

# Previsione del prezzo mediano delle abitazioni tramite tecniche bayesiane

Beatrice Malvezzi, Adalgisa Perrelli, Francesca Reina, Nicole Santero

9 Settembre 2025

## Abstract

---

Il presente lavoro affronta il problema della previsione del prezzo mediano delle abitazioni in California attraverso un approccio integrato che combina metodi statistici bayesiani con tecniche di analisi spaziale.

---

## 1 Introduzione e obiettivi

Il dataset considerato fa riferimento ai dati del censimento della California del 1990: ogni osservazione corrisponde a un “block group”, ovvero la più piccola unità geografica per cui l’Ufficio del Censimento degli Stati Uniti pubblica dati campione (un “block group” ha in genere una popolazione compresa tra 600 e 3000 abitanti). D’ora in avanti, per semplicità, si farà riferimento a questi “block group” con il termine “distretti”.

L’obiettivo del seguente lavoro è prevedere il prezzo mediano delle abitazioni per distretto: adottando un approccio bayesiano, verranno confrontate diverse specificazioni della prior per ottenere una posterior che fornisca il miglior modello di regressione lineare bayesiana. Successivamente, tramite l’implementazione di un modello di regressione lineare bayesiana a effetti misti si valuterà l’impatto del grado di vicinanza all’oceano sull’intercetta e sui coefficienti di regressione. Infine, mediante le tecniche di modellazione spaziale bayesiana si prevederà il prezzo mediano per un campione di abitazioni.

## 2 Descrizione del dataset

Il dataset *California Housing Prices* (Nugent, 2017) contiene 20640 osservazioni, ciascuna delle quali relativa a un distretto abitativo della California. Non sono presenti valori duplicati bensì valori mancanti che verranno trattati nel paragrafo seguente.

Le variabili elencate nella Tabella 3 descrivono le caratteristiche dell’abitazione e la sua posizione geografica.

## 3 Analisi esplorativa, pre-processing e model selection

Il dataset contiene missing values esclusivamente nella variabile `total_bedrooms`: vista la numerosità del dataset, si è deciso di eliminare le osservazioni incomplete senza procedere con l’imputazione.

È stata realizzata una mappa della California dove i punti presentano grandezza variabile a seconda della numerosità della popolazione per il distretto indicato, variabile `population`, e colore determinato dal prezzo mediano delle abitazioni nello stesso, variabile `median_house_value`. Punti di dimensioni maggiori corrispondono ad aree più densamente popolate e la scala cromatica passa da colori freddi a caldi all'aumentare del valore della variabile target (Figura 1).

Successivamente, è stata effettuata un'analisi esplorativa sui dati prendendo in considerazione le statistiche descrittive e il grafico delle distribuzioni univariate (Figura 2).

Le variabili `households`, `population`, `total_bedrooms` e `total_rooms` presentano una distribuzione fortemente asimmetrica con una coda lunga verso destra, ad indicare una possibile presenza di outliers. I boxplot realizzati hanno confermato questa ipotesi, rivelando la presenza di alcuni outliers per la variabile `total_rooms` poi rimossi dal dataset.

Dopo aver trasformato la variabile `ocean_proximity` tramite *one-hot-encoding*, dai boxplot realizzati appare evidente la relazione positiva esistente tra la variabile target e la prossimità alla costa (Figura 3).

Per eseguire la selezione delle variabili è stato adottato l'approccio *spike and slab*, una metodologia bayesiana che combina regolarizzazione e inferenza sui modelli. L'idea di base consiste nell'assegnare a ciascun coefficiente di regressione  $\beta_j$  una distribuzione a priori mista del tipo:

$$\beta_j \sim (1 - \pi_j) \delta_0 + \pi_j \mathcal{N}(0, \sigma^2), \quad (1)$$

dove  $\delta_0$  rappresenta lo *spike*, fortemente concentrato attorno a zero e associato all'ipotesi di irrilevanza della variabile, mentre  $\mathcal{N}(0, \sigma^2)$  costituisce lo *slab*, più diffuso e corrispondente all'ipotesi di rilevanza. Durante l'inferenza, tipicamente tramite campionamento MCMC, si calcolano le *posterior inclusion probabilities* per ciascuna covariata, che quantificano la probabilità di appartenenza allo *slab*. In questo modo la selezione non si limita a identificare un unico modello “ottimale”, ma fornisce anche una misura dell'incertezza associata all'inclusione delle variabili, favorendo modelli parsimoniosi e robusti. Dai risultati ottenuti è emerso che tre delle variabili dummy associate alla covariata `ocean_proximity` presentano una probabilità di inclusione prossima a zero. Per questo motivo, le modalità ‘ISLAND’, ‘NEAR BAY’ e ‘NEAR OCEAN’ sono state accorpate in un'unica categoria denominata ‘ON THE COAST’. Tale operazione ha inoltre contribuito a bilanciare la numerosità delle categorie risultanti.

Il correlogramma (Figura 4) ha evidenziato l'alta correlazione tra le variabili `households`, `population`, `total_rooms` e `total_bedrooms`.

Per identificare quale di queste covariate escludere, è stata applicata nuovamente la procedura di selezione basata sull'algoritmo *Random\_Forest*, che ha portato all'eliminazione della variabile `households`. Successivamente, una verifica della correlazione tra le covariate rimanenti ha suggerito di escludere anche `total_bedrooms`.

Per catturare l'informazione spaziale presente nel dataset, è stata creata la covariata `diag_coord` come somma delle coordinate geografiche `latitude` e `longitude`. Questa trasformazione ha lo scopo di rappresentare la posizione dei quartieri lungo la diagonale nord-ovest – sud-est dello stato, permettendo di sintetizzare in un'unica variabile l'effetto congiunto della latitudine e della longitudine. L'introduzione di `diag_coord` ha portato a un incremento significativo della correlazione con la variabile risposta `median_house_value`, suggerendo che la posizione spaziale combinata ha un ruolo rilevante nella determinazione dei valori abitativi.

La presenza di variabili con una distribuzione fortemente skewed, già evidenziata in precedenza, ha condotto all'applicazione di una tecnica di scaling: le covariate `housing_median_age`, `total_rooms`, `population`, `median_income` e `diag_coord` sono state standardizzate, mentre la variabile risposta è stata dapprima trasformata su scala logaritmica e successivamente standardizzata. In questo modo, per la variabile `median_house_value` si è giunti a una distribuzione approssimativamente Normale (Figura 5).

## 4 Regressione lineare bayesiana

La caratteristica essenziale dei metodi bayesiani è l'uso esplicito della probabilità per quantificare l'incertezza nelle inferenze basate sull'analisi statistica dei dati.

Per valutare l'effetto della scelta delle prior nel modello di regressione bayesiana, è buona pratica suddividere il dataset in due insiemi distinti: un train set, su cui si allena il modello e si esegue inferenza valutando la robustezza della stima, e un test set indipendente su cui si valuta la capacità predittiva e la bontà di adattamento del modello a dati non osservati in fase di stima.

I due elementi fondamentali del modello di regressione lineare bayesiana sono il modello probabilistico che descrive la distribuzione dei dati condizionatamente ai parametri e un'opportuna scelta della prior: non avendo informazioni pregresse sui dati, inizialmente verrà usata una prior non informativa e successivamente verranno proposte due specificazioni più informative.

### 4.1 Prior non informativa

Si consideri il seguente modello probabilistico:

$$Y_i | \mathbf{x}_i, \boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2) \quad \text{con } i = 1, \dots, n \quad (2)$$

dove:

- $Y_1, \dots, Y_n$  con  $Y_i \in \mathcal{Y} \subseteq \mathbb{R}$  è l'insieme delle variabili risposta;
- $\mathbf{x}_1, \dots, \mathbf{x}_n$  con  $\mathbf{x}_i \in \mathcal{X} \subseteq \mathbb{R}^p$  è l'insieme delle covariate;
- $\boldsymbol{\beta} \in \mathbb{R}^p$  è il vettore dei coefficienti di regressione;
- $\sigma^2$  è la varianza degli errori.

Il primo approccio consiste nell'adottare una prior non informativa impropria (vague prior) che lascia che siano le osservazioni a guidare l'inferenza:

$$\pi(\boldsymbol{\beta}, \sigma^2) \propto \sigma^{-2} \quad (3)$$

Si è dimostrato che con questa prior, la posterior congiunta fattorizza nel prodotto della posterior di  $\boldsymbol{\beta}$  condizionata a  $\sigma^2$  e nella posterior di  $\sigma^2$  che hanno entrambe distribuzione nota da cui è possibile campionare.

$$\boldsymbol{\beta} | \sigma^2, \mathbf{y}_{1:n}, \mathbf{x}_{1:n} \sim \mathcal{N}\left(\hat{\boldsymbol{\beta}}_{ML}, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}\right) \quad (4)$$

$$\sigma^2 | \mathbf{y}_{1:n}, \mathbf{x}_{1:n} \sim \mathcal{IG}\left(\frac{n-p}{2}, \frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{ML})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{ML})\right) \quad (5)$$

L'algoritmo implementato è quindi un collapsed Gibbs sampler: a differenza del Gibbs standard che prevede di aggiornare iterativamente ogni parametro campionando dalla sua distribuzione condizionata data l'attuale valore degli altri parametri, in questo caso il parametro  $\beta$  viene marginalizzato nel passo di aggiornamento di  $\sigma^2$  campionando quest'ultima da una distribuzione Inverse Gamma basata sui residui calcolati dalla stima ML di  $\beta$  e infine campiona  $\beta$  dalla sua distribuzione normale multivariata condizionata sull'ultimo  $\sigma^2$ .

