CAMBRIDGE
UNIVERSITY PRESS

Identifying probabilistic weather regimes targeted to a local-scale impact variable

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

[1]Department of Meteorology, University of Reading, Reading, UK
[2]Leipzig Institute for Meteorology, University of Leipzig, Leipzig, Germany
[3]Centre for Advanced Research Computing, University College London, London, UK
[4]European Centre for Medium-Range Weather Forecasts, Reading, UK
Corresponding author: Fiona R. Spuler; Email: f.r.spuler@pgr.reading.ac.uk

Abstract

Large-scale atmospheric circulation patterns, so-called weather regimes, modulate the occurrence of extreme events such as heatwaves or extreme precipitation. In their role as mediators between long-range teleconnections and local impacts, weather regimes have demonstrated potential in improving long-term climate projections as well as sub-seasonal to seasonal forecasts. However, existing methods for identifying weather regimes are not specifically designed to capture the relevant physical processes responsible for variations in the impact variable in question. This paper introduces a novel probabilistic machine learning method, RMM-VAE, for identifying weather regimes targeted to a local-scale impact variable. Based on a variational autoencoder architecture, the method combines non-linear dimensionality reduction with a prediction task and probabilistic clustering in one coherent architecture. The new method is applied to identify circulation patterns over the Mediterranean region targeted to precipitation over Morocco and compared to three existing approaches: two established linear methods and another machine-learning approach. The RMM-VAE method identifies regimes that are more predictive of the target variable compared to the two linear methods, both in terms of terciles and extremes in precipitation, while also improving the reconstruction of the input space. Further, the regimes identified by the RMM-VAE method are also more robust and persistent compared to the alternative machine learning method. The results demonstrate the potential benefit of the new method for use in various climate applications such as sub-seasonal forecasting, and illustrate the trade-offs involved in targeted clustering.

Impact Statement

This paper introduces a new machine learning method for identifying large-scale atmospheric circulation patterns, so-called weather regimes, that modulate a local-scale impact variable such as extreme precipitation. This has the potential to enhance the usefulness of regimes for various climate applications such as impact-based sub-seasonal to seasonal forecasting or downscaling of climate model output. Co-authored by researchers with respective backgrounds in meteorology and computer science, this paper is intended to introduce a new method in an accessible manner to researchers from both communities. Additionally, it aims to illustrate the similarities,

differences and trade-offs associated with novel machine learning methods compared to more established statistical approaches for dimensionality reduction and clustering in weather and climate science.

1. Introduction

1.1. Weather regimes as mediators between local impacts and long-range teleconnections

Large-scale atmospheric circulation modulates the occurrence of extreme events such as heavy precipitation and heat waves that cause devastating impacts on people and livelihoods across the planet. Understanding these dynamical drivers of local extreme impacts can both improve their near-term forecast skill for early-warning decisions (Coughlan de Perez et al., 2022; Gonzalez et al., 2022; Dunstone et al., 2023) and support the physical interpretation of future projected changes and associated uncertainty to identify robust adaptation pathways (Lemos et al., 2012; Shepherd et al., 2018).

Weather regimes, defined as persistent and recurrent circulation patterns, are one common approach to understanding the low-frequency variability of atmospheric circulation (Vautard, 1990; Ghil and Robertson, 2002; Hannachi et al., 2017). In many applications, weather regimes have proven particularly useful as discrete and interpretable mediators between long-range teleconnections in the climate system and local-scale impact variables (Yiou and Nogaj, 2004; Cassou, 2008; Beerli and Grams, 2019; Straus, 2022). For example, over the North Atlantic and European region, weather regimes have been shown to carry a predictability signal from tropical teleconnections such as the Madden-Julian Oscillation and the El-Niño Southern Oscillation (Lee et al., 2019; Gadouali et al., 2020), as well as from stratospheric polar vortex states (Charlton-Perez et al., 2018; Domeisen et al., 2020, while also modulating surface-level variables such as cold-extremes and precipitation (Ferranti et al., 2018; Pasquier et al., 2019).

Due to these teleconnection signals as well as their persistence, weather regimes can improve both the skill and usability of forecasts for extended-range lead times (Allen et al., 2021; Bloomfield et al., 2021. On climate timescales, weather regimes have been used to disentangle the dynamic and thermodynamic components of climate change for extreme event attribution and quantify the role of atmospheric internal variability in observed trends (Cattiaux et al., 2010; Horton et al., 2015; Terray, 2021), as well as to statistically downscale climate models (Ailliot et al., 2009; Maraun et al., 2010).

1.2. Mediterranean weather regimes and precipitation over Morocco

The present study investigates the potential of targeted weather regimes to capture precipitation extremes over Morocco. The country is vulnerable to both flooding driven by extreme rainfall, which has caused over 760 million USD in economic damages since 1950 (Delforge et al., 2023), as well as droughts that have threatened food security, agricultural livelihoods and compounded debt crises of the country (Tanarhte et al., 2024). Extreme precipitation events primarily occur in extended winter, between November and March, and can lead to different types of flooding events, ranging from gradual or flash floods of wadis (river valleys) to torrential flash floods of small mountain basins or flooding of urban areas (Loudyi et al., 2022).

Previous literature has investigated the dynamical drivers and precursors of wintertime extreme precipitation over the Western Mediterranean region, highlighting dynamically driven moisture flux from the Atlantic as a key driver (Ulbrich et al., 2012; Dayan et al., 2015) and positive anomalies in potential vorticity over the eastern Atlantic region as a precursor to extreme precipitation events (Toreti et al., 2016). Toreti et al. (2010) show that over the Western Mediterranean region, the negative geopotential height anomaly pattern associated with extreme precipitation is associated with an alignment of the subtropical jet with the African coastline and anomalous southwesterly surface to mid-tropospheric flow leading to large-scale ascending motions and instability over the Western Mediterranean region.

Weather regimes over both the North Atlantic and European, as well as the Mediterranean region, have been reported to modulate the occurrence of extreme precipitation over Morocco (Driouech et al., 2010; Pasquier et al., 2019; Gadouali et al., 2020; Mastrantonas et al., 2020; Giuntoli et al., 2022). For example, Mastrantonas et al. (2022a) demonstrate that Mediterranean weather regimes determine a significant increase in the probability of above 95th percentile precipitation. Using this information in a simple hybrid forecasting approach, they are able to slightly improve medium-range forecast skills over Morocco. Gadouali et al. (2020), on the other hand, identify seven wintertime weather regimes over the North Atlantic and show their association with precipitation over Morocco, along with their modulation by the Madden-Julian Oscillation. Recent findings by Chaqdid et al. (2023), investigating geopotential height, vertically integrated water vapor flux and wind speed anomalies associated with precipitation extremes over Morocco, however, indicate that these dynamical precursors might not be optimally resolved in weather regimes over either of the two regions, highlighting the scope for a more targeted approach in this region.

1.3. Research gap: identifying targeted weather regimes

Weather regimes are commonly identified using a combination of dimensionality reduction and clustering methods. While the dimensionality reduction step projects the high-dimensional data into a lower-dimensional subspace, the clustering subsequently identifies and assigns discrete regimes within this reduced space (Hannachi et al., 2017).

Following Michelangeli et al. (1995), principal component analysis (PCA, often referred to as Empirical Orthogonal Function or EOF analysis in atmospheric sciences (Hannachi et al., 2007) and k-means have established themselves as common choices for dimensionality reduction and clustering and have been applied in the relevant studies investigating weather regimes over the Western Mediterranean region (Gadouali et al., 2020; Mastrantonas et al., 2020). The advantage of this combination of methods is that they are easy to compute, understand, and interpret. However, they are not inherently more physically meaningful than other statistical dimensionality reduction and clustering methods. Alternative methods have been proposed in the literature on weather regimes, addressing, for example, the non-probabilistic nature of both methods which can lead to a loss of information on transitional states (Falkena et al., 2023). Figure 1 provides a summary of possible
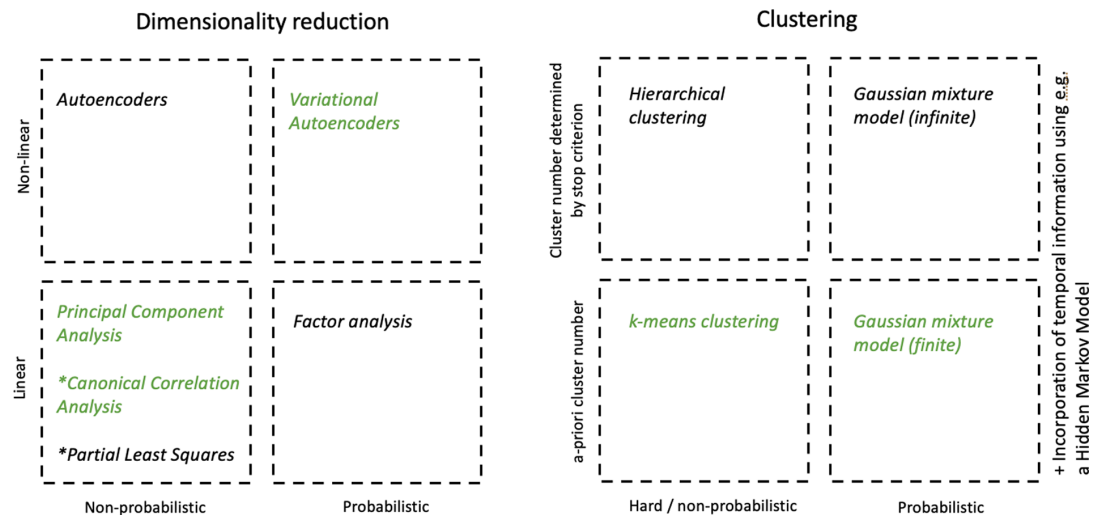

Figure 1. Illustration of selected methodological choices for dimensionality reduction and clustering based on Hannachi et al. (2017) and Murphy (2022). The methods highlighted in green are applied in this paper and described in more detail in Section 3. Methods highlighted with a star refer to joint dimensionality reduction methods between two high-dimensional spaces.

choices of dimensionality reduction and clustering methods for the identification of weather regimes. For a detailed discussion of existing methods for identifying weather regimes, we refer to Hannachi et al. (2017) and Straus et al. (2017).

The relationship between weather regimes and surface-level variables provides a key motivation for their investigation. However, most methods for identifying weather regimes are not specifically designed to capture variations of the impact variable in question, such as extreme precipitation over Morocco in this application. Therefore, available methods do not necessarily resolve the dynamical processes that modulate the relevant surface-level impact. Due to this limitation, studies investigating dynamical precursors of extremes using weather regimes might fail to capture relevant physical processes as well as potential predictors at subseasonal-to-seasonal lead times.

Recognizing the value of atmospheric patterns that are more informative of a local-scale variable, existing methods for identifying targeted atmospheric patterns are based on either pre-filtering data to extreme impact days (Rouges et al., 2023; Dorrington et al., 2024a, 2024b), clustering the impact variable directly (Ullmann et al., 2014; Bloomfield et al., 2020), or increasing the number of clusters to maximize the informativeness of regimes regarding the impact variable (Gadouali et al., 2020; Mastrantonas et al., 2020. However, these approaches compromise either regime persistence and robustness, and thereby their extended-range predictability, or the completeness of the representation of atmospheric dynamics. On the other hand, linear statistical methods for identifying related subspaces of two high-dimensional datasets such as canonical correlation analysis (CCA) have been applied, amongst others, by Vrac and Yiou (2010) in combination with k-means clustering to identify weather regimes targeted to rainfall over France. However, CCA, for example, identifies linear transformations such that the two reduced spaces are maximally correlated (Murphy, 2022), thereby projecting the data into partial subspaces and compromising the ability of the regimes to represent the full atmospheric phase space.

Both the optimal number of clusters (Straus et al., 2017; Dorrington and Strommen, 2020; Falkena et al., 2020), as well as the physical and statistical interpretation of the weather regimes (Stephenson et al., 2004; Hochman et al., 2021) have been subject to discussion, in particular, whether multimodality of the underlying probability density function is assumed. In this paper, weather regimes are here interpreted as statistical representations of the underlying physical processes that should be statistically robust and relevant to the intended use case, without making any stronger assumptions about the multimodality of the underlying probability density function.

1.4. Contribution

To address the research gaps outlined above, this paper presents a novel method for identifying probabilistic weather regimes targeted to a local-scale scalar impact variable based on a modified variational autoencoder architecture. The proposed method, called regression mixture model variational autoencoder (RMM-VAE), combines targeted dimensionality reduction with probabilistic clustering. This is achieved by integrating a regression into the dimensionality reduction step of the variational autoencoder (VAE) and regularizing the reduced space using a Gaussian mixture model. The method thereby aims to capture the dynamical processes that modulate the target variable while maintaining the physical robustness and persistence of the identified regimes.

