



UNIONE EUROPEA
Fondo Europeo di Sviluppo Regionale



Ministero dell'Università
e della Ricerca



*Mitigazione dei rischi naturali per la sicurezza e la mobilità
nelle aree montane del Mezzogiorno: Il contributo dei Collaboratori di Ricerca*

SVILUPPO E APPLICAZIONE DI UN MODELLO CLIMATICO INTEGRATO DI DOWNSCALING STATISTICO E MACHINE LEARNING

Responsabile Scientifico: Prof. Vito Telesca

Scuola di Ingegneria (SI – UNIBAS)

Relatore: Danilo Mazzilli

Ruolo: Contrattista

15 Gennaio 2024

INTRODUZIONE

Negli ultimi anni, la crescente consapevolezza che l'emissione di gas serra nell'atmosfera potrebbe portare ad un **aumento dell'intensità e della frequenza degli eventi estremi** ha messo in discussione le ipotesi di *casualità, stazionarietà e analisi del processo limitato alla scala del bacino*, che sono alla base degli studi idrologici/idraulici.

Si rende necessario dunque ampliare l'approccio metodologico che include processi deterministici su scala globale, come l'evoluzione temporale e spaziale delle strutture di circolazione atmosferica e i processi di trasporto dell'umidità su scala planetaria.

(Merz et al., 2014)

INTRODUZIONE

La Commissione Europea sottolinea l'importanza dello sviluppo di metodi e strumenti predittivi nel Libro Bianco sull'Adattamento ai Cambiamenti Climatici.

La metodologia centrale utilizzata nella valutazione delle conseguenze dei cambiamenti climatici sui processi idrologici consiste nei seguenti passi:

- **Valutare** le prestazioni dei Modelli di Circolazione Generale (GCM) confrontando le simulazioni con i dati generati dai modelli di rianalisi per selezionare il più adatto al nostro caso di studio.
- **Implementare** tecniche di downscaling che associano gli output dei GCM identificati ai regimi di pioggia osservati a scala locale.
- **Sviluppare** modelli che simulano a scala di bacino gli effetti dei cambiamenti climatici.

INTRODUZIONE

Si indagherà l'efficacia dei modelli di stacking per migliorare l'approccio già utilizzato in (Conticello et al., 2018) sviluppando un modello di classificazione a monte dell'NHMM, secondo schemi in grado di catturare la non linearità intrinseca delle dipendenze tra le precipitazioni e i predittori atmosferici, che aumenta sia l'efficacia nel riconoscere le fluttuazioni stagionali sia la capacità di catturare accuratamente gli estremi delle precipitazioni.

Lo scopo di questo documento è contribuire allo sviluppo di uno strumento per l'analisi e la previsione degli eventi idrologici estremi e non estremi a livello locale che tiene conto dell'impatto dei cambiamenti climatici e della non stazionarietà delle serie naturali.

In questo studio tale procedura verrà implementata dapprima per la regione Lazio e successivamente per la Basilicata.

DATI

Per la costruzione del modello sono stati utilizzati dati atmosferici e pluviometrici

È stata considerata la serie storica relativa ai campi di variabili atmosferiche, con una risoluzione spaziale di $0,28125^\circ \times 0,28125^\circ$, considerando latitudini tra 80N e 20S e longitudini tra -90W e 70E (dataset di rianalisi "ERA5" dell'ECMWF)

In dettaglio, le variabili atmosferiche considerate includono:

- 1) Altezza di geopotenziale a 850hPa (GPH850);**
- 2) Venti zionali e meridionali e umidità specifica misurata tra i livelli di 1000hPa e 300hPa.**

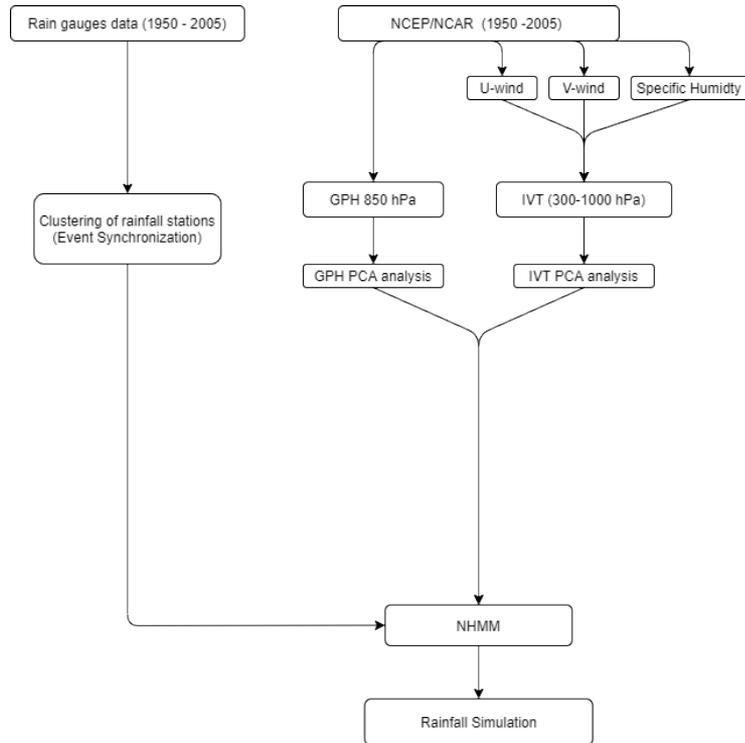
Il punto 2 è stato essenziale per calcolare il Trasporto Integrato del Vapore (IVT). Il calcolo dell'IVT ha permesso di quantificare il trasporto orizzontale di umidità nella colonna d'aria tra 1000 e 300 hPa.

