

CARATTERIZZAZIONE DI GRUPPI DI REPERTI FITTILI DELLA LAGUNA DI VENEZIA MEDIANTE ANALISI DISCRIMINANTE APPLICATA AI DATI CHIMICI

1. INTRODUZIONE

La subsidenza dell'area lagunare, oggetto da tempo di specifiche ricerche da parte dell'Istituto per lo Studio della Dinamica delle Grandi Masse del CNR, ha comportato, in epoca storica, il progressivo sprofondamento e talora la sommersione di antichi insediamenti abitativi. Indagini multidisciplinari sono particolarmente mirate agli ultimi due millenni in cui l'intensità del fenomeno ha avuto dei bruschi incrementi dovuti a diversificati interventi antropici (diversione dei fiumi, bonifiche, ecc. e, in tempi recenti, sfruttamento delle falde acquifere). Tra i vari approcci lo studio dei reperti archeologici consente di giungere alla datazione assoluta e relativa degli strati e quindi alla conoscenza degli andamenti spaziali e temporali del fenomeno stesso. In questo ambito si è già svolto un primo studio di datazione, applicando un metodo di analisi multivariata di tipo "fuzzy c-means clustering" a laterizi provenienti da vari siti archeologici lagunari, basata esclusivamente sui parametri formali (dimensioni geometriche) (CANAL, CAVAZZONI 1990).

Con il programma ora in corso ci si propone di estendere l'elaborazione, con ulteriori tecniche matematico-statistiche, ai dati provenienti dalle analisi chimiche di reperti archeologici fittili (laterizi e ceramiche) lagunari di cui sono noti sia il sito archeologico di provenienza che la datazione (da rilevanze archeologiche e fonti storico-archivistiche). Lo scopo è quello di caratterizzare chimicamente fittili ben datati e di riuscire a datare fittili di età sconosciuta per confronto dei dati analitici con quelli di fittili ben datati. Questo nell'ipotesi che nelle varie aree archeologiche ed in determinate epoche le argille impiegate per i cotti provengano preferibilmente da determinati bacini sedimentari.

In attesa di poter disporre di un congruo numero di dati da analisi chimiche recentemente avviate, si vuole, con questo lavoro, affrontare le problematiche relative ad una tecnica di analisi multivariata inerente non tanto alla formazione di gruppi, quanto alla loro caratterizzazione e differenziazione: l'analisi discriminante.

2. REPERTI FITTILI ANALIZZATI

Si sono presi in esame quattro gruppi di reperti fittili la maggior parte dei quali provenienti dalla laguna di Venezia di età medioevale e rinascimentale studiati in precedenti lavori (LAZZARINI *et al.* 1980; CALOGERO, LAZZARINI 1983; LAZZARINI 1987) e già suddivisi in gruppi sulla base di analisi storico-artistiche, mineralogiche e chimiche (valori medi).

Si tratta di quattro gruppi di reperti, tre dei quali formati da frammenti di ceramica e il quarto di laterizi nella quasi totalità di un tipo particolare detto "altinella", che indagini storiche fanno risalire ad una precisa epoca, il XII-XIII sec. (FAZIO *et al.* 1982).

Dei tre gruppi di ceramiche quelle del primo e del secondo sono state classificate come coeve ma di diversa area geografica (bizantine e veneziane arcaiche), quelle del secondo e del terzo gruppo sono tutte lagunari ma di epoche diverse (veneziane arcaiche e veneziane rinascimentali).

I quattro gruppi sono costituiti rispettivamente da:

1°) 14 ceramiche graffite bizantine di cui 9 provengono dalla laguna di Venezia e le altre sono campioni rappresentativi di 5 siti greci (MEGAW, JONES 1983; LAZZARINI, CALOGERO 1989);

2°) 13 ceramiche graffite veneziane arcaiche (XII-XIII sec.);

3°) 16 ceramiche graffite veneziane rinascimentali (XV-XVI sec.);

4°) 16 mattoni veneziani di cui 15 "altinelle" (XII-XIII sec.) e 1 tardo antico.

3. ANALISI MULTIVARIATA

L'analisi ha lo scopo di verificare se gruppi di diversi reperti fittili possano essere differenziati uno dall'altro sulla base dei valori dei loro costituenti chimici e se, di conseguenza, un nuovo reperto possa essere assegnato ad uno dei quattro gruppi sulla base di questi valori. I parametri o variabili su cui si basa l'analisi sono le concentrazioni, espresse in percentuale, degli elementi chimici rilevati in numero di dieci per i primi tre gruppi (ceramiche) e di otto (talvolta nove) per il quarto gruppo (mattoni). Un'analisi che coinvolga tutti quattro i gruppi è quindi solo possibile sulla base dei valori degli otto parametri comuni.

Le somme dei valori percentuali utilizzati non sono uguali a cento, neanche per il 4° gruppo poiché tra essi non è mai inclusa la percentuale di perdita al fuoco: si può quindi escludere una reciproca dipendenza lineare tra i dati. Tuttavia per alcuni anomali elementi del 4° gruppo le percentuali sono state ricalcolate affinché le loro somme fossero dello stesso ordine di grandezza delle altre.