The regimes identified by the RMM-VAE method are probabilistic as each datapoint is assigned probabilities of belonging to the different clusters, and the clusters themselves are fit as multidimensional Gaussian distributions. One advantage of probabilistic clusters compared to so-called hard clusters identified, for example, by k-means is that information on transitional states can be captured, leading to a more complete picture of reduced atmospheric dynamics (Falkena et al., 2023).

VAEs are a deep generative machine learning method introduced by Kingma and Welling (2013) and described in more detail in Section 3, that have shown promise in identifying non-targeted weather regimes (Baldo and Locatelli, 2022). The advantages of using a VAE architecture for identifying *targeted* weather regimes lie in their ability to generalize the linear dimensionality reduction conducted in PCA to

nonlinear transformations while offering the possibility of fitting an extendable probabilistic model in the dimensionality-reduced space. The approach presented here builds on previous machine learning architectures reported by Zhao et al. (2019a), abbreviated R-VAE (regression-VAE) that target the dimensionality reduction of a VAE without combining it with clustering, and Ye and Bors (2020) who fit a mixture model into the reduced space of a VAE.

The RMM-VAE method is applied to identify weather regimes over the Mediterranean region in extended winter (November to March) targeted to total precipitation over Morocco and is compared to two established linear approaches (PCA + k-means and CCA + k-means), as well as the R-VAE method (Zhao et al., 2019a) combined with k-means clustering (R-VAE + k-means) which is introduced in Section 3.

The performance of these four methods is analyzed in terms of the predictive skill of the resulting regimes with respect to the target variable, as well as their persistence and separability. These evaluation metrics are chosen to assess both the ability of the regimes to capture the relevant dynamical processes modulating the target variable and their physical robustness. To enhance the understanding and interpretability of the novel methods, we further investigate the reduced space—also called the latent space—along with the ability of the different dimensionality-reduction methods to reconstruct the input space from the reduced representation.

The remainder of this paper is structured as follows. Section 2 describes the data used, Section 3 provides a detailed description of the different methods, including the RMM-VAE method, and Section 4 reports details of the implementation and parameter choices. After comparing the results of the different methods for identifying targeted weather regimes in Section 5, and Section 6 discusses and concludes the findings of this paper.

2. Data

Atmospheric circulation patterns are investigated over the Mediterranean region (lat: 25°N–50°N; lon: 20°W–45°E, region shown in Figure 6) in extended winter (Nov–Mar) using ERA5 reanalysis data from 1940 to 2022 (Hersbach et al., 2020) of geopotential height at 500 hPa (z500) re-gridded to a resolution of 2.5° × 2.5°. The data is standardized by subtracting the climatological daily mean and dividing the result by the standard deviation at each grid point. While this breaks the geostrophic relationship between geopotential height and the horizontal wind field, the dimensionality-reduction and clustering methods do not make use of this relationship, and the choice was therefore deemed acceptable for this application.

For the initial application of the new method, ERA5 reanalysis data of daily total precipitation in a region over Morocco (lat: 30°N–36°N; lon: 11°W–0°E, resolution: 0.25° × 0.25°) is used as a target variable over the same period as a proxy for observations. Figure 2 shows the annual mean and 95th-percentile precipitation over the selected region. Precipitation data was normalized by applying a Box-Cox transformation, at each grid cell for CCA, and over the entire region for the two VAE methods (Box
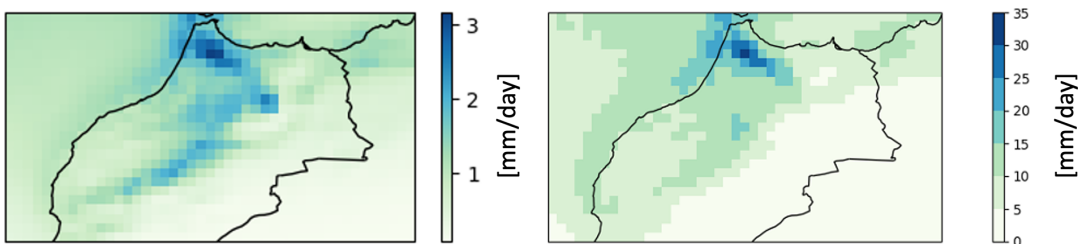

Figure 2. *Extended winter precipitation (November–March) over the selected region over Morocco based on ERA5 reanalysis data from 1940 to 2022. Left-side: daily mean precipitation. Right-side: 95th percentile of daily precipitation.*

and Cox, 1964. A three-day running average of daily total precipitation and a five-day running average of z500 were taken to reflect the duration of extreme precipitation events and associated weather systems during the observational period (Dayan et al., 2015; Loudyi et al., 2022).

3. Methods

This study compares four different methods for identifying weather regimes. All methods combine dimensionality reduction of the geopotential height space x into a latent space z with subsequent clustering. Except for the first method, PCA + k-means, all methods explicitly make use of the target variable t, precipitation over Morocco, in the identification of weather regimes.

3.1. Principal component analysis and k-means clustering (PCA + k-means)

The PCA + k-means method combines linear dimensionality reduction using PCA with subsequent clustering using k-means to identify weather regimes (Michelangeli et al., 1995). PCA provides a linear transformation of the data into a subspace z spanned by the orthogonal eigenvectors of the covariance matrix C_{xx} of the dataset x (Jolliffe and Cadima, 2016). k-means clustering is applied to iteratively partition the reduced space z into k sets with the objective of minimizing the within-cluster squared distance from the cluster center (Murphy, 2022). This method is implemented due to its prevalence in the current literature, including existing studies on Mediterranean weather regimes and precipitation over Morocco (Mastrantonas et al., 2020).

3.2. Canonical correlation analysis and k-means clustering (CCA + k-means)

The CCA + k-means method also combines a linear dimensionality reduction method with k-means clustering. CCA is a dimensionality reduction method that identifies linear transformations of two high-dimensional spaces, x and t, into respective subspaces such that the correlation between the projections of the variables onto their new basis vectors is maximized (Johnson and Wichern, 2013). The method is symmetric, meaning the spaces x and t are treated in the same manner. In contrast to the machine learning methods presented in this paper, CCA takes the full precipitation field as input, rather than the aggregate scalar total precipitation over the selected region. CCA is applied to the input geopotential height space x and target precipitation space t. Subsequently, weather regimes are identified by applying the k-means clustering algorithm to this dimensionality-reduced geopotential height space. This combination of methods is implemented to identify targeted clusters based on an established linear dimensionality reduction method.

3.3. Regression-variational autoencoder (R-VAE + k-means)

The R-VAE + k-means method is a targeted approach to the identification of weather regimes. The method is an extension of a VAE architecture that combines dimensionality reduction with a prediction task introduced by Zhao et al. (2019a) in a neuroscience application.

In a first application to climate data, this approach is amended to target the dimensionality reduction of the geopotential height input space x into the latent space z to the impact variable t, that is, total precipitation over Morocco. Weather regimes are identified by subsequentially clustering the reduced space z using the k-means clustering algorithm described in Section 3.1.

Section 3.3.1 introduces VAEs and variational inference. Any foundational technical statistical and machine learning terminology not explained directly in the text is highlighted in italics and introduced in more detail in a glossary in Appendix 2. Section 3.3.2 then provides a detailed description of the R-VAE method and its application to identify targeted weather regimes.

3.3.1. Variational autoencoders

Autoencoders can be interpreted as a non-linear extension of PCA implemented through an encoder and decoder neural network. In PCA, the encoder would correspond to the linear transformation of the high-dimensional input data into the dimensionality-reduced space of principal components, while the decoder

corresponds to the inverse transformation that reconstructs the input space using a selected number of principal components. Although autoencoders are more efficient at encoding the input data compared to PCA, the identified *latent* space is not necessarily continuous, which is an obstacle to subsequent clustering (Murphy, 2022).

VAEs are a generative machine learning architecture introduced by Kingma and Welling (2013). They extend the encoder-decoder architecture of autoencoders by fitting a probabilistic model of the data into the reduced space using Bayesian variational inference. By estimating the underlying probability distribution of the data in the latent space explicitly, the architecture allows for the generation of new samples from the encoded data, hence the term generative model. VAEs thereby provide a probabilistic and non-linear dimensionality reduction method and alternative to PCA.

The probabilistic graphical model underlying the VAE architecture is shown in Figure 4a. The model aims to identify a continuous latent space z that provides a dimensionality-reduced representation of the high-dimensional input space x. This statistical model with parameters θ can be fit using Bayesian inference based on Bayes theorem. However, this requires computing the *posterior probability* $p_\theta(z|x)$, which is in general computationally intractable. Therefore, the loss function of a VAE is derived using Bayesian variational inference (Murphy, 2023. Variational inference introduces a function q_ϕ from a selected distributional family with parameters ϕ to approximate the intractable posterior by minimizing the Kullback–Leibler (KL) divergence between the true and approximated posterior. Those terms of the KL-divergence that depend on the parameters of the model represent a lower bound to the likelihood of the data and are termed the evidence lower bound $\mathcal{L}(\theta,\phi|x)$. This evidence lower bound can then be minimized to provide the best variational estimate of the model, using the stochastic gradient variational bayes (SGVB) estimator in the case of VAEs. Kingma and Welling (2013) further introduce the reparameterization trick to generate samples from the probabilistic encoder $q_\phi(z|x)$ while still being able to *backpropagate* information through the network.

$$\mathcal{L}(x) = -\mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] + D_{KL}\big(q_\phi(z|x)|p_\theta(z)\big). \tag{1}$$

Equation 1 shows the resulting loss function that is minimized to fit a standard VAE. The first term represents the reconstruction loss of passing data points through the encoder and reconstructing it from its reduced representation. The second term represents the *regularization* loss, which penalizes the divergence of the fitted probability distribution in the latent space z from the *prior probability distribution* $p_\theta(z)$, which is often assumed to be a multivariate Gaussian with mean μ_z and standard deviation σ_z.

3.3.2. The R-VAE method

Zhao et al. (2019a) demonstrate that the standard VAE architecture can be extended to not only dimensionality-reduce the input space x but also predict a scalar target t variable that is subsequently used to *regularize* the latent space z. The method, termed R-VAE in this application, is illustrated in Figures 3 and 4b. Here, the inference model, shown in blue in Figure 3 and dashed lines in Figure 4b, not only estimates a dimensionality reduced space z from the input space x but also predicts the mean and variance of a target variable t. In the generative model, shown in green in Figure 3 and solid lines in Figure 4b, the target variable is then used to predict back into the latent space, thereby providing an additional regularization of the reduced space. The impact of this additional regularization when applying this model to the dimensionality reduction of geopotential height using precipitation over Morocco as a target variable is evaluated in Section 5.4.

The loss function of the model can be derived using the representation as a probabilistic graphical model shown in Figure 4b to provide the following factorization of the joint probability distributions:

$$p_\theta(x,z,t) = p_\theta(x|z)p_\theta(z|t)p_\theta(t), and \ q_\phi(z,t|x) = q_\phi(z|x)q_\phi(t|x). \tag{2}$$

The term $p_\theta(x|z)$ corresponds to the input space x reconstructed from the latent space z through the decoder network, and the term $p_\theta(z|t)$ to the latent space estimated from the predicted target variable t, while the prior distribution $p_\theta(t)$ here simply corresponds to the ground truth data of the target variable.
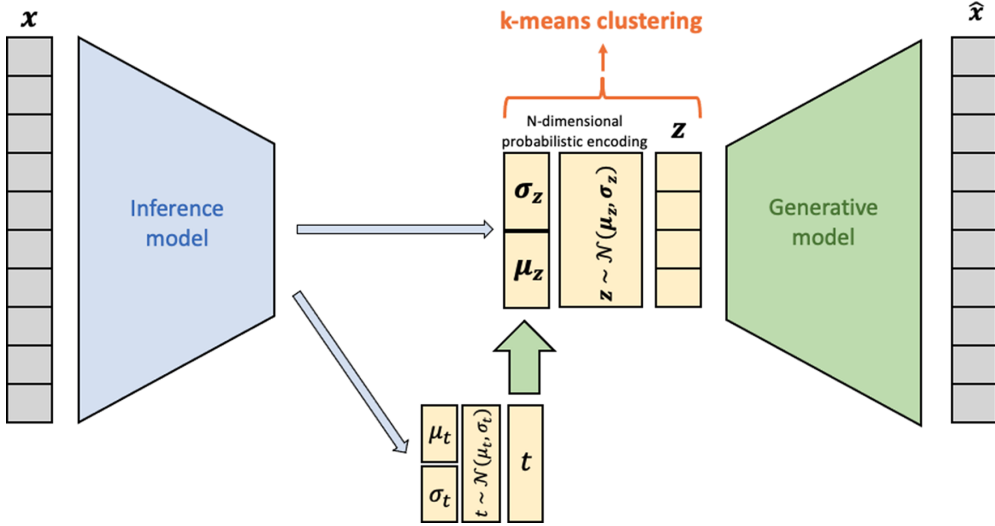
Figure 3. *Schematic diagram of the R-VAE + k-means method based on the architecture developed by Zhao et al. (2019a). The input data x is passed through the encoder network, shown here in blue, which outputs both an estimate of the latent space z and a prediction of the scalar target variable t. The target variable is then used to predict back into the latent space, thereby targeting the dimensionality reduction. The reduced space is subsequently clustered using k-means.*
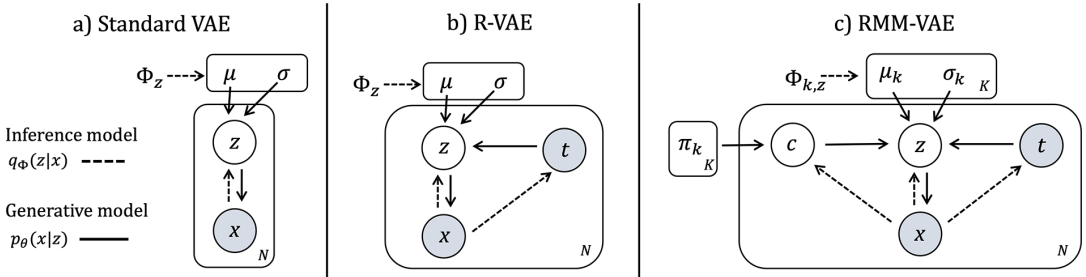

