

Dipartimento di Scienze Economiche, Matematiche e Statistiche

Università degli Studi di Foggia

I modelli geo-additivi per l'analisi del grado di salinità di un suolo

Barbara Cafarelli, Alessio Pollice

Quaderno n. 17/2005

Quaderno riprodotto al
Dipartimento di Scienze Economiche, Matematiche e Statistiche
nel mese di luglio 2005 e
depositato ai sensi di legge

Authors only are responsible for the content of this preprint.

Dipartimento di Scienze Economiche, Matematiche e Statistiche, Via IV
Novembre, 1, 71100 Foggia (Italy), Phone +39 0881-77.61.97, Fax +39 0881-
77.56.16

I modelli geo-additivi per l'analisi del grado di salinità di un suolo

Barbara Cafarelli

*Dipartimento di Scienze Economiche, Matematiche e Statistiche, Università degli Studi di Foggia, Via IV Novembre 1, 71100 Foggia,
b.cafarelli@unifg.it*

Alessio Pollice

*Dipartimento di Scienze Statistiche, Università degli Studi di Bari, Via Camillo Rosalba 53, 70124 Bari,
apollice@dss.uniba.it*

Riassunto: In questo studio un modello geo-additivo è stato utilizzato per analizzare la distribuzione spaziale del tasso di assorbimento del sodio nella zona costiera di Muravera-Villaputzu (CA) per individuare le cause che ne hanno determinato l'eccessivo grado di salinità, dannoso per le colture agrumicole praticate nella zona. I modelli geo-additivi rappresentano un'evoluzione del kriging universale e consentono di considerare esplicitamente i legami non lineari tra la risposta e le covariate e la correlazione spaziale descritta mediante una funzione di autocorrelazione.

1. Introduzione

L'elevato grado di salinità di un suolo può essere imputato a cause naturali (intrusione marina, lunghi periodi di siccità, ecc.) e a cause secondarie (uso improprio del terreno: eccessivo emungimento dai pozzi a fini irrigui, idropotabili e turistici, lavori di regimazione di fiumi, ecc.) e può avere ripercussioni negative in campo economico ed ambientale. Infatti la perdita di produttività, che deriva da una eccessiva presenza di sale nel suolo può innescare processi di desertificazione, poiché un ambiente via via più arido e privo di diversità biologica è destinato all'abbandono (Castrignanò e Puddu, 2004).

Da quanto detto risulta evidente l'importanza della determinazione dell'entità dell'eccesso di salinità di un suolo e delle cause che l'hanno generato, soprattutto nel contesto dell'agricoltura di precisione, strategia di gestione che, attraverso le tecnologie informatiche, utilizza dati di diversa natura come supporto alle decisioni nella conduzione della pratica agricola.

Alla base di questa strategia vi è lo studio delle caratteristiche fisiche, chimiche ed idrauliche dei suoli, utile alla comprensione dei processi che ne causano le variazioni nello spazio e nel tempo.

E' evidente come in questo contesto risultino particolarmente rilevanti le metodologie che consentono di tenere in debito conto la variabilità spaziale e temporale associata a tutti gli aspetti della produzione agricola.

L'implementazione di tecniche di gestione del suolo e di politiche di conservazione delle risorse naturali può dunque essere fortemente agevolata dall'uso delle metodologie proprie della Statistica Spaziale, che permettono altresì di calibrare le diverse pratiche operative in base al tipo di coltura praticata ed alle particolari condizioni del suolo, facilitando le scelte relative alla gestione differenziata dell'azienda agricola (Goovaerts, 1999).

La stima della variabilità spaziale dei processi caratterizzanti un determinato territorio e la conoscenza del grado di accuratezza di tali stime consentono di ottenere delle mappe spaziali attendibili relative alle diverse proprietà e caratteristiche dei suoli, sulle quali basare le eventuali decisioni o pianificazioni (Castrignanò, 1996).

Per sviluppare una conduzione agronomica localizzata, in presenza di aree aride o semiaride o rese tali dagli interventi dell'uomo, la corretta valutazione delle cause e dell'entità dell'eccesso di salinità diviene di fondamentale importanza. In ambito agronomico la costruzione di mappe relative alla distribuzione di alcuni indicatori della salinità (tasso di assorbimento del sodio, la conducibilità elettrica dell'estratto di saturazione) in presenza di altre variabili esplicative del fenomeno (altre caratteristiche dei suoli) è spesso ottenuta applicando il kriging universale (Pozdnyakova e Zhang, 1999). Tuttavia, la complessità delle relazioni esistenti tra i suddetti indicatori e le variabili antecedenti rende necessario che l'assunzione di linearità alla base di una corretta applicazione del suddetto predittore, sia verificata di volta in volta (Lemhann et al., 2003).

