

# Capitolo 1

## Descrizione del modello

In questo capitolo viene descritto nel dettaglio il modello STR-PDE per l'analisi di dati distribuiti in spazio e tempo, ed è calcolata la soluzione al problema di stima..

### 1.1 Caso senza covariate

#### 1.1.1 Dati e modello

Siano  $\{\underline{p}_i = (x_i, y_i); i = 1, \dots, n\}$  un insieme di  $n$  punti spaziali in un dominio limitato  $\Omega \subset \mathbb{R}^2$  e siano  $\{t_j; j = 1, \dots, m\}$  un insieme di  $m$  istanti temporali in un intervallo  $[0, T] \subset \mathbb{R}$ . In questi punti ed istanti osserviamo i dati: siano quindi  $z_{ij}$  i valori della variabile reale nel punto  $\underline{p}_i$  al tempo  $t_j$ .

Sarebbe possibile trattare i dati come provenienti da funzioni tempo-varianti in ognuno dei punti spaziali considerati, cioè che le osservazioni  $\{z_{ij}; i = 1, \dots, n; j = 1, \dots, m\}$  provengano da una funzione  $z_i(t)$  valutata all'istante  $t_j$ . Questa marginalizzazione può analogamente essere eseguita in spazio, supponendo che per ogni istante di tempo esista una funzione variante in spazio definita su  $\Omega$  che genera i dati. Tuttavia in questa analisi non si studiano questi modelli marginali, ma si suppone che le osservazioni  $\{z_{ij}; i = 1, \dots, n; j = 1, \dots, m\}$  siano generate da un campo spazio-temporale con l'aggiunta di rumore:

$$z_{ij} = f(\underline{p}_i, t_j) + \varepsilon_{ij} \quad i = 1, \dots, n \quad j = 1, \dots, m \quad , \quad (1.1)$$

dove  $\{\varepsilon_{ij}; i = 1, \dots, n; j = 1, \dots, m\}$  sono residui indipendenti identicamente distribuiti di media nulla e varianza  $\sigma^2$ . L'obiettivo del modello STR-PDE sarà la stima della funzione  $f(\underline{p}, t)$  dalle  $nm$  osservazioni a disposizione, minimizzando un funzionale  $J(f(\underline{p}, t))$  con la somma degli scarti quadratici tra le osservazioni e valori stimati, e termini di penalizzazione per la regolarità della funzione separati in spazio e tempo.

### 1.1.2 Definizione delle funzioni di base per $f(\underline{p}, t)$

Dall'analisi dei lavori analoghi già disponibili per modelli simili, anche se solo spaziali, si può dedurre che non è possibile dare una stima della funzione se non risulta espressa in una espansione di opportune funzioni di base. Infatti l'infinita possibilità di variazione di una funzione in un qualsiasi spazio funzionale non renderebbe possibile una stima computazionale del risultato. L'approccio scelto per la costruzione di  $f(\underline{p}, t)$  si basa sulla generalizzazione delle espressioni in funzione di base dei casi puramente spaziali o temporali.

Siano  $\{\varphi_k(t); k = 1, \dots, M\}$  un insieme di  $M$  funzioni di base definite sull'intervallo temporale  $[T_1, T_2]$  e  $\{\psi_l(\underline{p}); l = 1, \dots, N\}$  in insieme di  $N$  basi temporali definite sul dominio spaziale  $\Omega$ . Con le combinazioni lineari

$$\sum_{k=1}^M a_k \varphi_k(t) \quad \sum_{l=1}^N b_l \psi_l(\underline{p})$$

è possibile costruire, rispettivamente, funzioni varianti soltanto in tempo e in spazio. La funzione  $f(\underline{p}, t)$  nasce ipotizzando che i coefficienti costanti delle espansioni in base precedenti possano variare secondo la variabile che non è espressa dalle funzioni di base:

$$f(\underline{p}, t) = \sum_{k=1}^M a_k(\underline{p}) \varphi_k(t) \quad (1.2)$$

$$f(\underline{p}, t) = \sum_{l=1}^N b_l(t) \psi_l(\underline{p}) . \quad (1.3)$$

Si ha quindi che  $\{a_k(\underline{p}); k = 1, \dots, M\}$  sono i coefficienti spazio-varianti dell'espansione in basi di tempo e  $\{b_l(t); l = 1, \dots, N\}$  sono i coefficienti tempo-varianti dell'espansione in basi spaziali. In questa costruzione la funzione  $f(\underline{p}, t)$  possiede entrambe queste rappresentazioni.

Sono state ricavate due espressioni equivalenti per  $f$ , ma come si potrà notare in seguito questo è coerente, e sarà ricavata un'espressione più specifica. Per il momento occorre soltanto aggiungere condizioni per i coefficienti appena introdotti, che saranno necessarie per la costruzione del funzionale di penalizzazione  $J(f(\underline{p}, t))$ :

$$a_k(\underline{p}) \in H_{n_0}^2 \quad \forall k = 1, \dots, M$$

e

$$b_l(t) \in H^2 \quad \forall l = 1, \dots, N ,$$

dove  $H_0^2(\Omega) = \{h \in H^2(\Omega) | \partial_\nu h = 0 \text{ su } \partial\Omega\}$ , spazio incluso in  $H_2(\Omega)$  che garantisce alle funzioni  $a_k(\underline{p})$  di avere condizioni di Neumann alla frontiera.

