



Controlli di qualità sulle serie termometriche

Dr. Gabriele Cola, Prof. Luigi Mariani

*Dipartimento Produzioni Vegetali, Sezione Agronomia, Facoltà di Agraria
Università degli Studi di Milano*



Premessa

I controlli di qualità sono stati eseguiti quale azione preliminare, sia alla scelta delle stazioni termometriche storiche, che alla successiva analisi climatologica dei dati, allo scopo di ottenere indicazioni circa la potenzialità del data set disponibile a rappresentare coerentemente nello spazio e nel tempo l'andamento della variabile meteorologica oggetto dello studio.

L'indagine, analizzando separatamente le temperature massime e le minime, ha preso in considerazione l'intero data set disponibile, considerando quindi i dati di 107 stazioni dell'ex Ufficio Idrografico operative in Veneto per il periodo 1955-2004, ed ha comportato l'applicazione di una tecnica di cross validation procedendo come di seguito descritto:

1. operando sui dati misurati, per ogni giorno della serie ciascun valore è stato ricostruito sulla base degli altri valori misurati disponibili, adottando un metodo di spazializzazione deterministico (sistema a media pesata con peso inversamente proporzionale al quadrato delle distanze su dati previamente resi omogenei rispetto alla quota, utilizzando il gradiente termico medio giornaliero frutto di regressione lineare sui dati) ed ottenendo così un dato di tipo misurato_ricostruito
2. si è creata una tabella in cui si sono stoccati tutti i dati di tipo misurato_ricostruito
3. la tabella dei dati di tipo misurato_ricostruito è stata confrontata con la tabella originale dei dati misurati e si è ricavato l'errore di ricostruzione applicando una serie di indici statistici (Rmse, R^2 , Crm, ecc).
4. Gli indici di errore sono stati spazializzati in modo da verificare se l'errore presenta strutture spaziali significative.

Un'operazione analoga a quella sopra descritta è stata compiuta anche sulle serie di osservazioni termometriche del periodo 1993 - 2008 (dati ARPAV) in modo da verificare se i livelli degli indici considerati siano comparabili rispetto a quelli delle serie dell'ex Ufficio Idrografico.

Gli indici di performance del modello di ricostruzione

La matrice fittizia dei valori simulati è stata confrontata con la serie dei dati misurati (dati originari), ricavando alcuni indici in grado di descrivere l'accordo tra i dati misurati e simulati (errore di ricostruzione).

In particolare sono stati applicati i seguenti indici di performance:

Coefficiente di determinazione (R^2)

$$R^2 = \frac{(\sum_i (S_i - \bar{O})(O_i - \bar{O}))^2}{\sum_i (S_i - \bar{O})^2 \sum_i (O_i - \bar{O})^2}$$

dove S_i rappresenta il dato simulato nel mese i , con $i=1, \dots, N$, O_i sono i dati osservati e \bar{O} rappresenta la media delle osservazioni. Il coefficiente di determinazione R^2 esprime la frazione della variabile misurata che è spiegata dal modello. Il suo valore si approssima ad uno per simulazioni accurate.

Errore Quadratico Medio (RMSE)

$$RMSE = \sqrt{\sum_i (S_i - O_i)^2 / N}$$

RMSE è una misura dell'ordine di grandezza dell'errore e cioè delle deviazioni dei valori simulati S_i rispetto agli osservati. Il valore ottimale per l'indice è zero.

Efficienza del modello secondo Nash-Sutcliffe (ME)

$$ME = 1 - \frac{\sum_i (O_i - S_i)^2}{\sum_i (O_i - \bar{O})^2}$$

ME è l'indice che quantifica il miglioramento del modello rispetto ad una situazione nominale di riferimento data dalla media delle osservazioni. Valori negativi indicano che la distanza tra i valori osservati e simulati $\sum_{i=1} (S_i - O_i)^2$ è maggiore della distanza tra i valori osservati e la loro media $\sum_{i=1} (O_i - \bar{O})^2$. In tal caso, la media delle osservazioni è uno stimatore migliore per la variabile studiata. Quando invece ME è positivo, le stime del modello sono migliori della media dei valori osservati. L'espressione tende ad 1 quando i valori simulati tendono verso quelli osservati.

Coefficiente di massa residua (CRM)

$$\text{CRM} = (\sum_i O_i - \sum_i S_i) / \sum_i O_i$$

Un valore maggiore di zero mostra che il modello globalmente sottostima la variabile ($O < S$). viceversa, un valore negativo di CRM indica che il modello tende a sovrastimare i valori della variabile ($O > S$) (Bechini L.,1999).

