

# Capitolo 1

## Simulazione del codice

Prima di applicare il modello allo studio della produzione di rifiuti nella provincia di Venezia sono state eseguite simulazioni su un caso noto e più semplice. Si è scelto di analizzare il dominio a forma di C e la corrispondente funzione spaziale  $g(\underline{p})$  descritti in CITAZIONE NECESSARIA e nel pacchetto R *mgcv* (in figura 1.1 la funzione  $g(\underline{p})$ ), e si è introdotta una variazione temporale deformando con il coseno:

$$f(\underline{p}, t) = g(\underline{p})\cos(t)$$

Su questo semplice caso sono stati eseguiti i primi tentativi per il modello STSR sia per il caso con covariate che senza covariate.

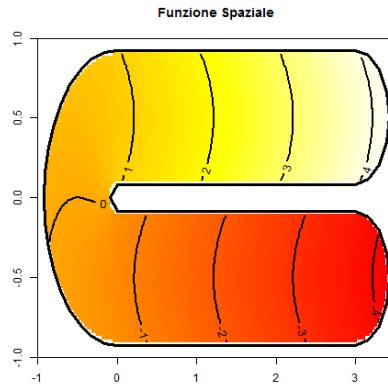


Figura 1.1: Funzione spaziale  $g(\underline{p})$

### 1.1 Triangolazione e istanti temporali

Nel dominio a forma di C non sono presenti punti spaziali definiti dalla natura del problema (come possono essere i comuni per la provincia di Venezia),

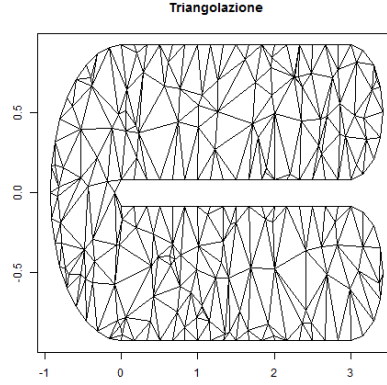


Figura 1.2: Triangolazione del dominio a forma di C

quindi è stato necessario ricavarli. Sono stati generati casualmente 150 punti all'interno del rettangolo  $(-1, +3.5) \times (-1, +1)$  e di questi sono stati considerati validi solo quelli che ricadevano all'interno del dominio. Non è stata usata la descrizione della frontiera presente in *mgcv*, ma una versione diversa che permette di avere punti anche nella parte rettilinea del bordo. In figura 1.2 è riportata la triangolazione ottenuta grazie al pacchetto R *RTriangle*. Come basi in spazio sono stati usati gli elementi finiti lineari definiti su questa triangolazione. In tutti gli esempi che seguiranno sarà considerata questa descrizione del dominio, che è formata da 241 punti, pari anche al numero di basi spaziali  $N$ . Di questi 108 sono di frontiera, e i restanti 133 corrispondono agli  $n$  punti sui quali saranno disponibili i dati.

Come intervallo temporale di variazione dei dati è stato scelto  $[0, 2\pi]$ , per sfruttare la periodicità del coseno. All'interno di questo intervallo sono stati ricavati 9 istanti temporali equidistanti tra di loro, quindi uno ogni  $\frac{\pi}{4}$ . Si è scelto di fissare come basi in tempo le *B-splines* cubiche. Il numero di basi  $M$  è stato scelto pari al numero di istanti temporali a disposizione  $m$ , quindi 9.

## 1.2 Caso senza covariate

Nei punti e negli istanti temporali disponibili i dati sono stati ricavati dalla funzione esatta sporcata dal rumore:

$$z_{ij} = g(\underline{p}_i) \cos(t_j) + \varepsilon_{ij}$$

con

$$\varepsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, 0.5^2) \quad \forall i \in 1 \dots n, \forall j \in 1 \dots m$$

Come primo passo è necessario scegliere i valori per  $\lambda$  ottimizzando l'indice  $\text{GCV}(\underline{\lambda})$  come riportato in RIMANDO NECESSARIO. In queste analisi

$\lambda_S$  e  $\lambda_T$  sono sempre espressi in potenze di 10. Per trovare buoni valori per i parametri si procede creando due insiemi discreti di variazione per  $\log_{10} \lambda_S$  e  $\log_{10} \lambda_T$  e minimizzando sui  $\underline{\lambda}$  corrispondenti al prodotto cartesiano tra di essi. Il procedimento viene iterato qualche volta (a causa dell'alto costo computazionale non è opportuno eseguire troppe iterazioni) rendendo la griglia sempre più fitta. In particolare, nel primo caso i valori sono distanziati di 1, poi (una volta che è possibile centrare gli intervalli in base al risultato precedente) di 0.25 e 0.125.

