

Roberto Fanfani, Mario Mazzocchi*

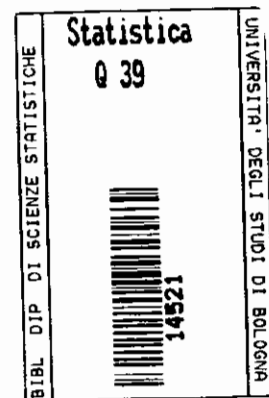
I metodi statistici per l'analisi
dei sistemi agricoli territoriali

Serie Ricerche 1999, n.2

*Dottorato di Ricerca in Economia e Politica Agraria



Dipartimento di Scienze Statistiche "Paolo Fortunati"
Università degli studi di Bologna



La presente ricerca è stata realizzata con il sostegno finanziario della Commissione Europea (dg VI – Agriculture and Fisheries) nell’ambito del programma di ricerca FAIR CT97-3403 “EUROTOOLS - Tools for evaluating EU agricultural policy at different decision levels” e con il contributo Murst 40% “Il sistema agroalimentare italiano e l’integrazione europea”

INDICE

PREFAZIONE	pag.	5
1. L’analisi dello sviluppo agricolo territoriale e i metodi statistici	pag.	6
2. Alcuni aspetti metodologici generali sulle tecniche utilizzate per le zonizzazioni	pag.	9
2.1 <i>L’interpretazione delle componenti principali</i>	pag.	11
2.2 <i>La Cluster Analysis</i>	“	13
3. L’analisi delle componenti principali “per blocchi” e “in blocco”	pag.	15
4. Un confronto empirico tra l’approccio “per blocchi” e “in blocco”	pag.	17
4.1 <i>Analisi in blocco</i>	“	17
4.2 <i>Analisi per blocchi</i>	“	19
4.3 <i>Un confronto</i>	“	20
4.4 <i>Un approccio “a stadi”</i>	“	24
5. Conclusioni	“	25
Appendice	“	27
<i>App. A – Elenco degli indicatori utilizzati</i>	“	29
<i>App. B – Interpretazione delle componenti</i>	“	30
<i>App. C – Le zonizzazioni secondo i diversi approcci</i>	“	32
<i>App. D – Descrizione cluster secondo gli indicatori</i>	“	35
Riferimenti bibliografici	“	41

Finito di stampare nel mese di Aprile 1999
presso le Officine Grafiche Tecnoprint
Via del Legatore 3, Bologna

I METODI STATISTICI PER L'ANALISI DEI SISTEMI AGRICOLI TERRITORIALI¹

PREFAZIONE

Lo studio dello sviluppo e delle trasformazioni dell'agricoltura è stato spesso accompagnato da una approfondita analisi territoriale a livello regionale e sub regionale, con l'individuazione di aree e zone caratterizzate da una sostanziale omogeneità al loro interno, ma che mettersero in risalto le differenziazioni esistenti fra le diverse aree.

Un precedente di queste analisi territoriali si può ritrovare nella definizione delle "zone agrarie", utilizzate dall'ISTAT per effettuare le stime della produzione agricola, basata su metodi estimativi. La definizione di queste zone era fondata essenzialmente sulla omogeneità delle rese dei cereali e del grano in particolare. Le "zone agrarie" avevano comunque dei difetti legati al fatto che distinguevano sostanzialmente le differenze tra pianura, collina e montagna, ma utilizzando i confini amministrativi includevano interi comuni in una zona o nell'altra.

Lo sviluppo delle analisi territoriali in agricoltura ha avuto senza dubbio una accelerazione in questo secondo dopoguerra, ad opera soprattutto di Rossi-Doria, che ha cercato di superare la visione dualistica fra Nord e Sud, partendo dalle prime analisi sulle differenziazioni interne del Mezzogiorno, fra le cosiddette "polpa" ed "osso". L'analisi di Rossi-Doria si è conclusa nel 1969 con la pubblicazione da parte dell'INEA della "Zonizzazione dell'agricoltura italiana". In questo lavoro l'individuazione delle principali zone agricole era fatta tenendo presente quattro variabili principali (riparto della superficie agraria e forestale, addetti agricoli, valore della PLV agricola e forestale, valore prodotto netto agricoltura), ma era sostanzialmente integrata dal giudizio degli esperti individuati a livello regionale. I lavori sono stati eseguiti a livello comunale, senza tenere conto dei limiti amministrativi. Successivamente le

¹ La responsabilità della ricerca è comune ai due autori. Roberto Fanfani ha curato la prefazione, Mario Mazzocchi ha curato la stesura dei paragrafi 2, 2.1, 2.2, 3, 4.1, 4.2, 4.3, 4.4 e 5. Il paragrafo 1 è frutto di una riflessione congiunta. Gli autori sono riconoscenti alla prof.ssa Elisa Montresor per le preziose indicazioni sulle tematiche dello sviluppo agricolo territoriale e al prof. Sergio Brasini per le utili osservazioni sui metodi statistici. Si ringraziano inoltre la dottoressa Alessandra Zanchini per la collaborazione alla stesura del paragrafo 2.2 e la dottoressa Cristina Brasili per le proficue e interessanti discussioni sull'argomento trattato.

analisi territoriali sono state approfondite con l'utilizzazione di metodologie statistiche multivariate, introdotte per giungere alle zonizzazioni dell'agricoltura italiana nell'ambito dei progetti CNR-Raisa (Cannata, 1989 e 1998).

La crescente importanza assunta dalle zonizzazioni è dovuta soprattutto all'attuazione delle politiche comunitarie. La prima direttiva sulle zone svantaggiate (268/75) definiva le aree alle quali venivano destinati gli incentivi della Politica Agricola Comunitaria e assegnate delle speciali "indennità compensative". Le politiche strutturali della Comunità si sono poi sviluppate alla fine degli anni '80, con la riforma dei fondi strutturali ed il ricorso ad un insieme limitato di indicatori per la definizione delle regioni interessate dall'obiettivo 1 (un solo criterio, un PIL pro capite regionale inferiore al 75% della media comunitaria) e dall'obiettivo 5b applicato a livello sub-regionale (PIL pro capite, importanza dell'occupazione e del reddito agricolo, dinamiche della popolazione). Con i regolamenti relativi ad Agenda 2000 gli interventi sulle aree rurali ricadono nel nuovo obiettivo 2, che utilizza criteri relativi alla densità di popolazione, al tasso di occupazione agricola, al tasso medio di disoccupazione e alle dinamiche demografiche.

1. L'analisi dello sviluppo agricolo territoriale e i metodi statistici

L'analisi dei sistemi territoriali a livello sub-regionale ha assunto una rilevanza crescente negli ultimi anni, specialmente come sostegno alle politiche strutturali dell'Unione Europea. Sta crescendo fortemente la domanda politica per strumenti interpretativi e - soprattutto - di metodi di valutazione dell'impatto dei piani di sviluppo e delle azioni a favore di determinate aree territoriali. Questo forte legame con esigenze immediate di carattere applicativo e di politica economica ha, però, relegato spesso in secondo piano l'approfondimento delle potenzialità interpretative degli strumenti metodologici tradizionalmente forniti ed usati dalla statistica. Il divario tra il grado di complessità raggiunto dalle analisi accademiche sulla fragilità economica e le esigenze dei "policy-makers" è evidente nell'essenzialità dei parametri indicati dalla Commissione Europea per l'individuazione delle aree di svantaggio rurale, sia quelle dell'obiettivo 5b dei fondi strutturali che quelle dell'obiettivo 2, come descritto nelle proposte di regolamento di Agenda 2000. Il "gap" si sta però riducendo progressivamente e il ricorso al partenariato fra Unione Europea e Regioni ha dato nuovi impulsi alla ricerca di tecniche affidabili per l'analisi dei

sistemi territoriali agricoli. Diventa però sempre più importante la leggibilità dei risultati e la replicabilità dei metodi nel tempo e nello spazio, mentre l'approfondimento metodologico potrebbe permettere il raggiungimento di risultati più dettagliati e con un contenuto informativo rispetto alle esigenze emerse.

Se da un lato la Commissione Europea e gli attori politici in genere hanno privilegiato le selezioni delle aree territoriali su cui intervenire su un numero limitato e immediato di indicatori, dall'altra le tecniche più sofisticate hanno il vantaggio di ridurre l'influenza di dati statistici non sempre affidabili. Copus e Crabtree (1992) individuano una serie di ragioni per cui l'analisi statistica è poco valorizzata nelle zonizzazioni "politiche" dell'agricoltura: scetticismo sulla rappresentatività dei dati, scarsa padronanza delle tecniche di statistica multivariata, desiderio di semplicità amministrativa e - soprattutto - flessibilità per bilanciarsi rispetto alle posizioni dei vari gruppi di interesse.

La necessità di una lettura dettagliata del territorio mediante tecniche di statistica multivariata è comunque evidente nella vasta letteratura sul tema. Il presente lavoro fa essenzialmente riferimento ai lavori curati da Cannata (1989), Boccafogli e Brasili (1998)², Cannata e Forleo (1998), che rappresentano i lavori più recenti sull'argomento³. La metodologia statistica utilizzata in queste analisi consiste nell'utilizzo congiunto di analisi delle componenti principali e cluster analysis. Partendo da un ampio insieme di indicatori che si assume completo per descrivere il sistema agricolo territoriale, l'analisi delle componenti principali (ACP) permette di sintetizzare tale insieme in un numero ridotto di componenti, che sia comunque sufficiente alla descrizione degli stessi sistemi. Tali componenti vengono poi utilizzate nella cluster analysis per identificare le aree agricole sul territorio, giungendo così alla zonizzazione.

Tale approccio metodologico è ormai consolidato nell'analisi dei sistemi territoriali in agricoltura, ma all'interno del disegno generale permangono comunque alcune varianti la cui rilevanza è ancora oggetto di ricerca. In particolare può variare il ruolo assunto dall'analisi delle com-

² In particolare il lavoro di Boccafogli e Brasili utilizza la tecnica della "fuzzy clustering", che permette di valutare il grado di appartenenza di ciascuna unità statistica al cluster al quale è stato assegnato e di controllare il livello di sfumatura della classificazione nei cluster.

