



UNIVERSITÀ DEGLI STUDI DI PADOVA

*Facoltà di Scienze Statistiche*

Corso di Laurea Specialistica  
in Statistica ed Informatica

TESI DI LAUREA

*Modelli Additivi con Serie Spazio-Temporali*

RELATORE: PROF. BRUNO SCARPA

LAUREANDO: MARCO RUFFA

ANNO ACCADEMICO 2007-2008



*Ai miei genitori*



# Indice

<b>1</b>	<b>Modelli Additivi</b>	<b>3</b>
1.1	Modelli additivi . . . . .	4
1.2	Gli stimatori di regressione non-parametrici . . . . .	7
1.2.1	Stimatori di regressione . . . . .	7
1.2.2	Regressione locale . . . . .	9
1.2.3	Stimatore di regressione pesata localmente ( <i>loess</i> ) . . . . .	10
1.2.4	Varianza e distorsione . . . . .	13
1.2.5	I gradi di libertà . . . . .	15
1.3	Stima del modello additivo . . . . .	17
1.3.1	Algoritmo di <i>backfitting</i> . . . . .	18
1.3.2	Inferenza sui modelli additivi . . . . .	19
1.4	Generalizzazioni . . . . .	22
1.5	Bootstrap . . . . .	23
1.5.1	Il bootstrap nell'analisi delle serie storiche . . . . .	25
1.5.2	Residual bootstrap . . . . .	28
<b>2</b>	<b>Dipendenza tra le osservazioni</b>	<b>29</b>
2.1	Dati autocorrelati . . . . .	29
2.1.1	L'autocorrelazione . . . . .	30
2.1.2	Lisciatori pesati . . . . .	31

2.1.3	Minimi quadrati generalizzati . . . . .	33
2.2	Rette di regressione con pesi autocorrelati e locali . . . . .	34
<b>3</b>	<b>La funzione di stima</b>	<b>37</b>
3.1	Funzione uni-dimensionale . . . . .	37
3.1.1	Sviluppo . . . . .	37
3.1.2	Verifica . . . . .	40
3.2	Funzione bi-dimensionale . . . . .	45
3.2.1	Sviluppo . . . . .	45
3.2.2	Verifica . . . . .	48
<b>4</b>	<b>Specificazione del modello</b>	<b>57</b>
4.1	I dati . . . . .	57
4.1.1	Descrizione dei dati . . . . .	57
4.2	Il modello . . . . .	59
4.3	Le stime . . . . .	60
<b>5</b>	<b>Conclusioni</b>	<b>67</b>
<b>A</b>	<b>Funzioni in R</b>	<b>69</b>
	<b>Bibliografia</b>	<b>79</b>

# Introduzione

Negli anni recenti, nella letteratura statistica ci sono diversi lavori che hanno per oggetto o fanno riferimento a tecniche di analisi di serie spatio-temporali. Molti sono i casi in cui si presentano questo genere di dati e ci si rese conto che le componenti spaziali e temporali presentano, nella maggior parte dei casi, importanti interazioni, da tenere in considerazione ai fini di una corretta descrizione del fenomeno oggetto di studio.

Un possibile metodo per stimare queste serie è utilizzare i modelli additivi con i quali è possibile stimare congiuntamente l'influenza di diverse variabili esplicative su una variabile risposta. In maniera più precisa questi modelli vengono utilizzati, per quanto riguarda la stima per processi spatio-temporali, inserendo tra le variabili esplicative una variabile che descrive il contributo dato dal tempo e una che descrive il contributo dato dallo spazio. In questo contesto è verosimile che le ipotesi di indipendenza dei residui non siano soddisfatte e quindi il processo presenterà autocorrelazione sia temporale che spaziale.

Il problema principale che si riscontra utilizzando i modelli additivi quando si hanno a disposizione dati autocorrelati è che gli stimatori non-parametrici usati in questi modelli non tengono conto della dipendenza presente nel processo e forniscono, perciò, stime che almeno in termini di efficienza non sembrano particolarmente attraenti.

Nel presente lavoro si propone una modifica degli stimatori non-parametrici di regressione inserendo nella procedura di stima anche le informazioni riguardanti l'eventuale dipendenza presente nei dati, in particolare si fa riferimento alla regressione locale.

Nei primi due capitoli si ricorda l'idea dei modelli additivi, dei dati autocorrelati e la tecnica *bootstrap* che verrà poi utilizzata per creare gli intervalli di confidenza e viene introdotto il nuovo tipo di lisciatore che considera l'autocorrelazione delle osservazioni. Nel terzo capitolo viene spiegato più dettagliatamente e implementato il nuovo stimatore e viene fatta una prova della sua capacità di stima con dei dati simulati confrontandolo con uno stimatore di tipo *loess* che non tiene conto delle informazioni riguardanti l'autocorrelazione. Questo viene compiuto sia per quanto riguarda la componente temporale (univariata) che per la componente spaziale (bivariata). Nel quarto capitolo viene studiato un caso reale riguardante la velocità del vento in Irlanda. La stima della serie spazio-temporale viene effettuata e confrontata con due modelli additivi il primo usa come lisciatori il nuovo stimatore proposto in questa tesi, mentre il secondo utilizza lisciatori *loess*. Nell'ultimo capitolo si forniscono le conclusioni di questo studio e si forniscono delle idee si possono realizzare per continuarlo e migliorarlo.

# Capitolo 1

## Modelli Additivi

Per riprodurre l'andamento di una struttura di dati spazio-temporali un possibile modello è del tipo

$$y(t, s) = f_1(t) + f_2(s) + \varepsilon(t, s) \quad (1.1)$$

dove la  $y$  è la variabile risposta,  $f_1(t)$  rappresenta il contributo dato dal tempo e  $f_2(s)$  il contributo dato dallo spazio (questo termine potrà essere bivariato se si utilizzano le coordinate cartesiane per indicizzare lo spazio) alla variabile risposta e  $\varepsilon(t, s)$  è il termine di errore che solitamente si considera indipendente e identicamente distribuito con media zero e varianza  $\sigma^2$ . Le  $f_j$  sono funzioni arbitrarie, una per ogni variabile esplicativa, le variabili possono dipendere anche da due o più variabili, quindi alcune componenti possono essere funzioni multidimensionali.

Questo tipo di modello è utile per studiare casi come la velocità del vento per un certo periodo in diversi siti, per monitorare il livello di inquinamento oppure in agricoltura per ottimizzare la produzione di colture.

Non conoscendo il meccanismo generatore dei dati si devono specificare queste funzioni. Se questo meccanismo fosse noto si potrebbe istituire un

modello parametrico ben definito, cosicché il modello sarebbe correttamente specificato e la distorsione sarebbe praticamente nulla. Il contesto in cui operiamo ci porta invece a procedere sostanzialmente secondo un approccio non parametrico.

## 1.1 Modelli additivi

Il modello (1.1) può essere visto come un modello additivo (Hastie e Tibshirani, 1990).

I modelli additivi si basano sull'idea di usare stimatori non-parametrici unidimensionali come componenti per la costruzione di una classe ristretta di modelli non parametrici per la regressione multipla (Scarpa, 1998).

Con dati di tipo spazio-temporale le due variabili esplicative sono il tempo e lo spazio le quali, per la loro stima, saranno associate a due lisciatori. Una caratteristica importante di questo tipo di osservazioni è l'autocorrelazione tra le varie osservazioni sia nel tempo che nello spazio, difatti è presumibile che osservazioni vicine (sia rispetto al tempo che allo spazio) siano correlate tra loro in qualche modo.

In generale, l'utilizzo dei modelli additivi del tipo (1.1) cerca di evitare una serie di problemi che sono caratteristici dell'utilizzo di stimatori non-parametrici  $p$ -dimensionali, come ad esempio:

- La cosiddetta “maledizione della dimensionalità” (Hastie, Tibshirani e Friedman 2001). All'aumentare del numero di variabili, i punti osservati si disperdono molto rapidamente e gli intorni di un punto determinati raggruppando una fissata percentuale di altri punti ad esso vicini, sono meno locali rispetto al caso di uno stimatore per una singola variabile; si avranno quindi maggiori distorsioni.

- Nel caso multidimensionale la necessità di assumere un qualche tipo di metrica che spesso è difficile da giustificare, specialmente quando le variabili sono misurate in diverse unità di misura o sono altamente correlate.
- L'elevata complessità computazionale (in genere  $\geq O(n^2)$  operazioni) quando gli stimatori sono multivariati.

La definizione generale di modello additivo viene definita nel modo seguente (Hastie e Tibshirani, 1990)

$$Y = \alpha + \sum_{j=1}^p f_j(X_j) + \varepsilon \quad (1.2)$$

dove i dati siano composti di  $n$  realizzazioni di una variabile aleatoria  $Y = [y_1, \dots, y_n]$  e di  $p$  variabili esplicative  $X_j = [x_{1j}, \dots, x_{nj}]$  con  $j = 1, \dots, p$  che possono essere determinate oppure possono essere misurazioni di variabili aleatorie e dove gli errori  $\varepsilon$  sono solitamente considerati indipendenti dalle  $X_j$  con media  $E(\varepsilon) = 0$  e varianza  $Var(\varepsilon) = \sigma^2$ .

Nel modello (1.2) si assume che  $E\{f_j(X_j)\} = 0$  in modo che ogni funzione non abbia costanti libere, che invece vengono raccolte tutte in  $\alpha$ .

Il modello additivo ha una motivazione a priori molto importante per l'analisi dei dati, ovvero nella (1.2) ogni variabile viene rappresentata separatamente. Questa peculiarità è comune anche al modello lineare. Una volta adattato il modello additivo ai dati si possono disegnare separatamente le  $p$  funzioni relative alle diverse variabili per esaminare separatamente le stesse variabili nel descrivere la risposta.

Il modello additivo può essere considerato come una generalizzazione del modello lineare. Nel modello additivo le funzioni stimate possono essere considerate l'analogo dei parametri in un modello di regressione lineare. Per

quanto riguarda l'interpretazione nell'ambito dei modelli additivi bisogna porre attenzione particolare a quelle variabili che spiegano lo stesso aspetto della variabile risposta e che quindi possono sembrare non significative. Solo a questo punto è opportuno esaminare le funzioni adattate, sapendo che ognuna è davvero necessaria per costruire la superficie multivariata.

Un altro importante problema, è quello di determinare la forma delle funzioni  $f_j(\cdot)$  per  $j = 1, \dots, p$  che vengono utilizzate nel modello (1.2). Nel caso più generale si può non formulare alcun tipo di ipotesi iniziale sulla forma di tali funzioni, una possibile via è quella di usare stimatori non-parametrici per la regressione univariata. Tali stimatori si determinano nel caso semplice di regressione con una sola variabile esplicativa e cercano di stimare al meglio la funzione di regressione senza dover effettuare scelte pesanti sulla forma della funzione. In letteratura sono molte le proposte per stimatori di questo tipo e in seguito se ne presenterà uno ulteriore. Un possibile modo per stimare un modello additivo può essere quello di scegliere uno di questi stimatori per ciascuna variabile esplicativa e attraverso l'algoritmo di *backfitting* determinare la stima di ciascuna funzione (Friedman e Stuetzle, 1981).

Questo algoritmo consiste nello stimare una componente alla volta lasciando tutte le altre fissate e iterando tale processo in modo da stimare tutte le componenti. In questo modo se le stime correnti sono  $\hat{f}_k$ ,  $k = 1, \dots, p$ , allora  $\hat{f}_k$  viene aggiornata lasciando i residui parziali  $r_{ij} = y_i - \sum_{k \neq j} \hat{f}_k(x_{ij})$  sulla variabile  $x_{ij}$ . La procedura che implementa una tale idea verrà descritta con maggiore dettaglio in seguito.

Nella sezione (1.2) si presentano alcuni principali stimatori utilizzati per le stime delle funzioni  $f_j(\cdot)$ . Nella sezione (1.3), invece, si presentano in maggior dettaglio i modelli additivi, presentandone interpretazione, algoritmi di stima e utilizzo.

## 1.2 Gli stimatori di regressione non-parametrici

In questo paragrafo si presenta la logica per descrivere in maniera non parametrica l'andamento di una variabile risposta  $Y$  come funzione di una o più variabili esplicative  $X_1, X_2, \dots, X_p$ . Le stime ottenute per tale andamento presentano una minore variabilità rispetto alle osservazioni  $Y$ , questo in inglese è detto *smoother* che per l'appunto significa lisciatore. Un'importante proprietà dei lisciatori è la loro natura non-parametrica che permette di non assumere nessuna rigida forma di dipendenza tra la variabile risposta e le variabili esplicative.

Questi stimatori servono per ottenere tutte le componenti necessarie per poter applicare l'algoritmo di *backfitting*, come presentato nella sezione (1.3.1).

Considerando i problemi di tipo spazio-temporale, è pertanto utile sviluppare sia degli stimatori che utilizzino una sola variabile esplicativa, com'è il tempo, sia degli stimatori che considerino due variabili esplicative congiuntamente; il modo più semplice per trattare lo spazio è quello di considerarne le coordinate cartesiane rispetto a un sistema di riferimento assegnato, ottenendo così una coppia di coordinate per ciascun sito dello spazio.

### 1.2.1 Stimatori di regressione

Uno stimatore di regressione non parametrica o lisciatore (dall'inglese *smoother*) ha lo scopo di sintetizzare una variabile risposta  $y = [y_1, \dots, y_n]^T$  tramite le misurazioni nei punti  $x = [x_1, \dots, x_n]^T$ . Uno stimatore di  $y$  utilizzando  $x$  è la funzione  $\mathcal{S}(x_0 | x, y)$  che in ogni punto  $x_0$  stima la media di  $y$  condizionata a  $x$ . Solitamente si è interessati solo ai valori stimati nei punti osservati  $x_i$ ,  $i = 1, \dots, n$  e quindi lo stimatore risulta  $\hat{y}_i = \mathcal{S}(x_i | x, y)$ .

Se risulta che:

$$\mathcal{S}(x_0 \mid x, y_1 + \alpha y_2) = \mathcal{S}(x_0 \mid x, y_1) + \alpha \mathcal{S}(x_0 \mid x, y_2)$$

per ogni costante  $\alpha \in \mathbb{R}$  allora lo stimatore per la regressione non-parametrica (*smoother*) è detto lineare, questo implica che

$$\mathcal{S}(x_0 \mid x, y) = \sum_{i=1}^n s(i, x_0, x) y_i$$

per qualche insieme di pesi  $s(i, x_0, x)$ .

Nel caso in cui sia possibile scrivere lo stimatore come applicazione lineare definita da  $\hat{y} = Sy$  allora la matrice lisciante (*smoother matrix*)  $S$  dipende dalle  $x_1, x_2, \dots, x_n$  e dal particolare stimatore, ma non dipende da  $y$  (Hastie e Tibshirani, 1990).

Dato un algoritmo di lisciamiento lineare, si può determinare la corrispondente matrice  $S$ , lisciando i vettori della base unitaria. Il risultato del lisciamiento dell' $i$ -esimo vettore unitario è, infatti, l' $i$ -esima colonna di  $S$  (a volte chiamata “funzione di risposta a impulso” cfr. per esempio O’Sullivan, 1986). Si noti che una tale operazione non può essere fatta nel caso di uno stimatore non lineare per la regressione non-parametrica perché in quel caso le stime dipendono da  $y$  in maniera non-lineare.

Gli stimatori lineari possiedono alcune caratteristiche come la velocità e semplicità di calcolo e la simmetria della matrice di lisciamiento.

Nel caso in cui le variabili esplicative siano due o più, bisogna risolvere ulteriori problemi. Specificare un lisciatore può essere fatto in diverse maniere a seconda della famiglia scelta e anche in diversi modi pur scegliendo un unico tipo stimatore.

È importante osservare che può essere sufficiente determinare un metodo di stima  $\hat{y}_i$  nei punti osservati  $i = 1, \dots, n$  e poi estenderlo a uno stimatore in ogni punto del dominio di  $x$  usando per esempio l’interpolazione lineare.

### 1.2.2 Regressione locale

Considerando un modello del tipo

$$y = f(x) + \varepsilon$$

dove si è interessati ad esaminare la relazione tra la variabili  $x$  e  $y$ , e  $\varepsilon$  è il termine di errore casuale non osservato. Si può assumere, senza perdita di generalità, che la media degli errori sia zero. Ci si limita a stimare la  $f(x)$  presumendo per quest'ultima solo alcune condizioni di regolarità e solo in corrispondenza di un punto  $x_0$  fissato.

Se la  $f(x)$  è una funzione derivabile con derivata continua in  $x_0$  allora in base allo sviluppo in serie di Taylor  $f(x)$  è localmente approssimabile con una retta passante per il punto  $(x_0, f(x_0))$ , cioè

$$f(x) = \underbrace{f(x_0)}_{\alpha} + \underbrace{f'(x_0)}_{\beta}(x - x_0) + c$$

dove con  $c$  si indica una grandezza di ordine inferiore a  $|x - x_0|$  (Bertsch e Dal Passo, 2001).

Questo afferma che una qualunque funzione  $f(x)$ , sufficientemente regolare, può essere approssimata localmente da una retta. Utilizzando questa idea in ambito statistico, si cerca di stimare la  $f(x)$  in un intorno di  $x_0$  utilizzando  $n$  coppie di osservazioni  $(x_i, y_i)$  per  $i = 1, \dots, n$ .

Si introduce ora un criterio per pesare le osservazioni in base alla loro distanza da  $x_0$

$$\min_{\alpha, \beta} \sum_{i=1}^n \{y_i - \alpha - \beta(x_i - x_0)\}^2 w_i \quad (1.3)$$

dove i pesi  $w_i$  sono scelti maggiori quando la distanza tra  $x_0$  e  $x_i$  è più piccola. Il metodo di stima risultante viene chiamato regressione locale e si

riferisce a un'ottica di pesi locali intorno a  $x_0$  basati sul criterio dei minimi quadrati pesati. Una volta risolto il problema di minimo della (1.3), la stima derivante è  $\hat{f}(x_0) = \hat{\alpha}$ .

