Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche

Metodi statistici per l'analisi di dati spaziali categoriali: un'applicazione su dati di utilizzazione del suolo

Relatore: Prof.ssa Giuliana Cortese
Dipartimento di Scienze Statistiche

Correlatore: Dott. Paolo Girardi
DAIS - Università Ca' Foscari Venezia

Laureando: Francesco Gugole
Matricola N. 1241715

Anno Accademico 2021/2022

A nonna Luisa.

Indice

Introduzione

La recente crescita nella disponibilità di dati referenziati tramite coordinate, oltre alla parallela facilità nell'accesso ad una moltitudine di informazioni, ha reso possibile lo sviluppo di una branca della statistica dedicata all'analisi di fenomeni e strutture difficilmente indagabili senza considerarne la natura spaziale.
L'ingente quantità di informazione fornita da strumentazioni sempre più evolute, inoltre, necessita di adeguati sistemi informatici che supportino tali moli di dati, spesso complessi, per poterne estrarre significati utili in qualsivoglia processo decisionale.
In generale, l'analisi dei dati spaziali fonda i suoi metodi a partire dalla cosiddetta "prima legge della geografia" (Tobler 1970): questa, infatti, postula che "tutto sia correlato con tutto il resto, ma oggetti vicini siano più correlati di oggetti lontani". Questa legge, del tutto intuitiva, si pone alla base di numerosi metodi di modellazione e dei fondamentali concetti di autocorrelazione spaziale e dipendenza spaziale. È in questo contesto e con queste sfide che nasce l'analisi dei dati spaziali o, per meglio dire, dei dati spazialmente dipendenti.
Gli approcci allo studio di fenomeni calati nello spazio sono molteplici e molti di essi combinano tecniche e conoscenze derivanti sia dalla teoria statistica (spesso generalizzazioni di metodi appartenenti all'analisi di serie storiche) che dai rispettivi campi di applicazione: non sono solo le scienze naturali come la geologia, la geografia e la climatologia ad avvalersi di tali strumenti, ma anche la medicina, l'economia e diverse attività informatiche tra cui, ad esempio, il processamento delle immagini. È altresì vero che, allo stato attuale, la maggior parte delle applicazioni si realizza nel campo geospaziale, soprattutto grazie all'accessibilità a numerosi dati satellitari; in tali casi, ci si riferisce all'ambito dell'analisi di dati spaziali con il nome di "geostatistica".

È molto frequente, quando si trattano dati aventi natura spaziale, imbattersi in misure rilevate su scale continue, certamente importanti in un gran numero di applicazioni. Si pensi, come accennato in precedenza, a fenomeni meteorologici quali precipitazioni, temperature e rilevazioni climatiche in genere piuttosto che a studi topografici relativi ad altitudine o profondità marine. In sintesi, i dati spaziali "continui" sono caratterizzati dall'assenza di confini definiti e da transizioni progressive e "lisce" tra i diversi valori appartenenti al supporto della variabile di interesse.
Talvolta, però, l'informazione disponibile e gli scopi di ricerca sono maggiormente compatibili con problemi a cui, nell'analisi statistica e nel *data mining*, ci si riferisce col nome di "classificazione". Sono questi i casi in cui le rilevazioni vengono espresse sottoforma di variabili dicotomiche ad indicare la presenza (o assenza) di un determinato carattere oppure, generalizzando la casistica, di variabili categoriali. Si noti come, in questo secondo caso, sia possibile imbattersi sia in variabili categoriali sconnesse, sia in variabili categoriali ordinali. Risulta evidente come la presenza di una scala ordinale fornisca maggiore informazione e possa semplificare il lavoro dell'analista.
Nel presente lavoro, l'attenzione viene posta proprio sui dati spaziali categoriali e sui modelli atti a descriverli ed effettuare previsioni. Dopo un capitolo introduttivo dedicato ai dati spaziali e alle loro caratteristiche intrinseche (§2), si illustrano i principali metodi di analisi utilizzati attualmente (§3). Successivamente (§4) viene testato il modello di riferimento per dati categoriali, ovvero quello che fa utilizzo dei cosiddetti "transiogrammi", mediante opportune simulazioni in diverse condizioni e tramite il confronto con un altro metodo di ricostruzione delle immagini. Prima di concludere (§5), il modello viene applicato a diversi casi di studio per valutarne l'utilità in contesti reali.

1 Tipi di dati spaziali

Nella letteratura sia passata che recente (Cressie 1993; Schabenberger e Gotway 2017) è frequente la distinzione che viene operata tra i diversi tipi di dati spaziali: dati geostatistici, dati *lattice* (o regionali) e dati che individuano motivi e schematicità (i cosiddetti *point pattern*).
Le tre tipologie differiscono sia per i differenti campi di applicazione che per la differente natura della rilevazione effettuata e consentono o, per meglio dire, impongono metodologie di analisi adeguate a seconda della disponibilità. Per facilitare la comprensione del seguito della trattazione, viene introdotto il concetto di *processo spaziale*, definito in d dimensioni attraverso la seguente scrittura:

$$\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d\}, \tag{1.1}$$

dove Z rappresenta la variabile di interesse, $\mathbf{s}$ il set di coordinate spaziali e D il dominio. È molto frequente che il numero di dimensioni d sia pari a 2 o 3. I tre tipi di dati spaziali elencati precedentemente differiscono, sostanzialmente, per le caratteristiche del dominio D, come descritto nelle seguenti sezioni. Vale la pena sottolineare come sia il dominio a discriminare la tipologia dei dati e non tanto la natura continua o discreta della variabile dipendente.
Poiché il presente lavoro si concentra su fenomeni a variabile risposta categoriale, i seguenti esempi insistono su tali situazioni.

1.0.1 Dati geostatistici

Nel caso dei dati geostatistici, il dominio D definito poc'anzi è continuo e fissato. Per dominio "fissato" si intende l'immutabilità di questo tra un'osservazione del processo e l'altra. La continuità mette in evidenza come un'osservazione della variabile di interesse $Z(\mathbf{s})$ sia rilevabile in ogni punto dello spazio. Tali dati, quindi, mostrano mutamenti "continui" all'interno del dominio considerato. Ed è proprio la continuità del dominio a rendere impossibile un campionamento completo dello spazio; per questo motivo, uno degli obiettivi nell'analisi dei dati geostatistici è quello di ricostruire, mediante previsioni, la superficie relativa alla variabile Z. Tale tipo di dati è particolarmente frequente nelle applicazioni meteorologiche e geografiche.
Un esempio chiarificatore può essere individuato nell'analisi delle precipitazioni rilevate presso le stazioni meteorologiche sparse sull'intero territorio dello stato di Washington, negli Stati Uniti d'America (National Centers for Environmental Information 2021).

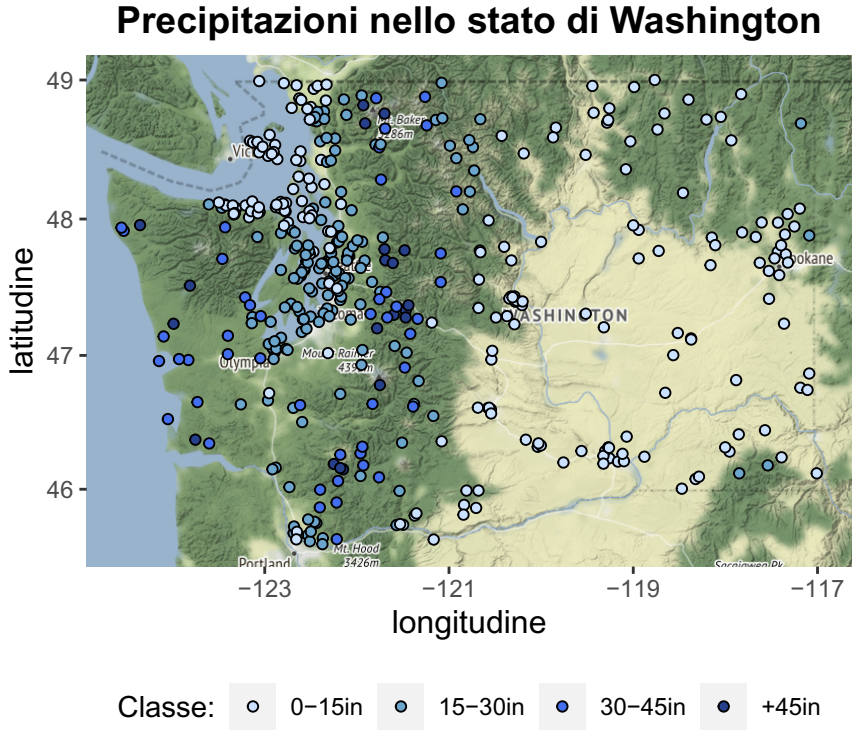

Figura 1.1: precipitazioni nello stato di Washington, USA in pollici. Rilevamento effettuato su 427 stazioni con almeno 110 rilevazioni nel periodo gennaio-aprile 2021. Categorizzazione in 4 classi.

I dati utilizzati provengono dai National Centers for Environmental Information (NCEI), ovvero l'autorità che collabora con la National Oceanic and Atmospheric Administration (NOAA) fornendo dati per lo studio di eventi e fenomeni idrologici e atmosferici e sono stati categorizzati in 4 classi, rappresentate in figura dai diversi colori. Nel caso specifico, il dataset presenta numerose stazioni meteorologiche di rilevamento; ne vengono considerate solo alcune, selezionando quelle che nel periodo considerato (gennaio-aprile 2021) hanno raccolto dati per almeno 110 giorni.
Come si evince da una semplice analisi visuale, la zona dei rilievi e quella prossimale alla costa è sensibilmente più piovosa di quella più pianeggiante e situata nell'entroterra. In particolare, si evidenzia una cintura di centraline che hanno rilevato alti livelli di precipitazioni attorno alla zona individuata dallo stretto di Puget. Un lieve incremento del fenomeno meteorologico di interesse è osservabile anche in corrispondenza dei rilievi più orientali dello stato.

1.0.2 Dati regionali (o lattice)

I dati regionali (o *lattice*) appartengono a domini D fissati e discreti e, spesso, sono frutto di aggregazioni o conteggi. Poiché in questo caso le misure sono effettuate su aree, il concetto di "distanza" può non risultare di facile comprensione; è per questo motivo che è frequente identificare un luogo rappresentativo per ciascuna regione, individuato tramite coordinate precise.
Numerosi esempi possono essere elencati, tra cui il seguente relativo alla classificazione sismica dei comuni appartenenti alla regione Sicilia. Il dataset contenente le informazioni di interesse è stato ottenuto dal sito del Dipartimento della Protezione Civile ed è aggiornato ad aprile 2021. Il rischio sismico di un singolo comune viene espresso da un numero intero compreso tra 1 e 4, decrescente con il rischio di scosse telluriche (Dipartimento della Protezione Civile 2021). La rappresentazione grafica seguente esprime tale rischio con l'ausilio di colori diversi per ogni classe.
Anche in questo caso è facile individuare, nonostante l'elevata sismicità media della regione, le zone ritenute particolarmente a rischio, ovvero il territorio intorno allo stretto di Messina e la valle del Belice nella parte sud-occidentale dell'isola.

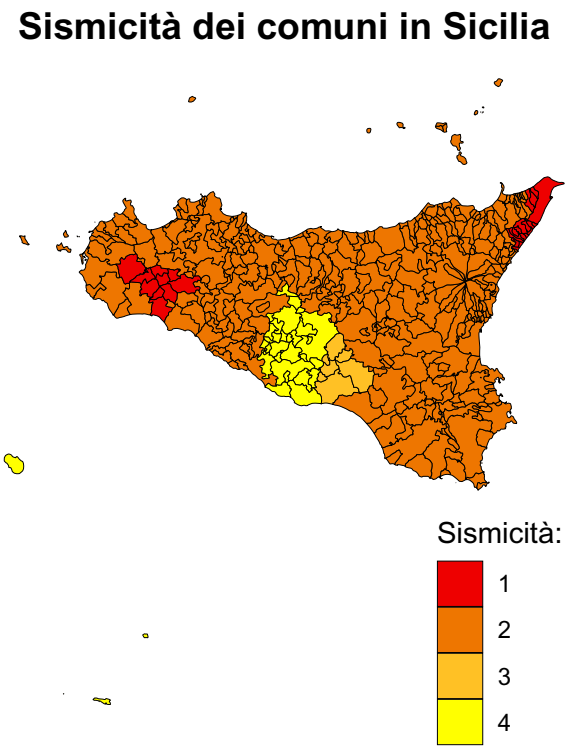

Figura 1.2: sismicità dei comuni siciliani, suddivisi nelle 4 possibili categorie (1: sismicità alta; 2: medio-alta; 3: medio-bassa; 4: bassa).

1.0.3 Point pattern

A differenza dei due casi precedenti, il dominio dei dati *point pattern* è casuale. In altre parole, l'ubicazione della realizzazione della variabile di interesse non è nota a priori, ma viene registrata nel momento in cui si manifestano occorrenze dell'evento studiato. Il campione, di conseguenza, aumenta con il manifestarsi spontaneo o meno del fenomeno nello spazio.

Un esempio di tali dati viene mostrato mediante il dataset fornito dall'Istituto Nazionale di Geofisica e Vulcanologia (INGV) relativo ai terremoti d'interesse per il territorio italiano di magnitudo maggiore o uguale a 4.0 nel periodo tra l'anno 1000 e il 2019 (Istituto Nazionale di Geofisica e Vulcanologia 2021). Per motivi di interpretabilità grafica, la rappresentazione viene limitata agli eventi sismici di magnitudo momento strumentale (MMS) maggiore o uguale a 4.0 e verificatisi a partire dall'anno 1980.

La cartina mostra un'elevatissima frequenza di eventi tellurici lungo la dorsale apenninica, in particolare nelle aree appartenenti ai parchi nazionali dei Monti Sibillini

e del Gran Sasso. Una seconda area di interesse viene individuata più a sud, in corrispondenza del parco nazionale d'Abruzzo, Lazio e Molise. I terremoti esterni a tali zone risultano sporadici e di intensità non particolarmente rilevante.

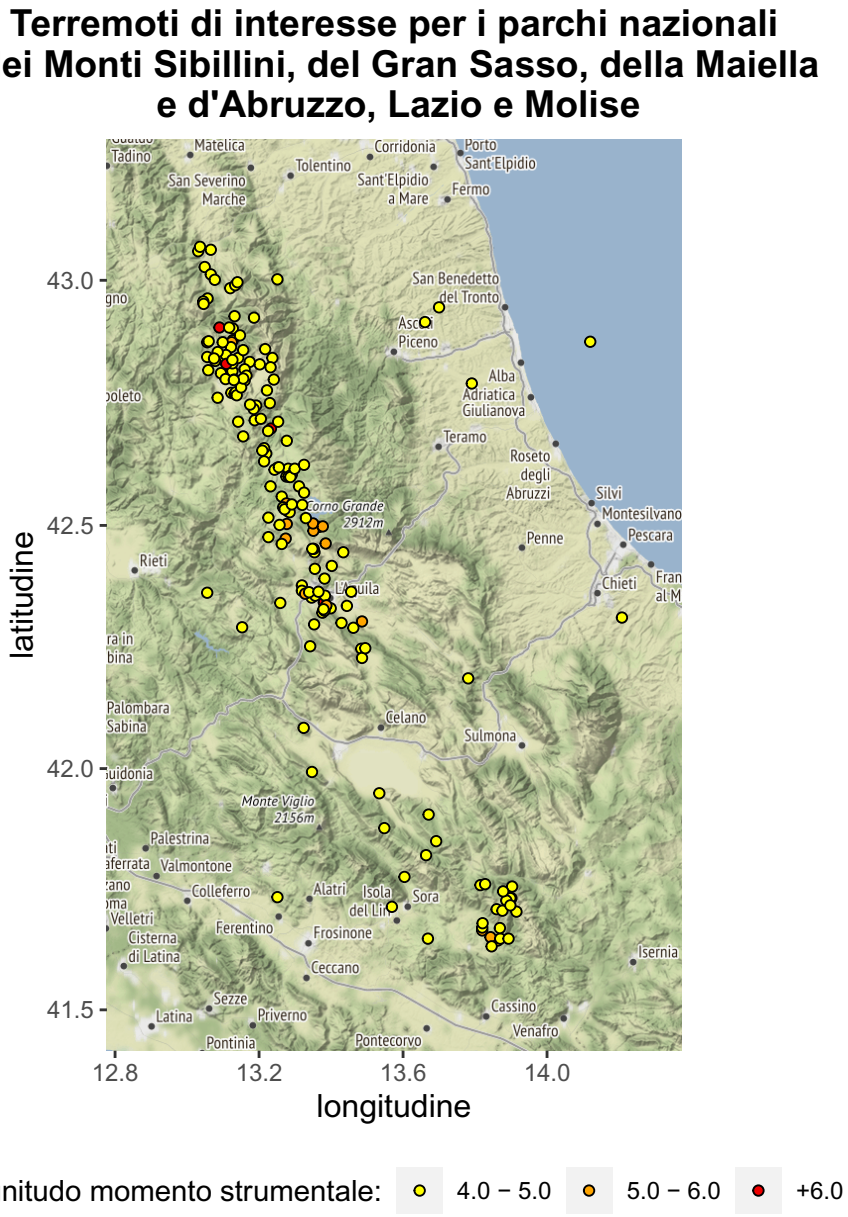

Figura 1.3: eventi sismici di interesse per quattro parchi nazionali italiani (parco nazionale dei Monti Sibllini, parco nazionale del Gran Sasso, parco nazionale della Maiella e parco nazionale d'Abruzzo, Lazio e Molise) di magnitudo momento strumentale maggiore o uguale a 4.0 per il periodo 1980-2019.

1.1 Stazionarietà

Una proprietà importante da valutare nel contesto dell'analisi dei dati spaziali è la stazionarietà.
Si supponga che il processo spaziale considerato abbia una media finita. Utilizzando le definizioni introdotte nella sezione precedente, la media può essere scritta come $\mu(\mathbf{s}) = \mathrm{E}[Z(\mathbf{s})]$. Il processo $Z(\mathbf{s})$ viene detto "gaussiano" se $Z = (Z(\mathbf{s}_1), ..., Z(\mathbf{s}_n))^T$ ha distribuzione Normale multidimensionale, dove n indicizza il set di coordinate $\mathbf{s}$ per ogni osservazione registrata.
In generale, la stazionarietà dei processi spaziali può essere di due tipi: stazionarietà forte (detta anche "stretta") e stazionarietà debole (o "di secondo ordine").
Un processo spaziale definito come in precedenza si dice *stazionario in senso forte* se, per ogni insieme di n punti $\{\mathbf{s}_1, ..., \mathbf{s_n}\}$ e per ogni $\mathbf{h} \in \mathbb{R}^d$, la distribuzione di $(Z(\mathbf{s}_1), ..., Z(\mathbf{s}_n))^T$ è la stessa di $(Z(\mathbf{s}_1 + \mathbf{h}), ..., Z(\mathbf{s}_n + \mathbf{h}))^T$, dove $\mathbf{h}$ definisce la distanza tra due realizzazioni del processo in uno spazio d-dimensionale.
Al contrario, un processo si dice *stazionario in senso debole* se $\mu(\mathbf{s}) = \mu$ (ovvero se la media è costante) e se:

$$\mathrm{Cov}[Z(\mathbf{s}), Z(\mathbf{s}+\mathbf{h})] = \mathrm{E}[(Z(\mathbf{s}) - \mathrm{E}[Z(\mathbf{s})])(Z(\mathbf{s}+\mathbf{h}) - \mathrm{E}[Z(\mathbf{s})])] = C(\mathbf{h}) \qquad (1.2)$$

per ogni $\mathbf{h} \in \mathbb{R}^d$, dove $C(\mathbf{h})$ è una funzione di covarianza detta anche "covariogramma". In altre parole, la stazionarietà debole implica che la relazione tra due punti del processo possa essere espressa in termini di una funzione di covarianza C dipendente solo dalla distanza $\mathbf{h}$ che li separa e non dalle coordinate di questi. Tale risultato, unitamente al fatto che $\mathrm{Cov}[Z(\mathbf{s}), Z(\mathbf{s}+\mathbf{0})] = \mathrm{Var}[Z(\mathbf{s})] = C(\mathbf{0})$, implica che un processo debolmente stazionario abbia anche varianza costante in ogni punto dello spazio.
Come è facilmente intuibile, la stazionarietà forte implica la stazionarietà debole; non vale il contrario, salvo nel caso di un processo gaussiano.
La funzione di covarianza $C(\mathbf{h})$, nel caso di processi stazionari di secondo ordine, manifesta le seguenti proprietà:

1. $C(\mathbf{0}) \geqslant 0$;

2. $C(\mathbf{h}) = C(-\mathbf{h})$;

3. $C(\mathbf{0}) \geqslant |C(\mathbf{h})|$;

4. $C(\mathbf{h}) = \text{Cov}[Z(\mathbf{s}), Z(\mathbf{s}+\mathbf{h})] = \text{Cov}[Z(\mathbf{0}), Z(\mathbf{h})]$.

A partire dalla funzione di covarianza è possibile definire anche la funzione di autocorrelazione:

$$R(\mathbf{h}) = C(\mathbf{h})/C(\mathbf{0}), \tag{1.3}$$

che assume valori compresi tra -1 e 1 (inclusi). Un terzo tipo di stazionarietà è definibile nel caso in cui il processo non sia nemmeno debolmente stazionario, ma lo sia il processo sulle differenze $Z(\mathbf{s}) - Z(\mathbf{s}+\mathbf{h})$; in questo caso si parla di processo *intrinsecamente stazionario*. È possibile mostrare come un processo stazionario in senso debole sia anche intrinsecamente stazionario, mentre non vale la relazione in senso opposto. Più formalmente, un processo è intrinsecamente stazionario se $\text{E}[Z(\mathbf{s})] = \mu$ e se:

$$\frac{1}{2}\text{Var}[Z(\mathbf{s}) - Z(\mathbf{s}+\mathbf{h})] = \gamma(\mathbf{h}), \tag{1.4}$$

dove la funzione $\gamma(\mathbf{h})$ è detta "semivariogramma" o, nella forma $2\gamma(\mathbf{h})$ semplicemente "variogramma". La stima e l'utilizzo di variogrammi (e in particolare di transiogrammi) verranno illustrati nel capitolo 3, dopo aver introdotto altri concetti utili alla trattazione di fenomeni con variabile risposta categoriale. È inoltre possibile stabilire una relazione tra il semivariogramma e la funzione di covarianza:

$$\begin{aligned}
\text{Var}[Z(\mathbf{s}) - Z(\mathbf{s}+\mathbf{h})] &= \text{Var}[Z(\mathbf{s})] + \text{Var}[Z(\mathbf{s}+\mathbf{h})] - 2\text{Cov}[Z(\mathbf{s}), Z(\mathbf{s}+\mathbf{h})] \\
&= 2(\text{Var}[Z(\mathbf{s})]) - 2C(\mathbf{h}) \\
&= 2(C(\mathbf{0}) - C(\mathbf{h})) \\
&= 2\gamma(\mathbf{h})
\end{aligned}$$

Metodi e modelli statistici relativi a processi debolmente stazionari, quindi, possono essere approcciati da entrambi i punti di vista. Va tuttavia sottolineato come, nel caso di processi intrinsecamente ma non debolmente stazionari, il parametro $C(\mathbf{h})$ non esista; sono questi i casi in cui si rende necessario lavorare utilizzando $\gamma(\mathbf{h})$. Inoltre, il

variogramma non richiede la conoscenza della media, mentre per il calcolo della funzione di covarianza si rende necessaria.

1.2 Isotropia e anisotropia

Nel paragrafo precedente, la notazione relativa al semivariogramma evidenziava come il valore di tale funzione dipendesse dal vettore di separazione **h** solo in termini della sua lunghezza; in tali casi, il variogramma viene definito *isotropico*. Questa assunzione viene molto spesso utilizzata per semplicità rappresentativa e interpretativa, ma risulta di scarsa rilevanza ai fini pratici.
Nel caso in cui l'assunzione non sia valida, ovvero nelle situazioni in cui la funzione sia variabile non solo con la lunghezza del vettore **h** ma anche con la sua direzione, il variogramma si dice *anisotropico*. Questo caso rappresenta il più comune nelle diverse applicazioni, con particolare riferimento a quelle geologiche e minerarie.
L'assunzione di isotropia risulta particolarmente comoda, in pratica, quando non c'è ragione di ipotizzare *pattern* spaziali del fenomeno di interesse legati alla direzione.

2 Il transiogramma

Per la trattazione del *transiogramma* si rende necessaria una breve e preventiva panoramica del *variogramma*, ovvero dello strumento equivalente nel caso di variabili di interesse definite con supporto continuo.
Come messo in evidenza nel capitolo precedente, e riprendendo quanto ivi scritto, il variogramma rappresenta una misura di dipendenza spaziale al variare della distanza che intercorre tra due punti appertenenti al dominio di rilevazione.

2.1 Variogramma

Si considerino due punti nello spazio con coordinate $\mathbf{s_1}$ e $\mathbf{s_2}$ a cui sono associate, rispettivamente, le misure z_1 e z_2, realizzazioni del processo $Z(\mathbf{s})$. Il variogramma viene costruito a partire da uno scatterplot nel quale si rappresentano sulle ascisse le distanze $|\mathbf{s_1} - \mathbf{s_2}|$ tra i due punti e, sulle ordinate, le quantità $\frac{1}{2}(z_1 - z_2)^2$. Sotto l'assunzione di media costante, uno stimatore per il variogramma empirico può essere ottenuto tramite il metodo dei momenti. Il relativo stimatore assume la forma (Matheron 1962):

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2, \tag{2.1}$$

con $h \in \mathbb{R}^d$ e $N(\mathbf{h})$ è definito come:

$$N(\mathbf{h}) = \{(\mathbf{s}_i, \mathbf{s}_j) : \mathbf{s}_i - \mathbf{s}_j = \mathbf{h};\ i, j = 1, ..., n\}, \tag{2.2}$$

e indica le coppie di punti separate da un vettore di lunghezza $\mathbf{h}$; $|N(\mathbf{h})|$ è la cardinalità di tale insieme di coppie di punti. Poiché in molti contesti, a causa della

distribuzione irregolare dei dati, non è comune osservare numerose coppie di punti separate dalla stessa distanza, si considerano delle regioni di tolleranza.
Dal momento che è possibile osservare anisotropia (ovvero comportamenti diversi a seconda della direzione considerata), è frequente la costruzione di scatterplot per classi di direzione (solitamente non meno di quattro); le direzioni vengono scelte a partire dalla teoria sottostante al fenomeno considerato.
L'interpretazione di un variogramma viene condotta a partire dal riconoscimento e dalla quantificazione di tre elementi: *range*, *soglia* e *nugget effect*. In alcuni casi, il grafico può avere un andamento sempre crescente all'aumentare di $\mathbf{h}$; in altre, invece, è possibile che il grafico si stabilizzi su un certo valore (la "soglia") a significare che, oltre una certa distanza, $Z(\mathbf{s})$ e $Z(\mathbf{s}+\mathbf{h})$ sono incorrelate. Tale distanza è detta "range". Inoltre, mentre alcuni fenomeni generano un variogramma che parte dall'origine, in certe situazioni il grafico parte da un valore diverso da zero sulle ordinate. Questa discontinuità (si ricorda che $\gamma(\mathbf{0}) = 0$) può essere dovuta a diverse cause, tra cui componenti del fenomeno di interesse con un raggio inferiore a quello del passo di campionamento o errori di misura; ci si riferisce ad essa con il nome di "effetto nugget". Naturalmente il variogramma empirico non fornisce un'indicazione sulla dipendenza spaziale ad ogni lag, ma solo in corrispondenza dei punti per cui è possibile calcolarlo. È altresì evidente come in un qualsiasi contesto di analisi questo sia limitante. Per questo motivo vengono stimati modelli che consentano di ottenere maggiori informazioni dai dati e che si rendono necessari per lo sviluppo di metodi più complessi. Tali modelli differiscono per la forma della funzione interpolante e, in particolare, nella modellazione ai bassi lag, dove la presenza di un eventuale "nugget" rende più delicata l'interpretazione del fenomeno. Tra i modelli più utilizzati rientrano quello lineare, quello sferico, il modello esponenziale, il modello gaussiano, quello di Matérn e il modello legge di potenza. Nella relativa letteratura (Chilès e Delfiner 1999; Banerjee, Carlin e Gelfand 2003), tali modelli vengono spesso proposti nei termini della funzione di covarianza, da cui risulta semplice ricavarne la forma diretta anche per il variogramma.