³ Una zonizzazione nazionale dell'agricoltura a livello comunale è stata inoltre prodotta dal CAIRE (1995), con un approccio intermedio: partendo da un ampio insieme di indicatori, i comuni vengono distinti in aree marginali e non utilizzando dei valori soglia.

ponenti principali, che oltre ad essere un utile strumento preliminare alla zonizzazione vera e propria, può diventare in se stessa un'efficace metodologia per l'interpretazione del territorio e per una valutazione "gerarchica" del ruolo assunto dai singoli indicatori. In tal senso l'ACP può essere una tecnica appropriata per la "traduzione" dei risultati di analisi complesse in strumenti semplici e di facile lettura.

Può quindi essere interessante valutare le diverse modalità di applicazione dell'ACP, in particolare verificare i risultati conseguibili con le versioni "in blocco" o "per blocchi". L'analisi in blocco, utilizzata da Cannata (1989) nella zonizzazione dei sistemi agricoli territoriali italiani negli anni '80 (SATI80), prevede l'inserimento in un passaggio unico di tutti gli indicatori considerati. Nei lavori più recenti si è, invece, privilegiato l'approccio per blocchi (SATI90, Cannata, 1998), nel quale l'ACP si svolge in due passaggi: nel primo gli indicatori vengono suddivisi in gruppi (blocchi) omogenei secondo un dato schema di classificazione (ad esempio blocchi di variabili legate allo sviluppo economico generale, strutturale, demografico, ecc.), nel secondo viene applicata l'ACP all'interno dei singoli blocchi. Le differenze tra i due approcci sono state presentate da Anania e Tarsitano (1995) e sono riprese da Forleo (1998). I due metodi non mostrano differenze sostanziali se le correlazioni tra indicatori appartenenti a blocchi diversi sono trascurabili, ma in generale, come sottolinea sempre Forleo (1998) non sono state ancora identificate le situazioni in cui un approccio risulta superiore all'altro. In particolare, la situazione in cui indicatori appartenenti a blocchi diversi sono uniti da legami deboli, è difficilmente generalizzabile. Nell'adottare il metodo per blocchi si richiede dunque estrema cura da parte del ricercatore nel verificare che tale condizione sia effettivamente realizzata.

Il presente lavoro si propone di dare una valutazione delle "performance" dei due metodi, utilizzando come base di partenza i dati comunali dell'Emilia-Romagna. Si cercherà inoltre di evidenziare il ruolo dell'ACP come strumento interpretativo capace di offrire una lettura socioeconomica alle componenti stesse e valutare, complessivamente, quali tra gli indicatori originali siano più utili a spiegare le differenze tra i sistemi agricoli territoriali.

Da un punto di vista più propriamente statistico l'obiettivo è anche quello di quantificare le differenze secondo le diverse varianti metodologiche, per cercare di spiegare in quali situazioni un metodo sia preferibile all'altro. Resta invece da indagare la rilevanza della scelta iniziale degli

indicatori specifici, ma tale obiettivo esula dagli scopi del presente lavoro e viene affrontato solo marginalmente.

La rappresentatività ed importanza dei vari indicatori utilizzati in questo ambito di ricerca e l'informazione in essi contenuta è discussa in maniera più approfondita nei lavori sopra citati, mentre contributi significativi sono stati prodotti da OECD (1994 e 1996), FAO (1990), Montresor (1998), Fanfani e Montresor (1998).

Nel presente lavoro gli aspetti metodologici generali vengono trattati nel paragrafo 2, con particolare attenzione all'interpretazione delle componenti (paragrafo 2.1) e all'applicazione della cluster analysis (par. 2.2). Nel paragrafo 3 viene introdotta la distinzione tra analisi delle componenti principali "per blocchi" e "in blocco". Le implicazioni delle differenze metodologiche tra i due approcci vengono discusse con l'ausilio di un'applicazione alla zonizzazione dell'agricoltura con dati comunali dell'Emilia-Romagna nel paragrafo 4. Nello stesso paragrafo viene inoltre introdotta la possibilità di un approccio "a stadi".

2. Alcuni aspetti metodologici generali sulle tecniche utilizzate per le zonizzazioni

La formulazione teorica dell'analisi delle componenti principali più nota è quella data da Hotelling (1933), mentre l'indirizzo metodologico si fa risalire già a Pearson (1901)⁴. Rimandando a Krzanowski (1988) per la formalizzazione analitica, in sintesi l'ACP è una tecnica statistica multivariata che permette di trasformare un insieme di p indicatori rilevato su n unità statistiche in un insieme più ridotto di variabili che siano comunque in grado di spiegare una proporzione significativa della variabilità presente nei dati originali⁵.

Le variabili ottenute e selezionate o "componenti principali" sono una combinazione lineare degli indicatori originali e, contrariamente ad essi,

⁴ Pearson, K. (1901), "On lines and planes of closest fit to systems of points in space", *Philosophical Magazine*, 2, pp. 559-571

Hotelling, H. (1933), "Analysis of a complex of statistical variables into principal components", *Journal of Educational Psychology*, 24, pp. 417-441

⁵ Come è noto tale proporzione dipende dal metodo utilizzato nella scelta del numero di componenti principali da considerare. In questa analisi si è adottato il criterio di Guttman-Kaiser, secondo il quale si trattengono le componenti i cui autovalori siano superiori o uguali ad 1. In generale tale criterio ha portato alla selezione di un numero di componenti capace di spiegare oltre il 70% della variabilità dei dati originali.

non sono correlate tra loro. Come è noto l'ACP può essere eseguita sia sulla matrice di varianze e covarianze che su quella di correlazione derivate dalla matrice dei dati. I due metodi non portano agli stessi risultati e generalmente si applica il secondo, che non risente degli effetti di scala dovuti alle diverse unità di misura degli indicatori. In pratica si ricorre ad una standardizzazione degli indicatori originali. Tale passaggio si rende necessario per evitare che indicatori con maggiore variabilità esercitino un'influenza distorcente sull'estrazione delle componenti principali. E' comunque essenziale verificare il metodo adottato dai vari software quando non sia possibile – come spesso accade – effettuare una scelta a priori. Per la presente analisi è stato utilizzato il software SAS, con la realizzazione di programmi che permettono, partendo dalla matrice degli indicatori, di giungere alla zonizzazione finale secondo i diversi approcci metodologici qui descritti. Tale aspetto non è secondario nel valutare la potenziale applicabilità e ripetibilità della metodologia, che si riduce all'utilizzo di una singola applicazione informatica, anche in ambiti non strettamente scientifici.

L'applicazione dell'ACP sulla matrice di correlazione permette il passaggio dagli indicatori originali alle componenti principali. Il passaggio immediatamente successivo è quello di calcolare i punteggi assunti dalle singole unità statistiche (in questo caso i comuni) nelle singole componenti. Nel caso si operi sulla matrice di correlazione, le k componenti principali (con $k < p$) sono date dalle seguenti combinazioni lineari, in termini matriciali

$$Y = \hat{X}A \quad (1)$$

dove Y è la matrice $n \times k$ contenente i punteggi assunti dalle n unità statistiche nelle k componenti
 A è la matrice vettore $p \times k$ dei coefficienti normalizzati
 \hat{X} è la matrice $n \times p$ dei dati standardizzati

I punteggi della matrice Y sono i valori che verranno poi utilizzati per la cluster analysis. La varianza dei punteggi di ogni singola componente principale è uguale all'autovalore della componente stessa. In alcuni

software i punteggi vengono automaticamente standardizzati⁶. L'utilizzo di punteggi standardizzati nella cluster analysis presuppone che tutte le componenti abbiano la stessa varianza e conseguentemente lo stesso peso nella zonizzazione e ciò può portare a serie distorsioni. Il problema della standardizzazione dei punteggi assume invece un significato diverso quando le componenti provengono da più blocchi di indicatori, come avviene nell'analisi "per blocchi" e come viene discusso più in dettaglio nel paragrafo 3.

2.1 L'interpretazione delle componenti principali

Anche prima di effettuare la cluster analysis per ottenere la zonizzazione finale, l'interpretazione economica delle componenti e il confronto tra i punteggi assunti dai singoli comuni può fornire informazioni utili alla lettura dello sviluppo agricolo territoriale. L'interpretazione delle componenti deriva dall'osservazione del segno e della rilevanza dei coefficienti contenuti nella matrice A . Ad esempio, poniamo che il punteggio del comune i -esimo nella componente j -esima sia dato dalla seguente combinazione lineare

$$y_{i,j} = a_{1,j} \hat{x}_{i,1} + a_{2,j} \hat{x}_{i,2} + a_{3,j} \hat{x}_{i,3} + a_{4,j} \hat{x}_{i,4} \quad (2)$$

assumendo per semplicità che l'insieme originale degli indicatori sia costituito da soli 4 elementi, ad esempio:

$\hat{x}_{i,1}$ è il Pil pro-capite (standardizzato, del comune i -esimo)

$\hat{x}_{i,2}$ è la proporzione di addetti in agricoltura

$\hat{x}_{i,3}$ è la densità di popolazione

$\hat{x}_{i,4}$ è la proporzione di addetti nell'industria

Assumendo che la prima componente risulti del tipo:

⁶ Tale situazione è particolarmente frequente nei software che utilizzano una routine valida anche per l'analisi fattoriale, nella quale l'analisi delle componenti principali viene considerata come caso particolare.

$$y_{i,1} = 0,7 \cdot \hat{x}_{i,1} - 0,4 \cdot \hat{x}_{i,2} + 0,3 \cdot \hat{x}_{i,3} + 0,5 \cdot \hat{x}_{i,4} \quad (3)$$

Evidentemente un comune con un elevato PIL pro-capite, densamente popolato e con una proporzione preponderante di addetti nell'industria assumerà un punteggio alto in tale componente, mentre i comuni più rurali figureranno in fondo alla graduatoria dei punteggi della stessa componente. Pertanto la prima componente descritta rappresenterebbe una sorta di indice di "urbanità-ruralità". Inoltre, poiché le variabili $\hat{x}_{i,j}$ sono standardizzate e hanno quindi la stessa variabilità e lo stesso ordine di grandezza, il peso delle singole variabili sul valore delle componenti è legato direttamente al valore assoluto del coefficiente. Se si considerano i coefficienti elevati al quadrato, essendo la loro somma vincolata a 1 per costruzione, si otterrà la proporzione in cui ciascuna variabile pesa sul valore della componente.