Figure 4. *Different variational autoencoder models represented as probabilistic graphical models using plate notation. The inference model corresponds to the encoder and the generative model to the decoder of the architecture. a) A standard VAE with input variable x, latent variable z and prior Φ_z on the parameters μ and σ of the multivariate Gaussian distribution of z. b) The R-VAE method with an additional target variable t, and c) the RMM-VAE method with probabilistic cluster assignment c regularized by the prior π_k. In all panels, dashed lines indicate the inference model and solid lines the generative model.*

The term $q_\phi(t|x)$ corresponds to the regression of the target variable from the input space x, and the term $q_\phi(z|x)$ to the latent space estimated from the input space using the encoding network.

The loss function of this modified VAE can then be derived as the KL divergence of the two probability distributions. For the full derivation, see Zhao et al. (2019a) and Appendix A.

$$\mathcal{L}(x) = -D_{KL}\big(q_\phi(z,t|x)|p_\theta(z,x,t)\big)$$
$$= \mathbb{E}_{q_\phi(z|x)}\big[\log p_\theta(x|z)\big] - \mathbb{E}_{q_\phi(t|x)}\big[D_{KL}\big(q_\phi(z|x)|p_\theta(z|t)\big)\big] - D_{KL}\big(q_\phi(t|x)|p_\theta(t)\big). \tag{3}$$

As in the standard VAE described in the previous section, the first term represents the reconstruction loss of the dimensionality reduction. The third term represents the regression loss term, penalizing divergence of the predicted target variable $q_\phi(t|x)$ from the ground truth data $p(t)$. The second term uses the predicted target variable to regularize the dimensionality-reduced input space by penalizing the divergence between the two estimates of the latent space z: one that is based on the dimensionality reduction of the original geopotential height data, $q_\phi(z|x)$, and one that is predicted from the precipitation target variable, $p_\theta(z|t)$.

In this application, the two components of the inference model, $q_\phi(z|x)$ and $q_\phi(t|x)$, are parametrized as N-dimensional Gaussian distributions and estimated using non-linear functions with parameters ϕ. Similarly, the probabilistic decoder is parametrized as a Gaussian and modeled as a nonlinear function with parameters θ. Under the assumption that the decoder captures the nonlinearity of the generative model, a linear model for $p_\theta(z|t) \sim \mathcal{N}(at, \mathbb{I})$ is implemented, where a is a vector of unit norm. This constrains the number of parameters the model has to fit overall.

3.4. Regression-mixture model variational autoencoder (RMM-VAE)

In the novel RMM-VAE method, we extend the R-VAE architecture to directly fit a Gaussian mixture model into the reduced space of the VAE, instead of fitting a single multidimensional Gaussian, which is subsequently clustered using k-means as in the R-VAE + k-means approach. The RMM-VAE method thereby integrates probabilistic clustering and targeted dimensionality reduction in a single coherent statistical model. A conceptual advantage of this is the ability of the model to represent the different aims of targeted clustering—identifying physically robust as well as predictive clusters—as terms in a coherently derived loss function, hence allowing for their statistical interpretation and an investigation of their trade-offs.

To derive the RMM-VAE method, the single multidimensional Gaussian chosen to regularize the latent space in the R-VAE method is replaced by k multidimensional Gaussians with mean μ_k and standard deviation σ_k. In addition, the probabilities c_{ik} of the datapoint x_i belonging to cluster k are estimated. This builds on architectures combining VAEs with mixture models presented, for example, by Ye and Bors (2020), Zhao et al. (2019b), and Jiang et al. (2017). Gaussian mixture models themselves are an established probabilistic clustering method (Murphy, 2022) that has been used to identify probabilistic weather regimes (Straus et al., 2017, Baldo and Locatelli, 2022). Figures 4c and 5 illustrate the method.
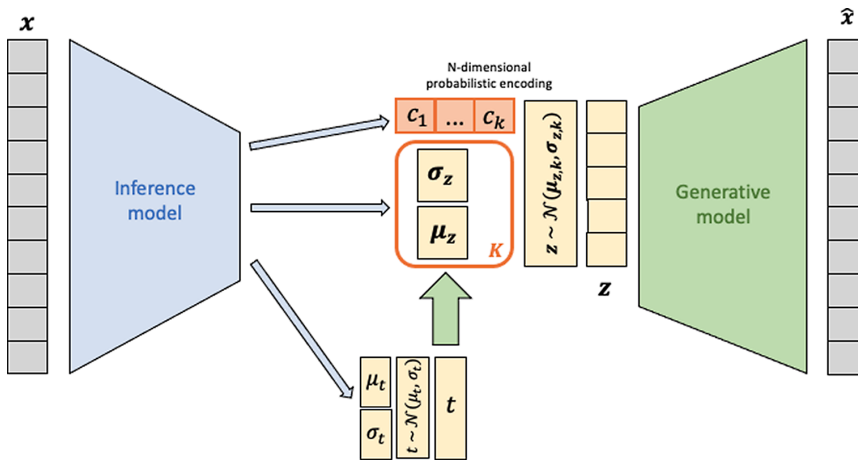

Figure 5. *Schematic diagram of the proposed RMM-VAE approach. In contrast to the R-VAE method, the encoder network, shown again in blue, outputs not only an estimate of the latent space z and a prediction of the scalar target variable t, but also a probabilistic cluster assignment of the data point c_k. The method thereby combines a regression VAE (R-VAE) with probabilistic clustering using mixture models (MM) in a coherent statistical model.*

The inference model, shown again in blue in Figure 5 and dashed lines in Figure 4c, estimates not only the dimensionality reduced space z and target variable z but also the cluster probabilities c_k.

The conditional independence assumptions embedded in the corresponding graphical model shown in Figure 4c are used to re-write the joint probability distribution of the model and derive the loss function of the model. For the full derivation of the loss function, see Appendix A.

$$
\begin{aligned}
\mathcal{L}(x) = & -D_{KL}\big(q_\phi(z,c,t|x)|p_\theta(x,z,t,c)\big) \\
= & \sum_k q_\phi\big(c^k|x\big)\Big[\mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - \mathbb{E}_{q_\phi(t|x)}\big[D_{KL}\big(q_\phi(z|x)|p(z|t)\big)\big] \\
& -D_{KL}\big(q_\phi(t|x)|p(t)\big) - D_{KL}\big(q_\phi(z|x)|p_\theta(z|c^k)\big)\Big] - D_{KL}\big(q_\phi(c^k|x)|p(c^k)\big).
\end{aligned}
\tag{4}
$$

The first three terms in brackets correspond to the terms of the R-VAE loss function for an individual mixture component: the reconstruction loss, the divergence of the estimated target variable $q_\phi(t|x)$ from the ground truth data, and the divergence between the latent spaces generated from the target variable $p(z|t)$ and the latent space encoded from the input data $q_\phi(z|x)$), all weighted by the cluster assignment $q_\phi(c^k|x)$. The fourth term of the loss function minimizes the divergence between the cluster mean k and the latent space estimated from the input data, again weighted by the probability of cluster k occurring, $q_\phi(c^k|x)$. The final term regularizes the cluster assignment $q_\phi(c^k|x)$ to the previous cluster occurrence frequency $p(c^k)$.

The components of the inference model $q_\phi(z|x)$ and $q_\phi(t|x)$, and generative model $p_\theta(x|z)$ and $p_\theta(z|t)$, are parametrized as in the previous method. $p(c^k)$ is a categorical distribution populated by the occurrence frequency of the different clusters updated at each step. This occurrence frequency is used as prior to the probabilistic cluster assignment of an individual day $q_\phi(c^k|x)$. Individual mixture components $p(z|c^k)$ are modelled as Gaussians with mean μ and the identity covariance matrix. The latter choice is made to constrain the number of parameters and avoid model overfitting.

4. Experiments

The encoders and decoders of both VAE methods are implemented using three dense layers of decreasing dimensionality of 128, 64, and 32, respectively. A batch size of 128 and the ReLU activation function are chosen. For 100 epochs, the model is trained on iterative train-test splits using k-fold cross-validation, and subsequently fitted again using a random weights initialization on the entire dataset. For the implementation of the neural network architectures, the Python package keras (Chollet et al., 2015) was used. For a number of the evaluation metrics as well as the implementation of the two linear methods, the Python package scikit-learn (Pedregosa et al., 2011) was employed.

Table 1 provides an overview of the compared methods and relevant hyperparameters. A 10-dimensional latent space was implemented for all the methods, and cluster numbers between 4 and 10 were investigated based on the understanding that the correct number of clusters will depend on the use case and cannot be determined in a general sense in all regions (Straus et al., 2017). The sensitivity of the results to both these choices was investigated and, in the case of the cluster number, the sensitivity of the clusters to sub-sampling of the input data for different choices of k is shown in Appendix B. Based on these results, k = 5 was identified as a reasonable choice of cluster number to visualize in the results section where required.

For both VAE methods, the inclusion of a hyperparameter β based on Higgins et al. (2017) is investigated. The hyperparameter is multiplied with the respective first terms in equations 3 and 4 representing reconstruction loss, thereby changing the weight of the reconstruction objective in the loss function. Two values of β are explored for each of the two methods, whereby v1 ($\beta = 1$) corresponds to the original loss function without the inclusion of an additional hyperparameter, and v2 ($\beta < 1$) decreases the importance of the reconstruction loss term in the loss function. Selected values for β are shown in Table 1.

Table 1. *Overview of methods for the identification of weather regimes and associated parameter choices*

Method	Version	β	Target variable
PCA + k-means	–	–	None
CCA + k-means	–	–	t – full daily precipitation field at 0.25° resolution
R-VAE + k-means	v1	1	t – spatially averaged daily precipitation (scalar)
	v2	0.1	"–"
RMM-VAE	v1	1	"–"
	v2	0.5	"–"

5. Results

Section 5.1 presents the weather regimes identified over the Mediterranean region using the different methods, alongside the conditional changes in the probability of extreme precipitation over Morocco. Next, the predictive skill of the regimes with respect to the target variable (Section 5.2), and their physical robustness (Section 5.3) are evaluated and contrasted. To gain further insight and interpretability of the methods, we investigate the respective dimensionality-reduced spaces in Section 5.4 and the reconstructed input spaces in Section 5.5.

5.1. Cluster centers and characteristics

Figure 6 shows the weather regimes, defined as the centers of the identified clusters, along with the associated change in the conditional probability of extreme precipitation over Morocco, defined as the exceedance of the grid cell-specific 95th percentile.

The weather regimes identified using the PCA + k-means method correspond to those found over the Mediterranean region in other publications using this method such as Giuntoli et al. (2022) and Mastrantonas et al. (2022b). While the number of regimes investigated varies, both publications identify the meridional patterns observed in regimes 3 and 4, a high geopotential height anomaly (regime 5 – termed Mediterranean high in Mastrantonas et al., 2022b), and the western low anomalies seen in regimes 1 and 2, (termed Iberian and Biscay Low in Mastrantonas et al., 2022b). Regime 1, associated with a geopotential height low over the west of Europe, increases the probability of extreme precipitation by a factor of three to four, while the other regimes show no or marginal increases in the probability of extreme precipitation. This is consistent with the results found by Mastrantonas et al. (2020). Overall, the regimes identified by PCA + k-means have roughly similar frequencies of occurrence.

