



**AGRI@STAT - VERSO
UN NUOVO SISTEMA DI
STATISTICHE AGRICOLE**

*Firenze
30-31 maggio 2005*

Atti del convegno

UNIVERSITÀ DEGLI STUDI DI FIRENZE
Dipartimento di Statistica "Giuseppe Parenti"
sede: Viale Morgagni 59, 50134 Firenze
Tel.: +39.055.4237211 - Fax: +39.055.4223560
sito web: <http://www.ds.unifi.it/>

CONVEGNO DI STUDI
AGRI@STAT - Verso un nuovo sistema di statistiche agricole

Comitato scientifico

Antonio Giusti (Presidente)
Università degli Studi di Firenze
Leonardo Casini
Università degli Studi di Firenze
Andrea Giommi
Università degli Studi di Firenze
Monica Pratesi
Università degli Studi di Pisa
Massimo Sabbatini
Università degli Studi di Cassino

Comitato organizzatore

Alessandra Petrucci (Presidente)
Università degli Studi di Firenze
Chiara Bocci
Università degli Studi di Firenze
Emilia Rocco
Università degli Studi di Firenze
Nicola Salvati
Università degli Studi di Pisa

Segreteria organizzativa

Chiara Bocci
Dipartimento di Statistica "Giuseppe Parenti"
Università degli Studi di Firenze
tel: +39.055.4237263
fax: +39.055.4223560
e-mail: agri@ds.unifi.it



**AGRI@STAT - VERSO
UN NUOVO SISTEMA DI
STATISTICHE AGRICOLE**

Firenze
30-31 maggio 2005

Atti del convegno

a cura di Chiara Bocci, Andrea Giommi, Alessandra Petrucci



Dipartimento di Statistica “Giuseppe Parenti”
Università degli Studi di Firenze

Organizzazione del convegno
Dipartimento di Statistica “Giuseppe Parenti”,
Università degli studi di Firenze

Il Comitato Organizzatore ringrazia:

- L’Azienda di Promozione Turistica di Firenze
- I Partecipanti alla Tavola Rotonda
- I Presidenti ed i Relatori delle diverse Sessioni

che hanno reso possibile la realizzazione del Convegno e la pubblicazione del presente volume

© Dipartimento Istituzioni, Metodi Quantitativi e Territorio
Università degli Studi di Cassino
Via S. Angelo Loc. Folcara
03043 Cassino (FR)
tel. 0776/2994705
<http://dimet.eco.unicas.it>

Cassino, 2006

ISBN 88-901551-2-4

INDICE

INTRODUZIONE	1
<i>Andrea Giommi</i>	
SESSIONE I	3
L'informazione territoriale	
<i>Barbara Cafarelli, Alessio Pollice</i>	5
I modelli geo-additivi per l'analisi del grado di salinità di un suolo	
<i>Roberto Benedetti, Paolo Postiglione, Giuseppe Espa</i>	15
Metodi statistici per la stima della copertura del suolo con dati a terra e da satellite	
SESSIONE II	27
Famiglia e lavoro	
<i>Luca Bartoli, Laura Palombo</i>	29
Struttura e comportamenti lavorativi delle famiglie agricole	
<i>Velia Bartoli</i>	41
Il rischio di morte per i lavoratori nel settore agricolo in Italia	
SESSIONE III	51
Unità di Firenze	
<i>Iacopo Bernetti, Chiara Franciosi, Ginevra Virginia Lombardi</i>	53
Impatto della riforma della Politica Agricola Comunitaria sull'uso del suolo e la multifunzionalità del settore agricolo	
<i>Chiara Bocci, Alessandra Petrucci, Emilia Rocco</i>	67
L'utilizzo della vicinanza spaziale a fini predittivi mediante la regressione geografica: un'applicazione alla stima per piccole aree in ambito agricolo.	

SESSIONE IV	77
Archivi e/o censimenti?	
<i>Bruno Massoli</i>	79
Verso il Censimento Agricolo 2010	
<i>Giuseppe Garofalo, Orietta Patacchia</i>	91
Il Registro statistico delle unità del settore dell'agricoltura: problematiche e soluzioni possibili per la sua realizzazione	
SESSIONE V	99
La qualità dei dati	
<i>Aurora De Santis, Marcello D'Orazio, Giuseppe Lecardane, Giovanna Brancato</i>	101
Aree di intervento per il miglioramento dell'accuratezza dei dati sulla consistenza degli allevamenti	
<i>Ugo Guarnera, Orietta Luzi</i>	109
Valutazione del trattamento degli errori di misura e di risposta nell'indagine SPA	
<i>Piero Demetrio Falorsi, Matteo Mazziotta, Fabrizio Solari</i>	119
La stima degli errori di misura nelle indagini sull'agricoltura: uno studio sull'indagine di controllo della SPA 2003	
SESSIONE VI	131
Indicatori ambientali e criteri di intervento	
<i>Giampaola Bellini</i>	133
Agricultural practices and agrienvironmental indicators: data availability	
<i>Rocio Cándor Golec, Marina Vitullo, Riccardo De Lauretis</i>	141
Contribution of ISTAT statistics to the national air emission inventory for the agriculture sector	
<i>Giampaola Bellini, Teresa Di Sarro</i>	153
Criteri d'intervento nella difesa fitosanitaria dei fruttiferi	
<i>Luca Salvati, Tomaso Ceccarelli, Antonio Brunetti</i>	159
Desertificazione, cambiamenti climatici e agricoltura in Italia: primi risultati di un modello di valutazione del rischio di desertificazione	