In questo lavoro si vuole sperimentare l'applicazione dell'analisi discriminante su quattro gruppi costituiti da un numero scarso di elementi rispetto al

numero di parametri considerati e non soddisfacenti del tutto alle condizioni statistiche richieste per la validità dell'analisi stessa.

Questo nella convinzione che, anche in tali condizioni, l'analisi discriminante possa conservare validità qualora i gruppi abbiano una "forte struttura", cioè siano fortemente differenziati dalle loro medie multivariate (CATLING *et al.* 1980, 68), come nel nostro caso.

Per verificare la "forte struttura" dei gruppi si è applicata preliminarmente l'analisi delle componenti principali, in sigla PCA, al complesso dei dati dei quattro gruppi.

Per lo scopo che ci si propone è preferibile trasformare i valori originali dei parametri in "scarti" dalla comune media multivariata. Questi scarti o scostamenti dalla media, saranno capaci di riprodurre i gruppi originali, quanto più le medie multivariate dei gruppi risultano differenziate tra loro. Tali scarti sono stati inoltre standardizzati, cioè divisi per la loro deviazione standard, commisurandoli così inversamente con la dispersione dei dati originali.

La PCA sostituisce alle variabili originali un numero minore di nuove variabili (combinazioni lineari delle vecchie variabili) dette componenti principali, rispetto alle quali è resa massima la varianza complessiva del sistema. Quando il complesso dei dati ha una distribuzione normale i punteggi su ciascuna componente (cioè i valori rispetto alle nuove variabili), essendo somme ponderate di molte variabili, tendono a loro volta ad essere distribuiti normalmente, addensandosi attorno alla media. Se invece vi sono vari centri di addensamento, come si suppone che nel nostro caso siano le medie dei quattro gruppi, dall'analisi emerge una struttura a "clusters".

Nei diagrammi cartesiani delle prime due o tre componenti principali possono essere rappresentati, e quindi verificati visivamente, i gruppi e l'importanza delle variabili originali nella loro formazione.

L'analisi discriminante, a differenza della PCA, è stata applicata ai dati dei quattro gruppi considerati separatamente: essa trasforma le variabili in un numero minore di variabili rispetto alle quali è possibile individuare relazioni tra i gruppi, valutare l'importanza delle variabili originali nella discriminazione tra i gruppi e classificare nuovi elementi. Le nuove variabili, dette variabili canoniche, sono combinazioni lineari delle vecchie variabili; rispetto ad esse è reso massimo il rapporto della matrice di varianza-covarianza tra i gruppi [B] con quella entro i gruppi [W]. Esse hanno la forma:

$$y = b_1x_1 + b_2x_2 + \dots + b_px_p \quad (1)$$

dove i coefficienti b_i sono gli elementi degli autovettori corrispondenti agli autovalori della matrice $[W^{-1}B]$ e le x_i sono le p variabili originali.

Se i gruppi sono g , con p variabili, le variabili canoniche sono in numero di $p^* = \min(g-1, p)$.

Applicando ai dati la trasformazione (1) si ottengono i valori rispetto alle nuove variabili, che considerati come coordinate, possono essere rappresentati graficamente a coppie nel piano cartesiano. Rispetto a ciascuna coppia di variabili si può così ottenere una rappresentazione grafica della posizione reciproca dei gruppi, dei loro centri (medie multivariate) e valutare rispetto a quale variabile canonica meglio si realizza la discriminazione tra essi. Un nuovo elemento può essere assegnato al gruppo dal cui centro la sua distanza sarà minore che dagli altri.

I "pesi" delle variabili originali, cioè il loro contributo relativo nella discriminazione tra i gruppi, relativamente a ciascuna variabile canonica, sono rappresentati dai prodotti $b_i \delta_i$, dove δ_i sono le radici quadrate delle devianze della matrice [B], opportunamente scalati su ciascun autovalore. Poiché i loro rapporti reciproci sono invarianti rispetto alle trasformazioni di scala risulta inutile applicare particolari trasformazioni lineari alle variabili che presentano differenti scale metriche, come suggeriscono alcuni autori (MAXWELL 1977, 98).

Un ulteriore procedimento applicato alle nuove variabili fornisce, per ogni gruppo, una funzione discriminante che consente a sua volta di attribuire nuovi elementi ai vari gruppi. Ciascuna funzione discriminante, caratterizzante un gruppo, è espressa da:

$$g_i(x) = \ln[F_i(x)p_i] \quad i = 1, 2, \dots, g$$

dove F_i indica la funzione di densità normale multivariata e p_i la probabilità bayesiana "a priori". Ai soli fini della classificazione di un nuovo elemento campionario ad uno dei gruppi il procedimento può essere applicato direttamente alle variabili originali. Nell'applicazione si è ritenuto opportuno usare le trasformate canoniche in quanto esse permettono una più articolata interpretazione dei dati in termini di importanza delle variabili originali ai fini della discriminazione. Questo aspetto può avere fini operativi, nel senso di risparmio nel numero delle analisi chimiche da esaminare ai fini della classificazione.