2.2 Variogramma indicatore

Poiché il focus del seguente lavoro è sui dati categoriali, prima di introdurre i metodi dedicati si rende necessario illustrare anche il *variogramma indicatore*.
Per una qualsiasi variabile continua $Z(\mathbf{s})$ è possibile definire una o più funzioni indicatrici associate a corrispondenti soglie z:

$$I(\mathbf{s}; z) = \mathbb{1}_{Z(\mathbf{s}) < z} = \begin{cases} 1 & \text{se } Z(\mathbf{s}) < z, \\ 0 & \text{altrimenti} \end{cases} \tag{2.3}$$

A partire dalle funzioni indicatrici è possibile esprimere la forma del variogramma indicatore (Bárdossy 1997):

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (I(\mathbf{s}_i) - I(\mathbf{s}_j))^2. \tag{2.4}$$

Ogni variabile $I(\mathbf{s})$ può essere considerata come una variabile casuale di Bernoulli che prende valore 1 con probabilità p e valore 0 con probabilità $q = 1 - p$. N variabili casuali di Bernoulli indipendenti seguono una distribuzione Binomiale che, sotto le specifiche condizioni imposte dal teorema centrale del limite, viene approssimata soddisfacentemente da una distribuzione Normale. Nel caso della statistica spaziale, tali variabili casuali bernoulliane sono tra loro correlate; si assume, quindi, che possano essere approssimate da una distribuzione Normale multivariata (Pardo-Igúzquiza, Grimes e Teo 2006).
Il variogramma così definito permette lo sviluppo e la modellazione del cosiddetto "indicator kriging", una variante per variabili categoriali del ben più noto e diffuso *kriging* semplice; la difficile interpretazione fisica nel caso di dati geologici o simili e l'incapacità di cogliere asimmetria nella distribuzione spaziale delle variabili categoriali, però, lo rende uno strumento poco utilizzato e, in certi contesti, superato. La classe di modelli costituita dal kriging e dalle sue varianti relative ai dati categoriali verrà trattata più approfonditamente nel capitolo 4.

2.3 Transiogramma

A causa della grande diffusione nell'applicazione delle catene markoviane unidimensionali, per lungo tempo lo strumento utilizzato per descrivere variazioni temporali e spaziali in dati geografici è stato la matrice delle probabilità di transizione (spesso indicata con l'acronimo TPM - *Transition Probability Matrix*) (Li 2007b). Queste matrici si rivelano molto utili nel rappresentare le relazioni tra le diverse categorie della variabile di interesse, ma difettano nella rappresentazione della dipendenza spaziale ai diversi lag.
Il transiogramma viene introdotto come strumento capace di stimare probabilità di transizione per diversi lag spaziali e rappresenta una buona misura per poter quantificare l'autocorrelazione intra-classe e la cross-correlazione inter-classe. Anche i punti critici propri del variogramma vengono superati: il transiogramma, infatti, fornisce uno stumento grafico di semplice lettura e interpretazione.

Formalmente, un transiogramma si definisce come una funzione unidimensionale della probabilità di transizione per diversi valori della distanza $\mathbf{h}$:

$$p_{ij}(\mathbf{h}) = Pr[Z(\mathbf{s}+\mathbf{h}) = j | Z(\mathbf{s}) = i]. \tag{2.5}$$

Nella fattispecie, $p_{ij}(\mathbf{h})$ rappresenta la funzione della probabilità di transizione della variabile Z dallo stato i allo stato j al variare di $\mathbf{h}$. Solitamente i viene denominata "classe di testa" e j "classe di coda".
La probabilità $p_{ii}(\mathbf{h})$ si riferisce alla probabilità di permanenza nello stesso stato di partenza (dato il lag spaziale $\mathbf{h}$); il diagramma così generato prende il nome di *autotransiogramma*. La probabilità $p_{ij}(\mathbf{h})\,(i \neq j)$, invece, si definisce *cross-transiogramma*.
La variabile casuale Z, nel seguito, sarà ancora assunta come debolmente stazionaria; questa assunzione, unitamente all'ipotesi ergodica, consente di stimare i transiogrammi direttamente dai dati spaziali.
Nel seguito si farà riferimento a due tipi di transiogramma: *transiogrammi teorici* e

transiogrammi empirici (in letteratura indicati rispettivamente come *idealized* e *real-data transiograms*). A prescindere dal tipo, però, le proprietà di base sono le stesse per tutti i transiogrammi:

1. non-negatività: $p_{ij}(\mathbf{h}) \geq 0$;;
2. somma unitaria: $\sum_{j=1}^{n} p_{ij}(\mathbf{h}) = 1, \ \forall\, i$;
3. assenza di *nugget effect*: $p_{ij} = 0$ nel caso di cross-transiogrammi ($i \neq j$) e $p_{ii} = 1$ nel caso di autotransiogrammi (nel caso di classi mutuamente esclusive);
4. asimmetria: tipicamente, $p_{ij}(\mathbf{h}) \neq p_{ji}(\mathbf{h}), \ \forall\, i \neq j$;
5. irreversibilità: se i transiogrammi sono calcolati unidirezionalmente, $p_{ij}(\mathbf{h}) \neq p_{ij}(-\mathbf{h}), \ \forall\, i \neq j$.

Anche nel caso dei transiogrammi è possibile e spesso consigliabile considerare situazioni anisotrope, anche se rimane possibile stimare transiogrammi multidirezionali e omnidirezionali.

2.3.1 Transiogramma teorico

I transiogrammi teorici, pur non fornendo informazioni dettagliate come quelli ricavati dai dati reali, si rendono comunque utili nell'analisi dei dati spaziali categoriali grazie alla facilità interpretativa e di calcolo. Le correlazioni tra classi, infatti, sono facilmente rilevabili dalla sola osservazione di questi diagrammi che, in qualche modo, rappresentano una versione stilizzata e regolare dei transiogrammi empirici.
Il metodo più comune utilizzato per il calcolo dei transiogrammi teorici si avvale della matrice delle probabilità di transizione e della proprietà di Markov di primo ordine. Tale proprietà afferma che la distribuzione condizionata di uno stato futuro dati gli stati passati e quello presente è indipendente dalla storia passata e dipende esclusivamente dallo stato attuale. Formalmente (Li 2007b):

$$Pr[Z(m+1) = \alpha | Z(m) = \beta, Z(m-1) = \gamma, ..., Z(0) = \omega] =$$
$$= Pr[Z(m+1) = \alpha | Z(m) = \beta] = p_{\beta\alpha},$$

dove $Z(0), ..., Z(m+1)$ è una sequenza di stati spaziali della variabile Z; α, β, γ e ω sono stati della variabile casuale nello spazio degli stati; $p_{\beta\alpha}$ è la probabilità di transizione dallo stato β allo stato α.
Una catena markoviana di primo ordine può essere rappresentata da una matrice TPM. Per esempio, a un processo a tre stati può essere associata una matrice come la seguente:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}.$$

Se n e m rappresentano due distanze in una sequenza spaziale e $m < n$, la proprietà di cui sopra può essere espressa nei termini dell'*equazione di Chapman-Kolmogorov per processi omogenei*:

$$p_{ik}(n) = \sum_j [p_{ij}(m)p_{jk}(n-m)], \tag{2.6}$$

dove, ancora una volta, $p_{ik}(n)$ è la probabilità di transizione dallo stato i allo stato k ad una distanza n. Con le probabilità espresse nella matrice TPM, l'equazione di Chapman-Kolmogorov si può riscrivere come segue:

$$\mathbf{P}(n) = \mathbf{P}(m)\mathbf{P}(n-m), \tag{2.7}$$

e in uno spazio in cui n e m sono "step" spaziali e imponendo $m = 1$ si perviene a:

$$\mathbf{P}(n) = \mathbf{P}(1)\mathbf{P}(n-1). \tag{2.8}$$

Applicando ricorsivamente la (2.8), si ottiene una forma ancora più compatta:

$$\mathbf{P}(n) = \mathbf{P}(1)^n. \tag{2.9}$$

Al crescere di n, quindi della distanza, le probabilità di transizione contenute nella matrice $\mathbf{P}(n)$ diventano stazionarie; in questo modo, inoltre, viene generato un diagramma che descirve le probabilità di transizione da ogni stato ad un altro al variare del lag spaziale $\mathbf{h}$, diciamo $p_{ij}(\mathbf{h})$. Se la matrice $\mathbf{P}(1)$ viene stimata su un'area

sufficientemente grande in rapporto alla grandezza "media" dei poligoni che compongono la mappa, il transiogramma che ne consegue avrà una soglia uguale alla proporzione della classe di coda j.
I transiogrammi teorici possiedono delle proprietà specifiche, ma che differiscono tra autotransiogrammi e cross-transiogrammi. Un autotransiogramma teorico, indicato con $p_{ii}(\mathbf{h})$, ha origine nel punto (0,1) e decresce gradualmente all'aumentare del lag spaziale fino al raggiungimento della soglia c_i. In altre parole:

$$\lim_{\mathbf{h}\to\infty} p_{ii}(\mathbf{h}) = c_i = p_i. \tag{2.10}$$

Similmente a quanto detto nella sezione §3.1, il valore in corrispondenza del quale l'autotransiogramma si stabilizza è detto "range di autocorrelazione", indicato con a_i.
Un cross-transiogramma teorico, indicato con $p_{ij}(\mathbf{h})$, ha origine nel punto (0,0) e cresce progressivamente fino allo stabilizzarsi al valore soglia c_{ij}. Come in precedenza, quindi:

$$\lim_{\mathbf{h}\to\infty} p_{ij}(\mathbf{h}) = c_{ij} = p_j. \tag{2.11}$$

Il lag spaziale in corrispondenza del quale il cross-transiogramma si stabilizza è detto "range di cross-correlazione", indicato con a_{ij}, valore che rappresenta la distanza entro la quale due classi sono correlate.
Per quanto riguarda l'autotransiogramma, è possibile ricostruire la dimensione media del poligono della relativa classe tracciando la tangente al grafico passante per il punto (0,1) fino all'asse delle ascisse e verificandone il punto di intersezione.
In conclusione, va posto in evidenza il rischio di incorrere in "effetti di confine" nel caso in cui i transiogrammi fossero calcolati a partire da aree ridotte. Se una classe è più frequentemente vicina ai bordi della suddetta area, è probabile che i poligoni che la compongono siano incompleti con la conseguente imprecisione nella stima delle probabilità di transizione. Il risultato pratico è una rilevazione errata, tipicamente più bassa, del numero di transizioni che coinvolgono quella classe rispetto all'effettivo.

2.3.2 Transiogramma empirico

Fino a questo momento si è fatto riferimento esclusivamente a catene markoviane di primo ordine. Le catene markoviane di ordine superiore al primo sono rappresentazioni di processi non-markoviani, ovvero di processi per i quali le probabilità di transizione dipendono da porzioni più consistenti della "storia" del processo, non limitandosi allo stato più prossimo. Nelle applicazioni spaziali, un processo non-markoviano si traduce in situazioni in cui lo stato spaziale attuale non dipende solo dagli stati ad esso adiacenti, ma anche da quelli non adiacenti in una sequenza spaziale di stati.
In questo contento, i processi markoviani di primo ordine vengono ipotizzati per la loro semplicità e non per le effettive proprietà dei dati. Si può dire più correttamente che, se i transiogrammi calcolati a partire dai dati reali sono molto simili a quelli teorici, allora i dati sono "consistenti" con l'ipotesi di markovianità.
Ad ogni modo, la differenza sostanziale con i transiogrammi teorici risiede nella capacità, da parte dei transiogrammi empirici, di cogliere maggiormente caratteristiche e peculiarità della dipendenza spaziale.
I transiogrammi empirici, a loro volta, possono essere ricondotti a due categorie: *transiogrammi esaustivi* e *transiogrammi sperimentali*.
I transiogrammi esaustivi sono, sinteticamente, delle rappresentazioni semplificate della realtà spaziale di una determinata area, per quanto accurate e complete di numerose informazioni dettagliate; la rappresentazione rimane semplificata perché una resa perfetta è impraticabile a causa dell'impossibilità di campionare un'area in ogni punto (in qualche modo, *censirla*). Tali modelli vengono usati per estrarre informazioni utili sulle relazioni tra le classi del territorio (nel contesto geospaziale a cui ci stiamo riferendo) o per ipotizzare strutture e *pattern* in zone dalla simile conformazione.
I transiogrammi empirici, invece, vengono costruiti a partire dalla rilevazione di pochi dati in un'area dalle grandi dimensioni. Il diagramma, in questo caso, è formato da punti sparsi che non forniscono la stessa quantità di informazione dei transiogrammi esaustivi. La soluzione, come nel caso dei variogrammi trattati in precedenza, è la stima di un modello tramite interpolazione. La stima dei transiogrammi può essere effettuata contando il numero di transizioni da uno stato (diciamo i) ad ogni altro (diciamo j) al

variare del lag **h**. La formula è la seguente:

$$p_{ij}(\mathbf{h}) = \frac{F_{ij}}{\sum_{j=1}^{n} F_{ij}(\mathbf{h})}, \tag{2.12}$$

dove n è il numero di stati possibili per la variabile categoriale di interesse. Quello dei transiogrammi empirici rappresenta, naturalmente, il caso più frequente.
Per quanto concerne la modellizzazione a partire dalla nuvola di punti descritta poc'anzi (come anche descritto riferendosi al variogramma), la scelta della funzione più appropriata e la specificazione dei parametri che la governano sono aspetti cruciali; diverse funzioni, infatti, sia in termini di forma che di valore dei parametri, ipotizzano e comunicano dipendenze ed eterogeneità spaziali diverse. Una tabella riassuntiva dei modelli utilizzati più frequentemente è riportata di seguito (Li 2007b):

Modello	**Funzione**
per autotransiogrammi	
Lineare	$p_{ii}(\mathbf{h}) = 1 - (1 - p_i)\frac{\mathbf{h}}{a_i},\ \mathbf{h} < a_i$ $p_{ii}(\mathbf{h}) = p_i,\ \mathbf{h} \geqslant a_i$
Sferico	$p_{ii}(\mathbf{h}) = 1 - (1 - p_i)[1.5(\frac{\mathbf{h}}{a_i}) - 0.5(\frac{\mathbf{h}}{a_i})^3],\ \mathbf{h} < a_i$ $p_{ii}(\mathbf{h}) = p_i,\ \mathbf{h} \geqslant a_i$
Esponenziale	$p_{ii}(\mathbf{h}) = 1 - (1 - p_i)[1 - \exp(-3\frac{\mathbf{h}}{a_i})]$
Gaussiano	$p_{ii}(\mathbf{h}) = 1 - (1 - p_i)\{1 - \exp[-3(\frac{\mathbf{h}}{a_i})^2]\}$
Coseno-Esponenziale	$p_{ii}(\mathbf{h}) = 1 - (1 - p_i)[1 - \exp(-3\frac{\mathbf{h}}{a_i})\cos(b\mathbf{h})]$
Coseno-Gaussiano	$p_{ii}(\mathbf{h}) = 1 - (1 - p_i)\{1 - \exp[-3(\frac{\mathbf{h}}{a_i})^2]\cos(b\mathbf{h})\}$
per cross-transiogrammi	
Lineare	$p_{ij}(\mathbf{h}) = p_j\frac{\mathbf{h}}{a_{ij}},\ \mathbf{h} < a_{ij}$ $p_{ij}(\mathbf{h}) = p_j,\ \mathbf{h} \geqslant a_{ij}$
Sferico	$p_{ij}(\mathbf{h}) = p_j[1.5(\frac{\mathbf{h}}{a_{ij}}) - 0.5(\frac{\mathbf{h}}{a_{ij}})^3],\ \mathbf{h} < a_{ij}$ $p_{ij}(\mathbf{h}) = p_j,\ \mathbf{h} \geqslant a_{ij}$
Esponenziale	$p_{ij}(\mathbf{h}) = p_j[1 - \exp(-3\frac{\mathbf{h}}{a_{ij}})]$
Gaussiano	$p_{ij}(\mathbf{h}) = p_j\{1 - \exp[-3(\frac{\mathbf{h}}{a_{ij}})^2]\}$
Coseno-Esponenziale	$p_{ij}(\mathbf{h}) = p_j[1 - \exp(-3\frac{\mathbf{h}}{a_{ij}})\cos(b\mathbf{h})]$
Coseno-Gaussiano	$p_{ij}(\mathbf{h}) = p_j\{1 - \exp[-3(\frac{\mathbf{h}}{a_{ij}})^2]\cos(b\mathbf{h})\}$

Tabella 2.1: alcune funzioni matematiche utilizzate per la modellazione di autotransiogrammi e cross-transiogrammi. Legenda: a_i: range di autocorrelazione; a_{ij}: range di cross-correlazione; p_i: proporzione della classe i-esima; $b = 2\pi/\lambda$; λ: lunghezza d'onda della funzione coseno.

Si noti come, a differenza di quanto valido per i variogrammi, l'effetto nugget non venga incluso. I transiogrammi calcolati in contesti multi-classe, infatti, come detto in

precedenza, non prevedono tale effetto.
Il limite dei modelli matematici riportati in tabella, per quanto capaci di descrivere soddisfacentemente un gran numero di situazioni, è quello di non essere in grado di interpretare alcuni comportamenti peculiari riscontrabili in certe situazioni. Questo problema si verifica particolarmente nel caso di transiogrammi empirici con un picco o evidenti effetti non markoviani.

3 Modelli per dati categoriali

Si consideri nuovamente il processo spaziale definito nel capitolo 2, ovvero:

$$\{Z(\mathbf{s}) : \mathbf{s} \in D \supset \mathbb{R}^d\}, \tag{3.1}$$

dove Z è la variabile categoriale di interesse che può assumere K valori, ognuno corrispondente ad una delle classi mutuamente esclusive.
Per modellare il processo spaziale $Z(\mathbf{s})$ si ricorre alla stima della probabilità di realizzazione di una classe per il punto avente coordinate $\mathbf{s}_0$, condizionatamente ai punti campionati. In altre parole si cerca una stima per la seguente quantità:

$$Pr\{Z(\mathbf{s}_0) \,|\, z(\mathbf{s}_1), ..., z(\mathbf{s}_N)\}. \tag{3.2}$$

I metodi e i modelli utilizzati al giorno d'oggi possono essere raggruppati in tre classi (Cao 2016): approcci basati su variabili latenti, geostatistica multipunto e approcci a integrazione probabilistica. Nel seguito i primi due verranno presentati brevemente lasciando maggior spazio al terzo approccio, più aderente alla presente trattazione.

3.1 Metodi basati su variabili latenti

Uno dei possibili modi di considerare i dati spaziali categoriali è quello di vederli come "il risultato di discontinuità nei processi fisici continui sottostanti (latenti, inosservabili)" (Cao 2016). Questa classe di metodi fa uso di variabili latenti correlate per tener conto degli effetti spaziali. Il fulcro di questi approcci risiede nel fare inferenza sulle variabili latenti a partire dai dati categoriali disponibili; le variabili latenti, infatti, non possono

essere osservate direttamente.
Ad avvalersi di tali variabili vi sono diversi metodi, tra cui la simulazione gaussiana troncata (TGS - *Truncated Gaussian Simulation*) e i modelli lineari generalizzati ad effetti misti (GLMM - *Generalized Linear Mixed Models*).

3.1.1 Simulazione gaussiana troncata

Utilizzando il metodo TGS, le diverse aree (differenziate per classe di appartenenza) non vengono simulate direttamente; dapprima, infatti, si simula da una distribuzione Normale e, in seguito, questa viene trasformata nella variabile categoriale mediante l'utilizzo di soglie.
Definendo $Q(\mathbf{s})$ la realizzazione di una variabile casuale Normale e $\mathbb{1}_{F_1}(\mathbf{s})$, la trasformazione di $Q(\mathbf{s})$ in variabile categoriale (2 classi) è rappresentabile come segue (Armstrong et al. 2011):

$$\mathbb{1}_{F_1}(\mathbf{s}) = 1 \iff -\infty \leqslant Q(\mathbf{s}) < \mathrm{t}_1,$$
$$\mathbb{1}_{F_2}(\mathbf{s}) = 1 \iff \mathrm{t}_1 \leqslant Q(\mathbf{s}) < \infty,$$

dove $\mathbf{s}$ è l'insieme di coordinate che individuano un'osservazione nello spazio. È possibile rappresentare più di due classi attraverso la specificazione di molteplici soglie. In particolare, per rappresentare K classi, è necessario specificare $K-1$ soglie $(\mathrm{t}_1, \mathrm{t}_2, ..., \mathrm{t}_{\mathrm{K}-1})$.
Il metodo appena descritto si rivela adatto alla trattazione di variabili categoriali ordinali, ma può essere generalizzato a variabili categoriali sconnesse per consentire la specificazione di adiacenze tra classi più complesse. Per farlo, si ricorre alla cosiddetta *simulazione plurigaussiana troncata* (TPS) che, in breve, consiste nel simulare molteplici campi gaussiani, possibilmente correlati. In questi casi, il calcolo delle soglie richiede la definizione di una partizione delle *facies*, ovvero una rappresentazione schematica in N dimensioni dei rapporti che intercorrono tra le classi. Ad esempio, nel caso in cui si simulino due campi gaussiani, tale partizione si riduce ad un quadrato composto da K rettangoli N-dimensionali, ciascuno rappresentante una classe; i

rettangoli adiacenti definiscono le classi che possono essere confinanti nella simulazione.

3.1.2 Modelli Lineari Generalizzati ad Effetti Misti

Un'altra classe di modelli che si avvale di campi gaussiani casuali è quella dei modelli lineari generalizzati ad effetti misti. Essi rappresentano un'estensione dei tradizionali modelli lineari generalizzati; tale estensione si ottiene attraverso l'introduzione di variabili latenti non osservabili e la specificazione della relazione che intercorre tra la variabile di interesse e il campo gaussiano attraverso un'opportuna funzione di legame. Innanzitutto si assume che la variabile di interesse segua una distribuzione Multinomiale con K classi, ovvero:

$$Z(\mathbf{s}) \sim Mn(1, \pi(\mathbf{s})), \tag{3.3}$$

dove $\boldsymbol{\pi}(\mathbf{s}) = (\pi_1(\mathbf{s}), \ldots, \pi_K(\mathbf{s}))^T$ è il vettore delle probabilità marginali corrispondenti ad ognuna delle K possibili classi e vale che $\sum_{k=1}^{K} \pi_k(\mathbf{s})$.
Per modellare variabili categoriali assumendo che le rispettive classi siano indipendenti, una scelta appropriata risulta essere il modello lineare generalizzato (GLM) con funzione di legame logistica multinomiale. L'introduzione discussa nel presente paragrafo prevede l'inserimento di K variabili latenti $u(\mathbf{s}, k), k = 1, \ldots, K$ tra loro indipendenti. La generica funzione di legame relativa al punto i-esimo di coordinate $\mathbf{s}$ diventa (Cao, Kyriakidis e Goodchild 2011):

$$log\frac{Pr(Z(\mathbf{s}_i) = k)}{Pr(Z(\mathbf{s}_i) = k^*)} = \beta_0^k + u(\mathbf{s}_i, k), \tag{3.4}$$

dove β_0^k è il predittore lineare (nel caso specifico, solo l'intercetta) dipendente dalla classe k considerata, mentre k^* è la classe di riferimento scelta arbitrariamente e comunemente detta "baseline". Si assume, inoltre, che
$\mathbf{u}(\mathbf{s}, k) = (u_1(\mathbf{s}_1, k), \ldots, u_K(\mathbf{s}_N, k))^T$ sia un campo casuale gaussiano (*Gaussian Random Field*, GRF) specificato attraverso una funzione di media e una funzione di covarianza definita positiva. Vale la pena sottolineare come i K campi casuali gaussiani siano indipendenti, fatto che si traduce nella nullità della funzione di covarianza nel caso in cui $k \neq k^{'}$, con $k^{'}$ generica classe diversa da k.

Applicando il teorema di Bayes si ottiene la funzione di probabilità di appartenenza ad una determinata classe k, cioè:

$$Pr(Z(\mathbf{s}_i) = k|\mathbf{u}(\mathbf{s}_i)) = \frac{\exp(\beta_0^k + u(\mathbf{s}_i, k))}{\sum_{k'=1}^{K} \exp(\beta_0^k + u(\mathbf{s}_i, k))}, \tag{3.5}$$

detta "funzione soft-max". A partire da questa è possibile ottenere la funzione di probabilità predittiva per il generico punto di interesse di coordinate $\mathbf{s}_0$.
Il ricorso all'inferenza bayesiana si rende necessario a causa della difficoltà nella specificazione delle funzioni di verosimiglianza e dell'elevato costo computazionale.

3.2 Geostatistica multipunto

La classe di approcci che risponde al nome di "geostatistica multipunto" si propone di modellare la dipendenza spaziale e di simulare le relative superfici attraverso il condizionamento a molteplici punti campionati nello spazio e il riconoscimento di pattern spaziali individuati a priori.

