



L A M P O

Lombardy-based
Advanced Meteorological
Predictions and Observations

V. Guglieri - S. Barindelli - M. Sangiorgio - E. Mehdipour

Nowcasting con reti neurali

Il sistema LAMPO

Progetto di:



POLITECNICO
MILANO 1863

In collaborazione con:



Fondazione
Politecnico
di Milano



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Finanziato da:

Fondazione
CARIPLO



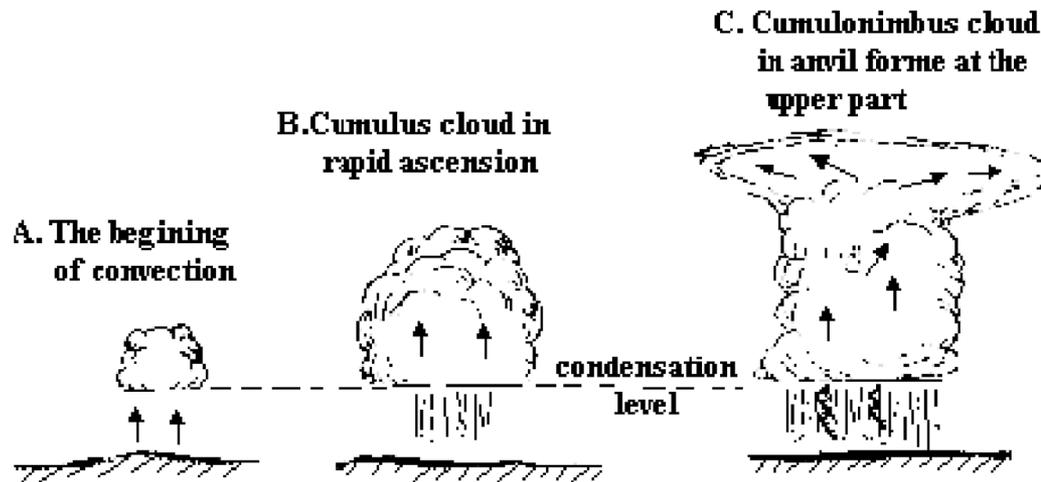


la predizione di piogge convettive

il fenomeno da predire è:

- fortemente localizzato nello spazio
- fortemente localizzato nel tempo

i modelli numerici di predizione spesso non riescono a prevedere tali fenomeni in quanto hanno una risoluzione spaziale e temporale troppo bassa



Lorle fasi di sviluppo di una pioggia convettiva



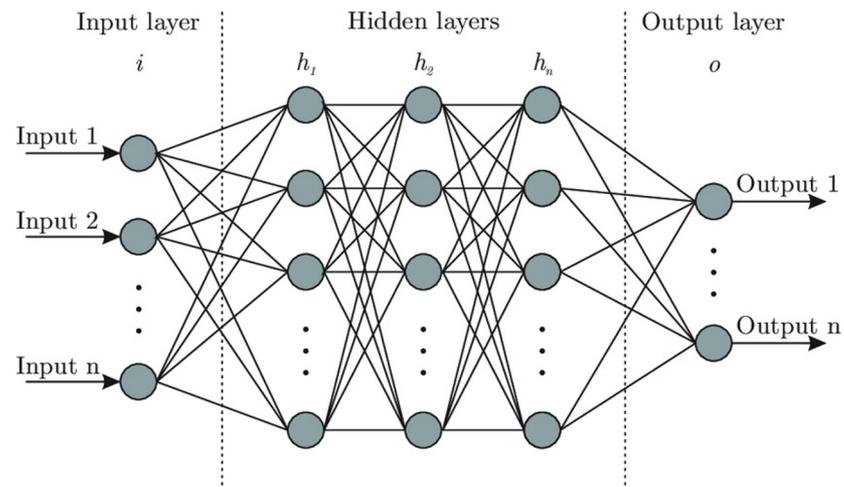
il sistema LAMPO

requisiti

- basato solo su osservazioni locali e 'dense'
- che utilizzi anche l'informazione GNSS sulla distribuzione (spazio-temporale) di vapore acqueo in atmosfera

soluzione

utilizzo di un sistema non fisicamente basato in grado di interpretare il comportamento fortemente non lineare dei dati





i dati LAMPO

osservazioni dirette da sensori meteorologici e prodotti dell'elaborazione di osservazioni GNSS e radar

