

ANALISI STATISTICHE E GEOSTATISTICHE CON PYARCHINIT: PRIMA SPERIMENTAZIONE

1. INTRODUZIONE

Le analisi di distribuzione spaziale su record archeologici, integrati con tecniche geostatistiche, hanno un ruolo fondamentale nell'archeologia preistorica soprattutto per approfondire le conoscenze dei livelli archeologici di frequentazione umana. In questo lavoro si presenta un approccio metodologico che integra tecniche statistiche multivariate, database e GIS, tutto in un'unica soluzione software. Dal momento che l'archeologia, dal punto di vista delle geoscienze, è una disciplina intrinsecamente spaziale, abbiamo utilizzato un approccio quantitativo spaziale come la geostatistica multivariata, che in genere è applicato al settore minerario e ambientale, per lo studio di un importante sito archeologico dell'Italia settentrionale (Grotta di Fumane, Verona), gestendo le informazioni quantitative sul campo come variabili regionalizzate. Così facendo abbiamo valutato interessanti correlazioni spaziali, imponendo modelli di co-regionalizzazioni lineari su semivariogrammi sperimentali e cross-semivariogrammi. Inoltre, per l'analisi della distribuzione spaziale dei dati, abbiamo proseguito con la creazione di uno strumento che integra diverse librerie di R, come gstat, implementate in un DBMS PostgreSQL con estensione PostGIS.

Le analisi eseguite sulla fauna di Grotta di Fumane mostrano interessanti modi per gestire spazialmente le categorie studiate, aprendo nuove prospettive per studi quantitativi archeologici. Inoltre, è importante sottolineare che le analisi statistiche e geostatistiche effettuate sono da considerare sperimentazioni dal punto di vista sia informatico, sia metodologico/archeologico, in quanto i risultati ottenuti servono ad una comprensione maggiore dei dati e del sito che si analizza e anche come corredo della documentazione di base. Queste sperimentazioni sono in continua evoluzione al fine di ricercare la metodologia migliore da adottare.

2. IL PLUGIN PYARCHINIT

PyArchInit è un plugin scritto per QGIS e realizzato in Python. Nasce come progetto di Tesi di Laurea di Luca Mandolesi nel 2005 e, a partire dal 2010, è stato oggetto della Tesi di Dottorato di ricerca di Enzo Cocca, con l'implementazione di nuove funzionalità e miglioramenti del software. In particolare, sono state apportate modifiche alla struttura del database a livello di relazioni ed entità, e di conseguenza sono state inserite funzioni trigger per la scheda archeozoologica (MANDOLESI, COCCA 2013). I database utilizzati da PyArchInit sono PostgreSQL

con estensione spaziale PostGIS e il database nativo di QGIS Spatialite. L'estensione spaziale PostGIS può essere installata sia in versione 1.x che 2.x.

3. LA SCHEDA ARCHEOZOLOGICA

La scheda archeozoologica è stata creata appositamente (COCCA 2015) e adattata al contesto del sito Grotta di Fumane per la gestione della fauna terrestre e l'avifauna: solo in questa scheda sono inserite, infatti, funzioni speciali per le analisi geostatistiche. Al momento questa scheda si adatta solamente al database PostgreSQL a causa di alcune funzionalità gestionali a livello di trigger che non è possibile implementare in Spatialite. La schermata principale è costituita da otto campi: sito, US, area, periodo, fase, coordinate x, y, z. I primi cinque sono da riempire obbligatoriamente, in quanto chiavi esterne che permettono le relazioni con la scheda sito, con la tabella geometrica pyarcheozoo e la tabella alfanumerica fauna. Inoltre la scheda è divisa in sette sezioni: parametri di connessione, inserimento dati, grafici, statistica univariata, statistica multivariata, geostatistica, tool.

3.1 *Parametri di connessione*

La sezione dedicata ai parametri di connessione è stata inserita per favorire la connessione del database direttamente con il programma statistico R. I parametri necessari per la connessione sono: host, user, password, porta, nome del database. Il sistema sfrutta un modulo Python chiamato pyper, che serve per far dialogare R con il linguaggio Python, mentre utilizza la libreria RPostgreSQL per collegare R con il database. Inoltre va specificata la path dove salvare le immagini prodotte e la risoluzione in pixel (Fig. 1). Una volta impostati questi parametri il database è connesso automaticamente.