## 4.2 Unit Information Prior

Il passaggio successivo consiste nell'adottare una prior poco informativa mantenendo il modello (1), la cosiddetta unit information prior: questa prior apporta una leggera informazione aggiuntiva e in questo contesto è “equivalente” ad un piccolo gruppo di 5 osservazioni. Così facendo viene garantita l'esistenza finita di media e varianza per il parametro  $\sigma^2$ . In particolare:

$$\boldsymbol{\beta} | \sigma^2 \sim \mathcal{N}\left(\hat{\boldsymbol{\beta}}_{OLS}, s_{OLS}^2 (\mathbf{X}^T \mathbf{X})^{-1}\right) \quad (6)$$

$$\sigma^2 \sim \mathcal{IG}\left(\frac{\nu_0}{2}, \frac{\nu_0 * s_{OLS}^2}{2}\right) \quad \text{con } \nu_0 = 5 \quad (7)$$

L'analisi MCMC è stata effettuata via STAN, con 4 catene da 20000 iterazioni e 10000 di burn-in.

## 4.3 Prior poco informativa con shrinkage

Nel modello di regressione lineare bayesiana con shrinkage gerarchico la variabile risposta segue la distribuzione riportata nel modello (1).

Per i coefficienti  $\beta_j$  è stata scelta una prior normale condizionata:

$$\beta_j | \sigma^2, \lambda \sim \mathcal{N}\left(\beta_{0j}, \frac{\sigma^2}{\lambda}\right), \quad (8)$$

con  $\beta_{0j} = 0$  e  $\lambda \sim \text{Gamma}(1, 1)$ . Questa formulazione introduce shrinkage, spingendo i coefficienti verso zero in modo analogo alla regressione Ridge, ma con intensità di penalizzazione stimata dai dati tramite  $\lambda$ .

La varianza degli errori  $\sigma^2$  è stata modellata con una prior Inverse-Gamma:

$$\sigma^2 \sim \mathcal{IG}(a_0, b_0), \quad (9)$$

con  $a_0 = 12$  e  $b_0 = 11$ . Questi valori sono stati scelti in modo tale da garantire che il valore atteso e la varianza di  $\sigma^2$  siano definiti e coerenti con la varianza empirica della variabile risposta.

La stima bayesiana è stata effettuata tramite campionamento MCMC con Stan, usando 4 catene, 5000 iterazioni di cui 3000 di warm-up.

## 4.4 Mixed Model

I modelli misti rappresentano una naturale estensione dei modelli di regressione lineare e dei modelli lineari generalizzati, particolarmente utili quando i dati presentano una struttura gerarchica o di raggruppamento. In questo contesto, oltre alle covariate esplicative

standard (effetti fissi), è possibile introdurre degli effetti casuali (random effects) che permettono di modellare la variabilità specifica dei gruppi. Ad esempio, nel dataset *California Housing*, le osservazioni possono essere suddivise in base alla variabile `ocean_proximity`, e risulta ragionevole supporre che l'impatto di alcune covariate (quali, ad esempio, l'età media delle abitazioni o il reddito mediano), così come l'intercetta, possa variare tra i diversi gruppi geografici.

Formalmente, un modello lineare misto può essere scritto come

$$y_i = x_i^\top \beta + u_i^\top \gamma_{c_i} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2), \quad (10)$$

dove  $x_i$  e  $u_i$  sono rispettivamente i vettori delle covariate con effetti fissi e random,  $\beta$  è il vettore dei coefficienti fissi,  $\gamma_{c_i}$  rappresenta i coefficienti (inclusa l'intercetta) specifici del gruppo  $c_i$  e  $\sigma^2$  è la varianza residua.

Nel nostro caso, le distribuzioni a priori sono state specificate come

$$\beta \sim \mathcal{N}(\mathbf{0}, I_p), \quad \gamma_j \sim \mathcal{N}(\mathbf{0}, I_q), \quad \sigma^2 \sim \mathcal{IG}(3, 2), \quad (11)$$

dove  $I_p$  e  $I_q$  sono matrici identità di dimensione rispettivamente pari al numero di covariate con effetti fissi e random. A posteriori, grazie alla scelta di distribuzioni a priori coniugate, le distribuzioni condizionali complete dei parametri appartengono a famiglie note (Normale e Inverse-Gamma), con parametri aggiornati che possono essere calcolati esplicitamente. In particolare, nel caso del modello misto lineare con errori gaussiani, si ha

$$\beta \mid y, X, U, c, \gamma, \sigma^2 \sim \mathcal{N}(b_n, \Sigma_n), \quad \gamma_j \mid y, X, U, c, \beta, \sigma^2 \sim \mathcal{N}(\tau_n, \Lambda_n), \quad (12)$$

$$\sigma^2 \mid y, X, U, c, \beta, \gamma \sim \mathcal{IG}(a_n, b_n), \quad (13)$$

con i parametri aggiornati dati da

$$\begin{aligned} \Sigma_n &= \left( \Sigma_0^{-1} + \frac{1}{\sigma^2} X^\top X \right)^{-1}, & b_n &= \Sigma_n \left( \Sigma_0^{-1} b_0 + \frac{1}{\sigma^2} X^\top (y - U\gamma) \right), \\ \Lambda_n &= \left( \Psi_0^{-1} + \frac{1}{\sigma^2} U_j^\top U_j \right)^{-1}, & \tau_n &= \Lambda_n \left( \Psi_0^{-1} \tau_0 + \frac{1}{\sigma^2} U_j^\top (y_j - X_j \beta) \right), \\ a_n &= a_0 + \frac{n}{2}, & b_n &= b_0 + \frac{1}{2} \sum_{i=1}^n (y_i - x_i^\top \beta - u_i^\top \gamma_{c_i})^2. \end{aligned}$$

Questa formulazione rende immediata l'implementazione di uno schema di Gibbs sampling, in cui si aggiornano iterativamente i parametri campionando dalle loro distribuzioni condizionali complete. Per ridurre l'autocorrelazione tra campioni successivi, è stato inoltre effettuato *thinning*, mantenendo un campione ogni 20 generati. In tal modo è possibile stimare congiuntamente effetti fissi e casuali, ottenendo non solo stime puntuali ma anche l'intera distribuzione a posteriori di ciascun parametro.

## 4.5 BART

Il modello Bayesian Additive Regression Trees (BART) è un approccio bayesiano non parametrico molto flessibile che esprime la variabile risposta come:

$$Y_i = f(x_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2), \quad (14)$$

dove la funzione sconosciuta  $f(x)$  è rappresentata come somma di  $m$  alberi di regressione:

$$f(x) = \sum_{j=1}^m g(x; T_j, M_j),$$

con  $T_j$  che descrive la struttura dell'albero  $j$  e  $M_j$  i valori stimati ai nodi terminali.

Le prior utilizzate sono quelle proposte da Chipman et al. (2010) e implementate di default nella funzione `wbart()`:

$$p(\text{split at depth } d) = \alpha(1 + d)^{-\beta}, \quad \alpha = 0.95, \quad \beta = 2, \quad (15)$$

che penalizza alberi troppo profondi;

$$\mu_{jk} | T_j \sim \mathcal{N}(0, \sigma_\mu^2), \quad \sigma_\mu = \frac{0.5}{k\sqrt{m}}, \quad (16)$$

per regolarizzare i valori ai nodi terminali;

$$\sigma^2 \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu\lambda}{2}\right), \quad (17)$$

che assegna una distribuzione plausibile alla varianza degli errori.

L'inferenza si basa su un algoritmo MCMC di Bayesian backfitting: a ogni iterazione, gli alberi vengono aggiornati uno alla volta calcolando i residui rispetto agli altri; la struttura di ciascun albero è aggiornata con Metropolis-Hastings, i valori ai nodi terminali sono campionati via Gibbs sampling, e  $\sigma^2$  è aggiornato da una prior Inverse-Gamma.

## 5 Criteri per la valutazione della bontà del modello

### 5.1 Interpretazioni risultati MCMC

Vengono ora introdotte delle premesse che saranno valide per qualsiasi tipo di analisi affrontata in questo lavoro.

Quando si allena un modello sul train set, si considera buono un modello se sono valide le seguenti condizioni:

- ISTOGRAMMA: la distribuzione a posteriori dei parametri deve risultare simmetrica, unimodale e non presentare code pesanti; invece, a posteriori la varianza dell'errore,  $\sigma^2$ , deve presentare andamento tipico di una Inverse-Gamma;
- TRACEPLOT: si ha convergenza se la catena oscilla rapidamente intorno a un valore centrale ben definito assumendo la tipica forma “a tubo” e copre tutto il dominio plausibile dei parametri;
- NAIVE SE e TIME SERIES SE: se sono molto simili e in particolare hanno stesso ordine di grandezza, il campione si comporta come indipendente dalla distribuzione target;
- EFFECTIVE SIZE: più è vicino all'ampiezza campionata, migliore è l'indipendenza del campione e maggiore è l'efficienza del campionamento;
- GEWEKE: se il valore è compreso nell'intervallo [-2,2] allora la catena è ergodica cioè ha raggiunto la condizione di stazionarietà.

## 5.2 Confronto tra modelli sul train set

Sulla base delle premesse appena riportate e dei grafici presentati in Appendice (Sezioni Prior non informativa, Prior poco informativa: UIP e Prior poco informativa con shrinkage), si può concludere che i modelli con prior non informativa, UIP e con shrinkage portano a risultati soddisfacenti e molto simili in termini di convergenza. Per quanto concerne il Mixed Model (Sezione Mixed Model in Appendice) l'analisi dei risultati ottenuti mostra che tra le covariate considerate, la posizione geografica si conferma come il predittore fisso più robusto, con un effetto significativo e stabile, mentre il reddito mediano evidenzia un impatto positivo più marcato soprattutto nelle aree costiere, suggerendo un'interazione importante tra localizzazione e caratteristiche socio-economiche. Inoltre, la presenza di eterogeneità tra i gruppi giustifica l'utilizzo di un modello lineare misto con random slopes, che consente di catturare meglio le differenze strutturali tra zone costiere, interne e intermedie. Per quanto riguarda la diagnostica MCMC, le catene mostrano buona convergenza complessiva, con risultati particolarmente stabili per la varianza residua, sebbene per alcuni coefficienti sia consigliabile un incremento del burn-in e del thinning per migliorare il mescolamento.