Operativamente un modo per scegliere i pesi è porre

$$w_i = \frac{1}{h} \omega\left(\frac{x_i - x_0}{h}\right) \quad (1.4)$$

dove la funzione  $\omega(\cdot)$  viene chiamata nucleo ed è una densità simmetrica all'origine e  $h$ , il parametro di lisciamiento, rappresenta un fattore di scala.

La funzione di minimo (1.3) dipende dai pesi  $w_i$  che a loro volta dipendono da  $h$ ,  $\omega(\cdot)$  e  $x_0$ . Anche fissati il parametro di lisciamiento e il nucleo, i pesi dipendono da  $x_0$  (1.4) e, quindi, anche il problema di minimo nella 1.3 dipende da  $x_0$ . Si vuole, però, stimare la  $f(x)$  per diverse scelte di  $x$ , il che porta a altrettante operazioni di minimizzazione. Si può mostrare che la stima per un generico punto  $x$  può essere calcolata in modo lineare in  $y$ .

Solitamente si stima la  $f(x)$  in un intero insieme di  $m$  valori e quindi si possono calcolare tutte le  $m$  stime mediante un sistema di operazioni matriciali del tipo

$$\hat{f}(x) = S_h y$$

dove  $S$  è una matrice  $m \times n$ , detta, matrice di lisciamiento,  $x$  è il vettore delle ascisse dove stimiamo  $f$  e  $\hat{f}(x)$  è il corrispondente vettore delle stime.

### 1.2.3 Stimatore di regressione pesata localmente (*loess*)

Ora descriviamo dettagliatamente lo stimatore basato sulla regressione locale pesata, proposto da Cleveland (1979), comunemente detto *loess* (*locally weighted regression*).

Questo metodo consiste nello stimare una retta di regressione pesata localmente per il punto  $x_0$  usando  $k$  vicini più vicini (*nearest neighbour*), può essere calcolato in un numero successivo di passi:

1. si identifichi l'insieme composto dai  $k$  vicini più vicini (*nearest neighbour*) chiamato  $N(x_0)$
2. si calcoli la distanza del *nearest neighbour* più lontano da  $x_0$  come

$$\Delta(x_0) = \max_{N(x_0)} |x_0 - x_i|$$

3. si assegnino i pesi  $w_i$  ad ogni punto in  $N(x_0)$ , usando la funzione tricubo

$$W\left(\frac{|x_0 - x_i|}{\Delta(x_0)}\right)$$

dove

$$W(u) = \begin{cases} (1 - u^3)^3, & 0 \leq u < 1; \\ 0 & \text{altrimenti.} \end{cases}$$

4. lo stimatore  $s(x_0)$  viene calcolato nel punto  $x_0$  determinando il polinomio di regressione ai minimi quadrati pesati di  $y$  su  $x$  nell'intorno  $N(x_0)$ , usando i pesi calcolati al punto 3.

Il numero di *nearest neighbour* da considerare nell'intorno  $N(x_0)$  viene espresso come percentuale o *span* dei punti utilizzati per lo stimatore in  $x_0$ , questo è il parametro di lisciamiento.

Cleveland (1979) ha anche discusso l'uso di una regressione robusta in ogni intorno per proteggersi da eventuali valori anomali. Tale algoritmo lascia ripetutamente i dati e ad ogni iterazione pesa in maniera minore i punti con residui grandi. L'utilizzo di quest'algoritmo rende però lo stimatore non lineare.

Un altro metodo per costruire gli intorno dei veri punti  $x_i$  sono i vicini più vicini simmetrici (*symmetric nearest neighbour*). Si prende un numero  $w$  compreso tra 0 e 1, e si assume che  $[wn]$  sia dispari. L'intorno consiste di  $[wn]$  punti,  $([wn]-1)/2$  a sinistra e a destra di  $x_i$  e il punto  $x_i$  stesso. Il numero  $w$  controlla il livello di liscio ottenuto dalla stima risultante, in inglese viene chiamato *span*. Maggiore è lo *span* e più lisce tendono a essere le funzioni. È da osservare che gli intorno vengono troncati vicino ai punti di frontiera (Hastie e Tibshirani, 1990).

Di norma i vicini più vicini sono preferibili ai vicini più vicini simmetrici, perché in un intorno con un fissato numero di punti, la distanza media dei punti da  $x_0$  è minore nel primo caso (a meno che i predittori siano diversamente spazati). In generale questo comporta una minore distorsione.

Un tale stimatore allora risulta particolarmente diffuso in quanto riassume la proprietà di entrambe le metodologie di fondo usate per ottenere stimatori non-parametrici. Condivide con gli stimatori di tipo *nearest neighbour* l'abilità di adattare il loro parametro di liscio alla densità locale delle variabili esplicative, mentre ha le stesse capacità di liscio i dati dei metodi basati sul *kernel*.

La regressione pesata locale si generalizza in maniera diretta al caso di due o più dimensioni (Cleveland e Devlin, 1988). Una volta definita la distanza in  $\mathbb{R}^p$ , l'intorno e la funzione di peso sono esattamente assegnati nella stessa maniera del caso unitario, e si può calcolare un piano locale anziché una retta.

Questo tipo di stimatore richiede la scelta di un parametro di liscio, in pratica viene scelto o in maniera empirica, attraverso ispezione visiva della curva, oppure utilizzando qualche altro metodo automatico, come per esempio la convalida incrociata (*cross-validation*), come presentato nel paragrafo

successivo.

Un tale stimatore richiede un carico computazionale pari a  $O(n^2)$  operazioni, dato che i pesi devono essere ricalcolati per ogni intorno.

### 1.2.4 Varianza e distorsione

Volendo studiare la distribuzione degli stimatori si osserva che al decrescere della distorsione necessariamente la varianza cresce e viceversa (Azzalini e Scarpa, 2004).

In questa sezione si assume che  $y_i = f_i + \varepsilon_i$ , dove  $f_i$  è la vera funzione e gli errori  $\varepsilon_i$  sono incorrelati con valore atteso 0 e varianza comune  $\sigma^2$ . Si assume pure che il parametro di liscio sia fissato con qualche valore specifico.

In genere lo stimatore *loess* risulta distorto per  $f$  arbitrarie, e si può rappresentare la distorsione del campione  $n$ -variato  $x_i$ , con la relazione  $d = f - Sf$  perchè fissato il parametro di *span* il metodo *loess* risulta lineare.

Usando il metodo delle rette di regressione pesate localmente, diminuendo il valore dello *span* la stima effettuata si adatterà meglio i dati, ma se cresce troppo questo adattamento finisce con il rincorrere nelle oscillazioni casuali dei dati, il che si traduce in un aumento della varianza, senza guadagno rilevante per la distorsione. Questa situazione è detta *sovra-adattamento* del modello ai dati e comporta un eccesso di *ottimismo* nella valutazione dell'errore. Al contrario se il valore dello *span* è troppo elevato (vicino a 1) la stima risulterà troppo liscia adattandosi poco ai dati, questo porterà ad avere una varianza elevata (Azzalini e Scarpa, 2004).

La presenza della componente di distorsione è essenzialmente dovuta alla mancata conoscenza del meccanismo generatore dei dati. Se questo fosse noto, si potrebbe istituire un modello ben definito e la distorsione sarebbe

nulla o al più trascurabile, come nel caso di modelli correttamente specificati (Azzalini e Scarpa, 2004).

Per ottenere stimatori non-distorti si può definire la funzione da stimare come valore atteso dello stimatore. procedendo in questo modo diversi stimatori possono stimare diverse quantità in un dato problema. Un altro approccio è considerare cosa succede asintoticamente, ci sarà generalmente distorsione asintotica. Se tuttavia l'ammontare di lisciamiento è decrescente a un appropriato tasso con cui si avvicina al limite, allora sotto opportune condizioni di regolarità, le stime dovrebbero essere consistenti. La matrice di varianze e covarianze dello stimatore lineare  $\hat{y} = Sy$  è semplicemente (Stone 1977)

$$\text{cov } \hat{y} = SS^T \sigma^2$$

Come si è accennato, la varianza e la distorsione sono entità in conflitto tra loro, che non possiamo minimizzare contemporaneamente. Dobbiamo quindi fare una scelta di compromesso tra *distorsione e varianza*. Tale fenomeno, noto nella letteratura di lingua inglese come *bias-variance trade off*, è legato, per gli stimatori non-parametrici, essenzialmente al parametro di lisciamiento.

Per trovare il compromesso ideale tra la varianza e la distorsione esistono varie tecniche, tra le quali

**Insieme di stima e di prova:** Questa tecnica consiste nel prendere una parte dei dati, selezionata casualmente, per la stima dei vari modelli candidati che differiscono rispetto al parametro di lisciamiento, ed è detta insieme di stima; la restante parte è usata per valutare le loro prestazioni, e quindi scegliere il preferibile.

**Metodo della convalida incrociata:** In questo caso si dividono i dati in  $p$  parti casuali con  $1/p$  di dati in ciascuna, e poi a rotazione usiamo  $p - 1$

porzioni per la stima e una per la verifica. È utile se si hanno a disposizione pochi dati e serve per eliminare, almeno in parte, l'arbitrarietà del metodo precedente.

**Criteri basati sull'informazione:** Non sempre è appropriato il confronto fra le log-verosimiglianze per stimare un modello. Quando il modello stesso non è prefissato e lo si sceglie entro un insieme di modelli alternativi bisogna anche tener conto del diverso numero di parametri utilizzati. Questi criteri introducono una penalizzazione in tal senso. Un altro vantaggio è che possono essere applicati anche a famiglie di modelli non annidati, diversamente del test del rapporto di verosimiglianza.

**Selezione automatica delle variabili:** La realizzazione di procedimenti automatizzati per la selezione di variabili si utilizzano quando i modelli sono tutti della stessa tipologia e cambia solo la lista delle variabili coinvolte. Questa procedura automatizzata individua un sottoinsieme di variabili che minimizza un certo criterio di selezione scelto. Il suo difetto è l'onere computazionale enorme a causa della stima di molti modelli. Questa tecnica può essere implementata in vari modi: selezione passo-a-passo, selezione in avanti o selezione all'indietro (per ulteriore dettagli Azzalini e Scarpa, 2004).

### 1.2.5 I gradi di libertà

Dopo aver stimato dei modelli è utile conoscere quanti sono i “gradi di libertà” utilizzati, una nozione che si ispira alla regressione lineare parametrica (Hastie e Tibshirani, 1990).

Ci sono almeno tre possibili definizioni di gradi di libertà che dipendo-

no dal contesto in cui vengono usate, derivate per analogia dal modello di regressione lineare.

- $tr(SS^T)$
- $tr(2S - S^T S)$
- $tr(S)$

Considerando l'ultima definizione, un'interpretazione della statistica  $C_p$  di Mallows (Mallows, 1973) corregge la somma dei quadrati degli scarti in modo da renderla non distorta rispetto al vero errore quadratico medio per la previsione, sommando la quantità  $2\hat{\sigma}^2$  che è una stima non distorta di  $\sigma^2$  e  $p = tr(S)$  è il numero di parametri. I gradi di libertà dello stimatore in questo contesto sono la traccia di  $S$ .

Le tre definizioni di gradi di libertà riportate precedentemente possono essere usate per determinare il valore del parametro di lisciamiento, producendo una curva con circa quel numero di gradi di libertà utile per determinare tale parametro in una classe di stimatori, quando la scelta automatica è impraticabile.

Per stimatori simmetrici con autovalori della matrice di lisciamiento  $\theta_i$  non negativi i tre modi portano a risultati asintoticamente equivalenti. Quindi in questo caso si può scegliere come definizione  $tr(S)$ , tenendo presente che è anche la più semplice da calcolare.

Difatti, considerando la seguente relazione

$$tr(S) = \sum_{i=1}^n \theta_i$$

il calcolo dei gradi di libertà, in questo caso, richiede solo la diagonale di  $S$ .

Si noti, inoltre, che se  $S$  è una matrice di proiezione simmetrica i tre metodi coincidono.

Purtroppo nessuna delle definizioni date in precedenza è attraente dal punto di vista computazionale. In particolare, sarebbe conveniente usare  $tr(S_j) - 1$  per selezionare quanto si vuole lisciare la  $j$ -esima variabile esplicativa inclusa nel modello, o anche come approssimazione dei gradi di libertà dell'errore senza la  $j$ -esima variabile, per il confronto tra modelli. Si sottrae uno perché c'è una costante ridondante in  $p - 1$  dei  $p$  termini del modello; in generale si sottrarrà la dimensione di  $\bigcap_j \mathcal{M}_1(S_j)$ , dove  $\mathcal{M}_1$  è lo spazio degli autovalori per il  $j$ -esimo lisciatore (Hastie e Tibshirani, 1990). Buja, Hastie e Tibshirani (1989), tramite piccoli esperimenti di simulazione, hanno scoperto che sommando i singoli gradi di libertà si ha una buona approssimazione dei veri gradi di libertà. Le uniche eccezioni si hanno quando le variabili esplicative presentano correlazione estremamente alta o quando viene usato un parametro di lisciamiento molto piccolo.

### 1.3 Stima del modello additivo

Il modello additivo fornisce un logico sviluppo del modello lineare standard, permettendo l'uso di stimatori di funzioni arbitrarie (quindi, non solo lineari) delle variabili esplicative. Tale argomento è stato trattato in dettaglio da Hastie e Tibshirani (1986), Green e Yandell (1985), Green, Jennison e Seheult (1985), Engle, Granger, Rice e Weiss (1986).

Si può considerare anche le interazioni tra le variabili, definendo una nuova variabile come funzione di due o più variabili esistenti (per es. il prodotto di due variabili), si possono introdurre alcuni selezionati tipi di interazione nel modello additivo.

### 1.3.1 Algoritmo di *backfitting*

Esistono una varietà di metodi efficienti per risolvere il sistema di un modello additivo, ovviamente metodi che dipendono sia dal numero che dal tipo di stimatori non-parametrici usati.

Il metodo di Gauss-Seidel, applicato ai blocchi formati dalle componenti  $f_1, \dots, f_p$  cerca di sfruttare la particolare struttura del sistema.

L'algoritmo di *backfitting* è sostanzialmente una variante a questo metodo e consiste in una procedura iterativa per risolvere il sistema di un modello additivo. Tale algoritmo può essere descritto nei seguenti passi:

1. *Inizializzazione*:  $f_i = f_i^{(0)}$ ,  $i = 1, \dots, p$  dove  $f_i^{(0)}$  è una qualsiasi funzione della  $i$ -esima variabile, per esempio la media
2. *Ciclo*  $i = 1, \dots, p, 1, \dots, p, \dots$

$$f_j^{(i)} = \mathcal{S}_j \left( y - \sum_{k \neq j} f_k^{(i-1)} \right)$$

*fino a* : ciascuna funzione  $f_j^{(i)}$  è uguale alla  $f_j^{(i-1)}$ .

Prima di richiamare il principale risultato relativo alla convergenza del *backfitting* bisogna dare la nozione di concurrità.

Prima ricordiamo che con il termine “collinearità” ci si riferisce alla dipendenza lineare tra variabili esplicative come causa di degenerazione, mentre il termine “concurrità” è stato usato (Buja, Donnell e Stuetzle, 1986) per descrivere dipendenza non-lineare che porta alla degenerazione in modelli additivi. In senso tecnico, la concurrità si riconduce alla collinearità in un modello in cui viene applicata una trasformazione (non-lineare) delle variabili esplicative, ad esempio, una regressione additiva dove si effettua una trasformazione lineare e quadratica ai predittori.

Si consideri l'algoritmo di *backfitting* con gli stimatori simmetrici  $S_j$ ,  $j = 1, \dots, p$ , con autovalori in  $[0,1]$ . Se  $S_j$  non presenta concurrità, si può mostrare che l'algoritmo di *backfitting* converge a una sola soluzione del sistema indipendentemente dalla funzione di partenza. Al contrario, se c'è concurrità, l'algoritmo di *backfitting* converge a una delle soluzioni del sistema, ma la funzione di partenza determina la soluzione finale.

È possibile giustificare la motivazione delle stesse equazioni normali muovendo da altri e diversi punti di partenza e seguendo diversi approcci, per esempio Buja, Hastie e Tibshirani (1989) ricavando le equazioni normali come soluzione di un problema ai minimi quadrati penalizzati.

Un altro approccio, un pò astratto, per definire e stimare modelli di regressione non-parametrica, di cui i modelli additivi sono un caso particolare, viene presentato da Chen, Gu e Wahba (1989); Hastie e Tibshirani (1990) hanno ripreso tale approccio e lo hanno esemplificato nel caso di modelli additivi. Tale approccio si basa sugli spazi di Hilbert che riproducono il nucleo ed è sostanzialmente una versione più approfondita su tali spazi dei minimi quadrati penalizzati.

Per quanto riguarda questa tesi non si ritiene opportuno scendere in maggiori dettagli su questi metodi, anche se sono utili per dare una giustificazione all'algoritmo del *backfitting* appena presentato.

### 1.3.2 Inferenza sui modelli additivi

In questa sezione si assume che  $y_i = f(x_i) + \varepsilon_i$  dove  $f(x_i)$  è la vera funzione di regressione e gli errori  $\varepsilon_i$  sono incorrelati con media zero e varianza  $\sigma^2$ .