3.2 *Inserimento dati*

Questa sezione prevede l'inserimento dei dati riguardanti i singoli ossi rinvenuti e studiati (Fig. 2). Le voci sono state estrapolate dal database preesistente e i dati sono inseriti nella tabella nominata fauna che attraverso dei trigger compila automaticamente altre tabelle che serviranno per il conteggio statistico. Le voci "ordine", "famiglia", "classe" e "specie" sono dei campi precompilati e descrivono tutti gli elementi finora ritrovati. Al momento il sistema non prevede l'inserimento di nuove specie, in quanto le relazioni tabellari create non permettono modifiche se non manuali. Il sistema è in via di sviluppo per adattarlo a tutti i contesti di scavo archeologico.

3.3 *Grafici e statistica univariata*

La statistica descrittiva univariata ha come obiettivo lo studio della distribuzione di ogni variabile, singolarmente considerata, all'interno della



Fig. 1 – Scheda archeozoologica: sezione Parametri di connessione.

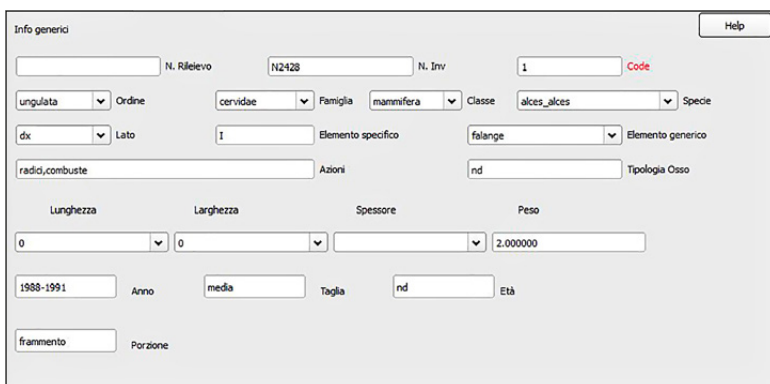


Fig. 2 – Scheda archeozoologica: sezione Inserimento dati.

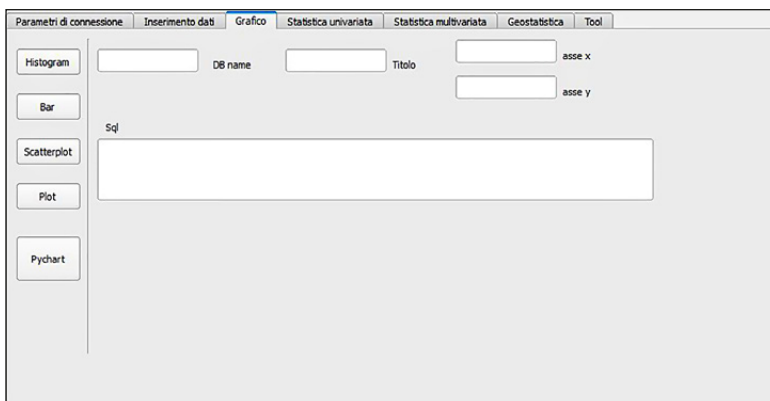


Fig. 3 – Scheda archeozoologica: sezione Grafico.

popolazione (analisi per colonna), mentre la statistica descrittiva multivariata si occupa dello studio della distribuzione di due o più variabili congiuntamente considerate. In questa sezione si possono produrre analisi di frequenza sotto forma di istogrammi su ciascuna specie presente nel database con gli indici di posizione. C'è la possibilità di attribuire al grafico un titolo, un nome per l'asse x e un nome per l'asse y. Inoltre si possono creare cinque tipi di grafici a seconda dell'esigenza, utilizzando una query da scrivere manualmente nel campo sql e quindi interrogando direttamente il database (Fig. 3). Si è voluto creare questa sezione in modo tale da rendere flessibile e particolareggiata l'estrapolazione dei dati statistici. I grafici che si possono creare sono istogrammi, bar, scatterplot, plot, grafici a torta e confronti a quattro variabili con istogrammi di frequenza, curve di densità e curve cumulative.

