

Un metodo di postelaborazione modellistica basato su machine learning per la previsione dei temporali

Daniele D'Alessandro, Anna Fornasiero, Andrea Selvini, Davide Cesari
Arpae-SIMC, Servizio Idro Meteo Clima dell'Emilia-Romagna, Bologna, Italia

Introduzione

I temporali sono fenomeni convettivi che si sviluppano generalmente su scale spazio-tempo limitate e ai quali possono essere associate condizioni meteorologiche estreme come ad esempio forti raffiche di vento, intense grandinate, elevate quantità di pioggia e attività elettrica sostenuta. Proprio la ridotta estensione nello spazio e nel tempo rende la previsione dei temporali eccezionalmente incerta, sia in termini di localizzazione che di intensità dei massimi di precipitazione prodotti.

Obiettivi

In questo lavoro è stato applicato un sistema di apprendimento basato su machine learning ad un insieme di predittori modellistici meteorologici con l'obiettivo di formulare una previsione dei fenomeni temporaleschi con maggiore accuratezza rispetto all'uso dell'uscita diretta del modello stesso.

Fase di apprendimento

Per la previsione di fenomeni temporaleschi sul territorio nazionale è stato utilizzato il sistema **XGBoost** (eXtreme Gradient Boosting) alimentato da uno storico dati costituito da predittori e predittando, essenziale per la fase di apprendimento dell'algoritmo.

Predittori → serie di indici meteorologici derivati dall'archivio delle analisi tridimensionali del modello previsionale COSMO:

•Avv.Geop. 500 hPa •L.I. •V.V.300 hPa •V.V.700 hPa •K Index •TWC •Jet 250 hPa •Jet 925 hPa •V.W.S. 500-950 hPa •V.W.S. 700-1000 hPa
•CAPE •CIN •MCSindex •R.H. 500 hPa •Avv.T 500 hPa •Avv.Td 850 hPa •Vort.Rel. 500 hPa

Predittando → dataset di fulminazioni registrate dalla rete Lampinet. I dati sono stati considerati sia in termini di numero di scariche per macroarea che in termini di densità spaziale media di fulminazioni per macroarea.

Periodo di analisi → Novembre 2014 - Maggio 2017

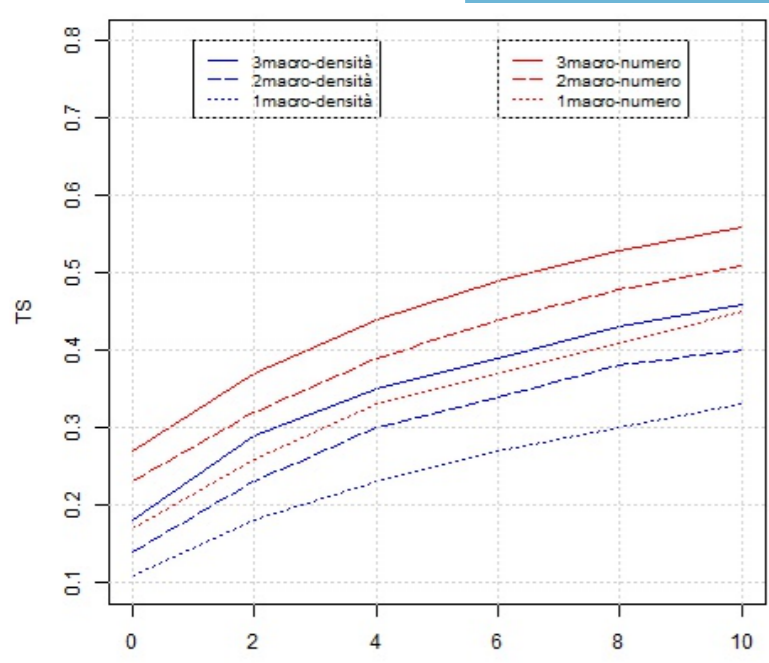
Aggregazione dati → nel tempo su intervalli orari e nello spazio su un insieme di aree nazionali di allertamento (area media 3.000 km²);

