

Working Paper Series, N. 7, May 2005



Department of Statistical Sciences
University of Padua
Italy

UNIVERSITÀ
DEGLI STUDI
DI PADOVA
DIPARTIMENTO
DI SCIENZE
STATISTICHE

Erosione della costa di Holderness: costruzione di un modello gerarchico a effetti casuali per dati censurati

Claudia Furlan

Department of Statistical Sciences
University of Padua
Italy

Abstract: An awareness of the possible cliff position at some future date is fundamental to coastal planning and shoreline management, for example to avoid development in vulnerable areas. At the beginning, to predict cliff recession rates deterministic methods have been investigated. Then, recession predictions are being expressed in probabilistic terms. However, so far, only easy models were developed.

We consider the cliff erosion at the Holderness Coast: since 1951 a monitoring program started in 118 stations along the coast, providing an invaluable, but often censored, source of information.

We built hierarchical random effect models, taking account of the known dynamics of the process and including the censored information.

Capire e prevedere il tipo di erosione di una costa è essenziale per redigere un'adeguata pianificazione del territorio, volta ad evitare situazioni catastrofiche per quei centri urbani che sono o saranno in una situazione di pericolo.

Da principio, in letteratura, furono proposti dei metodi deterministici per predire i tassi di erosione costiera. Successivamente, le previsioni sulle recessioni furono espresse in termini probabilistici. Tuttavia, fino ad ora, non è stato sviluppato un modello che permetta di descrivere adeguatamente le dinamiche dell'erosione costiera.

In questo articolo è stata considerata l'erosione della costa di Holderness: dal 1951 è stato avviato un programma di monitoraggio in 118 postazioni, lungo la costa, che fornisce un'inestimabile, ma spesso censurata, fonte di informazione. Si sono costruiti alcuni modelli gerarchici a effetti casuali che potessero tener conto delle dinamiche di base del processo erosivo e inglobare l'informazione censurata.

Keywords: Hierarchical Model, Monte Carlo Markov Chain, Reversible Jump Monte Carlo Markov Chain, Coastal Erosion, Holderness Coast.

Final version (2005-05-05)

Indice

1	Introduzione	2
2	L'erosione costiera	3
2.1	Il caso della costa di Holderness	4
2.2	Alcuni approcci di previsione utilizzati finora	4
3	Descrizione dei dati	5
4	Dallo studio dell'erosione alla modellistica	8
4.1	Dinamiche del fenomeno	8
4.2	Diversa strutturazione dei dati	11
4.3	Resistenze e danni cumulati	13
5	Un modello gerarchico bayesiano a effetti casuali	13
5.1	Modellazione dell'intensità delle frane tramite la distribuzione Gamma	13
5.1.1	Alcune simulazioni	15
5.2	Modellazione dell'intensità delle frane tramite la distribuzione di Pareto Generalizzata	16
5.2.1	Alcune simulazioni	18
6	Algoritmo MCMC	19
6.1	Distribuzione Gamma per le intensità delle frane	20
6.2	Distribuzione di Pareto Generalizzata per le intensità delle frane	21
7	Studi di simulazione	21
7.1	Tramite la distribuzione Gamma per l'intensità delle frane	22
7.2	Tramite la distribuzione di Pareto Generalizzata per l'intensità delle frane	22
8	Applicazione al Sito 33	24
8.1	Utilizzo di una soglia	25
8.2	Utilizzo dei dati originari	27
9	Cambio della relazione tra danni cumulati e frane: dall'esponenziale all'identità	30
9.1	Studio di simulazione	32
9.2	Applicazione al Sito 33	34
10	Uso dell'informazione censurata dei dati mancanti	36
11	Prossimi sviluppi	43

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
<http://www.stat.unipd.it>

Corresponding author:
Claudia Furlan
tel: +39 049 827 4192
furlan@stat.unipd.it

Erosione della costa di Holderness: costruzione di un modello gerarchico a effetti casuali per dati censurati

Claudia Furlan

Department of Statistical Sciences

University of Padua

Italy

Abstract: An awareness of the possible cliff position at some future date is fundamental to coastal planning and shoreline management, for example to avoid development in vulnerable areas. At the beginning, to predict cliff recession rates deterministic methods have been investigated. Then, recession predictions are being expressed in probabilistic terms. However, so far, only easy models were developed.

We consider the cliff erosion at the Holderness Coast: since 1951 a monitoring program started in 118 stations along the coast, providing an invaluable, but often censored, source of information.

We built hierarchical random effect models, taking account of the known dynamics of the process and including the censored information.

Capire e prevedere il tipo di erosione di una costa è essenziale per redigere un'adeguata pianificazione del territorio, volta ad evitare situazioni catastrofiche per quei centri urbani che sono o saranno in una situazione di pericolo.

Da principio, in letteratura, furono proposti dei metodi deterministici per predire i tassi di erosione costiera. Successivamente, le previsioni sulle recessioni furono espresse in termini probabilistici. Tuttavia, fino ad ora, non è stato sviluppato un modello che permetta di descrivere adeguatamente le dinamiche dell'erosione costiera.

In questo articolo è stata considerata l'erosione della costa di Holderness: dal 1951 è stato avviato un programma di monitoraggio in 118 postazioni, lungo la costa, che fornisce un'informazione inestimabile, ma spesso censurata, fonte di informazione.

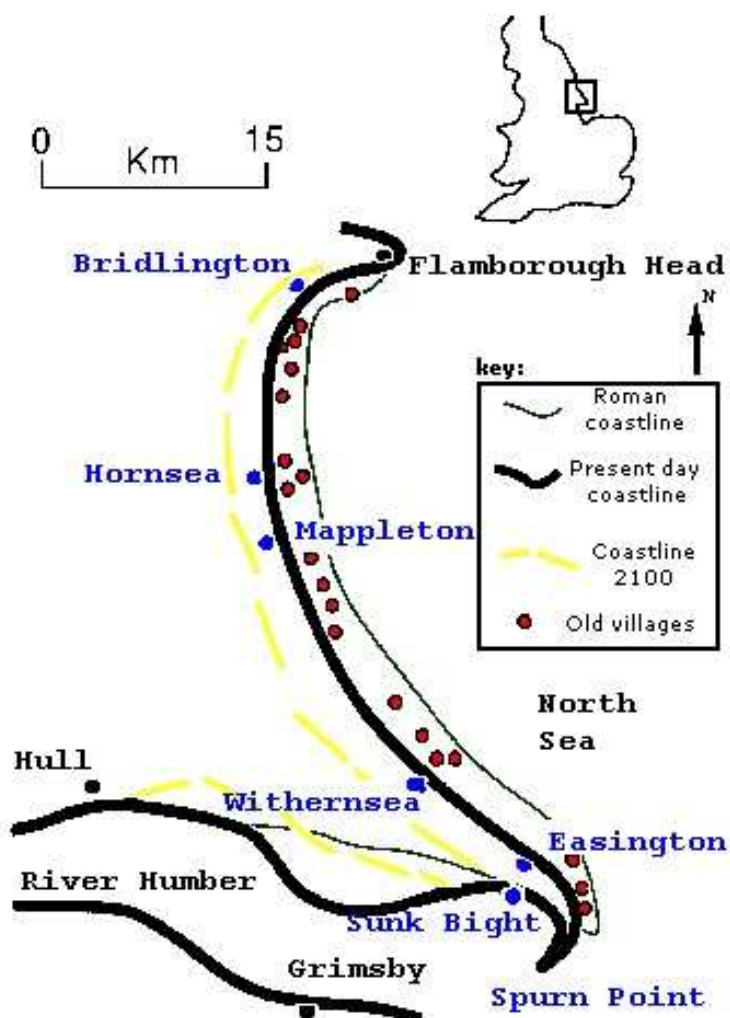
Si sono costruiti alcuni modelli gerarchici a effetti casuali che potessero tener conto delle dinamiche di base del processo erosivo e inglobare l'informazione censurata.

Keywords: Hierarchical Model, Monte Carlo Markov Chain, Reversible Jump Monte Carlo Markov Chain, Coastal Erosion, Holderness Coast.

1 Introduzione

La costa di Holderness si colloca nell'est dello Yorkshire, in Gran Bretagna, estendendosi da Bridlington (a nord) a Spurn Point (a sud), per un totale di 45 chilometri. Essa è costituita da scogliere, formatesi durante l'Era glaciale, di altezza che va dai 3 ai 35 metri, ed è una delle coste a più alto tasso di erosione in Europa. Negli ultimi 2000 anni la linea costiera è arretrata di circa 400 metri e dall'epoca dell'Impero

Figura 1: La costa di Holderness: situazione all'epoca dell'Impero Romano (linea continua sottile), odierna (linea continua spessa) e prevista per il 2100 (linea tratteggiata in grigio) con l'attuale tasso di erosione. I tondini neri indicano la posizione degli antichi agglomerati urbani scomparsi.



Romano sono scomparsi nel mare più di 30 agglomerati urbani. Se non si interverrà si prevede che nel 2100 le principali città costiere saranno destinate a scomparire, come si può vedere nel Grafico 1. Queste informazioni sono rese disponibili nel sito web <http://www.nerc.ac.uk/lois/coast1.shtm>.

2 L'erosione costiera

L'erosione è quel processo per il quale le rocce costiere vengono frantumate dall'azione del mare e trasportate in mare, o lungo la costa, dalle onde e dal vento.

L'azione delle onde dipende dalla loro dimensione, che è in relazione al vento e al tratto di mare che devono percorrere prima di raggiungere la costa più vicina: maggiori sono la distanza da percorrere e la velocità del vento e più le onde sono impetuose.

Quando un'onda si avvicina al fondale poco profondo vicino alla costa, l'attrito con il letto del mare provoca un rallentamento della sua base, mentre la parte superiore dell'onda stessa si muove più velocemente, aumentando in altezza e rompendosi sulla spiaggia.

Le onde possono essere costruttive o distruttive. Quelle costruttive sono basse in proporzione alla lunghezza e sono caratterizzate da un forte sciabordio e una debole risacca; quelle distruttive sono alte in proporzione alla lunghezza e sono caratterizzate da un debole sciabordio e una forte risacca.

L'erosione è più forte quando le onde si rompono ai piedi delle scogliere perché, in questo caso, è la base della scogliera che viene erosa, ivi formando una nicchia. Più la nicchia aumenta di dimensione e maggiore è il peso della roccia sopra di essa, portandola al franamento.

Queste informazioni si possono trovare nel sito web

<http://www.geography.learnontheinternet.co.uk/topics/waves.html>.

2.1 Il caso della costa di Holderness

Un milione di anni fa, la costa dello Yorkshire era formata da un tratto di scogliere di calcare e si trovava a più di 32 chilometri a ovest di dove si trova ora. Durante l'Era glaciale si addossarono a queste scogliere dei depositi di argilla, formando il nuovo litorale. È proprio questa argilla mista a sabbia che è facilmente erodibile e i detriti che si formano sono facilmente spazzati via dal mare, anziché fornire protezione alle scogliere contro l'effetto delle onde. È stimato che ogni anno 10^6m^3 di detriti vengano assorbiti dal mare, e di questi solo $0.28 \cdot 10^6 \text{m}^3$ siano trasportati lungo la costa dalle onde, mentre il rimanente venga portato al largo. Il tasso medio di erosione della costa è di 1.82m annui (Pethick and Legget, 1993). Tuttavia, la costa non si erode tutta allo stesso tasso. La parte nord è caratterizzata da scogliere alte e ripide, cavità e archi: esse sono il risultato di un calcare relativamente più resistente e, quindi, con minor tendenza a franare. La parte sud di questa costa, invece, è caratterizzata da depositi glaciali non consolidati formati da argilla, sabbia e ghiaia. Anche qui la linea costiera è costituita da scogliere, ma la loro altezza è minore di quelle della parte nord, visto che la loro composizione porta a continue frane.

Le violente tempeste marine che si scatenano lungo la costa giocano un ruolo molto importante nel fenomeno dell'erosione costiera, perché danneggiano le rocce fino a farle cadere (Hall *et al.*, 2002).

2.2 Alcuni approcci di previsione utilizzati finora

Capire e prevedere il tipo di erosione di una costa è essenziale per redigere un'adeguata pianificazione del territorio, volta ad evitare situazioni catastrofiche per quei centri urbani che sono o saranno in una situazione di pericolo. Nel mondo ci sono

parecchie coste che sono o che dovrebbero essere tenute sotto controllo; ad esempio, sono stati riscontrati tassi medi annui di erosione costiera superiore a 1 metro in molte nazioni tra cui Danimarca, Germania, Unione Sovietica, Giappone, Nuova Zelanda, Canada, Gran Bretagna e Stati Uniti.

Da principio, in letteratura, furono proposti dei metodi deterministici per predire i tassi di erosione costiera, relazionandoli all'energia delle onde, al livello dell'acqua e alla resistenza delle rocce (Kamphuis, 1987). Questi metodi deterministici possono essere criticati perché non contemplano l'incertezza del processo erosivo. Per esempio, gli aumenti delle recessioni costiere spesso capitano in concomitanza alle tempeste marine, ma una frana non è la conseguenza necessaria dell'arrivo di una tempesta: per cadere la rupe deve essere già in stato di deterioramento. Successivamente, le previsioni sulle recessioni furono espresse in termini probabilistici: per un esempio si veda Hall *et al.* (2002). Tuttavia, fino ad ora, non è stato sviluppato un modello che permetta di descrivere adeguatamente le dinamiche dell'erosione, in particolare la possibile relazione tra l'intensità delle frane e il tempo che intercorre tra esse.

3 Descrizione dei dati

Per programmare l'utilizzo del litorale e per evitare di creare dei centri urbani in zone a rischio, è fondamentale prevedere quale sarà la posizione delle scogliere nel futuro. Per questo motivo il Consiglio del Distretto locale ha avviato, dal 1951, un programma di misurazione della recessione costiera. Lungo la costa furono installati 118 postazioni intervallate le une dall'altre di 500 metri e posizionate a una distanza tra 50 e 100 metri dalla costa (Pethick, 1996). La prima postazione a nord si trova a Sewerby e l'ultima a sud si trova a Spurn. Le misurazioni riguardano la recessione della parte alta delle scogliere e rappresentano la distanza della rupe dalla relativa postazione: dal 1951 al 1992 la frequenza di rilevazione è annuale (usualmente Marzo) e dal 1993 al 2002 semestrale (usualmente Marzo e Settembre).

Il dataset, fornito da Jim Hall del dipartimento di Ingegneria Civile di Newcastle (UK), rappresenta un'inestimabile fonte di informazione sia temporale che spaziale. I dati, tuttavia, debbono essere trattati con una certa cautela per i seguenti motivi:

1. nello stesso dataset coesistono misurazioni annuali e semestrali e, al momento, quelle semestrali sono state unificate in annuali. Il problema dell'unificazione sorge quando uno dei due valori è mancante; in questo caso, è stato considerato come annuale il valore non mancante. Questa scelta non porta ad una eccessiva distorsione. Infatti, se il valore mancante è il primo (quello di marzo per intendersi) la recessione indicata nel secondo valore (quello di settembre) sarà cumulativa di tutto l'anno e l'informazione resta completa. Quando, invece, è il valore di settembre a mancare si ha una distorsione dell'informazione, perché si registrano, per quell'anno, le erosioni avvenute solo fino a Marzo; tuttavia, tale distorsione risulta lieve perché la maggior parte della recessione avviene durante l'inverno a causa delle violente tempeste marine.

2. Esistono alcuni errori sistematici: le misurazioni non registrano, infatti, eventi di frequenza più alta di quella annuale o semestrale e su una scala spaziale inferiore ai 500 metri (Pethick, 1996).
3. Originariamente le misurazioni erano registrate nel sistema di misura inglese e probabilmente si è verificata della confusione passando al sistema metrico, anche per la conversione dei valori registrati negli anni antecedenti (Pethick, 1996).
4. Il tasso dei valori mancanti è alto e spesso la mancanza di misurazioni si verifica per più anni consecutivi.
 Fortunatamente, anche dai dati mancanti si possono trarre delle informazioni: per esempio, se di un anno manca la misurazione della recessione, il valore dell'anno seguente indicherà la recessione cumulata in quei due anni. Questo vale anche per un blocco di valori mancanti: la prima misurazione disponibile indicherà la recessione cumulata durante quegli anni in cui non si hanno avuti i valori. Di conseguenza, se un periodo di valori mancanti è seguito da uno zero, si possono sostituire tali valori mancanti con uno zero; così facendo si recupera dell'informazione e si riduce il tasso di valori mancanti.
 Queste informazioni vengono invece perse quando il valore mancante esprime la situazione di postazione danneggiata (per esempio, per una frana). In questa situazione, che si verifica per lo più nelle prime 31 postazioni, i valori mancanti rappresentano una situazione particolare da cui non si può evincere nulla e che va a falsare anche il totale della recessione costiera in quel sito.
 I due tipi di valori mancanti, appena descritti, hanno due diverse codifiche nel dataset.
5. È impossibile pensare che la precisione di rilevazione sia la stessa per una così vasta gamma di postazioni e dal 1951 fino ad oggi.

Il dataset si presenta piuttosto complicato per le situazioni che vi compaiono e che sono state brevemente descritte qui sopra. Occorre, quindi, conoscerlo in ogni dettaglio per poter sviluppare un qualsiasi modello. Si è deciso di approfondire, oltre alla presenza dei dati mancanti, i seguenti due aspetti per capire le analogie e le differenze tra siti:

1. l'attività erosiva media e totale (fino all'ultima rilevazione disponibile);
2. il fenomeno di pausa dall'attività erosiva, che è caratterizzato dalla presenza di 0 o più 0 consecutivi, che chiameremo *blocchi di 0*.