## 5.3 Confronto previsioni sul test set

Modello	MAE	RMSE	$R^2$
Non informativo	0.453277	0.595022	0.648715
Prior UIP	0.453281	0.595018	0.648719
Prior con shrinkage	0.453298	0.595037	0.648698
Mixed Model	0.4396461	0.5814272	0.664584
BART	0.361066	0.493247	0.758608

Tabella 1: Confronto delle prestazioni predittive dei modelli

Dalle metriche riportate nella Tabella soprastante possiamo notare come per i primi tre modelli la performance non vari in modo significativo. Tale risultato porta a concludere che l'impatto delle prior scelte sulla posterior è scarso in quanto la verosimiglianza domina, vista l'elevata ampiezza del campione simulato. Inoltre, l'aggiunta di effetti casuali all'interno del Mixed Model comporta un leggero miglioramento della prestazione, riuscendo a catturare le differenze nei valori delle covariate rispetto al gruppo geografico di appartenenza delle osservazioni. In ultimo, il modello BART, modellizzando una relazione non lineare tra la variabile target e le altre, risulta essere il migliore tra i cinque.

Nella Figura 24 sono riportati i grafici dei valori osservati contro quelli predetti per ciascun modello, con intervalli di credibilità al 95%. In tutti i casi le previsioni seguono l'andamento della bisettrice. Emerge nuovamente la superiorità del modello BART, che presenta intervalli di credibilità più stretti e una minore dispersione rispetto agli altri.

È stata realizzata infine la mappa della California (Figura 23) che mostra la distribuzione geografica delle previsioni ottenute con il modello BART. Si osserva il pattern atteso: prezzi mediani delle abitazioni più elevati lungo la costa e nelle aree metropolitane, e valori più contenuti nelle zone interne.

## 6 Confronto tra modelli spaziali: regressione bayesiana e Kriging

Mentre le analisi precedenti hanno fornito un quadro complessivo attraverso approcci di prior non informative e debolmente informative, Mixed Model e BART applicati all'intero dataset, la complessità computazionale richiede un approccio complementare per quanto riguarda l'analisi spaziale.

Nella seguente sezione viene quindi considerato un campione stratificato proporzionale rappresentativo del dataset, composto da 1015 osservazioni (Figura 25). Il campionamento è stato effettuato sulla base della suddivisione in contee fornita dalla libreria `tigris` di R. Nonostante tale suddivisione sia riferita all'anno 2024, risulta coerente con quella del 1990, anno in cui è stato svolto il censimento. Questo procedimento consente di rendere computazionalmente fattibile l'analisi spaziale pur mantenendo la rappresentazione geografica.

Il campione stratificato è stato successivamente suddiviso in training e test set in proporzione 70/30 per poter valutare le prestazioni predittive dei modelli.

L'analisi si articola in due diverse fasi: modellazione bayesiana spaziale su dati puntuali e Kriging.

### 6.1 Specificazione dei modelli bayesiani spaziali

Innanzitutto è necessario definire un modello per dati puntuali bayesiano che tenga conto della dipendenza spaziale. I dati spaziali di tipo puntuale sono realizzazioni della variabile aleatoria  $Y(\mathbf{s})$ , dove  $\mathbf{s} \in D \subset \mathbb{R}^2$  denota una generica posizione spaziale all'interno del dominio di interesse  $D$ .

Il modello di regressione implementato è:

$$Y(\mathbf{s}) = \mathbf{x}^T(\mathbf{s})\boldsymbol{\beta} + \omega(\mathbf{s}) + \epsilon(\mathbf{s}) \quad (18)$$

Dove:

- $\omega(\mathbf{s})$  cattura i residui dell'associazione spaziale non spiegata dalle covariate, con distribuzione indicizzata dai parametri  $\sigma^2$  (varianza del processo) e  $\phi$  (parametro di correlazione spaziale);
- $\epsilon(\mathbf{s})$  rappresenta la variabilità locale a distanze minori della distanza minima tra i punti osservati, con distribuzione indicizzata da  $\tau^2$  (effetto nugget).

Siano  $X$  una matrice  $n \times p$  la cui  $i$ -esima riga è  $\mathbf{x}^T(\mathbf{s}_i)$  e  $\boldsymbol{\omega} = (\omega(\mathbf{s}_1), \dots, \omega(\mathbf{s}_n))^T$ . La specificazione del modello è completa integrando il random effect  $\boldsymbol{\omega}$ , ottenendo il modello marginale:

$$\mathbf{Y} | \boldsymbol{\beta}, \tau^2 \sim \mathcal{N}(X\boldsymbol{\beta}, \sigma^2 H(\phi) + \tau^2 I_n) \quad (19)$$

$$\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{b}_0, \boldsymbol{\Lambda}_0) \quad \text{dove} \quad \boldsymbol{\beta}_0 = \mathbf{0}_p, \boldsymbol{\Lambda}_0 = 4\mathbf{I}_p \quad (20)$$

$[H(\phi)]_{ij} = \rho(\|\mathbf{s}_i - \mathbf{s}_j\|, \phi)$  definisce la matrice di correlazione spaziale basata sulla distanza euclidea tra le posizioni.

Per completare la specificazione bayesiana, si assegnano prior inverse-gamma ai parametri di varianza  $\boldsymbol{\theta} = (\sigma^2, \tau^2, \phi)$ :

$$\sigma^2 \sim \mathcal{IG}(a_\sigma, b_\sigma) \quad \text{dove } a_\sigma = 3, b_\sigma = 1.5 \quad (21)$$

$$\tau^2 \sim \mathcal{IG}(a_\tau, b_\tau) \quad \text{dove } a_\tau = 3, b_\tau = 1.5 \quad (22)$$

$$\phi \sim \mathcal{IG}(a_\phi, b_\phi) \quad \text{dove } a_\phi = 3, b_\phi = 2 \quad (23)$$

Sono tutte prior debolmente informative che consentono ai dati di guidare l'inferenza senza porre vincoli troppo restrittivi sui parametri, mantenendo allo stesso tempo la stabilità numerica del processo di stima.

La previsione spaziale in una nuova posizione  $s_0$  dato un vettore di predittori  $\mathbf{x}_0 = \mathbf{x}(s_0)$  si ottiene generando una distribuzione normale con momenti

$$E[Y_0|\mathbf{y}] = \mathbf{x}_0^T \boldsymbol{\beta} + \boldsymbol{\gamma}^T (\sigma^2 H(\phi) + \tau^2 I_n)^{-1} (\mathbf{y} - X\boldsymbol{\beta}) \quad (24)$$

$$\text{Var}(Y_0|\mathbf{y}) = \sigma^2 + \tau^2 - \boldsymbol{\gamma}^T (\sigma^2 H(\phi) + \tau^2 I_n)^{-1} \boldsymbol{\gamma} \quad (25)$$

Per quantificare la rilevanza della struttura spaziale è stato implementato anche un modello di riferimento che non considera la dipendenza spaziale e mantiene soltanto la componente di regressione lineare. Di seguito questo modello verrà chiamato *ridotto* per distinguerlo dal *modello completo* che include anche la componente spaziale.

## 6.2 Metriche di valutazione della performance

Per confrontare i due modelli bayesiani si è utilizzato il *Bayes Factor*. Dati due modelli  $M_1$  e  $M_2$ , per convenzione si dice che nei casi in cui  $BF_{12} > 1$  oppure  $\log BF_{12} > 10$  c'è un'evidenza fortemente positiva a favore del modello  $M_1$ .

Per estendere il confronto in termini predittivi anche ai modelli di Kriging sono state calcolate le metriche MAE (Mean Absolute Error), RMSE (Root Mean Square Error) e  $R^2$  confrontando i valori previsti con quelli osservati nel test set.

## 6.3 Discussione e risultati

Il *modello bayesiano completo* presenta parametri spaziali con le stime riportate nella Tabella 33. Il valore elevato di  $\phi$  suggerisce una correlazione spaziale che decresce in modo rapido con la distanza, indicando un'influenza spaziale localizzata. Il rapporto tra  $\sigma^2$  e  $\tau^2$ , pari a 1.2, suggerisce che la componente di rumore ha quasi lo stesso impatto del segnale spaziale strutturato.

Per quanto riguarda i coefficienti delle covariate, si osservano delle differenze rispetto al *modello ridotto* (Tabella 35), in particolare il coefficiente di `median_income` risulta inferiore, suggerendo che la componente spaziale "assorbe" parte dell'effetto di questa componente. I coefficienti del modello con le sole covariate mostrano valori coerenti con le aspettative economiche, con le variabili dummy `ocean_lt1h` e `ocean_inland` che mostrano effetti negativi significativi sui prezzi.

Le statistiche di Geweke per entrambi i modelli (Tabelle 34 e 36) sono comprese nell'intervallo  $[-2, +2]$  indicando la convergenza delle catene MCMC (Figure in Appendice nella sezione Analisi Spaziale). La differenza tra Naive SE e Time-Series SE è generalmente minima per entrambi i modelli, mostrando un'autocorrelazione limitata nelle catene MCMC e quindi una buona efficienza del campionamento.

La Tabella 32 riporta il confronto formale tra i due modelli bayesiani tramite il **log Bayes Factor**. Il valore 113.4799 esprime un'evidenza molto forte a favore del *modello ridotto*, contro l'inclusione della componente spaziale, indicando che la complessità aggiunta

non è giustificata dai dati osservati. La superiorità del *modello ridotto* si può attribuire alla capacità delle variabili socio-economiche di catturare la variabilità spaziale. Inoltre, alcune variabili come le dummies di `ocean_proximity`, `median_income` e `diag_coord` potrebbero incorporare informazioni spaziali significative, rendendo la componente spaziale esplicita ridondante. Infine, il campionamento stratificato del dataset potrebbe aver ridotto artificialmente l'autocorrelazione spaziale naturale.

