

# Analisi Spaziale Bayesiana dei Prezzi delle Abitazioni a Milano

Addis Gaia, Nobili Marco, Tusetti Lucrezia

23 aprile 2025

## Abstract

---

Il presente studio si concentra sulla previsione del prezzo al metro quadro degli immobili residenziali milanesi, tramite modelli di regressione spaziale in ambito Bayesiano.

L'analisi ha previsto la specificazione e la stima di tre modelli Bayesiani basati su dati point-referenced: (i) un modello con sola struttura spaziale, (ii) un modello con sole covariate e (iii) un modello con entrambe le componenti. In parallelo, è stato implementato un Kriging Universale come benchmark frequentista. La selezione delle covariate è stata effettuata tramite approccio spike-and-slab.

Le metriche di errore calcolate su un test set confermano la superiorità predittiva del modello che integra informazioni spaziali e covariate, sottolineando il ruolo chiave della dipendenza spaziale nella modellazione dei prezzi immobiliari in ambito urbano.

Infine, l'analisi è stata estesa al contesto areale, costruendo modelli spaziali distinti per tipologia di appartamento (mono-, bi-, tri- e quadrilocali), basati su prezzi medi per quartiere. Questo approccio ha permesso di fornire una lettura complementare e aggregata del mercato immobiliare milanese.

---

## 1 Introduzione

La città di Milano è uno dei principali centri economici e sociali d'Italia, ma la sua estensione territoriale risulta relativamente contenuta rispetto alla sua rilevanza. L'elevata attrattività della città genera una domanda abitativa in costante crescita, che si scontra con la scarsa disponibilità di aree edificabili, contribuendo così a un progressivo aumento dei prezzi degli immobili.

Negli ultimi anni, il mercato immobiliare milanese ha infatti attirato un'attenzione crescente, soprattutto per quanto riguarda l'aumento dei costi, le difficoltà di accesso per alcune categorie di cittadini e il timore, più volte sollevato, di una possibile bolla speculativa.

### 1.1 Obiettivi

In questo scenario, la possibilità di stimare con precisione il prezzo al metro quadro degli immobili assume un'importanza strategica, sia per gli operatori del settore sia per le istituzioni.

Il presente lavoro si propone di affrontare questo problema attraverso l'uso di modelli Bayesiani spaziali, che consentono di integrare in modo coerente le caratteristiche dell'abitazione con la sua posizione geografica all'interno della città. Infatti, la localizzazione di un immobile ha un'influenza diretta e non trascurabile sul suo valore, e tali modelli permettono di catturare le dipendenze geografiche che spesso sfuggono ai modelli di regressione tradizionali.

## 1.2 Descrizione del dataset

Il dataset utilizzato per l'analisi (Tabella 1) raccoglie informazioni aggiornate a Febbraio 2023 su abitazioni in vendita nella città di Milano. Ogni osservazione rappresenta un'unità abitativa, caratterizzata da variabili strutturali ed energetiche. La componente spaziale è stata incorporata tramite la variabile `geometry`, che contiene le coordinate geografiche degli immobili nel sistema Gauss-Boaga (EPSG:3003), adatto all'analisi spaziale sul territorio italiano. Tali coordinate sono state ottenute tramite un'operazione di georeferenziazione a partire dagli indirizzi contenuti nel dataset. È inoltre stato associato ad ogni abitazione il codice relativo alla zona OMI (Osservatorio del Mercato Immobiliare) a cui appartiene. Le osservazioni esterne ai confini comunali sono state escluse.

VARIABILE	DESCRIZIONE	MODALITÀ/FORMATO
<code>price</code>	Prezzo di vendita dell'immobile (in euro)	numeric
<code>m2</code>	Superficie in metri quadrati	numeric
<code>rooms</code>	Numero di stanze dell'abitazione	integer
<code>bathrooms</code>	Numero di bagni dell'abitazione	integer
<code>condominium expenses</code>	Spese condominiali (in euro)	numeric
<code>year of build</code>	Anno di costruzione dell'immobile	integer
<code>condition</code>	Stato conservativo, con quattro livelli ordinati: da ristrutturare, abitabile, ristrutturata, nuova	factor
<code>energy efficiency</code>	Classe energetica (da A a G)	factor
<code>car parking</code>	Numero di posti auto inclusi	integer
<code>elevator</code>	Presenza dell'ascensore: sì, no	factor
<code>floor level</code>	Piano su cui si trova l'abitazione	integer
<code>heating centralized</code>	Tipologia di riscaldamento: centralizzato, autonomo	factor
<code>neighborhood</code>	Quartiere in cui è situato l'immobile	character
<code>address</code>	Via in cui è situato l'immobile	character
<code>geometry</code>	Coordinate in formato Gauss-Boaga	coord

Tabella 1: Descrizione delle variabili e formato dei dati

A partire dalle variabili `price` e `m2` si crea la variabile dipendente  $y$  come:  $y = \log\left(\frac{\text{price}}{m^2}\right)$ . Tale trasformazione ha un duplice fine: uniformare il confronto tra immobili, esprimendo il valore in termini di prezzo al metro quadro anziché prezzo totale, e rendere la variabile dipendente più simile a una distribuzione Normale, migliorando l'efficacia e la robustezza del modello.