SESSIONE VII	167
Unità di Cassino	
<i>Marcello De Rosa, Carlo Russo, Massimo Sabbatini</i>	169
Le strategie emergenti delle imprese agricole: il settore primario fra multifunzionalità e industrializzazione	
<i>Luca Bartoli, Marcello De Rosa</i>	185
Aspetti economici e demografici nella dinamica delle famiglie agricole	
SESSIONE VIII	197
Unità di Pisa	
<i>Monica Pratesi, Nicola Salvati</i>	199
Regressione M-quantilica per la stima della funzione di ripartizione per piccola area. Il caso della produzione di olive in Toscana	
<i>Stefano Marchetti, Monica Pratesi, Nicola Salvati</i>	211
La stima per piccole aree non rappresentate nel campione: il caso della produzione di uva nei comuni della Toscana	
SESSIONE IX	221
Nuove prospettive	
<i>Felicetta Carillo, Adele Coppola, Giuseppe Marotta</i>	223
La riforma della Politica Agricola Comunitaria: un percorso metodologico per l'analisi dell'impatto in Campania	
<i>Mario Adua</i>	235
Riflessioni e proposte per il prossimo Censimento Agricolo del 2010	
<i>Marco Platania, Donatella Privitera</i>	245
L'applicazione delle tecnologie informatiche in campo agricolo: un'analisi in Calabria	
<i>Aurora De Santis, Cecilia Manzi, Stefano Salvi</i>	257
Due nuovi progetti nelle statistiche zootecniche dell'ISTAT: l'indagine di consistenza del bestiame avicunicolo e uno studio di fattibilità sull'utilizzo dell'anagrafe bovina	
<i>Piero Augusto Nasuelli, Vittorio Vetrano</i>	269
La riforma della PAC e le rilevazioni statistiche nel comparto delle produzioni animali	

SESSIONE X	281
Modelli e metodi di stima	
<i>Annalisa Pallotti</i>	283
Modelli di alberi di regressione in agricoltura: un'applicazione con i dati sull'utilizzo di prodotti fitosanitari su singola coltivazione	
<i>Marco Ballin, Mauro Scanu, Paola Vicard</i>	295
Uso di reti bayesiane per la costruzione di stimatori in popolazioni finite	
<i>Edoardo Pizzoli, Giuseppe Lecardane</i>	303
Modelli di simulazione in agricoltura: un'applicazione con i dati dell'indagine sui risultati economici delle aziende agricole (REA), dell'indagine sulla struttura e produzione delle aziende agricole (SPA) e della contabilità nazionale	
<i>Teresa Amodio, Giuseppe Marotta</i>	317
Una proposta metodologica per l'analisi dei sistemi territoriali agroalimentari e rurali (STAR)	
 SESSIONE XI	 329
La multifunzionalità	
<i>Renato Salvatore</i>	331
Metodi di ottimizzazione per la valutazione della multifunzionalità delle aziende agricole	
<i>Concetta Cardillo, Carlo Russo, Massimo Spigola</i>	343
Il paradigma Strategia-Struttura nelle imprese agricole a conduzione diretta: un'analisi esplorativa delle strategie emergenti	
<i>Loredana De Gaetano</i>	355
Dall'azienda agricola all'impresa agricola: il caso della multifunzionalità delle aziende agricole fiorentine	
<i>Concetta Cardillo, Annalisa Grassi, Massimo Greco</i>	365
La multifunzionalità, punto di forza o debolezza del sistema agricolo nazionale?	
<i>Edoardo Pizzoli, Veronica Rondinelli</i>	377
Multifunzionalità dell'agricoltura e agricoltura biologica: una analisi della struttura e delle performance aziendali sulla base dei dati dell'indagine sui risultati economici delle aziende agricole (REA) e dei dati statistici e amministrativi disponibili	

SESSIONE XII	389
Aspetti economici	
<i>Andrea Brugnoli</i>	391
Delimitazione del campo di osservazione dei censimenti e delle indagini sulla struttura e produzioni delle aziende agricole negli stati membri dell'unione europea	
<i>Giacomo Branca, Antonio Macrì</i>	403
Strutture agrarie e forme giuridiche: i problemi della rilevazione a fini statistici delle proprietà collettive e delle connesse forme di gestione.	

