

Problem Statement

- High Impact Hazards:** The Indian Himalayas are intensely susceptible to catastrophic, highly localised cloudbursts and flash floods driven by steep orographic gradients.
- Observational Deficit:** Extreme elevations and complex terrain create a severely data-sparse environment, severely limiting the density of ground truth meteorological monitoring.
- NWP bottleneck:** Traditional Numerical Weather and Prediction (NWP) Models require huge computational resources and walltime to generate sub-km scale data which limits the operational feasibility.
- Deterministic ML Trap:** While data driven models are computationally efficient, but they optimise for mean centric errors and tend to just smoothen out the precipitation fields and miss the localised extreme intensity peaks required for hazard forecasting.



Datasets

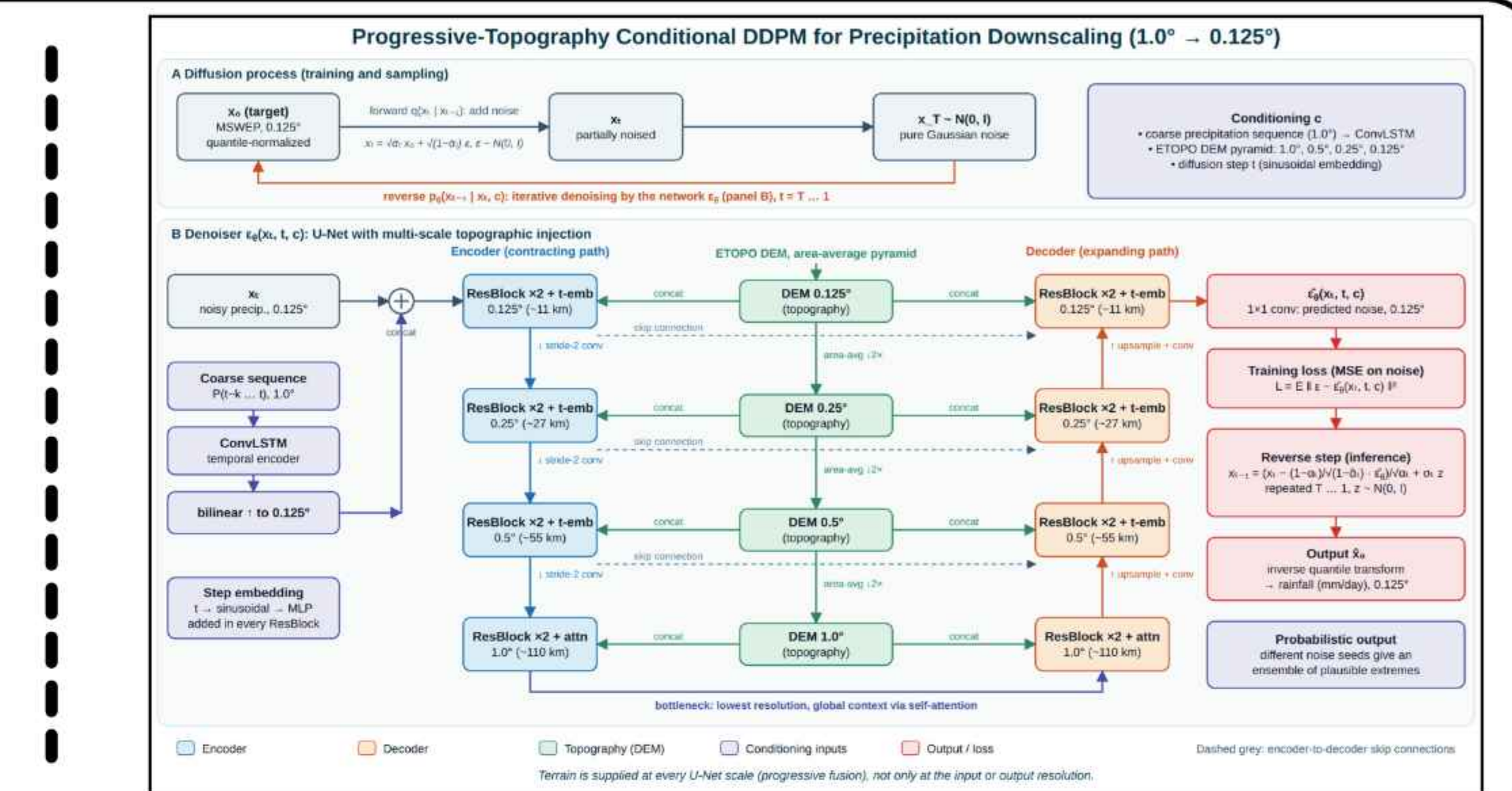
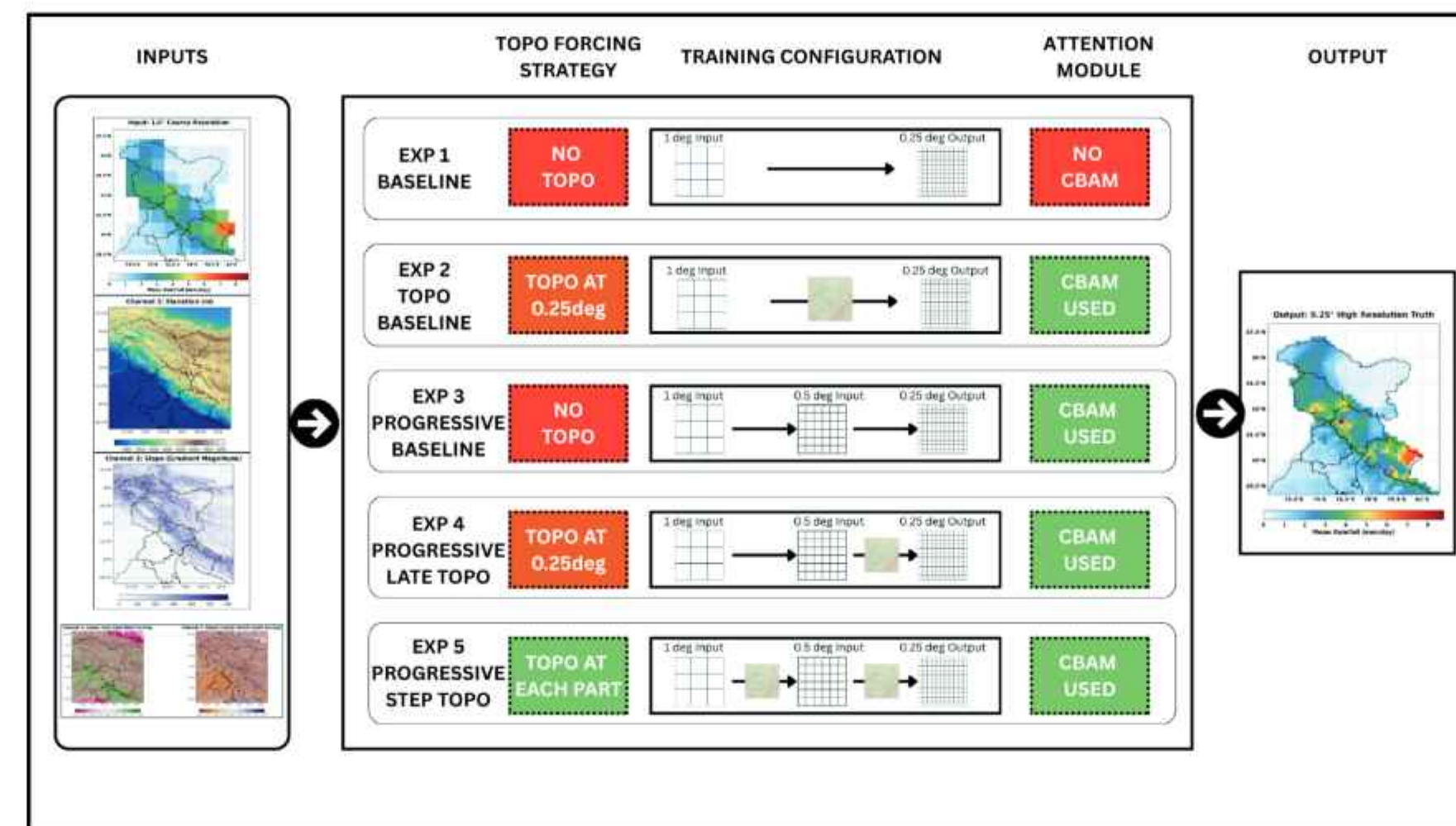
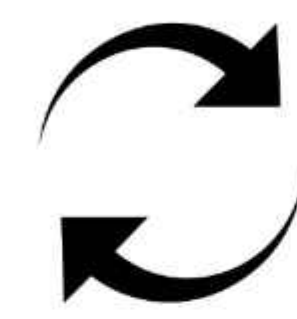
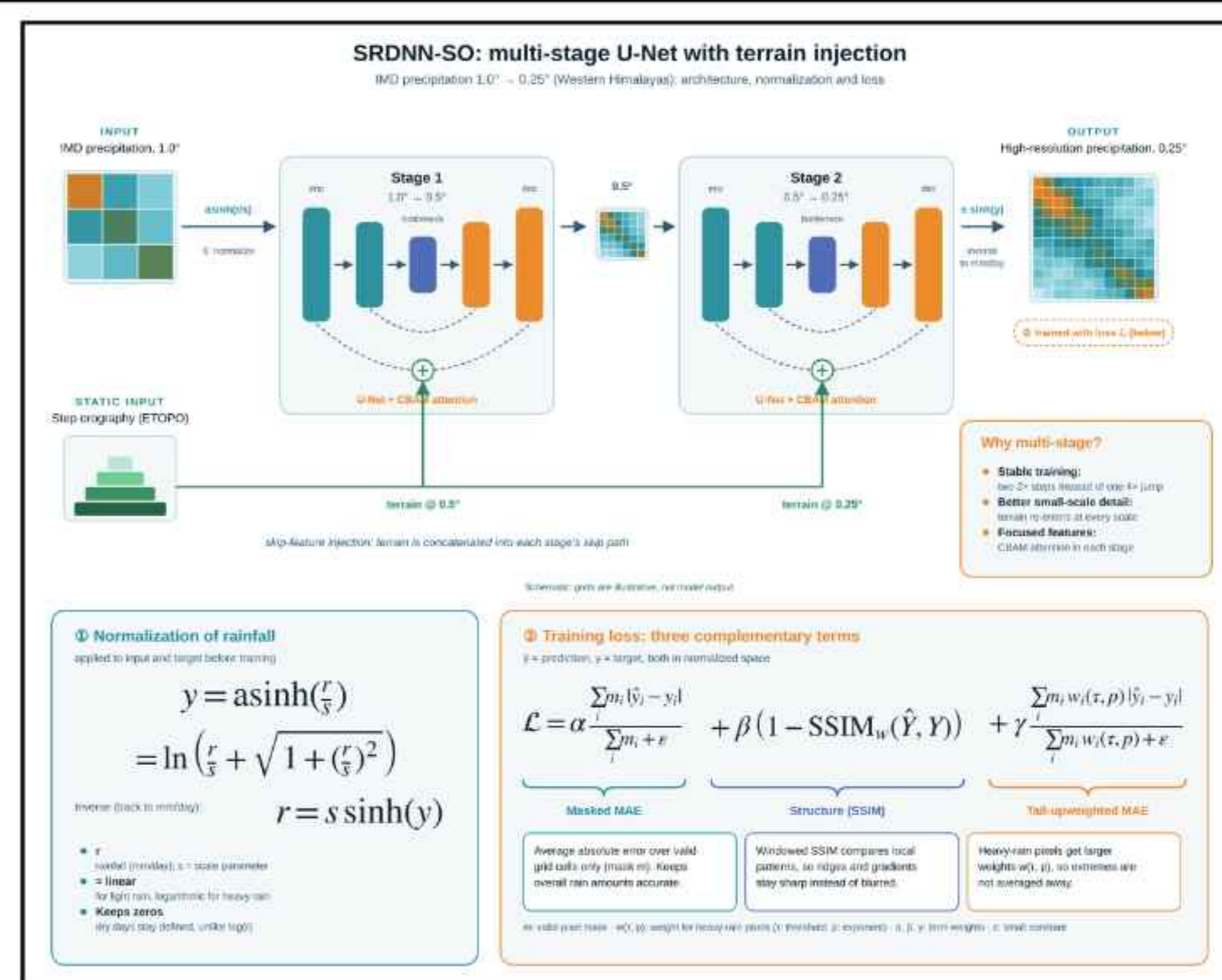
Training Datasets

