

A CLUSTERING ANALYSIS OF MESONET STATION NETWORK FOR RAIN AND TEMPERATURE IN FRIULI VENEZIA GIULIA

ANALISI E CLUSTERIZZAZIONE DI UNA RETE DELLE STAZIONE METEOROLOGICHE IN FRIULI VENEZIA GIULIA

Andrea Cicogna^{1*}, Agostino Manzato¹

¹ Osservatorio Meteorologico Regionale (OSMER) ARPA-FVG, via Oberdan 18/a I-33040 Visco (UD)

*andrea.cicogna@meteo.fvg.it

Abstract

This work analyzes the rain and temperature data collected during 1995-2011 by 49 mesonet stations managed by OSMER - ARPA FVG to find objectively a minimum subset of stations (38 for rain and 32 for temperature) that have a significantly different attributes. That can be useful both for climatological purposes than for data quality check of the different station observations. As attributes to build the metric (distance among stations) from the meteorological variables have been considered the Mean, Standard Deviation, percentiles and the absolute value of the linear correlation coefficient ($|r|$). When more attributes are intersected together it is much more difficult to reduce the minimum number of stations needed. To discriminate stations in homogeneous groups two clustering techniques have been used. One is partitive (K-means), while the other is hierarchical (Ward-hierarchical). It has been found that there are more differences in the results varying the attributes used to define the metric than changing the clustering technique.

Keywords: station clustering, daily temperature, daily rain.

Parole chiave: clusterizzazione stazioni, temperatura giornaliera, pioggia giornaliera.

Introduzione

Il fine del presente lavoro è quello di cercare di individuare, nella esistente rete meteorologica gestita da ARPA-OSMER, un sottoinsieme minimo di stazioni, le cui misure differiscono tra loro in modo significativo e capaci di descrivere, per quanto possibile compiutamente, la climatologia della temperatura e delle piogge del Friuli Venezia Giulia. Tale suddivisione si basa sulla conoscenza quantitativa dei fenomeni meteorologici indagati, così le misure pluviometriche e termometriche diventano la base su cui costruire le metriche per il clustering, come indicato da vari autori, ad es. Leber *et al.* (1995), Bunkers *et al.* (1996), Fowell (1997), Mimmack *et al.* (2007) e Lund and Li (2009).

Materiali e Metodi

Sono stati analizzati i dati di 49 stazioni della rete gestita da ARPA-OSMER. In particolare sono stati esaminati i dati giornalieri di pioggia (quantità di pioggia giornaliera – RRtot, giorno piovoso definito come RRgg=1 se RRtot≥1 mm, altrimenti RRgg=0) (Fig. 2) e i dati giornalieri di temperatura (temperatura minima, media, massima ed escursione termica ovvero Tmin, Tmed, Tmax e Tesc) (Fig. 5) per tutti i dati disponibili tra il 1995 e il 2011.

Quali attributi per costruire le metriche (misure di distanza tra le stazioni) sono state utilizzate la media dei valori giornalieri, la dev.st. e, limitatamente a RRgg e RRtot, il 90° (Q90) e il 99° percentile (Q99). Gli attributi così definiti per ogni stazione sono stati standardizzati e quindi combinati tra loro per costruire le seguenti metriche:

- Media vs. deviazione standard per ognuna delle 6 variabili studiate. La metrica tra le stazioni diventa la distanza sul piano media - dev.st. standardizzate
- Media di RRgg (praticamente la frequenza della pioggia) vs. il 90° e il 99° percentile di RRtot. Queste metriche mettono in relazione il numero di giorni di pioggia con le piogge più intense.
- Valore assoluto dell'indice di correlazione ($|r|$) di ogni stazione verso le altre.

Mentre le metriche a e b lavorano su scale temporali lunghe (adatte a studi climatici), la metrica c è sensibile a scale temporali brevi e quindi più adatta a studi di quality check.