Per quanto riguarda il tasso dei valori mancanti, dato che non tutte le postazioni furono monitorate dal 1951, si è deciso di non considerare come dati mancanti gli anni precedenti all'effettiva partenza del monitoraggio.

Analizzando il Grafico 2, si capisce come il tasso di valori mancanti sia piuttosto alto un pò in tutti i siti e la sua media è pari a 23.5%. In particolare, i tassi che vanno dal sito 11 al sito 31 sono sempre più alti del valore medio. La spiegazione sta nel fatto che per la maggior parte di questi siti non sono state prese le misurazioni tra

Figura 2: Tasso di valori mancanti per ogni postazione. La prima postazione si trova a nord della costa e l'ultima a sud. Il tasso medio tra le postazioni è 23.5% ed è rappresentato dalla linea orizzontale tratteggiata.

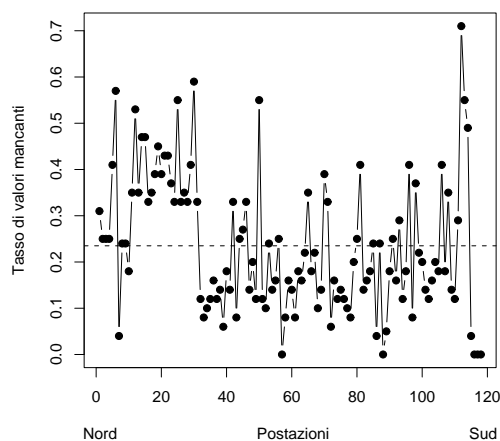
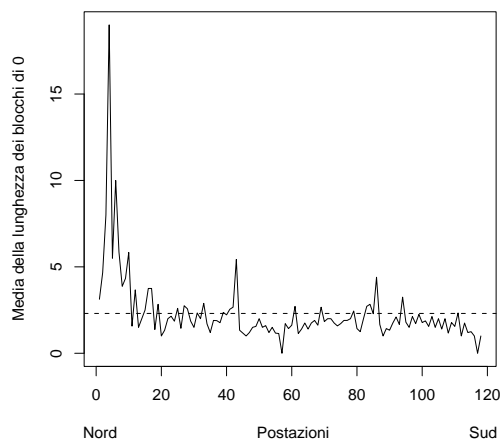


Figura 3: Media della lunghezza dei blocchi di 0 per ogni postazione da nord a sud. La media tra i siti è 2.3 anni ed è indicata dalla linea orizzontale tratteggiata.



gli anni 1971 e 1983; inoltre, in questo ultimo anno le postazioni furono danneggiate con la conseguente perdita di tutte le informazioni.

Nel Grafico 3 vengono riportate le lunghezze medie dei blocchi di zero. Si può notare come tali medie siano più alte del valore medio (2.3 anni) per i primi siti del

nord e che dal sito 44 si abbia la tendenza a stare sotto tale valore medio. Sembra, quindi, che al nord le frane siano in genere più intervallate da anni di pausa di quelle del sud; questo va a conferma della composizione geologica (più compatta al nord e più franante al sud) discussa nella Sezione 2.1.

Per il calcolo di tali medie i valori mancanti hanno giocato un ruolo fondamentale; si poteva scegliere, infatti, tra le seguenti due possibilità:

1. considerare un blocco di 0 come una serie di 0 tra due valori che non fossero mancanti, tralasciando, quindi, anche eventuali blocchi all'inizio e alla fine delle serie storiche dei siti;
2. non tener conto dei dati mancanti e contare i blocchi di 0 presenti nelle serie storiche.

I due metodi sono piuttosto equivalenti dal sito 32 in poi, ma nei siti precedenti inducono delle forti differenze nei risultati. Per cogliere l'importanza della questione può essere utile il seguente esempio: il sito 3 comincia con un blocco di 21 zeri, seguiti da 10 valori mancanti e finisce con un blocco di 4 zeri. Con il primo metodo si tralascia sia il blocco iniziale che quello finale e la media che ne risulta è 3.5. Con il secondo metodo, invece, la media è 8. Viene scelto il secondo metodo, perché il primo porta a sottostimare eccessivamente la lunghezza media dei blocchi di 0.

Il Grafico 4 conferma l'ipotesi di una struttura spaziale per l'erosione costiera nella costa di Holderness. Nella prima riga vengono riportati i tassi di 0 e il numero di 0 per ogni sito, i cui valori medi tra i siti sono rispettivamente 38.39% e 14.38%. Nel calcolo dei tassi di 0, i denominatori corrispondono al numero di anni dalla prima misurazione sottratto il numero di anni con valori mancanti.

Nella seconda riga vengono riportati le medie delle erosioni annuali e i totali delle recessioni costiere fino all'ultima rilevazione disponibile, per ogni sito. I valori medi tra i siti sono rispettivamente 1.28m e 57.73m. Nel calcolo delle medie delle erosioni annuali il denominatore è il numero di anni dalla prima misurazione. I valori mancanti non hanno influenza perché, dopo essi, la misurazione è cumulativa del periodo non registrato, a meno che la postazione non venga danneggiata con una conseguente sottostima della media.

La parte nord è caratterizzata da erosioni più piccole e più distanziate nel tempo mentre la parte sud da erosioni più forti e più ravvicinate.

Nel Grafico 5 sono riportati i valori massimi delle erosioni annuali per ogni sito; il valore medio tra le postazioni è 9.95m. Sembra esserci un trend positivo muovendosi verso sud; di conseguenza, l'erosione costiera aumenta muovendosi verso sud.

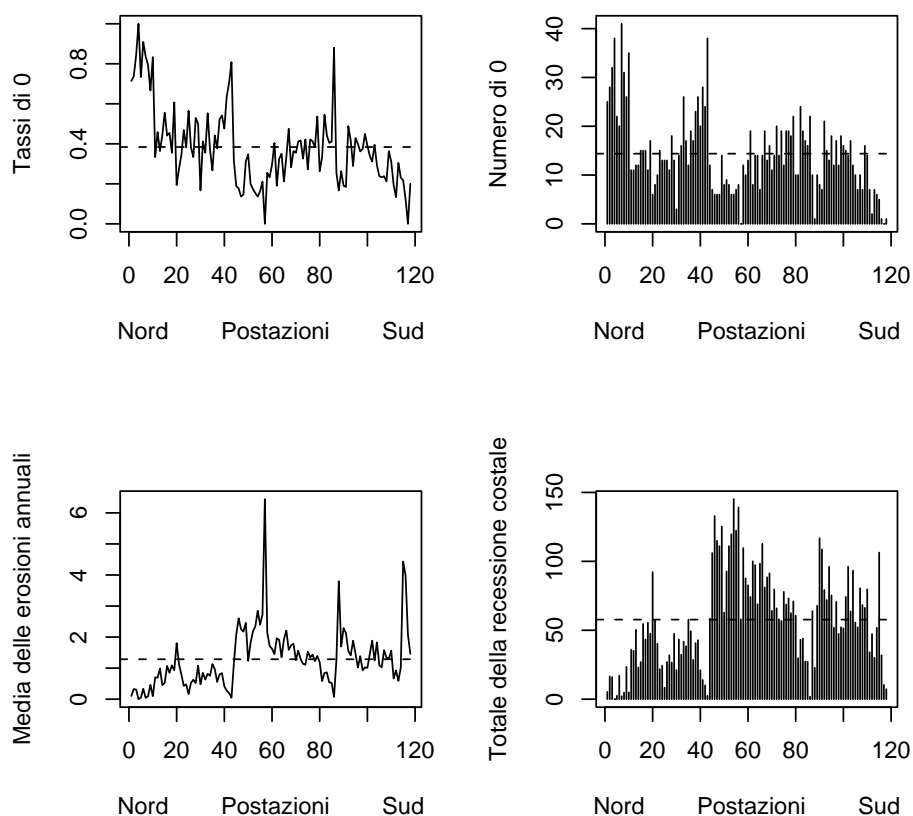
Per quanto riguarda i valori massimi, occorre ricordare che nelle serie storiche le erosioni registrate dopo valori mancanti sono cumulative e quindi sono più alti.

4 Dallo studio dell'erosione alla modellistica

4.1 Dinamiche del fenomeno

Il fenomeno dell'erosione è molto complesso: si è deciso, quindi, di tener conto solo degli aspetti più importati, per riuscire a formulare un modello iniziale. Ora,

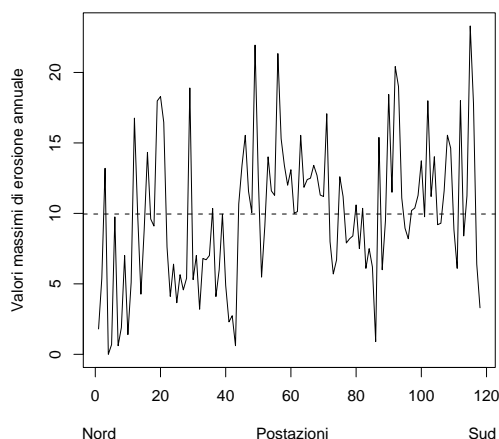
Figura 4: Blocchi di 0 e di frane. Le linee tratteggiate indicano i valori medi. Nella prima riga sono rappresentati i tassi di 0 per sito (il valore medio è pari a 38.39%) e il numero totale di 0 per sito (il valore medio è pari a 14.38). Nella seconda riga, invece, sono rappresentati le medie delle erosioni annuali (il valore medio è 1.28m) e i totali delle erosioni costiere (il valore medio è 57.73m).



in sintesi, verranno descritte le caratteristiche principali del fenomeno dell'erosione costiera per la costa di Holderness (vedasi il Grafico 6):

- le rocce hanno una forza coesiva che cerca di impedire alle rupi di franare (*effetto geologico*).
- Esse sono continuamente soggette a danni, sia per il movimento del mare che per le tempeste marine, che sono particolarmente forti e frequenti durante l'inverno (*effetto meteorologico*).
- L'erosione diviene il risultato di una competizione tra due processi: la forza

Figura 5: Valori massimi di erosioni annuali per le postazioni da nord a sud. Il valore medio tra i siti è 9.95m ed è indicato con la linea orizzontale tratteggiata.



coesiva della roccia (*effetto geologico*) e l'effetto danneggiante del mare (*effetto meteorologico*).

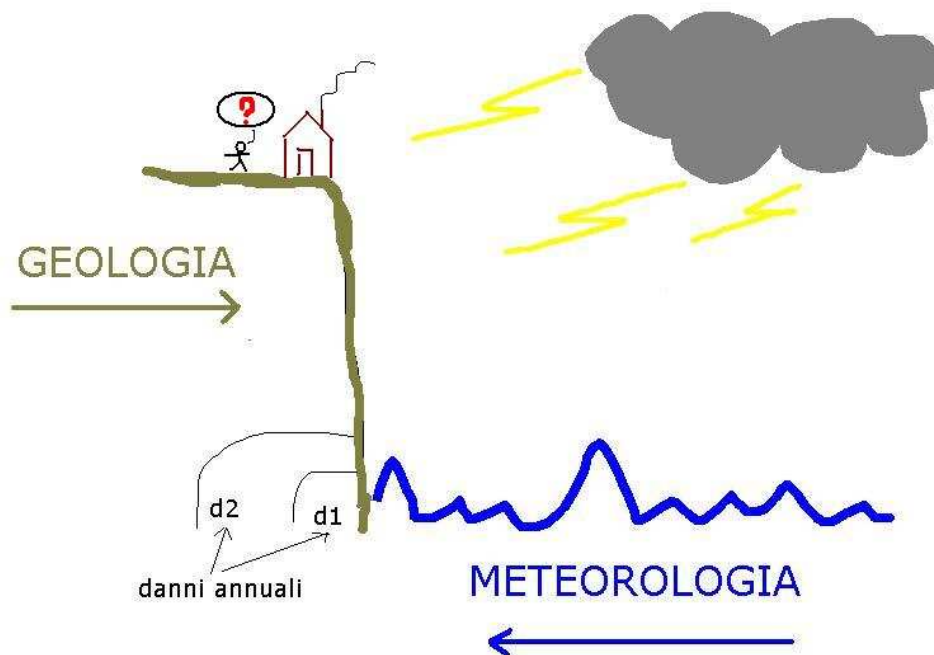
- Il moto ondoso, rinvigorito dall'effetto meteorologico, scontrandosi contro la base delle scogliere, forma una cavità che si ingrandisce con il tempo (*danno cumulato* negli anni).
- Quando il peso della roccia sopra la cavità erosa è eccessivo si verifica la frana.

Dopo queste premesse, si passi ad osservare la serie storica di un sito per estrapolare informazioni dai dati e integrarla con i concetti già visti sull'erosione; questo è un buon metodo per specificare un modello da cui si possano simulare le dinamiche dell'erosione vedendo se corrispondono alle dinamiche osservate. Nel Grafico 7 viene rappresentato il monitoraggio del Sito 33, che si trova a Low Skirlington (parte nord della costa). Viene scelto questo sito perché i valori mancanti sono solo 4, di cui 3 all'inizio della serie storica: è possibile, quindi, eliminare le prime osservazioni per le prime analisi, senza dover tener conto dei valori mancanti. Inoltre, esso cattura le caratteristiche dei siti della parte nord della costa. Osservando la dinamica della spezzata, si può notare come dopo i periodi di pausa, rappresentati dagli 0 o da valori molto prossimi a 0¹, ci sia una certa evidenza ad osservare frane di intensità elevata. Per questo sito, sembra che:

1. per ogni anno di pausa che passa dall'ultima frana la cavità alla base della scogliera aumenta,
2. più grande è la cavità, più grande è la massa sopra il vuoto e quindi più grande è il danno cumulato,

¹Su questo punto si ritornerà nella Sezione 8.

Figura 6: L'erosione come il risultato di due forze contrastanti. L'effetto del mare insieme, a quello meteorologico, danneggia la roccia formando una cavità alla base che si ingrandisce sempre più, fino a che il peso sopra il vuoto è eccessivo e si verifica la frana.



3. l'entità della frana sembra dipendere dall'entità del danno cumulato.

L'alta resistenza della roccia² implica che sia necessario un alto numero di anni per poter osservare una frana, con un conseguente alto danno cumulato; queste rocce, quindi, hanno un'inclinazione a franare meno frequentemente.

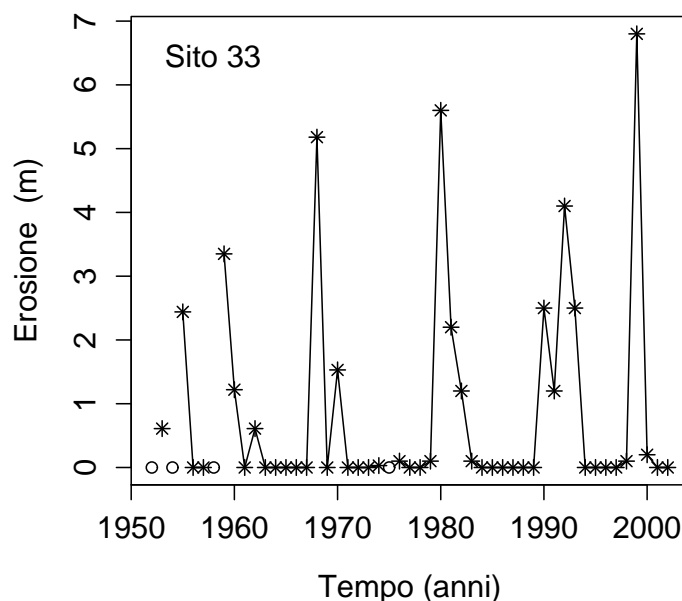
4.2 Diversa strutturazione dei dati

Le serie storiche, così come sono state costruite, non permettono di modellare le frane (valori positivi) diversamente dal tempo di attesa tra di esse (blocchi di zero). Per tenere in considerazione la differenza concettuale tra una misurazione di valore zero e una misurazione di valore positivo si è deciso di assegnare una diversa strutturazione ai dati.

Si chiami F_w , $w = 1951, \dots, 2002$, la registrazione avvenuta nell'anno w per un particolare sito: questa serie storica rappresenta i dati che si hanno a disposizione. Si ipotizzi, quindi, di essere nell'anno 1951 in un'ipotetica postazione e che si verifichi

²Relativamente all'entità della resistenza della parte sud della costa.

Figura 7: Serie storica del Sito 33. Negli anni in corrispondenza dei tondini non si hanno avute le rilevazioni e ivi la spezzata si interrompe.



una frana di entità 4m: tale valore, secondo il programma di monitoraggio condotto, verrebbe registrato come F_{1951} . Alternativamente, si potrebbe scegliere di indicarla con S_1 , cioè come l'entità della prima frana avvenuta. Quanti anni sono passati dalla precedente frana? Essendo l'inizio della serie storica si applica la convenzione che l'ultima frana sia successa l'anno prima, cioè il 1950. Quindi, non essendoci stati anni di pausa tra le due frane, il numero di anni associati alla prima frana è $N_1 = 0$. Si supponga, ora, che dopo il 1951 ci siano voluti 2 anni di pausa prima di osservare la successiva frana, di entità pari a 7m; secondo il programma di monitoraggio si avrebbe: $F = (F_{1951}, F_{1952}, F_{1953}, F_{1954}) = (4, 0, 0, 7)$. Alternativamente, come fatto prima, si può dire che la seconda frana è stata di entità 7, cioè $S_2 = 7$ e che gli anni di pausa sono stati $N_2 = 2$. Ora, i dati in F possono essere scritti come $S = (4, 7)$, $N = (0, 2)$. Se l rappresenta il numero di frane verificatesi durante il periodo temporale 1951-2002, i vettori S, N saranno di lunghezza pari a l . Si è, quindi, riusciti a formulare un modo alternativo per la registrazione degli eventi, che permette di isolare i due processi di erosione e del tempo di attesa, entrambi rilevanti nello studio della dinamica della recessione costiera.