3.4 Statistica multivariata

Spesso è interessante studiare due o più variabili simultaneamente. L'osservazione contemporanea di due o più variabili sugli individui di una collettività consente di evidenziare le eventuali relazioni esistenti fra le variabili e di valutare il loro grado di correlazione. In questa sezione si è applicata una funzione (Figg. 4-5) che si occupa dell'analisi di due variabili congiuntamente considerate; in particolare, risulta interessante sapere se, e in quale modo, le due variabili si influenzano o se, al contrario, sono indipendenti dall'altra in base ad alcuni indici in grado di interpretare il tipo di legame esistente tra due variabili. In definitiva, si tratta di raccogliere i dati in una tabella a doppia entrata (o tabella di contingenza) in grado di mostrare congiuntamente le modalità dei due caratteri.

In questa sezione è possibile produrre una matrice di correlazione, che permette di analizzare tutte le variabili prese per coppie e di definire se all'aumentare dell'una variabile aumenta anche l'altra. Se le coppie di variabili avranno un indice di correlazione (ρ) maggiore-uguale a 0.5 (coppie segnate in rosso nel grafico) e un p.value con il valore basso saranno correlate, questa sarà la prova statistica che indica una relazione di buona validità del parametro ρ con coefficiente di significatività pari al 95%. La matrice di correlazione può essere applicata al momento alle tabelle `tot_famiglia` e `tot_ordine`.

3.5 Geostatistica

La sezione geostatistica si divide in due gruppi. Nel primo gruppo sono state inserite le funzioni per la costruzione di un semivariogramma (Fig. 6). Questo tipo di statistica permette di fare delle correlazioni spaziali su coppie di variabili, fino a un massimo di 13, che possono essere scelte dai menù a

```

from pyperimpon *
from pyarchinit_folder_installation impon *
for i in range(len(sef.DATATA_LIST)):
    temp_dataset = 0
    try:
        imp_dataset = (sef.DATATA_LIST[i].fma)
        conname += (sef.DATATA_LIST[i].fma) #contaggio totale
        dataset.append(imp_dataset)
    except:
        pass

r = R()
r.library(RPostgreSQL)
r.library(Rtibble)
r.con <- dbDriver("PostgreSQL")
con = r$con <- dbConnect(drv, host="127.0.0.1", port="5432", passwd="postgres", user="postgres")
#(sef.con.currentText(), sef.sef.db.conn(), sef.sef.port.currentText(), sef(sef.password.sef()), sef(sef.user.currentText()))
eval(con)
g = "Kjnp g(" %A %d correlation matrix png", width=2500, height=2500, pointsize=200) %
#(sef.sef.inaEdi_tesp_gw.scale.mat(), sef(sef.DATATA_LIST[i].fma))
eval(g)
h = "Kjarchaeology::table <- dbGetQuery(con, 'select * from tot_specie_terrestre where fase = %d') %
#(sef.DATATA_LIST[i].fma)
eval(h)
r("

panel.hist <- function(x, ...) {
  usr <- par("usr")[1:2]
  par(mfrow=c(usr[1], 2), 1:5)
  h <- hist(x, plot=FALSE)
  breaks <- h$breaks; nb <- length(breaks)
  y <- h$counts; y <- y/max(y)
  mcl(breaks[=B], 0, breaks[-1], y, col="cornsilk2", ...)
}

panel.cor <- function(x, y, digits=3, prefix="", cex.cor) {
  usr <- par("usr")[1:2]

```

Fig. 4-5 – Blocco di codice per la creazione della matrice di correlazione delle prime 10 specie presenti nella tabella tot_specie_terrestre (in pyarchinit_Archeozoology_mainapp.py linea 2327-2399).

