Implicit learning of convective organization explains precipitation stochasticity

Sara Shamekh [1,†], Kara D. Lamb[1], Yu Huang[1], and Pierre Gentine[1]

[1]Department of Earth of Environmental Engineering, Columbia University, New York, NY, USA
[*]corresponding author: Sara Shamekh (ss6287@columbia.edu)

ABSTRACT

Accurate prediction of precipitation intensity is of crucial importance for both human and natural systems, especially in a warming climate more prone to extreme precipitation. Yet climate models fail to accurately predict precipitation intensity, particularly extremes. One missing piece of information in traditional climate model parameterizations is sub-grid scale cloud structure and organization, which affects precipitation intensity and stochasticity at the grid scale. Here we show, using storm-resolving climate simulations and machine learning, that by implicitly learning sub-grid organization, we can accurately predict precipitation variability and stochasticity with a low dimensional set of variables. Using a neural network to parameterize coarse-grained precipitation, we find mean precipitation is predictable from large scale quantities only; however, the neural network cannot predict the variability of precipitation ($R^2 \sim 0.4$) and underestimates precipitation extremes. Performance is significantly improved when the network is informed by our novel organization metric, correctly predicting precipitation extremes and spatial variability ($R^2 \sim 0.95$). The organization metric is implicitly learned by training the algorithm on high-resolution precipitable water, encoding organization degree and humidity amount at the subgrid-scale. The organization metric shows large hysteresis, emphasizing the role of memory created by sub-grid scale structures. We demonstrate this organization metric can be predicted as a simple memory process from information available at the previous time steps. These findings stress the role of organization and memory in accurate prediction of precipitation intensity and extremes and the necessity of parameterizing sub-grid scale convective organization in climate models to better project future changes in the water cycle and extremes.

Introduction

The global mean accumulated precipitation set by the energy balance of the atmosphere is typically close to the observed value in global circulation models (GCM)[1]. Moreover, average precipitation as a function of precipitable water (PW), the vertically integrated water vapor content in an atmospheric column, exhibits a simple power-law behavior[2], increasing rapidly with PW. However, precipitation shows large variability around this value (Figure 1). In other words, given the large scale quantities, precipitation exhibits strong stochasticity. However, climate models fail to accurately reproduce precipitation statistics; they rain too often and too little (the so-called "drizzle problem"), and significantly underestimate extreme precipitation[3]. This underestimation has been associated with the parameterization of deep convection and microphysics in GCM's[4,5]. Typical mass-flux convection schemes use a model of convective updrafts, under the assumption of quasi-equilibrium (QE) with the larger (resolved) scale climate state[6]. Similarly, cloud cover parameterization and their impact on radiation sometimes use representations of aggregation, but these remain largely ad hoc[7]. These approaches effectively assume a mean state of sub-grid-scale convection and cloudiness, ignoring any pattern of or interaction between convective clouds and organization.

In both observations and storm-resolving models (with a horizontal resolution of a few kilometers), convective clouds form different modes and degrees of organization that can persist from a few hours to several days and impact precipitation significantly[8,9]. For instance, convection growing in a moist environment, which is the case with organized convection[10], has a longer lifetime and can grow deeper. Because the air entrained at the edge of convection is moist, it does not reduce the buoyancy of convective plumes. This mechanism also creates a moisture memory that is self-reinforcing, such that regions that experience deep convection and precipitation are more likely to have convective precipitation in the near future[10]. Furthermore, precipitation efficiency, defined as the ratio of surface precipitation to the cloud condensate, is larger for organized convection resulting in an enhancement in surface precipitation[11]. Precipitation efficiency is inversely related to rain re-evaporation, which is larger in a dry environment. Forecasting an accurate precipitation efficiency is crucial for a more realistic prediction of precipitation extremes in a warming climate[12]. Furthermore, inter-cloud interactions have been shown to change precipitation statistics: convection triggered by the collision of cold pools, areas of relatively low temperature generated by the evaporation of rain or melting of ice, is more likely to result in extreme precipitation[13]. Cold pools play a crucial role in the modification of the atmospheric moisture field, the depth and width of deep convection, as well as convective organization[14–17], especially on a

sub-daily time scale[18]. These emphasize the key role of climate model subgrid scales (100km) on precipitation.

Therefore, subgrid scale organization may be one source of stochasticity regulating precipitation intensity at the climate scale. The failure of traditional deterministic parameterization to accurately reproduce the statistical behavior of precipitation is potentially rooted in their negligence of subgrid-scale organization. Parameterizing sub-grid-scale organization has previously been suggested as a potential approach to circumvent the so-called "entrainment" problem[19]. Mapes and Neale (2011) included a prognostic and empirical “organization" variable in their convective parameterization, with the effect that sub-grid-scale structure led to deeper and stronger convection than in the unorganized case. Their prognostic “organization" parameter was a single empirical dimensionless scalar, tuned such that it had a steady-state value on the order of unity and a timescale on the order of a few hours.

In this work, we use machine learning to develop a precipitation parameterization that incorporates data-driven organization metrics for its prediction. Machine learning techniques such as deep neural networks provide a powerful opportunity for developing new parameterizations, given the availability of high resolution simulations of the atmosphere. Neural networks (NN's) can closely approximate the underlying function that relates observed quantities (or inputs) to the target quantities. Machine learning has recently proven useful for parameterizing sub-grid scale processes, such as convection and precipitation, in climate models by training neural networks to emulate higher resolution simulations where processes such as convection are explicitly resolved[4,20–23]. Neural networks can emulate unresolved processes from resolved variables, as they are powerful in capturing the nonlinear behavior of a physical system, given they have access to the necessary predictors. Machine learning has also shown tremendous skill in analyzing images and their complex spatial structure. Here we take advantage of this capacity of NN's to test two hypotheses, namely to:

- Evaluate to what extent the mean and variability of precipitation is predictable using only resolved quantities.
- Investigate whether informing the neural network with subgrid-scale organization information improves its prediction of precipitation.

Our neural network is a simple fully connected feed-forward network (Figure 2.b) that receives coarse-scale variable as input and predicts coarse-scale mean precipitation as well as its standard deviation (assuming the process to be stochastic). Due to the stochastic nature of precipitation, we prefer to predict the probability distribution instead of one single precipitation value for each given large scale condition. Thus we treat precipitation as a random variable, and target its first and second moment (mean and variance). We train our neural network using two-dimensional fields of global storm resolving simulations coarse-grained to typical climate model resolution, of 100 km, which we refer to as large-scale quantities. The high-resolution simulation was produced using the System for Atmospheric Modeling (SAM)[24], as a part of the DYnamics of the Atmosphere general circulation Modeled on Non-hydrostatic Domains (DYAMOND) Phase 2 Inter-comparison Project[25] and is publicly available. The original resolution of SAM-DYAMOND is 4.2 km and the coarse-grained resolution is 100x100 km^2 (Figure 2.a). The impact of the resolution on the results is shown in table S2 and figure S3.

Predicting precipitation from large scale quantities

To investigate the first hypothesis, i.e. the predictability of precipitation using large-scale quantities only, we use a neural network depicted in Figure 2.b which we refer to as Baseline NN.

Figure 3.a shows the precipitation predictability when the NN uses as input PW, SST, specific humidity, and temperature. The two latter variables inform the baseline-NN about the boundary layer conditions which has been shown to be important for the prediction of precipitation triggering[26]. The baseline NN recovers the critical behavior of mean precipitation accurately and its rapid transition past this critical point. However, it can not explain the precipitation variability observed in the global storm resolving simulations and its performance, measured by R^2 across all samples, is very close to zero (0.01). Low R^2 reveals that even though the baseline-NN captures some spread, it does not find a significant relationship between inputs and precipitation. Furthermore, R^2 computed at each bin of precipitable water (3.a green line) does not exceed 0.4.

We run further tests with our baseline NN, in which we also include total cloud cover (at the coarse scale) as input of the network. In climate models, total cloud cover is a parameterized variable and not directly related to precipitation, so that being included as input of the NN it could provide hints about the condensed water, which is directly used for the parameterization of precipitation. This inclusion (not shown) improves the prediction negligibly (R^2 = 0.06) but emphasizes that the mean cloud cover does not provide the relevant information for the prediction of precipitation. Additionally, further analysis confirms that including Convective Available Potential Energy (CAPE) or Convective INhibition (CIN) as predictors does not improve this prediction (not shown). Comparing the probability density function of precipitation predicted by the baseline-NN with the true precipitation (Figure 3.b) reveals that the model also fails to predict the tail of the distribution, in line with climate models' prediction of precipitation and their difficulty in representing extremes.