4.3 Resistenze e danni cumulati

Si è parlato spesso della resistenza della roccia, del danno cumulato provocato dal mare e di come la frana si verifichi quando il danno cumulato vince sulla resistenza della roccia. Non si conoscono, però, queste quantità; tuttavia, si potrebbe inserirle in un modello come delle quantità ignote.

Si continui l'esempio della Sezione 4.2; si immagini di essere nel 1950 e che si sia verificata la frana fittizia, di cui si è parlato: un attimo dopo il verificarsi di questa frana la roccia produrrà una resistenza che viene indicata con R_1 . Tale forza sarà quella che dovrà essere sconfitta per poter osservare la frana del 1951, che è stata indicata con S_1 . Visto che la frana fittizia e quella del 1951 sono consecutive, vuol dire che un solo danno annuale d_{11} è bastato per sconfiggere R_1 e quindi si identificherà con il danno cumulato D_1 . Notare come i pedici di N, R, D corrispondano a quello di S .

L'aspetto importante è che ad ogni frana la competizione tra resistenza e danno cumulato si azzerava e si riparte da capo. Di conseguenza, le frane sono indipendenti tra loro, ma dipendenti dal passato per un numero di anni pari al periodo di pausa precedente.

Si consideri la seconda frana: un attimo dopo il verificarsi di S_1 , nel 1951, si è sviluppata R_2 , la resistenza che dovrà essere sconfitta dal danno cumulato D_2 , necessario per osservare la frana S_2 del 1954. Visto che gli anni di pausa tra S_1 e S_2 sono stati 2, si avranno 3 danni annuali d_{21}, d_{22}, d_{23} e il danno cumulato sarà la loro somma $D_2 = d_{21} + d_{22} + d_{23}$. E così via per il resto della serie storica.

Quindi, ad ogni frana $S_i, i = 1, \dots, l$ della serie storica corrisponderanno N_i, R_i, D_i . Si può anche quantificare la differenza tra il danno cumulato e la resistenza come $X_i = D_i - R_i, i = 1, \dots, l$.

Si veda il Grafico 8 per una migliore visualizzazione delle dinamiche del processo di lotta tra resistenza e danno cumulato, nel caso che la frana i -esima accada dopo 2 anni di pausa dalla precedente (come la frana del 1954 in questo esempio).

5 Un modello gerarchico bayesiano a effetti casuali

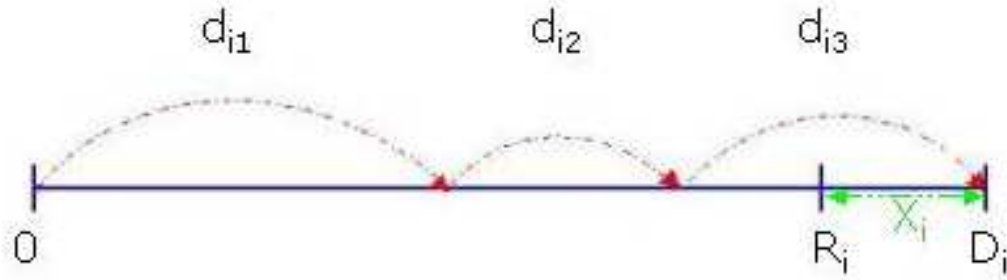
Formulare un modello permette di approssimare la realtà e di riprodurre le caratteristiche tramite delle simulazioni, se la sua struttura è sufficientemente buona e articolata da consentire un'adeguata interpretazione e previsione del fenomeno.

5.1 Modellazione dell'intensità delle frane tramite la distribuzione Gamma

È il momento di formalizzare le ipotesi fatte finora, assegnando delle distribuzioni alle quantità che si vogliono considerare. Vista la complessità delle dinamiche si è pensato ad un modello gerarchico, in un ambiente bayesiano.

In particolare, la distribuzione Gamma per l'intensità delle frane è una delle possibili distribuzioni su supporto positivo; non sembrava una scelta inopportuna osservando l'istogramma delle frequenze relative del Sito 33 (vedasi Grafico 10). Tuttavia, nella

Figura 8: Processo latente delle resistenze e dei danni cumulati. Esempio in cui la frana arriva dopo 3 anni dalla precedente: bisogna aspettare il tempo necessario perché il danno cumulato superi la resistenza.



Sezione 5.2 verrà presa in considerazione la distribuzione di Pareto Generalizzata. Siano $i = 1, \dots, l$ e $j = 1, \dots, N_i$, allora:

1. LIVELLO DEI DATI.

- $S_i \sim Ga(e^\delta, e^{-(\alpha-\delta+\beta D_i)})$: le frane sono quantità positive e dipendono da D_i . In questo modo il valor medio di S_i non dipende dal parametro di forma δ , ma esso è l'esponenziale di una combinazione lineare in D_i .
- $N_i \sim Pois(R_i)$: le dinamiche tra N_i e d_{ij} , illustrate nel Grafico 8 ricordano il processo degli arrivi di Poisson in cui N_i è il numero di arrivi e d_{ij} l'intensità degli eventi. È sensato pensare che più sia alta la resistenza da superare e più anni di attesa siano necessari per il verificarsi della frana.

2. LIVELLO DEGLI EFFETTI CASUALI.

- $R_i \sim Exp(\lambda)$: le resistenze sono quantità positive e si suppone non dipendano dal tempo.
- $d_{ij} \sim Exp(1)$: tutti gli arrivi nel processo di Poisson devono essere indipendenti ed avere l'identica distribuzione. Il parametro della distribuzione è stato fissato a 1 perché, essendo l'erosione il risultato di una competizione tra due forze (geologica e meteorologica), occorre fissare la scala di una delle due.

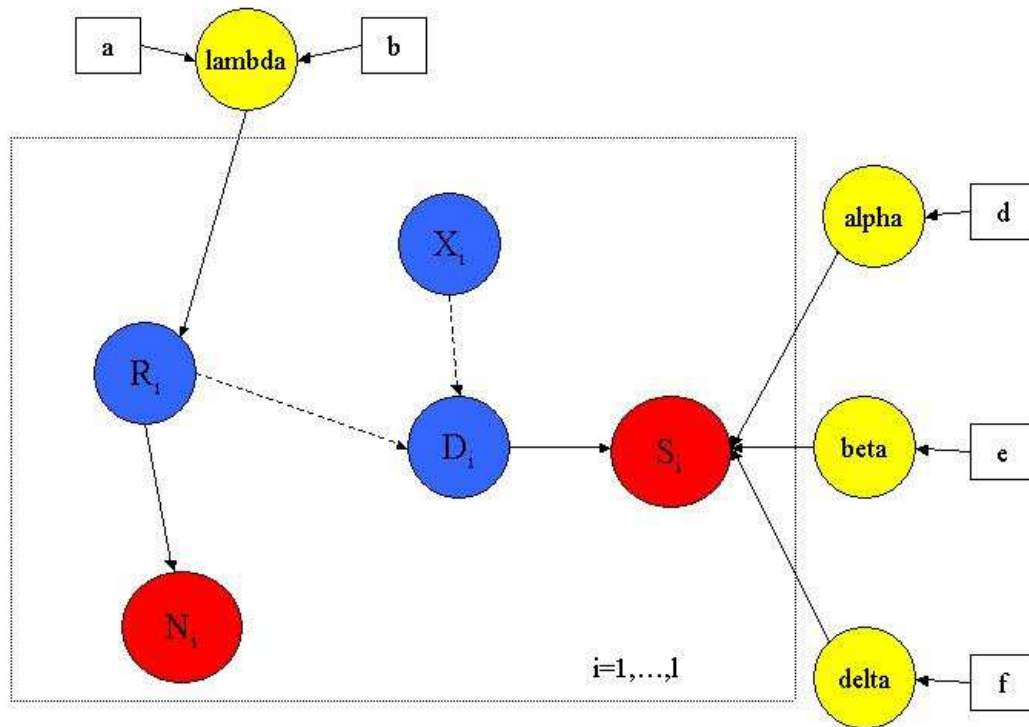
Per proprietà del processo di Poisson si ha che:

$$X_i = D_i - R_i \sim Exp(1).$$

3. LIVELLO DEGLI IPERPARAMETRI.

- $\lambda \sim Ga(a, b)$: descrive le caratteristiche geologiche del sito, in quanto condiziona la distribuzione delle resistenze e quindi la presenza dei blocchi di 0.

Figura 9: Rappresentazione grafica del modello gerarchico bayesiano con effetti casuali. Le frecce continue rappresentano una causalità, mentre quelle tratteggiate un'operazione logica (in particolare $D_i = R_i + X_i$)



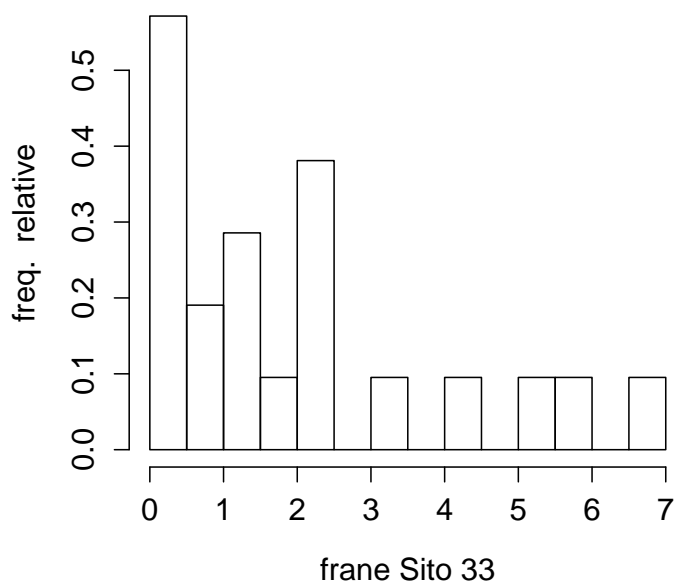
- $\alpha \sim N(0, d)$: descrive l'influenza delle caratteristiche del sito sull'intensità delle frane.
- $\beta \sim N(0, e)$: descrive l'influenza dei danni cumulati sull'intensità delle frane.
- $\delta \sim N(0, f)$: è un parametro di forma e serve per capire la forma della coda delle distribuzioni delle intensità S_i .

Le costanti a , b , d , e , f sono tali da rendere le distribuzioni vaghe. Tale modello ha una rappresentazione grafica osservabile nella Figura 9.

5.1.1 Alcune simulazioni

Nel Grafico 11 viene proposta una simulazione di 30 frane per i seguenti valori degli iperparametri: $\lambda = 1$, $\alpha = 0.6$, $\beta = 0.2$ e $\delta = 0.4$. Si notino le caratteristiche: il valore di λ è medio e dà quindi luogo a blocchi di zero di diversa lunghezza, ma non eccessiva; la tendenza ad osservare alti valori di frane in relazione a un lungo periodo di riposo non è molto evidente ed infatti il valore di β è vicino allo zero; sono presenti frane consecutive simulando un andamento crescente o decrescente tra

Figura 10: Istogramma dell'intensità delle frane del Sito 33 (non vengono considerate le frane uguali a 0).



esse, come si era visto nel Sito 33.

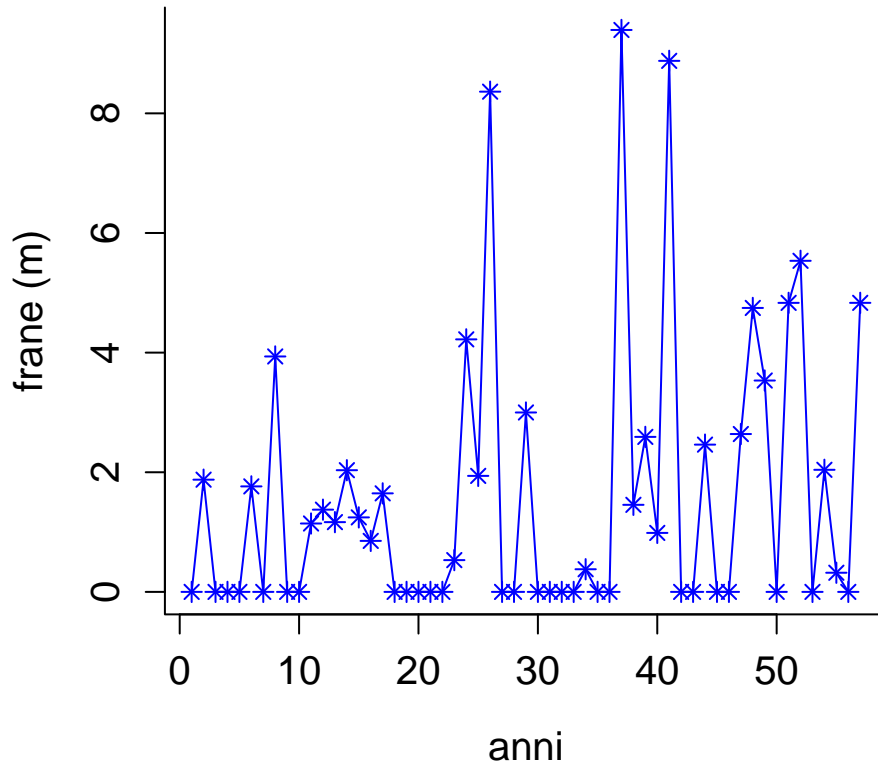
Per vedere più in dettaglio l'effetto dei parametri, verranno proposte altre simulazioni in cui verrà cambiato un solo parametro alla volta, rispetto alla configurazione di base del Grafico 11. Nel Grafico 12 si può notare come il diminuire di λ aumenti il numero di blocchi di zero e quindi la lunghezza della serie; di conseguenza, aumentando i danni cumulati aumenta anche la probabilità di valori più alti dopo lunghi blocchi di zero (si confrontino le ordinate). L'aumento di α , invece, determina un aumento costante nel valor medio delle frane.

Nel Grafico 13 vengono analizzati gli effetti di β e δ . Si noti come, l'aumentare di β aumenta l'influenza dei danni cumulati, provocando un aumento del valor medio delle frane soprattutto dopo un blocco di zeri (a volte poco credibile, essendoci una frana di 500m circa). All'aumentare di δ , invece, la coda della distribuzione cambia e le intensità delle frane sono concentrate su valori più piccoli, coerentemente con la parte costante del livello medio dato da α , limitando gli effetti di β .

5.2 Modellazione dell'intensità delle frane tramite la distribuzione di Pareto Generalizzata

La distribuzione di Pareto Generalizzata si è sviluppata nel contesto della teoria dei valori estremi per la modellizzazione dei valori eccedenti una soglia. In questo

Figura 11: Simulazione di 30 frane per i seguenti valori degli iperparametri: $\lambda = 1$, $\alpha = 0.6$, $\beta = 0.2$ e $\delta = 0.4$, utilizzando la distribuzione Gamma per le intensità.



contesto, il suo utilizzo sarebbe giustificato dal fatto che un'erosione può essere interpretata come un evento estremo sopra la soglia dello 0, cioè della stabilità della roccia. Inoltre, come si vedrà nella Sezione 8, verrà sperimentata una soglia superiore a 0. La distribuzione di Pareto Generalizzata è una distribuzione più flessibile della distribuzione Gamma, in quanto il parametro di forma permette code troncate, esponenziali e polinomiali.

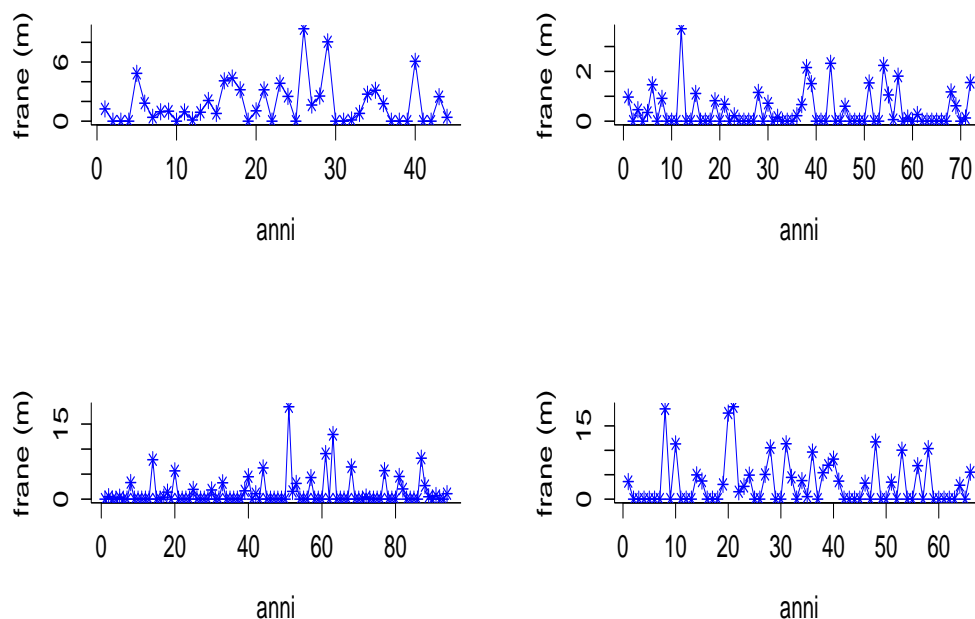
Si può, quindi, modificare la distribuzione delle intensità delle frane nel seguente modo:

$$S_i \sim Pa(\sigma_i, \xi), \quad i = 1, \dots, l$$

dove $\sigma_i = (1 - \xi)e^{-(\alpha + \beta D_i)}$. La densità che ne risulta è:

$$p(s_i) = \frac{1}{\sigma_i} \left(1 + \frac{\xi}{\sigma_i} s_i\right)^{-1-1/\xi} \quad s_i > 0, \quad \xi \neq 0, \quad \left(1 + \frac{\xi}{\sigma_i} s_i\right) > 0 \quad (1)$$

Figura 12: Simulazioni di 30 frane con l'utilizzo della Gamma per l'intensità. In senso antiorario da sinistra in alto: $\lambda = 2$, $\lambda = 0.5$, $\alpha = 1.5$, $\alpha = -0.5$.



per ogni $i = 1, \dots, l$. Si veda Coles (2001) per approfondimenti su tale distribuzione.