Costruisci un semivariogramma

	ID	formula	ID	formula
<input type="text"/> Cutoff	<input type="checkbox"/> calcinati	calcinati	<input type="checkbox"/> pdi	bos_bison
<input type="text"/> Pail	<input type="checkbox"/> calcinati	bos_bison	<input type="checkbox"/> stambecco	bos_bison
<input type="text"/> Sph	<input type="checkbox"/> camoscio	bos_bison	<input type="checkbox"/> stria	bos_bison
<input type="text"/> Range	<input type="checkbox"/> capriolo	bos_bison	<input type="checkbox"/> conidi	bos_bison
<input type="text"/> Nugget	<input type="checkbox"/> cervo	bos_bison	<input type="checkbox"/> ursidi	bos_bison
<input type="text"/> 0.1 Lag	<input type="checkbox"/> combusto	bos_bison	<input type="checkbox"/> megacero	bos_bison
<input type="text"/> 0.1 Kappa	<input type="checkbox"/> cori	bos_bison		

Fig. 6 – Scheda archeozoologica: sezione Geostatistica.

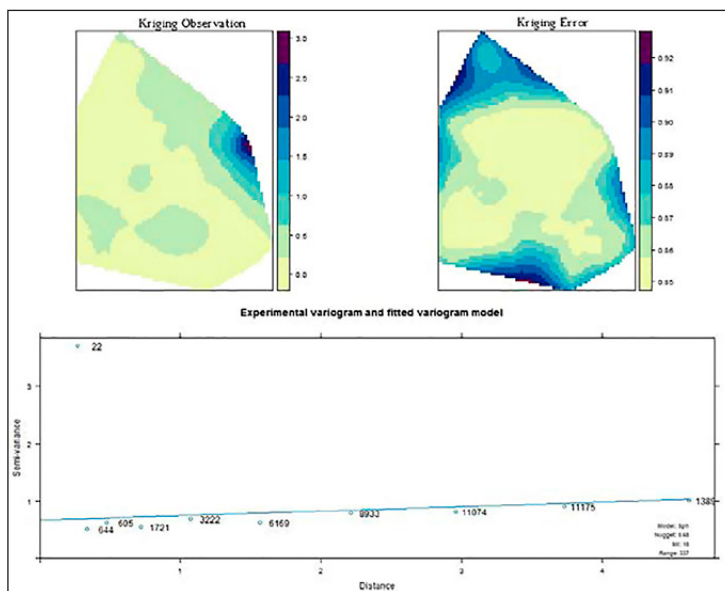


Fig. 7 – Semivariogramma sperimentale, fase 1.

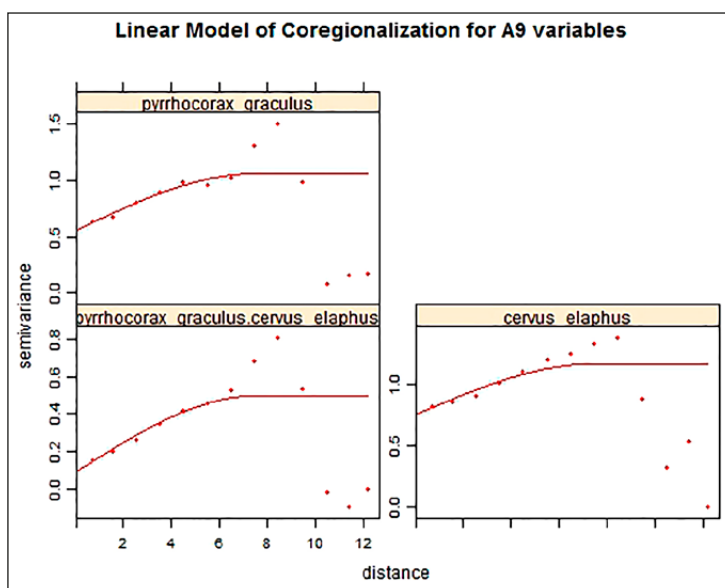


Fig. 8 – Semivariogramma sperimentale, fase 2.