CCA + k-means, on the other hand, identifies multiple regimes associated with an increase in the conditional probability of extreme precipitation by a factor of three or more in different regions of Morocco (regimes 1–3). The spatial patterns of extreme precipitation appear to be modulated by the location of the geopotential height low around Morocco. In contrast, high geopotential height anomalies dominate over the western Mediterranean in the two regimes associated with a lower-than-average probability of extreme precipitation. The regimes associated with extreme precipitation (regimes 1–3) have a slightly lower frequency of occurrence than the other two (regimes 4–5). It can be observed that the anomalies that primarily define the CCA + k-means cluster centers are located in the Western Mediterranean region, which will be further investigated in Section 5.5.

In contrast to the different spatial patterns of extreme precipitation associated with different weather regimes identified using CCA + k-means, the R-VAE + k-means method identifies a single regime associated with a five to six times increase in the probability of extreme precipitation. Furthermore, we find that the cluster center of regime 2, which occurs on almost 30% of days, shows little mean z500 anomaly, meaning it is quite close to a climatologically average day in extended winter. This indicates that while the method is able to identify a regime that is highly informative regarding the occurrence of
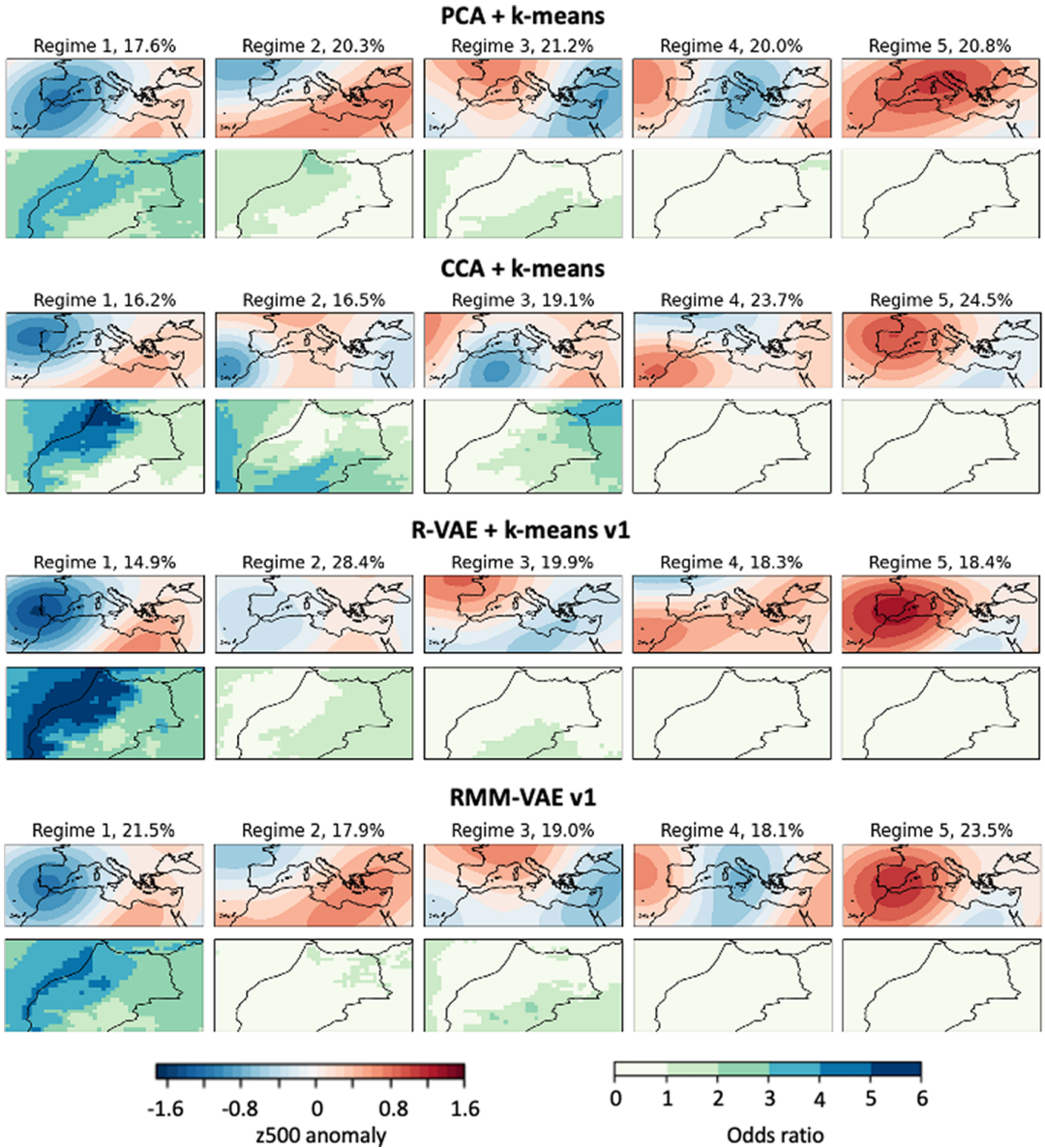
Figure 6. *Identified weather regimes (top rows) and corresponding odds ratios of extreme precipitation (bottom rows) for the four different methods with the number of clusters specified as k = 5. The regime frequencies are given in percent. The odds ratio of extreme precipitation corresponds to the ratio of the probability of the climatological 95th percentile of precipitation at the grid cell conditional on that weather regime, divided by the unconditional probability of 95th percentile of precipitation (i.e., 0.05). The weather regimes are ordered in decreasing order of total precipitation during the days assigned to this cluster by the respective method.*

extreme precipitation, it might not be able to identify structure in the full atmospheric phase space. This will be further investigated in Sections 5.4 and 5.5.

The regimes identified by the RMM-VAE method appear to strike a balance between the baseline method PCA + k-means, and the purely targeted method R-VAE + k-means. On the one hand, the method identifies a single regime associated with a higher probability increase of extreme precipitation compared
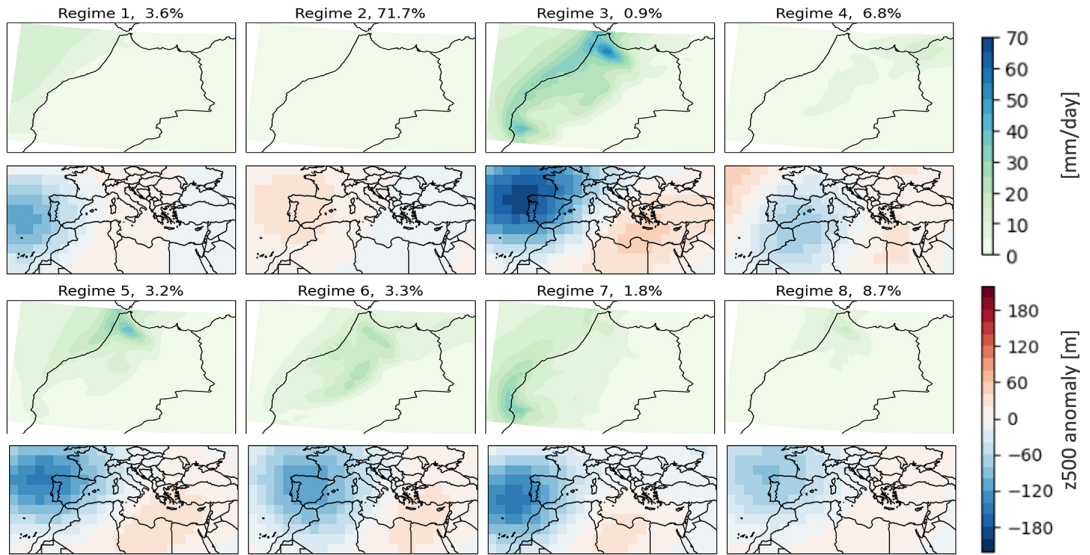
Figure 7. *Precipitation clusters computed on precipitation reanalysis data without pre-processing using the k-means clustering algorithm for k = 8 (top rows) and corresponding z500 anomalies (bottom rows).*

to PCA + k-means, similar to R-VAE + k-means. On the other hand, the resulting cluster centers are visually more similar to the PCA + k-means cluster centers.

To gain further insight into the dynamical precursors of extreme precipitation, we cluster the precipitation field directly using k-means clustering and investigate the corresponding average geopotential height anomalies, shown in Figure 7. As noted in Section 1, clustering the impact variable directly compromises regime persistence and leads to an incomplete representation of the large-scale dynamics, and is therefore shown for illustration purposes only, but not compared as an alternative targeted clustering method.

In agreement with the odds ratios associated with the weather regimes shown in Figure 6, the location and intensity of the precipitation events appear to be modulated by the location and intensity of a geopotential height low off the coast of Spain, consistent with existing literature on dynamical drivers of extreme precipitation in the Western Mediterranean discussed in Section 1.2. Comparing these patterns with the weather regimes shown in Figure 6, we find that the two VAE methods cluster the different patterns of z500 anomalies identified in Figure 7 in one single weather regime, while CCA + k-means disaggregates some of the different spatial patterns of extreme precipitation into different weather regimes. This finding is consistent with the way the different methods incorporate the target variable: while CCA makes use of the entire precipitation field as input data and can hence extract more information about its spatial structure, the VAE methods only receive a single scalar target variable, total precipitation, as input, and are therefore not able to separate different spatial patterns.

5.2. *Evaluating the skill of weather regimes in predicting the target variable*

To evaluate the ability of the weather regimes to capture the dynamical processes responsible for modulating extremes in precipitation over Morocco, we evaluate an empirical prediction of the target variable using the identified weather regime assignment.

The prediction is calculated by multiplying the probability of the weather regimes assigned by the respective method with the conditional probability of the target variable given that weather regime. To assess the ability of the weather regimes to capture both the body and tail of the target variable

distribution, both the skill of the weather regimes in predicting precipitation terciles and the exceedance of the 95th percentile are evaluated. The skill of this prediction is analyzed using the Ranked Probability Score (RPS), a strictly proper scoring rule to measure the accuracy of a probabilistic prediction of mutually exclusive discrete outcomes widely used in forecast evaluation (Gneiting and Raftery, 2007). Skill scores were also used by Schiemann and Frei (2010) to quantify the surface impact of circulation types. The RPS is defined as

$$RPS = \frac{1}{N}\sum_{n=1}^{N}\sum_{j=1}^{m}\left(\delta_{i_n j} - p_j\right)^2 = -\frac{1}{N}\sum_{n=1}^{N}\left(1 - 2p_{i_n} + \sum_{j=1}^{m}p_j^2\right), \tag{5}$$

where m is the number of forecast categories and N is the number of timesteps. $\delta_{i_n j}$ 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. The corresponding skill score (RPSS) is calculated with respect to a reference forecast, chosen here to be the climatology over the entire period, and defined as

$$RPSS = 1 - \frac{RPS_{forecast}}{RPS_{climatology}}. \tag{6}$$

A RPSS of 1 indicates a perfect forecast, while values close to zero indicate little, or no skill compared to the reference forecast.

The resulting RPSS is calculated for precipitation terciles (Figure 8a), and for extreme precipitation (Figure 8b). The predictive skill is overall higher for terciles compared to the 95th percentile threshold, which is to be expected. In both cases, all targeted methods outperform PCA + k-means (blue line), highlighting the potential of improving the predictive skill of standard weather regime definitions. R-VAE + k-means (green and red lines) performs best, followed by RMM-VAE (purple and brown lines) up to k = 8, after which it is slightly outperformed by CCA + k-means.

The better performance of R-VAE can be attributed to RMM-VAE having more objectives to achieve simultaneously as it aims to identify probabilistic clusters while also disentangling the latent space with respect to the target variable. For both VAE methods, increasing the importance of the prediction objective in the loss function (v2) further boosts the predictive skill, although not consistently across cluster numbers.
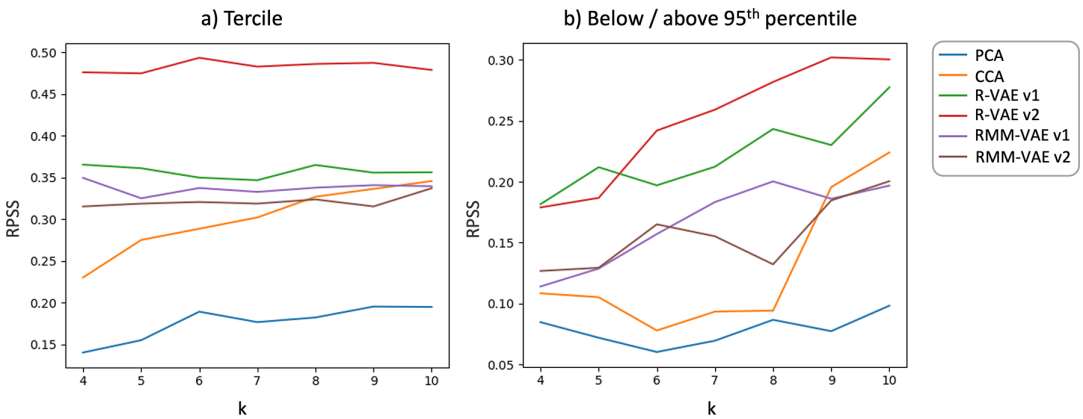

Figure 8. *Ranked Probability Skill Score of an empirical prediction of total precipitation over Morocco using the weather regimes, shown for different numbers of weather regimes k. a) Skill score for the prediction of the tercile of the precipitation distribution, and b) Skill score for extreme precipitation, defined as a binary prediction above or below the 95th percentile. For probabilistic clustering, the skill score is computed using the most likely cluster at the given data point. The higher the RPSS, the more predictive the weather regimes are of precipitation over Morocco.*

Increasing the number of clusters does not improve the skill except for CCA + k-means. For the two VAE methods, this result is likely because the targeted dimensionality reduction already groups data points with similar precipitation amounts in the reduced space, as discussed in section 5.4. Unlike the other methods, CCA takes the full precipitation field as input. The ability to extract spatial information about the precipitation field might enable the predictive skill to increase with cluster number.