METODOLOGIA

È stata effettuata una comparazione tra due metodi per la simulazione delle altezze delle precipitazioni.

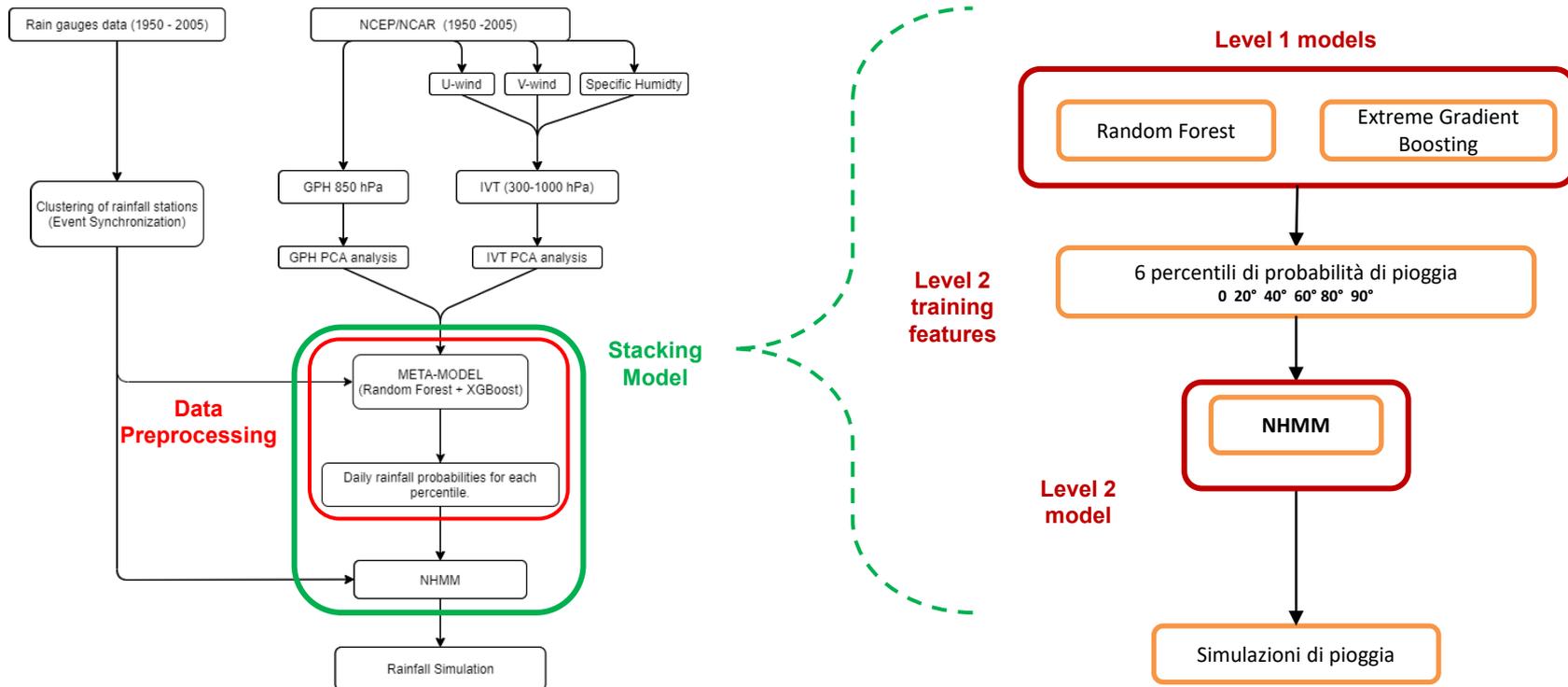
- Il primo è il metodo classico utilizzato in precedenti lavori di ricerca che impiega direttamente predittori atmosferici nel NHMM.
- Il secondo, invece, rappresenta un nuovo metodo che prevede l'uso di un modello di Stacking a monte del NHMM. Per la simulazione delle precipitazioni, questo non sfrutta più predittori atmosferici, ma le probabilità giornaliere di pioggia relative a sei percentili derivanti dall'output dello Stacking.

METODOLOGIA



Schema proposto da Conticello et al.2018 (Conticello et al., 2018)

METODOLOGIA



DATA PROCESSING

✓ Clustering

- Classificazione delle precipitazioni
- Analisi PCA
- Modello Stacking
- Modelli di Livello-1: Random Forest e Extreme Gradient Boosting
- Modello di Livello-2: NHMM

DATA PROCESSING

La suddivisione in cluster di stazioni con condizioni meteorologiche locali simili si effettua utilizzando le tecniche di Event Synchronization e Modularity. Questa metodologia è stata applicata alle serie temporali con una finestra temporale pari a ± 1 giorno garantendo che gli HPEs, registrati in diversi pluviometri e che mostrano un alto grado di sincronizzazione tra loro, siano causati dalle stesse caratteristiche meteorologiche.