To summarize, the ensemble mean precipitation is significantly predictable from coarse-scale quantities (PW, SST and boundary layer information) by a neural network; however, the variability in precipitation cannot be captured. These findings

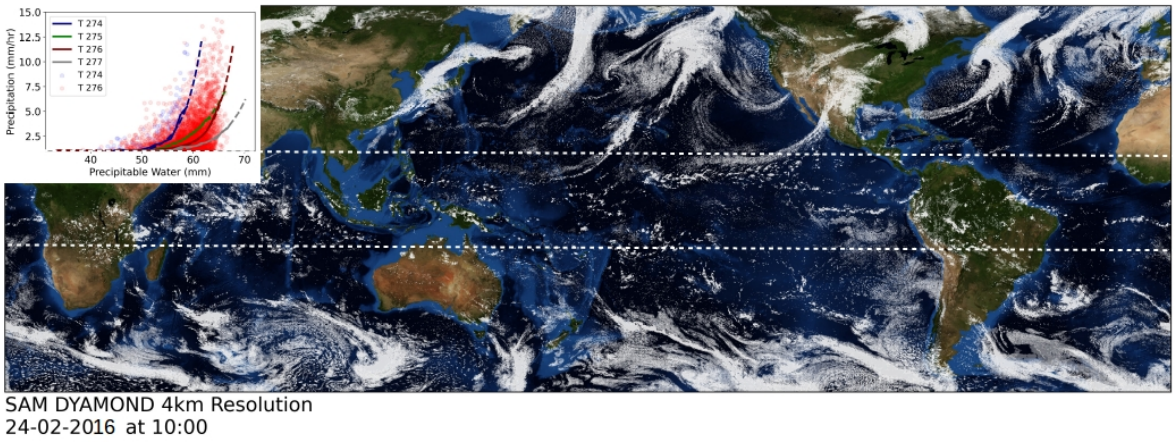

Figure 1. Global storm resolving model Snapshot of a cloud scene on 24 February 2016 from SAM as part of the DYAMOND dataset. 10 days of the tropical regions (displayed between the two white dashed lines) from this simulation are used for this analysis. The inset plot shows precipitation versus precipitable water for 10 days of SAM simulations. Lines show the ensemble average precipitation conditionally averaged by 0.3-mm bins of precipitable water and for 1-K bins of free tropospheric temperature. Scatter dots show the spread in precipitation for each bin of precipitable water and averaged free tropospheric temperature across the simulation domain and time period.

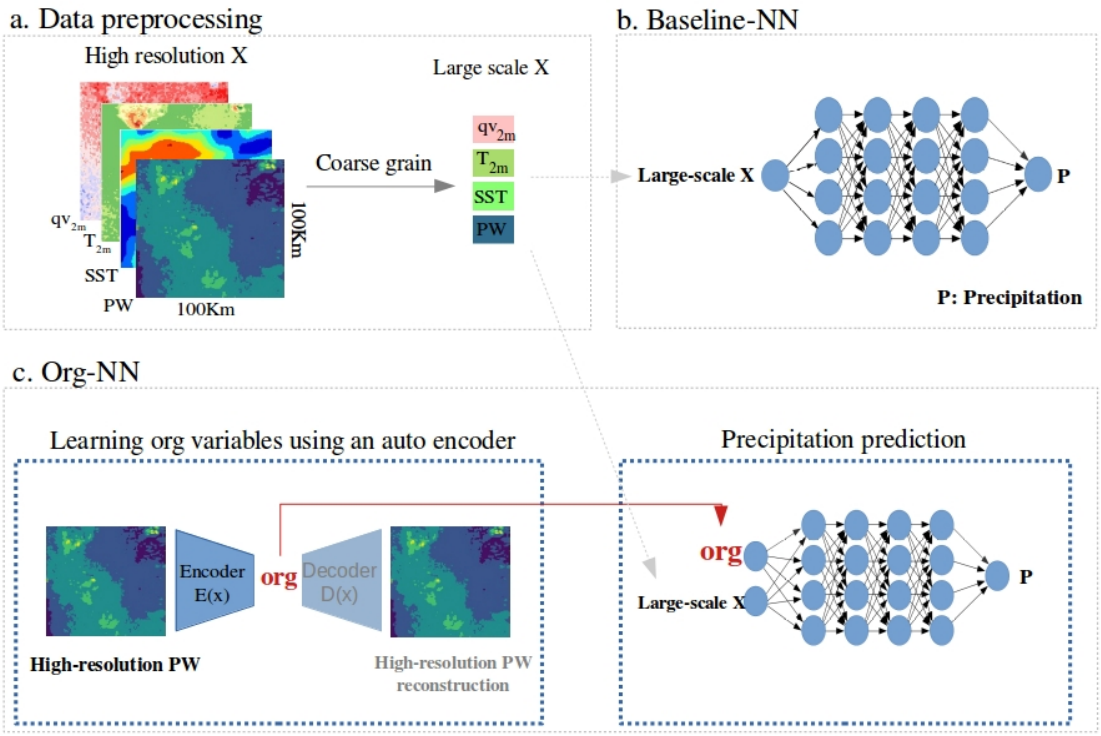

Figure 2. Overview of proposed framework for parameterizing precipitation a. Coarse-graining the high resolution data. b. Baseline-NN architecture: this network receives large scale variables (e.g. SST and PW) as input and predicts precipitation. c. Org-NN architecture: left panel shows the auto-encoder that receives the high-resolution PW as input and reconstructs it after passing it through a bottleneck. The right panel shows the neural network that predicts precipitation. The input to this network is the large-scale variables (as for the baseline network) as well as org extracted from the auto-encoder. The two blocks are trained simultaneously.