Per completare l'analisi e validare i risultati ottenuti, sono stati implementati anche due approcci di Kriging: Ordinario e Universale.

La seguente tabella riporta le metriche di valutazione per tutti e quattro i modelli considerati:

Modello	MAE	RMSE	$R^2$
Bayesiano completo	1.0691	1.7129	0.4134
Bayesiano ridotto	0.5291	0.6745	0.6127
Kriging Ordinario	0.5547	0.7572	0.5149
Kriging Universale	0.5653	0.7896	0.4782

Tabella 2: Confronto delle prestazioni predittive dei modelli

I risultati confermano la superiorità del modello bayesiano con le sole covariate con i valori migliori di MAE, RMSE e  $R^2$ . Il Kriging Ordinario risulta comunque un buon modello alternativo, superando sia il modello bayesiano completo, sia il Kriging Universale.

Il *modello ridotto* supera il Kriging Ordinario in quanto ques'ultimo modella solo la struttura spaziale di `median_house_value` ignorando le covariate, mentre le variabili socio-economiche si rivelano predittori più forti della sola dipendenza spaziale. Il Kriging ordinario ha una performance superiore a quella *modello bayesiano completo*, confermando la presenza di struttura spaziale nei dati, ma evidenziando anche problemi di identificabilità.

La performance del Kriging Universale risulta solo leggermente peggiore rispetto a quella del Kriging Ordinario confermando l'ipotesi che le covariate spiegano già efficacemente la variabilità spaziale e che la loro combinazione con la struttura spaziale offre benefici limitati rispetto alla sola modellazione spaziale.

L'analisi complessiva dei risultati porta a concludere che le covariate sono predittori più potenti quando considerate da sole nel *modello bayesiano ridotto*, mentre la struttura spaziale isolata fornisce informazioni moderatamente utili, come evidenziato dai risultati del Kriging Ordinario. La combinazione di entrambe le componenti non offre benefici aggiuntivi né nel contesto bayesiano né nel caso del Kriging. La componente spaziale residua risulta quindi principalmente rumore piuttosto che segnale strutturato, introducendo complessità piuttosto che miglioramenti nella previsione.

## 7 Considerazioni finali

In conclusione di quanto risultato sul dataset completo, il modello BART si è rivelato il più performante grazie alla sua capacità di catturare relazioni non lineari tra la variabile target e le covariate mentre il Mixed Model ha evidenziato l'utilità dell'aggiunta di effetti casuali per risaltare la variazione spaziale delle variabili esplicative.

Per quanto riguarda i risultati per il dataset ottenuto tramite campionamento stratificato proporzionale, essi hanno mostrato che un approccio non spaziale, basato su covariate

selezionate, si rivela più efficace e parsimonioso rispetto ai modelli computazionalmente più complessi che incorporano la dipendenza spaziale in modo esplicito.

## Riferimenti bibliografici

- Chipman, H. A., George, E. I., and McCulloch, R. E. (2010). Bart: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266–298.
- Corradin, R. (2024/25). Corso di Bayesian Statistical Modeling. <https://rccorradin.github.io/BSM/>.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis*. CRC Press, 3rd edition.
- Nugent, C. (2017). California Housing Prices. <https://www.kaggle.com/datasets/camnugent/california-housing-prices>.

## Appendice

### Pre-processing

Variabile	Descrizione	Tipo
longitude	Una misura di quanto a Ovest si trova un'abitazione: più il valore è alto, più è localizzata a Ovest	Numerica
latitude	Una misura di quanto a Nord si trova un'abitazione: più il valore è alto, più è localizzata a Nord	Numerica
housing_median_age	Età media di una casa all'interno di un distretto: un valore basso indica un edificio più recente	Numerica
total_rooms	Numero totale di stanze in un distretto	Numerica
total_bedrooms	Numero totale di camere da letto in un distretto	Numerica
population	Numero totale di persone residenti in un distretto	Numerica
households	Numero totale di nuclei familiari (gruppo di persone che risiedono in un'unità abitativa) per distretto	Numerica
median_income	Reddito medio dei nuclei familiari all'interno di un distretto (misurato in decine di migliaia di dollari USA)	Numerica
ocean_proximity	Posizione dell'abitazione rispetto all'oceano. Assume i valori: <ul style="list-style-type: none"> <li>• &lt;1H OCEAN: località a meno di un'ora di macchina dalla costa</li> <li>• INLAND: località interne lontane dalla costa</li> <li>• ISLAND: località sulle isole</li> <li>• NEAR BAY: località vicino a una baia</li> <li>• NEAR OCEAN: località costiere sull'Oceano Pacifico</li> </ul>	Categorica
median_house_value	Variabile target. Prezzo mediano delle abitazioni all'interno di un distretto (in dollari USA)	Numerica

Tabella 3: Descrizione delle variabili del dataset

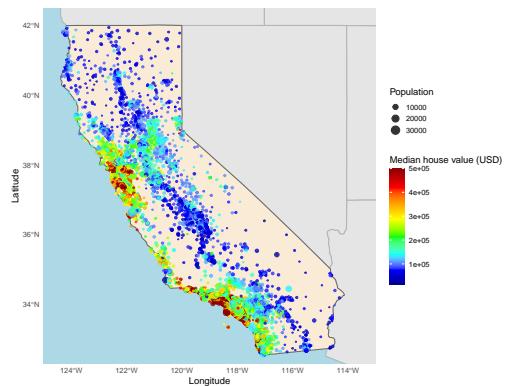


Figura 1: Popolazione e costo mediano delle abitazioni per distretto abitativo in California

