

LE CLASSIFICAZIONI NUMERICHE PER LA CORRELAZIONE DEI SUOLI

Roberto Barbetti, socio A.I.P.

C.R.A.- A.B.P., Centro di ricerca per l'Agrobiologia e la Pedologia, Piazza M. d'Azeglio, 30, 50121, Firenze.

Email: barbetti@soilmaps.it

Introduzione

In Italia già dal lontano 1972 Sanesi, Mancini ed altri padri della pedologia riunitisi in una tavola rotonda discutevano sull'impiego delle classificazioni numeriche automatizzate basate sulla statistica multivariata per la correlazione dei suoli.

In quegli anni l'inadeguatezza degli elaboratori, il carattere prevalentemente locale dei rilevamenti ed infine con lo sviluppo di lì a poco delle moderne classificazioni tassonomiche (Soil Taxonomy e FAO) portarono questi sistemi di calcolo ad essere accantonati dalla comunità pedologica.

Ai giorni nostri con la raccolta dei dati storici da parte degli enti competenti e con lo sviluppo dei sistemi informativi pedologici nazionali di tutto il mondo la questione della classificazione numerica applicata ai suoli è tornata di attualità.

Metodi

L'approccio classico della attività di correlazione per la creazione delle tipologie di suolo (STS) avviene solitamente tramite interrogazione dei database con query multiparametriche o più semplicemente mediante filtri applicati in sequenza fino ad isolare record con i caratteri di selezione comuni e facendo prevalere le classificazioni tassonomiche come discriminante per la creazione dei raggruppamenti.

Avvalendosi solo ed esclusivamente di questa tecnica però, l'attività del correlatore può diventare difficoltosa qualora si analizzino contemporaneamente più variabili su molti record.

Ad esempio chi ha esperienza di cartografia pedologica da 1:250.000 fino scale minori 1:1.000.000 sa che non sempre è facile far corrispondere una omogenea classificazione tassonomica, oltre il primo livello, degli individui di suolo appartenenti ad una certa STS.

Questo aspetto può giustificare l'integrazione di strumenti di analisi aggiuntivi per gestire dataset medio-grandi, anche perché la correlazione è un'attività di tipo ciclico, che può essere rivista con l'implemento di nuovi rilevamenti, permettendo di ridurre il grado di incertezza associato alla tipologia stessa.

il CRA-ABP Centro per l'Agrobiologia e la Pedologia di Firenze si è impegnato negli ultimi anni e soprattutto con il progetto "Badasuoli" banca dati pedologica d'Italia 1:1.000.000 a sviluppare un gestionale per database "ISIS" (Italian Soil Information System) che pur basandosi interamente su query SQL è minuziosamente integrato da funzioni e routine in linguaggio VBA (Visual Basic for

Application) per rendere più veloce e più facile l'attività di correlazione.

Citando solo alcune di queste facilitazioni diciamo che si può usufruire di query e maschere precostituite dove si possono effettuare multiselezioni di profili per eseguire statistiche descrittive "on the fly"; e la possibilità di connettersi al GIS per eseguire query di tipo spaziale.

Solo Dal 2007 con l'uscita di Osaca (Outil Statistique d'Aide à la Cartogénèse Automatique) la comunità scientifica può disporre di un software per la statistica multivariata cluster analysis interamente dedicato ai suoli.

Osaca è una Web Application Java Client-Server scaricabile o immediatamente eseguibile dall'European Soil Portal (<http://eusoils.jrc.ec.europa.eu/>) dove è possibile testare il dataset di esempio allegato.

Il software in questione è ben documentato ma all'atto pratico risulta abbastanza ostico e con alcune limitazioni. I dati di input devono essere necessariamente file .csv con rigide regole di formattazione, specificando il tipo di variabile numerica nelle intestazioni dei campi; accetta un limitato numero di variabili (12) e soprattutto non permette la presenza di valori nulli all'interno della tabella.

La tecnica di cluster analysis applicata, è di tipo K-means e propone vari opzioni di calcolo per la distanza fra i profili a seconda che si voglia enfatizzare maggiormente la differente successione degli orizzonti oppure lo spessore di questi.

L'operazione successiva crea la matrice di dissimilarità, dove i valori delle distanze possono essere tipo euclideo o di tipo Manhattan dopodiché segue la procedura di agglomerazione.