## 2 Cenni teorici ai modelli di regressione spaziali Bayesiani

### 2.1 Selezione delle covariate

Al fine di ridurre la complessità del modello e i tempi computazionali, si è applicato il metodo spike-and-slab per la selezione delle covariate rilevanti.

Si introducono quindi le variabili aumentate  $\underline{\gamma}^T = (\gamma_1, \dots, \gamma_p)$ ,  $\gamma_j \in \{0, 1\}$ ,  $\forall j = 1, \dots, p$ , dove:

$$\gamma_j = \begin{cases} 1 & \text{se la } j\text{-esima variabile viene inclusa nel modello} \\ 0 & \text{se la } j\text{-esima variabile viene esclusa dal modello} \end{cases}$$

Le variabili  $\underline{\gamma}$  vengono utilizzate per costruire una prior per ogni coefficiente di regressione, che sia una mistura della componente slab e della componenti spike.

In particolare, il modello specificato è:

$$\begin{aligned} \beta_j | \gamma_j &\stackrel{\text{ind}}{\sim} \gamma_j N(0, c_j^2 \tau_j^2) + (1 - \gamma_j) N(0, \tau_j^2) \\ \gamma_j | \theta_j &\stackrel{\text{ind}}{\sim} \text{Bernoulli}(\theta_j) \\ \theta_j &\stackrel{\text{ind}}{\sim} \pi(\theta_j) \end{aligned}$$

dove  $c_j > 1$  e grande a piacere,  $j = 1, \dots, p$ .

Sulla base dell'informazione empirica, dopo aver implementato iterativamente numerosi modelli, si selezionano tramite Highest Posterior Probability le covariate che risultano più utili alla regressione e che vengono pertanto utilizzate come predittori. Si calcola quindi:

$$\hat{\underline{\gamma}} = \arg \max_{\underline{\gamma} \in \{0,1\}^p} \pi(\underline{\gamma} | y_{1:n}, \underline{x}_{1:n}) \approx \arg \max_{\underline{\gamma} \in \{0,1\}^p} \frac{1}{R} \sum_{r=1}^R \mathbb{I}_{[\underline{\gamma}^{(r)} = \underline{\gamma}]}.$$

### 2.2 Dati puntuali

I dati spaziali di tipo puntuale sono una realizzazione della variabile aleatoria  $Y(\underline{s})$ , dove  $\underline{s} \in D \subset R^2$  e denota una generica posizione spaziale all'interno del dominio di interesse  $D$ .

Un modello Bayesiano di regressione che tenga conto della dipendenza spaziale è dato da:

$$Y(\underline{s}) = \underline{x}^T(\underline{s})\underline{\beta} + \omega(\underline{s}) + \epsilon(\underline{s})$$

Dove:

- $\omega(\underline{s})$  è il random effect, un processo gaussiano spaziale che cattura i residui dell'associazione spaziale e la cui distribuzione è indicizzata dai parametri  $\sigma^2$  e  $\phi$ ;
- $\epsilon(\underline{s})$  rappresenta la variabilità a distanze minori della distanza minima tra i punti rilevati, la cui distribuzione è indicizzata da  $\tau^2$ .

Il modello applicato nel lavoro è di tipo marginale ed è il risultato dell'integrazione dei random effects. In particolare, la specificazione di tale modello è:

$$\begin{aligned} \underline{Y} | \underline{\beta}, \tau^2, \sigma^2, \phi &\sim N(\underline{X}\underline{\beta}, \sigma^2 H(\phi) + \tau^2 I_n) \\ \underline{\beta} &\sim N(\underline{b}_0, \Lambda_0) \end{aligned}$$

dove  $[H(\phi)]_{ij} = \rho(\|\underline{s}_i - \underline{s}_j\|, \phi)$ . Inoltre:

$$\begin{aligned} \sigma^2 &\sim IG(a_\sigma, b_\sigma) \\ \tau^2 &\sim IG(a_\tau, b_\tau) \\ \phi &\sim IG(a_\phi, b_\phi) \end{aligned}$$

Si denota con  $y_0 = Y(\underline{s}_0)$  il punto in cui si vuole fare previsione, date le covariate  $\underline{x}_0 = \underline{x}(\underline{s}_0)$ .

Considerando i dati distribuiti in modo Gaussiano, la previsione viene ottenuta generando da una distribuzione Normale i cui momenti sono:

$$\begin{aligned} E[Y_0 | \underline{y}] &= \underline{x}_0^T \underline{\beta} + \underline{\gamma}^T (\sigma^2 H(\phi) + \tau^2 I_n)^{-1} (\underline{y} - \underline{X}\underline{\beta}) \\ Var(Y_0 | \underline{y}) &= \sigma^2 + \tau^2 - \underline{\gamma}^T (\sigma^2 H(\phi) + \tau^2 I_n)^{-1} \underline{\gamma}. \end{aligned}$$

### 2.3 Dati areali

Nel contesto dei dati spaziali di tipo areale,  $Y(s)$  è l'osservazione del fenomeno, dove  $s \in D$  e il dominio  $D$  è partizionato in un numero finito di unità areali. Ad ogni area  $s_i$ , per  $i = 1, \dots, n$ , è associato un valore osservato  $Y(s_i)$ .