In questo lavoro si sono studiate le cause dell'elevata salinità dei terreni della piana costiera di Muravera-Villaputzu (CA), zona rinomata per le colture agrumicole, in cui è stato registrato un notevole calo produttivo negli ultimi anni. Allo scopo è stato utilizzato un modello di dipendenza di tipo geo-additivo, non essendo plausibile l'assunzione di linearità tra l'indicatore di salinità prescelto ed alcune delle covariate rilevate.

2. Materiali e metodi

2.1 I dati di Muravera-Villaputzu

La rilevazione di campioni di suolo è stata effettuata in corrispondenza di 192 localizzazioni georeferenziate situate nella zona costiera di Muravera - Villaputzu. All'area di studio è stata sovrapposta una griglia a maglie quadrate di lato pari a 1 km, ogni quadrato è stato poi suddiviso in 16 subcelle quadrate di lato pari a 250 m, al cui interno sono stati effettuati i prelievi di suolo in modo che ve ne fosse almeno uno ogni 100 ha.

I punti campionari così individuati sono stati georeferenzati secondo il sistema Gauss-Boaga (Figura 1).

I campioni di suolo sono stati prelevati fino ad una profondità di 40 cm mediante una trivella e per ciascun campione sono state effettuate le seguenti misurazioni di laboratorio:

- caratteristiche fisiche del suolo o frazioni tessiturali (*limo grosso, limo fine, sabbia grossa, sabbia fine, argilla*);

- caratteristiche chimiche: livello di acidità (pH) e basi di scambio (ca , mg , k , na);
- misure del livello di salinità del suolo: conducibilità elettrica dell'estratto di saturazione, tasso di assorbimento del sodio (SAR), capacità di scambio cationico (CSC) e cationi del complesso di scambio (CS).

I dati analizzati sono stati rilevati e forniti dal Centro Regionale Agrario Sperimentale della Sardegna e l'analisi è stata condotta in collaborazione con l'Istituto Sperimentale Agronomico di Bari.

Per descrivere e analizzare la distribuzione spaziale del grado di salinità del suolo si è scelto il tasso di assorbimento del sodio (SAR), una variabile ricavata dalle quattro basi di scambio e considerata un indicatore efficiente del quantitativo di sali solubili presenti nel terreno ed in particolare del sodio scambiabile. La conoscenza della distribuzione spaziale del sodio scambiabile permette infatti di definire delle pratiche agricole di correzione degli eccessi di salinità mediante irrigazione o applicazioni di gesso idrato (Pozdnyakova e Zhang, 1999).

L'analisi è stata condotta sulla trasformata logaritmica del SAR .

Dall'analisi esplorativa della distribuzione spaziale del $\log(SAR)$ si nota che i livelli più elevati di sodio scambiabile si registrano nella zona più vicina al mare dell'area considerata (Figura 1) e che questi decrescono all'aumentare della distanza dalla costa: si passa da valori superiori a 32 del SAR in prossimità della costa, a valori inferiori a 5 dello stesso indicatore per distanze attorno ai 2000m. È bene sottolineare che queste variazioni, che sembrerebbero indicare una dipendenza del grado di salinità di questo suolo dalla distanza dal mare, si verificano con una certa continuità, fatta eccezione per un'area a nord-est e una più a sud dove i valori del SAR risultano più elevati ed una zona sulla costa dove i valori sono invece molto più bassi rispetto alle aree circostanti.

La presenza di tali aree, il cui livello di salinità, non essendo ascrivibile alla distanza dal mare, non trova apparente giustificazione in cause naturali, ha spinto a considerare la possibilità che essa possa dipendere dalle suddette cause secondarie (Castrignanò e Puddu, 2004).

2.2 I modelli geo-additivi.

La complessità della relazione tra il tasso di assorbimento del sodio e le variabili esplicative rilevate, ha portato a preferire l'utilizzo di un approccio semiparametrico, tramite un modello di dipendenza geo-additivo.