La formula principale che descrive l'approccio è:

$$Q^c = \frac{s^c(x|y) + s^c(y|x)}{m_x + m_y}$$

- Q rappresenta la sincronizzazione tra due serie temporali binarie, denominate x e y , in una finestra temporale τ ;
- $s(x|y)$ e $s(y|x)$ sono il numero di volte in cui gli eventi della serie temporale y hanno influenzato la serie temporale x e viceversa;
- m_x e m_y sono il numero di eventi nelle serie temporali x e y .

DATA PROCESSING

- Clustering
- ✓ **Classificazione delle precipitazioni**
- Analisi PCA
- Modello Stacking
- Modelli di Livello-1: Random Forest e Extreme Gradient Boosting
- Modello di Livello-2: NHMM

DATA PROCESSING

Gli accumuli giornalieri sono stati trasformati in sei classi di precipitazione relative a **6 percentili**, elaborati per ogni cluster di stazioni omogenee.

La Classe 0 rappresenta valori non-piovisi o altezze di precipitazione giornaliera inferiori a 1 mm, per le altre si passa dal 20° al 90° percentile:

- 1) classe 0: No pioggia
- 2) classe 1: 20° percentile
- 3) classe 2: 40° percentile
- 4) classe 3: 60° percentile
- 5) classe 4: 80° percentile
- 6) classe 5: 90° percentile;

DATA PROCESSING

- Clustering
- Classificazione delle precipitazioni
- Analisi PCA**
- Modello Stacking
- Modelli di Livello-1: Random Forest e Extreme Gradient Boosting
- Modello di Livello-2: NHMM

DATA PROCESSING

I dati dei predittori atmosferici sono stati analizzati tramite la tecnica **PCA (Principal Component Analysis)** per **ridurre le dimensioni dei campi di input atmosferici** per un addestramento più semplice e una strutturazione del modello. La capacità di ridurre la dimensionalità dei dati consente di accelerare le analisi per elaborazioni più veloci e gestibili, inoltre i risultati sono di facile visualizzazione, la trasformazione dei dati in spazi a due o tre dimensioni rende possibile una rappresentazione grafica intuitiva. Mitigazione della multicollinearità, capacità di rilevare pattern e tendenze dominanti nei dati e riduzione del rumore sono ulteriori fattori positivi che rendono questa tecnica decisamente efficace.

DATA PROCESSING

- Clustering
- Classificazione delle precipitazioni
- Analisi PCA
- Modello Stacking**
- Modelli di Livello-1: Random Forest e Extreme Gradient Boosting
- Modello di Livello-2: NHMM

DATA PROCESSING

I campi atmosferici ridotti grazie alla tecnica PCA costituiranno l'input del modello Stacking.

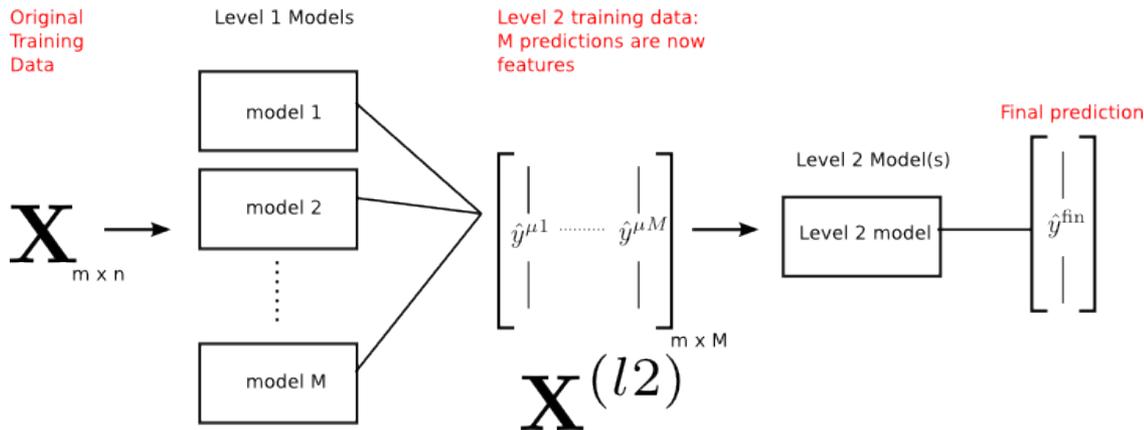
Lo Stacking è un algoritmo di apprendimento automatico basato su ensemble che apprende come combinare al meglio le previsioni provenienti da diversi modelli di apprendimento automatico che offrono buone prestazioni. Nei metodi statistici e di apprendimento automatico, i metodi di ensemble utilizzano diversi algoritmi di apprendimento per ottenere una performance predittiva migliore rispetto a quella che potrebbe essere ottenuta da ciascuno dei singoli algoritmi di apprendimento.

L'architettura di un modello di Stacking coinvolge due o più modelli di base e un meta-modello che combina le previsioni dei modelli di base.

- **Modelli di Livello-1 (o Modelli di Base):** Modelli adattati sui dati di addestramento e le cui previsioni vengono compilate.
- **Modello di Livello-2 (o Meta-Modello):** Modello che apprende come combinare al meglio le previsioni dei modelli di base.

DATA PROCESSING

- I dati di addestramento iniziali X hanno m osservazioni e n campi.
- Ci sono M modelli diversi addestrati su X .
- Ogni modello fornisce previsioni per l'outcome (y) che vengono poi inserite in un secondo set di dati di addestramento $X^{(2)}$.
- Le M previsioni diventano caratteristiche per questi dati di secondo livello.
- Un modello di secondo livello (o modelli) può quindi essere addestrato su questi dati per produrre gli outcome finali che saranno utilizzati per le previsioni.