- IMD Gridded RF Data (1901-2025)** (0.25deg X 0.25deg)
- Multi-Source Weighted-Ensemble Precipitation (MSWEP) (1979-2025)** (0.1deg X 0.1deg)
- NOAA ETOPO 2022 Digital Elevation Model** (30" X 30")

Independent Validation Datasets

- IMD-GPM Rainfall Data (2020-2025)** (0.25deg X 0.25deg)
- IMERG Satellite Data (2020-2025)** (0.1deg X 0.1deg)
- MERA Reanalysis (NCMRWF)** (4km X 4km)

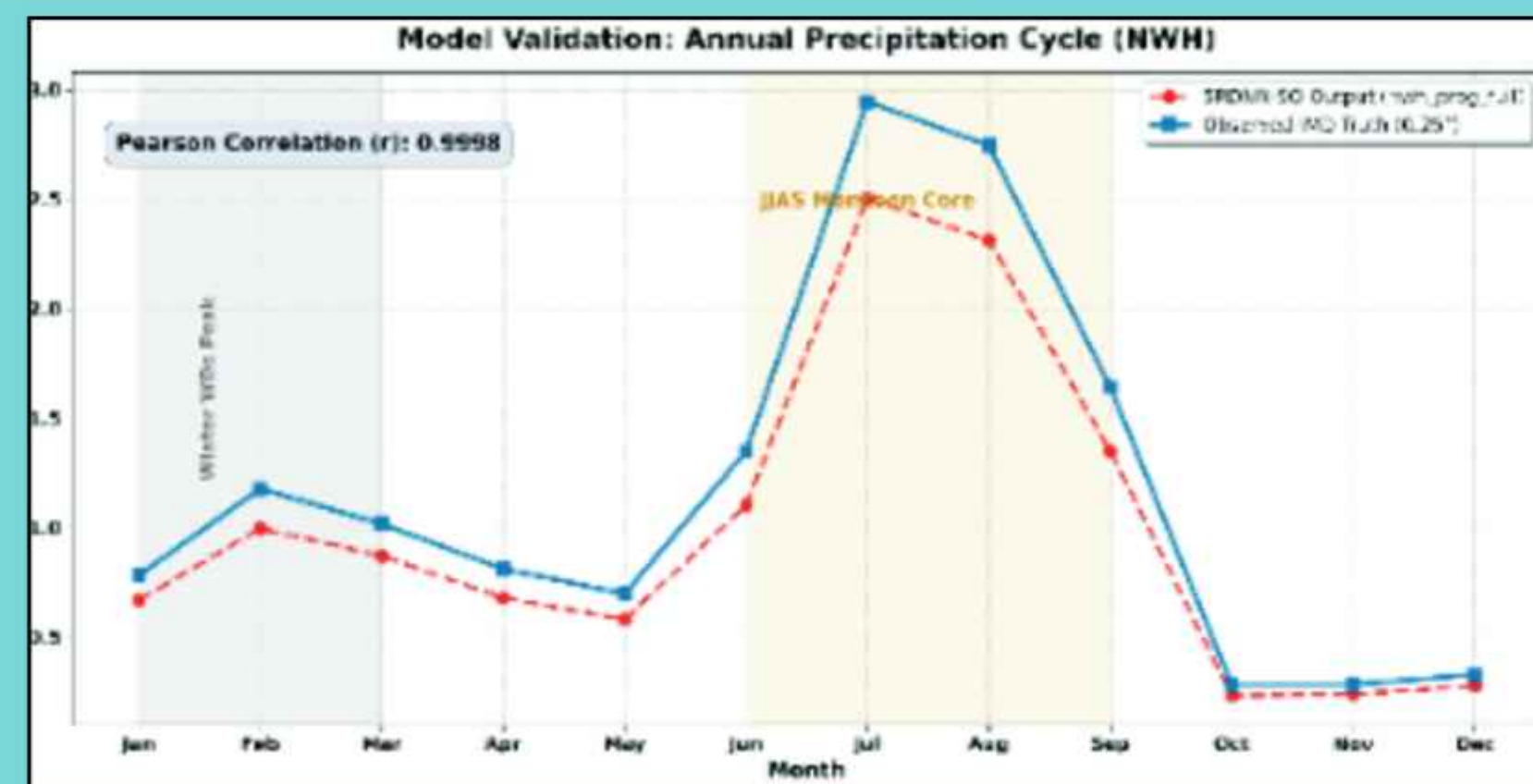
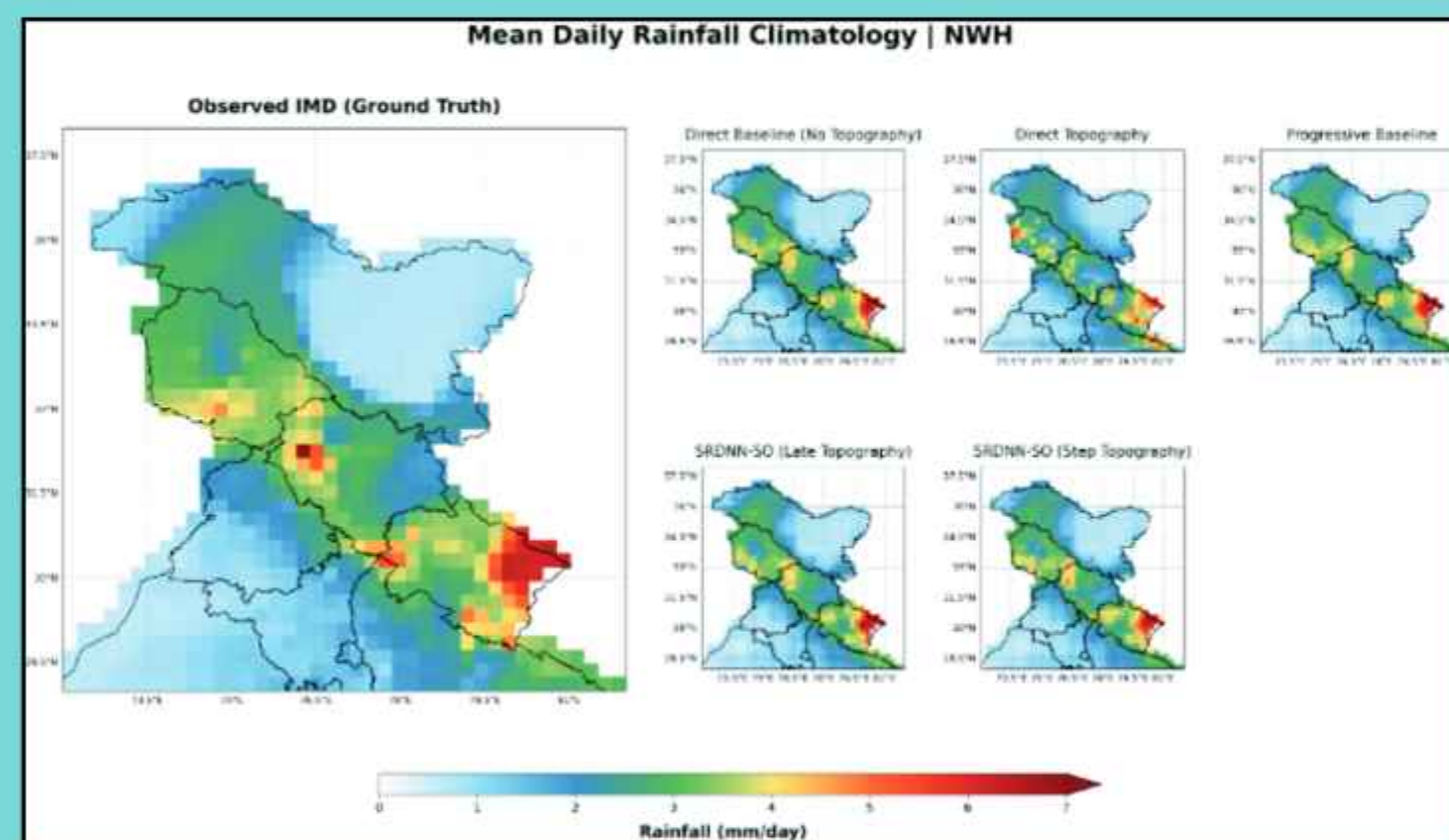
Methods



Results

Does Terrain Help ?