Variabili	Sorgente	Periodo considerato	Frequenza	# di sensori
Temperatura	ARPA Lombardia	2010 - 2020	10 min	151
Pressione Atmosferica	ARPA Lombardia	2010 - 2020	10 min	61
Umidità relativa	ARPA Lombardia	2010 - 2020	10 min	114
Direzione del vento	ARPA Lombardia	2010 - 2020	10 min	113
Velocità del vento	ARPA Lombardia	2010 - 2020	10 min	95
ZTD da sensori geodetici	SPIN Network	2010 - 2020	30 sec	6
ZTD da sensori a basso costo	LAMPO	2019 e 2020	30 sec	9
Radar Storm Tracking	MeteoSwiss	2010 - 2020	5 min	-



deep learning

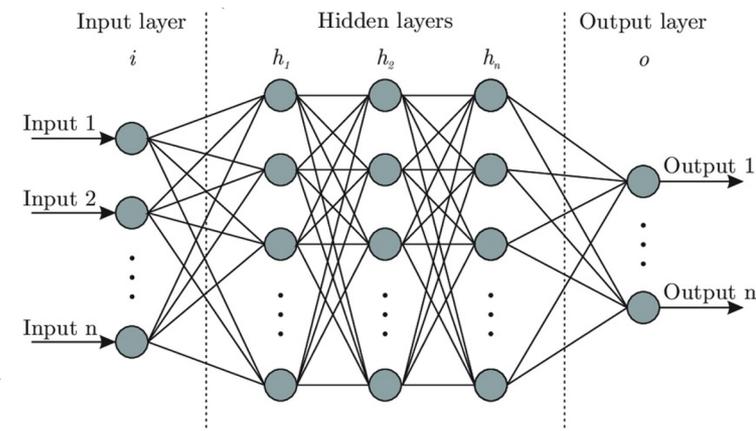
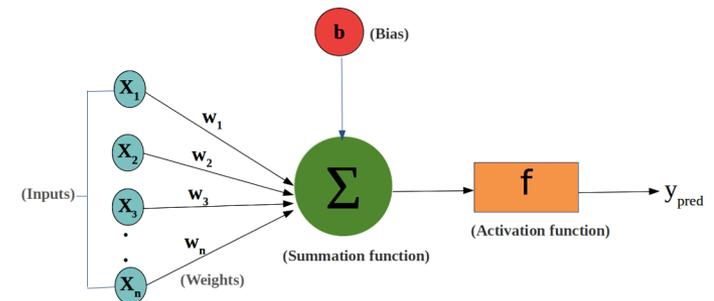
LAMPO si basa su un metodo di machine learning che effettua una classificazione binaria supervisionata attraverso una **FULLY CONNECTED FEED FORWARD NEURAL NETWORK**

Le reti neurali sono costituite da strati di neuroni o nodi. I nodi di ogni strato sono connessi a quelli dei strati adiacenti

Forward Propagation: le osservazioni subiscono delle trasformazioni (non lineari) nei nodi, e i valori che ne risultano vengono opportunamente pesati prima di passare allo strato successivo. Lo strato finale combina tutti i valori trasformati e opportunamente pesati per ottenere l'output/predizione.

Cost function: la predizione ottenuta su un dataset di training viene confrontata con il valore noto dell'output e una funzione di costo fornisce una misura della 'bontà' della predizione.

Backward Propagation: In base al risultato ottenuto si torna indietro a modificare opportunamente i pesi.





tuning di parametri e iperparametri

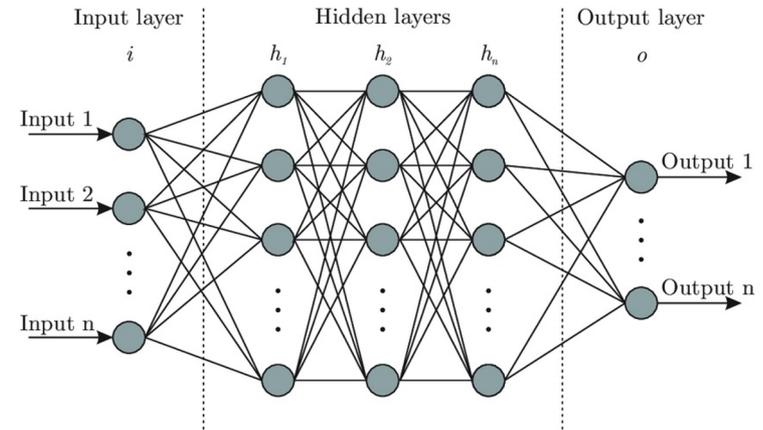
è la fase di training del modello basata sulle osservazioni di input e sul corrispondente output che in questa fase è noto

parametri:

- pesi e bias
- definiti durante la fase detta di training vero e proprio
- per LAMPO il training dataset va dal 2010 al 2016

iperparametri:

- sono quelli relativi alla struttura della rete: # di *hidden layers*, # di neuroni, dimensione dei *mini-batch*, learning rate
- La selezione del modello migliore avviene sempre sulla base dei dati, ma si usa un dataset detto di validazione
- per LAMPO il dataset di validazione è costituito dai dati del 2017



valutazione delle prestazioni:

- infine si stabiliscono le prestazioni del modello prescelto su un dataset non utilizzato nella sua definizione (dataset di test)
- per LAMPO è costituito dai dati del 2018



questioni ed esperimenti

relativi alla scelta e la successiva definizione del modello, ma anche relativi all'impatto dei dati di input sulla predizione

01

Scelta del modello: lineare o non lineare? (Sangiorgio, Barindelli et al, 2019a)

02

L'utilizzo del vapore acqueo da GNSS migliora la predizione o no? (Sangiorgio, Barindelli et al, 2019b)

03

Come ridurre la numerosità dei dati di input, tenendo conto anche della presenza di dati irregolari (con 'buchi')? (Sangiorgio, Barindelli et al, 2020)

04

Come affrontare il problema di un sistema che dispone di un learning dataset con pochi eventi di pioggia intensa (fenomeno poco frequente), dataset non bilanciato?

05

E' possibile migliorare le prestazioni del modello rimescolando i dati?

06

Qual è il contributo della rete LAMPO (con ricevitori ravvicinati) alla predizione?

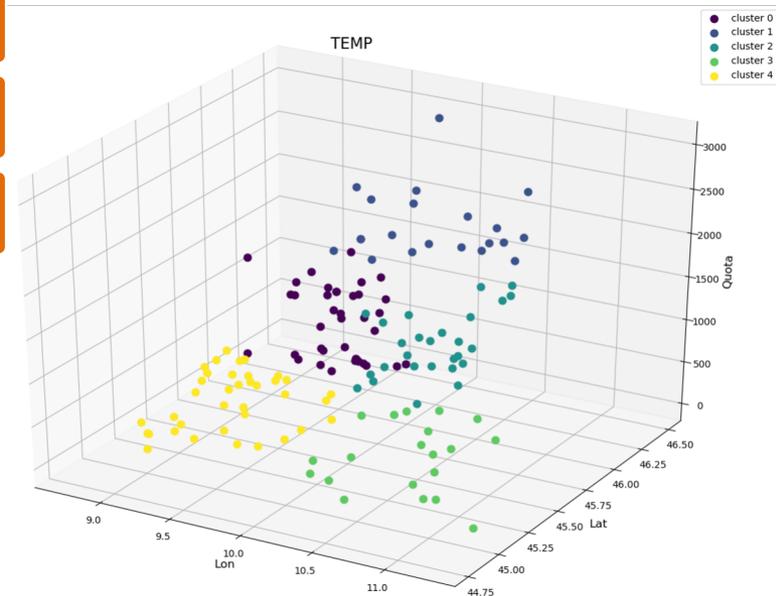
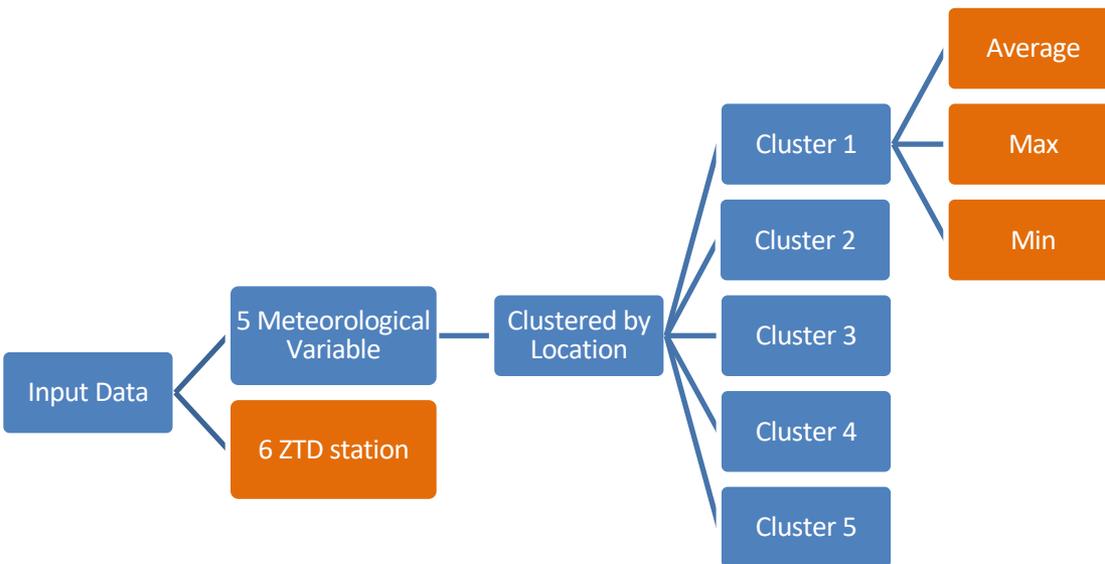