DATA PROCESSING

- Clustering
- Classificazione delle precipitazioni
- Analisi PCA
- Modello Stacking
- Modelli di Livello-1: Random Forest e Extreme Gradient Boosting**
- Modello di Livello-2: NHMM

DATA PROCESSING

Lo Stacking è appropriato quando diversi modelli di machine learning sono efficaci su un dataset ma lo sono in modi diversi. È una buona idea utilizzare una gamma di modelli che fanno ipotesi molto diverse su come risolvere il compito di modellazione predittiva. In questo caso si è utilizzato il **bagging (Random Forest)** per ridurre la varianza delle previsioni e l'approccio **boosting (Extreme Gradient Boosting)** per produrre un modello robusto meno incline agli errori dei suoi componenti.

L'output consiste in probabilità di pioggia giornaliera, divise in 6 classi, ognuna associata a un percentile specifico in modo che la somma delle probabilità di pioggia giornaliera di tutte e sei sia uguale a 1.

Questo output è stato utilizzato come input per il modello NHMM.

DATA PROCESSING

- Clustering
- Classificazione delle precipitazioni
- Analisi PCA
- Modello Stacking
- Modelli di Livello-1: Random Forest e Extreme Gradient Boosting
- Modello di Livello-2: NHMM**

DATA PROCESSING

L'NHMM è un modello doppiamente stocastico in cui vengono generate serie temporali multivariate, utilizzando una specifica distribuzione, condizionalmente a uno stato nascosto identificato.

Uno stato nascosto rappresenta una particolare configurazione meteorologica che influenza la probabilità e la quantità di precipitazione in un certo numero di luoghi simultaneamente.

Sono stati addestrati modelli diversi per determinare le migliori variabili di input e i loro parametri. Ogni modello identifica cinque stati nascosti, sono stati selezionati come ottimali proprio cinque stati nascosti e tramite l'algoritmo di Viterbi si è identificata la sequenza più probabile di stati nascosti associati alla sequenza di osservazioni.

DATA PROCESSING

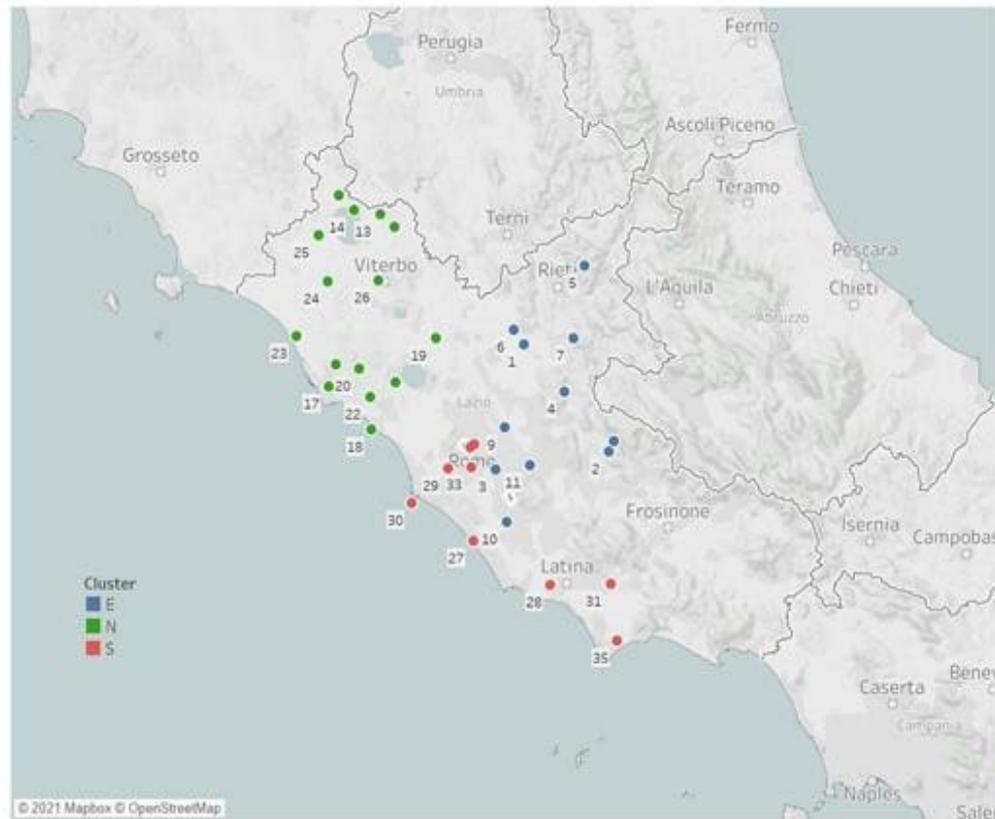
Sono stati identificati gli stati che presentavano le caratteristiche di un evento meteorico **estremo**. Per ogni cluster, verranno elaborate cento simulazioni di pioggia utilizzando i modelli creati in questo processo.

I risultati ottenuti sono stati poi confrontati con l'uso diretto dei predittori atmosferici all'interno dell'NHMM senza applicare lo Stacking.