Le condizioni statistiche per l'applicabilità dell'analisi discriminante richiederebbero, oltre alla diversità dei valori medi delle variabili nei vari gruppi, verificata nel nostro caso, anche l'uguaglianza delle matrici di varianza-covarianza dei vari gruppi e la normalità multivariata dei dati. Inoltre per una stima attendibile dei parametri (vettori delle medie e matrice di varianza-covarianza) si dovrebbe disporre di campioni casuali di notevoli dimensioni per ogni gruppo.

Dal punto di vista statistico i nostri quattro gruppi sono costituiti da un numero scarso di reperti. Nella loro costituzione inoltre si è tenuto conto, oltre che di criteri tecnico-stilistici, anche dei parametri chimici su cui si basa questa analisi: ciò esclude che gli oggetti costituenti i gruppi siano stati scelti casualmente (ma, d'altra parte, consente ai gruppi di avere una "forte struttura").

Neppure le matrici di varianza-covarianza sono uguali. Molti autori (MORRISON 1976, 137; DAVIS 1986, 485) tuttavia ammettono che l'analisi discriminante può essere significativa anche in tali condizioni anomale.

La riduzione della dimensionalità ottenuta con le variabili canoniche può essere ulteriormente utilizzata nell'operazione di confronto tra i gruppi in termini geometrici. Un confronto tra i gruppi mediante l'analisi delle componenti principali (PCA) applicata a ciascuno di essi può fornire informazioni soltanto su singularità di tipo lineare. Applicando invece l'analisi delle componenti principali quadratiche, in sigla QPCA, a ciascun gruppo, relativamente alle variabili canoniche, è possibile rilevare relazioni anche di tipo non lineare nella struttura dei gruppi. In particolare la componente corrispondente al più piccolo autovalore individua la forma geometrica che maggiormente si adegua alla struttura dei singoli gruppi (GNANADESIKAN 1977, 53-61).

Limitando l'attenzione alle prime due variabili canoniche (y_1, y_2) ogni gruppo viene rappresentato da un'equazione del tipo

$$z = a_1 y_1 + a_2 y_2 + a_3 y_1 y_2 + a_4 y_1^2 + a_5 y_2^2$$

L'algoritmo delle QPCA applicato alla quintupla di variabili ($y_1, y_2, y_1 y_2, y_1^2, y_2^2$) rendendo massima la varianza di z per ciascun gruppo consente di ricavare i coefficienti a_i per le quattro forme quadratiche, di riconoscerne la natura e di rappresentarle nel piano (y_1, y_2).

4. RISULTATI E LORO DISCUSSIONE

L'analisi PCA applicata al complesso dei quattro gruppi di dati individua le nuove variabili corrispondenti agli assi principali della dispersione dei dati nello spazio (iperspazio).

Poiché tali assi sono quelli lungo i quali sono massime le correlazioni tra le variabili, l'analisi evidenzia, in primo luogo, le strutture di correlazione del complesso di dati.

Così la prima componente, che spiega un 58% della varianza complessiva, è massimamente correlata con Ca, Ti, Si e Mg e, in misura un poco inferiore con Al e K: dall'esame della matrice di correlazione si nota che tali elementi sono tra loro sensibilmente correlati. Alla seconda componente (15% di varianza totale) sono correlati soltanto Na e Fe (in misura inferiore) che però tra loro non sono correlati.

Il Na, soprattutto, non risulta correlato con alcuna altra variabile ed è quindi di particolare importanza come variabile indipendente nel sistema. Complessivamente nel piano cartesiano delle due prime componenti i quattro gruppi risultano ben evidenziati (Fig. 1) ad eccezione del terzo che risulta diviso in due sottogruppi caratterizzati dai valori alti o bassi del Ca. Solo alcuni elementi ap-