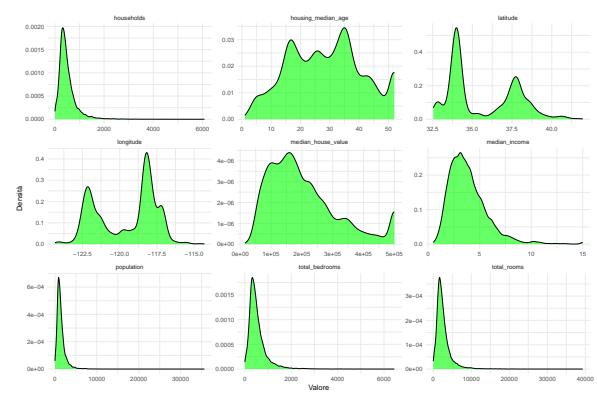


Figura 2: Densità delle variabili quantitative incluse nel dataset

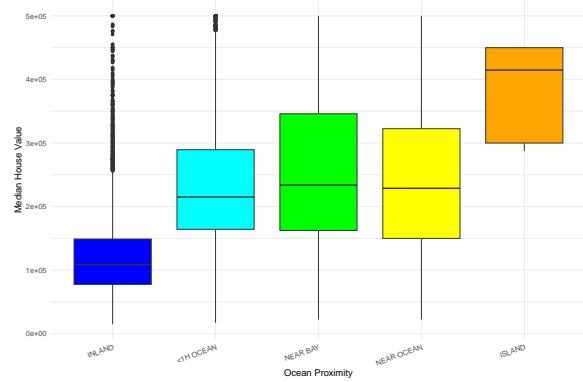


Figura 3: Boxplot del valore mediano delle abitazioni per distanza dall'oceano

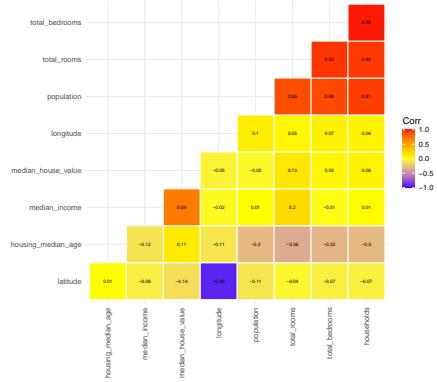


Figura 4: Correlogramma delle variabili di partenza

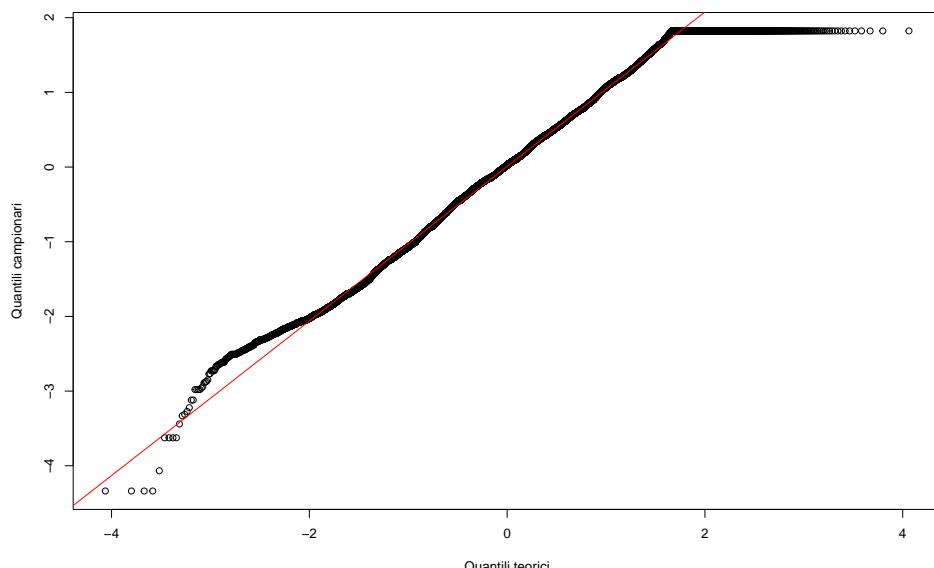
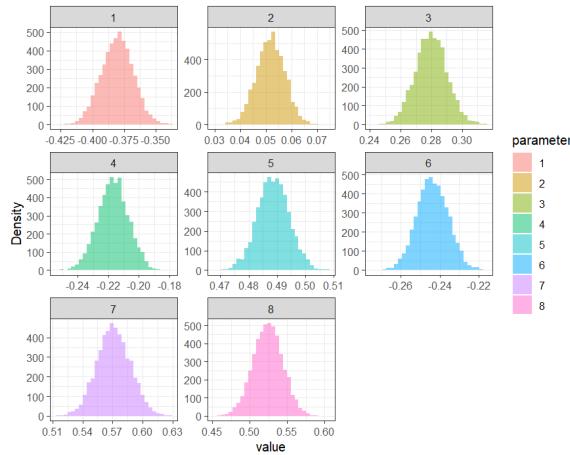
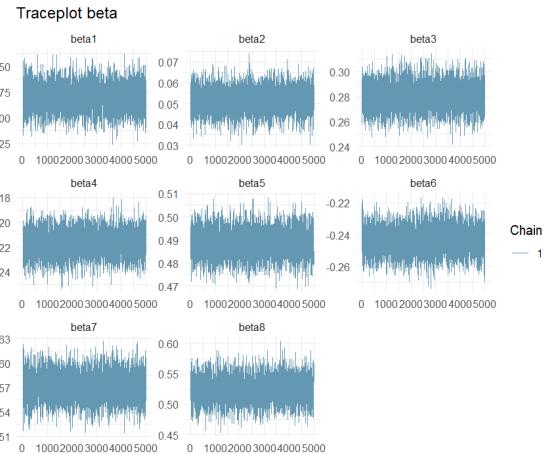
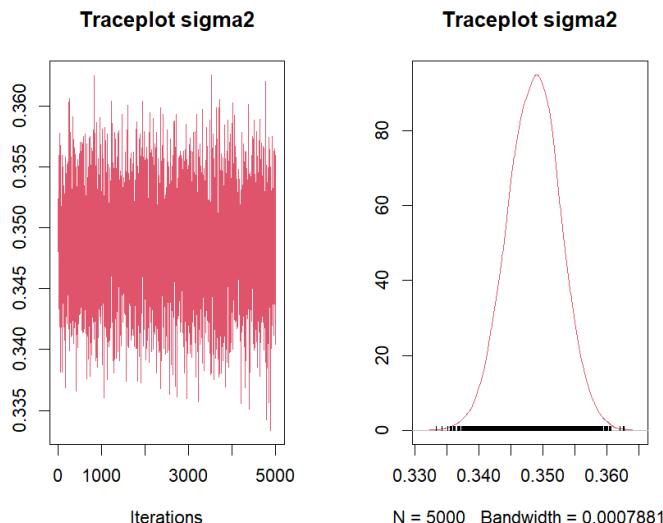


Figura 5: Q-Q plot della variabile risposta a seguito della trasformazione su scala logaritmica e standardizzazione

## Prior non informativa


 Figura 6: Distribuzione a posteriori delle  $\beta$ 

 Figura 7: Traceplot  $\beta$ 

 Figura 8: Traceplot e distribuzione a posteriori  $\sigma^2$ 

	Naive SE	Time-series SE
beta[1]	1.775e-04	1.775e-04
beta[2]	7.888e-05	7.888e-05
beta[3]	1.456e-04	1.456e-04
beta[4]	1.417e-04	1.417e-04
beta[5]	7.947e-05	7.780e-05
beta[6]	1.142e-04	1.142e-04
beta[7]	2.378e-04	2.378e-04
beta[8]	2.799e-04	2.799e-04

 Tabella 4: Stime degli errori standard per  $\beta$ 

	Naive SE	Time-series SE
$\sigma^2$	5.775e-05	5.598e-05

 Tabella 5: Errore standard per  $\sigma^2$

	beta[1]	beta[2]	beta[3]	beta[4]	beta[5]	beta[6]	beta[7]	beta[8]
ESS	5000.000	5000.000	5000.000	5000.000	5217.097	5000.000	5000.000	5000.000

Tabella 6: Effective Sample Size per i parametri  $\beta$

Parametro	ESS
$\sigma^2$	5321.96

Tabella 7: Effective Sample Size per  $\sigma^2$

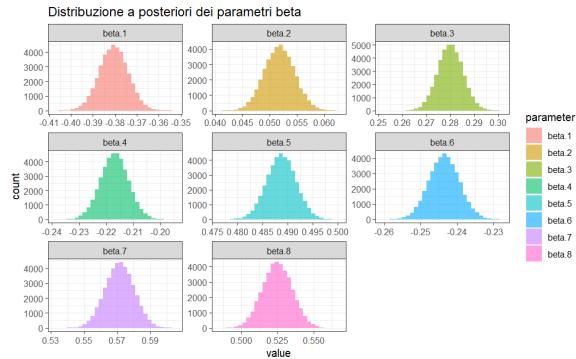
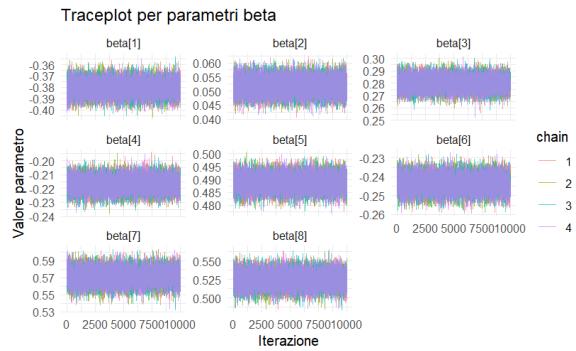
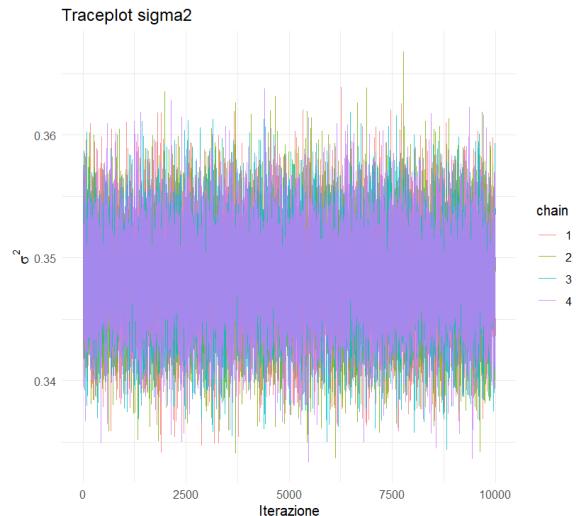
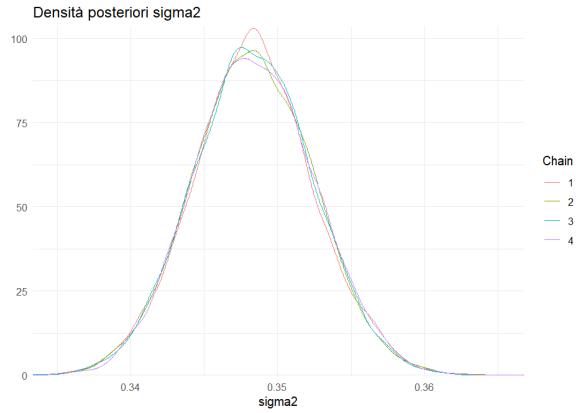
beta[1]	beta[2]	beta[3]	beta[4]	beta[5]	beta[6]	beta[7]	beta[8]
0.29679	0.09872	-0.17798	0.32748	0.93560	0.16480	-0.37861	-0.89849

Tabella 8: Statistiche di Geweke per i parametri  $\beta$

Parametro	Geweke
$\sigma^2$	0.7213

Tabella 9: Valore di Geweke per  $\sigma^2$

## Prior poco informativa: UIP


 Figura 9: Distribuzione a posteriori  $\beta$ 

 Figura 10: Traceplot  $\beta$ 

 Figura 11: Traceplot  $\sigma^2$ 

 Figura 12: Distribuzione a posteriori  $\sigma^2$ 

	Naive SE	Time-series SE
beta[1]	3.260e-05	5.344e-05
beta[2]	1.397e-05	1.472e-05
beta[3]	2.645e-05	3.595e-05
beta[4]	2.567e-05	3.485e-05
beta[5]	1.455e-05	1.702e-05
beta[6]	2.089e-05	3.159e-05
beta[7]	4.387e-05	7.037e-05
beta[8]	5.179e-05	8.372e-05

 Tabella 10: Stime degli errori standard per  $\beta$ 

	Naive SE	Time-series SE
chain:1	4.092e-05	4.423e-05
chain:2	4.096e-05	4.351e-05
chain:3	4.048e-05	4.252e-05
chain:4	4.089e-05	4.261e-05

 Tabella 11: Errori standard per  $\sigma^2$

	beta[1]	beta[2]	beta[3]	beta[4]	beta[5]	beta[6]	beta[7]	beta[8]
ESS	14898.36	36179.01	21792.24	21886.26	29392.08	17542.24	15590.74	15343.36

Tabella 12: Effective Sample Size per i parametri  $\beta$

	chain 1	chain 2	chain 3	chain 4
ESS	8560.616	8862.185	9063.949	9207.614

Tabella 13: Effective Sample Size per  $\sigma^2$

	beta[1]	beta[2]	beta[3]	beta[4]	beta[5]	beta[6]	beta[7]	beta[8]
chain 1	-2.8861	1.2058	1.9754	-1.8753	-0.8284	2.7034	3.2059	2.8722
chain 2	0.2171	1.4729	0.6756	-0.7826	-0.2453	-0.1774	-0.2713	0.0297
chain 3	0.0933	1.2861	0.1377	0.6599	0.5962	0.3487	0.1457	0.1712
chain 4	-0.2274	0.4170	-0.5140	-0.0818	0.0938	0.2318	0.1420	-0.1717

