

DELIMITAZIONE DI AREE METEOROLOGICHE OMOGENEE CON IL METODO FUZZY C-MEANS E LORO VISUALIZZAZIONE MEDIANTE TECNICA RGB.

Guastaferro F.¹, Castrignanò A.¹, De Benedetto D.¹, Sollitto D.¹

¹ C.R.A.- S.C.A. (Unità di Ricerca per i Sistemi Colturali degli Ambienti caldo-aridi)– Bari francesca.guastaferro@entecra.it

Abstract

Lo scopo di questo studio è quello di caratterizzare, localizzare e visualizzare aree con condizioni climatiche molto simili fra loro, utilizzando tecniche di clustering multivariato. In particolare si è impiegato il metodo “Fuzzy c-means”, che produce una classificazione continua, assegnando ad ogni punto un vettore descrivente il grado di appartenenza a ciascuna classe. Per la visualizzazione si è impiegata una tripletta di colori RGB, che permette di combinare i fattori in studio in un’unica mappa.

Le variabili considerate per questo lavoro sono la pioggia cumulata e le medie delle temperature massime e minime su scala mensile, interpolate mediante la metodologia dell’inverso del quadrato della distanza. Per ridurre il numero, si è impiegata l’analisi delle componenti principali (PCA), seguita da una rotazione attraverso la trasformazione ortogonale “VARIMAX”. La fuzzy c-means è stata quindi applicata alle prime tre componenti principali e il risultato è stato quello di suddividere l’area in 3 regioni meteo omogenee. Utilizzando il sistema RGB, si è prodotta una mappa che, attraverso la tonalità di colore, ha permesso di valutare, anche visivamente, la combinazione delle condizioni climatiche all’interno di ciascuna regione meteo.

Introduzione

Le regioni omogenee o clusters sono aree caratterizzate da condizioni ambientali (climatiche, topografiche ed edafiche) simili fra loro: attraverso una procedura di “Cluster analysis” è possibile suddividere l’area di studio in porzioni di territorio, da sottoporre eventualmente ad una gestione differenziata. Questa metodologia raggruppa individui simili in classi distinte, chiamate appunto “cluster”, nello spazio ad N-dimensioni, dove ogni individuo è caratterizzato da N proprietà (Lark, 1998). Attualmente esistono numerosi algoritmi che applicano questa procedura; quello utilizzato in questo lavoro, è il metodo di classificazione “fuzzy c-means”, per identificare aree meteorologiche omogenee. Si è preferito questo approccio perché restituisce una classificazione spaziale continua, che meglio si adatta a rappresentare la variabilità dei fenomeni naturali. Anziché assegnare ad ogni pixel una ed una sola classe, viene associato un vettore, i cui elementi esprimono il grado di condivisione di ciascuna classe. Il metodo, inoltre, consente di valutare l’incertezza della classificazione, derivante dalla geometria della rete di monitoraggio e dalla qualità dell’informazione meteorologica. E’ possibile, inoltre, visualizzare l’eterogeneità intrinseca a queste aree “omogenee”, impiegando una tripletta di colori RGB (rosso – verde – blu) che con un’unica tonalità di colore, permette di rappresentare la combinazione di condizioni meteorologiche presenti (Castrignanò *et al.*, 2009).

Materiali e metodi

L’area oggetto di studio fa parte della Capitanata (provincia di Foggia) ed è monitorata principalmente da due reti meteorologiche: Servizio Idrografico (SI) e Consorzio di Bonifica (CB). Per questo studio sono state considerate 14 stazioni meteorologiche; alcune, ricadenti all’esterno dell’area di studio, sono state ugualmente prese in esame, perché ritenute sufficientemente limitrofe. Le variabili climatiche analizzate sono state: la pioggia cumulata mensile e le medie mensili delle temperature massime e minime. Queste variabili sono state interpolate ai nodi di una griglia quadrata di passo 100 m, utilizzando

