Supplemental Material

Journal of the Atmospheric Sciences
Data-Driven Modeling of Stratiform and Convective Rain Area
https://doi.org/10.1175/JAS-D-25-0178.1

Data-driven Modeling of Stratiform and Convective Rain Area

Sara Shamekh, Pedro Angulo-Umana, Paul O'Gorman

February 17, 2026

A Additional Methods

A.1 Conditional Normalizing Flow: Probabilistic Formulation

This section summarizes the probabilistic formulation underlying the conditional normalizing flow (CNF) used in this study. The CNF is employed to approximate the conditional distribution

$$p(\mathbf{A} \mid \mathbf{X}_{\mathrm{LS}}),$$

where $\mathbf{A} = (A_{\mathrm{shallow}}, A_{\mathrm{deep}}, A_{\mathrm{stratiform}}) \in \mathbb{R}^3$ denotes the observed rain-area fractions and $\mathbf{X}_{\mathrm{LS}} \in \mathbb{R}^{12}$ represents the large-scale environmental predictors.

The CNF introduces a latent variable $\mathbf{z} \in \mathbb{R}^3$ drawn from a simple base distribution, chosen here as a standard multivariate normal,

$$p_{\mathbf{Z}}(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

An invertible and differentiable transformation

$$\mathbf{z} = f_\phi(\mathbf{A}, \mathbf{X}_{\mathrm{LS}}), \qquad \mathbf{A} = f_\phi^{-1}(\mathbf{z}, \mathbf{X}_{\mathrm{LS}}),$$

maps between the rain-area space and the latent space, with conditioning on $\mathbf{X}_{\mathrm{LS}}$ entering exclusively through the transformation.

This construction permits exact evaluation of the conditional density via the change-of-variables formula,

$$\log p_\phi(\mathbf{A} \mid \mathbf{X}_{\mathrm{LS}}) = \log p_{\mathbf{Z}}(f_\phi(\mathbf{A}, \mathbf{X}_{\mathrm{LS}})) + \log \left| \det \left(\frac{\partial f_\phi(\mathbf{A}, \mathbf{X}_{\mathrm{LS}})}{\partial \mathbf{A}} \right) \right|,$$

which defines the objective optimized during training.

A2. CNF Architecture and Training Details

The transformation f_ϕ is parameterized as a stack of conditional masked autoregressive flow (MAF) layers [2]. MAF exploits an autoregressive factorization over the components of $\mathbf{A}$, allowing flexible modeling of joint dependencies while retaining a tractable Jacobian determinant.

The CNF used in this study consists of four MAF layers, each implementing an affine autoregressive transformation in the forward direction $\mathbf{A} \mapsto \mathbf{z}$ of the form

$$z_k = \frac{A_k - \mu_k(A_{<k}, \mathbf{h})}{\sigma_k(A_{<k}, \mathbf{h})},$$

where μ_k and σ_k are outputs of a masked neural network, k indexes the components of $\mathbf{A}$, and $A_{<k} = (A_1, \ldots, A_{k-1})$. The autoregressive index refers to variable ordering and does not represent time.

The term "masked autoregressive" refers to the way statistical dependence between the components of $\mathbf{A}$ is structured within each flow layer. The masking enforces that the transformation applied to the k-th component depends only on the preceding components $A_{<k}$ and not on A_k itself or on any subsequent components. As a result, each rain-area component is modeled conditionally on those that precede it in a prescribed ordering, yielding a sequential factorization of the joint distribution across components.

This autoregressive structure serves two purposes. First, it guarantees that the transformation remains invertible, since each component can be solved for uniquely given the previous ones. Second, it leads to a triangular Jacobian matrix, whose determinant is given by the product of the diagonal terms, allowing the conditional density to be evaluated exactly and efficiently. Importantly, the ordering implied by the autoregressive index is purely a modeling device and does not correspond to temporal evolution or causality.

To reduce sensitivity to variable ordering and to increase expressiveness, permutation layers are inserted between successive MAF blocks. These permutations reorder the components of $\mathbf{A}$, ensuring that each rain-area component can condition on all others across the full stack of flow layers. Random permutations are used in this work.

Conditioning on the large-scale environment is implemented via a learned embedding

$$\mathbf{h} = g(\mathbf{X}_{\mathrm{LS}}),$$

where $g(\cdot)$ is a feedforward neural network with two hidden layers. The embedding dimension is set to 8. The context vector $\mathbf{h}$ is computed once per input $\mathbf{X}_{\mathrm{LS}}$ and provided to each MAF layer, where it modulates the scale and shift parameters of the affine transformations.

