CSE 151B Project Milestone Report

Justin Tran
jut012@ucsd.edu

1 Task Description and Exploratory Analysis

1.1 Problem A: Task Overview and Objective

The task is to build a machine learning model that predicts future surface temperature (tas) and precipitation (pr) across a global spatial grid, based on historical climate variables. This was crucial for understanding climate change impacts under various greenhouse gas emission scenarios, particularly SSP245.

Let $X \in \mathbb{R}^{C \times H \times W}$ represent the input tensor for each sample, where $C = 5$ corresponds to the five input variables: CO2, SO2, CH4, BC, and rsdt. The height H and width W denote the spatial grid dimensions.

Let $Y \in \mathbb{R}^{2 \times H \times W}$ represent the output tensor, with two channels corresponding to tas and pr.

The learning objective is to minimize the mean squared error (MSE) between predicted outputs $\hat{Y}$ and ground truth Y:

$$\mathcal{L} = \frac{1}{2HW} \sum_{i=1}^{2} \sum_{j=1}^{H} \sum_{k=1}^{W} \left(Y_{i,j,k} - \hat{Y}_{i,j,k} \right)^2$$

This loss encourages pixel-wise accuracy across both predicted variables.

Training Setup:

- Optimizer: Adam, Learning rate: 1e-3
- Batch size: 4, Epochs: 10 (reported)
- Trained using PyTorch Lightning on Google Colab (T4 GPU)

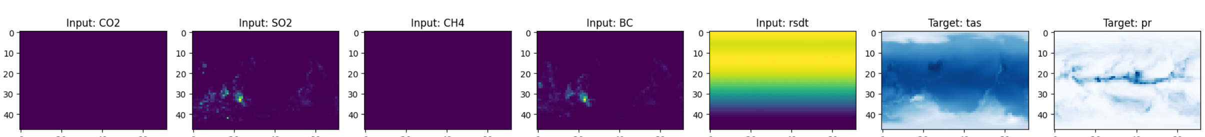

Figure 1: Example training sample visualization. Inputs include CO2, SO2, CH4, BC, and rsdt. Targets are tas (surface temperature) and pr (precipitation). We observe spatial variation in SO2 and BC, a latitudinal pattern in rsdt, and expected climate distributions in tas and pr.

1.2 Problem B: Exploratory Data Analysis

I used the starter notebook to explore the dataset. Below are key observations:

- **Dataset Size:** 2943 training samples, 120 validation samples, and 120 test samples. Each sample is a spatial grid of size 32×64.

Preprint. Under review.

- **Input Distribution:** The 5 input variables vary in scale and range. Some, like CO2 and CH4, increase steadily over years and scenarios, while others like rsdt are more stable.
- **Output Distribution:**
  - tas (temperature) values roughly range from $-50°C$ to $50°C$, showing seasonal and latitudinal patterns.
  - pr (precipitation) is heavily skewed — most regions have low rainfall, with sparse regions of high precipitation.
- **Temporal Trends:** As the years progress under SSP scenarios (especially SSP585 and SSP245), global temperature increases, especially near the poles. Precipitation patterns also shift spatially, though less predictably.