Definiamo  $W$  come la matrice di adiacenza  $n \times n$  che descrive come le aree sono disposte e connesse tra di loro. Si ha pertanto che:

$$w_{ij} = \begin{cases} 1 & \text{se le aree } i \text{ e } j \text{ condividono una porzione di perimetro} \\ 0 & \text{altrimenti} \end{cases}$$

La specificazione del modello è data da:

$$\begin{aligned} Y_i | \mu_i, \nu^2 &\stackrel{\text{ind}}{\sim} f(y_i | \mu_i, \nu^2) \\ g(\mu_i) &= \underline{x}_i^T \underline{\beta} + \phi_i \\ \underline{\beta} &\sim N(\underline{b}_0, \Lambda_0) \\ \nu^2 &\sim IG(a_\nu, b_\nu) \\ (\phi_1, \dots, \phi_n) &\sim CAR(W, \rho, \tau^2) \\ \rho &\sim Beta(a_\rho, b_\rho) \\ \tau^2 &\sim IG(a_\tau, b_\tau) \end{aligned}$$

Dove  $i = 1, \dots, n$  e dove, nel caso più generale,  $g(\mu_i)$  è ottenuto come  $g(\mu_i) = \underline{x}_i^T \underline{\beta} + \phi_i + h(O_i)$ .

## 2.4 Bayes Factor

Per confrontare le performance dei modelli, si utilizza il Bayes Factor. Dati due modelli  $M_1$  e  $M_2$ , esso è definito come:

$$BF_{12} = \frac{\frac{P(M_1|y_{1:n})}{P(M_2|y_{1:n})}}{\frac{P(M_1)}{P(M_2)}} = \frac{m_1(y_1, \dots, y_n | \underline{x}_{1,1:n})}{m_2(y_1, \dots, y_n | \underline{x}_{2,1:n})}$$

Dove  $P(M_j)$  è la probabilità a priori associata al modello  $j$ -esimo e  $P(M_j|y_{1:n})$  quella a posteriori. Inoltre:

$$m_j(y_1, \dots, y_n | \underline{x}_{j,1:n}) = \int_{\Theta} \prod_{i=1}^n f(y_i | \underline{x}_{j,i}, \underline{\theta}_j) \pi_j(\underline{\theta}_j) d\underline{\theta}_j \quad j = \{1, 2\}$$

Quando  $BF_{12} > 1$  l'evidenza empirica è a favore di  $M_1$ , altrimenti si preferirà  $M_2$ .

## 3 Risultati

Si sono dunque applicati i metodi descritti nel capitolo precedente al dataset in questione.

### 3.1 Selezione delle covariate

In primo luogo, si è effettuata una procedura di selezione delle covariate. A tal fine, si è selezionata un'osservazione per ogni quartiere e si è considerato il relativo prezzo al metro quadro in scala logaritmica. La scelta di considerare soltanto un tale sottoinsieme dei dati è stata guidata da necessità computazionali, in quanto le variabili prese in considerazione in questa procedura sono numerose e i tempi di calcolo non sarebbero altrimenti risultati gestibili.

Si è dunque utilizzato l'approccio spike-and-slab con la seguente specificazione:

$$\begin{aligned} c_j^2 &= 30^2 = 900 \quad \forall \quad j = 1, \dots, p \\ \kappa_j &= 0.02 \quad \forall \quad j = 1, \dots, p \\ \tau_{V S_j}^2 &= \frac{\kappa_j^2 (c_j^2 - 1)}{2 \log(c_j) c_j^2} = 5.87 \times 10^{-5} \quad \forall \quad j = 1, \dots, p \\ \theta_j &\stackrel{\text{ind}}{\sim} \text{Beta}(5, 2) \quad \forall \quad j = 1, \dots, p \end{aligned}$$

Si scelgono inoltre le seguenti prior per la componente spaziale del modello:

$$\begin{aligned} \sigma^2 &\sim \text{IG}(2.1, 1) \\ \tau^2 &\sim \text{IG}(2.1, 1) \\ \phi &\sim \text{IG}(2.1, 10^{-5}) \end{aligned}$$

Si è quindi costruita una catena MCMC di lunghezza 1500, le cui prime 800 iterazioni sono state scartate come burn-in.

A partire da essa, si è valutata la miglior combinazione delle variabili tramite HPD.

In particolare, il modello visitato più volte dalla catena è risultato essere quello composto dalle seguenti covariate: **intercetta, condition, car parking, elevator, heating centralized**.

Si è dunque deciso di considerare questo sottoinsieme delle covariate per tutte le restanti analisi, sia per quelle relative ai dati point-referenced sia per quelle relative ai dati areali.

### 3.2 Modello di regressione spaziale Bayesiano con covariate

Si è innanzitutto diviso il dataset in training e test. Per farlo, sono state scelte casualmente due osservazioni per quartiere (in alcuni casi una, in quanto ne era presente una soltanto) e sono state inserite nel training set. Tutte le restanti osservazioni sono state inserite nel test set. L'analisi è stata quindi condotta su 271 osservazioni e i risultati prodotti sono stati utilizzati per effettuare stime su 1505 diverse abitazioni.

Si è dunque costruito un modello basato sui dati point-referenced con una specificazione analoga a quella utilizzata per la selezione delle covariate:

$$\begin{aligned}\underline{\beta} &\sim \text{MultiNorm}(\min\{\underline{Y}\}, 0, \dots, 0)^T, 10^6 I_n) \\ \sigma^2 &\sim \text{IG}(2.1, 1) \\ \tau^2 &\sim \text{IG}(2.1, 1) \\ \phi &\sim \text{IG}(2.1, 10^{-5})\end{aligned}$$

Si è inoltre scelto un correlogramma Gaussiano, ovvero del tipo  $\rho(\|\underline{s}_i - \underline{s}_j\|, \phi) = e^{-\phi^2 \|\underline{s}_i - \underline{s}_j\|^2}$ .