Questi modelli sono costituiti dalla somma di una componente semiparametrica nella forma di un modello additivo, che serve per esprimere il legame non lineare tra la risposta e le variabili esplicative, e di un modello lineare ad effetti misti (LME), con cui si esplicita la correlazione spaziale dei valori osservati.

Siano s_i e y_i , $i = 1, \dots, n$, rispettivamente i valori osservati di una variabile antecedente e della variabile risposta in corrispondenza della localizzazione spaziale \mathbf{x}_i , $\mathbf{x} \in \mathfrak{R}^2$, un modello geo-additivo può essere definito attraverso la seguente formulazione generale:

$$y_i = \beta_0 + f(s_i) + \beta_1^T \mathbf{x}_i + S(\mathbf{x}_i) + \varepsilon_i \quad (1)$$

dove $f(\cdot)$ è una funzione *smooth*, $S(\mathbf{x})$ è un processo stocastico gaussiano stazionario del secondo ordine con valore atteso nullo e varianza costante pari a σ_x^2 . Nella (1) la componente additiva, quella ad effetti misti e quella erratica ε , sono tra loro indipendenti.

Al fine di ottenere una versione della (1) che sia stimabile utilizzando le procedure relative ai modelli a effetti misti implementate nei software statistici standard, si è considerata una versione *low rank* sia della componente additiva sia di quella a effetti misti.

Per la prima componente si è utilizzata una funzione *penalized spline* della forma:

$$f(s_i) = \beta_s s_i + \sum_{k=1}^{K_s} u_k (s_i - \kappa_k)_+$$

considerando i coefficienti u_k dei nodi κ_k come degli effetti casuali (Wand, 2003). Il vettore di effetti casuali $\mathbf{u} = (u_1, \dots, u_{K_s})^T$ con $E(\mathbf{u}) = 0$ e $\text{cov}(\mathbf{u}) = \sigma_s^2 \mathbf{I}$ contiene tutti i coefficienti dei K_s nodi, mentre $\mathbf{Z}_s = ((s_i - \kappa_k)_+)_{1 \leq i \leq n, 1 \leq k \leq K_s}$ è la matrice disegno ($t_+ = \max(0, t)$, $t \in R$).

Allo scopo di ottenere un'analogia versione *low rank* della componente a effetti misti si è resa necessaria l'individuazione di un secondo insieme di nodi $\kappa_1, \dots, \kappa_K$ rappresentativo delle localizzazioni spaziali $\mathbf{x}_1, \dots, \mathbf{x}_n$ e la definizione delle matrici di correlazione $\mathbf{\Omega}_x = (h(\|\mathbf{x}_i - \kappa_k\|/\phi))_{1 \leq i \leq n, 1 \leq k \leq K}$ e $\mathbf{\Omega}_\kappa = (h(\|\kappa_k - \kappa_{k'}\|/\phi))_{1 \leq k, k' \leq K}$, costruite tramite la funzione di correlazione appartenente alla famiglia di Matérn (Stein, 1999),

$$h(r) = \sigma_x^2 (1 + |r|/\phi) \exp(-|r|/\phi) \quad (2)$$

La matrice $\mathbf{Z}_x = \mathbf{\Omega}_x \mathbf{\Omega}_\kappa^{-1/2}$ è considerata come matrice disegno riferita al vettore di effetti casuali $\tilde{\mathbf{u}}$ con $E(\tilde{\mathbf{u}}) = 0$ e $\text{cov}(\tilde{\mathbf{u}}) = \sigma_x^2 \mathbf{I}$.

Alla luce della formulazione *low rank* delle componenti additiva e a effetti misti, ponendo nella (1) $\mathbf{X} = (\mathbf{1}, s_i, \mathbf{x}_i^T)_{1 \leq i \leq n}$, $\tilde{\mathbf{u}} = (\mathbf{u}^T, \tilde{\mathbf{u}}^T)^T$ e $\mathbf{Z} = (\mathbf{Z}_s \mid \mathbf{Z}_x)$, si ottiene la formulazione *low rank* del modello geo-additivo (Kammann e Wand, 2003)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\tilde{\mathbf{u}} + \boldsymbol{\varepsilon} \quad \text{con} \quad \text{cov}\begin{pmatrix} \tilde{\mathbf{u}} \\ \boldsymbol{\varepsilon} \end{pmatrix} = \begin{bmatrix} \sigma_s^2 \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_x^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_\varepsilon^2 \mathbf{I} \end{bmatrix} \quad (3)$$

Nella (3) si assume che la componente erratica $\boldsymbol{\varepsilon}$ abbia media nulla, varianza costante pari a σ_ε^2 , e $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ per $i \neq j$, $j = 1, \dots, n$ e che la componente erratica, quella spazialmente correlata e quella additiva siano tra loro incorrelate. Inoltre la suddetta formulazione può essere facilmente estesa per comprendere più di un effetto non-lineare.