### 1.1.3 Funzionale di penalizzazione $J_{\lambda}(f(\underline{p}, t))$

Per poter stimare  $f(\underline{p}, t)$  si introduce la minimizzazione di un funzionale non formato solamente dagli scarti quadratici tra le osservazioni e le stime negli  $nm$  punti disponibili. Sono inclusi in esso anche altri due termini, che derivano dalla penalizzazione di opportune derivate in spazio e tempo per poter garantire regolarità alla funzione.

Analogamente a quanto fatto il 1.1.2, per costruire tale funzionale si considerano inanzitutto i problemi marginali in spazio e tempo.

Per una funzione spazio-variante sono disponibili più alternative per penalizzare la regolarità, ma in questo caso si considera la radice quadrata la norma  $L^2$  del laplaciano. La stessa scelta è stata fatta in altre pubblicazioni come [5], [6] e [7]. Quindi se  $g(\underline{p}) : \Omega \mapsto \mathbb{R}$  è una funzione spazio-variante, allora si può definire la penalizzazione in spazio come

$$J_S(g(\underline{p})) = \int_{\Omega} \left( \Delta g(\underline{p}) \right)^2 d\underline{p} .$$

Analogamente in tempo si avrà la penalizzazione della radice quadrata della norma  $L^2$  della derivata seconda. Se  $h(t) : [T_1, T_2] \mapsto \mathbb{R}$ , allora si avrà

$$J_T(h(t)) = \int_{T_1}^{T_2} \left( \frac{\partial^2 h(t)}{\partial t^2} \right)^2 dt .$$

Grazie alle ipotesi di regolarità introdotte in sez. 1.1.2 tali penalizzazioni possono essere applicate ai coefficienti degli sviluppi delle eq. 1.2 e 1.3. Per questo motivo si definisce:

$$\begin{aligned} J_{\lambda}(f(\underline{p}, t)) = & \sum_{i=1}^n \sum_{j=1}^m (z_{ij} - f(\underline{p}_i, t_j))^2 + \\ & + \lambda_S \sum_{k=1}^M J_S(a_k(\underline{p})) + \lambda_T \sum_{l=1}^N J_T(b_l(t)) , \end{aligned}$$

cioè

$$\begin{aligned} J_{\lambda}(f(\underline{p}, t)) = & \sum_{i=1}^n \sum_{j=1}^m (z_{ij} - f(\underline{p}_i, t_j))^2 + \\ & + \lambda_S \sum_{k=1}^M \int_{\Omega} \left( \Delta(a_k(\underline{p})) \right)^2 d\underline{p} + \lambda_T \sum_{l=1}^N \int_{T_1}^{T_2} \left( \frac{\partial^2 b_l(t)}{\partial t^2} \right)^2 dt , \quad (1.4) \end{aligned}$$

dove  $\lambda_S > 0$  e  $\lambda_T > 0$  sono i parametri di smoothing che stabiliscono il peso della penalizzazione della regolarità della funzione rispettivamente in spazio e tempo.

### 1.1.4 Modello ideale

Il funzionale di penalizzazione riportato in eq. 1.4 è il risultato di una costruzione associata alla generalizzazione delle penalizzazioni marginali in spazio e tempo, ma idealmente si può ricondurre al seguente:

$$\begin{aligned} \tilde{J}_{\underline{\Lambda}}(f) = & \sum_{i=1}^n \sum_{j=1}^m (z_{ij} - f(\underline{p}_i, t_j))^2 + \\ & + \lambda_S \int_{T_1}^{T_2} \int_{\Omega} (\Delta f(\underline{p}, t))^2 d\underline{p} dt + \lambda_T \int_{\Omega} \int_{T_1}^{T_2} \left( \frac{\partial^2 f(\underline{p}, t)}{\partial t^2} \right)^2 dt d\underline{p}, \end{aligned} \quad (1.5)$$

dove  $J_S$  e  $J_T$  sono applicati direttamente alla funzione  $f(\underline{p}, t)$  e sono integrati, rispettivamente, sull'intervallo spaziale e il dominio spaziale.

Se si applica la forma per  $f(\underline{p}, t)$  dell'eq. 1.2 nel termine di penalizzazione del laplaciano in spazio, allora si può ritrovare:

$$\begin{aligned} \int_{T_1}^{T_2} \int_{\Omega} (\Delta f(\underline{p}, t))^2 d\underline{p} dt &= \int_{T_1}^{T_2} \int_{\Omega} \left( \Delta \left( \sum_{k=1}^M a_k(\underline{p}) \varphi_k(t) \right) \right)^2 d\underline{p} dt \\ &= \int_{T_1}^{T_2} \int_{\Omega} \left( \sum_{k=1}^M \Delta a_k(\underline{p}) \varphi_k(t) \right)^2 d\underline{p} dt \\ &= \int_{T_1}^{T_2} \int_{\Omega} \left( \sum_{k=1}^M \Delta a_k(\underline{p}) \varphi_k(t) \right) \left( \sum_{h=1}^M \Delta a_h(\underline{p}) \varphi_h(t) \right) d\underline{p} dt \\ &= \int_{T_1}^{T_2} \int_{\Omega} \left( \sum_{k=1}^M \sum_{h=1}^M \Delta a_k(\underline{p}) \Delta a_h(\underline{p}) \varphi_k(t) \varphi_h(t) \right) d\underline{p} dt \\ &= \sum_{k=1}^M \sum_{h=1}^M \int_{\Omega} \Delta a_k(\underline{p}) \Delta a_h(\underline{p}) d\underline{p} \int_{T_1}^{T_2} \varphi_k(t) \varphi_h(t) dt. \end{aligned} \quad (1.6)$$

