Stochastic emulator algorithm details

Building a machine learning climate emulator

Harvard Earth and Planetary Science 226, Eli Tziperman

1. Noise-conditioned predictions

Let S_t denote the three physical state fields at month t: SST, D20, and zonal wind stress. Let X_t denote the complete deterministic input to an emulator, consisting of S_t together with the calendar and the model-specific mask or positional features. The stochastic versions of all four emulators append one independent Gaussian-noise field $Z_t \sim \mathcal{N}(0, I)$ to the state, calendar, and positional inputs. The noise has the same spatial dimensions as one physical input field, such as SST: it contains one independent $\mathcal{N}(0, 1)$ value at each spatial location and is concatenated as one additional input feature. Thus it is one noise field, not one field per physical variable. For a given input state and noise realization, each model evaluation produces one stochastic forecast realization:

$$\widehat{S}_{t+1}^{(k)} = f_\theta(X_t, Z_t^{(k)}). \tag{1}$$

Here k indexes independent noise realizations; during training, $k = 1, 2$. Keeping the input state and trained network weights fixed, different noise realizations produce different forecast realizations. Together, these realizations represent the model's predictive distribution.

2. Training

For every one-month training pair (X_t, S_{t+1}), the code uses two independent realizations of the stochastic field and minimizes the energy score over ocean grid points

$$\mathcal{L} = \frac{1}{2\sqrt{3N}}\left(\|\widehat{S}_{t+1}^{(1)} - S_{t+1}\|_2 + \|\widehat{S}_{t+1}^{(2)} - S_{t+1}\|_2\right) - \frac{1}{2\sqrt{3N}}\|\widehat{S}_{t+1}^{(1)} - \widehat{S}_{t+1}^{(2)}\|_2. \tag{2}$$

Here N is the number of ocean grid points, and $\|\cdot\|_2$ is the Euclidean norm over all three fields and those points. The first term rewards both stochastic forecast realizations for accurately predicting the next-month state; the negative pairwise term rewards appropriate spread and discourages collapse to one deterministic forecast. Training retains the weights with the lowest validation energy score.

3. Ensemble forecasts

The initial state is used for all ensemble members. At each forecast month, every member receives a new independent noise field, advances through the model, is masked over land, and feeds its own prediction into the next step while the calendar advances.