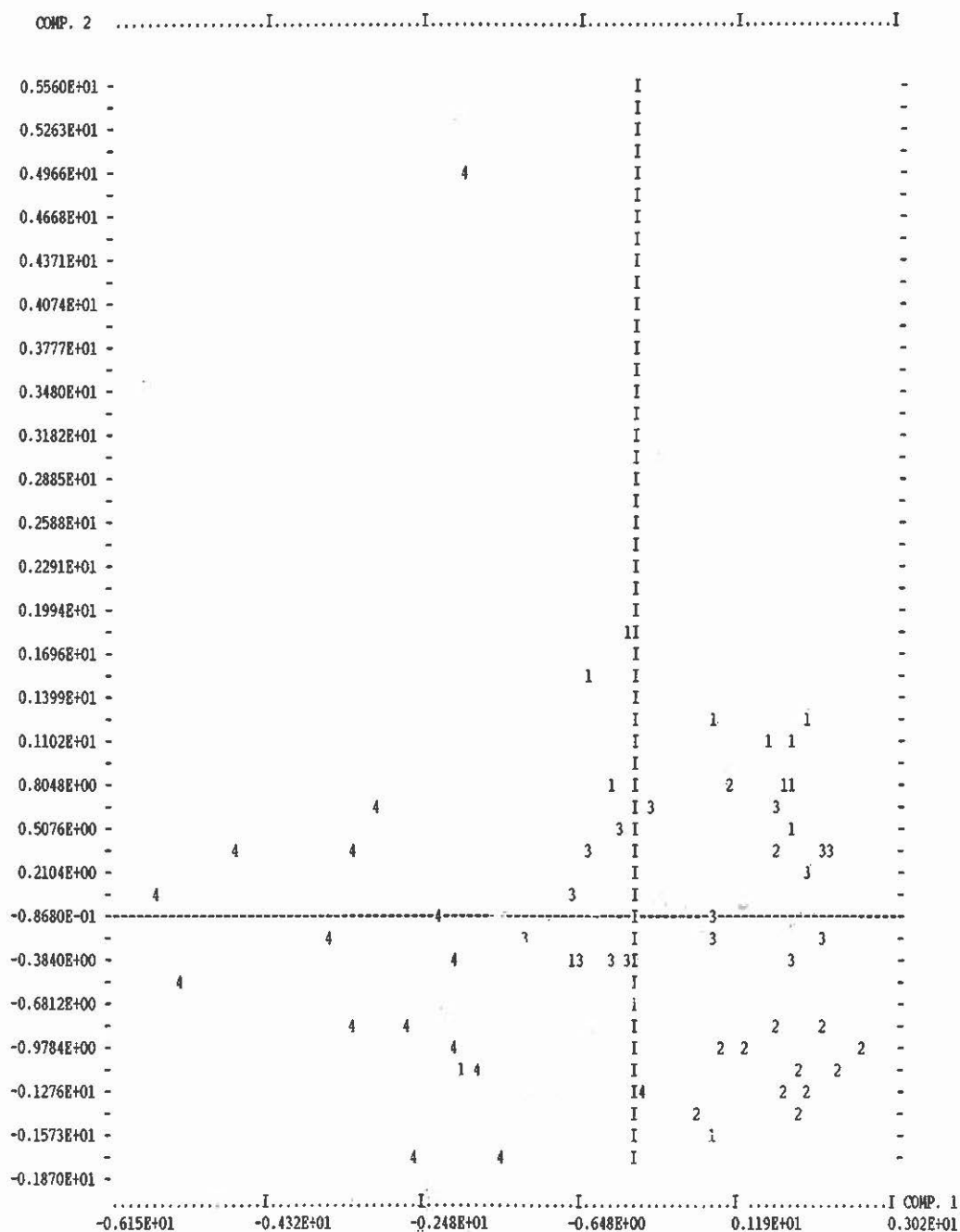


Fig. 1 — Analisi PCA. Nel piano cartesiano delle prime due componenti principali sono rappresentati i quattro gruppi di reperti coi simboli 1, 2, 3, 4.

paiono anomali. Si può quindi parlare di una "forte struttura" dei gruppi. Con questo si vuol dire che gli scarti dalla media multivariata, relativamente alla deviazione standard di ciascuna variabile, hanno una struttura discontinua con diversi centri di addensamento per i vari gruppi.

Con l'analisi discriminante si sono ottenute tre variabili canoniche.

La significatività dei tre autovalori λ_j è verificata per mezzo del test di Bartlett, che si distribuisce come una χ^2 con $p + g - 2j$ gradi di libertà (MAXWELL 1977). Essi risultano tutti significativi:

autovalore	gradi di libertà	χ^2
$\lambda_1 = 4,14$	10	84,95
$\lambda_2 = 2,30$	8	61,15
$\lambda_3 = 0,71$	6	27,86

Le prime due variabili canoniche che complessivamente hanno un potere discriminante pari al 90% del totale possono rappresentare con buona approssimazione il sistema.

I pesi relativi delle variabili originali nella formazione delle tre variabili canoniche sono rispettivamente:

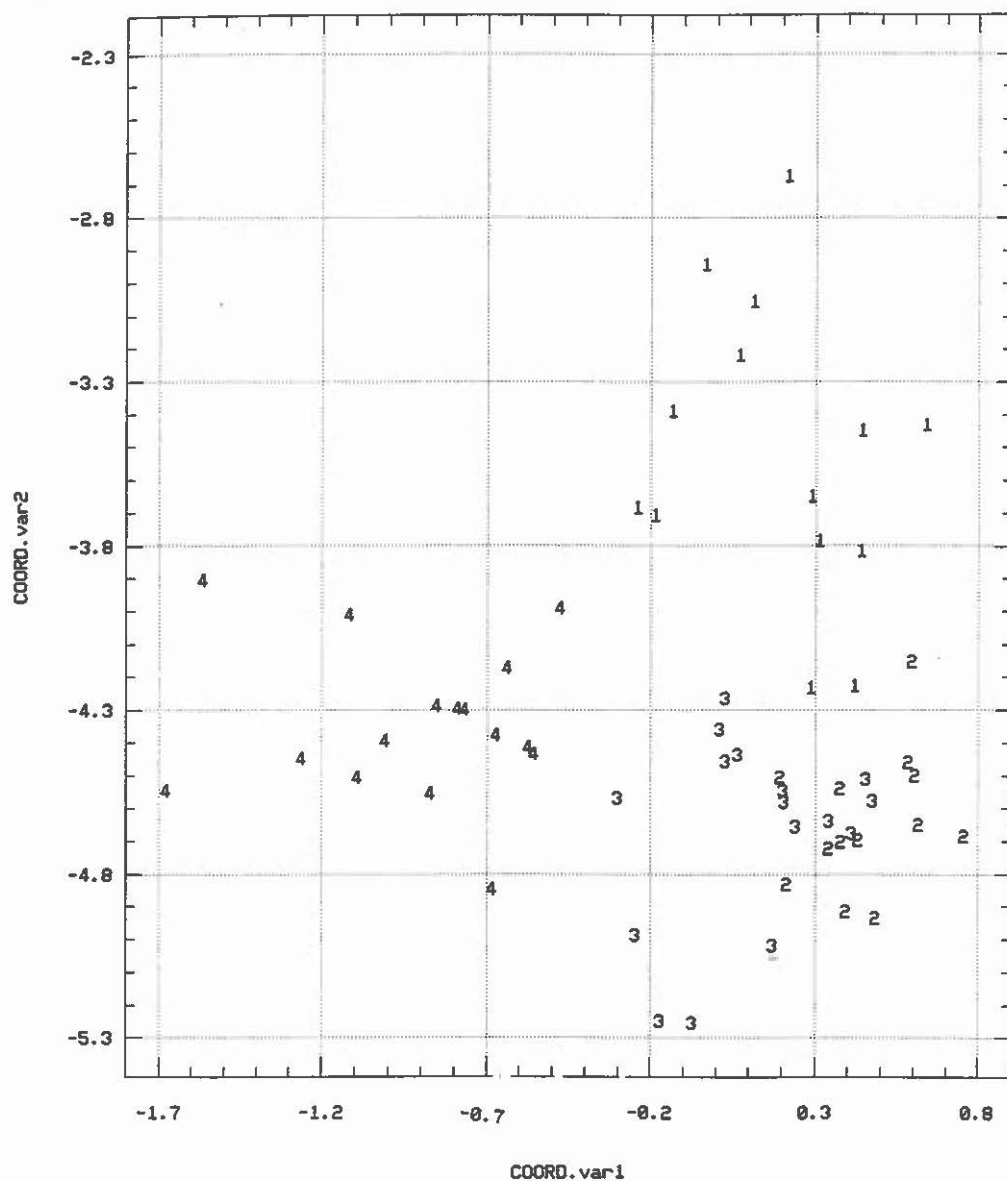
Ca	Fe	Ti	K	(Mg	Na	Si	Al)
-0,82	0,72	-0,52	0,39	-0,20	-0,19	0,22	-0,08
Al	Fe	Si	Mg	(K	Ca	Ti	Na)
-0,41	0,35	-0,30	-0,27	-0,17	-0,16	0,15	-0,00
Ca	Al	Si	Mg	(K	Fe	Ti	Na)
-0,38	-0,35	-0,33	-0,29	0,09	-0,06	-0,06	-0,01

La rappresentazione grafica dei risultati è riportata nelle Figg. 2 e 3. Nel piano cartesiano delle prime due variabili canoniche (Fig. 2) sono rappresentati coi simboli 1, 2, 3, 4 i punti appartenenti ai quattro gruppi. Le coordinate di ciascun punto sono i punteggi relativi alle prime due variabili canoniche.

Si osserva che la prima variabile canonica, costituita principalmente dalle variabili Ca, Fe, Ti, discrimina il gruppo 4 dai primi tre, mentre il gruppo 1 è discriminato dagli altri tre gruppi dalla seconda variabile canonica che è soprattutto costituita dall'Al e dal Fe.

Nella rappresentazione tridimensionale relativa alle tre variabili canoniche (Fig. 3) i gruppi 2 e 3 appaiono discriminati dalla terza variabile, costituita principalmente da Ca, Al, Si.

Per un esame più dettagliato dei parametri atti alla discriminazione, sono stati sottoposti all'analisi discriminante i gruppi a due a due. L'analisi discrimi-