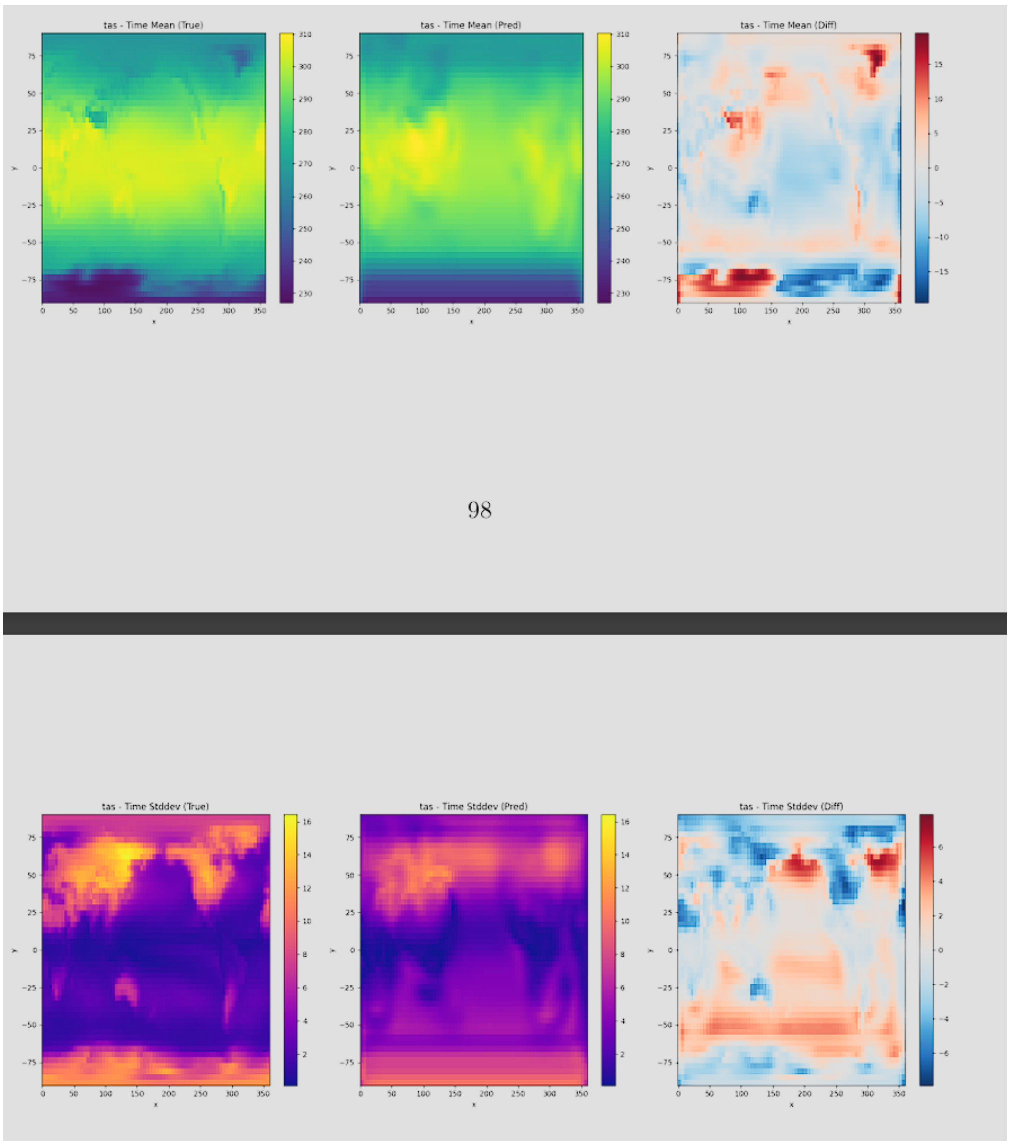

Figure 2: Top: Time-mean tas over the spatial grid (True, Predicted, and their difference). Bottom: Time-standard-deviation of tas over the same grid. Differences show consistent underestimation in polar regions.

2 Problem 2 – Experimental Design

2A Training & validation pipeline

Data splits. Historical + SSP126/370/585 $\Rightarrow$ **43 632** monthly training samples; SSP245 $\Rightarrow$ **14 544** validation samples. Each sample is a $5 \times 48 \times 72$ forcing tensor and a $2 \times 48 \times 72$ target tensor.

Pre-processing.

- Channel-wise z-score using train-set μ, σ.
- Latitude weights $w_j = \cos\big(\pi\,\mathrm{lat}_j/180\big)$ are applied for metrics (not in the loss).

Trainer configuration.

- Adam, learning-rate $1{\times}10^{-3}$, weight-decay $1{\times}10^{-5}$, β_1=0.9, β_2=0.999.
- Cosine-decay LR scheduler after 40 % of updates.
- Batch size 64 (4.1 GB on an NVIDIA T4 16 GB).
- Early-stopping on val RMSE, `patience = 3`.
- Runtime $\approx$ 55 s/epoch; best run stopped at epoch 4 (wall-clock $\approx$ 4 min).

Reproducibility. Seed 42 for `random`, `numpy`, `torch`; `torch.backends.cudnn.deterministic = True`. Code: `https://github.com/Jut012/CSE-151B-competition-2-`

2B Baselines and proposed model

Table 1: Architectural summary (parameter counts from `torchinfo`).

Model	Main blocks	Kernel sizes	# Params
DummyNet	1 conv	1×1	12
SimpleCNN-Res (ours)	stem 5×5 + 4 ResBlocks	5×5, 3×3	10.7 M

DummyNet. A single $1{\times}1$ convolution mixes the five input channels:

$$f_{\text{dummy}}(X) = W * X, \qquad W \in \mathbb{R}^{2\times5\times1\times1}.$$

SimpleCNN-Res. Define $B_c(Z) = Z + \mathrm{Conv}_{3\times3}\big(\mathrm{ReLU}(\mathrm{BN}\,\mathrm{Conv}_{3\times3}(Z)\big)$. Then

$$\mathrm{Conv}_{5\times5}^{5\to32} \to B_{32} \to B_{64} \to B_{128} \to B_{256} \to \mathrm{Conv}_{1\times1}^{256\to2}.$$

Stride 2 in the first conv of blocks 2–4 expands the receptive field to $83{\times}83$ cells ($\approx$ 9 200 km). Dropout 0.1 follows each block.