### Distorsione e consistenza

Una stima del modello additivo sarà tipicamente distorta, a meno che non vengano fatte alcune rigide assunzioni. Assumendo che  $f(x)$  sia un polinomio di grado prefissato, allora l'appropriata stima ai minimi quadrati sarà non-distorta. Invece assumendo che  $f$  sia arbitraria le stime additive saranno distorte come nel caso univariato. La presenza della componente di distorsione è essenzialmente dovuta alla mancata conoscenza del meccanismo generatore dei dati.

La consistenza asintotica è un argomento tipicamente più semplice da affrontare. O si assume che il modello additivo sia non distorto oppure si studia la consistenza sulle proiezioni della  $f$  sullo spazio delle stime additive.

### Errori standard

Dalle sezioni precedenti si è osservato che ogni funzione stimata in un modello additivo è il risultato di una mappa lineare o di un lisciatore applicato su  $y$ . Questo significa che le formule per la varianza e i gradi di libertà sviluppate precedentemente nel caso univariato possono essere riprese e applicate al modello additivo. In presenza di convergenza si può esprimere  $\hat{f}_j$  come  $S_j y$  per qualche matrice  $S_j$  di dimensione  $n \times n$ . Se gli errori sono indipendenti e identicamente distribuiti, allora la  $cov(\hat{f}_j) = S_j S_j^T \sigma^2$  dove  $\sigma^2 = var(Y_i)$ .

Si può calcolare direttamente la matrice lisciante  $S$ , ad eccezione di casi molto particolari come il modello semiparametrico. Generalmente il miglior approccio sembra essere quello di applicare la procedura di *backfitting* ad ognuno degli  $n$  vettori  $n$ -dimensionali che creano le colonne della matrice identità  $I_n$ . Il risultato del *backfitting* applicato all' $i$ -esimo vettore unitario produce la stima di un vettore che costituirà la  $i$ -esima colonna di  $S_j$ . Similmente la somma delle colonne della matrice  $S_j$  sarà la  $i$ -esima colonna della

matrice  $S$ . Le bande di confidenza si possono costruire usando  $\pm$  due volte la radice quadrata degli elementi diagonali di  $\sigma^2 \hat{S}_j \hat{S}_j^T$ .

Poiché l'algoritmo di *backfitting* è calcolabile in  $O(kn)$  operazioni per stimatori di complessità  $O(n)$ , questa procedura richiede complessivamente  $O(kn^2)$  operazioni. Dunque  $k = pmC$ , dove  $p$  pari al numero di variabili esplicative,  $m$  al numero di iterazioni dell'algoritmo di *backfitting* e  $C$  è una costante propria del particolare stimatore scelto (Hastie e Tibshirani, 1990).

Si supponga che il modello additivo sia corretto, cioè  $Y_i = f_+(X_i) + \varepsilon_i$ , e che la stima di  $Sf_+ = g_+$  sia  $\hat{f}_+ = Sy$ . Inoltre un approssimazione di una quantità pivot per  $g_+$  è

$$\nu(g_+) = (\hat{f}_+ - g_+)^T (SS^T \hat{\sigma}^2)^{-1} (\hat{f}_+ - g_+)$$

Si assuma di avere una stima del parametro di dispersione  $\hat{\sigma}^2$  e la distribuzione  $G$  di  $\nu$ , allora possiamo costruire un insieme di confidenza simultaneo di livello  $1 - \alpha$  per tutte le funzioni componenti:

$$C(g_1, \dots, g_p) = \{g_1, \dots, g_p; \nu(g_+) \leq G_{1-\alpha}\}$$

Ovviamente questo risultato è fortemente legato alle ipotesi di indipendenza e identica distribuzione degli errori.

### Analisi della devianza

La teoria della distribuzione asintotica per i modelli lineari è ben nota. Se  $\eta_1$  e  $\eta_2$  sono due modelli lineari con  $\eta_1$  annidato in  $\eta_2$ . Allora sotto appropriate condizioni di regolarità e assumendo che  $\eta_1$  sia corretto,  $D(\hat{\eta}_1; \hat{\eta}_2) = D(y; \hat{\eta}_1) + D(y; \hat{\eta}_2)$  ha una distribuzione assintotica pari a un  $\chi^2$  con gradi di libertà pari alla differenza della dimensione dei due modelli. Questo risultato

è essenzialmente usato per confrontare i due modelli. Tale quantità ha distribuzione asintotica  $\chi^2$  con gradi di libertà pari alla differenza della dimensione dei due modelli. Se il parametro di dispersione è ignoto si può derivare un test  $F$  approssimato in maniera analoga (Hastie e Tibshirani 1990).

Per modelli non-parametrici e additivi ha senso usare la devianza per assegnare i diversi modelli e confrontare le loro differenze.

Un altro approccio è usare il *bootstrap* per stimare la distribuzione della devianza per un effetto, sotto ipotesi nulla che tale effetto sia assente. Sebbene sia utile per giustificare la più semplice approssimazione del  $\chi^2$ , una tale simulazione è molto onerosa in tempi di calcolo.

## 1.4 Generalizzazioni

È possibile generalizzare i modelli additivi con la stessa logica con cui sono stati generalizzati i modelli lineari (McCullagh e Nelder, 1983). Si tratta cioè di considerare gli effetti della variabili esplicative sulla scala lineare, ma la distribuzione delle variabili risposta come pure funzione che serve da legame tra le variabili esplicative e tale distribuzione, si assume appartenga a una classe più generale. Nel caso dei modelli additivi, la generalizzazione ai Modelli Additivi Generalizzati si specifica sostituendo la componente lineare dei Modelli Lineari Generalizzati  $\alpha + \sum_j \beta_j X_j$  con la consueta componente additiva  $\alpha + \sum_j f_j(X_j)$  e mantenendo la classe più generale per la distribuzione delle variabili risposta e della funzione legame.

Per ulteriori dettagli sui Modelli Additivi Generalizzati, per quanto riguarda la generalizzazione dell'algoritmo di *backfitting*, chiamata *local scoring*, e per quanto riguarda analisi statistiche si stima e di verifica di ipotesi si fa riferimento a Hastie e Tibshirani (1990). Nel caso considerato in questa

tesi non si ritiene opportuno aggiungere maggiori dettagli.

## 1.5 Bootstrap

L'idea base del *bootstrap* fu introdotta nel 1979 da Bradley Efron proponendo una metodologia basata su calcoli informatici finalizzati alla stima dello scarto quadratico medio di una stima  $\hat{\theta}$ . Il *bootstrap* è, quindi, una di quelle metodologie statistiche che nell'ambito anglosassone vengono dette *computer intensive*; trova cioè la sua ragione d'essere nella possibilità di disporre di adeguati strumenti di calcolo elettronico. A ciò si aggiunga che il metodo *bootstrap* ha il vantaggio di fornire una serie ampia di esemplificazioni, in articoli pubblicati su riviste a diffusione internazionale, in merito all'inferenza anche per funzioni molto complesse, difatti non risulta influenzata dalla complessità matematica dello stimatore impiegato. È un metodo generale per ottenere informazioni circa la variabilità e la distribuzione di statistiche campionarie  $\hat{\theta}$ ; quindi permette di stimare i limiti di confidenza del parametro  $\theta$  della popolazione, quando non si possiedono informazioni sulla sua distribuzione.

Se il campione è formato da  $k$  dati, l'idea di base è di estrarre da esso per campionamento semplice con ripetizione molti campioni di  $k$  osservazioni, chiamati “campione *bootstrap*” allo scopo di trovare la probabilità che la misura in oggetto cada all'interno di intervalli predeterminati.

Il campione *bootstrap* è nient'altro che il campione originario nel quale, per effetto dell'estrazione con ripetizione, alcuni dati sono ripetuti ed altri sono assenti, per mantenere lo stesso numero d'osservazioni. È proprio la modalità di estrazione, fondata sulla ripetizione, a generare la variabilità nelle stime; poiché è richiesto che i campioni abbiano tutti lo stesso numero

d'osservazioni, se si estraesse senza ripetizione sarebbero tutti identici.

Ognuno di questi campioni *bootstrap* di  $k$  osservazioni può contenere due o più valori identici, con l'ovvia esclusione d'altri valori che sono contenuti nel campione originale. Ognuno di questi permette di ottenere una stima della statistica desiderata.

La distribuzione della misura statistica calcolata è trattata come una distribuzione costruita a partire da dati reali (cioè della popolazione) e fornisce una stima dell'accuratezza statistica.

Per ciascuno dei  $B$  campioni *bootstrap* si calcola lo stimatore  $\theta$  desiderato ottenendo una replica *bootstrap*. Con  $B$  campioni di repliche casuali, si ottiene la successione di stime o repliche *bootstrap*, che sono la realizzazione della variabile casuale  $T$ .

La funzione di ripartizione empirica dei  $B$  valori  $\theta$  ottenuti fornisce una stima accurata delle caratteristiche della variabile casuale  $T$ . L'approssimazione è tanto più precisa quanto più  $B$  è elevato.

L'utilizzo del *bootstrap* porta ad ottenere stime dei parametri e misure della distorsione e della varianza alternative, e alla costruzione di intervalli di confidenza. L'approccio classico a questa categoria di procedure inferenziali prevede l'uso di statistiche pivotali e delle loro distribuzioni esatte o asintotiche.

L'intervallo di confidenza può essere determinato con il metodo percentile. Sulla base della distribuzione dei  $B$  campioni *bootstrap*, si individuano i valori che escludono una quota  $\alpha/2$  nei due estremi della distribuzione. Ad esempio, con 1000 repliche e se  $\alpha$  è uguale a 0.05 è sufficiente considerare l'intervallo che esclude i 25 valori inferiori e i 25 superiori.

Il metodo del percentile ordina per rango i valori ottenuti con le  $k$  repliche *bootstrap*, individuando i valori estremi entro i quali si collocano,  $1 - \alpha$

(in percentuale) valori della distribuzione, con  $\alpha/2$  in ognuno dei due estremi. Fondato sui ranghi, il metodo ovviamente non varia anche con trasformazione dei dati. Come condizioni di validità richiede che la distribuzione *bootstrap* non sia distorta; nel caso di piccoli campioni, ha un'accuratezza limitata.

Il metodo del percentile non sempre è corretto: l'intervallo di confidenza che comprende il 95% dei  $k$  valori calcolati non sempre contiene il 95% delle volte, però, a differenza di altri metodi, genera un intervallo asimmetrico.

### 1.5.1 Il bootstrap nell'analisi delle serie storiche

La metodologia bootstrap “classica” presenta una limitazione in quanto si può applicare nel caso in cui il campione sia i.i.d. Cao (1999). Al di fuori di questo contesto la metodologia *bootstrap* comporta problemi di “inconsistenza”, vale a dire la variabilità degli stimatori *bootstrap* cresce con l'aumentare dei campioni *bootstrap* estratti, ciò può portare a stime errate e diventa pertanto difficilmente proponibile il suo impiego nell'analisi delle serie storiche in quanto in questo caso si presenta una struttura di correlazione tra le successive osservazioni.

Già dagli inizi degli anni ottanta la ricerca ha proposto varianti al *bootstrap classico* in modo da conservare la struttura originaria dei dati durante il processo di ricampionamento.

Una delle procedure *bootstrap* per risolvere questo tipo di problemi è il *residual bootstrap*, che verrà spiegato in seguito.

Freedman e Peters nel 1984 proposero un'altra tecnica, il *recursive bootstrap* basato sui residui che si può utilizzare in modo valido qualora la serie storica sia rappresentata da un modello  $ARMA(p; q)$  con  $p$  e  $q$  noti. Finché il modello coinvolge variabili dipendenti non legate vengono generati campio-

ni della variabile indipendente in modo ricorsivo preservando la dipendenza della serie.

Nei casi in cui la struttura seriale non sia specificabile se ne può tenere conto utilizzando per esempio il *block bootstrap* o (*blockwise bootstrap*) (Carlstein (1986)) alla base del quale vi è l'idea di ricampionare non singole osservazioni ma blocchi di dati, mantenendo al loro interno lo stesso grado di dipendenza.

Nel *wild bootstrap* (originariamente chiamato *weighted bootstrap* da Wu, 1986) per ogni osservazione  $i$  del vettore dei residui dei minimi quadrati si genera la distribuzione  $\hat{F}_i$  da questa funzione di probabilità:

$$\hat{F}_i = \begin{cases} \frac{(1-\sqrt{5})\varepsilon_i}{2} & \text{con prob. } \frac{1+\sqrt{5}}{2\sqrt{5}} \\ \frac{(1+\sqrt{5})\varepsilon_i}{2} & \text{con prob. } 1 - \frac{1+\sqrt{5}}{2\sqrt{5}} \end{cases}$$

Importante da notare che la media di  $\hat{F}_i$  è zero, la sua varianza è  $\varepsilon_i^2$  e il momento terzo è  $\varepsilon_i^3$ . A questo punto si generano i vari residui  $\varepsilon_b$  come campioni presi dalla distribuzione  $\hat{F}_i$  che ha una distribuzione approssimata ai veri residui  $\varepsilon$ .

Una variante di quest'ultimo metodo è stata proposta da Kunsch (1989), essa è nota con il nome di *moving block bootstrap* ed è applicabile a serie stazionarie in genere.

Alternative al *moving block bootstrap* sono il *circular block bootstrap* e lo *stationary block bootstrap* introdotti da Politis e Romano (1992), Politis e Romano (1994).

Il *circular block bootstrap* ha lo scopo di centrare attorno alla media campionaria la distribuzione degli stimatori ricavati dagli pseudo dati al fine di ottenere stime non distorte. Lo *stationary block bootstrap* si differenzia dal

primo in quanto permette di conservare la stazionarietà della serie originale anche negli pseudo dati.

Una delle più recenti soluzioni proposte è il *matched block bootstrap* (Carlstein et al. (1998)). Il *matched block bootstrap* è una versione del *block bootstrap* in cui invece di estrarre i blocchi completamente a caso, assegnando uguale probabilità di estrazione ad ogni blocco e ricostruendo la stocastica indipendenza con il meccanismo del rimpiazzo, le probabilità di estrazione dei blocchi risultano dipendenti dal blocco precedentemente estratto, l'unione tra blocchi è regolata da una matrice di transizione.

Altri autori, tra cui Findley (1986) e Stine (1987), hanno preso in esame la costruzione di intervalli di previsione per modelli autoregressivi. Per ciascuna replica  $X_t^*$   $t = 1, \dots, n$ , generata dall'esperimento *bootstrap*, viene adattato un modello autoregressivo di ordine  $p$  sulla base del quale si estrapolano i valori:  $X_{n+k}^*$ . In tal modo, gli errori di previsione sono resi osservabili, essendo forniti da:  $X_{n+k} - X_{n-k}^*$ , ed è quindi possibile pervenire, iterando l'estrazione dei campioni bootstrap, alla stima dell'errore quadratico medio di previsione e alla costruzione degli di previsione.

In tale direzione, un altro metodo per ottenere intervalli di confidenza per modelli autoregressivi tenendo conto della variabilità della stima e dell'ordine dei parametri è stato introdotto da Masarotto (1990), considerando diversi criteri per l'identificazione automatica e forme alternative di stimatori.

Un altro metodo, che si basa sullo studio appena descritto, è stato proposto da Grigoletto (1998) che si basa sulla proposta di Masarotto (1990), permettendo sostanzialmente di ridurre la varianza della stima della distribuzione.

### 1.5.2 Residual bootstrap

Nell'analisi che verrà descritta in seguito si è deciso di utilizzare la tecnica del *residual bootstrap* (Efron e Tibshirani, 1986 e 1993).

Questo approccio consiste nel ricampionamento sull'insieme dei residui ottenuti dalla stima fatta. In termini formali il modello di stima è:

$$y = X\hat{\beta} + \varepsilon$$

e quindi

$$\varepsilon = y - X\hat{\beta}$$

dove  $\hat{\beta}$  è il vettore della stima dei minimi quadrati.

Per i residui si assume una funzione empirica di distribuzione che assegna la stessa probabilità ad ognuno, cioè ogni residuo avrà probabilità  $1/N$  (con  $N$  osservazioni) di essere riutilizzato nello schema di ricampionamento, assumendo che i residui siano indipendenti. Si costruisce una ripetizione *bootstrap* da un insieme di residui estratti casualmente (con ripetizione) in combinazione con i parametri stimati nel primo stadio.

$$y_b = \varepsilon_b + X\hat{\beta}$$

dove:  $\varepsilon_b$  è il vettore dei residui ricampionati,  $\hat{\beta}$  è la stima del primo stadio,  $X$  è la matrice delle osservazioni sulle variabili esplicative,  $y_b$  è la stima *bootstrap* della variabile risposta.

A questo punto si può ricalcolare  $\hat{\beta}_b$  per ogni campione, ottenendo una stima *bootstrap*. Questo afferma che la stima *bootstrap*  $\hat{F}$  è la funzione di distribuzione empirica dei residui dei minimi quadrati.

# Capitolo 2

## Dipendenza tra le osservazioni

### 2.1 Dati autocorrelati

Nei processi di tipo spazio-temporale è presumibile assumere che ci sia dipendenza tra le osservazioni. Se non si considera questa caratteristica il rischio è quello di avere una stima fortemente distorta (per eccesso o per difetto) degli errori sui parametri di regressione. Per correggere questo andamento è possibile ricorrere a tecniche che permettono di avere una stima corretta delle variabili di regressione, specificando un preciso schema di correlazione fra le varie rilevazioni.

Il principale effetto di dipendenza è dovuto a correlazioni spaziali o temporali fra le varie osservazioni. Ad esempio fenomeni climatici o economici di solito presentano dipendenza temporale, nel senso che il presente dipende dal passato, mentre in uno studio agrario è facile che vi sia dipendenza spaziale nel senso che parcelle vicine tendono ad assomigliarsi ai fini dell'esperimento.

