



Ricostruzione dati mancanti

Prof. Antonio Berti ⁽¹⁾, Dr. Gianmarco Tardivo ⁽²⁾

*⁽¹⁾ Dipartimento di Agronomia Ambientale e Produzioni vegetali
Università degli Studi di Padova*

*⁽²⁾ Centro Meteorologico di Teolo (PD)
ARPAV, Dipartimento Regionale Sicurezza del Territorio*



All'interno delle serie giornaliere pluriennali di dati di temperatura, usate per la costituzione di questo atlante, vi sono sparsi casualmente piccoli intervalli di dati mancanti; infatti, nel corso degli anni di attività di una rete di stazioni di misura, possono accadere dei malfunzionamenti sia agli strumenti (in questo caso, termometri) sia ai loro sistemi di trasmissione. In tali circostanze possono trascorrere ore o giorni prima che il personale tecnico completi un intervento di riparazione.

Il tempo in cui lo strumento non funziona si risolve in perdita di valori di temperatura registrati. La ricostruzione dei dati mancanti o errati diviene quindi un passaggio necessario per le analisi climatologiche basate su serie storiche.

Le tecniche più usate per tali scopi sono le *between-station*, *within-station* e *regression-based* (Allen et al. 2001) con una preferenza per le tecniche che adottano il *fitting* multilineare.

Anche le tecniche geostatistiche come il *kriging* possono essere impiegate per la ricostruzione dei dati mancanti, tuttavia Ian et al. (1998) evidenziano come l'approccio multilineare sia generalmente più soddisfacente sia per la qualità della ricostruzione che per la maggiore facilità di gestione del sistema di ricostruzione stesso. La ricostruzione di molti intervalli di giorni mancanti con metodi *kriging* risulterebbe infatti molto impegnativa dal punto di vista computazionale (WMO, 1983 e 2007). Tale metodo, infatti, richiede lo studio del semivariogramma per ogni singola ricostruzione, rendendo praticamente impossibile una completa automazione del sistema. La World Meteorological Organisation (WMO) ha analizzato diversi approcci di regressione multipla per la ricostruzione dei dati (WMO, 2007), considerando in particolare metodi semplici basati su regressioni lineari.

Nel caso della ricostruzione dei dati delle serie storiche del Veneto, si è scelto di sviluppare un approccio originale basato su regressioni multilineari. Il metodo ricostruisce ogni intervallo di dati mancanti di una stazione tenendo conto dei dati contemporanei presenti delle stazioni circostanti e della relazione storica esistente fra la stazione con l'intervallo mancante e queste altre stazioni. Ciò permette di tener conto sia della morfologia spaziale che temporale della variabile temperatura nel territorio studiato.

Rispetto alle più semplici tecniche basate sulla regressione tra la stazione con i dati mancanti e una o più stazioni considerate come riferimento, il metodo impiegato ha il vantaggio concettuale di ricercare il set di stazioni con il migliore adattamento in tutto l'ambito spaziale disponibile, permettendo di ridurre l'effetto di variazioni locali come, ad esempio, la presenza di zone abitate nelle vicinanze o fattori morfologici specifici (orientamento nelle zone montane) ed effettuando la ricerca sia tra i dati precedenti al periodo mancante che tra quelli posteriori. Pur presentando una maggiore complicazione dal punto di vista computazionale, quindi, il metodo impiegato permette una migliore ricostruzione dei dati.

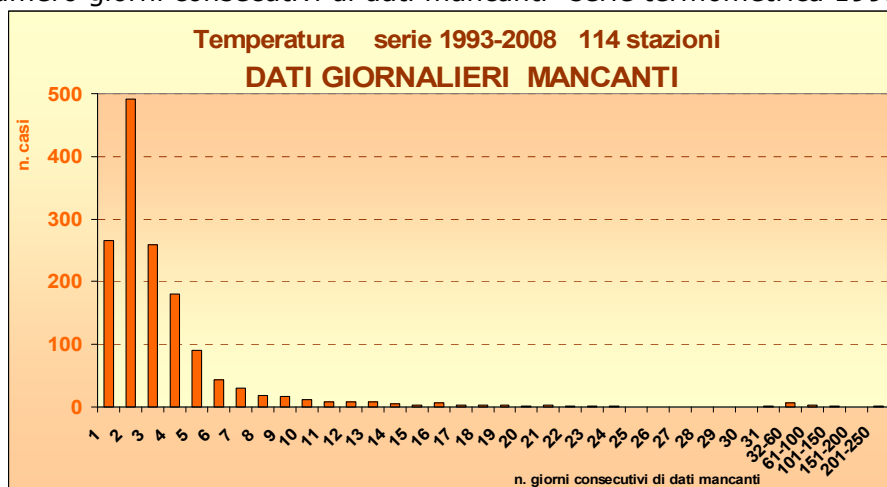
Si sono svolti diversi tipi di ricostruzioni facendo variare opportunamente i parametri che entrano in gioco in tale metodo, si è poi valutato il metodo eseguendo un rilevante numero di prove *cross-validation*. Nel complesso il 95° percentile della distribuzione dei valori medi degli scarti tra dati reali e ricostruiti è variato tra 0.056 °C e 0.285 °C. Le inversioni (cioè, per un giorno ricostruito, la presenza di temperature massime minori delle medie o di medie minori delle minime) sono apparse trascurabili, ammontando al 19 casi su 625308 giorni, con un errore massimo di 0,927 °C

