Capitolo 1

Geostatistica per l'analisi di fenomeni temporali

Tradizionalmente, nell'analisi delle serie storiche non si ricorre a metodologie geostatistiche, le quali, di solito, vengono applicate per analizzare, attraverso il variogramma, la relazione spaziale tra le osservazioni campionate ad alcune localizzazioni di un dominio e prevedere il corrispondente fenomeno spaziale (Chilés J.P. e Delfiner P., 1999; Journel A.G. e Huijbregts C.J., 1981; Matheron G., 1963; Posa D. e De Iaco S., 2009).
Tuttavia, in tale contesto, può risultare conveniente integrare gli strumenti geostatistici con quelli solitamente utilizzati, al fine di affrontare problematiche di stima in condizione di interpolazione, oltre che di estrapolazione.
Nel presente capitolo, dopo aver illustrato le principali caratteristiche di un fenomeno temporale, saranno descritte le differenze fondamentali tra la Geostatistica e l'analisi temporale classica, nonché illustrata l'utilità degli strumenti geostatistici in tale contesto. In particolare, sarà sottolineato il ruolo del variogramma nella modellizzazione e previsione di fenomeni temporali.

1.1 Caratteristiche dei fenomeni temporali

Nella maggior parte dei settori delle scienze applicate, quali l'economia (produzione, reddito, ...), la finanza (quotazioni azionarie, tassi di cambio, ...), l'ingegneria ambientale (concentrazione di inquinanti nell'aria, nell'acqua,...) e la demografia (popolazione residente, popolazione straniera,...), l'analisi delle dinamiche temporali dei fenomeni che intervengono in tali contesti, diviene di fondamentale importanza. Essi si sviluppano nel tempo e sono denominati processi temporali.
Se per ciascun tempo $t \in T$, si definisce una variabile aleatoria X_t, in grado di descrivere le possibili manifestazioni di un fenomeno al tempo t, allora l'insieme delle variabili aleatorie X_t, con $t \in T$, definite su uno stesso spazio di probabilità, è denominato processo stocastico ed è indicato con la notazione (Chatfield C., 2004)

$$\{X_t, t \in T \subseteq \mathbb{R}\}.$$

Assegnato un processo stocastico X_t, è possibile distinguere:

- una componente deterministica (*trend*), indicata con μ_t;
- una componente residua, indicata con Y_t.

Il modello fondamentale per un processo stocastico X_t, può essere espresso come segue:

$$X_t = \mu_t + Y_t.$$

È opportuno sottolineare che la componente strutturale viene solitamente individuata ricorrendo a strumenti matematici (ad esempio il metodo dei minimi quadrati), mentre per la componente residua si richiede l'analisi della correlazione temporale.
L'insieme delle osservazioni, effettuate sequenzialmente nel tempo è definito "serie storica" e può essere considerato come un campionamento ad istanti di tempo consecutivi, di solito equispaziati, della più lunga sequenza di variabili aleatorie che compongono il processo stocastico. Tale campione può essere costituito da osservazioni ad un numero finito di istanti temporali oppure da un'osservazione continua su un intervallo di tempo.
Nel caso in cui tali osservazioni siano relative ad istanti di tempo t equispaziati, una serie storica viene definita come una successione finita di valori x_t, $t = 1, 2, ..., N$, ed è indicata come segue:

$$\{x_t,\ t = 1, 2, ..., N\}.$$

Gli aspetti che caratterizzano una serie storica sono

a) l'ordine temporale,

b) lo schema di campionamento,

c) la periodicità ed il *trend* nel lungo periodo,

d) la dipendenza tra le osservazioni.

