Learning predictable and informative dynamical drivers of extreme precipitation using variational autoencoders

Fiona R. Spuler[1, 2], Marlene Kretschmer[1, 3], Magdalena Alonso Balmaseda[4], Yevgeniya Kovalchuk[5], and Theodore G. Shepherd[1]

[1]Department of Meteorology, University of Reading, Reading, UK
[2]The Alan Turing Institute, London, UK
[3]Leipzig Institute for Meteorology, Leipzig University, Leipzig, Germany
[4]European Centre for Medium-Range Weather Forecasts, Reading, UK
[5]Centre for Advanced Research Computing, University College London, London, UK

Correspondence: Fiona R. Spuler (f.r.spuler@pgr.reading.ac.uk)

Abstract.

Large-scale atmospheric dynamics modulate the occurrence of extreme precipitation events and provide sources of predictability of these events on timescales ranging from days to decades. In the midlatitudes, regional dynamical drivers are frequently represented as discrete, persistent and recurrent circulation regimes. However, available methods identify circulation regimes which are either predictable but not necessarily informative of the relevant local-scale impact studied, or targeted to a local-scale impact but no longer as predictable. In this paper, we introduce a generative machine learning method based on variational autoencoders for identifying probabilistic circulation regimes targeted to spatial patterns of precipitation. The method, CMM-VAE, combines targeted dimensionality reduction and probabilistic clustering in a coherent statistical model and extends a previous architecture published by the authors to allow for categorical target variables. We investigate the trade-off between regime informativeness of local precipitation extremes and predictability of the regimes at subseasonal lead times. In an application to study drivers of extreme precipitation over Morocco, we find that the targeted CMM-VAE regimes are more informative of the impact variable of interest, compared to two well-established linear approaches, while maintaining the predictability of conventional non-targeted circulation regimes in subseasonal hindcasts, hence resolving the trade-off identified in previous studies. Furthermore, the targeted regimes and their predictability are physically interpretable in terms of known subseasonal teleconnections relevant to the region, the Madden-Julian Oscillation and variability of the stratospheric polar vortex. The proposed method therefore allows to identify predictable, interpretable and locally relevant representations of regional dynamical drivers given a target variable of interest. These results highlight the potential of the method for a variety of applications, ranging from subseasonal forecasting to attribution and statistical downscaling.

1 Introduction

Extreme events such as heatwaves and extreme precipitation cause devastating impacts on lives and livelihoods around the world every year. Improving the forecasts of these extremes at timescales ranging from days to decades, in particular in the

context of a changing climate, can support societal resilience through measures such as improved early-warning systems, forecast-based financing, and robust climate change adaptation (Coughlan de Perez et al., 2019; Lemos et al., 2012).

The occurrence and predictability of extreme events is often modulated by regional dynamical drivers such as the North Atlantic Oscillation over north-western Europe or the Caribbean Low-Level Jet over Central America and the Caribbean (García-Martínez and Bollasina, 2020; Scaife et al., 2008). These regional dynamical drivers are frequently predictable at extended lead times themselves, and can furthermore be modulated by teleconnections from low-frequency modes of variability in the climate system such as the El-Niño Southern Oscillation, which act as sources of predictability at longer timescales (Ferranti et al., 2018; Le et al., 2023; Mariotti et al., 2020; Saggioro et al., 2024).

Regional dynamical drivers and associated teleconnections have therefore been used to improve extreme event predictions on a range of timescales. At medium-range lead times up to 15 days ahead, forecasts conditional on regional dynamical drivers have demonstrated improved skill (Allen et al., 2021; Mastrantonas et al., 2022; Rouges et al., 2024). At subseasonal-to-seasonal (S2S) lead times, reduced representations of dynamical drivers and empirical models of teleconnections to other modes of variability have been leveraged to improve forecast skill (Bach et al., 2024; de Fondeville et al., 2023; Bommer et al., 2025; Baker et al., 2018; Kretschmer et al., 2017), as well as identify and explain so-called windows of opportunity of higher-than-average forecast skill (Dunstone et al., 2023; Mariotti et al., 2020). On climate timescales, large-scale drivers have been used to gain a physical understanding of climate model uncertainty as well as conditional predictability by building plausible storylines of future change (Harvey et al., 2023; Mindlin et al., 2023; Shepherd et al., 2018).

A well-established approach to representing regional dynamical drivers, such as jet variability in the midlatitudes, is the identification of recurrent and persistent patterns of atmospheric circulation, so-called weather regimes (Ghil and Robertson, 2002; Hannachi et al., 2017). These regimes are commonly identified using a combination of linear dimensionality reduction and non-probabilistic clustering (see e.g. Michelangeli et al. (1995)). Over several regions, this approach has been shown to identify weather regimes that are persistent and predictable by dynamical forecast models (Dorrington et al., 2022; Falkena et al., 2022; Rouges et al., 2024; Straus, 2022) as well as useful for understanding teleconnections from, for example, the Madden-Julian Oscillation (MJO) or Stratospheric Polar Vortex (SPV) (Cassou, 2008; Domeisen et al., 2020).

While these conventional weather regimes are designed to capture the main features of the circulation over a given region such as the North Atlantic, they do not necessarily disentangle the dynamical patterns that modulate extreme impacts in a specific country or area of interest (Bloomfield et al., 2020; Wiel et al., 2019; Vrac and Yiou, 2010; Mastrantonas et al., 2020). However, the ability of regional dynamical drivers to improve forecasts across different timescales depends on both their informativeness of the local extreme impact of interest, as well as their own predictability and ability to represent the atmospheric phase space. Therefore, conventional circulation regimes, despite their predictability, do not necessarily represent suitable regional dynamical drivers for any given target variable as they lack informativeness of the local impact. Different methods have been proposed to identify regimes targeted to a local impact, such as clustering the impact variable directly (Bloomfield et al., 2020; Ullmann et al., 2014) or filtering for extreme impact days prior to clustering (Dorrington et al., 2024). However, while the resulting regimes are more informative of the impact studied, they were shown to compromise regime predictability (Bloomfield et al., 2021).
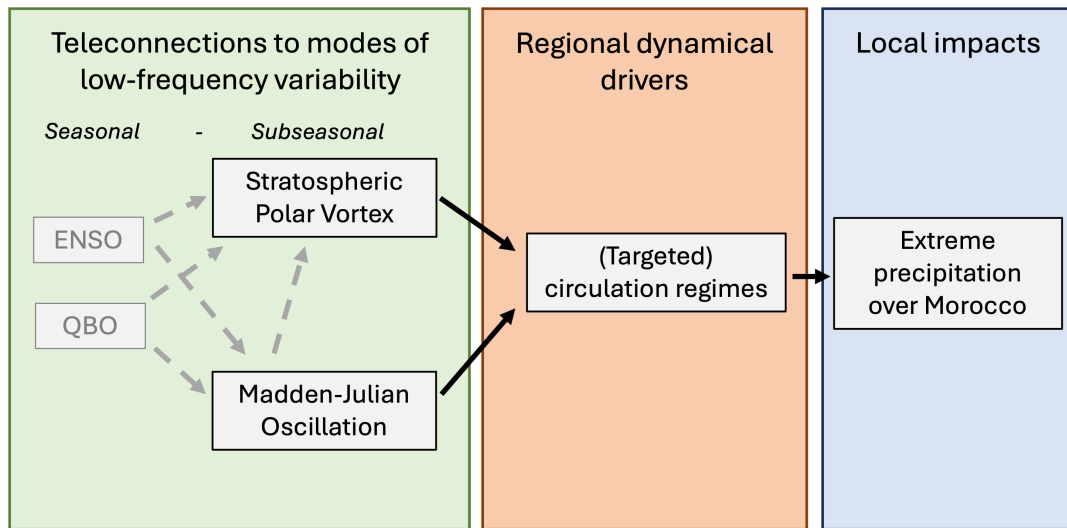
Figure 1. Graphical illustration of teleconnections (green box) from the SPV and MJO at subseasonal timescales, their mediation by the targeted circulation regimes (orange box) and associated impact on extreme precipitation over Morocco (blue box).

In this study, we introduce a probabilistic machine learning method for identifying targeted regimes and investigate the ability of this method to balance the trade-off between regime informativeness and predictability. The method, termed Categorical Mixture Model Variational Autoencoder (CMM-VAE), is based on a variational autoencoder architecture and combines targeted dimensionality reduction and probabilistic clustering in a single coherent statistical model that is fit using Bayesian variational inference. The method builds on a previous method introduced in Spuler et al. (2024a) which was found to identify regimes that are more informative of the chosen target variable while still being persistent and representative of the entire atmospheric phase space over the region. These promising results motivate the further investigation of the ability of the approach to balance the trade-off between regime predictability and informativeness. The CMM-VAE method enables the application of the approach presented in Spuler et al. (2024a) to spatial patterns of extreme precipitation, which can provide information that is more useful at local scales compared to the spatially averaged precipitation used in Spuler et al. (2024a). The CMM-VAE method is described in detail in section 2.2.3.

We apply the method to study circulation regimes targeted to precipitation over Morocco, as well as their predictability in subseasonal hindcasts and associated teleconnections (see Figure 1 for an overview). With most of the rainfall occurring in extended winter, the country is vulnerable to both extreme rainfall, which leads to flooding and is the focus of this study, as well as drought, which impacts agricultural livelihoods and overall macroeconomic stability (Loudyi et al., 2022). Previous studies have shown that extreme precipitation events over Morocco are associated with dynamically driven moisture flux from the Atlantic. This can occur through an alignment of the subtropical jet with the African coastline and anomalous south-westerly surface to mid-tropospheric flow, leading to large-scale ascending motions and instability over the Western Mediterranean region (Dayan et al., 2015; Khouakhi et al., 2022; Toreti et al., 2010). These regional dynamical drivers of precipitation over

Morocco have been studied in terms of both North Atlantic and Mediterranean circulation regimes (Driouech et al., 2010; Mastrantonas et al., 2020; Tramblay et al., 2012). While certain regimes over both regions, such as the negative phase of the North Atlantic Oscillation (NAO), are associated with an increase in the probability of extreme precipitation, the dynamical mechanisms described above and analysed in terms of distinct patterns in Chaqdid et al. (2023) are not clearly captured in the regimes over either region. This motivates the application of the introduced machine learning method to this region, as a first case study of whether the approach is able to identify dynamical drivers that are informative of precipitation over Morocco but which also present an interpretable and predictable partitioning of the large-scale atmospheric phase space.

In terms of timescales, we choose to evaluate teleconnections and predictability at subseasonal to seasonal lead times. At these lead times, regional dynamical drivers of precipitation extremes in Morocco have been shown to be modulated by both the Madden-Julian Oscillation (MJO) (Gadouali et al., 2020) and variability of the northern-hemisphere stratospheric polar vortex (SPV) (Zhang et al., 2024). The MJO is a leading mode of global subseasonal variability that modulates deep tropical convection and thereby acts as a source of Rossby waves leading to teleconnections to extratropical regions (Lee et al., 2020; Roundy et al., 2010). Variability in the SPV, on the other hand, has been shown to influence subseasonal forecast skill over the European and Mediterranean region, with weaker SPV states leading to an equatorward shift of the tropospheric eddy-driven jet and associated storm tracks (Kidston et al., 2015; Kretschmer et al., 2018).

The contribution of this paper is threefold. The first contribution is to introduce the CMM-VAE method, short for Categorical Mixture Model Variational Autoencoder, which enables the application of the method previously presented in Spuler et al. (2024a) to study spatial patterns of precipitation extremes as target variables. Furthermore, we work with a more realistic precipitation dataset (CHIRPS) instead of the reanalysis-based precipitation data used in Spuler et al. (2024a).

The second contribution is to use the CMM-VAE method to identify and analyse targeted regime representations of the dynamical drivers of extreme precipitation over Morocco. We compare the informativeness of these targeted regimes to conventional non-targeted circulation regimes identified using Principal Component Analysis (PCA) and k-means clustering, as well as targeted clusters identified using a linear targeted method, Canonical Correlation Analysis, again combined with k-means clustering.

The third contribution of the paper is to investigate the predictability of these targeted circulation regimes, compared to conventional, non-targeted, regimes. We first analyse the ability of subseasonal dynamical reforecasts to predict the targeted regimes which provides an assessment of the predictability of the regimes conditional on state-of-the-art dynamical models. We then evaluate the predictability of the regimes in reanalysis data, conditional on teleconnections from two relevant modes of subseasonal variability, the MJO and SPV. We analyse conditional predictability in terms of the information-theoretical metrics of conditional entropy and mutual information. Next to presenting another line of evidence for the predictability of the regimes, this investigation shows whether the targeted CMM-VAE regimes can be interpreted as representing physical processes that are modulated by large-scale drivers and can hence be used for further applications such as statistical downscaling as well as to improve the understanding of dynamical processes over the region.