Si noti che la scelta di un così piccolo valore di  $b_\phi$  è giustificata dalla necessità di gestire la scala di riferimento delle coordinate delle osservazioni. Poichè è stata utilizzata una georeferenziazione tramite sistema di riferimento Gauss-Boaga, le coordinate per ogni abitazione sono nell'ordine di grandezza dei milioni di metri, ed è stato quindi necessario centrare la Gamma Inversa in valori molto bassi per non ottenere problemi numerici nel calcolo di  $\rho(d, \phi)$ .

Si è dunque costruita una catena MCMC di lunghezza 2000, di cui 1000 iterazioni sono state considerate per il burn-in.

Si sono successivamente calcolate le previsioni puntuali sulle osservazioni inserite nel test set. Per diagnosticare la bontà dei risultati, sono stati calcolati RMSE, MAE e MAPE per ognuna di esse. I valori ottenuti sono riportati in Tabella 2 al termine di questo capitolo, in cui viene effettuato un confronto tra i risultati dei diversi modelli.

Per avere un'idea grafica del funzionamento del modello, si è scelta una casa da prendere come referenza per valutare l'andamento del suo prezzo in una fitta griglia di punti ricoprente l'intera area della città di Milano. In particolare, si è scelta una casa in una condizione buona/abitabile, senza posti auto inclusi, in uno stabile con l'ascensore e con riscaldamento autonomo. L'andamento del suo prezzo al metro quadro è riportato nella Figura 1, con un range che varia da un minimo di 2086 €/m<sup>2</sup> ad un massimo di 9691 €/m<sup>2</sup>. Come era possibile attendersi, si nota un chiaro aumento dei prezzi nella zona centrale della città ed un leggero aumento in zone periferiche ma vicine alle stazioni delle metropolitane (M5 a Nord, M2 a Sud-Ovest).

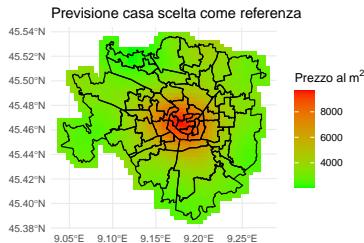


Figura 1: Previsione del prezzo al m<sup>2</sup> della casa scelta come referenza

### 3.3 Modello di regressione spaziale Bayesiano senza covariate

Per valutare quanto la presenza delle covariate sia utile a spiegare la diversità di prezzi, si è deciso di costruire un modello basato solo sulla dipendenza spaziale. Per farlo, si sono considerate delle prior del tutto analoghe a quelle utilizzate nel precedente punto, con un'unica differenza relativa al fatto che il coefficiente  $\beta$  è ora composto da un'unico elemento, per cui si considera una distribuzione  $N(\min\{Y\}, 10^6)$ .

Anche qui si sono poi calcolate le previsioni sulle osservazioni di test e sui punti del precedente grigliato, su cui sono stati ottenuti risultati analoghi.

### 3.4 Modello di regressione Bayesiano

Per completare l'analisi Bayesiana di questi dati, si è deciso di costruire anche un semplice modello di regressione Bayesiano, che non sfrutta quindi la componente spaziale che si ha a disposizione.

Anche in questo caso sono state calcolate le previsioni sul dataset di test, mentre non è stato possibile eseguire l'analisi sul grigliato in quanto si sarebbero ottenuti valori coincidenti (in quanto nel grigliato si considerano le covariate come costanti nei diversi punti, ma qui non è presente una componente spaziale che li possa diversificare).

### 3.5 Kriging

Si è infine deciso di costruire anche un modello proveniente da un framework frequentista per poter effettuare una valutazione più completa dei modelli considerati. In particolare, si è deciso di considerare un Kriging Universale (in quanto è chiara la non stazionarietà nei dati) che sfrutta anche le covariate che abbiamo a disposizione. Questo modello dovrà quindi essere confrontato con il primo dei modelli Bayesiani costruiti.

Anche qui sono state usate le osservazioni del training set per stimare i pesi del Kriging e sono poi state effettuate previsioni sul test set.

### 3.6 Confronto tra modelli

Si è iniziato confrontando i modelli sulla base della loro capacità previsiva. Per valutarla, sono stati calcolati gli indici RMSE, MAE e MAPE per tutte le previsioni effettuate sulle osservazioni del test set. I risultati ottenuti sono riportati in Tabella 2:

Indice	Modelli Bayesiani			
	Spazio+Variabili	Spazio	Variabili	Kriging
RMSE	1440.40	1564.23	2435.83	1411.93
MAE	977.95	1077.70	1771.13	986.00
MAPE	17.09	18.87	30.85	17.25

Tabella 2: Confronto tra gli errori nei diversi modelli

I risultati ottenuti sono del tutto in linea con le aspettative:

- I modelli che includono sia le covariate che la componente spaziale performano molto meglio degli altri. Le differenze tra il modello Bayesiano ed il Kriging sono trascurabili: performano entrambi molto bene.
- I modelli che includono solo una delle due componenti hanno un fitting peggiore. In particolar modo, rimuovendo le informazioni relative alla posizione, si perde molta della capacità previsiva del modello.