Concludendo questa breve recensione su Osaca possiamo dire che può offrire un valido aiuto per la correlazione di suoli a partire da zero; ovvero quando per un dataset non sono note o non ben definite le tipologie pedologiche.

Quando invece si tratta di sottoporre ad analisi molti subset ecco che il procedimento può divenire troppo lento.

Con la recentissima uscita dell'applicazione AQP (Algorithm for Quantitative Pedology) creato dai ricercatori della UC Davis Soil Resource Laboratory della California, la classificazione numerica applicata ai suoli sembra aver fatto un notevole passo in avanti.

Trattasi di una libreria del software statistico open source R che offre enormi potenzialità garantite dalla dinamicità del mezzo con la moltitudine dei pacchetti aggiuntivi disponibili e con una crescente comunità di sviluppo e utilizzo.

Questa libreria contiene attualmente delle funzioni che vanno dalla rappresentazione grafica verticale ed orizzontale dei suoli, alla gestione dei pattern di diffrazione sperimentali XRD (X-ray diffraction) ed appunto alla comparazione dei profili.

La base teorica che accomuna Osaca ed AQP è rappresentata dal fatto che ogni singolo orizzonte di

suolo è considerato come un vettore costituito dai valori dei singoli parametri dove è possibile applicare l'aritmetica vettoriale. Un profilo è quindi rappresentato da una lista di uno o più vettori. AQP non presenta limiti sul numero di parametri da utilizzare e non si limita a leggere i file .csv, interfacciandosi efficacemente a tutte le piattaforme di database conosciute.

Ad esempio per importare dati da MSAccess basta caricare la libreria RODBC per connettersi alle tabelle, ed eseguire SQL Query direttamente in ambiente R.

I dati vengono importati come dataframe che sostanzialmente rappresenta il formato tabellare letto da R.

Anche per AQP c'è la necessità di gestire i formati dei campi (casting esplicito) e i nomi delle intestazioni, ma il tutto può essere eseguito in maniera efficace e speditiva se una volta per tutte si raccolgono in sequenza tutti comandi per le operazioni ripetitive, creando un file .txt e richiamabile successivamente in modalità script.

La trasformazione dei dataframe ad oggetti SoilProfileList (funzione initProfileList) permette al software di saper leggere i singoli profili con i colori munsell decodificati, gli orizzonti genetici e le profondità associate che possono essere plottate in grafici. Con il comando `compare_profile` si creano le matrici di dissimilarità o di distanza tra gli individui suolo che rappresentano la base di partenza per la creazione dei dendrogrammi di cluster analysis. La possibilità poi, di associare i profili plottati ai dendrogrammi rappresenta uno standard di aiuto ai correlatori mai raggiunto finora.

Il procedimento per la creazione delle matrici di dissimilarità è diversa rispetto ad Osaca.

in AQP si prescinde dalla suddivisione in orizzonti genetici, confrontando i suoli per singole "fette" o slices definibili dall'utente e che per default sono di spessore di 1 cm. Le dissimilarità tra tutti gli individui vengono create in forma matriciale come sommatoria delle differenze delle singole slices dove viene applicato un coefficiente di decadimento (K).

Questo fattore determina come diminuisce il peso delle slice nel calcolo delle distanze con l'aumentare della profondità. K andrebbe tarato sul dataset su cui si intende operare, ma fortunatamente i ricercatori della UC Davis hanno testato differenti valori che possono essere adatti per le nostre esigenze. Ad esempio un valore di $K=1$ fa sì che le slices abbiano tutte lo stesso peso fino alla massima profondità di indagine, mentre un valore di 0,1 rimuove ogni influenza al di sotto dei 30 cm; il che può essere utile se ad esempio si vogliono confrontare solamente gli orizzonti lavorati.

Anche ISIS propone il concetto di distanza tra gli individui con l'applicazione di due indici di somiglianza paesaggistico e analitico. Questi parametri sono stati utilizzati nel progetto Badasuoli e sono stati molto utili per l'individuazione dei profili caposaldo e contemporaneamente degli outliers. I suoli caposaldo scelti sono quelli che hanno un indice di diversità minore rispetto al profilo

modale.

Il profilo modale è l'oggetto che descrive le caratteristiche della STS: è diviso in orizzonti funzionali, ognuno con le proprie statistiche descrittive: moda, media, deviazione standard e numerosità campionaria derivate dai profili.