The remainder of the paper is structured as follows. Section 2 introduces the CMM-VAE method for identifying targeted regimes (2.2), and presents the data and methods used to capture spatial patterns of extreme precipitation over Morocco (2.1),
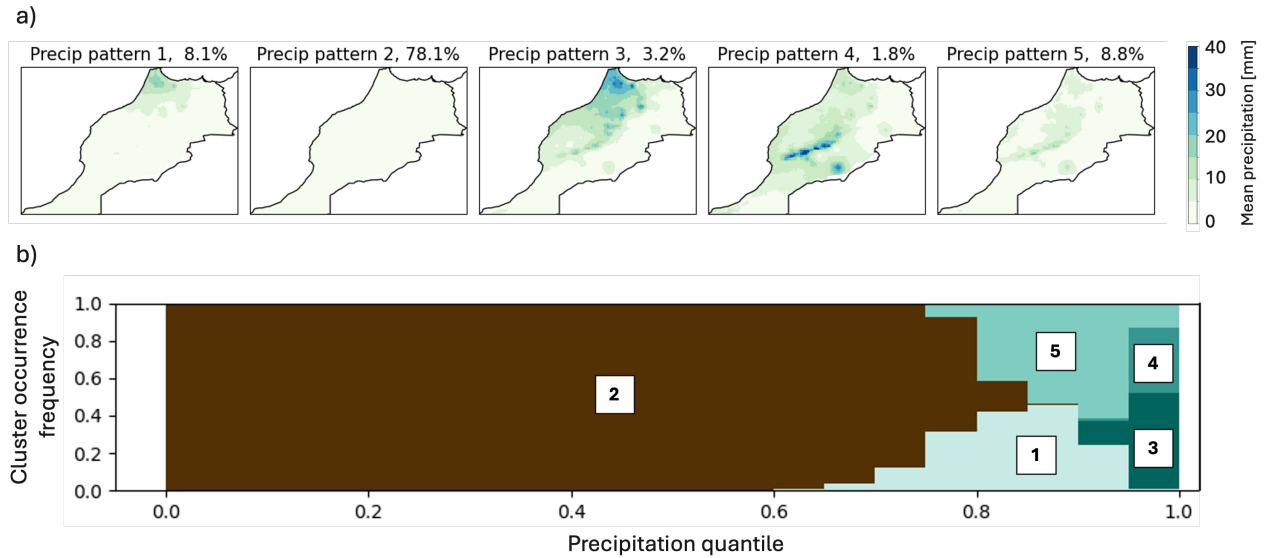
Figure 2. Top row: spatial patterns of 3-day precipitation events over Morocco in extended winter (November - March) identified using k-means clustering. The percentage number in the heading indicates the occurrence probability over all days. Bottom row: occurrence probability of clusters in different quantiles of total precipitation over Morocco. The vertical axis represents the percentage of days in a given precipitation quantile that are assigned to a specific pattern.

as well as to analyse the MJO and SPV teleconnections and skill in subseasonal dynamical reforecasts (2.3). Section 3 presents the results of this study: the application of the CMM-VAE method to identify circulation regimes targeted to precipitation over Morocco (3.1), the evaluation of the forecast skill of these regimes in subseasonal hindcasts (3.2), as well as subseasonal teleconnections to the MJO and SPV in reanalysis data (3.3) Section 4 presents a discussion of the results, a conclusion and an outlook.

2 Data and Methods

2.1 Extreme precipitation over Morocco as local target variable

As target variable, we consider extreme precipitation over Morocco in extended winter (November-March). To this end, CHIRPS v2.0 (Funk et al., 2015) precipitation data is averaged over three days at each grid cell, as this timescale captures the duration of most extreme precipitation events over the observational period (Loudyi et al., 2022). The CHIRPS v2.0 dataset combines in-situ station data with satellite data for all longitudes and 50°S - 50°N and is available from 1981 to present at 0.05° resolution. The dataset was chosen as it has a more realistic representation of precipitation compared to the reanalysis data used in Spuler et al. (2024a) and shows good performance over Africa compared to other available gridded rainfall datasets (Dinku et al., 2018; Maidment et al., 2017).

125   To capture precipitation extremes, we analyse the 95th percentile at each grid cell as well as of spatially averaged precipitation over Morocco. Moreover, we compute the dominant spatial patterns of precipitation by applying k-means clustering on the precipitation data described above, to be able to capture different dynamical drivers of precipitation over different regions (Chaqdid et al., 2023). Results for different choices of k were assessed and a cluster number of 5 was chosen as the minimal number that represents the most prevalent distinct spatial patterns of precipitation over the region. The resulting precipitation

130 patterns shown in Figure 2 contain information about both common spatial patterns of precipitation (top row) as well as the extremality of these precipitation events (bottom row): Pattern 2 summarises all days associated with no or little precipitation, while patterns 3 and 4, represent most days above the 95th percentile of total precipitation over Morocco. On the other hand, patterns 1 and 3 (and likewise 4 and 5) represent related spatial patterns but different levels of extremality.

2.2   Identifying (targeted) circulation regimes as regional dynamical drivers

135 Atmospheric circulation patterns are investigated using geopotential height data at 500 hPa (z500) over the East Atlantic and Mediterranean region (20°N - 80°N; 50°W - 30°E) in extended winter (November - March) based on ERA5 reanalysis data from 1981 to 2022 (Hersbach et al., 2020) re-gridded to a resolution of 2.5° x 2.5°. The geopotential height data is standardized by subtracting the climatological daily mean and dividing the result by the standard deviation across grid points.

    This choice of region was based on multiple considerations. Previous studies found the North Atlantic to be the key moisture

140 source for precipitation over Morocco, and existing literature identifies the NAO as one of the dynamical drivers of precipitation over Morocco (Driouech et al., 2010; Khouakhi et al., 2022; Tramblay et al., 2012). Furthermore, we found that the anomalies related to circulation regimes targeted to precipitation over Morocco over the Mediterranean region analysed in Spuler et al. (2024a) extend to the North Atlantic region. However, key results of this paper were found not to be sensitive to the choice of region.

145   Table 1 provides an overview of the different methods for identifying circulation regimes used in this study which are described in detail below.

2.2.1   Principal Component Analysis and k-means clustering (PCA + k-means)

Principal Component Analysis (PCA, commonly referred to as Empirical Orthogonal Function analysis in atmospheric science) combined with k-means clustering have established themselves as a common choice for determining non-targeted circulation

150 regimes (Charlton-Perez et al., 2018; Michelangeli et al., 1995), including over the Atlantic and Mediterranean regions (Giuntoli et al., 2022; Mastrantonas et al., 2020). While other approaches exist (Hannachi et al., 2017), this two-step approach (hereafter abbreviated PCA + k-means) is used as a baseline here against which to benchmark the targeted method. PCA is a linear dimensionality reduction method that projects a higher-dimensional input space into a reduced space spanned by the orthogonal eigenvectors of the covariance matrix of the data (e.g. Jolliffe and Cadima (2016)). K-means clustering is then

155 applied to the reduced space of principal components and partitions the data into k sets that minimize the mean within-cluster squared distance from the respective cluster centre. PCA is implemented using the eofs Python package (Dawson, 2016) and the first 15 principal components are retained, while k-means clustering is applied using the Python sklearn implementation.

Table 1. Overview of methods used for identifying (targeted) circulation regimes.

Abbreviation	Method	Target variable
PCA + k-means	Linear dimensionality reduction using Principal Component Analysis and subsequent k-means clustering	None
CCA + k-means	Linear dimensionality reduction using regularized Canonical Correlation Analysis and subsequent k-means clustering	CHIRPS precipitation over Morocco aggregated by administrative regions (see (Loudyi et al., 2022), Figure 2.1).
CMM-VAE	Categorical Mixture Model - Variational Autoencoder: nonlinear dimensionality reduction and probabilistic clustering	Spatial patterns of extreme precipitation over Morocco shown in Figure 2.

2.2.2 Regularized Canonical Correlation Analysis with k-means clustering (CCA + k-means)

Canonical Correlation Analysis (CCA) is a linear dimensionality reduction method that jointly identifies respective linear transformations of two high-dimensional spaces onto subspaces that maximize the correlation between the projections of the variables onto their new basis vectors (Johnson and Wichern, 2013). The method has previously been applied to identify targeted circulation regimes (Vrac and Yiou, 2010), and is therefore implemented here, in combination with k-means clustering, to provide a well-established linear targeted method against which to compare methods based on nonlinear variational autoencoders. However, both the precipitation and geopotential height field were found to be too high dimensional for traditional CCA to algorithmically converge. We therefore aggregated precipitation at the district level over Morocco and implemented a ridge regularization parameter for the geopotential height field (Vinod, 1976). The ridge regularization penalizes the number of dimensions used and has been shown to address numerical convergence issues of CCA when applied to collinear data. The regularized CCA method is implemented using the Python package cca-zoo (Chapman and Wang, 2021). Due to the aggregation of the precipitation field, only the first 10 canonical covariates could be computed and were used for subsequent clustering.

2.2.3 CMM-VAE: a nonlinear targeted method based on variational autoencoders

We introduce a novel method, referred to as CMM-VAE, for identifying probabilistic and targeted circulation regimes based on a variational autoencoder, building on the RMM-VAE method previously introduced in Spuler et al. (2024a).

A variational autoencoder (VAE) is a deep generative machine learning method used for non-linear dimensionality reduction, that is to find a reduced representation of a high-dimensional input space. Autoencoders can be interpreted as a non-linear extension of PCA implemented through an encoder and decoder neural network. *Variational* autoencoders extend the encoder-decoder architecture of autoencoders by fitting a probabilistic model into the reduced space using Bayesian variational inference
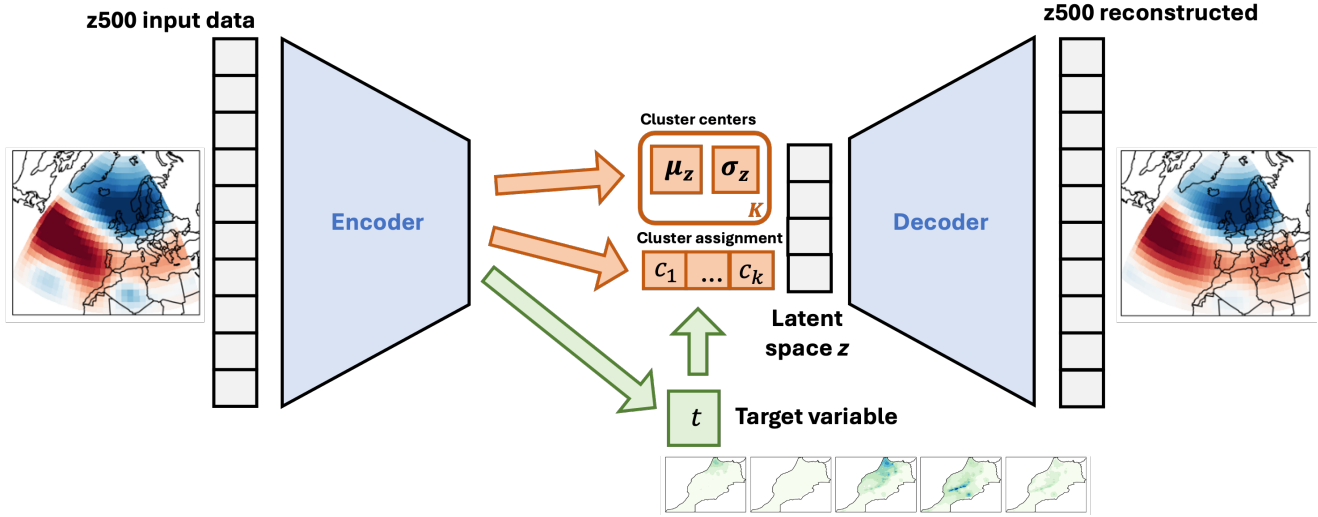
Figure 3. Graphical illustration of the variational autoencoder architecture underlying the CMM-VAE method. Normalized geopotential height data is input into the encoder which is a neural network of three dense layers of decreasing dimensionality. In the latent space, the method fits k multivariate Gaussian distributions with means μ and standard deviations σ and the cluster assignments c_i of individual days (orange arrows), as well as a regression to the target variable t which is used to regularize the latent space (green arrows). The decoder mirrors the encoder in architecture and reconstructs the original input data from the model fit in the latent space.

(Kingma and Welling, 2013). The method is referred to as generative because the probabilistic reduced space is continuous and can be used to simulate new realizations of the resulting regimes.

To identify targeted circulation regimes, Spuler et al. (2024a) extend the baseline VAE architecture in two ways to develop a method called Regression Mixture Model Variational Autoencoder (RMM-VAE). One is to fit a Gaussian mixture model (i.e. a mixture of several Gaussian distributions) into the reduced space to identify probabilistic circulation regimes, instead of a single multivariate Gaussian distribution. The method thereby combines dimensionality reduction and probabilistic clustering in a coherent statistical model (orange arrows in Figure 3). In contrast, other conventional methods for identifying weather regimes implement dimensionality reduction and clustering separately and often conduct a 'hard', as opposed to a probabilistic, cluster assignment of individual days. The probabilistic cluster assignment implemented in the RMM-VAE method, retains more information on transitional states between clusters.

The second modification introduced in the targeted VAE method is to use the encoder of the architecture to predict the chosen target variable, which, in our case, is precipitation over Morocco. The predicted target variable is then used to inform (i.e. regularize) the latent space to obtain clusters that contain a more coherent response in terms of the target variable (green arrows in Figure 3). Spuler et al. (2024a) show that this regularization organises the latent space in terms of the target variable, that is disentangles the reduced dimension associated with changes in the target variable. This leads to the identification of

circulation regimes that are more coherent in terms of their precipitation response. These modifications to the original VAE architecture are introduced by deriving a modified loss function of the architecture.

The CMM-VAE method extends the RMM-VAE method in the following way. The underlying loss function derived for the RMM-VAE method required the target variable to be a scalar Gaussian, which limits the applicability of the method. Here we derive a modification of the loss function which enables the application of the method to higher-dimensional categorical target variables and is therefore called CMM-VAE (Categorical Mixture Model - Variational Autoencoder). Instead of a linear regression in the prediction component of the encoder, the CMM-VAE method fits a higher-dimensional logistic regression. Furthermore, the subsequent regularization of the latent space is modified. Instead of implementing the regularization as a regression from the continuous target variable to the latent space, which is how the loss function of the RMM-VAE is derived, the CMM-VAE method predicts the categorical cluster assignment from the (also categorical) target variable. This enables the regularization of the latent space using a categorical target variable. The loss function derived for the CMM-VAE architecture, as well as a more detailed explanation of differences to the RMM-VAE method, can be found in Appendix A.