Per quanto riguarda i primi tre modelli stimati, si è inoltre utilizzato il Bayes Factor (in scala logaritmica) per fare un confronto a coppie tra di essi. I risultati ottenuti sono riportati in Tabella 3.

Confronto	Log Bayes Factor
Spazio+Var vs SoloSpazio	19.61
Spazio+Var vs SoloVar	145.62
SoloSpazio vs SoloVar	126.01

Tabella 3: Log Bayes Factor a coppie

Valutando tali valori, possiamo affermare con sicurezza che si preferisce in ogni caso il modello con spazio e variabili, mentre tra gli altri due modelli si preferisce quello che considera soltanto la componente spaziale. Questo risultato è in linea con quanto ottenuto in termini di performance previsiva, anche in termini di ordine di grandezza della differenza.

### 3.7 Modello per dati areali

Per avere una visione completa del caso in esame, si è infine deciso di costruire un modello considerando i dati in formato areale. Per farlo, si è aggiunta una variabile di allocazione utile per assegnare ogni osservazione ad uno specifico quartiere di Milano. Si è poi considerato il prezzo medio al metro quadro (sempre in scala logaritmica) per ogni quartiere e si è costruito un modello su tale valore.

In particolare, per produrre una stima da un punto di vista diverso dal precedente, si è deciso di considerare soltanto la componente spaziale e di dividere i dati in quattro categorie sulla base del numero di locali.

Si sono quindi costruiti quattro diversi modelli (rispettivamente per monolocali, bilocali, trilocali e quadri-locali), con la medesima specificazione:

$$\begin{aligned}
 Y_i | \mu_i, \nu^2 &\stackrel{\text{ind}}{\sim} N(\mu_i, \nu^2) \\
 \mu_i &= \underline{x}_i^T \underline{\beta} + \phi_i \\
 \beta &\sim N(0, 10^3) \\
 \nu^2 &\sim IG(2.1, 1) \\
 (\phi_1, \dots, \phi_n) &\sim CAR(W, \rho, \tau^2) \\
 \rho &\sim Beta(1, 1) \\
 \tau^2 &\sim IG(2.1, 1)
 \end{aligned}$$

Sono quindi state ottenute delle stime del prezzo medio al metro quadro per tipologia di appartamento per ogni zona di Milano. I risultati ottenuti sono riportati in Appendice.

Si sono infine confrontati i risultati ottenuti in ognuno dei quattro modelli tramite un boxplot.

Dalla Figura 2 notiamo che la variabilità nelle stime sembra essere più elevata nei monolocali e che il prezzo mediano al metro quadro (seppur leggermente) si abbassa man mano che aumenta il numero di locali, come è ragionevole aspettarsi.

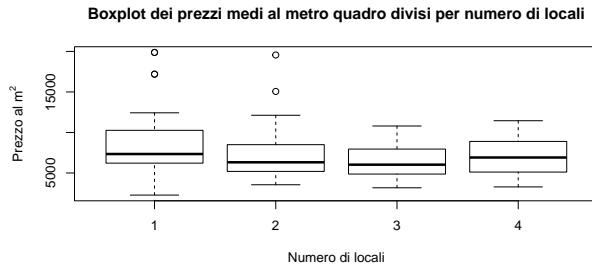


Figura 2: Boxplot divisi per numero di locali

## 4 Conclusioni

L'analisi svolta ha avuto come obiettivo la previsione del prezzo al metro quadro degli immobili residenziali nella città di Milano, mediante modelli di regressione spaziale in ambito Bayesiano. A partire da un dataset aggiornato a Febbraio 2023, contenente variabili strutturali, energetiche e geografiche per ciascun immobile, è stata costruita una variabile dipendente logaritmica, al fine di normalizzare i dati e rendere più efficace la modellazione.

Sono stati stimati e confrontati tre modelli Bayesiani point-referenced: uno con sola componente spaziale, uno con sole covariate e uno che integra entrambe le fonti di informazione. A completamento, è stato implementato anche un modello di Kriging Universale come riferimento frequentista. La selezione delle

covariate è stata effettuata tramite approccio spike-and-slab Bayesiano. I modelli sono stati valutati sulla base delle metriche RMSE, MAE, MAPE e del Bayes Factor.

I risultati mostrano come il modello congiunto, che integra sia le covariate che la struttura spaziale, fornisca le migliori prestazioni predittive, confermando l'importanza della dipendenza spaziale per cogliere l'eterogeneità territoriale del mercato immobiliare. Anche il Kriging ha mostrato performance comparabili, ma l'approccio Bayesiano si distingue per la maggiore flessibilità e per la possibilità di integrare conoscenza a priori.

L'analisi è stata successivamente estesa al contesto areale, costruendo modelli spaziali separati per tipologia di appartamento (in base al numero di locali), con l'obiettivo di fornire una lettura aggregata del mercato immobiliare su base di quartiere. I risultati confermano le tendenze attese, come una maggiore variabilità nei prezzi per i monolocali e un lieve calo del prezzo mediano al metro quadro all'aumentare del numero di stanze.

Nel complesso, il lavoro mostra come i modelli spaziali in un framework Bayesiano possano costituire uno strumento utile per la previsione di dati di questo tipo. In questa analisi si sono utilizzate informazioni a priori molto vaghe, ma i modelli proposti potrebbero essere arricchiti con conoscenze derivanti da esperti del settore; una migliore specificazione delle prior potrebbe infatti raffinare la capacità descrittiva e previsiva degli stessi.