Ripartizione dataset → 80% training set - 20% validation set

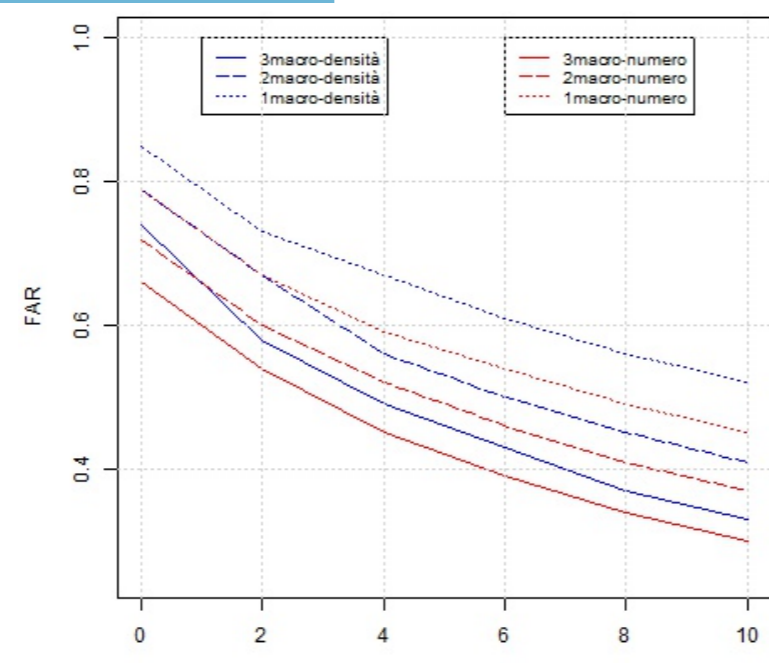


Aree nazionali di allertamento

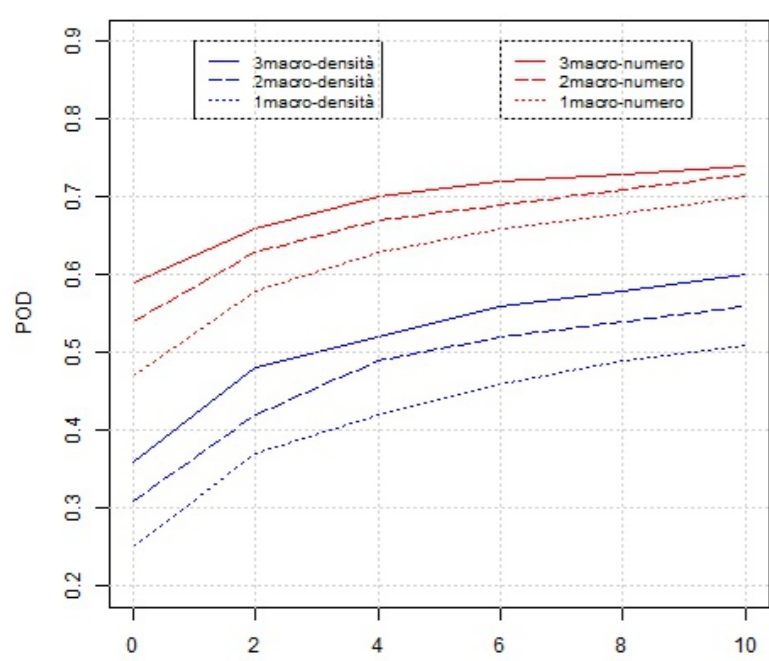
Score sulla nazione



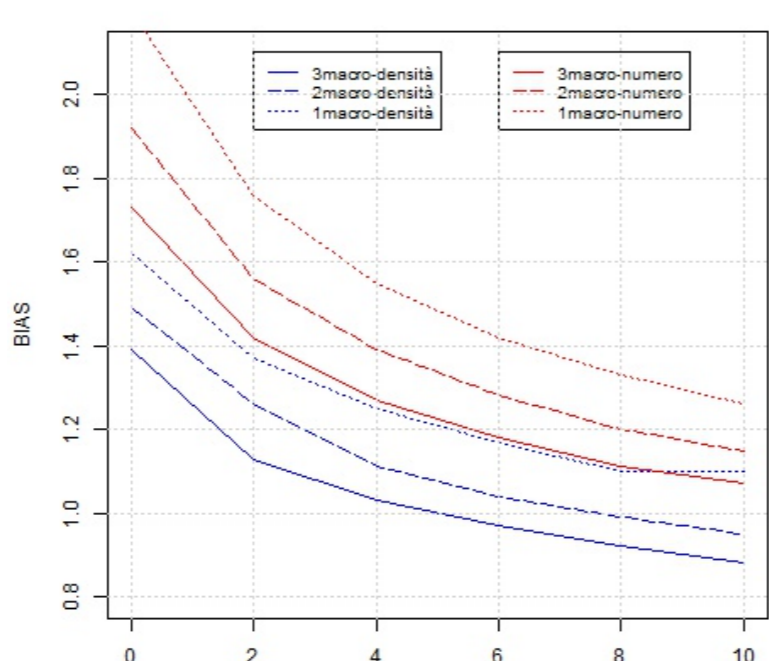
Threat Score in funzione del gap spazio-temporale utilizzato per la validazione