5.3. *Evaluating the physical robustness of the weather regimes*

Existing literature on targeted weather regimes finds that patterns that are more informative of a local-scale target variable risk compromising the physical robustness of regimes, which is associated with their persistence and subseasonal predictability. To evaluate the robustness of the targeted regimes, we assess the persistence and the separability of the identified circulation patterns as proxies for their physical robustness.

The separability of the clusters is assessed using the silhouette score (Rousseeuw, 1987). The higher the silhouette score, the better the clusters are separated from one another, while a silhouette score close to zero indicates that the separation between different clusters is not statistically significant. The statistical robustness of the patterns to sub-sampling is also evaluated and shown in Appendix B.

Figure 9a shows the distribution of mean cluster persistence across $k = 5$ clusters. Mean persistence across clusters is highest for PCA, followed by RMM-VAE. However, the spread of the distribution of persistence across clusters is lower for PCA and CCA compared to the two VAE methods, in particular RMM-VAE. This result indicates that while all five PCA + k-means clusters have around the same average persistence, there are some clusters with a longer and some with a shorter average persistence identified by the VAE methods. These results are qualitatively similar for other choices of k (not shown). Sample time series of the cluster assignments in different methods are shown in Appendix C. Similarly, all targeted clusters perform worse in terms of cluster separability (Figure 9b) compared to the baseline method PCA + k-means. RMM-VAE outperforms the other targeted methods, while R-VAE performs the worst overall.

Overall, these results indicate that RMM-VAE identifies more coherent and robust clusters compared to the other targeted methods, in particular, the R-VAE + k-means method. This result indicates that there is a benefit in performing the probabilistic clustering in one coherent statistical model and predicting the
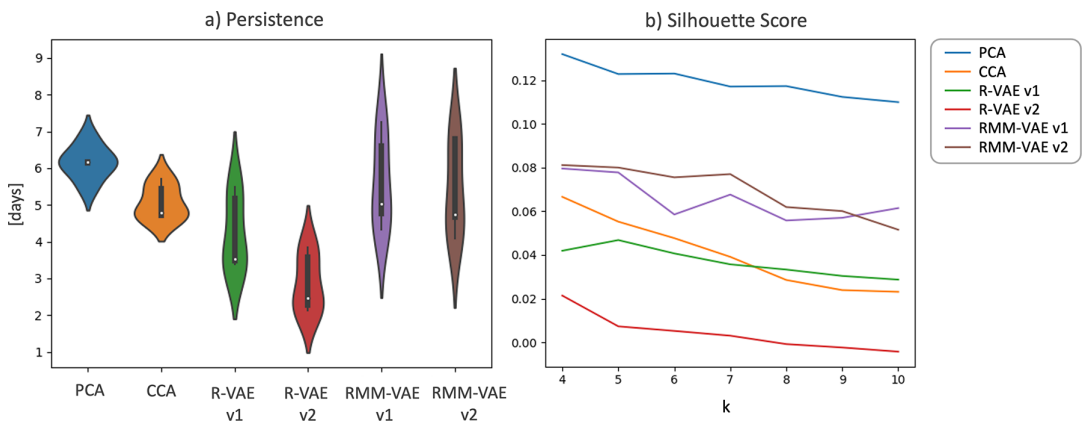

Figure 9. *Regime persistence and separability. a) Distribution of mean persistence across $k = 5$ regimes. Violin plots show the kernel density estimation of the distribution, the distribution median (white point) as well as the interquartile range (black box). The ranking of different methods in terms of persistence remains the same for different choices of k. b) Silhouette score for a range of cluster numbers k. The silhouette score defined as the mean silhouette coefficient (b-a)/max(a,b), where a is the average intra-cluster distance and b is the average inter-cluster distance that is the average distance between all clusters.*

target variable and dimensionality reducing the input space in a single step, as opposed to separating the two steps, as in the R-VAE + k-means method.

5.4. Structure of the dimensionality-reduced spaces

To enhance the interpretability of the methods, we investigate the dimensionality-reduced spaces identified by different methods along with the associated cluster assignment.

All methods identify a 10-dimensional dimensionality-reduced space of geopotential height data of either principal components, canonical variates, or multidimensional Gaussian distributions. To visualize this 10-dimensional representation in two dimensions, we use t-distributed stochastic nearest neighbor embedding (t-SNE) (van der Maaten and Hinton, 2008), a method commonly applied in the machine learning community for the visualization of high-dimension datasets. The resulting visualization, shown in Figure 10, provides an intuition for how the target variable is distributed in the dimensionality-reduced space (top row), and how different clustering methods capture this distribution (bottom row). The t-SNE method preserves nearest neighbors but projects the high-dimensional data onto dimensions that can no longer be interpreted in terms of the physical units of the original input space, therefore the axes in Figure 10 do not have a unit associated with them. Different values of the perplexity parameter, which determines the number of neighbors considered for each point, were tested, and a value of 10 was chosen as it shows representative results.

We find that both targeted VAE methods, R-VAE and RMM-VAE, disentangle the dimension of the geopotential height dataset associated with variations in precipitation over Morocco, the scalar target variable t (dark dots in Figure 10, top row). This aligns with the findings presented by Zhao et al. (2019a) demonstrating a similar type of disentanglement when applying the R-VAE method to the studied neuroscience application. Interpreting the disentanglement in the context of the dynamical drivers of extreme precipitation shown in Figures 6 and 7, the disentangled dimension can be seen to represent the location and depth of the geopotential height anomaly over the Western Mediterranean shown to modulate precipitation over Morocco.

The PCA latent space, on the other hand, shows less organization with respect to the target variable, as expected, since the target variable is not part of the dimensionality reduction. In contrast, the CCA latent
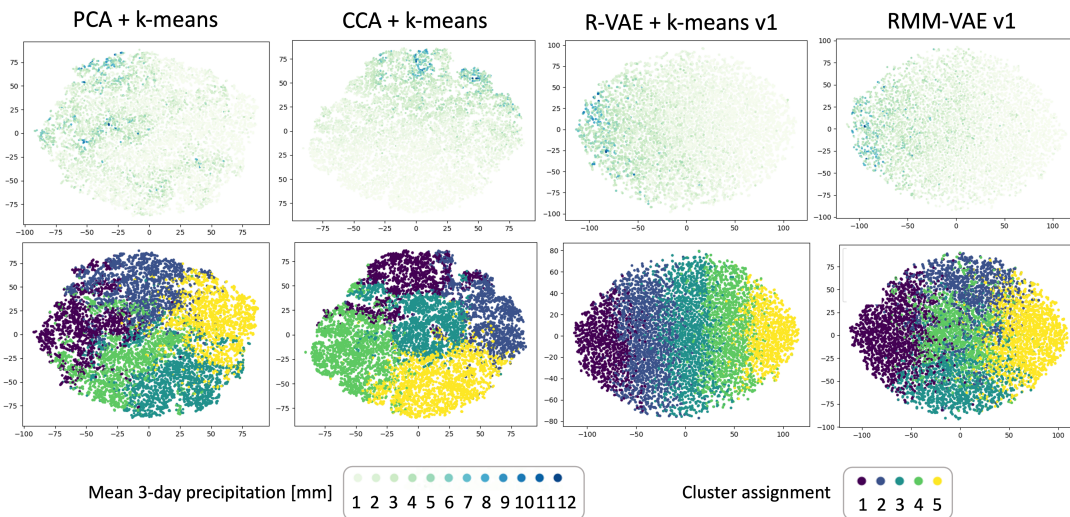

Figure 10. *Visualization of the 10-dimensional latent spaces in two dimensions using t-distributed stochastic nearest-neighbor embedding (t-SNE). Embedded data points are coloured according to the value of the target variable, total mean precipitation (top row), and according to the cluster they are subsequently assigned to (bottom row).*

space appears to have more structure compared to PCA, though not as aligned as the latent spaces identified by the R-VAE and RMM-VAE methods.

Coloring the data points according to their assigned cluster shown in the bottom row of Figure 10, we find that the R-VAE and RMM-VAE methods achieve improved predictive skill regarding the target variables shown in Figure 8 by first grouping data points associated with a similar precipitation impact in the dimensionality reduction step, and subsequently assigning them to one cluster.

The R-VAE method, which carries out the targeted dimensionality reduction and clustering in two separate steps, identifies clusters in 'bands' along the dimension associated with the target variable. The RMM-VAE method, on the other hand, which fits probabilistic clusters while simultaneously disentangling the dimension associated with the target variable, organizes the clusters in a way that appears more consistent with the target of minimizing the distance between points in one cluster, which is also the case for PCA + k-means and CCA + k-means. This difference provides an interpretation and explanation as to why the RMM-VAE method is able to balance informativeness with respect to the target variable with cluster robustness better than R-VAE, although the latter provides a higher predictive skill.

5.5. Ability to reconstruct the input spaces

To evaluate the performance of the dimensionality reduction performed by different methods, we compare the geopotential height data reconstructed from the dimensionality-reduced spaces to the original input data. In the case of PCA, for example, an input data point corresponding to a z500 anomaly pattern is compared to that same data point reconstructed using the first 10 principal components. In the case of the VAE methods, the input data point is compared to that same data point after passing it through the encoder and decoder of the model.

The performance of the dimensionality reduction can then be assessed by computing the mean squared error (MSE) between input and reconstructed data. The lower this value, the more information about the input space is captured by the dimensionality reduction method. Figure 11 shows the distribution of this error across all data points.

Both VAE methods, in particular the respective v1 methods (with $\beta = 1$) have a lower and less widely distributed MSE compared to both PCA and CCA. This means that despite targeting their respective latent spaces to an impact variable, the VAE methods still outperform the two linear methods in terms of representing the atmospheric dynamics in a dimensionality-reduced space.

When increasing the importance of the prediction objectives in the VAE v2 models (with $\beta < 1$), both R-VAE v2 and RMM-VAE v2 still outperform PCA and CCA but perform worse than their respective v1 counterparts. The RMM-VAE v2 ($\beta = 0.5$) method performs slightly better than R-VAE v2 ($\beta = 0.1$), which is consistent with the different values for β chosen to ensure convergence of the model. This result shows that a trade-off between targeting the dimensionality reduction and reconstructing the full phase space exists: while both v2 methods perform better in the task of predicting the target variables, this comes at the cost of a loss of skill in the dimensionality reduction.

Investigating the reconstruction of an individual data point, shown in Figure 12, we find that both v2 methods and CCA focus the dimensionality reduction on the Western Mediterranean region surrounding Morocco. This explains the worse performance of the two v2 methods compared to their respective v1 versions, as well as the worse performance of CCA compared to PCA in Figure 11. The result highlights that some methods for identifying targeted weather regimes such as CCA come at the cost of only representing the dynamics of a partial subspace. The two VAE v1 methods on the other hand, appear to be able to balance this trade-off.