03 come ridurre la numerosità dei dati di input?

risolvendo anche il problema di dati con 'buchi'

più variabili vengono considerate, più dev'essere complessa la rete e più dati servono per ottenere prestazioni interessanti

i dati sono stati raggruppati sulla base della loro posizione riducendo il numero delle variabili meteorologiche di input a 5x3 valori per ciascuna epoca

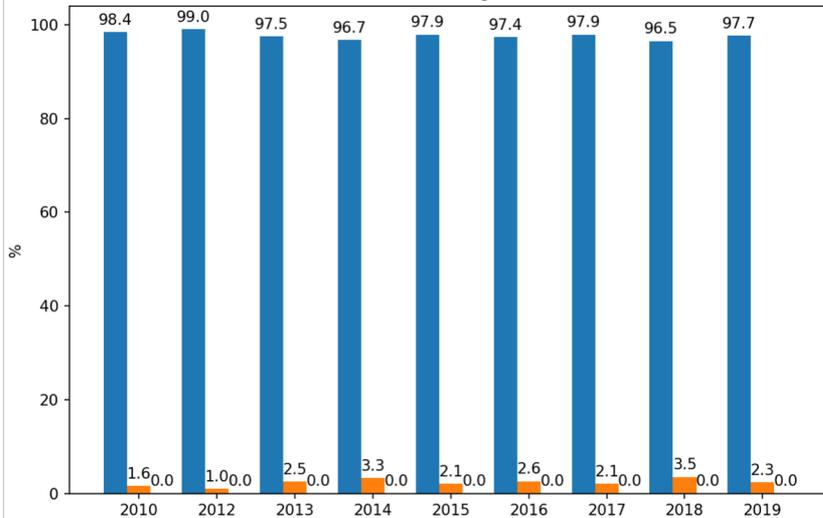




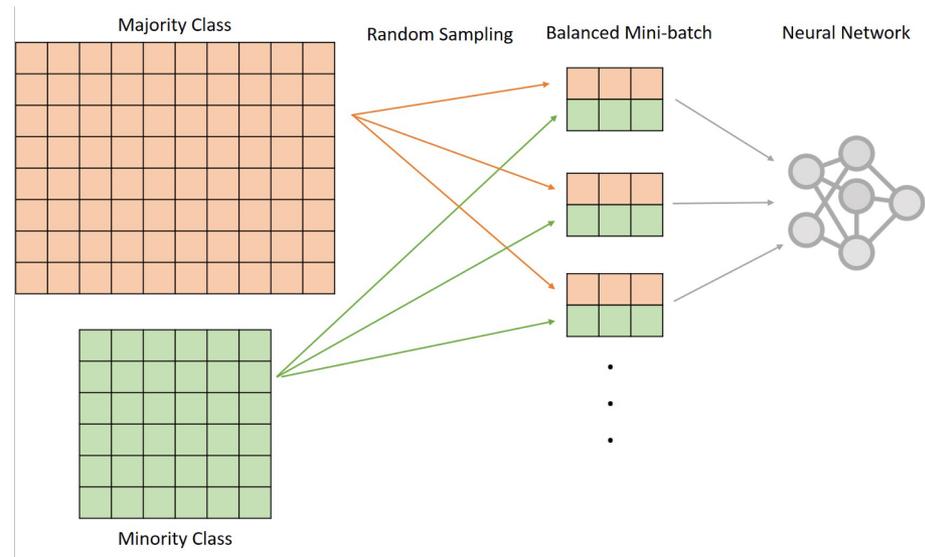
04 come risolvere la questione del dataset non bilanciato?

