TOWARDS A MACHINE LEARNING BASED MULTIMODEL FOR PRECIPITATION FORECAST OVER THE ITALIAN PENINSULA

Politecnico di Torino
Department of Environment, Land and Infrastructure Engineering

[1,2]LUCA MONACO, [1]Roberto Cremonini, [2]Prof. Francesco Laio
[1]ARPA Piemonte (Dipartimento Rischi Naturali e Ambientali), [2]Politecnico di Torino (DIATI)
Contacts: luca.monaco@polito.it, luca.monaco@arpa.piemonte.it

RESEARCH QUESTION

Is a machine learning based multimodel ensemble a viable solution to predict deterministic and probabilistic precipitation fields?

BACKGROUND

Many weather models available (NWPs)

NWPs present limitations by definition

Ever growing increase in observational data reliability

Precipitation forecast is fundamental for the National Italian Civil Protection' system

MACHINE LEARNING BASED MULTIMODEL ENSEMBLE OF DETERMINISTIC NWPs FOR PRECIPITATION FORECAST
[Monaco et al., ongoing]

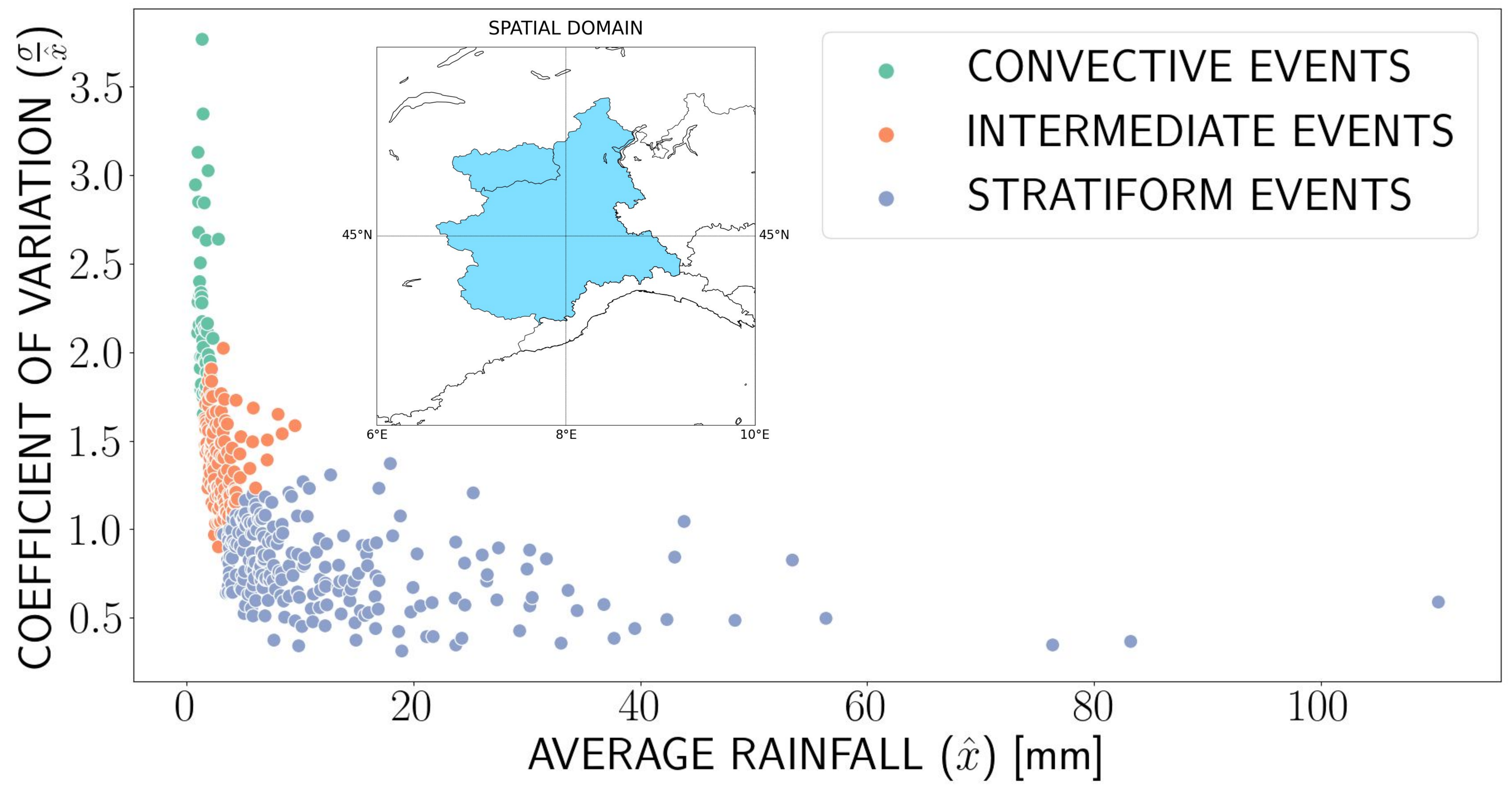
EXPERIMENTS

01	Non Negative Least Square (NNLSs)	• NNLS 1: different cells, shared weights • NNLS 2: different cells, different weights
02	Fully connected Neural Networks (FCNNs)	• **Deterministic output**: FCNN 3 layers • 50-500-1500-5000 neurons
03	Convolutional Neural Networks (CNNs)	• **Deterministic output**: CNN 4 layers, plain and 20% dropout (CNN 20% DO); U-NET; • **Probabilistic output**: U-NET 20% Montecarlo Dropout (MC DO U-NET).