6. Discussion and conclusion

In this paper, we present a novel machine learning method, RMM-VAE, for identifying weather regimes targeted to a scalar impact variable. The method combines the different objectives of targeting weather regimes - predicting the target variable and identifying robust dynamical patterns - in a coherent
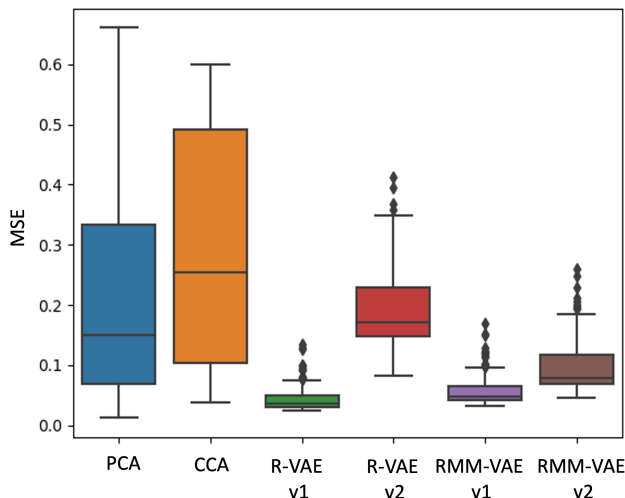
Figure 11. *Distribution of the reconstruction loss, assessed using the mean squared error (MSE) between original input data and reconstructed data for all data points. The thin line in the boxes corresponds to the mean of the distribution, while the boxes extend to quartiles of the dataset. The whiskers extend to points that lie within 1.5 inter-quartile ranges of the opposite quartile, so the lower quartile in the case of the upper bound of the whiskers and the upper quartile in the case of the lower bound of the whiskers. Observations outside this range are displayed as black points.*
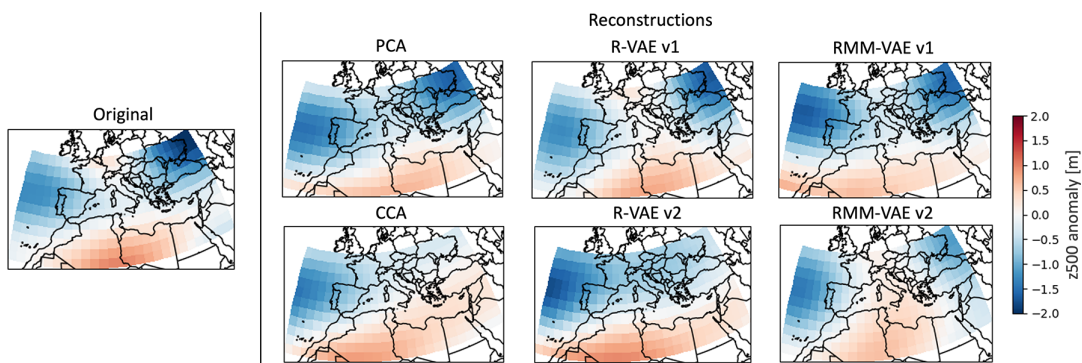
Figure 12. *Gridded and normalized z500 anomalies, as detailed in Section 3, on an example day 1940-01-04, showing the original data on the left and the reconstructions using different methods on the right.*

probabilistic model for the first time. The model integrates non-linear and targeted dimensionality reduction with probabilistic clustering using Gaussian mixture models in a modified VAE architecture described in Section 3.4.

The RMM-VAE method is applied to identify weather regimes over the Mediterranean region targeted to precipitation over Morocco. Results are compared to three alternative approaches: PCA combined with k-means clustering as the currently established standard practice for identifying weather regimes (Bloomfield et al., 2020; Giuntoli et al., 2022), CCA combined with k-means as an established statistical method to relate two high-dimensional input spaces, and R-VAE combined with k-means clustering Zhao et al. (2019a) which is a precursor of the RMM-VAE method.

Overall, we find that the novel RMM-VAE method is able to improve the predictive skill of the identified probabilistic weather regimes with respect to the target variable while maintaining higher regime robustness

and persistence compared to the other targeted methods. The method thereby balances the different objectives of targeted clustering well, and better than the other methods assessed in this study.

Evaluating the identified weather regimes, we find that all the targeted methods resolve the dynamical drivers of extreme precipitation better than the non-targeted baseline PCA + k-means approach (Figure 6). Moreover, we find that the R-VAE method performs best in predicting the target variable from the weather regimes assignment, followed by RMM-VAE up to a certain cluster number. Investigating the persistence and separability of the regimes as proxies for their physical robustness, we find that while all targeted methods perform worse than PCA, the RMM-VAE method performs best among the targeted methods (Figure 9).

By investigating the dimensionality-reduced spaces estimated by the different methods (Figure 10), we find that the two VAE methods disentangle the dimension of the geopotential height field associated with variations in precipitation over Morocco. However, while the R-VAE + k-means method subsequently clusters geopotential height data in bands along this disentangled dimension, integrating the probabilistic clustering with the targeted dimensionality reduction in the RMM-VAE method appears to identify more coherent clusters.

Analyzing the reconstruction of the input space from the reduced representations, we find that both VAE methods outperform the two linear methods (PCA + k-means and CCA + k-means) in terms of reconstruction loss (Figure 11). While a more efficient reconstruction of the input space is expected from a generic VAE architecture due to the possibility of fitting a non-linear encoding function (Murphy, 2023, this result is not obvious here, given that the presented RMM-VAE method has the two additional objectives of disentangling a scalar target variable and fitting probabilistic clusters.

The results highlight two trade-offs in identifying weather regimes targeted to a local-scale impact variable. First, regimes that are more predictive of an impact variable risk compromising their physical robustness are assessed here through regime persistence and separability (compare Figures 8 and 9). Although the R-VAE + k-means method identifies the most targeted regimes, the method performs worst in terms of cluster persistence and separability. This loss of physical robustness can imply that the predictability of the regimes themselves is reduced, which is undesirable for their use in applications such as extended-range forecasting. This trade-off was encountered by Bloomfield et al. (2021), who found that clustering their target variable directly maximized information about the impact (Bloomfield et al., 2020) but came at the cost of significantly reduced regime predictability (Bloomfield et al., 2021). The second trade-off is that more predictive clusters can be achieved by focusing the dimensionality reduction on a subset of the input space, as we see in Figure 12 in the case of CCA, as well as the two VAE methods if the reconstruction loss is down-weighted using the β parameter. However, the resulting regimes do not contain information on the full atmospheric phase space, which implies a loss of information on transition dynamics. This trade-off is implicitly encountered in methods that filter input days to the occurrence of extremes, thereby manually subsetting the input space (Rouges et al., 2023).

The RMM-VAE method performs well in navigating these trade-offs. Among the targeted clustering methods, it identifies the most robust clusters, while performing second-best only to the other VAE method in terms of predictive skill. In contrast, the CCA + k-means method identifies clusters that attain a similarly low separability score as the R-VAE + k-means method, while achieving a significantly lower predictive skill. This result highlights the benefit of using a probabilistic machine learning method to fit a non-linear dimensionality reduction that also allows incorporating a prediction target in a statistically coherent model. Furthermore, the RMM-VAE method makes the trade-offs involved in identifying targeted weather regimes explicit and allows their expression as part of the loss function of the model (equation 4), rather than addressing them implicitly through the choice of cluster number (Gadouali et al., 2020) or pre-filtering (Rouges et al., 2023). In addition, the RMM-VAE method identifies probabilistic weather regimes that can give valuable information on transitional states, to be investigated in future work.

The proposed RMM-VAE method has limitations that should be addressed in future work. For example, we find that by using the full precipitation field as input, CCA is able to identify different

spatial patterns of extreme precipitation associated with different regimes. This stands in contrast to the VAE methods that are trained using total precipitation over Morocco as a scalar target variable, and therefore, primarily disentangle the dynamical drivers of rainfall over regions along the coast that contribute more to total rainfall. To address this, the RMM-VAE method could be further developed to incorporate higher-dimensional target variables.

The RMM-VAE method could also be applied to other regions and target variables, including more realistic and decision-relevant impact variables. In particular, precipitation based on ERA5 reanalysis data, the target variable used here, has known biases, especially in the pre-satellite era (Lavers et al., 2022). While the impact variable explored here is 3-day precipitation over Morocco, any impact variable that has a justifiable link to the large-scale meteorological variables, such as renewable energy supply or the number of people impacted by an extreme event, could be used. Furthermore, the sensitivity of the regimes to pre-processing steps such as the choice of geographical region and the lowpass filter applied to the data could also be further investigated. Future work could also assess the predictability of the regimes themselves, their decadal variability, as well as their relationship to known teleconnections. This would improve the understanding of their usefulness for applications such as sub-seasonal to seasonal forecasting, dynamical adjustment and statistical downscaling of climate models.

The RMM-VAE method presented in this paper contributes a novel probabilistic machine learning method to statistically relate large-scale atmospheric dynamics to regional extremes in local-scale impact variables. The method shows promise in identifying weather regimes that disentangle the dynamical drivers of the target variable while maintaining the physical robustness of the regimes better than other methods, indicating its potential usefulness for a range of climate applications. The results also give further insight into the trade-offs involved in targeting weather regimes to a local impact variable. Overall, this contribution aims to highlight the benefits of motivating and guiding method development in machine learning with a physical research question and understanding of atmospheric dynamics, hopefully contributing to the further development of suitable machine-learning methods in this field.

References

Ailliot P, **Thompson C and Thomson P** (2009) Space–time modelling of precipitation by using a hidden markov model and censored gaussian distributions. *Journal of the Royal Statistical Society Series C: Applied Statistics* **58**(3).

Allen S, **Evans GR**, **Buchanan P and Kwasniok F** (2021) Incorporating the North Atlantic Oscillation into the post-processing of MOGREPS-G wind speed forecasts. *Quarterly Journal of the Royal Meteorological Society* **147**(735), 1403–1418. 10.1002/qj.3983.

Baldo A and Locatelli R (2022) A probabilistic view on modelling weather regimes. *International Journal of Climatology* **1**(21). 10.1002/joc.7942.

Beerli R and Grams CM (2019) Stratospheric modulation of the large-scale circulation in the Atlantic–European region and its implications for surface weather events. *Quarterly Journal of the Royal Meteorological Society* 145(725), 3732–3750. 10.1002/qj.3653.

Bloomfield HC, **Brayshaw DJ and Charlton-Perez AJ** (2020) Characterizing the winter meteorological drivers of the European electricity system using targeted circulation types. *Meteorological Applications* 27(1), e1858. 10.1002/met.1858.

Bloomfield HC, **Brayshaw DJ**, **Paula LMG and Charlton-Perez A** (2021) Pattern-based conditioning enhances sub-seasonal prediction skill of European national energy variables. *Meteorological Applications* 28(4). 10.1002/met.2018.

Box GEP and Cox DR (1964) An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* 26(2), 211–252.

Cassou C (2008) Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation. *Nature* 455 (7212), 523–527. 10.1038/nature07286.

Cattiaux J, **Vautard R**, **Cassou C**, **Yiou P**, **Masson-Delmotte V and Codron F** (2010) Winter 2010 in Europe: A cold extreme in a warming climate. *Geophysical Research Letters* 37(20). 10.1029/2010GL044613.

Chaqdid A, **Tuel A**, **Fatimy AE and Moçayd NE** (2023) Extreme rainfall events in Morocco: Spatial dependence and climate drivers. *Weather and Climate Extremes* 40, 100556. 10.1016/j.wace.2023.100556.

Charlton-Perez AJ, **Ferranti L and Lee RW** (2018) The influence of the stratospheric state on North Atlantic weather regimes. *Quarterly Journal of the Royal Meteorological Society* 144(713), 1140–1151. 10.1002/qj.3280.

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

Coughlan de Perez E, **Harrison L**, **Berse K**, **Easton-Calabria E**, **Marunye J**, **Marake M**, **Murshed SB and Zauisomue E-H** (2022) Adapting to climate change through anticipatory action: The potential use of weather-based early warnings. *Weather and Climate Extremes* 38, 100508. 10.1016/j.wace.2022.100508.

Dayan U, **Nissen K and Ulbrich U** (2015) Review Article: Atmospheric conditions inducing extreme precipitation over the eastern and western Mediterranean. *Natural Hazards and Earth System Sciences* 15(11), 2525–2544. 10.5194/nhess-15-2525-2015.

Delforge D, et al. (2023) EM-DAT: The Emergency Events Database. *Preprint.* https://doi.org/10.21203/rs3.rs-3807553/v1.

Domeisen DIV, **Grams CM and Papritz L** (2020) The role of North Atlantic–European weather regimes in the surface impact of sudden stratospheric warming events. *Weather and Climate Dynamics* 1(2), 373–388. 10.5194/wcd-1-373-2020.

Dorrington J and Strommen KJ (2020) Jet speed variability obscures Euro-atlantic regime structure. *Geophysical Research Letters* 47(15), e2020GL087907. 10.1029/2020GL087907.

Dorrington J, **Grams C**, **Grazzini F**, **Magnusson L and Vitart F** (2024a) Domino: A new framework for the automated identification of weather event precursors, demonstrated for European extreme rainfall. *Quarterly Journal of the Royal Meteorological Society* 150(759), 776–795. 10.1002/qj.4622.

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

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

Dunstone N, **Smith DM**, **Hardiman SC**, **Davies P**, **Ineson S**, **Jain S**, **Kent C**, **Martin G and Scaife AA** (2023) Windows of opportunity for predicting seasonal climate extremes highlighted by the Pakistan floods of 2022. *Nature Communications* 14(1), 6544. 10.1038/s41467-023-42377-1.

Falkena SKJ, **de Wiljes J**, **Weisheimer A and Shepherd TG** (2023) A Bayesian Approach to Atmospheric Circulation Regime Assignment. *Journal of Climate* 36(24), 8619–8636. 10.1175/JCLI-D-22-0419.1.