Intervalli per $\log_{10} \lambda_S$ e $\log_{10} \lambda_T$	Miglior valore
$\log_{10} \lambda_S \in \{-5, -4, \dots, +1\}$	$\underline{\lambda} = (10^0, 10^{-3})$
$\log_{10} \lambda_T \in \{-5, -4, \dots, +1\}$	
$\log_{10} \lambda_S \in \{-1, -0.75, \dots, +1\}$	$\underline{\lambda} = (10^{-0.5}, 10^{-3.25})$
$\log_{10} \lambda_T \in \{-4, -3.75, \dots, -2\}$	
$\log_{10} \lambda_S \in \{-1, -0.875, \dots, +0\}$	$\underline{\lambda} = (10^{-0.375}, 10^{-3.25})$
$\log_{10} \lambda_T \in \{-3.75, -3.625, \dots, -2.75\}$	

Tabella 1.1: Analisi di  $\text{GCV}(\underline{\lambda})$ 

L'analisi è stata eseguita con  $\underline{\lambda} = (10^{-0.375}, 10^{-3.25})$  e la stima della funzione si è rivelata molto buona.

In figura 1.4 sono riportati i confronti tra funzione reale e stimata nei primi istanti di tempo (la scala di colori è stata resa uniforme tra tutti i grafici). Si può notare come la funzione stimata sia effettivamente molto simile a quella reale.

Inoltre in figura 1.5 si ha il confronto dell'evoluzione temporale nel punto  $(3, -0.5)$ . La curva reale è nota, ed è una cosinusoide moltiplicata per  $g(3, -0.5) = -3.785398$  (funzione blu nel grafico, contiene anche i dati raccolti nei vari istanti di tempo). La curva stimata è molto vicina e non è una perfetta interpolazione dei dati. Questo è un ottimo risultato, poichè si è riusciti a cogliere la variazione temporale del fenomeno senza sovrastimare ed interpolare i dati (cosa che sarebbe successa con valori di  $\lambda_T$  più bassi).

### 1.3 Caso con covariate

Nel problema della stima della funzione  $f(\underline{p}, t) = g(\underline{p})\cos(t)$  non sono presenti covariate, quindi è stato necessario generarle aggiungendo al caso precedente. Ad ogni dato è aggiunta la covariata e il rumore come segue:

$$z_{ij} = g(\underline{p}_i)\cos(t_j) + \beta w_{ij} + \varepsilon_{ij}$$

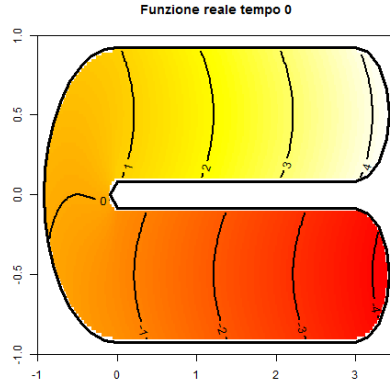
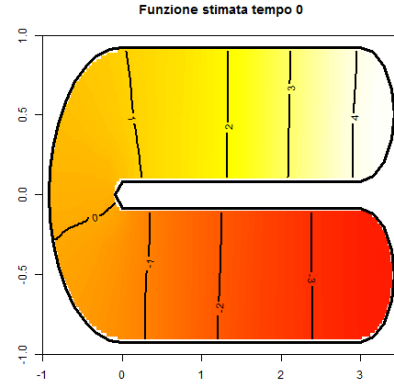
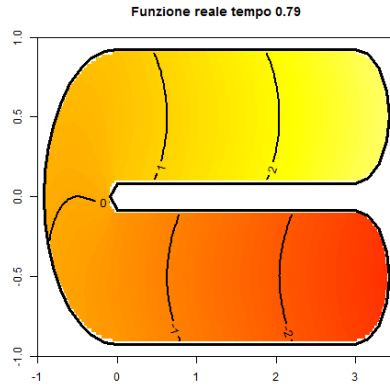
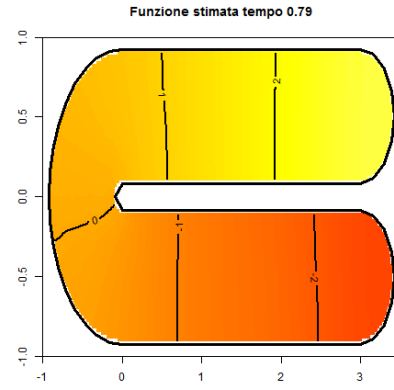
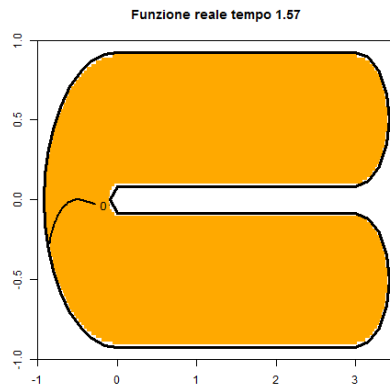
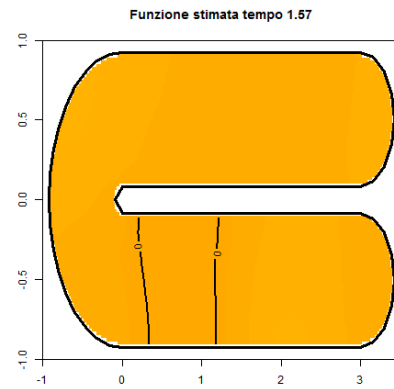
(a) Funzione reale a  $t = 0$ (b) Funzione stimata a  $t = 0$ (c) Funzione reale a  $t = \frac{\pi}{4}$ (d) Funzione stimata a  $t = \frac{\pi}{4}$ (e) Funzione reale a  $t = \frac{\pi}{2}$ (f) Funzione stimata a  $t = \frac{\pi}{2}$ 

