Vol. 100. Amsterdam, The Netherlands: Elsevier.

Wilks, D.S. (2015) Multivariate ensemble model output statistics using empirical copulas. *Quarterly Journal of the Royal Meteorological Society*, 141, 945–952.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Mockert, F., Grams, C.M., Lerch, S., Osman, M. & Quinting, J. (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

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4840 by Karlsruher Institut F., Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License