The encoder and decoder of the CMM-VAE architecture were implemented using three dense neural network layers of decreasing dimensionalities (256, 128 and 64) and a ReLU activation function. The architecture was implemented using the Python library keras (Chollet et al., 2015). The models were iteratively trained for 150 epochs with a batch size of 128 and evaluated on different train-test splits in a k-fold cross-validation approach. The best-performing weights were then used to encode the entire dataset. A latent space of dimensionality 15 was selected.

2.3 Predictability metrics

2.3.1 The predictive skill of regimes in subseasonal hindcast experiments

The skill of subseasonal dynamical reforecasts in predicting the occurrence of the (targeted) circulation regimes is analysed using a lower-resolution reforecast experiment using the 47r3 cycle of the ECMWF IFS (CY47R3_LR) developed by Roberts et al. (2023) ranging back to 1980. This lower-resolution reforecast was shown to predict circulation regimes over the region sufficiently well to justify the trade-off between lower resolution and extended time period.

The 11-member ensemble forecasts of geopotential height at 500hPa up to lead times of 47 days, initialized on the 1st, 8th, 15th and 22nd of each month, were downloaded through MARS for the period 1980-2020. Start dates covering the extended winter period November - March were selected (i.e. from 22/09 to 22/03). The reforecasts were pre-processed to match ERA5 data: after selecting the region over which the reanalysis data was analyzed (20-80°N / 50°W-30°E) and re-gridding reforecasts to a resolution of 2.5°x2.5°, the climatological mean was subtracted, and the result was divided by the standard deviation across grid cells. Both the mean and standard deviation were calculated for each day of the year and lead time independently across ensemble members. Finally, for each ensemble member, a rolling window mean of five days was calculated to correspond to the window length chosen in the reanalysis data.

Following these pre-processing steps, the reforecasts were projected onto the circulation regimes calculated in the reanalysis data. For the two linear methods, PCA and CCA, data for each ensemble member, lead time and initialization date was

projected onto the principal component or canonical covariates pre-computed on reanalysis data. Subsequently, the k-means clustering fitted on reanalysis data was applied to the projected reforecasts to predict cluster assignments of reforecasts. For the variational autoencoder method, the VAE trained on reanalysis data was used to predict the latent space, cluster assignment and reconstruction of reforecasts for each ensemble member, lead time and initialization date.

The forecast skill of the circulation regimes in subseasonal dynamical reforecasts was evaluated using (1) the Brier skill score extended to a multi-category forecast and (2) the area under the Receiver Operating Characteristic (ROC-AUC). The Brier score is a strictly proper scoring rule defined as $BS = \frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{m} (\delta_{i_n j} - p_j)^2$ (Gneiting and Raftery, 2007), where m is the number of forecast categories and N is the number of timesteps. δ_{ij} is the Kronecker delta which equals 1 if the observation i at timestep n corresponds to category j, and 0 otherwise, and p_j the forecast probability of category j. Based on the Brier score, the Brier skill score was calculated with respect to the skill score of a climatological forecast in the following way: BSS = 1 - BS_forecast / BS_climatology. To compare the performance across methods, the score was calculated over all regimes since the predictive skill of individual regimes cannot be directly compared between methods due to the non-correspondence of individual regimes. The ROC curve shows the hit rate of the forecast over the false alarm rate as a function of the threshold (that a forecast must exceed to define a hit) extended to a multi-category forecast.

2.4 Conditional predictability in reanalysis data based on information theory

For the characterisation of the MJO, the real-time multivariate (RMM) MJO index was used (Wheeler and Hendon, 2004). This index is based on the first two principal components of combined fields of daily anomalies in 15°S-15°N outgoing longwave radiation, zonal winds at 850 and 200hPa, and the removal of the interannual variability by linear regression against the SST time series reflecting ENSO, which gives an RMM1 and RMM2 index. These two indices can be plotted in a phase diagram, where an amplitude larger than 1 represents the occurrence of an MJO event, and the angle assigns a day to MJO phases 1 to 8, which reflect the propagation of the MJO from Africa over the Indian Ocean and the Maritime Continent to the Western Pacific. MJO indices were accessed from the Australian Bureau of Meteorology and calculated based on Gottschalck et al. (2010) and Wheeler and Hendon (2004).

To investigate the tropospheric impacts of weak and strong polar vortex states, the zonally averaged zonal wind at 60°N and 100 hPa was calculated from December to March and divided into terciles over the entire season to reflect weak, neutral and strong vortex conditions based on ERA5 reanalysis data. The pressure level of 100 hPa was chosen to capture the downward impact of the stratospheric variability. This variable and index has been used in previous studies, including Charlton-Perez et al. (2018).

Subseasonal teleconnections from both the MJO and SPV are themselves known to be modulated by seasonal modes of variability such as the QBO and ENSO (Lee et al., 2019; Toms et al., 2020). While these are not directly investigated here, the seasonal intermittency of the subseasonal teleconnections is assessed using a block-bootstrapping approach which provides an estimate of the robustness of the subseasonal teleconnection across seasons (Roberts et al., 2023). Furthermore, the modulation of the SPV by the MJO (Garfinkel et al., 2014) is not assessed.

The predictability of the (targeted) regimes given these teleconnections was then evaluated based on information theory, an approach which has been applied in climate science and machine learning (DelSole, 2004; Fang et al., 2024; Runge et al., 2012). In particular, the conditional entropy and mutual information between the regimes and the two subseasonal teleconnections are evaluated (Murphy, 2022). Given two variables X and Y, their individual entropies $H(X)$ and $H(Y)$ measure the average uncertainty inherent in the possible outcomes of the variable (Murphy, 2022), calculated as follows $H(X) = -\sum_{x \in X} p(x) \log p(x)$. Here, Y is taken to be the regime and X the phase or tercile of the large-scale driver (MJO or SPV). Based on previous literature, we assume that knowing the phase or tercile of the large-scale driver (i.e. X) will give us information about, or reduce the uncertainty in Y. This can be formalized as the metric of conditional entropy $H(Y|X)$ that quantifies the amount of uncertainty remaining about the target variable Y given that X is observed, and therefore provides an estimate of conditional predictability. Given two discrete variables, their conditional entropy is calculated as follows: $H(Y|X) = \mathbb{E}_{p(X)}[H(p(Y|X)] = -\sum_x p(x) \sum_y p(y|x) \log p(y|x)$. Subtracting the conditional entropy from the uncertainty inherent in the variable Y gives a symmetric measure of information shared between two variables, that is, their mutual information: $I(X,Y) = H(Y) - H(Y|X) = H(X) - H(X|Y)$. Here, mutual information is adjusted to account for the mutual information that would be detected between two independent sets of clusters (Vinh et al., 2010).

Conditional entropy provides an estimate of the average conditional predictability of the regional dynamical drivers given the large-scale teleconnections over lead times, irrespective of the specific model chosen to make this prediction. However, under distributional assumptions, conditional entropy has been shown to provide a lower bound to mean squared error achievable in a regression model aiming to predict Y from X (Vinh et al., 2010). Since conditional entropy does not take into account the uncertainty in the atmospheric regimes themselves, both conditional entropy and adjusted mutual information were evaluated. In contrast to metrics such as momentary information transfer proposed by Runge et al. (2012), evaluating mutual information and conditional entropy here does not disentangle the information transfer from X to Y at a specific lead time τ from the information transfer at lead time $\tau - \epsilon$ combined with autocorrelation of Y during the time ϵ. However, disentangling this difference is not considered relevant for our present purpose.

3   Results

3.1   Characteristics of the circulation regimes and their informativeness precipitation over Morocco

Figure 4 shows the circulation regimes identified by the different methods, alongside the odds ratio of extreme precipitation at each grid cell during the days assigned to the respective regimes.

The PCA + k-means method identifies the regimes expected from previous literature: the two phases of the NAO (regimes 1 and 3), Scandinavian Blocking (regime 2), an Atlantic low and the Atlantic Ridge (regimes 4 and 5) (Falkena et al., 2020; Michelangeli et al., 1995). The negative phase of the NAO is associated with a moderate increase in the probability of extreme precipitation, which is in line with existing literature that finds a correlation between the NAO- and wet conditions over Morocco due to a southward shift of the North Atlantic storm tracks (Driouech et al., 2021; Tramblay et al., 2012).
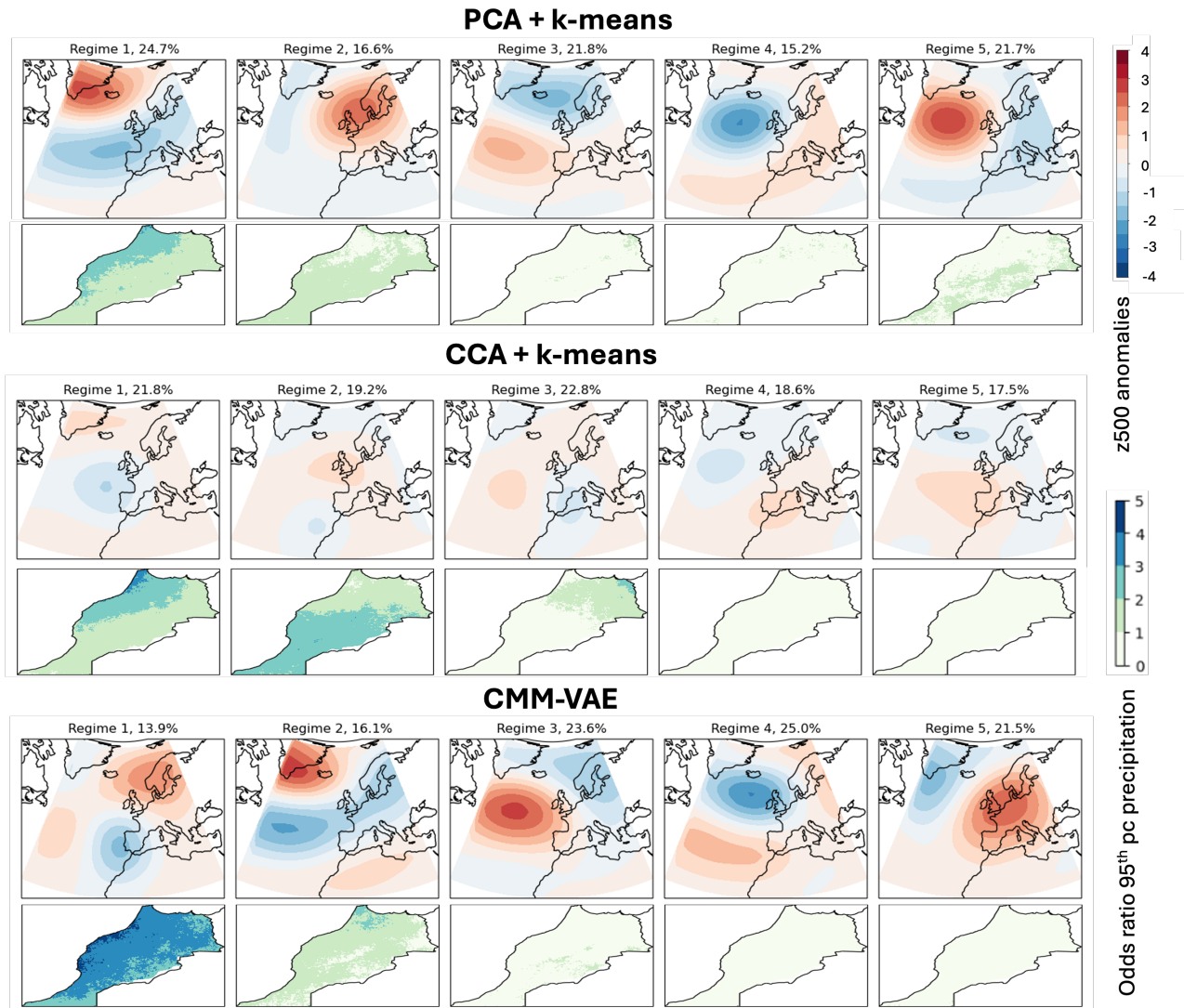
Figure 4. Identified circulation regimes (top rows) and corresponding odds ratios of extreme precipitation (bottom rows) for the four different methods with the number of clusters specified as $k = 5$. The regime frequencies are given in percent. The odds ratio of extreme precipitation corresponds to the ratio of the probability of the climatological 95th percentile of precipitation at the grid cell conditional on that circulation regime, divided by the unconditional probability of 95th percentile of precipitation (i.e. 0.05). The regimes are ordered in decreasing order of total precipitation during the days assigned to this cluster by the respective method.
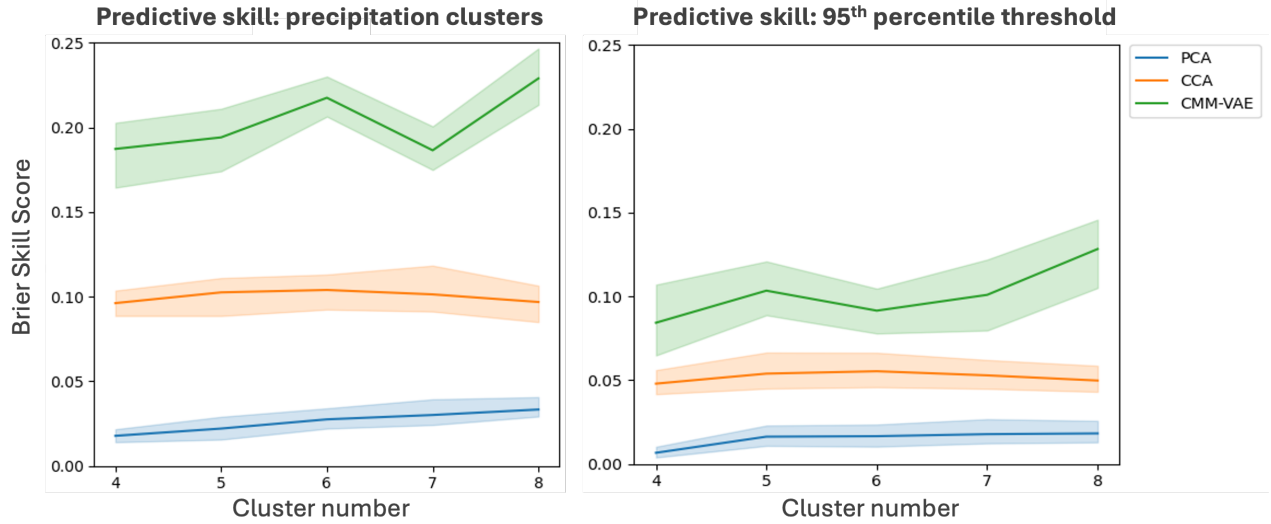
Figure 5. Informativeness of the regimes of exceedance of the 95th percentile of total precipitation over Morocco (right) and precipitation clusters shown in Figure 2 (left), evaluated using the Brier Skill Score. 95% confidence interval computed based on bootstrap procedure with n=50.