3.2.1 Metodo "*di Strebelle*"

Al metodo a cui si sta facendo riferimento non è riferito alcun nome particolare; per questo, nel corso della presente trattazione, vi ci si riferirà con la dicitura di "metodo di Strebelle", facendo seguito al lavoro pubblicato dallo stesso Strebelle nel 2002 (Strebelle 2002).
L'idea del metodo è quella di utilizzare immagini predefinite, dette *di training*, che riproducono determinati pattern spaziali al fine di riconoscerli nei dati campionati. Questo genere di approccio (come tutti i metodi basati sulla geostatistica multipunto) consente di non specificare la forma di un campo casuale sottostante e permette di alleviarne il costo computazionale associato; inoltre, il metodo consente maggior flessibilità ed efficienza.

3.2.2 Cumulanti spaziali

Alternativamente è stato proposto un metodo basato sui cumulanti, ovvero statistiche di ordine maggiore che combinano l'informazione portata dai momenti. Nella fattispecie, i cumulanti sono proprio combinazioni di momenti, definiti in modo tale da fornire informazioni via via più dettagliate relativamente alla distribuzione di riferimento (Dimitrakopoulos, Mustapha e Gloaguen 2010).
Anche questo metodo si fonda sul riconoscimento di pattern spaziali stilizzati che rappresentano, in uno spazio n-dimensionale, i cumulanti stessi.

3.3 Metodi di integrazione probabilistica

Poiché nel contesto della geostatistica multipunto, soprattutto in presenza di grosse moli di dati, è frequente incorrere in problemi sia di complessità che computazionali nella stima della probabilità condizionata, i cosiddetti "approcci a integrazione probabilistica" propongono la scomposizione di tale condizionamento in una combinazione di misure di dipendenza spaziale che coinvolgono due o tre punti alla volta. Nello specifico, con una notazione simile a quella utilizzata da Cao (2016), la relazione che intercorre tra il punto di coordinate $\mathbf{s}_0$ e uno qualsiasi dei suoi vicini avente coordinate $\mathbf{s}_i$ può essere espressa nei termini di una misura di continuità spaziale $\delta(Z(\mathbf{s}_0), Z(\mathbf{s}_i))$. Formalmente:

$$Pr\{Z(\mathbf{s}_0) = k \,|\, z(\mathbf{s}_1), ..., z(\mathbf{s}_N)\} = f(\delta(Z(\mathbf{s}_0) = k, z(\mathbf{s}_1)), ..., \delta(Z(\mathbf{s}_0) = k, z(\mathbf{s}_N))).$$

La scelta della misura δ, naturalmente, ricopre un ruolo importante nel presente contesto. Essa, solitamente, quantifica la similarità tra le coppie di punti considerate; per questo motivo vengono spesso scelti strumenti quali il variogramma (nel caso di variabili continue) e il transiogramma (nel caso di variabili categoriali).
Una delle specificazioni più comuni per la specificazione della funzione f è rappresentata dalla combinazione lineare delle diverse misure δ, nello specifico:

$$Pr\{Z(\mathbf{s}_0) = k \,|\, z(\mathbf{s}_1), ..., z(\mathbf{s}_N)\} = \sum_{i=1}^{N} \lambda_i \, \delta\{Z(\mathbf{s}_0) = k, z(\mathbf{s}_i)\}, \tag{3.6}$$

dove i λ_i fungono da pesi per le misure δ in modo da tener conto della dipendenza spaziale tra un punto e suoi vicini più prossimi. Modelli per dati categoriali che sfruttano questo genere di struttura sono l'*indicator kriging* e la sua estensione, ovvero l'*indicator cokriging*. Alternativamente può essere specificata una forma moltiplicativa:

$$Pr\{Z(\mathbf{s}_0) = k \,|\, z(\mathbf{s}_1), ..., z(\mathbf{s}_N)\} = \prod_{i=1}^{N} \delta\{Z(\mathbf{s}_0) = k, z(\mathbf{s}_i)\}^{\omega_i}, \tag{3.7}$$

dove gli ω_i sono pesi con funzione equivalente ai λ_i specificati nella versione addittiva. Attraverso l'applicazione di una funzione monotona quale il logaritmo, infatti, è possibile riscrivere la (3.7) nella forma (3.6):

$$\log(Pr\{Z(\mathbf{s}_0) = k \,|\, z(\mathbf{s}_1), ..., z(\mathbf{s}_N)\}) = \sum_{i=1}^{N} \omega_i \log(\delta\{Z(\mathbf{s}_0) = k, z(\mathbf{s}_i)\}). \tag{3.8}$$

Questo tipo di formulazione si riscontra frequentemente come risultato della fattorizzazione di Bayes e trova applicazioni nei metodi basati sui cosiddetti *Markov Chain Random Fields* o sulla massima entropia bayesiana.

3.3.1 Kriging, Cokriging e Indicator Kriging

Nella trattazione di variabili categoriali, l'approccio di analisi tradizionalmente più utilizzato è l'*indicator kriging*. Prima di esplorare il metodo nel dettaglio, si rende necessaria una breve introduzione al kriging semplice, ovvero al metodo da cui discendono tutte le varianti che ne portano il nome.

3.3.1.1 Kriging

Il *kriging* è un metodo di interpolazione spaziale che nasce dal lavoro svolto da Georges Matheron nel suo "Trattato di geostatistica applicata" (Matheron 1962) e prende il nome dall'ingegnere minerario sudafricano Danie Krige.
Le diverse specificazioni di tale metodo sono molteplici, ma nel presente paragrafo ci si concentrerà brevemente solo sul *kriging semplice*, il *kriging ordinario* e il *kriging universale*.

Il *kriging semplice* si basa su due importanti assunzioni semplificatrici (Wackernagel 2003):

- la media μ del processo $Z(\mathbf{s})$ è da considerarsi nota e costante sull'intero dominio, ovvero $\mu = \mathrm{E}[Z(\mathbf{s})], \forall\, \mathbf{s} \in D$;
- il processo $Z(\mathbf{s})$ viene assunto debolmente stazionario (non è necessario che sia gaussiano) con funzione di covarianza $C(\mathbf{h})$ nota e dipendente esclusivamente dalla distanza $\mathbf{h}$ che separa due punti qualsiasi.

L'assunzione sulla media consente di modellare esclusivamente gli scarti da questa, garantendo stime più precise. Va altresì notato che il caso di media nota e costante sia spesso irrealistico o non assumibile a priori.
In questo caso, la forma del predittore $Z^*(\mathbf{s})$ per il processo $Z(\mathbf{s})$ in un punto non campionato $\mathbf{s}_0$ è la seguente:

$$Z^*(\mathbf{s}_0) = \mu + \sum_{i=1}^{N} \lambda_i \left(Z(\mathbf{s}_i) - \mu\right). \tag{3.9}$$

Con la media nota e costante, il predittore $Z^*(\mathbf{s}_0)$ soddisfa automaticamente la proprietà di non distorsione e di conseguenza non è necessario richiedere che i pesi λ_i per le differenti osservazioni sommino a 1:

$$\mathrm{E}[Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)] = \mu + \sum_{i=1}^{N} \lambda_i \left(\mathrm{E}[Z(\mathbf{s}_i)] - \mu\right) - \mathrm{E}[Z(\mathbf{s}_0)] = \mu - \mu = 0. \tag{3.10}$$

Dal momento che lo stimatore risulta non distorto, la varianza dell'errore di stima, definito come $Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)$ può essere espressa semplicemente come:

$$\begin{aligned} \sigma_\varepsilon^2 &= Var(Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)) = \mathrm{E}[(Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0))^2] = \\ &= \mathrm{E}[(Z^*(\mathbf{s}_0))^2 + (Z(\mathbf{s}_0))^2 - 2Z^*(\mathbf{s}_0)Z(\mathbf{s}_0)] = \\ &= -\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i\, \lambda_j\, \gamma(\mathbf{h}_{ij}) - \gamma(\mathbf{0}) + 2\sum_{i=1}^{N} \lambda_i\, \gamma(\mathbf{h}_{i0}), \end{aligned} \tag{3.11}$$

$\forall\, j = 1, \ldots, N$, e dove $\gamma(\mathbf{h}_{ij})$ è il valore assunto dal variogramma in corrispondenza della distanza $\mathbf{h}$ che intercorre tra i punti i e j. Poiché si è interessati a minimizzare

tale quantità, si procede calcolando le derivate prime rispetto ai pesi λ_i e ponendole uguali a zero. Si ottiene:

$$\sum_{j=1}^{N} \lambda_j \, \gamma(\mathbf{h}_{i0}) = \gamma(\mathbf{h}_{i0}), \ \forall \, i = 1, \ldots, N. \tag{3.12}$$

Il sistema di equazioni lineari associato consente di trovare univocamente i pesi ottimali:

$$\begin{bmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \\ \vdots \\ \hat{\lambda}_N \end{bmatrix} = \begin{bmatrix} \gamma(\mathbf{h}_{11}) & \gamma(\mathbf{h}_{12}) & \ldots & \gamma(\mathbf{h}_{1N}) \\ \gamma(\mathbf{h}_{21}) & \gamma(\mathbf{h}_{22}) & \ldots & \gamma(\mathbf{h}_{2N}) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(\mathbf{h}_{N1}) & \gamma(\mathbf{h}_{N2}) & \ldots & \gamma(\mathbf{h}_{NN}) \end{bmatrix}^{-1} \begin{bmatrix} \gamma(\mathbf{h}_{10}) \\ \gamma(\mathbf{h}_{20}) \\ \vdots \\ \gamma(\mathbf{h}_{N0}) \end{bmatrix}. \tag{3.13}$$

Un'altra quantità di interesse è la varianza di stima σ^2_{SK}, calcolabile in ogni punto di previsione $\mathbf{s}_0$:

$$\sigma^2_{\mathrm{SK}}(\mathbf{s}_0) = -\gamma(\mathbf{0}) + \sum_{i=1}^{N} \hat{\lambda}_i \, \gamma(\mathbf{h}_{i0}). \tag{3.14}$$

Essa si rivela particolarmente utile nei casi in cui i punti campionati sono irregolarmente distribuiti nello spazio. Grazie ad essa, infatti, è possibile costruire intervalli di confidenza e delle "mappe di variabilità di kriging" che forniscono un'indicazione sulla precisione delle stime.
Più realisticamente, in molte circostanze, non è possibile assumere note le componenti del processo spaziale $Z(\mathbf{s})$. È in questo contesto che le assunzioni semplificatrici del kriging semplice vengono abbandonate in favore del *kriging ordinario*. Le assunzioni su cui si fonda il metodo sono le seguenti:

- la media μ del processo $Z(\mathbf{s})$ è costante ma ignota;

- il processo $Z(\mathbf{s})$ viene ipotizzato intrinsecamente stazionario con variogramma $\gamma(\mathbf{h})$.

La condizione di non distorsione, in questo caso, viene garantita imponendo un vincolo

di somma unitaria sui pesi λ_i, ovvero $\sum_{i=1}^{N} \lambda_i = 1$:

$$\begin{aligned} \mathrm{E}[Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)] &= \mathrm{E}\left[\sum_{i=1}^{N} \lambda_i\, Z(\mathbf{s}_i) - Z(\mathbf{s}_0) \sum_{i=1}^{N} \lambda_i\right] = \\ &= \sum_{i=1}^{N} \lambda_i\, \mathrm{E}[Z(\mathbf{s}_i) - Z(\mathbf{s}_0)] = 0. \end{aligned} \tag{3.15}$$

Ancora una volta, minimizzando la varianza dell'errore di stima, si perviene al sistema di equazioni lineari associate al kriging ordinario. In forma matriciale:

$$\begin{bmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \\ \vdots \\ \hat{\lambda}_N \\ \hat{\mu} \end{bmatrix} = \begin{bmatrix} \gamma(\mathbf{h}_{11}) & \gamma(\mathbf{h}_{12}) & \dots & \gamma(\mathbf{h}_{1N}) & 1 \\ \gamma(\mathbf{h}_{21}) & \gamma(\mathbf{h}_{22}) & \dots & \gamma(\mathbf{h}_{2N}) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(\mathbf{h}_{N1}) & \gamma(\mathbf{h}_{N2}) & \dots & \gamma(\mathbf{h}_{NN}) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma(\mathbf{h}_{10}) \\ \gamma(\mathbf{h}_{20}) \\ \vdots \\ \gamma(\mathbf{h}_{N0}) \\ 1 \end{bmatrix}. \tag{3.16}$$

La varianza di stima σ^2_{OK} del kriging ordinario nel generico punto $\mathbf{s}_0$ diventa:

$$\sigma^2_{\mathrm{OK}}(\mathbf{s}_0) = \hat{\mu} - \gamma(\mathbf{0}) + \sum_{i=1}^{N} \hat{\lambda}_i\, \gamma(\mathbf{h}_{i0}). \tag{3.17}$$

Nel contesto del kriging, un'altra possibilità è rappresentata dalla possibile presenza di una media non stazionaria (non costante) del processo spaziale, ma variabile a seconda dell'ubicazione considerata. Una possibile soluzione è quella di guardare al processo come composto da due parti: una combinazione lineare di funzioni deterministiche atta a descrivere il *trend* non stazionario (indicata con $\mu(\mathbf{s})$) e una parte stocastica relativa ai residui (indicata con $Y(\mathbf{s}) = Z(\mathbf{s}) - \mu(\mathbf{s})$). Schematicamente, le assunzioni alla base del *kriging universale* possono essere riscritte come segue:

- il processo può essere decomposto in una parte deterministica associata al trend non stazionario e in una parte stocastica a rappresentare i residui rispetto al trend, ovvero $Z(\mathbf{s}) = \mu(\mathbf{s}) + Y(\mathbf{s})$;

- $Y(\mathbf{s})$ è da considerarsi come un processo intrinsecamente stazionario avente media nulla ($\mathrm{E}[Y(\mathbf{s})] = 0$) e "variogramma residuale" $\gamma_Y(\mathbf{h})$;

- siano $f_0, f_1, \ldots, f_L$ ($L \in \mathbb{N}$) funzioni deterministiche dipendenti dalle coordinate del dominio spaziale. Si assume che il trend non stazionario $\mu(\mathbf{s})$ sia combinazione lineare di tali funzioni valutate punto per punto, ovvero che:

$$\mu(\mathbf{s}) = \sum_{l=0}^{L} a_l \, f_l(\mathbf{s}),$$

dove L è il numero di funzioni costituenti la combinazione lineare; si supponga anche $f_0(\mathbf{s}) = 1$ costante e siano i coefficienti a_l tutti diversi da 0.

A partire dalla scomposizione della parte deterministica appena descritta, è possibile riscrivere il processo spaziale esplicitando la combinazione lineare tra le funzioni f_l di trend e i relativi coefficienti:

$$Z(\mathbf{s}_i) = \mu(\mathbf{s}_i) + Y(\mathbf{s}_i) = \sum_{l=0}^{L} a_l \, f_l(\mathbf{s}_i) + Y(\mathbf{s}_i), \tag{3.18}$$

e in forma matriciale:

$$\mathbf{Z} = \begin{bmatrix} Z(\mathbf{s}_1) \\ Z(\mathbf{s}_2) \\ \vdots \\ Z(\mathbf{s}_N) \end{bmatrix} = \begin{bmatrix} 1 & f_1(\mathbf{s}_1) & \ldots & f_L(\mathbf{s}_1) \\ 1 & f_1(\mathbf{s}_2) & \ldots & f_L(\mathbf{s}_2) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & f_1(\mathbf{s}_N) & \ldots & f_L(\mathbf{s}_N) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_L \end{bmatrix} + \begin{bmatrix} Y(\mathbf{s}_1) \\ Y(\mathbf{s}_2) \\ \vdots \\ Y(\mathbf{s}_N) \end{bmatrix} = F\mathbf{a} + \mathbf{Y}. \tag{3.19}$$

Si noti come, nel caso in cui $L = 0$, ci si stia riconducendo al caso del kriging ordinario, con la presenza di una media ignota ma costante a_0.
Il predittore del kriging universale per un punto qualsiasi del dominio spaziale può essere scritto in conformità con quanto esplicitato nel caso del kriging semplice e del kriging ordinario, tenendo conto di quanto appena riportato relativamente al trend non costante:

$$Z^*(\mathbf{s}_0) = \sum_{i=1}^{N} \lambda_i \, Z(\mathbf{s}_i) = \sum_{i=1}^{N} \lambda_i \left(\sum_{l=0}^{L} a_l \, f_l(\mathbf{s}_i) + Y(\mathbf{s}_i) \right). \tag{3.20}$$

È agevole dimostrare come, date le assunzione sulla natura deterministica di $\mu(\mathbf{s})$ e sui coefficienti a_l, la non distorsione del predittore sia soddisfatta se e solo se
$\sum_{i=1}^{N} \lambda_i \, f_l(\mathbf{s}_i) = f_l(\mathbf{s}_0), \; l = 0, \ldots, L.$ Poiché nel seguito sarà conveniente esprimere gli

oggetti nelle forme vettoriali e matriciali, lo stesso risultato può essere riscritto come:

$$\boldsymbol{\lambda}^T F = \mathbf{f}_0^T. \tag{3.21}$$

La varianza del termine d'errore ha la seguente formulazione:

$$\sigma_\varepsilon^2 = -\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i\,\lambda_j\,\gamma_Y(\mathbf{h}_{ij}) + 2\sum_{i=1}^{N} \lambda_i\,\gamma_Y(\mathbf{h}_{i0}) = -\boldsymbol{\lambda}^T\Gamma_Y\boldsymbol{\lambda} + 2\boldsymbol{\lambda}^T\gamma_Y^*, \tag{3.22}$$

e può essere minimizzata utilizzando, ancora una volta, il metodo dei moltiplicatori di Lagrange, sotto il vincolo di non distorsione. Il sistema di equazioni lineari associate al kriging universale diventa, quindi:

$$\begin{cases} \sum_{j=1}^{N} \lambda_j\,\gamma(\mathbf{h}_{ij}) + \sum_{l=0}^{L} \omega_l\,f_l(\mathbf{s}_i) = \gamma(\mathbf{h}_{i0}), & i = 1,\dots,N \\ \sum_{j=1}^{N} \lambda_j\,f_l(\mathbf{s}_j) = f_l(\mathbf{s}_0), & l = 0,\dots,L \end{cases} \tag{3.23}$$

riscrivibile in forma matriciale come segue:

$$\begin{cases} \Gamma_Y\boldsymbol{\lambda} + F\boldsymbol{\omega} = \boldsymbol{\gamma}_{Y,0} \\ F^T\boldsymbol{\lambda} = \mathbf{f_0} \end{cases} \iff \begin{bmatrix} \Gamma_Y & F \\ F^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\omega} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma}_{Y,0} \\ \mathbf{f_0} \end{bmatrix}, \tag{3.24}$$

forma da cui è semplice ricavare il vettore di pesi del kriging universale e il vettore dei parametri di Lagrange $\boldsymbol{\omega}$. Si noti che per garantire l'unicità della soluzione del sistema è necessario che la matrice F sia a rango pieno.
Il predittore, a questo punto, può essere espresso nei termini usuali, ovvero come combinazione lineare tra i pesi stimati poc'anzi e il valore assunto dal processo nei punti campionati:

$$Z^*(\mathbf{s}_0) = \sum_{i=1}^{N} \lambda_i\,Z(\mathbf{s}_i). \tag{3.25}$$

La varianza di stima nel caso del kriging universale può essere calcolata mediante la seguente formula:

$$\sigma^2_{\mathrm{UK}} = \sum_{i=1}^{N} \lambda_i\,\gamma_Y(\mathbf{h}_{i0}) + \sum_{l=0}^{L} \omega_l\,f_l(\mathbf{s}_0). \tag{3.26}$$

3.3.1.2 Cokriging

Il *cokriging* consente di incorporare informazioni contenute in variabili ausiliarie. In altre parole, la previsione riguardante la variabile di interezze Z nel punto di coordinate $\mathbf{s_0}$, diciamo $Z_1(\mathbf{s}_0)$, viene calcolata introducendo nel modello variabili che mostrano correlazione con la risposta, indicate nel seguito come $Z_2, \ldots, Z_P$.
In questo caso, lo stimatore assume la seguente forma:

$$Z^*(\mathbf{s}_0) = \sum_{i=1}^{N} \sum_{j=1}^{P} \lambda_{ij}\, Z_j(\mathbf{s}_i), \tag{3.27}$$

sotto i vincoli di $\sum_{i=1}^{N} \lambda_{i1} = 1$ e $\sum_{i=1}^{N} \lambda_{ij} = 0,\ 1 \leqslant j \leqslant P$, per garantire la non distorsione. I pesi $\boldsymbol{\lambda}$ e il vettore di medie $\boldsymbol{\mu}$ relative a ciascuna variabile possono essere calcolati mediante un sistema lineare analogo a quelli visti in precedenza:

$$\begin{bmatrix} \lambda_{11} \\ \vdots \\ \lambda_{N1} \\ \lambda_{12} \\ \vdots \\ \lambda_{NP} \\ \mu_1 \\ \mu_2 \\ \vdots \\ \mu_P \end{bmatrix} = \begin{bmatrix} \Gamma_{11} & \ldots & \Gamma_{1P} & \mathbf{1} & 0 & \ldots & 0 \\ \Gamma_{21} & \ldots & \Gamma_{2P} & 0 & \mathbf{1} & \ddots & \vdots \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & 0 \\ \Gamma_{P1} & \ldots & \Gamma_{PP} & 0 & 0 & \ldots & \mathbf{1} \\ \mathbf{1}^T & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & \mathbf{1}^T & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & \ldots & \mathbf{1}^T & 0 & 0 & \ldots & 0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_{11}(\mathbf{h}_{10}) \\ \vdots \\ \gamma_{11}(\mathbf{h}_{N0}) \\ \gamma_{12}(\mathbf{h}_{10}) \\ \vdots \\ \gamma_{1P}(\mathbf{h}_{N0}) \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{3.28}$$

dove $\mathbf{1}^T = (1, \ldots, 1) \in \mathbb{R}^n$ e le matrici $\Gamma_{ij},\ 1 \leqslant i, j \leqslant P$, sono tutte quadrate di dimensione N e contengono i valori del variogramma calcolato tra tutte le coppie di punti e per tutte le diverse coppie di variabili Z prese in considerazione.

3.3.1.3 Indicator Kriging

L'idea su cui si basa la tecnica denominata *indicator kriging* non consiste nello stimare un solo variogramma sulla base della distribuzione intera della variabile di interesse,

bensì diversi modelli considerando, di volta in volta, soglie diverse a partire dai variogrammi indicatori introdotti nella sezione §{3.2}.
Nel caso più generale, la variabile di interesse viene "divisa" in R soglie che possono essere rappresentate con la notazione seguente: $q \in \{q_1, \ldots, q_R\}$. È in contesti come quello appena descritto che si parla di *multiple indicator kriging* (MIK). Spesso, tali soglie, vengono scelte in corrispondenza dei quantili della distribuzione della variabile risposta; in questo modo si viene a creare un dataset composto da variabili dicotomiche.
Per procedere alla stima, viene applicato un modello di kriging ordinario per ognuna delle soglie imposte. Così facendo si perviene alla stime della funzione di ripartizione per ogni possibile punto di previsione $\mathbf{s}_0$ e per ogni soglia z_r (Cressie 1993):

$$I^*(\mathbf{s}_0; z) = \sum_{i=1}^{N} \lambda_{ir}\, I(\mathbf{s}_i; z_r),\ r = 1, \ldots, R. \tag{3.29}$$

Il sistema lineare che consente di calcolare i pesi λ_{ir} è strutturalmente equivalente a quello visto nel caso del cokriging.
Nel caso più semplice del metodo appena descritto si considera una sola soglia z. In questo contesto, la previsione può essere condotta mediante il seguente stimatore:

$$I^*(\mathbf{s}_0; z) = \sum_{i=1}^{N} \lambda_i\, I(\mathbf{s}_i; z), \tag{3.30}$$

col vincolo di somma a 1 dei pesi λ_i.