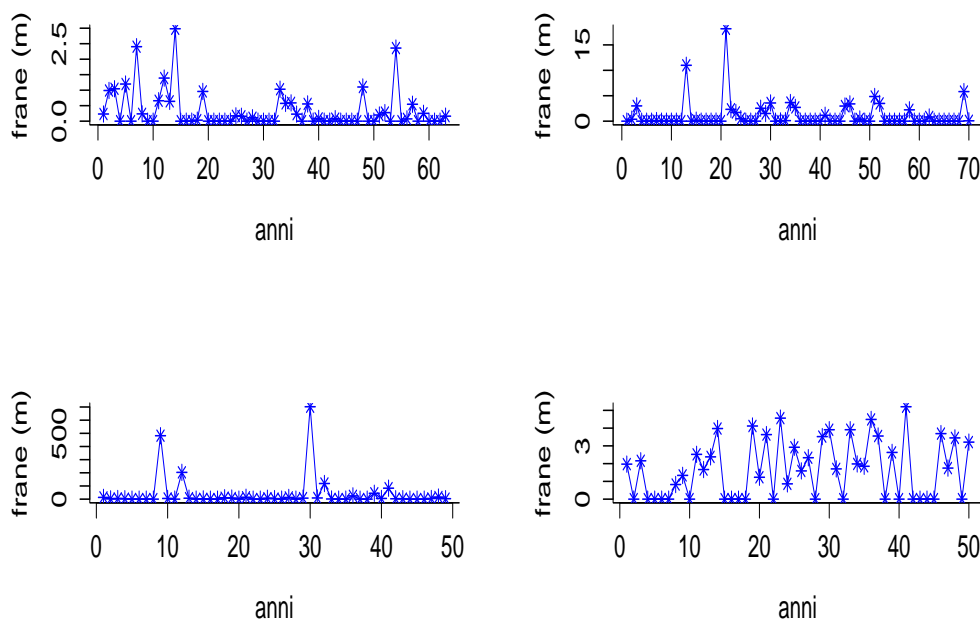
Con tale parametrizzazione il valore atteso delle intensità delle frane è lo stesso sia che si utilizzi la Pareto Generalizzata che la Gamma (quando $\xi < 0$, ma è proprio la situazione che ci si aspetta in questa applicazione). Di conseguenza, le stime dei valori attesi a posteriori di α e β dovrebbero essere molto simili tra i due modelli, in quanto hanno la stessa interpretazione.

In entrambe le specificazioni, i parametri di forma e di scala sono stati resi positivi tramite la funzione esponenziale. Tale scelta è stata fatta per comodità: l'esponenziale permetteva di non dover imporre restrizioni ulteriori sui parametri. In tal modo, però, per ogni frana, la potenza del relativo danno cumulado incide fortemente sulla variabilità della distribuzione dell'intensità, dando del credito, a volte, a valori poco realistici. Questo aspetto, verrà approfondito nella Sezione 9.

5.2.1 Alcune simulazioni

Nel Grafico 11 vengono proposte due simulazioni di 30 frane ciascuna. Entrambe hanno gli stessi valori per i primi tre iperparametri e sono uguali ai corrispondenti valori della configurazione di base usata con la Gamma (Sezione 5.1): $\lambda = 1$, $\alpha = 0.6$, $\beta = 0.2$. Per fare un confronto di come può cambiare la coda della distribuzione

Figura 13: Simulazioni di 30 frane con l'utilizzo della Gamma per l'intensità. In senso antiorario da sinistra in alto: $\beta = -0.5$, $\beta = 1$, $\delta = 1.5$, $\delta = -0.5$.



delle frane, vengono cambiati i valori di ξ . La simulazione in alto corrisponde a $\xi = -0.4$ e quella in basso a $\xi = 0.4$.

6 Algoritmo MCMC

La complessità della struttura parametrica e il contesto bayesiano indirizzano ad usare le tecniche MCMC (si veda Gilks *et al.* (1996) per i dettagli) per la stima delle densità a posteriori dei parametri.

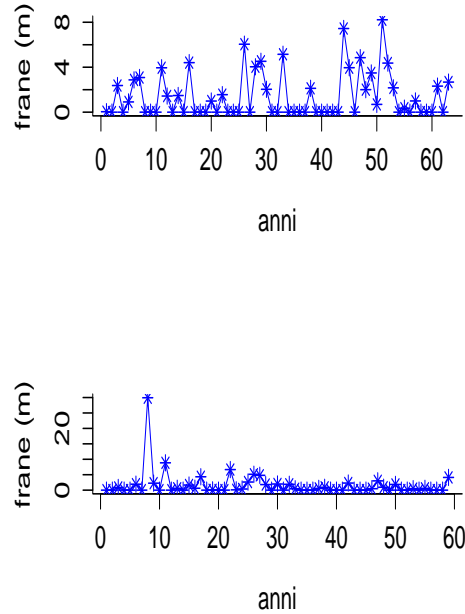
L'algoritmo MCMC utilizzato prevede un passo di Gibbs-sampling per l'aggiornamento di λ e passi di Metropolis-Hastings per l'aggiornamento dei rimanenti parametri: infatti, solo per λ ci si può avvalere delle proprietà delle famiglie coniugate.

Per ridurre la correlazione tra α e β , si è operata la centratura dei danni cumulati, che ha portato alla definizione del parametro α^* in sostituzione ad α :

$$E[S_i] = \exp(\alpha + \beta D_i) = \exp(\alpha^* + \beta(D_i - \bar{D}))$$

dove $\bar{D} = (1/l) \sum_{i=1}^l D_i$ e $\alpha^* = \alpha + \beta \bar{D}$. Poi, per ottenere la catena di α , basta fare la trasformazione inversa partendo da α^* .

Figura 14: Simulazioni di 30 frane con l'utilizzo della Pareto Generalizzata per l'intensità. In alto: $\xi = -0.4$; in basso $\xi = 0.4$.



6.1 Distribuzione Gamma per le intensità delle frane

Qui, di seguito, sono riportati i dettagli dell'algoritmo; si tenga conto che il pedice $(t-1)$ sta ad indicare il valore assunto all'iterazione $(t-1)$ -esima e p sta ad indicare la proposta del parametro all'iterazione t -esima:

1. $\lambda_t \sim Ga(a + l, b + \sum_{i=1}^l R_i)$ $i = 1, \dots, l$, dove $a = b = 1/100$ (nel Gibbs-sampling la proposta viene accettata con probabilità pari a 1).
2. $R_{ip} \sim Exp(\lambda + 1)$ $i = 1, \dots, l$; si tratta di un Independent Metropolis-Hastings, in quanto il valore proposto non dipende dal valore precedente della catena. Tale valore viene accettato con probabilità A_{R_i} .
3. $X_{ip} \sim Exp(1)$; anche qui si tratta di un Independent Metropolis-Hastings. Tale valore viene accettato con probabilità A_{X_i} .
4. $\alpha_p^* \sim N(\alpha_{(t-1)}^*, ha)$, dove $ha = 0.05$ e regola il passo nel processo di aggiornamento del parametro. Tale valore viene accettato con probabilità A_A .
5. $\beta_p \sim N(\beta_{(t-1)}, hb)$, dove $hb = 0.05$. Tale valore viene accettato con probabilità A_B .

6. $\delta_p \sim N(\delta_{(t-1)}, hd)$, dove $hd = 0.05$. Tale valore viene accettato con probabilità A_D .

Tenendo conto che i parametri sono stati aggiornati secondo l'ordine appena illustrato, le probabilità di accettazione utilizzate nei passi di Metropolis-Hastings sono:

$$\begin{aligned}
 A_{R_i} &= \min \left\{ 1, \frac{R_{ip}^{N_i} f_S(s_i | R_{ip}, X_{i(t-1)}, \alpha_{(t-1)}, \beta_{(t-1)}, \delta_{(t-1)})}{R_{i(t-1)}^{N_i} f_S(s_i | R_{i(t-1)}, X_{i(t-1)}, \alpha_{(t-1)}, \beta_{(t-1)}, \delta_{(t-1)})} \right\} \\
 A_{X_i} &= \min \left\{ 1, \frac{f_S(s_i | R_{it}, X_{ip}, \alpha_{(t-1)}, \beta_{(t-1)}, \delta_{(t-1)})}{f_S(s_i | R_{it}, X_{i(t-1)}, \alpha_{(t-1)}, \beta_{(t-1)}, \delta_{(t-1)})} \right\} \\
 A_A &= \min \left\{ 1, \frac{\sum_{i=1}^l f_S(s_i | R_{it}, X_{it}, \alpha_p, \beta_{(t-1)}, \delta_{(t-1)})}{\sum_{i=1}^l f_S(s_i | R_{it}, X_{it}, \alpha_{(t-1)}, \beta_{(t-1)}, \delta_{(t-1)})} \right\} \\
 A_B &= \min \left\{ 1, \frac{\sum_{i=1}^l f_S(s_i | R_{it}, X_{it}, \alpha_t, \beta_p, \delta_{(t-1)})}{\sum_{i=1}^l f_S(s_i | R_{it}, X_{it}, \alpha_t, \beta_{(t-1)}, \delta_{(t-1)})} \right\} \\
 A_D &= \min \left\{ 1, \frac{\sum_{i=1}^l f_S(s_i | R_{it}, X_{it}, \alpha_t, \beta_t, \delta_p)}{\sum_{i=1}^l f_S(s_i | R_{it}, X_{it}, \alpha_t, \beta_t, \delta_{(t-1)})} \right\}
 \end{aligned}$$

dove f_S è la densità della distribuzione Gamma che regola l'intensità delle frane.

6.2 Distribuzione di Pareto Generalizzata per le intensità delle frane

In caso di utilizzo della Pareto Generalizzata, il procedimento dell'algoritmo è uguale a quello descritto nella Sezione 6.1; l'unica differenza è che il parametro di forma si indica con ξ al posto di δ , e che si deve utilizzare la densità della Pareto Generalizzata al posto della Gamma.

7 Studi di simulazione

Per utilizzare questa procedura, in entrambe le specificazioni del modello, occorre prima svolgere degli studi con dei dati simulati per vedere se si ottengono catene di Markov stabili attorno ai veri valori. Dei numerosi studi di simulazione condotti, se ne riporta uno solo (per ogni modello) per non appesantire eccessivamente la presentazione.

7.1 Tramite la distribuzione Gamma per l'intensità delle frane

Sono state generate 500 frane con i valori degli iperparametri utilizzati nella Sezione 5.1.1, nella configurazione di base, seguendo lo schema del modello considerato nella Sezione 5.1:

- $\lambda = 1$
- $\alpha = 0.6$
- $\beta = 0.2$
- $\delta = 0.4$.

Quindi è stata utilizzata la procedura illustrata nella Sezione 6.1 con un numero di iterazioni pari a 10000. Per la stima dei valori attesi a posteriori, il periodo di burn-in è stato definito pari alle prime 5000 iterazioni.

Nella Figura 15, vengono presentate le catene di Markov per gli iperparametri, il confronto tra le resistenze simulate e quelle stimate e così pure per i danni cumulati. Le stime dei valori attesi a posteriori e gli intervalli di credibilità al 95% sono riportati nella Tabella 1. I risultati si possono considerare soddisfacenti: le catene hanno un buon mixing e si comportano piuttosto stabilmente attorno ai veri valori degli iperparametri. L'aderenza delle stime delle resistenze e dei danni cumulati ai relativi veri valori, dipende in parte dalla forza della relazione tra l'intensità delle frane e i valori dei danni cumulati: cioè da β . In questo caso è stato adoperato un valore piuttosto basso di β , in modo da non fare uno studio di simulazione nel caso più favorevole. L'aderenza ai veri valori è buona, tenendo conto che non sono stati osservati.

7.2 Tramite la distribuzione di Pareto Generalizzata per l'intensità delle frane

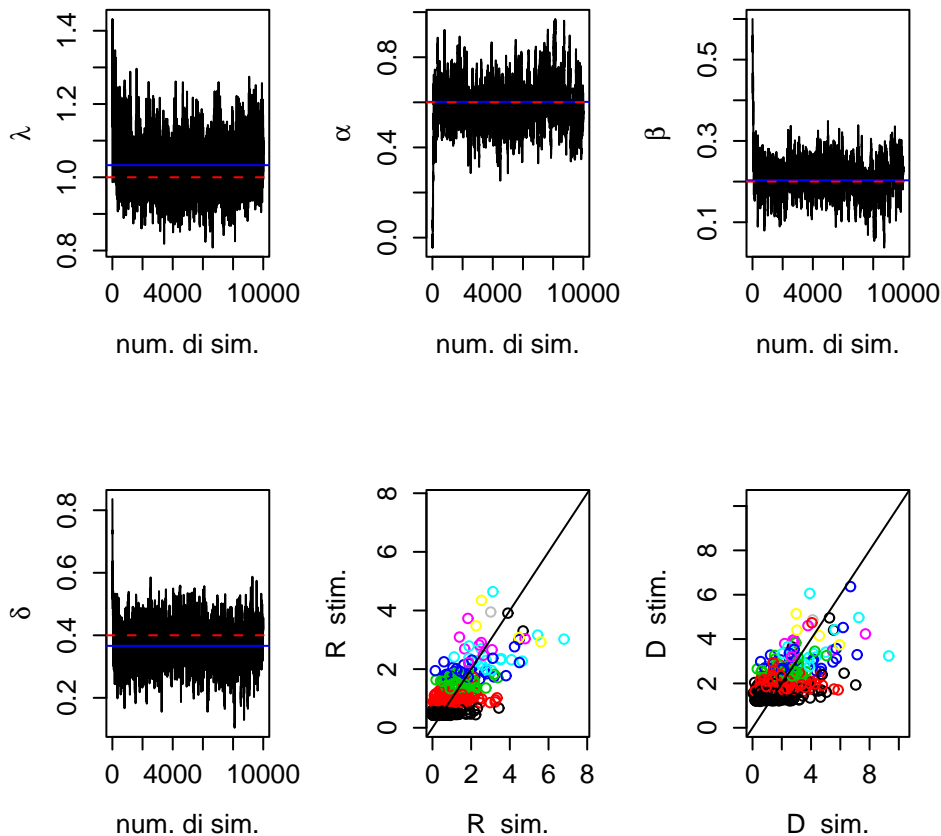
Si riporta, come per il caso della Gamma, uno studio di simulazione tra quelli eseguiti. Seguendo lo schema del modello considerato nella Sezione 5.2, sono state simulate 100 frane con i valori degli iperparametri della prima configurazione della Sezione 5.2.1:

- $\lambda = 1$
- $\alpha = 0.6$
- $\beta = 0.2$
- $\xi = -0.4$.

Successivamente, è stata utilizzata la procedura illustrata nella Sezione 6.2 con un numero di iterazioni pari a 10000. Per la stima dei valori attesi a posteriori, il periodo di burn-in è stato fissato pari alle prime 5000 iterazioni.

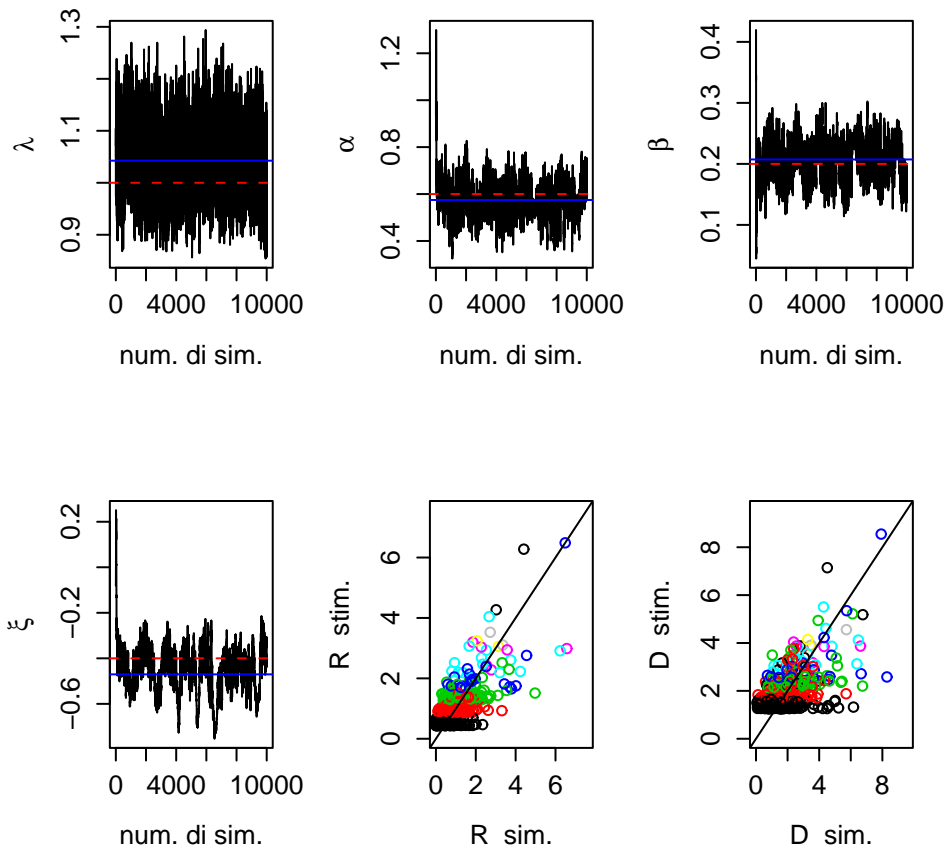
Nella Figura 16, vengono presentate le catene di Markov per gli iperparametri, il confronto tra le resistenze simulate e quelle stimate e così pure per i danni cumulati. Le stime dei valori attesi a posteriori e gli intervalli di credibilità al 95% sono

Figura 15: Studio di simulazione utilizzando la distribuzione Gamma per il controllo della convergenza delle catene per gli iperparametri del modello. I veri valori sono rappresentati dalle linee grigie tratteggiate, mentre le stime dei valori attesi a posteriori sono rappresentate dalle linee grigie continue. Gli ultimi due grafici rappresentano i confronti tra le stime e i veri valori delle resistenze e dei danni cumulati.



riportati nella Tabella 1. I risultati si possono considerare abbastanza soddisfacenti: le catene hanno un discreto mixing e si comportano piuttosto stabilmente attorno ai veri valori dei parametri. Probabilmente occorre trovare una riparametrizzazione che renda i parametri meno correlati tra loro, ma in questo caso verrebbe meno l'uguaglianza dei valori attesi tra le due specificazioni (per $\xi < 0$), che sembra essere piuttosto importante in questa fase di confronti. Inoltre, la precisione nella stima dei valori attesi a posteriori degli iperparametri, comuni ai due modelli, non viene alterata cambiando la distribuzione delle intensità delle frane da Gamma a Pareto Generalizzata. Si confrontino, a tal proposito, i relativi intervalli di credibilità nella Tabella 1.