False Alarm Ratio in funzione del gap spazio-temporale utilizzato per la validazione

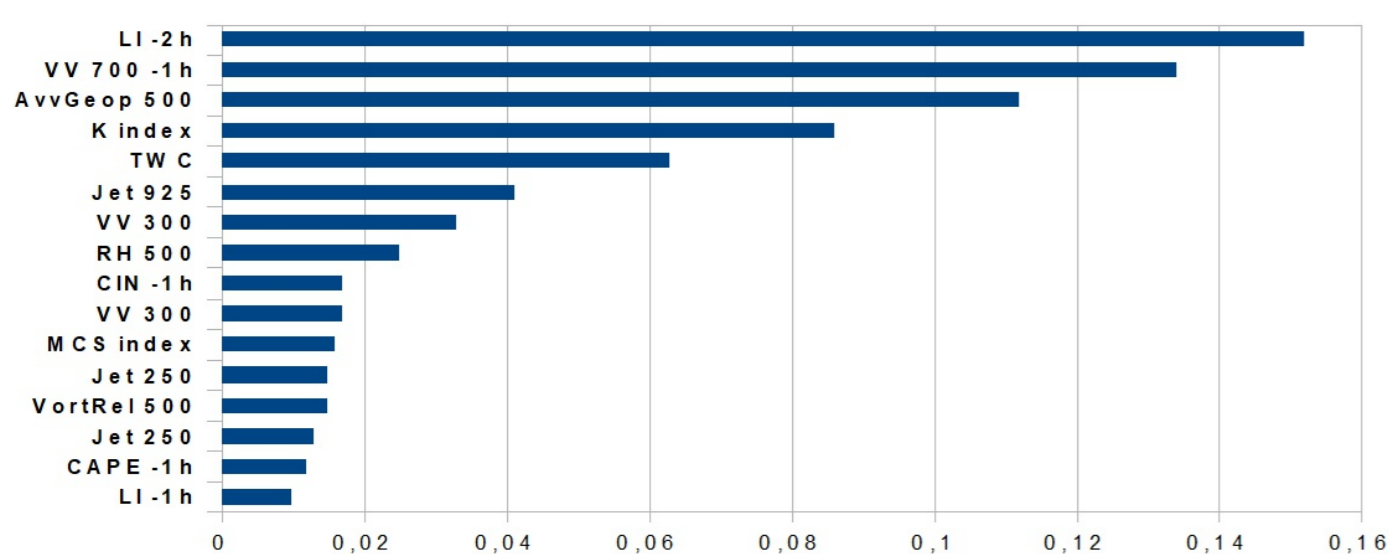


Probability Of Detection in funzione del gap spazio-temporale utilizzato per la validazione



Bias in funzione del gap spazio-temporale utilizzato per la validazione

Peso predittori



I predittori con maggior peso nella previsione dei temporali sono risultati essere: **Lifted Index, Velocità Verticale a 700 hPa, Avvezione di Geopotenziale a 500 hPa e K index**

Numero Macroarea	$\Delta t(h)$												
	0	2	4	6	8	10	12						
1	0,17	0,26	0,33	0,37	0,41	0,45	0,47						
2	0,23	0,32	0,39	0,44	0,48	0,51	0,54						
3	0,27	0,37	0,44	0,49	0,53	0,56	0,59						
4	0,30	0,41	0,47	0,52	0,56	0,59	0,61						
5	0,33	0,43	0,50	0,55	0,58	0,61	0,64						
6	0,35	0,45	0,52	0,56	0,60	0,63	0,65						