290   The CMM-VAE method also identifies the negative phase of the NAO and finds it to be associated with a moderate increase in extreme precipitation (CMM-VAE regime 2). However, CMM-VAE identifies another dynamical pattern associated with an even higher increase of extreme precipitation over Morocco (CMM-VAE regime 1): this regime is related to a Scandinavian Blocking alongside a localized low around the western coast of the Iberian Peninsula and Morocco. The dynamical pattern represented by this additional regime is consistent with the geopotential height anomalies during extreme precipitation events

295   analysed in previous publications (Chaqdid et al., 2023; Toreti et al., 2010). Dynamically, it relates to (south-)westerly mid-tropospheric flow and associated moisture transport from the Atlantic found to drive precipitation over Morocco (Dayan et al., 2015; Khouakhi et al., 2022). The associated low-level zonal wind and streamfunction anomalies shown in Appendix B indicate a split jet configuration. In contrast, the non-targeted PCA + k-means regimes do not show this additional pattern and do not resolve an increase in extreme precipitation associated with the Scandinavian Blocking regime as it lacks the resolution of the

300   localized low off the coast of the Iberian Peninsula.

Aside from disentangling this additional regime modulating extreme precipitation over Morocco, the CMM-VAE method identifies regimes similar to those found in the non-targeted PCA + k-means clustering approach: the NAO+ and Atlantic Ridge regimes look relatively similar, while the CMM-VAE method identifies a slightly southward shifted Scandinavian Blocking regime that is associated with a positive geopotential height anomaly over the entire Mediterranean region. The targeted regimes

305   are statistically well separated and show an overall only slightly reduced persistence compared to the non-targeted regimes (see Appendix B).

The CCA + k-means method projects the geopotential height data into a subspace which is maximally correlated with precipitation over Morocco, thereby identifying targeted clusters which are associated with an increase in extreme precipitation over Morocco but which show less structure in the rest of the atmospheric phase space.

310 Overall, we find that the CMM-VAE is able to identify a regime representation of dynamical conditions over the region that is more informative of precipitation extremes over Morocco by disentangling a dynamical driver not identified in conventional PCA + k-means regimes. Informativeness is here diagnosed using the skill of the regimes in predicting the target variable. In contrast to CCA, the CMM-VAE method identifies more structure overall in the atmospheric phase space, and therefore regimes which are more persistent and statistically robust (see Appendix B).

315 To further quantify the informativeness of the regimes of precipitation over Morocco, we construct a forecast of extreme precipitation and precipitation clusters based on the regime occurrence and the associated conditional probability of the target variable. We compare the skill of this forecast for the different methods (Figure 5). The higher the skill, the stronger the link between circulation regimes and precipitation over Morocco.

We find that the CMM-VAE method outperforms PCA + k-means and CCA + k-means in terms of predicting both the 320 precipitation clusters (Figure 5, left) as well as the exceedance of 95th percentile precipitation (Figure 5, right), hence identifying circulation regimes that are more informative of the extreme precipitation over Morocco and confirming the analysis of odds ratios shown in Figure 4. The skill is overall higher for predicting the precipitation cluster assignment compared to the threshold exceedance of 95th percentile precipitation.

The cluster number for further investigation, $k = 5$, was selected on the basis of the robustness of cluster centers to subsam-325 pling analyzed in Spuler et al. (2024a). In sensitivity checks performed, it was found that the principle results presented in this paper are not sensitive to the choice of cluster number.

3.2 How well are the circulation regimes predicted in subseasonal hindcasts?

We now investigate the skill of dynamical subseasonal hindcasts in predicting the circulation regimes. The previous section showed that the CMM-VAE method identifies regimes that are more informative of the local-scale impact, namely extreme 330 precipitation over Morocco. The aim of the analysis performed in this section is to investigate if the CMM-VAE regimes are as predictable as those identified with the conventional PCA + k-means approach. Together with the enhanced informativeness of the target variable, this would provide evidence that the CMM-VAE method can identify more suitable representations of regional dynamical drivers that can help improve predictions of precipitation over Morocco across a range of timescales.

The forecast skill of the different regimes in subseasonal hindcasts is assessed using two evaluation metrics: the Brier Skill 335 Score (BSS) and the Area Under the Curve of the Receiver Operator Characteristic (ROC-AUC). The ROC-AUC, shows the hit rate of the forecast over the false alarm rate as a function of a threshold extended to a multi-category forecast and has a similar interpretation to the resolution term of the BSS. Results are shown in Figure 6.

The targeted CMM-VAE regimes are found to be as predictable in terms of both BSS and ROC-AUC as the non-targeted regimes identified using PCA + k-means. The CCA + k-means regimes are overall less predictable in both metrics. Skill drops 340 below zero, i.e. below climatological skill, due to the imperfect climatological calibration of the reforecasts in all methods.
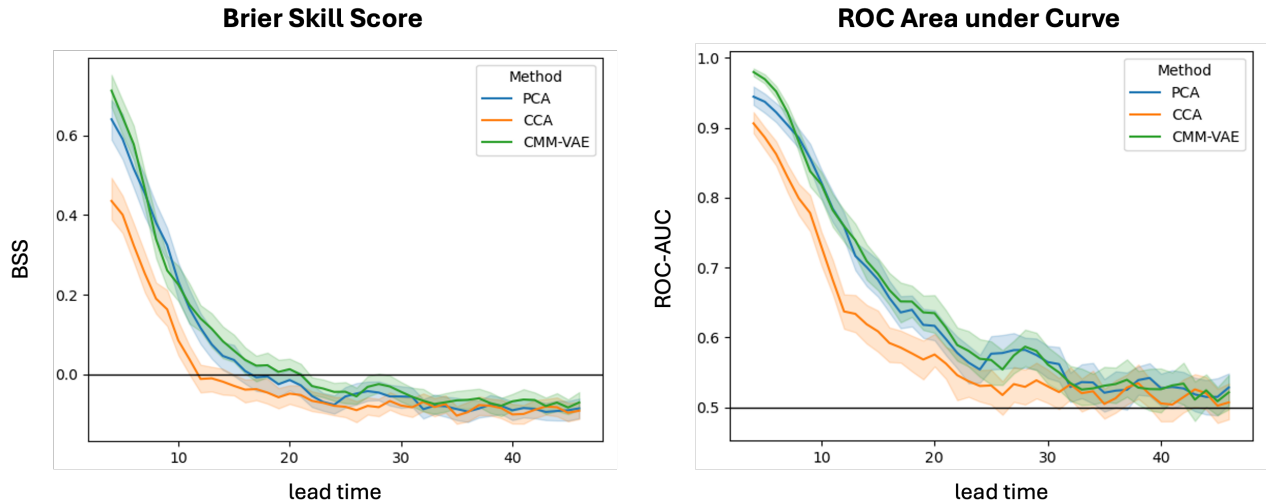
Figure 6. Brier Skill Score (left) and ROC AUC for circulation regime assignment predicted by the subseasonal hindcasts. Confidence interval based on a bootstrapping procedure with n=100.

The BSS can be further decomposed into terms representing the reliability, i.e. calibration or conditional bias, the resolution of the forecast, and the observational uncertainty (Stephenson et al., 2008). We analyse this decomposition for the different regimes to understand the similar performance of CMM-VAE and PCA + k-means regimes in terms of overall skill score (see Appendix C). We find that the CMM-VAE regimes perform slightly worse in terms of resolution but slightly better than PCA + k-means regimes in terms of reliability, i.e. the reliability and resolution of both methods are within the confidence interval of the respective other and there is a small difference in the mean.

3.3 Teleconnections between subseasonal modes of variability and circulation regimes in reanalysis data

In this section, we investigate the predictability of the targeted regimes in reanalysis data, given two known subseasonal teleconnections relevant to the region: the MJO and variability in the SPV. Predictability here is understood in the information theoretical sense as the amount of information shared between two sets of variables - subseasonal modes of variability and the targeted circulation regimes. This is assessed by first analyzing changes in the conditional probabilities of regime occurrence, and then building on this to evaluate adjusted mutual information and conditional entropy of the two sets of variables as information theoretical measures of predictability. This analysis provides insight into whether the predictability of the targeted circulation regimes in subseasonal reforecasts is also physically interpretable in terms of large-scale dynamical drivers.
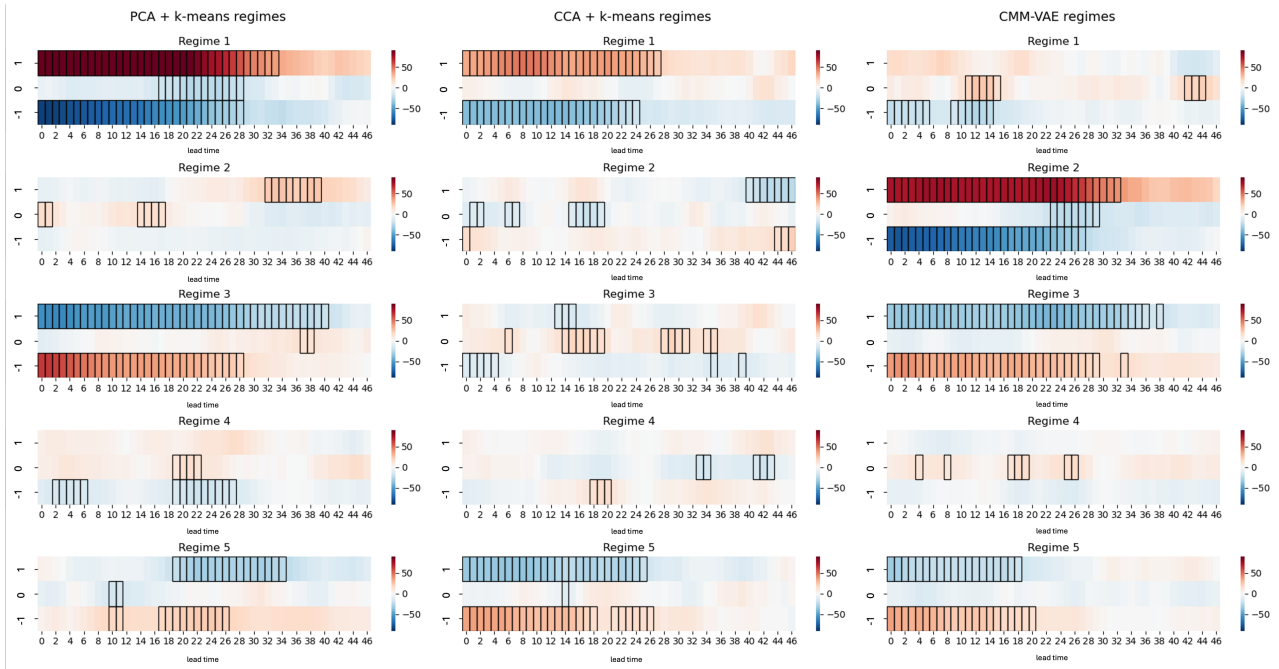
Figure 7. Change in conditional probability of the different circulation regimes (= absolute difference between conditional and unconditional probability, i.e. if a regime occurs around 10% of days and shows a 10% increase here, it becomes twice as likely) given or following weak (-1), neutral (0) and strong (1) states of the stratospheric polar vortex for lags up to 47 days. Statistically significant changes in the conditional probability are indicated using a black rectangle around the cell. These are calculated using a block-bootstrapping approach that samples entire DJFM seasons from the data with n=1000.

3.3.1 Teleconnections from the stratospheric polar vortex

Figure 7 shows the change in the probability of the different circulation regimes, following weak, neutral or strong states of the polar vortex (labeled -1, 0 and 1 respectively). The circulation regimes are ordered as in Figure 4, i.e. by the occurrence of precipitation in each regime, from high to low.