Figura 16: Studio di simulazione utilizzando la distribuzione di Pareto Generalizzata per il controllo della convergenza delle catene per gli iperparametri del modello utilizzato. I veri valori sono rappresentati dalle linee tratteggiate grigie, mentre le stime dei valori attesi a posteriori sono rappresentati dalle linee grigie continue. Gli ultimi due grafici rappresentano i confronti tra le stime e i veri valori delle resistenze e dei danni cumulati.



8 Applicazione al Sito 33

Dopo l'esito positivo degli studi di simulazione, si può passare ad applicare l'algoritmo MCMC al Sito 33, di cui si è parlato nella Sezione 4. Si distingueranno sempre i due casi: l'uso della distribuzione Gamma e quella di Pareto Generalizzata nella modellazione dell'intensità delle frane. In tal modo, si potranno anche confrontare le modalità di convergenza delle catene markoviane passando dai dati simulati ai dati reali.

Per la presenza dei valori mancanti all'inizio della serie storica (vedasi Figura 7), sono stati considerati gli eventi solo dal 1961; l'ultima frana considerata, invece, è

del 2000. All'interno della serie storica, poi, c'è la presenza di un valore mancante che viene preceduto da una frana di 3 cm e viene seguito da una di 10 cm. Questo significa che la frana di 10 cm rappresenta la recessione costiera di due anni.

Gli eventi inferiori a 10 centimetri sono poco presenti nel dataset: essi rappresentano, infatti, il 5% di tutte le frane registrate nell'intero studio. Il loro valore potrebbe essere paragonato a quello di uno 0, per un paio di motivi:

- è impossibile sapere se tutti gli operatori coinvolti nel progetto arrivassero a segnalare gli eventi con tale precisione;
- le misurazioni vengono fatte nella parte alta della scogliera, mentre il fenomeno dell'erosione si verifica ai piedi delle stesse, generando una distorsione tra il fenomeno e la sua misurazione. Bisognerebbe riuscire a distinguere gli eventi delle frane proprie dalle piccole cadute di materiale, dovute ad altri motivi.

Nel caso del Sito 33, tali eventi di piccolissima entità si trovano dopo una lunga serie di 0. La loro presenza influenzerebbe, soprattutto, la stima degli iperparametri α e β : questi eventi porterebbero l'algoritmo ad "imparare" che in presenza di un alto danno cumulato si verificano delle frane di intensità quasi nulla. Tuttavia, questo sarebbe in contraddizione con le caratteristiche geologiche e dinamiche della parte della costa nord a cui appartiene il sito. Considerato che 10 cm oltre ad essere il quantile al 5% delle frane dell'intero dataset è anche quello del Sito 33, si potrebbero uguagliare a zero le frane di intensità inferiore o uguale. In poche parole, si introdurrebbe una soglia pari a 10 cm e si considererebbe l'intensità delle frane come la differenza dell'intensità originaria dalla soglia. Questo modo di procedere verrà illustrato nella Sezione 8.1, mentre nella Sezione 8.2 si procederà ad utilizzare l'algoritmo di stima ai valori originari.

8.1 Utilizzo di una soglia

Nella Figura 17 vengono rappresentate le catene degli iperparametri del modello quando si usa la distribuzione Gamma per modellare l'intensità delle frane, mentre nella Figura 18 quando si usa la distribuzione di Pareto Generalizzata. Nella Tabella 2, invece, vengono riportati le stime dei valori attesi a posteriori e gli intervalli

Iperpar.	Vero valore	Gamma		Pareto	
		val. att.	int. cred.	val. att.	int. cred.
λ	1	1.033	(0.909, 1.165)	1.042	(0.927, 1.167)
α	0.6	0.602	(0.407, 0.822)	0.575	(0.444, 0.715)
β	0.2	0.203	(0.115, 0.289)	0.207	(0.142, 0.263)
δ	0.4	0.366	(0.230, 0.496)	–	–
ξ	–0.4	–	–	–0.471	(–0.703, –0.294)

Tabella 1: Stime dei valori attesi a posteriori e intervalli di credibilità al 95% per gli iperparametri in entrambe le specificazione del modello tramite la distribuzione Gamma e la distribuzione di Pareto Generalizzata, nello studio di simulazione.

di credibilità al 95%, per entrambi i casi. La stabilità delle catene sembra essere raggiunta in entrambi i casi, infatti le stime dei valori attesi a posteriori e gli intervalli di credibilità per gli iperparametri in comune ai due modelli λ, α, β sono molto simili. L'utilizzo della distribuzione di Pareto Generalizzata porta, tuttavia, a un mixing meno buono. In particolare, il parametro di forma ξ non viene identificato con precisione: sembrerebbe che qualsiasi forma della coda andasse bene, come si può notare dalla catena e dall'intervallo di credibilità. La spiegazione potrebbe stare nel fatto che i dati a disposizione sono troppo pochi per poter inferire sulla coda della loro distribuzione: la Pareto Generalizzata a differenza della Gamma offre molte possibilità, ma la penuria di dati fa sì che non si possa avere la certezza di una scelta sicura. La stima del valore atteso a posteriori di ξ non è nient'altro che una mediazione tra le varie possibilità.

Per problemi di spazio, non vengono riportate le stime delle resistenze e dei danni cumulati: esse sono simili in entrambi i modelli.

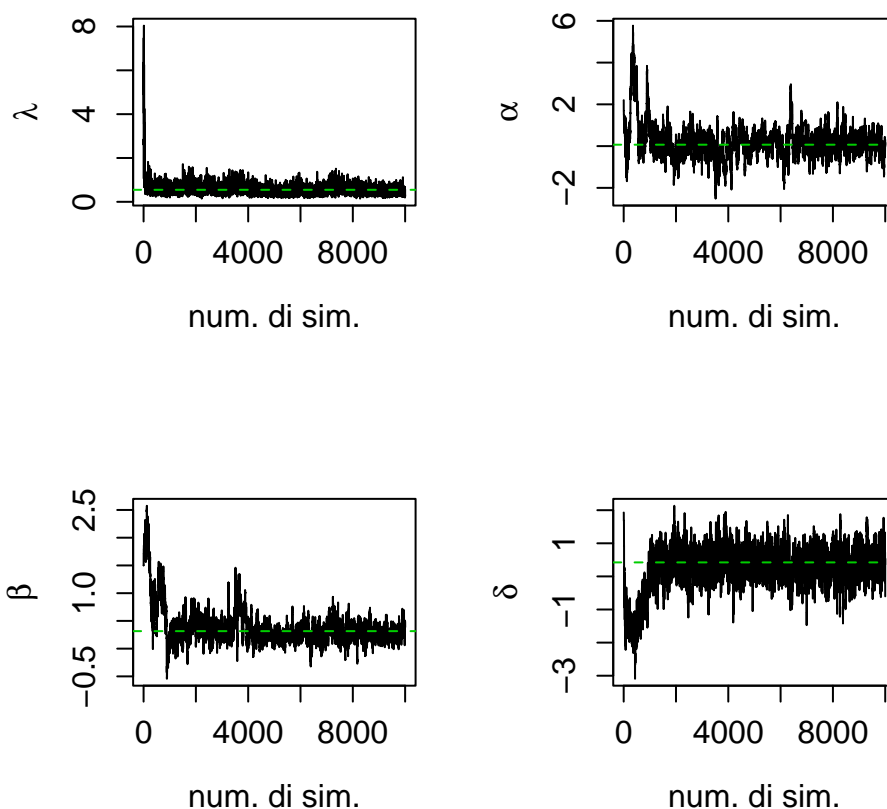
Ora, per valutare l'adattamento del modello ai dati vengono simulate, per ognuno dei due modelli, 100 frane utilizzando le stime dei valori attesi a posteriori degli iperparametri. Nella Figura 19, per le frane osservate e quelle simulate tramite i due metodi, si trovano la stima delle densità, il Probability plot e il Quantile plot; inoltre, vi si trova il Quantile plot per gli anni di pausa osservati e simulati. I dati sono pochi e ovviamente non permettono di capire con certezza se i modelli usati consentono un buon adattamento; tuttavia, l'adattamento appare abbastanza buono e i due modelli sembrano comportarsi equivalentemente nel dominio dei valori osservati. L'unica limitazione è che le code stimate delle distribuzioni dell'intensità delle frane nei due modelli, danno una certa credibilità a dei valori poco plausibili in natura.

In definitiva, tra la distribuzione Gamma e quella di Pareto Generalizzata, è preferibile la Gamma, in quanto il parametro di forma ξ della Pareto Generalizzata non riesce a catturare le informazioni sulla forma della coda e la stima del suo valore atteso a posteriori non è nient'altro che una compensazione di tutte le possibilità. Inoltre, tale stima essendo negativa implica una forma troncata della distribuzione delle frane; tuttavia, questo non è risultato sufficiente per non avere dei valori

Iperpar.	Gamma		Pareto	
	val. att.	int. cred.	val. att.	int. cred.
λ	0.547	(0.255, 0.992)	0.536	(0.262, 0.915)
α	0.068	(-1.248, 1.207)	0.070	(-0.876, 1.033)
β	0.315	(0.009, 0.535)	0.293	(0.044, 0.621)
δ	0.423	(-0.004, 0.811)	—	—
ξ	—	—	-1.307	(-3.152, 0.267)

Tabella 2: Stime dei valori attesi a posteriori e intervalli di credibilità al 95% per gli iperparametri in entrambe le specificazione del modello tramite la distribuzione Gamma e la distribuzione di Pareto Generalizzata, applicati al Sito33, con l'uso di una soglia di 10cm.

Figura 17: Catene Markoviane degli iperparametri del modello in cui si specifica la distribuzione dell'intensità delle frane tramite la distribuzione Gamma, nell'applicazione al Sito 33 con l'uso di una soglia di 10cm. Le linee tratteggiate orizzontali rappresentano la stima dei valori attesi a posteriori.



simulati eccessivamente estremi.

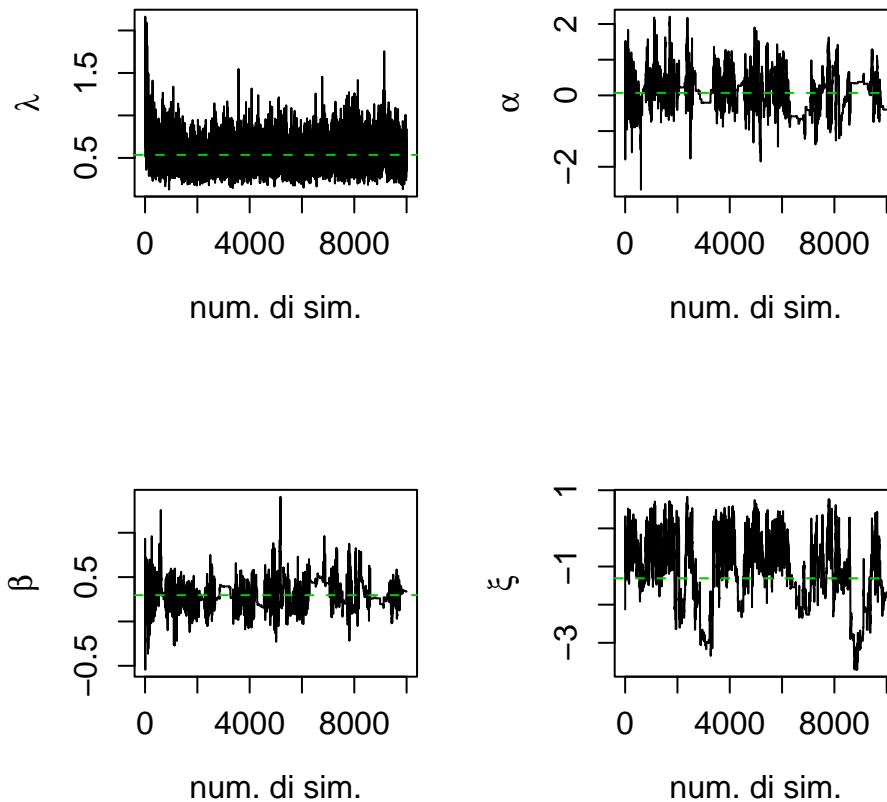
8.2 Utilizzo dei dati originari

Si passi, ora, a vedere cosa succede alle stime delle densità a posteriori degli iperparametri, quando si considerano i valori originari delle frane, senza, quindi, utilizzare alcuna soglia.

Si ricordi che si è deciso di utilizzare la distribuzione Gamma per modellare l'intensità delle frane.

Le stime dei valori attesi a posteriori e gli intervalli di credibilità sono riportati nella Tabella 3. Si confrontino tali risultati con quelli ottenuti con l'introduzione della soglia (Tabella 2): le stime dei valori attesi a posteriori di α e β cambiano notevol-

Figura 18: Catene Markoviane degli iperparametri del modello in cui si specifica la distribuzione dell'intensità delle frane tramite la distribuzione di Pareto Generalizzata, nell'applicazione al Sito 33 con l'uso di una soglia di 10cm. Le linee tratteggiate orizzontali rappresentano la stima dei valori attesi a posteriori.

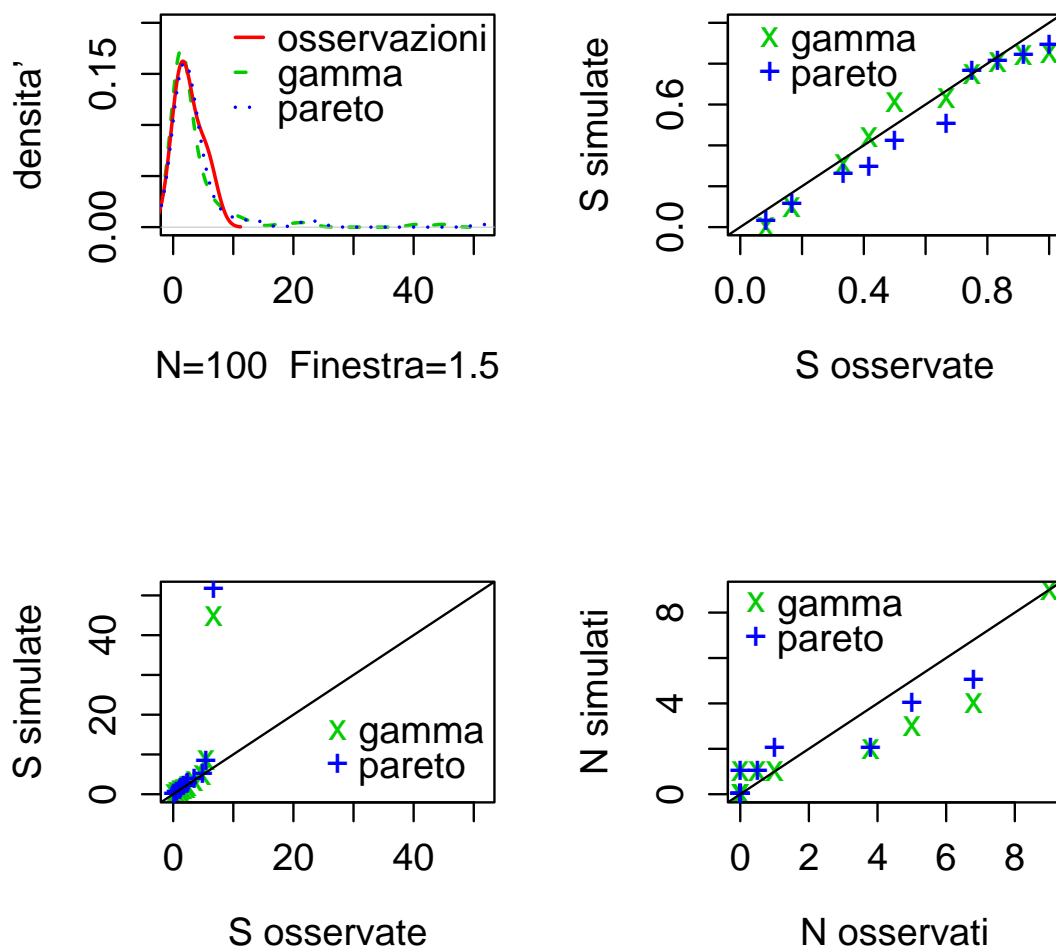


mente tra i due casi.