Questo termine è equivalente a quello proposto in 1.4 se le basi temporali siano ortonormali, poichè in tal caso l'ultimo integrale in eq. 1.6 sarebbe 1 se  $k = h$  e 0 altrimenti.

Allo stesso modo, se si sostituisce la forma per  $f(\underline{p}, t)$  dell'eq. 1.3 nella

penalizzazione ideale dell'eq. 1.5 si ottiene:

$$\begin{aligned}
\int_{\Omega} \int_{T_1}^{T_2} \left( \frac{\partial^2 f(\underline{p}, t)}{\partial t^2} \right)^2 dt d\underline{p} &= \int_{\Omega} \int_{T_1}^{T_2} \left( \frac{\partial^2 \sum_{l=1}^N b_l(t) \psi_l(\underline{p})}{\partial t^2} \right)^2 dt d\underline{p} \\
&= \int_{\Omega} \int_{T_1}^{T_2} \left( \sum_{l=1}^N \frac{\partial^2 b_l(t)}{\partial t^2} \psi_l(\underline{p}) \right)^2 dt d\underline{p} \\
&= \int_{\Omega} \int_{T_1}^{T_2} \left( \sum_{l=1}^N \frac{\partial^2 b_l(t)}{\partial t^2} \psi_l(\underline{p}) \right) \left( \sum_{h=1}^N \frac{\partial^2 b_h(t)}{\partial t^2} \psi_h(\underline{p}) \right) dt d\underline{p} \\
&= \int_{\Omega} \int_{T_1}^{T_2} \left( \sum_{l=1}^N \sum_{h=1}^N \frac{\partial^2 b_l(t)}{\partial t^2} \frac{\partial^2 b_h(t)}{\partial t^2} \psi_l(\underline{p}) \psi_h(\underline{p}) \right) dt d\underline{p} \\
&= \sum_{l=1}^N \sum_{h=1}^N \int_{T_1}^{T_2} \frac{\partial^2 b_l(t)}{\partial t^2} \frac{\partial^2 b_h(t)}{\partial t^2} dt \int_{\Omega} \psi_l(\underline{p}) \psi_h(\underline{p}) d\underline{p}.
\end{aligned} \tag{1.7}$$

La stessa osservazione del caso precedente vale anche ora: se le basi in spazio sono ortonormali, si ritrova la penalizzazione proposta in 1.4.

Quindi in questo lavoro è proposto un modello che risulterà essere computazionalmente semplice ma non perfettamente equivalente a questo modello ideale. Le basi proposte, infatti, non sono ortonormali. Comunque sia, gli insiemi di basi che saranno adattati sono sparsi, cioè i termini  $\int_{T_1}^{T_2} \varphi_k(t) \varphi_l(t) dt$  e  $\int_{\Omega} \psi_l(\underline{p}) \psi_k(\underline{p}) d\underline{p}$  sono diversi da zero solo per poche coppie di indici.

### Discretizzazione dei termini di penalizzazione di $J_{\lambda}(f(\underline{p}, t))$

Così come è scritto in 1.4, il funzionale  $J_{\lambda}(f(\underline{p}, t))$  non è ancora adatto ad essere trattato computazionalmente. Per poter avere una forma che renda semplice la stima,  $f(\underline{p}, t)$  e  $J_{\lambda}(f(\underline{p}, t))$  saranno nuovamente discretizzati.

Il primo caso da analizzare è l'integrale del laplaciano dei coefficienti spazio-varianti dell'eq. 1.2 in modo analogo a quanto fatto in [6]. Fissato  $k$ , l'integrale

$$\int_{\Omega} \left( \Delta(a_k(\underline{p})) \right)^2 d\underline{p}$$

può essere semplificato introducendo la funzione  $g_k(\underline{p})$  come segue:

$$\int_{\Omega} g_k(\underline{p}) v(\underline{p}) d\underline{p} = \int_{\Omega} \Delta(a_k(\underline{p})) v(\underline{p}) d\underline{p} \quad \forall v(\underline{p}) \in L^2(\Omega). \tag{1.8}$$

Non è difficile verificare che se  $g_k(\underline{p})$  rispetta l'equazione precedente, per l'arbitrarietà di  $v$  allora

$$\int_{\Omega} \left( \Delta(a_k(\underline{p})) \right)^2 d\underline{p} = \int_{\Omega} \Delta(a_k(\underline{p})) g_k(\underline{p}) d\underline{p}. \tag{1.9}$$