We find that the influence of the SPV on precipitation over Morocco appears to be primarily modulated via the NAO with an increase in the probability of the NAO- (PCA regime 1 and CMM-VAE regime 2) following weak SPV states. On the other hand, the conditional probability of European blocking and the Atlantic Ridge regime (CMM-VAE regimes 3 and 5) is reduced following weak vortex states. This result is in line with established findings on weaker SPV states leading to an equatorward shift of the tropospheric eddy-driven jet and associated storm tracks (Kidston et al., 2015; Kretschmer et al., 2018). The localized low associated with a Scandinavian Blocking (CMM-VAE regime 1), on the other hand, does not appear to be significantly modulated by the SPV. The CCA + k-means regimes appear to be less strongly modulated by the SPV, although we find a slight increase in the probability of the regime associated with the strongest increase in extreme precipitation following weak vortex states which is in line with the dynamical mechanisms found for the PCA + k-means and CMM-VAE regimes.
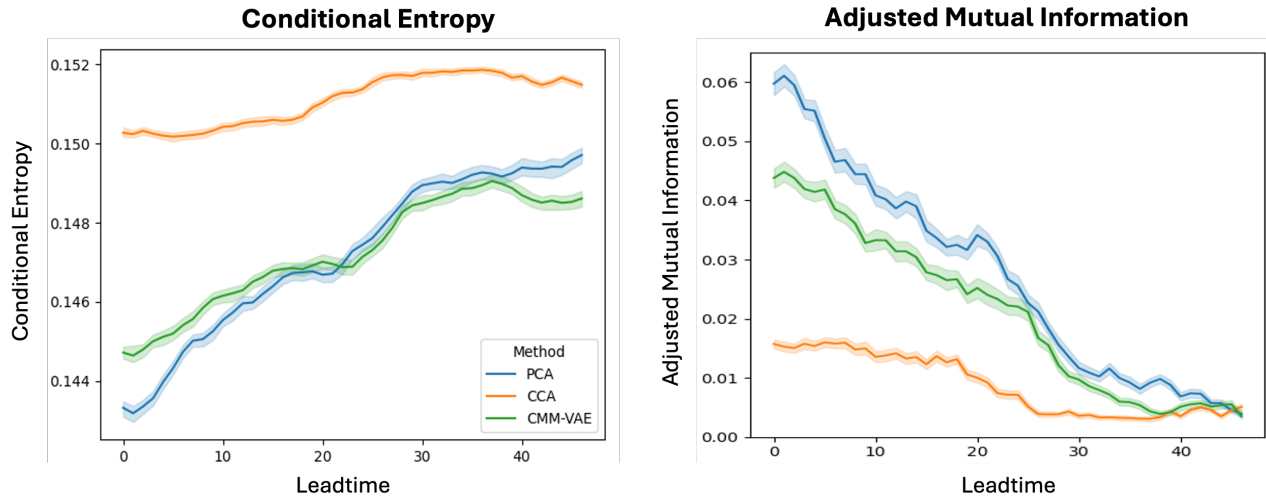
Figure 8. Left: Conditional entropy, i.e. the average uncertainty in the circulation regime given that the state of the polar vortex is known, averaged across the regimes. Lower values are better. Right: Adjusted Mutual Information. i.e. shared information between the circulation regime and the state of the stratospheric polar vortex which in addition to conditional entropy also accounts for the uncertainty in the regimes themselves, averaged across regimes. Higher values are better.

To quantify this difference in conditional predictability of the regimes given teleconnection from the SPV, we analyze the mutual information between the different regimes and SPV states, as well as the conditional entropy of the regimes given the

370 SPV state. These metrics provide an information theoretical assessment of predictability that is here aggregated across regimes. Results are shown in Figure 8.

We find that conditional entropy, which quantifies the average uncertainty in the targeted circulation regime remaining given the state of the polar vortex is known, is similar for the PCA + k-means and CMM-VAE regimes and higher for the CCA regimes. Mutual information between the regimes and SPV states, which also takes into account the uncertainty in the regimes

375 themselves, is highest for the PCA + k-means regimes and lowest for the CCA regimes. The difference between conditional entropy and mutual information is due to the fact that the entropy of the PCA + k-means regimes appears to be larger than that of the CMM-VAE regimes which can vary depending on the lead time analysed in the hindcasts.

These results show that the CMM-VAE and PCA + k-means regimes are more predictable given knowledge of the SPV compared to the CCA + k-means regimes, hence capturing the downward impact of this stratospheric teleconnection better.

380 The downward impact of the SPV assessed by these two metrics is found to decrease somewhat monotonically over time.

3.3.2   Teleconnections from the Madden-Julian oscillation

Figure 9 shows the change in probability of the different circulation regimes, following different phases of the MJO. The results highlight the oscillatory nature of the MJO, with the impact of different MJO phases on individual regimes propagating over
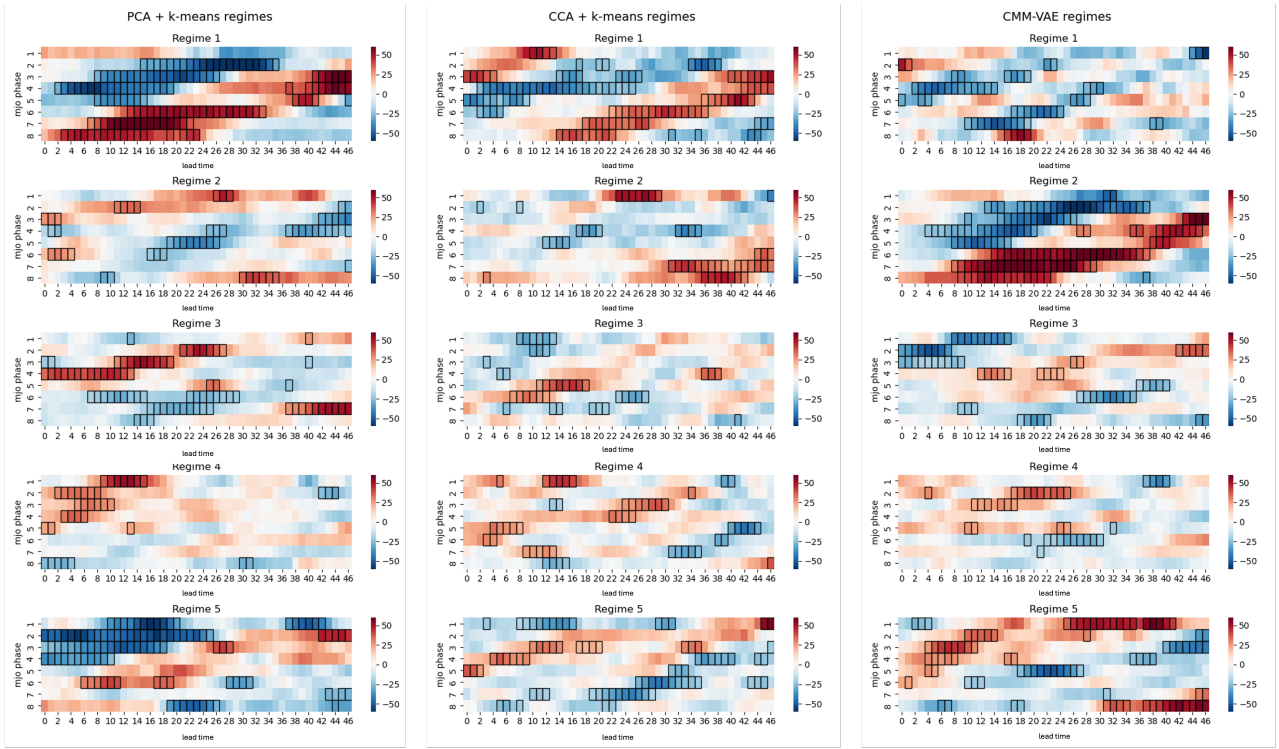
Figure 9. As in Figure 8 but for MJO phases.

lead times. Furthermore, large and significant changes in the conditional probabilities of individual regimes are found even at
lead times up to 47 days, whereas stratospheric impacts tend to decay earlier (see Figure 7).

CMM-VAE regime 2 and PCA regime 1 which are associated with a negative NAO pattern, are modulated most strongly by
the MJO, with a decrease in occurrence probability following MJO phases 1-4 and an increase following phases 6-8. This is
consistent teleconnections between the MJO and NAO reported in the literature on the North Atlantic region (Cassou, 2008),
as well as Morocco specifically (Gadouali et al., 2020). CMM-VAE regime 1, which is associated with the highest increase in
the probability of extreme precipitation over Morocco, shows some modulation by the MJO.

Results for the conditional entropy and adjusted mutual information shown in Figure 10 highlight that the PCA + k-means
and CMM-VAE regimes capture the dynamical teleconnection mechanisms between MJO and circulation over the Mediter-
ranean region slightly better than CCA + k-means regimes, and hence are more predictable given knowledge of the MJO phase.
In contrast to the teleconnection from the SPV, predictability from the MJO oscillates over lead times, and for the CMM-VAE
regimes shows the lowest level of conditional entropy (highest level of mutual information) for lead times of 47 days.
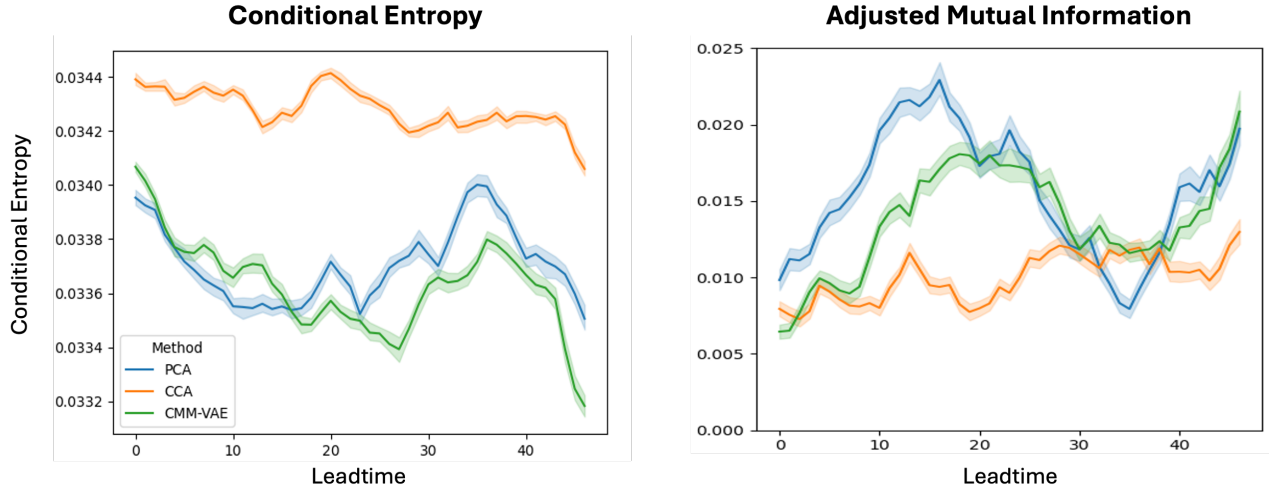
Figure 10. As in Figure 7 but for MJO phases.

4   Discussion and Conclusions

This paper introduces a novel method, the Categorical Mixture Model Variational Autoencoder (CMM-VAE) to identify regional dynamical drivers of a chosen impact variable in the form of targeted circulation regimes. Compared to two well-established linear methods for identifying circulation regimes, we find the targeted CMM-VAE regimes are more informative
400   of the impact variable of interest while maintaining their predictability in subseasonal hindcasts and dynamical interpretability. Applying the method to study drivers of precipitation over Morocco, we find that the method is able to disentangle an additional circulation pattern as a dynamical driver of extreme precipitation.

Through a regularized variational autoencoder architecture and modified loss function, the CMM-VAE method extends the method previously presented in Spuler et al. (2024a) to a higher-dimensional categorical target variable. This new method
405   enables the identification of probabilistic circulation regimes targeted to spatial patterns of extreme precipitation over Morocco. The identified regimes are compared to regimes identified using Principal Component Analysis and k-means clustering (PCA + k-means) as a baseline non-targeted method, and Canonical Correlation Analysis and k-means clustering (CCA + k-means) as a linear targeted clustering method.

The CMM-VAE method identifies a probabilistic partitioning of the atmospheric phase space that better disentangles dy-
410   namical patterns modulating extreme precipitation over Morocco (Figure 4), thereby enhancing the informativeness of the resulting regimes (Figure 5). The additional regime identified by the CMM-VAE method, which is not found in the PCA or CCA + k-means regimes, is associated with a Scandinavian blocking together with a localised cut-off low off the coast of Morocco. This dynamical pattern is consistent with previous literature investigating dynamical drivers of extreme precipitation over Morocco and the Western Mediterranean (Chaqdid et al., 2023; Toreti et al., 2010).

415   Investigating the skill of dynamical subseasonal hindcasts in predicting the circulation regimes, we find that the targeted CMM-VAE regimes are as predictable as the baseline non-targeted PCA + k-means regimes in subseasonal hindcasts, and more predictable than the regimes identified using CCA + k-means (Figure 6). This is a significant result compared to previous studies, which showed a trade-off between identifying locally informative patterns and regimes that are predictable at subseasonal lead times (Bloomfield et al., 2021). The result imply that in this region, the CMM-VAE method is able to identify

420 a representation of regional dynamical drivers that balances and even resolves the trade-off between informativeness of local impacts and subseasonal predictability. The lower predictability of the targeted CCA + k-means regimes can be attributed to the fact that the method projects the data into a correlated subspace but does not capture the structure in the rest of the phase space as well (see Figure 4).

    The ability of this probabilistic machine learning method to strike a balance between local informativeness and predictability

425 of the targeted regimes can be attributed to several factors. One is the efficiency of neural networks in identifying a nonlinear transformation function that encodes the information in a more informative reduced space and therefore enables the subsequent regularisation, i.e. targeting, of the dimensionality reduction. The second is that the loss function derived using variational inference represents the different objectives of targeted clustering - such as representation of the full phase space and informativeness of the target variable - in a coherent statistical model that can be jointly optimized.

