

Capitolo 1

Venezia

Come è già stato brevemente accennato in precedenza, l'applicazione scelta per il modello STR-PDE riguarda i dati della produzione di rifiuti urbani nel periodo di anni dal 1997 al 2011 nella provincia di Venezia. Per rifiuti urbani si intendono rifiuti domestici, prodotti in locali, aree pubbliche, parchi o giardini, spiagge o provenienti dalla pulizia delle strade o di altri luoghi pubblici. Non sono conteggiati i rifiuti speciali (tra cui ad esempio gli industriali, agricoli o provenienti da attività commerciali o di costruzione) o pericolosi (per i quali esistono programmi di smaltimento particolari).

In realtà i dati raccolti dall'Agenzia Regionale per la Prevenzione e Protezione Ambientale del Veneto (Arpav) riguardano tutto il Veneto. Tuttavia è stata analizzata solo la provincia di Venezia per due motivi. Inanzitutto l'interesse particolare per la zona della laguna veneta, in cui è rilevante il ruolo del turismo (come si potrà notare in seguito). Inoltre considerare tutto il Veneto aumenta notevolmente le dimensioni delle matrici in gioco e causa una grossa spesa computazionale per la ricerca della soluzione. Quindi è stato scelto di concentrarsi su un dominio più piccolo ma nel quale è possibile notare più facilmente le particolarità dell'andamento del fenomeno della produzione dei rifiuti urbani.

Per ogni comune della provincia di Venezia e per ogni anno è disponibile il numero di rifiuti totali raccolti in tonnellate e la popolazione residente. La popolazione è certamente un valore influente per la produzione di rifiuti, perciò la quantità di riferimento non sarà il valore dei rifiuti totali raccolti in ogni anno per comune, ma il valore pro capite.

Le coordinate spaziali dei comuni sono la longitudine e la latitudine, disponibili on line¹. Nel caso dei comuni con dato replicato (sez 1.2.2), sono state scaricate da Google Maps.

¹<http://www.dossier.net/utilities/coordinate-geografiche/>

1.1 L'inclusione dell'effetto del turismo

L'inclusione della popolazione residente nella risposta tramite la scelta di usare i valori pro capite è necessaria, poichè permette di depurare la risposta da una variabile che per sua natura la influenzerebbe. Ma sarebbe un errore fermarsi solo alla popolazione residente, poichè anche i turisti rappresentano una componente non trascurabile di produzione di rifiuti urbani.

Nella provincia di Venezia sono presenti molte zone di elevata attrazione turistica. La più rilevante di queste è Venezia, ma si hanno anche zone balneari (come Lido di Venezia, Cavallino-Treporti, Jesolo, San Michele al Tagliamento, Bibione, ecc...). L'informazione scelta per sintetizzare l'attività turistica è il numero di posti letto totali del comune, valore disponibile grazie all'applicativo dell'Istat *Atlante Statistico dei Comuni*² per ogni anno. Il totale dei posti letto per comune è la somma di vari tipi di attività non solamente alberghiere (ad esempio sono conteggiati anche esercizi complementari, bed & breakfast, campeggi) e saranno considerati normalizzati per la popolazione residente per uniformità con la risposta. I valori ricavati saranno inseriti nel modello come possibile covariata.

1.2 Trattamento del dominio

Per poter studiare il problema a livello computazionale occorre avere una buona approssimazione della frontiera della regione. Questa è disponibile nel pacchetto R *raster* che descrive dati geografici di moltissime zone del mondo sia a livello nazionale che locale (nel caso italiano province e comuni) tramite poligoni molto precisi.

Una volta scaricata la provincia di Venezia si è riscontrato subito un problema: la regione è composta da un insieme di 101 poligoni distinti (a causa delle numerose isole di cui è composta la laguna di Venezia) e ogni poligono ha un alto numero di vertici (ad esempio, la prima delle due regioni corrispondenti all'entroterra aveva 10538 vertici). Non è possibile analizzare il problema su un territorio così descritto, perciò è stata necessaria una analisi iniziale della frontiera per ridurne la complessità.

Oltre all'entroterra (composto da due poligoni) stati scelti solo le più importanti isole della laguna veneta: Venezia, Murano, Lido di Venezia e Pellestrina (più rilevanti a livello di popolazione e turismo). I poligoni sono stati semplificati in tutti i casi e uniti tra loro con ponti dove era possibile. Tra le isole collegate solamente via mare con il resto del territorio sono state simulati ponti in corrispondenza delle trafficate linee di trasporto pubblico con traghetto.

²<http://www.istat.it/it/archivio/113712>

1.2.1 Regression splines

Per ridurre l'elevato numero di vertici di ognuno dei poligoni considerati si è scelto di ricorrere ad un'analisi di smoothing di dati funzionali. Ad ogni poligono è associata una coppia di funzioni: la latitudine e la longitudine rispetto all'ascissa curvilinea (disponibili per punti, corrispondenti ai vertici) che sono state rappresentate in basi e valutate in un numero molto inferiore di punti, con i quali si è costruita la nuova definizione della regione.