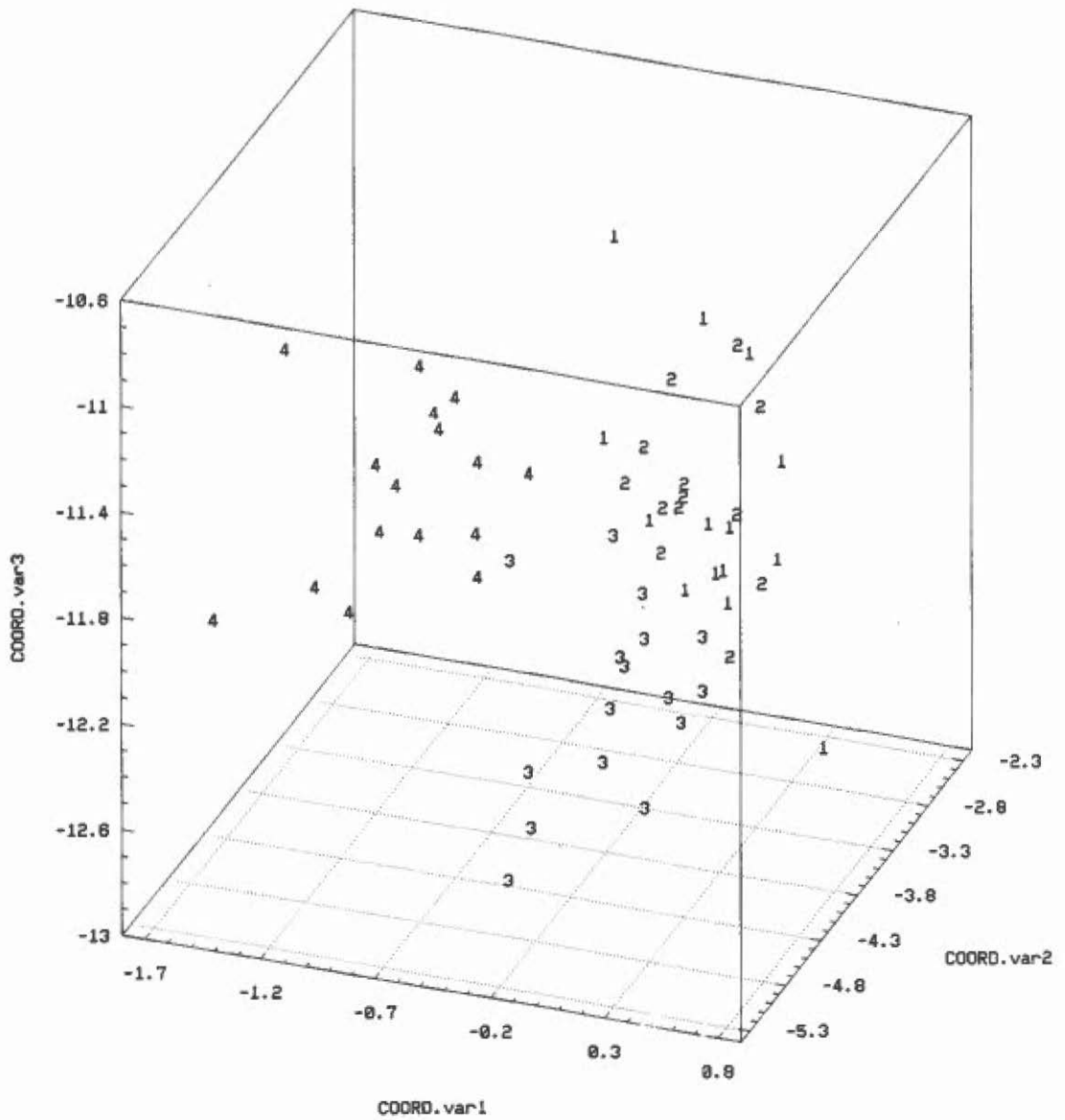


Fig. 3 — Analisi discriminante. I quattro gruppi sono rappresentati, coi simboli 1, 2, 3, 4, nello spazio individuato dalle tre variabili canoniche.

nante è stata applicata alle seguenti coppie:

1° - 2° tra ceramiche di provenienza bizantina e ceramiche di produzione veneziana entrambe dei sec. XII e XIII,

2° - 3° tra ceramiche di produzione veneziana del XII e XIII sec. e ceramiche di analoga produzione dei sec. XV e XVI,

2° - 4° tra ceramiche e mattoni (altinelle) entrambi veneziani dei sec. XII e XIII (ad eccezione di un mattone tardo antico).

Sono state calcolate le funzioni discriminanti per ciascuna delle tre coppie. I pesi, in ordine decrescente secondo l'importanza delle variabili nell'operazione di discriminazione, sono riportati di seguito.

Tra il primo e il secondo gruppo:

Ca	Si	Fe	(K	Al	Mg	Ti	Na)
-1,06	-0,44	-0,35	0,23	-0,16	0,13	-0,13	0,09

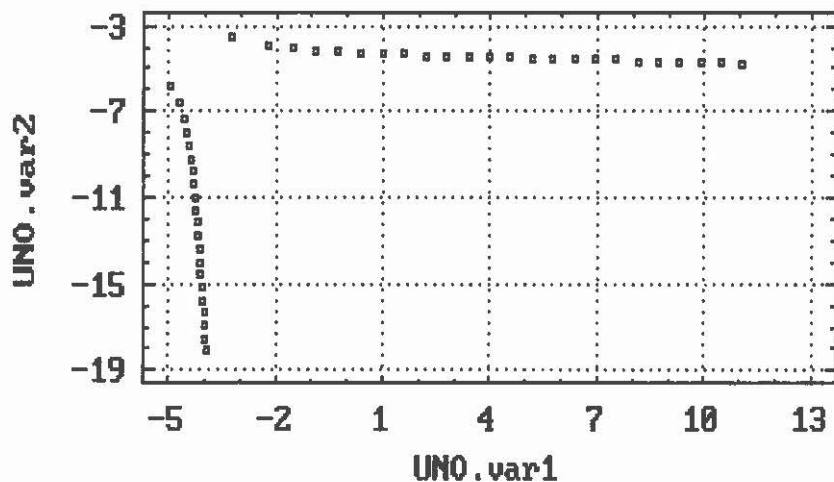
Tra il secondo e il terzo:

Mg	Ca	Si	(Na	Al	Ti	Fe	K)
0,82	0,72	0,51	0,34	0,20	0,08	-0,08	-0,01

Tra le ceramiche e i mattoni, gruppi 2 e 4:

Si	Al	K	(Ti	Ca	Na	Fe	Mg)
0,95	0,67	0,60	-0,59	-0,43	-0,08	0,08	0,03

A verifica della validità delle variabili canoniche nella discriminazione tra



Figg. 4-7 — Analisi QPCA. I quattro gruppi sono rappresentati da forme quadratiche nel piano delle prime due variabili canoniche.

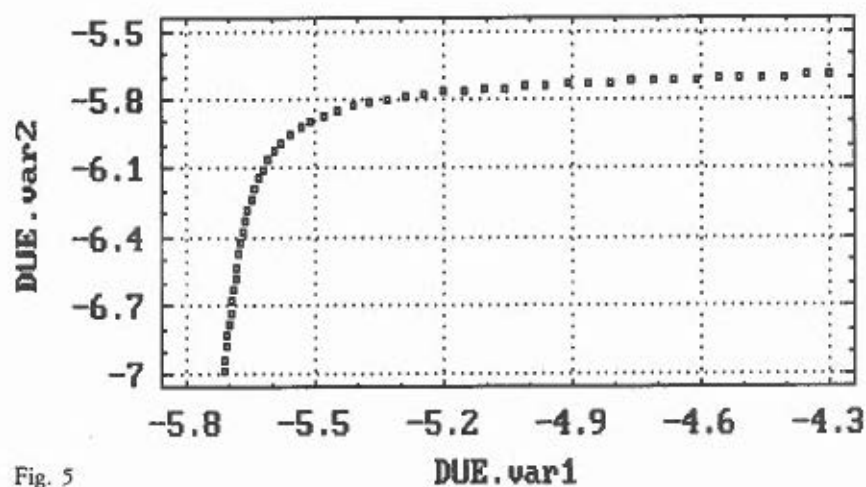


Fig. 5

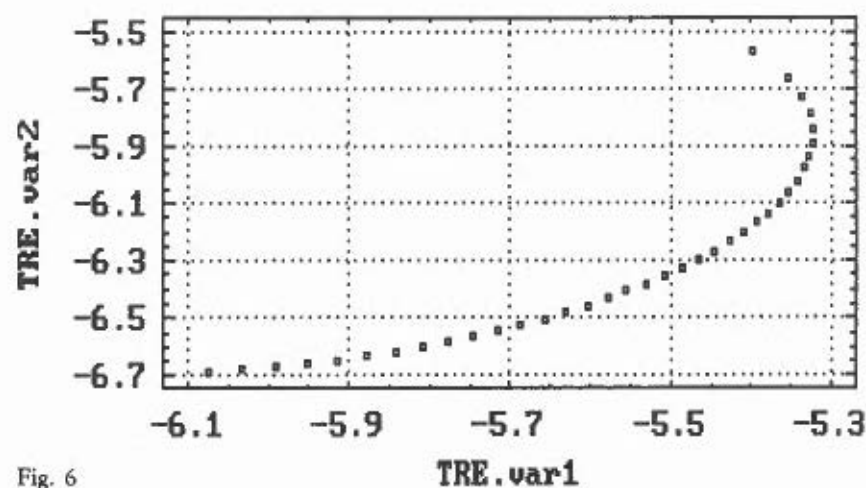


Fig. 6

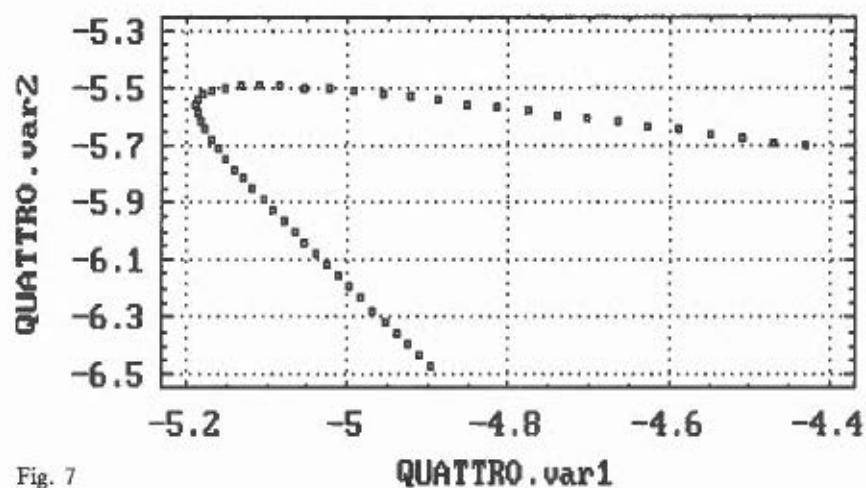


Fig. 7

i quattro gruppi si sono riclassificati tutti gli elementi dei quattro gruppi mediante quattro funzioni discriminanti applicate alle tre variabili canoniche, con uguale probabilità "a priori" di appartenenza a ciascun gruppo.