Per adattare un modello additivo ad un processo spazio-temporali, si può inserire tra le variabili esplicative una variabile che indica il tempo e una che indica lo spazio. In questo contesto è verosimile che le ipotesi di indipenden-

za dei residui non siano soddisfatte: il processo presenterà autocorrelazione temporale e spaziale.

### 2.1.1 L'autocorrelazione

L'interdipendenza o autocorrelazione è la conseguenza della presenza di una relazione funzionale tra ciò che accade in un determinato punto e ciò che accade in altri punti. In altre parole, le caratteristiche di un determinato fenomeno in una regione in un determinato momento non sono spiegate unicamente da determinanti interne alla stessa ma anche da alcune peculiarità proprie delle altre regioni e in altri momenti, più o meno vicini nello spazio e nel tempo (Wei 2006).

In presenza di reali fenomeni di interazione spazio-temporale, l'eterogeneità è correlata alla mancanza di stabilità nello spazio e nel tempo delle relazioni o dei comportamenti analizzati. Dal punto di vista della modellizzazione questo implica che le forme funzionali ed i parametri variano a seconda della loro localizzazione geografica e nella loro collocazione temporale e non sono omogenei in tutto l'insieme dei dati.

Per correggere questo andamento è possibile ricorrere a tecniche che permettono di avere una stima corretta delle variabili di regressione, specificando un preciso schema di correlazione fra le varie rilevazioni.

In un modello del tipo (1.1) in caso di correlazione si assume che

$$\varepsilon \sim N(0, \Sigma)$$

con  $\Sigma$ , matrice di covarianza d'errore, simmetrica, semidefinita positiva e indipendenti dal tempo (o spazio), cioè dipendono solo dalla distanza tra le osservazioni (sia nel senso spaziale che temporale). Nel caso di regressione ordinaria  $\Sigma$  coincide con la matrice identità; nel caso di regressione pesata ta-

le matrici ha elementi fuori diagonale nulli e elementi diversi sulla diagonale principale. Le condizioni di stazionarietà assicurano che la funzione di autocorrelazione diminuisca all'aumentare dell'ampiezza del ritardo (Johnston, 1993). Un processo stazionario è un processo stocastico in cui la funzione di densità di probabilità di una qualche variabile casuale non cambia né nel tempo né nello spazio. Quindi anche i parametri media e varianza non cambiano nel tempo e nello spazio, di conseguenza la distribuzione del processo casuale mantiene gli stessi attributi ovunque.

### 2.1.2 Lisciatori pesati

In qualche istanza si può voler dare pesi differenti alle osservazioni. Questo capita spesso nel caso in cui il modello assuma eteroschedasticità delle varianze, cioè

$$Y_i = f(x_i) + \epsilon_i$$

Dove  $\text{var}(\epsilon_i) = \sigma_i^2$ . Si potrebbe avere che l' $i$ -esima osservazione abbia peso  $1/\sigma_i^2$ , assumendo le varie  $\sigma_i$  note. Se le varianze non sono note bisogna usare qualche tipo di tecnica iterativa di liscio pesato.

Per ogni tipo di liscio comune usato per costruire un modello additivo è facile includere i pesi.

Nel caso di un liscio lineare qualsiasi, che può essere rappresentato con la sua matrice di pesi  $S$ , al quale si vogliono incorporare i pesi  $w_i$  normalizzandoli in modo che la loro somma valga uno, sembra naturale ridefinire i pesi del liscio come  $w_j S_{ij}$ , e poi rinormalizzare in modo che le righe sommino a uno. Formalmente si dovrebbe procedere come segue. Data  $W = \text{diag}(w)$ , una matrice diagonale dove nella diagonale entrano i pesi  $w_i$ , e poi si prende  $D$  come  $D = \text{diag}(Sw)$ . Successivamente il liscio pesato

sarà

$$D^{-1}SW$$

Nel caso di uno stimatore a regressione pesata localmente (*loess*), ha senso, semplicemente, costruire una nuova matrice di pesi moltiplicando la matrice dei pesi delle osservazioni e la matrice dei pesi del nucleo locale, usando la risultante come matrice di pesi nella stima dei minimi quadrati.

Un'altra possibile strada è definire il lisciatore pesato attraverso i minimi quadrati penalizzati, questa è la procedura standard definita per lo stimatore *spline* pesato. Il criterio dei minimi quadrati pesati penalizzati (Hastie e Tibshirani, 1990)

$$(y - f)^T W (y - f) + \lambda f^T K f$$

È minimizzato da  $\hat{f} = Sy$  dove

$$s = (W + \lambda K)^{-1} W$$

Effettivamente il calcolo di  $\hat{f}$  può essere compiuto approssimativamente allo stesso modo dello stimatore *spline* non pesato, con soltanto un piccolo aumento della complessità computazionale dell'algoritmo.

Un altro spiacevole problema è che la matrice dei pesi degli stimatori  $W$  generalmente non è diagonale. Per gli stimatori *spline* altri stimatori lineari simmetrici, può essere usato l'approccio dei minimi quadrati penalizzati, l'unico inconveniente è l'aumento della complessità computazionale. Nel caso dello stimatore a regressione pesata localmente (*loess*) possiamo definire un generico stimatore pesato che può essere ottenuto usando un'appropriata partizione di  $W$  come matrice di peso per ogni stima locale (Hastie e Tibshirani, 1990).

### 2.1.3 Minimi quadrati generalizzati

Con riferimento al modello 1.2 si ipotizza che sia  $\text{var}(\varepsilon) = \sigma^2 I$ . Questa ipotesi può cadere o perchè gli errori sono correlati fra loro oppure perchè la varianza non è costante. Scrivendo  $\text{var}(\varepsilon) = \sigma^2 \Sigma$ , la matrice  $\Sigma$  esprime la relazione fra gli errori. Nel caso di varianza non costante ed errori non correlati  $\Sigma$  sarà diagonale, mentre per errori correlati  $\Sigma$  avrà elementi fuori dalla diagonale diversi da zero.

Questi modelli vengono trattati con la tecnica dei minimi quadrati generalizzati (Generalized Least Squares, GLS) che minimizzano la quantità:

$$\varepsilon^T \varepsilon = (y - X\beta)^T \Sigma^{-1} (y - X\beta)$$

La soluzione delle equazioni normali è ora:

$$\begin{aligned} \hat{\beta} &= (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} y \\ \text{var}(\hat{\beta}) &= (X^T \Sigma^{-1} X)^{-1} \sigma^2 \end{aligned} \quad (2.1)$$

Dato che si può sempre scrivere  $\Sigma = S S^T$  con  $S$  matrice triangolare (decomposizione di Choleski) si ha:

$$(y - X\beta)^T S^{-T} S^{-1} (y - X\beta) = (S^{-1}y - S^{-1}X\beta)^T (S^{-1}y - S^{-1}X\beta)$$

cioè la regressione GLS è equivalente a regredire  $S^{-1}X$  contro  $S^{-1}y$ . Infatti:

$$y = X\beta + \varepsilon \quad (2.2)$$

a questo punto si ipotizza che esista una matrice  $L$ , non singolare tale che

$$\Sigma^{-1} = L^T L$$

cosicchè

$$\Sigma^{-1} = L^{-1}(L^T)^{-1}$$

moltiplicando entrambi i membri della 2.2 per  $L$  si ha il modello nelle variabili trasformate

$$y' = Ly = X'\beta + \varepsilon' = L(X\beta + \varepsilon)$$

conduce a una nuova regressione lineare. Esaminando la varianza dei nuovi errori  $\varepsilon'$  si ottiene:

$$\text{var}(\varepsilon') = \text{var}(S^{-1}\varepsilon) = S^{-1}(\text{var}(\varepsilon))S^{-T} = S^{-1}\sigma^2SS^TS^{-T} = \sigma^2I$$

e quindi, per il modello trasformato, le ipotesi sugli errori sono soddisfatte.

Inoltre lo stimatore basato sui minimi quadrati generalizzati gode, così come lo stimatore basato sui minimi quadrati ordinari, della proprietà di correttezza

$$E[\hat{\beta}] = E[\beta + (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\varepsilon] = \beta$$

## 2.2 Rette di regressione con pesi autocorrelati e locali

Se si vuole stimare un modello additivo in presenza di errori  $\varepsilon$  autocorrelati ci si accorge che le stime, almeno in termini di efficienza, non sembrano particolarmente attraenti, perchè gli stimatori non parametrici univariati generalmente usati in tale algoritmo non tengono conto della dipendenza presente nel processo.

In letteratura un possibile metodo per la risolvere questo problema, è stato proposto da Scarpa (1998) dove viene modificato l'algoritmo di *backfitting* in modo che si tenga conto dall'autocorrelazione.

## 2.2. RETTE DI REGRESSIONE CON PESI AUTOCORRELATI E LOCALI 35

Un metodo proposto da Lehmann, McC. Overton, Leathwick (2002) chiamato GRASP (generalized regression analysis and spatial prediction) gioca due importanti ruoli nell'ambito ecologico. Primo, sviluppa le relazioni tra la variabile di interesse e le variabili ambientali. Secondo, fornisce un metodo per costruire la previsione spaziale usando i punti di misurazione dei dati del campione.

Un altro studio per ridurre l'effetto dell'autocorrelazione è stato proposto da Segurado, Araújo e Kunin (2006) nel quale si usa un approccio a modello nullo per quantificare quest'effetto usando variabili prova generate da randomizzazioni toroidali della struttura spaziale dei dati ambientali, generando campioni per le variabili originarie.

Oppure, sono state studiate anche delle regole ecologiche e delle funzioni ecosistemistiche necessarie se si vuole migliorare la stima classica, si veda Guisan, Lehmann, Ferrier, Austin, McC.Overton, Aspinall e Hastie (2006). Qui viene data particolare importanza all'intuito ecologico e all'abilità statistica dell'analizzatore rispetto al metodo statistico usato e si puntualizza che in ambito spaziale si deve dare più importanza all'effetto dell'autocorrelazione.

In questo lavoro si propone una modifica degli stimatori non-parametrici di regressione inserendo nella procedura di stima anche le informazioni riguardanti l'eventuale dipendenza presente negli errori; le stime ottenute con tale algoritmo modificato e le loro proprietà statistiche vengono poi confrontate con quelle ottenute dallo stimatore usuale.

Si consideri ora un modello di tipo (1.2) in cui, però, gli errori  $\varepsilon$  siano autocorrelati tra loro. Si assuma anche che la forma funzionale della struttura di autocorrelazione sia nota per esempio attraverso la formulazione di un modello per il processo degli errori  $\varepsilon_i$ .

Tra i vari metodi nella regressione multipla si è deciso di utilizzare i

modelli additivi per diverse ragioni. I modelli additivi stimano curve di risposta con stimatori non-parametrici invece che con termini parametrici. Questo permette di esplorare in modo più accurato variabili risposta con andamenti ambientali, e permette di stimare modelli statistici migliori in accordo con la teoria ecologica.

I modelli additivi sono stati usati in numerose applicazioni per funzioni ecologiche: da Yee and Mitchell (1991); per confrontarli con i modelli lineari generalizzati (Austin and Meyers, 1996; Bio e al., 1998; Franklin, 1998); in confronto alle reti neurali (Brosse and Lek, 2000); per assegnare l'effetto dei cambiamenti climatici (Leathwick e al., 1996; Lehmann, 1998);

In questa tesi, per risolvere questo problema, si è deciso di costruire e utilizzare come stimatore non-parametrico si usa una regressione lineare pesata tenendo conto dei vicini più vicini (*nearest neighbour*) e anche pesata tenendo conto dell'autocorrelazione tra le osservazioni. Questa idea si basa sul voler migliorare l'efficienza degli stimatori *loess* aggiungendo una componente per l'autocorrelazione.

# Capitolo 3

## La funzione di stima

In questo capitolo si presenta e si prova la funzione di stima basata sulle rette di regressione con pesi autocorrelati e locali.

Considerando le serie spazio-temporali, analizzeremo, sicuramente, una variabile che descrive il tempo e una che descrive lo spazio, la prima di uni-dimensionale e la seconda di bi-dimensionale. In termini pratici significa che si ha bisogno di due funzioni, distinte per la dimensionalità della variabile esplicativa, entrambe basate sulle idee proposte nelle sezioni (1.3.1) e (2.2).

Questa funzione e le simulazioni per fare le varie prove e verifiche sono state sviluppate nell'ambiente applicativo R (rif. Appendice).

### 3.1 Funzione uni-dimensionale

#### 3.1.1 Sviluppo

Per sviluppare questa funzione si è tenuto conto solo della variabile esplicativa temporale, cioè ci si riferisce solo al caso in cui la nostra funzione di stima sia del tipo

$$y(t) = f(t) + \varepsilon(t) \tag{3.1}$$

dove la  $y$  è la variabile risposta,  $f(t)$  è il contributo dato dal tempo e  $\varepsilon(t)$  è il termine di errore che sarà autocorrelato, come presentato nel capitolo precedente.

Il primo problema che si è riscontrato è stato quello di decidere che tipo di autocorrelazione prendere in considerazione. Da altri studi fatti sui dati che verranno analizzati nel capitolo successivo si è osservato che la dipendenza temporale può essere ben descritta da un modello autoregressivo a ritardo uno che presuppone le seguenti ipotesi:

- Gli  $n$  valori della variabile dipendente  $y_t$  sono misurati a intervalli costanti;
- I residui del modello sono legati fra loro dalla semplice regola autoregressiva:

$$\varepsilon_t = \alpha\varepsilon_{t-1} + \eta_t \tag{3.2}$$

dove  $\eta_t \sim N(0, \sigma^2)$  è una componente di “rumore bianco“ e  $\alpha$  il coefficiente di autoregressione che lega una osservazione alla precedente.

Ovviamente di solito il valore di  $\alpha$  non è noto. Una sua stima può essere ottenuta stimando vari modelli, confrontandoli tra loro e decidendo quale valore del parametro  $\alpha$  adatta meglio il modello ai dati.

A questo punto si crea una matrice per i pesi dell'autocorrelazione simmetrica e che resterà costante per ogni iterazione dalla funzione perché l'autocorrelazione dipende solo dalla differenza temporale e non dal momento considerato. Tale matrice  $n \times n$  risulta essere

$$\begin{pmatrix} 1 & \alpha^1 & \alpha^2 & \dots & \alpha^{n-1} & \alpha^n \\ \alpha^1 & 1 & \alpha^1 & \dots & \alpha^{n-2} & \alpha^{n-1} \\ \alpha^2 & \alpha^1 & 1 & \dots & \alpha^{n-3} & \alpha^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ \alpha^{n-1} & \alpha^{n-2} & \alpha^{n-3} & \dots & 1 & \alpha^1 \\ \alpha^n & \alpha^{n-1} & \alpha^{n-2} & \dots & \alpha^1 & 1 \end{pmatrix} \quad (3.3)$$

Successivamente, per ogni punto osservato, si crea una matrice diagonale di pesi. Dato il punto di stima si calcola la sua distanza da ogni punto del campione. Da queste distanze si calcolano i pesi utilizzando come parametro di lisciamiento  $h$ . Anche il valore di  $h$ , come quello di  $\alpha$ , non è noto e si può stimare confrontando i vari modelli.

A questo punto si devono unire in qualche modo le due matrici dei pesi che sono state calcolate. Il modo scelto per mettere insieme le matrici è quello di fare un prodotto matriciale tra le due

$$W_{complessivo} = W_{autocorrelazione} W_{nearest\ neighbour}$$

La vera e propria stima viene calcolata tramite una regressione lineare con il metodo dei minimi quadrati generalizzati con i pesi che sono stati appena calcolati. Successivamente si deve iterare questa procedura ad ogni punto del campione per avere una stima complessiva. Cioè, per ogni ripetizione, si crea una nuova matrice dei pesi *nearest neighbour* e si calcola la stima di quel determinato punto, risultando una procedura onerosa in termini computazionali.

### 3.1.2 Verifica

Come insieme di prova si usa una simulazione di 100 dati autocorrelati del tipo AR(1) (3.2) con parametro di autocorrelazione pari a 0,8 . Questi dati vengono prima stimati con lo stimatore appena descritto e poi con uno stimatore di tipo *loess*. Successivamente si effettua un confronto tra le varie stime fatte.

I dati simulati sono rappresentati in figura 3.1.

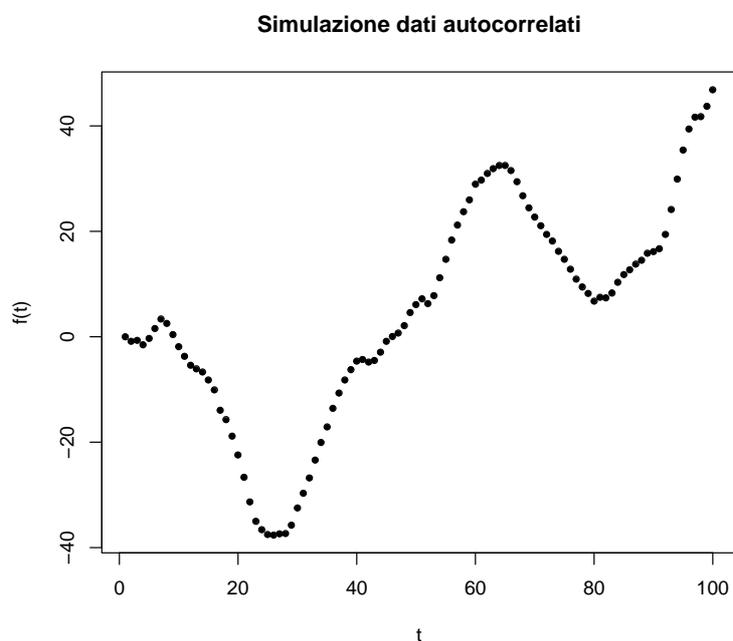


Figura 3.1: Simulazione AR(1) con parametro 0.8

Come prima verifica si fanno tre stime con il vero valore del parametro di autoregressione e cambiando l'ampiezza della finestra per il calcolo dei pesi *nearest neighbour*.