Il vantaggio fondamentale di questa formulazione della (1) consiste nella possibilità di ottenere le stime dei parametri del modello tramite REML, utilizzando la funzione `lme` di S-PLUS (libreria `nlme` di R), la funzione `spm` della libreria SemiPar di R o la `PROC MIXED` di SAS. Inoltre la rappresentazione *low rank*, fa sì che il numero delle funzioni di base utilizzate non aumenti con l'ampiezza campionaria (Hastie, 1996).

L'ammontare di *smoothing* del modello (3) con riferimento alla componente additiva e a quella spazialmente correlata è quantificato in base ai rapporti σ_e^2/σ_s^2 , e σ_e^2/σ_x^2 , calcolabili utilizzando le varianze stimate con REML.

Questi parametri di *smoothing* intervengono nella determinazione dei gradi di libertà della componente non lineare e permettono di controllare il *trade-off* tra l'adattamento del modello ai valori osservati (valori prossimi allo zero dei parametri di *smoothing*) e lo *smoothness* dello stesso (valori elevati dei suddetti parametri).

2.3 Analisi della distribuzione spaziale della salinità.

Al fine di individuare le cause dei livelli elevati di sodio scambiabile nell'area oggetto di studio è stato stimato un primo modello di dipendenza geo-additivo (3) in cui si è

- esplicitata la relazione non lineare tra il $\log(SAR)$ e il pH nella parte additiva,
- rappresentata la correlazione spaziale attraverso un modello lineare ad effetti misti,
- incluso l'effetto lineare dei cationi del complesso di scambio e della capacità di scambio cationico sul $\log(SAR)$.

Il numero di nodi equispaziati necessari per rappresentare l'effetto non lineare del pH , risultato pari a 23, è stato determinato come

$$K = \max\left(5, \min\left(\frac{1}{4} \times \text{numero di valori di } pH \text{ distinti}, 35\right)\right)$$

(Ngo et al., 2004) mentre il numero dei nodi per la componente spaziale, risultato pari a 48, e le relative coordinate geografiche sono state individuate mediante l'algoritmo di *space filling*, implementato nella funzione `default.knots.2D` della libreria SemiPar di R (Wand. et al., 2005).

Per la struttura di correlazione spaziale si è utilizzata la (2) con un valore del range ϕ pari a 5.636m, calcolato convenzionalmente come distanza massima tra i punti campionati (Kammann e Wand, 2003).

Questa scelta permette di ridurre il numero dei parametri del modello e non risulta particolarmente influente sui valori delle stime. Peraltro per l'applicazione in questione la determinazione del range è meno importante di quella delle componenti di varianza e dei parametri di *smoothing*, da cui dipendono i gradi di libertà della componente non-lineare e l'adattamento del modello stesso ai valori osservati (Nychka, 2000 e Kammann e Wand, 2003).

Il modello descritto è stato stimato tramite il metodo REML, utilizzando la funzione `spm` della libreria `SemiPar` di R.

Le stime degli effetti fissi della componente lineare del modello, fatta eccezione per l'intercetta, sono risultate significative (Tab. 1).

