JAMES | Journal of Advances in Modeling Earth Systems®

RESEARCH ARTICLE
10.1029/2022MS003130

Special Section:
Machine learning application to Earth system modeling

Key Points:

- A Variational Encoder Decoder (VED) can predict sub-grid-scale thermodynamics from the coarse-scale climate state
- The VED's latent space can distinguish convective regimes, including shallow/deep/no convection
- The VED's latent space reveals the main sources of convective predictability at different latitudes

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

G. Behrens,
Gunnar.Behrens@dlr.de

Citation:

Behrens, G., Beucler, T., Gentine, P., Iglesias-Suarez, F., Pritchard, M., & Eyring, V. (2022). Non-linear dimensionality reduction with a variational encoder decoder to understand convective processes in climate models. *Journal of Advances in Modeling Earth Systems*, *14*, e2022MS003130. https://doi.org/10.1029/2022MS003130

Received 12 APR 2022
Accepted 11 JUL 2022

Non-Linear Dimensionality Reduction With a Variational Encoder Decoder to Understand Convective Processes in Climate Models

Gunnar Behrens[1,2], Tom Beucler[3], Pierre Gentine[2,4], Fernando Iglesias-Suarez[1], Michael Pritchard[5], and Veronika Eyring[1,6]

[1]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany, [2]Department of Earth and Environmental Engineering, Columbia University, New York, NY, USA, [3]Institute of Earth Surface Dynamics, University of Lausanne, Lausanne, Switzerland, [4]Earth Institute and Data Science Institute, Columbia University, New York, NY, USA, [5]Department of Earth System Science, University of California Irvine, Irvine, CA, USA, [6]University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany

Abstract Deep learning can accurately represent sub-grid-scale convective processes in climate models, learning from high resolution simulations. However, deep learning methods usually lack interpretability due to large internal dimensionality, resulting in reduced trustworthiness in these methods. Here, we use Variational Encoder Decoder structures (VED), a non-linear dimensionality reduction technique, to learn and understand convective processes in an aquaplanet superparameterized climate model simulation, where deep convective processes are simulated explicitly. We show that similar to previous deep learning studies based on feed-forward neural nets, the VED is capable of learning and accurately reproducing convective processes. In contrast to past work, we show this can be achieved by compressing the original information into only five latent nodes. As a result, the VED can be used to understand convective processes and delineate modes of convection through the exploration of its latent dimensions. A close investigation of the latent space enables the identification of different convective regimes: (a) stable conditions are clearly distinguished from deep convection with low outgoing longwave radiation and strong precipitation; (b) high optically thin cirrus-like clouds are separated from low optically thick cumulus clouds; and (c) shallow convective processes are associated with large-scale moisture content and surface diabatic heating. Our results demonstrate that VEDs can accurately represent convective processes in climate models, while enabling interpretability and better understanding of sub-grid-scale physical processes, paving the way to increasingly interpretable machine learning parameterizations with promising generative properties.

Plain Language Summary Deep neural nets are hard to interpret due to their hundred thousand or million trainable parameters without further postprocessing. We demonstrate in this paper the usefulness of a network type that is designed to drastically reduce this high dimensional information in a lower-dimensional space to enhance the interpretability of predictions compared to regular deep neural nets. Our approach is, on the one hand, able to reproduce small-scale cloud related processes in the atmosphere learned from a physical model that simulates these processes skillfully. On the other hand, our network allows us to identify key features of different cloud types in the lower-dimensional space. Additionally, the lower-order manifold separates tropical samples from polar ones with a remarkable skill. Overall, our approach has the potential to boost our understanding of various complex processes in Earth System science.

1. Introduction

Earth System Models (ESM) are essential tools to investigate projected changes in precipitation patterns due to different climate scenarios. However, the atmospheric component of traditional ESMs, an atmospheric Global Circulation Model (GCM) with resolution of around 100 km, cannot directly simulate precipitation-generating convective processes, as they occur on scales of a few kilometers, so on much smaller length scales than the grid resolution (Bony et al., 2015; Randall et al., 2003). Therefore, GCMs rely on parameterizations to represent the effect of convective sub-grid-scale processes on the large-scale resolved state (Randall, 2013; Randall et al., 2003). However, the models exhibit large persisting systematic biases such as the presence of a Double Inter Tropical Convergence Zone (ITCZ, a zonal band of strong precipitation in the tropics that forms the ascending branch of the Hadley cell), uncertainties in shortwave cloud radiative forcing, or an over- or underestimation of cloud cover in certain regions as can be seen in recent phases of the Coupled Model Intercomparison Project (CMIP, Bock et al., 2020).

A suitable and well-established alternative is the use of Storm Resolving Models (SRM, grid size ~ 4 km), where a large fraction of convective processes (e.g., deep convection) is directly simulated. SRMs alleviate a number of issues in GCMs, such as the diurnal cycle, the representation of convective aggregation, or the variability and intensity of precipitation (Stevens et al., 2020). Nevertheless, SRM simulations still rely on parameterizations of fine-scale processes, which strongly affect precipitation formation (microphysics) and the onset of convective processes (turbulence). Due to the high computational costs of SRM runs, model iterations are limited to a sequence of months up to a few years, and long-term climate projections will remain unfeasible within the next decade on such fine resolutions.

Recent advances in Machine Learning (ML) have shown great potential in learning sub-grid-scale convective processes and are a promising approach to improve parameterizations in GCMs. Deep feed-forward neural nets showed great accuracy learning convective processes explicitly represented in a superparameterized aquaplanet simulation (Gentine et al., 2018), and successfully replaced the physics package of the GCM leading to stable prognostic ML-based simulations (Rasp et al., 2018). The ML-based model showed substantial improvements simulating both the mean climate and its variability, as represented by the superparameterized model, compared to the host GCM with conventional parameterizations. Similar advances in the prognostic skill, where the ML approach is coupled to the dynamical core of a circulation model, of global precipitation distributions on an aquaplanet were achieved by Yuval and O'Gorman (2020), with random forests and neural networks. Beyond aquaplanets, Mooers et al. (2021) showed that feed-forward neural nets also skilfully reproduce the SuperParameterization (SP) in the presence of real topography based on offline tests, where the ML approach is evaluated against test data, without implementing the resulting representation back into the GCM. Han et al. (2020) likewise demonstrated the potential to learn SP with residual neural nets based on real topography data. X. Wang et al. (2021) showed the possibility to achieve a stable decade-long hybrid ML-Community Atmosphere Model run with real topography based on an ensemble of multiple residual neural nets that separately emulate convective heating, moistening, downwelling solar radiation and radiative fluxes affected by convection, albeit with some distortions of time-mean tropical rainfall bands.

Most of these studies used neural net architectures with several hidden layers and hundreds of thousands or millions of degrees of freedom (weights and biases of the networks), with the exception of Yuval and O'Gorman (2020). Therefore, for all reviewed cases, quantifying the influence of one large-scale climate variable (input) on an emulated sub-grid-scale variable (output) remains challenging without well-suited state-of-the-art interpretability or attribution methods, especially for such high-dimensional regression tasks (Mamalakis et al., 2021). This is due to these machine learning algorithms' large internal variability, and clearly limits the trustworthiness of the reproduced sub-grid-scale variables and the reproduced variability.

In this context, it is natural to wonder whether the use of lower-order models with a smaller latent manifold might prove a promising strategy to overcome the reliance on computationally expensive attribution methods. Our goal is to simplify the interpretation of reproduced convective processes and to provide physical interpretation of the learned relationships, and more generally to reduce the effective dimensionality and build trust in the estimated emulation of convective processes. The purpose of this study is thus to explore whether Variational (Auto) Encoder (VAEs) Decoder structures (Kingma & Welling, 2014) can realistically reproduce convective processes, while enhancing the interpretability of the complex interaction between convection and driving large-scale conditions.

VAEs have only begun to be explored in the atmospheric sciences. Initially Alberdi et al. (2018), showed the potential of VAEs to compress non-linear and chaotic data of the Lorenz 96' model (Lorenz, 1996), which is an ansatz for the turbulent convective nature of the atmosphere. VAEs proved to be powerful tools for the identification of different phases of Northern Hemispheric polar vortex in reanalysis data (Krinitskiy et al., 2019). They demonstrated the applicability of VAEs for common spatio-temporal climate data sets and their advantage compared to more standard linear approaches like Empirical Orthogonal Functions. A further step toward the use of VAEs has been the objective self-supervised classification of convective regimes based on the fine (kilometer-scale) details of explicitly resolved updrafts in global simulations (Mooers et al., 2020). Their VAE identified different tropical convective regimes based on embedded cloud-scale vertical velocity profiles within the embedded subdomains of a global SP simulation. They further showed that their VAE is a powerful approach to detect anomalies like tropical convective extremes and geographically rare forms of dry, continental convection in climate data sets with strong spatio-temporal variability (Mooers et al., 2020).

Here, we use a variational network to investigate the effective dimensionality of the convective parameterization problem, as well as interpret its latent space to delineate convective regimes and large-scale drivers of convective processes. Previous studies without ML approaches explored the interaction of convection and the large-scale climate conditions (Derbyshire et al., 2004) or convective regimes (Frenkel et al., 2012, 2013, 2015; Huaman & Schumacher, 2018) mostly in the tropics and subtropics. While the art and science of interpreting latent spaces is in its infancy, we will demonstrate that one promising method is to leverage the generative modeling capabilities by direct manipulation of the latent manifold. For our variational network, this reveals different convective regimes and how they are connected to driving large-scale conditions (temperature, specific humidity and radiative processes). For instance, we explore whether the geographic region of a GCM sample can be inferred solely based on its latent space position.

The paper is organized as follows. Section 2 describes the climate simulation and machine learning approach used in this study. Section 3 focuses in its first part on the deterministic skill of a variational network for sub-grid-scale SP variables (decoding capabilities) and in the second part on the physical interpretability and meaningfulness of the resulting latent space (encoding capabilities of large-scale climate conditions and convective processes). Section 4 leverages the variational network's latent space to explore the drivers of different convective regimes. Finally, Section 5 provides a discussion and summary of the enhanced interpretability of convective processes via variational networks, as well as an outlook of such generative ML approaches in the context of new hybrid climate models.

2. Data and Methods

2.1. Data: Superparameterized Aquaplanet Simulation

We use a 2-year aquaplanet simulation of the SuperParameterized Community Atmosphere Model v3.0 (SPCAM) (Collins et al., 2006; Khairoutdinov et al., 2005) under the configuration of Pritchard and Bretherton (2014) in which Sea Surface Temperatures (SST) were imposed following a realistic zonally symmetric distribution (Andersen & Kuang, 2012). The SST maximum in the tropics is slightly displaced to 5° N and decreases meridionally toward the poles to reduce exact equatorial symmetry. The solar forcing is fixed to Austral Summer conditions (no seasonal variability), but includes diurnal variability. The model has a coarse horizontal resolution corresponding to a typical grid size of 300 km near the equator. The vertical axis extends from the surface to ∼ 40 km (3.5 hPa) following a hybrid coordinate with 30 levels (22 levels below 100 hPa). The GCM uses a 30-min time step. Following Pritchard et al. (2014), the SP component consists of 8 nested 2D columns oriented meridionally on the same vertical axis and with a sub-grid size of 4 km (Grabowski, 2001; Khairoutdinov & Randall, 2001). Deep convection is explicitly resolved every 20 s and a Smagorinsky 1.5-order turbulence closure, and a one-moment microphysics parameterization (Khairoutdinov & Randall, 2003) are used. SPCAM in this configuration yields a realistic reproduction of the ITCZ and tropical wave-spectra with a pronounced Madden-Julian-Oscillation (MJO)-like signal, as well as improved precipitation distributions compared to the host GCM (CAM, Pritchard et al., 2014). However, this SPCAM setup neglects momentum transport, and for our approach, we sidestep the SP of cloud ice and water sources and sinks and instead emulate their radiative consequences through the total diabatic heating, as in Rasp et al. (2018).

2.2. Model: Variational Encoder Decoder

We develop a VED (see schematic in Figure 1) to holistically learn sub-grid-scale processes in SPCAM. VAEs traditionally reproduce their inputs, for example, learning a mapping from large-scale variables to themselves. Here, our goal is to map large-scale to sub-grid-scale variables. Therefore, we adopt a VED architecture to include the emulation of sub-grid-scale variables. We include convection, turbulence, and radiation by simultaneously predicting the total diabatic heating and moistening tendencies alongside a decoded reconstruction of the relevant input data that summarize local large-scale state information prior to radiative-convective adjustment. Compared to deep feed-forward neural nets, the VED enhances the interpretability of convective processes and

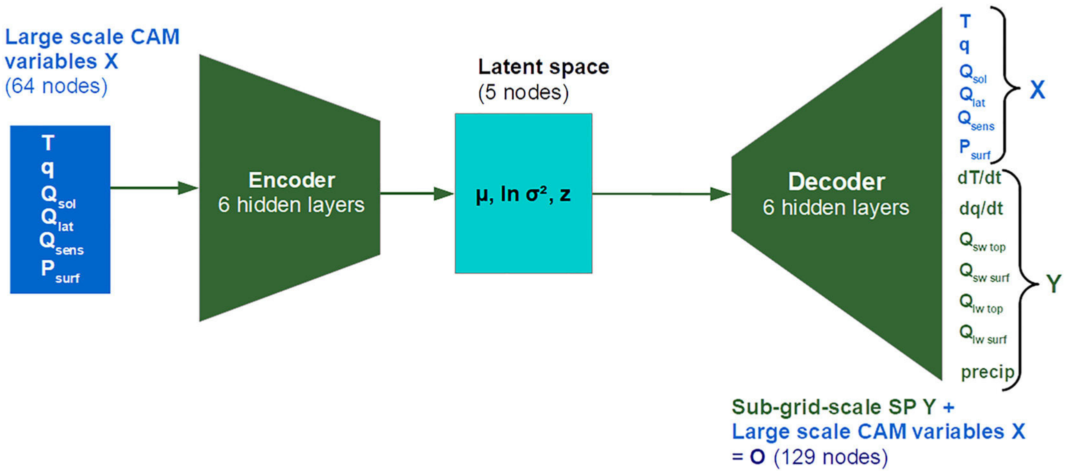

Figure 1. Schematic of the constructed Variational Encoder Decoder (VED) which uses large-scale Community Atmosphere Model (CAM) variables to investigate simulated sub-grid-scale convective processes of SP. The latent space consists of mean $\boldsymbol{\mu}$, a logarithmic variance $\ln\boldsymbol{\sigma}^2$ layer and the latent variables **z**. The output data **O** of the decoder includes a reconstruction of the input data **X** to the encoder to encourage a latent space that can additionally compress the large-scale climate variables, in addition to their mapping to the target sub-grid-scale fields **Y**.