3.3.2 Markov Chain Random Field

Un altro approccio che sta trovando crescente consenso nell'analisi dei dati georeferenziati si basa sulle catene e sui campi casuali di Markov.
Alcuni dei metodi basati su tali processi prevedono l'impiego di molteplici (spesso 2 o 3) catene di Markov e l'assunzione di piena indipendenza tra queste. Un problema estremamente frequente e limitante, però, riguarda la sistematica sottostima delle classi a realizzazione meno probabile e conseguentemente la sovrastima delle classi più probabili.
Più recentemente, è stato proposto un metodo che prevede l'impiego di un'unica catena

di Markov e che non necessita dell'assunzione di indipendenza; questo metodo risponde al nome di *Markov Chain Random Field* (MCRF) e si basa su campi casuali markoviani (MRF) (Li 2007a). A differenza dei campi usuali, MCRF è costruito sulla base di un'unica "catena direzionale" in uno spazio avente, in ogni punto, una distribuzione di probabilità condizionata che dipende esclusivamente dai vicini più prossimi nelle diverse direzioni. Tale catena markoviana si può muovere nello spazio di interesse sia casualmente che seguendo percorsi specifici prestabiliti. In altre parole, MCRF può essere definito come una catena di Markov "spaziale" che si muove nel dominio secondo determinate regole probabilistiche di transizione dipendenti dalla direzione; in ogni posizione avente classe ignota, è possibile effettuare una stima basandosi solamente sui vicini più prossimi, aventi classe nota, nelle diverse direzioni.
Utilizzando la notazione illustrata finora, è possibile esprimere la distribuzione condizionata di probabilità di appartenere ad una determinata classe per un generico punto di coordinate $\mathbf{s}_0$:

$$Pr(Z(\mathbf{s}_0)|Z(\mathbf{s}_1),\ldots,Z(\mathbf{s}_N)) = Pr(Z(\mathbf{s}_0)|Z(\mathbf{s}_1),\ldots,Z(\mathbf{s}_m)), \tag{3.31}$$

dove N indicizza i punti campionati e, quindi, noti e m il numero di punti "vicini" considerati.
Utilizzando il teorema di Bayes è possibile decomporre la formula scritta poc'anzi nel seguente prodotto di probabilità condizionate:

$$\begin{aligned} Pr(Z(\mathbf{s}_0)|Z(\mathbf{s}_1),\ldots,Z(\mathbf{s}_m)) &= \frac{Pr(Z(\mathbf{s}_m),\ldots,Z(\mathbf{s}_2),Z(\mathbf{s}_1),Z(\mathbf{s}_0))}{Pr(Z(\mathbf{s}_1),\ldots,Z(\mathbf{s}_m))} = \\ &= c \cdot Pr(Z(\mathbf{s}_m)|Z(\mathbf{s}_0),Z(\mathbf{s}_{m-1}),\ldots,Z(\mathbf{s}_1))\ldots \\ &\ldots Pr(Z(\mathbf{s}_2)|Z(\mathbf{s}_0),Z(\mathbf{s}_1)) \cdot Pr(Z(\mathbf{s}_0)|Z(\mathbf{s}_1)), \end{aligned} \tag{3.32}$$

con

$$c = \frac{Pr(\mathbf{s}_1)}{Pr(Z(\mathbf{s}_1),\ldots,Z(\mathbf{s}_m))}$$

costante poiché non dipende dall'ignota posizone $\mathbf{s}_0$. È possibile semplificare il calcolo facendo ricorso all'assunzione di *indipendenza condizionale*. Questa assume che, dato un punto di coordinate $\mathbf{s}_0$, i suoi vicini più prossimi $\mathbf{s}_1,\ldots,\mathbf{s}_m$ nelle diverse direzioni siano

condizionatamente indipendenti, ovvero:

$$Pr(Z(\mathbf{s}_i)|Z(\mathbf{s}_0),\ldots,Z(\mathbf{s}_m)) = Pr(Z(\mathbf{s}_i)|Z(\mathbf{s}_0)). \tag{3.33}$$

In questo modo vengono considerate solo interazioni tra coppie di punti, ciascuna definita dal punto ignoto e di interesse $\mathbf{s}_0$ e da uno dei punti noti posizionati in prossimità. A partire dall'assunzione di indipendenza condizionale, quindi, l'equazione (3.32) può essere semplificata, ottenendo la forma che segue:

$$\begin{aligned} Pr(Z(\mathbf{s}_0)|Z(\mathbf{s}_1),\ldots,Z(\mathbf{s}_m)) = c \cdot Pr(Z(\mathbf{s}_m)|Z(\mathbf{s}_0)) \ldots \\ \ldots Pr(Z(\mathbf{s}_2)|Z(\mathbf{s}_0)) \cdot Pr(Z(\mathbf{s}_0)|Z(\mathbf{s}_1)). \end{aligned} \tag{3.34}$$

È possibile utilizzare le informazioni fornite dai transiogrammi impiegando, al posto delle probabilità condizionate a coppie di punti, le probabilità di transizione dipendenti dai diversi lag spaziali. La nuova espressione diventa:

$$\begin{aligned} Pr(Z(\mathbf{s}_0) = k_0|Z(\mathbf{s}_1) = k_1,\ldots,Z(\mathbf{s}_m) = k_m) = c \cdot p_{k_0,k_m}(\mathbf{h}_m)\ldots p_{k_0,k_2}(\mathbf{h}_2) \cdot p_{k_1,k_0}(\mathbf{h}_1) = \\ = \frac{\prod_{i=2}^{m} p_{k_0,k_i}(\mathbf{h}_i) \cdot p_{k_1,k_0}(\mathbf{h}_1)}{\sum_{f=1}^{N}(\prod_{i=2}^{m} p_{f,k_i}(\mathbf{h}_i) \cdot p_{k_1,f}(\mathbf{h}_1))}, \end{aligned} \tag{3.35}$$

dove $p_{k_0,k_i}(\mathbf{h}_i)$ è la probabilità di transizione, nella direzione i e per la distanza $\mathbf{h}$, dallo stato k_0 (relativo al punto di interesse di coordinate $\mathbf{s}_i$) allo stato k_i ($i \neq 0$); $\mathbf{s}_1$ è il punto prossimale da cui (o attraverso cui) la catena spaziale si muove per giungere a $\mathbf{s}_0$; k_0, k_i e f sono stati del processo; $\mathbf{h}_i$ è la distanza tra il punto di interesse e il vicino $\mathbf{s}_i$. L'equazione ricavata fornisce l'espressione generale della distribuzione di probabilità condizionata della catena spaziale Z in ogni posizione ignota del campo casuale. La forma esatta della formula dipende dal numero di vicini più prossimi al punto di stima nelle diverse direzioni e alle rispettive distanze.

3.3.3 Massima entropia bayesiana

L'obiettivo dell'approccio a massima entropia bayesiana risulta essere, ancora una volta, la determinazione della probabilità di realizzazione di ciascuna classe in un generico punto di coordinate $\mathbf{s}_0$, relativamente a una variabile categoriale avente natura spaziale.

Nel presente contesto si cerca un'approssimazione della probabilità condizionata che si manifesti una determinata classe nel punto di coordinate $\mathbf{s}_0$, note le categorie negli N punti campionati; tale approssimazione viene ricavata utilizzando solo probabilità univariate e bivariate. La probabilità condizionata può essere riscritta in questo modo:

$$p_{k_0|k_1,\ldots,k_N} = \frac{p_{k_0,k_1,\ldots,k_N}}{p_{k_1,\ldots,k_N}} = \frac{p_{k_0,k_1,\ldots,k_N}}{\sum_{k_0=1}^{K} p_{k_0,k_1,\ldots,k_N}} = \frac{p_{k_0} p_{k_1,\ldots,k_N|k}}{\sum_{k_0=1}^{K} p_{k_0} p_{k_1,\ldots,k_N|k_0}}. \tag{3.36}$$

L'approssimazione viene portata trasformando la probabilità condizionata di cui sopra in un prodotto di probabilità condizionate bivariate attraverso la cosiddetta "assunzione Naïve di Bayes":

$$\overline{p}_{k_1,\ldots,k_N|k_0} = \prod_{i=1}^{N} p_{k_i|k_0}(\mathbf{h}_i), \tag{3.37}$$

ovvero si sta assumendo indipendenza tra gli stati in due punti qualsiasi dello spazio, dato quello nel punto di previsione $\mathbf{s}_0$. La probabilità condizionata approssimata, quindi, diventa:

$$\overline{p}_{k_0|k_1,\ldots,k_N} = \frac{p_{k_0} \prod_{i=1}^{N} p_{k_i|k_0}(\mathbf{h}_i)}{\sum_{k_0=1}^{K} p_{k_0} \prod_{i=1}^{N} p_{k_i|k_0}(\mathbf{h}_i)} = \frac{p_{k_0}^{1-N} \prod_{i=1}^{N} p_{k_0,k_i}(\mathbf{h}_i)}{\sum_{k_0=1}^{K} p_{k_0}^{1-N} \prod_{i=1}^{N} p_{k_0,k_i}(\mathbf{h}_i)}, \tag{3.38}$$

equazione nota come "classificatore Naïve di Bayes".
Per illustrare più dettagliatamente il metodo, si rende necessaria una sintetica introduzione sul concetto di "entropia" in ambito statistico e della teoria dell'informazione. L'entropia di una distribuzione di probabilità P finita avente stati $\mathbf{Y} \in \mathcal{Y}$ viene indicata come:

$$H(P) = \sum_{\mathbf{Y} \in \mathcal{Y}} P(\mathbf{Y}) \ln(P(\mathbf{Y})) = -\mathrm{E}_P[\ln(P(\mathbf{Y}))].$$

La divergenza di Kullback-Leibler (detta anche "entropia relativa") è una nozione fortemente imparentata: essa, infatti, esprime quanto sono differenti due distribuzioni di probabilità.
Il metodo che ricade sotto il nome di *massima entropia bayesiana* (BME), si articola in due fasi che coinvolgono la quantità illustrata poc'anzi:

1. viene trovata la distribuzione di probabilità congiunta $\widetilde{p}_{k_0,k_1,\ldots,k_N}$, soluzione del

principio di massima entropia soggetta ai vincoli univariati e bivariati, ovvero a:

$$\sum_{k_0=1}^{K} \cdots \sum_{k_N=1}^{K} \widetilde{p}_{k_0,k_1\ldots,l_N} \mathbf{1}_{[Y_k=i]} = p_i$$

$$\sum_{k_0=1}^{K} \cdots \sum_{k_N=1}^{K} \widetilde{p}_{k_0,k_1,\ldots,l_N} \mathbf{1}_{[Y_k=i;\, Y_{k'}=j]} = p_{ij}(\mathbf{h}),$$

dove $i, j = 1, \ldots, K$.

2. viene calcolata la probabilità condizionata della categoria k_0 nel punto di coordinate $\mathbf{s}_0$:

$$\widetilde{p}_{k_0|k_1,\ldots,k_N} = \frac{\widetilde{p}_{k_0,k_1,\ldots,k_N}}{\widetilde{p}_{k_1,\ldots,k_N}}.$$

È possibile dimostrare che la distribuzione di probabilità congiunta a massima entropia assume la seguente forma:

$$\widetilde{p}_{k_0,k_1,\ldots,k_N} = \mu \prod_{i \in \{k_0,k_1,\ldots,k_N\}} \lambda_i \prod_{j \in \{k_0,k_1,\ldots,k_{N-1}\}} \prod_{j' \in \{k_{j+1},\ldots,k_N\}} \nu_{j,j'}, \tag{3.39}$$

dove μ, $\boldsymbol{\lambda} = (\lambda_{k_0}, \lambda_{k_1}, \ldots, \lambda_{k_N})$ e $\boldsymbol{\nu} = (\nu_{k,k_1}, \nu_{k,k_2}, \ldots, \nu_{k_{N-1},k_N})$ sono calcolati in modo da corrispondere alle probabilità univariate e bivariate. L'interesse, però, è rivolto alla probabilità condizionata $\widetilde{p}_{k|k_1,\ldots,k_N}$ piuttosto che alla congiunta. Sostituendo l'equazione (2) nella (3.39) si ottiene la seguente forma:

$$\widetilde{p}_{k_0|k_1,\ldots,k_N} = \frac{\lambda_{k_0} \prod_{i=1}^{N} \nu_{k_0,k_i}}{\sum_{k_0=1}^{K} \lambda_{k_0} \prod_{i=1}^{N} \nu_{k_0,k_i}}. \tag{3.40}$$

Poiché il calcolo dei parametri sopracitati si rivela essere, spesso, molto oneroso in termini computazionali, è possibile restringere il problema considerando solo vincoli sulle N probabilità tra il punto di previsione $\mathbf{s}_0$ e i punti campionati. A tale soluzione ci si riferisce col nome di "BME ristretta" (Allard, D'Or e Froidevaux 2011), poiché rappresenta un'approssimazione computazionalmente vantaggiosa della "BME completa". La versione ristretta consente di esplicitare la probabilità condizionata in

forma chiusa:

$$\begin{aligned} \overline{p}_{k_0|k_1,\ldots,k_N} &= \frac{\overline{p}_{k_0,k_1,\ldots,k_N}}{\overline{p}_{k_1,\ldots,k_N}} = \\ &= \frac{p_{k_0}^{1-N} \prod_{i=1}^{N} p_{k_0,k_i}(\mathbf{h}_i)}{\sum_{k_0=1}^{K} p_{k_0}^{1-N} \prod_{i=1}^{N} p_{k_0,k_i}(\mathbf{h}_i)}. \end{aligned} \tag{3.41}$$

Si noti come la soluzione proposta nel metodo BME sia simile per struttura e componenti a quella ottenuta col metodo MCRF.

4 Simulazioni

Nella presente sezione si procederà ad osservare la qualità del modello di ricostruzione spaziale che utilizza le catene di Markov Monte Carlo e la massima entropia bayesiana per effettuare previsioni. I test vengono eseguiti simulando griglie le cui celle assumono la funzione deii pixel delle immagini satellitari prese in considerazione nel capitolo successivo. Le simulazioni vengono condotte a partire da diversi scenari, distinti per numero di categorie considerate, proporzione di dati disponibili e grado di dipendenza spaziale.
Nella prima parte del capitolo viene presentato il modello teorico utilizzato per la simulazione di mappe stilizzate attraverso griglie di pixel e contraddistinte da correlazione spaziale. Successivamente si esplora in dettaglio il metodo simulativo adottato, illustrando qualche esempio e i risultati fondamentali per la valutazione del metodo. Viene anche proposto il confronto con l'algoritmo "Iterated Conditional Modes" (ICM), proposto nella letteratura di settore da Julian Besag (Besag 1986).

4.1 Modello di Potts

Per la simulazione di mappe in griglia che possano risultare di interesse per il presente lavoro è necessario ricorrere a un metodo che consenta di rappresentare la correlazione spaziale che intercorre tra le diverse aree che caratterizzano la zona di interesse.
Per questo scopo è vantaggioso ricorrere ad un modello i cui concetti fondamentali sono mutuati dalla meccanica statistica e dalla termodinamica, ovvero il "modello di Potts". Tale modello consente di modellare la dipendenza spaziale tra le diverse categorie attraverso un campo casuale di Markov, la cui distribuzione viene specificata nei

termini della distribuzione della variabile d'interesse rilevata in ciascuno dei siti $\mathbf{s}_i, i = 1, \ldots, N$ condizionatamente ai valori osservati nel suo intorno (Gaetan, Girardi e Pastres 2017). Lo schema dei vicini facenti parte dell'intorno del punto $\mathbf{s}_0$ può assumere diverse specificazioni, dette anche "ordini". In Figura 4.1 vengono riportati gli ordini più frequentemente considerati:

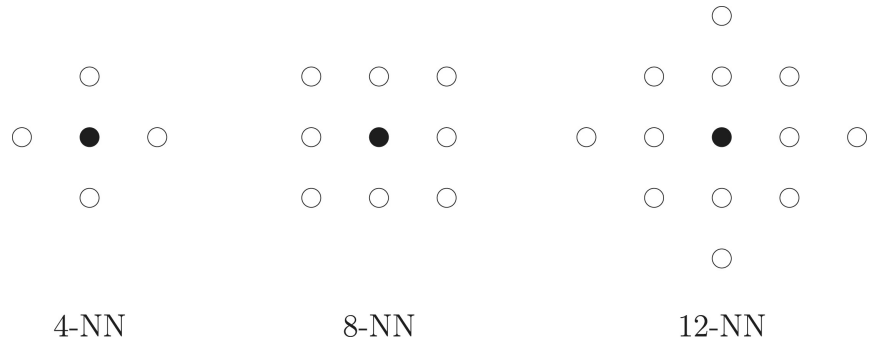

Figura 4.1: schema dell'intorno in campi casuali markoviani di primo, secondo e terzo ordine su griglie bidimensionali regolari. Viene anche indicato il numero di vicini facenti parte dell'intorno specificato nei tre diversi casi (Gaetan, Girardi e Pastres 2017).

Dalla figura si può notare come nei casi di MRF (*Markov Random Fields*) di primo ordine vengano considerati solo i pixel adiacenti a quello di interesse nelle quattro direzioni cardinali; nei campi di Markov di secondo ordine si considerano anche i pixel confinanti diagonalmente; i MRF di terzo ordine aggiungono nei vicini considerati altri 4 pixel, uno per ciascuna delle direzioni nord, est, sud e ovest. Nel prosieguo, per semplicità di notazione, l'insieme dei punti vicini a quello generico di coordinate $\mathbf{s}_i$ verrà indicato con $\partial\mathbf{s}_i = \{l : i \sim j\}$, dove con $i \sim j$ si indica la vicinanza dei punti di coordinate $\mathbf{s}_i$ e $\mathbf{s}_j, i \neq j$ (Gaetan, Girardi e Pastres 2017).
A questo punto è possibile indicare formalmente la distribuzione di probabilità condizionata a cui si faceva riferimento in precedenza per il punto di interesse di coordinate $\mathbf{s}_0$:

$$Pr(Z(\mathbf{s}_0) = k \,|\, z(\partial\mathbf{s}_0) = k_{\partial\mathbf{s}_0}), \tag{4.1}$$

dove $k_{\partial\mathbf{s}_0}$ indica le classi note dei punti osservati nell'intorno del punto $\mathbf{s}_0$.
Il modello di Potts viene definito come segue:

$$Pr(Z(\mathbf{s}_0) = k \,|\, z(\partial\mathbf{s}_0) = k_{\partial\mathbf{s}_0}) = \frac{\exp\{\beta v_{0,k}\}}{\sum_{k=1}^{K} \exp\{\beta v_{0,k}\}}, \tag{4.2}$$

dove $v_{0,k} = \sum_{j \in \partial \mathbf{s}_i} I(z(\mathbf{s}_j) = k)$ e β è un parametro di regolarizzazione che misura il grado di dipendenza spaziale. In particolare, in meccanica statistica, esso è inversamente proporzionale alla "temperatura assoluta": maggiore è il valore di β, minore sarà la "temperatura" e, di conseguenza, il sistema manifesterà maggior stabilità; l'opposto vale per valori decrescenti del parametro. La "maggior stabilità" a cui si fa riferimento, nel presente contesto si traduce in aree più estese e definite che manifestano la medesima classe. Viceversa, un valore basso di β implica alta temperatura e alto disordine (Besag 1986).

4.2 Modello "Multinomial Categorical Simulation"

Il modello che sfrutta le probabilità di transizione tra stati calcolate mediante la massima entropia bayesiana viene illustrato e proposto in un articolo di Allard, D'Or e Froidevaux (Allard, D'Or e Froidevaux 2011); in seguito, applicandolo a un caso di ricostruzione del sottosuolo, un articolo di Sartore, Fabbri e Gaetan (Sartore, Fabbri e Gaetan 2016) lo ha ripreso, descrivendo i passaggi simulativi consentiti tramite l'impiego del pacchetto "spMC" del software R.
Nel seguito, vengono illustrate le scelte e i passaggi principali relativi alla stima e all'impiego di tale modello nel caso di interesse, ovvero la ricostruzione di mappe bidimensionali perturbate.
Inizialmente, utilizzando i dati noti campionati nello spazio, vengono costruiti i "transiogrammi empirici", ovvero le funzioni che modellano la probabilità di passaggio di stato all'aumentare del lag spaziale **h** che separa due punti. Oltre a tali probabilità, si calcolano anche le proporzioni relative a ciascuna categoria rilevata nel dataset di riferimento. Una volta compiute queste operazioni, ad ogni transiogramma empirico viene associato un "transiogramma teorico" attraverso un modello esponenziale (Tabella 2.1), in modo da ottenere una probabilità di transizione tra stati ad ogni lag spaziale d'interesse.
Il passaggio successivo, ovvero quello relativo alla previsione nei punti ignoti, viene condotto utilizzando una catena di Monte Carlo, i cui nodi sono i punti della griglia. La previsione vera e propria viene effettuata tramite la massima entropia bayesiana (Allard,

D'Or e Froidevaux 2011, sezione §3.3.3); la formula specifica viene riportata in seguito:

$$\begin{aligned}\overline{p}_{k_0|k_1,\ldots,k_N} &= Pr\left(Z(\mathbf{s}_0) = k_0 \mid \bigcap_{i=1}^{N} Z(\mathbf{s}_i) = k_i\right) = \\ &= \frac{p_{k_0}\prod_{i=1}^{N} p_{k_i|k_0}(\mathbf{h}_i)}{\sum_{k_0=1}^{K} p_{k_0}\prod_{i=1}^{N} p_{k_i|k_0}(\mathbf{h}_i)} = \frac{p_{k_0}^{1-N}\prod_{i=1}^{N} p_{k_0,k_i}(\mathbf{h}_i)}{\sum_{k_0=1}^{K} p_{k_0}^{1-N}\prod_{i=1}^{N} p_{k_0,k_i}(\mathbf{h}_i)},\end{aligned} \tag{4.3}$$

Al pixel di coordinate $\mathbf{s}_0$ per cui interessa effettuare una previsione viene assegnata la categoria a cui corrisponde la probabilità condizionata stimata più alta.
Va inoltre specificato che per la simulazione di un singolo pixel non ci si limita a considerare un ristretto insieme di vicini, ma tutte le osservazioni disponibili concorrono alla stima del valore di interesse.
Il modello è implementato nel pacchetto "spMC" di R e utilizzabile mediante la funzione "sim_mcs()" che, in input, richiede i transiogrammi empirici e teorici stimati in precedenza.

4.3 Algoritmo ICM

L'algoritmo ICM viene proposto per la prima volta in un articolo di Julian Besag del 1986 (Besag 1986) nell'ambito della regolarizzazione e ricostruzione di immagini disturbate o "rumorose". Poiché nelle prove di simulazione presentate in seguito viene applicato e confrontato con il modello basato sulla massima entropia, ne viene brevemente presentato il funzionamento nella presente sezione.
Utilizzando la notazione dello stesso Besag, sia S una regione bidimensionale composta da elementi di forma rettangolare quali sono i pixel. Una possibile conformazione di S viene indicata con $x = (x_1, x_2, \ldots, x_N)$, dove il generico x_i si riferisce alla colorazione del pixel i-esimo. Con x^* viene indicata la vera e ignota regione, la quale è conveniente vedere come un vettore di variabili casuali $X = (X_1, X_2, \ldots, X_N)$, dove X_i è la variabile che assegna il colore al pixel i-esimo. Infine, con y_i ci si riferisce all'osservazione i-esima e con y al relativo vettore, realizzazione del vettore casuale $Y = (Y_1, Y_2, \ldots, Y_N)$.
Vengono effettuate due due assunzioni:

1. data una configurazione di x, le variabili $Y_1, Y_2, \ldots, Y_N$ sono condizionatamente

indipendenti e ogni variabile $Y_i, i = 1, \ldots, N$, ha la stessa funzione di densità condizionata nota $f(y_i|x_i)$ e dipendente solo da X_i. Di conseguenza, la distribuzione condizionata di y noto x è:

$$\log(y|x) = \prod_{i=1}^{N} f(y_i|x_i); \tag{4.4}$$

2. la vera struttura di colori x^* è realizzazione di un campo casuale di Markov con distribuzione $\{p(x)\}$.

Come scritto nella sezione precedente, e facendo seguito alla "prima legge della geografia" di Tobler, l'interesse si concentrerà sui campi le cui distribuzioni condizionate sono localmente dipendenti, ovvero, per ogni x vale la (4.1).
L'algoritmo ICM risulta essere particolarmente vantaggioso grazie sia al suo scarso carico computazionale che alla capacità di ignorare gli effetti di larga scala di $\{p(x)\}$. Definendo $\hat{x}$ come la stima provvisoria di x^*, l'obiettivo dell'algoritmo è quello di aggiornare la stima $\hat{x}_i$ del pixel i-esimo sfruttando tutta l'informazione disponibile. Una scelta naturale sembra essere quella di scegliere, per lo stesso pixel i-esimo, il colore che ha massima probabilità, condizionatamente ai dati osservati y e alla ricostruzione corrente negli altri punti, ovvero $\hat{x}_{S\setminus i}$. Grazie al teorema di Bayes, di conseguenza, si ottiene la seguente probabilità condizionata:

$$Pr(x_i|y, \hat{x}_{S\setminus i}) \propto f(y_i|x_i)p_i(x_i|\hat{x}_{\partial \mathbf{x}_i}). \tag{4.5}$$

Per quanto concerne la specificazione del campo casuale di Markov, $\{p(x)\}$, nel caso di categorie sconnesse, Besag suggerisce di utilizzare la seguente specificazione:

$$p_i(x_i = k|\partial x_i) \propto \exp\{\alpha_k + \beta u_i(k)\}, \tag{4.6}$$

dove $u_i(k)$ è il numero di vicini del pixel i-esimo aventi colore k. I coefficienti α_k e β vengono scelti arbitrariamente; il coefficiente β, in particolare, risulta essere unico e comune a tutte le categorie. Gli α_k, nel caso di categorie intercambiabili, vengono posti tutti uguali a zero, il che conduce alla specificazione della seguente distribuzione di

probabilità condizionata:

$$p_i(x_i = k|\partial x_i) = \frac{\exp\{\beta u_i(k)\}}{\sum_{k=1}^{K} \exp\{\beta u_i(k)\}}. \tag{4.7}$$

4.4 Simulazioni con modello di Potts

Inizialmente vengono predisposti gli oggetti utili alla simulazione delle osservazioni. Il primo oggetto necessario alla simulazione è la matrice in cui in ogni riga viene indicato un punto definito del raster e in colonna gli otto pixel che lo circondano (due per ognuna delle quattro direzioni cardinali principali, considerando gli intorni di secondo ordine definiti in Figura 4.1). La dipendenza spaziale viene simulata attraverso l'aggiornamento iterativo dell'intera griglia; più in particolare, ad ogni iterazione, i pixel vengono aggiornati utilizzando uno schema "a scacchiera a righe alternate" (Figura 4.2) in modo da modificare parallelamente quattro blocchi di celle.

I valori simulati, spazialmente correlati, vengono inseriti all'interno di un "raster", ovvero l'oggetto che consente di rappresentare un'immagine digitalmente. Esso, in buona sostanza, altro non è che una griglia ortogonale composta da celle che, nel lessico proprio della computer grafica, vengono chiamate "pixel". I raster simulati hanno dimensione pari a 10×10.
A livello pratico e a ciascuna iterazione, per ogni pixel viene contato il numero di vicini appartenente a ciascuna classe. Questi valori, una volta pesati attraverso una "matrice spaziale" scelta arbitrariamente e trasformati con la funzione esponenziale, vengono convertiti in probabilità. Un'estrazione multinomiale, a questo punto, assegna la classe al pixel in questione. La ripetizione della procedura per un numero sufficiente di volte consente di creare mappe caratterizzate da dipendenza spaziale. L'intensità di quest'ultima viene specificata attraverso il "coefficiente inverso di temperatura", indicato con β: all'aumentare del coefficiente, la temperatura diminuisce e il processo risulta più stabile e converge in un numero basso di iterazioni; viceversa scegliendo un β contenuto. La simulazione di una singola mappa viene fermata dopo poche iterazioni (cinque, nel caso specifico) in modo da evitare la convergenza, garantendo la presenza di tutte le categorie specificate e ottenendo la dipendenza spaziale desiderata.

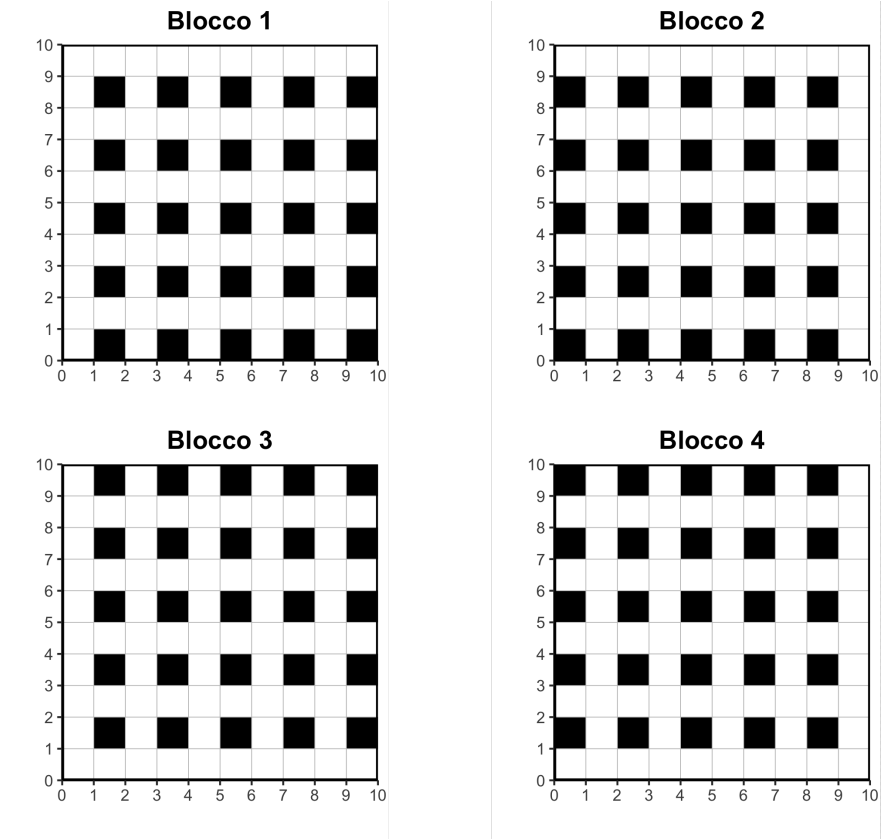

Figura 4.2: schema di aggiornamento "a scacchiera a righe alternate" per l'aggiornamento iterativo dei pixel e la generazione di griglie con dipendenza spaziale. I pixel "neri" vengono aggiornati simultaneamente. Ogni blocco viene aggiornato 5 volte.