Come si può facilmente intuire, l'interpretazione delle componenti è raramente immediata come nel caso descritto a titolo esemplificativo. I problemi sono molteplici, il numero degli indicatori è generalmente molto più ampio (29 nell'analisi in blocco SATI80, 38 nell'analisi in blocco effettuata in questo lavoro) e spesso i segni dei coefficienti forniscono indicazioni contraddittorie. Inoltre l'interpretazione è raramente univoca e tende a diventare più difficoltosa quando ci si appresta a spiegare le ultime componenti, quelle con una proporzione minore di variabilità spiegata. Un metodo utile per semplificare l'analisi e sottrarsi alle tentazioni di avventurarsi in interpretazioni soggettive è quello di "filtrare" le variabili proprio attraverso il peso che ogni singolo coefficiente esercita sul valore della componente. Non esistono criteri standard per tale operazione, che lascia comunque uno spazio residuo all'arbitrarietà del ricercatore. In questo lavoro si è adottata la regola può essere di considerare gli indicatori i cui coefficienti elevati al quadrato maggiori o uguali a 0,05, cioè quelli il cui apporto al punteggio della componente fosse almeno pari al 5%, definendo tale proporzione come percentuale:

$$s_{l,j} = a_{l,j}^2 \cdot 100 \quad (4)$$

dove $a_{l,j}$ (con $l=1, \dots, p$) è il coefficiente del l -esimo indicatore nella componente j .

Lo stratagemma consente di dare una prima rapida interpretazione alle componenti, che può poi essere sottoposta a verifica o revisione guardando eventualmente il segno e la significatività di altri indicatori. In generale gli $s_{l,j}$ forniscono un'indicazione sull'intensità dell'azione di un indicatore sulla componente, mentre la direzione deve essere interpretata guardando i segni dei coefficienti $a_{l,j}$.

Un'altra indicazione che l'ACP può fornire, riguarda l'apporto complessivo dato dai singoli indicatori alle componenti principali selezionate. In pratica se gli $s_{l,j}$ forniscono una misura del peso di ogni singolo indicatore all'interno di una componente, la semplice media \bar{s}_j restituisce una misura, seppure naïf, dell'importanza dell'indicatore sul complesso delle componenti principali. Tale media presuppone, però, che tutte le componenti pesino allo stesso modo, mentre sembra più appropriata una media ponderata alla proporzione di variabilità spiegata da ciascun autovalore:

$$w_j = \sum_{j=1}^k s_{l,j}^2 \cdot \frac{Var(Y_j)}{Var(Y)} \quad (5)$$

dove $Var(Y_j)$ è la varianza della componente j -esima (uguale all'autovalore j -esimo)

$Var(Y)$ è la varianza complessiva delle k componenti selezionate (uguale alla somma dei primi k autovalori)

Secondo tale misura si possono individuare gli indicatori più rappresentativi nella descrizione del fenomeno e raccogliere informazioni utili, nel caso l'obiettivo sia una riduzione dell'insieme originale di indicatori.

2.2 La Cluster Analysis

Una volta individuate le componenti principali si procede all'aggregazione vera e propria dei comuni mediante cluster analysis. Anche per quanto riguarda la cluster analysis esistono però più varianti, che possono portare a risultati sostanzialmente diversi (cfr. Aldenderfer e Blashfield (1984)), per cui è importante identificare a priori la metodologia più adatta all'obiettivo. Un problema fondamentale, ancora non risolto univocamente, è quello dell'individuazione del numero ottimale dei gruppi.

Come sottolinea Fabbri (1997), il criterio di aggregazione non può prescindere dall'informatività del risultato rispetto agli obiettivi originali dell'analisi: qualsiasi tecnica giudicata statisticamente valida, deve sempre essere confermata da un risultato soddisfacente anche in termini interpretativi.

La prima distinzione è quella tra metodi gerarchici e non gerarchici, che si contrappongono per il fatto che una scelta fatta a un dato livello dell'algoritmo possa o meno essere rivista al passaggio successivo. I metodi gerarchici, che non permettono di rivedere la classificazione, hanno il vantaggio di individuare il numero di gruppi in modo non soggettivo, ma mediante l'uso di statistiche ad hoc. Nei metodi non gerarchici, invece, l'assegnazione del numero di cluster deve essere decisa a priori dal ricercatore, introducendo in questo modo un inevitabile elemento di soggettività.

E' possibile tentare di compensare la rigidità del risultato derivante da tecniche gerarchiche, con la flessibilità di un metodo non gerarchico. A questo proposito si può operare inizialmente con una tecnica gerarchica, che permetta di individuare in modo più oggettivo il numero di cluster, e successivamente applicare il metodo non gerarchico (utilizzando il numero di cluster suggerito dalla precedente applicazione) per ottenere la composizione di ciascun gruppo. Anche le tecniche gerarchiche sono comunque molteplici e conducono a indicazioni divergenti. Per quanto riguarda le tecniche di zonizzazione, i metodi gerarchici più robusti per individuare il numero di cluster sono quelli del legame medio e l'algoritmo di Ward, che presentano caratteristiche più desiderabili rispetto agli altri⁷. Nel caso presentato in questo articolo l'algoritmo di Ward non forniva indicazioni univoche sul numero di cluster, per cui la scelta è caduta sul metodo del legame medio.

Per individuare effettivamente il numero di cluster si utilizza generalmente il valore del rapporto tra la varianza tra i gruppi e quella entro i gruppi, detto pseudo F, per tutte le potenziali partizioni. Il numero di cluster corrispondente ad un valore di ottimo locale della pseudo F, che soddisfi anche esigenze di informatività, è quello da utilizzare per l'applicazione della tecnica non gerarchica. Un altro criterio utilizzato per stabi-

lire il numero ottimale di gruppi è quello del Cubic Clustering Criterion. Una descrizione delle due statistiche è riportata in Brasini et al. (1993).

Il metodo non gerarchico utilizzato per la formazione dei gruppi è quello delle k-medie (o di McQueen). Date n unità e determinato il numero k di gruppi, viene effettuata una prima partizione e poi si procede assegnando ciascuna delle rimanenti $n-k$ unità al gruppo al cui centroide sono più vicine. Dopo ogni assegnazione, il centroide del nuovo cluster che si è venuto a formare viene ricalcolato.

Riassumendo le considerazioni sopra riportate, l'approccio individuato per effettuare la zonizzazione vera e propria mediante cluster analysis è il seguente:

1. Utilizzare preliminarmente un approccio gerarchico, precisamente quello del legame medio o di Ward, per determinare il numero di cluster. Per la scelta del numero ottimale di cluster si osservano le statistiche pseudo-F e il Cubic Clustering Criterion
2. Una volta stabilito il numero di gruppi omogenei (gradi di sviluppo agricolo territoriale) si effettua la zonizzazione mediante il metodo non gerarchico aggregativo delle k-medie
3. Seguendo il principio di informatività si può scegliere di aggregare le zone con un numero ridottissimo di elementi (meno di 3) al cluster più vicino

3. L'analisi delle componenti principali "per blocchi" e "in blocco"

L'analisi "per blocchi" presuppone una suddivisione preliminare degli indicatori in blocchi o macrodeterminanti. Tale tecnica ha il vantaggio di effettuare una serie di ACP su un numero limitato di indicatori della stessa specie, facilitando l'interpretazione delle componenti, ma comporta una serie di scelte che aumentano il grado di arbitrio del ricercatore. Uno dei vantaggi nell'utilizzare tale approccio potrebbe essere quello di evitare che alcuni indicatori, considerati importanti dal ricercatore, ma correlati ad altre variabili, siano "nascosti" nella fase interpretativa. Si pensi ad esempio ad una variabile legata all'inquinamento e all'ambiente, che con ogni probabilità sarà fortemente legata al grado di industrializzazione. Con la scelta di agire per blocchi si presuppone che ogni macrodeterminante debba essere rappresentata nella cluster analysis. La conseguenza è quella di amplificare l'effetto di alcune variabili, tra loro correlate, ma appartenenti a blocchi diversi.

⁷ Il metodo del "legame singolo" è stato escluso per i notevoli problemi legati all'effetto a catena, i metodi del "centroide" e del legame "completo" per la sensibilità alla presenza di outliers, che potrebbe pregiudicare l'esito dell'intera indagine. I metodi che appaiono più robusti sono quello del "legame medio" e l'algoritmo di Ward.

Come già ricordato, studi metodologici precedenti hanno mostrato l'equivalenza tra i due metodi di analisi quando le correlazioni tra indicatori appartenenti a blocchi diversi sono trascurabili. In tale caso viene però a mancare il vantaggio di salvaguardare la rappresentatività di certi indicatori, in quanto essi non rischiano di essere annullati da forti correlazioni con altri.

Inoltre, ma tale osservazione è riproponibile in termini diversi per l'analisi in blocco, diventa determinante la ripartizione degli indicatori tra i blocchi: il peso di ciascuna macrodeterminante sarà proporzionale al numero di indicatori in essa classificati. Un aspetto essenziale è il modo in cui le componenti estratte da blocchi diversi peseranno nella cluster analysis. Se non si opera alcuna standardizzazione sui punteggi, ciascuna componente peserà in base alla propria varianza. Tale varianza è, come si è visto, uguale all'autovalore della componente, che a sua volta è fortemente legato al numero di indicatori contenuti nel blocco di indicatori⁸. In questo caso l'approccio per blocchi introdurrebbe dunque una distorsione, dando più rilevanza alle componenti estratte dai blocchi con il maggior numero di indicatori, che già sono più numerose. Se, invece, si opera la standardizzazione dei punteggi delle singole componenti estratte, tutte le componenti avranno lo stesso peso e quindi ogni macrodeterminante influirà sulla cluster analysis in maniera direttamente proporzionale al numero di componenti da essa estratti. Quest'ultima scelta sembra più coerente con quella iniziale di operare sulla matrice di correlazione, che presuppone proprio che tutti gli indicatori abbiano la stessa importanza. Resta la considerazione che, per qualunque tipo di analisi, la scelta degli indicatori, non solo in termini qualitativi, ma anche come numerosità rispetto all'aspetto considerato, riveste un ruolo essenziale.

Nell'analisi "per blocchi" rimane perciò un certo grado di arbitrarietà nella suddivisione degli indicatori e non esiste la possibilità di una interpretazione "trasversale" delle componenti, cioè non è possibile considerare congiuntamente i coefficienti di indicatori appartenenti a macrodeterminanti diverse. Ciò è invece possibile utilizzando l'analisi "in blocco", per cui ogni componente è una combinazione lineare di tutti gli indicatori originali. Le difficoltà di interpretazione dovute all'elevato numero di coefficienti possono essere ridotte utilizzando il metodo descritto nel paragrafo precedente. L'efficacia dei due tipi di analisi non è

stata oggetto di particolari studi, per cui può essere utile applicare entrambe e cercare di verificarne le differenze, come nel caso qui riportato dei 38 indicatori dell'Emilia-Romagna.