Quest'ultima assunzione non è presente nell'analisi statistica classica che viene sviluppata, essenzialmente, sulla base di un campione casuale di osservazioni indipendenti, per cui ogni dato è denso di informazioni sulla variabile casuale che lo ha generato, ma non aggiunge alcuna informazione sull'osservazione successiva.
Lo studio di n osservazioni temporali di un fenomeno X consente, invece, di prevedere valori futuri per tale variabile. Pertanto, al fine di evidenziare ed utilizzare la connessione che caratterizza una successione di dati temporali, l'analisi delle serie storiche fa riferimento alla teoria dei processi stocastici (Bloomfield P., 2000; Box G.E.P. e Jenkins G.M., 1976; Brockwell P.J., Davis R.A., 1987).
In particolare, assegnato un processo stocastico $\{X_t,\, t \in T\}$, esso è caratterizzato dalla seguente funzione di ripartizione congiunta

$$F_{t_1,t_2,...,t_n}(x_1, x_2, ..., x_n) = Prob(X_{t_1} \leq x_1, X_{t_2} \leq x_2, ..., X_{t_n} \leq x_n). \tag{1.1}$$

Per tale processo è possibile definire i seguenti momenti del primo e secondo ordine, come di seguito riportato (Piccolo D., 1990).

- *Momento del primo ordine:*

    1. Valore atteso
    $$\mu_t = E(X_t), \qquad t \in T$$
    ovvero il valore medio del processo.
    Se si assume che tale valore sia costante, $\mu_t = \mu$, allora la stima sarà $\hat{\mu} = \dfrac{1}{n}\displaystyle\sum_{t=1}^{n} x_t$.

- *Momenti del secondo ordine*

    1. Varianza temporale, definita per ogni $t \in T$,
    $$\sigma_t^2 = Var(X_t) = E(X_t - \mu_t)^2.$$
    Se si assume che la varianza sia costante, $\sigma_t^2 = \sigma^2$, allora la stima sarà $\hat{\sigma}^2 = \frac{1}{n-1}\displaystyle\sum_{t=1}^{n}(x_t - \hat{\mu})^2$.
    Tuttavia, vi sono modelli per serie storiche con varianza non costante (*GARCH models*).

2. Covarianza temporale o autocovarianza, ovvero

$$C(t,s) = Cov(X_t, X_s) = E[(X_t - \mu_t)(X_s - \mu_s)], \qquad \forall t, s \in T.$$

3. Coefficiente di autocorrelazione temporale, ovvero

$$\rho(X_t, X_s) = \frac{Cov(X_t, X_s)}{\sqrt{Var(X_t)Var(X_s)}} \qquad \forall t, s \in T.$$

4. Variogramma

$$2\gamma(t_i, t_j) = Var(X_{t_i} - X_{t_j}),$$

essendo $\gamma(t_i, t_j) = \dfrac{Var(X_{t_i} - X_{t_j})}{2}$ il semivariogramma.

In particolare, l'autocovarianza per un processo stocastico $\{X_t,\, t \in T\}$ deve essere una funzione semi-definita positiva, per cui deve essere soddisfatta la seguente condizione:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C(t,s) \geq 0$$

$$\forall n \in \mathbb{N}, \qquad \forall t, s \in T,\ a_i \in \mathbb{R},\ i = 1, 2, ..., n.$$

Tale condizione assicura che la varianza di una qualsiasi combinazione lineare $Y = \sum_{i=1}^{n} a_i X_t$ sia non negativa, ovvero

$$Var\left[\sum_{i=1}^{n} a_i X_t\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C(t,s) \geq 0.$$

Un processo stocastico $\{X_t, t \in T\}$ è stazionario in senso stretto se le distribuzioni finito-dimensionali sono invarianti sotto una traslazione arbitraria di istanti temporali, ovvero se:

$$F_{t_1, t_2, ..., t_n}(x_1, x_2, ..., x_n) = F_{t_1+k, t_2+k, ..., t_n+k}(x_1, x_2, ..., x_n)$$

$$\forall k \in \mathbb{N}, \forall x_i \in \mathbb{R}, i = 1, 2, ..., n.$$

È opportuno evidenziare che se la serie storica è stazionaria in senso stretto, la funzione di distribuzione della variabile aleatoria è la stessa in ogni punto del dominio. Inoltre, la funzione di ripartizione congiunta dipende solo dalla distanza tra gli elementi del dominio e non dai loro valori. Naturalmente, ciò non significa che una particolare realizzazione si presenterà uguale ad un'altra realizzazione.