All parameters, including those of the flow and the context network, are trained jointly by maximizing the conditional log-likelihood described in SM1. Sampling from the trained CNF proceeds by drawing $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and applying the inverse transformation $\mathbf{A} = f_\phi^{-1}(\mathbf{z}, \mathbf{X}_{\mathrm{LS}})$.

A2.1 Training Procedure

The CNF is trained end-to-end by minimizing the negative conditional log-likelihood over a training dataset of paired observations

$$\left\{\mathbf{A}^{(i)}, \mathbf{X}_{\mathrm{LS}}^{(i)}\right\}_{i=1}^{N}.$$

Prior to training, the full dataset is randomly partitioned into disjoint training, validation, and test sets, comprising 70%, 15%, and 15% of the samples, respectively. The training set is used to optimize model parameters, the validation set to monitor convergence and perform early stopping, and the test set is reserved for out-of-sample evaluation.

Model optimization is performed using the Adaptive Moment Estimation (Adam) optimizer [1], with an initial learning rate of 10^{-3}, decayed multiplicatively by a factor of 0.98. Training is carried out with a batch size of 128 and a maximum of 500 epochs. Early stopping is applied based on validation loss, with a patience of 10 epochs.

A2.2 Sampling and Estimation of Conditional Moments

After training, the CNF can be sampled conditionally on a given large-scale state $\mathbf{X}_{\mathrm{LS}}^{(i)}$ by drawing independent latent samples $\mathbf{z}^{(j)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and applying the inverse transformation,

$$\widehat{\mathbf{A}}^{(i,j)} = f_\phi^{-1}\left(\mathbf{z}^{(j)}, \mathbf{X}_{\mathrm{LS}}^{(i)}\right).$$

Repeating this procedure for $j = 1, \ldots, M$ yields Monte Carlo samples $\widehat{\mathbf{A}}^{(i,j)}$ drawn from the learned conditional distribution $p_\phi(\mathbf{A} \mid \mathbf{X}_{\mathrm{LS}}^{(i)})$.

Conditional moments are then estimated by Monte Carlo averaging over the generated samples. For example, the conditional mean shallow rain-area fraction is approximated as

$$\overline{A}_{\mathrm{shallow}}^{(i)} = \frac{1}{M} \sum_{j=1}^{M} \widehat{A}_{\mathrm{shallow}}^{(i,j)},$$

with analogous expressions for deep and stratiform rain, as well as for higher-order conditional moments if required.

A2.3 Use in Equation Discovery

The symbolic regression step operates exclusively on flow-derived conditional moments, rather than on instantaneous realizations of the rain-area vector $\mathbf{A}$. Specifically, equation discovery seeks functional relationships of the form

$$\overline{\mathbf{A}} = F(\mathbf{X}_{\mathrm{LS}}),$$

where $\overline{\mathbf{A}}$ denotes conditional means estimated from Monte Carlo samples generated by the trained CNF.

Restricting equation discovery to conditional moments reduces sensitivity to sampling noise and short-term variability, allowing the inferred equations to represent systematic and robust dependencies of rain-area statistics on the large-scale environment. This approach separates stochastic variability, captured by the CNF, from deterministic structure, captured by the symbolic regression model.

A3. Training strategy

All machine-learning models are trained using a standard *train/validation/test* split, obtained via a *random partitioning* of the available data. The training set is used to optimize model parameters, the validation set is used only to monitor generalization, and the test set is held out entirely for final performance evaluation.

Models are implemented in `PyTorch` and optimized using stochastic gradient–based methods. Overfitting is prevented using *early stopping*, whereby training is halted when the validation loss does not improve for a prescribed number of epochs. The model state corresponding to the best validation performance is retained.

Model performance is quantified using the coefficient of determination, R^2, defined as

$$R^2 \;=\; 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}, \tag{1}$$

where y_i are the reference values, $\hat{y}_i$ the model predictions, and $\bar{y}$ the mean of the reference data. An $R^2 = 1$ indicates perfect agreement, while $R^2 = 0$ corresponds to performance no better than predicting the mean. The metric can become negative when the model performs worse than this baseline, indicating poor predictive skill.

References

[1] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[2] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. *Advances in neural information processing systems*, 30, 2017.