L'UTILIZZO DELLA VICINANZA SPAZIALE A FINI PREDITTIVI MEDIANTE LA REGRESSIONE GEOGRAFICA: UN'APPLICAZIONE ALLA STIMA PER PICCOLE AREE IN AMBITO AGRICOLO

Chiara Bocci

*Dipartimento di Statistica "G. Parenti",
Università degli Studi di Firenze
bocci@ds.unifi.it*

Alessandra Petrucci

*Dipartimento di Statistica "G. Parenti",
Università degli Studi di Firenze
alex@ds.unifi.it*

Emilia Rocco

*Dipartimento di Statistica "G. Parenti",
Università degli Studi di Firenze
rocco@ds.unifi.it*

Abstract: Per i fenomeni che si presentano sul territorio con caratteristiche di non stazionarietà, un modello "globale" quale il modello di regressione lineare classico può indurre a conclusioni sbagliate. Di qui la necessità di ricorrere a tecniche, quale il modello di regressione geografica pesata (GWR), che consentono di modellare una relazione tra variabili che varia nello spazio. La regressione geografica pesata, nel lavoro, viene utilizzata per stimare, a livello comunale, la superficie investita a vite nella provincia di Firenze. A tal fine si impiega l'informazione relativa alla collocazione sul territorio delle aziende agricole introdotta dal Quinto Censimento Italiano dell'Agricoltura, insieme ai dati provenienti da un campione con caratteristiche analoghe a quello dell'indagine corrente sulla struttura e produzione delle aziende agricole (SPA) condotta dall'ISTAT.

Parole chiave: stima per piccole aree, dati spaziali, non stazionarietà spaziale, regressione geografica pesata

1. Introduzione

La possibilità di collocare le aziende agricole sul territorio è uno degli elementi innovativi che ha caratterizzato il Quinto Censimento Italiano dell'Agricoltura del 2000. Questa ulteriore informazione apre nuove sfide e pone nuovi interrogativi sull'analisi statistica dei dati raccolti. In particolare la georeferenziazione di ciascuna azienda

agricola, come evidenziato anche in lavori precedenti (Lipizzi, 2004; Bocci *et al.*, 2005), può costituire una informazione particolarmente utile per l'analisi di molti fenomeni di interesse in campo agricolo.

In particolare, con questo contributo si vuole evidenziare l'utilità della collocazione spaziale delle aziende agricole per l'analisi di fenomeni ad esse connessi che sono non stazionari sul territorio. L'applicazione a tali fenomeni di un modello "globale", quale il modello di regressione lineare classica, che assume l'indipendenza del fenomeno dalla collocazione spaziale dell'azienda su cui è rilevato può indurre a conclusioni sbagliate e generare residui con una notevole correlazione spaziale. Da queste considerazioni nasce l'idea di ricorrere a tecniche statistiche in grado di modellare il fenomeno tenendo direttamente conto della sua variabilità spaziale. Ovviamente l'applicabilità di tali tecniche è subordinata alla disponibilità di informazioni sulla collocazione sul territorio delle unità della popolazione di studio.

Nel lavoro si prende in esame una specifica tecnica, nota come Geographically Weighted Regression (GWR), che modella localmente la non stazionarietà riscontrata nel fenomeno oggetto d'analisi. Tale tecnica viene utilizzata per ottenere la stima a livello comunale della superficie investita a vite nella provincia di Firenze, impiegando i dati provenienti da un campione con caratteristiche analoghe al campione dell'indagine corrente sulla struttura e produzione delle aziende agricole (SPA) condotta dal ISTAT.

Il lavoro è organizzato come segue: nel secondo paragrafo viene riportata una sintesi del metodo GWR. Nel terzo paragrafo vengono presentati i risultati e viene discussa l'applicazione del metodo GWR alla stima della superficie coltivata a vite. Infine l'ultimo paragrafo contiene le conclusioni e le indicazioni del lavoro futuro.

2. La regressione geografica pesata

La regressione geografica pesata (GWR) estende il tradizionale modello di regressione inserendo nel modello stesso la relazione spaziale tra le unità. Specificatamente trasforma il modello di regressione globale

$$y_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} + \varepsilon_i \quad i = 1, \dots, n$$

nel modo seguente:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} + \varepsilon_i, \quad i = 1, \dots, n \quad (1)$$

ovvero viene sostituito al k -esimo parametro globale β_k , il k -esimo parametro locale $\beta_k(u_i, v_i)$ che è funzione delle coordinate (u_i, v_i) dell' i -esima unità di studio.

E' importante sottolineare che i parametri locali dipendono solo dalle coordinate dell'unità e non da altre informazioni ad essa relative e possono essere pertanto stimati in qualsiasi punto dello spazio, anche se non è campionato; possiamo quindi, in modo più generale, definire il generico parametro $\beta_k(u_i, v_i)$ come una realizzazione nel punto i della funzione continua $\beta_k(u, v)$.

I parametri locali $\beta_k(u_i, v_i)$ sono stimati usando la procedura dei minimi quadrati pesati. I pesi w_{ij} , per $j = 1, \dots, n$, in ciascun punto (u_i, v_i) sono ottenuti come funzione continua della distanza tra l'unità i e l'unità j .