Non Eventi
Eventi Estremi

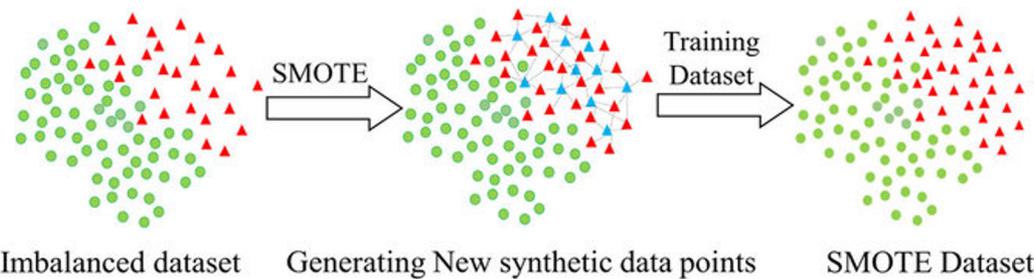
Distribution of events Percentage



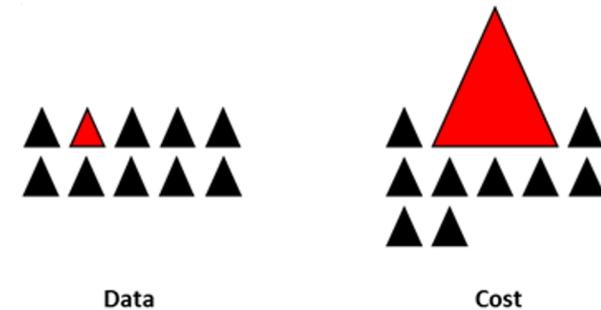
2. Balanced Mini-Batch



Synthetic Minority Oversampling Technique (SMOTE)



3. Cost-Sensitive Training

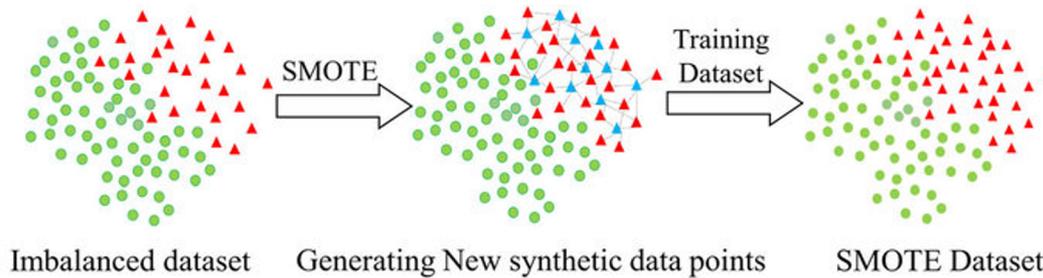


● Majority class data points ▲ Minority class data points ▲ Synthetic minority class data points



04 come risolvere la questione del dataset non bilanciato?