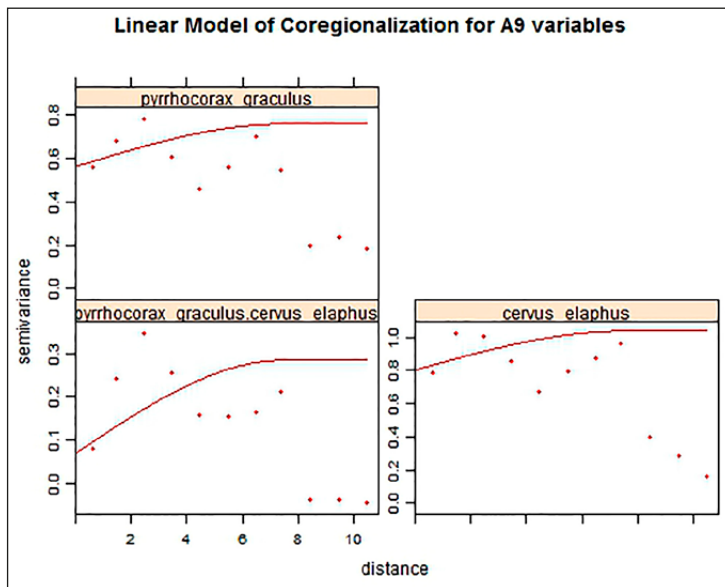


Fig. 9 – Distribuzione del *Cervus elaphus*, fase 1.

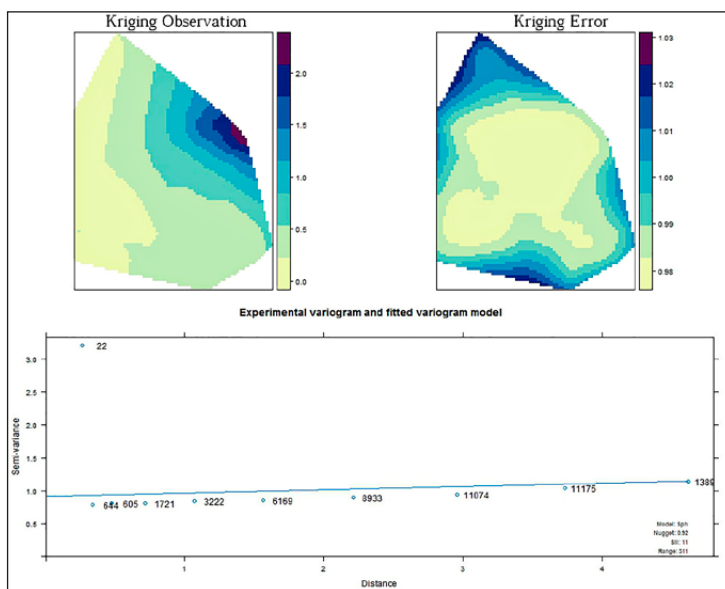


Fig. 10 – Distribuzione del *Pyrrhocorax graculus*, fase 1.

tendina. Per questa funzione si utilizza il modulo `gstat` di R. I parametri da utilizzare sono il `cutoff`, `partial sill`, `model`, `range`, `nugget effect`, `lag`, `kappa`. Una volta costruito il semivariogramma è possibile visualizzarlo anche come mappa.

Nel secondo gruppo invece è possibile costruire delle mappe di distribuzione per singola variabile. Con questo metodo si è riusciti a definire tutte le distribuzioni delle singole variabili delle specie per fasi. I parametri da inserire sono gli stessi usati per la costruzione del semivariogramma. Essi possono essere lasciati invariati, e in questo caso saranno presi i valori in automatico, oppure si possono inserire manualmente nel momento in cui la deviazione standard non tende allo zero. Infatti, il sistema produrrà tre file di testo dove verranno scritti i risultati dell'errore medio (ME), dell'errore quadratico medio (MSE) e la deviazione standard (MSDR). La mappa di distribuzione sarà salvata in raster `.tiff`, georiferita e tramite lo strumento `clipper` sarà possibile ritagliarne i bordi, in modo tale da farla combaciare con il perimetro dello scavo e quindi della parte indagata (nel nostro caso il poligono perimetrale della grotta).

4. CONCLUSIONI

Le analisi statistiche effettuate sulla fauna di Grotta di Fumane consentono di fare alcune considerazioni importanti per le successive analisi spaziali. A seconda del tipo di raggruppamento che si è evidenziato durante queste analisi si può così porre in luce in modo sempre più accurato il grado di dettaglio che si vuole raggiungere.