Per avere una rappresentazione tramite funzioni di base di queste funzioni sono state provate più tecniche, ma la scelta definitiva è ricaduta sulle *Regression Splines* cubiche senza penalizzazione della derivata seconda. Infatti i risultati non sono stati migliori negli altri casi a causa della zona interna alla laguna di Venezia, fortemente frastagliata: penalizzare la derivata seconda eliminava troppe asperità presenti sulle coste del territorio, mentre con *Kernel Smoothing* sono state ricavate regioni che, dopo la triangolazione, presentavano troppi triangoli composti solamente da punti di frontiera (e quindi senza dati) rispetto agli altri metodi.

Una volta fissato un ragionevole numero di punti con cui descrivere la regione sono stati eseguiti più tentativi per decidere il miglior numero di basi necessario per descrivere le due funzioni. Il criterio di scelta è stato complesso, poichè sono stati esclusi i valori che generavano intersezioni nella nuova descrizione della regione e comuni esterni alla frontiera. Ma la scelta del miglior numero di basi per *Regression Splines* è ricaduta sul valore che, una volta eseguito lo smoothing della regione, causava la minor distanza tra i nuovi punti della regione e il poligono iniziale. In figura 1.1 è riportato il risultato dello smoothing sul primo poligono che descrive l'entroterra della provincia di Venezia (descritta con 100 punti, molto meno dei 10538 iniziali).

Dopo aver ripetuto l'analisi per ognuna delle isole scelte inizialmente la descrizione finale è stata ricavata unendo tra loro tutti i nuovi poligoni. In seguito è stata eliminata una zona costiera dell'entroterra della laguna di Venezia che, sebbene sia presente sia in *raster* che nei grafici di Google Maps, corrisponde ad una parte fangosa e paludosa e quindi disabitata. Non essendo possibile che su di essa siano prodotti rifiuti è stata tagliata dalla regione. Per questo motivo si troverà sempre una zona non analizzata sui grafici con mappe da Google Maps. In figura 1.2 è riportata la descrizione finale del dominio con i punti spaziali considerati (si consulti anche la sezione 1.2.2).

1.2.2 Scelte particolari tra i comuni

L'uso di valori pro capite per rifiuti e posti letto consente di replicare del comune anche su altri punti in cui risulta necessario. Ad esempio, le isole di Murano, Lido di Venezia e Pellestrina non sono sedi di comune, ma si riferiscono a Venezia. Quindi il dato di Venezia è stato replicato in queste

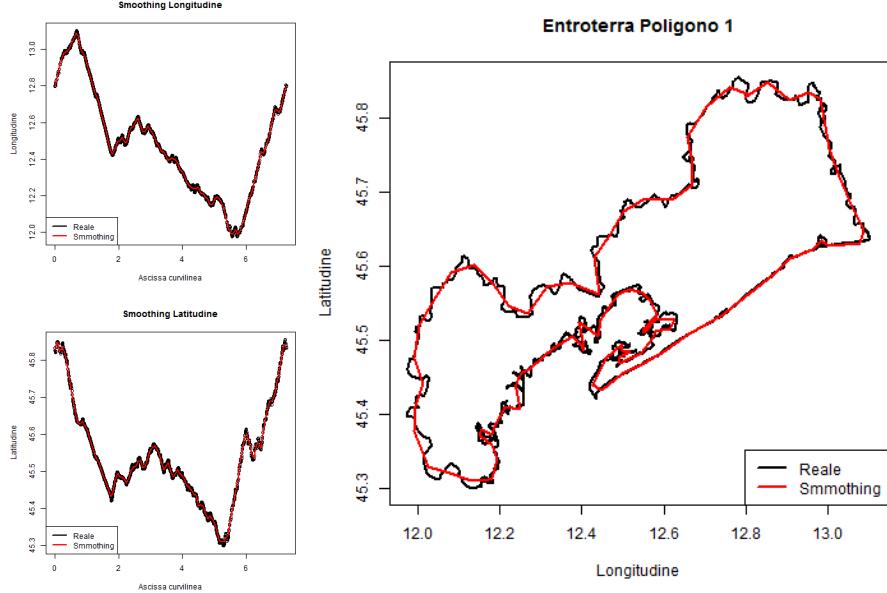


Figura 1.1: Smoothing con *Regression Splines* cubiche per il primo poligono dell'entroterra della provincia di Venezia

isole ad ogni anno, per avere un valore di riferimento in quanto zone distaccate. Come si può notare in figura 1.2(b) anche nell'isola di Venezia è stato duplicato il dato, per avere una triangolazione senza troppi triangoli composti solo da punti di frontiera in una zona della provincia di Venezia di particolare rilevanza.

Un caso particolare riguarda il comune di Cavallino-Treporti, che è stato istituito nel 1999 da una parte dei territori del comune di Venezia. La separazione all'interno dei dati, però, è presente dal 2002. Di conseguenza prima di questo anno il dato in Cavallino-Treporti è una replica del dato di Venezia.