4. Un confronto empirico tra l'approccio "per blocchi" e "in blocco"

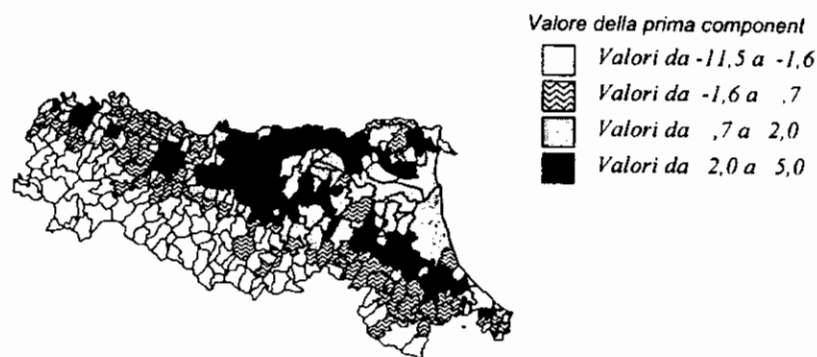
4.1 Analisi in blocco

Inserendo i 38 indicatori elencati nell'appendice A e derivati da un'applicazione di Montresor, Mazzocchi (1999) si ottengono 11 componenti principali con autovalore superiore a 1, che complessivamente spiegano circa il 77% della variabilità totale degli indicatori. Utilizzando le statistiche s_{ij} descritte nella formula (4) e fissando una soglia del 5%, sono stati individuati gli *indicatori caratterizzanti* ed è stato inquadrato un primo profilo. Per approfondire la descrizione della componente sono poi stati analizzati indicatori meno significativi in termini di varianza spiegata, ma utili a confermare o rivedere la descrizione del profilo. L'interpretazione delle 11 componenti è riportata nell'appendice B, dove viene indicata molto sinteticamente una chiave di lettura della componente. In alcuni casi si tratta di un "contrasto", cioè non c'è un'interpretazione univoca, quanto una forte differenza tra le caratteristiche dei comuni a punteggi alti e quelli a punteggi bassi secondo determinate variabili. Infine, per certe componenti, la lettura avviene convenzionalmente in senso inverso, cioè le aree che rispondono meglio alla lettura data sono quelle con i punteggi più bassi. Tali componenti sono indicate da un segno (-).

Come è evidente, l'interpretazione di alcune componenti non è particolarmente felice in termini informativi. Per altre, in particolare per le prime, l'interpretazione è più immediata. La prima componente, indicata come "livello di sviluppo agricolo", è caratterizzata da indicatori legati alla produttività del terreno e del lavoro (RLS/SAU, RLS/ULA), al livello di anzianità della popolazione (indici di vecchiaia e di dipendenza), al tasso di attività femminile e ad utilizzi delle superfici agricole tipici delle aree montane (% superficie a boschive, % SAU a prati e pascoli). Nella seconda componente risultano rilevanti tutti gli indicatori relativi alla dimensione delle aziende agricole più un indicatore legato all'intensità di lavoro (ULA/SAU).

⁸ Operando sulla matrice di correlazione, infatti, la somma degli autovalori delle componenti è uguale al numero di indicatori.

Fig. 4.1 – Distribuzione dei punteggi per comune della prima componente principale



Tav. 4.1 – Classificazione degli indicatori secondo la loro rappresentatività nell'ACP in blocco

Variabile	w_i	Variabile	w_i
1. RLSSAU	3,22	20. AZ50HA	2,60
2. SAUMED	3,03	21. VARAZ	2,59
3. INDALIME	2,99	22. PRATI	2,55
4. DIPEND	2,93	23. LAUREA	2,50
5. ULASAU	2,92	24. VITE	2,49
6. ADD100	2,91	25. ORTIVE	2,48
7. ADDIND	2,89	26. VARSUP	2,47
8. RLSULA	2,89	27. DENPOP	2,38
9. S2HA	2,85	28. BOVPRATI	2,37
10. BOVINI	2,84	29. FORAGGI	2,35
11. ADDAGR	2,80	30. DOCVITE	2,27
12. FRUTTI	2,75	31. DISOCC	2,20
13. CEREALI	2,74	32. DINPOP	2,13
14. BOSCHIVE	2,71	33. AZALLEV	2,12
15. AZ2HA	2,69	34. SUINI	2,12
16. S50HA	2,69	35. PIL	2,10
17. ADDTERZ	2,65	36. AVICOLI	1,99
18. ATTFEMM	2,64	37. VARSAU	1,90
19. VECCHIAI	2,61	38. TRATTSAU	1,89

Un'ulteriore verifica per l'interpretazione data può derivare dalla graduatoria dei punteggi assunti dai comuni in ogni singola componente. La prima componente, che nei punteggi alti dovrebbe essere caratterizzata da aree a forte produttività agricola, è rappresentata nella figura 4.1.

Considerando tutte e 11 le componenti estratte mediante ACP in blocco, i 341 comuni dell'Emilia-Romagna analizzati sono stati raggruppati in 8 cluster secondo l'approccio indicato nel paragrafo 2.2. La rappresentazione grafica degli otto cluster è riportata nell'appendice C.1, mentre i valori medi degli indicatori originali all'interno dei vari cluster elencati nell'appendice D.1. Inoltre, secondo il valore assunto dagli indicatori w_i descritti nella formula (5), l'importanza dei singoli indicatori nella costruzione delle componenti principali è riportata nella tavola 4.1.

4.2 Analisi per blocchi

Partendo dagli stessi indicatori e dalla stessa base dati, l'analisi delle componenti principali può essere svolta all'interno di gruppi di indicatori considerati omogenei, come nei lavori di Cannata (1998) e Boccafogli, Brasili (1998).

La suddivisione dei 38 indicatori in 5 macrodeterminanti è riportata nella tavola 4.2. La classificazione in blocchi è aperta a interpretazioni alternative e a suddivisioni diverse possono corrispondere zonizzazioni diverse, se le correlazioni tra indicatori appartenenti a blocchi diversi non sono trascurabili. In questo caso le componenti con autovalore maggiore di 1 estratte dai 38 indicatori sono complessivamente 14. Come discusso nel paragrafo 3, i punteggi delle componenti devono essere standardizzati, per evitare distorsioni. La cluster analysis gerarchica ha indicato un numero ottimale di cluster pari a 6. La mappa dei comuni dell'Emilia-Romagna aggregati secondo la zonizzazione per blocchi è riportata nell'appendice C.2, le medie degli indicatori nell'appendice D.2.

Una prima importante differenza tra questa analisi e quella "in blocco" si può sottolineare. Nell'analisi per blocchi (app. C.2) i comuni capoluogo di provincia appartengono tutti allo stesso gruppo, mentre in quella in blocco (app. C.1), Rimini e Forlì sono classificate in due gruppi a parte. Ciò potrebbe essere dovuto al fatto che nella suddivisione in blocchi, le variabili demografico-sociali (blocco 2) sono correlate all'indicatore del Pil pro-capite e alla percentuale di addetti in agricoltura (blocco 1), come appare relativamente ovvio. Essendo le componenti estratte separatamente, tale correlazione viene ignorata e il Pil pro-capite e la percentuale

di addetti in agricoltura sono in pratica sovrappesate nell'analisi. Di conseguenza i comuni urbani, che presentano un elevato Pil pro-capite ed un bassa percentuale di addetti in agricoltura, oltre che valori simili in alcune variabili demografiche (su tutte la densità di popolazione) tendono a venire omologati nell'analisi per blocchi. Volendo sottolineare questo potenziale problema, si è effettuata un'ulteriore analisi per blocchi, aumentandone il numero da 5 a 7 e separando gli indicatori contenuti nei blocchi 1 e 3, in modo da creare un effetto ancora maggiore in termini di correlazione indesiderata.

In particolare, per collegarsi all'esempio precedente, si sono isolati gli indicatori relativi all'industria e al settore agroalimentare, fortemente correlati agli altri indicatori socio-economici e quelli relativi alla produttività dell'agricoltura, correlati alla dimensione aziendale. In tal modo tali indicatori avranno un peso più alto nella zonizzazione.

In effetti si giunge ad un risultato molto differente. L'aumento del numero di blocchi ha portato a 15 il numero delle componenti principali estratte. La zonizzazione ottimale ha condotto all'identificazione di 8 cluster, come riportato nella figura in appendice C.3. Questa volta i comuni capoluogo di provincia tornano a separarsi, probabilmente a causa dell'aumentata rilevanza delle variabili legate all'industria e alla struttura agricola.

Le differenze rispetto all'analisi per blocchi precedente sono presumibilmente dovute all'aumento delle correlazioni tra i blocchi. Come si può vedere dalla tavola 4.3, le correlazioni medie tra i blocchi superiori a 0,2 sono diventate 8.

4.3 Un confronto

Un confronto oggettivo tra i due metodi non è semplice e, come già visto, le prestazioni dipendono strettamente dall'impostazione e dagli obiettivi della ricerca. I due esempi riportati di analisi per blocchi mostrano la sensibilità di questa analisi al problema delle correlazioni tra indicatori appartenenti a blocchi diversi.

Se si considera la capacità di differenziare i comuni secondo le variabili iniziali, un giudizio sul risultato può essere effettuato mediante una serie di test sulle medie. Si è sottoposta a verifica l'ipotesi di uguaglianza delle medie degli indicatori per tutte le coppie di gruppi con un semplice test *t* e si è calcolata la percentuale di test che risultano significativi.