Dalla figura 3.2 si può notare come le tre stime sono coerenti con la simulazione. Bisogna notare che per finestre troppo piccole (curva verde) la stima tende quasi a interpolare i punti, questa stima non va bene perchè si avrà un effetto della distorsione troppo elevato. Se, invece, la finestra è più ampia (curva blu) la stima tende a essere troppo liscia. La giusta stima sembrerebbe data dalla curva rossa.

Per vedere l'effetto che ha il parametro per l'autocorrelazione sulla stima, adesso verranno presentate altre stime con la finestra *nearest neighbour* fissa e si cambierà il parametro dell'autocorrelazione.

Dalla figura 3.3 si può notare che al variare del parametro di autocorre-

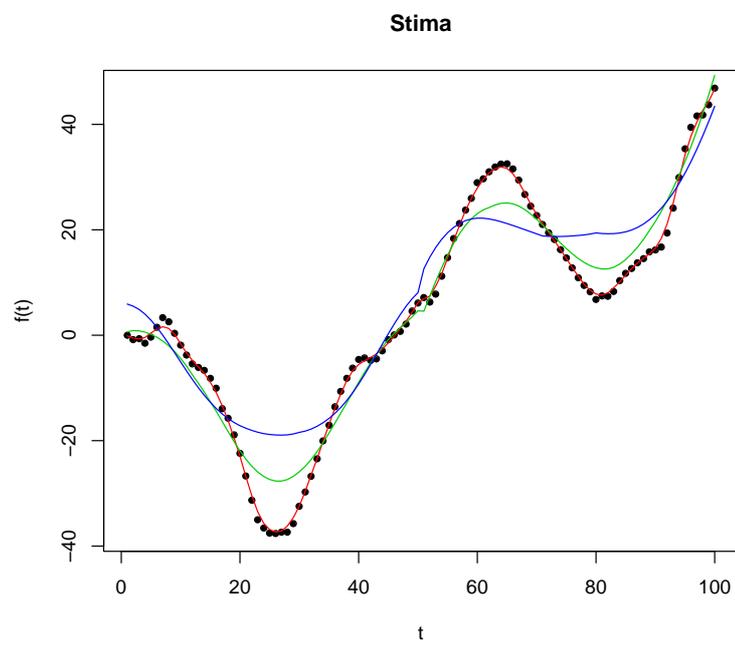


Figura 3.2: Stima con finestre *nearest neighbour* diverse; rosso  $h=1$ , verde  $h=5$ , blu  $h=10$

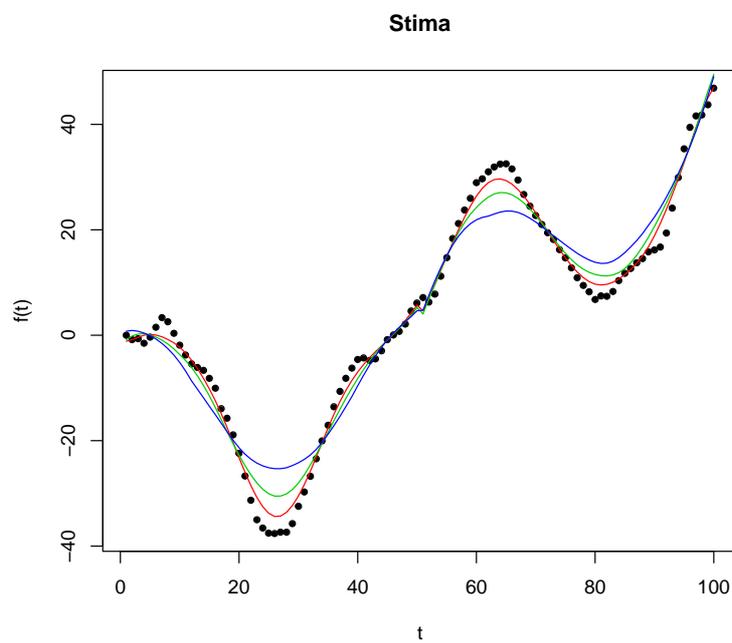


Figura 3.3: Stima diversi parametri di autocorrelazione; rosso  $\alpha=0.4$ , verde  $\alpha=0.7$ , blu  $\alpha=0.95$

lazione ci sono differenze nella stima, se il parametro dell'autocorrelazione è fissato troppo basso ci sarà un sovra-adattamento del modello ai dati (curva verde), viceversa ci si troverà ad avere una stima troppo liscia.

Dopo aver fatto queste considerazioni si deve poter confrontare lo stimatore con un metodo classico. A questo proposito si è pensato di confrontare la stima fatta dalla nuova funzione con un lisciatore di tipo *loess*.

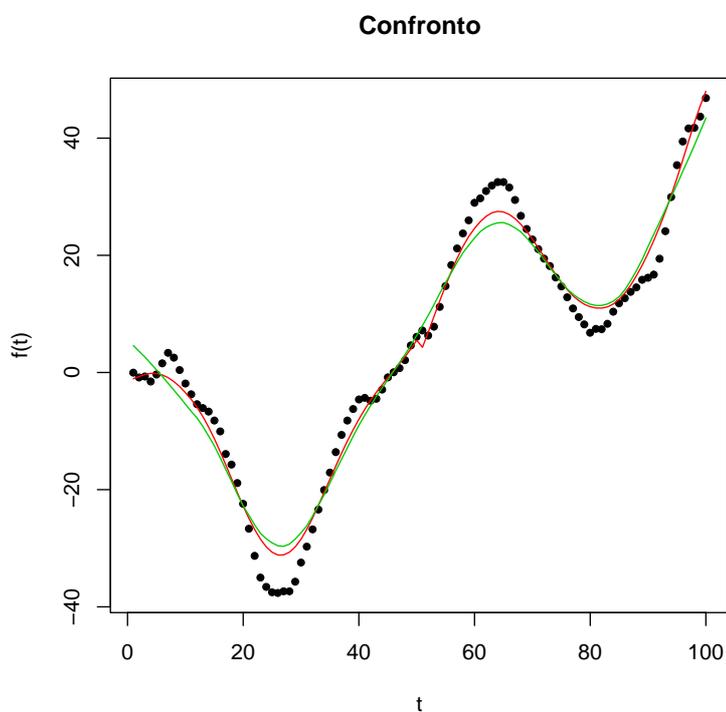


Figura 3.4: Confronto tra la nuova funzione (rosso) e uno stimatore *loess* (verde)

Dal grafico 3.4 si vede che le stime sono praticamente equivalenti, le uniche piccole differenze si notano alle frontiere. Un'altra caratteristica importante da far notare è gli intervalli di confidenza al 95% per queste ultime due stime.

Questi intervalli sono stati creati con la procedura del *residual bootstrap*

(1.5.2), usando il metodo dei percentili. Da notare che la banda di confidenza della nuova funzione è sempre minore di quella calcolata con il metodo basato sul *loess* (figura 3.5). Questo effetto è riconducibile all'autocorrelazione dei dati, peculiarità considerata solo nel caso del nuovo stimatore.

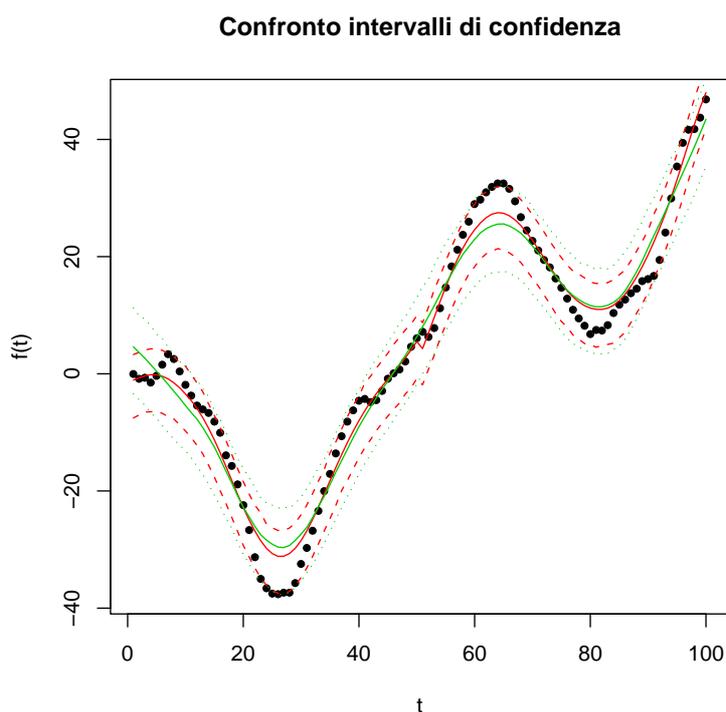


Figura 3.5: Confronto intervalli di confidenza della nuova funzione (rosso) e uno stimatore *loess* (verde)

## 3.2 Funzione bi-dimensionale

### 3.2.1 Sviluppo

Il caso con la variabile esplicative bi-dimensionale ha la stessa logica del caso uni-dimensionale, bisogna soltanto porgere attenzione per alcuni aspetti.

La funzione da stimare è del tipo

$$y(s) = f(s) + \varepsilon(s) \quad (3.4)$$

dove  $s$  è bi-dimensionale e identifica le coordinate spaziali,  $s = (s_1, s_2)$

Il concetto principale che sta alla base dell'autocorrelazione spaziale dei dati si basa sul fatto che osservazioni effettuate su punti vicini presentano effettivamente una minore variabilità rispetto ad osservazioni effettuate tra punti distanti. Quando le unità statistiche sono unità spaziali le variabili presentano generalmente autocorrelazione spaziale, per cui le unità non sono più tra loro indipendenti e perdono, almeno in parte, la loro scambiabilità (Streitberg, 1979).

In presenza di autocorrelazione spaziale, in una determinata area del territorio, si effettuano rilevazioni della variabile d'interesse in corrispondenza di diversi punti di rilevamento. Tali effetti, che si manifestano in modo continuo su tutto il territorio, si presentano ovviamente con modalità simili nelle aree vicine tra loro, mentre la somiglianza generalmente diminuisce al crescere della distanza tra le zone. Pertanto, la presenza di autocorrelazione spaziale porterebbe ad identificare, se esistono, aree del territorio omogenee sotto gli aspetti considerati in cui le singole unità spaziale sono tra loro "contigue".

Si considerano misure della distanza (o della similarità) fra le coppie di unità spaziali che tengono conto dell'informazione relativa alla presenza di autocorrelazione nelle variabili osservate, al fine di valutarne l'influenza sui risultati dell'analisi.

La contiguità viene definita tramite una matrice di pesi quadrata dell'ordine  $n \times n$  simmetrica, che definisce lo schema della connessioni o legami e delle loro intensità (Badaloni e Vinvi, 1988).

Per formalizzare il concetto di contiguità in un modo più complesso si può

definire una matrice di pesi più generalizzata  $W$ , dove i pesi  $w_{ij}$  denotano l'effetto della unità territoriale  $j$  sull'unità  $i$ . Il peso  $w_{ij}$  avrà valore zero se non esiste correlazione tra i punti  $i$  e  $j$  e avrà valore sempre maggiore all'aumentare della relazione tra i due punti.

Nel caso spaziale si fanno due assunzioni, quella di isotropia cioè che la correlazione tra i punti dello spazio non dipenda dalla direzione scelta, e quella di stazionarietà spaziale cioè che la correlazione fra due punti dipende esclusivamente dalla loro distanza e non dalla loro posizione nello spazio.

Dopo aver fatto queste assunzioni, l'approccio usato per modellare la dipendenza spaziale, di uso molto comune, è il metodo del covariogramma isotropico, cosicché la covarianza tra due punti qualsiasi di rilevazione dipende solo dalla distanza tra di essi.

Il covariogramma o funzione di covarianza stazionaria, è la rappresentazione della relazione che intercorre tra una variabile  $Z(s)$  e  $Z(s+d)$ , dove  $d$  è la distanza che intercorre tra le due variabili. Inoltre, per i dati analizzati, si fa in modo che il covariogramma dipenda solo dalla distanza  $d$  e non dalla direzione, in questo caso si parla di isotropia.

La funzione del covariogramma, nel caso di isotropia, è la seguente (Cressie 1993):

$$C(d) = \gamma^2 \rho(\|d\|)$$

Nel caso preso in esame si userà modello esponenziale e quindi risulterà:

$$C(d) = \gamma^2 \exp\left(-\frac{\|d\|}{\phi}\right)$$

Bisogna anche definire che tipo di misura per la distanza usare, tenendo conto che i dati si riferiscono a coordinate spaziali è opportuno utilizzare la

distanza euclidea. Presi due punti  $a$  e  $b$  con coordinate  $x$  e  $y$  la distanza euclidea tra i due punti è definita come

$$\|a - b\| = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2}$$

Con il covariogramma si crea la matrice dei pesi che tiene conto dell'autocorrelazione, i due parametri  $\gamma^2$  e  $\phi$  sono stimati, come nel caso unidimensionale, stimando vari modelli, confrontandoli tra loro e decidendo quali valori dei parametri adatta meglio il modello ai dati.

Per la matrice dei pesi relativa ai *nearest neighbour* la procedura è sostanzialmente uguale, cioè si genera una matrice diagonale  $n \times n$  dove nella diagonale vengono inseriti i pesi ricavati in funzione della distanza euclidea tra i vari siti e del parametro di liscio  $h$ . Anche il valore di  $h$  non è noto e si può stimare tramite confronto tra vari modelli.

A questo punto le due matrici si uniscono tra loro come nel caso precedente e si stima il modello esattamente come regressione lineare con il metodo dei minimi quadrati generalizzati con i pesi che sono stati appena calcolati e con variabile esplicative le coordinate spaziali.

A causa della doppia dimensionalità della variabile rispetto al caso precedente la procedura risulta maggiormente onerosa in termini computazionali.

### 3.2.2 Verifica

Per valutare l'efficacia dello stimatore appena descritto si simulano dei dati spaziali dipendenti tramite un covariogramma di tipo esponenziale descritto nel paragrafo precedente.

Questi dati vengono prima stimati con il lisciatore appena descritto e poi con uno stimatore di tipo *loess*. Successivamente si effettua un confronto tra le varie stime fatte.

I dati simulati sono rappresentati in figura 3.6.

**Simulazione dati autocorrelati**

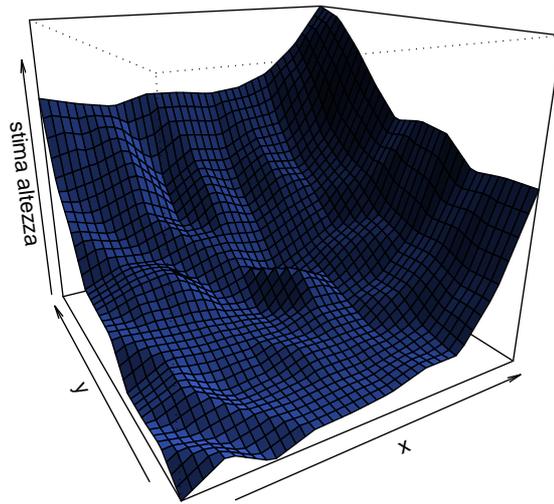


Figura 3.6: Dati Simulati

Prima di calcolare le varie stime bisogna far notare che il parametro  $\gamma^2$  non ha molta importanza perchè è soltanto un fattore di scala, molto più importanti sono il parametro  $\phi$  relativo alla forma della curva del covariogramma e il parametro relativo alla finestra di dati da prendere per la stima.

Successivamente si riportano alcuni stime dei dati.

Nella stima di figura 3.7 i parametri sono troppo grandi, sia per la finestra di stima che per l'autocorrelazione. In questo caso la stima risulta semplicemente un piano e rapportandola ai dati simulati, si può considerare troppo liscia.

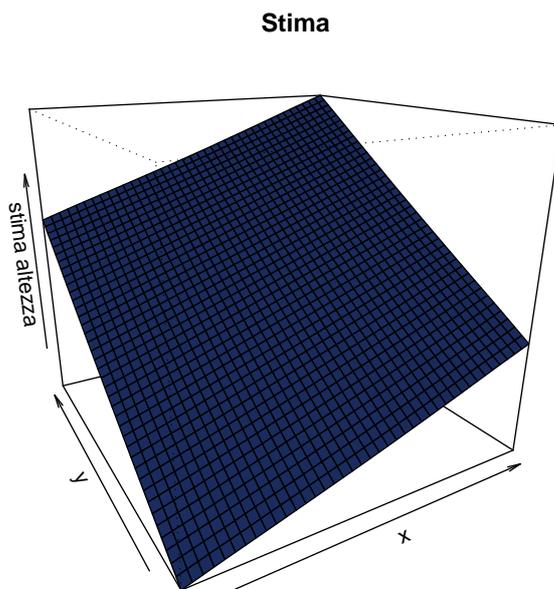


Figura 3.7: Stima con parametri  $h=5$ ,  $\phi = 3$  e  $\gamma = 3$

Invece la figura 3.8 presenta un sovra-adattamento dei dati al modello, che, come spiegato in precedenza, addattandosi troppo ai dati aumenta la

varianza senza guadagno rilevante per la distorsione. In questo caso, infatti, la stima sembra interpolare in maniera quasi esatta i dati.

