

TECNICHE DI POST-ELABORAZIONE DI TEMPERATURA MINIMA A CONFRONTO PER UN'AREA ALPINA

E. Eccel⁽¹⁾, R. Barbiero⁽²⁾, D. Cesari⁽³⁾, L. Ghielmi⁽¹⁾, P. Granitto⁽¹⁾ e F. Grazzini⁽³⁾

(1) IASMA – Istituto Agrario di S. Michele all'Adige – Centro Sperimentale. (2) PAT – Provincia Autonoma di Trento – Meteotrentino. (3) ARPA Emilia – Romagna

Abstract

La previsione della temperatura al suolo è intimamente legata alla posizione topografica che i modelli meteorologici numerici attribuiscono a ciascun punto della propria griglia. L'applicazione di una correzione fissa, punto per punto, non risolve però il problema a scala temporale giornaliera, ed in particolare per quanto riguarda i valori minimi, perché il profilo termico notturno è fortemente dipendente dalle condizioni meteorologiche. Ai fini di una buona previsione quantitativa della temperatura minima a una risoluzione spaziale utile, si sono testati algoritmi di *downscaling* delle uscite grezze dei modelli che tengono conto delle variabili meteorologiche influenti. Si sono confrontati diversi approcci: un semplice modello multilineare, l'applicazione di reti neurali e l'uso di un modello multi-albero ("foresta stocastica"). Sono presentati e commentati i risultati del confronto della post-elaborazione per l'area trentina dei modelli operativi ECMWF e LAMI.

Introduzione

La temperatura minima in ambiente montano è fortemente determinata dall'orografia. Non solo sussistono forti differenze altitudinali tra area ed area, ma anche la morfologia influisce sulla distribuzione dell'aria raffreddata dall'irraggiamento notturno (Gassmann e Mazzeo, 2001). I modelli meteorologici rimangono il miglior strumento prognostico a disposizione del previsore, tuttavia esiste una forte discrepanza tra l'orografia del modello e quella reale, che può raggiungere puntualmente quasi i 1000 m con la risoluzione tipica dell'ECMWF (griglia di 70x40 km²). Da ciò emerge l'utilità di un modello di *downscaling*, che consenta di convertire l'output grezzo del modello in una previsione quantitativamente calibrata e riferita alle aree di interesse. Lo scopo del lavoro è la creazione di procedure di post-elaborazione statistica dei modelli meteorologici che tengano conto di tutte le variabili meteorologiche potenzialmente influenti sulla previsione delle minime.

Materiali e metodi

Sono stati presi in considerazione due modelli meteorologici: il modello di circolazione globale (GCM) del Centro Europeo ECMWF (risoluzione 0.5°), separatamente per la corsa delle ore 12 e 00 UTC, e l'implementazione italiana del LAM (modello ad area limitata) "Lokal Model" tedesco, il LAMI (risoluzione 0.0625°).

Gli output per la corsa delle 12 erano disponibili per un periodo compreso tra 01/2001 e 10/2005, mentre per la corsa delle 00 il periodo era compreso tra 01/2004 e 10/2005. La corsa delle 12 è stata impiegata anche nel solo periodo 2004-05 per ottenere un confronto con la corsa delle 00.

La post-elaborazione dell'output grezzo si è orientata su due modelli non lineari: le reti neurali (NN) e la "foresta stocastica" (*random forest*), RF (Breiman, 2001). Per confronto, si sono però valutati anche metodi più

semplici, come la correzione dell'errore medio o la regressione multilineare.

L'approccio di calcolo impiegato è il MOS (model output statistics), che impiega come predittori le variabili in output previste alle ore 06 UTC nei nodi di griglia circostanti l'area in oggetto (9 per l'ECMWF e 21 per il LAMI) e come predittando la temperatura media nell'area misurata alla stessa ora a 2 m. Si è scelta l'area della media Val d'Adige, mediando le temperature minime delle stazioni di S. Michele, Trento sud e Rovereto, di proprietà IASMA (circa 40 km di sviluppo lineare, quote da 170 a 210 m s.l.m.). L'area è di particolare interesse per la suscettibilità alle gelate primaverili ed è climaticamente omogenea. Per quanto riguarda il modello multilineare, si è impiegato un algoritmo di selezione dei predittori "a gradino" tramite il criterio di Akaike.

L'algoritmo a rete neurale (NN) impiega una configurazione del tipo "multilayer perceptron" (Rumelhart and Mc Clelland, 1986) ed è implementato dal software "Statistica Neural Networks".

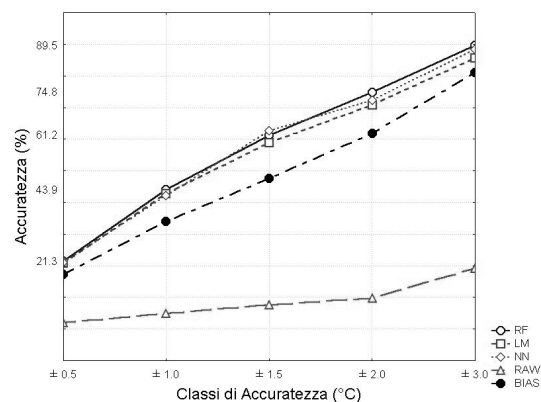


Fig. 1 – Accuratezza a confronto per i diversi modelli

L'algoritmo di "foresta stocastica" (RF) è un modello multi-albero (1000 in questa applicazione), dove ciascun albero è un regressore. La presenza di molti alberi rende la previsione assai più accurata che facendo uso di un solo albero "medio". Per entrambi i sistemi non lineari l'intero

insieme dei casi (giorni) è stato suddiviso casualmente in una parte per l'addestramento dell' algoritmo e una per la verifica. Si è usata l'implementazione per il software "R", disponibile liberamente su R-CRAN.