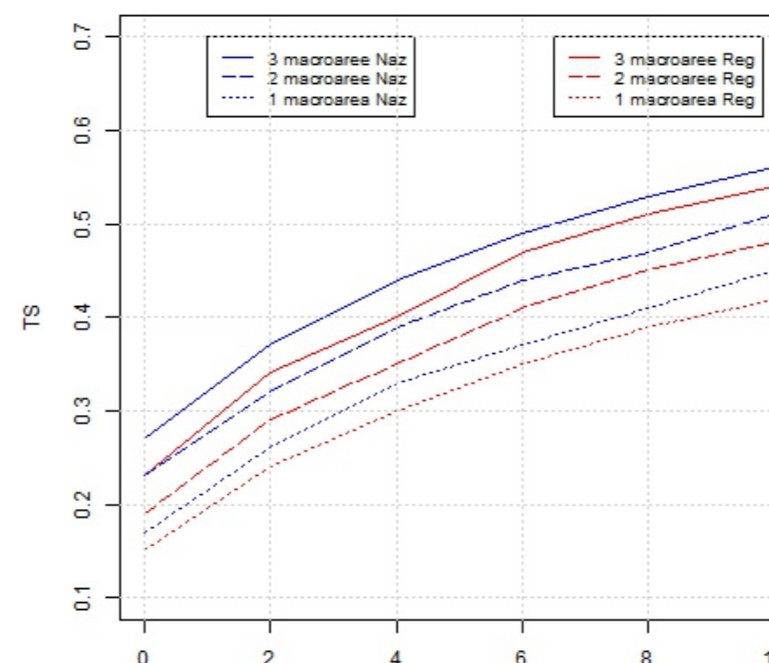
Numero Macroarea	$\Delta t(h)$												
	0	2	4	6	8	10	12						
1	0,79	0,67	0,59	0,54	0,49	0,45	0,42						
2	0,72	0,60	0,52	0,46	0,41	0,37	0,34						
3	0,66	0,54	0,45	0,39	0,34	0,30	0,27						
4	0,62	0,50	0,41	0,36	0,31	0,27	0,24						
5	0,59	0,46	0,39	0,32	0,28	0,24	0,21						
6	0,56	0,44	0,36	0,30	0,25	0,22	0,19						

Variatione spazio-tempo di TS, FAR, POD e BIAS

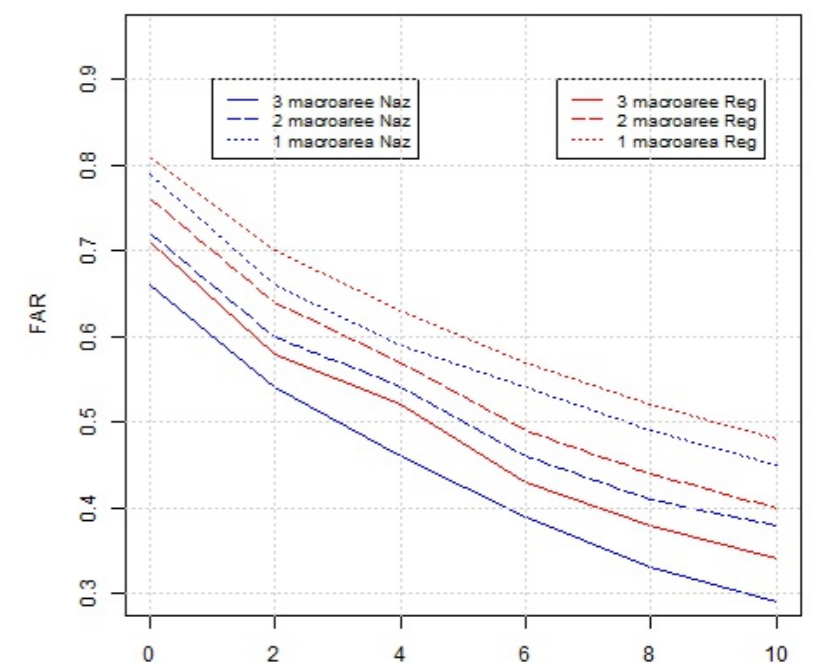
Numero Macroarea	$\Delta t(h)$												
	0	2	4	6	8	10	12						
1	0,47	0,58	0,63	0,66	0,68	0,70	0,71						
2	0,54	0,63	0,67	0,69	0,71	0,73	0,73						
3	0,59	0,66	0,70	0,72	0,73	0,74	0,75						
4	0,61	0,68	0,71	0,73	0,74	0,75	0,76						
5	0,63	0,69	0,72	0,74	0,75	0,76	0,77						
6	0,65	0,70	0,73	0,74	0,76	0,77	0,77						

Numero Macroarea	$\Delta t(h)$												
	0	2	4	6	8	10	12						
1	2,20	1,76	1,55	1,42	1,33	1,26	1,22						
2	1,92	1,56	1,39	1,28	1,20	1,15	1,11						
3	1,73	1,42	1,27	1,18	1,11	1,07	1,03						
4	1,62	1,35	1,21	1,13	1,07	1,03	1,00						
5	1,54	1,29	1,17	1,09	1,04	1,00	0,97						
6	1,48	1,25	1,14	1,07	1,01	0,98	0,95						