Dopo aver simulato i dati in griglia, tramite campionamento casuale, gli stessi vengono suddivisi in insieme di stima e insieme di verifica: il primo set di dati si rende utile per la stima delle probabilità di transizione e per la definizione del modello predittivo. Le previsioni, condotte sui pixel appartenenti al set di verifica e confrontate con il loro vero valore, sono utili per determinare l'efficacia del metodo in analisi. Tale confronto viene operato mediante l'impiego di un'opportuna misura di similarità descritta nel seguito (l'indice ARI). Va specificato nuovamente come la quantità di dati utilizzati per la stima delle probabilità di transizione (e, di conseguenza, la quantità relativa all'insieme di verifica) sia variabile nelle simulazioni, in modo da poter studiare la qualità predittiva del modello al variare dell'informazione disponibile.
Successivamente vengono stimati i transiogrammi empirici per ciascuna combinazione di classi e, sulla base di questi, i transiogrammi teorici che saranno utili alla stima e alla

previsione effettive.
Le griglie così costruite, con una certa proporzione di pixel noti e un'altra di pixel su cui interessa effettuare le stime, si prestano per la valutazione dei due metodi predittivi: il modello basato sulla stima massima entropia bayesiana (MCS) e l'algoritmo ICM. Una volta effettuate le previsioni in corrispondenza dei pixel "mancanti", la bontà delle stesse viene calcolata utilizzando l'indice di Rand corretto, in modo da mantenere costante il valore atteso nel confronto tra due classificazioni costruite casualmente; la versione proposta è quella di Hubert e Arabie (Hubert e Arabie 1985). Più formalmente, l'indice corretto a valore atteso costante viene costruito a partire dalla seguente formulazione:

$$ARI = \frac{\text{Indice} - \text{Indice atteso}}{\text{Valore massimo} - \text{Indice atteso}}.$$

Considerando una tabella di contingenza che confronta le categorie osservate con quelle stimate dai modelli:

	Previsto				
Osservato	P_1	P_2	$\dots$	P_K	
O_1	$n_{1,1}$	$n_{1,2}$	$\dots$	$n_{1,K}$	a_1
O_2	$n_{2,1}$	$n_{2,2}$	$\dots$	$n_{2,K}$	a_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
O_K	$n_{K,1}$	$n_{K,2}$	$\dots$	$n_{K,K}$	a_K
	b_1	b_2	$\dots$	b_K	n

e indicizzando con i le classi dei valori osservati (O_i) e con j le classi dei valori previsti (P_j), la formulazione estesa dell'indice di Rand corretto diventa:

$$ARI = \frac{\sum_{ij}\binom{n_{ij}}{2} - [\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}]/\binom{n}{2}}{\frac{1}{2}[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}] - [\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}]/\binom{n}{2}}. \tag{4.8}$$

Le simulazioni vengono ripetute facendo variare il numero di categorie considerate nella

mappa, il "parametro di temperatura" β definito nel paragrafo §4.1, la proporzione di dati mancanti da ricostruire. Più specificatamente, vengono considerati tutti casi incrociando $K = 3, 4, 5$ categorie, $\beta = (0.2, 0.6)$ e $\pi = (0.60, 0.50.0.40, 0.30, 0.20)$ proporzioni di dati mancanti.

4.4.1 K = 3 categorie

Le prime simulazioni vengono effettuate considerando mappe caratterizzate da 3 categorie non ordinate. Come specificato in precedenza, viene messo a confronto il modello predittivo MCS con l'algoritmo di regolazione Iterative Conditional Modes proposto da Besag. Per ognuno di questi modelli si considerano due scenari, distinti per dal valore del "coefficiente di temperatura" β: si è scelto di testare un valore basso ($\beta = 0.2$) corrispondente a una temperatura "alta" e un valore alto ($\beta = 0.8$) corrispondente a una temperatura bassa. Per ogni modello e per ogni β, è stata testata la qualità predittiva del modello mediante l'utilizzo dell'indice ARI in diverse condizioni di disponibilità informativa. Nel caso di $K = 3$ categorie, sono state utilizzate le seguenti proporzioni di dati mancanti: 60%, 50%, 40%, 30% e 20%.
La matrice spaziale scelta per questo caso è la seguente:

$$S_3 = \begin{bmatrix} 4 & 1.2 & 0.8 \\ 1.2 & 4 & 1.2 \\ 0.8 & 1.2 & 4 \end{bmatrix}$$

Per ciascuna combinazione di modello, parametro di temperatura e proporzione di dati mancanti sono state replicate 50 simulazioni e su queste, dopo aver calcolato l'indice di Rand corretto, sono state determinate media, mediana, deviazione standard, minimo e massimo. Nella seguente tabella vengono riportati i risultati di simulazione:

	β	% dati mancanti	media	Me	sd	min	max
MCS	0.2	60%	0.265	0.240	0.171	0.000	0.625
		50%	0.311	0.279	0.189	0.000	0.938
		40%	0.343	0.331	0.175	0.000	0.870
		30%	0.376	0.381	0.220	0.000	1.000
		20%	0.458	0.415	0.244	0.000	1.000
	0.8	60%	0.523	0.523	0.262	0.000	1.000
		50%	0.586	0.586	0.193	0.141	1.000
		40%	0.602	0.642	0.286	0.000	1.000
		30%	0.720	0.761	0.257	0.000	1.000
		20%	0.736	0.782	0.234	0.284	1.000
ICM	0.2	60%	0.250	0.228	0.148	0.018	0.661
		50%	0.319	0.294	0.207	0.000	1.000
		40%	0.333	0.341	0.134	0.099	0.676
		30%	0.376	0.363	0.224	0.026	1.000
		20%	0.458	0.400	0.271	0.000	1.000
	0.8	60%	0.525	0.526	0.211	0.000	0.888
		50%	0.579	0.610	0.173	0.195	1.000
		40%	0.696	0.736	0.197	0.000	1.000
		30%	0.784	0.783	0.182	0.235	1.000
		20%	0.789	0.801	0.187	0.407	1.000

Tabella 4.1: principali statistiche descrittive relative all'indice di Rand corretto (ARI) calcolato sugli insiemi di verifica e utilizzando le previsioni effettuate con il modello basato sulla massima entropia bayesiana (MCS) e con l'algoritmo ICM (ICM) ($K = 3$ categorie, 50 simulazioni per combinazione).

Va specificato che nelle simulazioni con $\beta = 0.2$ e il 50% di dati mancanti, in due casi l'insieme di verifica è risultato perfettamente omogeneo nella previsione delle classi; di conseguenza, l'indice ARI è risultato non definibile e le simulazioni utili al calcolo delle statistiche descrittive sono state 48. Lo stesso discorso va considerato anche per il caso di $\beta = 0.8$ e il 20% di dati mancanti, dove le simulazioni utili sono state 49.

Ai fini interpretativi, risulta più agevole osservare gli stessi risultati espressi in forma grafica con l'ausilio di opportuni boxplot appaiati, divisi per proporzione di dati mancanti e valore del parametro β:

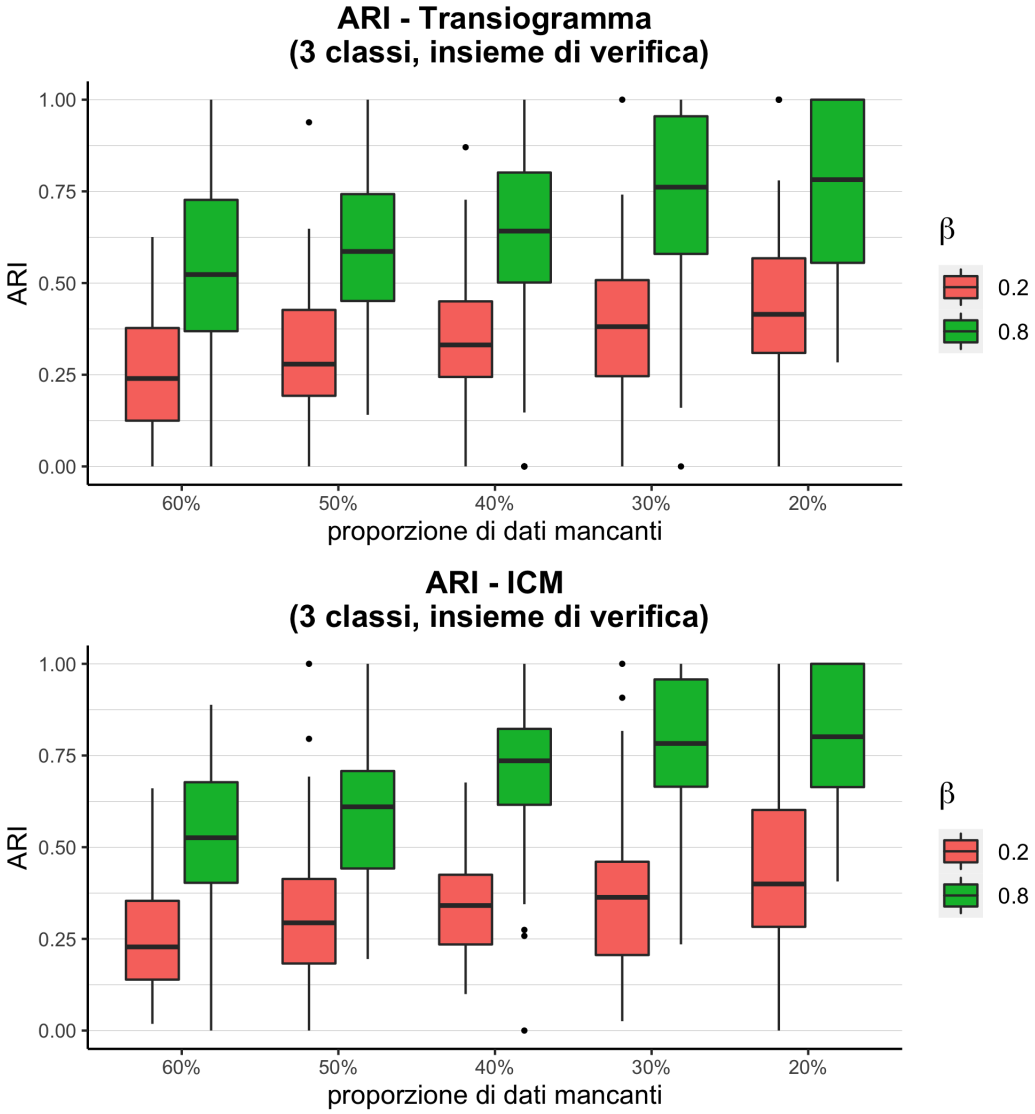

Figura 4.3: rappresentazione, mediante boxplot, delle statistiche riportate in 4.1 relativamente alle simulazioni con $K = 3$ classi.

Dall'osservazione grafica, l'algoritmo ICM sembra farsi preferire leggermente, producendo risultati mediamente migliori e più stabili in termini di deviazione standard calcolata sull'indice ARI. Nel seguito vengono riportati due esempi di simulazioni.

4.4.1.1 3 categorie, $\beta = 0.2$

Negli esempi di simulazione che seguono vengono considerate griglie di dimensione 10×10 con 3 categorie; in tutti questi esempi illustrativi viene considerata una proporzione di dati mancanti pari al 60% e si procede alla ricostruzione mediante previsioni sia utilizzando il modello basato sulla massima entropia bayesiana, sia quello basato sull'algoritmo ICM. In questo primo caso si è imposto $\beta = 0.2$. Vengono riportati in Figura 4.4 il transiogramma del relativo modello e in Figura 4.5 l'immagine originale, i valori campionati casualmente e considerati noti e le ricostruzioni operate con i due metodi, ovvero il modello MCS e l'algoritmo ICM.
È possibile osservare come entrambi i metodi riescano a riconoscere, anche se non con estrema precisione, le diverse zone. Tale imprecisione è imputabile anche e soprattutto all'alta "temperatura" indotta dal relativo coefficiente. Per completezza di trattazione, gli indici di Rand corretti calcolati sugli insiemi di verifica nelle due ricostruzioni sono, rispettivamente, $ARI_{TR} = 0.258$ e $ARI_{ICM} = 0.368$.

4.4.1.2 3 categorie, $\beta = 0.8$

Il secondo esempio, per il quale viene utilizzato un valore di $\beta = 0.8$, chiarisce anche l'aspetto della scelta della temperatura, inversamente proporzionale a β. Dapprima, in Figura 4.6, viene riportato il transiogramma nelle sue versioni empirica e teorica e, successivamente in Figura 4.7, le mappe simulate e ricostruite con i metodi proposti. La ricostruzione risulta essere molto più precisa grazie alla regolarità delle categorie di partenza. Il modello ottenuto a partire metodo MCS appare essere migliore nella zona in basso a destra della mappa, mentre l'algoritmo ICM ottiene previsioni migliori nella zona dove confinano tutte e tre le categorie. L'indice ARI calcolato con il metodo Multinomial Categorical Simulation è pari a $ARI_{TR} = 0.785$, mentre quello relativo all'algoritmo ICM è $ARI_{ICM} = 0.816$, in chiaro aumento rispetto ai casi in cui viene utilizzata una temperatura più alta che comporta maggior entropia.

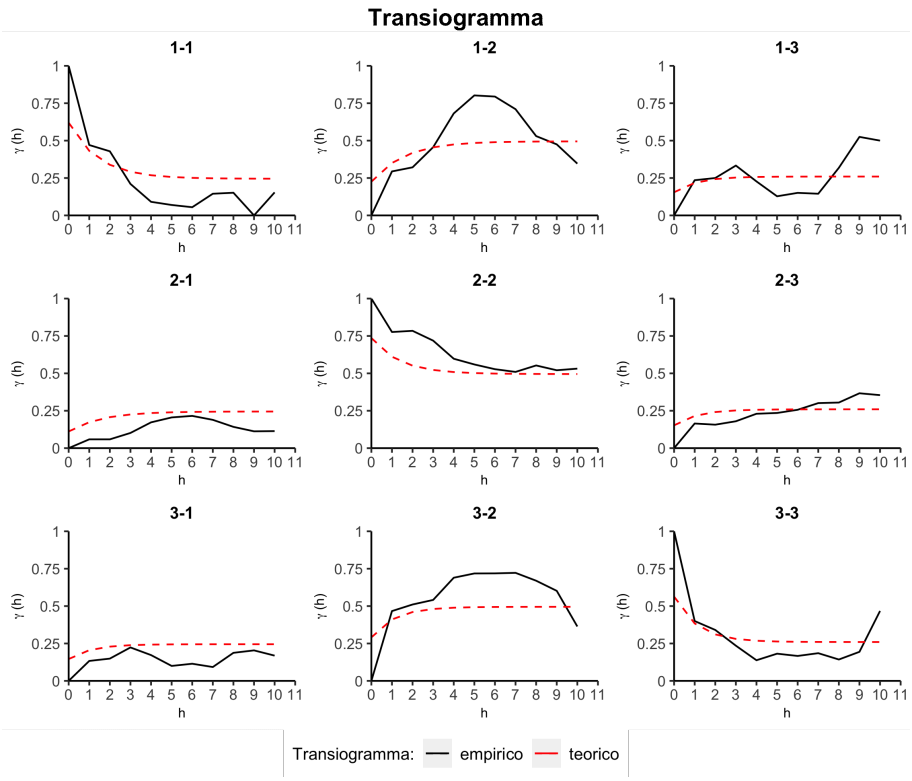

Figura 4.4: transiogramma esponenziale empirico (nero) e teorico (rosso) della probabilità di cambiamento di stato al variare della distanza **h** per il relativo esempio con $K = 3$ categorie e $\beta = 0.2$.

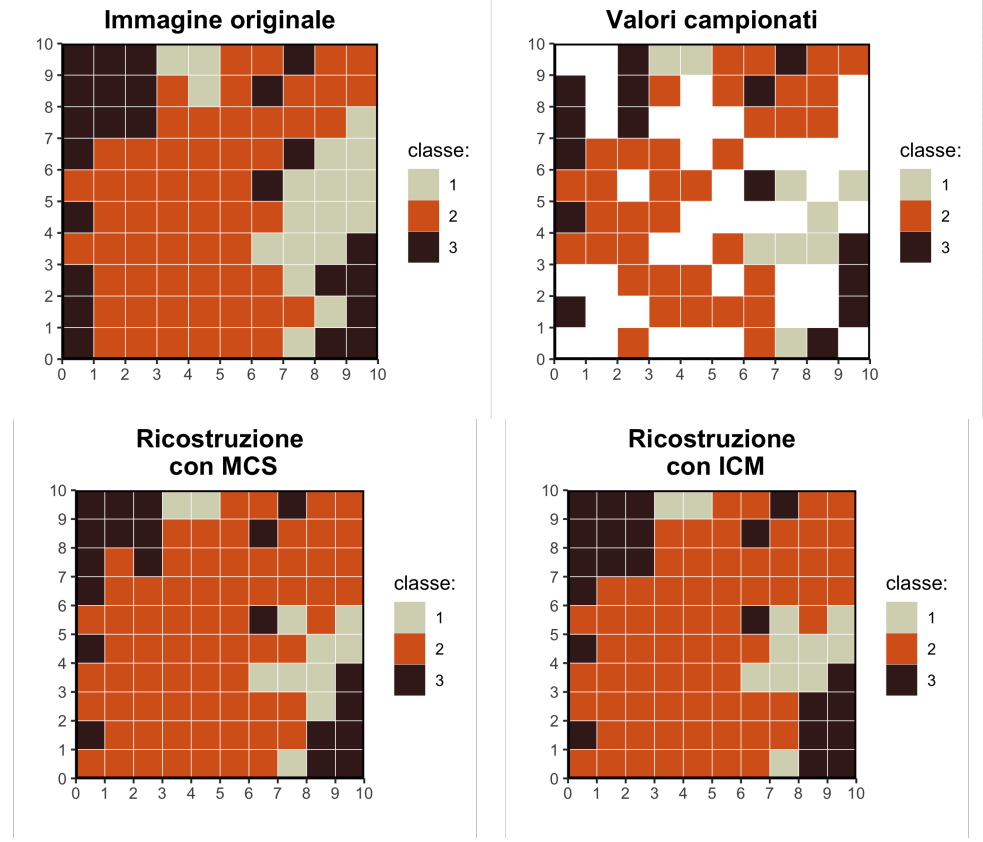

Figura 4.5: esempio di ricostruzione di una griglia con $K = 3$ categorie, utilizzando un "coefficiente di temperatura" pari a $\beta = 0.2$. L'immagine originale, a cui vengono tolti valori casualmente, viene ricostruita utilizzando il metodo MCS e l'algoritmo ICM.

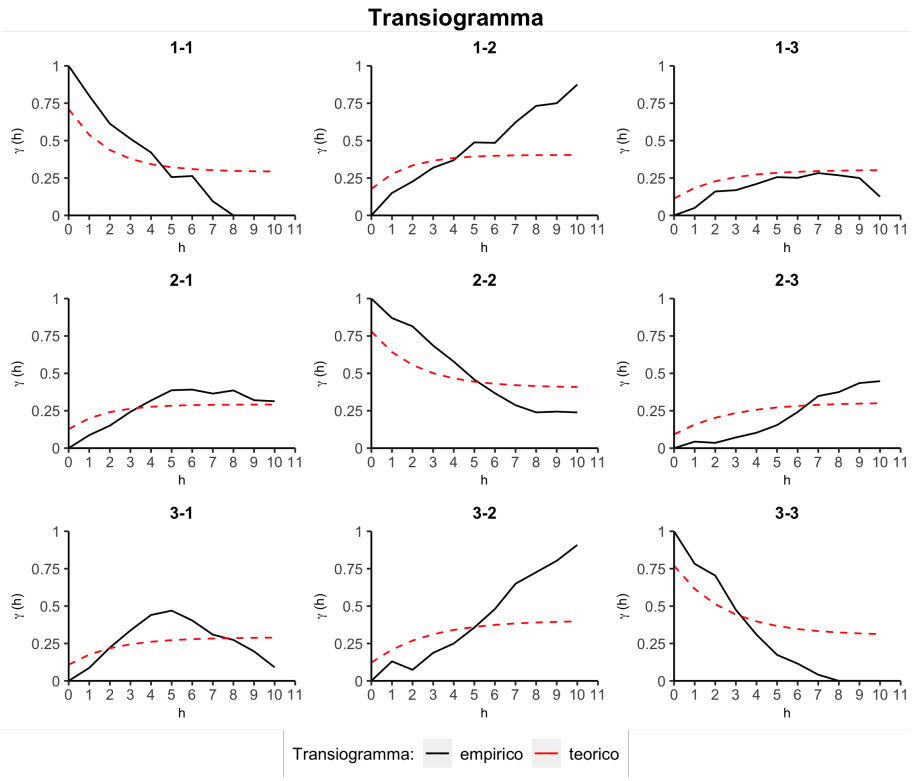

Figura 4.6: transiogramma esponenziale empirico (nero) e teorico (rosso) della probabilità di cambiamento di stato al variare della distanza **h** per il relativo esempio con $K = 3$ categorie e $\beta = 0.8$.

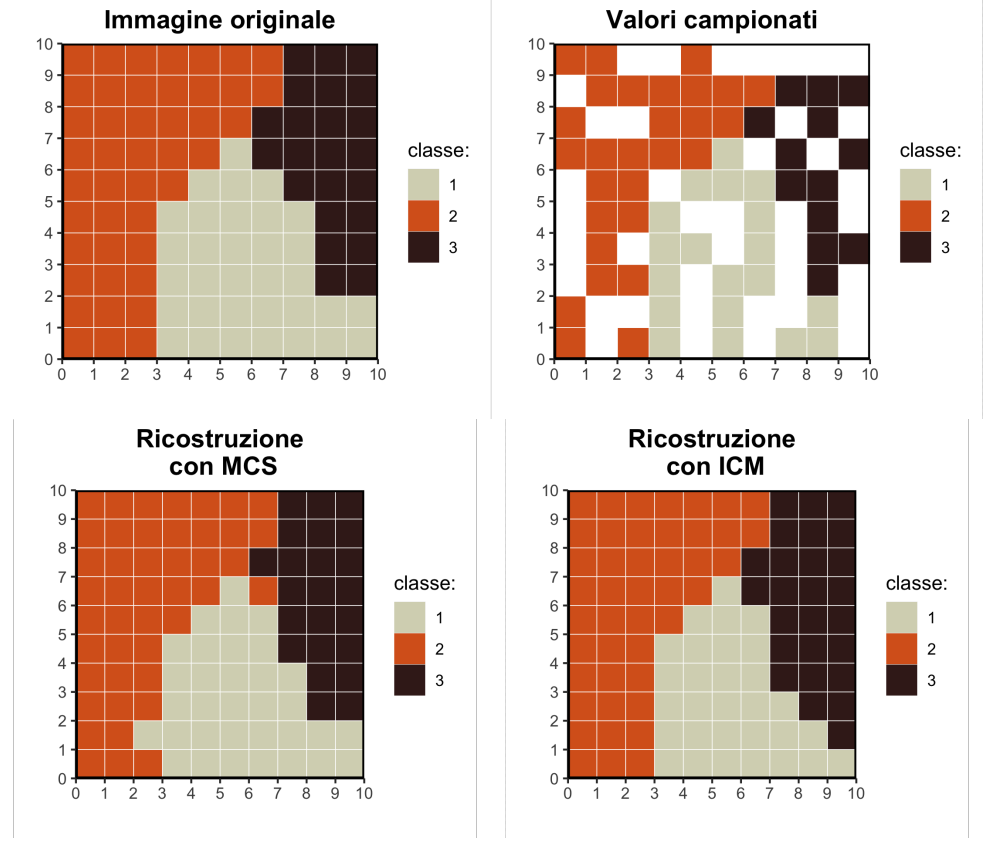

Figura 4.7: esempio di ricostruzione di una griglia con $K = 3$ categorie, utilizzando un "coefficiente di temperatura" pari a $\beta = 0.8$. L'immagine originale, a cui vengono tolti valori casualmente, viene ricostruita utilizzando il metodo MCS e l'algoritmo ICM.

4.4.2 K = 4 categorie

Le simulazioni relative alle mappe caratterizzate dalla compresenza di $K = 4$ categorie seguono lo schema descritto in precedenza per il caso di $K = 3$. La dimensione dei raster, i coefficienti β considerati e le proporzioni di dati mancanti, infatti, rimangono i medesimi. Le statistiche di sintesi, similmente, sono quelle viste in precedenza:

	β	% dati mancanti	media	Me	sd	min	max
MCS	0.2	60%	0.170	0.143	0.130	0.006	0.732
		50%	0.188	0.178	0.121	0.010	0.499
		40%	0.185	0.173	0.115	0.009	0.467
		30%	0.240	0.231	0.171	0.006	0.841
		20%	0.279	0.240	0.213	0.000	0.777
	0.8	60%	0.514	0.505	0.197	0.000	0.895
		50%	0.619	0.687	0.214	0.096	0.974
		40%	0.627	0.621	0.182	0.296	1.000
		30%	0.651	0.660	0.214	0.196	1.000
		20%	0.705	0.714	0.223	0.228	1.000
ICM	0.2	60%	0.141	0.107	0.142	0.003	0.730
		50%	0.173	0.161	0.109	0.004	0.395
		40%	0.176	0.164	0.123	0.000	0.572
		30%	0.236	0.183	0.181	0.002	0.857
		20%	0.275	0.207	0.235	0.000	1.000
	0.8	60%	0.519	0.515	0.149	0.212	0.842
		50%	0.610	0.604	0.157	0.280	0.954
		40%	0.659	0.628	0.155	0.256	0.928
		30%	0.717	0.700	0.170	0.238	1.000
		20%	0.782	0.830	0.196	0.354	1.000

Tabella 4.2: principali statistiche descrittive relative all'indice di Rand corretto (ARI) calcolato sugli insiemi di verifica e utilizzando le previsioni effettuate con il modello basato sulla massima entropia bayesiana (MCS) e con l'algoritmo ICM (ICM) ($K = 4$ categorie, 50 simulazioni per combinazione).