Grazie alla la formula di Green ed alle condizioni di Neumann per  $a_k(\underline{p})$ , si possono semplificare gli integrali come segue:

$$\begin{aligned}\int_{\Omega} \Delta(a_k(\underline{p}))g_k(\underline{p})d\underline{p} &= - \int_{\Omega} \nabla a_k(\underline{p})\nabla g_k(\underline{p})d\underline{p} \\ \int_{\Omega} \Delta(a_k(\underline{p}))v(\underline{p})d\underline{p} &= - \int_{\Omega} \nabla a_k(\underline{p})\nabla v(\underline{p})d\underline{p} .\end{aligned}$$

Per poter calcolare analiticamente questi integrali è necessario introdurre l'uso delle basi spaziali  $\{\psi_l(\underline{p}); l = 1, \dots, N\}$  per le funzioni  $a_k$ ,  $g_k$  e  $v$ . Siano quindi:

$$a_k(\underline{p}) = \sum_{l=1}^N c_{lk}\psi_l(\underline{p}) \quad g_k(\underline{p}) = \sum_{l=1}^N g_{lk}\psi_l(\underline{p}) \quad v(\underline{p}) = \sum_{l=1}^N v_l\psi_l(\underline{p}) .$$

Per semplificare le notazioni saranno usati i seguenti vettori:

$$\underline{c}_k = \begin{bmatrix} c_{1k} \\ c_{2k} \\ \dots \\ c_{Nk} \end{bmatrix} \quad \underline{g}_k = \begin{bmatrix} g_{1k} \\ g_{2k} \\ \dots \\ g_{Nk} \end{bmatrix} \quad \underline{v} = \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_N \end{bmatrix}$$

e gli analoghi per le funzioni di base e le loro derivate parziali:

$$\begin{aligned}\underline{\psi} &= \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_n \end{bmatrix} \\ \underline{\psi}_x &= \begin{bmatrix} \partial\psi_1/\partial x \\ \partial\psi_2/\partial x \\ \vdots \\ \partial\psi_n/\partial x \end{bmatrix} \quad \underline{\psi}_y = \begin{bmatrix} \partial\psi_1/\partial y \\ \partial\psi_2/\partial y \\ \vdots \\ \partial\psi_n/\partial y \end{bmatrix} .\end{aligned} \tag{1.10}$$

Grazie all'uso delle funzioni di base e a ciò che è stato ottenuto dall'applicazione della formula di Green, le relazioni 1.8 e 1.9 diventano:

$$\underline{g}_k \left( \int_{\Omega} \underline{\psi} \underline{\psi}^T \right) \underline{v} = -\underline{c}_k \left( \int_{\Omega} (\underline{\psi}_x \underline{\psi}_x^T + \underline{\psi}_y \underline{\psi}_y^T) \right) \underline{v} \quad \forall \underline{v} \in \mathbb{R}^N$$

$$\int_{\Omega} \left( \Delta(a_k(\underline{p})) \right)^2 d\underline{p} = -\underline{c}_k \left( \int_{\Omega} (\underline{\psi}_x \underline{\psi}_x^T + \underline{\psi}_y \underline{\psi}_y^T) \right) \underline{g}_k ,$$

quindi, se si introducono le matrici (analogamente a quanto fatto in [6])

$$R_0 = \int_{\Omega} \underline{\psi} \underline{\psi}^T$$

$$R_1 = \int_{\Omega} (\underline{\psi}_x \underline{\psi}_x^T + \underline{\psi}_y \underline{\psi}_y^T)$$

si trova in definitiva che

$$\int_{\Omega} \left( \Delta(a_k(\underline{p})) \right)^2 d\underline{p} = \underline{c}_k^T R_1 R_0^{-1} R_1 \underline{c}_k = \underline{c}_k^T P_S \underline{c}_k \quad (1.11)$$

Si noti che la matrice  $P_S$  non dipende da  $k$ , e può essere considerata la stessa per tutte le funzioni  $a_k(\underline{p})$ . Inoltre è simmetrica, poichè  $R_0$  e  $R_1$  lo sono.

Grazie all'introduzione della discretizzazione in basi spaziali

$$a_k(\underline{p}) = \sum_{l=1}^N c_{lk} \psi_l(\underline{p})$$

è stato possibile ridurre la penalizzazione con l'integrale del quadrato di  $\Delta a_k(\underline{p})$  alla valutazione di una forma quadratica che non cambia con le funzioni  $a_k$ . Ma si ha anche un'altra conseguenza: per l'equivalenza ipotizzata tra le espressioni di  $f(\underline{p}, t)$  in 1.2 e 1.3, allora è necessario che i coefficienti tempo-varianti dell'espansione in basi spaziali assumano la seguente forma:

$$b_l(t) = \sum_{k=1}^M c_{lk} \varphi_k(t) .$$

Si ritrova quindi anche in questo caso l'espansione in funzioni di base.