APPLICAZIONE E RISULTATI PER LA REGIONE LAZIO

Per questo studio è utilizzato un archivio di 55 anni (1951-2005) di rilevazioni giornaliere di precipitazione, raccolte da 36 stazioni relative a differenti bacini della regione del Lazio.

La figura sottostante mostra i tre cluster di pluviometri. I cluster sono situati nelle aree costiere settentrionali e meridionali e nella parte montuosa orientale della regione. E' chiara l'omogeneità spaziale.



APPLICAZIONE E RISULTATI PER LA REGIONE LAZIO

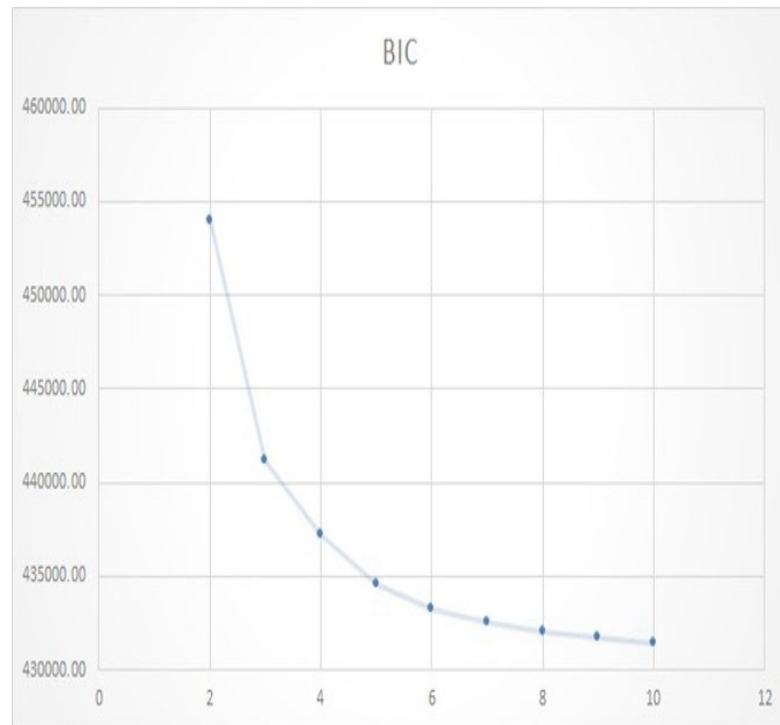
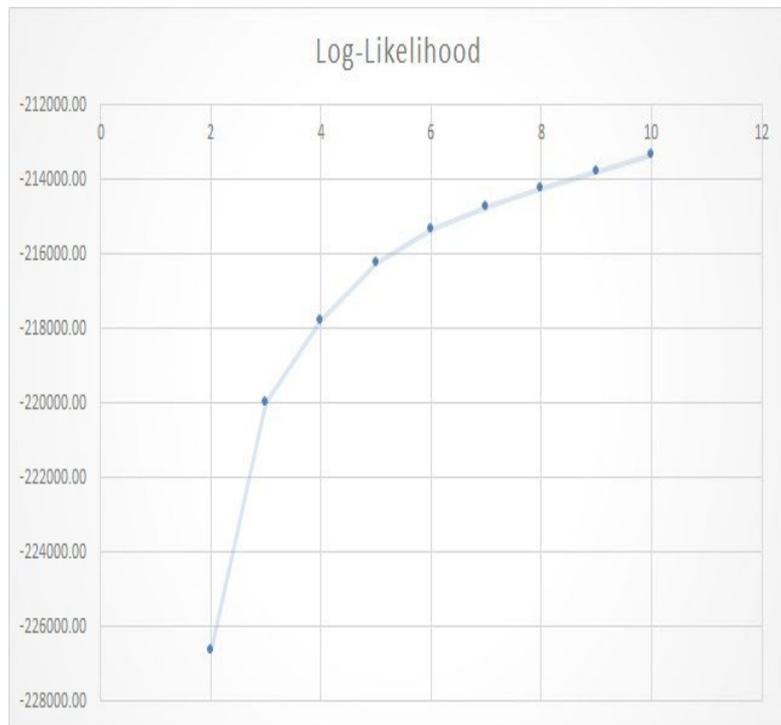
La tecnica **PCA** ha permesso di ottenere 20 componenti del GPH e 13 dell'IVT.

Le 33 PCAs così elaborate, che rappresentano il 90% della varianza totale del campione, costituiscono i predittori utilizzati all'interno del modello Stacking.

I risultati preliminari del modello NHMM mostrano che cinque stati nascosti erano ottimali per tutti e tre i cluster.

Come si può vedere dai due grafici sottostanti, l'incremento percentuale delle prestazioni del modello, data dalla pendenza della curva, diminuisce all'aumentare del numero di stati nascosti; si è scelto di utilizzare cinque stati nascosti a causa del compromesso tra prestazioni e tempo di calcolo.