1.2.3 Triangolazione del dominio

La triangolazione è stata prodotta tramite il pacchetto R *RTriangle*. Poichè nella zona ad est il numero di capoluoghi di comune (e quindi di nodi della triangolazione) è minore rispetto al resto della regione, è stata fissata un'area massima per i triangoli generati da *RTriangle*. Questo ha reso la triangolazione più fitta anche dove non lo sarebbe stata, e garantirà una stima della risposta più precisa nella zona balneare ad est, che come si potrà notare in seguito ha una grande importanza per la distribuzione dei rifiuti. Affinchè questo sia possibile sono stati aggiunti nuovi punti spaziali, che restano senza dato per tutta l'analisi. In fig. 1.3 si ha la triangolazione finale che sarà usata da ora in avanti.



Figura 1.2: Frontiera e punti spaziali per la provincia di Venezia

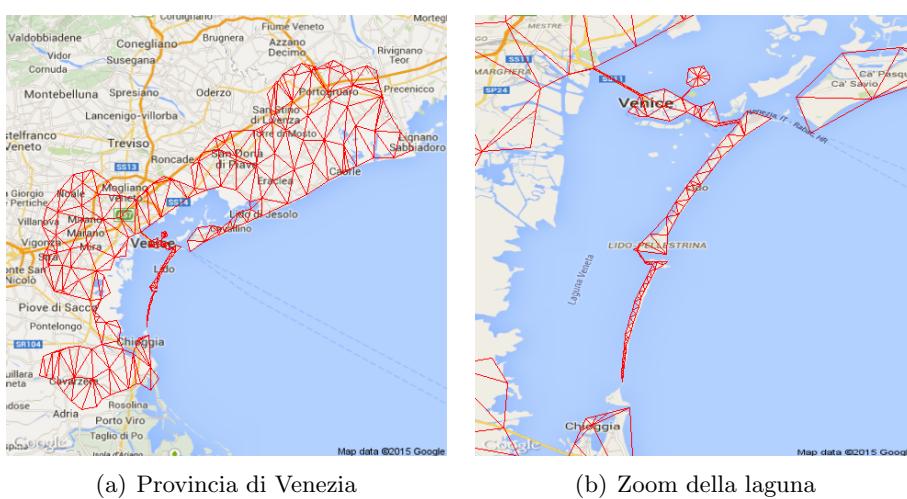


Figura 1.3: Triangolazione della provincia di Venezia

In conclusione, il dominio è descritto da 414 punti (41 con dati, 373 di frontiera o aggiunti dalla triangolazione) e da 475 triangoli.

1.3 Analisi preliminare dei dati

Prima di eseguire l'analisi, è opportuno avere una visualizzazione dei dati sul dominio della provincia di Venezia. In fig. 1.4 sono riportati i bubbleplot dei dati, che permettono di avere un'idea del fenomeno in spazio e tempo.