Sia

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0(u_1, v_1) & \beta_1(u_1, v_1) & \cdots & \beta_p(u_1, v_1) \\ \vdots & \vdots & \ddots & \vdots \\ \beta_0(u_n, v_n) & \beta_1(u_n, v_n) & \cdots & \beta_p(u_n, v_n) \end{bmatrix} \quad (2)$$

la matrice dei parametri locali. Ciascuna riga si ottiene come

$$\hat{\boldsymbol{\beta}}(i) = (\mathbf{X}^T \mathbf{W}(i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(i) \mathbf{y}, \quad (3)$$

dove $i = 1, \dots, p$ indica l' i -esima riga della matrice (2), \mathbf{X} è la matrice delle variabili esplicative, \mathbf{y} è la variabile dipendente e $\mathbf{W}(i)$ è la matrice $n \times n$ dei pesi w_{ij} (spatial weighting matrix) della forma

$$\mathbf{W}(i) = \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{bmatrix}. \quad (4)$$

Lo stimatore in (3) è uno stimatore dei minimi quadrati pesati, ma la matrice dei pesi $\mathbf{W}(i)$ non è costante ma deve essere ricalcolata per ogni i . Infatti, per ogni unità i i pesi sono funzione della vicinanza delle altre unità ad i e le unità più vicine hanno maggiore peso nella stima dei parametri $\boldsymbol{\beta}(i)$ rispetto a quelle più lontane.

La scelta della struttura dei pesi $\mathbf{W}(i)$ è quindi un punto fondamentale della procedura GWR. Ovviamente nel caso in cui il sistema dei pesi sia tale che

$$w_{ij} = 1 \quad \forall i, j$$

il modello coincide con quello di regressione tradizionale.

Un primo passo per tener conto della struttura locale del fenomeno è quello di escludere tutti i punti che hanno rispetto all'unità i una distanza maggiore di un valore prefissato d , assegnando una funzione dei pesi del tipo:

$$w_{ij} = \begin{cases} 1 & \text{se } d_{ij} < d \\ 0 & \text{altrimenti,} \end{cases} \quad (5)$$

dove d_{ij} è la distanza tra i e j .

Questa funzione di peso, tuttavia, essendo discontinua, presenta l'inconveniente di provocare un cambiamento repentino della stima dei coefficienti di regressione a seconda del gruppo di dati definito dalla distanza d .

Un modo per ovviare a questo problema è quello di specificare una funzione di peso di tipo continuo. Una funzione delle più comuni è la funzione Gaussiana

$$w_{ij} = \exp\left[-1/2(d_{ij}/b)^2\right], \quad (6)$$

dove il parametro b (*bandwidth*) definisce l'insieme di osservazioni attorno al punto i che contribuiscono alla stima locale. In un caso estremo, corrispondente a $b = 0$, solo le unità che hanno le stesse coordinate di i vengono considerate. All'estremo opposto, ovvero quando $b = \infty$, la stima non è più locale, ma globale.

Una funzione alternativa è la bi-quadratica

$$w_{ij} = \begin{cases} \left[1 - (d_{ij}/b)^2\right]^2 & \text{se } d_{ij} < b \\ 0 & \text{altrimenti,} \end{cases} \quad (7)$$

particolarmente utile perché produce una funzione continua con peso simile a quello gaussiano fino alla distanza b dal punto i e con peso zero per ogni unità con distanza maggiore di b (Fotheringham *et al.*, 2002).

Le funzioni di peso fin qui esaminate sono fisse in termini di forma e di dimensione su tutto lo spazio, presentano infatti un *bandwidth* b fisso. Un problema che può nascere con questo tipo di funzioni è che se le unità non sono tutte equamente distribuite nella regione di analisi le stime prodotte potrebbero essere distorte nelle zone molto dense di osservazioni e eccessivamente variabili nelle zone con poche osservazioni.

Entrambi i problemi possono essere risolti utilizzando una funzione *kernel* variabile nello spazio. In questo modo il *bandwidth* si adatterà in funzione della densità delle unità: risulterà più piccolo nelle zone con alta densità di unità e più grande nelle zone a bassa densità.

Un metodo per costruire una funzione di peso adattiva è quello di considerare fisso non il *bandwidth*, ma il numero di vicini N che voglio includere nella stima dei parametri locali.

Una funzione bi-quadratica con *bandwidth* adattivo, che tenga quindi conto solo degli N -esimi vicini più vicini ad i è

$$w_{ij} = \begin{cases} \left[1 - (d_{ij}/b)^2\right]^2 & \text{se } j \text{ è uno dei } N - \text{esimi vicini più vicini ad } i \text{ e} \\ & b \text{ è la distanza del } N - \text{esimo vicino} \\ 0 & \text{altrimenti.} \end{cases} \quad (8)$$

Questa volta, la calibrazione del modello coinvolge la scelta di N , cioè il numero di unità che devono essere incluse nella stima dei parametri locali.

La scelta del *bandwidth* o del numero di vicini deve tener conto del *trade-off* tra varianza e distorsione delle stime dei parametri $\beta(u,v)$: se si considerano solo punti troppo vicini al punto (u,v) si ottengono stime con poca variabilità ma molta distorsione, al contrario se si considerano anche punti molto distanti da (u,v) le stime sono sostanzialmente corrette ma molto più variabili.

Un criterio per scegliere il *bandwidth* più appropriato (o il valore più appropriato di N) può essere quello detto *cross-validation* che prevede di scegliere il valore b che minimizza la quantità

$$CV = \sum_{i=1}^n [y_i - \hat{y}_{\neq i}(b)]^2, \quad (9)$$

dove $\hat{y}_{\neq i}(b)$ è il valore predetto di y_i ottenuto escludendo l'osservazione i -esima dal processo di calibrazione.