Tabella 14: Statistiche di Geweke per i parametri  $\beta$

	chain 1	chain 2	chain 3	chain 4
$\sigma^2$	0.3732	0.1405	1.3046	-0.2787

Tabella 15: Statistiche di Geweke per  $\sigma^2$

## Prior poco informativa con shrinkage

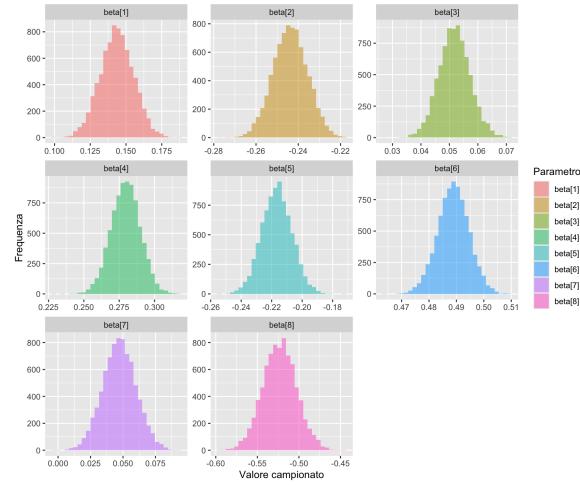


Figura 13: Distribuzione a posteriore  $\beta$

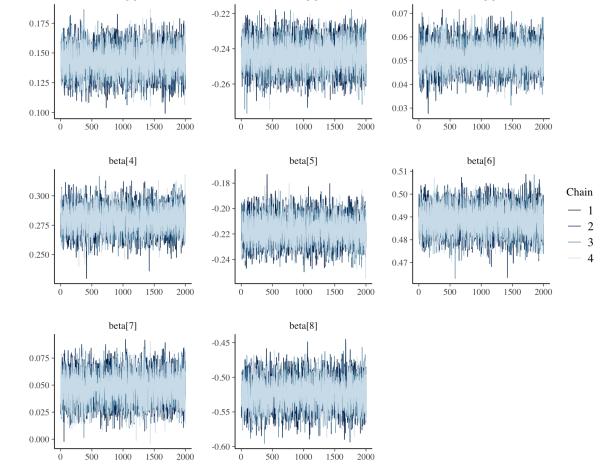


Figura 14: Traceplot  $\beta$

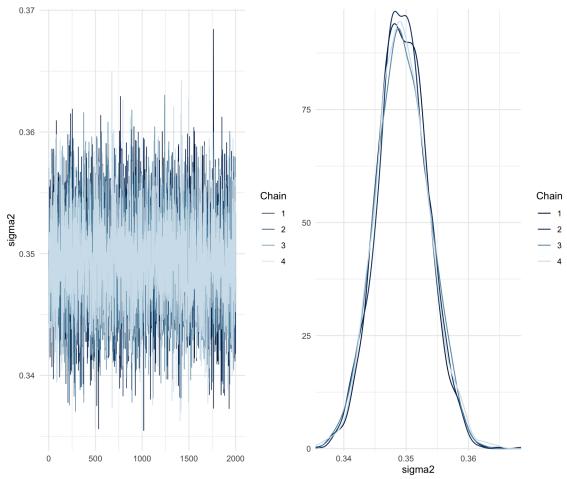


Figura 15: Traceplot e distribuzione a posteriore di  $\sigma^2$

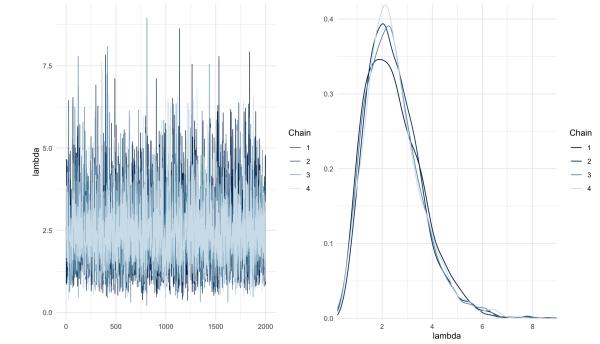


Figura 16: Traceplot e distribuzione a posteriore di  $\lambda$

	Naive SE	Time-series SE
beta[1]	1.312e-04	1.889e-04
beta[2]	9.100e-05	1.138e-04
beta[3]	6.152e-05	5.971e-05
beta[4]	1.153e-04	1.430e-04
beta[5]	1.109e-04	1.392e-04
beta[6]	6.395e-05	6.728e-05
beta[7]	1.441e-04	1.836e-04
beta[8]	2.267e-04	3.141e-04

Tabella 16: Stime degli errori standard per  $\beta$

	Naive SE	Time-series SE
chain:1	9.052e-05	9.052e-05
chain:2	9.061e-05	9.079e-05
chain:3	9.488e-05	9.488e-05
chain:4	9.506e-05	9.937e-05

Tabella 17: Errori standard per  $\sigma^2$

	Naive SE	Time-series SE
chain:1	0.02562	0.02449
chain:2	0.02402	0.02100
chain:3	0.02498	0.02418
chain:4	0.02472	0.02573

Tabella 18: Errori standard per  $\lambda$

	beta[1]	beta[2]	beta[3]	beta[4]	beta[5]	beta[6]	beta[7]	beta[8]
ESS	3974.399	5293.740	8513.618	5293.689	5118.855	7298.490	4955.089	4245.603

Tabella 19: Effective Sample Size per i parametri  $\beta$

	chain 1	chain 2	chain 3	chain 4
ESS	2000.000	1991.925	2000.000	1830.476

Tabella 20: Effective Sample Size per  $\sigma^2$

	chain 1	chain 2	chain 3	chain 4
ESS	2188.367	2614.870	2134.568	1846.914

Tabella 21: Effective Sample Size per  $\lambda$

	beta[1]	beta[2]	beta[3]	beta[4]	beta[5]	beta[6]	beta[7]	beta[8]
chain 1	0.27254	1.30764	1.20139	0.30726	-0.41655	-0.79415	-0.03617	-0.76877
chain 2	1.58270	0.34310	0.90270	1.00710	-1.84400	1.13540	-0.80630	-0.78560
chain 3	1.14840	0.08950	-0.12467	0.01852	0.45700	0.24562	-1.35898	-0.11256
chain 4	-0.21039	0.67263	-0.20132	0.62544	-0.36299	0.05613	0.01212	-0.60852