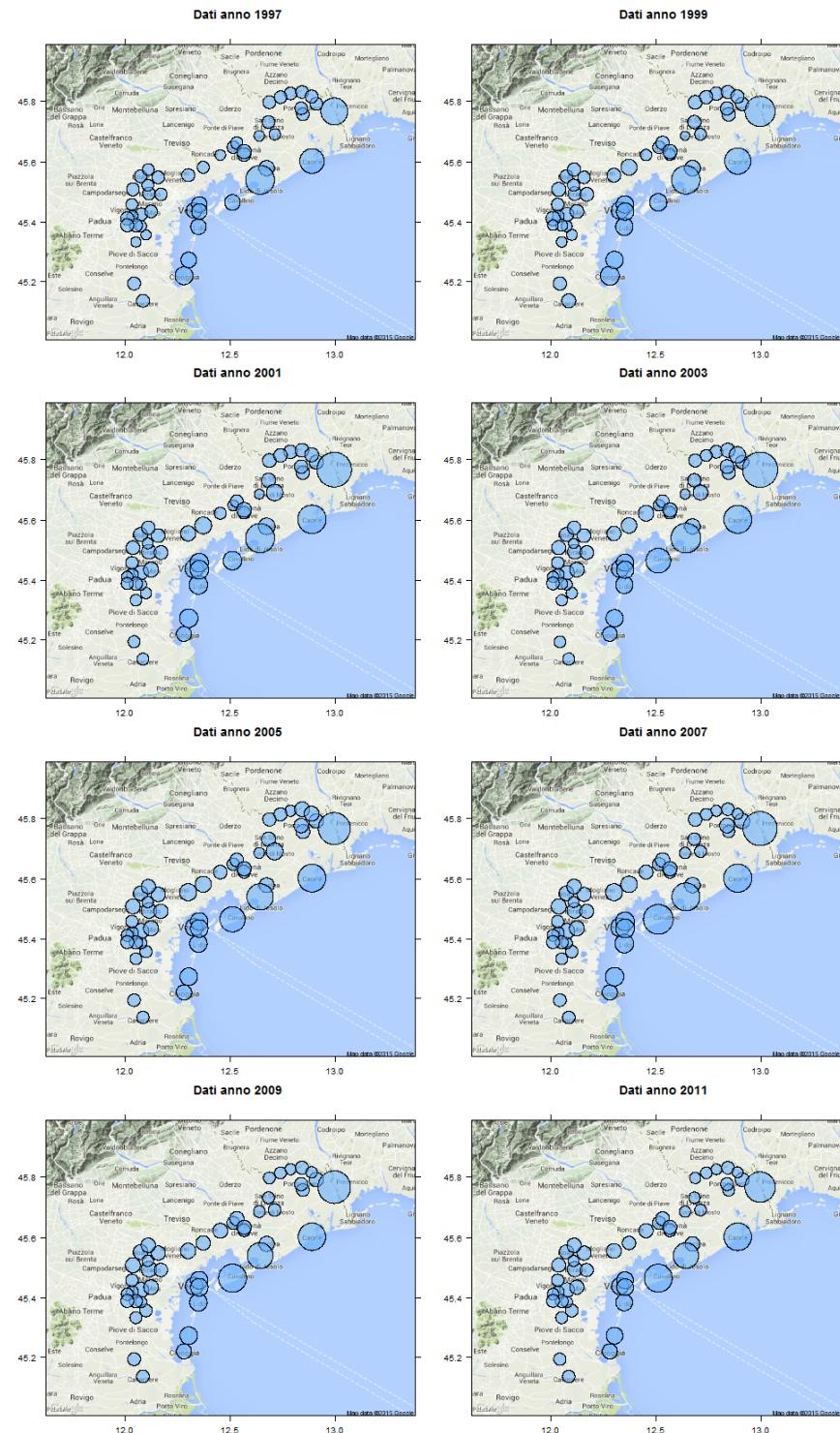


Figura 1.4: Bubbleplot delle misurazioni della produzione di rifiuti urbani pro capite ogni due anni dal 1997 al 2011.

Da questa analisi iniziale ci si aspetta che il risultato mostri una produzione di rifiuti più lata nella zona balneare della regione, e che non abbia grosse variazioni nel tempo, poichè i dati sono più o meno simili negli anni.

Gli stessi grafici sono stati ripetuti in fig. 1.5 per il numero di posti letto pro capite. In questo caso però, si può evidenziare un andamento diverso.