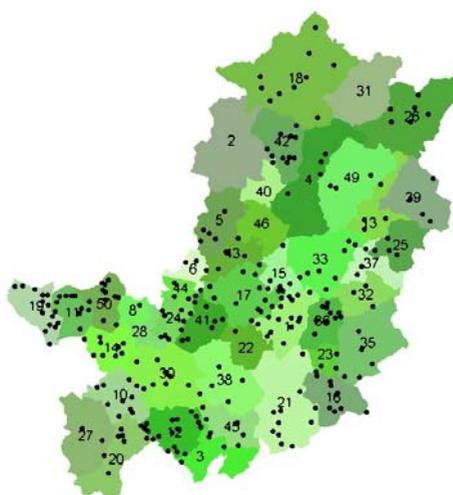
3. Applicazione

Per rispondere all'esigenza crescente di disporre a livello comunale di dati relativi alla superficie coltivata e alla produzione almeno per le principali colture che interessano il territorio, il modello GWR è stato utilizzato per stimare a livello comunale la superficie coltivata a vite nella provincia di Firenze. La vite insieme all'olivo, infatti, è la principale coltura in tale territorio. In questo lavoro ci limitiamo a prendere in esame la superficie coltivata mentre non consideriamo la produzione perché per quest'ultima non disponiamo di dati reali (di origine censuaria) o di altre stime attendibili con cui confrontare i nostri risultati.

Il modello è stato applicato ad un campione casuale semplice di 300 aziende (figura 1) selezionato dalla lista delle aziende agricole della provincia di Firenze quale risulta in base ai dati del quinto censimento dell'agricoltura. La dimensione campionaria è stata fissata in 300 aziende perché questa è la dimensione del campione utilizzato nell'indagine sulla produzione delle aziende agricole condotta dall'ISTAT nel 2003 (Ballin *et al.*, 2003). Il campione dell'indagine SPA, che viene condotta dall'ISTAT con cadenza biennale, infatti, potrebbe essere secondo noi quello da utilizzare in una potenziale applicazione reale della metodologia proposta. L'indagine SPA non utilizza un campione casuale semplice ma un campione stratificato, tuttavia i criteri di stratificazione definiti per tale indagine al fine di ottenere stime attendibili a livello regionale rendono il campione provinciale simile ad un campione casuale semplice.

Figura 1: Distribuzione del campione di 300 aziende sul territorio provinciale

Cod	Comune	Cod	Comune
01	Bagno a Ripoli	25	Londa
02	Barberino di Mugello	26	Marradi
03	Barberino Val D'elsa	27	Montaione
04	Borgo San Lorenzo	28	Montelupo Fiorentino
05	Calenzano	30	Montespertoli
06	Campi Bisenzio	31	Palazzuolo sul Senio
08	Capraia e Limite	32	Pelago
10	Castelfiorentino	33	Pontassieve
11	Cerreto Guidi	35	Reggello
12	Certaldo	36	Rignano sull'arno
13	Dicomano	37	Rufina
14	Empoli	38	San Casciano in Val di Pesa
15	Fiesole	39	San Godenzo
16	Figline Valdarno	40	San Piero a Sieve
17	Firenze	41	Scandicci
18	Firenzeuola	42	Scarperia
19	Fucecchio	43	Sesto Fiorentino
20	Gambassi Terme	44	Signa
21	Greve in Chianti	45	Tavarnelle Val di Pesa
22	Impruneta	46	Vaglia
23	Incisa in Val D'arno	49	Vicchio
24	Lastra a Signa	50	Vinci

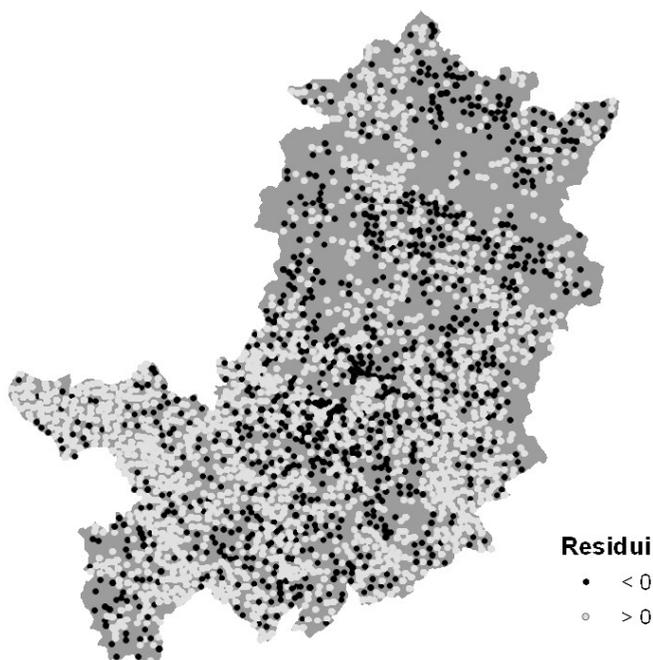


Dopo aver individuato, con una prima analisi esplorativa dei dati, come uniche variabili esplicative significative la superficie agricola utilizzata (SAU) e l'unità di dimensione europea¹ (UDE), per verificare la presunta non stazionarietà spaziale del fenomeno abbiamo applicato al campione un modello di regressione lineare classico. Nonostante i parametri del modello (tabella 1) siano risultati significativi, come ci aspettavamo la distribuzione dei residui non è casuale sul territorio (figura 2).