Falkena SKJ, **de Wiljes J**, **Weisheimer A and Shepherd TG** (2020) Revisiting the identification of wintertime atmospheric circulation regimes in the Euro-Atlantic sector. *Quarterly Journal of the Royal Meteorological Society* 146(731), 2801–2814. 10.1002/qj.3818.

Ferranti L, **Magnusson L**, **Vitart F and Richardson DS** (2018) How far in advance can we predict changes in large-scale flow leading to severe cold conditions over Europe? *Quarterly Journal of the Royal Meteorological Society* 144(715), 1788–1802. 10.1002/qj.3341.

Gadouali F, **Semane N**, **Muñoz Ág and Messouli** (2020) On the Link Between the Madden-Julian Oscillation, Euro-Mediterranean Weather Regimes, and Morocco Winter Rainfall. *Journal of Geophysical Research: Atmospheres* 125(8), e2020JD032387. 10.1029/2020JD032387.

Ghil M and Robertson AW (2002) Waves" vs. "particles" in the atmosphere's phase space: A pathway to long-range forecasting? *Proceedings of the National Academy of Sciences* 99(suppl_1), 2493–2500. 10.1073/pnas.012580899.

Giuntoli I, **Fabiano F and Corti S** (2022) Seasonal predictability of Mediterranean weather regimes in the Copernicus C3S systems. *Climate Dynamics* 58(7), 2131–2147. 10.1007/s00382-021-05681-4.

Gneiting T and Raftery AE (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* 102(477), 359–378. 10.1198/016214506000001437.

Gonzalez PLM, **Howard E**, **Ferrett S**, **Thomas HAF**, **Martínez-Alvarado O**, **Methven J and Woolnough SJ** (2022) Weather patterns in Southeast Asia: Enhancing high-impact weather subseasonal forecast skill. *Quarterly Journal of the Royal Meteorological Society.* 10.1002/qj.4378.

Hannachi A, **Jolliffe IT and Stephenson DB** (2007) Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology 27*(9), 1119–1152. 10.1002/joc.1499.

Hannachi A, **Straus DM**, **Christian LEF**, **Corti S and Woollings T** (2017) Low-frequency nonlinearity and regime behavior in the Northern Hemisphere extratropical atmosphere. *Reviews of Geophysics 55*(1), 199–234. 10.1002/2015RG000509.

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 RJ**, **Hólm E**, **Janisková M**, **Keeley S**, **Laloyaux P**, **Lopez P**, **Lupu C**, **Radnoti G**, **Patricia de Rosnay**, **Iryna Rozum**, **Freja Vamborg**, **Sebastien Villaume, and Jean-Noël Thépaut** (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society 146*(730). 10.1002/qj.3803.

Higgins I, **Matthey L**, **Pal A**, **Burgess C**, **Glorot X**, **Botvinick M**, **Mohamed S and Lerchner A** (2017) beta-VAE: Learning basic visual concepts with a constrained variational framework.

Hochman A, **Messori G**, **Quinting JF**, **Pinto JG and Grams CM** (2021) Do Atlantic-European Weather Regimes Physically Exist? *Geophysical Research Letters 48*(20), e2021GL095574. 10.1029/2021GL095574.

Horton DE, **Johnson NC**, **Singh D**, **Swain DL**, **Rajaratnam B and Diffenbaugh NS** (2015) Contribution of changes in atmospheric circulation patterns to extreme temperature trends. *Nature 522*(7557), 465–469. 10.1038/nature14550.

Jiang Z, **Zheng Y**, **Tan H**, **Tang B and Zhou H** (2017) Variational Deep Embedding: An Unsupervised and Generative Approach to Clustering. http://arxiv.org/abs/1611.05148. arXiv:1611.05148 [cs].

Johnson RA and Wichern DW (2013) *Applied Multivariate Statistical Analysis: Pearson New* International Edition. Pearson Higher Ed.

Jolliffe IT and Cadima J (2016) Principal component analysis: a review and recent developments. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences 374*(2065), 20150202. 10.1098/rsta.2015.0202.

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

Kullback S and Leibler RA (1951) On Information and Sufficiency. *The Annals of Mathematical Statistics 22*(1), 79–86. 10.1214/aoms/1177729694.

Lavers DA, **Simmons A**, **Vamborg F and Rodwell MJ** (2022) An evaluation of ERA5 precipitation for climate monitoring. *Quarterly Journal of the Royal Meteorological Society 148*(748), 3152–3165. 10.1002/qj.4351.

Lee RW, **Woolnough SJ**, **Charlton-Perez AJ and Vitart F** (2019) ENSO Modulation of MJO Teleconnections to the North Atlantic and Europe. *Geophysical Research Letters 46*(22), 13535–13545. 10.1029/2019GL084683.

Lemos MC, **Kirchhoff CJ and Ramprasad V** (2012) Narrowing the climate information usability gap. *Nature Climate Change 2* (11), 789–794. 10.1038/nclimate1614.

Loudyi D, **Hasnaoui MD and Fekri A** (2022) Flood Risk Management Practices in Morocco: Facts and Challenges. In Sumi T, Kantoush SA and Saber M (eds), *Wadi Flash Floods: Challenges and Advanced Approaches for Disaster Risk Reduction, Natural Disaster Science and Mitigation Engineering: DPRI reports*. Singapore: Springer, pp. 35–94.

van der Maaten L and Hinton G (2008) Visualizing Data using t-SNE. *Journal of Machine Learning Research 9*(86), 2579–2605.

Maraun D, **Wetterhall F**, **Ireson AM**, **Chandler RE**, **Kendon EJ**, **Widmann M**, **Brienen S**, **Rust HW**, **Sauter T**, **Themeßl M**, **Venema VKC**, **Chun KP**, **Goodess CM**, **Jones RG**, **Onof C**, **Vrac M and Thiele-Eich I** (2010) Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics 48* (3), RG3003. 10.1029/2009RG000314.

Mastrantonas N, **Herrera-Lormendez P**, **Magnusson L**, **Pappenberger F and Matschullat J** (2020) Extreme precipitation events in the Mediterranean: Spatiotemporal characteristics and connection to large-scale atmospheric flow patterns. *International Journal of Climatology 41*(4), 2710–2728. 10.1002/joc.6985.

Mastrantonas N, **Furnari L**, **Magnusson L**, **Senatore A**, **Mendicino G**, **Pappenberger F and Matschullat J** (2022a) Forecasting extreme precipitation in the central Mediterranean: Changes in predictors' strength with prediction lead time. *Meteorological Applications 29*(6), e2101. 10.1002/met.2101.

Mastrantonas N, **Magnusson L**, **Pappenberger F and Matschullat J** (2022b) What do large-scale patterns teach us about extreme precipitation over the Mediterranean at medium- and extended-range forecasts? *Quarterly Journal of the Royal Meteorological Society 148*(743), 875–890. 10.1002/qj.4236.

Michelangeli P-A, **Vautard R and Legras B** (1995) Weather Regimes: Recurrence and Quasi Stationarity. *Journal of the Atmospheric Sciences 52*(8), 1237–1256. 10.1175/1520-0469(1995)052<1237:WRRAQS>2.0.CO;2.

Murphy K (2022) Probabilistic Machine Learning: An Introduction.

Murphy K (2023) *Probabilistic Machine Learning: Advanced Topics*. MIT Press.

Pasquier JT, **Pfahl S and Grams CM** (2019) Modulation of Atmospheric River Occurrence and Associated Precipitation Extremes in the North Atlantic Region by European Weather Regimes. *Geophysical Research Letters 46*(2), 1014–1023. 10.1029/2018GL081194.

Pedregosa F, **Varoquaux G**, **Michel V**, **Thirion B**, **Grisel O**, **Blondel M**, **Prettenhofer P**, **Weiss R**, **Dubourg V**, **Vanderplas J**, **Passos A**, **Cournapeau D**, **Brucher M**, **Perrot M and Duchesnay E** (2011) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research 12*(0), 2825–2830.

Rouges E, **Ferranti L**, **Kantz H and Pappenberger F** (2023) European heatwaves: Link to large-scale circulation patterns and intraseasonal drivers. *International Journal of Climatology 1*(21). 10.1002/joc.8024.

Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics 20*(0), 53–65. 10.1016/0377-0427(87)90125-7.

Schiemann R and Frei C (2010) How to quantify the resolution of surface climate by circulation types: An example for Alpine precipitation. *Physics and Chemistry of the Earth, Parts A/B/C 35*(9), 403–410. 10.1016/j.pce.2009.09.005.

Shepherd TG, **Boyd E**, **Calel RA**, **Chapman SC**, **Dessai S**, **Dima-West IM**, **Fowler HJ**, **James R**, **Maraun D**, **Martius O**, **Senior CA**, **Sobel AH**, **Stainforth DA**, **Simon FBT**, **Trenberth KE**, van den Hurk BJJM, **Watkins NW**, **Wilby RL and Zenghelis DA** (2018) Storylines: an alternative approach to representing uncertainty in physical aspects of climate change. *Climatic Change 151*(3–4), 555–571. 10.1007/s10584-018-2317-9.

Stephenson DB, **Hannachi A and O'Neill A** (2004) On the existence of multiple climate regimes. *Quarterly Journal of the Royal Meteorological Society 130*(597), 583–605. 10.1256/qj.02.146.

Straus DM (2022) Preferred intra-seasonal circulation patterns of the Indian summer monsoon and active-break cycles. *Climate Dynamics 59*(5), 1415–1434. 10.1007/s00382-021-06047-6.

Straus DM, **Molteni F and Corti S** (2017) Atmospheric Regimes: The Link between Weather and the Large-Scale Circulation. In Christian LEF and O'Kane TJ (eds), *Nonlinear and Stochastic Climate Dynamics*. Cambridge: Cambridge University Press, pp. 105–135. 10.1017/9781316339251.005.

Tanarhte M, **De Vries AJ**, **Zittis G and Chfadi T** (2024) Severe droughts in North Africa: A review of drivers, impacts and management. *Earth-Science Reviews 250*, 104701. 10.1016/j.earscirev.2024.104701.

Terray L (2021) A dynamical adjustment perspective on extreme event attribution. *Weather and Climate Dynamics 2*(4), 971–989. 10.5194/wcd-2-971-2021.

Toreti A, **Xoplaki E**, **Maraun D**, **Kuglitsch FG**, **Wanner H and Luterbacher J** (2010) Characterisation of extreme winter precipitation in Mediterranean coastal sites and associated anomalous atmospheric circulation patterns. *Natural Hazards and Earth System Sciences 10*(5), 1037–1050. 10.5194/nhess-10-1037-2010.

Toreti A, **Giannakaki P and Martius O** (2016) Precipitation extremes in the Mediterranean region and associated upper-level synoptic-scale flow structures. *Climate Dynamics 47*(5), 1925–1941. 10.1007/s00382-015-2942-1.

Ulbrich U, **Lionello P**, **Belusic D**, **Jacobeit J**, **Knippertz P**, **Kuglitsch FG**, **Leckebusch GC**, **Luterbacher J**, **Maugeri M**, **Maheras P**, **Nissen K**, **Pavan V**, **Pinto J**, **Saaroni H**, **Seubert S**, **Toreti A**, **Xoplaki E and Ziv B** (2012) Climate of the Mediterranean: Synoptic Patterns, *Temperature, Precipitation, Winds, and Their Extremes*, 301–346.

Ullmann A, **Fontaine B and Roucou P** (2014) Euro-Atlantic weather regimes and Mediterranean rainfall patterns: present-day variability and expected changes under CMIP5 projections. *International Journal of Climatology 34*(8), 2634–2650. 10.1002/joc.3864.

Vautard R (1990) *Multiple Weather Regimes over the North Atlantic: Analysis of Precursors and Successors. 118*(10).

Vrac M and Yiou P (2010) Weather regimes designed for local precipitation modeling: Application to the Mediterranean basin. *Journal of Geophysical Research 115*(D12), D12103. 10.1029/2009JD012871.

Ye F and Bors AG (2020) Mixtures of Variational Autoencoders. In *2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6, Paris, France, IEEE. 10.1109/IPTA50016.2020.9286619.

Yiou P and Nogaj M (2004) Extreme climatic events and weather regimes over the North Atlantic: When and where? *Geophysical Research Letters 31*(7). 10.1029/2003GL019119.

Zhao Q, **Adeli E**, **Honnorat N**, **Leng T, and Pohl KM.** (2019a) Variational AutoEncoder For Regression: Application to Brain Aging Analysis. *Medical image computing and computer-assisted intervention : MICCAI... International Conference on Medical Image Computing and Computer-Assisted Intervention*, *11765*: 823–831. 10.1007/978-3-030-32245-8_91.

Zhao Q, **Honnorat N**, **Adeli E and Pohl KM** (2019b) Variational Autoencoder with Truncated Mixture of Gaussians for Functional Connectivity Analysis. *Information processing in medical imaging 11492*, 867–879. 10.1007/978-3-030-20351-1_68.