I. Synthetic Minority Oversampling Technique (SMOTE)



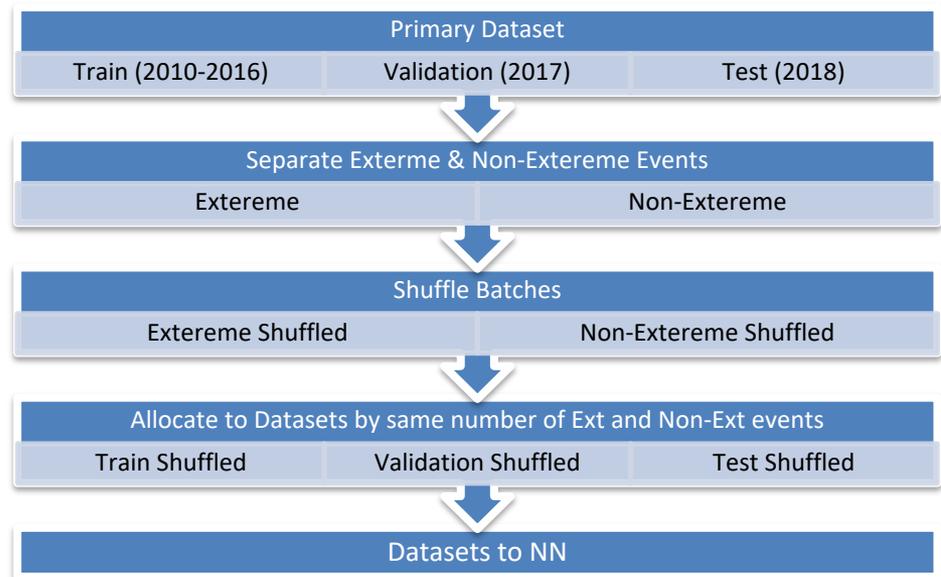
● Majority class data points ▲ Minority class data points ▲ Synthetic minority class data points

Esperimento:	Training set accuracy	Validation set accuracy	Test set accuracy	# training samples	Nr di parametri	rapporto
I8) tutte le variabili, classificazione dei dati di input in base alla posizione (coordinate normalizzate), min/media/max per ogni cluster	0.91	0.82	<u>0.81</u>	49587	10592	4.68

nell'81% dei casi il sistema predice correttamente



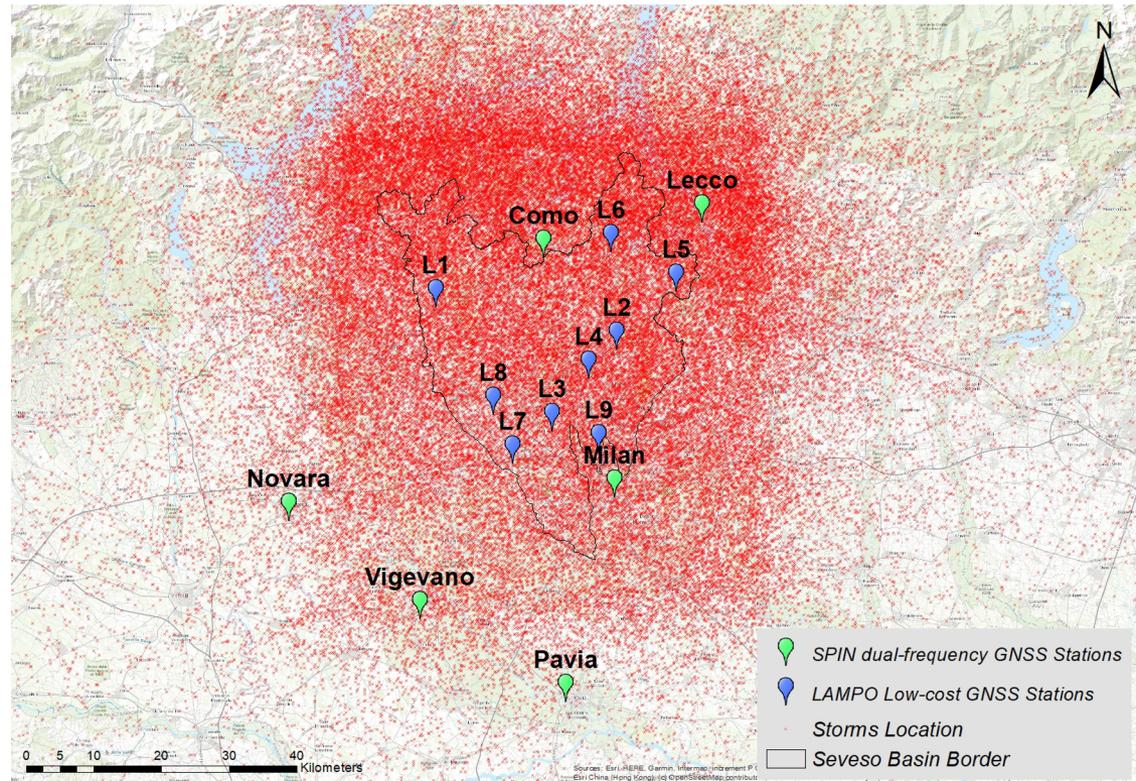
05 è possibile migliorare le prestazioni del sistema rimescolando i dati?