Non resta altro che discretizzare anche  $\int_{T_1}^{T_2} \left( \frac{\partial^2 b_l(t)}{\partial t^2} \right)^2 dt$ . Ma dopo aver introdotto l'uso delle funzioni di base, allora se si definisce

$$P_T = \begin{bmatrix} \int_{T_1}^{T_2} \varphi_1''(t) \varphi_1''(t) dt & \int_{T_1}^{T_2} \varphi_1''(t) \varphi_2''(t) dt & \dots & \int_{T_1}^{T_2} \varphi_1''(t) \varphi_M''(t) dt \\ \int_{T_1}^{T_2} \varphi_2''(t) \varphi_1''(t) dt & \int_{T_1}^{T_2} \varphi_2''(t) \varphi_2''(t) dt & \dots & \int_{T_1}^{T_2} \varphi_2''(t) \varphi_M''(t) dt \\ \vdots & \vdots & \dots & \vdots \\ \int_{T_1}^{T_2} \varphi_M''(t) \varphi_1''(t) dt & \int_{T_1}^{T_2} \varphi_M''(t) \varphi_2''(t) dt & \dots & \int_{T_1}^{T_2} \varphi_M''(t) \varphi_M''(t) dt \end{bmatrix}$$

e il vettore

$$\underline{c}_l = \begin{bmatrix} c_{l1} \\ c_{l2} \\ \dots \\ c_{lM} \end{bmatrix} ,$$

si ritrova:

$$\int_{T_1}^{T_2} \left( \frac{\partial^2 b_l(t)}{\partial t^2} \right)^2 dt = \underline{c}_l^T P_T \underline{c}_l .$$

Anche la matrice  $P_T$  è simmetrica.

In conclusione, si può notare come la parte di penalizzazione per la regolarizzazione di  $f$  in 1.4 sia diventata un'unica forma quadratica. Per mostrarlo, si introduce il vettore

$$\underline{c} = \begin{bmatrix} c_{11} \\ \vdots \\ c_{1M} \\ c_{21} \\ \vdots \\ c_{2M} \\ \vdots \\ c_{NM} \end{bmatrix}$$

e la matrice  $P$ , definita con opportuni prodotti di Kronecker come segue:

$$P = \lambda_S (P_S \otimes I_M) + \lambda_T (I_N \otimes P_T) ,$$

dove  $I_M$  and  $I_N$  sono matrici identità di dimensioni  $M \times M$  e  $N \times N$  rispettivamente. Allora si avrà:

$$\begin{aligned} & + \lambda_S \sum_{k=1}^M \int_{\Omega} \left( \Delta(a_k(\underline{p})) \right)^2 d\underline{p} + \lambda_T \sum_{l=1}^N \int_{T_1}^{T_2} \left( \frac{\partial^2 b_l(t)}{\partial t^2} \right)^2 dt \\ & \lambda_S \sum_{k=1}^M \underline{c}_k^T P_S \underline{c}_k + \lambda_T \sum_{l=1}^N \underline{c}_l^T P_T \underline{c}_l = \underline{c}^T P \underline{c} . \quad (1.12) \end{aligned}$$

A causa della simmetria dei termini con cui è costruita, anche la matrice  $P$  è simmetrica.

### 1.1.5 Soluzione del problema di stima

Grazie a quanto ricavato nel paragrafo precedente, la parte di penalizzazione delle derivate del funzionale  $J_{\lambda}(f(\underline{p}, t))$  si è ridotta ad una forma quadratica. Ma questo è stato possibile grazie all'espressione in funzione di base per i coefficienti delle eq. 1.2 e 1.3:

$$a_k(\underline{p}) = \sum_{l=1}^N c_{lk} \psi_l(\underline{p}) \quad b_l(t) = \sum_{k=1}^M c_{lk} \varphi_k(t) .$$

Essendo 1.2 e 1.3, allora in definitiva:

$$f(\underline{p}, t) = \sum_{l=1}^N \sum_{k=1}^M c_{lk} \psi_l(\underline{p}) \varphi_k(t), \quad (1.13)$$



cioè la funzione da stimare è la combinazione lineare di tutti i possibili prodotti incrociati tra le funzioni di base in tempo e spazio. Questa formulazione può essere considerata la definitiva per la funzione  $f(\underline{p}, t)$ , e permette di poter identificare la funzione con il vettore dei suoi coefficienti  $\underline{c}$ . Inoltre, ne consegue che è possibile scrivere in modo più agevole il funzionale  $J_{\lambda}(f(\underline{p}, t))$ .

Siano definiti il vettore dei valori osservati

$$\underline{z} = \begin{bmatrix} z_{11} \\ \vdots \\ z_{1m} \\ z_{21} \\ \vdots \\ z_{2m} \\ \vdots \\ z_{nm} \end{bmatrix} \quad (1.14)$$

e le matrici  $\Phi$  (con le valutazioni delle basi spaziali nei punti  $\{\underline{p}_i; i = 1, \dots, n\}$ ) e  $\varphi$  (con le valutazioni delle basi temporali  $\{t_j; j = 1, \dots, m\}$ ):

$$\Psi = \begin{bmatrix} \psi_1(\underline{p}_1) & \psi_2(\underline{p}_1) & \dots & \psi_N(\underline{p}_1) \\ \psi_1(\underline{p}_2) & \psi_2(\underline{p}_2) & \dots & \psi_N(\underline{p}_2) \\ \vdots & \vdots & \dots & \vdots \\ \psi_1(\underline{p}_n) & \psi_2(\underline{p}_n) & \dots & \psi_N(\underline{p}_n) \end{bmatrix}$$

$$\Phi = \begin{bmatrix} \varphi_1(t_1) & \varphi_2(t_1) & \dots & \varphi_M(t_1) \\ \varphi_1(t_2) & \varphi_2(t_2) & \dots & \varphi_M(t_2) \\ \vdots & \vdots & \dots & \vdots \\ \varphi_1(t_m) & \varphi_2(t_m) & \dots & \varphi_M(t_m) \end{bmatrix}.$$