la metodologia dell’inverso del quadrato della distanza. Lo strumento software utilizzato per condurre l’analisi spaziale è stato ArcGis 9.1. Per produrre una classificazione spaziale continua si è applicato un approccio di clustering del tipo “fuzzy c-means” (Fridgen *et al.*, 2004). L’algoritmo calcola, per ogni pixel, un vettore di valori di “membership” per ciascuna classe, che indica il grado di condivisione fra le classi. L’insieme dei valori ottenuti di memberships può essere sottoposto ad un processo di “defuzzifying”, assegnando a ciascun pixel la classe corrispondente al valore più alto della funzione di membership. L’approccio fuzzy si basa sulla minimizzazione ai minimi quadrati della distanza media del singolo punto dal centroide del cluster a cui appartiene. La distanza euclidea dovrebbe essere utilizzata solo per variabili indipendenti statisticamente e con uguale varianza; condizioni queste raramente soddisfatte nella pratica. Dato anche il numero elevato di variabili (38 in quanto alle 36 variabili meteorologiche sono state aggiunte anche le coordinate spaziali), il che rende praticamente impossibile l’inversione della matrice di varianza-covarianza, richiesta dall’algoritmo di clustering, oltre alle caratteristiche di intercorrelazione ed eteroschedasticità delle variabili, si è proceduto ad una riduzione del numero delle variabili e ad una loro ortostandardizzazione. A tale scopo è stata applicata l’analisi delle componenti principali (PCA) alla matrice di correlazione, seguita da una rotazione delle componenti attraverso la trasformazione ortogonale “VARIMAX”, utilizzando la procedura FACTOR del pacchetto di software statistico SAS / STAT, v 9.1. Solo le componenti con autovalori superiori ad 1 sono state utilizzate per la classificazione, effettuata con il software MZA, Management Zone Analyst. Il numero ottimale dei cluster è determinato sulla base del criterio di minimizzazione di due indici: l’indice di “fuzziness” (FPI) e quello di entropia standardizzata (NCE). FPI ($0 \leq FPI \leq 1$) è una misura del grado di condivisione fra le classi: un valore vicino a 1 indica un elevato grado di condivisione, mentre 0 corrisponde al caso di classi distinte, non

sovrappontesi. NCE ($0 \leq NCE \leq 1$) stima il grado di confusione nella classificazione: un valore vicino a 1 indica una forte disorganizzazione, mentre 0 riflette un'elevata organizzazione.

Risultati

Dall'analisi delle 36 mappe, 12 per ciascuna variabile (non mostrate), si rileva: un gradiente crescente in direzione NW - SE per le temperature massime; per le temperature minime, la zona NE è caratterizzata da valori generalmente più bassi; i totali mensili pluviometrici mostrano un'estrema variabilità, anche se la parte N è generalmente più piovosa di quella S. La classificazione fuzzy è stata applicata alle prime 3 componenti principali ruotate con autovalore maggiore di 1, che cumulamente spiegano circa l'85% della varianza totale. La rotazione VARIMAX ha permesso l'interpretazione delle 3 componenti principali; sulla prima pesano principalmente e positivamente le temperature massime, mentre sulla seconda le temperature minime e sulla terza le piogge del periodo verno-primaverile. Possiamo, quindi, con una perdita di informazione pari a circa il 15% della varianza totale, descrivere il campo meteorologico dell'area in esame mediante tre indici ortogonali: il primo, che potremmo definire delle temperature massime, il secondo delle temperature minime ed il terzo delle piogge. La classificazione è stata ripetuta facendo variare il numero di classi fra 2 e 6. In fig. 1 è riportato l'andamento dei due indici. Come si può notare, i due grafici non raggiungono un minimo assoluto, ma dopo un minimo locale, in corrispondenza dei cluster 3 e 4, tendono a diminuire monotonamente.

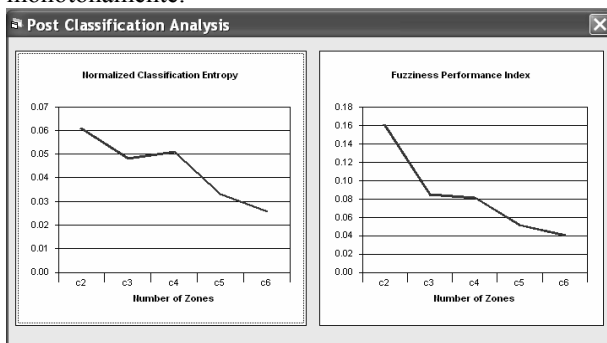


Fig. 1 Indici di NCE e FPI ottenuti con il software MZA

Optando per il criterio di maggiore semplicità, si è scelto il numero 3 di classi come quello ottimale (fig. 2).