La matrice di riclassificazione nei quattro gruppi:

0,86	0,07	0,07	0,00
0,00	1,00	0,00	0,00
0,00	0,00	1,00	0,00
0,00	0,00	0,00	1,00

rileva solo un doppio errore relativo al primo gruppo i cui elementi solo in piccola parte, pari al 7%, vengono attribuiti rispettivamente al 2° e al 3° gruppo. La discriminazione a coppie presenta un solo errore di riclassificazione dal gruppo 1 al 2. Gli errori di classificazione, dovuti in entrambi i casi al primo gruppo, sono giustificati dalla possibile diversa manifattura d'origine dei relativi reperti, riuniti soltanto dalla definizione "bizantini".

L'analisi delle componenti principali quadratiche è stata applicata alle prime due variabili canoniche in ciascun gruppo. Le forme quadratiche nel piano (y_1 , y_2) corrispondenti a 1°, 2°, 4° gruppo risultano iperboli, solo il 3° gruppo, a cui corrisponde un'ellissi, ha una configurazione che si avvicina a quella di una distribuzione normale multivariata (Figg. 4-7). Ciò fa emergere una differenza funzionale oltre che parametrica nelle relazioni tra le variabili del terzo gruppo rispetto agli altri gruppi.

SILVIA CAVAZZONI, GIANCARLO TARONI

C.N.R., Istituto per lo Studio
della Dinamica delle Grandi Masse - Venezia

LORENZO LAZZARINI

Dipartimento Scienze della Terra
Università « La Sapienza » - Roma

BIBLIOGRAFIA

- CANAL E., CAVAZZONI S. 1990, *Antichi insediamenti antropici nella laguna di Venezia: analisi multivariata di tipo "fuzzy c-means clustering"*, « Archeologia e Calcolatori », 1, 165-177.
- CALOGERO S., LAZZARINI L. 1983, *Caratterizzazione chimico-fisica di ceramiche graffite bizantine e veneziane arcaiche trovate nella laguna veneta*, « Faenza », 69, 1-2, 60-70, Tav. XI.
- CATLING H.W., CHERRY J.F., JONES R.E., KILLEN J.T. 1980, *The linear B inscribed stirrup jars and west Crete*, Ann. of British School of Athens, 75.
- DAVIS J.C. 1986, *Statistics and Data Analysis in Geology*, New York, John Wiley & Sons, II edition.
- FAZIO G.S., HREGLICH L., LAZZARINI U., PIREDDA V., VERITÀ M. 1982, *Le altinelle a Venezia*:

- problemi storici, caratterizzazione chimico-fisica, cause di deterioramento*, in *Il mattone di Venezia*, Venezia, Comune di Venezia e C.N.R.-ISDGM.
- GNANADESIKAN R. 1977, *Methods for Statistical Data Analysis of Multivariate Observations*, New York, John Wiley & Sons, 53-61.
- LAZZARINI L. 1987, *Analisi chimico-mineralogiche su alcune ceramiche tardomedioevali del ritrovamento della Misericordia, Venezia*, « *Archeologia Veneta* », 10, 233-236.
- LAZZARINI L., CALOGERO S., BURRIESCI N., PETRERA M. 1980, *Chemical, Mineralogical and Mossbauer Studies of Venetian and Paduan Renaissance Sgraffito Ceramics*, « *Archaeometry* », 22, 1, 57-68.
- LAZZARINI L., CALOGERO S. 1989, *Early local and imported Byzantine sgraffito ware in Venice: a characterization and provenience study*, in *Proceedings of the 25th Inter. Symp., Archaeometry*, Elsevier, 571-584.
- MAXWELL A.E. 1977, *Multivariate Analysis in Behavioural Research*, London, Chapman and Hall.
- MEGAW A.H.S., JONES R.E. 1983, *Byzantine and allied pottery: a contribution by chemical analysis to problems of origin and distribution*, "The Annual of British School of Archaeology at Athens", 78, 235-263, Tavv. 24-30.
- MORRISON D.F. 1976, *Metodi di analisi statistica multivariata*, Milano, C.E.A.

ABSTRACT

Multivariate discriminant analyses are applied to the chemical composition data of four groups of fictile archaeological findings from the lagoon of Venice. Three groups are constituted of ceramic fragments of different manufacture, local or Byzantine imported and of different epochs, Middle Ages or Renaissance. The fourth group is formed of special Middle Age bricks ("altinelle"). Our aim is to find first the chemical characteristic of each group and, on this basis, to classify other unidentified findings. After having verified the structure of the four groups by means of PCA analysis, three discriminant canonical variates are determined: the relative weights of the original variables represent their discriminant power. The groups are also analyzed in pairs and their relative discriminating variables are found. Discriminant function on the basis of the three new variates are calculated for each of the four groups and used for their verification. Finally a quadratic PCA analysis applied to the first two variates of each group allows, by means of recognition of their geometric form, to also point out non linear relationships.