how they are connected to the driving large-scale climate via its latent space of reduced dimensionality. Regarding the input fields (**X**), we closely mirror the established precedent of Rasp et al. (2018) by using profiles of specific humidity **q(p)** in $\frac{kg}{kg}$ and temperature **T(p)** in K on 30 vertical levels each, as extracted from the end of the host model dynamics or the beginning of the physics package. **X** additionally includes the scalar values of solar insolation $\mathbf{Q}_{sol}$ in $\frac{W}{m^2}$, surface latent heat flux $\mathbf{Q}_{lat}$ in $\frac{W}{m^2}$ and surface sensible heat flux $\mathbf{Q}_{sens}$ in $\frac{W}{m^2}$, and surface pressure $\mathbf{P}_{surf}$ in Pa. That is, **X** is a concatenation of these two vectors and four scalars, [**q(p)**, **T(p)**, $\mathbf{Q}_{sol}$, $\mathbf{Q}_{lat}$, $\mathbf{Q}_{sens}$, $\mathbf{P}_{surf}$], into a 64-element input vector. The VED is trained to predict **O**, which combines the reconstruction of the same large-scale input data (as described above) with the sub-grid-scale process rate output fields targeted by Rasp et al. (2018) **Y** (i.e., a parameterization): vertical profiles of total diabatic specific humidity tendency **dq(p)/dt** in $\frac{kg}{kg \times s}$ and total diabatic temperature tendency **dT(p)/dt** in $\frac{K}{s}$ defined on 30 pressure levels, as well as scalar values for shortwave and longwave radiative heat fluxes at the model top ($\mathbf{Q}_{sw\ top}$ and $\mathbf{Q}_{lw\ top}$) and at the surface ($\mathbf{Q}_{sw\ surf}$ and $\mathbf{Q}_{lw\ surf}$) in $\frac{W}{m^2}$, and precipitation rate **precip** in $\frac{m}{s}$. The full predicted vector **O** = [**dq(p)/dt**, **dT(p)/dt**, $\mathbf{Q}_{sw\ top}$, $\mathbf{Q}_{sw\ surf}$, $\mathbf{Q}_{lw\ top}$, $\mathbf{Q}_{lw\ surf}$, **precip**, **q(z)**, **T(z)**, $\mathbf{Q}_{sol}$, $\mathbf{Q}_{lat}$, $\mathbf{Q}_{sens}$, $\mathbf{P}_{surf}$] has a dimension of 129.

As it will be the main ML model used in this study, we henceforth abbreviate the VED structure simultaneously predicting sub-grid-scale convective processes and large-scale climate conditions to "VED" for simplicity. A prior experiment with a $\mathrm{VED}_{X \rightarrow Y}$ that was trained on **X** to predict **Y**, similar to the established precedent of Rasp et al. (2018), does not encode the large-scale climate variables **X** as much in its latent space compared to VED. This limited our ability to gain insight into convective predictability with $\mathrm{VED}_{X \rightarrow Y}$ (see Section S.3A and Figure S16 in Supporting Information S1 for details). In contrast the combined reproduction of sub-grid-scale processes and large-scale climate variables with VED together with our generative modeling method allows us to explore convective regimes and corresponding large-scale climate conditions.

The encoding part of the VED (Encoder) consists of six hidden layers, which progressively reduce the dimensionality from 463 nodes in the first hidden layer down to five nodes (the latent variables) in the latent space. These values were chosen following a formal hyperparameter search (see Section S.1 in Supporting Information S1). We will test the sensitivity of emulations of the VED with respect to the number of latent nodes in Section 3 in detail. In the following we will refer to one distinct latent variable in the context of the network architecture as "latent node." While we will use the notation "latent space" for the manifold spanned by all latent variables. Within this latent space, the mean $\boldsymbol{\mu}$ and logarithmic variance $\ln\boldsymbol{\sigma}^2$ are computed for each node, where $\boldsymbol{\sigma}$ is the standard deviation of the posterior (Kingma & Welling, 2014). Then a so-called "reparameterization trick" (Kingma & Welling, 2014) is utilized to map the original distribution based on $\boldsymbol{\mu}$ and $\ln\boldsymbol{\sigma}^2$ onto an isotropic gaussian distribution. We used the $\ln\boldsymbol{\sigma}^2$ instead of $\boldsymbol{\sigma}^2$ for the construction of the network to simplify the reparameterization and the computation of the VED loss. The resulting latent variables **z** (5 dimensions) are used to investigate convective processes and drivers of convective predictability. Henceforth we will use the notation "latent dimension" to describe the subspace spanned by one particular latent variable. We will show in Section 4 that characteristic convective regimes and large-scale climate states are encoded in **z**. The latent variables **z** are the only input fed to the decoding part of the VED (Decoder), which reconstructs both large-scale and sub-grid-scale fields. In the decoder, the dimensionality is progressively increased to 463 in the last hidden layer before the 129-node output layer. We use the Rectified Linear Unit (ReLU) as activation function of all hidden layers of the Encoder and Decoder except for the Decoder output layer, where we use an Exponential Linear Unit (ELU) based on prior hyperparameter testing (see S.1). In the latent space, $\boldsymbol{\mu}$ and $\ln \boldsymbol{\sigma}^2$ are linearly activated, whereas for the latent variables **z** we call the reparameterization function. In summary, the Encoder and Decoder of the VED consist of 388,440 and 418,469 total trainable parameters, respectively.

We train the VED over 40 epochs (number of iterations through training data), during which the weights and biases are updated to minimize the VED loss function (see Equation 1).

$$\text{VED loss} = \text{reconstruction loss} + \lambda \ \text{KL loss} \tag{1}$$

The loss function is the sum of a reconstruction and a Kullback-Leibler (KL, Equation 3) loss term. The first term measures the Mean Square Error (MSE, Equation 2) between the predicted ($\mathbf{O}^{emul}$) and the ground truth data ($\mathbf{O}$).

$$\text{reconstruction loss} = \frac{1}{M} \times \frac{1}{N} \sum_{i=1}^{(M=129)} \sum_{j=1}^{(N=\text{batch size})} \left(O_{ij} - O_{ij}^{emul}\right)^2 \tag{2}$$

The KL loss term can be interpreted as a regularizer of the resulting latent distributions (Kingma & Welling, 2014), which penalizes the complexity in the latent space based on the KL divergence.

$$\text{KL loss} = \frac{1}{2} \times \frac{1}{N} \sum_{j=1}^{(N=\text{batch size})} \sum_{k=1}^{(K=\text{latent space width})} \left[-1 - \ln \sigma_{jk}^2 + \mu_{jk}^2 + \sigma_{jk}^2\right] \tag{3}$$

$$\lambda \in \mathbb{R}_+ \tag{4}$$

We apply a KL annealing approach that multiplies the KL loss term by an annealing factor $\boldsymbol{\lambda}$ with initial value 0. The annealing factor then grows after a certain epoch during the training process (Alemi et al., 2018). This generally improves the reproduction capabilities of VAEs due to lowering the impact of the regularizing KL term (Mooers et al., 2020), avoiding a posterior collapse (Alemi et al., 2018), which negatively affects training. During a training step a 2D batch (dimensions 714 × 64) of 714 samples, the batch size, is fed into the VED to optimize the weights and biases. We use Adam as the VED's optimizer (Kingma & Ba, 2014). The purpose of an optimizer is to improve the networks performance (minimization of the networks loss function in our case) during the training process based on stochastic gradient descent. We choose this particular optimizer to follow the same strategy like in the preceding study of Rasp et al. (2018). The learning rate (the applied down-gradient step to optimize the loss) has an initial value of 0.00074594 based on a formal hyperparameter tuning and is divided by factor 5 after every seventh epoch over the course of the training. The batch size and the initial learning rate were chosen based on a formal hyperparameter search. Further optimized hyperparameters and a description of the hyperparameter search can be found in Table S1 and Section S.1 in Supporting Information S1. The chosen hyperparameters represent a suitable local minimum for the optimization of the VED architecture but should not be considered as the optimal hyperparameter setting.

2.3. Benchmarking

To benchmark the performance of our VED, we construct three reference networks with different architectures. The first reference network is an Encoder Decoder (ED). The ED closely mirrors the architecture of the VED except that there is no KL regularization, meaning that the calculation of $\ln \boldsymbol{\sigma}^2$ and $\boldsymbol{\mu}$ is omitted. Furthermore, the ED's loss function only relies on the reconstruction loss. The second reference network, LR, is a further simplification of the ED, for which linear activations are used, which can be viewed as an equivalent to a principal component regression except that the latent space is not orthogonal. That is, the LR network can be interpreted as the combination of linear dimensionality reduction and regression modules. We use a reference deep Artificial Neural Net (reference ANN) with its original output normalization based on Rasp et al. (2018), which was proven to be a skilful emulator of SPCAM. Note that to reproduce Rasp et al. (2018), meridional wind profiles were used as input fields to construct and train the reference ANN network. As an additional baseline model, we implement a linear version of our reference ANN. Similar to the reference ANN, this "Reference Linear Model" uses 256 nodes and nine hidden layers but replaces all of the ANN's activation functions with the identity function (i.e., passing the values unchanged). Finally, we constructed one further VED structure and a conditional VAE in the runup of this study, which are presented in Section S.3 in Supporting Information S1 together with their strengths and limitations. Our goal is to strike a balance between the successful emulation of the target sub-grid-scale output data **Y** with compression, and the usefulness of scientific interpretation for convective processes and large-scale climate states. The VED we have chosen (see Figure 1) is optimal on these fronts.

We split the SPCAM simulation into space-time shuffled training, unshuffled validation and unshuffled test data sets spanning 3 months (~ 4,400 time steps) each. The input data **X** is normalized by subtracting the mean of each variable at each vertical level and dividing by the range between minimum and maximum of the resulting anomalies. Furthermore, we normalize the output of the VED, ED and LR as described in Section S.1 in Supporting Information S1. The output normalization, that is scaling to the same order of magnitude, allows us to achieve comparable reproduction skills across all fields. We show the impact of the existing differences of the VED output normalization and the reference ANN output normalization (Rasp et al., 2018) on the evaluation of mean reproduction skills of the networks in Section S.2 in Supporting Information S1.

In the next section we will evaluate the performance of the VED with respect to common reproduction metrics, and discuss the interpretability of the information encapsulated in the latent space.

3. Evaluation of the VED

In this section, we assess the predictive skill of the VED, and compare its mean regimes/statistics and tropical variability against reference networks. Furthermore, we evaluate the interpretability of the VED's latent space with respect to climate and convective variables. With this analysis, we are investigating the overall decoding (reproduction) and encoding (dimensionality reduction, interpretability) abilities of the VED to learn convective processes.

3.1. Mean Regimes and Statistics

We start by evaluating the accuracy of the VED predictions to assess the impact of its dimensionality reduction on the overall performance. We use the Mean Squared Error (MSE) to assess the performance of the VED predictions across sub-grid-scale fields **Y** for the training, validation, and test sets based on our VED output normalization. Overall, the VED shows good reproduction skills (see Table S4 in Supporting Information S1). The VED (test MSE = 0.165) clearly outperforms the linear model LR (test MSE = 0.243) in all data sets. The difference in predictive skills between VED and ED (test MSE = 0.165) is negligible. However, both networks express increased but comparable MSE with respect to reference ANN (test MSE = 0.135), in spite of the reference ANN having a substantially larger dimensionality (no latent manifold with a dramatic dimensionality reduction down to five nodes). These results are robust to the choice of output normalization (VED's vs. reference ANN's, Rasp et al., 2018), as demonstrated in the Section S.2 in Supporting Information S1.

In the following, we explore whether a latent space of 5 nodes is a good compromise between accuracy to reproduce convective processes and physical interpretability in the latent space. Figure 2 shows the VED performance (MSE) on test, validation, and training data as a function of the latent space width. We find a substantial sensitivity of the VED's performance to the latent space width - smaller width results in reduced accuracy associated with increased dimensionality reduction. Even for a latent space of two nodes, the VED has a higher predictive skill than the Reference Linear Model, confirming the necessity of using nonlinear models to faithfully represent sub-grid-scale processes. Moreover, the VED's performance is converging toward the reference ANN for larger latent space widths (8 nodes). A latent space of 5 nodes results in a small reduction of predictive skills compared to the "wider" latent space (Figure 2), indicated by a MSE decrease of only ≈ 0.012 between a latent space of 5 nodes and 8 nodes. Additionally, we will show later (in Section 4) that such a latent space width enables the characterization of realistic convective regimes and drivers of convective processes on specific nodes. This suggests that the overlap between different nodes is small. Despite this small overlap, we will show in Section 4 that the resulting five latent nodes govern both SP convective processes and CAM climate states in most cases. For larger

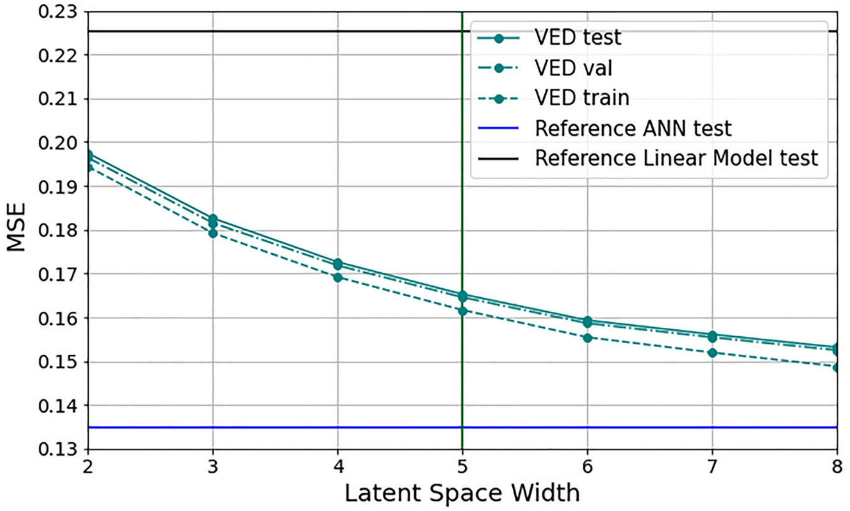

Figure 2. Mean squared error as a function of Latent Space Width of the Variational Encoder Decoder (VED) for test (solid cyan), validation (dashed-dotted cyan) and training data set (dashed cyan curve) using our VED output normalization. The horizontal solid blue/black line represents the mean square error scores of the reference Artificial Neural Net (reference ANN) of Rasp et al. (2018) / a linear version of this network (Reference Linear Model) on test data with fixed layer width of 256 nodes in the nine hidden layers.

latent space widths of 6 nodes and more, the interpretability of resulting convective regimes gets more challenging due to the decaying impact of one latent node, or increasingly concurring influences between the nodes on SP convective processes or CAM climate variables. To summarize, regardless of how the output data are normalized (see Figure S1 in Supporting Information S1), the VED performs better than the Reference Linear Model and approaches the performance of the fully connected reference ANN as the latent space width increases.