Figura 1.3: Stime della funzione  $f(\underline{p}, t)$  ad alcuni istanti di tempo, caso senza covariate

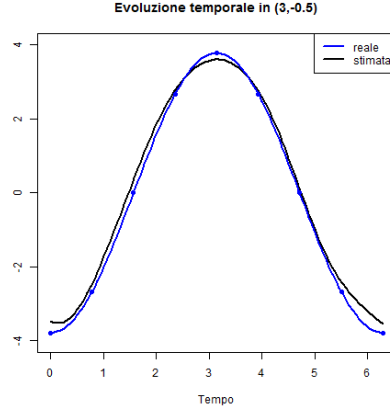


Figura 1.4: Evoluzione temporale nel punto fissato  $\underline{p} = (3, -0.5)$

dove covariate e rumore sono generate da due normali tra loro indipendenti:

$$w_{ij} \stackrel{\text{iid}}{\sim} N(0, 1) \quad \forall i \in 1 \dots n, \forall j \in 1 \dots m$$

$$\varepsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, 0.5^2) \quad \forall i \in 1 \dots n, \forall j \in 1 \dots m$$

e  $\beta$  è fissato a 1. Se il modello è buono, c'è da aspettarsi che la parte di funzione stimata senza covariate sia vicina a  $f(\underline{p}, t)$  e che  $\hat{\beta}$  si avvicini a 1.

Anche in questo caso è necessaria una analisi preliminare per fissare i valori per  $\lambda$  ottimizzando l'indice  $\text{GCV}(\underline{\lambda})$  in RIMANDO NECESSARIO, che nel caso con covariate si differenzia dal precedente solo per la forma della *smoothing matrix*. In tabella 1.2 sono riportati i risultati ricavati secondo lo stesso approccio del caso senza covariate.

Intervalli per $\log_{10} \lambda_S$ e $\log_{10} \lambda_T$	Miglior valore
$\log_{10} \lambda_S \in \{-5, -4, \dots, +1\}$	$\underline{\lambda} = (10^0, 10^{-4})$
$\log_{10} \lambda_T \in \{-5, -4, \dots, +1\}$	
$\log_{10} \lambda_S \in \{-1, -0.75, \dots, +1\}$	$\underline{\lambda} = (10^{0.25}, 10^{-3.75})$
$\log_{10} \lambda_T \in \{-5, -4.75, \dots, -3\}$	
$\log_{10} \lambda_S \in \{-0.25, -0.125, \dots, +0.75\}$	$\underline{\lambda} = (10^{-0.125}, 10^{-3.25})$
$\log_{10} \lambda_T \in \{-4.25, -4.125, \dots, -3.25\}$	

Tabella 1.2: Analisi di  $\text{GCV}(\underline{\lambda})$

I risultati dello studio di  $\text{GCV}(\underline{\lambda})$  si sono rivelati molto simili al caso senza covariate riportato in 1.1, lasciando pensare che la stima della funzione  $f(\underline{p}, t)$  sia molto simile al caso precedente (infatti, a meno di covariate e rumore, coincidono).

Eseguendo l'analisi dei dati con  $\underline{\lambda} = (10^{-0.125}, 10^{-3.25})$  questa ipotesi è confermata, e si trova una buona stima della funzione. In figura 1.6 si hanno i grafici dei primi istanti di tempo (si ricorda che la funzione tracciata non contiene la parte spiegata dalle covariate, ma solo la stima di  $f(\underline{p}, t)$ ).