La matrice spaziale scelta, nel caso di 4 categorie, ha la seguente conformazione:

$$S_4 = \begin{bmatrix} 4 & 1.6 & 1.2 & 0.8 \\ 1.6 & 4 & 1.6 & 1.2 \\ 1.2 & 1.6 & 4 & 1.6 \\ 0.8 & 1.2 & 1.6 & 4 \end{bmatrix}.$$

Gli stessi risultati e le relative conclusioni possono essere rappresentati graficamente attraverso boxplot appaiati come quelli riportati in seguito:

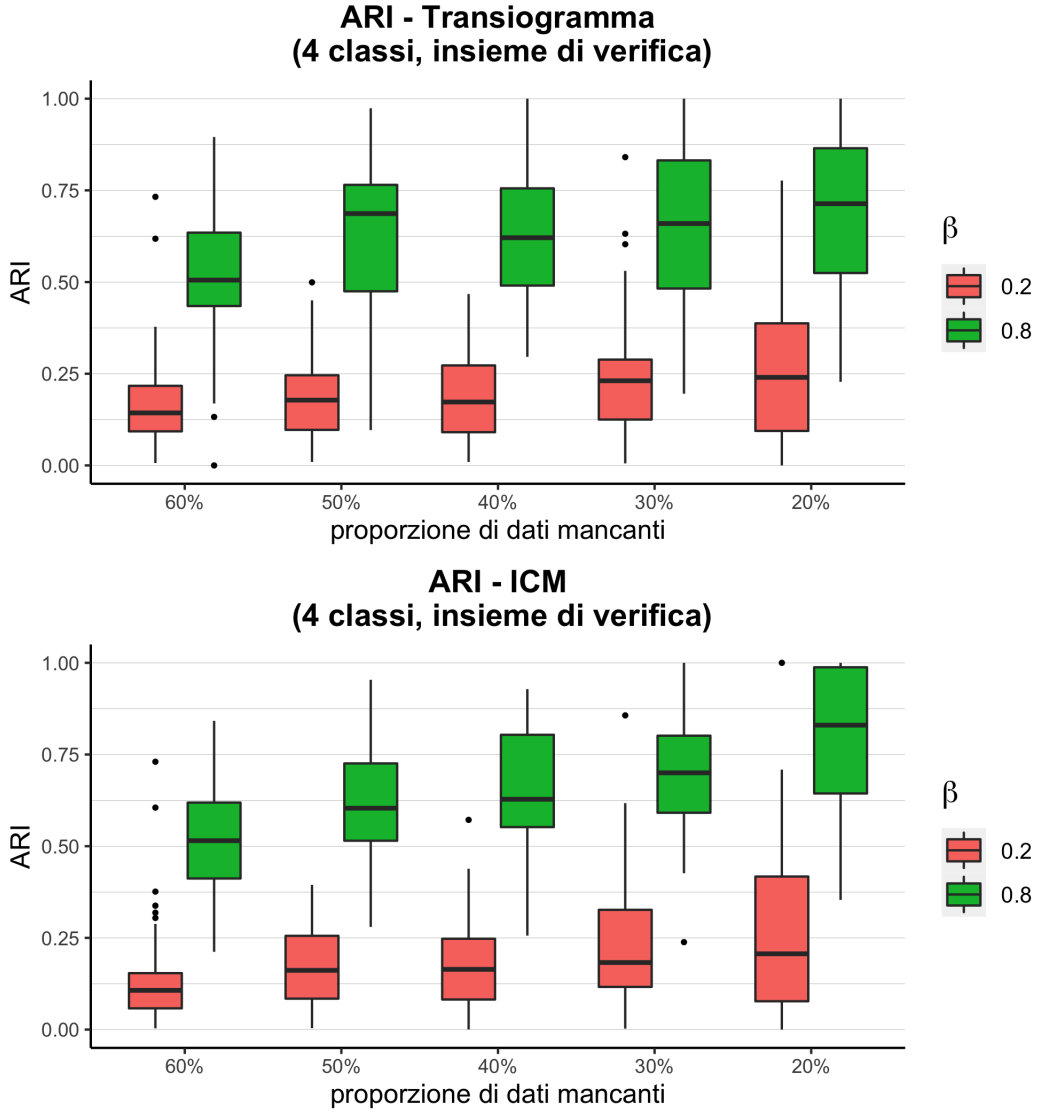

Figura 4.8: rappresentazione, mediante boxplot, delle statistiche riportate in 4.2 relativamente alle simulazioni con $K = 4$ classi.

Anche nel caso di $K = 4$ categorie è possibile notare come l'algoritmo ICM registri performance leggermente migliori rispetto al modello basato sulla massima entropia bayesiana. Rispetto al caso precedente, dove la differenza rimaneva comunque non molto marcata, nel contesto in analisi tale difformità risulta ancor meno evidente; per temperature alte ($\beta = 0.2$), in particolare, i due modelli sembrano essere sostanzialmente equivalenti.
Come nel caso con 3 categorie, vengono riportati due esempi (uno per ognuno dei valori di β considerati) di ricostruzione in ambiente simulativo.

4.4.2.1 4 categorie, $\beta = 0.2$

Utilizzando una temperatura più alta ($\beta = 0.2$), la ricostruzione risulta, naturalmente, più difficoltosa quale che sia il metodo considerato. In Figura 4.9 è raffigurato il transiogramma relativo all'esempio, la cui rappresentazione in raster è riportata in Figura 4.10.
Il modello che impiega la massima entropia prevede meglio i punti ignoti presenti nella parte alta del raster; viceversa per il metodo basato sull'algoritmo ICM. Gli indici di Rand corretti risultano essere pari a $ARI_{TR} = 0.315$ nel caso del metodo MCS e a $ARI_{ICM} = 0.193$ nel caso dell'algoritmo di regolarizzazione ICM.
Si nota già con l'inserimento di una sola categoria aggiuntiva la maggiore difficoltà predittiva, dovuta anche dalla presenza sporadica di qualche pixel isolato.

4.4.2.2 4 categorie, $\beta = 0.8$

Scegliendo una temperatura più bassa, invece, le categorie si "raggruppano" e si avvicinano a convergenza più rapidamente. Questo permette previsioni più precise, come nel caso illustrato nel seguito. Ancora una volta vengono riportati i grafici del transiogramma esponenziale e le relative mappe simulate, campionate e ricostruite, rispettivamente in Figura 4.11 e in Figura 4.12.
Entrambi i modelli ottengono risultati soddisfacenti. L'eccezione di qualche pixel sparso che assume colori non coerenti con l'immagine originaria, infatti, non intacca il risultato complessivo di una buona previsione globale della mappa. L'indice ARI per il modello

ottenuto tramite il modello spaziale bayesiano è $ARI_{TR} = 0.538$; con l'algoritmo Iterated Conditional Modes è $ARI_{ICM} = 0.602$.

Come accennato nell'illustrazione del caso precedente, in generale, l'introduzione di ulteriori categorie comporta una maggiore difficoltà predittiva; si riscontra ancora, invece, il miglioramento nella qualità delle previsioni al diminuire della "temperatura" utilizzata in fase di simulazione della griglia.

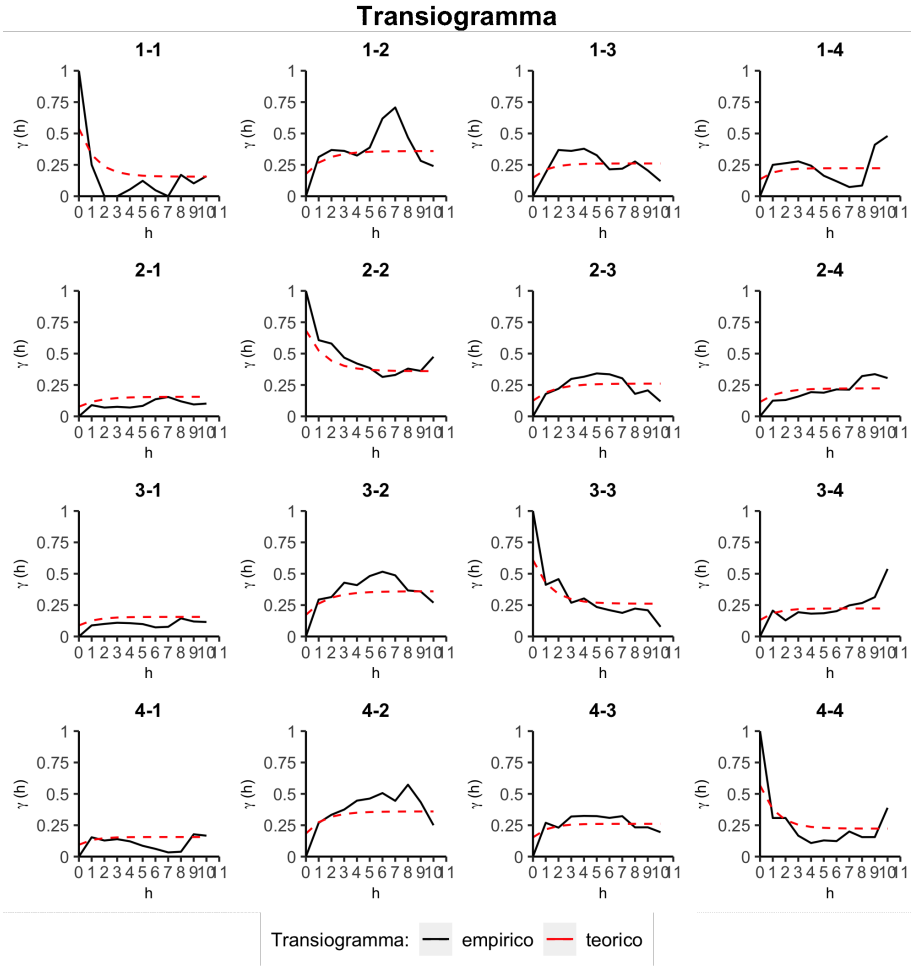

Figura 4.9: transiogramma esponenziale empirico (nero) e teorico (rosso) della probabilità di cambiamento di stato al variare della distanza **h** per il relativo esempio con $K = 4$ categorie e $\beta = 0.2$.

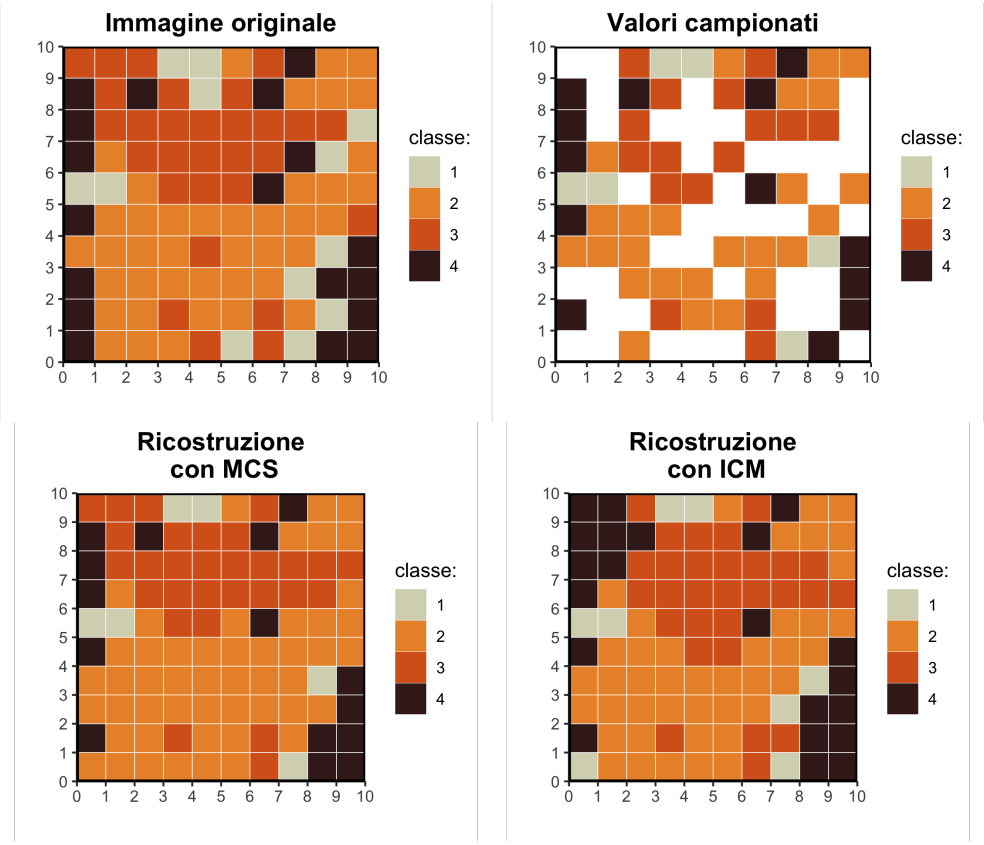

Figura 4.10: esempio di ricostruzione di una griglia con $K = 4$ categorie, utilizzando un "coefficiente di temperatura" pari a $\beta = 0.2$. L'immagine originale, a cui vengono tolti valori casualmente, viene ricostruita utilizzando il metodo MCS e l'algoritmo ICM.

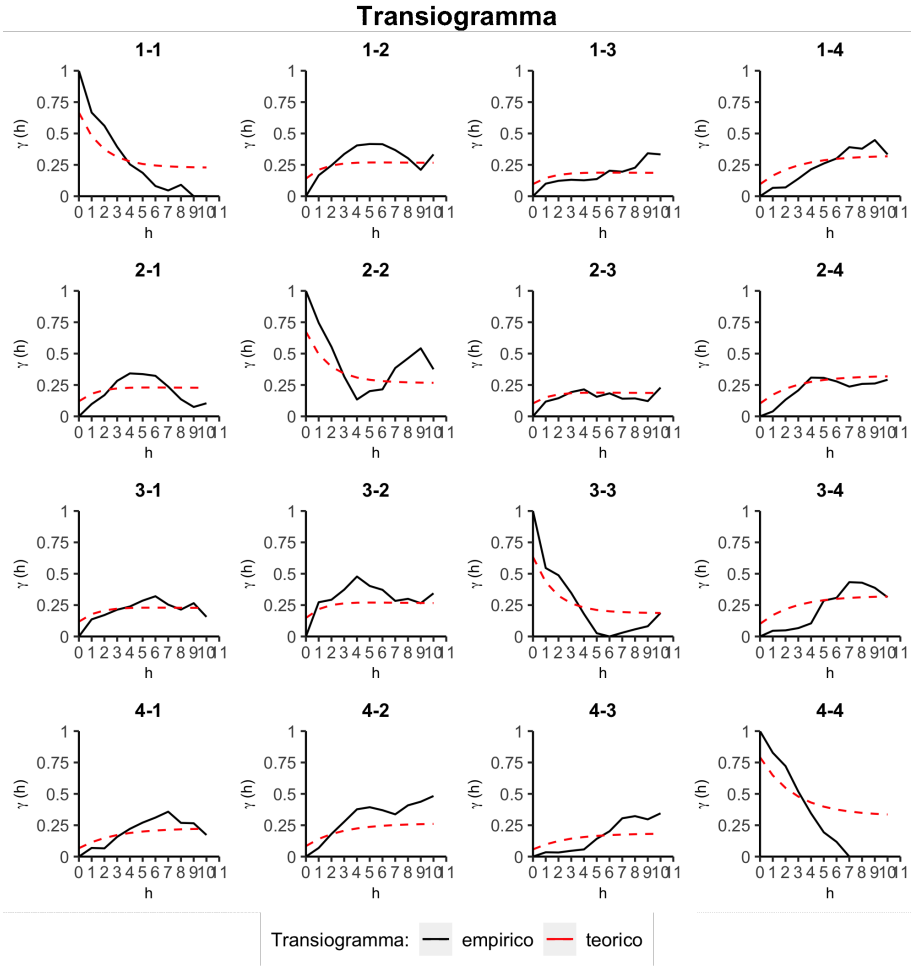

Figura 4.11: transiogramma esponenziale empirico (nero) e teorico (rosso) della probabilità di cambiamento di stato al variare della distanza **h** per il relativo esempio con $K = 4$ categorie e $\beta = 0.8$.

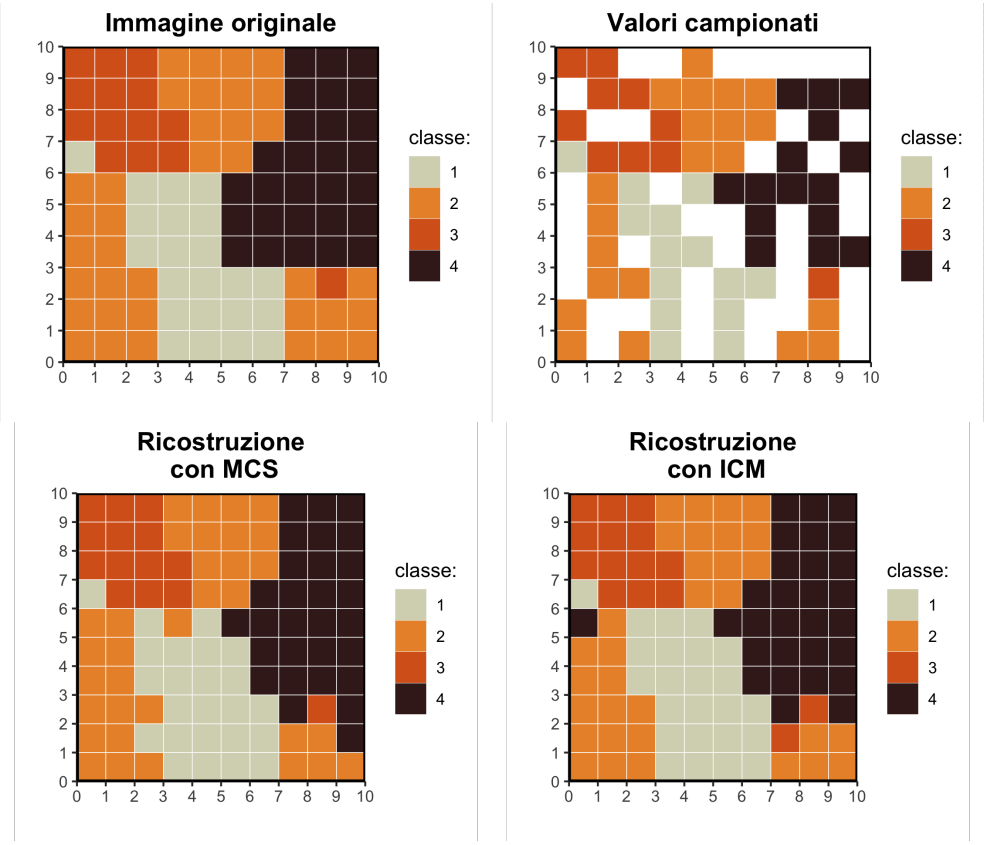

Figura 4.12: esempio di ricostruzione di una griglia con $K = 4$ categorie, utilizzando un "coefficiente di temperatura" pari a $\beta = 0.8$. L'immagine originale, a cui vengono tolti valori casualmente, viene ricostruita utilizzando il metodo MCS e l'algoritmo ICM.

4.4.3 K = 5 categorie

Le prove simulative, non diversamente da quanto effettuato nei due casi precedenti con 3 e 4 categorie, consistono di 50 prove per ciascuna combinazione di modello scelto, parametro di temperatura β e proporzione di dati mancanti. L'unica differenza risiede nella percentuale di dati noti e ignoti; nel caso di 5 categorie e di transiogrammi esponenziali, infatti, il numero minimo di osservazioni necessarie alla simulazione è pari a 55; ne consegue che le diverse proporzioni di dati mancanti prese in considerazione, in questo caso, sono del 45%, 40%, 30% e 20%.
I risultati di simulazione sono i seguenti:

	β	% dati mancanti	media	Me	sd	min	max
MCS	0.2	45%	0.084	0.058	0.072	0.000	0.305
		40%	0.090	0.064	0.079	0.006	0.285
		30%	0.109	0.082	0.101	0.001	0.423
		20%	0.122	0.093	0.113	0.002	0.484
	0.8	45%	0.575	0.574	0.180	0.155	1.000
		40%	0.609	0.598	0.164	0.306	0.942
		30%	0.703	0.730	0.165	0.288	1.000
		20%	0.674	0.679	0.222	0.017	1.000
ICM	0.2	45%	0.085	0.068	0.064	0.001	0.303
		40%	0.091	0.082	0.072	0.001	0.279
		30%	0.131	0.105	0.137	0.000	0.701
		20%	0.135	0.095	0.148	0.000	0.697
	0.8	45%	0.630	0.631	0.159	0.297	1.000
		40%	0.626	0.615	0.150	0.254	1.000
		30%	0.723	0.727	0.180	0.329	1.000
		20%	0.760	0.789	0.182	0.294	1.000

Tabella 4.3: principali statistiche descrittive relative all'indice di Rand corretto (ARI) calcolato sugli insiemi di verifica e utilizzando le previsioni effettuate con il modello basato sulla massima entropia bayesiana (MCS) e con l'algoritmo ICM (ICM) ($K = 5$ categorie, 50 simulazioni per combinazione, 100 con il 45% di dati mancanti).

Dal punto di vista metodologico, si sottolinea come il numero di simulazioni effettuate nello scenario con il 45% di dati mancanti sia stato pari a 100, invece delle usuali 50. In un caso, nel contesto definito da $\beta = 0.8$ e con il 20% di dati mancanti, si riscontra nuovamente il problema di omogeneità nel set di verifica; le simulazioni utili per tale combinazione sono, quindi, 49.
Ancora una volta viene riportata anche la versione grafica dei risultati riportati precedentemente:

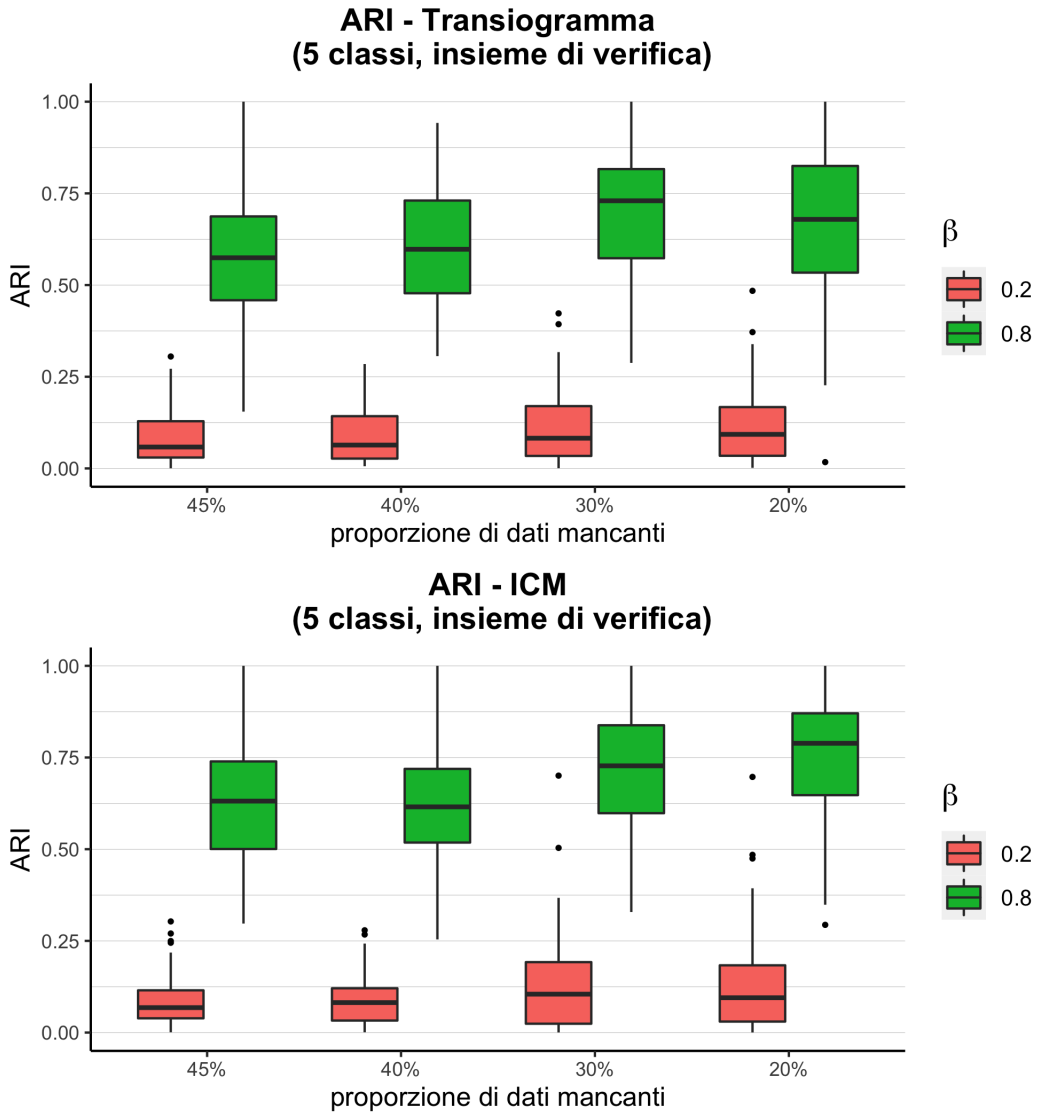

Figura 4.13: rappresentazione, mediante boxplot, delle statistiche riportate in 4.3 relativamente alle simulazioni con $K = 5$ classi.

Anche in questo caso i risultati appaiono del tutto comparabili, con l'indice ARI che sottolinea il deciso miglioramento predittivo nel caso di categorie ben definite e compatte.
Si sottolinea come la matrice spaziale scelta, nel caso di 5 categorie, abbia la seguente conformazione:

$$S_5 = \begin{bmatrix} 4 & 2 & 1.6 & 1.2 & 0.8 \\ 2 & 4 & 2 & 1.6 & 1.2 \\ 1.6 & 2 & 4 & 2 & 1.6 \\ 1.2 & 1.6 & 2 & 4 & 1.6 \\ 0.8 & 1.2 & 1.6 & 2 & 4 \end{bmatrix}.$$

Vengono proposti due ulteriori esempi di simulazione su raster, utilizzando $K = 5$ categorie e gli usuali valori del parametro di temperatura β.

4.4.3.1 5 categorie, $\beta = 0.2$

Nel primo caso, con $\beta = 0.2$, si nota immediatamente come le aree definite dalle diverse classi siano del tutto disomogenee, preludio a una previsione certamente difficoltosa; le mappe son riportate in Figura 4.15. In Figura 4.14 viene anche evidenziata la matrice dei transiogrammi empirici e teorici relativi alle probabilità di transizione.
Come facilmente preventivabile, il numero di pixel previsti correttamente diminuisce rispetto ai casi con 3 o 4 categorie. Nonostante ciò, alcuni pattern riescono comunque ad essere riprodotti, in particolare in corrispondenza delle zone più omogenee all'interno della mappa.
Il criterio di valutazione delle previsioni è nuovamente l'ARI. Considerando il modello basato sulla massima entropia bayesiana si ottiene $ARI_{TR} = 0.087$, mentre con l'algoritmo ICM risulta $ARI_{ICM} = 0.095$, in evidente calo rispetto a tutte le situazioni precedenti.