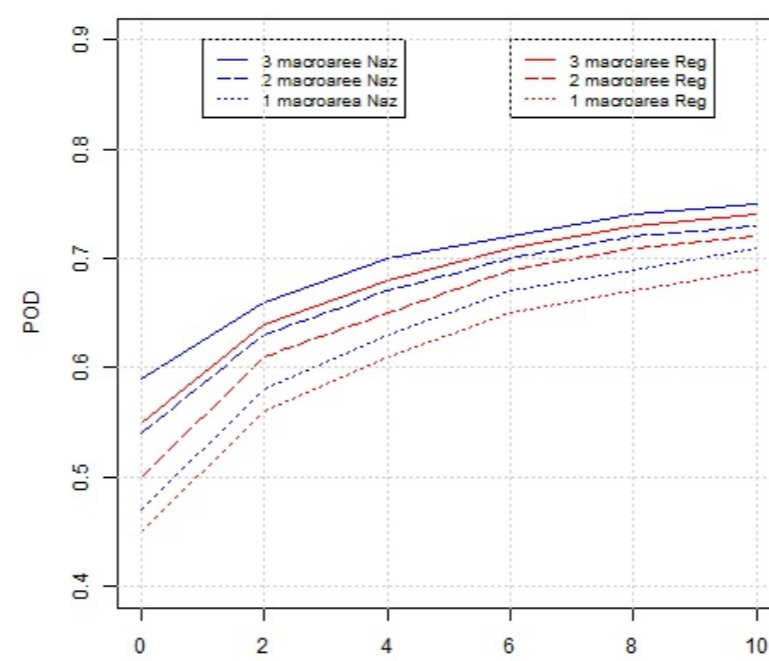
Score sull'Emilia-Romagna



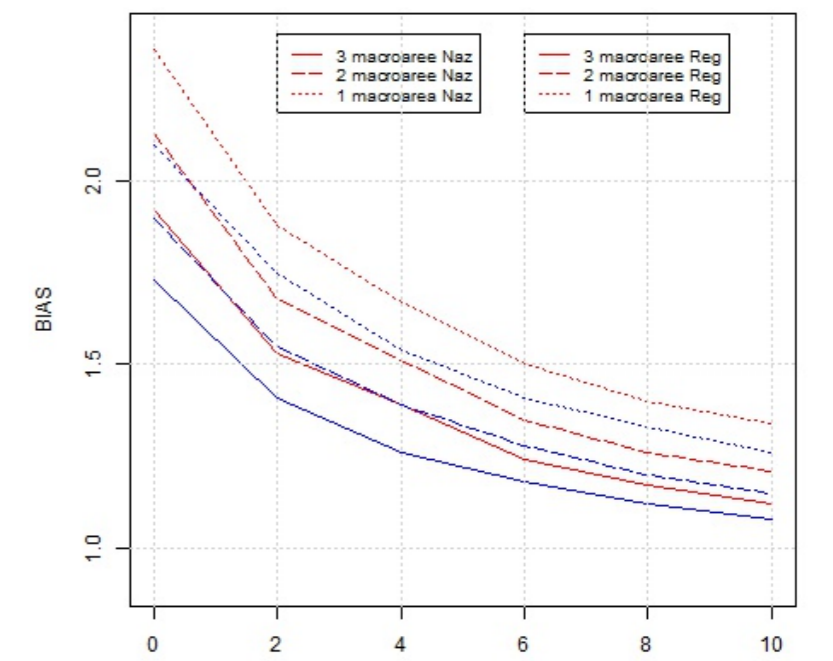
Threat Score



False Alarm Ratio



Probability Of Detection



Bias

Dataset più piccolo → minore apprendimento → risultati peggiori

Conclusioni e sviluppi futuri

Conclusioni

- il numero di scariche per macroarea risulta essere il predittando che porta a risultati migliori;
- la qualità dei risultati è strettamente connessa alla quantità di dati utilizzati per allenare l'algoritmo;
- gli score ottenuti mostrano una buona capacità di previsione di occorrenza di temporali, sebbene con lacune evidenti sulle scale spazio-temporali limitate.

Sviluppi futuri

- ampliamento del dataset per un miglior apprendimento dell'XGBoost;
- sviluppo di un sistema di previsione multi-class per permettere una distinzione dei temporali in termini di intensità;
- analisi della correlazione esistente tra i dati di fulminazione ed i dati di grandine stimati da radar (con particolare riguardo ai rapporti CG+/CG- e CG/IC).