430   All methods for identifying regimes studied here require choosing the region over which to cluster atmospheric circulation, as well as the number of clusters k, a priori. The North Atlantic region was chosen based on previous literature highlighting the importance of dynamical drivers from the North Atlantic, but regimes were also analysed for circulation anomalies over a smaller Mediterranean region. The number of clusters was chosen based on the sensitivity of the cluster centres to subsampling analysed in Spuler et al. (2024a), but results for other cluster numbers were also computed (e.g. Figure 5. We find

435 that the improved informativeness of the CMM-VAE regimes, as well as their equal predictability in subseasonal hindcasts, are robust to both the choice of k as well as the choice of region.

    We also investigate and explain this predictability in terms of subseasonal teleconnections relevant to the region, the MJO and variability in the SPV. Conditional predictability given these teleconnections is analysed based on mutual information and conditional entropy, two information-theoretical measures of predictability. In line with the analysis of predictive skill in

440 subseasonal hindcast data, the CMM-VAE regimes show similar levels of conditional predictability as the non-targeted PCA + k-means regimes. Conditional predictability of the regimes given the SPV is higher during strong or weak, as opposed to neutral, vortex states, and decays over subseasonal lead times. The conditional predictability of the regimes given the MJO, on the other hand, shows a clear oscillation across subseasonal lead times. This result highlights potential windows of opportunity for subseasonal forecast skill in predicting precipitation extremes over Morocco.

445   Furthermore, the regimes disentangle distinct dynamical mechanisms through which extreme precipitation over Morocco is modulated by the MJO and variability in the SPV. The results suggest that the impact of both the MJO and SPV on precipitation over Morocco is mediated primarily via the NAO, while the CMM-VAE regime associated with a localised geopotential low pattern and Scandinavian Blocking, which is the one associated with the highest increase in the probability of extreme precipitation, does not show a strong link to the SPV and is somewhat modulated by the MJO. This result highlights that the targeted

CMM-VAE regimes - which are statistically optimised based on the local-scale variable - also represent physical processes that are modulated by large-scale drivers and can be used to understand the modulation of the frequency of precipitation extremes over Morocco by low-frequency modes of internal variability in the climate system.

While the focus of this paper is on predictability and teleconnection relationships at subseasonal lead times, these findings are relevant to seasonal timescales and studies of regional climate change. In subsequent work, targeted regimes could be investigated in future climate projections and used to condition bias adjustment and downscaling approaches (Dorrington et al., 2022; Maraun et al., 2010; Spuler et al., 2024b). The method could be further tested in applications to other regions and target variables, and refined using recently proposed modifications to the loss function of the model to improve its performance in predicting extremes (Wessel et al., 2025). Furthermore potential amplifying effects with local drivers of precipitation could be investigated. Finally, the conditional predictability of the regimes given the MJO and SPV could be investigated in hindcast data and used to identify windows of forecast opportunity for extreme precipitation over Morocco.

Code and data availability. All data used in this study is stored and available for download in Zenodo: https://zenodo.org/records/14534652. The code will be made publicly available upon final publication.

Author contributions. Conceptualization: F.S., M.K., T.S., M.B. Methodology: F.S., M.K. Investigation, software, data curation: F.S. Visualisation: F.S., M.K. Supervision: M.K., T.S. Writing original draft: F.S., M.K., T.S; Writing review and editing: F.S., M.K., T.S., Y.K., M.B. All authors approved the final submitted draft.

Competing interests. The authors declare no competing interests exist.

Acknowledgements. The authors thank Jakob Wessel for useful discussions and feedback. ERA5 reanalysis data (Hersbach et al., 2020) was downloaded from the Copernicus Climate Change Service (C3S) (2023). The results contain modified Copernicus Climate Change Service information 2020. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains. Subseasonal hindcasts were accessed through MARS, the ECMWF meteorological archive. The authors thank Chris Roberts for guidance on accessing the subseasonal hindcasts through MARS.

References

Allen, S., Evans, G. R., Buchanan, P., and Kwasniok, F.: Incorporating the North Atlantic Oscillation into the post-processing of MOGREPS-G wind speed forecasts, Quarterly Journal of the Royal Meteorological Society, 147, 1403–1418, https://doi.org/10.1002/qj.3983, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3983, 2021.

Bach, E., Krishnamurthy, V., Mote, S., Shukla, J., Sharma, A. S., Kalnay, E., and Ghil, M.: Improved subseasonal prediction of South Asian monsoon rainfall using data-driven forecasts of oscillatory modes, Proceedings of the National Academy of Sciences, 121, e2312573 121, https://doi.org/10.1073/pnas.2312573121, publisher: Proceedings of the National Academy of Sciences, 2024.

Baker, L. H., Shaffrey, L. C., and Scaife, A. A.: Improved seasonal prediction of UK regional precipitation using atmospheric circulation, International Journal of Climatology, 38, e437–e453, https://doi.org/10.1002/joc.5382, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.5382, 2018.

Bloomfield, H. C., Brayshaw, D. J., and Charlton-Perez, A. J.: Characterizing the winter meteorological drivers of the European electricity system using targeted circulation types, Meteorological Applications, 27, e1858, https://doi.org/10.1002/met.1858, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/met.1858, 2020.

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

Bommer, P. L., Kretschmer, M., Spuler, F. R., Bykov, K., and Höhne, M. M.-C.: Deep Learning Meets Teleconnections: Improving S2S Predictions for European Winter Weather, https://doi.org/10.48550/arXiv.2504.07625, arXiv:2504.07625 [cs], 2025.

Cassou, C.: Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation, Nature, 455, 523–527, https://doi.org/10.1038/nature07286, number: 7212 Publisher: Nature Publishing Group, 2008.

Chapman, J. and Wang, H.-T.: CCA-Zoo: A collection of Regularized, Deep Learning based, Kernel, and Probabilistic CCA methods in a scikit-learn style framework, Journal of Open Source Software, 6, 3823, https://doi.org/10.21105/joss.03823, 2021.

Chaqdid, A., Tuel, A., El Fatimy, A., and El Moçayd, N.: Extreme rainfall events in Morocco: Spatial dependence and climate drivers, Weather and Climate Extremes, 40, 100 556, https://doi.org/10.1016/j.wace.2023.100556, 2023.

Charlton-Perez, A. J., Ferranti, L., and Lee, R. W.: The influence of the stratospheric state on North Atlantic weather regimes, Quarterly Journal of the Royal Meteorological Society, 144, 1140–1151, https://doi.org/10.1002/qj.3280, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3280, 2018.

Chollet, F., , et al.: Keras, https://keras.io, 2015.

Coughlan de Perez, E., van Aalst, M., Choularton, R., van den Hurk, B., Mason, S., Nissan, H., and Schwager, S.: From rain to famine: assessing the utility of rainfall observations and seasonal forecasts to anticipate food insecurity in East Africa, Food Security, 11, 57–68, https://doi.org/10.1007/s12571-018-00885-9, 2019.

Dawson: eofs: A Library for EOF Analysis of Meteorological, Oceanographic, and Climate Data, https://openresearchsoftware.metajnl.com/articles/10.5334/jors.122, 2016.

Dayan, U., Nissen, K., and Ulbrich, U.: Review Article: Atmospheric conditions inducing extreme precipitation over the eastern and western Mediterranean, Natural Hazards and Earth System Sciences, 15, 2525–2544, https://doi.org/10.5194/nhess-15-2525-2015, publisher: Copernicus GmbH, 2015.

de Fondeville, R., Wu, Z., Székely, E., Obozinski, G., and Domeisen, D. I. V.: Improved extended-range prediction of persistent stratospheric perturbations using machine learning, Weather and Climate Dynamics, 4, 287–307, https://doi.org/10.5194/wcd-4-287-2023, publisher: Copernicus GmbH, 2023.

510   DelSole, T.: Predictability and Information Theory. Part I: Measures of Predictability, https://journals.ametsoc.org/view/journals/atsc/61/20/1520-0469_2004_061_2425_paitpi_2.0.co_2.xml, section: Journal of the Atmospheric Sciences, 2004.

Dinku, T., Funk, C., Peterson, P., Maidment, R., Tadesse, T., Gadain, H., and Ceccato, P.: Validation of the CHIRPS satellite rainfall estimates over eastern Africa, Quarterly Journal of the Royal Meteorological Society, 144, 292–312, https://doi.org/10.1002/qj.3244, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3244, 2018.

515   Domeisen, D. I. V., Grams, C. M., and Papritz, L.: The role of North Atlantic–European weather regimes in the surface impact of sudden stratospheric warming events, Weather and Climate Dynamics, 1, 373–388, https://doi.org/10.5194/wcd-1-373-2020, publisher: Copernicus GmbH, 2020.

Dorrington, J., Strommen, K., and Fabiano, F.: Quantifying climate model representation of the wintertime Euro-Atlantic circulation using geopotential-jet regimes, Weather and Climate Dynamics, 3, 505–533, https://doi.org/10.5194/wcd-3-505-2022, publisher: Copernicus

520   GmbH, 2022.

Dorrington, J., Wenta, M., Grazzini, F., Magnusson, L., Vitart, F., and Grams, C.: Precursors and pathways: Dynamically informed extreme event forecasting demonstrated on the historic Emilia-Romagna 2023 flood, https://doi.org/10.5194/egusphere-2024-415, 2024.

Driouech, F., Déqué, M., and Sánchez-Gómez, E.: Weather regimes—Moroccan precipitation link in a regional climate change simulation, Global and Planetary Change, 72, 1–10, https://doi.org/10.1016/j.gloplacha.2010.03.004, 2010.

525   Driouech, F., Stafi, H., Khouakhi, A., Moutia, S., Badi, W., ElRhaz, K., and Chehbouni, A.: Recent observed country-wide climate trends in Morocco, International Journal of Climatology, 41, E855–E874, https://doi.org/10.1002/joc.6734, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.6734, 2021.

Dunstone, N., Smith, D. M., Hardiman, S. C., Davies, P., Ineson, S., Jain, S., Kent, C., Martin, G., and Scaife, A. A.: Windows of opportunity for predicting seasonal climate extremes highlighted by the Pakistan floods of 2022, Nature Communications, 14, 6544,

530   https://doi.org/10.1038/s41467-023-42377-1, number: 1 Publisher: Nature Publishing Group, 2023.

Falkena, S. K., de Wiljes, J., Weisheimer, A., and Shepherd, T. G.: Revisiting the identification of wintertime atmospheric circulation regimes in the Euro-Atlantic sector, Quarterly Journal of the Royal Meteorological Society, 146, 2801–2814, https://doi.org/10.1002/qj.3818, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3818, 2020.

Falkena, S. K., de Wiljes, J., Weisheimer, A., and Shepherd, T. G.: Detection of interannual ensemble forecast signals over the North

535   Atlantic and Europe using atmospheric circulation regimes, Quarterly Journal of the Royal Meteorological Society, 148, 434–453, https://doi.org/10.1002/qj.4213, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.4213, 2022.

Fang, X., Dijkstra, H., Wieners, C., and Guardamagna, F.: A Nonlinear Full-Field Conceptual Model for ENSO Diversity, https://doi.org/10.1175/JCLI-D-23-0382.1, section: Journal of Climate, 2024.

Ferranti, L., Magnusson, L., Vitart, F., and Richardson, D. S.: How far in advance can we predict changes in large-scale flow leading to severe

540   cold conditions over Europe?, Quarterly Journal of the Royal Meteorological Society, 144, 1788–1802, https://doi.org/10.1002/qj.3341, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3341, 2018.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J.: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, Scientific Data, 2, 150 066, https://doi.org/10.1038/sdata.2015.66, number: 1 Publisher: Nature Publishing Group, 2015.

Gadouali, F., Semane, N., Muñoz, , and Messouli, M.: On the Link Between the Madden-Julian Oscillation, Euro-Mediterranean Weather Regimes, and Morocco Winter Rainfall, Journal of Geophysical Research: Atmospheres, 125, e2020JD032 387, https://doi.org/10.1029/2020JD032387, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020JD032387, 2020.

García-Martínez, I. M. and Bollasina, M. A.: Sub-monthly evolution of the Caribbean Low-Level Jet and its relationship with regional precipitation and atmospheric circulation, Climate Dynamics, 54, 4423–4440, https://doi.org/10.1007/s00382-020-05237-y, 2020.

Garfinkel, C. I., Benedict, J. J., and Maloney, E. D.: Impact of the MJO on the boreal winter extratropical circulation, Geophysical Research Letters, 41, 6055–6062, https://doi.org/10.1002/2014GL061094, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2014GL061094, 2014.

Ghil, M. and Robertson, A. W.: "Waves" vs. "particles" in the atmosphere's phase space: A pathway to long-range forecasting?, Proceedings of the National Academy of Sciences, 99, 2493–2500, https://doi.org/10.1073/pnas.012580899, 2002.

Giuntoli, I., Fabiano, F., and Corti, S.: Seasonal predictability of Mediterranean weather regimes in the Copernicus C3S systems, Climate Dynamics, 58, 2131–2147, https://doi.org/10.1007/s00382-021-05681-4, 2022.

Gottschalck, J., Wheeler, M., Weickmann, K., Vitart, F., Savage, N., Lin, H., Hendon, H., Waliser, D., Sperber, K., Nakagawa, M., Prestrelo, C., Flatau, M., and Higgins, W.: A Framework for Assessing Operational Madden–Julian Oscillation Forecasts, https://doi.org/10.1175/2010BAMS2816.1, section: Bulletin of the American Meteorological Society, 2010.