## Appendice

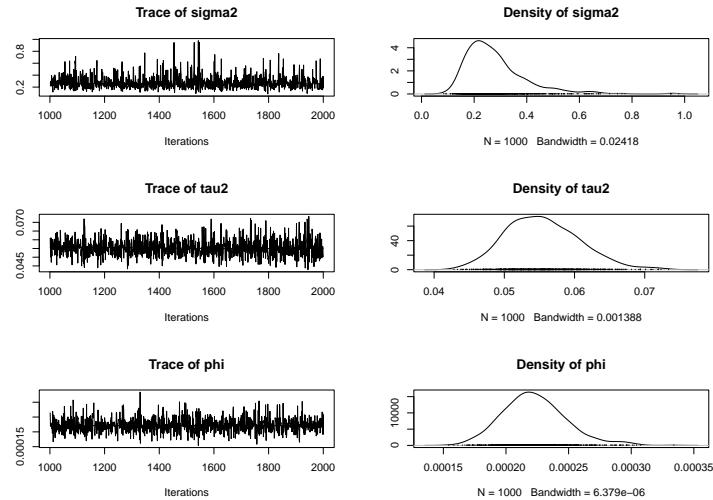


Figura 3: Catene MCMC modello spaziale con covariate

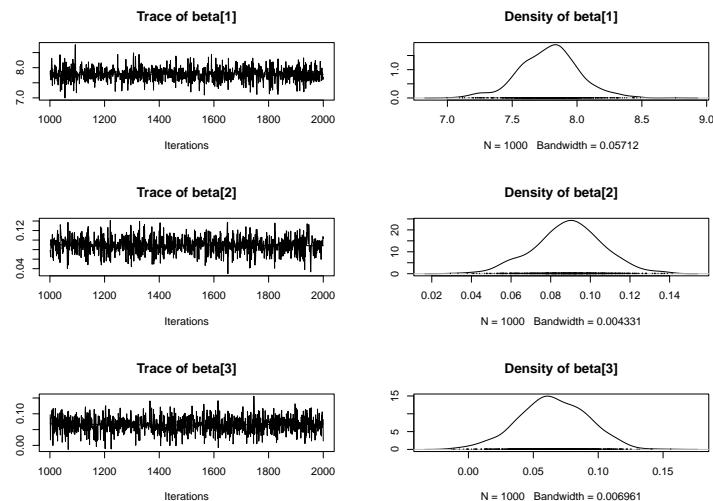


Figura 4: Catene MCMC modello spaziale con covariate

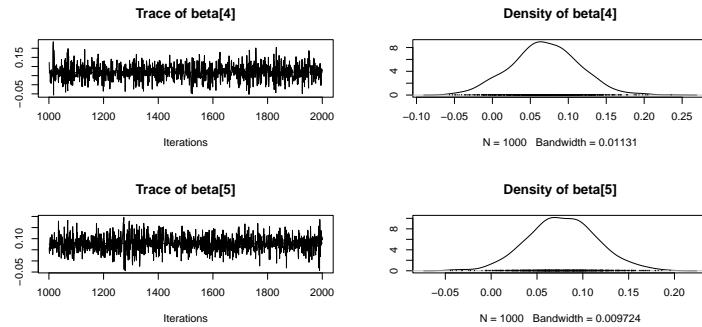


Figura 5: Catene MCMC modello spaziale con covariate

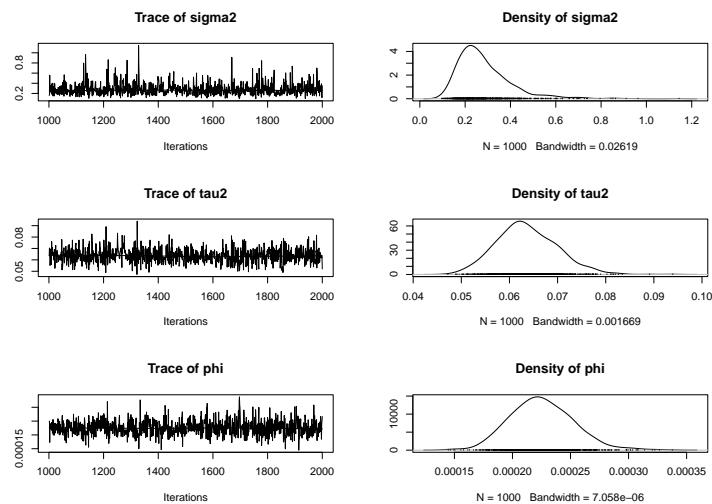


Figura 6: Catene MCMC modello spaziale senza covariate

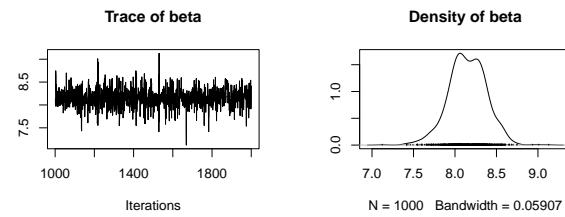


Figura 7: Catene MCMC modello spaziale senza covariate

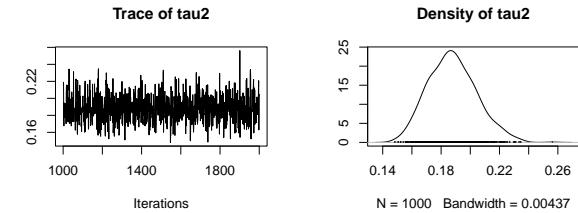


Figura 8: Catene MCMC modello con sole covariate

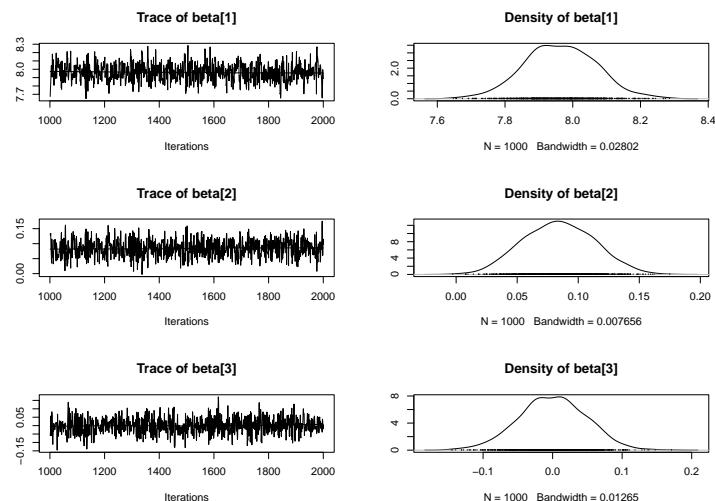