Tabella 1: Parametri stimati con il modello di regressione lineare classico

Parametri Stimati				
Variabile	Stima	S.E.	Valore t	Pr > t
Intercetta	-31.764	17.319	-1.83	0.0676
SAU	-0.064	0.015	-4.16	<.0001
UDE	19.473	1.260	15.46	<.0001

Figura 2: Distribuzione dei residui del modello di regressione lineare classico sul territorio provinciale



Sostenuti da tale risultato siamo passati a definire non un unico modello di regressione, ma un modello per ciascuna azienda non inclusa nel campione, ovvero abbiamo applicato la tecnica GWR per la quale ogni modello è diverso dall'altro per i

¹ L'unità di dimensione europea (UDE) di una azienda agricola corrisponde a 1200 € per ciascuna unità di Reddito Lordo Standard (RLS). Il RLS è definito a sua volta come la differenza tra il valore standard della produzione e l'importo standard di alcuni costi specifici. Tale differenza viene determinata per ogni singola produzione vegetale e animale in ciascuna regione. Il RLS totale dell'azienda agricola corrisponde alla somma dei valori ottenuti per ogni produzione moltiplicando il RLS unitario per il numero di unità rispettive (ISTAT, 2000).

valori assunti dai parametri che non sono unici ma realizzazione di una funzione continua e sono stimati per ciascuna azienda utilizzando un insieme di aziende “vicine” incluse nel campione.

Per misurare la vicinanza tra le aziende è stata utilizzata la funzione di distanza euclidea e sono state impiegate come coordinate spaziali dell’azienda quelle relative al centroide del foglio di mappa in cui cade il centro aziendale. Pur consapevoli dell’approssimazione, la scelta è stata obbligata dal fatto che il dato censuario si limita a segnalare il foglio di mappa in cui cade il centro aziendale e non le precise coordinate dello stesso.

Inoltre non è stato assegnato a ciascuna azienda vicina lo stesso peso ma per determinare il peso da assegnare a “ogni vicina” è stata utilizzata una funzione *kernel* bi-quadratica con *bandwidth* variabile basata sul criterio dei vicini più vicini (funzione (8)). Per la scelta del numero dei vicini è stato utilizzato il criterio *cross-validation*, visto in (9), che ha individuato come numero di vicini ottimale 17.

Dopo aver stimato per ciascuna azienda non campionata i parametri del corrispondente modello di regressione e averli utilizzati per predire la superficie coltivata a vite dall’azienda, le stime della superficie comunale sono state ottenute aggregando i valori delle aziende site nel comune medesimo.

I risultati (figura 3 e tabella 2 – colonne E e F) anche se non completamente positivi sono comunque incoraggianti:

- la media rispetto ai comuni dell’errore di stima valutato mediante la differenza relativa tra la stima ottenuta nel modo appena descritto e il dato censuario è pari circa al 34%;
- per oltre il 50% dei comuni tale errore è inferiore al 10%;
- per il 77% dei comuni l’errore è inferiore al 20%.

Gli stessi calcoli effettuati utilizzando il modello globale (tabella 2 – colonne G e H) portavano invece a delle stime comunali inaccettabili.

Figura 3: Differenza relativa tra le stime ottenute dal modello GWR e il dato censuario