I dati mancanti della serie termometrica 1993-2008

I dati della la serie termometrica 1993-2008, costituita da 114 stazioni, presentano nel loro complesso 6021 x 3 (min. med. max.) giorni di dati mancanti; le interruzioni di funzionamento nelle serie risultano essere 1480 X 3 (**Fig.1**). Il caso più frequente è quello di un'interruzione di 2 giorni consecutivi con 419 casi (33% delle fallanze). L'80% delle interruzioni rientrano nelle classi da 1 a 4 giorni consecutivi di dati

mancanti. Il 98% di tali interruzioni non supera i 20 giorni consecutivi e l'86% non supera i 5 giorni. L'interruzione massima trattata, occorsa alla stazione di S. Giovanni Illarione (VR), è di 213 giorni consecutivi.

Figura 1 – Numero giorni consecutivi di dati mancanti- serie termometrica 1993-2008



I dati mancanti della serie termometrica 1955-2004

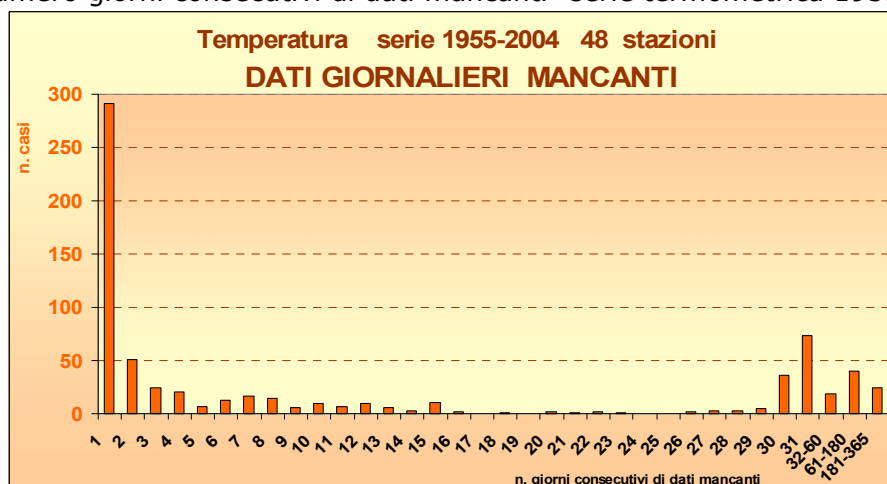
I dati delle 48 stazioni scelte per costituire la serie termometrica 1955-2004 presentano nel loro complesso circa 225.350 x 2 (min. e max.) giorni di dati mancanti; le interruzioni di funzionamento nelle serie risultano essere 777 X 2.

La categoria di giorni consecutivi di dati mancanti maggiormente rappresentate sono quelle di 1 giorno con 291 casi (37% delle interruzioni) e quelle di un mese con 118 casi (15% delle interruzioni) (**Fig.2**).

Le discontinuità più problematiche ed "impattanti" sullo studio climatologico sono però quelle relative a più anni continui di dati mancanti.

La casistica di tali discontinuità varia dai 19 casi di assenza di circa 2 anni continui di dati alla assenza, in due casi, di 30 anni continui di dati.

Figura 2 – Numero giorni consecutivi di dati mancanti- serie termometrica 1955-2004



Bibliografia

Allen R.J., De Gaetano A.T.,J. Bartolini G. 2001. Estimating missing daily temperature extremes using an optimized regression approach. International Journal of Climatology 21: 1305-1319. DOI: 10.1002/joc.679.
 WMO, 1983. Guide to climatological practices. 2nd edition.
 WMO, 2007. Guide to climatological practices. 3rd edition (draft).