Figura 9: Catene MCMC modello con sole covariate

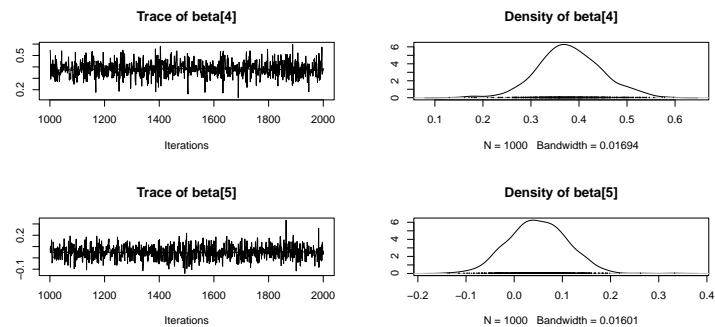
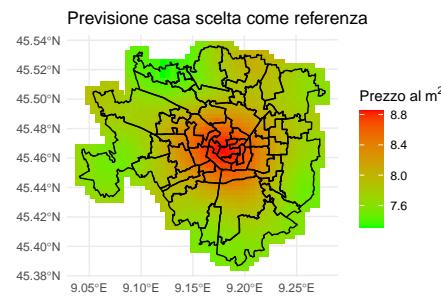
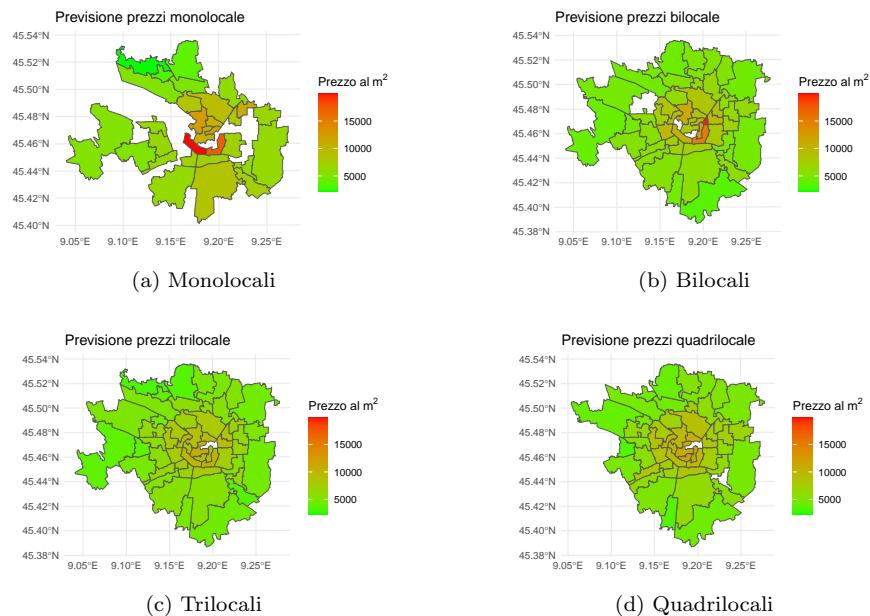


Figura 10: Catene MCMC modello con sole covariate

Modello	intercetta	condition	car parking	elevator	heating centralized
Spazio+Variabili	7.79	0.09	0.07	0.07	0.08
Solo Spazio	8.15	-	-	-	-
Solo Variabili	7.96	0.08	0.00	0.38	0.05

Tabella 4: Coefficienti di regressione stimati nei diversi modelli

Figura 11: Previsione del prezzo al m<sup>2</sup> della casa scelta come referenza nel modello solo spazialeFigura 12: Previsione areale del prezzo al m<sup>2</sup> per diverse tipologie di immobili