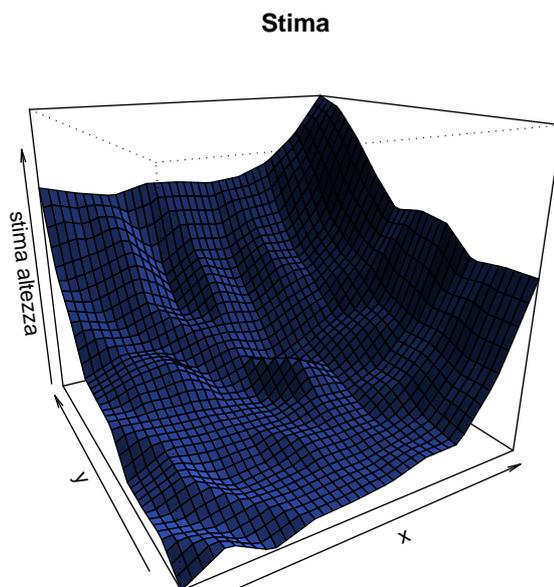


Figura 3.8: Stima con parametri  $h=0.5$ ,  $\phi = 0.4$  e  $\gamma = 1$

La figura 3.9 risulta una stima buona della simulazione, ha un andamento simile ai dati senza incorrere in oscillazioni casuali di essi. Questa figura e la 3.10, ottenuta da una stima basata sul metodo *loess*, non presentano differenze significative.

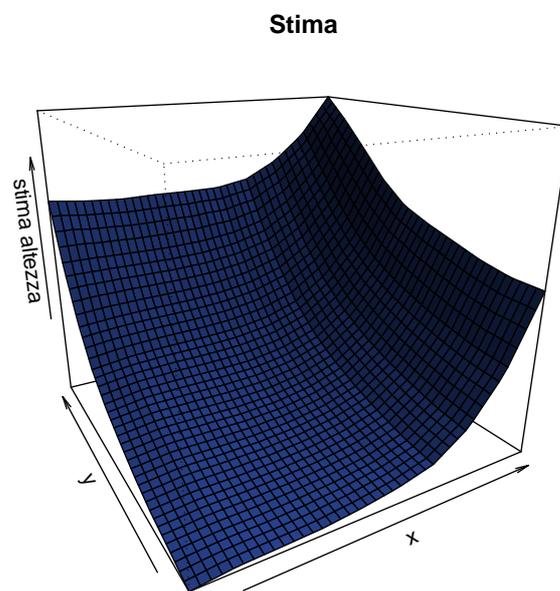


Figura 3.9: Stima con parametri  $h=0.8$ ,  $\phi = 0.7$  e  $\gamma = 1$

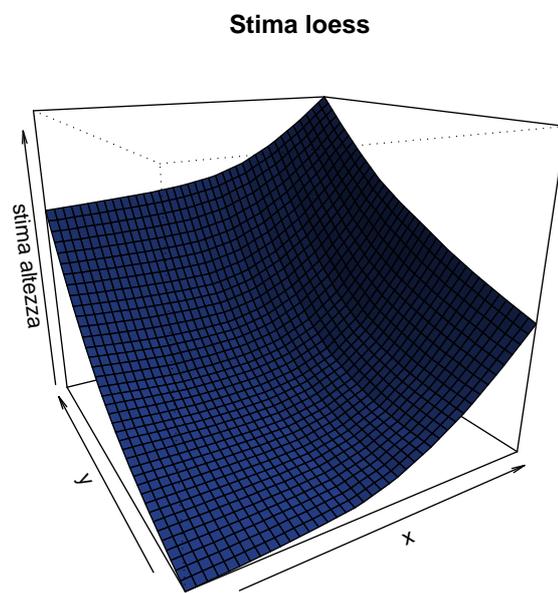


Figura 3.10: Stima metodo *loess*

Anche in questo caso si vuole rapportare le varianze delle due stime, per fare questo si è deciso di realizzare dei grafici per poter raffigurare questa eventuale differenza. I 100 dati simulati sono stati raggruppati in 25 gruppi ognuno di 4 osservazioni. Per ciascuna di queste 25 sotto-aree si è costruito un boxplot per raffigurare la variabilità in quella zona dello spazio. Questi boxplot visualizzano la media e i percentili del 2,5% e del 97,5% calcolati tramite la procedura *residual bootstrap* (cap 1.5.2), usando il metodo dei percentili.

I grafici di figura 3.11 si riferiscono alla nuova funzione, invece la figura 3.12 si riferisce alla stima effettuata con il *loess*.

Per poter fare un confronto le scale dei vari boxplot sono le stesse, da questi grafici si nota che la variabilità della nuova stima rispetto al *loess* è decisamente minore.

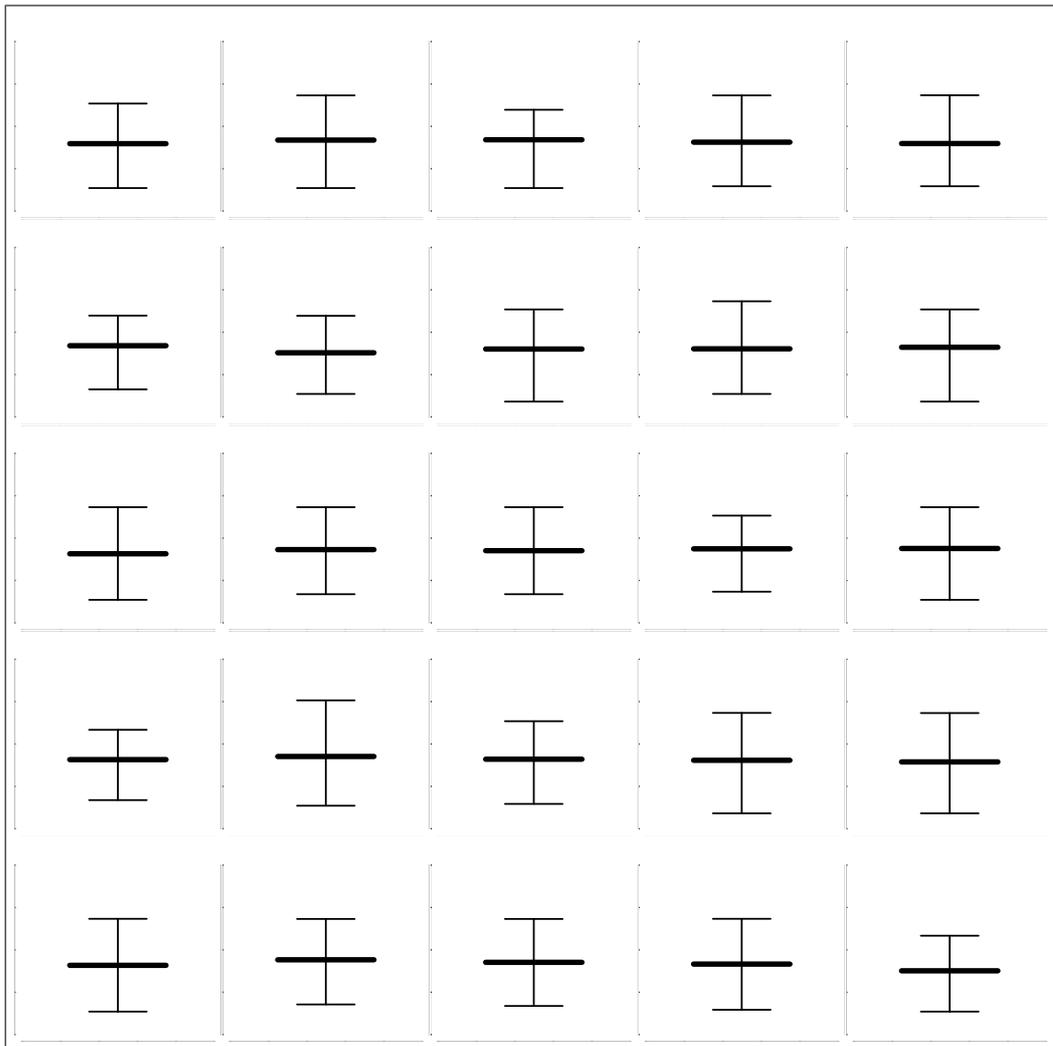


Figura 3.11: Box plot effettuati usando la nuova stima

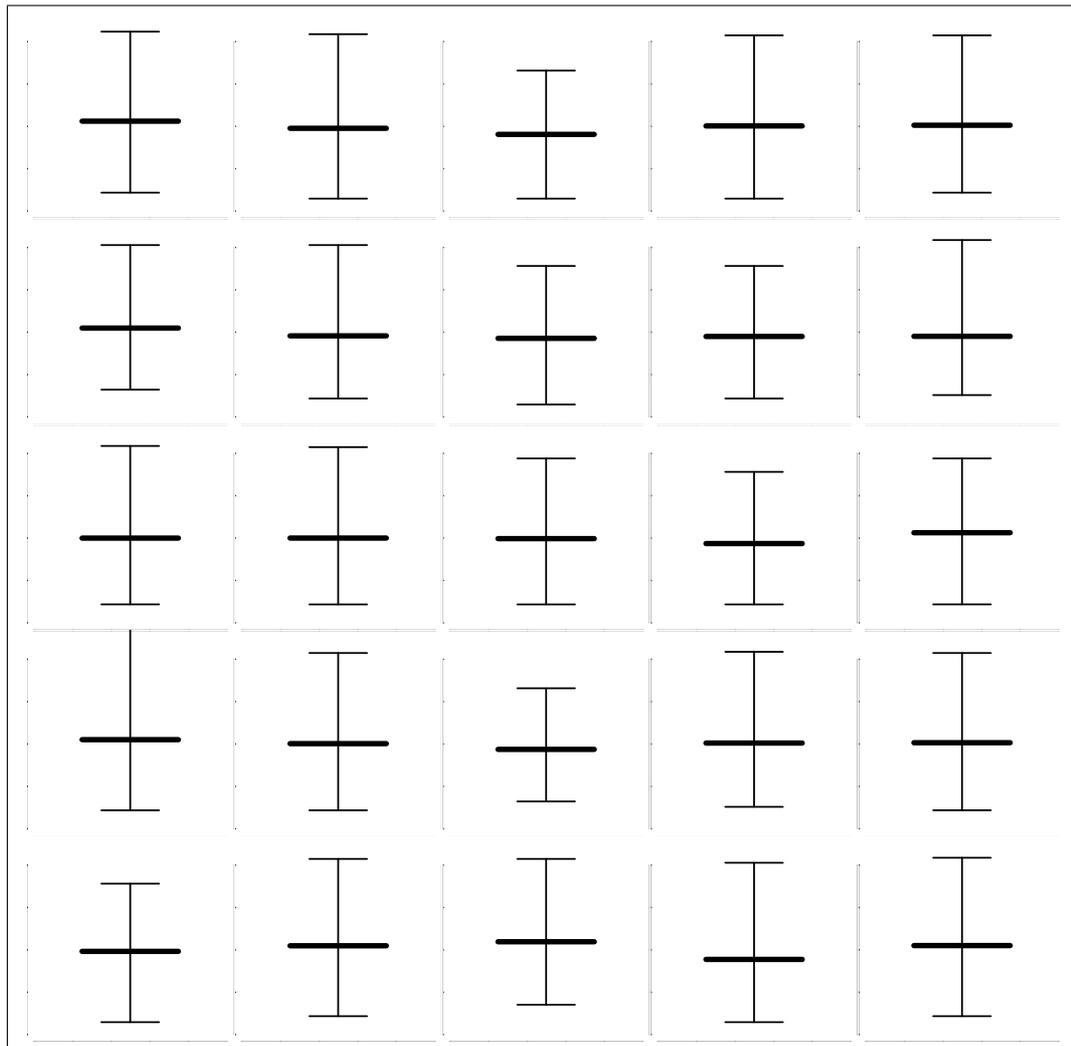


Figura 3.12: Box plot effettuati usando la stima *loess*

# Capitolo 4

## Specificazione del modello

In questo capitolo si effettuerà un'analisi su una serie spazio-temporale relativa alla velocità del vento in Irlanda, allo scopo di verificare il miglioramento apportato dalla stima del nuovo lisciatore presentato.

In questo capitolo prima si presentano i dati, poi il modello, subito dopo le stime effettuate e successivamente il confronto tra le stime.

### 4.1 I dati

#### 4.1.1 Descrizione dei dati

L'insieme di dati che si considera riguarda la rilevazione oraria della velocità del vento in 12 stazioni meteorologiche in Irlanda (4.1) nel periodo compreso tra il 1961 e il 1978. La velocità del vento è stata registrata in nodi, che sono poi stati convertiti in metri al secondo per le analisi. Per gli scopi necessari a questo lavoro si è deciso di utilizzare unicamente le medie giornaliere per ogni sito per un periodo di rilevazione di un anno, verrà utilizzato l'anno 1978.

Semplici analisi esplorative effettuate da Haslett e Raftery (1978) suggerir-

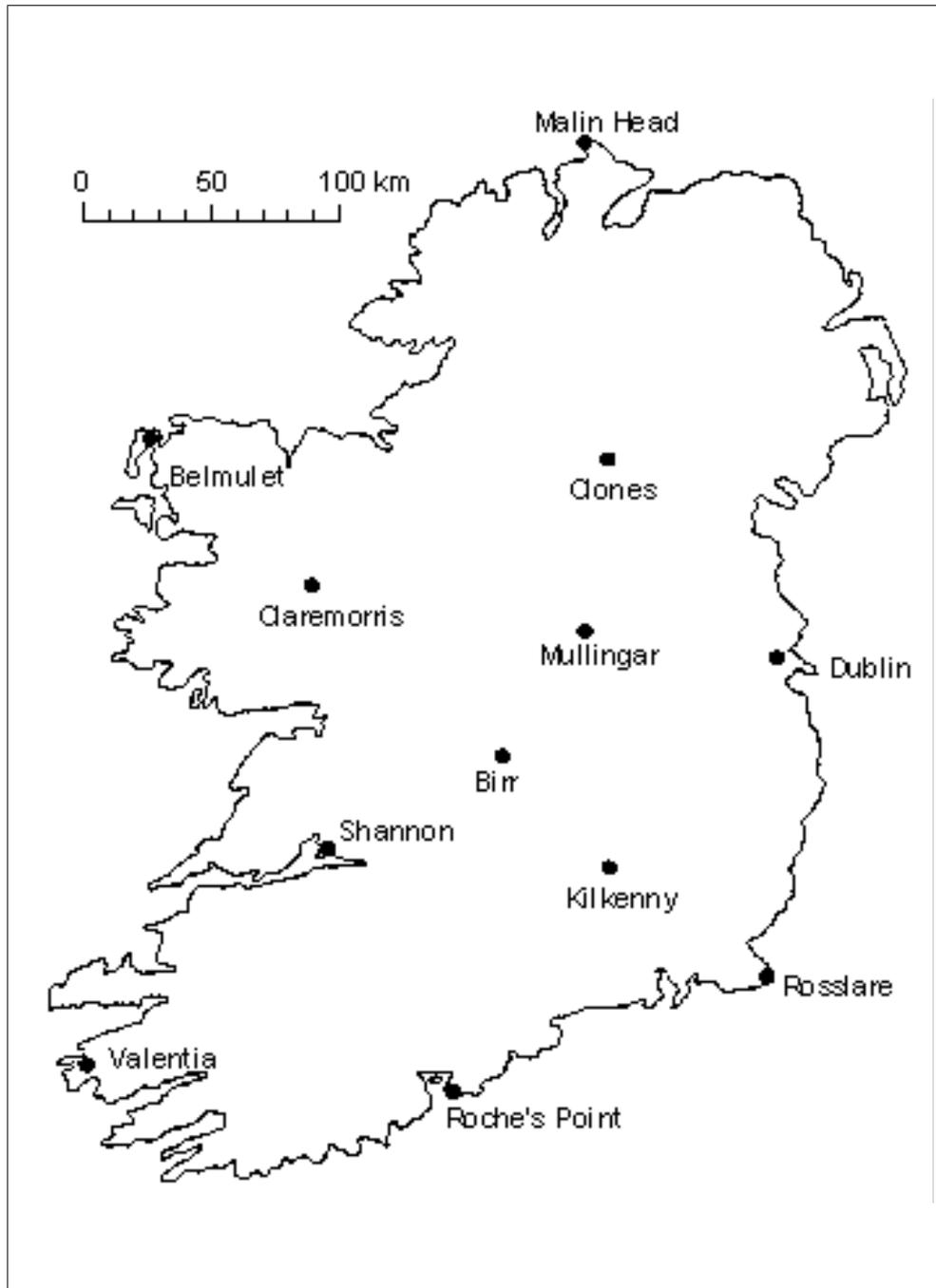


Figura 4.1: Mappa dell'Irlanda con contrassegnati i siti di rilevamento

scono di calcolare la radice quadrata delle velocità dei venti per stabilizzarne la varianza sia nel tempo che nello spazio e correggere l'asimmetria che appare nelle distribuzioni marginali rendendole approssimativamente normali. Gli stessi autori mostrano che la struttura di covarianza spaziale tra le osservazioni può essere approssimata ragionevolmente bene da un covariogramma esponenziale (Cressie, 1993). Per quando riguarda la struttura di autocorrelazione temporale si osserva che la dipendenza può essere descritta da un modello autoregressivo di ordine uno. Oltre allo studio effettuato da Haslett e Raftery (1989) sui dati relativi alla velocità del vento in Irlanda da 1961 al 1978, ne sono stati effettuati altri sempre con lo scopo di studiare i modelli spazio-temporali (Smyth, 1994; Scarpa, 1998; de Luna e Genton, 2005; Gneiting, Genton e Guttorp, 2005; Bouette, Chassagneux, Sibai, Terron, Charpentier, 2006).