Tav. 4.2 - Analisi per blocchi con 5 macrodeterminanti: indicatori, correlazioni medie e numero di componenti estratte

Blocco	Gruppo	Indicatori	1	2	3	4	5	N. comp.
1	Sviluppo economico	PIL, ADDAGR, ADDIND, ADDTER, DISOCC, INDALIME, ADD100	0,39	0,26	0,13	0,12	0,15	3
2	Demografico-sociali	DENPOP, VECCHIAI, ATTFEMM, DIPEND, LAUREA	0,26	0,48	0,16	0,25	0,20	2
3	Struttura agricola	SAUMED, AZ2HA, AZ50HA, S2HA, S50HA, TRATTSAU, ULASAU, RLSSAU, RLSULA	0,13	0,16	0,44	0,10	0,19	3
4	Dinamiche	VARPOP, VARS AU, VARAZ, VARSUP	0,12	0,25	0,10	0,42	0,14	1
5	Specializzazione agricola	BOSCHIVE, CEREALI, FORAGGI, PRATI, ORTIVE, FRUTTI, VITE, VITEDOC, AZALLEV, BOVINI, SUINI, AVICOLI, BOVPRATI	0,15	0,20	0,19	0,14	0,28	5

Tav. 4.3 - Analisi per blocchi con 7 macrodeterminanti: indicatori, correlazioni medie e numero di componenti estratte

Blocco	Gruppo	Indicatori	1	2	3	4	5	6	7	N. comp.
1	Sviluppo mico	econo- PIL, ADDAGR, ADDTER, DISOCC	0,43	0,29	0,11	0,13	0,15	0,28	0,13	2
2	Demografico- sociali	DENPOP, VECCHIAI, ATTFFEMM, DIPEND, LAUREA	0,29	0,48	0,10	0,25	0,20	0,21	0,25	2
3	Dimensione cola	agri- SAUMED, AZ2HA, AZ50HA, S2HA, S50HA	0,11	0,10	0,68	0,08	0,12	0,11	0,25	2
4	Dinamiche	VARPOP, VARSAU, VARAZ, VARSUP	0,13	0,25	0,08	0,42	0,14	0,10	0,14	1
5	Specializzazione agricola	BOSCHIVE, CEREALI, FORAGGI, PRATI, ORTIVE, FRUTTI, VITE, VITEDOC, AZALLEV, BOVINI, SUINI, AVICOLI, BOVPRATI	0,15	0,20	0,12	0,14	0,28	0,14	0,28	5
6	Sviluppo industria e agroalimentare	ADDIND, INDALIME, ADD100	0,28	0,21	0,11	0,10	0,14	0,62	0,18	1
7	Produttività cola	agri- TRATTSAU, RLSSAU, RLSULA, ULASAU	0,13	0,25	0,25	0,14	0,28	0,18	0,53	2

Lo stesso confronto è stato effettuato sulle componenti principali e i risultati sono riportati nella tavola 4.4. Tale confronto è, però, limitato all'eterogeneità tra i gruppi formati. Per avere indicazioni anche sull'omogeneità entro i gruppi rispetto agli indicatori originali o alle componenti, vengono riportati i valori della Pseudo F ricalcolati sia rispetto alla matrice dei dati originali che sulle componenti principali.

Il valore della Pseudo F⁹ è calcolato secondo la seguente formula:

$$F = \frac{Dev(B)/(g-1)}{Dev(W)/(n-g)} \quad (6)$$

dove *Dev(B)* rappresenta la devianza tra i cluster
Dev(W) rappresenta la devianza entro i cluster
n rappresenta il numero di unità statistiche
g rappresenta il numero di gruppi

Tav. 4.4 - Test sulle cluster analysis ottenute: confronti tra le medie dei gruppi e Pseudo F

Metodo	Componenti principali		Indicatori originali	
	Test t sulle medie ¹	Pseudo F	Test t sulle medie ¹	Pseudo F
Analisi in blocco	61,4%	44,5	65,9%	29,3
Analisi con 5 blocchi	71,0%	42,3	68,9%	32,1
Analisi con 7 blocchi	68,6%	33,8	64,6%	26,1

1) % dei test risultati significativi ad un livello di confidenza del 95%

Sotto il profilo delle componenti principali, la tavola 4.4 mostra come in questo caso l'analisi per blocchi, sebbene abbia permesso di aumentare la percentuale di medie significativamente diverse rispetto all'analisi in blocco, porti ad un raggruppamento in cluster meno efficiente dal punto di vista del rapporto tra eterogeneità tra i gruppi ed omogeneità interna ai gruppi. La pseudo-F cala ancora più nettamente quando si introduce ulteriore correlazione nell'analisi con 7 blocchi. Se invece ci si riferisce agli indicatori originali, opportunamente standardizzati, si può notare come

⁹ Il test, tratto da Brasini et al. (1993) è stato formulato originariamente da Calinski e Harabasz (1974).

l'analisi con 5 blocchi abbia ottenuto la migliore performance sia in termini di test sulle medie che di pseudo F, mentre quella con 7 è risultata meno efficace dell'analisi in blocco.

E' comunque necessario tenere sempre presente che le componenti principali estratte in blocco sono tra loro non correlate, mentre tra quelle estratte per blocchi esistono alcune correlazioni altamente significative. Per dare un'idea delle implicazioni di tale problema si può riportare brevemente il risultato dell'analisi condotta direttamente sugli indicatori originali senza effettuare ACP e ignorando dunque le correlazioni tra gli indicatori. Nella zonizzazione risultata ottimale secondo il test pseudo F (con 5 gruppi), circa l'86% dei comuni tendeva a raggrupparsi in un unico cluster, mentre i primi due cluster riunivano addirittura il 97% dei comuni. Neppure la cluster con 10 gruppi, che presentava un altro picco nel valore della pseudo F, ha portato a risultati accettabili: 65% dei comuni in un unico gruppo, 93% nei primi tre gruppi.

4.4 Un approccio "a stadi"

Una soluzione per conciliare l'esigenza del ricercatore di dare rilevanza a un particolare insieme di variabili ed il rispetto dell'obiettivo statistico di inserire componenti non correlate nella cluster analysis potrebbe essere nell'applicazione del metodo in blocco a due o più stadi successivi.

Ad esempio si potrebbero utilizzare in un primo stadio solo gli indicatori relativi alla specializzazione e alla struttura agricola, per giungere ad una zonizzazione che tenga conto solo dell'attitudine produttiva del territorio. Se poi l'obiettivo fosse quello di verificare l'interazione con l'industria agroalimentare, si potrebbero utilizzare tutti gli indicatori relativi alla concentrazione e specializzazione dell'agroindustria in un secondo stadio ed eventualmente solo in quelle zone che rivestono un particolare interesse dal punto di vista economico (potenziali distretti agroindustriali).

Un approccio come quello appena descritto non eliminerebbe le possibilità di utilizzare le conoscenze a priori del fenomeno oggetto di studio e di isolare gruppi di indicatori considerati particolarmente rilevanti dal punto di vista economico. Al tempo stesso, suddividendo l'analisi in stadi successivi, si avrebbe il vantaggio di eliminare i problemi statistici dell'analisi per blocchi e di rendere più trasparente l'effetto delle scelte del ricercatore.

5. Conclusioni

Nel presente lavoro sono state riportate una serie di riflessioni sui metodi statistici utilizzati nell'analisi territoriale ed in particolare sulle differenze tra gli approcci "in blocco" e "per blocchi". Le considerazioni emerse sono, conseguentemente, principalmente legate all'oggettività dell'analisi in termini puramente statistici, mentre la caratterizzazione economica dei risultati esula dagli obiettivi di questo studio. In particolare è evidente l'influenza della scelta degli indicatori originali sull'intero procedimento.

Il confronto tra i due metodi nel caso dell'Emilia-Romagna e dei 38 indicatori considerati ha messo in luce le potenziali distorsioni del metodo "per blocchi" dovute alla presenza di correlazioni significative tra indicatori appartenenti a blocchi diversi e alla ripartizione degli indicatori tra i vari blocchi. In questo caso cade, infatti, la proprietà di non correlazione tra le componenti principali utilizzate nella cluster analysis, perciò viene meno l'utilità stessa dell'analisi delle componenti principali. Del resto l'applicazione diretta della cluster analysis direttamente sugli indicatori, cioè ignorando volutamente tutte le correlazioni, non porta a risultati accettabili dal punto di vista informativo. Viene inoltre sottolineata la rilevanza della scelta, nell'analisi "per blocchi", se utilizzare punteggi standardizzati o non standardizzati per le componenti e le implicazioni teoriche alla base di tale scelta. L'approccio più coerente con le assunzioni alla base delle analisi territoriali è quello di standardizzare i punteggi nel caso dell'analisi "per blocchi", al contrario di quanto avviene in quella "in blocco".

L'analisi "in blocco" elimina invece il passaggio più o meno soggettivo della suddivisione in macrodeterminanti e consente anche un'interpretazione "trasversale" delle componenti. A tal fine nel paragrafo 2.1 è stato proposto un metodo per la semplificazione dell'interpretazione delle componenti. Nello stesso paragrafo viene descritto come sia possibile ottenere una valutazione gerarchica della rilevanza degli indicatori originali nelle componenti principali selezionate.

D'altra parte non è improbabile che l'applicazione di un'analisi in blocco non porti a risultati interpretabili dal punto di vista economico. Per valorizzare le conoscenze a priori del fenomeno e rispettare le caratteristiche desiderabili in termini statistici è stato proposto nel par. 4.4 un approccio a stadi, che espliciti le scelte del ricercatore nel dare risalto a determinate macrodeterminanti dello sviluppo territoriale agricolo.

Per avere una valutazione più completa degli aspetti evidenziati in questo lavoro, sarà utile allargare la base di dati territoriali ad altre regioni con caratteristiche agricole diverse. In tal modo sarà anche possibile avere informazioni più attendibili sulla rilevanza degli indicatori considerati nella individuazione dei sistemi agricoli territoriali.

