## Supplemental Information

We wish to provide additional details about our decomposition discussed in Section 2 and the choices and performance behind our unsupervised machine learning framework from Section 3. We start by discussing information on the hyper-parameters choices for the clustering that leads to our dynamical regime shifts, ’s. As a justification for our use of a Variational Autoencoder, we provide a comparison of the performance against a simpler linear baseline. Finally, we also include additional steps showing how we decompose the full spatial variance of heavy precipitation for analysis as well as plotting all the terms of the full decomposition.

### *A. K-Means Clustering Approach*

We apply the K-Means Clustering algorithm to partition the latent space of our VAE and analyze which physical properties can be clustered in this reduced order  space. This approach first randomly assigns centroids, C, to locations in the  space (note we actually use the more modern k++ algorithm [5] to maximize the initial distances between the centroids). Latent representations of each sample , in the test dataset of size , are assigned to their nearest centroid. The second stage of the algorithm moves the centroid to the mean of the assigned cluster. The process repeats until the sum of the square distances (or the Inertia, ) between the latent space data points and the centroids are minimized [22, 23] such that:

 (5.1)

in which  is the mean of the given samples belonging to a cluster l for the total number of cluster centers C. We always calculate ten different K-means initializations and then select the initialization with the lowest inertia. This process allows us to derive the three data-driven convection regimes within SPCAM highlighted in Fig 3.

We qualitatively choose an optimal number of cluster centroids (centers),  by incorporating domain knowledge rather than a traditional approach relying on the rate of decrease in  as  increases or a single quantitative value such as a Silhouette Coefficient [34] or Davies-Bouldin Index [7]. More specifically, we identify the maximum number of “unique clusters”. We define a “unique cluster” of convection as a group in the latent space where the typical physical properties (vertical structure, intensity, and geographic domain) of the vertical velocity fields are not similar to the physical properties of another group elsewhere in the latent space. Empirically this exercise enables us to create three unique regimes of convection (Fig 3). When we increase  above three, we get sub-groups of “Deep Convection” without differences in either vertical mode, intensity, or geography. Thus we don’t consider  to be physically meaningful for our purposes.

Because we seek to contrast common clusters between different climates, we do not use Agglomerative (hierarchical) Clustering unlike other recent works that cluster compressed representations of clouds from ML models [8, 20]. Using the K-means approach, we can save the cluster centroids at the end of the algorithm. This provides a basis for cluster assignments for latent representations of out-of-sample test datasets when we use a common encoder as in Section B. More specifically, we only use the cluster centroids to get label assignments in other latent representations. We don’t move the cluster centroids themselves once they have been optimized on the original test dataset (the second part of the K-means algorithm). Keeping the center of the clusters the same between different types of test data ensures we can objectively contrast cluster differences through the lens of the common latent space.

### *B. VAE Benchmarking and Performance Evaluation*

We train our VAE on 160,000 unique vertical velocity fields and use an additional 125,000 samples to validate and optimize the model hyperparameters. Finally, we leverage 1,000,000 vertical velocity fields in the test dataset for robust analysis. The high count in the test dataset is necessary both due to the high spatio-temporal correlations common in meteorological data but also because of the geographic conditioning in our analysis – we need enough samples at each lat/lon grid cell, not just globally. To determine whether our data are nonlinear enough to warrant the use of a VAE we also train a baseline model of the same architecture but with all activation functions replaced by “linear”. The fact that the VAE reconstructs the vertical velocity snapshots with both lower error and a higher degree of structural similarity suggests significant non-linearity is involved in compressing and rebuilding the 2D fields (Tab 1 and Tab 2). This problem is therefore well suited for the non-linear dimensionality reduction of the VAE encoder and less so for linear models.

|  |  |  |  |
| --- | --- | --- | --- |
| Mean Squared Error | | | |
| Model | Training Set | Validation Set | Test Set |
| VAE |  |  |  |
| Linear Baseline |  |  |  |

Table 1: The MSE of both of our models (“linear baseline” and VAE) calculated across training/validation/test data. For both training and test data, we see low reconstruction errors, suggesting satisfactory skill and generalization ability. Overall, the VAE outperforms the “linear” baseline

|  |  |  |  |
| --- | --- | --- | --- |
| Structural Similarity Index Metric | | | |
| Model | Training Set | Validation Set | Test Set |
| VAE |  |  |  |
| Linear Baseline |  |  |  |

Table 2: The mean SSIM [36] of both of our models across training/validation/test data. The models both generalize well to our test data. Again, the VAE outperforms the “linear baseline”

Tab 1 and Tab 2 show 160,000 is enough training samples to create reconstructions of high-resolution vertical velocity fields with both a low MSE and a high degree of overall structural similarity. Though there is a small amount of overfitting, we see that performance remains strong for a test dataset containing multiple species of convection from all parts of the tropics ranging from deserts to rainforests; oceans to continents.