Tabella 22: Statistiche di Geweke per i parametri  $\beta$

	chain 1	chain 2	chain 3	chain 4
$\sigma^2$	1.02152	-0.12155	-1.05700	-0.00635

Tabella 23: Statistiche di Geweke per  $\sigma^2$

	chain 1	chain 2	chain 3	chain 4
$\lambda$	1.44870	-0.22995	-0.04411	-0.77659

Tabella 24: Statistiche di Geweke per  $\lambda$

## Mixed Model

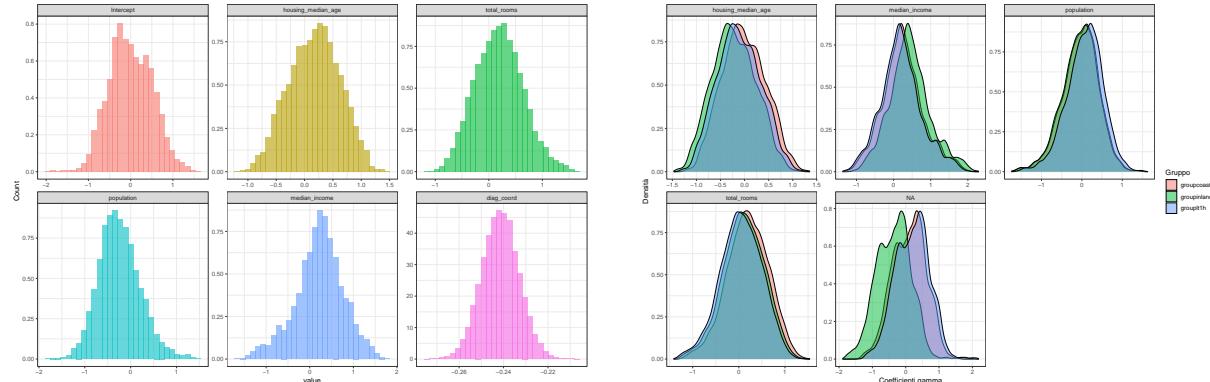


Figura 17: Distribuzione a posteriori dei coefficienti  $\beta$

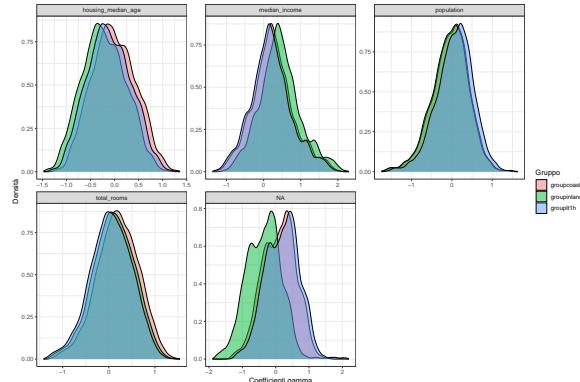


Figura 18: Distribuzione a posteriori dei coefficienti  $\gamma$

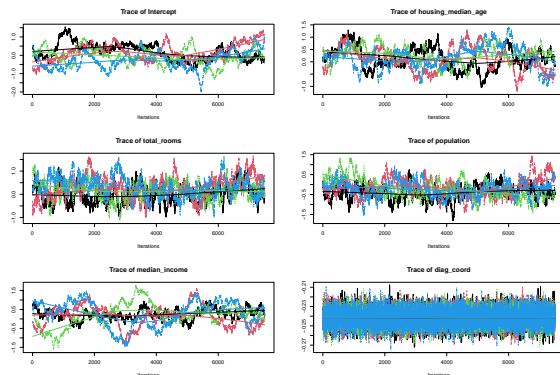


Figura 19: Traceplot dei coefficienti  $\beta$

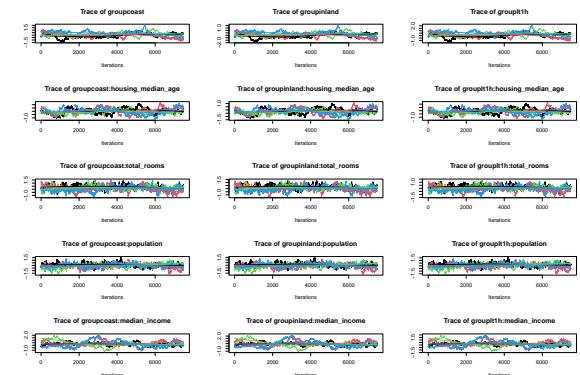


Figura 20: Traceplot dei coefficienti  $\gamma$

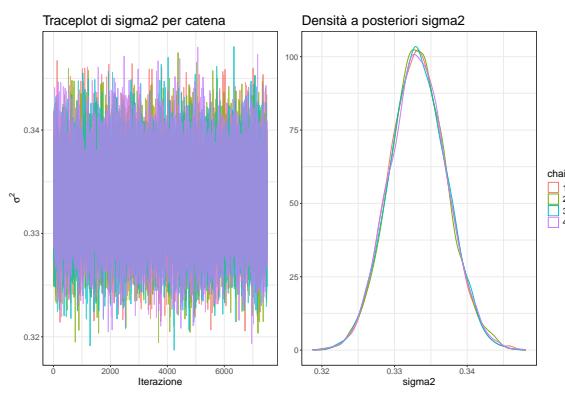


Figura 21: Traceplot e distribuzione a posteriori di  $\sigma^2$

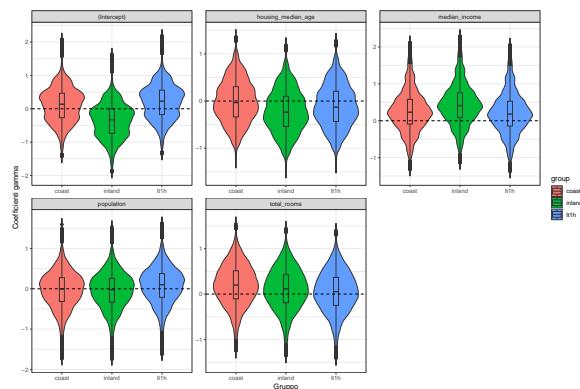


Figura 22: Violin-plot dei coefficienti  $\gamma$

	Naive SE	Time-series SE
beta[1]	2.931e-03	9.382e-02
beta[2]	2.589e-03	6.926e-02
beta[3]	2.587e-03	3.733e-02
beta[4]	2.609e-03	3.918e-02
beta[5]	3.188e-03	1.071e-01
beta[6]	4.807e-05	4.807e-05

Tabella 25: Stime degli errori standard per i coefficienti  $\beta$

	Naive SE	Time-series SE
gamma[1]	0.002930	0.09395
gamma[2]	0.002931	0.09370
gamma[3]	0.002931	0.09233
gamma[4]	0.002589	0.07005
gamma[5]	0.002589	0.06974
gamma[6]	0.002589	0.06965
gamma[7]	0.002591	0.03738
gamma[8]	0.002587	0.03734
gamma[9]	0.002587	0.03733
gamma[10]	0.002612	0.04300
gamma[11]	0.002610	0.03942
gamma[12]	0.002609	0.03987
gamma[13]	0.003187	0.10729
gamma[14]	0.003187	0.10695
gamma[15]	0.003188	0.10695

Tabella 26: Stime degli errori standard per i coefficienti  $\gamma$

	Naive SE	Time-series SE
chain:1	4.497e-05	4.405e-05
chain:2	4.518e-05	4.414e-05
chain:3	4.526e-05	4.526e-05
chain:4	4.559e-05	4.559e-05

Tabella 27: Stime degli errori standard per  $\sigma^2$

	beta[1]	beta[2]	beta[3]	beta[4]	beta[5]	beta[6]
chain:1	-0.0604	2.2952	-0.4700	-0.4002	0.6278	-0.7736
chain:2	-3.05545	0.29266	-4.80676	0.09422	-0.69642	-0.08007
chain:3	1.7993	-0.4091	0.2180	4.8699	-2.9594	0.3036
chain:4	-0.7515	-0.1718	1.8841	3.2581	0.4009	-0.1106

Tabella 28: Statistiche di Geweke per i coefficienti  $\beta$

	gamma[1]	gamma[2]	gamma[3]	gamma[4]	gamma[5]	gamma[6]	gamma[7]	gamma[8]	gamma[9]	gamma[10]	gamma[11]	gamma[12]	gamma[13]	gamma[14]	gamma[15]
chain:1	0.05965	0.06360	0.05758	-2.26873	-2.34782	2.31696	0.44191	0.47293	0.44522	0.36303	0.39030	0.39976	-0.63421	-0.64009	-0.64233
chain:2	2.99591	3.09718	3.06940	-0.29924	-0.29977	-0.29409	4.83523	4.81616	4.97137	-0.09967	-0.10067	-0.09486	0.69300	0.68293	0.69822
chain:3	-1.7962	-1.7662	-1.8005	0.3999	0.4026	0.4111	-0.2121	-0.2272	-0.2116	-4.9450	-4.8338	-4.8528	2.8904	2.8603	2.8538
chain:4	0.7785	0.7777	0.7599	0.1817	0.1674	0.1694	-1.8966	-1.9152	-1.9223	-3.3649	-3.3745	-3.2865	-0.3891	-0.4219	-0.3962

Tabella 29: Statistiche di Geweke per i coefficienti  $\gamma$

	chain:1	chain:2	chain:3	chain:4
$\sigma^2$	-0.8426	0.6376	0.9953	1.127

Tabella 30: Statistiche di Geweke per  $\sigma^2$

## BART

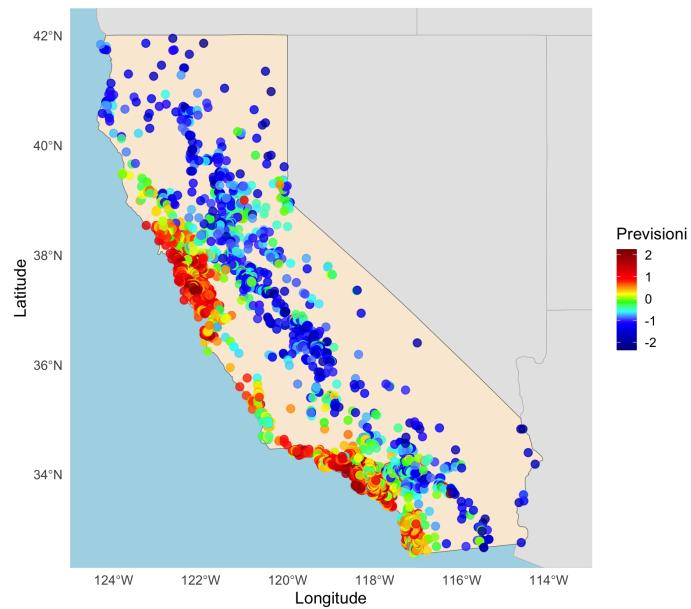


Figura 23: Mappa California con valori previsti

## Valori reali vs Valori previsti bayesiane con intervalli di credibilità al 95%

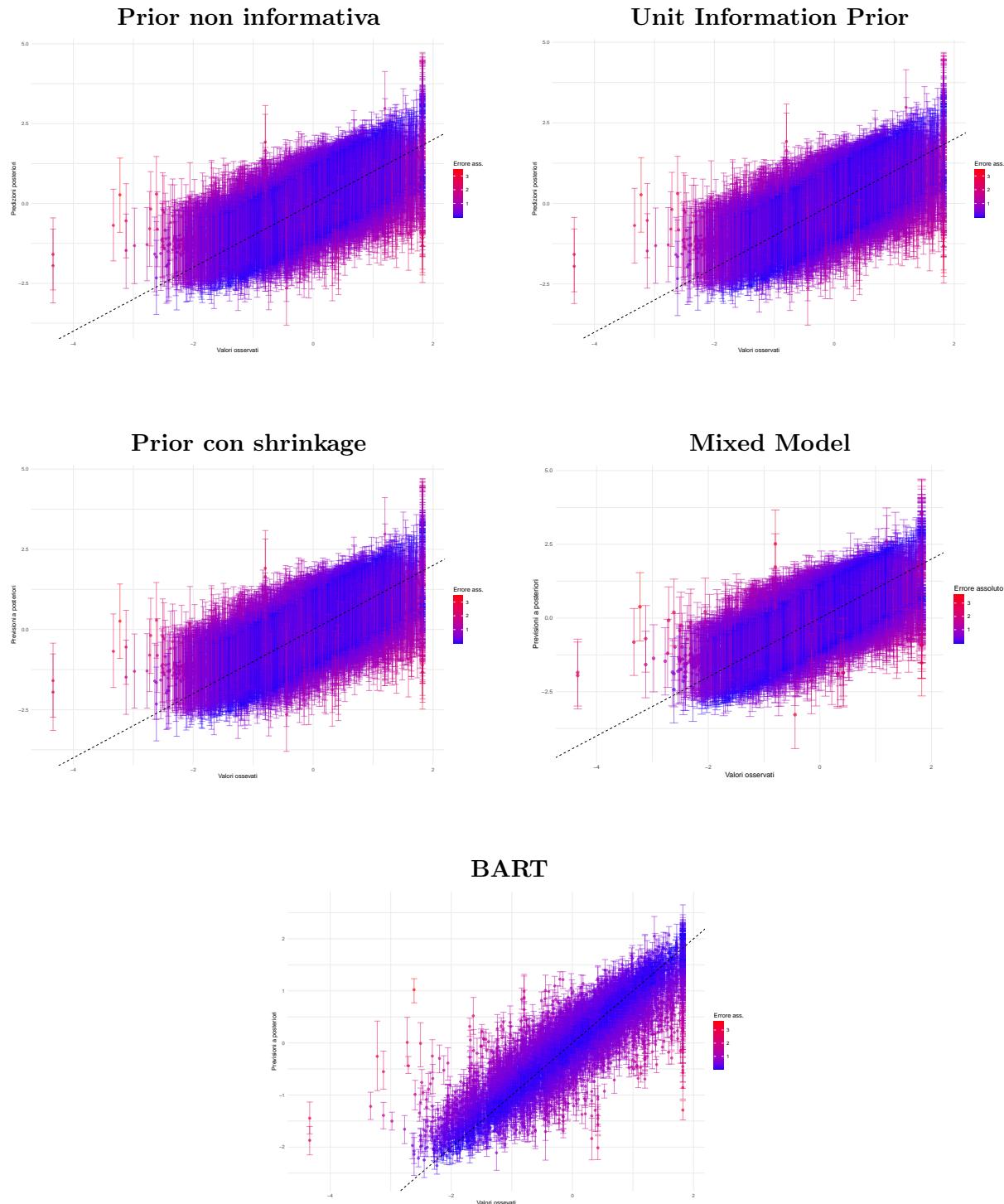


Figura 24: Confronto tra i modelli

## Analisi Spaziale

Coefficiente	Variabile
beta[1]	Intercetta
beta[2]	total_rooms
beta[3]	population
beta[4]	median_income
beta[5]	ocean_lt1h
beta[6]	ocean_inland
beta[7]	diag_coord