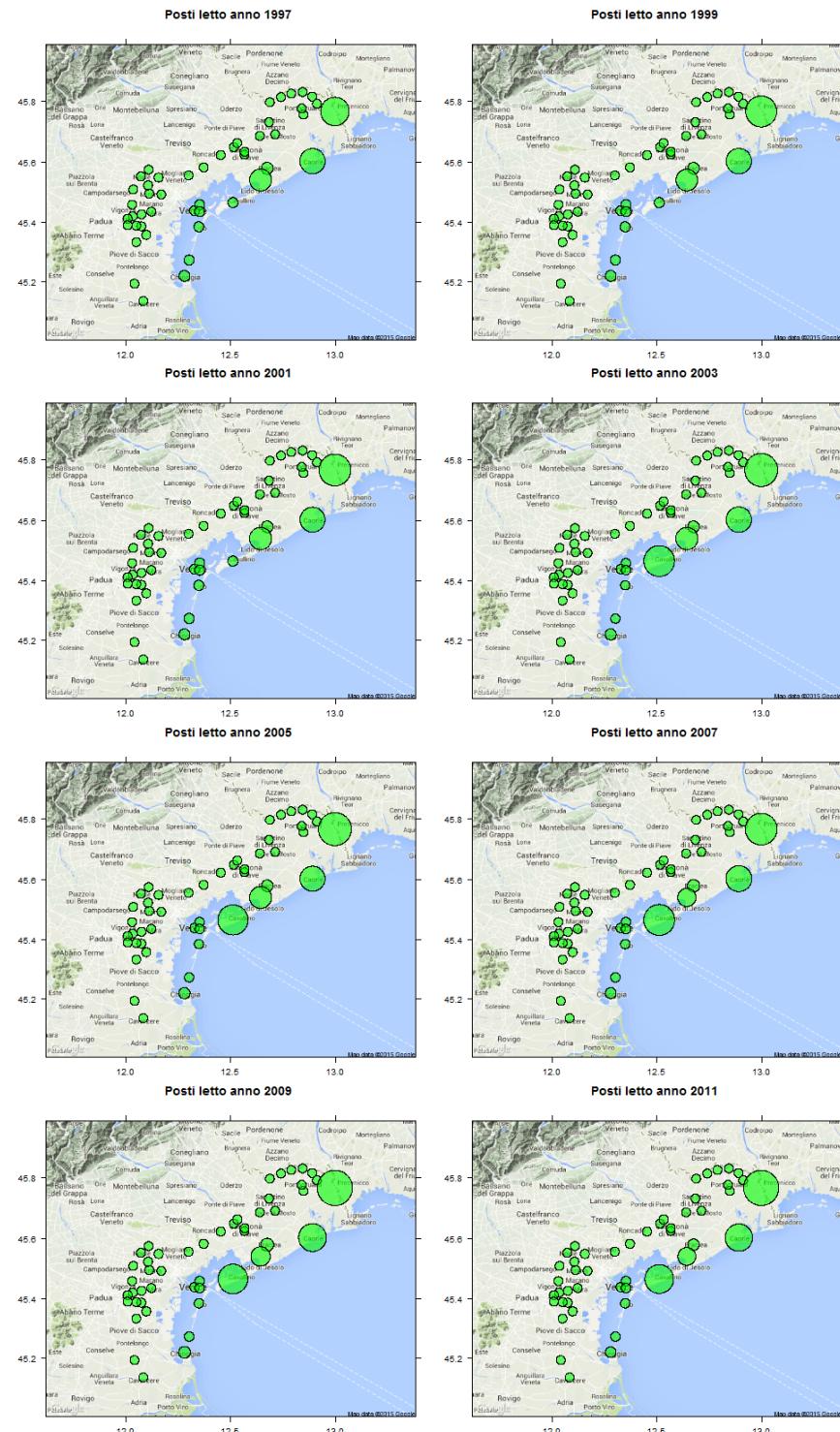


Figura 1.5: Bubbleplot delle misurazioni della produzione di rifiuti urbani pro capite ogni due anni dal 1997 al 2011.

Si può notare che il numero di posti letto assume valori maggiori nella zona balneare rispetto al resto della regione. Contrariamente a quello che ci si aspettava, non si hanno molto posti letto nell'isola di Venezia. Invece si ha una grossa variazione nel comune di Cavallino-Treporti dopo la separazione dal comune di Venezia (si nota che il valore aumenta dopo il 2002). Questo indica che i posti letto sono in realtà numerosi in Cavallino-Treporti, ma quando era unito al comune di Venezia, questo effetto non poteva essere notato in modo così dettagliato. Ci si può quindi aspettare che la funzione stimata presenti un comportamento differente in Cavallino-Treporti nelle due situazioni.

1.4 Applicazione del modello senza covariate

1.4.1 Analisi preliminare e ricerca del miglior λ

Le basi in spazio scelte per l'applicazione del modello in questo caso sono gli elementi finiti lineari. Ad ognuno dei punti spaziali (interni o di frontiera) è associata una funzione di base, coerentemente con la triangolazione prodotta. Di conseguenza, si ha $N = 414$ mentre il numero di punti con dati n è minore ed è pari a 41.

In tempo, esattamente come nel caso del dominio a forma di C, sono state scelte come funzioni di base le B-splines cubiche. L'intervallo temporale per la descrizione del dominio è [1997, 2011], e i dati sono disponibili con cadenza annuale. Anche in questo caso assumiamo che il numero di basi sia pari al numero di istanti temporali a disposizione, quindi $M = m = 15$.

Prima di calcolare i risultati dell'analisi occorre fissare i parametri λ_S e λ_T . Il procedimento è perfettamente analogo a quello ricavato nel caso del dominio a forma di C, minimizzando la quantità $GCV(\lambda)$ in (CITAZIONE NECESSARIA). In tabella sono disponibili i risultati.

Intervalli per $\log_{10} \lambda_S$ e $\log_{10} \lambda_T$	Miglior valore
$\log_{10} \lambda_S \in \{-12, -11, \dots, +1\}$	$\lambda = (10^{-10}, 10^0)$
$\log_{10} \lambda_T \in \{-8, -7, \dots, +1\}$	
$\log_{10} \lambda_S \in \{-11, -10.5, \dots, -9\}$	$\lambda = (10^{-9.5}, 10^0)$
$\log_{10} \lambda_T \in \{-1, -0.5, \dots, +1\}$	
$\log_{10} \lambda_S \in \{-10, -9.875, \dots, -9\}$	$\lambda = (10^{-9.625}, 10^0)$
$\log_{10} \lambda_T \in \{-0.5, -0.375, \dots, +0.5\}$	

Tabella 1.1: Analisi di $GCV(\lambda)$

1.4.2 Risultati

La produzione dei rifiuti è stata analizzata con $\lambda = (10^{-9.625}, 10^0)$. In fig. 1.6 sono riportati i risultati ottenuti nei 15 anni a disposizione. Questi grafici (grazie anche alla visualizzazione sulle mappe di Google Maps) permettono di studiare il profilo della funzione nei vari istanti temporali, e avere un'idea dell'evoluzione della produzione di rifiuti a livello geografico.