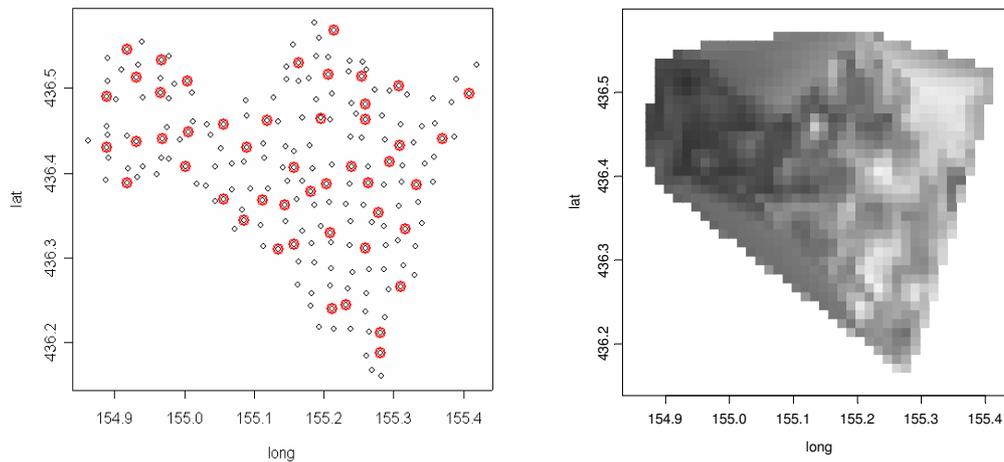


Figura 1: Localizzazioni campionate e posizionamento dei nodi; distribuzione spaziale del $\log(SAR)$.

Anche la parte non lineare del modello è risultata significativa, infatti i valori dei gradi di libertà (*df*) e dei parametri di *smoothing* (Tab. 1), mostrano la necessità di includere nel modello gli effetti non lineari del *pH* (Fig. 2) e, soprattutto, della componente spaziale (Wand et al., 2005).

Componente lineare				
Covariate	Coefficienti	Std.Error	Ratio	p-value
Intercetta	-609.50000	9.862e+02	-0.618	0.5369
<i>CSC</i>	0.02635	7.552e-03	3.490	0.0005
<i>CS</i>	0.01721	1.128e-03	15.260	0.0000
Componente non lineare				
Covariate	df	Parametri di smoothing	N° nodi	
<i>pH</i>	3.289	2.2200	23	
<i>long, lat</i>	13.130	0.2416	48	

Tabella 1: Stime degli effetti fissi della componente lineare del modello e dei gradi di libertà e del numero di nodi riferiti alla componente non lineare.

I valori delle previsioni ottenute tramite il modello stimato mostrano un andamento conforme a quello dei dati osservati (Fig 3). Anche l'analisi dei residui del modello stimato non evidenzia alcuna struttura spaziale degli stessi.

L'utilizzo delle componenti tessiturali del suolo (*sabbia*, *limo* ed *argilla*) come ulteriori variabili esplicative nel modello geo-additivo non ha portato risultati

significativi (Tab. 2). Sembra pertanto che la distribuzione del logaritmo del SAR nell'area in questione sia influenzata, piuttosto che dalla struttura del suolo, dalla componente geografica ed in particolare dalla longitudine, altresì assimilabile alla distanza dal mare.