Codice	Nome Zona	Monolocale	Bilocale	Trilocale	Quadrilocale
B01	Cadorna, Monti, Boccaccio	12083.6	-	10799.0	11090.5
B02	Castello, Melzi D'Eril, Sarpi	12083.6	9910.1	9619.0	11406.9
B05	Turati, Moscova, Repubblica	10689.0	10096.0	8265.8	11449.5
B06	Venezia, Manin, Montenapoleone	-	12122.0	-	-
B08	Masnaghi, Porta Vittoria, Porta Romana	11861.8	11255.3	8636.8	9737.4
B11	Porta Ticinese, Porta Genova, Magenta	16966.3	19561.5	-	-
B12	Scala, Manzoni, Vittorio Emanuele, San Babila	20096.0	15061.9	9312.2	10514.5
B13	Brera, Duomo, Cordusio, Torino	-	-	10583.7	9803.4
B14	Missori, Italia, Vetra, S. Eufemia	-	10238.3	9936.3	9345.2
C01	Fiera, Diaz, Fontana, Europa	8475.6	6281.2	6996.8	7127.9
C02	Cenisio, Procaccini, Firenze	-	8494.0	8002.3	7964.9
C03	Fiera Milano, Cesare, Sempione	-	7593.0	8286.0	8895.0
C04	Amendola, Monte Rosa, Buonarroti	-	8926.4	9500.1	10233.2
C05	Pagano, Monti, Wagner	-	9793.7	7720.4	8751.5
C06	Piemonte, Washington, Cimarosa	-	7294.7	7654.2	7328.6
C07	Solari, Napoli, Savona	7164.1	7146.2	6855.8	7078.2
C08	Naviglio Grande, Arcole, San Gottardo	7368.9	7430.4	7906.5	8389.2
C09	Tabacchi, Sarfatti, Crema	7420.0	7926.3	7445.2	7497.6
C10	Libia, XXII Marzo, Indipendenza	-	9448.8	9266.6	8467.9
C11	Regina Giovanna, Pisacane, Castel Morrone	8267.8	8594.4	8070.7	7178.7
C12	Abruzzi, Eustachi, Plinio	9212.0	8479.1	7951.2	9148.2
D01	Stazione Centrale, Gioia, Zara	5295.6	4559.3	5132.9	4710.2
D02	Musocco, Varesina, Certosa	4811.3	5277.1	5751.8	4565.8
D03	Bovisa, Dergano, Imbonati	3315.4	4789.6	3912.7	4595.2
D04	Largo Boccioni, Aldini, Lopez	3753.4	4322.3	3454.8	4140.1
D05	Bovisasca, Affori, P. Rossi	-	5049.6	4841.2	5302.0
D06	Niguarda, Ornato	-	5194.1	4867.6	5112.8
D07	Fulvio Testi, Bicocca, Ca' Granda	-	5351.5	5394.4	5264.0
D08	Monza, Precotto, Gorla	-	6504.9	5889.9	7467.3
D09	Zara, Istria, Marche	6273.1	6695.1	5537.5	6043.7
D10	Loreto, Turro, Padova	-	4463.9	4814.4	4919.7
D11	P.Lambro, Feltre, Udine	11240.8	5881.8	6915.7	6909.1
D12	Aspromonte, Porpora, Teodosio	-	6319.6	6237.6	7491.3
D13	Leonardo Da Vinci, Gorini	6694.8	6362.6	4969.8	6860.6
D14	Lambrate, Rubattino, Folli	-	6947.8	6786.9	6175.6
D15	Argonne, Viale Corsica	6495.7	4710.2	4437.0	4834.7
D16	Forlanini, Mecenate, Pontenasco	-	4947.4	5541.5	6079.4
D17	Ortomerato, Molise, Piacenza	6136.6	4944.6	6585.5	-
D18	Boncompagni, Toffetti, Bacchiglione	7457.6	5811.4	3449.2	5589.6
D19	Omero, Gabriele Rosa, Brenta	-	3546.8	4430.0	4380.8
D20	Ronchetti, Chiaravalle, Ripamonti	8711.1	5770.2	5498.8	6192.6
D21	Barona, Famagosta, Faenza	6265.2	5437.6	5622.9	5074.4
D22	San Cristoforo, Ronchetto, Ludovico Il Moro	-	5653.5	5326.2	6150.6
D23	Giambellino, Farini, Frattini	6541.8	5815.2	6560.8	8112.0
D24	San Siro, Tripoli, Brasilia	6530.7	5693.5	5917.4	6000.0
D25	Siena, Innocenti, Bisceglie	5332.3	5431.9	4748.8	4753.1
D26	Lorentaggio, Inganni, Bisceglie	-	4532.3	3714.3	3281.1
D27	Novara, San Carlo, Amati	6482.6	5714.9	6025.6	5631.4
D28	Segesta, Capece Latro, Areusa	-	-	6315.3	6240.3
D29	Ippodromo, Caprilli, Monte Stella	-	-	4305.5	4927.2
E01	Raggio Q., Romano, Cagnola	5491.2	4140.5	3764.0	-
E02	Gallaratese, Lampugnano, Figino	-	4714.4	5037.0	4060.7
E03	Missaglia, Chiesa Rossa, Gratosoglio	-	4581.6	4836.5	3370.2
E04	Quarto Oggiaro, Roserio, Amoretti	2298.2	-	3174.0	-

Tabella 5: Previsione del prezzo al m<sup>2</sup> divisa per zona