B. Additional Figures

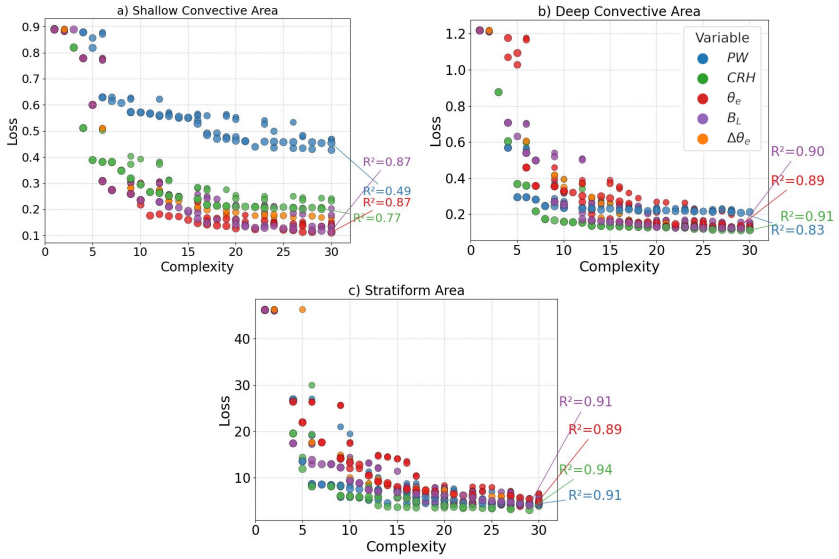

Figure S1: Pareto fronts plots for various PySR runs. Each marker is a candidate equation; the abscissa is expression complexity (parse–tree nodes) and the ordinate is validation loss (MSE). Colours denote the primary predictor family retained by the formula: precipitable water (PW), column–relative humidity (CRH), equivalent potential temperature (θ_e), plume buoyancy metrics (B_L), and deficit from saturation ($\Delta\theta_e$). For further details on the inputs combination see section 3.c. Panels show (a) shallow-convective, (b) deep-convective, and (c) stratiform area. Annotation shows the best R^2 for each set of experiments.

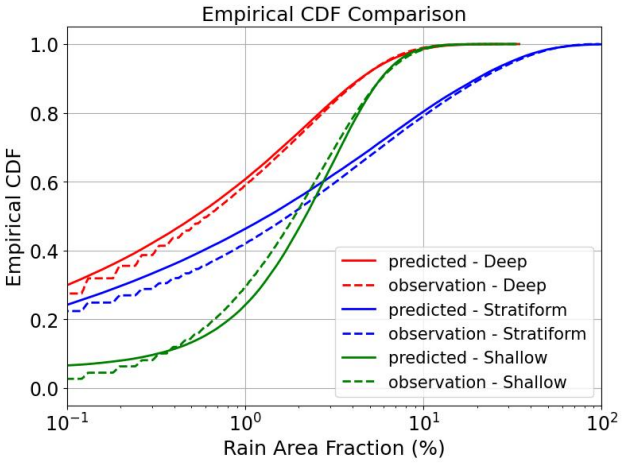

Figure S2: Empirical CDFs of Rain-Area Fractions for Deep, Stratiform, and Shallow Rain Types. Shown are the marginal cumulative distribution functions for observed (dashed lines) and CNF predicted (solid lines) rain-area fractions across rain types. A two-sample Kolmogorov–Smirnov (KS) test confirms that the distributions differ significantly (KS statistic $\approx$ 0.06–0.10; p $\ll$ 0.01), indicating minor but statistically significant deviations: the model slightly overestimates mid-range areas for deep convection and shifts shallow-rain fractions toward larger values.