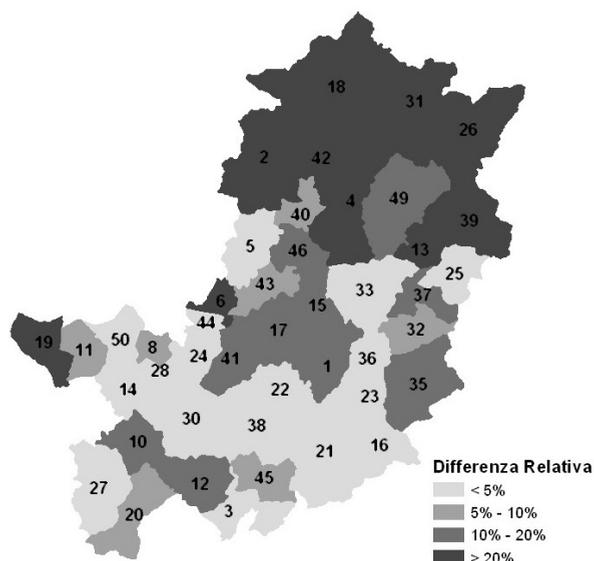
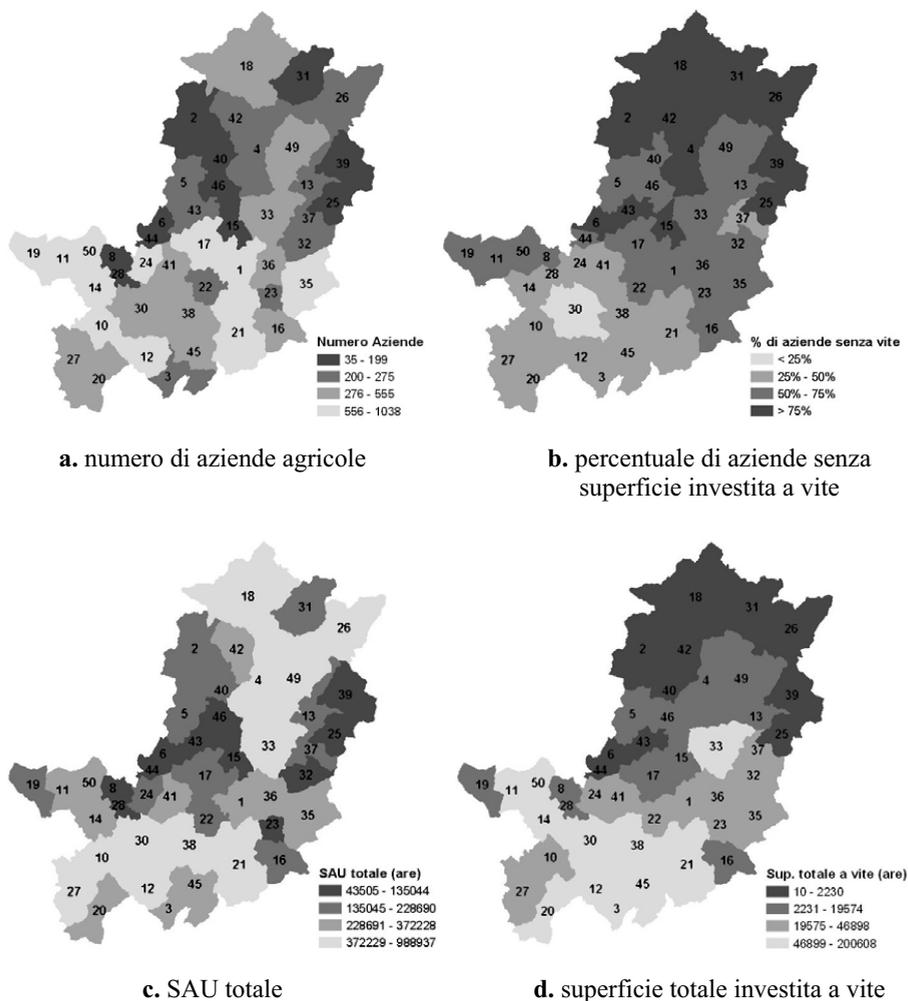


Tabella 2: Stime della superficie totale coltivata a vite per i comuni della provincia di Firenze

A	B	C	D	E	F	G	H
Comune	Codice ISTAT	Numero Aziende (Cens 2000) (*)	Superficie totale a Vite (Cens 2000) (*)	Stima GWR (N = 17)	Differenza Relativa %	Stima Reg. Globale	Differenza Relativa %
Bagno a Ripoli	01	700	31315	25661	-18.05%	62354	99.12%
Barberino di Mugello	02	72	1742	1201	-31.05%	13220	658.90%
Barberino Val d'Elsa	03	251	90099	91342	1.38%	75857	-15.81%
Borgo San Lorenzo	04	270	4786	7441	55.48%	32733	583.94%
Calenzano	05	230	6125	5949	-2.87%	19225	213.87%
Campi Bisenzio	06	199	697	1303	87.00%	7868	1028.81%
Capraia e Limite	08	161	18283	19391	6.06%	14872	-18.66%
Castelfiorentino	10	598	46898	53774	14.66%	36890	-21.34%
Cerreto Guidi	11	1038	100549	106161	5.58%	43321	-56.92%
Certaldo	12	673	106311	120210	13.07%	63642	-40.14%
Dicomano	13	232	13166	10438	-20.72%	13802	4.83%
Empoli	14	845	68826	70485	2.41%	41364	-39.90%
Fiesole	15	162	4490	5324	18.57%	20072	347.03%
Figline Valdarno	16	351	11702	11144	-4.77%	16062	37.25%
Firenze	17	736	8304	9410	13.32%	61182	636.78%
Firenzuola	18	401	10	90	801.06%	3603	35933.48%
Fucecchio	19	915	16407	24136	47.11%	92	-99.44%
Gambassi Terme	20	345	74620	70064	-6.11%	51396	-31.12%
Greve in Chianti	21	678	200560	209962	4.69%	137137	-31.62%
Impruneta	22	275	23434	24117	2.92%	36691	56.57%
Incisa in Val d'Arno	23	236	20259	20721	2.28%	12751	-37.06%
Lastra a Signa	24	612	41653	40036	-3.88%	31851	-23.53%
Londa	25	183	2230	2172	-2.60%	1441	-35.40%
Marradi	26	223	1692	1025	-39.44%	19719	1065.43%
Montaione	27	305	35967	36305	0.94%	31718	-11.81%
Montelupo Fiorentino	28	136	19574	19349	-1.15%	15378	-21.44%
Montespertoli	30	467	197531	197007	-0.27%	146577	-25.80%
Palazzuolo sul Senio	31	90	49	77	56.82%	8030	16287.31%
Pelago	32	209	41115	44638	8.57%	32162	-21.78%
Pontassieve	33	476	72828	74417	2.18%	80093	9.98%
Reggello	35	821	27928	31314	12.12%	35520	27.18%
Rignano sull'Arno	36	555	46102	44122	-4.29%	38240	-17.05%
Rufina	37	243	28873	24051	-16.70%	17042	-40.97%
San Casciano in V. di P.	38	515	183840	179739	-2.23%	149668	-18.59%
San Godenzo	39	101	61	111	81.73%	3407	5485.06%
San Piero a Sieve	40	35	1024	942	-7.99%	12219	1093.25%
Scandicci	41	445	34775	39772	14.37%	47120	35.50%
Scarperia	42	254	2116	1468	-30.65%	23920	1030.45%
Sesto Fiorentino	43	232	1896	1788	-5.69%	12927	581.80%
Signa	44	80	1756	1820	3.63%	10430	493.96%
Tavarnelle Val di Pesa	45	284	60431	54990	-9.00%	46731	-22.67%
Vaglia	46	81	3396	3863	13.76%	2235	-34.19%
Vicchio	49	326	7551	6353	-15.87%	34326	354.58%
Vinci	50	761	96181	93446	-2.84%	57253	-40.47%