As a complementary metric to evaluate the performance of the VED, we use the coefficient of determination $\mathbf{R^2}$ (Equation 5).

$$\mathbf{R^2} = 1 - \frac{\mathbf{MSE}}{\mathbf{Var}} \tag{5}$$

$$\mathbf{MSE} = \frac{1}{\mathbf{P}} \sum_{t=1}^{\mathbf{P}} \left(\mathbf{Y_t} - \mathbf{Y_t^{\mathit{emul}}}\right)^2 \tag{6}$$

$$\mathbf{Var} = \frac{1}{\mathbf{P}} \sum_{t=1}^{\mathbf{P}} \left(\mathbf{Y_t} - \frac{1}{\mathbf{P}} \sum_{t=1}^{\mathbf{P}} \mathbf{Y_t}\right)^2 \tag{7}$$

It is defined as the difference of 1 and the ratio between the Mean Squared Error (**MSE**) and the true Variance (**Var**) of the data, where **P** is the length of the time series, **t** is the respective time step and **Y** / $\mathbf{Y}^{\mathit{emul}}$ are the true value of the test data / VED prediction. We constructed at first the time series of all output variables **O** from the test data set or predictions and computed the respective coefficients of determination in each grid cell (64 points in latitude × 128 points in longitude = 8,192) of all layers. We selected the global sub-grid heating and moistening fields at 700 hPa for the evaluation of the VED's $\mathbf{R^2}$ (Figure 3).

We choose **dq/dt** and **dT/dt** fields at this pressure level because of the limited skill in fitting lower tropospheric convective processes with neural nets that has been reported across multiple investigations, and which has been speculated to be associated with an underrepresentation of stochastic variability linked to shallow and deep convection (Gentine et al., 2018; Rasp et al., 2018; Mooers et al., 2021; X. Wang et al., 2021). Both networks, VED and reference ANN, exhibit similar emulation skill patterns for heating and moistening tendencies, including the skill deficits for low-level moistening tendencies in the tropics, as seen in previous studies. Overall, we see a decreased reproduced variability with the VED ($\mathbf{R}^2_{\text{global mean}}$ = 0.57 / 0.42 for **dT/dt / dq/dt**; 35% and 22% of horizontal grid cells for temperature and specific humidity tendencies with $\mathbf{R}^2 > 0.7$, respectively) compared to the

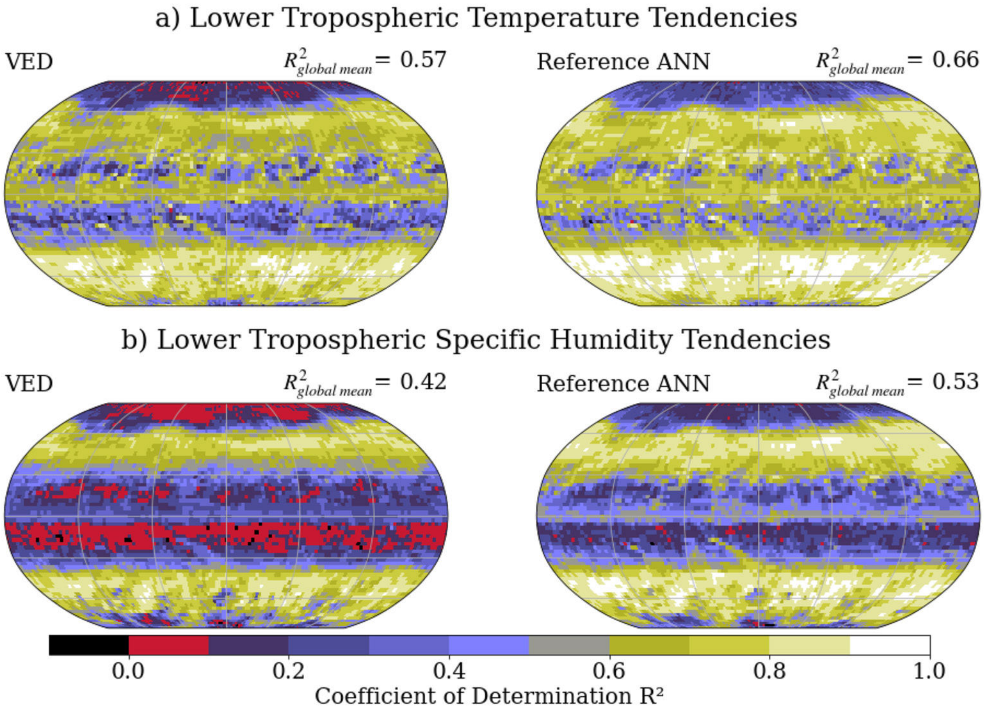
a) Lower Tropospheric Temperature Tendencies
VED
$R^2_{global\ mean} = 0.57$
Reference ANN
$R^2_{global\ mean} = 0.66$
b) Lower Tropospheric Specific Humidity Tendencies
VED
$R^2_{global\ mean} = 0.42$
Reference ANN
$R^2_{global\ mean} = 0.53$
0.0
0.2
0.4
0.6
0.8
1.0
Coefficient of Determination R²

Figure 3. Coefficient of determination $\mathbf{R^2}$ of lower tropospheric temperature tendencies (a) and lower tropospheric specific humidity tendencies (b) at 700 hPa for the Variational Encoder Decoder (left) and reference Artificial Neural Net (right column). The global mean $\mathbf{R^2}$ of each field is indicated in the upper right above every subplot.

reference ANN ($\mathbf{R}^2_{\text{global mean}}$ = 0.66, 0.53 for **dT/dt**, **dq/dt**; 51% and 35% of horizontal grid cells with $\mathbf{R^2} > 0.7$ for temperature and specific humidity tendencies, respectively). The VED shows regions of high reproduction skill for both, temperature and specific humidity tendencies along the mid-latitude storm tracks (~ 45° N/S, $\mathbf{R^2} \sim 0.7$) and in the ITCZ region near the equator (ascending branch of Hadley Cell associated with deep convection, $\mathbf{R^2} \sim 0.6$). Both networks exhibit weaker prediction skill of specific humidity and temperature tendencies near the descending branches of the Hadley Cell (subtropical highs ~ 20° N/S) associated with an underestimation of (shallow) convective variability. Mooers et al. (2021) also found comparably weaker reproduction skill of their neural net in this region. Recently P. Wang et al. (2022) showed that the reproduction of moistening tendencies in the subtropics can be improved by using non-local features from adjacent grid cells as additional inputs of the neural net. Nevertheless, the VED shows good reproduction skill associated with convective processes in the lower troposphere compared to the reference ANN, despite its strongly reduced dimensionality in the latent space. This suggests that the information from large-scale climate variables **X** that is relevant for the prediction of sub-grid-scale convective processes **Y** is closer to five (our latent space's dimensionality) than 64 (the input vector length). In other words, this means that the number of large-scale variables needed to skillfully emulate sub-grid-scale processes is far smaller than the number of original input variables of the superparameterization. This is consistent with assumptions made by reduced-complexity models, such as the lower-dimensional multi-cloud model (Frenkel et al., 2012) or the quasi-equilibrium tropical circulation model (Neelin & Zeng, 2000).

3.2. Tropical Variability

Current ESMs exhibit large biases in tropical precipitation and associated patterns (Bock et al., 2020). These regional uncertainties can be attributed to the fact that many ESMs struggle to reproduce tropical intra-seasonal variability like the Madden Julian Oscillation (MJO, an eastward propagating pattern of clustered deep convection in the Indo-Pacific Region; Zhang, 2005). SPCAM yields a more realistic reproduction of the MJO compared to the traditional convective parametrization of CAM (Khairoutdinov et al., 2005). Furthermore, the governing tropical variability is largely reproducible with deep learning approaches (Rasp et al., 2018). Here, we investigate the ability of the VED to not distort the high-frequency tropical variability (15° N to 15° S) as simulated by

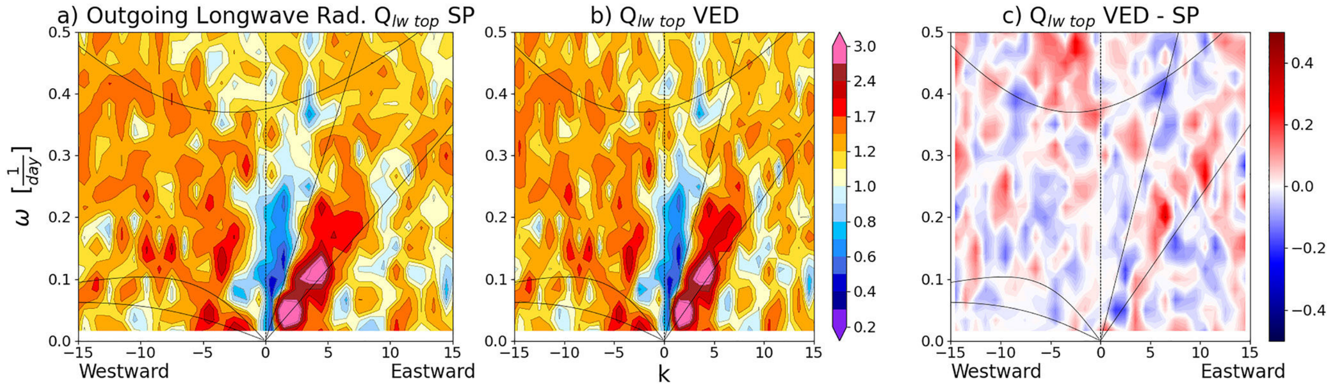

Figure 4. Wheeler Kiladis diagram based on tropical outgoing longwave radiation [15° N - 15° S] of superparameterization (SP) (a), of Variational Encoder Decoder (VED) predictions (b) and the absolute difference of spatio-temporal wave spectra VED-SP (c) for 1 year of SP simulations.

SPCAM compared to the reference ANN. For this analysis, we use the entire second year of the SPCAM simulation to identify driving tropical variability with frequency lower than $\frac{1}{30}$ days^{-1}. This second SP year includes the 3-month sequence of the validation data set but has no overlap with the training data set.

Figure 4 shows the Wheeler-Kiladis diagrams, diagnosing the equatorial symmetric component (zonal wave numbers **k**) of outgoing longwave radiation ($\mathbf{Q}_{lw\ top}$) with respect to its frequency $\boldsymbol{\omega}$ for both SPCAM (Figure 4a) and VED (Figure 4b). Eastward propagating, non-dispersive Kelvin waves ($\boldsymbol{\omega}^{-1}$ ∼ 8–20 days, **k** ∼ 2–5) and the MJO ($\boldsymbol{\omega}^{-1}$ ∼ 30 days, **k** = 1) are not distorted by the VED. The resulting differences in the reproduced spatio-temporal variability with respect to SPCAM are generally confined within −0.2 to 0.2 (unit-less values) (Figure 4c), which amounts to a relative error of roughly 20%, and are not associated with a damping or absence of general features in $\boldsymbol{\omega}$-**k** space.

Although the reference ANN shows slightly better reproduction skill (see Figure S3 in Supporting Information S1), the VED and also ED (see Figure S2 in Supporting Information S1) can realistically reproduce not only mean regimes and characteristics of convective processes but also the associated variability even with its strongly reduced dimensionality on only 5 latent nodes.

Next, we evaluate our main interest –the physical interpretability of the VED with respect to convective processes– by exploring the information encapsulated in its latent space. We will show in the following sections that the representation of general convective processes is actually much lower dimensional than potentially envisioned.

3.3. Interpretability via Latent Space Exploration

In this section, we investigate convective processes and large-scale climate states captured in the latent space of the VED. This will give us a first impression of general drivers of convective predictability encapsulated in the latent manifold and will show the potential to study convective processes with only five latent nodes. Latent spaces of VAEs behave to some extent as a non-linear equivalent of a Principal Component Analysis, PCA (e.g., Rolinek et al., 2019), due to a skilful lower-dimensional encoding of information fed into the network. Therefore, we test whether the latent space of the VED retains a meaningful lower dimensional representation of convective processes like we would expect from a traditional PCA.

Human visualization of the full five latent dimensions (5 nodes, 5D) in a 2D schematic requires some additional dimensionality reduction. For visualization purposes, we therefore use a PCA to first compress the 5D manifold into a 2D lower-dimensional embedded space, which allows a visual inspection of the encapsulated information. The resulting 2D PCA representation contains 82% of the total variance of the VED's latent space. Figure 5 shows the first (*x*-axis) and second leading PC (*y*-axis) of the compressed latent space for 1 million randomly sampled points. The manifold, which is spanned by the two leading PCs, is then divided into a regular grid of size 50 (PC 1) × 50 (PC 2) cells. Tracking each selected sample allows us to characterize the embedded information for both convection and large-scale climate states. This permits us to compute conditional averages of these convection related variables in each grid cell of the 2D PCA compressed manifold.

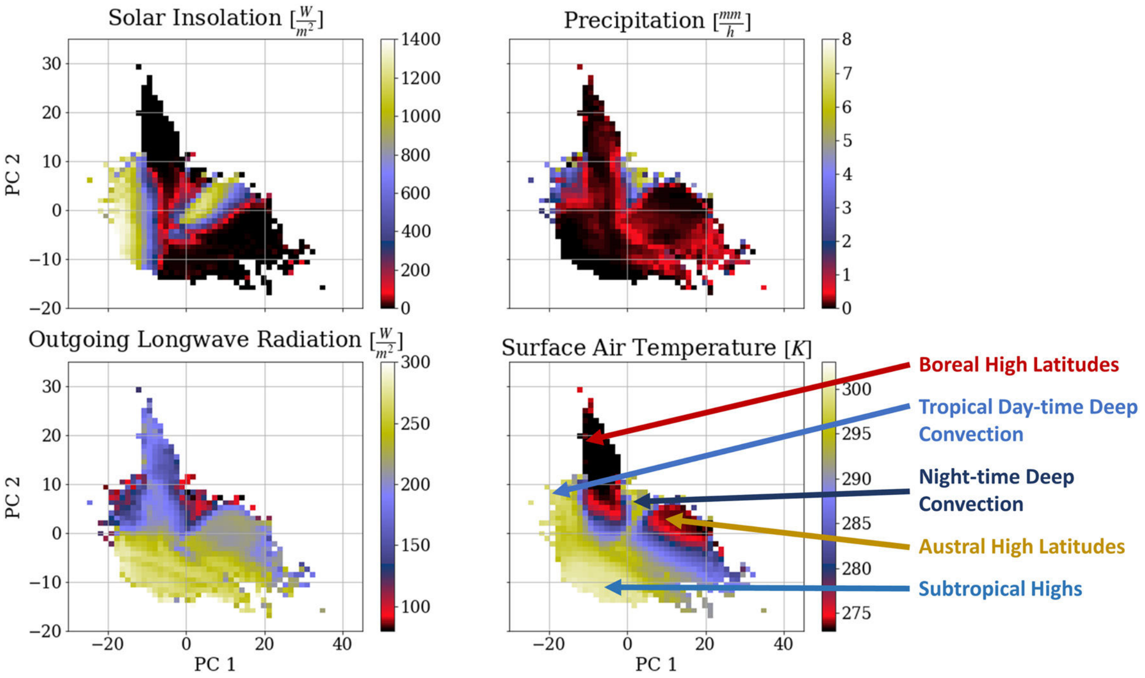

Figure 5. The 2D Principal Component Analysis (PCA)-compressed latent space of the Variational Encoder Decoder (VED) and associated conditional averages of solar insolation (upper left), precipitation (upper right), outgoing longwave radiation (lower left) and surface air temperature (lower right panel) of projected SP test data (see color scheme in each subplot). The *x*-axis / *y*-axis in all subplots indicates the first / second leading PC of the 5D latent space, which have a combined "explained variance" of around 0.82. The arrows in the lower right subplot indicate the position of characteristic samples from different geographic regions inside the 2D PCA-compressed latent space of the VED mentioned in the text.