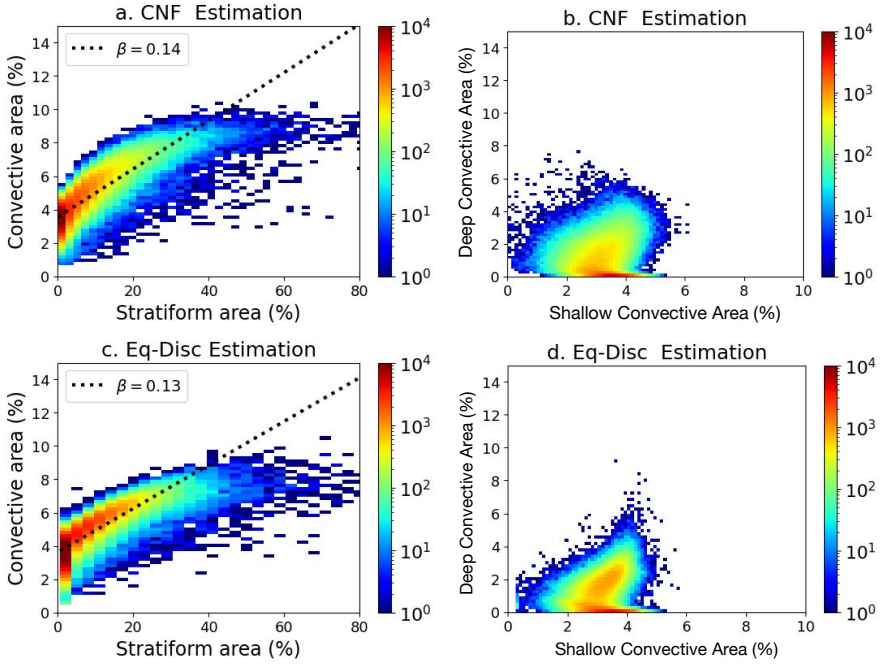

Figure S3: Panels show two-dimensional histograms of the mean predicted rain-area fractions, conditioned on environmental variables, from the CNF model (top) and the equation-discovery (Eq-Disc) model (bottom). Left panels (a, c) depict total convective area (deep + shallow) versus stratiform area; right panels (b, d) show deep versus shallow convective area. Color indicates frequency on a logarithmic scale. In panels (a) and (c), the black dotted line represents a linear regression between total convective and stratiform area, with the estimated slope shown in the legend. Similar to Fig. 4 but for latitude band 20S-20N.

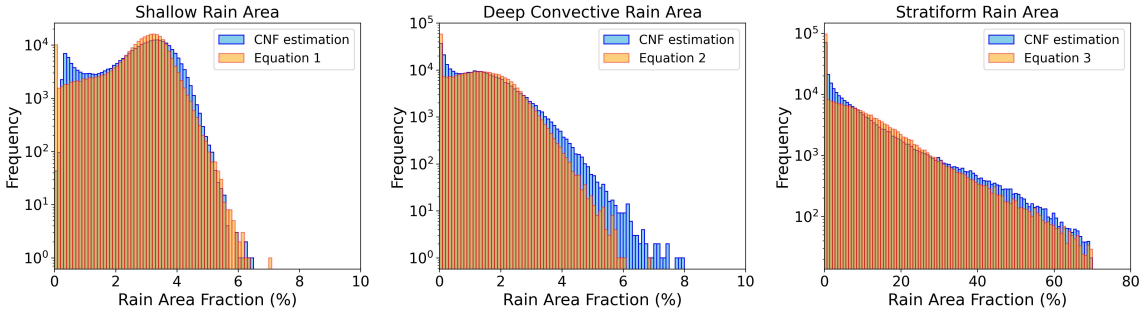

Figure S4: Comparison of the conditional distributions of rain area fractions for shallow (left), deep convective (middle), and stratiform (right) precipitation. Blue color shows estimates from the CNF, while orange color shows predictions from the equations discovered via symbolic regression. Across all rain types, the equations accurately reproduce the overall distributions and successfully capture the shape and tails.

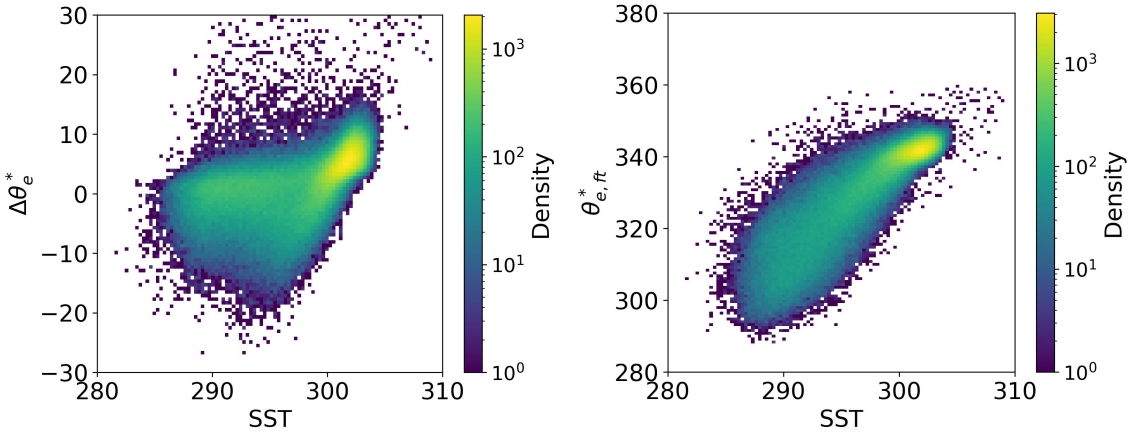

Figure S5: Joint distributions of SST with two measures of thermodynamic state. Left: Lower-tropospheric stability, expressed as $\Delta\theta_e^* = \theta_{e,\mathrm{pbl}}^* - \theta_{e,\mathrm{ft}}^*$. Right: Lower free-tropospheric saturation equivalent potential temperature, $\theta_{e,\mathrm{ft}}^*$. Colors indicate density on a logarithmic scale.