4.4.3.2 5 categorie, $\beta = 0.2$

Viene proposto un ultimo esempio con $\beta = 0.8$. In questo secondo caso le aree sono decisamente più definite e già con una semplice valutazione grafica è possibile affermare che il metodo MCS fornisce previsioni migliori rispetto a quelle ottenute con l'algoritmo

ICM, come mostrato dalle ricostruzioni presenti in Figura 4.17. Il transiogramma, nelle sue versioni empirica e teorica, risulta avere la rappresentazione grafica riportata in Figura 4.16.

Nell'esempio considerato, la differenza sostanziale è ben visibile nella parte più in basso della griglia, in corrispondenza di quella che dovrebbe essere una zona omogenea, le cui previsioni ci si può attendere che siano tutte relative alla "classe 5"; con l'algoritmo ICM, invece, questa zona risulta estremamente variabile. Gli ARI calcolati nei due casi risultano essere pari a $ARI_{TR} = 0.568$ e $ARI_{ICM} = 0.546$.

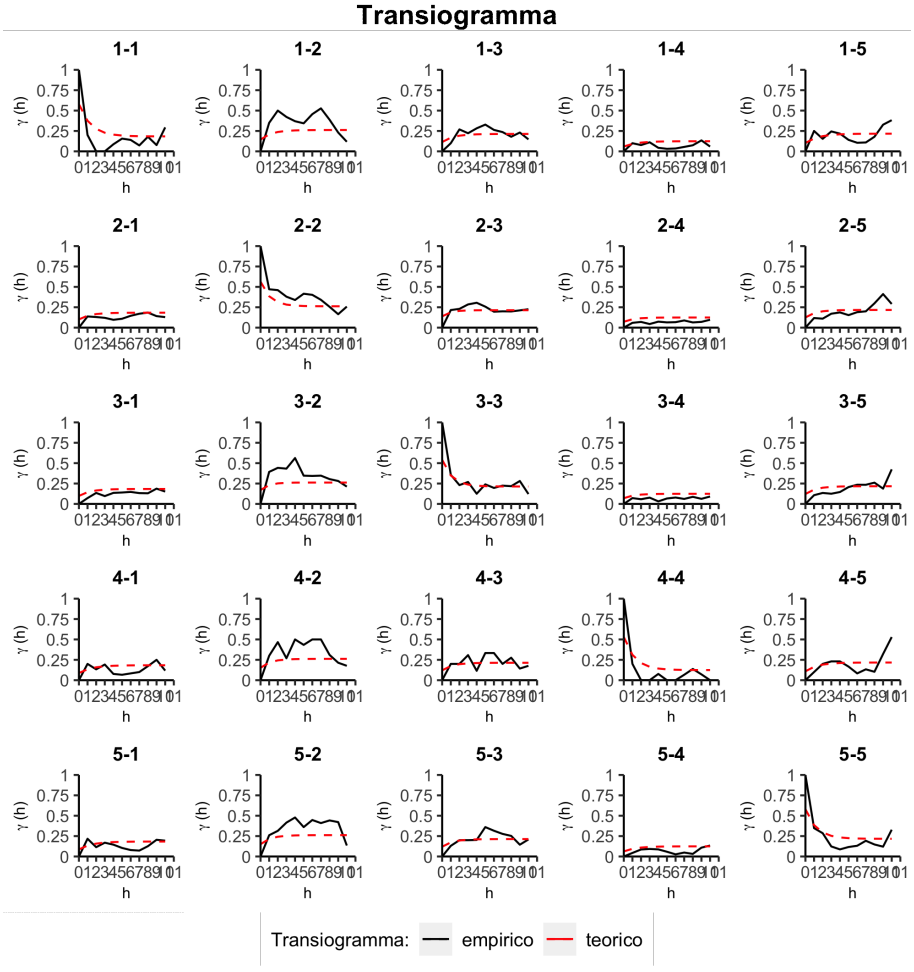

Figura 4.14: transiogramma esponenziale empirico (nero) e teorico (rosso) della probabilità di cambiamento di stato al variare della distanza **h** per il relativo esempio con $K = 5$ categorie e $\beta = 0.2$.

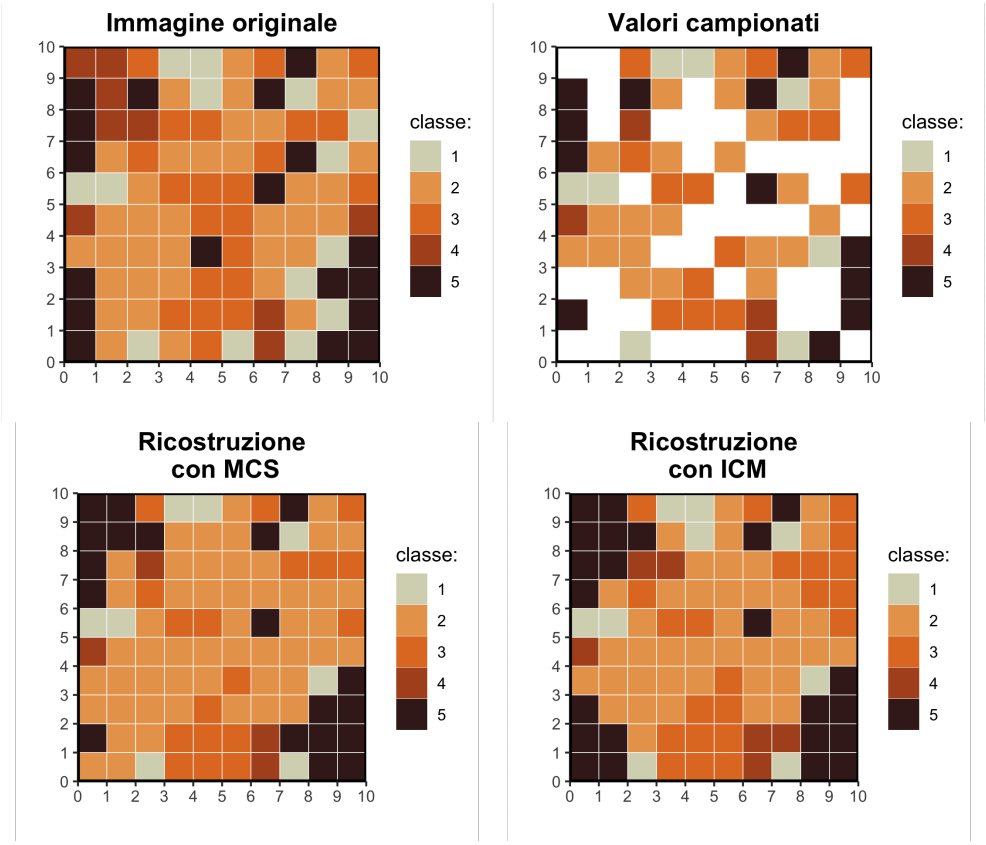

Figura 4.15: esempio di ricostruzione di una griglia con $K = 5$ categorie, utilizzando un "coefficiente di temperatura" pari a $\beta = 0.2$. L'immagine originale, a cui vengono tolti valori casualmente, viene ricostruita utilizzando il metodo MCS e l'algoritmo ICM.

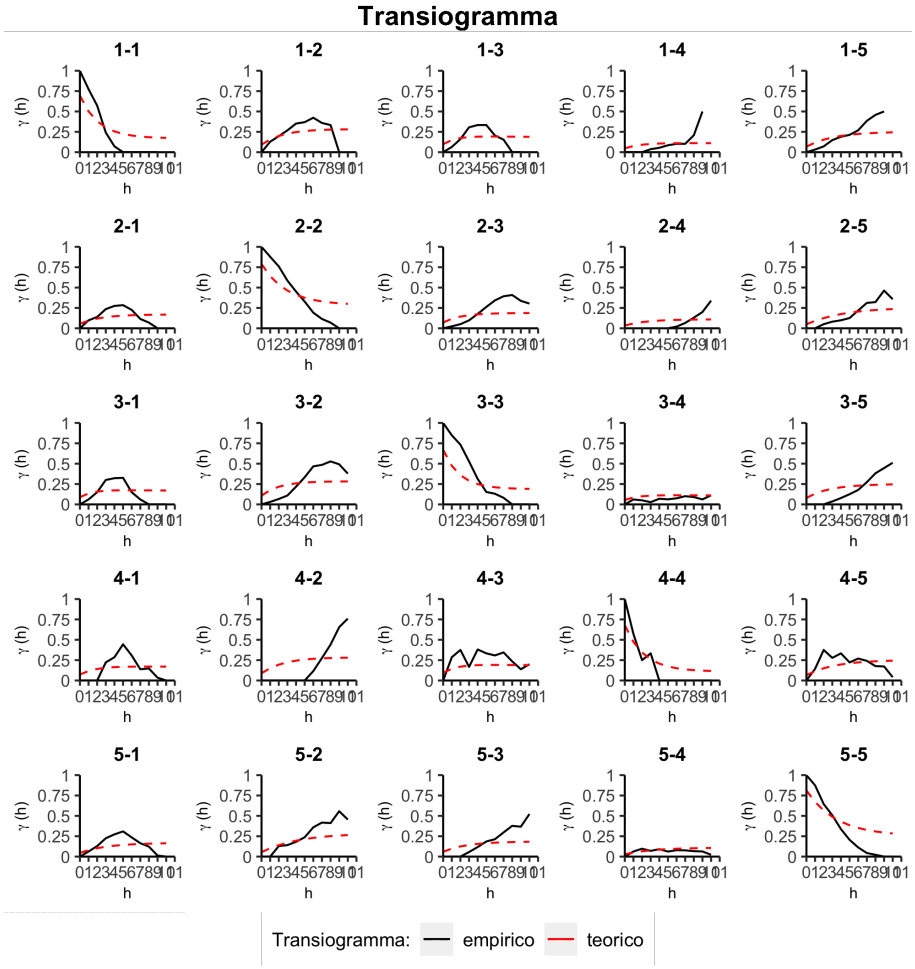

Figura 4.16: transiogramma esponenziale empirico (nero) e teorico (rosso) della probabilità di cambiamento di stato al variare della distanza **h** per il relativo esempio con $K = 5$ categorie e $\beta = 0.8$.

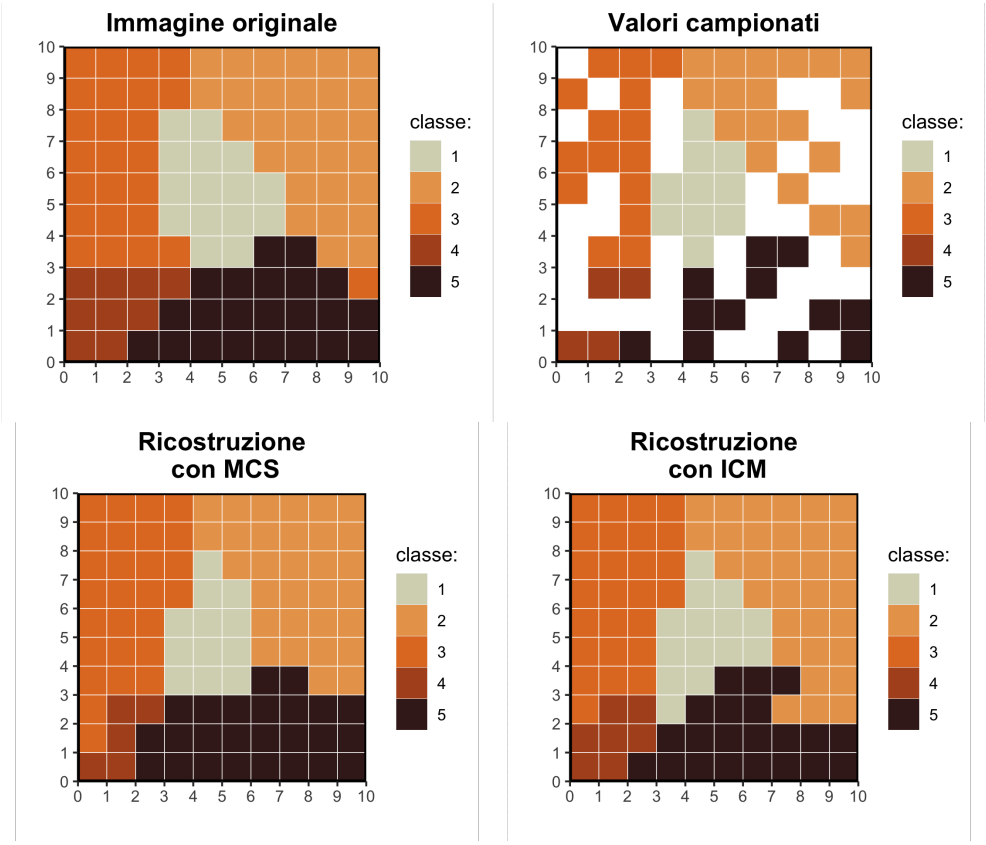

Figura 4.17: esempio di ricostruzione di una griglia con $K = 5$ categorie, utilizzando un "coefficiente di temperatura" pari a $\beta = 0.8$. L'immagine originale, a cui vengono tolti valori casualmente, viene ricostruita utilizzando il metodo MCS e l'algoritmo ICM.

4.5 Commento sulle simulazioni

In conclusione, è possibile produrre qualche commento relativamente alle simulazioni effettuate poc'anzi.
Innanzitutto i due approcci, quello parametrico che utilizza in partenza la stima dei transiogrammi per interpretare la dipendenza spaziale del processo ed effettuare previsioni sfruttando la massima entropia bayesiana (MCS) e quello non-parametrico proposto da Besag che risponde al nome di "Iterated Conditional Modes", forniscono risultati non dissimili in fase predittiva. In particolare è possibile osservare e confermare tendenze preventivabili precedentemente alle simulazioni, come la miglior capacità predittiva al diminuire della temperatura (inversamente proporzionale a β) scelta in fase di simulazione della mappa e al diminuire della proporzione di dati mancanti.
Si nota anche una crescente difficoltà nell'ottenere simulazioni omogenee e affidabili al crescere del numero K di classi considerate. Da questo punto di vista nel presente lavoro si sono considerati tre scenari, con 3, 4 o 5 possibili categorie; in una griglia delle dimensioni di 10×10 pixel risulta naturale osservare una maggior "entropia" al crescere della temperatura e all'aumentare delle categorie rilevate.
Va evidenziato come, dal punto di vista strettamente numerico, l'algoritmo ICM sembri farsi preferire, pur in modo non sensibile. Si sottolinea, però, che il modello basato sulla massima entropia bayesiana proposto viene implementato nella sua forma più semplice, senza considerare possibili effetti anisotropici che, probabilmente, incrementerebbero la qualità delle previsioni.

5 Applicazioni a dati reali

Nel presente capitolo viene applicato a casi reali il modello esplorato nelle sezioni precedenti e relativo alla ricostruzione di immagini mediante il metodo *Multinomial Categorical Simulation*.
Il primo esempio proposto riguarda l'isola di La Palma, nell'arcipelago delle Canarie; l'immagine satellitare è caratterizzata da un'alta presenza di nuvole che nascondono il perimetro dell'isola. In questo primo caso l'obiettivo è duplice: ricostruire il confine tra terra e mare e riconoscere le caratteristiche e le trame del suolo dell'isola.
In un altro esempio si propone una seconda immagine satellitare riferibile agli incendi scoppiati in California nel corso del 2021. La zona di interesse si trova a nord-est della città di Sacramento, dove la presenza di nubi causate da tali incendi ha mascherato il suolo sottostante. Ancora una volta l'obiettivo è quello di ricostruire il terreno sottostante a partire dalle zone non coperte.
Il terzo caso applicativo si concentra sulla ricostruzione dell'utilizzo del suolo della provincia di Verona a partire dalla relativa indagine campionaria "LUCAS 2018 (*Land Use / Cover Area Frame Survey*)", condotta da Eurostat.
Prima di trattare i casi applicativi, viene presentato l'algoritmo di clustering utilizzato per la riduzione della complessità delle immagini satellitari analizzate.

5.1 Algoritmo CLARA

L'algoritmo CLARA (*Clustering LARge Application*) si rende utile nei casi in cui sia d'interesse operare un cosiddetto *hard clustering* delle unità statistiche a disposizione. Nel caso specifico le unità sono rappresentate da pixel di diverso colore; l'obiettivo è quello di riproporre la stessa immagine sfruttando un numero contenuto di colori, scelto arbitrariamente in partenza.
Più nel dettaglio, CLARA opera il clustering sfruttando i medoidi ma, invece di utilizzare tutte le unità presenti nel dataset, si avvale solo di un campione rappresentativo per definirli. L'algoritmo, in buona sintesi, consiste dei seguenti passaggi:

1. vengono creati casualmente dei sottoinsiemi del dataset originario della grandezza specificata;
2. ad ognuno di questi sottoinsiemi viene applicato l'algoritmo PAM (*Partitioning Around Medoids*) in modo da individuare, per ciascuno, i K_C medoidi rappresentativi. Ogni osservazione del dataset viene assegnata al medoide più vicino;
3. per la valutazione della bontà del clustering viene calcolata una misura di dissimilarità tra le osservazioni e e il medoide più vicino a ciascuna;
4. per contenere l'errore di campionamento, le operazioni di cui sopra vengono ripetute per un numero specificato di volte. La partizione finale scelta sarà quella con la più bassa misura di dissimilarità.

L'algoritmo CLARA, nella presente trattazione, sarà utilizzato per ridurre la complessità delle immagini. Infatti, ogni pixel potrà assumere solo una delle K_C tonalità possibili, con K_C scelto arbitrariamente.

5.2 Isola di La Palma, Canarie

Il primo esempio di applicazione del modello di previsione mediante massima entropia bayesiana riguarda l'isola di La Palma, facente parte dell'arcipelago delle isole Canarie, nell'Oceano Atlantico.
L'immagine satellitare di partenza, ottenuta dal sito britannico di Sky News (Sky News 2021), è caratterizzata da una presenza ingente di nuvole, distribuite sulla quasi totalità delle coste dell'isola; tali nubi non consentono di osservare nel dettaglio tutta la superficie di La Palma, cosa che può essere di interesse in determinate situazioni, possibilmente emergenziali. La Figura 5.1 illustra il caso di studio:

Figura 5.1: immagine originale dell'isola di La Palma, Canarie, ottenuta dal sito web di Sky News.

L'obiettivo, quindi, è quello di ricostruire il perimetro dell'isola e cercare di identificarne i pattern fisici che la contraddistinguono.
Data l'alta definizione dell'immagine e considerando la capacità computazionale a disposizione, si è scelto di ridurre le dimensioni della stessa aumentando la dimensione dei singoli pixel aggregandone di adiacenti e "mediando" il colore. L'immagine di partenza per le analisi, in questo caso, assume dimensioni pari a 110×100 pixel).

Dal punto di vista algoritmico, la prima operazione è di *hard clustering*, consistente nell'assegnazione di ogni pixel ad una delle K_C classi possibili, con K_C scelto arbitrariamente. In questo caso, viene imposto $K_C = 20$; in altre parole, l'immagine verrà ricomposta utilizzando solo 20 colori selezionati dall'algoritmo scelto. Per tale operazione si è utilizzato l'algoritmo CLARA (*Clustering LARge Application*) che consente di maneggiare grandi quantità di dati attraverso il campionamento dell'immagine e l'utilizzo di *medoidi*. La nuova immagine in Figura 5.2 composta da soli 20 colori risulta come segue:

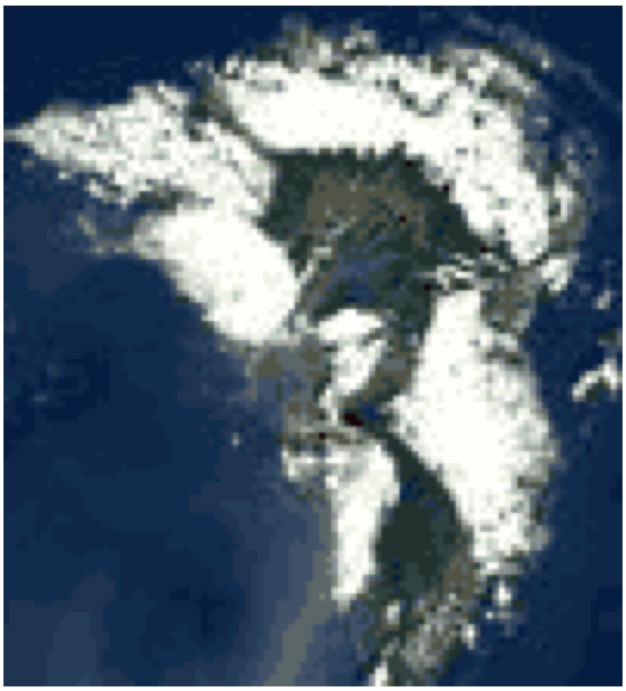

Figura 5.2: immagine ricostruita dopo l'impiego dell'algoritmo CLARA e utilizzando 20 cluster.

L'immagine ora composta da 11000 pixel di $K_C = 20$ colori diversi viene ulteriormente semplificata attraverso il raggruppamento dei colori in 4 classi; si nota, infatti, che la figura è composta da altrettanti colori, fondamentalmente: verde, marrone, blu e bianco. Questo secondo clustering viene operato tramite l'aggregazione soggettiva dei colori. I pixel assegnati alla classe corrispondente al colore "bianco" vengono quindi considerati come quelli mancanti e su cui è interessante effettuare una previsione. Il campione su cui stimare il modello basato sulla massima entropia, di conseguenza, sarà composto

dalle restanti unità statistiche. Successivamente vengono stimati i transiogrammi e le relative previsioni sull'intera mappa; è giusto specificare che il numero di classi considerate per la ricostruzione, a questo punto, è $K = 3$ (blu per il mare, verde per la superficie erbosa/boschiva, marrone per i rilievi).
L'immagine che mostra le tre categorie utili e i pixel "mancanti" (Figura 5.3) è riportata di seguito insieme a quella contenente la previsione globale di interesse:

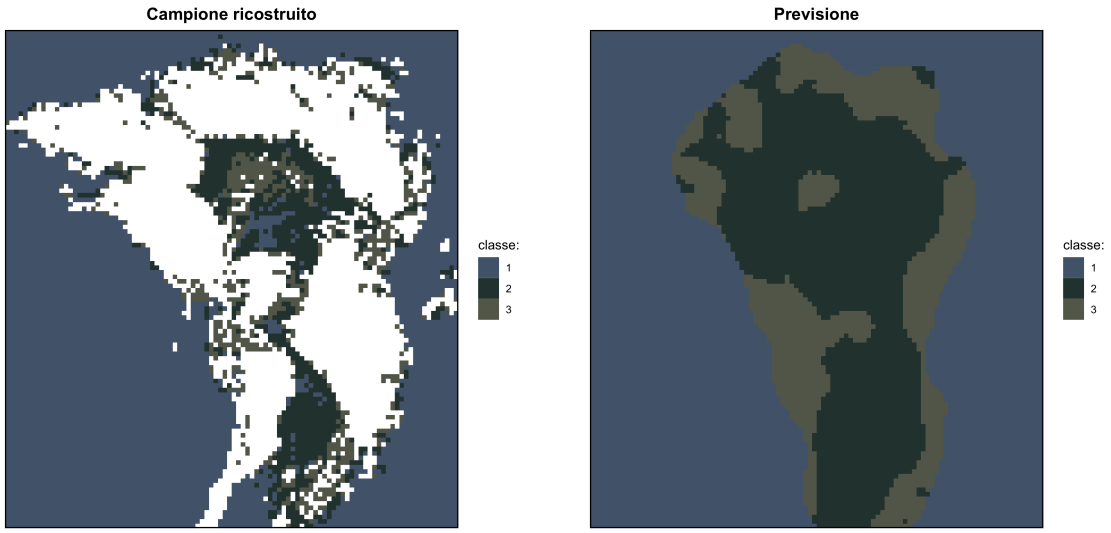

Figura 5.3: Immagine campionata e relativa previsione effettuata con il metodo MCS.

Per poter effettuare una valutazione grafica della bontà delle previsioni in questo caso applicativo, viene riportata anche l'immagine satellitare dell'isola in condizioni di tersità. Si sottolinea la provenienza di tali immagini: nello specifico si tratta del sito web di Google Earth, il quale mette a disposizione fotografie satellitari non disturbate e dalla risoluzione utile per la conduzione di opportuni confronti tra osservazioni e previsioni. In Figura 5.4 è riportata l'immagine dell'isola di La Palma:

Figura 5.4: immagine satellitare di confronto, ottenuta dal sito web di Google Earth.

Come si può notare, i pattern fisici principali vengono riprodotti dalle previsioni, come la zona vulcanica circolare nella parte settentrionale o la sagoma delle aree verdi che si sviluppano intorno a questa e per tutta l'estensione dell'isola.
Si sottolinea anche come, a differenza dei casi simulativi considerati nel capitolo precedente, la distribuzione dei dati mancanti sia, in questo caso, peculiare. Infatti, le nubi rappresentano zone di previsione piuttosto estese piuttosto che pixel sparsi nell'immagine.

5.3 Incendi in California

La seconda immagine perturbata presa in analisi raffigura una porzione di territorio a nord-est di Sacramento, California, che include il lago Tahoe. La foto satellitare, scattata nel corso del 2021 e ottenuta dal sito web di "SciTech Daily", mostra una discreta porzione del suolo coperta dai fumi provocati dai numerosi incendi che hanno segnato la regione nel corso dell'anno (SciTech Daily 2021).
Nella fattispecie, si osserva come la nube occupi larga parte della porzione destra dell'immagine, coprendo per buona parte lo stesso lago e la zona occupata dalla città di Carson City, più a est. Di seguito, in Figura 5.5, viene rappresentata l'immagine originale, punto di partenza per la previsione della composizione del suolo al di sotto della nube di fumo:

Figura 5.5: immagine originale dell'incendio "Caldor" nella zona a nord-est di Sacramento, California.

