



## **Validazione dati**

*Prof. Antonio Berti, Prof. Maurizio Borin*

*Dipartimento di Agronomia Ambientale e Produzioni vegetali  
Università degli Studi di Padova*



L'analisi delle serie storiche deve essere fondata su metodi statistici rigorosi. Prima di fare questo, però, è indispensabile procedere alla **verifica di consistenza** (correttezza) dei dati disponibili, in mancanza della quale le serie dati non sono in grado di fornire informazioni utili per la ricostruzione del clima o addirittura generano valutazioni inesatte. Le serie di dati meteorologici, infatti, possono contenere disomogeneità ed errori dello stesso ordine di grandezza, o talora addirittura maggiori, dei segnali a lungo termine che le analisi si propongono di evidenziare. Inoltre esse possono essere incomplete (quando, ad esempio, mancano valori in modo casuale nel corso del tempo) o discontinue (quando per periodi prolungati il dato è mancante). I dati vanno sottoposti a controlli nel corso dei quali ad ogni dato controllato va associata una etichetta di validità del tipo: 0 = dato in attesa di validazione; 1 = dato validato; 2 = dato sospetto; 3 = dato errato. Ultimati i controlli di consistenza potrebbe risultare necessario procedere alla **ricostruzione dei dati errati** (ed eventualmente di quelli sospetti) in vista delle elaborazioni successive.

Per la verifica di consistenza è importante far ricorso a software di controllo dei dati meteorologici operanti a partire da dati di stazioni climaticamente omogenee ovvero da dati della stessa stazione che precedano o seguano quello errato. Nel presente lavoro si è fatto ricorso al software gratuito Climatica (Danuso, 2001) nato per la gestione dei dati meteorologici.

Il software dispone di funzionalità diverse tra cui la possibilità di controllo e di validazione dei dati, la conversione di formato delle date o delle unità di misura delle variabili, la produzione di indici statistici e climatici. Consente di effettuare operazioni generali di gestione sui database climatici (merge e append), sui metadati (inserimento e modifica), sulle variabili (cancellazione, generazione e ricalcolo) e sui record (ricerca, ordinamento, cancellazione, inserimento). Tali operazioni, preliminari e fondamentali per ogni elaborazione successiva, sono automatizzate per permettere all'utente di evidenziare i dati mancanti o inattendibili presenti nel dataset di lavoro. La tipologia di dato (misurato/ricostruito) ed il grado di attendibilità sono mantenute, per ciascuna variabile climatica, attraverso l'impiego di variabili di supporto.

La valutazione della qualità dei dati viene realizzata secondo criteri interni (range, coerenza interna, ecc.) o esterni (confronto con dati di località vicine) utilizzando una tecnologia a reti neurali. CLIMATICA permette di analizzare statisticamente i dati meteorologici con l'ausilio di procedure statistiche (media, mediana, deviazione standard, minimo e massimo, ecc.). Dispone inoltre di un ambiente di calcolo che dispone di numerose funzioni matematiche nonché di funzioni per la generazione di numeri casuali. Sono disponibili inoltre procedure per l'interpolazione e ricostruzione dei dati mancanti come la regressione lineare multipla, la regressione non-lineare, la creazione di reti neurali e le "moving statistics" (media mobile, ecc.). Le operazioni di gestione dei dataset possono essere effettuate sia attraverso finestre di dialogo che con l'uso diretto di comandi che permettono l'automazione delle procedure ripetitive (Danuso, 2005).

### **Controllo, validazione e ricostruzione dei dati**

Per ognuna delle stazioni venete i data set sono stati controllati e validati e poi ricostruiti i dati mancanti. I dati ricostruiti sono stati in seguito validati nuovamente con Climatica al fine di verificare scrupolosamente l'accuratezza della ricostruzione degli stessi.

### **Bibliografia**

Danuso F., Sandra M., 2005. CLIMATICA: un software per la gestione ed elaborazione delle informazioni climatiche. Rivista di Agrometeorologia, anno 10, n. 1, 24-25.