Partendo dal raggruppamento in classi, passando attraverso i gruppi di ordine e famiglia, sono state evidenziate le specie tra fauna terrestre e avifauna che hanno un grado di correlazione. Sulla base di questi risultati è stato possibile effettuare un'analisi geostatistica e confrontare se esiste una correlazione spaziale tra le variabili prese in considerazione (Figg. 7-8). Dal risultato ottenuto si evidenzia, ad esempio, che il *Pyrrhocorax Graculus* ha una correlazione spaziale con il *Cervus Elaphus* in un range di sette metri e mezzo e in qualsiasi punto in cui si osservi la grotta. Creando una mappa di distribuzione per le due specie, si nota inoltre che entrambe sono localizzate nello stesso luogo e cioè sono distanti dalle strutture di combustione e dai focolari (Figg. 9-10).

ENZO COCCA

Dipartimento Africa Asia e Mediterraneo
Università L'Orientale di Napoli
enzo.ccc@gmail.com

LUCA MANDOLESÌ

Ad Arte snc.
mandoluca@gmail.com

BIBLIOGRAFIA

- COCCA E. 2014, *Kojtepa 2013: The use of 3D for the drawings of excavation. A methodological approach*, «Newsletter di Archeologia CISA», 5, 1-20.
- COCCA E. 2015, *Il GIS nell'ambito di sistemi innovativi per la gestione del dato archeologico. Sviluppo e implementazione di un sistema gestionale e analitico con strumenti open source di banche dati archeologiche. Caso studio Grotta di Fumane*, Tesi di Dottorato di ricerca in Scienze e Tecnologie per l'Archeologia e i Beni Culturali, Università di Ferrara, XXV ciclo, a.a. 2015.
- MANDOLESI L., COCCA E. 2013, *Pyarchinit: gli sviluppi dopo ArchoFOSS 2009*, in M. SERLORENZI (ed.), *ArchoFOSS. Free, Libre and Open Source Software e Open Format nei processi di ricerca archeologica. Atti del VII Workshop (Roma 2012)*, «Archeologia e Calcolatori», Suppl. 4, 128-138.
- PEBESMA E.J. 2004, *Uni- and multivariable geostatistical modelling, prediction and simulation* (<http://www.gstat.org/bin/gstat.pdf/>).
- PERESANI M. 2008, *Age of the final Middle Palaeolithic and Uluzzian levels at Fumane Cave, Northern Italy, using 14C, ESR, 234 U/230 Th and thermoluminescence methods*, «Journal of Archaeological Science», 35, 2986-2996.
- ROMANDINI M., NANNINI N., TAGLIACCOZZO A., PERESANI M. c.s., *The ungulate assemblage from layer A9 at Grotta di Fumane, Italy: A zooarchaeological contribution to the reconstruction of Neanderthal ecology*, «Quaternary International», 1-17, in corso di stampa.
- VALENTI M. 2000, *La piattaforma GIS dello scavo. Filosofia di lavoro e provocazioni, modello dei dati e soluzione GIS*, «Archeologia e Calcolatori», 11, 93-109.

ABSTRACT

The analysis of the spatial distribution of archaeological records combined with geostatistical techniques plays a primary role in the field of prehistoric archaeology because this methodological approach can be used to understand if a real or chance spatial correlation exists between archaeological records. This essay presents an approach to this issue which makes use of statistics, databases and GIS open source tools in a one-solution instrument based on QGIS desktop environment, in order to study an important archaeological site in Northern Italy (Fumane Cave, Verona). Since archaeology is an intrinsically spatial discipline, the team chose a spatial quantitative approach like the multivariate geostatistics (which is usually used for mining and environmental analysis), by managing the quantitative fieldwork information as regionalized variables. In this way, interesting spatial correlations were evaluated by imposing linear coregionalization models on experimental semivariograms and cross-semivariograms. Moreover, the team conducted this analysis by creating a tool that integrates several powerful R libraries such as Gstat for spatial distribution analysis of data, implemented in DBMS PostgreSQL with PostGIS extension.