Nel caso con la soglia, α assume un valore vicino allo 0, mentre β un valore positivo: l'intensità delle frane viene, quindi, imputata quasi esclusivamente alla grandezza del relativo danno cumulato e di conseguenza all'azione delle tempeste.

Nel caso senza soglia, è α che assume un valore positivo e β vicino allo zero: sembra che il sito tenda a franare senza tener conto dell'effetto della meteorologia. Tuttavia, questa ultima interpretazione è in contrasto con quella che è la struttura geologica della parte nord della costa, dove la roccia è formata da materiale più resistente (rispetto alla parte sud), con minor tendenza a franare; sono, nella zona nord, le tempeste marine a giocare un ruolo fondamentale nel verificarsi delle frane, in quanto deteriorano le rocce fino a farle cadere.

Figura 19: Studio dell'adattamento dei due modelli (Gamma e Pareto Generalizzata) tramite simulazioni, ai dati del Sito 33. In senso orario da in alto a sn: confronto delle densità stimate tra frane simulate ed osservate, Probability plot e Quantile plot; infine, Quantile plot per gli anni di pausa.



Nel Grafico 20 viene analizzata una simulazione di 100 frane, ottenute con le stime dei valori attesi a posteriori degli iperparametri. Dall'alto verso il basso e da sinistra verso destra sono rappresentati, per le frane osservate e simulate: le stime delle densità, il Probability plot e il Quantile plot; inoltre, il Quantile plot per gli anni di pausa osservati e simulati.

L'aderenza sembra essere soddisfacente. Si osservi, in particolare, il grafico dei

quantili per le frane: il valore più grande non è così estremo come nel modello con la soglia³. Sembrerebbe, quindi, che il modello senza soglia funzioni meglio. Tuttavia, si sono riscontrate delle incoerenze logiche nella stima dei valori attesi a posteriori di alcuni iperparametri. Il modello deve fallire in qualche altra parte. Potrebbe essere che la funzione esponenziale, che lega il valore atteso delle frane al danno cumulato, non sia coerente con quello che succede in natura. Infatti, nel caso non si usi la soglia, l'influenza dei danni cumulati viene drasticamente ridotta in quanto la stima del valore atteso a posteriori di β è praticamente nullo e non si sono verificati eventi fuori da un dominio di valori plausibili. Invece, nel caso con la soglia, è β il parametro di forza nella distribuzione dell'intensità delle frane; i danni cumulati, cioè, assumono la maggiore influenza nella determinazione del valore medio delle frane. Di conseguenza, per la presenza della funzione esponenziale, è sufficiente che sia simulato un danno cumulato leggermente più grande degli altri che la distribuzione da cui si simulerà la relativa frana avrà un valore atteso di gran lunga maggiore.

9 Cambio della relazione tra danni cumulati e frane: dall'esponenziale all'identità

La comodità della funzione esponenziale, usata nel valore atteso delle distribuzioni delle intensità, deve essere abbandonata per la poca credibilità dei valori estremi che induce. Una proposta alternativa potrebbe essere quella di sostituirla con la funzione identità in modo da ottenere il seguente valore atteso:

$$E(S_i) = \alpha + \beta D_i. \quad (2)$$

In tal modo, un incremento unitario nel danno cumulato genera un incremento lineare nel valore atteso dell'intensità della frana. Per rispettare la precedente scrittura del valore atteso, si è deciso di scrivere la distribuzione delle frane così come segue:

$$S_i \sim Ga(\delta, \delta/(\alpha + \beta D_i)). \quad (3)$$

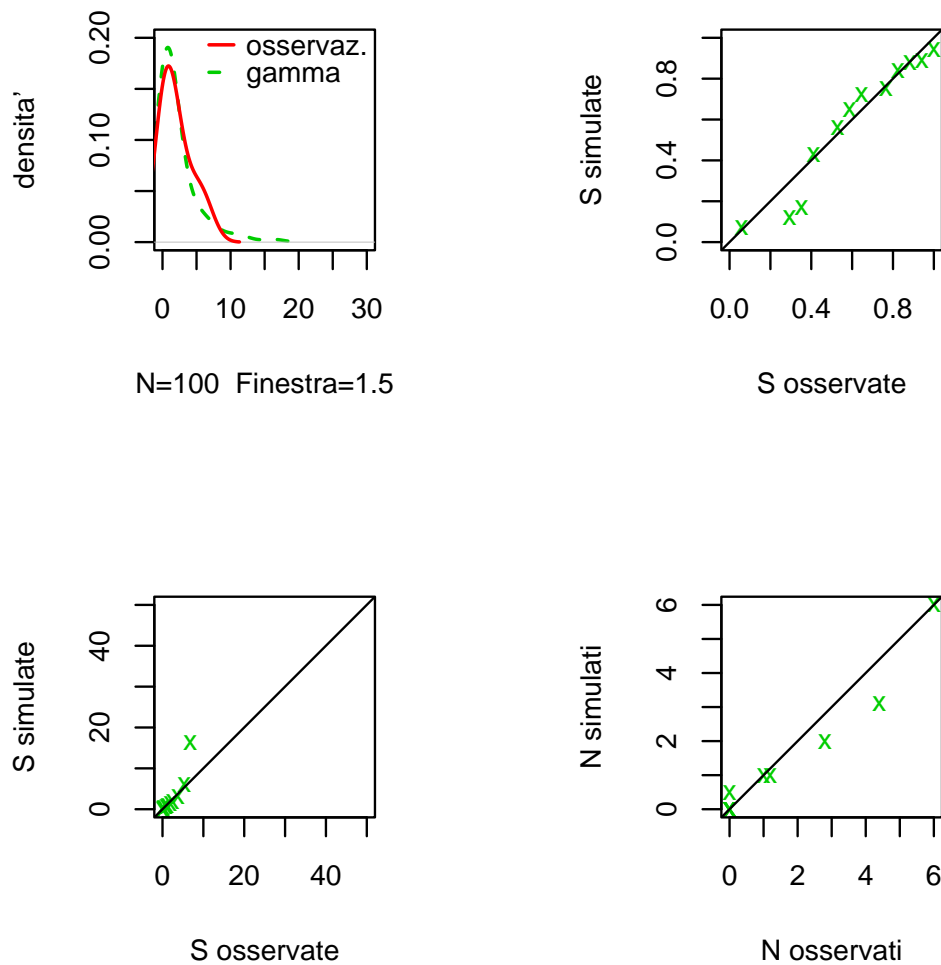
Si studi l'effetto dei parametri sui valori delle frane, come si è fatto per la funzione esponenziale. Nel Grafico 21 viene rappresentata una simulazione di 30 frane con

³Per facilitare il confronto, è stato lasciato lo stesso dominio nelle ordinate.

Iperpar.	val. att.	int. cred.
λ	0.829	(0.434, 1.399)
α	0.851	(-0.801, 2.659)
β	-0.053	(-0.924, 0.640)
δ	-0.507	(-1.159, 0.194)

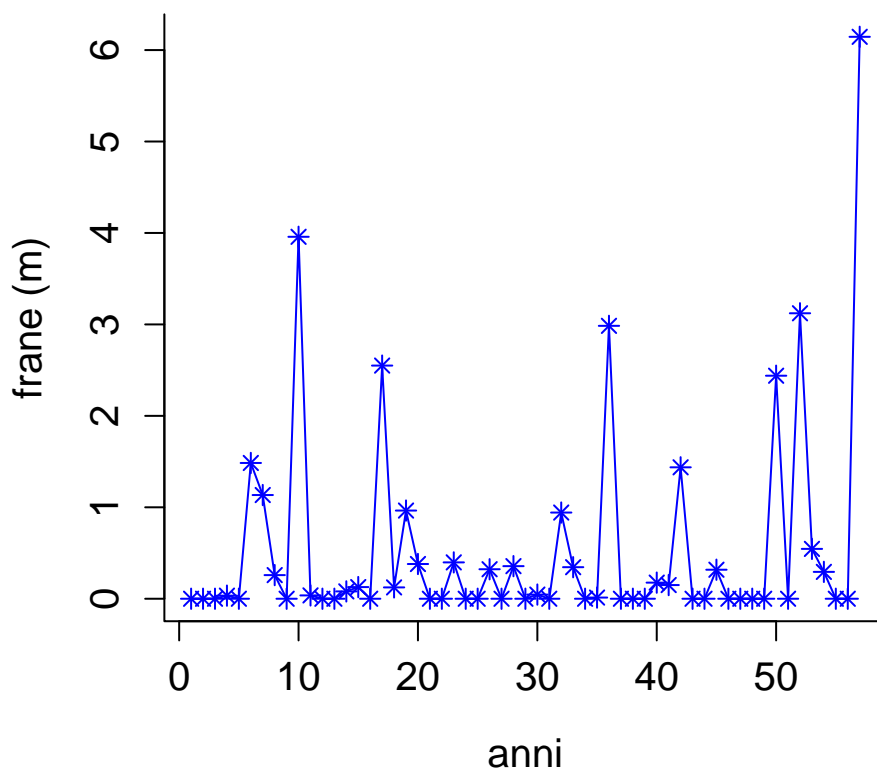
Tabella 3: Stime dei valori attesi a posteriori e intervalli di credibilità al 95% per gli iperparametri, con la distribuzione Gamma per le intensità delle frane, applicati al Sito33.

Figura 20: Studio dell'adattamento del modello (distribuzione Gamma per le frane) tramite simulazioni, ai dati del Sito 33. In senso orario da in alto a sn: confronto delle densità stimate tra frane simulate ed osservate, Probability plot e Quantile plot; infine, Quantile plot per gli anni di pausa.



l'utilizzo dei seguenti valori: $\lambda = 1$, $\alpha = 0.6$, $\beta = 0.5$ e $\delta = 0.4$. Il valore di β è maggiore di quello utilizzato nel relativo studio per la funzione esponenziale e non genera valori poco credibili come succedeva prima. Nel Grafico 22 sono rappresentati gli effetti dei cambiamenti di λ e α e nel Grafico 23 quelli per β e δ . L'unico cambiamento che si è verificato con la funzione identità è l'uso più flessibile di β e la maggiore credibilità dei valori che vengono generati. È proprio il risultato che si

Figura 21: Simulazione di 30 frane per i seguenti valori dei parametri: $\lambda = 1$, $\alpha = 0.6$, $\beta = 0.5$ e $\delta = 0.4$, utilizzando la distribuzione Gamma per le intensità e la funzione identità per il valore atteso.



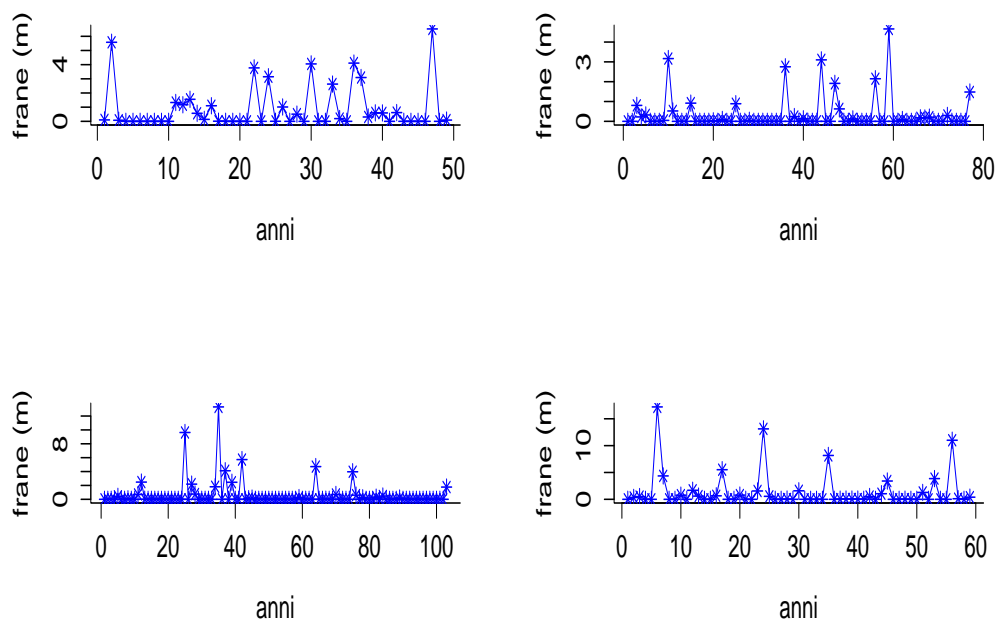
voleva ottenere, perchè la rimanente struttura del modello sembrava essere piuttosto soddisfacente.

9.1 Studio di simulazione

A questo punto è sufficiente rifarsi, sia per le simulazioni che per l'algoritmo di stima, alla struttura usata nel caso della funzione esponenziale, con l'accortezza che i parametri di forma e di scala di S_i devono essere positivi. In definitiva, occorre sempre porre il vincolo di positività su δ e $(\alpha + \beta D_i)$, per ogni $i = 1, \dots, l$.

Prima di tutto è stato svolto uno studio di simulazione per vedere se l'algoritmo riesce a identificare con sicurezza tutti i parametri. Sono stati utilizzati gli stessi valori proposti nello studio di simulazione con la densità Gamma e la funzione esponenziale della Sezione 7.1. Si veda la Figura 24 per un riscontro grafico dell'a-

Figura 22: Simulazioni di 30 frane con l'utilizzo della Gamma per l'intensità e la funzione identità per il valore atteso. In senso antiorario da sinistra in alto: $\lambda = 2$, $\lambda = 0.5$, $\alpha = 1.5$, $\alpha = -0.5$.

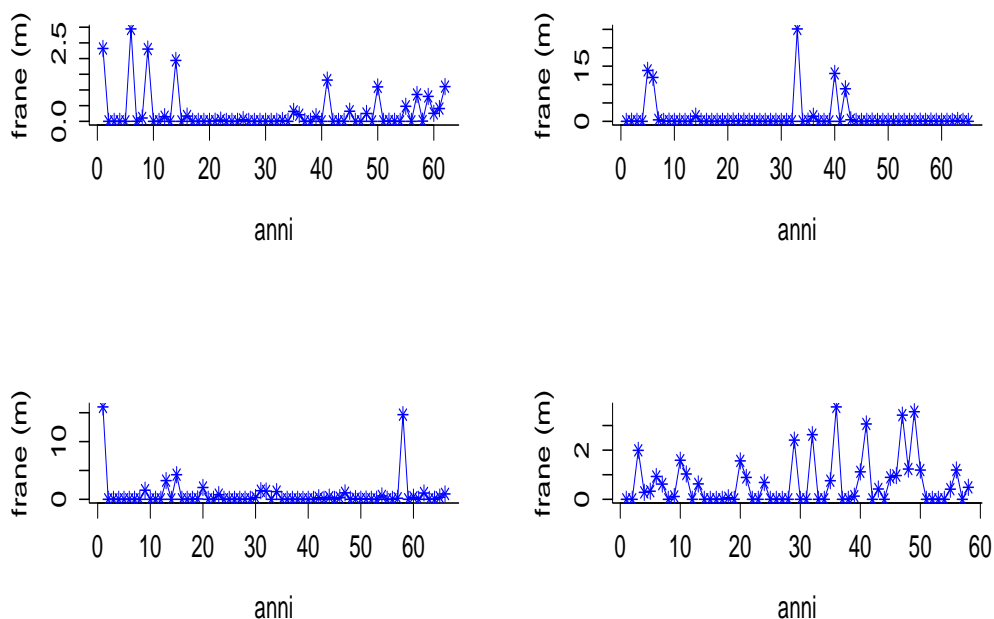


derenza delle stime dei parametri ai veri valori e la Tabella 4 per le stime dei valori attesi a posteriori degli iperparametri e i relativi intervalli di credibilità. I risultati si possono definire piuttosto soddisfacenti.

Iperpar.	Vero valore	val. att.	int. cred.
λ	1	0.971	(0.822, 1.148)
α	0.6	0.571	(0.057, 1.239)
β	0.2	0.287	(-0.031, 0.612)
δ	0.4	0.385	(0.331, 0.441)

Tabella 4: Studio di simulazione: stime dei valori attesi a posteriori e intervalli di credibilità al 95% per gli iperparametri, con l'uso della distribuzione Gamma e funzione identità nel valore atteso.