Gli indici rappresentativi dell'errore compiuto nella ricostruzione dei singoli dati giornalieri di temperatura massima e minima per i periodi indagati sono stati riportati sia in tabelle per stazione giorno (per l'individuazione di singoli dati aberranti) sia in tabelle riepilogative per stazione (per l'individuazione delle stazioni con andamenti significativamente non correlati con le stazioni di contorno).

RISULTATI

I risultati delle analisi di cross-validation sono riassunti in estrema sintesi in tabella 1. In particolare gli ottimi valori degli indici di performance per tutte le classi altitudinali 0-250 e 250-500 (ME superiore a 0.85, RMSE attorno a 2 e caratterizzata da una ridotta variabilità) indicano che gli algoritmi di ricostruzione potranno operare fornendo buoni risultati.

Tab. 1 - Sintesi dei risultati di cross-validation riferiti alle stazioni afferenti alle diverse classi altitudinali.

| classe altitudinale (m) | numero stazioni | RMSE media | RMSE dev. st. | R2 media | R2 dev. st. | ME media | ME dev. st. |
|-------------------------|-----------------|------------|---------------|----------|-------------|----------|-------------|
| 0-250 | 57 | 1.89 | 0.48 | 0.94 | 0.03 | 0.93 | 0.04 |
| 250-500 | 13 | 2.18 | 0.41 | 0.93 | 0.02 | 0.90 | 0.03 |
| 500-1000 | 16 | 2.42 | 0.94 | 0.93 | 0.03 | 0.87 | 0.09 |
| 1000-1500 | 13 | 2.40 | 0.86 | 0.93 | 0.03 | 0.85 | 0.14 |
| 1500-2000 | 8 | 2.50 | 0.56 | 0.92 | 0.04 | 0.86 | 0.06 |
| >2000 | 0 | - | - | - | - | - | - |

I risultati della cross validation ottenuti per le singole stazioni sono stati utilizzati per escludere dall'elaborazione quelle che presentavano i valori più scadenti per gli indici di performance. Si è in tal modo giunti al numero finale di 48 stazioni da utilizzare nel corso dello studio.

CONCLUSIONI

Riepilogando le suddette analisi sono state fondamentali per:

- evidenziare alcuni errori di attribuzione delle coordinate e delle quote delle stazioni storiche;
- guidare una serie di controlli mirati a singoli dati anomali che hanno consentito l'individuazione di errori di digitalizzazione, palesi errori di stampa o di misurazione

➤ guidare la scelta delle 48 stazioni storiche utilizzate per lo svolgimento dello studio

Dall'analisi effettuata sulle serie storiche dell'ex Ufficio Idrografico (1955-2004) è emerso, per quanto attiene alle temperature massime, i livelli di correlazione fra dati misurati e dati ricostruiti ottenuti in sede di cross validation sono mediamente più bassi rispetto a quelli forniti dalla rete ARPAV per il periodo 1993-2008. Tale fenomeno può da un lato indicare una minore accuratezza di alcune stazioni Idrografico ma potrebbe essere altresì stato prodotto da una maggior grado di "peculiarità" di alcune stazioni, il che va a costituire un elemento negativo ai fini di un'indagine mesoclimatica se frutto di effetti micrometeorologici oppure positivo se frutto invece di fenomeni attivi a scala maggiore.

Nel caso invece delle temperature minime, più soggette delle massime all'influenza dei processi a microscala legati ad esempio all'uso del suolo, la cross validation manifesta una significativa presenza di stazioni con bassi livelli di correlazione misurato-simulato in entrambe le serie considerate (Ufficio Idrografico e ARPAV). Tuttavia la distribuzione nello spazio delle stazioni a bassa correlazione è assai diversa nei due casi, manifestando una concentrazione nell'area alpina e prealpina nelle serie 1993-2008.

Bibliografia

Bechini L.,Ducco G., Donatelli M., Stein A., 2000. Modelling, interpolation and stochastic simulation of global solar radiation, *Agriculture, ecosystems and environment*, 81 (2000), 29-42.

Goovaerts P., 1997. *Geostatistics for Natural Resources Evaluation*, Oxford University Press, Applied Geostatistics Series, 483 p.

Lanzante J.R., 1996. Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data, *International Journal of Climatology*, Vol. 16, 1197-1226 (1996).

Mariani L., 2001. *Agrometeorologia.*, Clesav, Milano, 292 pp.