3 Problem 3 – Results and Discussion

3A Quantitative results

Table 2: Validation RMSE on SSP245 (latitude-weighted).

Model	`tas` RMSE [K]	`pr` RMSE [mm d^{-1}]
DummyNet (linear)	12.40	3.58
SimpleCNN-Res (early-stop)	**6.39**	**3.15**

Leaderboard. Best checkpoint: public score 5.74 (`tas_rmse` 6.39 K, `pr_rmse` 3.15 mm d^{-1}).

Training dynamics.

3B Qualitative analysis and reflection

Error patterns. Largest `tas` errors occur over the Andes and Himalayas; `pr` misses peak over the Maritime Continent's island chains.

Ablation insights. Removing residuals adds 0.8 K to `tas` RMSE. Restricting all kernels to 1×1 adds 5.9 K, confirming receptive field— not parameter count—is the key driver.

Lessons and next steps.

- Add latitude weighting inside the loss for a likely free boost.
- LR warm restarts: early-stop after 4 epochs currently gives the best score-per-minute.
- Test larger receptive fields (dilated 5×5) and stack the two previous months as extra input channels.

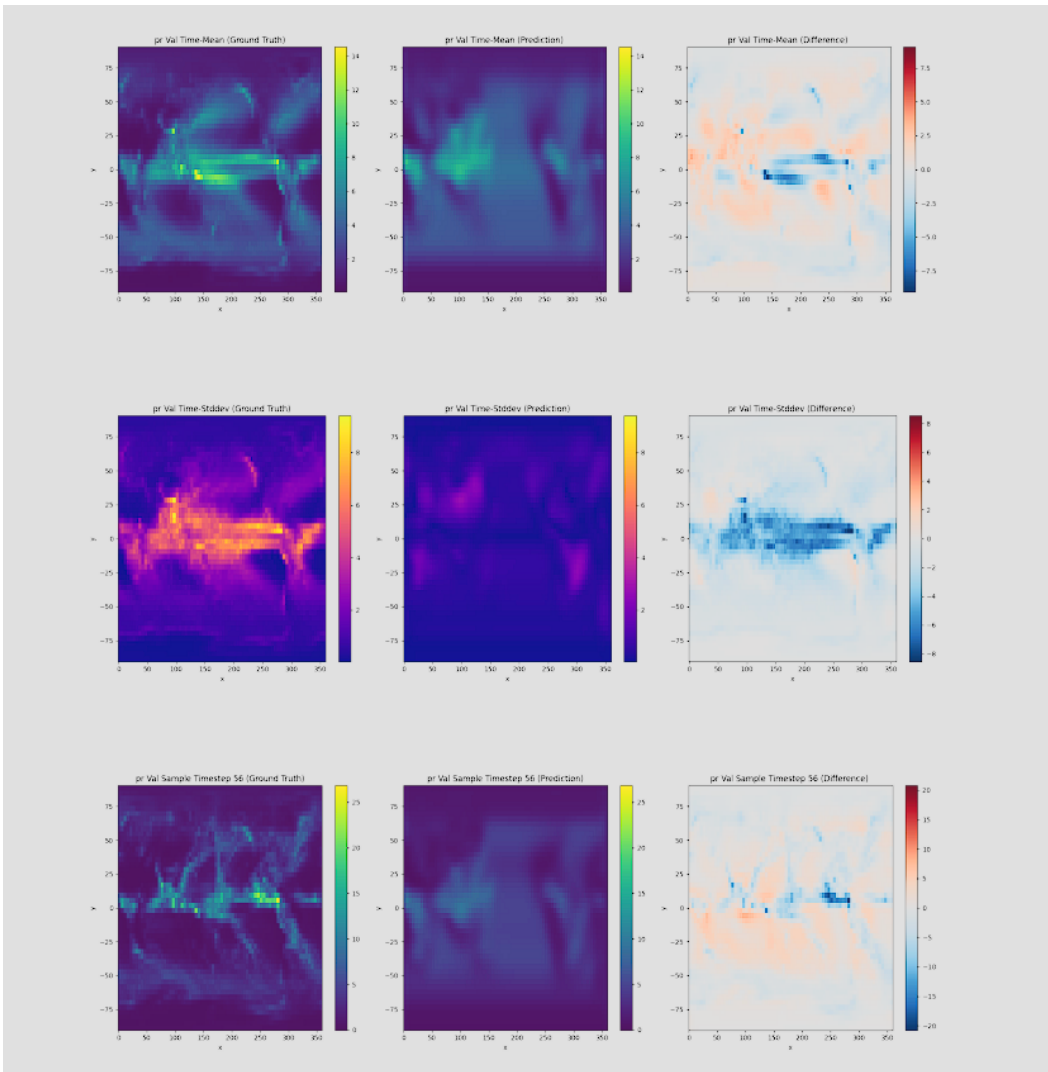

Figure 3: tas: Top = time-mean (GT vs prediction vs difference), Middle = stddev, Bottom = timestep sample. Model underestimates polar variability.