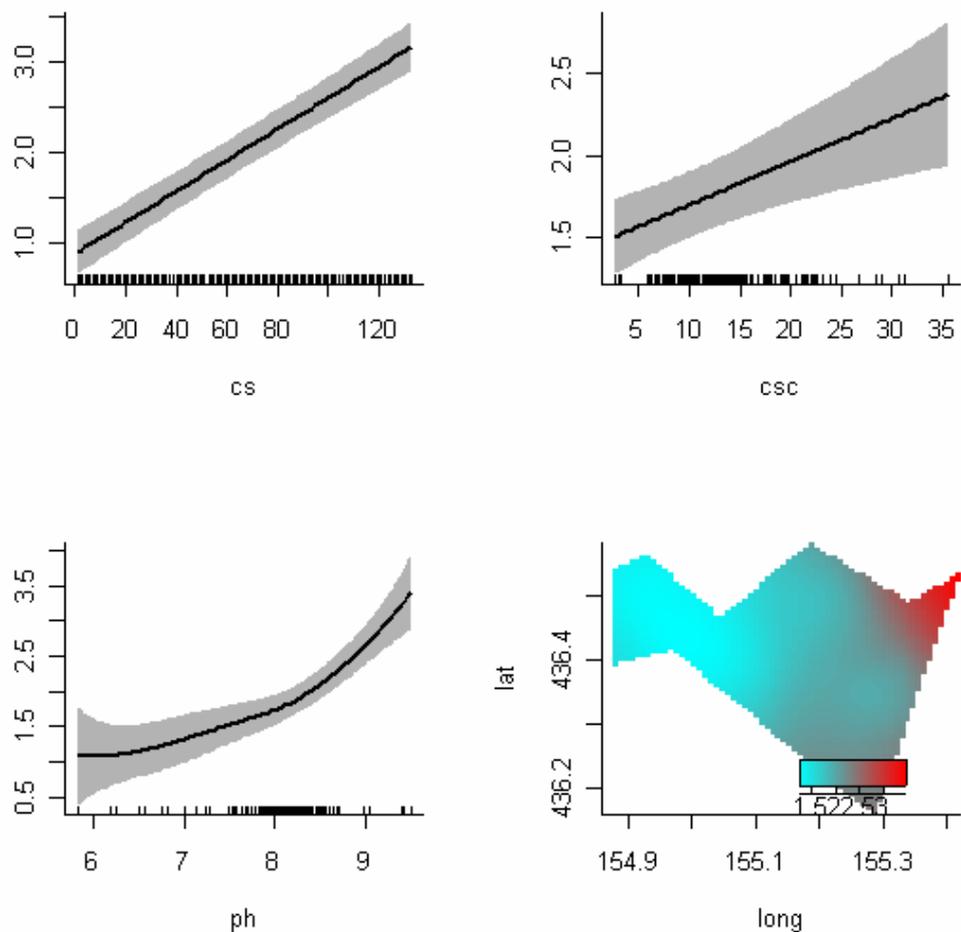


Figura 2: Effetti lineari del *CS* e del *CSC* ed effetti non-lineari del *pH* e della componente spaziale sul $\log(\text{SAR})$; bande di confidenza corrispondenti a ± 2 volte la deviazione standard delle stime

Questo porta a concludere che il calo produttivo nella coltura agrumicola registrato nell'ultimo decennio nella zona costiera di Muravera-Villaputzu, non è ascrivibile ad un uso improprio del terreno da parte dell'uomo bensì a fenomeni naturali quale quello dell'intrusione marina e ai periodi siccitosi che hanno caratterizzato l'area in questione nell'ultimo decennio.

3. Conclusioni

In questo lavoro è stata analizzata la distribuzione spaziale del tasso di assorbimento di sodio di un suolo utilizzato per colture agrumicole e le sue

relazioni con le altre variabili rilevate al fine di individuare le cause che hanno agito nel processo di salinizzazione della piana costiera di Muravera-Villaputzu, nella Sardegna sud-orientale. In quest'area, nell'ultimo decennio, si sono verificati fenomeni di intrusione marina e di contaminazione della falda freatica con acque

Componente lineare				
Covariate	Coefficienti	Std.Error	Ratio	p-value
Intercetta	-5.454e+02	8.723e+02	-0.6252	0.5320
<i>CSC</i>	3.849e-02	1.106e-02	3.4810	0.0005
<i>CS</i>	1.696e-02	1.132e-03	14.9800	0.0000
<i>SABBIA</i>	6.828e-03	1.377e-02	0.4957	0.6202
<i>LIMO</i>	5.854e-00	1.382e-02	0.4237	0.6719
<i>ARGILLA</i>	7.828e-03	1.373e-02	0.5702	0.5686
Componente non lineare				
Covariate	df	Parametri di smoothing	N° nodi	
<i>pH</i>	3.424	2.0540	23	
<i>long, lat</i>	11.560	0.2793	48	

Tabella 2: Stime degli effetti fissi della componente lineare del modello e dei gradi di libertà e del numero di nodi riferiti alla componente non lineare.

salmastre e un notevole calo delle produzioni agrumicole, caratterizzato dall'abbandono di parte delle superfici coltivate e dalla diminuzione delle produzioni unitarie.

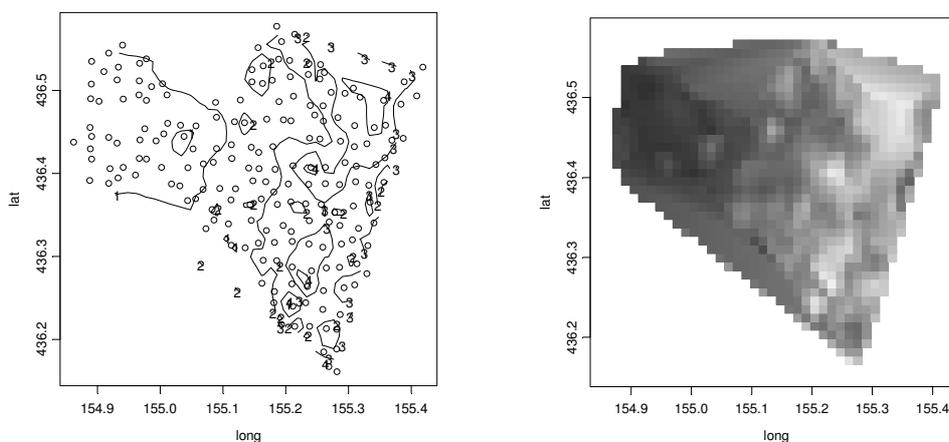


Figura 3: Distribuzione spaziale stimata di log(SAR)