APPLICAZIONE E RISULTATI PER LA REGIONE LAZIO



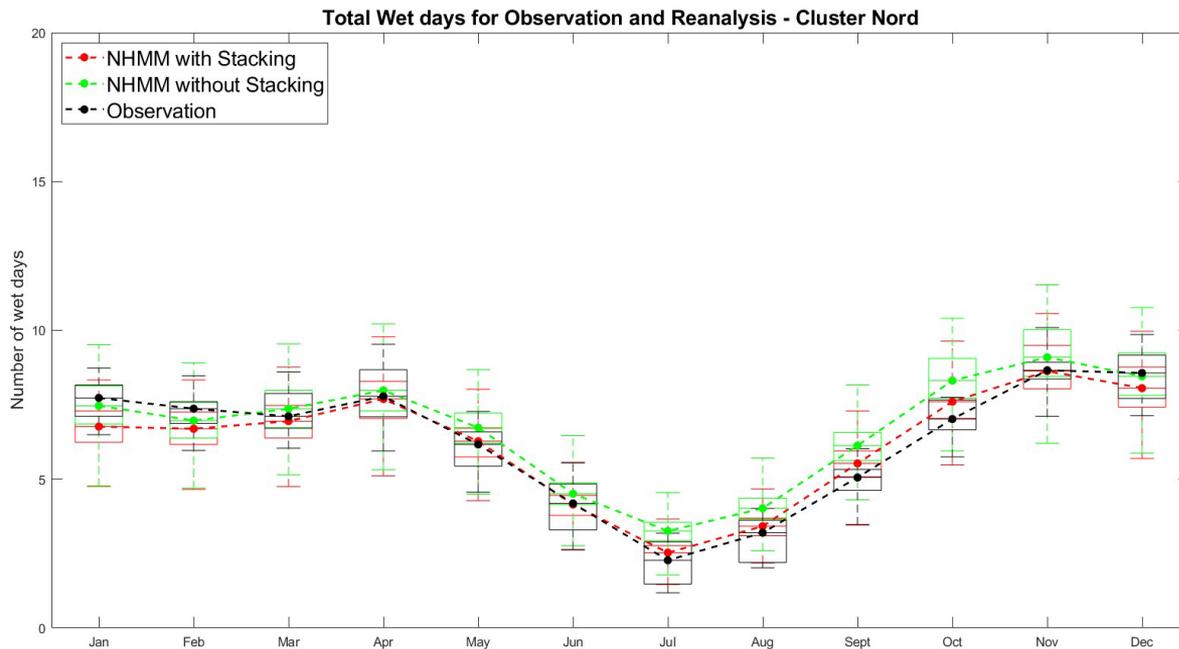
APPLICAZIONE E RISULTATI PER LA REGIONE LAZIO

Attraverso le simulazioni, è stato possibile mettere a confronto i risultati prodotti dai modelli di previsione, **sia con l'utilizzo dello stacking che senza**, rispetto ai dati effettivi di precipitazioni rilevati dalle stazioni pluviometriche

APPLICAZIONE E RISULTATI PER LA REGIONE LAZIO

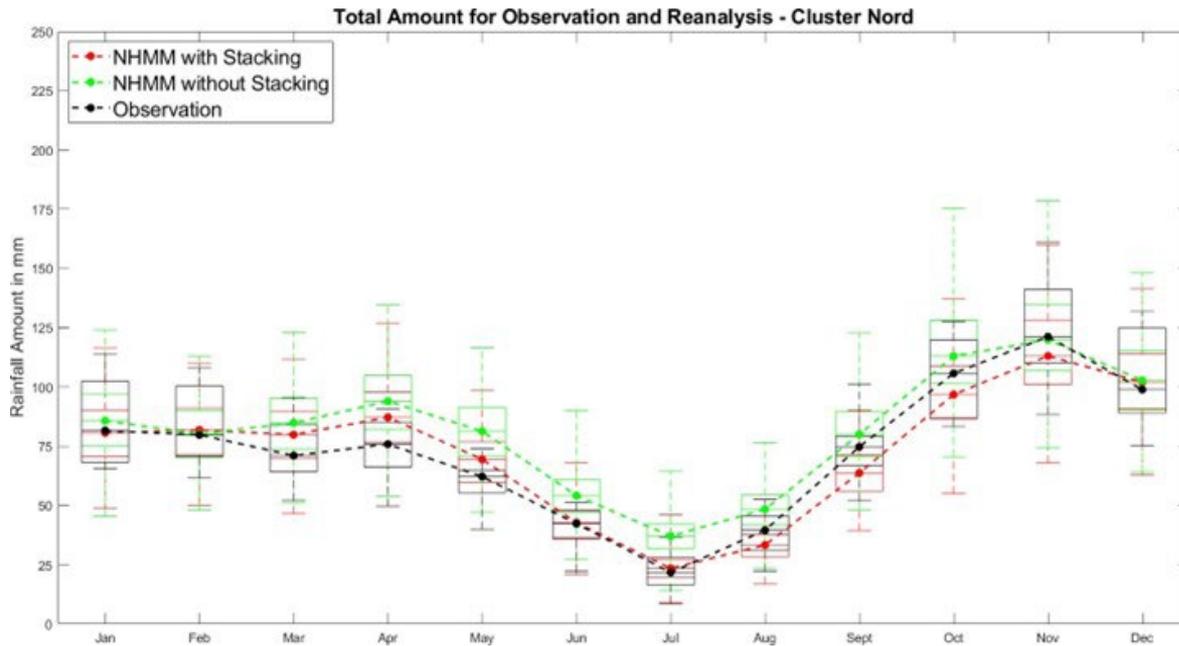
Per tutti i cluster, il modello con stacking migliora le prestazioni di previsione rispetto al modello senza stacking.