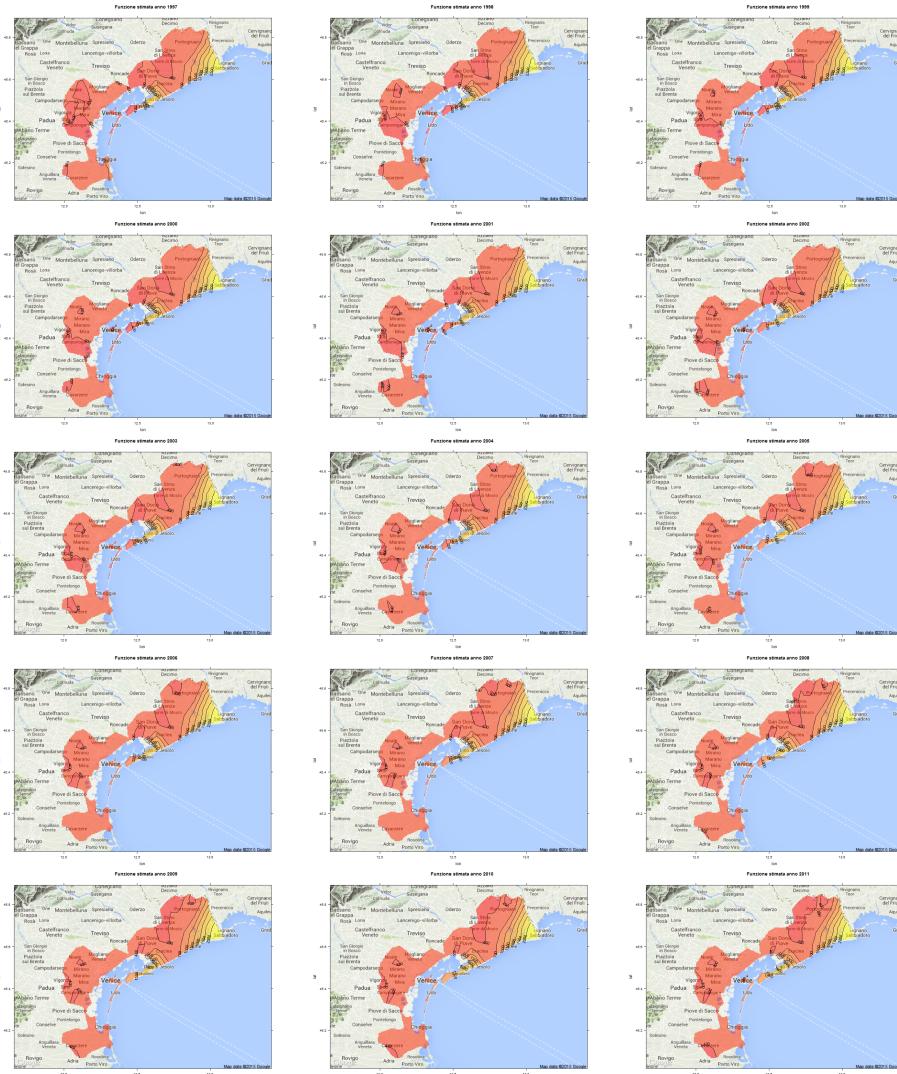


Figura 1.6: Stima della funzione spazio-temporale della produzione dei rifiuti nella provincia di Venezia a tempi fissati: dal 1997 al 2011.

Come si può notare dai grafici, ci sono due zone ad elevato valore di produzione dei rifiuti pro capite, corrispondenti a zone di elevato interesse turistico per le spiagge presenti. Nella parte ad est della provincia di Venezia, tra le altre, si hanno le località turistiche di Bibione (confinante con Lingano Sabbiadoro, che però è oltre il confine veneto) e Caorle. Scendendo verso sud-ovest si nota che anche Jesolo causa un innalzamento della funzione che descrive la risposta.

La produzione dei rifiuti è particolarmente alta in queste zone turistiche, e contrariamente a ciò che si poteva immaginare, molto di più di Venezia. Quindi già adesso possiamo immaginare che il turismo sarà rilevante nell'analisi della produzione di rifiuti, e la causa non sarà Venezia (dove la produzione non è eccessivamente diversa dagli altri comuni) ma le località balneari.

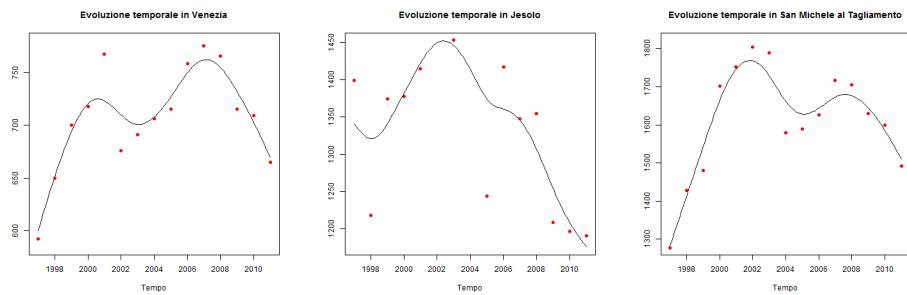


Figura 1.7: Stima della funzione spazio-temporale della produzione dei rifiuti nella provincia di Venezia in alcuni punti spaziali fissati, corrispondenti a Venezia, Jesolo e San Michele al Tagliamento.

In fig. ?? Sono riportati gli sviluppi temporali stimati in alcuni comuni. In particolare sono stati scelti Venezia, Jesolo e San Michele al Tagliamento. I punti rossi corrispondono al dato misurato. Si può notare come la funzione stimata spieghi bene l'andamento tracciato dai dati iniziali senza cadere nell'eccessiva interpolazione.

1.5 Applicazione del modello con covariate

1.5.1 Analisi preliminare e ricerca del miglior λ

Le basi in spazio e in tempo sono esattamente le stesse del caso precedente. Prima di eseguire l'analisi, come al solito, occorre cercare buoni valori per λ . Dalla tabella si possono seguire le iterazioni eseguite e il risultato finale.