Allora, se si definisce la matrice  $B$  tramite il prodotto di Kronecker

$$B = \Psi \otimes \Phi,$$

si può facilmente dire che

$$\begin{bmatrix} f(\underline{p}_1, t_1) \\ \vdots \\ f(\underline{p}_1, t_m) \\ f(\underline{p}_2, t_1) \\ \vdots \\ f(\underline{p}_2, t_m) \\ \vdots \\ f(\underline{p}_n, t_m) \end{bmatrix} = B \underline{c}. \quad (1.15)$$

Quindi è possibile dare una forma definitiva al funzionale di penalizzazione:

$$J_{\lambda}(\underline{c}) = (\underline{z} - B\underline{c})^T (\underline{z} - B\underline{c}) + \underline{c}^T P \underline{c} , \quad (1.16)$$

e per ricavare la stima della funzione  $f(\underline{p}, t)$  sarà sufficiente ricavare il vettore dei coefficienti  $c_k l$  risolvendo il problema di minimo:

$$\hat{\underline{c}} = \arg \min_{\underline{c} \in \mathbb{R}^{NM}} J_{\lambda}(\underline{c}) .$$

Grazie alla formulazione ricavata in 1.15, allora basta derivare per ottenere la soluzione al problema di stima. Grazie alla simmetria di  $P$ , si ha:

$$\frac{\partial}{\partial \underline{c}} J = -2B^T \underline{z} + 2(B^T B + P)\underline{c} ,$$

che posta uguale a zero, porta all'equazione

$$(B^T B + P)\underline{c} = B^T \underline{z}$$

e in conclusione

$$\hat{\underline{c}} = (B^T B + P)^{-1} B^T \underline{z} .$$

Il problema di stima è stato risolto, e la soluzione si ricava solo dalla risoluzione di un sistema lineare, seppur di grandi dimensioni (la matrice  $(B^T B + P)$  ha dimensioni  $NM \times NM$ , e già dagli esempi si potranno notare dimensioni elevate).

L'ultimo elemento da definire è la *smoothing matrix*  $S$ , usata per conoscere i valori stimati dal modello direttamente da quelli osservati:

$$\hat{\underline{z}} = B\hat{\underline{c}} = B(B^T B + P)^{-1} B^T \underline{z} = S \underline{z} .$$

### 1.1.6 Proprietà statistiche di $\hat{\underline{c}}$

Il modello di partenza indicato in 1.1 può essere scritto anche in forma matriciale

$$\underline{z} = B\underline{c} + \underline{\varepsilon} . \quad (1.17)$$

A causa delle proprietà statistiche di  $\underline{\varepsilon}$

$$\mathbb{E}[\underline{\varepsilon}] = \underline{0} \quad \text{Var}[\underline{\varepsilon}] = \sigma^2 I_{nm}$$

sia ha

$$\mathbb{E}[\underline{z}] = B\underline{c} \quad \text{Var}[\hat{\underline{c}}] = \sigma^2 I_{nm}$$

e quindi è immediato ricavare per lo stimatore  $\hat{\underline{c}}$  (grazie alle proprietà simmetria di  $S$ ):

$$\mathbb{E}[\hat{\underline{c}}] = (B^T B + P)^{-1} B^T B \underline{c} \quad \text{Var}[\underline{z}] = \sigma^2 (B^T B + P)^{-1} B^T B (B^T B + P)^{-1} .$$

Non è stata ipotizzata la gaussianità per  $\underline{\varepsilon}$ , ma se fosse ipotizzata anche  $\hat{\underline{c}}$  sarebbe gaussiano. Grazie a questa ulteriore ipotesi sarebbe possibile elaborare (con una data significatività) una regione di confidenza per  $\hat{\underline{c}}$  e quindi una banda di confidenza per la funzione stimata  $f$ .

## 1.2 Caso con covariate

Il modello si estende facilmente se si prevede che il dato possa essere influenzato da covariate. Il modello di (1.1) diventa:

$$z_{ij} = \underline{w}_{ij}^T \underline{\beta} + f(\underline{p}_i, t_j) + \varepsilon_{ij} \quad i = 1, \dots, n \quad j = 1, \dots, m, \quad (1.18)$$

dove  $\underline{w}_{ij}$  è il vettore delle  $p$  covariate associate a  $z_{ij}$  e  $\underline{\beta}$  è il vettore dei coefficienti di regressione. Di conseguenza, il funzionale discreto di 1.15 diventa:

$$J = (\underline{z} - W\underline{\beta} - B\underline{c})^T (\underline{z} - W\underline{\beta} - B\underline{c}) + \underline{c}^T S \underline{c},$$

dove  $W$  è la matrice  $nm \times p$  con i vettori  $\{\underline{w}_{ij}; i = 1, \dots, n; j = 1, \dots, m\}$ .