APPENDICE

App. A - Gli indicatori utilizzati nell'analisi sullo sviluppo territoriale agricolo dell'Emilia-Romagna

Variabile	Descrizione	Fonte
ADD100	% addetti in ind. alim. con più di 99 addetti	Elab. ISTAT Cens. Industria
ADDAGR	% addetti in agricoltura su totale addetti	CAIRE
ADDIND	% addetti industria su totale addetti	CAIRE
ADDERZ	% addetti terziario su totale addetti	CAIRE
ATTFEMM	Tasso di attività femminile	Elab. ISTAT Cens. Popolaz.
AVICOLI	Numero medio di capi avicoli per ha di SAU	Elab. ISTAT Cens. Agricoltura
AZ2HA	% aziende con SAU al di sotto di 2 ha	Elab. ISTAT Cens. Agricoltura
AZ50HA	% aziende con SAU oltre 50 ha	Elab. ISTAT Cens. Agricoltura
AZALLEV	% aziende con allevamenti	Elab. ISTAT Cens. Agricoltura
BOSCHIVE	% sup. agricola totale a boschive	Elab. ISTAT Cens. Agricoltura
BOVINI	Numero medio di capi bovini per ha di SAU	Elab. ISTAT Cens. Agricoltura
BOVPRATI	Bovini /ha SAU a prati, pascoli e foraggiere	Elab. ISTAT Cens. Agricoltura
CEREALI	% SAU a cereali	Elab. ISTAT Cens. Agricoltura
DENPOP	Residenti per km ² 1991	Elab. ISTAT Cens. Popolaz.
DINPOP	Variazione % pop. residente 1981-1991	Elab. ISTAT Cens. Popolaz.
DIPEND	Indice di dipendenza	Elab. ISTAT Cens. Popolaz.
DISOCC	Tasso di disoccupazione	ISTAT
DOCVITE	% SAU a vite DOC sul totale SAU a vite	Elab. ISTAT Cens. Agricoltura
FORAGGI	% SAU a foraggi	Elab. ISTAT Cens. Agricoltura
FRUTTI	% SAU a coltivazioni frutticole	Elab. ISTAT Cens. Agricoltura
INDALIME	Addetti industria alimentare per unità locale	Elab. ISTAT Cens. Industria
LAUREA	% laureati su pop. in età scolare	Elab. ISTAT Cens. Popolaz.
ORTIVE	% SAU a coltivazioni ortive	Elab. ISTAT Cens. Agricoltura
PIL	PIL pro-capite	CAIRE
PRATI	% SAU a prati e pascoli	Elab. ISTAT Cens. Agricoltura
RLSSAU	Reddito Lordo Standard per ha di SAU	CAIRE
RLSULA	Reddito Lordo Standard per ULA	CAIRE
S2HA	% SAU delle aziende con SAU sotto i 2 ha	Elab. ISTAT Cens. Agricoltura
S50HA	% SAU delle aziende con SAU oltre i 50 ha	Elab. ISTAT Cens. Agricoltura
SAUMED	SAU media per azienda	Elab. ISTAT Cens. Agricoltura
SUINI	Numero medio di capi suini per ha di SAU	Elab. ISTAT Cens. Agricoltura
TRATTSAU	Numero di trattrici per ha di SAU	Elab. ISTAT Cens. Agricoltura
ULASAU	Unità di lavoro agricolo per ha di SAU	CAIRE
VARAZ	Variazione % num. aziende 1982-1990	Elab. ISTAT Cens. Agricoltura
VARSAU	Variazione % SAU 1982-1990	Elab. ISTAT Cens. Agricoltura
VARSUP	Variazione sup. agricola totale 1982-1990	Elab. ISTAT Cens. Agricoltura
VECCHIAI	Indice di vecchiaia	Elab. ISTAT Cens. Popolaz.
VITE	% SAU a vite	Elab. ISTAT Cens. Agricoltura

Fonte: Montresor, Mazzocchi [1999]

App. B - Interpretazione delle componenti principali estratte mediante ACP in blocco

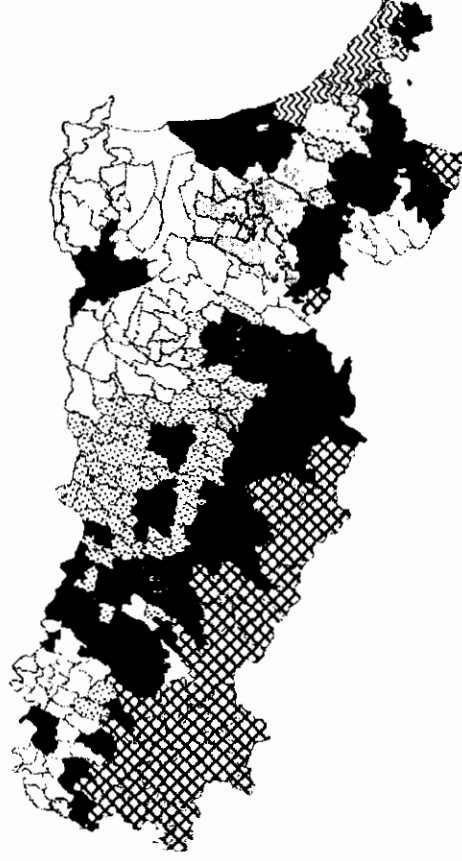
Componente	Indicatori "caratterizzanti"	a _{ij}	s _{ij}	Componente	Indicatori "caratterizzanti"	a _{ij}	s _{ij}
Livello di sviluppo agricolo	BOSCHIVE PRATI RLSSAU RLSULA VECCHIAI ATTFEMM DIPEND	-0,30 -0,25 0,29 0,23 -0,24 0,24 -0,30	8,8 6,2 8,3 5,3 5,9 5,7 9,3	Livello di urbanità (contrasto con ruralità)	PIL DENPOP ADDAGR BOVPRATI RLSSAU LAUREA ADDERZ	0,24 0,28 -0,38 -0,23 -0,26 0,33 0,33	5,6 7,6 14,2 5,3 6,9 10,6 10,7
Variabilità spiegata dalla componente 1: 21%				Variabilità spiegata dalla componente 4: 6%			
Livello di frammentazione delle strutture agricole	SAUMED AZ2HA AZ50HA S2HA S50HA ULASAU	-0,40 0,34 -0,33 0,38 -0,29 0,31	15,8 11,8 10,7 14,1 8,5 9,7	Integrazione agroindustriale	ORTIVE FRUTTI INDALIME VECCHIAI LAUREA VARAZ ADD100	-0,22 0,25 0,34 0,29 0,30 -0,23 0,39	5,0 6,4 11,7 8,3 9,0 5,1 15,6
Variabilità spiegata dalla componente 2: 13%				Variabilità spiegata dalla componente 5: 5%			
Livello presenza allev. bovini e suini (-)	DISOCC FORAGGI ORTIVE BOVINI SUINI ADDIND ADDERZ	0,38 -0,35 0,26 -0,35 -0,27 -0,23 0,24	14,1 12,3 6,8 12,5 7,4 5,2 5,9	Livello di presenza di allevamenti bovini intensivi (agricoltura sviluppata)	DENPOP ORTIVE DOCVITE AZALLEV BOVINI BOVPRATI ADDERZ VARAZ	0,26 0,29 -0,25 0,25 0,34 0,25 0,26 -0,26	7,0 8,6 6,1 6,3 11,5 6,2 6,9 7,0
Variabilità spiegata dalla componente 3: 9%				Variabilità spiegata dalla componente 6: 5%			

App. B (segue) - Interpretazione delle componenti principali estratte mediante ACP in blocco

Componente	Indicatori "caratterizzanti"	a _{ij}	s _{ij}	Componente	Indicatori "caratterizzanti"	a _{ij}	s _{ij}
Livello industrializzazione (abbandono attività agric.)	ADDAGR AZ2HA FRUTTI LAUREA ADDIND VARSAU VARSUP	-0,23 0,32 -0,28 -0,24 0,35 0,35 -0,28	5,4 10,3 8,1 5,8 12,6 12,5 7,6	Tendenza all'aumento della sup. agricola e aziende (presenza ind. agroalim.)	INDALIME DINPOP ATTFEMM VARAZ VARSAU ADD100	0,23 -0,28 -0,24 0,31 0,40 0,25	5,3 8,1 5,6 9,4 16,3 6,4
Variabilità spiegata dalla componente 7: 4%				Variabilità spiegata dalla componente 10: 3%			
Tendenza a concentrazione agricola (assenza ind. agroalim.)	S50HA DOCVITE SUINI INDALIME VARSAU ADD100	0,27 0,37 0,29 -0,38 0,28 -0,34	7,4 13,4 8,7 14,7 7,7 11,4	Arete viticole DOC e allevamenti, diminuzione aziende e sup. agricola (-)	PRATI VITE DOCVITE AZALLEV VARAZ VARSAU	0,24 -0,34 -0,33 -0,38 0,34 0,32	5,8 11,8 11,1 14,1 11,7 9,9
Variabilità spiegata dalla componente 8: 4%				Variabilità spiegata dalla componente 11: 3%			
Estensivizzazione agricola Agricoltura con allevamenti (presenza ind. agroalim.)	AZ50HA CEREALI PRATI AZALLEV AVICOLI ULASAU INDALIME LAUREA ADD100	0,27 -0,29 0,24 0,31 0,41 0,26 0,28 -0,24 0,23	7,1 8,7 5,9 9,6 17,2 6,7 7,6 5,8 5,5	Componenti estratte: 11			
Variabilità spiegata dalla componente 9: 4%				TOT. VARIANZA SPIEGATA			77%

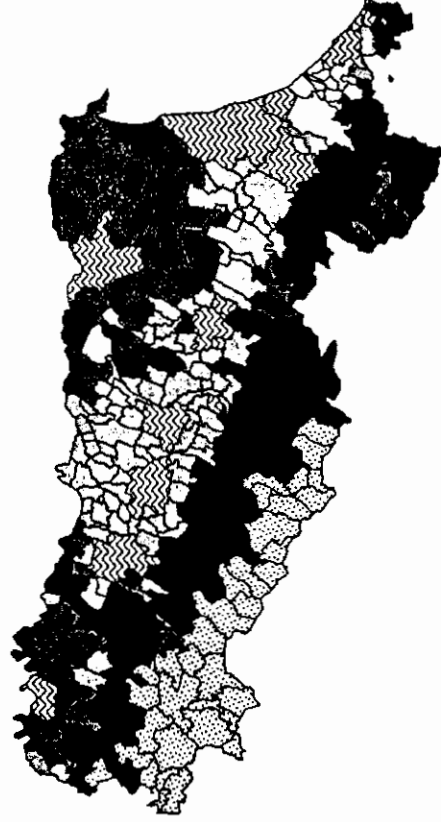
App. C.1 – Zonizzazione dello sviluppo agricolo in Emilia-Romagna (ACP “in blocco”)

Cluster analysis (ACP in blocco)
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8