Intervalli per $\log_{10} \lambda_S$ e $\log_{10} \lambda_T$	Miglior valore
$\log_{10} \lambda_S \in \{-5, -4, \dots, +1\}$	$\underline{\lambda} = (10^0, 10^{-3})$
$\log_{10} \lambda_T \in \{-5, -4, \dots, +1\}$	
$\log_{10} \lambda_S \in \{-1, -0.75, \dots, +1\}$	$\underline{\lambda} = (10^{-0.5}, 10^{-3.25})$
$\log_{10} \lambda_T \in \{-4, -3.75, \dots, -2\}$	
$\log_{10} \lambda_S \in \{-1, -0.875, \dots, +0\}$	$\underline{\lambda} = (10^{-0.375}, 10^{-3.25})$ 6
$\log_{10} \lambda_T \in \{-3.75, -3.625, \dots, -2.75\}$	

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

1.5.2 Risultati

Le analisi sono state eseguite con $\underline{\lambda} = (10^{-0.375}, 10^{-3.25})$. Come si può notare dai grafici in fig. 1.8, dove è riportata la stima della funzione $f(\underline{p}, t)$ in ogni punti senza l'aggiunta della parte spiegata dalla covariata, nelle zone dove la produzione di rifiuti è massima si hanno ora i valori più bassi.

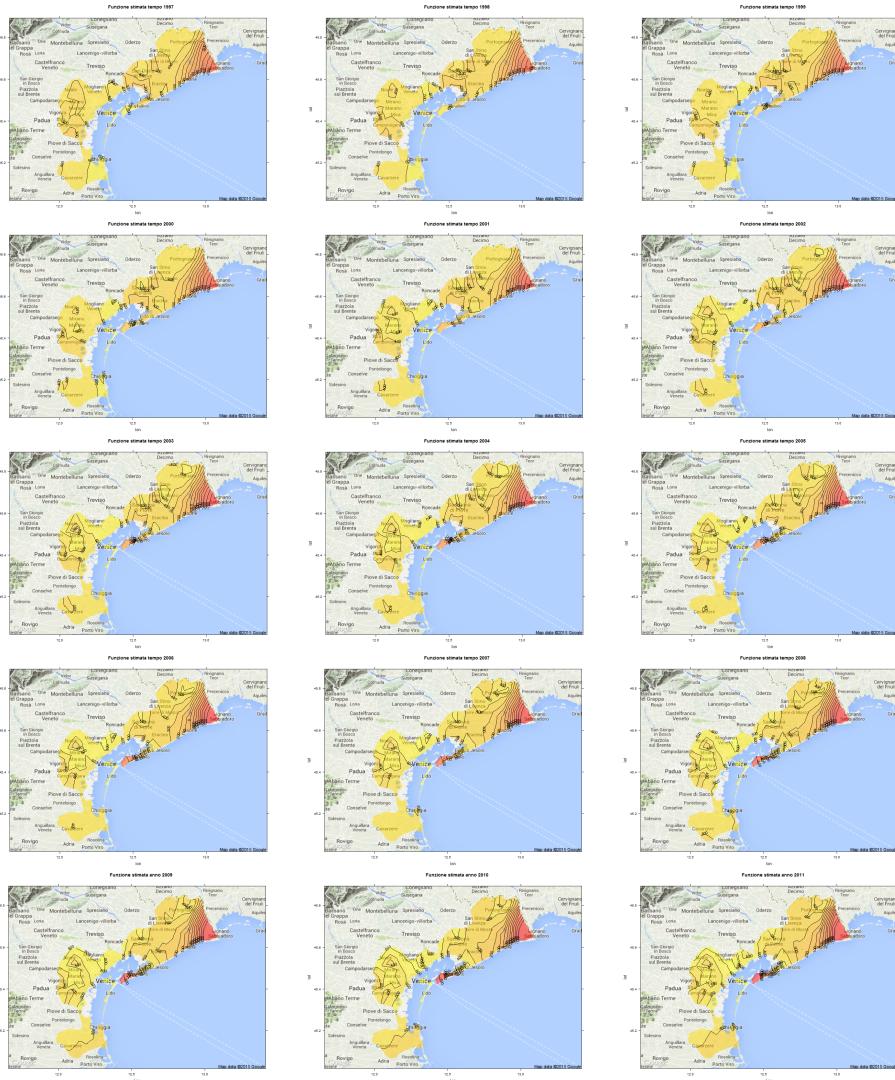


Figura 1.8: Estimated spatio-temporal field of the Venice waste data with the STSR method without covariates at fixed time instants: from year 1997 to year 2011.