Analogo al caso precedente è l'analisi dell'evoluzione temporale della funzione in un punto fissato. La curva stimata spiega l'andamento della reale, senza ricadere troppo nell'interpolazione.

Dai grafici precedenti si può trarre come conclusione che la stima della parte funzionale della risposta sia effettivamente una buona approssimazione della reale. Tuttavia occorre verificare anche che il contributo delle covariate sia ben riconosciuto dal modello, e per questo basta controllare il valore stimato di  $\beta$ . Si ha:

$$\hat{\beta} \approx 1.005 ,$$

valore vicinissimo al reale.

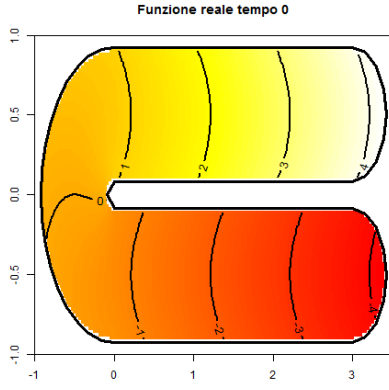
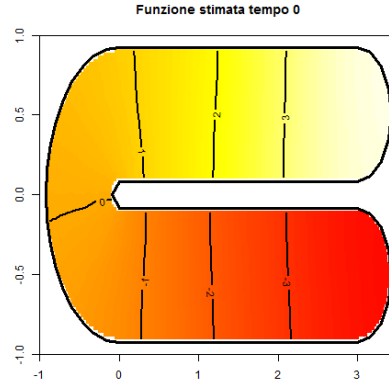
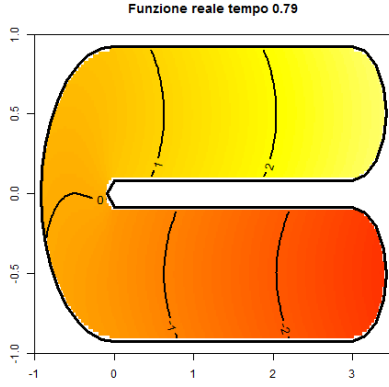
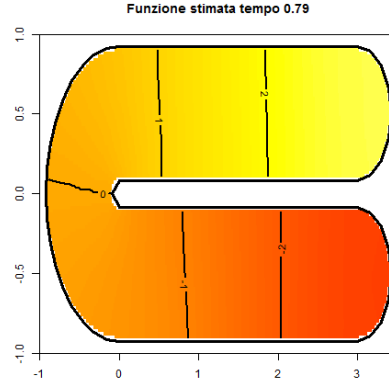
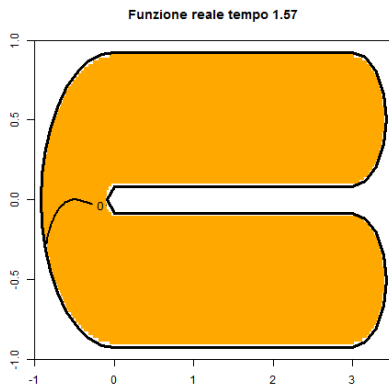
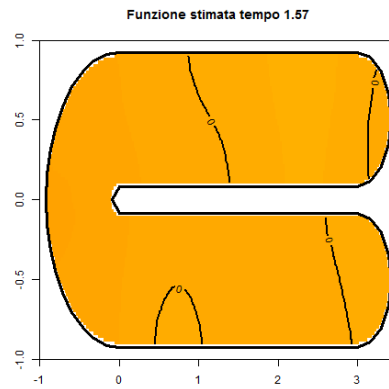
(a) Funzione reale a  $t = 0$ (b) Funzione stimata a  $t = 0$ (c) Funzione reale a  $t = \frac{\pi}{4}$ (d) Funzione stimata a  $t = \frac{\pi}{4}$ (e) Funzione reale a  $t = \frac{\pi}{2}$ (f) Funzione stimata a  $t = \frac{\pi}{2}$ 

Figura 1.5: Stime della funzione  $f(\underline{p}, t)$  ad alcuni istanti di tempo, caso con covariate

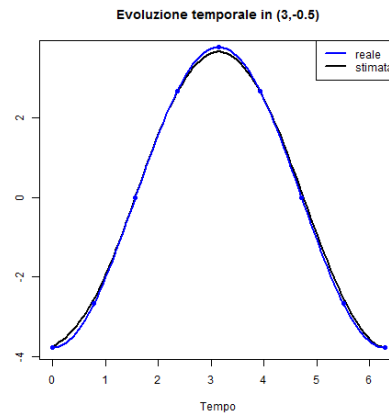


Figura 1.6: Evoluzione temporale nel punto fissato  $\underline{p} = (3, -0.5)$