A. Appendix - loss function derivations for R-VAE and RMM-VAE

A.1. R-VAE

$$
\begin{aligned}
\mathcal{L}(x) &= -D_{KL}\big(q_\phi(z,t|x)|p_\theta(z,x,t)\big) \\
&= -\int_{z,t} q_\phi(z|x)q_\phi(t|x)\log\frac{q_\phi(z|x)q_\phi(t|x)}{p_\theta(x|z)p_\theta(z|t)p_\theta(t)} \\
&= \int_z q_\phi(z|x)\log p_\theta(x|z) - \int_{z,t} q_\phi(z|x)q_\phi(t|x)\log\frac{q_\phi(z|x)}{p_\theta(z|t)} - \int_t q_\phi(t|x)\log\frac{q_\phi(t|x)}{p_\theta(t)} \\
&= \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - \mathbb{E}_{q_\phi(t|x)}\big[D_{KL}\big(q_\phi(z|x)|p_\theta(z|t)\big)\big] - D_{KL}\big(q_\phi(t|x)|p_\theta(t)\big).
\end{aligned}
\tag{A.1}
$$

A.2. RMM-VAE

$$
\begin{aligned}
\mathrm{L}(x) =\ & -D_{KL}(q_\phi(z,c,t|x)|p_\theta(x,z,t,c)) \\
=\ & -\sum_k \int_{z,t} q_\phi(z|x)q_\phi(t|x)q_\phi(c^k|x)\log\frac{q_\phi(z|x)q_\phi(t|x)q_\phi(c^k|x)}{p_\theta(x|z)p_\theta(z|t)p_\theta(z|c^k)p(c^k)} \\
=\ & \sum_k q_\phi(c^k|x)[\int_z q_\phi(z|x)\log p_\theta(x|z) - \int_z q_\phi(z|x)\log p_\theta(x|z) \\
& -\int_{z,t} q_\phi(z|x)q_\phi(t|x)\log\frac{q_\phi(z|x)}{p_\theta(z|t)} \\
& -\int_t q_\phi(t|x)\log\frac{q_\phi(t|x)}{p(t)} - \int_z q_\phi(z|x)\frac{q_\phi(z|x)}{p_\theta(z|c^k)} - \log\frac{q_\phi(c^k|x)}{p(c^k)}] \\
=\ & \sum_k q_\phi(c^k|x)[\mathrm{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - \mathrm{E}_{q_\phi(t|x)}[D_{KL}(q_\phi(z|x)|p(z|t))]] \\
& -D_{KL}(q_\phi(t|x)|p(t)) - D_{KL}(q_\phi(z|x)|p_\theta(z|c^k)) - D_{KL}(q_\phi(c^k|x)|p(c^k)).
\end{aligned}
\tag{A.2}
$$

B. Appendix - sensitivity analysis

To analyse the robustness of the results, the sensitivity of the results to two pre-processing steps is analysed:

- Averaging z500 over 3 days vs 5 days: qualitatively similar results, very similar predictive skill, ordering of latent space and reconstruction error. 3-day clusters have a slightly decreased persistence and silhouette score, as expected.
- Selecting a slightly larger geographical region (westward shift and overall larger region tried). This influences all evaluated metrics, however, the ranking of different methods in the different evaluation categories remains the same.

So while these choices do affect the precise values of different evaluation metrics, they do not affect the overall ordinal findings when comparing the methods.

Furthermore, the sensitivity of cluster centers to subsampling the data is analyzed. This is important because it is not desirable for the cluster centers to change drastically with new data. Figure A-1 shows the anomaly correlation coefficient (ACC) between the cluster centers computed from subsamples of the data and the cluster centers computed from the full data. We find that the sensitivity of the cluster centers to subsampling depends on the method and cluster number. Overall, the PCA + k-means and R-VAE + k-means methods produce very robust clusters (ACC > 0.95) to subsampling up to a cluster number of 5. Beyond $k = 5$ the robustness to subsampling degrades consistently for R-VAE + k-means, while it peaks again at $k = 8$ for PCA + k-means. Although the robustness of cluster centres is overall slightly lower for the RMM-VAE method, the most robust cluster numbers are consistent with the PCA + k-means method. The cluster centres identified by the CCA + k-means methods are the least robust to subsampling, with an ACC smaller than 0.7 for all cluster numbers analysed. This analysis justifies illustrating the resulting cluster centres for $k = 5$ in the results section and highlights that the choice of k depends on the method chosen.
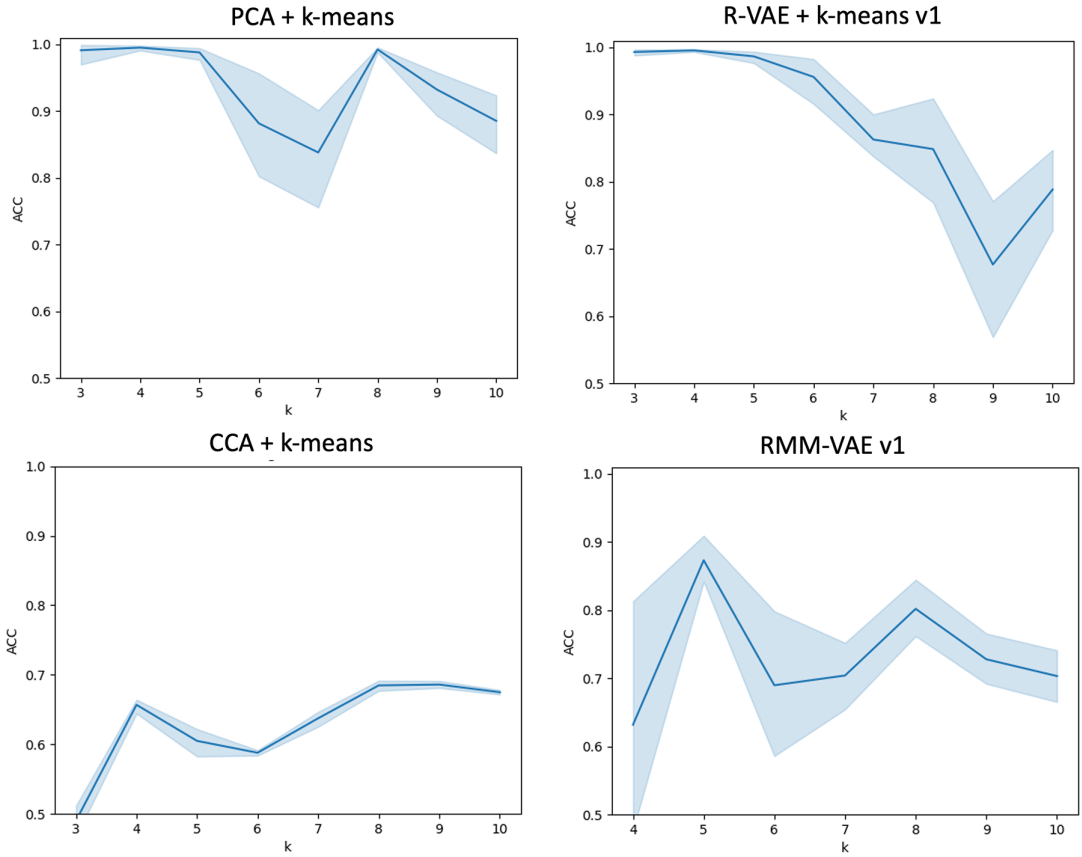
Figure A-1. *50 subsamples containing 80% of data points each are created, and cluster centres are computed. The two sets of cluster centres are then paired by matching centres with the lowest Anomaly Correlation Coefficient (ACC). The lowest of these maximum ACC values is recorded, corresponding to the ACC of the least well-correlated cluster pair.*
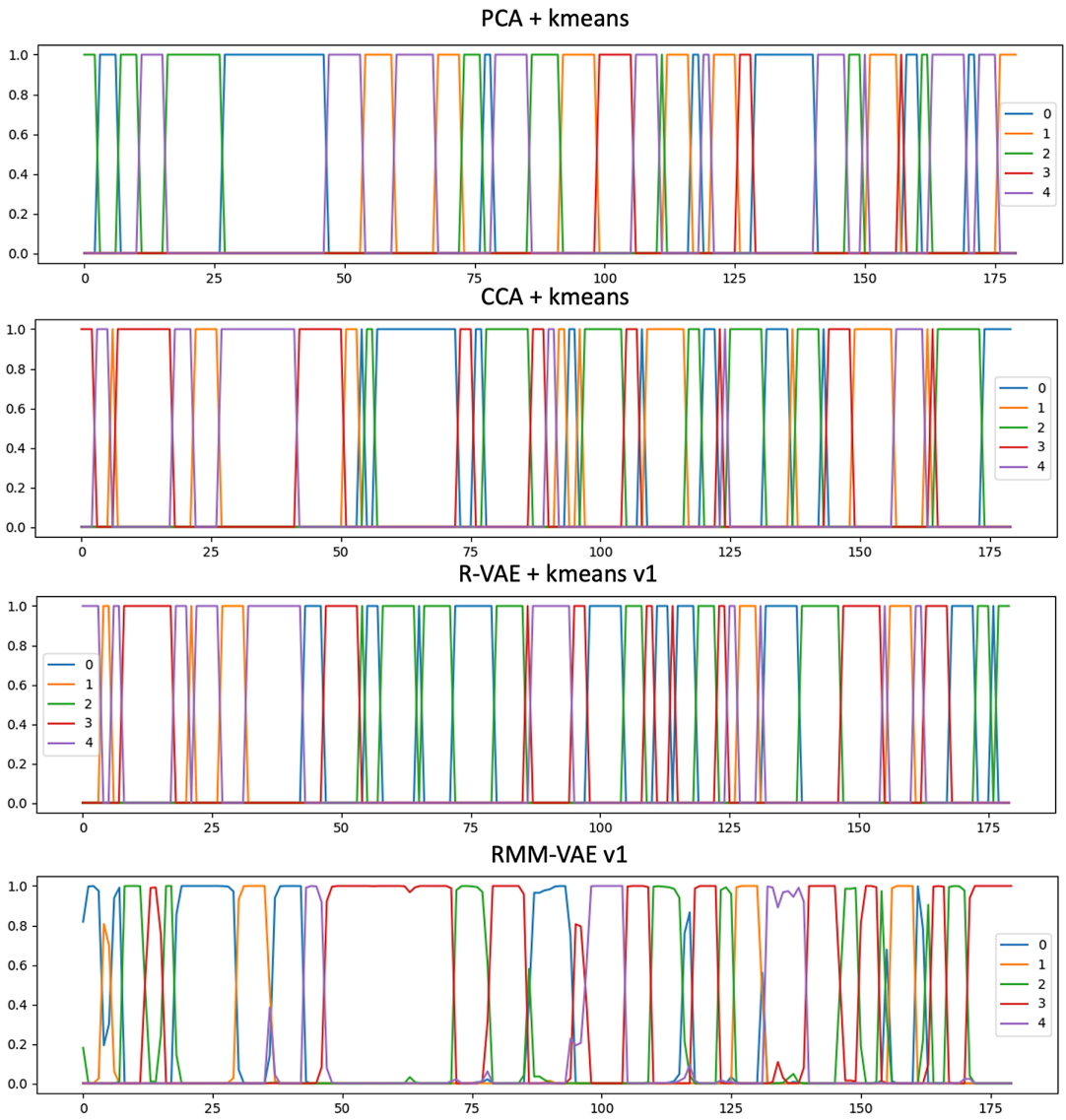
C. Appendix - sample cluster timeseries

Figure A-2. *Sample time series of cluster assignment in different methods.*

D. Statistical glossary

Table A-1. *Glossary of selected statistical and machine learning terminology based on Murphy (2022)*

Backpropagation	Algorithm to calculate the gradient of a loss function and implement gradient descent used to train neural networks. Part of the two-step cycle used to train neural networks: during the forward pass, the output of the neural network is computed given the current weights of the network. During the backward pass or backpropagation, the weights of the hidden layers of the network are adjusted to reduce the loss function.
Divergence	Distance metric between two probability distributions p and q, required to satisfy $D(p,q) \geq 0$ with equality iff $p = q$, symmetry and triangle inequality. The Kullback–Leibler divergence can be interpreted as the information lost by representing p with q, and is defined as the difference between the negative entropy of p and the crossentropy between p and q (Kullback and Leibler, 1951).
Latent space	Also known as embedding space. A d-dimensional vector space that encodes information about a higher-dimensional space in a meaningful representation.
Prior probability	In Bayesian statistics, the prior probability refers to the assumed probability of a variable before selected data or evidence are considered.
Posterior probability	In Bayesian statistics, the posterior probability corresponds to the conditional probability of statistical model parameters given the likelihood of the data and the selected prior.
Regularization	Penalty term added when fitting a statistical model used to prevent overfitting.