Phase 1: IMD 1.0deg $\rightarrow$ 0.25deg Deterministic U-Net



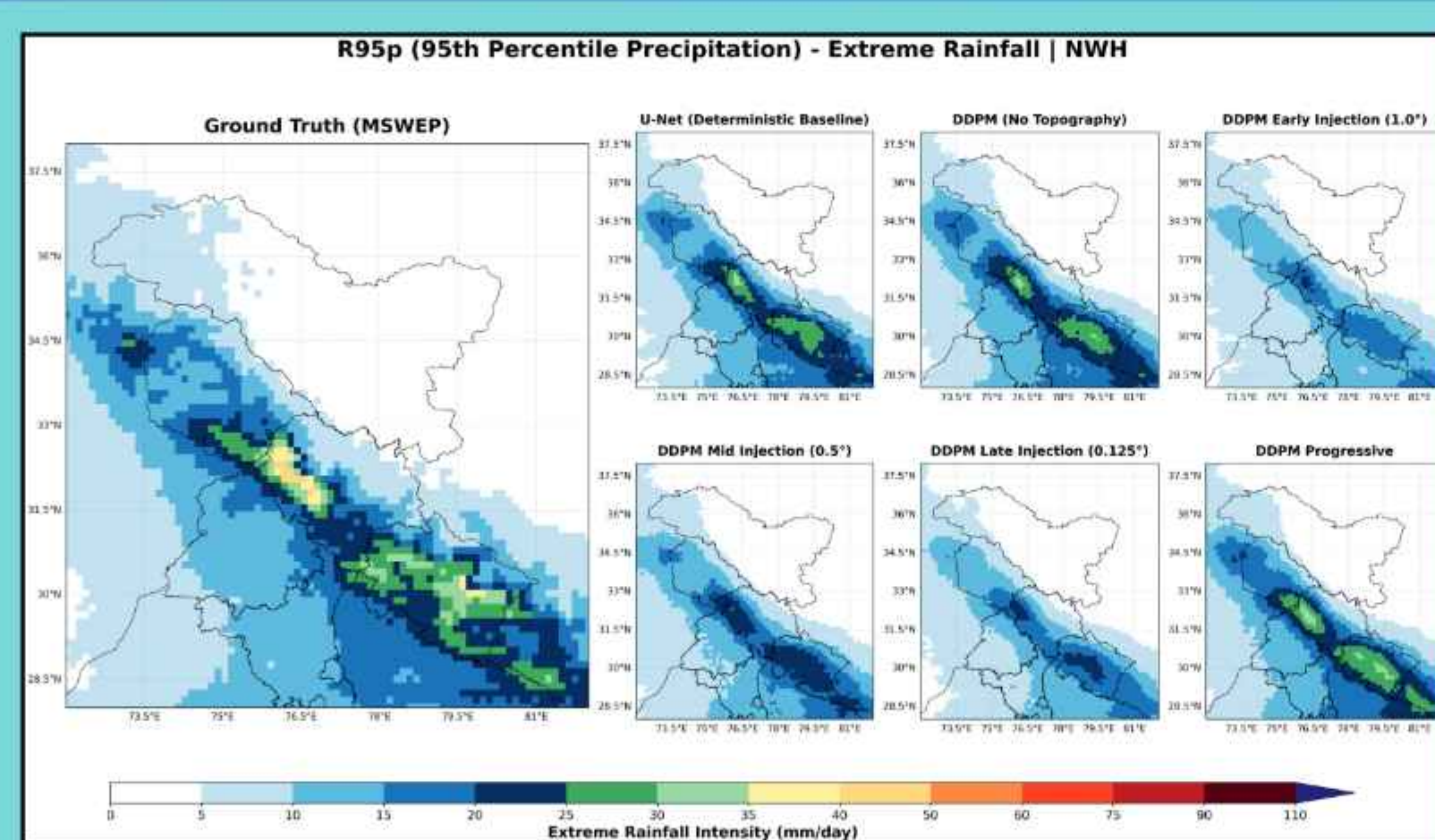
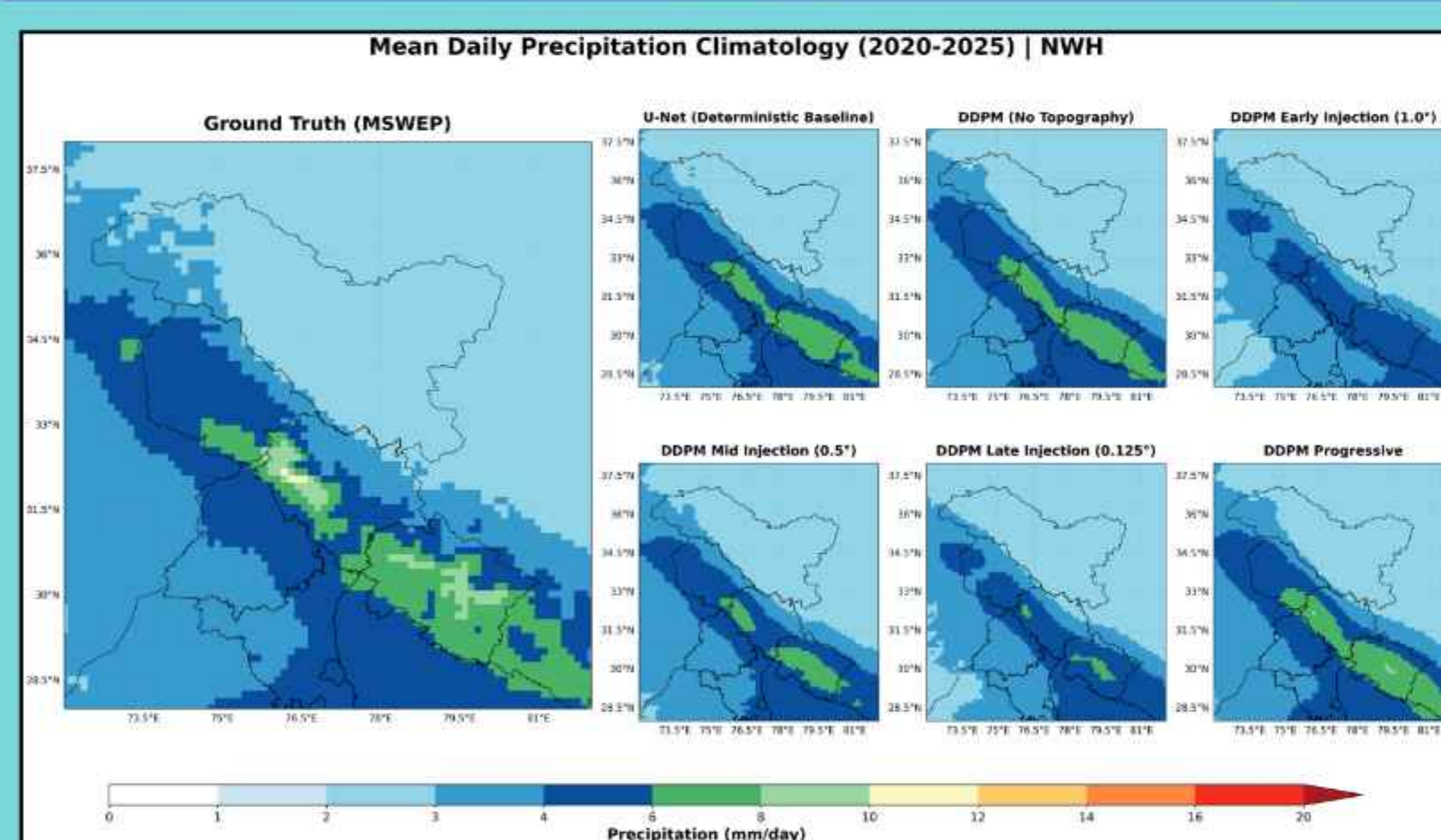
What we Observe !!