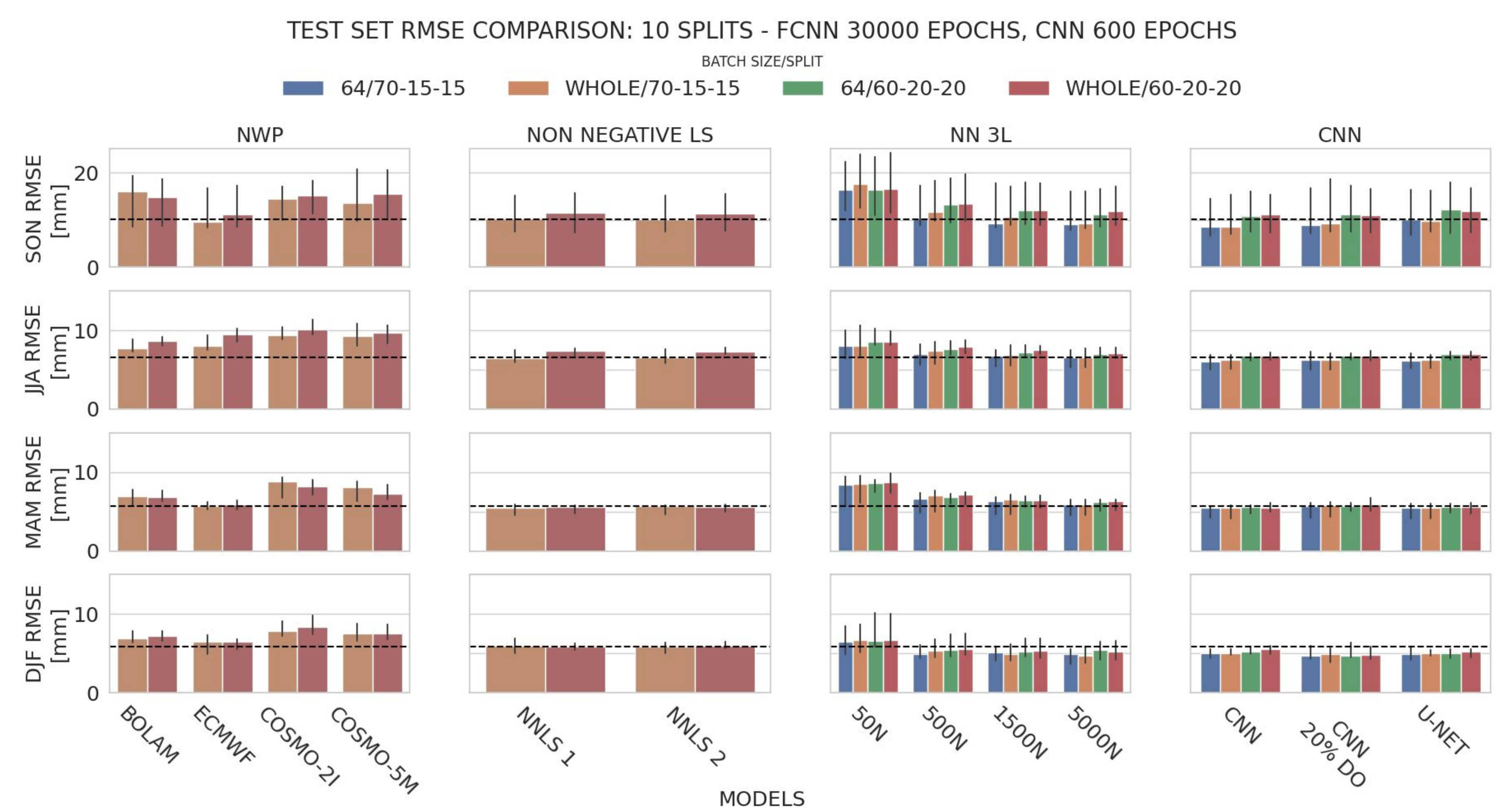
EXPERIMENTS SETUP

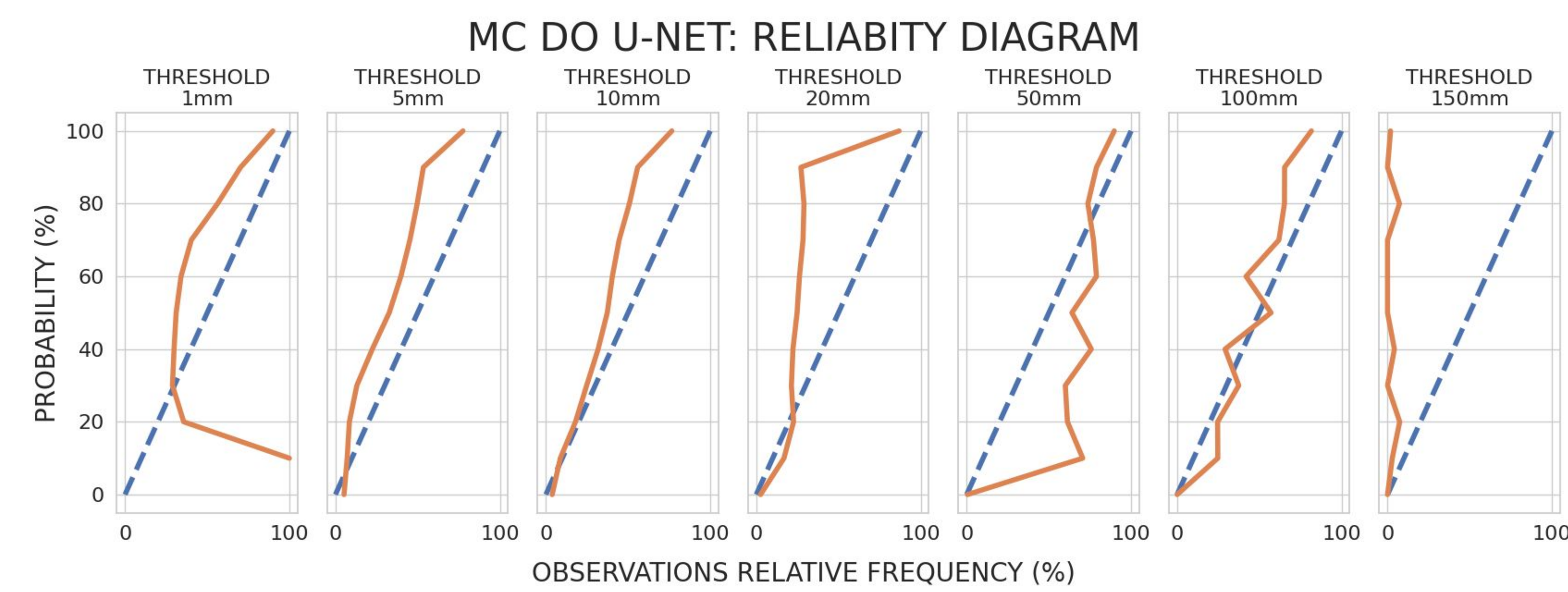
01	Dataset (I)	• 406 days from 2018 to 2022 with a significant precipitation signal • 10 x 70/15/15 split, 10 x 60/20/20 split
02	Dataset (II)	• Deterministic input NWPs: BOLAM-ISAC-CNR, ECMWF-IFS, COSMO-2I, COSMO-5M • Observations: interpolated rainfall gauges values by Optimal Interpolation
03	Space and Time	• Spatial domain: Piedmont, Valle d'Aosta • Time step and horizon: 24h

PRE-PROCESSING: DATASET CLUSTERING



RESULTS





SUMMARY

Question Is a machine learning based multimodel ensemble a viable solution to predict deterministic and probabilistic precipitation fields?

Task Compare deterministic post processed forecast with Non negative Least Square, compute Reliability Diagrams for MC DO U-NET at different thresholds.

Results Deterministic CNNs outperforms NNLS in every season except spring. MC DO U-NET is not reliable at each considered threshold.

Discussion We have to increase the dataset dimension to get objectively unbiased results.

FOLLOW THIS GUIDE IF I'M NOT AROUND

Direct model output forecasts by Numerical Weather Prediction models (NWPs) present some limitations caused by errors mostly due to sensitivity to initial conditions, sensitivity to boundary conditions and deficiencies in parametrization schemes (i.e. orography).

Post-processing can help to remove those errors. In this work we propose a machine learning based multimodel approach [Zhang and Ye, 2021] of deterministic precipitation forecasts, that provides in output both deterministic and probabilistic postprocessed precipitation forecasts.

We consider 24h forecasts over 24h in Piedmont and Valle d'Aosta. Two NNLSs are considered with both shared and different weights for different grid cells, and we use FCNNs and CNNs as neural networks. We adopt dropout within CNNs training, to try to take overfitting into account. We also use Montecarlo Dropout in U-NET, to get probabilistic forecasts.

We use a dataset composed by 406 days from 2018 to 2022 with a relevant precipitation signal, which is clustered into convective, intermediate and stratiform events with k-means. After that, these clusters are used to uniformly split the dataset into training/validation/test set with 70/15/15 and 60/20/20 proportions 10 times each, to give robustness to the deterministic output analysis. We provide the probabilistic output analysis for just 1 of the 70/15/15 splits.

Deterministic CNNs outperform NNLs in every season except spring, MC DO U-NET provides good reliability for the probability of exceeding 5 to 100mm, but fails for 150mm.