Considerando le medie delle variabili relative ai centroidi dei cluster (non riportate), l'area corrispondente al cluster 2 è caratterizzata da una maggiore piovosità e da temperature più basse. Tali caratteristiche si attenuano negli altri due cluster, ove il clima risulta più temperato; in particolare, il cluster 1 più meridionale risulta essere quello meno piovoso e più caldo. Volendo rappresentare in un'unica mappa l'informazione relativa al livello di condivisione fra le diverse classi, alla funzione di appartenenza relativa a ciascuna classe è stata assegnata la scala di tonalità di uno stesso colore: rosso per la prima classe, verde per la seconda e blu per la terza. Combinando le tre mappe, mediante il sistema RGB, si è ottenuta un'unica mappa (fig. 3).

Le aree con tonalità del rosso sono indicative della prevalenza dei caratteri specifici per la classe 1 (aridità), del verde per la classe 2 (piovosità) e del blu per la classe 3, (caratteri di transizione). Le aree più biancastre, in cui i tre colori si mescolano in maniera approssimativamente omogenea, indicherebbero quelle più incerte o di transizione, a cui non è possibile assegnare univocamente i caratteri distintivi di una data classe. Nel nostro caso una maggiore sovrapposizione dei tre colori e, quindi, una maggiore nebulosità (fuzziness) si può notare al confine tra due cluster contigui e nell'angolo SW.

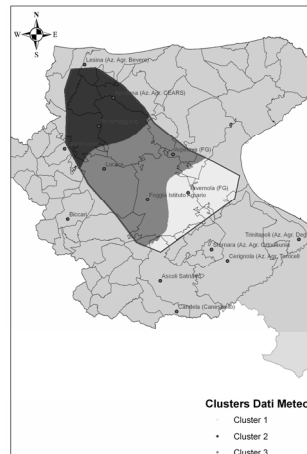


Fig. 2 – Classificazione meteorologica dell'area

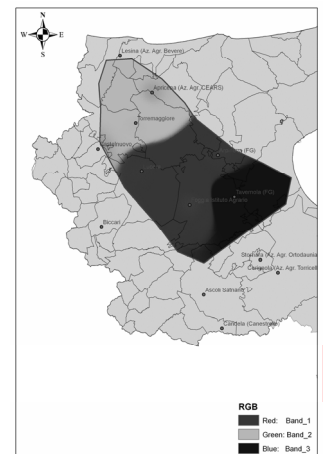


Fig. 3 Combinazione delle tre mappe utilizzando il sistema RGB

Conclusioni

Mediante la tecnica fuzzy è stato possibile ottenere una classificazione spaziale continua, che permette anche di valutare l'incertezza insita nella classificazione, a causa, fra l'altro, della scarsità e qualità dei dati. L'innovazione della rappresentazione RGB consiste nel visualizzare in un'unica mappa le informazioni relative al grado di eterogeneità climatica esistente all'interno dei cluster. La metodologia applicata presenta un notevole grado di flessibilità e può essere applicata a qualsiasi tipo di dati ambientali.

Ringraziamenti

Questo lavoro è stato realizzato grazie al progetto "Supporti decisionali per la conservazione e la gestione delle risorse idriche in aree vulnerabili del Sud Italia" – AQUATER" (Responsabile Dr. Michele Rinaldi – Responsabile U.O. Dr. Annamaria Castrignanò).

Bibliografia

- Castrignanò A., Guastaferro F., De Benedetto D., Moneta A., Basso B., Troccoli A., Pisante M., 2009. Accettato per: Joint International Agricultural Conference, Wageningen, July 6/7/8, 2009.
- Fridgen J.J., Kitchen N.R., Sudduth K.A., Drummond S.T., Wiebold W.J., Fraisse C.W., 2004. Mangement Zone Analyst (MZA): Software for Subfield Management Zone Delineation. *Agron. Journal* 96: 100-108.
- Hargrove W.W., Hoffman F.M., 2005. Potential of multivariate quantitative method for delineation and visualization of ecoregions. *Environmental Management* vol. 34, suppl. 1: pp. S39-S60.
- Lark R.M., 1998. Forming spatially coherent regions by classification of multivariate data: An example from the analysis of maps of crop yield. *Int. J. Geogr. Inf. Sci.* 12:83-98.