Hannachi, A., Straus, D. M., Franzke, C. L. E., Corti, S., and Woollings, T.: Low-frequency nonlinearity and regime behavior in the Northern Hemisphere extratropical atmosphere, Reviews of Geophysics, 55, 199–234, https://doi.org/10.1002/2015RG000509, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2015RG000509, 2017.

Harvey, B., Hawkins, E., and Sutton, R.: Storylines for future changes of the North Atlantic jet and associated impacts on the UK, International Journal of Climatology, 43, 4424–4441, https://doi.org/10.1002/joc.8095, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.8095, 2023.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3803, 2020.

Johnson, R. A. and Wichern, D. W.: Applied Multivariate Statistical Analysis: Pearson New International Edition, Pearson Higher Ed, ISBN 978-1-292-03757-8, google-Books-ID: xCipBwAAQBAJ, 2013.

Jolliffe, I. T. and Cadima, J.: Principal component analysis: a review and recent developments, Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 374, 20150 202, https://doi.org/10.1098/rsta.2015.0202, 2016.

Khouakhi, A., Driouech, F., Slater, L., Waine, T., Chafki, O., Chehbouni, A., and Raji, O.: Atmospheric rivers and associated extreme rainfall over Morocco, International Journal of Climatology, 42, 7766–7778, https://doi.org/10.1002/joc.7676, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.7676, 2022.

Kidston, J., Scaife, A. A., Hardiman, S. C., Mitchell, D. M., Butchart, N., Baldwin, M. P., and Gray, L. J.: Stratospheric influence on tropospheric jet streams, storm tracks and surface weather, Nature Geoscience, 8, 433–440, https://doi.org/10.1038/ngeo2424, publisher: Nature Publishing Group, 2015.

Kingma, D. P. and Welling, M.: Auto-Encoding Variational Bayes, http://arxiv.org/abs/1312.6114, arXiv:1312.6114 [cs, stat], 2013.

Kretschmer, M., Runge, J., and Coumou, D.: Early prediction of extreme stratospheric polar vortex states based on causal precursors: Prediction of extreme vortex states, Geophysical Research Letters, 44, 8592–8600, https://doi.org/10.1002/2017GL074696, 2017.

585 Kretschmer, M., Cohen, J., Matthias, V., Runge, J., and Coumou, D.: The different stratospheric influence on cold-extremes in Eurasia and North America, npj Climate and Atmospheric Science, 1, 1–10, https://doi.org/10.1038/s41612-018-0054-4, number: 1 Publisher: Nature Publishing Group, 2018.

Le, P. V. V., Randerson, J. T., Willett, R., Wright, S., Smyth, P., Guilloteau, C., Mamalakis, A., and Foufoula-Georgiou, E.: Climate-driven changes in the predictability of seasonal precipitation, Nature Communications, 14, 3822, https://doi.org/10.1038/s41467-023-39463-9,
590 number: 1 Publisher: Nature Publishing Group, 2023.

Lee, J. C. K., Lee, R. W., Woolnough, S. J., and Boxall, L. J.: The links between the Madden-Julian Oscillation and European weather regimes, Theoretical and Applied Climatology, 141, 567–586, https://doi.org/10.1007/s00704-020-03223-2, 2020.

Lee, R. W., Woolnough, S. J., Charlton-Perez, A. J., and Vitart, F.: ENSO Modulation of MJO Teleconnections to the North Atlantic and Europe, Geophysical Research Letters, 46, 13 535–13 545, https://doi.org/10.1029/2019GL084683, _eprint:
595 https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019GL084683, 2019.

Lemos, M. C., Kirchhoff, C. J., and Ramprasad, V.: Narrowing the climate information usability gap, Nature Climate Change, 2, 789–794, https://doi.org/10.1038/nclimate1614, number: 11 Publisher: Nature Publishing Group, 2012.

Loudyi, D., Hasnaoui, M. D., and Fekri, A.: Flood Risk Management Practices in Morocco: Facts and Challenges, in: Wadi Flash Floods: Challenges and Advanced Approaches for Disaster Risk Reduction, edited by Sumi, T., Kantoush, S. A., and Saber, M., Natural Disaster
600 Science and Mitigation Engineering: DPRI reports, pp. 35–94, Springer, Singapore, ISBN 9789811629044, https://doi.org/10.1007/978-981-16-2904-4_2, 2022.

Maidment, R. I., Grimes, D., Black, E., Tarnavsky, E., Young, M., Greatrex, H., Allan, R. P., Stein, T., Nkonde, E., Senkunda, S., and Alcántara, E. M. U.: A new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa, Scientific Data, 4, 170 063, https://doi.org/10.1038/sdata.2017.63, number: 1 Publisher: Nature Publishing Group, 2017.

605 Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Reviews of Geophysics, 48, RG3003, https://doi.org/10.1029/2009RG000314, 2010.

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., Dirmeyer, P. A., Ferranti, L., Johnson, N. C., Jones, J.,
610 Kirtman, B. P., Lang, A. L., Molod, A., Newman, M., Robertson, A. W., Schubert, S., Waliser, D. E., and Albers, J.: Windows of Opportunity for Skillful Forecasts Subseasonal to Seasonal and Beyond, Bulletin of the American Meteorological Society, 101, E608–E625, https://doi.org/10.1175/BAMS-D-18-0326.1, publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, 2020.

Mastrantonas, N., Herrera-Lormendez, P., Magnusson, L., Pappenberger, F., and Matschullat, J.: Extreme precipitation events in the Mediter-
615 ranean: Spatiotemporal characteristics and connection to large-scale atmospheric flow patterns, International Journal of Climatology, 41, 2710–2728, https://doi.org/10.1002/joc.6985, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.6985, 2020.

Mastrantonas, N., Furnari, L., Magnusson, L., Senatore, A., Mendicino, G., Pappenberger, F., and Matschullat, J.: Forecasting extreme precipitation in the central Mediterranean: Changes in predictors' strength with prediction lead time, Meteorological Applications, 29, e2101, https://doi.org/10.1002/met.2101, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/met.2101, 2022.

620   Michelangeli, P.-A., Vautard, R., and Legras, B.: Weather Regimes: Recurrence and Quasi Stationarity, Journal of the Atmospheric Sciences, 52, 1237–1256, https://doi.org/10.1175/1520-0469(1995)052<1237:WRRAQS>2.0.CO;2, publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, 1995.

Mindlin, J., Vera, C. S., Shepherd, T. G., and Osman, M.: Plausible Drying and Wetting Scenarios for Summer in Southeastern South America, Journal of Climate, -1, https://doi.org/10.1175/JCLI-D-23-0134.1, publisher: American Meteorological Society Section: Journal

625   of Climate, 2023.

Murphy, K.: Probabilistic Machine Learning: An Introduction, 2022.

Roberts, C. D., Balmaseda, M. A., Ferranti, L., and Vitart, F.: Euro-Atlantic Weather Regimes and Their Modulation by Tropospheric and Stratospheric Teleconnection Pathways in ECMWF Reforecasts, Monthly Weather Review, 151, 2779–2799, https://doi.org/10.1175/MWR-D-22-0346.1, publisher: American Meteorological Society Section: Monthly Weather Review, 2023.

630   Rouges, E., Ferranti, L., Kantz, H., and Pappenberger, F.: Pattern-based forecasting enhances the prediction skill of European heatwaves into the sub-seasonal range, Climate Dynamics, https://doi.org/10.1007/s00382-024-07390-0, 2024.

Roundy, P. E., MacRitchie, K., Asuma, J., and Melino, T.: Modulation of the Global Atmospheric Circulation by Combined Activity in the Madden–Julian Oscillation and the El Niño–Southern Oscillation during Boreal Winter, Journal of Climate, https://doi.org/10.1175/2010JCLI3446.1, section: Journal of Climate, 2010.

635   Runge, J., Heitzig, J., Marwan, N., and Kurths, J.: Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy, Physical Review E, 86, 061 121, https://doi.org/10.1103/PhysRevE.86.061121, 2012.

Saggioro, E., Shepherd, T. G., and Knight, J.: Probabilistic causal network modelling of Southern Hemisphere jet sub-seasonal to seasonal predictability, Journal of Climate, -1, https://doi.org/10.1175/JCLI-D-23-0425.1, publisher: American Meteorological Society Section: Journal of Climate, 2024.

640   Scaife, A. A., Folland, C. K., Alexander, L. V., Moberg, A., and Knight, J. R.: European Climate Extremes and the North Atlantic Oscillation, Journal of Climate, 21, 72–83, https://doi.org/10.1175/2007JCLI1631.1, publisher: American Meteorological Society Section: Journal of Climate, 2008.

Shepherd, T. G., Boyd, E., Calel, R. A., Chapman, S. C., Dessai, S., Dima-West, I. M., Fowler, H. J., James, R., Maraun, D., Martius, O., Senior, C. A., Sobel, A. H., Stainforth, D. A., Tett, S. F. B., Trenberth, K. E., van den Hurk, B. J. J. M., Watkins, N. W., Wilby, R. L., and

645   Zenghelis, D. A.: Storylines: an alternative approach to representing uncertainty in physical aspects of climate change, Climatic Change, 151, 555–571, https://doi.org/10.1007/s10584-018-2317-9, 2018.

Spuler, F. R., Kretschmer, M., Kovalchuk, Y., Balmaseda, M. A., and Shepherd, T. G.: Identifying probabilistic weather regimes targeted to a local-scale impact variable, Environmental Data Science, 3, e25, https://doi.org/10.1017/eds.2024.29, 2024a.

Spuler, F. R., Wessel, J. B., Comyn-Platt, E., Varndell, J., and Cagnazzo, C.: ibicus: a new open-source Python package and comprehensive

650   interface for statistical bias adjustment and evaluation in climate modelling (v1.0.1), Geoscientific Model Development, 17, 1249–1269, https://doi.org/10.5194/gmd-17-1249-2024, publisher: Copernicus GmbH, 2024b.

Stephenson, D. B., Coelho, C. a. S., and Jolliffe, I. T.: Two Extra Components in the Brier Score Decomposition, https://doi.org/10.1175/2007WAF2006116.1, section: Weather and Forecasting, 2008.

Straus, D. M.: Preferred intra-seasonal circulation patterns of the Indian summer monsoon and active-break cycles, Climate Dynamics, 59,

655   1415–1434, https://doi.org/10.1007/s00382-021-06047-6, 2022.

Toms, B. A., Barnes, E. A., Maloney, E. D., and van den Heever, S. C.: The Global Teleconnection Signature of the Madden-Julian Oscillation and Its Modulation by the Quasi-Biennial Oscillation, Journal of Geophysical Research: Atmospheres, 125, e2020JD032 653, https://doi.org/10.1029/2020JD032653, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020JD032653, 2020.

Toreti, A., Xoplaki, E., Maraun, D., Kuglitsch, F. G., Wanner, H., and Luterbacher, J.: Characterisation of extreme winter precipitation in Mediterranean coastal sites and associated anomalous atmospheric circulation patterns, Natural Hazards and Earth System Sciences, 10, 1037–1050, https://doi.org/10.5194/nhess-10-1037-2010, publisher: Copernicus GmbH, 2010.

Tramblay, Y., Badi, W., Driouech, F., El Adlouni, S., Neppel, L., and Servat, E.: Climate change impacts on extreme precipitation in Morocco, Global and Planetary Change, 82-83, 104–114, https://doi.org/10.1016/j.gloplacha.2011.12.002, 2012.

Ullmann, A., Fontaine, B., and Roucou, P.: Euro-Atlantic weather regimes and Mediterranean rainfall patterns: present-day variability and expected changes under CMIP5 projections, International Journal of Climatology, 34, 2634–2650, https://doi.org/10.1002/joc.3864, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.3864, 2014.

Vinh, N. X., Epps, J., and Bailey, J.: Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance, 2010.

Vrac, M. and Yiou, P.: Weather regimes designed for local precipitation modeling: Application to the Mediterranean basin, Journal of Geophysical Research, 115, D12 103, https://doi.org/10.1029/2009JD012871, 2010.

Wessel, J. B., Ferro, C. A. T., Evans, G. R., and Kwasniok, F.: Improving probabilistic forecasts of extreme wind speeds by training statistical post-processing models with weighted scoring rules, https://doi.org/10.1175/MWR-D-24-0151.1, section: Monthly Weather Review, 2025.

Wheeler, M. C. and Hendon, H. H.: An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction, Monthly Weather Review, 132, 1917–1932, https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2, 2004.

Wiel, K. v. d., Bloomfield, H. C., Lee, R. W., Stoop, L. P., Blackport, R., Screen, J. A., and Selten, F. M.: The influence of weather regimes on European renewable energy production and demand, Environmental Research Letters, 14, 094 010, https://doi.org/10.1088/1748-9326/ab38d3, publisher: IOP Publishing, 2019.

Zhang, C., Zhang, J., Xia, X., and Li, D.: Impact of Arctic Stratospheric Polar Vortex on Mediterranean Precipitation, https://doi.org/10.1175/JCLI-D-23-0469.1, section: Journal of Climate, 2024.