Figure S1: We plot the sample nearest to the centroid in the latent space of each of our three regimes of convection before being passed through the VAE encoder (a,c,e) and what those convection samples look like after being decoded by the VAE (b,d,f). We use this as a proxy for the centroids themselves, which at the high dimension of our latent space (1000) are themselves on a different manifold. Through this proxy, we can gleam the nature of the convection in each type of regime and how the VAE latent space organizes itself. Qualitatively, we see the VAE is able to reconstruct different forms of convection across different spatial scales.

Qualitatively, we also show decoded snapshots nearest the centroids from each of our three regimes of tropical convection. Overall reconstruction quality is good and we can see spatial structures from a variety of scales reproduced through the decoder of the VAE (Fig. S1 a,c,e vs. b,d,f). There is some under-prediction of the most intense up and downdrafts of convection (Fig. S1b), but this is to be expected as our loss is optimized to promote disentanglement of the latent space over reconstruction accuracy. If our goal was reconstruction accuracy instead of interpretability, we could add a statistical constraint to the loss function as we did in [25] to better reconstruct extremes and small-scale variance.

### *C. Full Decomposition of the Spatial Variance of Heavy Precipitation*

We derive the decomposition of the spatial variance of heavy precipitation in four steps. First, we multiply Eq. (2.7) by :

 (5.2)

For convenience, we use  to denote the first term of Eq. 5.2. We then take its spatial anomaly by applying the spatial anomaly operator :

 (5.3)

where we have used the fact that is uniform in space ( and ). Squaring Eq 5.3 yields:

 (5.4)

where the cross-terms CT can be decomposed into cross-terms involving spatial shifts in regime probability:

 (5.5)

cross-terms involving changes in how each regime produces precipitation:

 (5.6)

and additional cross-terms:

 (5.7)

Taking the spatial mean  of Eq 5.4 and noting that the spatial variance is defined as the spatial mean of the squared spatial anomaly, we derive the following decomposition:

 (5.8)

where we have introduced the decomposition’s numerical residual, which helps us assess which terms are significant. Grouping the terms irrelevant to the comparison between regime spatial shifts and intra-regime changes into a single term, , mathematically defined as:

 (5.9)

we recover Eq. (4.1) from the manuscript’s main text. For additional context on the significance of our decomposition, we plot all the terms in Fig. S2. We see that as in Fig. S2, our decomposition is most valid for high precipitation percentiles (percentiles where the residual (blue line) is of lesser magnitude than other quantities).

Figure S2: Derived from Eq. 5.8, we plot each term from the full decomposition for the variance in the change in heavy precipitation, . We focus primarily on precipitation percentiles 80-99, where our model is valid (the numerical residual, grey, is smaller than the key terms) and we have sufficient data (Fig. 3). Across these upper quantiles of precipitation, we find that the change in probability of convection type ( – red) is of greater importance than changes in the Dynamical Prefactors ( – blue). For additional context compared to Figure 4, we include all terms from Eq. 5.8

### *D. Supplemental Figures*

Figure S3: The shifts in different percentiles of precipitation with global warming, where we again stratified and plotted the data by latitude/longitude grid cell. As in Fig 3d we again remove the mean to highlight the dynamical pattern and see at what threshold the alignment with the VAE identified Deep Convection shifts (Fig 3c) is greatest. The top percentiles including (f-h) are pixelated because of a lack of samples that are out on the tail of the PDF.

Figure S4: The simple results of the simple regression model we use to predict patterns in heavy precipitation () using just the dynamic contributions,  and  identified by our unsupervised ML framework. We see our model works very well for high precipitation percentiles where the dynamic contributions are greatest and less well for lower percentiles where thermodynamics are also important.

Figure S5: From Eq. 2.7, we can decompose the changing spatial patterns (Fig 3f) into five terms, including probability changes in shallow convection (a), changes in deep convective precipitation (b), and the intercept of Dynamical Prefactor (c).