Figure 5 depicts the conditional averages of solar insolation ($\mathbf{Q}_{sol}$), precipitation (**precip**), outgoing longwave radiation ($\mathbf{Q}_{lw\ top}$), and surface air temperature ($\mathbf{T}_{surf}$) in the 2D PCA compressed latent space of the VED. Together the results show that distinct convective regimes are clearly separated in the latent space. More information on how the complex global superposition of distinct geographic convective regimes and large-scale processes in the latent space is contributed by separate latitudinal bands of the aquaplanet (tropics, boreal and austral mid latitudes) is provided in Section S.2 in Supporting Information S1. Therein Figure S4 in Supporting Information S1 shows the fixed SSTs of the simulation and S6 the regional decomposition of patterns in the VED's latent space. These two figures can aid as a reference guide for the following latent space exploration. We start the analysis by investigating the impact of the insolation $\mathbf{Q}_{sol}$ on the latent space position, including whether the expected diurnal cycle of convective processes in SPCAM (Khairoutdinov et al., 2005; Pritchard & Somerville, 2009) is manifested in the latent space of the VED. Indeed, solar insolation $\mathbf{Q}_{sol}$ reveals 2 distinct maxima with day-time conditions and 2 minima with night-time conditions, which are separated by diurnal transition zones, as expected from diurnally varying input and output data of SP. Cross-evaluating the conditional averages of solar insolation with $\mathbf{T}_{surf}$, one can diagnose that the 2D PCA compressed latent space of VED stores information that can be used to infer the geographic location of a sample. As an example, we can focus on the "fin-shaped" region (PC1 ∼ −8, PC2 ∼ 15) protruding from the top of the 2D PCA compressed latent space. Here the samples are characterized by anomalously cold ($273\ \mathrm{K} < \mathbf{T}_{surf} < 278\ \mathrm{K}$) climate conditions without solar insolation. Based on the fixed SST forcing (see Figure S4 in Supporting Information S1), the low surface air temperatures and the constant perpetual Austral Summer solar forcing, we can conclude that these samples originate from polar and subpolar latitudes in the Northern Hemisphere. Furthermore, we find a zone with day-time solar insolation ($\mathbf{Q}_{sol} > 700\ \frac{\mathrm{W}}{\mathrm{m}^2}$) and cold surface air temperature ($273\ \mathrm{K} < \mathbf{T}_{surf} < 280\ \mathrm{K}$) in the upper-right part of the latent space (PC1 ∼ 10, PC2 ∼ 5), which represents large-scale climate conditions that can be only found in the austral polar and subpolar latitudes in SPCAM test data.

We also explore the fingerprinting of precipitation on the latent space as a proxy for the strength of convective processes, since it is closely connected to convective moistening and convective heating (Emanuel, 1994; Lohmann et al., 2016). The 2D PCA compressed latent space of the VED reveals a good separation of samples with no or negligible precipitation, shallow convection with the formation of weak precipitation, and deep convective samples with intense precipitation (**precip** $> 10\ \frac{mm}{h}$ in the tropics, see Figure S6 in Supporting Information S1). We expect to see a clear separation between tropical deep convective samples and samples with no or negligible precipitation from the colder higher latitudes or the region of the subtropical highs in the 2D PCA compressed latent space due to the strong variation in the magnitude of convective processes with latitude as it is visible in Figure 5. If we now focus on the conditionally averaged plot of precipitation, two maxima are evident. The first precipitation maximum (PC1 ~ -15, PC2 ~ 5) is associated with day-time solar forcing, a minimum of outgoing long-wave radiation ($\mathbf{Q}_{lw\ top} < 150\ \frac{W}{m^2}$, which suggests high cloud tops in the upper half of the troposphere) and tropical surface air temperatures ($\mathbf{T}_{surf} > 295$ K). Therefore, this maximum originates from tropical day-time deep convective samples in SPCAM. The second maximum (PC1 ~ 5, PC2 ~ 5) exhibits slightly colder surface air temperatures, night-time conditions, decreased outgoing longwave radiation ($\mathbf{Q}_{lw\ top} \sim 100\ \frac{W}{m^2}$) and precipitation formation of more than $3\ \frac{mm}{h}$. It can be shown that this maximum originates from night-time deep convection from the tropics in its center and predominantly strong precipitating samples from the Northern and Southern extratropics along the left and right boundary, respectively.

Outgoing longwave radiation ($\mathbf{Q}_{lw\ top}$) is a good estimator for both the height of cloud tops based on the inferred brightness temperatures for convective samples or surface temperatures for non - or negligibly - convective samples. Based on the combination of high $\mathbf{Q}_{lw\ top}$ (no or negligible convection), no precipitation formation and anomalous warm surface temperatures ($\mathbf{T}_{surf} \sim 300$ K), one can conclude that samples from subtropical highs (the descending branch of the Hadley cell, with limited deep convective processes with large vertical extent in the free troposphere) are concentrated in the lower left part of the PCA compressed latent space (PC1 ~ -10, PC2 ~ -10) of the VED.

These results demonstrate how large-scale climate conditions and convective processes are connected and physically interpretable in the latent space (e.g., equivalence of **precip** maxima and $\mathbf{Q}_{lw\ top}$ minima), which illustrates the encoding power of the VED. Furthermore, the evaluated mean statistics support that the VED realistically reproduces convective processes and the associated variability despite a strong dimensionality reduction down to only five nodes in the latent space, which shows the decoding power of the network.

Similar reproduction abilities can be investigated for ED, but the physical interpretability of the resulting latent space is reduced compared to VED. The KL divergence used for the VED ensures an improved separation of latent modes. The effect can be seen in a larger number of centers of action in the ED's latent space and weaker gradients in the conditional average plots with respect to sub-grid-scale and climate variables, as can be seen in Figure S5 in Supporting Information S1 (ED vs. VED latent spaces) and S6 for the VED conditional average plot or S7 for the ED conditional average plot. Additionally, we tested the interpretability of the 2D PCA compressed latent space of a VED trained on **X** to emulate **Y**, in other words mirroring the input data and output data of SP (see Section S.3A in Supporting Information S1). In this case the latent space strongly focuses on the magnitude of heating or moistening tendencies, resembling a weak gradient from negligible convective processes toward strongly precipitating deep convective samples (see Figure S16 in Supporting Information S1). For large-scale climate variables like surface air temperature, the 2D PCA compressed latent space of a $\mathrm{VED}_{X\rightarrow Y}$ mostly distinguishes between warm conditions and cold conditions sorting samples from both poles close together in one minimum (see Figure S17 in Supporting Information S1), which makes the visual separation of austral and boreal polar latitudes nearly impossible. In contrast, VED shows a pronounced separation of austral and boreal polar samples and reveals distinct regimes of convective processes in its 2D PCA compressed latent space as seen in Figure 5, which is a clear advantage in interpretability of this network compared to $\mathrm{VED}_{X\rightarrow Y}$.

We further compared the interpretability of the 2D PCA compressed latent space of the VED against a traditional PCA on the large-scale input features **X**, as an unsupervised linear reference method. The first two leading PC's with respect to **X** show overall weak gradients in its lower-dimensional space for the conditional averages of solar insolation, outgoing longwave radiation and surface air temperature (Figure S8 in Supporting Information S1). The “centers of action” are less pronounced for the PCA on **X** compared to its equivalent in the latent space of VED seen in Figure 5. Especially the identification of deep convective samples is hardly possible inside the submanifold spanned by the two leading PC's of the large-scale variables as can be seen in Figure S8 in Supporting Information S1. In latitude - longitude plots (Figure S9 in Supporting Information S1) these two leading PC's resemble large-scale patterns with meridional gradients that show similarities with temperature or radiation fields but barely with sub-grid-scale variables. In contrast, the latent space of VED focuses on both large-scale and sub-grid-scale patterns. The first two latent variables are characterized by large-scale patterns connected to geographic variability and solar insolation (see Figure S10 in Supporting Information S1). The remaining three latent variables describe mostly sub-grid-scale convective processes, as can be seen in Figure S10 in Supporting Information S1.

The concept of the computation of conditional averages can be repeated also on 2D projections spanned by the original latent variables of the VED without a PCA as postprocessing step. An example of this more detailed latent space inspection can be found in Section S.2 in Supporting Information S1 (Figure S12 for precipitation, S13 for solar insolation and S14 for surface air temperature in Supporting Information S1).

As a next step, we will combine the reproduction skill and the encapsulated information in the latent space of the VED to investigate convective processes by identifying distinct large-scale drivers, associated convective regimes and geographic variability in detail.

4. Unveiling Drivers of Convective Processes in SPCAM Using Generative Modeling

In this section, we discuss the dominant drivers of convective processes encapsulated in the latent space of the VED using a generative modeling approach. We compute the marginal distributions of all 5 latent variables **z**. We focus on the 10th, 25th, 50th, 75th and 90th percentiles of the marginal distributions of the latent variables. Since most of these distributions are bi-modal (see Figures 6–10a), we select their median values as estimators for the intersect (origin) of the 5-dimensional **z**, instead of the mean. For all latent variables, the median is close to the mode value (peak value) of the marginal distributions. To generate the "median" climate conditions and associated convective processes from the "median" values of the latent variables, we construct a reference state $\mathbf{z}_{median}$ (Equation 8). $\mathbf{z}_{median}$ contains the median values for all five latent variables. This reference state is fed into the decoder of the VED to generate vertical heating, moistening, specific humidity, and temperature profiles. These vertical profiles represent the "median" state of convective processes and associated climate conditions.

$$\mathbf{z_{median}} = [\mathrm{median}\,(z_1)\,, \mathrm{median}\,(z_2)\,, \mathrm{median}\,(z_3)\,, \mathrm{median}\,(z_4)\,, \mathrm{median}\,(z_5)] \tag{8}$$

To investigate encapsulated convective regimes and large-scale climate states in the latent space of VED via generative modeling, we replace the median value with the different percentiles (perc (z_1) in Equation 9) along one specific marginal distribution. This analysis identifies how each latent node drives a variation of convective processes and large-scale climate states generated by the decoder and manifests in well-known convective regimes. The modified $\mathbf{z}_{translation}$ (Equation 9) can be seen as a latent forcing on the decoder, acting as a knob which amplifies or damps the associated convective features. Furthermore, $\mathbf{z}_{translation}$ influences the geographic variability of generated samples, allowing an interpolation from a tropical to a polar "background" climate state like a knob for the general volume of generated large-scale profiles. A clear separation between geographic versus convective modulation with a distinct $\mathbf{z}_{translation}$ is challenging and not the primary goal of our VED's decoder setup. The evaluation whether a distinct latent node drives more geographic than convective modulation necessarily involves an analysis of all generated variables - an interesting analysis trade-off revealed by this latent space exploration. $\mathbf{z}_{translation}$ can be geometrically interpreted as a translation along one distinct latent dimension in the 5-dimensional latent space. For instance, $\mathbf{z}_{translation}$ is applied as an example to latent node 1 perturbing the "median" conditions along this latent dimension, while keeping the median values for the four other dimensions:

$$\mathbf{z_{translation\,node\,1}} = \left[\mathrm{perc}\,(z_1)\,, \mathrm{median}\,(z_2)\,, \mathrm{median}\,(z_3)\,, \mathrm{median}\,(z_4)\,, \mathrm{median}\,(z_5)\right] \tag{9}$$

Applying a translation along one latent dimension while keeping the other latent variables fixed to their median values implicitly assumes that latent variables do not overly depend on each other. To test this independence, we calculate the Pearson correlation between all five latent variables using the entire test data set. The mean correlation coefficients between the latent dimensions are confined within ±0.35, except for a mean correlation of −0.74 between latent variables 2 and 5. The relatively large linear connection between latent variables 2 and 5 can be further explored by density plots using the 2D projection spanned by these latent variables, see Figure S11 in Supporting Information S1. While latent node 2 separates moist and warm from cold and dry tropospheric conditions, latent node 5 represents deep convection samples, which rely on anomalous wet and warm conditions in the troposphere. Therefore it is not surprising to see a pronounced anti-correlation between these nodes. This is