## 4.2 Il modello

Come riferimento al capitolo 1 si considera un modello additivo del tipo

$$y(t, s) = f(t) + g(s) + \varepsilon(t, s) \quad (4.1)$$

Si consideri ora un modello di tipo (4.1) dove la  $y$  è la velocità del vento,  $f(t)$  è il contributo dato dal tempo, dove la variabile  $t$  è univariata,  $g(s)$  è il contributo dato dallo spazio, dove la variabile  $s$  è bivariata, e  $\varepsilon(t, s)$  gli errori e siano considerati autocorrelati tra loro. Si assuma anche che la forma funzionale della struttura di correlazione sia nota per esempio attraverso la formulazione di un modello per il processo degli errori  $\varepsilon(t, s)$ . Una volta scelto come stimatore non-parametrico per le funzioni  $f(t)$  e  $g(s)$  lo stimatore basato sulle rette di regressione con pesi autocorrelati e locali (cap. 2.2), è

possibile un inserimento delle stime della variabilità della componente di errore  $\varepsilon(t, s)$  nella stima delle funzioni.

Si continuano a considerare, per semplificare il modello, le ipotesi di stazionarietà spaziale e temporale e isotropia nello spazio.

Nel nostro caso consideriamo modello autoregressivo di ritardo (o lag) uno (AR(1)) (cap. 3.1.1) per il tempo.

Diversamente, per lo spazio si considera in interazione descrivibile tramite un covariogramma isotropico esponenziale (cap. 3.2.1), con i due parametri  $\gamma^2$  e  $\phi$  che sono rispettivamente l'altezza e l'ampiezza della curva e vengono calcolati tramite stime ripetute.

Con il metodo delle rette di regressione con pesi autocorrelati e locali (cap. 2.2), l'algoritmo di *backfitting* (cap. 1.3.1) e il metodo dei minimi quadrati generalizzati (cap. 2.1.3) si stima il modello 4.1 tramite un modello additivo con lisciatori rette di regressione con pesi autocorrelati e locali. Grazie all'algoritmo di *backfitting* si possono calcolare le stime delle due funzioni, una per il tempo e una per lo spazio, congiuntamente.

Dopo aver eseguito varie stime, modificando i vari parametri dei lisciatori, si è arrivati alla stima che verrà descritta in seguito.

### 4.3 Le stime

Le stime effettuate tramite un modello additivo con lisciatori autocorrelati e pesati descritti in questa tesi sono rappresentate nelle figure seguenti. Vengono raffigurate le stime della componente temporale e spaziale prodotte dall'algoritmo di *backfitting* per quanto riguarda le due funzioni del modello 4.1 (fig. 4.2 e 4.3).

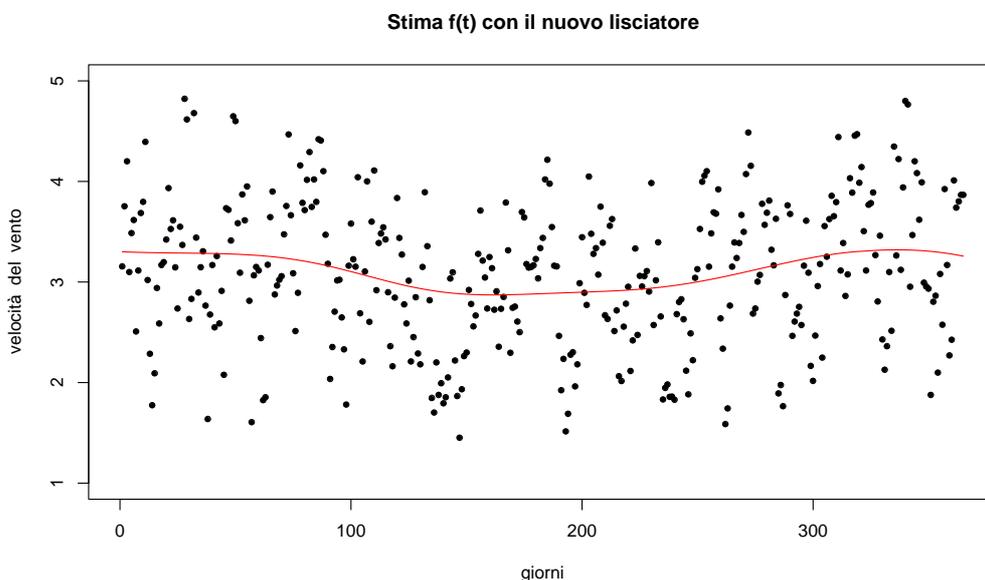


Figura 4.2: Stima della componente temporale tramite la nuova procedura

Per verificare l'effettivo miglioramento del metodo proposto in questa tesi si è deciso di metterlo in confronto con modello addittivo, con lisciatori di tipo *loess*.

Oltre a verificare la stima di dati si considera rilevante anche fornire una stima degli intervalli di confidenza dei due diversi metodi. Per realizzarli ci si basa sul metodo *bootstrap* (1.5).

In riferimento alla componente temporale rappresentiamo (fig. 4.4) le due stime con i rispettivi intervalli di confidenza al 95% tramite i percentili 2,5% e 97,5% calcolati con il *residual bootstrap* (1.5.2). Si nota che con un modello addittivo con il nuovo lisciatore, che considera l'autocorrelazione, si ottengono degli intervalli di confidenza minori rispetto alla stima effettuata con un modello addittivo con lisciatori di tipo *loess* che non considera questa autocorrelazione.

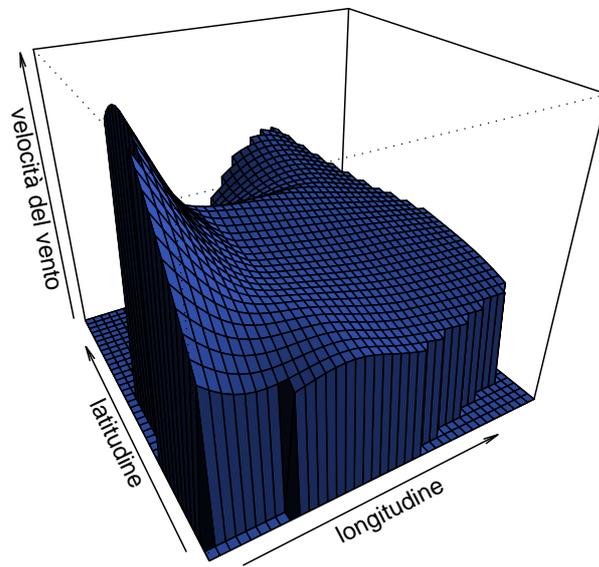
**Stima  $g(s)$  con il nuovo lisciatore**

Figura 4.3: Stima della componente spaziale tramite la nuova procedura

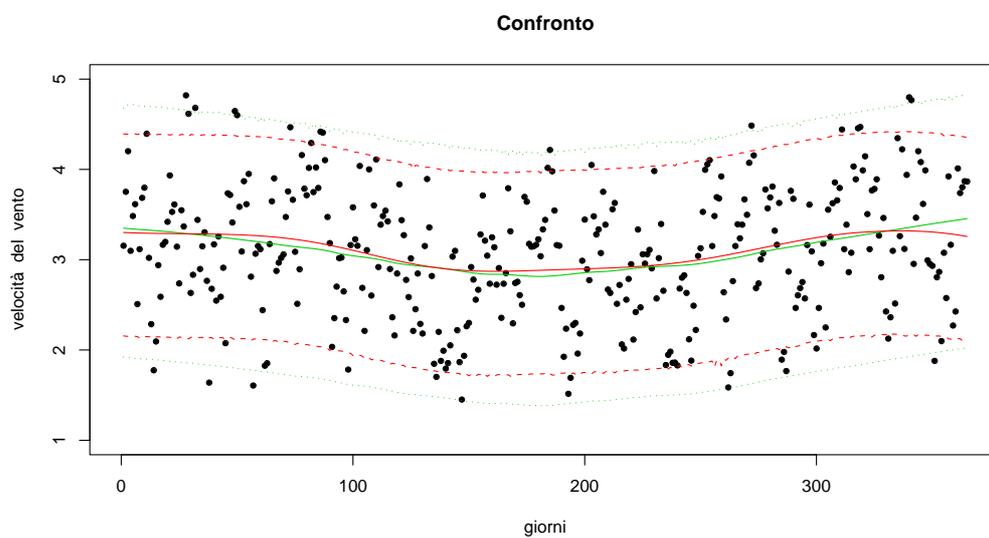


Figura 4.4: Confronto tra la stima della componente temporale tramite la nuova procedura (rosso) e un modello additivo (verde) con relativi intervalli di confidenza al 95%

Nelle figure 4.5 e 4.6 si mettono a confronto le stime della componente spaziale effettuate con un metodo additivo con il lisciatore autocorrelato proposto nella tesi e con lisciatori di tipo *loess*.

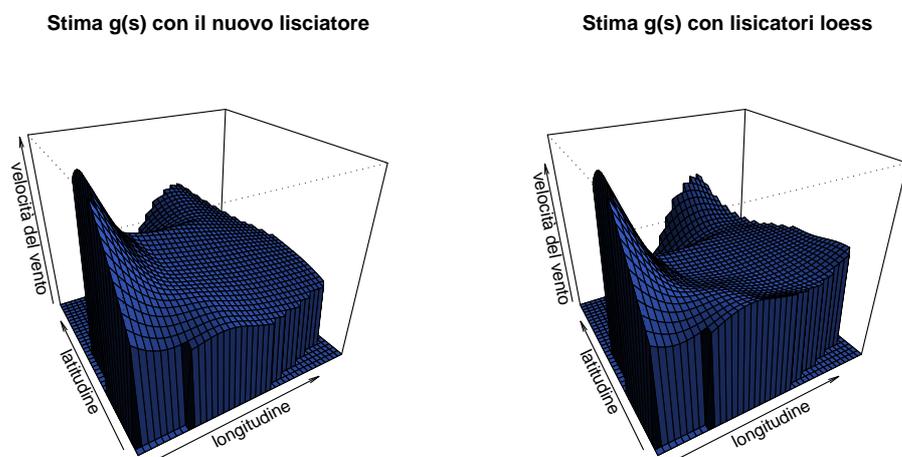


Figura 4.5: Confronto tra la stima della componente spaziale tramite un modello additivo con il nuovo lisciatore (sinistra) e con lisciatore *loess* (destra)

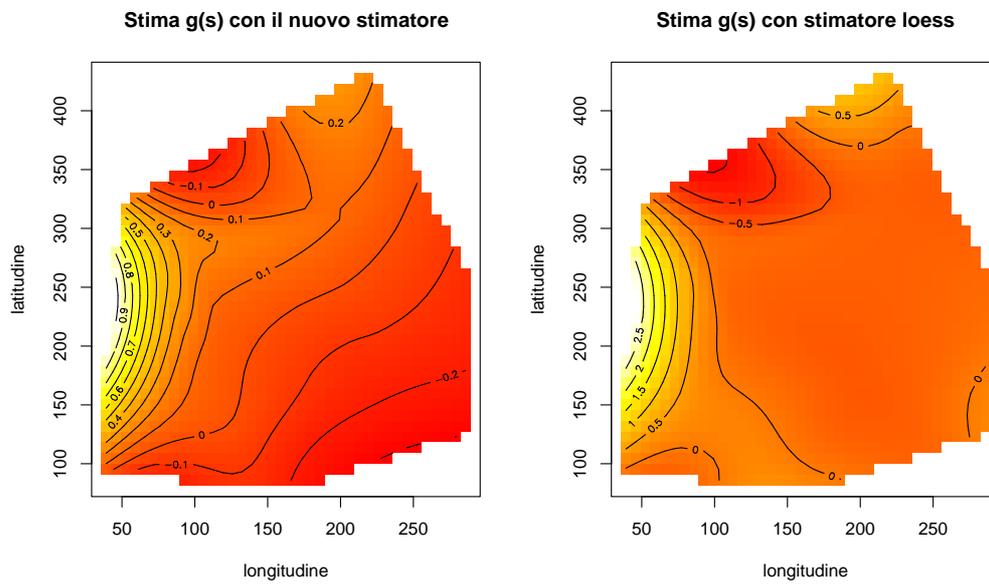


Figura 4.6: Confronto tra la stima della componente spaziale tramite un modello additivo con il nuovo lisciatore (sinistra) e con lisciatore *loess* (destra)

Anche le stime proposte per la componente spaziale risultano simili con entrambi i metodi. Come nel caso unidimensionale si propone una stima degli intervalli di confidenza dei due diversi metodi. Per realizzarli ci si basa ancora sul metodo *residual bootstrap* (1.5.2).

Si visualizza per ogni sito la propria variabilità con un intervallo di confidenza al 95% tramite un boxplot con contrassegnato la media e i percentili 2,5% e 97,5% (fig. 4.7).

Nel caso spaziale si nota che con il nuovo stimatore si ottengono degli intervalli di confidenza minori rispetto alla stima effettuata sempre con un modello additivo con lisciatori di tipo *loess* a causa dell'autocorrelazione tra le osservazioni, analogamente al caso unidimensionale.

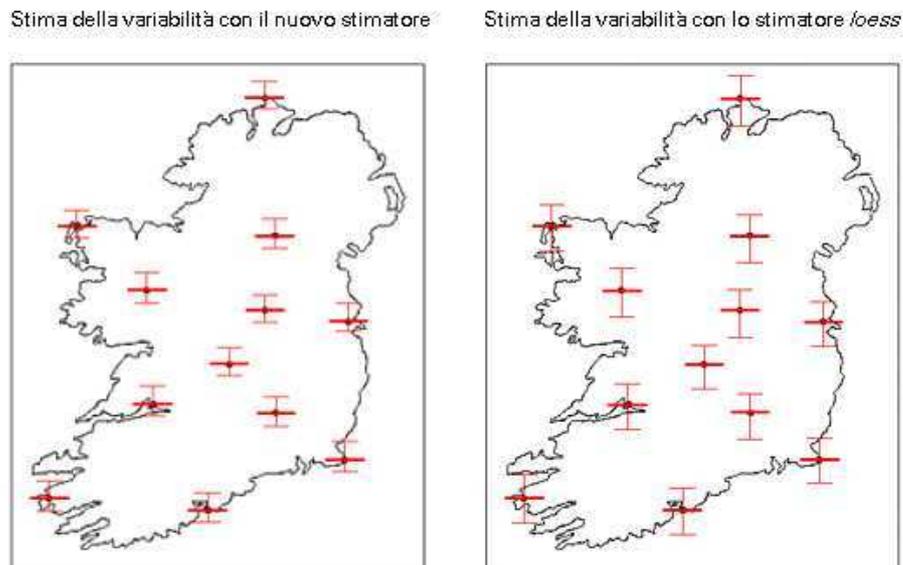


Figura 4.7: Confronto tra la variabilità della componente spaziale tramite la nuova procedura (sinistra) e un modello additivo (destra)

# Capitolo 5

## Conclusioni

Nel capitolo precedente è stato proposto il confronto tra due modelli additivi per una serie spazio-temporale. Nel primo modello il tipo di lisciatore utilizzato è quello proposto in questa tesi che tiene conto anche dell'auto-correlazione e nel secondo modello additivo gli stimatori sono di tipo *loess*. Da questo confronto si deduce che il nuovo metodo è equivalente per quanto riguarda la stima dei dati ma in termini di efficienza risulta migliore, difatti, produce degli intervalli di confidenza minori rispetto alla procedura basata sugli stimatori *loess*.

Per completare l'analisi proposta in precedenza si potrebbe cercare un metodo per la stima dei parametri. In riferimento al parametro  $h$ , ovvero il parametro per definire i pesi a seconda della distanza delle osservazioni si possono usare metodi come quello della convalida incrociata (1.2.4). Invece, per i parametri relativi all'autocorrelazione,  $\alpha$  per la componente temporale e  $\gamma$  e  $\phi$  per la componente spaziale, si dovrebbe calcolare la loro stima con altri metodi (ad esempio il metodo della verosimiglianza). Inoltre si potrebbero rendere meno specifiche le funzioni che descrivono l'autocorrelazione, per esempio utilizzare un modello autoregressivo di ordine  $p$  per il tempo, e per

lo spazio generalizzare il covariogramma facendo in modo di non soffermarci all'utilizzo di un covariogramma isotropico esponenziale ma servendosi anche altri modelli (ad esempio gaussiano, onda, sferico, ecc.). Un'ulteriore generalizzazione potrebbe essere data da non utilizzare delle ipotesi di stazionarietà e isotropia.

Un'altra strada da percorrere potrebbe essere quella di costruire la matrice dei pesi  $W$  in un sistema differente da quello utilizzato (per esempio inserendo un fattore di scala nella moltiplicazione delle due matrici di partenza).

Un'interessante continuazione, riguardate il metodo proposto in questo studio, sarebbe generalizzare tale metodologia per un modello additivo generalizzato, cioè aggiungere delle informazioni riguardanti l'autocorrelazione delle serie spazio-temporali a un modello additivo generalizzato, quindi considerando una classe più generale, non più gaussiana, per la distribuzione della variabile risposta come pure la funzione che serve da legame tra le variabili esplicative.