Appendix A: Details of the CMM-VAE method

680

The two regularized variational autoencoder methods, the Regression-Mixture Model Variational Autoencoder (RMM-VAE), introduced in Spuler et al. (2024a), and the CMM-VAE method introduced in this paper differ with respect to the way in which the latent space is regularized using the target variable. While in the RMM-VAE architecture, the target variable t is used to directly regularize the latent space z (central panel in Figure A1), this is not possible when working with higher dimensional

685 categorical target variables. In the CMM-VAE architecture, we therefore instead regularize the cluster assignment c using the target variable t. The regularization here means that the cluster assignment c is predicted from the target variable t, and that this prediction is used as a prior on the cluster assignment c_k. This is visualised in the graphical model shown in the right panel of Figure A1.
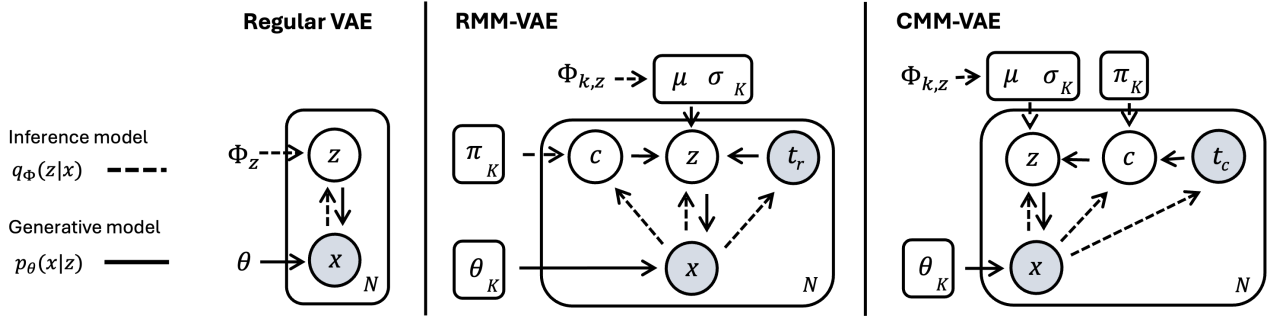
Figure A1. Representation of the statistical models underlying a regular variational autoencoder as graphical models used to derive the loss functions for the considered architectures, RMM-VAE and CMM-VAE. x represents the input z500 space, z - the identified latent space, c - the cluster assignments of individual data points, and t - the target variable. μ and σ are the parameters of the Gaussian distributions fitted into the latent space, π_k is the prior on the cluster occurrence frequency and θ_k are the parameters of the non-linear decoder.

This graphical model corresponds to the following decompositions of the inference and generative distributions.

690
$$q(z,c,t|x) = q(z|x) * q(c|x) * q(t|x) \text{ and } p(z,x,c,t) = p(x|z) * p(z|c) * p(c|t) * p(t) \tag{A1}$$

With these decompositions, we can now follow the standard procedure for Bayesian variational inference to derive the following loss function for the CMM-VAE architecture:

$$\mathcal{L}(x) = -D_{KL}(q_\phi(z,c,t\,|\,x)\,|\,p_\theta(x,z,t,c))$$
$$= \sum_k q_\phi(c^k\,|\,x) \Big[\mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x\,|\,z)] - \mathbb{E}_{q_\phi(z|x)}[D_{KL}(q_\phi(z\,|\,x)\,|\,p(z\,|\,c^k))]$$

695
$$- \mathbb{E}_{q_\phi(t|x)}[D_{KL}(q_\phi(c^k\,|\,x)\,|\,p_\theta(c^k\,|\,t))] \Big] - D_{KL}(q_\phi(t\,|\,x)\,|\,p(t)). \tag{A2}$$

The loss function can be interpreted as follows: $q(c^k|x)$ is the probability of an input datapoint being part of cluster k, predicted from the encoder network. $q(z|x)$ gives the probability distribution of latent space z given input data x, parameterized

as a multivariate Gaussian with mean μ and standard deviations σ, also predicted from the encoder network, as is $q(t|x)$ is the prediction of the precipitation class given input x. $p(z|c^k)$ is the latent variable as predicted from the cluster assignment c^k, $p(x|z)$ the latent space predicted from the decoder network, and $p(c^k|t)$ the cluster assignment predicted from the target variable which is used as prior to the cluster assignment.

The first term of the loss function corresponds to the reconstruction loss of the dimensionality reduction, the second term to cluster coherence, i.e. penalizes the distance of points in the latent space from their assigned cluster center. The third term can be interpreted as the regularization loss, and the last term as the prediction loss of the target variable t.
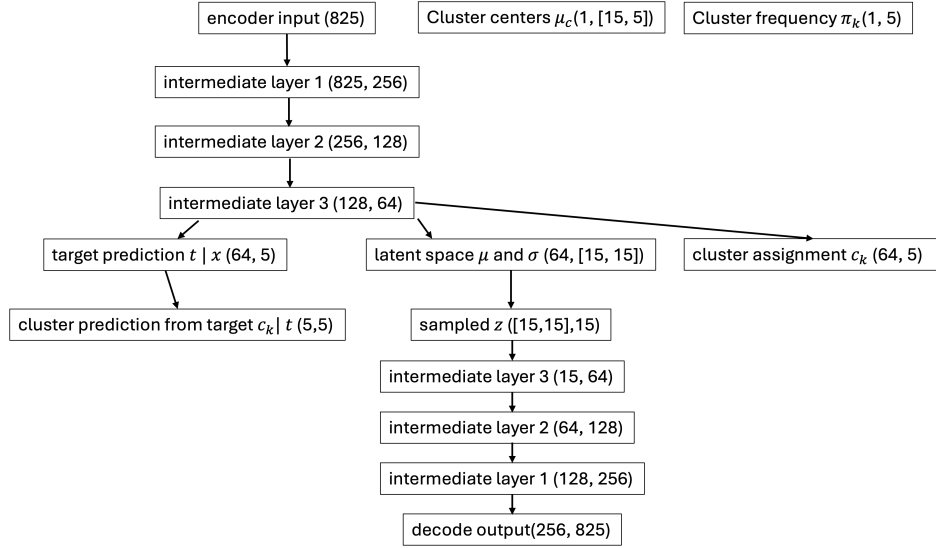
Figure A2. Graphical illustration of the layers implemented in the CMM-VAE architecture. For each layer, the number of input and output dimensions are respectively shown in the first and second number in brackets.

For a more detailed explanation of variational autoencoders and variational inference applied to study atmospheric circulation and the way in which the regularization acts on the latent space, we refer to Spuler et al. (2024a), where this type of architecture is first used to study target circulation regimes.

Appendix B: Further evaluation of the dimensionality reduction and statistical robustness of (targeted) regimes

In terms of reconstructing the original input space (Figure B1, left), the CMM-VAE method performs best by far, while CCA performs worst. This finding is in line with the results presented in Spuler et al. (2024a) for the RMM-VAE method.

In terms of the regime separability and persistence (Figure B1, center and right), the non-targeted PCA + k-means method performs best CCA performs worst according to both metrics, with a silhouette score being around 0, indicating overlapping and statistically non-robust clusters, and the lowest regime persistence. The CMM-VAE method show a similar regime persistence compared to PCA in some regimes, and a slightly lower persistence in others, indicating that the underlying dynamical
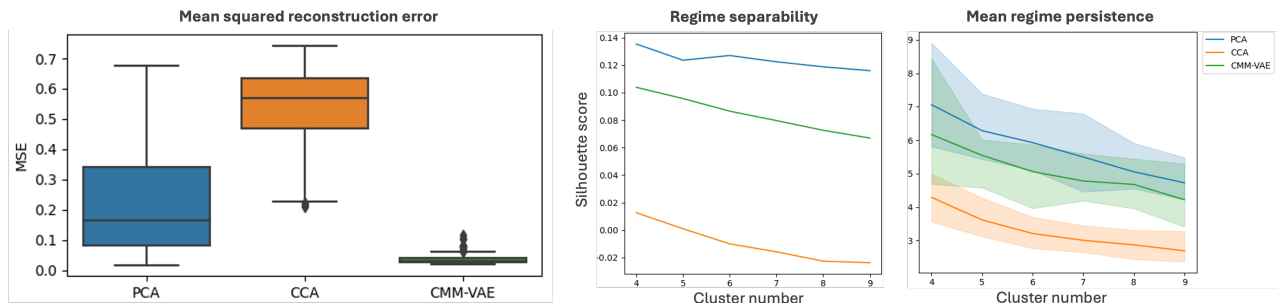
Figure B1. Local properties of the circulation regimes: (left) mean squared error between the z500 field reconstructed from the latent space and the original z500 input field for k=8 regimes. Little sensitivity of the result to the choice of cluster number detected: (centre) silhouette score of the clusters evaluated for different cluster numbers k; (right) mean regime persistence across the k regimes for different choices of cluster number k.
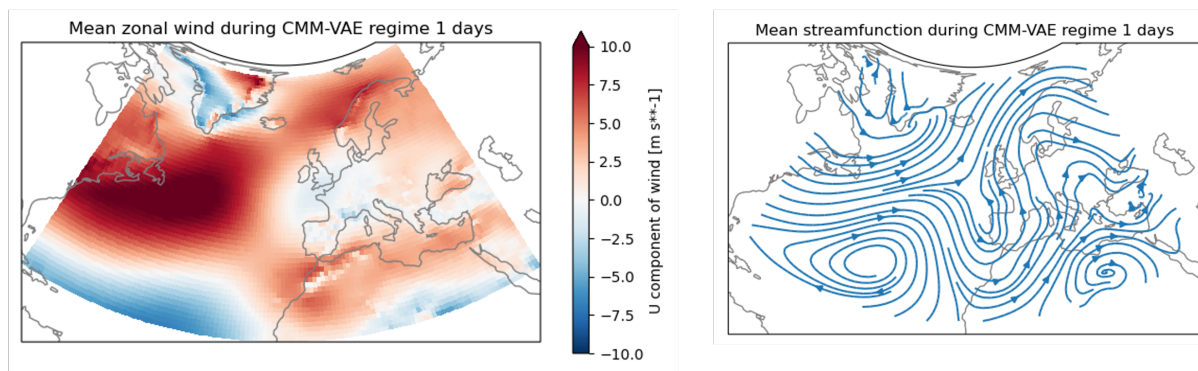
Figure B2. Mean anomalies in the zonal wind at 850hPa (left) and streamfunction calculated from zonal and meridional wind at 850hPa (right) during days associated with CMM-VAE regime 1.

processes modulating the target variable are not as persistent. In terms of regime separability, The CMM-VAE method shows a slightly lower silhouette score compared to the PCA + k-means method. However, a slight reduction would be expected from probabilistic clusters as the silhouette score does not consider regime probabilities and is calculated on the basis of the assignment of a data point to the most likely cluster.

Overall, these results are found to be consistent with the results presented for the RMM-VAE method in Spuler et al. (2024a)) and are therefore included in the appendix here for completeness purposes.

Appendix C: Extended forecast evaluation

The Brier Score can be decomposed into terms representing the reliability / probabilistic calibration, resolution and observational uncertainty of the forecast (Wilks 2019):
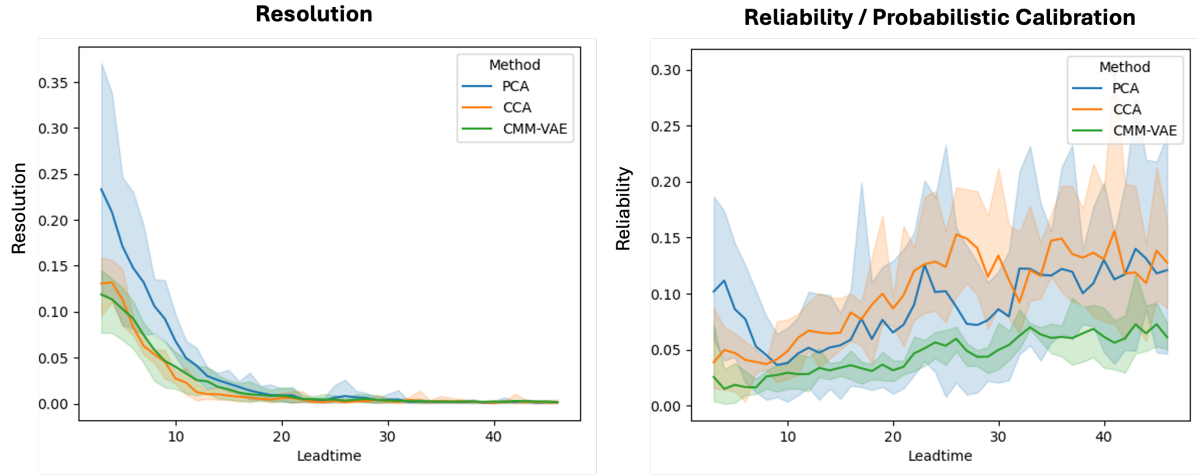
Figure C1. Decomposition of the Brier Skill Score in terms of resolution and reliability as described in the text above.

$$BS = \frac{1}{n}\sum_{i=1}^{I} N_i(y_i - \bar{o}_i)^2 - \frac{1}{n}\sum_{i=1}^{I} N_i(\bar{o}_i - \bar{o})^2 + \bar{o}(1 - \bar{o})^2 \qquad \text{(C1)}$$

725 Computing this analytical decomposition requires binning the forecast probabilities (which calculating the Brier score does not), which makes the results more unstable than the actual score. The results for n=12 bins are shown in Figure C1. The reliability represents the squared difference from the diagonal and assesses the probabilistic calibration, or conditional bias, of the forecast (lower is better), while the resolution term represents the difference from the climatological occurrence frequency of the regime - the larger the difference of the forecast to a climatological forecast, the higher the resolution term.

730 We find that CMM-VAE performs worse than PCA in terms of resolution but better than PCA in terms of reliability. An interesting difference between PCA and CMM-VAE methods is the non-probabilistic vs probabilistic cluster assignment. This means that the two methods are affected differently by the choice of binning and, therefore, the overall skill score shown in Figure 6 might be the more robust metric to use here. Second, the forecasts in this study are evaluated against a single 'true' observed cluster (for the CMM-VAE method, this corresponds to the cluster that is assigned the highest probability), which

735 might represent a disadvantage for the resolution of the CMM-VAE method.