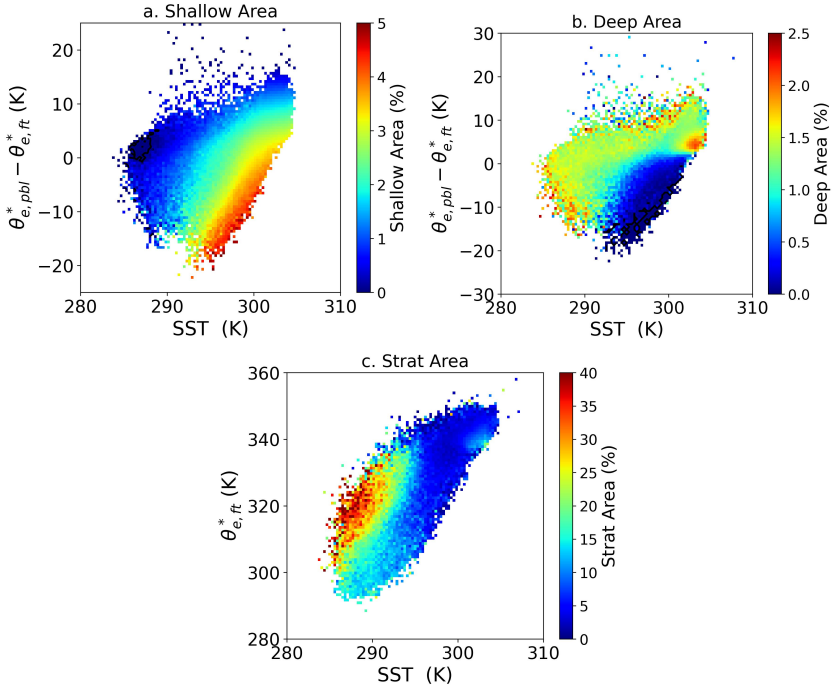

Figure S6: Phase–space distributions of rain area fractions across SST and thermodynamic state. (a) Shallow-convective area as a function of SST and lower-tropospheric stability, $\Delta\theta_e^* = \theta_{e,\mathrm{pbl}}^* - \theta_{e,\mathrm{ft}}^*$. (b) Deep-convective area as a function of SST and $\Delta\theta_e^*$. (c) Stratiform area as a function of SST and lower free-tropospheric saturation equivalent potential temperature, $\theta_{e,\mathrm{ft}}^*$. Colors denote the fractional area covered by each rain type (shallow, deep, or stratiform).

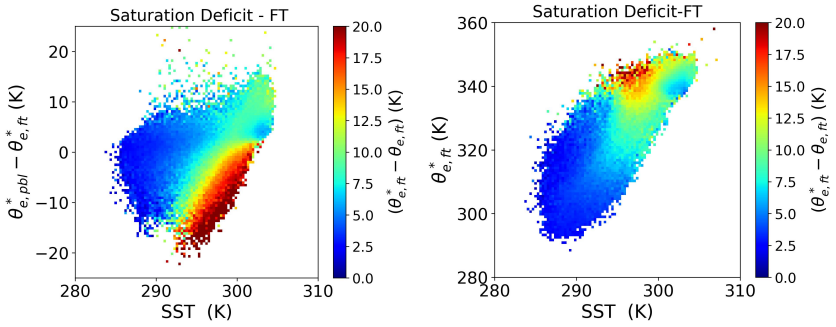

Figure S7: Phase–space distributions of column moisture diagnostics across SST and thermodynamic state. (a) Free-tropospheric dryness measured as $\Delta\theta_e^* = \theta_{e,\mathrm{pbl}}^* - \theta_{e,\mathrm{ft}}^*$, $\Delta\theta_e^* = \theta_{e,\mathrm{pbl}}^* - \theta_{e,\mathrm{ft}}^*$. (b) Same as (a) but shown in SST-$\theta_{e,ft}^*$ space for easier comparison with Fig. 9. Colors indicate the magnitude of $\Delta\theta_e^*$, with larger values on the colorbar corresponding to a drier free troposphere.