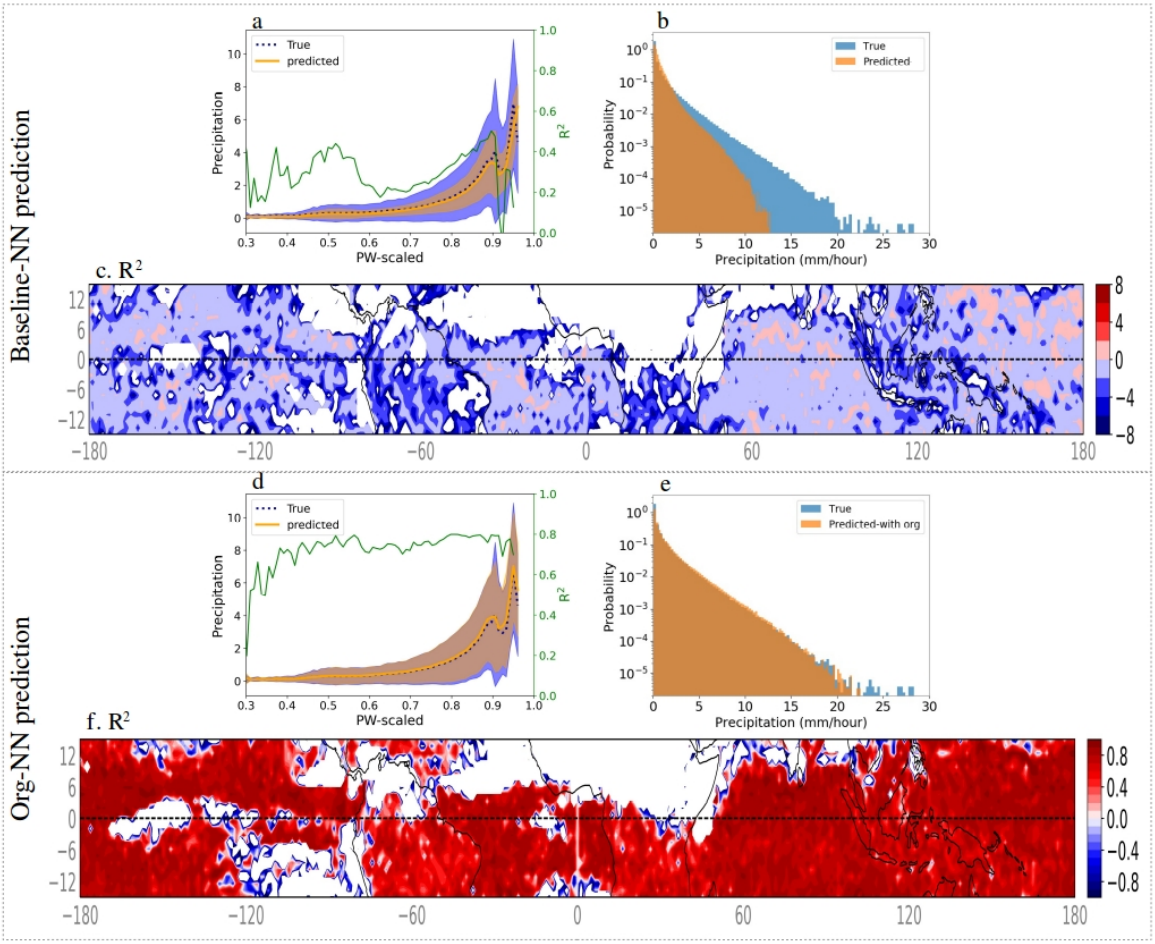

Figure 3. Performance of the NN (a-c) trained using only large scale variables, and (d-f) trained using large scale variables as well as *org* metric as input. a shows true (blue) and predicted (orange) precipitation versus bins of PW (1 mm) for training using only PW, SST, specific humidity, and temperature near the surface. Shading shows the standard deviation of precipitation. The green line plots the R^2 across PW bins. Panel b shows the pdf (probability density function) of precipitation for true (blue) and prediction from panel a (orange). Panel c displays the R^2 computed for each latitude and longitude location across time steps for panel a. Panel d-f show the same as a-c but predicted as by Org-NN, which includes the *org* metric in its inputs. Note that the color-bars in panel c and f have different ranges.

are robust to the choice of coarse-graining factor (not shown) and to NN complexity. This indicates that either the relevant information for an accurate prediction and replication of statistics is missing or that the variability in precipitation is totally stochastic and thus unpredictable. We investigate the next hypothesis, which is that precipitation variability is explainable by organization in the next section.

Org informed prediction of precipitation

One probable reason for the failure of the baseline-NN to predict precipitation variability might be due to the lack of information on sub-grid scale variability not present in coarse-scale variables so that additional sub-grid scale information needs to be included in the inputs of the NN. Here we specifically investigate the potential importance of sub-grid scale cloud patterns, i.e organization, and whether including this information in the neural network can improve its prediction. Over the last three decades, many studies have investigated the impact of organization on the dynamics and thermodynamics of convection as well as on precipitation. In parallel, more than 20 different metrics have been developed to quantify the degree of organization or more generally the cloud pattern[27]. Figure S1 shows the correlation between the time series of different metrics. These metrics capture some aspects of the organization and can potentially be used to inform the large-scale variables about subgrid-scale cloud patterns for predicting precipitation. However, these metrics have been designed for large domains ($>$ 200 km) while for climate-size domains (100-200 km) where convection might cluster to some degree, these metrics usually fail to provide

valuable information. Furthermore, each of these metrics targets a specific aspect of cloud organization that may not be directly relevant for predicting precipitation.

We instead turn this approach on its head and try to learn the implicit representation of organization most relevant to precipitation prediction. To do so we extract information relevant for predicting precipitation from a high-resolution field using a dimensionality reduction technique known as an auto-encoder. An auto-encoder (AE) is a powerful non-linear dimensionality reduction approach that has been originally developed for image processing[28]. The encoder part of the AE projects high-resolution inputs into a low dimension non-linear manifold that efficiently describes the data. This low dimension representation, i.e. a latent space, embeds the optimal information needed from the high-resolution field to reconstruct precipitation. Here we apply this AE to high-resolution PW and extract its latent representation (hereafter named *org*) to inform our precipitation predicting network about the organization. By coupling the AE with the NN which predicts precipitation, and learning the two networks in parallel, the AE directly receives feedback from the objective function of the NN through back propagation (the process of minimizing the loss function by adjusting the weights and biases of the NN). Thus, the AE is forced to extract relevant information that improves the prediction of precipitation. Figure 2.c shows the architecture of our network. We refer to this network as Org-NN. Importantly, we impose translation and rotation invariance of the loss function measuring the organization of the high-resolution field, as the organization of clouds should not depend on its exact location nor on the rotation of the field (See Methods). The left block of the schematic shows the auto-encoder that receives PW as input. The right block shows the fully connected feed-forward NN, with the same number of hidden layers and neurons as baseline-NN. This network receives the large-scale variables along with *org* and predicts precipitation. These two blocks are trained simultaneously. In other words, we train two networks end-to-end in parallel, one that predicts precipitation with SST, PW and 2-meter temperature and humidity (as previously) with the addition of organization latent variables *org*, and the other one that reduces the dimension of the high-resolution field into a few latent variables, the *org* variables. The dimension of *org* is set to 4. Further tests with different *org* dimensions are discussed in the method.

Figure 3.d-f shows the Org-NN prediction for PW. Org-NN demonstrates a significant improvement as compared to the baseline-NN. The R^2 of this prediction increases to 0.8 when computed across all data point. Reducing the number of org variables to 2 or changing the resolution does not significantly change the results as summarized in the table S2 and figure S3 in the supplementary material. R^2, computed for each bin of PW, is close to 0.80 for almost all bins except where precipitation is small (e.g. scaled-PW ~ 0.3).

We further quantify the Org-NN performance by comparing its probability density function (PDF) with the one of the true precipitation from the storm-resolving model(Figure 3.e). Org-NN fully captures the PDF including the tail of the distribution, which corresponds to the precipitation extremes. In climate models, hourly and sub-hourly precipitation extremes, which are dominated by deep convective precipitation, continue to be one substantial source of uncertainty.

Figure 3.f shows the R^2 of the Org-NN computed for each latitude-longitude grid across time steps. The white patches in this figure have precipitation smaller than 0.05 mm/hour so they are excluded from the input of the model. Org-NN has significantly larger R^2 (>0.8) except for regions that are close to the point which does not have precipitation larger than the threshold.

Prediction of the Org variables

Including *org* variables in the input of the neural network significantly improves the prediction of precipitation; however, in climate models, high-resolution fields are not available to be used for the prediction of *org*. One trivial solution for this problem would be to predict *org* variables from large-scale variables that are resolved by the climate model. Rephrased differently, can we apply a diagnostic model using coarse-scale variables as input to predict *org*? We examine this assumption using a neural network that receives large-scale quantities as input and predicts *org* variables. We find that, not surprisingly, *org* can not be predicted using large-scale quantities, further emphasizing its subgrid scale nature. In addition, if the *org* variables were predictable from only large-scale variables, the baseline-NN should have demonstrated greater skill (as discussed above).

From observations, we know that organization exhibits strong spatial and temporal memory. This means that *org* variables should carry some sort of memory that could be advected, potentially with the mean field. Mapes (2011) integrates this idea in his *org* metric by making it time-dependent and advective. We adopt the same idea and include *org* variables of previous time steps to predict the one for the current time. Mathematically this is expressed as:

$$org(t) = f(X_{lg}(t - mdt, t), org(t - ndt, t - dt), \theta) \tag{1}$$

where $org(t)$ is the latent variable at time t, f is a function approximated by a neural network, X_{lg} is large-scale variables, θ is the neural network parameters. dt is the time-frequency of the outputs and equal to 15 minutes for the DYAMOND data. m and n is the number of previous time steps that we want to include. To summarize, we predict org as a function of X_{lg} of the same and previous time steps and *org* of previous time steps.