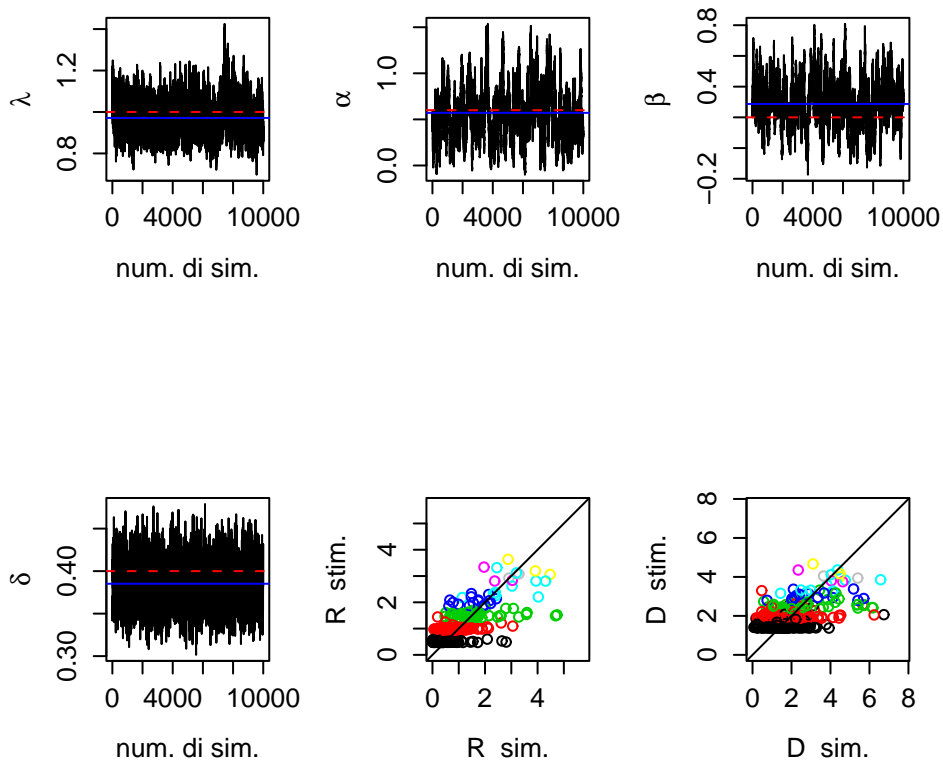
Figura 23: Simulazioni di 30 frane con l'utilizzo della Gamma per l'intensità e la funzione identità per il valore atteso. In senso antiorario da sinistra in alto: $\beta = 0$, $\beta = 1$, $\delta = 1.5$, $\delta = 0.1$.



9.2 Applicazione al Sito 33

Dopo aver osservato la funzionalità dello strumento di stima, si può passare ad applicare l'algoritmo MCMC ai dati del Sito 33 con l'utilizzo della soglia di 10cm, come spiegato nella Sezione 8.1. Nel Grafico 25 sono riportate le catene markoviane per gli iperparametri, mentre le stime dei valori attesi a posteriori e gli intervalli di credibilità al 95% sono riportati nella Tabella 5. Come si può vedere dal Grafico 25, non c'è una buona convergenza per δ . Questo è successo sebbene, per cercare di ridurre l'effetto della correlazione tra i parametri, sia stata utilizzata 1 iterazione ogni 10 delle 100000 totali. Ora, avendo visto che la convergenza è raggiunta con lo studio di simulazione, si desume che la quantità esigua di frane non permetta di identificare con ragionevole certezza la forma della coda della distribuzione. Evidentemente, la funzione esponenziale tutela questo aspetto in quanto il suo uso è il più naturale in matematica, per i problemi di positività dei parametri che si dovevano affrontare. In ogni caso, il messaggio che deriva dalle stime ottenute con la funzione identità è simile a quello delle stime della funzione esponenziale: si tratta di un sito in cui è probabile si osservano anni di pausa tra una frana e l'altra e l'entità delle frane sono spiegate per buona parte dai danni causati dall'effetto meteorologico: l'intervallo di

Figura 24: Studio di simulazione utilizzando la distribuzione Gamma e la funzione identità nel valore atteso di S_i , per il controllo della convergenza delle catene per gli iperparametri del modello. I veri valori sono rappresentati dalle linee grigie tratteggiate, mentre le stime dei valori attesi a posteriori sono rappresentate dalle linee grigie continue. Gli ultimi due grafici rappresentano i confronti tra le stime e i veri valori delle resistenze e dei danni cumulati.



credibilità al 95% per β non contiene lo zero, mentre quello per α sì. Come si può vedere dal Grafico 26, il vantaggio che ha portato l'adozione della funzione identità consiste nella plausibilità dei valori generati utilizzando le stime dei valori attesi a posteriori degli iperparametri. Quindi, anche quando non si riesce ad identificare con precisione tutti gli elementi del modello, si può dire che il modello sia valido nel suo complesso, perché globalmente riesce a riprodurre le caratteristiche del fenomeno oggetto di studio. Si può, quindi, ritenere che la scelta della funzione identità sia preferibile alla funzione esponenziale.

Nella Figura 27, viene riportata la parte della serie storica analizzata con le stime dei valori attesi a posteriori delle resistenze e dei danni cumulati, per ogni frana.

Si può notare come le frane di bassa entità siano accompagnate da valori bassi di resistenze e di danni cumulati, mentre il contrario accade per le frane di intensità più elevata.

Si osservino le quattro frane avvenute dopo l'anno 1990: esse sono consecutive e di intensità piuttosto elevata per il Sito 33. Il modello stima le relative resistenze con valori bassi, perché se le frane sono consecutive, vuol dire che la coesione della roccia, in ognuno di quegli anni, era bassa. Il comportamento del Sito è particolare perché, dopo le prime due frane, anziché osservare uno zero o una frana di più piccola intensità –come accade negli anni 1970 o 1980– si osserva una frana di più alta intensità; dopo questa si torna a scendere.

Per giustificare l'intensità elevata di queste ultime due frane, vengono stimati dei valori dei danni cumulati un pò più alti di quelli degli altri danni cumulati corrispondenti a frane consecutive, in modo da ottenere degli scarti un pò più alti dalle relative resistenze. La giustificazione che il modello trova, quindi, per come è stato costruito, è che in quegli anni la scogliera ha avuto per un tempo prolungato di tre anni (1991-1992-1993) una bassa coesione e negli ultimi due (1992-1993) ha subito un attacco un pò più forte da parte dei danni meteorologici.

È difficile ipotizzare, vista la complessità del fenomeno, se questa articolazione del modello sia appropriata o se esista una dipendenza delle frane dall'intensità delle precedenti, fattore che non è stato inserito. Gli esperti interpellati a tal proposito esprimevano la convinzione che, proprio per la complessità del fenomeno, non si potesse affermare che esista un'esplicita dipendenza. Si rimandano gli approfondimenti alla Sezione 11.

10 Uso dell'informazione censurata dei dati mancanti

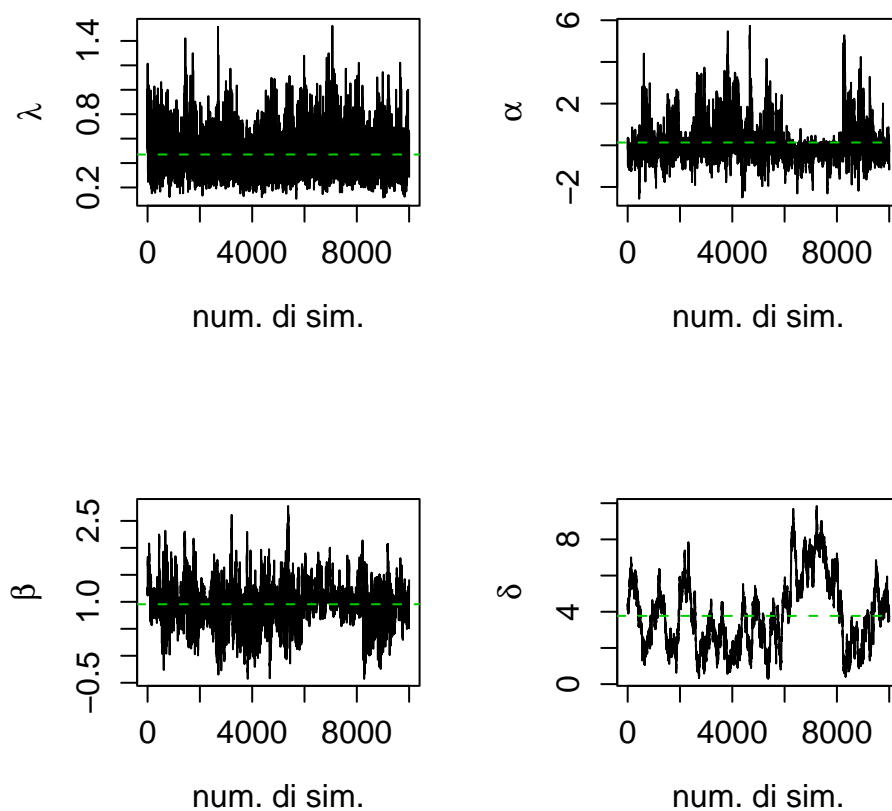
Nel dataset si trova un forte tasso di valori mancanti che impedisce di analizzare lunghi periodi di tempo: il modello in uso, infatti, richiede la continuità dei rilevamenti.

L'informazione censurata non è facile da utilizzare in quanto la prima registrazione che segue ad un periodo di dati mancanti riporta la recessione costiera totale avvenuta in tale periodo, ma le combinazioni di eventi che hanno portato a quella recessione sono molteplici. Si è pensato, allora, di immaginare i possibili scenari come a dei parametri e di utilizzare un algoritmo RJMCMC per recuperare l'infor-

Iperpar.	val. att.	int. cred.
λ	0.469	(0.226, 0.838)
α	0.132	(-1.176, 2.421)
β	0.955	(0.182, 1.765)
δ	3.771	(0.933, 8.014)

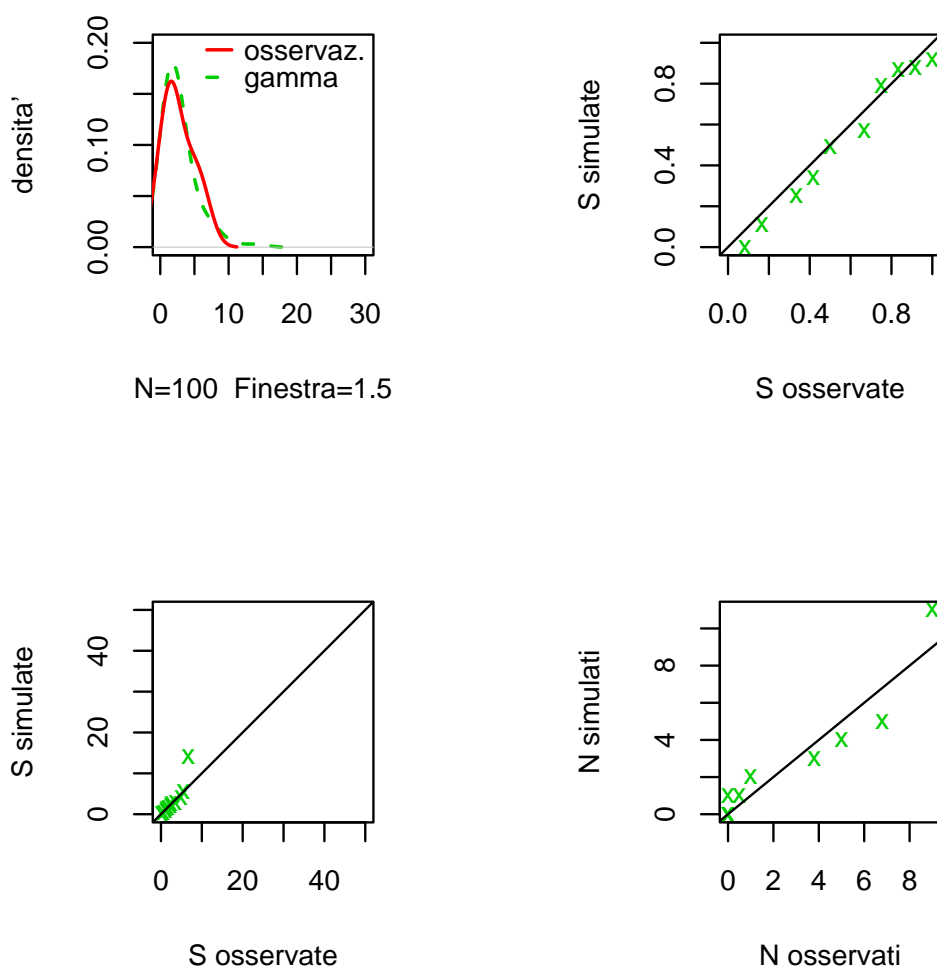
Tabella 5: Stime dei valori attesi a posteriori e intervalli di credibilità al 95% per gli iperparametri, con l'uso della funzione identità nel valore atteso (Sito33 con l'uso di una soglia di 10cm).

Figura 25: Catene Markoviane degli iperparametri del modello in cui si specifica la distribuzione dell'intensità delle frane tramite la distribuzione Gamma, con l'uso della funzione identità nel valore atteso, nell'applicazione al Sito 33, con l'uso di una soglia di 10cm. Le linee tratteggiate orizzontali rappresentano la stima dei valori attesi a posteriori.



mazione censurata.

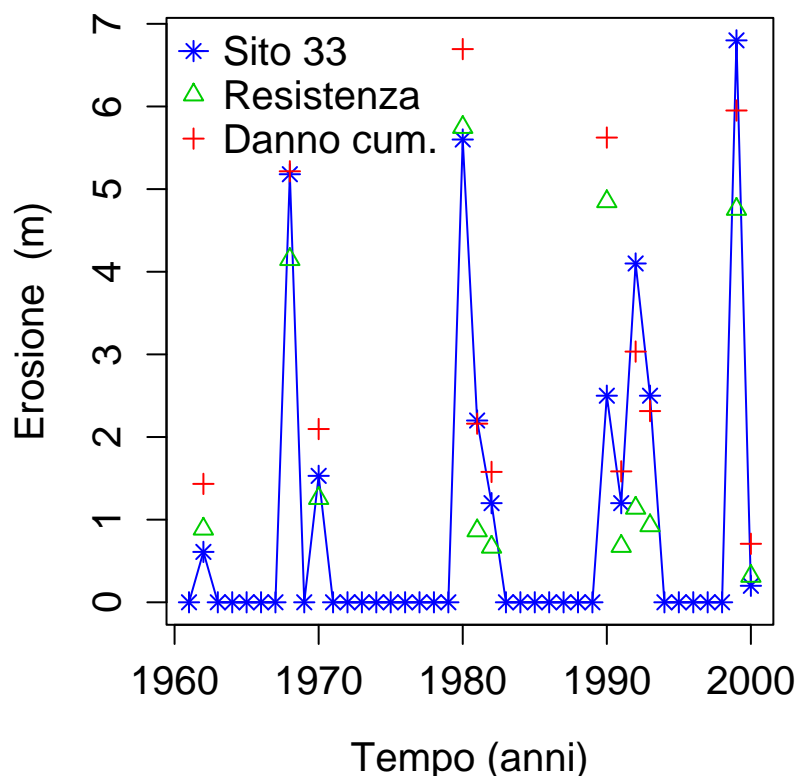
Figura 26: Studio dell'adattamento del modello (funzione identità) tramite simulazioni, ai dati del Sito 33 con l'uso di una soglia di 10cm. In senso orario da in alto a sn: confronto delle densità stimate tra frane simulate ed osservate, Probability plot e Quantile plot; infine, Quantile plot per gli anni di pausa.



Viene ora proposto un esempio per capire quante possibili configurazioni ci possano essere in conseguenza a un dato mancante. Si consideri il terzo valore mancante⁴ del Sito 33, rappresentato come il terzo pallino nella Figura 7 e corrispondente

⁴Il terzo valore mancante è quello più semplice da analizzare. I primi due, invece, corrispondenti agli anni 1952 e 1954, sono più complessi: essi sono molto vicini tra loro e le possibili configurazioni di uno influenzano le configurazioni dell'altro.

Figura 27: Rappresentazione della serie storica del sito 33, dall'anno 1961 al 2000. Per ogni frana, vengono riportate le stime dei valori attesi delle relative resistenze e danni cumulati.

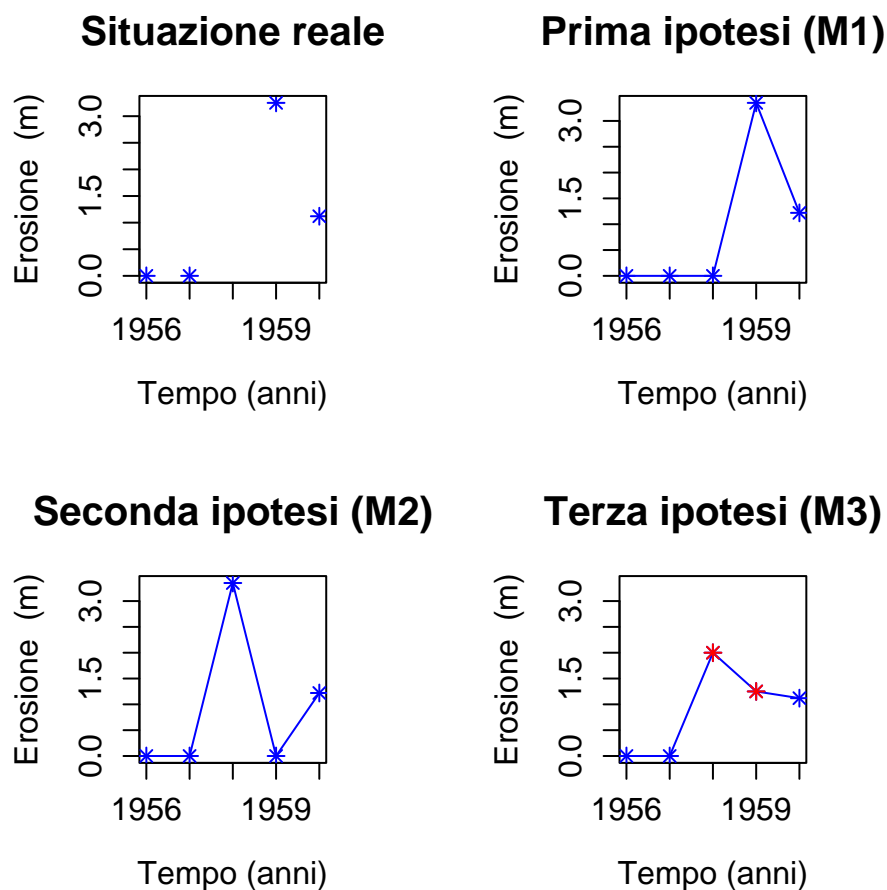


all'anno 1958. Il primo grafico in alto a sinistra della Figura 28 rappresenta la situazione attuale dei dati, per quella parte della serie storica che è sensibile a quello che è realmente successo nel biennio 1958-1959. Negli altri grafici della stessa Figura, vengono rappresentate le possibili configurazioni per tale biennio:

- La frana può essere avvenuta nell'anno del valore mancante (M1),
- in quello della rilevazione (M2),
- oppure è la somma di due frane (M3), accadute una per anno. Nel grafico i due valori sono stati fissati per uso illustrativo.