Esperimento:	Training set accuracy	Validation set accuracy	Test set accuracy	# training samples	# parametri	Ratio
33) dati rimescolati e creazione di insiemi con la stessa percentuale di eventi estremi e non estremi	0.93	0.85	0.87	49587	9602	5.16



06 qual è il contributo della rete LAMPO (con ricevitori ravvicinati) alla predizione?





06 qual è il contributo della rete LAMPO (con ricevitori ravvicinati) alla predizione?

LAMPO low-cost GNSS ZTD

Experiment:	Training set accuracy	Validation set accuracy	Test set accuracy	# training samples	# parametri	Ratio	# test samples	# Ext series in test
40) solo ZTD da rete LAMPO	0.70	0.71	0.71	16206	9002	1.8	793	11

LAMPO low-cost GNSS ZTD + Variabili meteorologiche standard

Experiment:	Training set accuracy	Validation set accuracy	Test set accuracy	# training samples	# parametri	Ratio	# test samples	# Ext series in test
36) ZTD LAMPO + dati meteo	0.98	0.92	0.87	19500	16832	1.16	691	15



in sintesi

- 01** includere le stime di vapore acqueo da GNSS nella predizione incrementa l'accuratezza di almeno il 4%
- 02** è possibile predire i temporali con mezz'ora di anticipo con una accuratezza dell'81%.
- 03** l'accuratezza può migliorare all'87% rimescolando i dati di training
- 04** l'utilizzo dei dati da rete GNSS locale è promettente

suggerimenti

- 01** la qualità dei dati di input è fondamentale per il successo della predizione
- 02** utilizzare architetture neurali più avanzate (ricorrenti o convoluzionali)



verso l'utilizzo operativo del sistema LAMPO

- 01** dati meteo sono disponibili in tempo quasi reale
- 02** prodotti di vapore GNSS possono essere aggiornati ogni ora e resi disponibili
- 03** il training della rete può essere aggiornato periodicamente in modo da includere le nuove osservazioni (i dati utili al training non sono richiesti in tempo reale, specialmente i dati da radar per definire la presenza di temporali nell'area considerata)
- 04** GNSS (prodotti) e dati meteo (osservazioni) vengono attualmente forniti con latenze diverse: modifiche sul modello di predizione o sui dati di input
- 05** si può ipotizzare l'emissione di un warning ogni ora che predice se nei prossimi 30 minuti il temporale ci sarà oppure no

altri esperimenti sullo stesso modello

sensibilità della predizione al variare della definizione di evento di pioggia intensa

hardware

la fase critica è quella del training, non quella di previsione



bibliografia

Sangiorgio, M., Barindelli, S., Guglieri, V., Biondi, R., Solazzo, E., Realini, E., Venuti, G., and Guariso, G. **(2019a)**. A comparative study on machine learning techniques for intense convective rainfall events forecasting. *Theory and Applications of Time Series Analysis*. Springer, Cham.

Sangiorgio, M., Barindelli, S., Biondi, R., Solazzo, E., Realini, E., Venuti, G., and Guariso, G. **(2019b)**. Improved extreme rainfall events forecasting using neural networks and water vapor measures. In 6th International conference on Time Series and Forecasting.

Sangiorgio, M., Barindelli, S., Guglieri, V., Venuti, G., and Guariso, G. **(2020)**. Reconstructing Environmental Variables with Missing Field Data via End-to-End Machine Learning. In International Conference on Engineering Applications of Neural Networks.

giovanna.venuti@polimi.it
matteo.sangiorgio@polimi.it
valerio.guglieri@gmail.com
stefano.barindelli@g-red.eu
ehsan.mehdipour@polimi.it



www.lampo.polimi.it



www.facebook.com/Lampo.Polimi



www.twitter.com/Lampo Polimi