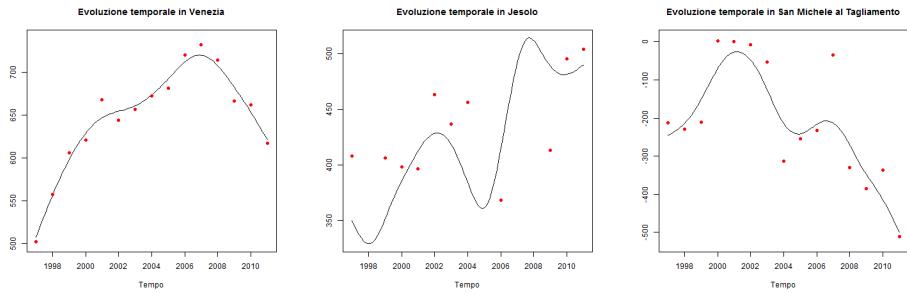


Figura 1.9: Stima della funzione spazio-temporale della produzione dei rifiuti nella provincia di Venezia in alcuni punti spaziali fissati, corrispondenti a Venezia, Jesolo e San Michele al Tagliamento.

In figura 1.9 si hanno anche i grafici della funzione stimata (senza il termine di covariate) in alcuni comuni. I punti tracciati in rosso rappresentano il dato, a cui è sottratto il termine di covariata tramite $\hat{\beta}$ stimato. Non può quindi essere considerato come la parte di dato generato esattamente dalla funzione, ma una sua approssimazione.

L'interpretazione del risultato in questo caso è legata all'alta affluenza turistica nelle zone balneari. Come si può notare dai grafici iniziali dalle fig. ?? e 1.5, su tutta la zona costiera non si hanno solo i massimi per i rifiuti pro capite, ma anche per il numero di posti letto pro capite. Quindi non soltanto la produzione di rifiuti è massima nella zona balneare (come già si notava dalla funzione stimata nel caso senza covariate) ma anche l'attività turistica. In questo caso

$$\hat{\beta} \approx 284.05 .$$

Più le covariate sono alte, più è sottratta informazione alla funzione $f(\underline{p}, t)$, poiché $\hat{\beta}$ è costante in spazio e tempo. Quindi la funzione ha un profilo quasi simile ad una traslazione del caso senza covariate nel resto del dominio (in questa zona il numero di posti letto pro capite, come si nota in fig., è quasi costante) e viene ribassato nella zona balneare, a causa dell'alto contributo della covariata.

Interessante è quanto accade nel comune di Cavallino-Treporti, dove il numero di posti letto pro capite subisce una grossa variazione nel tempo (come si nota dalla fig. 1.5).

Si può notare come la funzione cambi attorno all'anno 2002, in cui si ha la separazione del comune di Cavallino-Treporti da Venezia. Tuttavia, grazie alle proprietà di smoothing garantite dal modello, non si ha una variazione repentina, ma una curva che riesce a rendere perfettamente l'idea sul cambiamento della funzione.

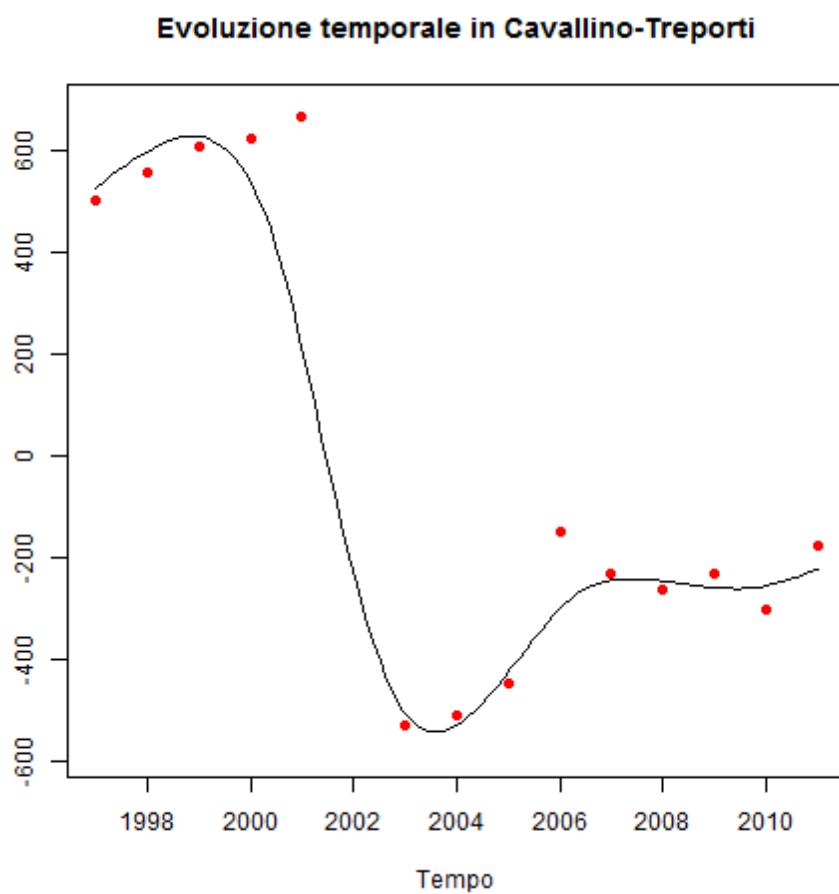


Figura 1.10: Andamento della funzione spazio-temporale nel comune di Cavallino-Treporti