- Terrain injection stage matters:** Model is sensitive to when and how you inject the topography information, Step Orography injection (SRDNN-SO) gives lowest MSE (5.49 mm/day), while late or early injection have highest MSE (~26.61 mm/day)
- Mimics the Intra-annual Variability:** Model is able to resolve and show the seasonal variability of precipitation in North-Western Himalayas.

But at 0.25deg (~27km), model is still smoothening extremes and underestimating the intensity

How about a generative approach ?

Phase 2: MSWEP 1.0deg $\rightarrow$ 0.125deg Conditional DDPM

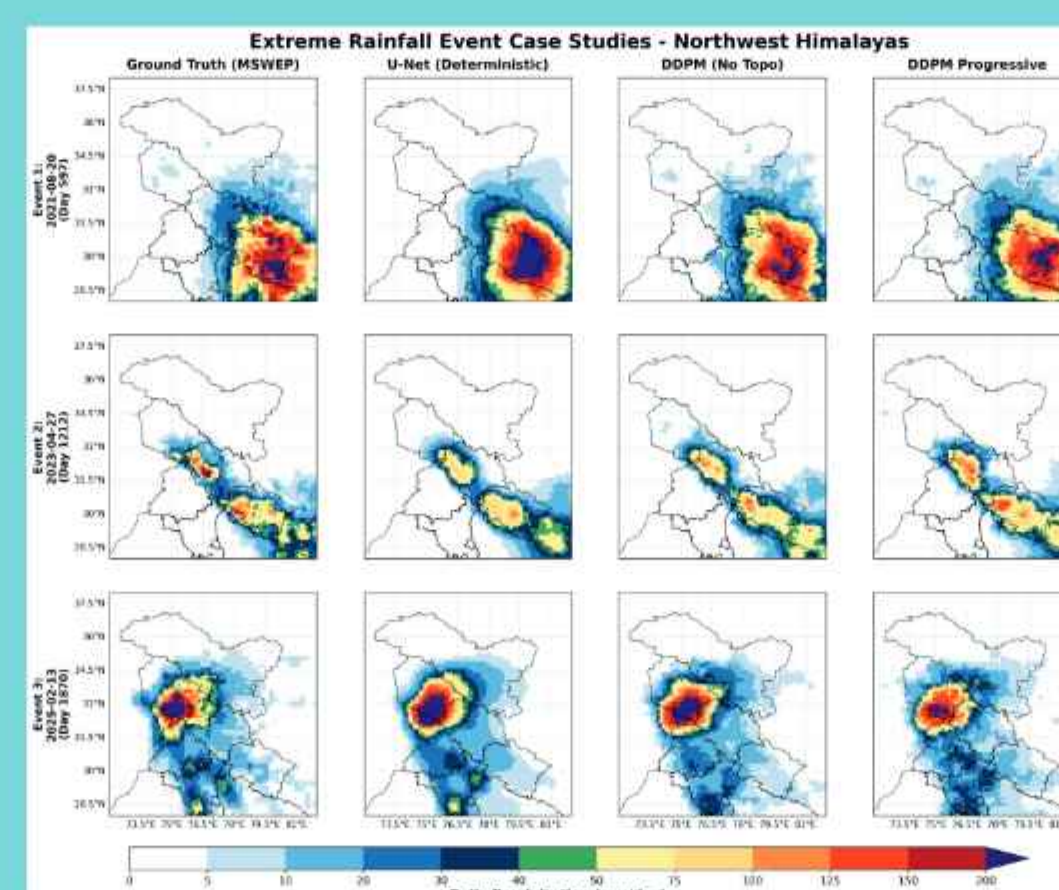
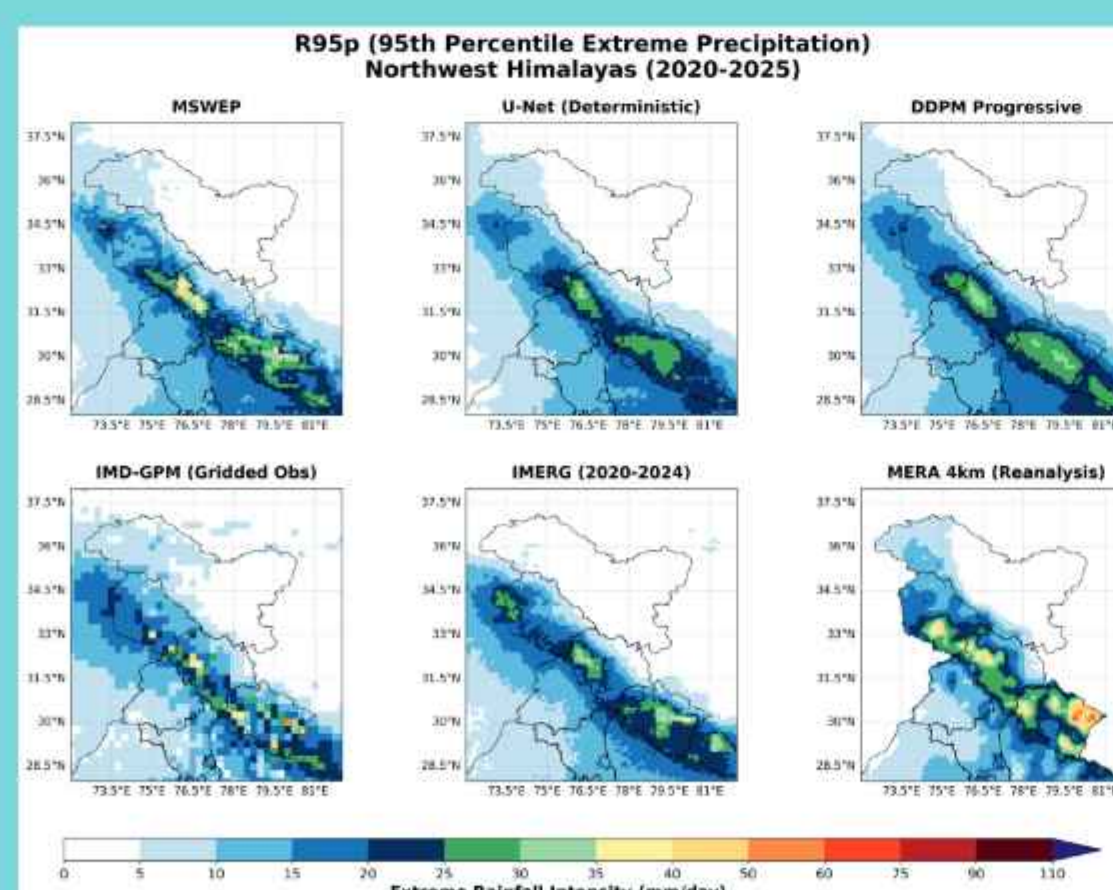
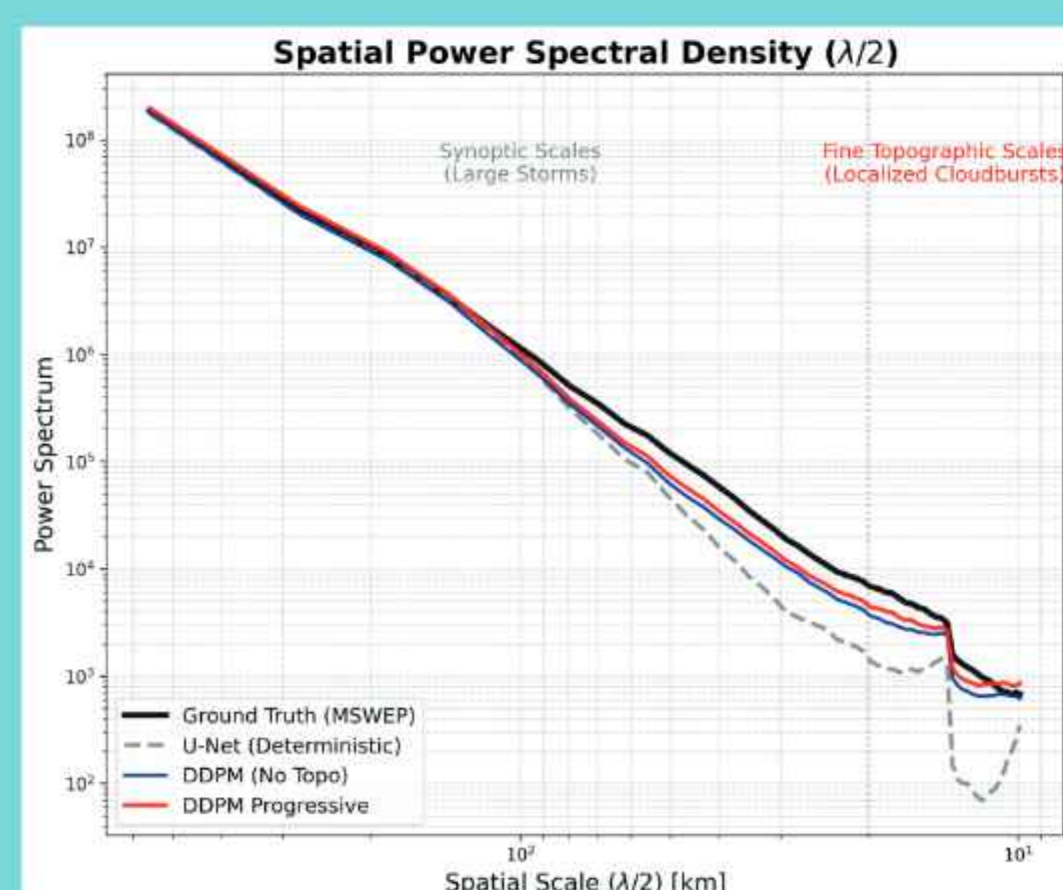