Anche in questo contesto, la definizione dell'immagine risulta essere troppo alta per poter essere considerata nella sua interezza. Per questo motivo se ne riduce la stessa arrivando ad una definizione di 100×100 pixel, dimensioni accessibili anche ai computer tradizionali.
Tramite l'algoritmo CLARA vengono identificati $K_C = 20$ cluster (colori) che

consentono la ricostruzione dell'immagine in modo approssimato ma sufficiente a permetterne il riconoscimento e la comprensione. L'immagine satellitare ricomposta in questo modo assume la conformazione mostrata in Figura 5.6:

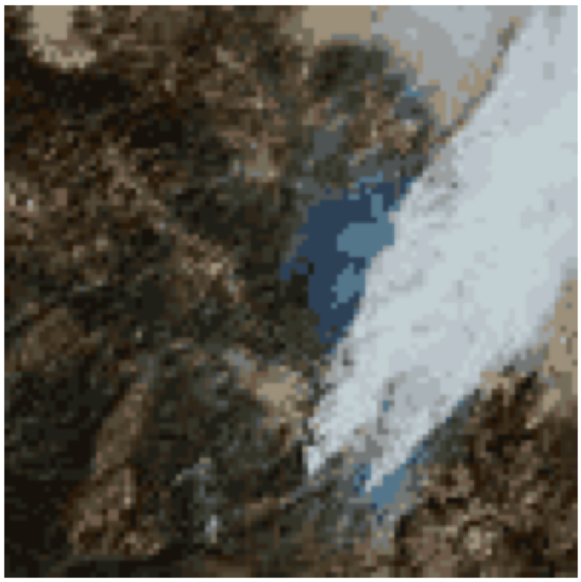

Figura 5.6: immagine ricostruita dopo l'impiego dell'algoritmo CLARA e utilizzando 20 cluster.

Le dimensioni considerate in precedenza producono un'immagine composta da 10000 pixel di tanti colori diversi quanti sono i K_C cluster specificati nell'implementazione dell'algortimo di *hard clustering*.
In questo caso, data la composizione più variegata del territorio rispetto al caso precedente dell'isola di La Palma, l'immagine ottenuta viene resa ancor più grezza attraverso il raggruppamento dei 20 colori in $K = 5$ categorie: verde per le zone erbose e boschive, due tonalità di marrone per zone più brulle o montuose, blu per l'acqua e bianco per le nubi provocate dagli incendi. I pixel di quest'ultima categoria verranno considerati come i dati mancanti in corrispondenza dei quali è interessante ricostruire la conformazione del suolo. Va evidenziato come, a fronte di una diminuzione nel numero di pixel da considerare rispetto al caso precedente, il costo computazionale del modello aumenti a causa del numero più elevato di categorie considerate.
L'immagine satellitare ricostruita utilizzando il raggruppamento in quattro categorie

(più quella ignota) e quella prevista utilizzando l'usuale modello basato sulla massima entropia bayesiana sono riportate in Figura 5.7, appaiate:

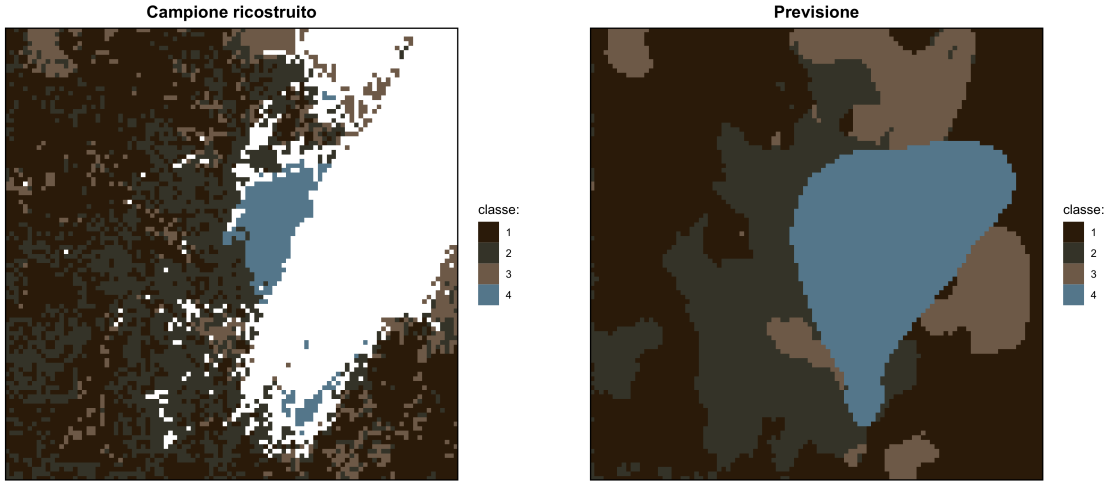

Figura 5.7: Immagine campionata e relativa previsione effettuata con il metodo MCS.

Anche in questo secondo caso viene riportata la mappa ottenuta da Google Earth priva di nubi e catturata in condizioni ottimali (Figura 5.8), utile per capire la qualità complessiva delle previsioni a livello di pattern e utilizzazione del suolo:

Figura 5.8: immagine satellitare di confronto, ottenuta dal sito web di Google Earth.

Osservando sia l'immagine nitida che quella di partenza, si nota come la ricostruzione delle zone montuose e boschive sia sufficientemente precisa. Lo stesso non si può dire per la forma e l'estensione del lago, la cui previsione risulta sensibilmente più grande delle reali dimensioni.
Va altresì notato come, ancora una volta, la zona per cui interessa effettuare previsioni è piuttosto concentrata, rendendo più complicata l'individuazione e la replicazione di trame superficiali.

5.4 Indagine LUCAS sulla provincia di Verona

L'indagine LUCAS è uno studio condotto da Eurostat riguardante la copertura e l'utilizzazione del suolo. Per il caso applicativo descritto nel seguito ci si riferirà alla rilevazione conclusa nel 2018 sull'intero suolo europeo (Eurostat 2018). Più specificamente, si considera la *copertura del suolo*, ovvero la copertura "fisica" osservata in corrispondenza della superificie terrestre. La variabile viene codificata in 8 livelli diversi a seconda delle caratteristiche del territorio: costruito, coltivato, boschivo, arbustivo, erboso, brullo, acquatico e paludoso.
L'area scelta su cui effettuare le previsioni tramite il modello basato sulla massima entropia bayesiana è quella relativa al territorio della provincia di Verona. L'indagine campionaria rileva la presenza di tutti i tipi di suolo possibili, eccezion fatta per le paludi. Va specificato, però, che zone arbustive, aree brulle o caratterizzate dalla presenza di acqua sono decisamente rare; di conseguenza, vengono considerate solo le restanti quattro categorie rilevate con maggior frequenza: aree costruite, coltivate, boschive ed erbose. Più nello specifico, vengono rimosse un paio di osservazioni relative sia alla categoria "territorio acquatico" che a quella "territorio arbustivo"; entrambe le classi presentano un'osservazione nella zona del Lago di Garda e nella parte meridionale della provincia. Nella cosiddetta "bassa veronese", inoltre, sono state registrate anche le otto osservazioni di "territorio brullo", anch'esse rimosse prima della stima dei transiogrammi.
Il campionamento, a livello tecnico, è stato effettuato considerando cerchi di raggio pari a 1.5 metri, con la possibilità di essere allargato a 20 metri nei casi in cui le classi osservate non fossero omogenee.
Questa scrematura rappresenta certamente una perdita di informazione, ma poiché l'area di interessa risulta estesa può essere ugualmente di interesse effettuare previsioni che forniscano un'idea generale delle caratteristiche della copertura del suolo nel veronese.
Va sottolineato anche come le rilevazioni non siano state effettuate su una griglia di punti precisa. Per questo motivo, le operazioni di stima e previsione che seguono sono state anticipate da una trasformazione dei dati in *raster* per poterle interpretare al

meglio. Per prima cosa viene riportata un'immagine che mostra i 243 dati campionati, suddivisi nelle quattro classi precedentemente elencate (Figura 5.9):

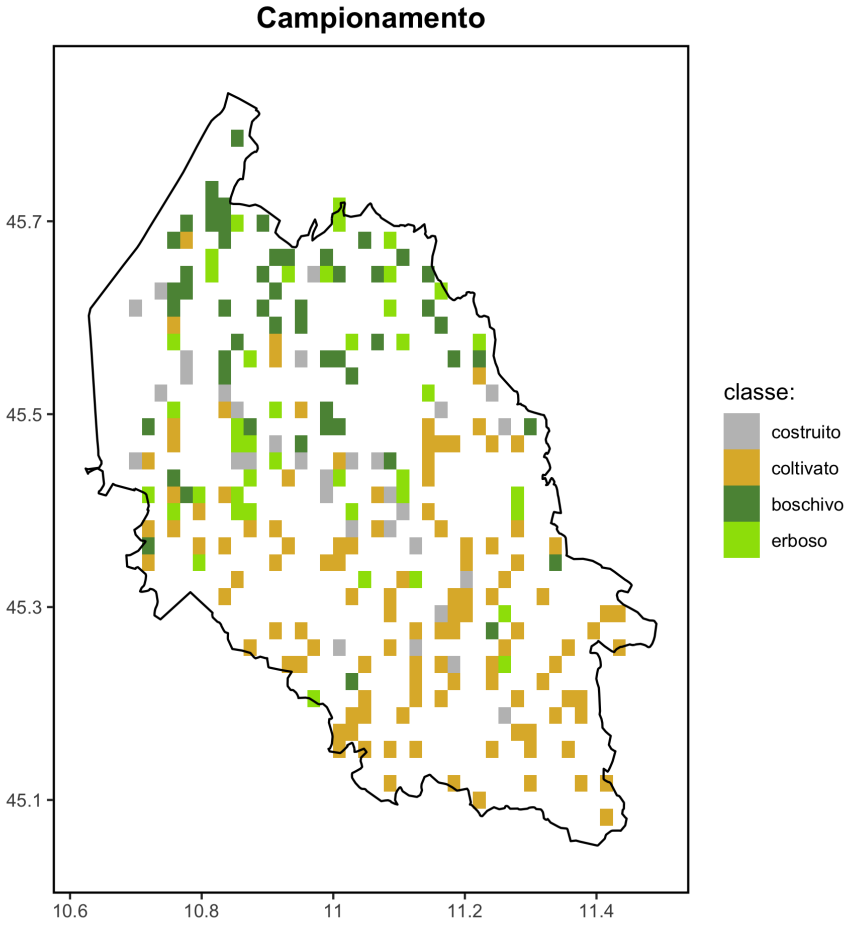

Figura 5.9: copertura del suolo rilevata in 255 luoghi campionati nella provincia di Verona nell'ambito dell'indagine LUCAS 2018 da parte di Eurostat.

Anche ad un'analisi puramente visiva risulta chiaro come l'area considerata sia caratterizzata da zone ben distinte, con particolare riferimento alle aree verdi nella zona a nord e a quelle coltivate nella parte meridionale; le aree urbane appaiono più sparse, fatto che, probabilmente, renderà più difficoltose le operazioni di previsione.
Il modello basato sulla massima entropia viene calcolato nella maniera usuale, ovvero stimando le probabilità di transizione attraverso il metodo della massima entropia e senza effetti anisotropici. Il campo casuale, similmente alle simulazioni svolte nel capitolo precedente, viene simulato con il metodo della simulazione categoriale

multinomiale. La griglia considerata ha dimensioni 50×50, per un totale di 2500 pixel. Le previsioni, private dei pixel esterni ai confini provinciali considerati, vengono mostrate in Figura 5.10:

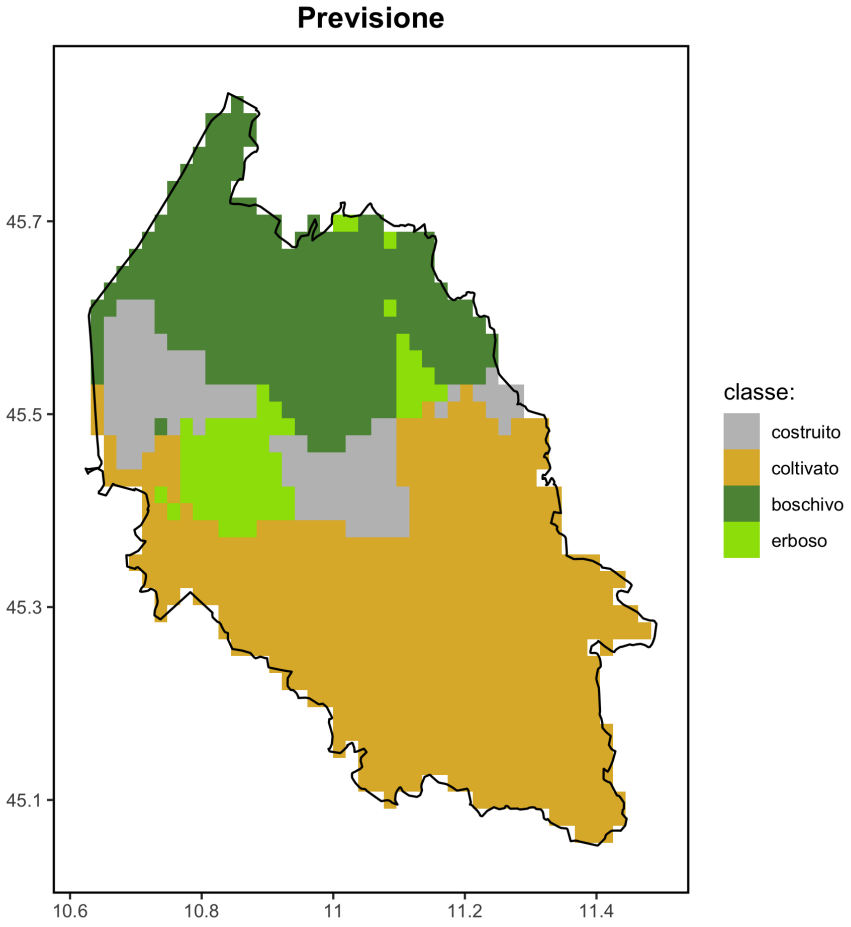

Figura 5.10: copertura del suolo prevista nell'intera provincia di Verona, utilizzando il metodo MCS.

Le previsioni confermano quanto ipotizzato in precedenza, ovvero la presenza di un'ampia area boschiva nella parte settentrionale (in corrispondenza della Lessinia, della Valpolicella e della Valpantena) e di una zona molto estesa di coltivazioni nella bassa veronese, in territorio pianeggiante. Una fascia di transizione intermedia alterna aree urbane (tra cui si distingue in posizione centrale la città di Verona) ad aree a prevalenza erbosa, difficilmente riconducibili a luoghi specifici. Per quanto riguarda le aree edificate, probabile che i pixel grigi a oriente identifichino San Bonifacio, ovvero

uno dei comuni più popolosi della provincia. Diverso il caso della zona a ovest, che potrebbe corrispondere ai comuni in prossimità del Lago di Garda (non considerato tra i dati campionati e nelle previsioni).
L'immagine satellitare di confronto viene riportata di seguito, in Figura 5.11. È comunque importante sottolineare come, da questa, sia difficile distinguere aree boschive da aree erbose.

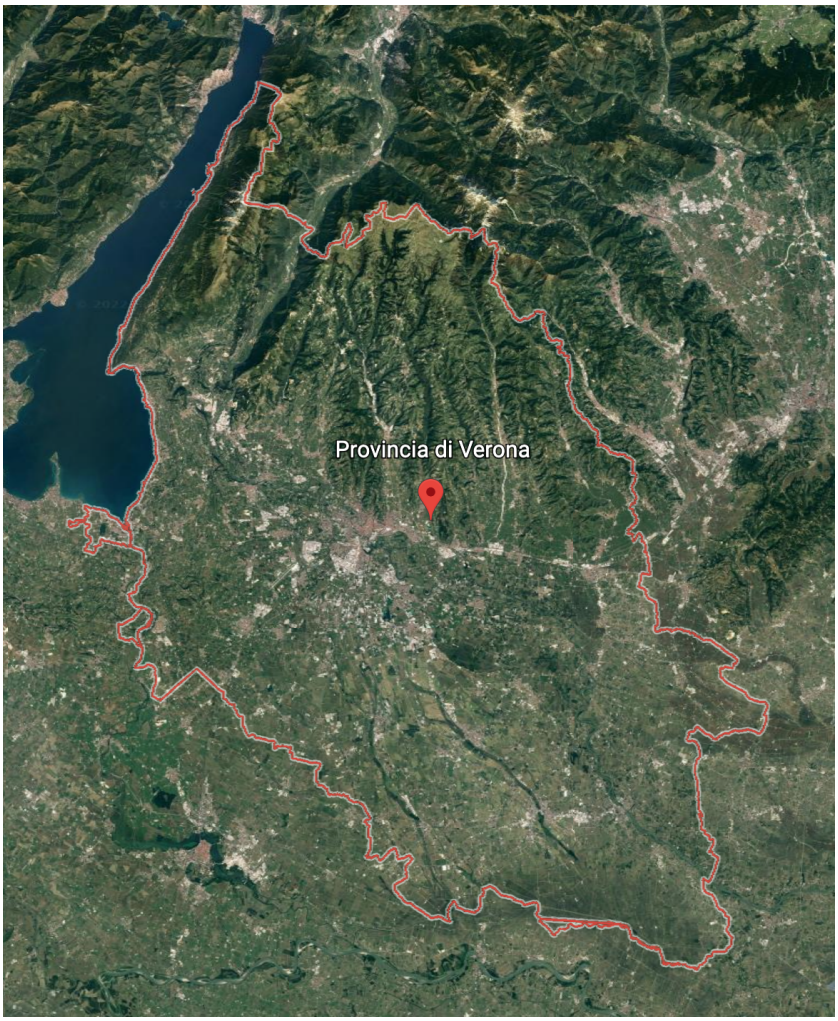

Figura 5.11: immagine satellitare della provincia di Verona ottenuta dal sito web di Google Earth.

Conclusioni

La crescente disponibilità di dati georeferenziati e immagini di superfici o volumi ha fatto incrementare anche la richiesta di opportune analisi descrittive e predittive a partire da contesti spaziali.
Nella moltitudine di dati e applicazioni possibili, il presente lavoro si è concentrato su aspetti specifici e spesso meno esplorati. In particolare, i riferimenti costanti sono stati la trattazione di dati categoriali, le applicazioni relative all'utilizzo e alla copertura del suolo e l'impiego di metodi provenienti dal *framework* dei campi causali di Markov.
La proposta metodologica si è sviluppata mediante l'impiego della massima entropia bayesiana e, più in generale, del metodo *Multinomial Category Simulation* per la modellazione della dipendenza spaziale tra le rilevazioni effettuate in uno spazio bidimensionale. Più nello specifico, è stato messo in evidenza come la stima e l'interpretazione dei transiogrammi, empirici prima e teorici dopo, forniscano un ottimo punto di partenza per la lettura della relazione spaziale che intercorre tra i punti dello spazio considerato. Per semplicità e dal momento che il pacchetto di stima "spMC" (Sartore, Fabbri e Gaetan 2016) del software R la implementa come opzione di default, è stata considerata solo la modellazione esponenziale dei diversi transiogrammi nell'arco di tutta la trattazione.
Per la stima dei parametri di transizione tra classi, inoltre, si è scelto di fare ricorso all'approccio della massima entropia bayesiana, proposto e illustrato nell'articolo di Allard, D'Or e Froidevaux (Allard, D'Or e Froidevaux 2011).
Il metodo della massima entropia, utilizzato in questo contesto per prevedere la conformazione della superficie considerata in condizioni di dati mancanti, è stato confrontato con un algortimo di ricostruzione e regolarizzazione di immagini perturbate

presente in letteratura a partire dagli anni Ottanta e proposto da Besag (Besag 1986): l'algoritmo *Iterated Conditional Modes*.
I risultati di simulazione, messi a confronto, mostrano una performance leggermente migliore per l'algoritmo ICM, ottenuto tramite un approccio non parametrico. Le semplificazioni considerate per il metodo MCS, però, lasciano pensare che opportune specificazioni di effetti direzionali possano migliorare i risultati ottenuti dall'ICM, soprattutto in condizioni di scarsità e "sparsità" di informazione disponibile. Per le simulazioni si sono considerati scenari diversi, variabili per numero di classi considerate, parametro di temperatura β e proporzione di dati mancanti. Tutti i risultati, valutati impiegando l'indice di Rand corretto, si sono rivelati concordanti con le attese: le simulazioni, infatti, mostrano miglioramenti al diminuire della "temperatura" considerata e all'aumentare della quantità di dati noti nei raster. In presenza di un numero maggiore di categorie, infine, si osserva un lieve peggioramento nella qualità predittiva dei modelli.
In conclusione, i risultati ottenuti testando il metodo Multinomial Categorical Simulation sui casi applicativi del capitolo 5 appaiono piuttosto soddisfacenti, considerando la semplicità del metodo e le assunzioni precedentemente elencate. Si sottolinea ancora una volta come nel caso dell'isola di La Palma e degli incendi in California, differentemente da quanto mostrato nelle simulazioni, le porzioni di dati mancanti siano compatte e piuttosto estese, fatto che rende più complicate le previsioni in queste zone. Nel caso dell'indagine LUCAS, la natura dei dati disponibili è più simile a quella mostrata nei casi simulati del capitolo 4; tenendo conto delle semplificazioni condotte in partenza, il metodo riconosce alcuni pattern specifici del territorio veronese, mostrando qualche difficoltà in zone di frontiera disomogenee. Tenendo conto della dimensione dell'area considerata, però, nel complesso i risultati sono soddisfacenti, dal momento che riescono a caratterizzare correttamente le macroaree della provincia di Verona.
In conclusione, la disponibilità di un modello parametrico per l'analisi della dipendenza spaziale nel contesto di dati categoriali permette di trattare con successo le informazioni disponibili e indicizzate da un sistema di coordinate. Trattandosi, inoltre, di un modello dalla grande facilità interpretativa e dalla flessibilità delle sue specificazioni, risulta

potenzialmente trasferibile in altri contesti applicativi, come quelli socio-economico, medico o geologico.

Per il prosieguo della trattazione relativa al metodo MCS, alcuni futuri sviluppi possibili possono riguardare generalizzazioni ed estensioni del modello, ad esempio introducendo effetti anisotropici, diversi tipi di transiogrammi teorici, funzioni non parametriche per la stima degli stessi e, possibilmente, l'introduzione di variabili esogene che consentano una previsione più accurata del processo di interesse.

Bibliografia

Allard, Denis, Dimitri D'Or e Roland Froidevaux (2011). "An efficient maximum entropy approach for categorical variable prediction". In: *European Journal of Soil Science* 62.3, pp. 381–393.

Armstrong, Margaret et al. (2011). *Plurigaussian simulations in geosciences*. Berlino: Springer Science & Business Media.

Banerjee, Sudipto, Bradley P. Carlin e Alan E. Gelfand (2003). *Hierarchical modeling and analysis for spatial data*. Boca Ratón: Chapman e Hall/CRC.

Bárdossy, András (1997). "Introduction to Geostatistics". In: *Institute of Hydraulic Engineering, University of Stuttgart*.

Besag, Julian (1986). "On the statistical analysis of dirty pictures". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 48.3, pp. 259–279.

Cao, Guofeng (2016). "Modeling uncertainty in categorical fields". In: *International Encyclopedia of Geography: People, the Earth, Environment and Technology*, pp. 1–11.

Cao, Guofeng, Phaedon C. Kyriakidis e Michael F Goodchild (2011). "A multinomial logistic mixed model for the prediction of categorical spatial data". In: *International Journal of Geographical Information Science* 25.12, pp. 2071–2086.

Chilès, Jean-Paul e Pierre Delfiner (1999). *Geostatistics: modeling spatial uncertainty*. New York: Wiley.

Cressie, Noel (1993). *Statistics for Spatial Data*. New York: J. Wiley e Sons.

Dimitrakopoulos, Roussos, Hussein Mustapha e Erwan Gloaguen (2010). "High-order statistics of spatial random fields: exploring spatial cumulants for modeling complex non-Gaussian and non-linear phenomena". In: *Mathematical Geosciences* 42.1, pp. 65–99.

Dipartimento della Protezione Civile (2021). *Classificazione sismica*. https://rischi.protezionecivile.gov.it/it/sismico/attivita/classificazione-sismica, ultimo accesso: 10/01/2022.

Eurostat (2018). *LUCAS micro data 2018 - Italy*. https://ec.europa.eu/eurostat/cache/lucas/IT_2018_20200213.CSV, ultimo accesso: 22/02/2022.

Gaetan, Carlo, Paolo Girardi e Roberto Pastres (2017). "Spatial clustering of curves with an application of satellite data". In: *Spatial statistics* 20, pp. 110–124.

Hubert, Lawrence e Phipps Arabie (1985). "Comparing partitions". In: *Journal of classification* 2.1, pp. 193–218.

Istituto Nazionale di Geofisica e Vulcanologia (2021). *Catalogo Parametrico dei Terremoti Italiani (CPTI15, v3.0)*. https://emidius.mi.ingv.it/CPTI15-DBMI15/, ultimo accesso: 11/01/2022.

Li, Weidong (2007a). "Markov chain random fields for estimation of categorical variables". In: *Mathematical Geology* 39.3, pp. 321–335.

Li, Weidong (2007b). "Transiograms for characterizing spatial variability of soil classes". In: *Soil Science Society of America Journal* 71.3, pp. 881–893.

Matheron, Georges (1962). *Traité de géostatistique appliquée*. Parigi: Editions Technip.

National Centers for Environmental Information (2021). *Riepilogo giornaliero di dati climatici per il periodo 01/01/2021 - 30/04/2021 nello stato di Washington, USA*. https://www.ncdc.noaa.gov/cdo-web/search, ultimo accesso: 20/12/2021.

Pardo-Igúzquiza, Eulogio, David I.F. Grimes e Chee-Kiat Teo (2006). "Assessing the uncertainty associated with intermittent rainfall fields". In: *Water resources research* 42.1.

Sartore, Luca, Paolo Fabbri e Carlo Gaetan (2016). "spMC: an R-package for 3D lithological reconstructions based on spatial Markov chains". In: *Computers & Geosciences* 94, pp. 40–47.

Schabenberger, Oliver e Carol A. Gotway (2017). *Statistical Methods for Spatial Data Analysis*. Boca Ratón: Taylor & Francis.

SciTech Daily (2021). *California Continues To Burn – More Than 7000 Wildfires, Burning Over 900,000 Hectares*. https://scitechdaily.com/california-continues-to-

burn-more-than-7000-wildfires-burning-over-900000-hectares/, ultimo accesso: 22/02/2022.

Sky News (2021). *La Palma volcano: New satellite images show violent eruption from space as lava flows across island*. https://news.sky.com/story/la-palma-volcano-new-satellite-images-show-violent-eruption-from-space-as-lava-flows-across-island-12431412, ultimo accesso: 22/02/2022.

Strebelle, Sebastien (2002). "Conditional simulation of complex geological structures using multiple-point statistics". In: *Mathematical geology* 34.1, pp. 1–21.

Tobler, Waldo R. (1970). "A Computer Movie Simulating Urban Growth in the Detroit Region". In: *Economic Geography* 46.2, pp. 234–240.

Wackernagel, Hans (2003). *Multivariate geostatistics: an introduction with applications*. Berlino: Springer Science & Business Media.