Per ricavare la soluzione occorre derivare questa espressione rispetto a  $\underline{\beta}$  e  $\underline{c}$ :

$$\frac{\partial}{\partial \underline{\beta}} J = -2W^T \underline{z} + 2W^T B \underline{c} + 2W^T W \underline{\beta},$$

$$\frac{\partial}{\partial \underline{c}} J = -2B^T \underline{z} + 2B^T W \underline{\beta} + 2(B^T B + P) \underline{c}.$$

Imponendo che le derivate siano uguali a zero si hanno le seguenti equazioni:

$$\begin{cases} W^T W \hat{\underline{\beta}} = W^T (\underline{z} - B \hat{\underline{c}}) \\ (B^T B + P) \hat{\underline{c}} = B^T (\underline{z} - W \hat{\underline{\beta}}) \end{cases}.$$

che ricordano le equazioni usate per la regressione e per il modello senza covariate, con la differenza che in questo caso a  $\underline{z}$  è sottratto in entrambi i casi la parte spiegata dal termine di modello a cui non si riferiscono  $\hat{\underline{\beta}}$  e  $\hat{\underline{c}}$  rispettivamente.

A questo punto si possono ricavare le soluzioni. Si ha:

$$\hat{\underline{c}} = [B^T B + P + B^T W (W^T W)^{-1} W^T B]^{-1} B^T [I - W (W^T W)^{-1} W^T] \underline{z} = A Q \underline{z}$$

con  $A = [B^T B + P + B^T W (W^T W)^{-1} W^T B]^{-1} B^T$  e  $Q = [I - W (W^T W)^{-1} W^T]$ , matrice molto importante nel caso di regressione lineare, poiché proietta il vettore dei dati nel sottospazio ortogonale allo spazio generato dalle colonne della matrice disegno, ricavando così il vettore dei residui. Questa matrice si ritrova anche in questo caso, ed è ancora valide le sue proprietà:

- $Q$  è idempotente, cioè  $Q Q = Q$ ;
- $Q$  è simmetrica;
- a causa del fatto che proietta nel sottospazio ortogonale di  $\text{Col}(W)$ ,  $Q W$  risulta essere la matrice nulla di opportune dimensioni.

Infine, la stima di  $\hat{\underline{\beta}}$  si ricava dalla stima ottenuta per  $\hat{\underline{c}}$ :

$$(W^T W)^{-1} W^T (I - BAQ) \underline{z}$$

In modo analogo al caso senza covariate, è necessario ricavare anche per questo caso la *smoothing matrix*:

$$\hat{\underline{z}} = B\hat{\underline{c}} + W\hat{\underline{\beta}} = [BAQ + W(W^T W)^{-1} W^T (I - BAQ)] \underline{z} = S \underline{z}.$$

### 1.2.1 Proprietà statistiche di $\hat{\underline{c}}$ e $\hat{\underline{\beta}}$

Anche in questo caso è possibile ricavare valore atteso e varianza degli stimatori ottenuti, ed è utile in quanto consente di verificare la significatività delle covariate. Per farlo però, è necessario avere la forma matriciale del modello indicato in 1.18:

$$\underline{z} = B\underline{c} + W\underline{\beta} + \underline{\varepsilon}. \quad (1.19)$$

Di nuovo si ha

$$\mathbb{E}[\underline{\varepsilon}] = \underline{0} \quad \text{Var}[\underline{\varepsilon}] = \sigma^2 I_{nm}$$

e di conseguenza

$$\mathbb{E}[\underline{z}] = B\underline{c} + W\underline{\beta} \quad \text{Var}[\underline{z}] = \sigma^2 I_{nm}.$$

Grazie a ciò e alle proprietà ricavate per la matrice  $Q$  è possibile ricavare che:

$$\mathbb{E}[\hat{\underline{c}}] = AQB\underline{c} \quad \text{Var}[\hat{\underline{c}}] = \sigma^2 AQA^T.$$

Per  $\hat{\underline{\beta}}$  i calcoli sono più complessi, ma si semplificano grazie alle proprietà indicate in precedenza per la matrice  $Q$ . Si ritrova:

$$\mathbb{E}[\hat{\underline{\beta}}] = \underline{\beta} + (W^T W)^{-1} W^T (I - BAB) \underline{c}$$

$$\text{Var}[\hat{\underline{\beta}}] = \sigma^2 (W^T W)^{-1} + \sigma^2 (W^T W)^{-1} W^T BAQA^T B^T W (W^T W)^{-1}.$$

Come nel caso senza covariate, anche ora si potrebbe stimare la gaussianità degli stimatori se fosse ipotizzata per  $\underline{\varepsilon}$ . Questo permette di elaborare intervalli di confidenza per le componenti di  $\underline{\beta}$  e di verificare la significatività delle covariate.

## 1.3 Stima di $\sigma^2$ e scelta dei parametri $\lambda_S$ e $\lambda_T$

Quanto riportato di seguito è valido indipendentemente dal fatto che siano inserite nel modello le covariate. Perciò è valido per entrambi i modelli proposti in precedenza.