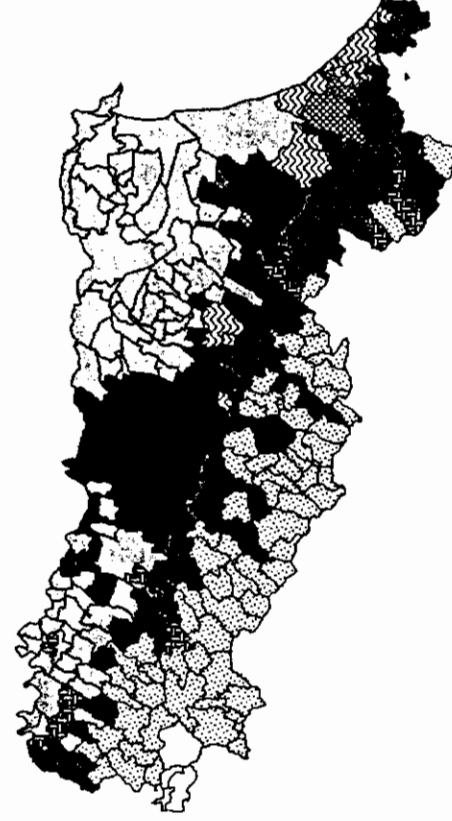
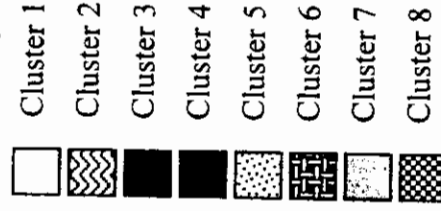
App. C.2 – Zonizzazione dello sviluppo agricolo in Emilia-Romagna (ACP “per blocchi” con 5 blocchi)

Cluster analysis (5 blocchi)
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6



App. C.3 – Zonizzazione dello sviluppo agricolo in Emilia-Romagna (ACP “per blocchi” con 7 blocchi)

Cluster Analysis (ACP 7 blocchi)



App. D.1 - Valori medi degli indicatori originali e delle componenti principali nei cluster (analisi in blocco)

Cluster	1	2	3	4	5	6	7	8	Emilia-Romagna
Numero comuni	18	55	88	9	30	50	68	23	341
ADD100	13,0	16,5	4,4	36,4	24,2	1,0	10,4	28,2	11,7
ADDAGR	8,6	14,9	18,8	2,9	18,4	31,9	10,0	19,0	17,4
ADDIND	32,2	43,2	38,7	31,7	42,1	22,6	54,6	37,5	39,9
ADDERZ	59,2	41,8	42,5	65,5	39,5	45,5	35,4	43,6	42,7
ATTFEMM	49,3	56,4	51,4	54,8	49,0	43,5	58,5	57,0	52,6
AVICOLI	89,2	12,3	21,4	7,1	8,3	3,8	12,2	82,5	21,7
AZ2HA	63,5	25,7	35,3	29,1	19,4	33,0	34,9	38,7	33,5
AZ50HA	0,2	2,5	1,5	3,3	8,3	1,0	1,7	0,8	2,2
AZALLEV	54,9	44,6	55,3	42,3	63,1	55,3	55,3	45,9	53,3
BOSCHIVE	2,0	2,0	23,4	5,8	13,4	45,5	3,9	3,1	15,5
BOVINI	0,2	0,3	0,8	0,8	1,0	0,5	1,4	0,2	0,8
BOVPRATI	1,2	3,0	1,3	2,0	2,9	0,5	3,7	2,2	2,1
CEREALI	39,4	40,7	21,7	33,1	33,8	8,0	26,0	18,3	25,7
DENPOP	667,8	152,6	79,1	1015,9	84,9	27,0	255,6	275,0	188,0
DINPOP	8,1	0,4	1,5	-3,4	1,2	-11,6	7,1	-0,7	0,6
DIPEND	41,0	43,6	49,5	40,0	45,7	67,6	41,5	46,1	48,4
DISOCC	13,1	8,2	6,7	6,9	7,8	7,9	5,2	6,8	7,3
DOCVITE	26,9	7,8	13,2	25,6	7,4	3,1	34,4	23,9	16,4
FORAGGI	14,2	12,2	52,7	28,1	33,3	38,1	33,8	10,8	33,0
FRUTTI	8,3	11,6	1,8	6,8	1,0	0,2	5,8	35,2	6,6
INDALIME	5,3	10,0	5,1	10,7	14,9	3,3	8,9	12,4	7,9
LAUREA	22,1	23,0	21,4	91,0	14,9	17,7	21,0	30,4	22,9
ORTIVE	7,9	5,5	0,8	4,0	8,2	0,1	1,4	1,6	2,7
PIL	24,5	24,6	22,2	33,5	25,4	16,6	30,6	24,5	24,3
PRATI	1,6	0,8	13,8	6,1	12,9	49,4	7,9	3,5	14,1
RLSSAU	3,9	3,8	1,7	3,2	2,5	0,7	3,9	5,3	2,8

App. D.1 (segue) - Valori medi degli indicatori originali e delle componenti princ. nei cluster (analisi in blocco)

Cluster	1	2	3	4	5	6	7	8	Emilia-Romagna
RLSULA	16,8	43,8	18,1	30,6	37,3	9,4	29,2	27,2	25,8
S2HA	19,3	2,3	4,7	2,8	0,9	4,5	4,4	6,7	4,7
S50HA	6,0	26,3	13,9	29,9	42,2	12,3	17,3	14,2	18,9
SAUMED	3,0	10,9	7,7	11,2	19,3	6,8	8,0	5,8	8,9
SUINI	1,4	0,5	1,3	1,7	1,0	0,4	4,2	1,0	1,6
TRATTSAU	0,2	0,2	0,1	0,1	0,1	0,1	0,2	0,2	0,2
ULASAU	0,2	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,1
VARAZ	-3,6	-12,1	-12,0	-14,0	-10,3	-24,9	-12,0	-5,9	-13,0
VARSAU	-4,3	-0,2	-4,2	-2,7	-0,4	26,4	-3,1	0,8	1,9
VARSUP	-1,1	-0,8	-4,2	-1,3	-2,1	-14,1	-2,6	1,6	-4,0
VECCHIAI	11,7	19,8	19,3	21,3	18,9	44,0	15,0	20,1	21,8
VITE	9,9	1,9	3,2	5,2	1,0	1,6	10,2	20,8	5,6
COMP1	1,7	1,7	-1,2	2,1	0,7	-5,2	2,2	2,2	0,0
COMP2	4,8	-1,2	0,1	0,2	-3,8	0,2	0,1	2,6	0,0
COMP3	2,2	1,5	-0,8	1,6	1,2	0,1	-1,9	0,9	0,0
COMP4	0,8	-0,5	0,5	3,5	0,3	-0,4	-0,2	-1,5	0,0
COMP5	-1,8	0,2	-0,7	2,3	-0,5	0,6	0,1	1,7	0,0
COMP6	1,1	-0,2	-0,2	1,6	0,5	0,2	-0,1	-1,2	0,0
COMP7	1,0	-0,1	-0,2	-1,3	0,2	0,0	0,5	-1,2	0,0
COMP8	-0,1	-0,5	-0,3	1,6	-0,2	0,0	0,7	0,1	0,0
COMP9	0,6	-0,9	-0,1	-1,4	1,3	0,1	0,0	0,8	0,0
COMP10	-0,1	-0,3	0,1	0,7	0,3	0,0	0,0	-0,1	0,0
COMP11	-0,5	0,3	0,1	-0,6	-0,3	0,0	0,1	-0,4	0,0

App. D.2 - Valori medi degli indicatori originali e delle componenti principali nei cluster (analisi con 5 blocchi)

Cluster	1	2	3	4	5	6	Emilia-Romagna
Numero comuni	102	34	69	105	14	17	341
ADD100	12,1	0,0	25,0	2,5	18,2	30,2	11,7
ADDAGR	11,2	29,3	17,6	21,7	16,5	3,4	17,4
ADDIND	51,7	22,1	41,1	36,3	34,1	28,0	39,9
ADDFERZ	37,1	48,7	41,3	42,1	49,4	68,7	42,7
ATTFEMM	57,9	41,5	52,6	50,9	54,5	52,5	52,6
AVICOLI	13,7	4,2	9,1	17,5	216,5	21,0	21,7
AZ2HA	32,9	33,8	23,4	35,8	55,0	45,6	33,5
AZ50HA	1,6	1,1	4,8	1,6	0,3	1,8	2,2
AZALLEV	54,1	55,8	51,9	54,6	51,7	41,9	53,3
BOSCHIVE	3,5	50,9	4,9	26,2	2,2	4,0	15,5
BOVINI	1,2	0,4	0,6	0,7	0,2	0,5	0,8
BOVPRATI	3,5	0,5	3,0	1,0	1,6	1,6	2,1
CEREALI	27,2	4,4	37,9	21,2	29,8	34,5	25,7
DENPOP	240,1	23,0	122,4	69,5	393,8	1034,9	188,0
DINPOP	5,0	-12,8	0,3	0,2	5,3	0,0	0,6
DIPEND	42,7	70,0	44,4	51,7	40,6	40,3	48,4
DISOCC	5,5	8,4	8,5	7,0	10,3	9,6	7,3
DOCVITE	27,8	0,8	4,9	15,1	27,7	24,0	16,4
FORAGGI	28,3	28,8	21,2	51,5	11,3	22,3	33,0
FRUTTI	9,6	0,0	7,7	2,3	22,8	9,8	6,6
INDALIME	9,0	3,3	13,4	4,4	8,7	9,0	7,9
LAUREA	23,2	17,3	19,3	20,6	19,0	65,3	22,9
ORTIVE	1,6	0,1	7,4	0,6	10,1	3,2	2,7
PIL	29,7	16,5	23,5	21,1	24,8	30,6	24,3
PRATI	6,4	63,9	4,3	15,0	2,2	4,3	14,1
RLSSAU	4,0	0,6	3,4	1,5	5,5	3,3	2,8

App. D.2 (segue) - Valori medi degli indicatori originali e delle componenti princ. nei cluster (analisi con 5 blocchi)