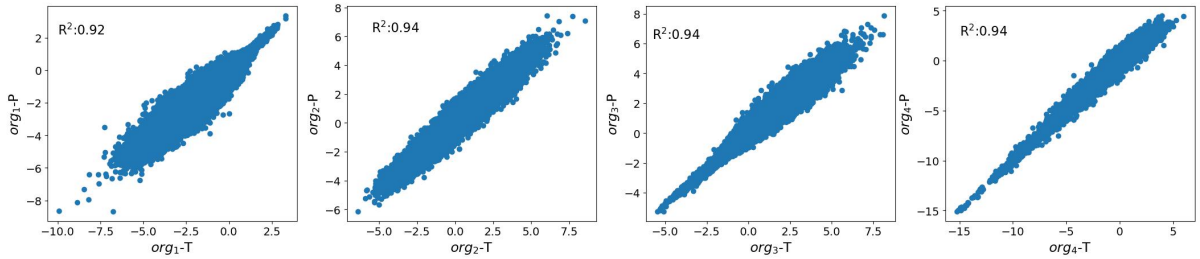

Figure 4. Predictability of *org* variables The figure shows the prediction of the 4 *org* variables using a neural network that receives large scale variables as well as *org* from the previous 4 times steps (Equation 1 with n = 4). The subscript P denotes the predictions by simple neural network and T denotes the true values extracted from the auto-encoder.

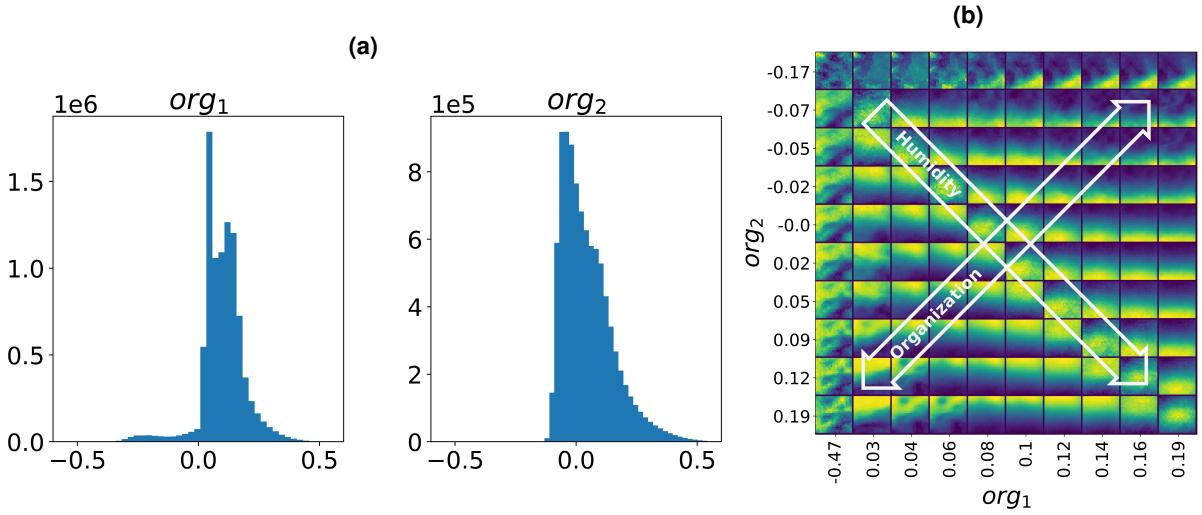

Figure 5. Latent space of *org* a) Histogram of org_1 and org_2 for the $n_{org} = 2$ case from predictions of the encoder trained on the DYAMOND simulations. b) Visualization of the latent representation for the $n_{org} = 2$ case. Color scaling is relative to the minimum and maximum values for PW in each reconstructed high resolution field to best show small scale contrast. Figure S4 shows the same figure, but with the color scale relative to the absolute contrast across all the reconstructed high resolution fields.

We find that setting m=0, making *org* dependent on only current-time large-scale variables, and n = 4 improves the prediction of *org* significantly (Figure 4), while setting n to 2 reduces the performance of the network (not shown). This indicates the importance of memory in subgrid-scale structure.

What does *org* measure?

Previous *org* metrics have been developed using statistical and geometric arguments[27], but here we have relied on a data-driven approach to develop an optimal metric for organization that is specifically targeting the predictability of precipitation at the grid-scale. Because neural networks have non-linear activation functions between layers, they are more challenging to interpret than linear models. To understand what the encoder is learning, we additionally train a decoder to reproduce the high-resolution fields from the latent representations. As discussed above, this decoder using a rotation-invariant loss function[29]; this imposes a constraint such that high resolution fields that are rotations of one another (by 90°, 180°, or 270°) are mapped to the same latent representation, as rotations should not be physically meaningful for predictions of precipitation.

To directly visualize the latent space, we first determine the observed distributions of the *org* parameters learned from the simulation data using the encoder. We denote these latent variables as org_1 and org_2 for $n_{org} = 2$, and org_1, org_2, org_3, and org_4 for $n_{org} = 4$. Figure 5a shows the distributions for the $n_{org} = 2$ case, and Figure S5 shows the distributions for $n_{org} = 4$ case. We then use the trained decoder to reconstruct high resolution fields sampled across the distribution of values observed in the training data set for the latent variables, effectively mapping out the latent space across the range of *org* parameters observed in the DYAMOND simulations. In the $n_{org} = 2$ case, the *org* values input to the decoder are the average values for each decile of the distributions of org_1 and org_2 obtained by encoding the high-resolution fields from the DYAMOND simulations to produce

a reconstruction that is characteristic of that portion of the latent space. Figure 5b shows the latent space visualization for the $n_{org} = 2$ case. For the $n_{org} = 4$ case, to visualize the latent representation, the *org* values input to the decoder are the average values for quantiles of six equal portions of the observed distributions. Visualization of the latent space for the $n_{org} = 4$ case in shown in the Supplemental Materials (Figure S6).

Mapping out the latent representations indicates that the learned *org* parameters are able to not only capture information about the mean values and variance of the high resolution fields, but also can capture specific details of its heterogeneity (in the $n_{org} = 2$ case). The *org* parameter is able to capture both the distribution of the high resolution PW field and the relative contrast between different areas (Figure 5b). In the $n_{org} = 4$ case, the *org* parameter additionally provides information about specific patterns and their relative translations and contrast (Figure S6).

Discussion

The results presented here suggest that coarse-scale variables (at a scale of climate models $\sim$100 km) are not sufficient predictors for accurate replication of precipitation statistics. This finding questions current climate model parameterizations of convection and precipitation that ignore any degree and mode of organization at the subgrid scale. Current climate model representation of convection is of an ensemble of unorganized, randomly distributed buoyant updrafts, without any interaction nor memory of the previous state of the system. This deficiency has been suggested as one major reason of well known climate model bias towards light rain[13,19,30], i.e. the so-called drizzle problem.

We have shown here that including convective organization subgrid scale information (in a neural network) significantly improves the prediction skill of precipitation, particularly the extremes of precipitation– a long-lasting issue for climate models. This finding suggests that precipitation stochasticity at climate model scale is linked to subgrid-scale structure and can be fully explained when this information is incorporated into the parameterization.

Our data-driven organization variables, *org*, extracted from high resolution field of PW were shown to carry the required information needed to accurately predict precipitation at the coarse scale. In addition, these organization metrics were shown to be predictable as a memory process informed by coarse-scale variables and thus can be implemented in coarse-scale climate models. The studies of self aggregation using small domain Cloud Resolving Models have emphasized the importance of memory for the survival of organization. In that case the memory has been attributed to moisture-convection feedback, moisture-radiation feedback, etc[9,10,31].

Further research on data-driven parameterizations of precipitation must answer the technical challenges such as online implementation of org-NN. While here we have found that using an *org* parameter with a size of $n_{org} = 4$ leads to the most improvement in the predictability of the precipitation stochasticity, $n_{org} = 2$ already demonstrates significant improvements over the baseline case. It also has the advantage of being more readily interpretable as capturing variability in organization and in the degree of relative humidity present in the high resolution fields. Including fewer additional predictors is preferable for developing future convective parameterizations that include *org*.

Our finding helps to guide and improve the representation of organization in climate models, one of the Grand Challenges on "Clouds, Circulation and Climate Sensitivity"[32]. We provide a pragmatic way to tackle two problems at once: the representation of cloud organisation and precipitation stochasticity in coarse-scale models. We also emphasize the key role of organization and its predictability on precipitation prediction.