L'analisi dei dendrogrammi (Fig. 1) risultanti dalle classificazioni gerarchiche di Ward ha permesso di stabilire il numero K dei cluster da utilizzare in ogni classificazione. In funzione della variabile e della metrica esaminata si è identificato il K (fatto variare tra 4 e 10) che massimizzasse l'incremento relativo (δk) della distanza di fusione (d_k):

$$\delta k = (d_k - d_{k+1})/d_{k+1}$$

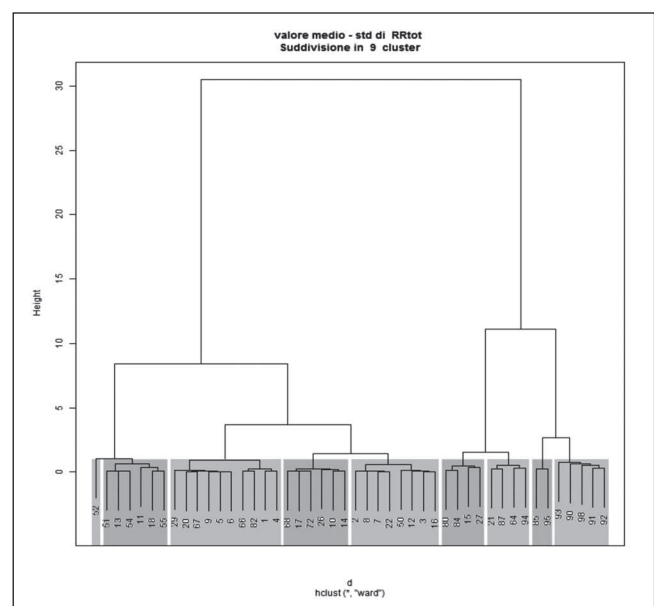


Fig. 1 - Dendrogram of RRtot to find 9 clusters (grey boxes) based on normalized mean and st.dev.

Fig. 1 - Dendrogramma di RRtot per individuare 9 gruppi (box grigi) in base a media e dev. st. normalizzate.

Tab. 1 - Number of clusters found optimizing the fusion distance of the Ward-hierarchical clustering.

Tab. 1 - N° di cluster identificati con l'analisi delle distanze di fusione dei dendrogrammi nelle classificazioni di Ward.

Variabile	Metriche			
	Media-dev.st	Media-Q90	Media-Q99	lrl
RRgg	9			7
RRtot	6	6	8	8
Tmin	8			6
Tmed	7			5
Tmax	6			5
Tesc	6			8

I risultati di tale analisi sono riportati in Tab. 1. Si può notare come il numero di cluster identificati varia da un minimo di 5 a un massimo di 9 in funzione delle variabili e delle metriche. Il numero dei cluster così determinato è poi stato utilizzato sia nella clusterizzazione partitiva che in quella gerarchica. Le clusterizzazioni sopra descritte sono state analizzate in 4 modi:

1. Analisi di tutte le singole clusterizzazioni effettuate con le diverse metriche e diversi metodi di clustering.
2. Per ognuna delle 6 variabili studiate sono state studiate le intersezioni delle clusterizzazioni ottenute con le diverse metriche solo con il metodo Ward-Hierarchical (il K-means dava risultati analoghi).
3. Analisi dell'intersezione delle clusterizzazioni effettuate su variabili afferenti alla stesso tipo di misura (RRgg, RRtt, Q90, Q99 da una parte e Tmin, Tmed, Tmax, Tesc dall'altra).
4. Analisi dell'intersezione di tutte le clusterizzazioni.

Infine per ridurre ulteriormente il numero dei raggruppamenti delle stazioni, si sono effettuate delle intersezioni a partire dalle classificazioni effettuate con un numero molto ridotto di cluster ($K \leq 4$) invece che il numero trovato massimizzando l'incremento della distanza di fusione.

Le classificazioni effettuate su ogni singola variabile, in funzione delle diverse metriche utilizzate, individuano raggruppamenti diversi. Per esempio per RRtot utilizzando come metriche media e Dev.st (Fig. 2) oppure lrl (Fig. 3) il numero dei gruppi e soprattutto le stazioni incluse nei singoli gruppi sono differenti. Utilizzando la prima metrica si può notare come si creano dei gruppi che ricordano la distribuzione climatica della pioggia media: mare, costa, bassa pianura e media pianura. Utilizzando invece come metrica di classificazione lrl si nota come i gruppi si differenziano in base alla vicinanza geografica. Non stupisce quindi che l'intersezione delle diverse classificazioni ottenute con diverse metriche dalle diverse variabili, individui poche stazioni che restano unite in uno stesso gruppo. Per quanto riguarda le variabili pluviometriche si individuano 38 gruppi (Fig. 4) partendo da un totale di 46 stazioni. Risultati analoghi si ottengono utilizzando i dati termometrici dove l'intersezione delle diverse classificazioni fatte per Tmin, Tmed, Tmax, Tesc individua 32 gruppi omogenei su 49 stazioni. Se invece si volessero effettuare forzatamente delle clusterizzazioni imponendo pochi gruppi ($K=2$, $K=3$ o $K=4$) ed effettuando quindi le intersezioni contemporaneamente su dati pluviometrici e termometrici si evidenzerebbe comunque la necessità di mantenere da un minimo di 26 (Fig. 5) ad un massimo di 36 stazioni rispetto alle attuali 49.