Cluster	1	2	3	4	5	6	7	8	Emilia-Romagna
RLSULA	30,3	8,6	42,8	16,7	19,7	24,4	25,8		
S2HA	4,2	4,3	1,8	4,9	16,2	10,4	4,7		
S50HA	16,2	15,4	33,2	14,6	7,5	19,4	18,9		
SAUMED	8,0	6,9	14,4	7,7	3,6	7,4	8,9		
SUINI	3,2	0,3	0,7	1,0	1,8	1,2	1,6		
TRATTSAU	0,2	0,1	0,1	0,1	0,2	0,2	0,2		
ULASAU	0,1	0,1	0,1	0,1	0,3	0,2	0,1		
VARAZ	-11,8	-27,6	-12,3	-12,0	-1,1	-8,8	-13,0		
VARS AU	-1,7	43,0	-0,4	-5,8	-1,3	0,1	1,9		
VARSUP	-1,6	-16,1	-1,7	-5,6	0,2	3,1	-4,0		
VECCHIAI	16,4	48,0	19,8	21,8	11,9	18,3	21,8		
VITE	11,1	0,4	1,2	3,8	9,2	8,0	5,6		
COMP1	0,7	-1,2	0,3	-0,4	-0,2	0,1	0,0		
COMP2	-0,5	0,2	0,3	-0,3	0,8	2,0	0,0		
COMP3	-0,4	0,4	0,6	0,0	0,1	-0,8	0,0		
COMP4	0,6	-2,0	0,1	-0,3	0,7	1,4	0,0		
COMP5	-0,2	0,6	-0,4	-0,2	-0,2	2,8	0,0		
COMP6	-0,1	-0,2	1,1	-0,2	-1,9	-0,6	0,0		
COMP7	0,6	-1,4	0,5	-0,8	1,5	0,4	0,0		
COMP8	-0,3	-0,3	0,3	-0,1	1,3	0,9	0,0		
COMP9	0,3	-1,6	0,1	0,0	0,8	0,5	0,0		
COMP10	0,6	-1,8	0,5	-0,6	1,2	0,6	0,0		
COMP11	0,5	-0,6	-0,2	0,1	-1,3	-0,5	0,0		
COMP12	0,6	0,2	-1,1	0,1	0,2	0,1	0,0		
COMP13	-0,2	0,1	-0,2	0,0	2,7	-0,5	0,0		
COMP14	0,3	1,1	0,2	-0,7	0,0	-0,4	0,0		

App. D.3 - Valori medi degli indicatori originali e delle componenti principali nei cluster (analisi con 7 blocchi)

Cluster	1	2	3	4	5	6	7	8	Emilia-Romagna
Numero comuni	4	17	77	76	67	10	85	5	341
ADD100	0,0	16,6	2,7	18,0	0,8	47,4	16,5	40,3	11,7
ADDAGR	41,2	6,3	17,6	11,2	28,0	17,7	15,6	13,7	17,4
ADDIND	16,4	28,8	40,4	51,3	26,7	48,3	42,2	39,3	39,9
ADDERZ	42,4	64,9	42,0	37,5	45,3	34,0	42,2	46,9	42,7
ATTFEMM	33,7	51,3	52,1	58,2	46,7	49,6	54,0	58,7	52,6
AVICOLI	0,7	55,2	29,4	11,5	4,1	18,8	9,4	411,2	21,7
AZ2HA	46,5	61,0	38,6	33,5	32,6	16,9	24,4	51,0	33,5
AZ50HA	1,6	0,4	2,1	1,5	1,0	12,2	3,0	0,3	2,2
AZALLEV	31,9	48,8	52,0	56,5	58,1	60,3	49,7	37,8	53,3
BOSCHIVE	58,9	2,5	19,3	3,0	42,8	18,2	2,7	1,8	15,5
BOVINI	0,3	0,2	0,6	1,4	0,6	1,0	0,7	0,1	0,8
BOVPRATI	0,3	0,9	1,1	3,7	0,7	2,3	3,2	2,1	2,1
CEREALI	1,7	36,0	26,1	23,4	9,2	28,9	39,1	1,8	25,7
DENPOP	10,0	955,9	103,7	278,1	34,2	75,8	156,6	469,3	188,0
DINPOP	-21,7	4,3	4,0	4,7	-7,2	2,7	-0,2	4,2	0,6
DIPEND	93,8	40,3	47,8	42,5	61,1	48,6	44,0	40,4	48,4
DISOCC	9,9	11,4	7,4	5,3	7,2	6,4	7,9	8,6	7,3
DOCVITE	0,0	23,1	27,0	29,1	2,1	4,6	7,0	22,6	16,4
FORAGGI	3,0	16,5	43,0	30,9	46,3	35,5	21,4	7,0	33,0
FRUTTI	1,1	11,8	4,0	11,2	0,5	1,2	8,1	30,7	6,6
INDALIME	1,1	6,2	5,2	10,3	3,6	27,9	9,4	16,2	7,9
LAUREA	18,0	39,5	22,7	24,6	17,2	16,3	24,1	19,5	22,9
ORTIVE	0,2	5,8	0,9	1,1	0,1	7,1	6,4	11,8	2,7
PIL	16,1	26,5	22,5	29,6	17,8	28,9	25,4	30,2	24,3
PRATI	91,4	2,0	11,5	7,7	39,4	19,8	1,1	2,3	14,1
RLSSAU	0,3	3,8	1,9	4,3	0,8	2,1	3,6	7,0	2,8

Cluster	1	2	3	4	5	6	7	8	Emilia-Romagna
RLSULA	4,2	17,2	19,2	29,3	10,2	36,8	42,3	23,2	25,8
S2HA	4,7	17,9	5,3	4,4	4,5	0,7	2,1	14,1	4,7
S50HA	23,7	8,6	19,3	15,6	11,1	50,5	26,0	12,8	18,9
SAUMED	6,4	3,6	7,9	7,8	7,0	22,5	12,1	4,1	8,9
SUINI	0,0	1,4	0,9	4,1	0,6	0,7	0,9	1,1	1,6
TRATTSAU	0,1	0,2	0,1	0,2	0,1	0,1	0,2	0,3	0,2
ULASAU	0,1	0,2	0,1	0,2	0,1	0,1	0,1	0,3	0,1
VARAZ	-43,5	-3,2	-9,2	-12,6	-20,9	-9,0	-12,1	-0,7	-13,0
VARSAU	467,2	-2,1	-3,8	-2,7	-11,2	5,4	-0,3	4,3	1,9
VARSUP	-27,7	1,6	-3,5	-2,2	-11,7	-0,1	-1,1	6,9	-4,0
VECCHIAI	118,9	14,9	18,3	16,1	32,3	19,4	19,5	12,3	21,8
VITE	0,5	9,9	7,5	11,5	1,2	0,7	1,6	8,6	5,6
COMP1	1,7	-0,6	0,1	-0,7	0,8	-0,2	-0,1	-0,5	0,0
COMP2	0,2	1,9	0,0	-0,6	0,1	-0,6	0,1	0,4	0,0
COMP3	-4,7	1,1	0,0	0,6	-1,1	-0,2	0,2	0,9	0,0
COMP4	2,7	1,6	-0,2	-0,1	0,1	-0,4	-0,2	-0,1	0,0
COMP5	-0,3	-1,7	-0,2	-0,2	-0,3	2,6	0,6	-1,2	0,0
COMP6	0,5	1,7	0,2	-0,3	-0,5	1,7	-0,2	1,1	0,0
COMP7	-4,0	0,7	0,3	0,2	-0,8	0,4	0,1	1,0	0,0
COMP8	-2,0	0,7	-0,2	0,6	-1,4	-0,2	0,6	1,5	0,0
COMP9	-1,4	-0,8	-0,1	0,8	-0,1	0,2	-0,1	-2,2	0,0
COMP10	0,4	0,0	0,4	0,8	0,0	-1,0	-1,0	0,3	0,0
COMP11	-0,7	0,3	0,0	0,0	0,1	0,4	-0,3	4,8	0,0
COMP12	2,3	-0,5	-0,9	0,5	0,1	0,2	0,2	0,2	0,0
COMP13	-1,0	-0,2	-0,3	0,5	-0,7	2,0	0,2	1,0	0,0
COMP14	-1,5	0,9	-0,5	0,8	-1,1	-0,4	0,4	2,4	0,0
COMP15	-0,7	-1,5	-0,2	-0,1	-0,5	1,1	1,1	-1,8	0,0

Riferimenti bibliografici

Aldenderfer, M.S., Blashfield R.K. (1984), *Cluster Analysis*, Sage Publications, Beverly Hills

Anania G., Tarsitano (1995), *Tecniche di analisi statistica multivariata per l'individuazione dei sistemi agricoli territoriali in Italia*, in I sistemi territoriali agricoli degli anni '90 (a cura di Cannata G.), CNR-Raisa, Rubbettino, Cosenza

Boccafogli F., Brasili C. (1998), *L'articolazione territoriale dello sviluppo agricolo in Emilia-Romagna*, La Questione Agraria, n. 70

Brasini, S., Tassinari, F., Tassinari, G. (1993), *Marketing e pubblicità: Approccio statistico all'analisi dei mercati di consumo*, Il Mulino, Bologna

Cannata G. (a cura di, 1989), *I sistemi agricoli territoriali Italiani*, CNR-IPRA, Angeli, Milano.

Cannata G., Forleo, M.B. (a cura di, 1998), *I sistemi agricoli territoriali delle regioni italiane*, CNR-IPRA

Cooperativa Architetti ed Ingegneri di Reggio Emilia - CAIRE (1995), *Atlante nazionale del territorio rurale*, CAIRE, Reggio Emilia

Copus A., B. Crabtree (1992), *Mapping Economic Fragility: an Assesment of the Objective 5b Boundaries in Scotland*, Journal of Rural Studies, 8(3)

Everitt, R.S (1993), *Cluster Analysis*, 3rd ed., Arnold, New York

Fabbris, L. (1997), *Statistica multivariata : analisi esplorativa dei dati*, Milano, McGraw-Hill Italia

Fanfani R., Montesor E. (1998), *L'articolazione dello sviluppo rurale e agroalimentare in Italia*. Atti della XXX Conferenza SIDEA,

"L'agricoltura italiana tra prospettive continentali e mediterranee", Napoli.

Forleo, M.B. (1998), *I sistemi agricoli territoriali delle regioni italiane. Aspetti metodologici ed operativi* in I sistemi agricoli territoriali delle regioni italiane (a cura di Cannata and Forleo)

Krzanowski, W.J. (1988), *Principles of Multivariate Analysis*, Oxford University Press

Mignani S., Montanari A. (1997), *Appunti di analisi statistica multivariata*, 2a edizione, Esculapio, Bologna

Montesor E., (1998), *Gli effetti regionali della politica comunitaria agricola*, Atti dei Seminari Jean Monnet.

Montesor E., Mazzocchi M. (1999), *Agricultural and rural development at regional level: an analytical approach*, paper presentato al 3° seminario Eurotools, Reading, 9-10 aprile 1999

OECD (1994), *Creating rural indicators for shaping territorial policy*

OECD (1996), *Territorial indicators of Employment - focusing on Rural Development*

Rossi-Doria, M. (1969), *Analisi zonale dell'agricoltura Italiana*, INEA

SAS Institute Inc., (1989) *SAS/STAT: User's Guide*