Methods

Data and preprocessing

The dataset was produced using the System for Atmospheric Modeling (SAM)[24], as a part the DYnamics of the Atmosphere general circulation Modeled on Non-hydrostatic Domains (DYAMOND) Phase 2 Inter-comparison Project of global storm-resolving models. The Phase 2 Inter-comparison project simulated 40 days during the northern-hemisphere wintertime, with the initial 10 days used as model spin-up and the last 30 days used for evaluation[25]. The spatial resolution of SAM-DYAMOND is about 4 km and the temporal resolution of 2D outputs is 15 minutes.

To prepare the data for our neural network we select the tropical band (20S-20N) and 10 days of simulation. We coarse grain these data to GCM-size grids ($\sim$100 km) as follows:

$$\overline{X}(i,j,k) = \frac{1}{L^2} \sum_{l=L(i-1)+1}^{l=Li} \sum_{m=L(j-1)+1}^{m=Lj} X(l,m,k), \tag{2}$$

where X is the field to be coarse grained, L is the averaging (coarse graining) factor, and i and j are the indices in the x and y directions. We test two coarse graining factors, L = 32 and L = 24. These two L correspond to resolutions of 130 km and 100 km, respectively. The results shown in the paper are for L = 24 except otherwise mentioned. The results are robust to changing L.

To provide the train, validation, and test data sets, we split the 10 days into 6, 2, and 2 days for train-validation and test, respectively. Furthermore, we only keep the samples with precipitation larger than a threshold (0.04 mm/hour) so that we concentrate only on predicting precipitation intensity rather than on precipitation triggering. The total number of samples is on the order of 10^8.

Neural network architecture

We use two neural networks: a feed-forward network that is informed with only large scale variables, which we refer to as the baseline NN. We use a second, novel architecture, which combines a feed-forward neural network with an auto-encoder. The auto-encoder extracts organization information from the high resolution PW field. Figure 2 shows these two networks. Both networks are implemented using the Tensorflow library version 2.7[33] and the hyper-parameters are tuned using the Sherpa hyperparameter tuning library library[34].

Baseline-NN

The baseline NN (2.a) is a fully connected feed-forward neural network with 4 hidden layers and [256,256,128,64] neurons. The learning rate is scheduled to decrease with epoch and initialized to 10^{-4}. The baseline NN has access to only large-scale variables and predicts precipitation. We run four tests in which the choice of input of the baseline NN differs and is as follow:

1. [PW, SST]
2. [PW, SST, specific humidity at 2m, temperature at 2 m]
3. [PW, SST, specific humidity at 2m, temperature at 2 m, total cloud cover]
4. [PW, SST, specific humidity at 2m, temperature at 2 m, total cloud cover, CAPE, CIN]

The baseline NN outputs expected precipitation as well as its standard deviation for each given sample.

Org-NN

The architecture of Org-NN is shown in Figure 2.b. The encoder part of the auto-encoder includes 3 one-dimensional convolutional layers followed by two fully connected layers. The input to this network is high resolution PW fields with dimension 24x24 grid points (or 100x100 km^2). The encoder output is *org* variables. The dimension of *org* is a hyper parameter of the network that we set to 4. The decoder has the inverse structure of the encoder. It receives *org* variables and reconstructs the original high resolution field (e.g. 24x24 grid points). The NN part of this network is similar to the baseline- NN except that the organization latent space, *org*, has been added to its input. The input to NN is [PW, SST, specific humidity at 2m, temperature at 2m, *org*]. Table S2 and Figure S3 summarizes how changing resolution (e.g. 32x32, 16x16 etc) and/or reducing *org* dimension affects the results. Changing the dimension of *org* to 2 does not affect the prediction of precipitation; however, it reduces the reconstruction accuracy.

Probabilistic loss function

Most regression neural networks use mean square error (MSE) as a loss function and, for a given input, predict a single target value. In our case, this means that given the large-scale variables we would predict a single value for precipitation. However, given the stochastic nature of precipitation, we here prefer to predict a probability distribution function for precipitation. Thus for each given observed variable (e.g. large scale condition) we predict an expected precipitation and its standard deviation, assuming a Gaussian distribution, following the strategy from Guillaumin (2021)[35], as follows:

$$P(\xi|\psi;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(\xi-\mu(\psi))^2}{2\sigma^2}\} \tag{3}$$

The log-likelihood then becomes:

$$l(\xi|\psi;\mu,\sigma) = -\frac{1}{2} log 2\pi\sigma^2 - \frac{(\psi-\mu(\psi,\theta))^2}{2\sigma^2} \tag{4}$$

We predict the mean, μ, and the standard deviation, σ, using a neural network with parameter θ:

$$l(\theta) = \sum_{i=1}^{n} \{-\frac{1}{2} log 2\pi f_2(\psi_i,\theta)^2 - \frac{(\psi-f_1(\psi_i,\theta))^2}{2f_2(\psi_i,\theta)^2}\} \tag{5}$$

where f_1 is the mean and f_2 the standard deviation predicted by the neural network. θ is the parameters of the network and ψ is the observed variables, in our case the large scale variables. The loss function is then defined as :

$$L(P,\hat{P}(\theta)) = -l(\theta) \tag{6}$$

Assuming a constant standard deviation in Equation 4 reduces the loss to mean square error.

For the auto-encoder, we use the mean square error (MSE) as loss. The emphasis of our work is on predicting precipitation and extracting *org* variables, but not accurately reconstructing the high resolution two-dimensional fields. Thus we give the MSE of the auto-encoder a smaller weight (0.3) so that org-NN is more concentrated on its principal task.

An organization metric should not be sensitive to the orientation of clouds. In other word it should be rotation invariant. A rotation invariant model maps an image and its rotated version to the same latent variables. We force this criteria in the loss function by adding the following term:

$$org = e(X_{NR}) \qquad org_R = e(X_R) \qquad loss_{RI} = (org - org_R)^2 \tag{7}$$

where *org* is the latent variables of auto-encoder, e is the encoder network, X_{NR} is the non-rotated high resolution input of the auto-encoder, and X_R is the rotated input.

Training and validation procedure

The input to each network is in the form of mini-batches so that we do not train on individual samples one by one, but rather on an ensemble of samples. Thus the shape of inputs to the baseline-NN and NN part of Org-NN are [n_{batch},n_f], where n_{batch} is the number of samples in each mini-batch and n_f is the number of field that we pass as input. For example, for the baseline NN where we pass PW and SST as input n_f equals 2. The shape of the input for the AE is [n_{batch},L_x,L_x,1] and L_x is the dimension of the high resolution field. We examine our model on 4 resolutions: 200 km, 130 km, 100 km, and 75 km that corresponds to L_x = 48, 32, 24, and 16 respectively. The model is trained on the mini batches of 128 samples, for 50 epochs. In order to prevent overfitting we implement early stopping with patience of 5 epochs. At each iteration, the network computes the loss averaged over samples in one mini-batch. This loss value is back propagated through the network and its derivative with respect to each neural network's parameter is computed. The neural network's parameters are then updated using the ADAM algorithm. This process is repeated over all mini-batches, which corresponds to one epoch. At the end of each epoch, the network's performance is validated using the validation data-set, which the network has not seen while being trained. The training-validation process continues until either the total number of epochs are reached or until the early stopping criteria is met. Here, the early stopping has patience equal to 5 epoch. This means that if the validation loss does not improve for five consecutive epochs, the network training stops. Early stopping was used to prevent the network's overfitting.