Sia nelle simulazioni del numero medio di giorni piovosi che nel caso delle altezze di pioggia, calcolate su scala mensile, i valori con stacking (in rosso) risultano più vicini ai valori osservati (in nero).



APPLICAZIONE E RISULTATI PER LA REGIONE LAZIO

Questa differenza è particolarmente evidente nella stagione estiva, dove il modello con stacking ha dimostrato una maggiore capacità nel prevedere l'andamento stagionale delle precipitazioni.



APPLICAZIONE E RISULTATI PER LA REGIONE LAZIO

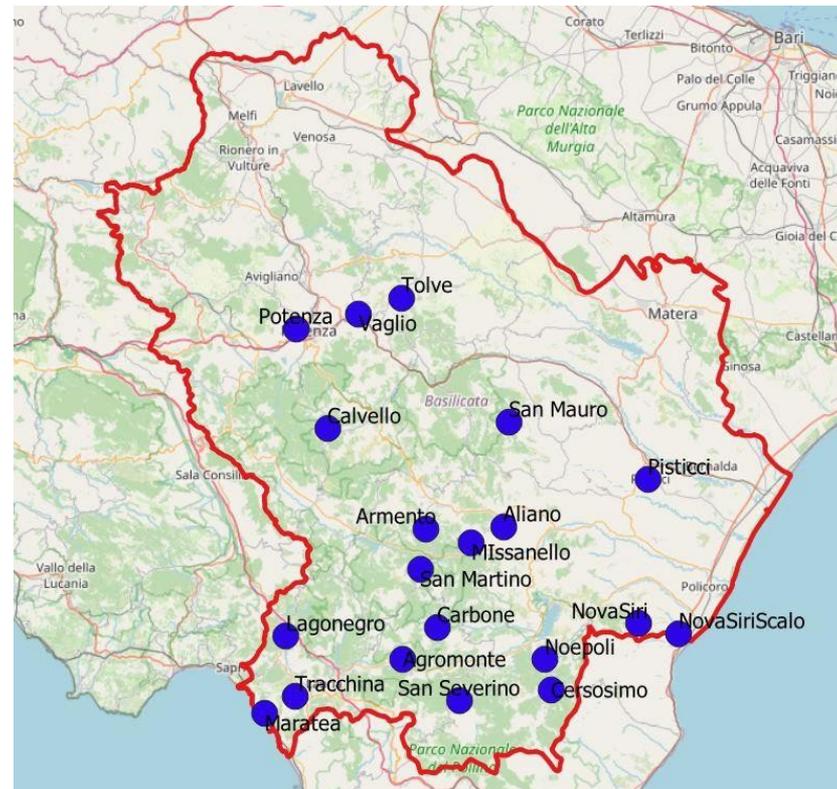
**La differenza tra i due modelli predittivi è stata quantificata
calcolando il MAE (errore assoluto medio)
rispetto ai valori osservati**

Il modello con lo stacking presenta un valore di **MAE**, in termini di millimetri di pioggia, significativamente inferiore (5,85 mm) rispetto al modello in cui non è stato utilizzato (9,10 mm)

APPLICAZIONE E RISULTATI PER LA REGIONE BASILICATA

Per l'addestramento e la validazione del modello, è stato impiegato un archivio di precipitazioni giornaliere relativo a un periodo di 50 anni (1951-2000), raccolto da 20 stazioni

Anche per la Regione Basilicata 5 stati nascosti rappresentano il bilanciamento ideale tra prestazioni di simulazione e tempo di elaborazione. Diversamente da quanto avvenuto per la Regione Lazio, le stazioni pluviometriche non sono state aggregate in cluster omogenei a causa del loro numero limitato.

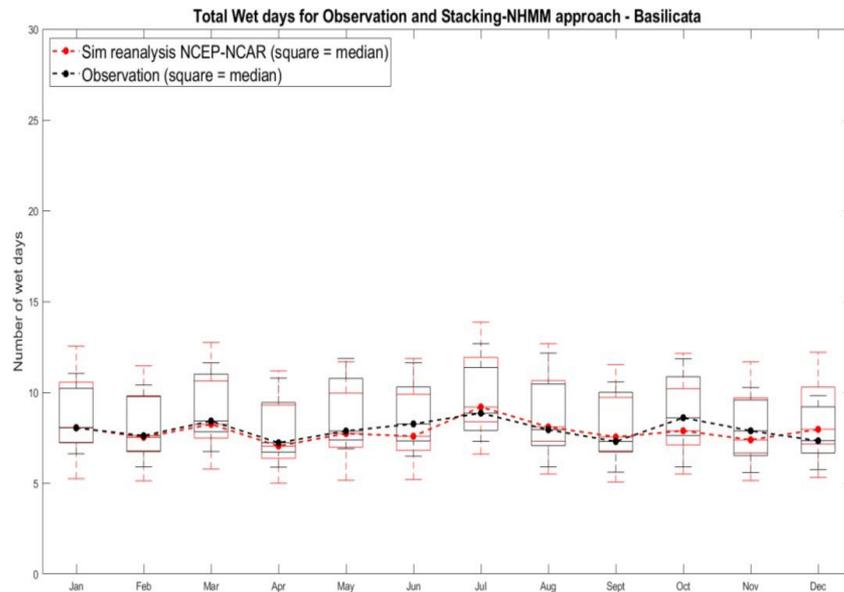
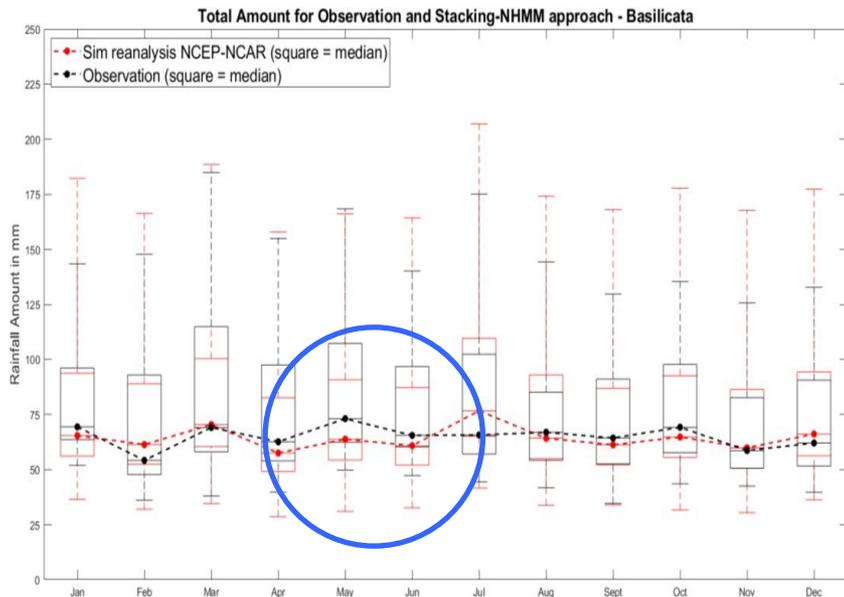