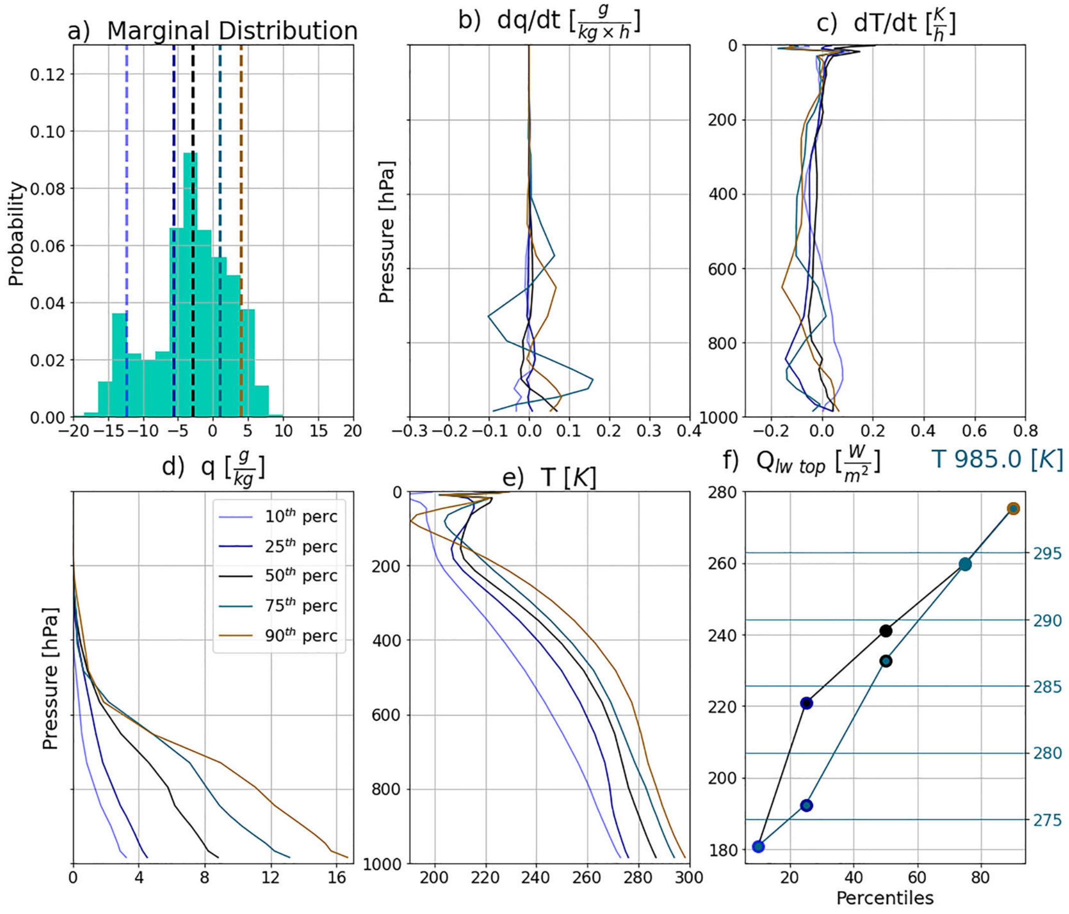

Figure 6. Marginal distribution of latent node 1 (a) and the resulting generated vertical profiles of specific humidity tendencies **dq/dt** (b), temperature tendencies **dT/dt** (c), specific humidity **q** (d) and temperatures **T** (e). The dashed lines in the marginal distribution plot represent the chosen percentiles (see legend in subplot d) and the resulting effect of the respective translation $\mathbf{z}_{translation}$ on the profiles is shown in the subplots. Furthermore, the longwave heat flux at the model top ($\mathbf{Q}_{lw\ top}$) and the surface air temperature ($\mathbf{T}_{surf}$/**T 985.0**) (f) are illustrated as function of the translation $\mathbf{z}_{translation}$ along the latent dimension 1. The marker-edge-color in panel f symbolize the respective percentiles of $\mathbf{z}_{translation}$. The black lines in subplots b-e indicate the generated reference state with $\mathbf{z}_{median}$.

a further evidence of the interpretability and meaningfulness of the VED's latent space, that is, learning physical processes in the lower-order manifold.

In the following we will use $\mathbf{z}_{translation}$ along all five latent dimensions to identify large-scale drivers of convective processes and different convective regimes in SPCAM. We use the notation "high $\mathbf{z}_{translation}$" to describe the cases when $\mathbf{z}_{translation} > \mathbf{z}_{median}$ and "low $\mathbf{z}_{translation}$" if $\mathbf{z}_{translation} < \mathbf{z}_{median}$. Figures 6–10 illustrate the marginal distribution along the respective latent dimensions (Panels a, where the dashed black line indicates the median value of each dimension, Equation 8). The other subplots of these figures show the generated vertical moistening, heating, specific humidity and temperature profiles (Panels b–e) of the decoder with respect to $\mathbf{z}_{median}$ (Equation 8) or $\mathbf{z}_{translation}$ (Equation 9 along a distinct latent dimension). Additionally, two sub-grid-scale and climate variables (Panels f), which are strongly affected by the applied latent forcing, are displayed as a function of $\mathbf{z}_{translation}$ for illustrative purposes. The marker-edge-color in the respective Panels *f* reveal the chosen percentiles. All other generated sub-grid-scale and large-scale climate variables are shown in Tables S7–S11 in Supporting Information S1. We investigate in the following that latent node 1 and latent node 2 focus on the large-scale climate (geographic) variability in **X** rather than on sub-grid-scale convective processes in **Y**. In contrast, latent nodes 3, 4 and 5 exhibit main characteristics of dominant convective regimes captured in **Y**.

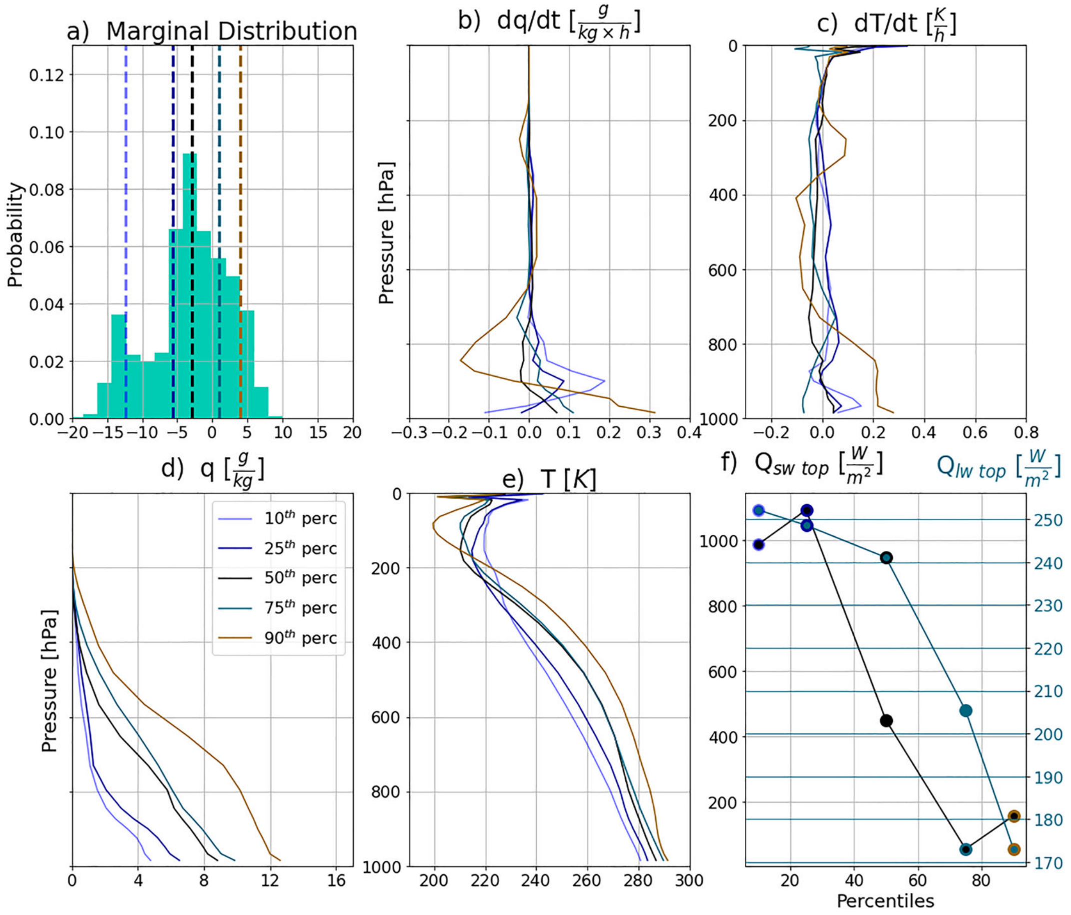

Figure 7. Marginal distribution of latent node 2 (a) and the resulting generated vertical profiles of specific humidity tendencies **dq/dt** (b), temperature tendencies **dT/dt** (c), specific humidity **q** (d) and temperatures **T** (e). The dashed lines in the marginal distribution plot represent the chosen percentiles (see legend in subplot d) and the resulting effect of the respective translation $\mathbf{z}_{translation}$ on the profiles is shown in the subplots. Furthermore, the shortwave heat flux at the model top ($\mathbf{Q}_{sw\ top}$) and the outgoing longwave heat flux ($\mathbf{Q}_{lw\ top}$) (f) are illustrated as function of the translation $\mathbf{z}_{translation}$ along the latent dimension 2. The marker-edge-color in panel f symbolize the respective percentiles of $\mathbf{z}_{translation}$. The black lines in subplots b-e indicate the generated reference state with $\mathbf{z}_{median}$.

4.1. Large-Scale Climate Variability Nodes

In this first part we demonstrate that latent nodes 1 and 2 capture mostly large-scale climate variability in **X**.

4.1.1. Latent Node 1: Global Temperature Variations

Global temperatures in the troposphere are dominated by the large meridional gradients from equatorial to polar latitudes mainly related to solar insolation differences between the tropics and extratropics.

The first latent node (Node 1) captures these global meridional temperature variations (Figure 6e), as suggested by the large spread of the surface temperature response to $\mathbf{z}_{translation}$, encompassing the tropics ($\mathbf{T}_{surf} \sim 298$ K, high $\mathbf{z}_{translation}$) and polar regions ($\mathbf{T}_{surf} \sim 273$ K, low $\mathbf{z}_{translation}$). Tropical regions are characterized by very moist conditions in the boundary layer ($\mathbf{q} > 10\,\frac{g}{kg}$), while being extremely dry at the poles ($\mathbf{q} \sim 1.5 - 3.5\,\frac{g}{kg}$), see Figure 6d. The strong connection between tropospheric temperatures or specific humidity and Node 1 can be shown with a linear correlation of globally concatenated temperature space-time series (of horizontal grid cells and time, featuring the large meridional gradients) and respective node space-time series. The resulting "linear explained variance" of temperature space-time series on Node 1 exceeds 0.5 (Figure S18 in Supporting Information S1), while the "linear explained variance" vanishes if the analysis is repeated on the time series for each horizontal grid cell (Figure S19 in Supporting Information S1, without the large meridional gradients). A detailed description how these two correlations metrics were computed can be found in Section S.4 in Supporting Information S1.

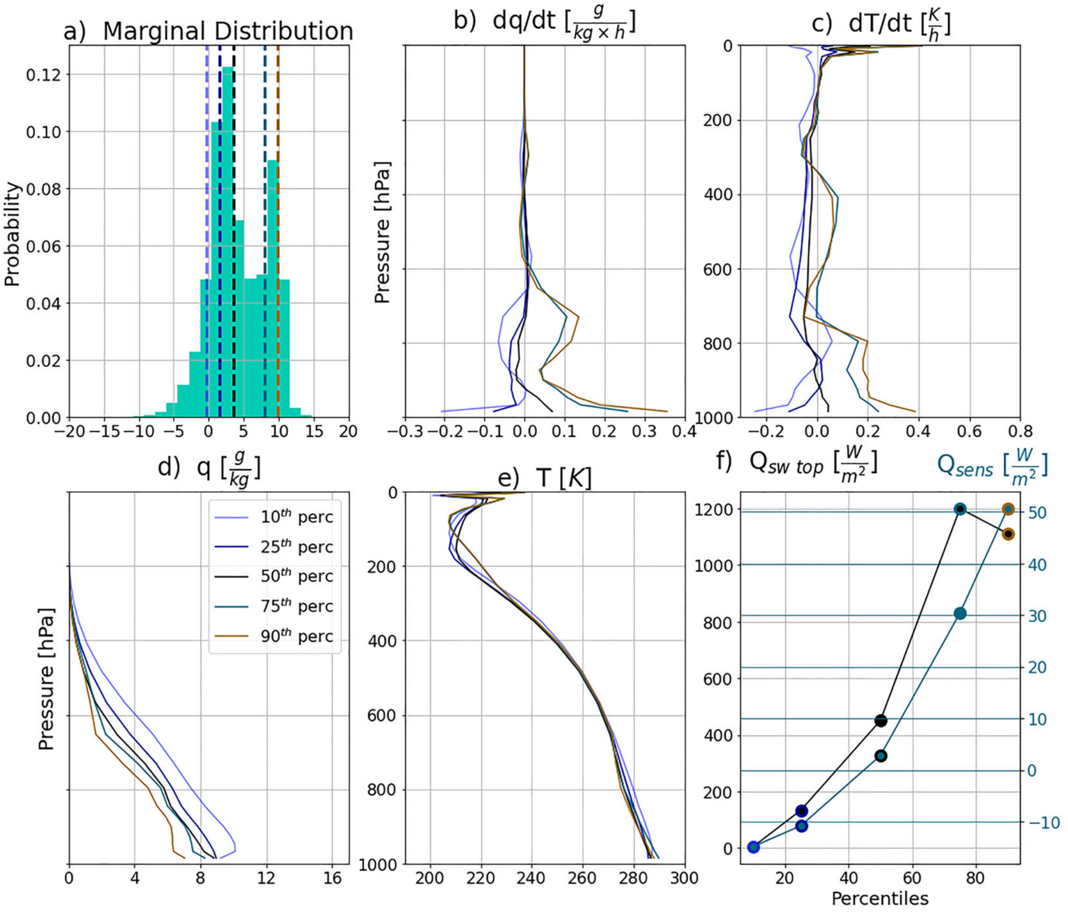

Figure 8. Marginal distribution of latent node 3 (a) and the resulting generated vertical profiles of specific humidity tendencies **dq/dt** (b), temperature tendencies **dT/dt** (c), specific humidity **q** (d) and temperatures **T** (e). The dashed lines in the marginal distribution plot represent the chosen percentiles (see legend in subplot d) and the resulting effect of the respective translation $\mathbf{z}_{translation}$ on the profiles is shown in the subplots. Furthermore, the shortwave heat flux at the model top ($\mathbf{Q}_{sw\ top}$) and the surface sensible heat flux ($\mathbf{Q}_{sens}$) (f) are illustrated as function of the translation $\mathbf{z}_{translation}$ along the latent dimension 3. The marker-edge-color in panel f symbolize the respective percentiles of $\mathbf{z}_{translation}$. The black lines in subplots b-e indicate the generated reference state with $\mathbf{z}_{median}$.

A physical interpretation of this response on the $\mathbf{z}_{translation}$ can be given based on the Clausius-Clapeyron relationship. A warmer atmosphere results in a near-exponentially higher saturation water vapor pressure, which in turn allows higher specific humidity content. Therefore, we see strongly coupled variations of temperature and specific humidity between the equator and the poles. In short, the first latent node represents these overarching large-scale meridional variations in tropospheric temperatures, influencing specific humidity, but is not necessarily linked to convective processes **Y**, but rather to large-scale conditions **X**, which are also part of the VED reconstruction.

4.1.2. Latent Node 2: Large-Scale Variability Along the Mid-Latitude Storm Tracks

Latent node 2 characterizes more the large-scale climate (and thus geographic) variability in **X** than focuses on a distinct convective regime. Latent dimension 2 (Node 2, Figure 7) clearly captures temperature and specific humidity variations in the troposphere, as can be seen in Figures 7d and 7e. Warmer and moister tropospheric conditions are associated with high $\mathbf{z}_{translation}$.