References

1. Allen, M. R. & Ingram, W. J. Constraints on future changes in climate and the hydrologic cycle. *Nature* **419**, 224—232, DOI: 10.1038/nature01092 (2002).
2. Neelin, J. D., Peters, O. & Hales, K. The transition to strong convection. *J. Atmospheric Sci.* **66**, 2367 – 2384, DOI: 10.1175/2009JAS2962.1 (2009).
3. Stephens, G. L. *et al.* Dreary state of precipitation in global models. *J. Geophys. Res. Atmospheres* **115**, DOI: https://doi.org/10.1029/2010JD014532 (2010). https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010JD014532.
4. Rasp, S., Pritchard, M. S. & Gentine, P. Deep learning to represent subgrid processes in climate models. *Proc. Natl. Acad. Sci.* **115**, 9684–9689 (2018).
5. Gettelman, A. *et al.* Machine learning the warm rain process. *J. Adv. Model. Earth Syst.* **13**, e2020MS002268, DOI: https://doi.org/10.1029/2020MS002268 (2021). E2020MS002268 2020MS002268, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020MS002268.
6. Arakawa, A. & Schubert, W. H. Interaction of a Cumulus Cloud Ensemble with the Large-Scale Environment, Part I. *J. Atmospheric Sci.* **31**, 674–701, DOI: 10.1175/1520-0469(1974)031\T1\textless{}0674:IOACCE\T1\textgreater{}2.0.CO;2 (1974).
7. Pincus, R., Barker, H. W. & Morcrette, J.-J. A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields. *J. Geophys. Res. Atmospheres* **108** (2003).
8. Holloway, C. E. *et al.* Observing convective aggregation. *Surv. Geophys.* **38**, 1199 – 1236, DOI: 10.1007/s10712-017-9419-1 (2017).
9. Muller, C. *et al.* Spontaneous aggregation of convective storms. *Annu. Rev. Fluid Mech.* **54**, 133–157, DOI: 10.1146/annurev-fluid-022421-011319 (2022). https://doi.org/10.1146/annurev-fluid-022421-011319.
10. Tompkins, A. M. Organization of tropical convection in low vertical wind shears: The role of water vapor. *J. Atmospheric Sci.* **58**, 529 – 545, DOI: 10.1175/1520-0469(2001)058<0529:OOTCIL>2.0.CO;2 (2001).
11. Da Silva, N. A., Muller, C., Shamekh, S. & Fildier, B. Significant amplification of instantaneous extreme precipitation with convective self-aggregation. *J. Adv. Model. Earth Syst.* **13**, e2021MS002607, DOI: https://doi.org/10.1029/2021MS002607 (2021). E2021MS002607 2021MS002607, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2021MS002607.
12. Li, R. L., Studholme, J. H. P., Fedorov, A. V. & Storelvmo, T. Precipitation efficiency constraint on climate change. *Nat. Clim. Chang.* **12**, 642–648, DOI: 10.1038/s41558-022-01400-x (2022).
13. Moseley, C., Hohenegger, C., Berg, P. & Haerter, J. O. Intensification of convective extremes driven by cloud–cloud interaction. *Nat. Geosci.* **9**, 748–752, DOI: 10.1038/ngeo2789 (2016).
14. Schlemmer, L. & Hohenegger, C. The formation of wider and deeper clouds as a result of cold-pool dynamics. *J. Atmospheric Sci.* **71**, 2842 – 2858, DOI: 10.1175/JAS-D-13-0170.1 (2014).
15. Schlemmer, L. & Hohenegger, C. Modifications of the atmospheric moisture field as a result of cold-pool dynamics. *Q. J. Royal Meteorol. Soc.* **142**, 30–42, DOI: https://doi.org/10.1002/qj.2625 (2016). https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2625.
16. Tompkins, A. M. Organization of tropical convection in low vertical wind shears: The role of cold pools. *J. Atmospheric Sci.* **58**, 1650 – 1672, DOI: 10.1175/1520-0469(2001)058<1650:OOTCIL>2.0.CO;2 (2001).
17. Haerter, J. O. Convective self-aggregation as a cold pool-driven critical phenomenon. *Geophys. Res. Lett.* **46**, 4017–4028, DOI: https://doi.org/10.1029/2018GL081817 (2019). https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018GL081817.
18. Haerter, J. O., Meyer, B. & Nissen, S. B. Diurnal self-aggregation. *npj Clim. Atmospheric Sci.* **3**, 30, DOI: 10.1038/s41612-020-00132-z (2020).
19. Mapes, B. & Neale, R. Parameterizing convective organization to escape the entrainment dilemma. *J. Adv. Model. Earth Syst.* **3**, DOI: https://doi.org/10.1029/2011MS000042 (2011). https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2011MS000042.
20. Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G. & Yacalis, G. Could machine learning break the convection parameterization deadlock? *Geophys. Res. Lett.* **45**, 5742–5751 (2018).
21. Brenowitz, N. D. & Bretherton, C. S. Spatially extended tests of a neural network parametrization trained by coarse-graining. *J. Adv. Model. Earth Syst.* **11**, 2728–2744 (2019).

22. Yuval, J. & O'Gorman, P. A. Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nat. communications* **11**, 1–10 (2020).

23. Mooers, G. *et al.* Assessing the potential of deep learning for emulating cloud superparameterization in climate models with real-geography boundary conditions. *J. Adv. Model. Earth Syst.* **13**, e2020MS002385, DOI: https://doi.org/10.1029/2020MS002385 (2021). E2020MS002385 2020MS002385, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020MS002385.

24. Khairoutdinov, M. F. & Randall, D. A. Cloud resolving modeling of the arm summer 1997 iop: Model formulation, results, uncertainties, and sensitivities. *J. Atmospheric Sci.* **60**, 607–625, DOI: 10.1175/1520-0469(2003)060<0607:CRMOTA>2.0.CO;2 (2003).

25. Stevens, B. *et al.* Dyamond: the dynamics of the atmospheric general circulation modeled on non-hydrostatic domains. *Prog. Earth Planet. Sci.* **6**, 61, DOI: 10.1186/s40645-019-0304-z (2019).

26. Muller, C. J., O'Gorman, P. A. & Back, L. E. Intensification of precipitation extremes with warming in a cloud-resolving model. *J. Clim.* **24**, 2784–2800 (2011).

27. Janssens, M. *et al.* Cloud patterns in the trades have four interpretable dimensions. *Geophys. Res. Lett.* **48**, e2020GL091001, DOI: https://doi.org/10.1029/2020GL091001 (2021). E2020GL091001 2020GL091001, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL091001.

28. Lee, K. & Carlberg, K. T. Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *J. Comput. Phys.* **404**, 108973, DOI: https://doi.org/10.1016/j.jcp.2019.108973 (2020).

29. Kurihana, T., Moyer, E., Willett, R., Gilton, D. & Foster, I. Data-driven cloud clustering via a rotationally invariant autoencoder. *IEEE Transactions on Geosci. Remote. Sens.* **60**, 1–25 (2021).

30. Tan, J., Jakob, C., Rossow, W. B. & Tselioudis, G. Increases in tropical rainfall driven by changes in frequency of organized deep convection. *Nature* **519**, 451–454 (2015).

31. Colin, M., Sherwood, S., Geoffroy, O., Bony, S. & Fuchs, D. Identifying the sources of convective memory in cloud-resolving simulations. *J. Atmospheric Sci.* **76**, 947–962 (2019).

32. Bony, S. *et al.* Clouds, circulation and climate sensitivity. *Nat. Geosci.* **8**, 261–268, DOI: 10.1038/ngeo2398 (2015).

33. Abadi, M. *et al.* {TensorFlow}: a system for {Large-Scale} machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, 265–283 (2016).

34. Hertel, L., Collado, J., Sadowski, P., Ott, J. & Baldi, P. Sherpa: Robust hyperparameter optimization for machine learning. *SoftwareX* **12**, 100591 (2020).

35. Guillaumin, A. P. & Zanna, L. Stochastic-deep learning parameterization of ocean momentum forcing. *J. Adv. Model. Earth Syst.* **13**, e2021MS002534 (2021).

Acknowledgements

SS and PG acknowledge funding from European Research council grant USMILE, from Schmidt Future project M2LiNES and from the National Science Foundation Science and Technology Center (STC) Learning the Earth with Artificial intelligence and Physics (LEAP). KDL acknowledges support from LEAP and DOE Grant DE-SC0022323 "Discovering Physically Meaningful Structures from Climate Extreme Data".