Nota: (*) Dato calcolato solo per 16802 aziende (le aziende presenti al Censimento 2000 erano 17045 ma di queste 243 non avevano il riferimento spaziale)

Figura 4: Dati dal Quinto Censimento dell'Agricoltura 2000 a livello comunale per la provincia di Firenze



Si può inoltre rilevare che i risultati peggiori si hanno in corrispondenza di quei comuni in cui è forte la componente agricola (figure 4.a e 4.c), ma è poco presente la coltura della vite (figure 4.b e 4.d).

4. Conclusioni

I risultati ottenuti non consentono di trarre delle considerazioni conclusive, tuttavia essi costituiscono il punto di avvio per uno studio più approfondito dell'applicazione di modelli locali ai dati agricoli. In particolare, due sono gli obiettivi principali che ora

interessa perseguire:

- affiancare alle stime un indice di variabilità ovvero individuare non solo delle stime puntuali ma anche un intervallo di stima;
- applicare e valutare un modello di regressione che utilizzi anche variabili ausiliarie fino ad ora non disponibili quali la stessa superficie coltivata a vite rilevata in altre occasioni, come ad esempio da precedenti censimenti; e/o variabili che descrivono le caratteristiche socio-demografiche ed economiche dell'area in cui l'azienda si trova.

Un altro aspetto da approfondire potrebbe essere l'uso di una più precisa collocazione spaziale dei dati. Nell'applicazione fatta è stata considerata come unità di osservazione l'azienda agricola e come sue coordinate spaziali non quelle effettive ma quelle relative al centroide del foglio di mappa in cui cade il centro aziendale. A questa approssimazione si aggiunge il fatto che il centro aziendale potrebbe essere lontano dall'area coltivata dall'azienda. Infatti il foglio di mappa del centro aziendale potrebbe non contenere tutti gli appezzamenti appartenenti all'azienda, i quali al contrario potrebbero ricadere in diversi fogli di mappa non necessariamente contigui (ISTAT, 2000). Pertanto un passo ulteriore potrebbe essere quello di utilizzare l'informazione relativa alla collocazione spaziale di tutti gli appezzamenti relativi all'azienda. Tale informazione è fornita solo parzialmente dalla sezione IX del censimento che associa a ciascun foglio di mappa solamente le superfici totali per le aggregazioni delle tipologie di utilizzo dei terreni definite nella sezione II. Va sottolineato inoltre che l'uso di tale informazione, qualora si rendesse disponibile, ci indurrebbe a considerare come unità di osservazione non l'azienda ma ogni singolo appezzamento che la costituisce e questo comporterebbe sicuramente un appesantimento nei calcoli che potrebbe essere tale da non essere giustificato dal presunto miglioramento dei risultati.

Riferimenti Bibliografici

- Ballin M., Salvi S. (2003), *Nota metodologica sul piano di campionamento adottato per l'indagine "Struttura e produzione delle aziende agricole 2003"*, ISTAT. On line: <http://www.istat.it/strumenti/rispondenti/imprese/strutturaazagr/strutturaazagr.htm> (ultima consultazione: luglio 2005).
- Bocci C. Petrucci A., Rocco E. (2005), L'uso di informazioni di tipo spaziale in ambito agricolo: un'applicazione relativa al comune di Fiesole, *Atti del convegno "Metodi d'Indagine e di Analisi per le Politiche Agricole"*, Università di Pisa, 21-22 ottobre 2004.
- Fotheringham A.S., Brunson C., Charlton M. (2002), *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, John Wiley & Sons, Chichester, UK.
- ISTAT (2000), *Piano generale del Quinto Censimento dell'Agricoltura: 22 ottobre 2000*, ISTAT, Roma.
- Lipizzi, F. (2004), L'integrazione dei disegni territoriali del Censimento della popolazione e degli edifici e del Censimento dell'agricoltura, *Atti del Convegno "L'Informazione Statistica e le Politiche Agricole (ISPA2004)"*, Università di Cassino, 6 maggio 2004.