What we Observe !!

- Means look alike but extremes play deciding factor:** Overall mean climatology is captured well by the variants but extreme precipitation is where the differences arise.
- Progressive model shows promise:** Model with progressive injection of topography shows physically more coherent results and anchors heavy rain to windward slopes and ridges instead of the valleys.

Does it capture extremes and hold up ?

Phase 2a: Case Studies and Comparison with Observation



What we Observe !!

- Deterministic System smoothenes the extremes:** U-Net loses power at higher resolution (smoothenes out), while DDPMs retain more power being closest to ground truth data.
- DDPMs capture EREs:** DDPM models (especially progressive topography) retain the intensity and footprint of Extreme Rainfall Event closer to reality.
- Independent Data Validation:** IMD-GPM, IMERG and MERA (4-km reanalysis) show the same ridge aligned extreme rainfall band.

Conclusion

Take Home Message: Giving the model progressive terrain information, and utilizing a generative rather than deterministic framework, successfully recovers the localized extreme rainfall cores that coarse datasets and MSE-optimized models smooth away.

Challenge

Himalayan precipitation events (especially cloudbursts) are deeply tied to steep terrain. Sparse observation networks make these catastrophic drivers difficult to track, while sub-kilometer NWP models are too computationally intensive for operational use.

Approach

Terrain aware downscaling approach yields better results in both deterministic as well generative models.

Performance

Step injection of terrain helps DDPM retain fine scale spectral power at higher resolutions. It also captures the footprint of extreme events better and agree with independent datasets (IMD-GPM, IMERG and MERA)

Impact

Sharper terrain faithful rainfall data (~11 km); a data-driven approach to complement costly NWP models for extreme rainfall event assessments and a step towards building a disaster resilient himalayan society.

Next Steps: Ensemble based uncertainty from multiple noise seeds and go further till sub-km scale downscaling

Acknowledgements

We thank IMD, NASA (GPM-IMERG), NCMRWF (MERA), MSWEP and NOAA (ETOPO) for making their data openly available for training and validation. We thank the Oak Ridge Leadership Computing Facility (OLCF), Oak Ridge National Laboratory, for the compute resources used to train the ML models. We are grateful to all members of the Weather and Climate Modelling Research Group (WeClimb), IISER Mohali, for their support.