Low $\mathbf{z}_{translation}$ characterizes cold and stable conditions during day-time ($\mathbf{Q}_{sw\ top} \sim 1000\ \frac{W}{m^2}$). These anomalous cold and dry conditions in the upper troposphere are associated with negligible convective processes, as diagnosed with a large outgoing longwave heat flux at the model top ($\mathbf{Q}_{lw\ top} \sim 240\ \frac{W}{m^2}$) and the formation of no precipitation (Table S8 in Supporting Information S1). Due to the large shortwave heat flux at the model top, the perpetual

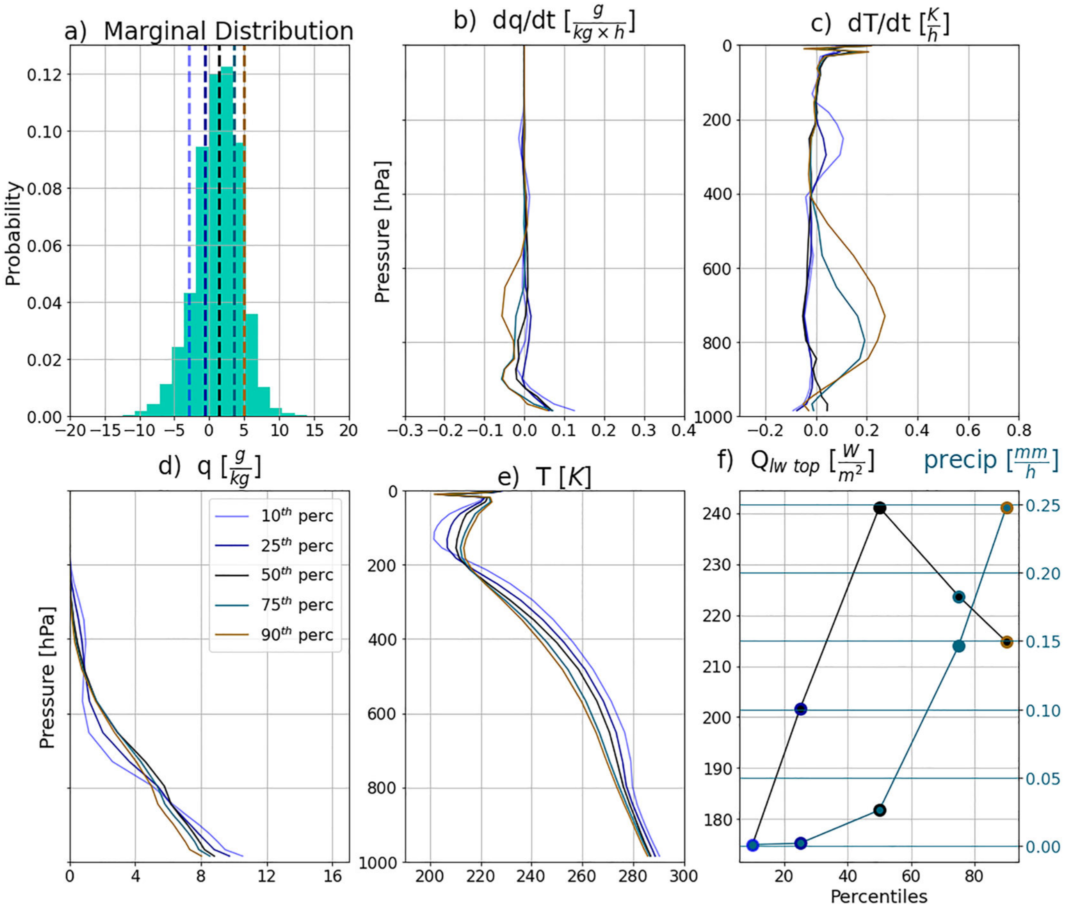

Figure 9. Marginal distribution of latent node 4 (a) and the resulting generated vertical profiles of specific humidity tendencies **dq/dt** (b), temperature tendencies **dT/dt** (c), specific humidity **q** (d) and temperatures **T** (e). The dashed lines in the marginal distribution plot represent the chosen percentiles (see legend in subplot d) and the resulting effect of the respective translation $\mathbf{z}_{translation}$ on the profiles is shown in the subplots. Furthermore, the longwave heat flux at the model top ($\mathbf{Q}_{lw\ top}$) and the precipitation rate (**precip**) (f) are illustrated as function of the translation $\mathbf{z}_{translation}$ along the latent dimension 4. The marker-edge-color in panel f symbolize the respective percentiles of $\mathbf{z}_{translation}$. The black lines in subplots b-e indicate the generated reference state with $\mathbf{z}_{median}$.

Austral Summer solar forcing and the low surface air temperatures ($\mathbf{T}_{surf} \sim 281$ K), low $\mathbf{z}_{translation}$ can be traced back to the austral mid-latitudes. Whereas high $\mathbf{z}_{translation}$ is linked to night-time conditions ($\mathbf{Q}_{sw\ top} < 200\ \frac{W}{m^2}$) with a warm, moist troposphere ($\mathbf{T}_{surf} \sim 291$ K). High $\mathbf{z}_{translation}$ is further characterized by mid-level convection ($\mathbf{Q}_{lw\ top} \sim 180 - 200\ \frac{W}{m^2}$) with intermediate precipitation formation (**precip** $\sim 0.12 - 0.15\ \frac{mm}{h}$, Table S8 in Supporting Information S1) associated with a warmer and moister upper troposphere and can be found in the subtropics on both hemispheres.

Our approach allows us to identify the main patterns of the large-scale climate state in **X**, which are main drivers of the general circulation and convection, besides convective processes in **Y** in the latent space. These convective processes are heavily modulated by **X**. Node 2 captures characterizing features of the large-scale meridional variability of specific humidity and temperatures between the mid latitudes and the subtropics (i.e., an essential driver of mid-latitude storm track dynamics; Bony et al., 2015). Latent dimension 2 is further influenced by the solar forcing. The clear separation between austral mid latitude temperature profiles on one side and samples from subtropical regions on the other side of the $\mathbf{z}_{translation}$ are further evidence that latent node 2 encapsulates a part of the geographic variability inside the latent space seen in Figure 5.

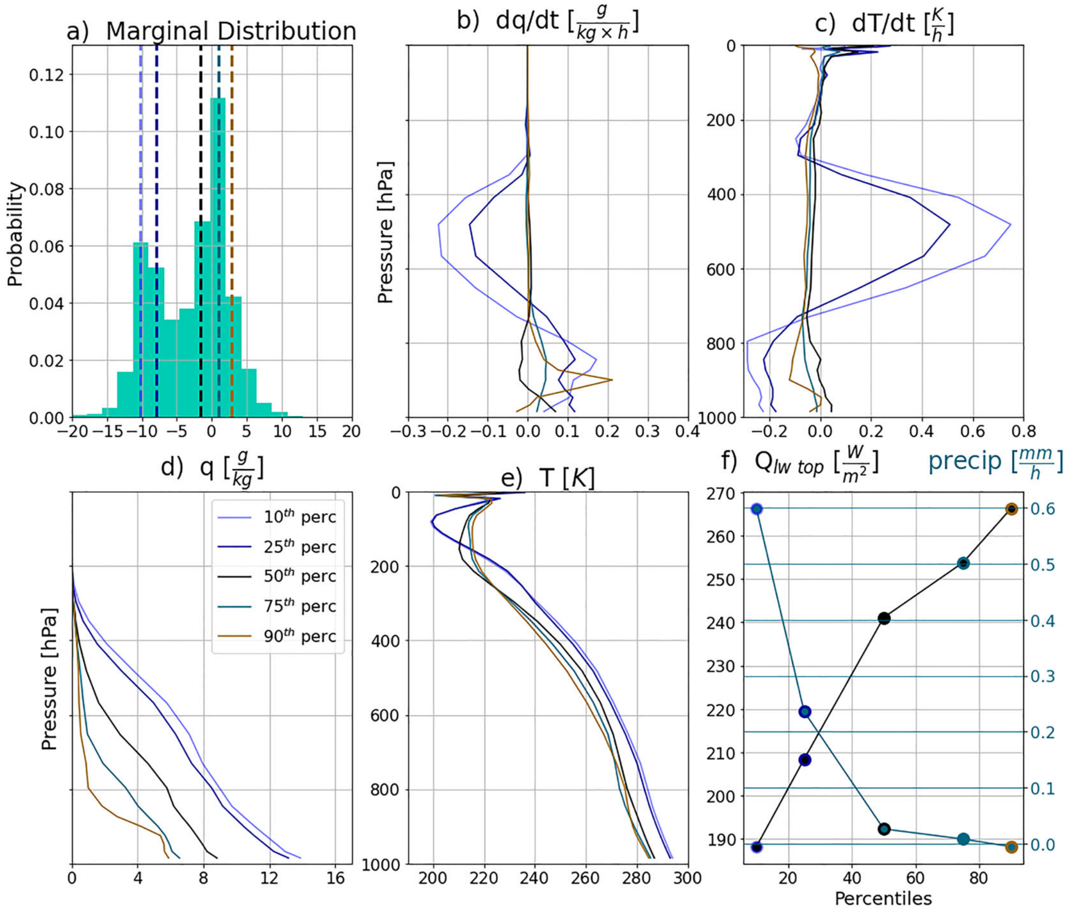

Figure 10. Marginal distribution of latent node 5 (a) and the resulting generated vertical profiles of specific humidity tendencies **dq/dt** (b), temperature tendencies **dT/dt** (c), specific humidity **q** (d) and temperatures **T** (e). The dashed lines in the marginal distribution plot represent the chosen percentiles (see legend in subplot d) and the resulting effect of the respective translation $\mathbf{z}_{translation}$ on the profiles is shown in the subplots. Furthermore, the longwave heat flux at the model top ($\mathbf{Q}_{lw\ top}$) and the precipitation rate (**precip**) (f) are illustrated as function of the translation $\mathbf{z}_{translation}$ along the latent dimension 5. The marker-edge-color in panel f symbolize the respective percentiles of $\mathbf{z}_{translation}$. The black lines in subplots b-e indicate the generated reference state with $\mathbf{z}_{median}$.

4.2. Convective Regime Nodes

Next, we will show that latent node 3, 4, 5 usefully characterize mostly distinct convective regimes in the sub-grid-scale process rate variables **Y**.

4.2.1. Latent Node 3: Shallow Convection

Shallow convective processes are one of the dominant cloud regimes investigated in observational studies (e.g., Huaman & Schumacher, 2018). Latent node 3 characterizes some of the main characteristics of shallow convective processes as revealed by its vertical profiles of specific humidity and temperature tendencies influenced by large-scale specific humidity and surface diabatic fluxes.

Figure 8 shows the marginal distribution of latent node 3 (Node 3), the generated vertical specific humidity and temperature tendencies, and the large-scale specific humidity and temperature profiles of the Decoder for $\mathbf{z}_{median}$, as well as the applied $\mathbf{z}_{translation}$. Furthermore, the generated shortwave heat flux ($\mathbf{Q}_{sw\ top}$) and surface sensible heat flux ($\mathbf{Q}_{sens}$) are displayed as a function of $\mathbf{z}_{translation}$. Along latent dimension 3, the specific humidity (**q**) decreases throughout the entire troposphere for increasing $\mathbf{z}_{translation}$, while surface diabatic fluxes (sensible heat flux $\mathbf{Q}_{sens}$ and latent heat flux $\mathbf{Q}_{lat}$, Table S9 in Supporting Information S1) increase. Likewise, the outgoing longwave radiation $\mathbf{Q}_{lw\ top}$ increases with increasing $\mathbf{z}_{translation}$ suggesting higher cloud tops and stronger convective processes for low $\mathbf{z}_{translation}$ (Table S9 in Supporting Information S1). In contrast, the intensity of shallow convection and outgoing longwave radiation decreases when $\mathbf{z}_{translation}$ increases (high $\mathbf{z}_{translation}$). Specific humidity tendencies (**dq/dt**) in the lower troposphere ($\mathbf{p} > 600$ hPa) react to $\mathbf{z}_{translation}$ in a bimodal way. They moisten, in combination with a strong positive surface diabatic forcing, the relatively dry ambient air in the lower troposphere above the reference conditions (high $\mathbf{z}_{translation}$), whereas the opposite is true for low $\mathbf{z}_{translation}$. In this case, negative **dq/dt** in combination with negative diabatic forcing lead to a drying of moist conditions in the lower troposphere. Precipitation is insensitive to $\mathbf{z}_{translation}$ due to the small vertical extent of convective moistening, confined below 600 hPa; this latent node evidently avoids deep convective regimes. The generated temperature profiles of $\mathbf{z}_{translation}$ along latent dimension 3 are characteristic of the subtropics and mid-latitudes in the SP simulation. The **dT/dt** profiles show slight variations near the surface due to $\mathbf{z}_{translation}$, while being insensitive in the middle troposphere. The fixed SST field (Figure S4 in Supporting Information S1) or the conditional averages of surface air temperatures (Figure S6 in Supporting Information S1) in certain regions can be used to gain a first visual orientation of the geographic origin of a generated sample. This first impression is complemented with a detailed search for such conditions in the SP test data. Furthermore, night-time conditions with small shortwave radiative heat flux at the model top $\mathbf{Q}_{sw\ top}$ and day-time conditions with high values of $\mathbf{Q}_{sw\ top}$ ($\mathbf{Q}_{sw\ top} \sim 1000\,\frac{W}{m^2}$) can be distinguished for low $\mathbf{z}_{translation}$ and high $\mathbf{z}_{translation}$, respectively.

Interestingly, the generated profiles and variables suggest that latent node 3 is mostly sorting information about sub-grid-scale processes **Y** within one sub-regime of **X**, rather than focusing on sorting the large-scale geographic variability in **X**. The strong response of **dq/dt** in the planetary boundary layer and adjacent layers, negligible precipitation formation and the characteristic temperature range between the subtropics and mid-latitudes, are key evidence that the latent node 3 encapsulates shallow convective processes. Shallow convection is influenced by the diurnal cycle, leading to a strengthening of shallow convective processes during the day and a weakening of these processes accompanied with a drying of the planetary boundary layer during the night, as it is supported by Figure 8.

4.2.2. Latent Node 4: Mid-Latitude Frontal Systems

Mid-latitude frontal systems are characterized by a large variety of convective regimes associated with the warm or cold front of these systems (Bony et al., 2015). On latent node 4 we discover certain characteristic features in sub-grid-scale profiles **X** and associated large-scale fields **Y**. These features allow us to draw links to distinctive convective regimes of mid-latitude cyclones based on their fingerprint in **X** and **Y**. Unlike the previous latent nodes, the response of the latent node 4 (Node 4, Figure 9) to the translation $\mathbf{z}_{translation}$ results in nearly constant solar insolation ($\mathbf{Q}_{sw\ top} \sim 440 - 450\,\frac{W}{m^2}$, see Table S10 in Supporting Information S1) and a narrow meridional band.