# Appendice A

## Funzioni in R

### Funzione unidimensionale

```
funz.unidim <- function(dati, x, h=1,
                        alpha=0, pesomin=0.005)
{
  if (abs(alpha)>1)
    stop(alfa non valido;non compreso tra -1:1)
  # controllo su parametro autocorrelazione
  n <- length(dati)
  pred <- rep(0, n)
  tmp <- rep(0, n)
  auto <- matrix(0, n, n)
  dati.pesomin<-NULL
  # definizioni di matrici e vettori
  auto <- pesi.auto.unid(x, dati, n, alpha)
  # pesi autocorrelati del tipo  $w[i, j] \leftarrow \alpha^{(j-k)}$ 
  for (i in 1 : n)
  {
    a <- x[i]
    vdrop <- NULL
```

```

w <- matrix(0, n, n)
tmp <- dnorm((x-a), mean=0, sd=h)
for (j in 1 : n)
  w[j,j]<-tmp[j]
# calcolo i pesi w dati dalla distanza (normale)
W <- auto%*%w
diagonale <- diag(W)
if (any(diagonale <= pesomin))
{
  dati.pesomin <- matr.pesomin.unid(dati, x, w,
                                   auto, W, n, pesomin)

  Wgls <- dati.pesomin$W
  datigls <- dati.pesomin$dati
  xgls <- dati.pesomin$x
  autogls <- dati.pesomin$auto
  wgls <- dati.pesomin$w
}
else
{
  Wgls <- W
  datigls <- dati
  xgls <- x
  autogls <- auto
  wgls <- w
}
# matrice dei pesi
eW <- eigen(Wgls,FALSE)
dgls <- eW$values
# autovalori
if(any(dgls <= 0))
{
  dati.gls <- matr.pesi0.unid(datigls, xgls,
                             wgls, autogls, dgls)
  # funzione che elimina i valori
  # vicini a 0 sulla matrice dei pesi
  Wgls <- dati.gls$W
  datigls <- dati.gls$dati

```

```

    xgls <- dati.gls$x
  }
  else
  {
    Wgls <- Wgls
    datigls <- datigls
    xgls <- xgls
  }
  fit <- lm.gls.as( datigls ~ xgls, W = Wgls)
  # calcolo la regressione lineare pesata
  pred[i] <- fit$coefficients[1] +
             a * fit$coefficients[2]
  # pred contiene i valori
  # di previsione per ogni punto
  }
pred
}

```

```

pesi.auto.unid <- function(x, dati, n, alpha)
{
  auto <- matrix(0, n, n)
  for (j in 1 : (n-1))
  {
    for(k in (j+1) : n)
    {
      auto[j,k] <- auto[k,j] <- alpha^(k-j)
    }
    auto[j,j] <- 1
  }
  auto[n,n] <- 1
  auto
}

```

```

matr.pesi0.unid <- function(dati, x, w, auto, d)
{
  n <- length(dati)
  dati.gls <- NULL
  while (any(d <= 0))
  {
    k <- NULL
    minw <- min(diag(w))
    for(j in 1 : n)
      if(minw==w[j,j])
        k <- c(k, j)
    w <- w[-k,-k]
    auto <- auto[-k,-k]
    dati <- dati[-k]
    x <- x[-k]
    W <- auto%*%w
    eW <- eigen(W, FALSE)
    d <- eW$values
    n <- n-length(k)
  }
  dati.gls$W <- W
  dati.gls$dati <- dati
  dati.gls$x <- x
  dati.gls
}

```

```

matr.pesomin.unid <- function(dati, x, w,
                              auto, W, n, pesomin)
{
  dati.pesomin <- NULL
  k <- NULL
  diagonale <- diag(W)
  for(j in 1 : n)
  {
    if (is.na(diagonale[j]))

```

```

        diagonale[j] <- 0
        if ((diagonale[j]<pesomin))
            k <- c(k, j)
    }
w <- w[-k,-k]
auto <- auto[-k,-k]
dati <- dati[-k]
x <- x[-k]
W <- auto%*%w
dati.pesomin$w <- w
dati.pesomin$W <- W
dati.pesomin$dati <- dati
dati.pesomin$x <- x
dati.pesomin$auto <- auto
dati.pesomin
}

```

## Funzione bidimensionale

```

funz.bidim <- function(dati, x, y, h=1,
                      phi=1, sigma=1, pesomin=0.005)
{
n <- length(y)
pred <- rep(0, n)
tmp <- rep(0, n)
auto <- matrix(0, n, n)
dati.pesomin <- NULL
auto <- pesi.auto.bid(x, y, n, phi, sigma)
# calcolo pesi autocorrelazione
for (i in 1 : n)
{
puntostima <- c(x[i],y[i])
vdrop<-NULL

```

```

w <- matrix(0, n, n)
pesi <- pesi.w.bid(x, y, n, puntostima, h)
# calcolo i pesi a dati dalla distanza (normale)
w <- pesi$w
distpunto <- pesi$distpunto
W <- auto%*%w
diagonale <- diag(W)
Wgls <- W
datigls <- dati
ygls <- y
xgls <- x
autogls <- auto
wgls <- w
if (any(diagonale <= pesomin))
{
  dati.pesomin <- matr.pesomin.bid (datigls, xgls,
                                   ygls, wgls, autogls, Wgls, n, pesomin)
  Wgls <- dati.pesomin$W
  datigls <- dati.pesomin$dati
  ygls <- dati.pesomin$y
  xgls <- dati.pesomin$x
  autogls <- dati.pesomin$auto
  wgls <- dati.pesomin$w
}
# matrice dei pesi
eW <- eigen(Wgls)
dgls <- eW$values
# autovalori
if(any(dgls <= 0))
{
  dati.gls <- matr.pesi0.bid(datigls, xgls,
                             ygls, wgls, autogls, dgls)
  # funzione che elimina i valori
  # vicini a 0 sulla matrice dei pesi
  Wgls <- dati.gls$W
  datigls <- dati.gls$dati
  ygls <- dati.gls$y
}

```

```

    xgls <- dati.gls$x
  }
  fit <- lm.gls.as(dati.gls ~ xgls + ygls, W = Wgls)
  # calcolo la regressione lineare pesata
  pred[i] <- fit$coefficients[1] +
            puntostima[1] * fit$coefficients[2]+
            puntostima[2] * fit$coefficients[3]
  # pred contiene i valori
  # di previsione per ogni punto
}
pred
}

```

```

matr.pesi0.bid <- function(dati, x, y, w, auto, d)
{
  n <- length(y)
  dati.gls <- NULL
  while (any(d <= 0))
  {
    k <- NULL
    minw <- min(diag(w))
    for(j in 1 : n)
      if(minw == w[j, j])
        k <- c(k, j)
    w <- w[-k, -k]
    auto <- auto[-k, -k]
    dati <- dati[-k]
    y <- y[-k]
    x <- x[-k]
    W <- auto%*%w
    eW <- eigen(W)
    d <- eW$values
    n <- n-length(k)
  }
  dati.gls$W <- W

```

```
dati.gls$dati <- dati
dati.gls$y <- y
dati.gls$x <- x
dati.gls
}
```

```
matr.pesomin.bid <- function(dati, x, y,
                             w, auto, W, n, pesomin)
{
  dati.pesomin <- NULL
  k <- NULL
  for(j in 1 : n)
  {
    diagonale <- diag(W)
    if (diagonale[j] <= pesomin)
      k <- c(k, j)
  }
  w <- w[-k, -k]
  dati <- dati[-k]
  y <- y[-k]
  x <- x[-k]
  auto <- auto[-k, -k]
  W <- auto%*%w
  dati.pesomin$w <- w
  dati.pesomin$dati <- dati
  dati.pesomin$W <- W
  dati.pesomin$y <- y
  dati.pesomin$x <- x
  dati.pesomin$auto <- auto
  dati.pesomin
}
```

```

distanza.euclidea <- function(a, b)
{
  dist <- sqrt((a[1]-b[1])^2+(a[2]-b[2])^2)
  dist
}

pesi.auto.bid <- function(x, y, n, phi, sigma)
{
  auto <- matrix(0, n, n)
  for (j in 1 : (n-1))
  {
    for(k in (j+1) : n)
    {
      dist <- distanza.euclidea(c(x[j],y[j]),
                                c(x[k],y[k]))
      auto[j,k] <- sigma * exp(-dist/phi)
      auto[k,j] <- auto[j,k]
    }
    auto[j,j] <- sigma
  }
  auto[n,n] <- sigma
  auto
}

pesi.w.bid <- function(x, y, n, puntostima, h)
{
  pesi <- NULL
  w <- matrix(0, n, n)
  for (j in 1 : n)
  {
    distpunto <- distanza.euclidea(puntostima,
                                    c(x[j],y[j]))
    w[j,j] <- dnorm(distpunto, mean=0, sd=h)
  }
}

```

```
}  
pesi$w <- w  
pesi$distpunto <- distpunto  
pesi  
}
```

# Bibliografia

- [1] Austin M.P., Meyers J.A. (1996), Current approaches to modeling the environmental niche of eucalypts: implication for management of forest biodiversity, *Forest Ecology and Management*, 85, pp. 95-106.
- [2] Azzalini A., Scarpa B. (2004), *Analisi dei dati e data mining*, Springer.
- [3] Badaloni M., Vinci E. (1998), Contributi all'Analisi dell'Autocorrelazione Spaziale, *Metron*, 46.
- [4] Bertsch M., Dal Passo R. (2001), *Elementi di Analisi Matematica*, Aracne Ed.
- [5] Bio A. M. F., Alkemade R., Barendregt A. (1998), Determining alternative models for vegetation response analysis: a non parametric approach, *Journal of Vegetation Science*, 9, pp. 5-16.
- [6] Brosse S., Lek S. (2000), Modelling roach (*Rutilus rutilus*) microhabitat using linear and nonlinear techniques, *Freshwater Biology*, 44, pp. 441-452.
- [7] Bouette J.C., Chassagneux J.F., Sibai D., Terron R., Charpentier A. (2006), Wind in Ireland : long memory or seasonal effect?, *Stochastic environmental research and risk assessment*, 20, pp. 141-151

- [8] Buja A., Hastie T. J., Tibshirani R. J. (1989), Linear smoothers and additive models, *Annals of Statistics*, 7, pp. 453-555.
- [9] Buja A., Donnell D.J., Stuetzle W. (1994), Analysis of Additive Dependencies and Concurvities Using Smallest Additive Principal Components, *The Annals of Statistics*, 22, pp. 1635-1673.
- [10] Cao R. (1999), An overview of bootstrap methods for estimating and predicting in time series, *Societad de Statistica e Investigation Operativa*, 8, 1, pp. 95-116.
- [11] Carlstein E. (1986), The use of subseries values for estimating the variance of a general statistics from a stationary sequence, *Annals of Statistics*, 14, pp. 1171-1179.
- [12] Carlstein E., Kim-Anh D., Hall P., Hersteberg T., Kunsch H. R. (1998), Matched-block bootstrap for dependent data, *Bernoulli*, 4, pp. 305-328.
- [13] Chatfield C. (1989), *Analysis of Time Series: an Introduction*, Chapman and Hall, London, (4th ed.).
- [14] Chen Z., Gu C., Wahba G. (1989), Comments to linear smoothers and Additive Models, *The Annals of Statistics*, 17, pp. 515-521.
- [15] Chernick M. R. (1999), *Bootstrap Methods: A Practitioner's Guide*, Wiley, John & Sons, Incorporated
- [16] Cleveland W. S. (1979), Robust locally weighted regression and smoothing scatterplots, *Journal of American Statistical Association*, 74, pp. 829-836.

- [17] Cleveland W. S., Devlin S. J. (1988), Locally-Weighted Regression: an approach to Regression Analysis by Local Fitting, *Journal of American Statistical Association*, 83, pp. 597-510.
- [18] Cressie n. (1993), *Statistics for Spatial Data* , Wiley.
- [19] De Luna X., Genton M. G. (2005), Predictive spatio-temporal models for spatially sparse environmental data, *Statistica Sinica*, 15, pp. 547-568.
- [20] Efron B. (1982), *An introduction to the bootstrap*, Chapman & Hall, New York.
- [21] Efron B., Tibshirani R. J. (1986), Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, *Statistical Science*, 1(1), pp. 54-77.
- [22] Efron B., Tibshirani R. J. (1993), *An introduction to the bootstrap*, Chapman & Hall, New York.
- [23] Engle R., Granger C., Rice J., Weiss A. (1986), Semi-parametric estimates of the relation between weather and electricity demand, *Journal of American Statistical Association*, 81, pp. 310-320.
- [24] Findley D. F. (1986), On bootstrap estimates of forecast mean square errors for autoregressive processes, *Computer Science and Statistics*, pp. 11-7.
- [25] Franklin J. (1998), Predicting the distribution of shrub species in southern California from climate and terrain-derived variables, *Journal of Vegetation Science*, 9, pp. 733-748.

- [26] Freedman D. A., Peters S. C. (1984), Bootstrapping an econometric model: some empirical results, *Journal of Business and Economic Statistics*, 2, pp. 150-158.
- [27] Friedman J., Stuetzle W. (1981), Projection pursuit regression, *Journal of American Statistical Association*, 76, pp. 817-823.
- [28] Gneiting T., Genton M. G., Guttorp P. (2005), Geostatistical Space-Time Models, Stationarity, Separability and Full Symmetry, *Technical Report Department of Statistics University of Washington*, 475.
- [29] Green P., Jennison C., Seheult A. (1985), Analysis of Field Experiments by Least Squares Smoothing, *Journal of the Royal Statistical Society*, 47, pp. 299-315.
- [30] Green P. J., Yandell B. S. (1985), Semi-parametric generalized linear models, *Proceedings of the International Conference on Generalized Linear Models*, 32, pp. 44-55.
- [31] Grigoletto M. (1998), Bootstrap prediction intervals for autoregressions: some alternatives, *International Journal of Forecasting*, 14, pp. 447-456.
- [32] Guisan A., Lehmann A., Ferrier S., Austin M., McC.Overton J., Aspinall R. e Hastie T. J. (2006), Making better biogeographical predictions of species' distributions, *Journal of Applied Ecology*, 43, pp. 386-392.
- [33] Haslett J., Raftery A.E. (1989), Space-time modelling with long-memory dependence: assessing ireland's wind power resource, *Journal of the Royal Statistical Society*, 38, pp. 1-50
- [34] Hastie T. J., Tibshirani R. J. (1986), Generalized Additive Models, *Statistical Science*, 1, pp. 297-310.

- [35] Hastie T. J., Tibshirani R. J. (1990), *Generalized Additive Models*, Chapman & Hall, New York.
- [36] Hastie T. J., Tibshirani R. J., Friedman J. (2001), *The Elements of Statistical Learning*, Springer Verlag, Berlin.
- [37] Johnston J., (1993), *Econometrica*, Franco Angeli.
- [38] Kunsch H. R. (1989), The jackknife and the bootstrap for general stationary observations, *Annals of Statistics*, 17, pp. 1217-1241.
- [39] Leathwick J.R., Whitehead D., McLeod M. (1996), Predicting changes in the composition of New Zealand's indigenous forests in response to global warming: a modeling approach, *Environmental Software*, 11, pp. 81-90.
- [40] Lehmann A. (1998), GIS modeling of submerged macrophyte distribution using generalized additive models, *Plant Ecology*, 139, pp. 113-124.
- [41] Lehmann A., Leathwick J.R., Overton J.McC. (2002), Assessing New Zealand fern diversity from spatial predictions of species assemblages, *Biodivers. Conserv.*, 11, pp. 2217-2238.
- [42] Li Y., Maddala G. S. (1996), Bootstrapping time series models (with coment), *Econometric Reviews*, 15, pp. 115-158.
- [43] Mallows C. L. (1973), Some Comments on  $C_p$ , *Technometrics*, 15, pp. 661-675.
- [44] Masarotto G. (1990), Bootstrap prediction intervals for autoregressions, *International Journal of Forecasting*, 6, pp. 229-239.

- [45] Masarotto G., Scarpa B. (1998), Analisi spazio-temporale di inquinanti atmosferici e posizionamento di centraline: Un approccio non parametrico, *Atti Riunione SIS*, 1, pp. 195-206.
- [46] McCullagh e Nelder (1989), *Generalized linear models*, Chapman and Hall, London.
- [47] O'Sullivan F. (1986), A statistical perspective on ill-posed inverse problems (with discussion), *Statistical Science*, 1, pp. 502-527.
- [48] Pascual L., Romo J. e Ruiz E. (2004), bootstrap predictive inference for arima processes, *Journal of time series analysis*, 25.
- [49] Politis D., Romano J. (1992), A circular block resampling procedure for stationary data, *Exploring the limits of bootstrap*, pp. 263-270.
- [50] Politis D., Romano J. (1992), A general resampling scheme for triangular arrays of  $\alpha$ -mixing random variables with application to the problem of spectral density estimation, *Annals of Statistics*, 20, pp. 1985-2007.
- [51] Politis D., Romano J. (1994), The stationary bootstrap, *Journal of American Statistical Association*, 89, 12, pp. 1303-1313.
- [52] Scarpa B. (1998), Modelli additivi generalizzati con autocorrelazione tra le osservazioni, *Atti Riunione SIS*, pp. 855-864.
- [53] Segurado P., Araújo M., Kunin W.E. (2006), Consequences of spatial autocorrelation on niche-based models, *Journal of Applied Ecology*, 43, pp. 433-444.
- [54] Schabenberger O., Gotway C. A. (2005), *Statistical Methods for Spatial Data Analysis* CRC Press.

- [55] Smyth G. K. (1994), A note on modelling cross correlations: hyperbolic secant regression, *Biometrika*, 81, pp. 396-402.
- [56] Stine R. A. (1987), Estimating properties of autoregressive forecasts, *Journal of The American Statistical Association*, 82, pp. 1072-1078.
- [57] Stone C. J. (1977), Consistent nonparametric regression, *Annals of Statistics*, 5, pp. 595-645.
- [58] Streitberg B. (1979), Multivariate Models of Dependent Spatial Data, *Statistical Analysis of Spatial Data*, pp. 139-177.
- [59] Wei W. W. (2006), *Time series analysis: univariate and multivariate methods*, Addison-Wesley, Reading, MA.
- [60] Wu C. F. J. (1986), Jackknife, bootstrap and other resampling methods in regression analysis, *Annals of Statistics*, 14(4), pp. 1261-1295.
- [61] Yee T. W., Mitchell N. D. (1991), Generalized Additive Models in plant ecology, *Journal of Vegetation Science*, 2, pp. 587-602.