Sempre più nella comunità scientifica si sente parlare di profilo medio o modale costruito su base statistica su profondità stabilite, sugli orizzonti genetici o sugli orizzonti funzionali dei profili.

Quando si creano carte derivate come la capacità d'uso è preferibile avere dati statistici di questo tipo rispetto a classi che rappresentano un certo range di variabilità per una certa caratteristica ma che necessitano di opportune maschere o report per essere decodificati.

Applicazioni

Attualmente il CRA-ABP sta utilizzando alcune funzioni di AQP per due finalità diverse.

La prima riguarda la revisione della banca dati Badasuoli: dato un certo profilo modale fornito dagli enti regionali ma mancante di profilo caposaldo, si associa "il più vicino" su quelli disponibili presso la banca dati del centro. Fig 1-2.

Il secondo per il progetto "Agrosceinari" in cui c'è la necessità di individuare una climosequenza; Ovvero monitorare suoli con le medesime condizioni evolutive di partenza, ma ubicati secondo un diverso gradiente climatico. fig.3

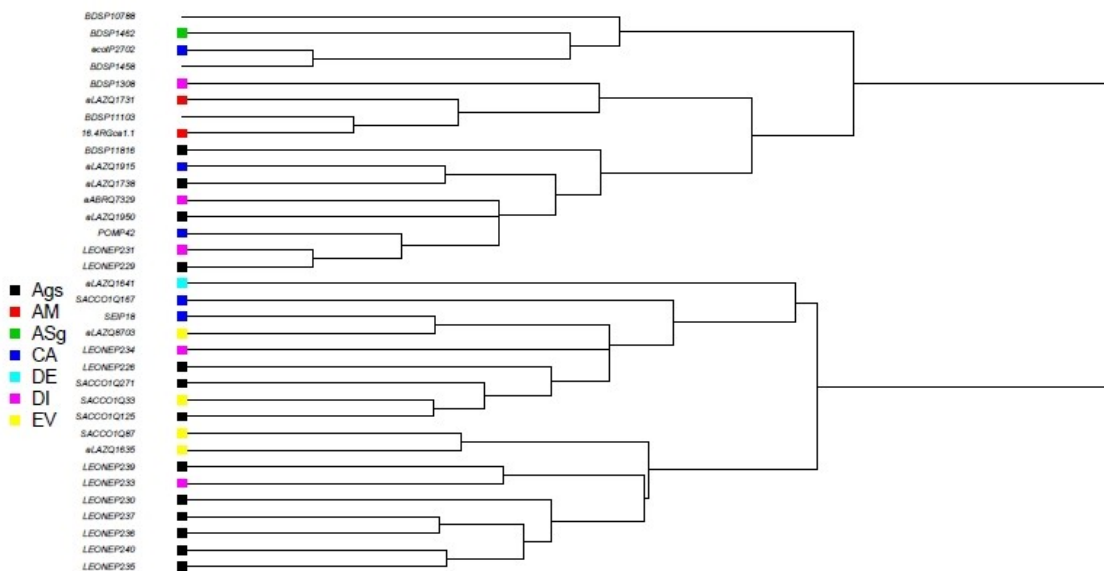


Fig. 2 dendrogramma di cluster analysis discriminato per litologia

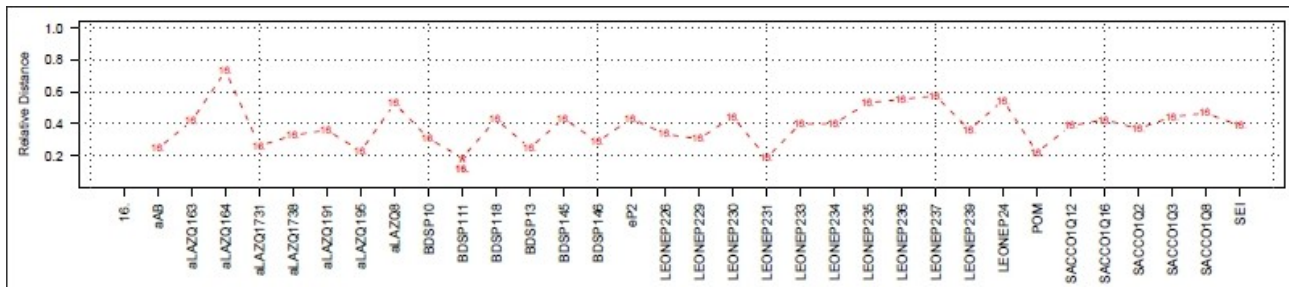


Fig. 3 grafico delle distanze relative tra il profilo modale e i singoli profili.