The generated surface temperature ranges from 286 to 290 K with varying $\mathbf{z}_{translation}$. This temperature range is common to mid-latitudes or the subtropics (e.g., see Figure S6 in Supporting Information S1) and can be found in the SPCAM simulations between 45° N/S and 25° N/S. Low $\mathbf{z}_{translation}$ corresponds to warmer and drier conditions in the free mid-troposphere between 800 hPa and 400 hPa, while moister conditions are found above and below. The anomalous moist conditions in the upper free troposphere are connected to a heating peak at 300 hPa (**dT/dt** $\sim 0.1\,\frac{K}{h}$, Figure 9c). Likewise, the difference between the shortwave heat flux at the model top and the surface is relatively small ($\mathbf{Q}_{sw\ top} - \mathbf{Q}_{sw\ surf} \sim 120 - 130\,\frac{W}{m^2}$, Table S10 in Supporting Information S1), which suggests optically thin clouds. Additionally, the outgoing long wave radiation is small ($\mathbf{Q}_{lw\ top} < 200\,\frac{W}{m^2}$) and no precipitation is formed. These conditions are characteristic of high cirrus-like convection.

On the other side, high $\mathbf{z}_{translation}$ shows relatively strong heating tendencies in the free troposphere (**dT/dt** $> 0.2\,\frac{K}{h}$, see Figure 9c) and drying conditions below 600 hPa down to the surface (**dq/dt** $\sim -0.1\,\frac{g}{kg \times h}$, Figure 9b). These conditions, along with moderate precipitation (**precip** $\sim 0.15 - 0.25\,\frac{mm}{h}$), higher outgoing longwave heat flux ($\mathbf{Q}_{lw\ top} > 200\,\frac{W}{m^2}$) and lower shortwave transmissivity ($\mathbf{Q}_{sw\ top} - \mathbf{Q}_{sw\ surf} \sim 170\,\frac{W}{m^2}$, Table S10 in Supporting Information S1) characterize mid-level cumulus convection.

Based on this evidence, we were able to show that latent node 4 focuses on sub-grid-scale convective processes in **Y**. The generated large-scale conditions exhibited by Node 4 are well-suited for these cirrus-like or cumulus convection regimes. In detail, latent node 4 shows a clear transition from a cirrus type convective regime (low $\mathbf{z}_{translation}$) to a cumulus type precipitating convective regime (high $\mathbf{z}_{translation}$) in mid-latitudes. This response is associated with frontal systems, which consist of high cirrus clouds in the surroundings of the warm front and cumulus convection along the cold front (Bony et al., 2015).

4.2.3. Latent Node 5: Deep Convection

Deep convection is the cloud regime with the largest vertical extent. It is characterized by especially strong convective heating and drying throughout almost the entire troposphere, as can be seen in Frenkel et al. (2015) and accompanied by anomalous intense precipitation (see Figure 5). The first mode of latent node 5 reveals general characteristics of a deep convective regime captured in generated sub-grid-scale variables $\mathbf{Y}$. The response of latent dimension 5 (Node 5) to $\mathbf{z}_{translation}$ shows either strong deep convection (first mode in Figure 10a) or stable conditions (second mode in Figure 10a) in the troposphere. A surface temperature of 293 K for low $\mathbf{z}_{translation}$ indicates subtropical regions (e.g., the surface temperature in the tropics is at least 3 K warmer in this SPCAM simulation). The warmer and moister troposphere for low $\mathbf{z}_{translation}$ is accompanied with strong heating and drying tendencies peaking at around 500 hPa of $0.5 - 0.7\,\frac{K}{h}$ and -0.15 to $-0.2\,\frac{g}{kg \times h}$ respectively. In this case, we observe intense precipitation formation up to $0.6\,\frac{mm}{h}$ and low outgoing longwave radiation ($\mathbf{Q}_{lw\ top} < 201\,\frac{W}{m^2}$). All these conditions are characteristics of subtropical deep convective events.

In contrast, high $\mathbf{z}_{translation}$ is associated with a mid latitude surface air temperature ($\mathbf{T}_{surf} \sim 5$ K colder than in the subtropics). A night-time (Table S11 in Supporting Information S1), dryer troposphere with very small or negligible heating and moistening tendencies (manifestation of stable conditions) throughout the troposphere is accompanied with relatively large outgoing long wave radiation ($\mathbf{Q}_{lw\ top} > 250\,\frac{W}{m^2}$) and no precipitation. Similar to latent node 3 and 4, latent node 5 comprises dominantly information about sub-grid-scale convective processes rather than large-scale geographic variability. Latent node 5 represents both deep convective events originating from the subtropics and mid-latitude stable conditions as can be already seen in the strong bimodality along the marginal distribution in Figure 10a.

5. Conclusion and Discussion

This study has shown how Variational Encoder Decoders (VEDs) can successfully machine learn a convective parameterization with considerable input compression while simultaneously enhancing the interpretability of deep learning methods, and enable better understanding of convective processes in climate models. We first showed that the VED is able to realistically reconstruct convective processes simulated by a superparameterized climate model, similar to previous studies with regular Artificial Neural Nets (ANNs) (Gentine et al., 2018; Rasp et al., 2018), but using automatically compressed input data. Furthermore, we demonstrated that the VED also enhances the interpretability of the relationship between large-scale climate fields and sub-grid-scale convective variables via its latent manifold, which is unfeasible via ANNs without attribution methods due to ANNs' large dimensionality (large number of hidden layers and nodes per layer). Our analysis is based on 9 months (equally split into training, validation and test data) of an aquaplanet simulation of the SuperParameterized Community Atmosphere Model (SPCAM). As shown in Figure 11a, the input variables of the VED resembled the large-scale climate fields (temperature, specific humidity and other thermodynamic drivers) from the general circulation model (CAM) passed onto the embedded cloud resolving model (SP). The latent space (lower dimensional manifold inside the network) of the VED had a dimensionality of five nodes, which is a small fraction of the dimensionality of the original input nodes information. To create an interpretable latent space, our optimal network reconstructed a combination of sub-grid-scale convective variables related to the SP component and large-scale climate variables associated with CAM. In comparison, as we have shown in the supplemental material, VEDs that attempt the traditional mapping from $\mathbf{X}$ to $\mathbf{Y}$ alone turn out to be less amenable to latent space exploration.

As a first step, we evaluated the reproduction performance of convective processes of the VED against a reference ANN (Rasp et al., 2018). The VED was capable of reconstructing the mean statistics of sub-grid-scale convective variables with an overall comparable, though slightly decreased, skill than the reference ANN despite the strong dimensionality reduction down to five latent nodes. This speaks to the dimensionality of information content required for a convective parameterization, and associated trade-offs. We found that compressing the input information did not overly distort the tropical wave spectrum. We showed that the choice of the latent space width is a critical hyperparameter for reproduction skills. Larger latent space widths ($\sim$ 8 nodes) yielded a reproduction performance of convective processes with almost the skill of the reference ANN, while smaller latent space widths ($\sim$ 2 nodes) still enabled an improved reproduction compared to a multi-dimensional linear regression

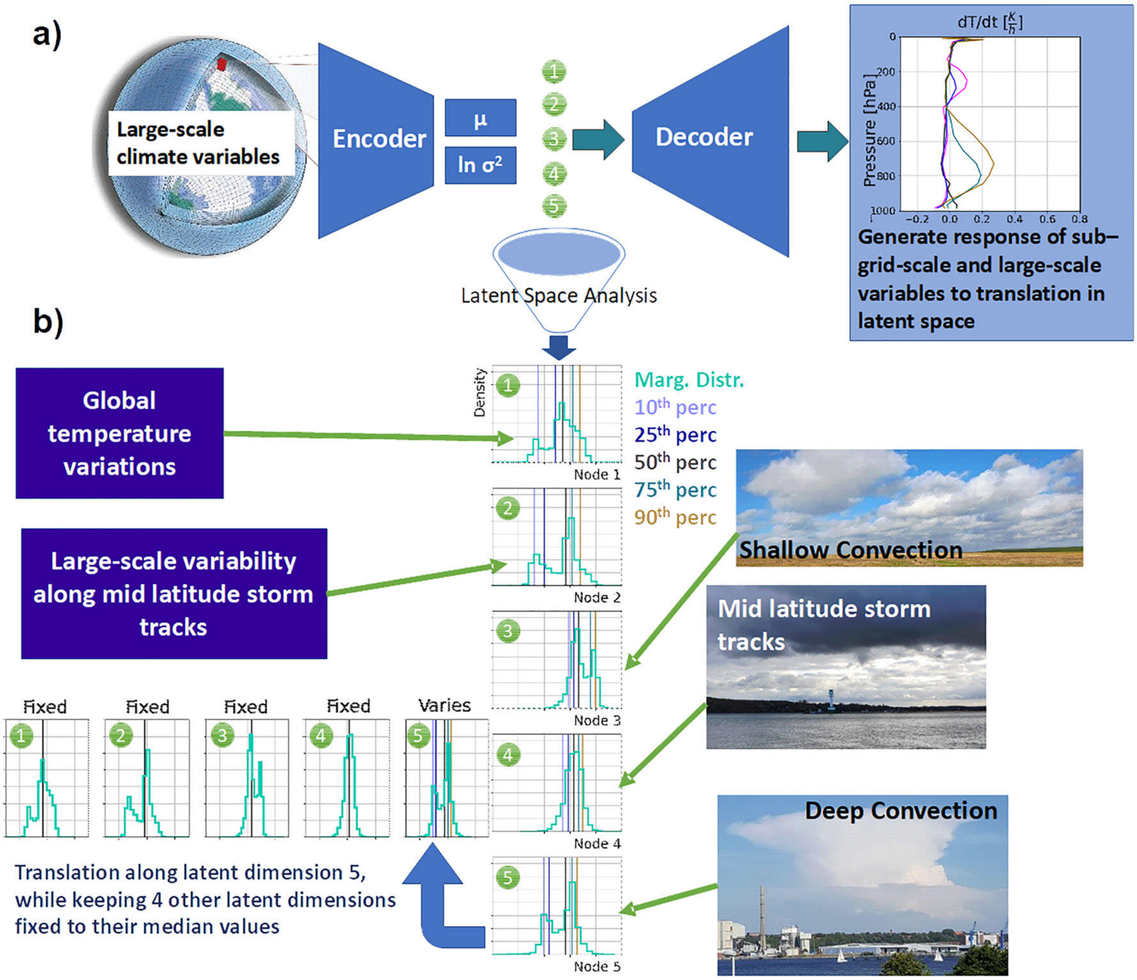
a)
Large-scale climate variables
Encoder
μ
ln σ²
1
2
3
4
5
Decoder
dT/dt [K/h]
Pressure [hPa]
0
200
400
600
800
1000
−0.2
0.0
0.2
0.4
0.6
0.8
Generate response of sub–grid-scale and large-scale variables to translation in latent space
Latent Space Analysis
b)
Global temperature variations
Large-scale variability along mid latitude storm tracks
Density
Node 1
Node 2
Node 3
Node 4
Node 5
Marg. Distr.
10th perc
25th perc
50th perc
75th perc
90th perc
Shallow Convection
Mid latitude storm tracks
Deep Convection
Fixed
Fixed
Fixed
Fixed
Varies
Translation along latent dimension 5, while keeping 4 other latent dimensions fixed to their median values

Figure 11. Schematic of the Variational Encoder Decoder (VED) setup (a) the investigated convective regimes and drivers of convective processes in the latent space of VED for each node (b). The translation along each latent dimension is shown in this example for latent node 5. The schematic of the large-scale atmospheric grid in (a) was adapted from Schneider et al. (2017).

baseline. We chose a latent space of five nodes as a sensible compromise between reproduction abilities of convective regimes and sensitivities separable in the latent manifold.

We began the analysis toward our main interest -latent space exploration with respect to physical interpretability–using traditional methods visualizing physical properties in a 2D projection of its leading PCs. This revealed that the VED distinguished day- and night-time conditions and varying strength of convective processes using the precipitation rate and outgoing longwave radiation as a proxy in its latent space (which was 2D compressed with a PCA for the purpose of visualization). The VED separated different global climate conditions and associated convective regimes from the poles to the equator in its latent space. The realistic reproduction of convective processes and climate conditions, along with the encapsulated information on geographic variability in an interpretable latent manifold, allowed a detailed analysis of governing drivers of convection and convective regimes with a VED.

Our latent exploration was then deepened by investigating convective processes and related drivers via a generative modeling approach, that is, forcing the decoder with the variability encapsulated along each latent dimension. The resulting temperature, specific humidity, heating, and moistening profiles successfully separated well-known large-scale driving climate conditions and convective regimes. Figure 11b summarizes the main results of this generative modeling approach. Overall, convective processes are controlled by large meridional gradients in temperature and specific humidity, from the equator to the poles, which were captured by the VED's Node 1 (Figure 11). We identified the large-scale climate variability in specific humidity and temperatures along the mid-latitude storm tracks (Node 2, Figure 11) as the other major driver of convective processes. Daytime stable, cold and dry tropospheric conditions suppress convective processes in the entire troposphere, whereas night-time unstable, warm and moist conditions in the troposphere drive precipitating mid-level convection. Apart from these large-scale nodes, the VED further reveals characteristics of distinct convective regimes on the remaining 3 latent nodes. The VED confined shallow convective processes below 600 hPa within its Node 3 (Figure 11); these processes are generally driven by surface diabatic fluxes and are predominantly originating from mid-latitudes and the subtropics. In anomalous dry conditions, positive surface diabatic heat fluxes during day-time enhance shallow convective processes associated with a convective moistening of the lower troposphere. The opposite is true in anomalous wet conditions during night-time. The mid-latitude storm tracks show large variability with respect to convective regimes associated with the eastward migrating frontal systems, features that were captured in the VED's latent space (Node 4, Figure 11). In the surroundings of the warm front high, optically thin, non-precipitating cirrus-like convection is found. In contrast lower, optically thick cumulus-like convection with intermediate precipitation formation is predominant near the cold front. Furthermore, deep convective regimes in the subtropics were clearly captured by the VED (Node 5, Figure 11). In this case, convective processes extend in the entire troposphere with a pronounced convective heating and drying near 500 hPa and are associated with intense precipitation. Opposing this extreme convective case, we found night-time, stable, cold and dry conditions in the free troposphere, which suppress convective processes on the other side of Node 5. Finally, while the interpretation of these convective regimes always required domain knowledge, the generative modeling approach simplified the analysis in comparison to other statistical analysis tools (e.g., correlations, clustering, attribution methods).