### 1.3.1 Stima di $\sigma^2$

Stimare la varianza dell'errore è necessario se si vuole fare inferenza sugli stimatori ed è molto semplice se si conoscono i gradi di libertà equivalenti del modello. Ma questi si ricavano dalla *smoothing matrix*:

$$\text{EDF} = \text{tr}(S) .$$

Con questo valore si calcola la stima della varianza, usando i residui e il numero totale di dati:

$$\hat{\sigma}^2 = \frac{1}{nm - \text{tr}(S)} (\underline{z} - \hat{\underline{z}})^T (\underline{z} - \hat{\underline{z}})$$

### 1.3.2 Parametri di smoothing

I parametri  $\lambda_S$  e  $\lambda_T$  hanno un ruolo rilevante nella stima della soluzione, poichè scelgono quanto peso dare allo smoothing della funzione in spazio e tempo. Quindi è opportuno che siano fissati accuratamente prima della stima della soluzione.

Secondo quanto indicato in [4], la scelta corretta si ha con il valore di  $\underline{\lambda}$  che realizza il minimo dell'indice di *generalized cross validation*

$$GCV(\underline{\lambda}) = \frac{nm}{nm - \text{tr}(S)} D(\hat{\underline{c}}, \hat{\underline{\beta}}) ,$$

dove  $D$  è la devianza del modello. Si ha:

$$D(\hat{\underline{c}}, \hat{\underline{\beta}}) = 2\sigma^2 (l_{\text{sat}} - l(\hat{\underline{c}}, \hat{\underline{\beta}})) ,$$

dove  $l$  è la logverosimiglianza del modello, che si ipotizza gaussiano, valutata rispettivamente nel suo massimo (valore di saturazione) e in corrispondenza dei valori stimati. Non è difficile dimostrare che, sia nel caso con covariate che senza covariate, si ha:

$$D(\hat{\underline{c}}, \hat{\underline{\beta}}) = (\underline{z} - \hat{\underline{z}})^T (\underline{z} - \hat{\underline{z}})$$

Di conseguenza, il miglior  $\underline{\lambda}$  può essere scelto come valore che minimizza

$$GCV(\underline{\lambda}) = \frac{nm}{nm - \text{tr}(S)} (\underline{z} - \hat{\underline{z}})^T (\underline{z} - \hat{\underline{z}}) .$$

si ha

$$\hat{f} = \arg \min_{\underline{c} \in \mathbb{R}^{NM}} J_{\underline{\lambda}}(f).$$

,  $\{c_{lk}; l = 1, \dots, N; k = 1, \dots, M\}$  sono coefficienti dell'espansione della funzione in basi separabili spazio-temporali. Le espressioni di  $f(\underline{p}, t)$  riportate spiegano che il campo spazio-temporale è il risultato della generalizzazione dei modelli marginali rispettivamente in tempo e in spazio. Perciò

la formulazione nell'eq. ?? deriva dall'introduzione della possibilità, per i coefficienti dell'espressione in basi di tempo in 1.2, di poter variare in spazio o analogamente, per i coefficienti dell'espressione in basi spaziali di 1.3, di poter variare in tempo. Ne consegue che anche essi siano esprimibili con le rispettive basi:

$$a_k(\underline{p}) = \sum_{l=0}^N c_{lk} \psi_l(\underline{p})$$

La formulazione finale della funzione  $f(\underline{p}, t)$  è quindi la combinazione lineare di tutti i possibili prodotti incrociati tra i due insiemi delle funzioni di base (eq. ??), ma concettualmente proviene dalla generalizzazione dei modelli marginali.

Il fondamento della modellizzazione del campo spazio-temporale sui modelli marginali in spazio e tempo introduce il concetto di separabilità più volte presente in questa formulazione. Come si potrà notare anche in seguito, anche i termini di penalizzazione saranno derivati dall'estensione di termini puramente spaziali o temporali.

# Bibliografia

- [1] Nicole H. Augustin, Verena M. Trenkel, Simon N. Wood, Pascal Lorange, *Space-time modelling of blue ling for fisheries stock management*, *Environmetrics*, 24, 109–119, (2013)
- [2] Laura Azzimonti, Laura M. Sangalli, Piercesare Secchi, Maurizio Domanin, Fabio Nobile, *Blood flow velocity field estimation via spatial regression with PDE penalization*, *Journal of the American Statistical Association*, (2015) DOI: 10.1080/01621459.2014.946036
- [3] Peter Craven, Grace Wahba, *Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation*, *Numerische Mathematik*, 31, 377–403, (1979)
- [4] Giampiero Marra, David L. Miller, Luca Zanin, *Modelling the spatio-temporal distribution of the incidence of resident foreign population*, *Statistica Neerlandica*, 66, 133–160, (2012)
- [5] Timothy O. Ramsay, *Spline smoothing over difficult regions*, *Journal of the Royal Statistical Society: Series B*, 64, 307–319, (2002)
- [6] Laura M. Sangalli, James O. Ramsay, Timothy O. Ramsay, *Spatial spline regression models*, *Journal of the Royal Statistical Society: Series B*, 75, 681–703, (2013)
- [7] Simon N. Wood, Mark W. Bravington, Sharon L. Hedley, *Soap film smoothing*, *Journal of the Royal Statistical Society: Series B*, 70, 931–955, (2008)