Supplementary Materials

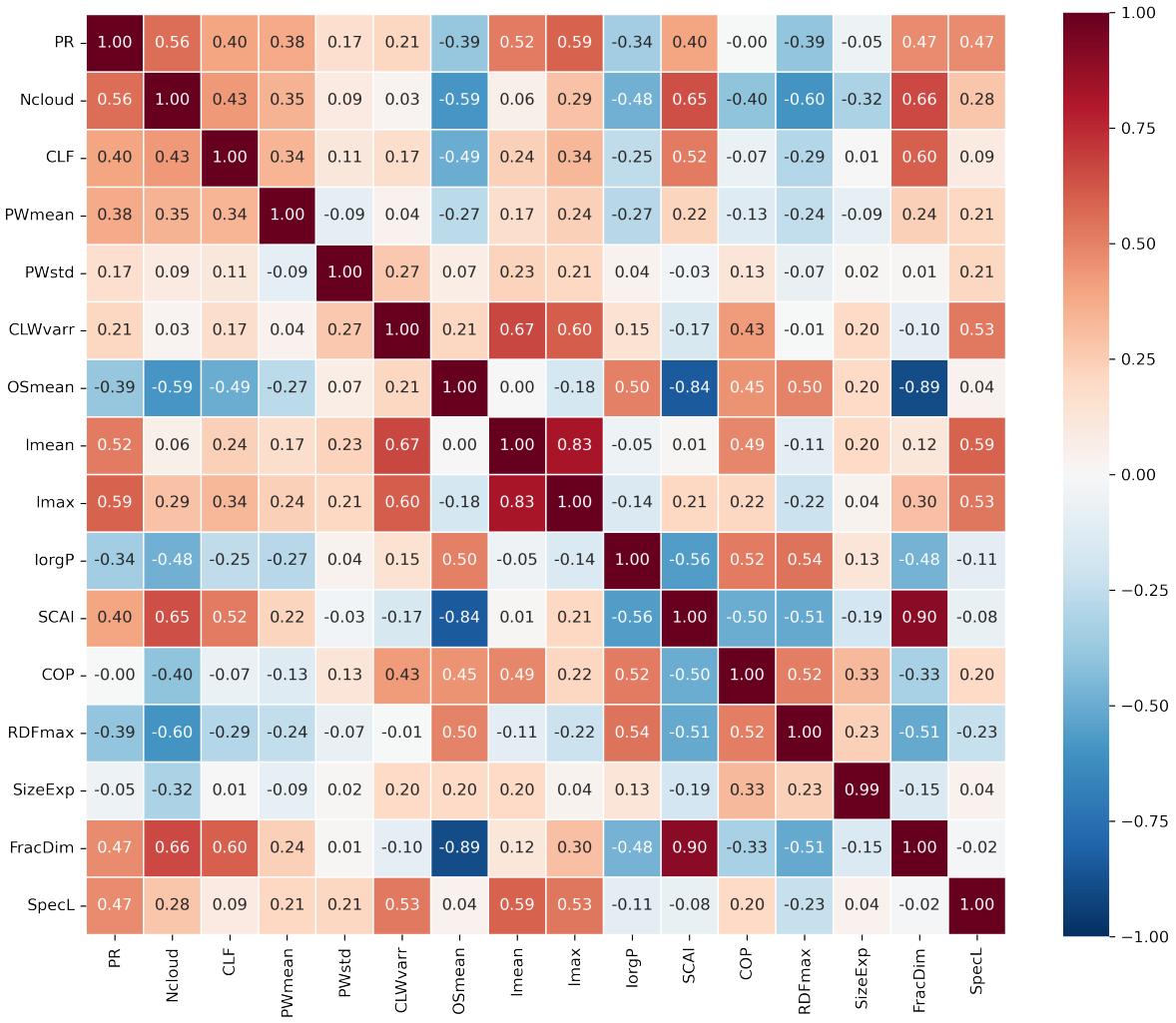

Figure S1. The heat-map shows pairwise correlations between pairs of previously defined organization metrics collected by Janssens et al. (2021)[27]. Simple descriptions for the metrics are provided with Table S1. Organization metrics have been measured for domains of 10x10 degree every 3 hours.

Table S1. Description of organization metrics used in Figure S1. Detailed definitions are provided in Janssens et al. (2021)[27]

Category	Name	Description
Statistical properties	PR	precipitation rate
	Ncloud	cloud object number
	CLF	cloud occurrence frequency
	PWmean	mean precipitable water
	PWstd	standard deviation of precipitable water
	CLWvarr	cloud liquid water variance ratio
Object-based metrics	OSmean	mean rectangular contiguous clear-sky area
	lmean	mean cloud object size
	lmax	max cloud object size
	IorgP	organization index
	SCAI	Simple Convective Aggregation Index
	COP	Convective Organisation Potential
	RDFmax	maximum value of the cloud object radial distribution function
Scale decomposition metrics	SizeExp	exponent of cloud size distribution (power law fit)
	FracDim	box-counting dimension of cloud boundaries
	SpecL	spectral length scale

Table S2. The impact of resolution and *org* dimension on the prediction of precipitation by the Org-NN model

Resolution (km^2)	grids	*org* size	R^2
70x70	16x16	4	0.82
100x100	24x24	4	0.82
100x100	24x24	2	0.80
135x135	32x32	4	0.81
135x135	32x32	2	0.79
200x200	48x48	4	0.78

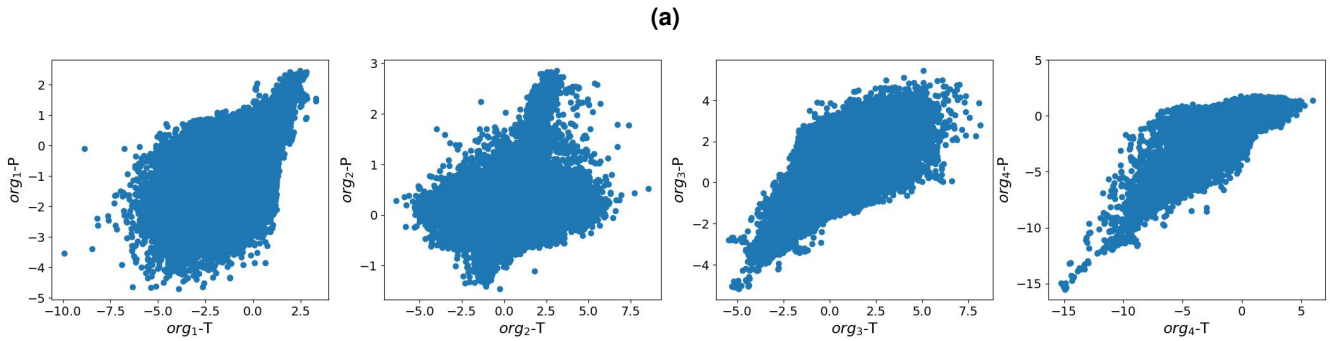

Figure S2. Prediction of *org* variables: Panel is Similar to Figure 4 but with only large scale variables as input.

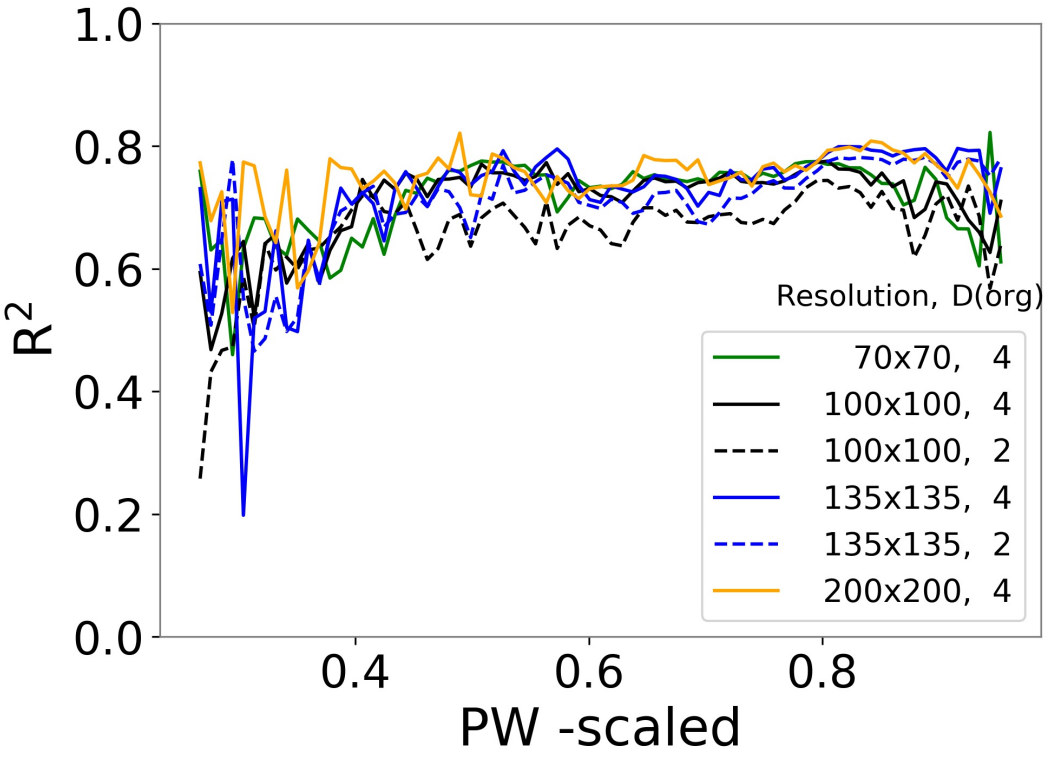

Figure S3. Panel shows the R^2 across PW bins for org-NN prediction trained on different resolutions of the inputs and *org* dimension.