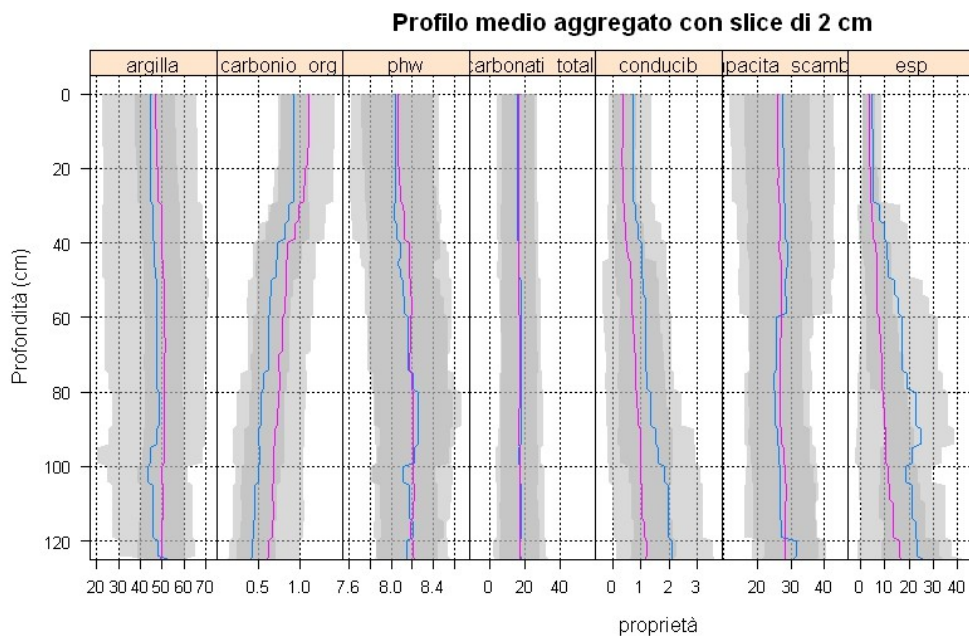


Fig. 3

La figura 3 riporta l'andamento medio della medesima tipologia di suolo 62.1VRcc1: Vertisuoli calcici della SoilRegion 62.2 su argille marine esaminata su due zone climatiche con indice di aridità diverso secondo la classificazione di Martonne. Pur mostrando un andamento molto simile per tutte le proprietà, sarà interessante studiare il maggiore contenuto di carbonio organico dell'una rispetto all'altra e una minore conducibilità elettrica ed ESP. (la linea blu rappresenta l'andamento dei suoli più aridi; la zona grigia rappresenta la variabilità). I campionamenti che verranno fatti in campo e le analisi di laboratorio diranno il resto.

Bibliografia

Carré F. and Jacobson M. Numerical classification of soil profile data using distance metrics
[Geoderma Volume 148, Issues 3-4](#), 15 January 2009, Pages 336-345

L'Abate, G. ,Barbetti R. , Costantini E.A.C. 2007 “CNCP 3.0, Dicembre 2007,
Applicazione per l'informatizzazione, la correlazione e la spazializzazione delle
osservazioni e delle unità pedologiche”. CD-Rom allegato al volume: E.A.C. Costantini (Ed.).
2007. “Linee guida dei metodi di rilevamento e informatizzazione dei dati pedologici”.
S.EL.CA. Srl, FIRENZE.

De Martonne, E. (1941) : Nouvelle carte mondiale de l'indice s'aridité. Météorol. 1941, 3-26

Documento web: <http://cran.r-project.org/web/packages/aqp/vignettes/aqp-vignette.pdf>

Documento web: <http://cran.r-project.org/web/packages/aqp/aqp.pdf>

Documento web <http://casoilresource.lawr.ucdavis.edu/drupal/node/916> New R Package 'aqp':
Algorithms for Quantitative Pedology.

Sanesi G. Wolf U. Applicazione di alcuni metodi di analisi multivariata alla cartografia del suolo
(1972) Società Italiana Scienza del Suolo Atti della tavola rotonda sul tema:

La cartografia dei suoli scopi-metodi- applicazioni pp.84

Sito web: <http://eussoils.jrc.ec.europa.eu/projects/OSACA/>