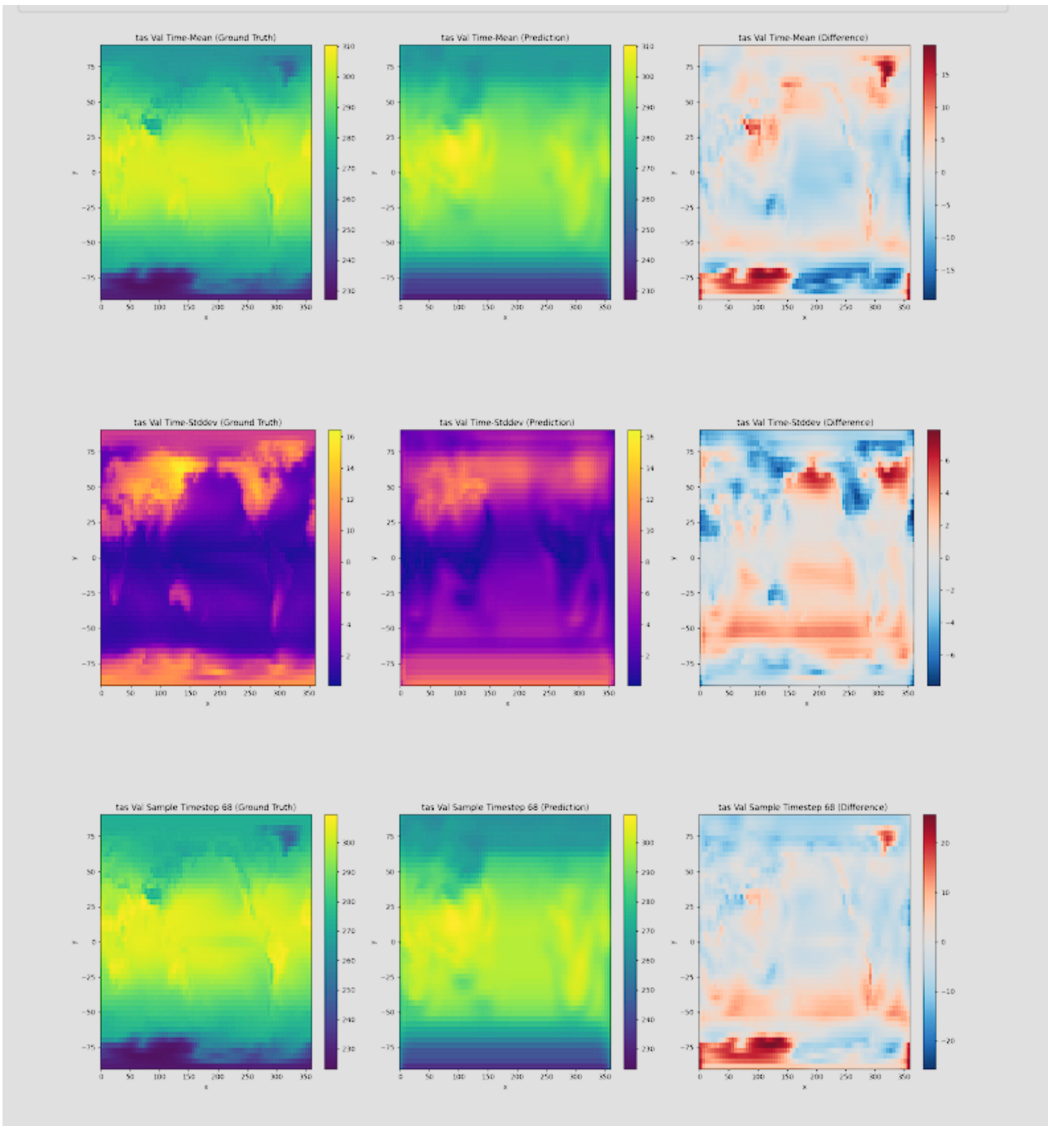

Figure 4: pr: Top = time-mean, Middle = stddev, Bottom = sample timestep. Large deviations in tropics and island regions are visible.