I risultati sono stati confrontati tra loro per mezzo dell'errore quadratico medio (RMSE) e dell'accuratezza, ossia la distribuzione degli errori residui in classi di ampiezza crescente.

MODELLO	Accuratezza %					Errore std (°C)
	± 0.5 °C	± 1.0 °C	± 1.5 °C	± 2.0 °C	± 3.0 °C	
ECMWF run12 2001-05	25	50	68	80	93	1.65
ECMWF run12 2004-05	27	51	71	80	93	1.63
ECMWF run00 2004-05	30	53	70	81	92	1.62
LAMI run00 2002-2005	24	47	65	77	91	1.76

Tab. 1 – Accuratezza ed errore standard a confronto nell'applicazione dell'algoritmo RF

Risultati

Si osserva una prestazione sensibilmente superiore per le tecniche NN e RF rispetto al modello lineare (LM), ma soprattutto rispetto all'output grezzo (RAW) e alla semplice correzione dell'errore medio (BIAS) – fig. 1, riferita al miglior punto di griglia utilizzabile. Per quanto riguarda i punti griglia, l'interpretazione dei risultati è difficoltosa. Nell'applicazione di RF punto per punto di griglia, si possono osservare, per il modello ECMWF, errori standard compresi tra 1.66 e 1.79 °C e indici di accuratezza sostanzialmente simili tra nodo e nodo. Dai risultati quindi non emerge una indicazione chiara sulla scelta di un punto griglia migliore di altri. Bisogna considerare poi che il numero di predittori potenziali è elevato ed è dato dal prodotto del numero di variabili per il numero di punti griglia. Si è quindi optato per RF come algoritmo da sviluppare e si è ritenuto necessario costruire una procedura oggettiva di selezione delle variabili e dei punti griglia. In quest'ottica si è deciso di utilizzare la misura dell'importanza di una variabile in RF e la sua capacità di funzionare senza incorrere in *overfitting* anche con un gran numero di variabili.

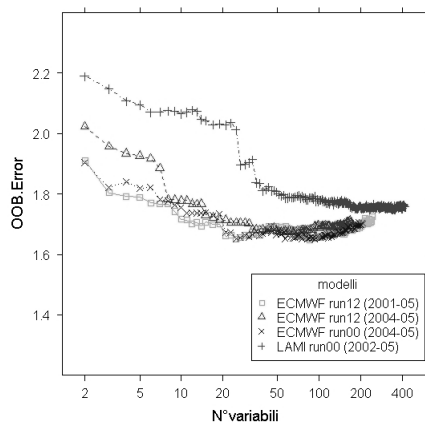


Fig. 2 – Errore del modello RF in funzione del numero di variabili considerate

Si è partiti con tutti i predittori nel MOS, si è calcolata la loro importanza e con una serie di cicli si è esclusa via via

la meno importante, fino all'ottenimento di un MOS limite con 2 sole variabili predittive (Diaz-Uriarte e Alvarez de Andrés, 2006 - fig. 2). Il numero ottimale di predittori varia da 64 a 82. I risultati per l'applicazione di RF agli output dei diversi modelli meteo sono riportati in tab. 1. Non sembrano esserci differenze particolarmente apprezzabili nell'utilizzare la corsa delle 00 piuttosto che quella delle 12 del giorno precedente. Considerando invece il LAMI, nonostante la griglia molto più fitta, e quindi la migliore aderenza con l'orografia reale, si sono riscontrati risultati lievemente peggiori che nel caso dell'ECMWF. Le variabili di maggior impatto nella post-elaborazione RF risultano le temperature a 2 m (soprattutto quelle dell'area di pianura a sud della Valdadige), la lunghezza della notte, l'umidità atmosferica a 850 hPa, la nuvolosità media.

Conclusioni

I risultati di questa indagine possono essere così riassunti.

- Una post-elaborazione delle previsioni grezze di temperatura minima condotte da modello meteo può migliorare l'affidabilità della previsione.
- La correzione è efficace anche apportando un semplice *offset* pari all'errore medio, punto per punto, ma migliora sensibilmente prendendo in considerazione altre variabili.
- All'interno dei tre modelli a molte variabili considerati, il migliore si è rivelato RF, ma con piccolo scarto con gli altri due.
- Non si notano miglioramenti apprezzabili utilizzando il modello LAMI, con risoluzione migliore rispetto ad ECMWF. Ciò pone qualche dubbio, dopo la recente introduzione di una griglia più fitta nel modello operativo ECMWF, sul possibile miglioramento della previsione di temperatura minima rispetto allo stato attuale.
- Anche in considerazione di ciò, si ritiene che la post-elaborazione degli output di modello abbia un limite difficile da superare in modo significativo; gli errori sono probabilmente da attribuire alla precisione stessa del modello (per esempio nella stima dell'entità nuvolosa o della sua evoluzione nel tempo) e portano ad effetti coerenti sulla previsione della temperatura minima; nell'errore non si apprezza quindi una natura sistematica, la quale, in quel caso, sarebbe correlabile ad altre variabili e quindi correggibile con post-elaborazione statistica.

Ringraziamenti

Il presente lavoro è stato eseguito nell'ambito del progetto GEPRI, finanziato dal Fondo Unico per la Ricerca della Provincia Autonoma di Trento.

Bibliografia

- Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1): 5-32.
- Diaz-Uriarte, R., Alvarez de Andrés, S., 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7:3.
- Gassmann, M.I. and Mazzeo, N.A., 2001. Nocturnal stable boundary layer height model and its application. *Atmospheric Research*, 57(4):247-259.
- Rumelhart, D.E., Mc Clelland, J., 1986. *Parallel Distributed Processing*. Vol 1. MIT Press, Cambridge.