APPLICAZIONE E RISULTATI PER LA REGIONE BASILICATA

Attraverso le simulazioni sono stati confrontati i risultati forniti dal modello Stacking con le misurazioni effettive di precipitazioni rilevate dalle stazioni.

Il modello ha mostrato un'ottima aderenza sia nelle altezze delle piogge che nel conteggio dei giorni piovosi, presentando errori contenuti. La differenza tra i due modelli predittivi è stata quantificata attraverso il **MAE (3,5 mm)**.

APPLICAZIONE E RISULTATI PER LA REGIONE BASILICATA



Le discrepanze più significative tra simulazione e osservazione delle altezze di pioggia sono state rilevate nei mesi di **Maggio e Giugno**, probabilmente a causa della variabilità spaziale dei dati durante l'estate, accentuata dalle differenze climatiche tra la costa Jonica (Est) e l'area del Lagonegrese (Ovest) della regione.

CONCLUSIONI E PROSPETTIVE FUTURE

L'indagine condotta su Lazio e Basilicata ha evidenziato l'efficacia e la rilevanza delle tecniche di clustering e modellazione delle precipitazioni in ambito scientifico.

L'utilizzo della tecnica PCA per i predittori GPH e IVT ha mostrato una notevole competenza nella riduzione dimensionale garantendo la conservazione di una parte significativa della varianza intrinseca dei dati.

La combinazione di questa metodica con l'adozione del modello NHMM, impostato con cinque stati nascosti, ha offerto un equilibrio ideale tra l'accuratezza delle previsioni e l'efficienza computazionale, suggerendo una direzione promettente per ricerche future.

CONCLUSIONI E PROSPETTIVE FUTURE

La Regione Basilicata, pur mostrando risultati promettenti nella presente ricerca, offre ulteriori spunti di approfondimento e sviluppo metodologico, soprattutto in merito alla **clusterizzazione delle stazioni pluviometriche**.

Nonostante l'attuale studio non abbia implementato una categorizzazione in cluster omogenei a causa del numero limitato di stazioni, si ritiene che un futuro lavoro in questa direzione potrebbe portare a ulteriori ottimizzazioni. Un terzo predittore a grande scala, specificamente la temperatura dell'aria, è tra i fattori futuri da considerare. La correlazione tra le variazioni di temperatura e i pattern di precipitazione è infatti ben documentata nella letteratura scientifica, si potrebbe così rafforzare la robustezza del modello.

CONCLUSIONI E PROSPETTIVE FUTURE

Per assicurare una copertura più completa del territorio regionale, si intende fare riferimento al **dataset ISPRA – SCIA**.

Questo dataset, caratterizzato da una griglia regolare di 10 km x 10 km, fornirà una risoluzione spaziale dettagliata delle precipitazioni, potenziando così la rappresentatività del modello sull'intero territorio regionale.

L'utilizzo di tale dataset consentirà non solo di comprendere meglio la distribuzione delle precipitazioni a livello locale, ma anche di integrare la modellazione su aree precedentemente non coperte o meno rappresentate.



UNIONE EUROPEA
Fondo Europeo di Sviluppo Regionale



Mitigazione dei rischi naturali per la sicurezza e la mobilità nelle aree montane del Mezzogiorno: Il contributo dei Collaboratori di Ricerca

Estratto dagli Atti del Convegno: Il contributo dei Collaboratori di Ricerca del progetto MITIGO presso l'Università della Basilicata

© 2024 Università degli Studi della Basilicata

Editrice Universosud – Potenza

ISBN 9791281551084



9 791281 551084

Pubblicazione realizzata con il cofinanziamento dell'Unione Europea – FESR, PON Ricerca e Innovazione 2014-2020.

www.ponricerca.gov.it