Repeating this analysis with an Encoder Decoder (ED) yielded almost identical reproduction capabilities compared to the VED, but the ED's latent space was significantly harder to interpret, with less pronounced center of actions for a given variable (see Figure S5 and S7 in Supporting Information S1). This hindered the identification of convective regimes or large-scale drivers of convective predictability within the latent space of ED. For example, although the ED captured a cirrus-like regime, no cumulus or deep convective regimes could be found with the generative modeling method. Likewise, the connection between large-scale climate variables was often less pronounced for the ED, which resulted in larger uncertainties of the geographic origin of a specific sample compared to the VED.

We discovered convective regimes with the VED that are in general agreement with existing work focused on tropical convection (Frenkel et al., 2012, 2013, 2015; Huaman & Schumacher, 2018). The specific humidity profile of the shallow convective regime of the VED was largely similar to the observed shallow convective latent heating profile in Huaman and Schumacher (2018) with a heating peak around 800 hPa. Furthermore, the heating profile of the mid-latitude cirrus-like regime of the VED compared well with that of the tropical stratiform regime shown in Frenkel et al. (2015) despite strong differences in the ambient conditions that led to their formation. Also the heating profiles of the mid-latitude cumulus regime of the VED and their tropical congestus expressed similarities in the lower troposphere with a pronounced convective heating peak above the boundary layer. Likewise, the VED's subtropical and tropical deep convection regime of Frenkel et al. (2015) were characterized by similar heating profiles. In our case, we identified these regimes solely based on SPCAM data in the latent space of the VED, where we did not prescribe the characteristics of each convective regime like it was done in the multi-cloud approach presented in Frenkel et al. (2012) and adapted from Khouider and Majda (2006). Furthermore, our approach was not based on inferred heating profiles via subclassing precipitation regimes (Stratiform, Convective, Shallow) as it was done for observational satellite products in Huaman and Schumacher (2018).

This work presented how convective processes, convective regimes, and large-scale drivers of convection in climate models can be investigated by leveraging generative Machine Learning (ML) approaches. Our approach enhanced the understanding of acting convective processes and the corresponding large-scale environment in which they form. As a next step, one could study cirrus-like or cumulus convection in detail by, for example, separating specific humidity and moistening tendencies related to the ice phase, linking how microphysical processes influence convection and are, in turn, affected by climate conditions (i.e., formation of ice phase, mixed phase or liquid phase clouds). Likewise, the development of regime-oriented ML-based convection parameterizations appears to be achievable with generative deep learning methods. Finally, VEDs could play an essential role in constructing new stochastic convection parameterizations, which could improve the representation of clouds and convection in Earth System Models (ESMs). Our results suggest that VED representations of climate processes can effectively combine statistical prediction with data-driven analysis, paving the way toward machine learning-based ESMs that remain interpretable, albeit through the yet mostly unfamiliar eccentricities of latent space exploration.

Data Availability Statement

The code used to train the VEDs, the conditional VAE and reference models, and to produce all figures of this manuscript is accessible in the following Github repository: https://github.com/EyringMLClimateGroup/behrens22james_SPCAM_VED, which is archived with Zenodo (https://zenodo.org/record/6925020#.YuKPRITP25c). The repository includes the Jupyter Notebooks, python files, conda environments used to reproduce all figures of the manuscript and attached supporting information. The text file https://github.com/EyringMLClimateGroup/behrens22james_SPCAM_VED/blob/master/List_of_Figures.txt illustrates where to find the code to reproduce each Figure in the Github repository. The above mentioned Github repository is a fork of Stephan Rasp's main repository published for Rasp et al. (2018), which can be found here: https://github.com/raspstephan/CBRAIN-CAM, archived using Zenodo (https://zenodo.org/record/1402384#.YajSg9BKiUk). The repository includes a helpful quickstart guide https://github.com/raspstephan/CBRAIN-CAM/blob/master/quickstart.ipynb to preprocess raw SPCAM data, train a neural network similar to reference ANN and how to evaluate it. An example of SPCAM data was archived on Zenodo for Rasp et al. (2018) and can be found here: https://zenodo.org/record/2559313#.YlVG0tPP25c. The full SPCAM raw data, of the order of several TBs, is archived on the GreenPlanet cluster at UC Irvine and available upon request. Additionally the preprocessed SPCAM data, of the order of 1 TB, used in this study is archived on DKRZ and is also available upon request.

Acknowledgments

We thank two anonymous reviewers and one internal reviewer for their helpful comments and suggestions, which improved our manuscript. Funding for this study was provided by the European Research Council (ERC) Synergy Grant “Understanding and modeling the Earth System with Machine Learning (USMILE)” under the Horizon 2020 research and innovation programme (Grant agreement No. 855187). Beucler acknowledges funding from the Columbia University sub-award 1 (PG010560-01). Gentine acknowledges funding from the National Science Foundation Science and Technology Center, Learning the Earth with Artificial intelligence and Physics, LEAP (Grant number 2019625). Pritchard acknowledges funding from the National Science Foundation Science and Technology Center, Learning the Earth with Artificial intelligence and Physics (LEAP, 2019625) and from the US Department of Energy Advanced Scientific Computing Research program (DE-SC0022331). This work used resources of the Deutsches Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA) under project ID 1179 (USMILE). Open access funding enabled and organized by Projekt DEAL.

References

Alberdi, X.-A. T., Reimers, C., Denzler, J., & Reichstein, M. (2018). SupernoVAE: VAE based kernel-pca for analysis of spatio-temporal Earth data. In *8th international workshop on climate informatics*.

Alemi, A., Poole, B., Fischer, I., Dillon, J., Saurous, R. A., & Murphy, K. (2018). Fixing a broken elbo. In *International conference on machine learning* (pp. 159–168).

Andersen, J. A., & Kuang, Z. (2012). Moist static energy budget of MJO-like disturbances in the atmosphere of a zonally symmetric aquaplanet. *Journal of Climate*, *25*(8), 2782–2804. https://doi.org/10.1175/JCLI-D-11-00168.1

Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., et al. (2020). Quantifying progress across different CMIP phases with the ESMValTool. *Journal of Geophysical Research: Atmospheres*, *125*(21), 1–28. https://doi.org/10.1029/2019JD032321

Bony, S., Stevens, B., Frierson, D. M., Jakob, C., Kageyama, M., Pincus, R., et al. (2015). Clouds, circulation and climate sensitivity. *Nature Geoscience*, *8*(4), 261–268. https://doi.org/10.1038/ngeo2398

Collins, W. D., Rasch, P. J., Boville, B. A., Hack, J. J., McCaa, J. R., Williamson, D. L., & Zhang, M. (2006). The dynamical simulation of the Community Atmosphere Model version 3 (CAM3). *Journal of Climate*, *19*(11), 2162–2183. https://doi.org/10.1175/JCLI3762.1

Derbyshire, S. H., Beau, I., Bechtold, P., Grandpeix, J. Y., Piriou, J. M., Redelsperger, J. L., & Soares, P. M. (2004). Sensitivity of moist convection to environmental humidity. *Quarterly Journal of the Royal Meteorological Society*, *130*(604), 3055–3079. https://doi.org/10.1256/qj.03.130

Emanuel, K. (1994). *Atmospheric convection*. Oxford University Press. Retrieved from https://books.google.de/books?id=VdaBBHEGAcMC

Frenkel, Y., Majda, A. J., & Khouider, B. (2012). Using the stochastic multicloud model to improve tropical convective parameterization: A paradigm example. *Journal of the Atmospheric Sciences*, *69*(3), 1080–1105. https://doi.org/10.1175/JAS-D-11-0148.1

Frenkel, Y., Majda, A. J., & Khouider, B. (2013). Stochastic and deterministic multicloud parameterizations for tropical convection. *Climate Dynamics*, *41*(5–6), 1527–1551. https://doi.org/10.1007/s00382-013-1678-z

Frenkel, Y., Majda, A. J., & Stechmann, S. N. (2015). Cloud-radiation feedback and atmosphere-ocean coupling in a stochastic multicloud model. *Dynamics of Atmospheres and Oceans*, *71*, 35–55. https://doi.org/10.1016/j.dynatmoce.2015.05.003

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, *45*(11), 5742–5751. https://doi.org/10.1029/2018GL078202

Grabowski, W. W. (2001). Coupling cloud processes with the large-scale dynamics using the clouds-resolving convection parameterization (CRCP). *Journal of the Atmospheric Sciences*, *58*(9), 978–997. https://doi.org/10.1175/1520-0469(2001)058<0978:ccpwtl>2.0.co;2

Han, Y., Zhang, G. J., Huang, X., & Wang, Y. (2020). A moist physics parameterization based on deep learning. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2020MS002076. https://doi.org/10.1029/2020ms002076

Huaman, L., & Schumacher, C. (2018). Assessing the vertical latent heating structure of the east Pacific itcz using the cloudsat cpr and trmm pr. *Journal of Climate*, *31*(7), 2563–2577. https://doi.org/10.1175/JCLI-D-17-0590.1

Khairoutdinov, M. F., Randall, D., & DeMott, C. (2005). Simulations of the atmospheric general circulation using a cloud-resolving model as a superparameterization of physical processes. *Journal of the Atmospheric Sciences*, *62*(7), 2136–2154. https://doi.org/10.1175/jas3453.1

Khairoutdinov, M. F., & Randall, D. A. (2001). A cloud resolving model as a cloud parameterization in the NCAR community climate system model: Preliminary results. *Geophysical Research Letters*, *28*(18), 3617–3620. https://doi.org/10.1029/2001GL013552

Khairoutdinov, M. F., & Randall, D. A. (2003). Cloud resolving modeling of the arm summer 1997 iop: Model formulation, results, uncertainties, and sensitivities. *Journal of the Atmospheric Sciences*, *60*(4), 607–625. https://doi.org/10.1175/1520-0469(2003)060<0607:crmota>2.0.co;2

Khouider, B., & Majda, A. J. (2006). A simple multicloud parameterization for convectively coupled tropical waves. part i: Linear analysis. *Journal of the Atmospheric Sciences*, *63*(4), 1308–1323. https://doi.org/10.1175/jas3677.1

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kingma, D. P., & Welling, M. (2014). Auto-encoding variational bayes. In *2nd international conference on learning representations. ICLR 2014–Conference track proceedings (Ml)*.

Krinitskiy, M. A., Zyulyaeva, Y. A., & Gulev, S. K. (2019). Clustering of polar vortex states using convolutional autoencoders. *CEUR Workshop Proceedings*, *2426*, 52–61.

Lohmann, U., Lüönd, F., & Mahrt, F. (2016). *An introduction to clouds: From the microscale to climate*. Cambridge University Press. Retrieved from https://books.google.de/books?id=FbpDDAAAQBAJ

Lorenz, E. N. (1996). Predictability: A problem partly solved. *Proc. seminar on predictability*, *1*.

Mamalakis, A., Ebert-Uphoff, I., & Barnes, E. A. (2021). Neural network attribution methods for problems in geoscience: A novel synthetic benchmark dataset. *arXiv preprint arXiv:2103.10005*.

Mooers, G., Pritchard, M., Beucler, T., Ott, J., Yacalis, G., Baldi, P., & Gentine, P. (2021). Assessing the potential of deep learning for emulating cloud superparameterization in climate models with real-geography boundary conditions. *Journal of Advances in Modeling Earth Systems*, *13*(5), e2020MS002385. https://doi.org/10.1029/2020ms002385

Mooers, G., Tuyls, J., Mandt, S., Pritchard, M., & Beucler, T. (2020). Generative modeling of atmospheric convection. *arXiv*. https://doi.org/10.1145/3429309.3429324

Neelin, J. D., & Zeng, N. (2000). A quasi-equilibrium tropical circulation model—Formulation. *Journal of the Atmospheric Sciences*, *57*(11), 1741–1766. https://doi.org/10.1175/1520-0469(2000)057<1741:aqetcm>2.0.co;2

Pritchard, M. S., & Bretherton, C. S. (2014). Causal evidence that rotational moisture advection is critical to the superparameterized madden–julian oscillation. *Journal of the Atmospheric Sciences*, *71*(2), 800–815. https://doi.org/10.1175/jas-d-13-0119.1

Pritchard, M. S., Bretherton, C. S., & DeMott, C. A. (2014). Restricting 32–128 km horizontal scales hardly affects the mjo in the superparameterized community atmosphere model v. 3.0 but the number of cloud-resolving grid columns constrains vertical mixing. *Journal of Advances in Modeling Earth Systems*, *6*(3), 723–739. https://doi.org/10.1002/2014ms000340

Pritchard, M. S., & Somerville, R. C. J. (2009). Assessing the diurnal cycle of precipitation in a multi-scale climate model. *Journal of Advances in Modeling Earth Systems*, *2*, 12. https://doi.org/10.3894/james.2009.1.12

Randall, D. A. (2013). Beyond deadlock. *Geophysical Research Letters*, *40*(22), 5970–5976. https://doi.org/10.1002/2013GL057998

Randall, D. A., Khairoutdinov, M., Arakawa, A., & Grabowski, W. (2003). Breaking the cloud parameterization deadlock. *Bulletin of the American Meteorological Society*, *84*(11), 1547–1564. https://doi.org/10.1175/BAMS-84-11-1547

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(39), 9684–9689. https://doi.org/10.1073/pnas.1810286115

Rolinek, M., Zietlow, D., & Martius, G. (2019). Variational autoencoders pursue pca directions (by accident). In *Proceedings ieee conf. on computer vision and pattern recognition (cvpr)* (pp. 12406–12415). Retrieved from http://openaccess.thecvf.com/content_CVPR_2019/papers/Rolinek_Variational_Autoencoders_Pursue_PCA_Directions_by_Accident_CVPR_2019_paper.pdf

Schneider, T., Teixeira, J., Bretherton, C., Brient, F., Pressel, K., Schär, C., & Siebesma, A. (2017). Climate goals and computing the future of clouds. *Nature Climate Change*, *7*(1), 3–5. https://doi.org/10.1038/nclimate3190

Stevens, B., Acquistapace, C., Hansen, A., Heinze, R., Klinger, C., Klocke, D., et al. (2020). The added value of large-eddy and storm-resolving models for simulating clouds and precipitation. *Journal of the Meteorological Society of Japan*, *98*(2), 395–435. https://doi.org/10.2151/jmsj.2020-021

Wang, P., Yuval, J., & O'Gorman, P. A. (2022). Non-local parameterization of atmospheric subgrid processes with neural networks. *arXiv preprint arXiv:2201.00417*.

Wang, X., Han, Y., Xue, W., Yang, G., & Zhang, G. J. (2021). Stable climate simulations using a realistic gcm with neural network parameterizations for atmospheric moist physics and radiation processes. *Geoscientific Model Development Discussions*, 1–35.

Yuval, J., & O'Gorman, P. A. (2020). Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature Communications*, *11*(1), 1–10. https://doi.org/10.1038/s41467-020-17142-3

Zhang, C. (2005). Madden-julian oscillation. *Reviews of Geophysics*, *43*(2). https://doi.org/10.1029/2004RG000158