Conclusioni

Il numero e la localizzazione delle stazioni della rete di monitoraggio termopluviometrico gestito da Arpa-Osmer non sembra essere riducibile senza perdere informazioni meteo-

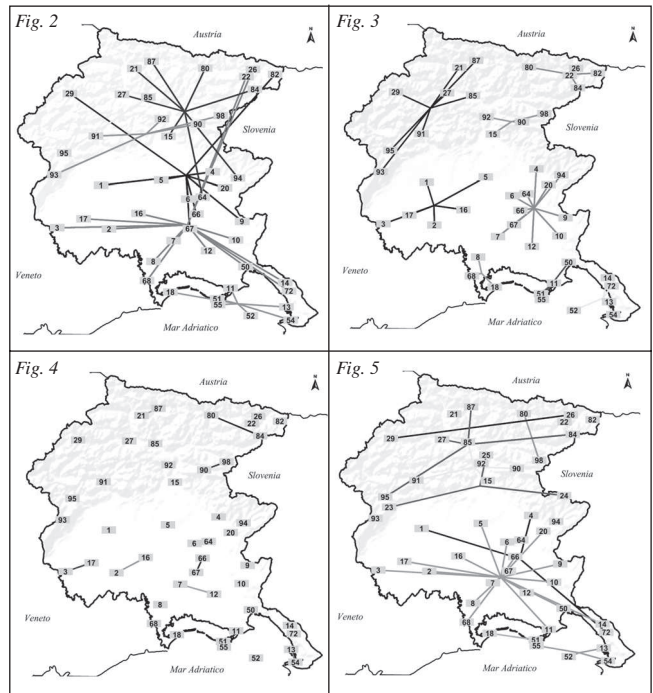


Fig. 2 - Map representation of Ward clustering for RRtot with mean and st.d.

Fig. 2 - Restituzione in mappa della clusterizzazione Ward per RRtot (con Media -Dev.st.).

Fig. 3 - Map representation of Ward clustering for RRtot with |r|.

Fig. 3 - Restituzione in mappa della clusterizzazione Ward per RRtot (con lrl).

Fig. 4 - Map representation of groups found intersecting the 6 Ward clustering shown in Tab.1 for RRgg and RRtot.

Fig. 4 - Restituzione in mappa dei gruppi individuati con la intersezione delle 6 clusterizzazioni Ward indicate in Tab. 1 per RRgg e RRtot.

Fig. 5 - Map representation of groups found intersecting all the clusters found imposing K=2 for all the variables and metrics.

Fig. 5 - Restituzione in mappa dei gruppi individuati con la intersezione di tutti i cluster ottenuti imponendo $K=2$ per tutte le variabili e metriche.

climatiche. In particolare il campo di pioggia, per sua natura molto più variabile nello spazio e nel tempo di quello termico, necessita di un'alta densità di punti di misura. Considerata l'estensione del Friuli Venezia Giulia (7845 km²) le stazioni pluviometriche passerebbero da una densità di 170 a 206 staz./km².

Bibliografia

- Bunkers M. J., Miller J. R. Jr., De Gaetano A. T., 1996. Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique. *Journal of Climate* 9, 130-146.
- Fovell R. G., 1997. Consensus Clustering of U.S. Temperature and Precipitation Data. *Journal of Climate* 10, 1405-1427.
- Leber D., Holawe F., Hausler H., 1995. Climatic classification of the Tibet Autonomous Region using multivariate statistical methods. *GeoJournal* 37, 451-472.
- Lund, R., Li B., 2009. Revisiting Climate Region Definitions via Clustering. *Journal of Climate* 22, 1787-1800.
- Mimmack G. M., Mason S. J., Galpin J. S., 2001. Choice of Distance Matrices in Cluster Analysis: Defining Regions. *Journal of Climate* 14, 2790-2797.