Tabella 31: Corrispondenza tra coefficienti  $\beta$  e variabili del modello

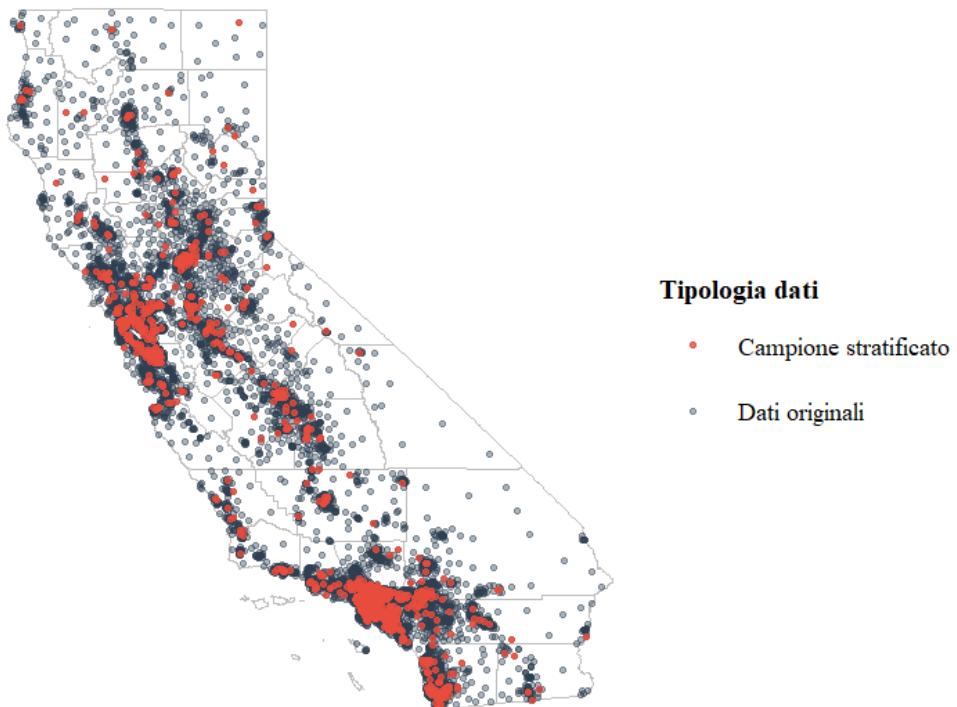


Figura 25: Confronto tra dati originali e campione stratificato proporzionale

Ridotto vs Completo	
log Bayes Factor	113.4799

Tabella 32: Confronto tra il modello con le sole covariate e il modello che considera anche la componente spaziale tramite log Bayes factor

Parametro	Mean	SD	Naive SE	Time-series SE
sigma2	0.2011	0.02762	0.0008734	0.0008734
tau2	0.1672	0.01187	0.0003754	0.0004018
phi	17.8009	2.02699	0.0640991	0.0640991

Parametro	Mean	SD	Naive SE	Time-series SE
beta[1]	0.1682	0.08825	0.0027908	0.003847
beta[2]	0.3108	0.04693	0.0014842	0.001931
beta[3]	-0.2516	0.04486	0.0014187	0.001923
beta[4]	0.3979	0.02307	0.0007296	0.000908
beta[5]	-0.1774	0.09852	0.0031156	0.003925
beta[6]	-0.5826	0.12697	0.0040150	0.005494
beta[7]	-0.2468	0.03974	0.0012566	0.001323

Tabella 33: Riassunti posteriori dei parametri del modello con covariate e componente spaziale

Parametro	Geweke	Parametro	Geweke
sigma2	0.5246	beta[1]	0.10632
tau2	0.5280	beta[2]	0.87839
phi	-0.2733	beta[3]	-1.22826
		beta[4]	-0.99338
		beta[5]	0.08458
		beta[6]	-0.44685
		beta[7]	0.93200

Tabella 34: Statistiche di Geweke per il modello con covariate e componente spaziale

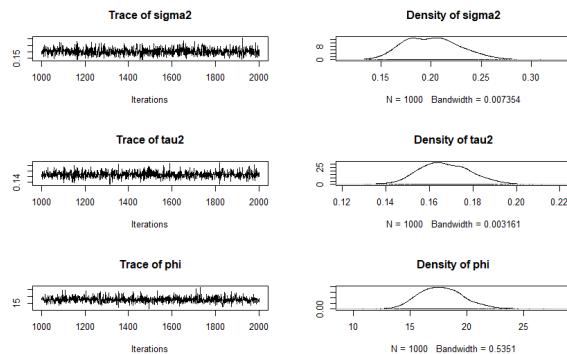


Figura 26: Catene MCMC (traceplot) e densità a posteriori dei parametri spaziali del modello con covariate e componente spaziale (2000 iterazioni, 1000 warm-up)

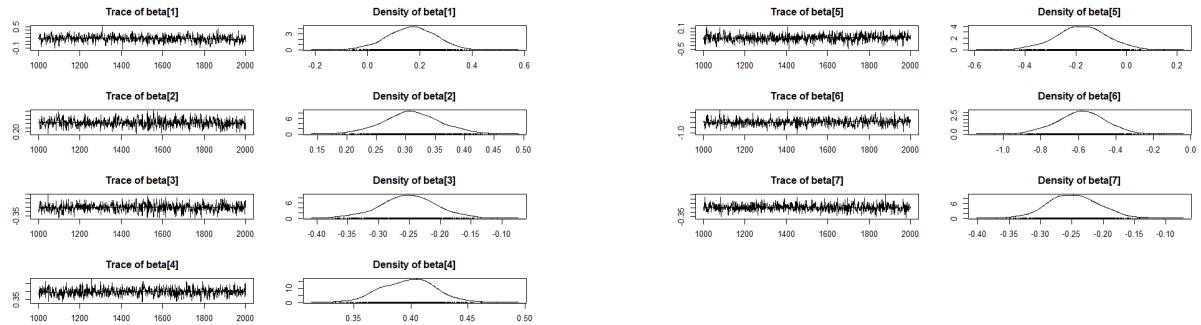


Figura 27: Catene MCMC e densità a posteriori dei coefficienti di regressione del modello con covariate e componente spaziale (2000 iterazioni, 1000 warm-up)

Parametro	Mean	SD	Naive SE	Time-series SE
beta[1]	0.18300	0.05410	0.0017108	0.0026037
beta[2]	0.29050	0.05700	0.0018025	0.0025237
beta[3]	-0.24893	0.05706	0.0018043	0.0024893
beta[4]	0.49634	0.02784	0.0008804	0.0009894
beta[5]	0.02413	0.05955	0.0018832	0.0025075
beta[6]	-0.54356	0.09012	0.0028499	0.0041643
beta[7]	-0.25841	0.03573	0.0011300	0.0014385

Tabella 35: Riassunti posteriori per il modello con le sole covariate

Parametro	Geweke
beta[1]	-0.71043
beta[2]	0.05963
beta[3]	0.02602
beta[4]	0.88301
beta[5]	-0.48515
beta[6]	0.56551
beta[7]	-0.72914

Tabella 36: Statistiche di Geweke per il modello con le sole covariate

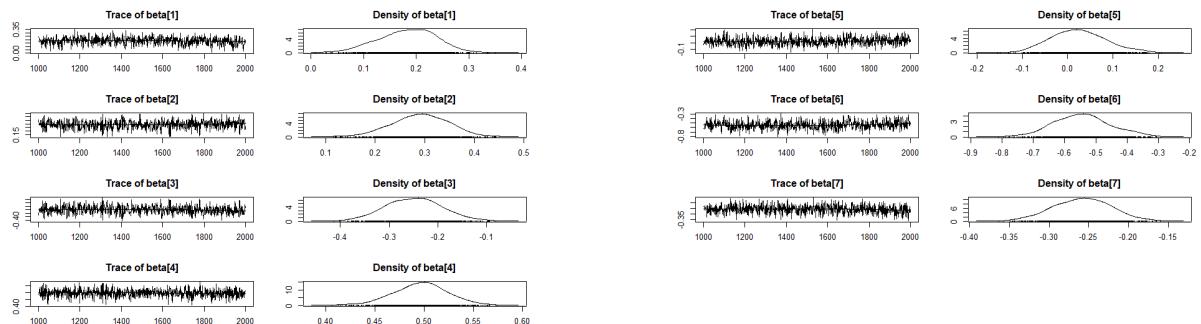


Figura 28: Catene MCMC e densità a posteriori dei coefficienti di regressione del modello con sole covariate (2000 iterazioni, 1000 warm-up)