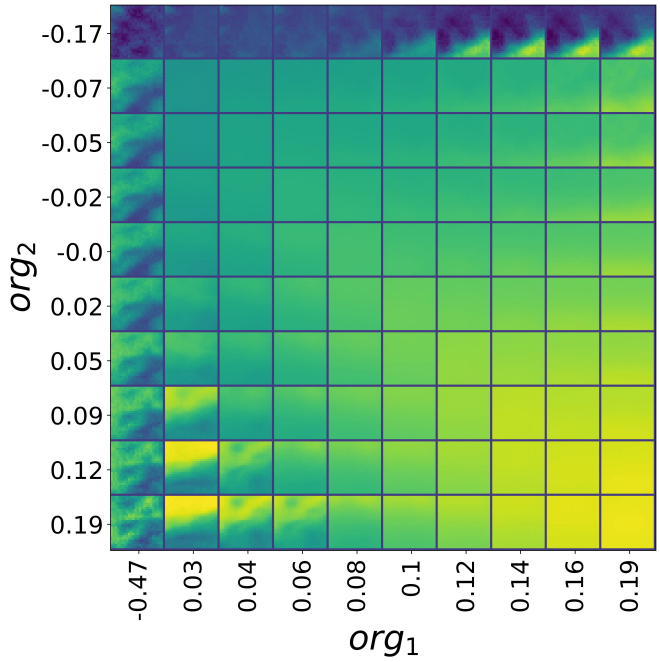

Figure S4. As in Figure 5b, but color scaling for PW is relative to the maximum and minimum across all the reconstructed high resolution fields, rather than each individual field.

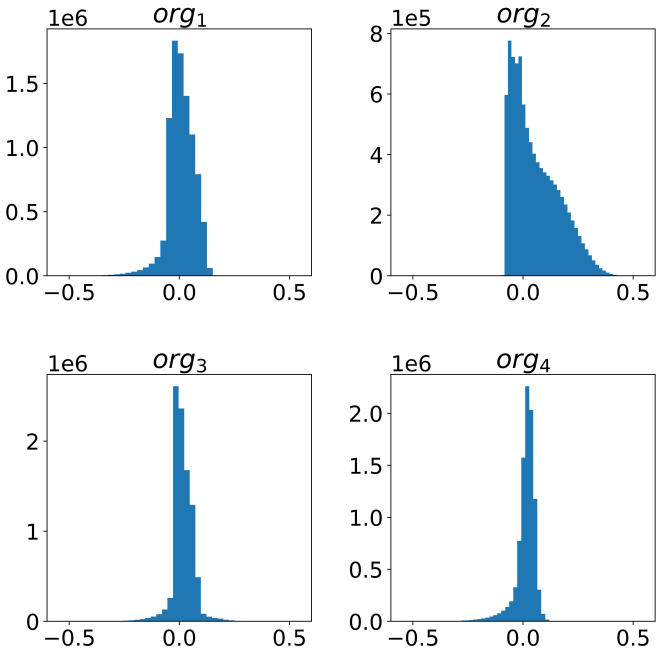

Figure S5. Histograms of *org* parameters from the DYAMOND simulations for the $n_{org} = 4$ case.

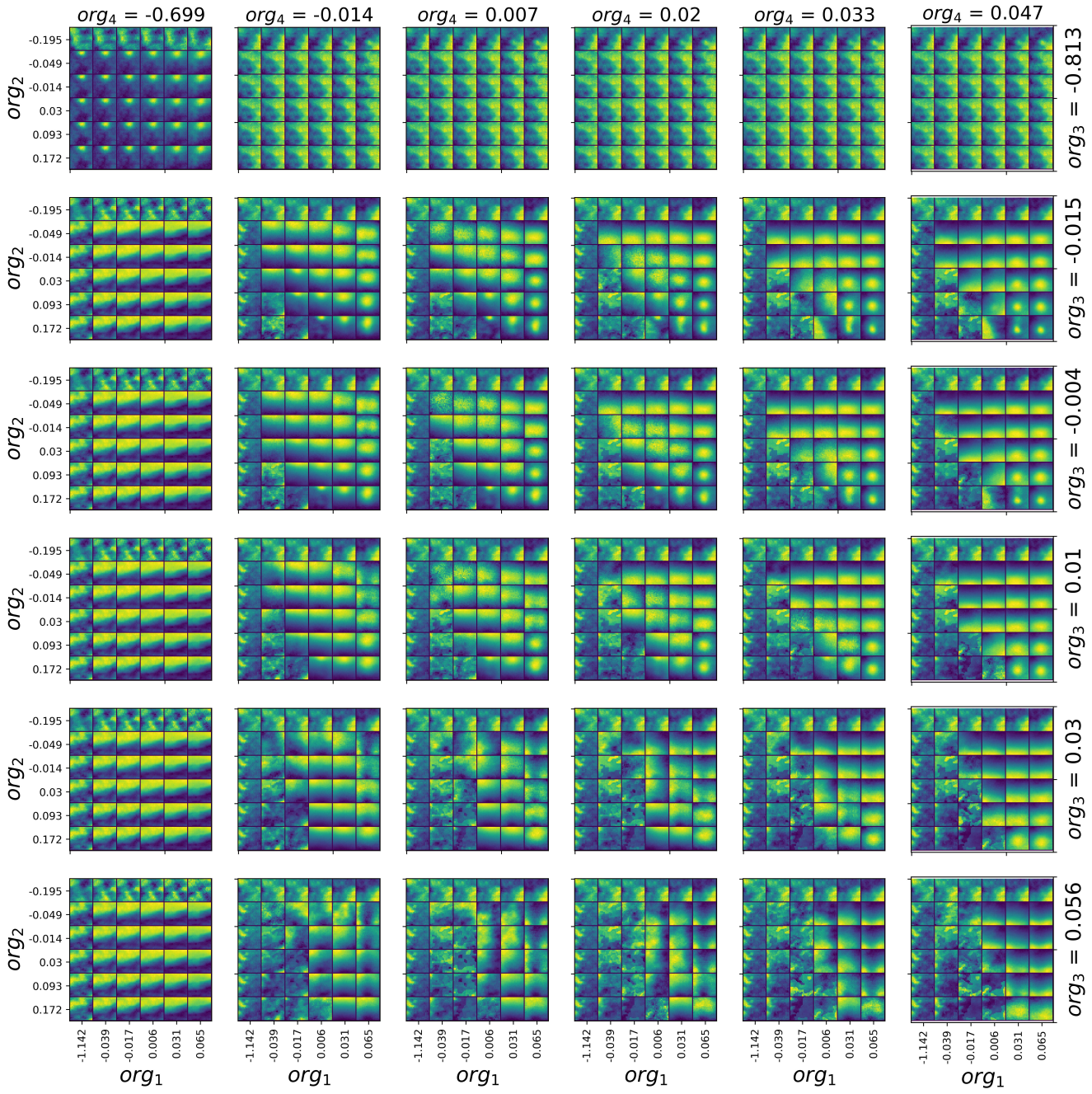

Figure S6. Visualization of the latent space for the $n_{org} = 4$ case. Each subplot shows the reconstructed high resolution fields of PW across the distribution of values for org_1 and org_2, with the values for org_3 and org_4 held fixed. Each row of plots uses the value for a different quantile of the distribution of org_3 (as indicated to the right of that row), while each column of plots uses values for different quantiles of the distribution of org_4 (as indicated in the column headers).