In un primo tempo si è pensato che il peggioramento qualitativo delle acque sotterranee potesse essere imputato a cause secondarie, ovvero ad interventi da

parte dell'uomo, quali quelli relativi ai lavori di regimazione del fiume Flumendosa, o agli sbarramenti realizzati a monte dell'area indagata (dighe dell'Alto e Medio Flumendosa) o all'apertura delle foci per le attività della pesca o ancora ai cospicui emungimenti dai pozzi per uso irriguo, turistico ed idropotabile (Puddu et al., 2002). Proprio questa ipotesi ha determinato nel Settore Utilizzazione del Territorio del CRAS (Centro Regionale Agrario Sperimentale di Cagliari) la volontà di verificare le cause effettive dell'eccessiva presenza di sale nel suolo in questione, tramite un progetto riguardante lo studio della salinizzazione dei terreni e delle falde di tutta la piana, a partire dall'anno 2000.

La distribuzione spaziale del logaritmo del tasso di assorbimento del sodio è stata analizzata utilizzando un modello di dipendenza geo-additivo che, in contrapposizione al kriging universale, ha permesso di considerare sia i legami lineari esistenti tra il tasso di assorbimento del sodio, i cationi di scambio e la capacità di scambio cationico, sia quelli non lineari riferiti al *pH* ed alla componente geografica.

I risultati ottenuti hanno portato ad escludere l'influenza della componente strutturale del suolo sull'andamento del fenomeno in esame, e a concludere che l'elevato grado di salinità del suolo, causa di un notevole calo produttivo della coltura agrumicola, non è legato a fattori di origine antropica, ma piuttosto a ben note cause naturali.

Bibliografia

- Castrignanò, A., 1996. Utilità della geostatistica in agraria. *Bollettino della SISS*, 7: 87-92.
- Castrignanò A. e Puddu R. (2004). Fattori fisico-chimici e topografici che influenzano la stima della salinizzazione e sodicizzazione dei suoli della piana costiera di Muravera-Villaputzu (Sardegna sud-orientale). *Atti del Convegno della Società Italiana della Scienza del Suolo, 2004*.
- Claeskens G., Aerts M. e Wand, M.P. (1999). Some results on penalized spline estimation in generalized additive and semiparametric models. *Proceedings of the 52nd Session of the International Statistical Institute, Helsinki*, pp. 207-208.
- Goovaerts, P., 1999. Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma*, 89: 1-45.
- Hastie T.J. (1996). Pseudosplines. *J. R. Statistic. Soc.*, B, 58, 379-396.
- Kamman E.E. e Wand M.P. (2003) Geoaddivitive models, *Applied Statistics*, 52, 1-18.
- Lemhann A., Overton J.McC., Leathwick J.R. (2003) GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling*, 160, pagg 165-183
- Ngo L. e Wand M.P. (2004) Smoothing with Mixed Model Software, *Journal of Statistical Software*, 9, 1
- Nychka (2000). Spatial process estimates as smoother. In *Smoothing and Regression* (ed. M. Schimek). Heidelberg: Springer.
- Pozdnyakova L.e Zhang R. (1999) Geostatistical Analyses of Soil Salinity in a Large Field, *Precision Agriculture*, 1, 153-165
- Puddu R., Sanna F., Fanni S., Fantola F., Manca D., Mereu G., Fadda N., Lai P., Onano M., 2002. Realizzazione di un sistema di monitoraggio per il controllo

del fenomeno di salinizzazione dei suoli: risultati della fase conoscitiva nella piana di Muravera-Villaputzu (Sardegna sud-orientale). *Bollettino SISS* 51, 1-2: 607-625.

Stein M.L. (1999). *Interpolation of Spatial Data – Some Theory for Kriging*. Springer.

Wand M.P., French J.L., Ganguli B. Kammann E.E., Stadenmayer J. e Zanobetti A. (2005). SemiPar 1.0. R package. [http:// cran.r-project.org](http://cran.r-project.org)

Wand, MP (2003). Smoothing and mixed models. *Computational Statistics*, 18, 223-249.