Si supponga che i tre modelli siano equiprobabili, cioè $P(M1) = P(M2) = P(M3) = 1/3$.

Figura 28: Situazione reale e possibili scenari dovuti al dato mancante nell'anno 1958.



Quando si passa a considerare l'intera serie storica del Sito 33, con tutti e tre i valori mancanti⁵ occorre usare un algoritmo RJMCMC per poter stimare le densità a posteriori delle possibili configurazioni per ogni dato mancante. Per la stima degli iperparametri α, β, δ e degli effetti casuali, la struttura rimane la stessa usata nella Sezione 9; ora si aggiungono tre passi in più perché, per ogni valore mancante, si deve formulare, e poi valutare tramite adeguate probabilità di accettazione, una proposta tra i modelli M1, M2, M3.

I modelli M1 e M2 presuppongono lo stesso numero di frane (e quindi di resistenze e di danni cumulati), in quanto identificano l'avvenimento di una sola frana. L'unica differenza è l'anno in cui prevedono sia successa tale frana. Quindi, quando si valuta uno spostamento tra questi due modelli, non ci sono cambiamenti di numerosità

⁵Visto che si utilizza la soglia di 10cm, il valore mancante corrispondente all'anno 1975 è stato posto uguale a 0, in quanto la recessione cumulata è proprio di 10cm. Di conseguenza, per ogni possibile configurazione le frane sarebbero di intensità inferiore alla soglia.

dello spazio parametrico; i cambi sono di natura deterministici perché è solo il vettore degli anni di pausa che cambia numericamente in alcuni elementi.

Il problema, invece, nasce quando si valuta uno spostamento tra uno di questi due modelli e M3: infatti, in M3 c'è una frana, una resistenza e un danno cumulato in più. Lo spazio parametrico di M3 è più grande, quindi, di 3 parametri.

Per capire con maggior precisione come avviene il RJMCMC, si riprenda la notazione di Green (1995) e si identifichi con X_t lo stato corrente del vettore dei parametri. Si ipotizzi, per un dato mancante, che potrebbe essere quello della Figura 28, di essere nel modello M1 (o equivalentemente M2) e di voler valutare il modello M3 (proposta). Allora, X_t è formato da $(\lambda_t, \alpha_t, \beta_t, \delta_t, D_t, R_t)$, dove D_t e R_t sono i vettori delle resistenze e dei danni cumulati all'iterazione t -esima ed è di lunghezza pari a $d_1 = 2l + 4$. Per passare al modello M3 si devono generare $r_1 = 3$ numeri casuali $u = (u_1, u_2, u_3)$, da una densità $g(u|X_t)$. Il primo numero casuale viene generato così:

$$u_1 \sim U(0, 1)$$

e serve per dividere l'intensità della frana cumulata s^* in due:

$$s_1^* = u_1 s^* \quad \text{e} \quad s_2^* = s^*(1 - u_1).$$

Con gli altri due numeri casuali vengono generati una resistenza e un danno cumulato associati alla frana s_1^* :

$$u_2 \sim Exp(\lambda_t) \quad \text{e} \quad u_3 \sim Exp(1)$$

che saranno rispettivamente $r^* = u_2$ e $d^* = u_2 + u_3$ (il danno cumulato viene calcolato come la resistenza più un eccesso).

I parametri del modello M3 saranno $Y = (\lambda_t, \alpha_t, \beta_t, \delta_t, D_t, R_t, s_1^*, r^*, d^*)^6$. La lunghezza di Y è pari a $d_2 = 2l + 7$.

Quando, invece, si torna indietro da M3 a M1 (o M2) non si devono generare numeri casuali ($r_2 = 0$), ma si devono collasare le due frane s_1^*, s_2^* in s^* e cancellare la resistenza e il danno cumulato associati a s_1^* . In questo modo:

$$d_1 + r_1 = 2l + 4 + 3 = d_2 + r_2 = 2l + 7 + 0$$

è soddisfatta la condizione di bilanciamento delle dimensioni. Ora manca solo di definire $\pi(Y)$ e $\pi(X_t)$ per poter calcolare la probabilità di accettazione del modello proposto M3, come è indicato in Green (1995). Esse sono la verosimiglianza in ciascun caso moltiplicata la probabilità a priori di tutti i parametri coinvolti e la a-priori del modello per lo scenario del valore mancante (che è sempre uguale a $1/3$). Come si può intuire, questo algoritmo non è di facile implementazione nè di facile utilizzo, perché i due vettori dei dati, su cui si calcola la verosimiglianza, cambiano a seconda delle configurazioni dei valori mancanti che si stanno valutando. Inoltre, tale procedura si complica maggiormente quando le configurazioni di un dato mancante influenzano quelle di un altro dato mancante. Di conseguenza, occorre scrivere un programma *ad hoc* per ogni valore mancante del sito e quindi per ogni sito.

⁶Solo s_1^* è casuale, perché s_2^* viene calcolata come differenza tra s^* e s_1^* .

Si passi, ora, a presentare i risultati dell'applicazione del RJMCMC al Sito 33. Nella Tabella 6 vengono riportate le stime dei valori attesi a posteriori degli iperparametri con i relativi intervalli di credibilità al 95%. Nella Figura 29 vengono, invece, riportate le catene markoviane per gli iperparametri. Come nella Sezione 9, sono state generate 100000 iterazioni e se ne sono tenute solo una ogni dieci. Confrontando i due grafici 25 e 29, si può subito capire come l'introduzione dell'informazione censurata dei primi nove anni della serie storica, abbia portato un deciso miglioramento nella convergenza di δ e nel mixing di α e β . Le stime dei valori attesi a posteriori di α e β sono anche cambiati sensibilmente (si confrontino le Tabelle 5 e 6): prendendo in considerazione tutta la serie storica è stata stimata maggiore la propensione a franare in base alle caratteristiche geologiche del sito. Tuttavia, resta sempre molto importante l'azione dei danni cumulati, infatti l'intervallo di credibilità di β è strettamente positivo. Su questo argomento torneremo tra breve.

Nella Figura 30 vengono rappresentate le stime delle densità a posteriori delle configurazioni per ogni valore mancante. Appare evidente, che il modello M3 è quello meno plausibile: questo, infatti, è un sito in cui esiste una tendenza ad osservare blocchi di zero, cioè anni di pausa tra una frana e l'altra. L'ipotesi, quindi, di smezzare le frane in due anni consecutivi, appare meno credibile.

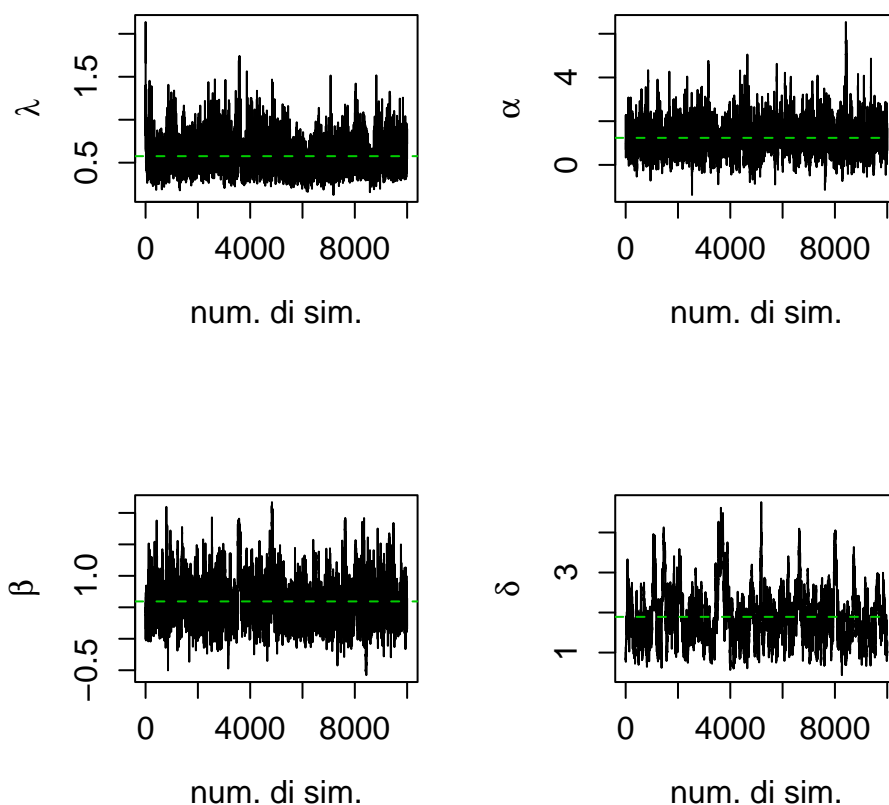
Nella Figura 31, viene rappresentato il Sito 33, utilizzando le configurazioni con più alta probabilità per ogni valore mancante: M2 per il primo, M1 per il secondo e M2 per il terzo. Si noti che l'algoritmo ha dato più credibilità a una struttura simile a quelle che si trovano nel resto della serie storica: ha costruito, infatti, una successione di due frane, come nel 1990, quattro anni di pausa e poi una frana di maggiore intensità delle precedenti. Da qui si capisce perché la stima del valore atteso a posteriori di α sia aumentato: sono aumentate il numero di frane consecutive e il loro livello è piuttosto alto; di conseguenza, il modello spiega questa circostanza aumentando il potere del parametro che regola la parte dell'intensità delle frane dovuta alle caratteristiche geologiche del roccia.

Nella Figura 32, vengono rappresentate le diagnostiche di 100 simulazioni, tramite le stime dei valori attesi a posteriori degli iperparametri, con le osservazioni del Sito 33 (dal 1961 in poi). La scala del Quantile plot, per le frane simulate ed osservate, rimane la stessa in tutto il capitolo per facilitare il paragone tra la funzione identità ed esponenziale. La funzione identità rimane la preferita delle due e le diagnostiche sembrano essere soddisfacenti.

Iperpar.	val. att.	int. cred.
λ	0.576	(0.298, 1.002)
α	1.237	(-0.023, 2.893)
β	0.593	(0.019, 1.432)
δ	1.892	(0.903, 3.585)

Tabella 6: Stime dei valori attesi a posteriori e intervalli di credibilità al 95% per gli iperparametri, nel caso del RJMCMC, con l'uso della funzione identità nel valore atteso (Sito33 con l'uso di una soglia di 10cm).

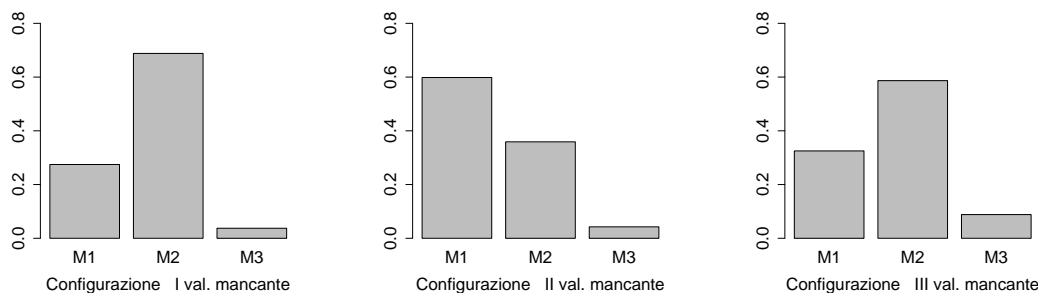
Figura 29: Utilizzo del RJMCMC per l'utilizzo dell'informazione censurata fornita dai tre valori mancanti del Sito 33 (per i dati superiori a 10cm). Vengono rappresentate le catene Markoviane degli iperparametri del modello in cui si specifica la distribuzione dell'intensità delle frane tramite la distribuzione Gamma con l'uso della funzione identità nel valore atteso. Le linee tratteggiate orizzontali rappresentano la stima dei valori attesi a posteriori.



11 Prossimi sviluppi

Come già anticipato nella Sezione 9 finora non è stata considerata l'ipotesi di una dipendenza tra frane successive, ma solamente tra le frane e i relativi periodi di attesa. Gli esperti davano, infatti, poco credito alla possibilità di stabilire un'esplicita dipendenza tra frane successive, per la complessità del fenomeno. Tuttavia, analizzando il Sito 33, sembra che la rupe per stabilizzarsi, dopo aver subito un danno, debba rilasciare una certa quantità di materiale; questo può avvenire in una volta sola oppure in più volte. Sarebbe forse utile, quindi, provare a inserire un parametro nella distribuzione dell'intensità delle frane che tenesse conto, per ogni frana, della

Figura 30: Rappresentazione della stima delle densità a posteriori delle configurazioni per i tre valori mancanti, nel Sito 33.



somma di quelle recessioni antecedenti ad essa, non interrotte da anni di pausa; in tal modo, si relazionerebbe l'intensità della frana allo stadio di assestamento della rupe, tramite la quantità di materiale rilasciato. In alternativa, si potrebbe inserire un parametro che tenga conto solo dell'intensità della frana precedente.

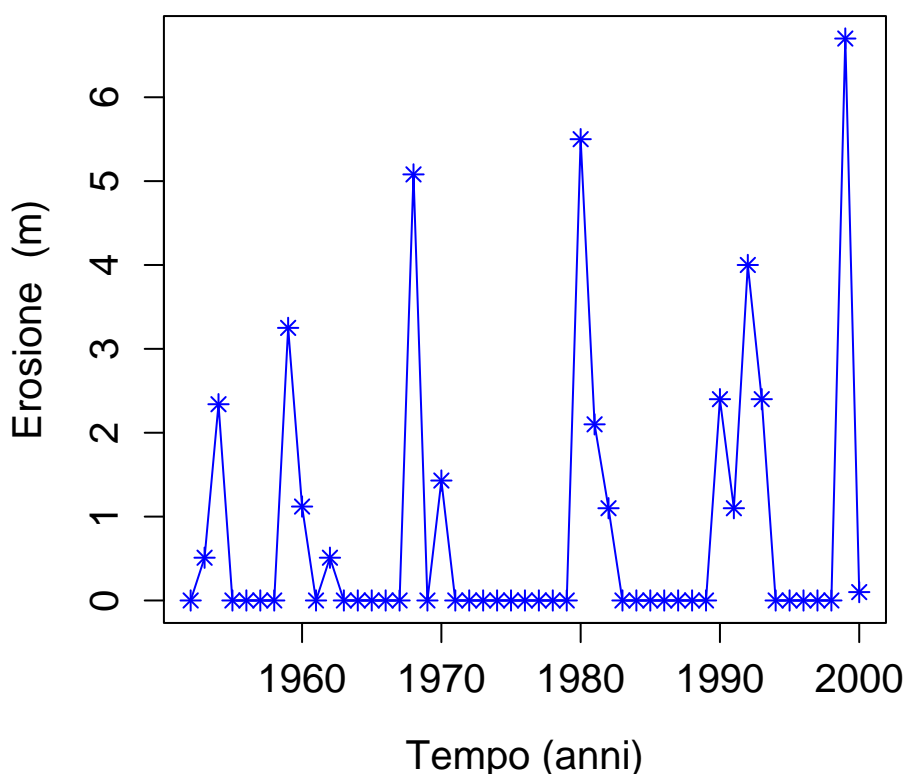
Dopo aver applicato l'algoritmo ad alcuni siti considerati individualmente, si potrebbe pensare a introdurre una relazione spaziale tra di essi. La costa, infatti, cambia le proprie caratteristiche andando da nord a sud ed è quindi naturale che siti vicini abbiano caratteristiche simili. La scrittura di un'algoritmo MCMC con inclusa una struttura spaziale riuscirebbe a trasportare l'informazione tra i siti, permettendo l'identificazione dei parametri anche per quei siti con brevi spezzoni completi di osservazioni. Inoltre, volendo utilizzare l'informazione censurata degli spezzoni delle serie storiche con i dati mancanti, si potrebbe scrivere un RJMCMC e in questo modo l'informazione potrebbe migliorare l'identificazione dei parametri del modello e l'affidabilità dell'inferenza.

Infine, si potrebbe pensare di impostare il lavoro in un altro modo da quello descritto in questa tesi. Infatti, l'aspetto cruciale di questo dataset è la presenza dei dati mancanti che danno luogo a spezzoni di dati con informazione censurata. Per evitare la scrittura di difficili algoritmi RJMCMC, si potrebbe pensare, allora, di interpretare le osservazioni come delle serie storiche a rilevazioni irregolari e modellare, di conseguenza, le recessioni cumulate in questi tempi irregolari, piuttosto delle singole frane.

12 Conclusioni

Il modello gerarchico a effetti casuali proposto sembra interpretare piuttosto bene il fenomeno dell'erosione: permette, infatti, di modellare entrambi i fenomeni di erosione e del tempo di attesa tra le frane. Inoltre, riesce a quantificare la differente influenza, riguardo l'intensità dell'erosione di un sito, della struttura geologica e dei danni meteorologici.

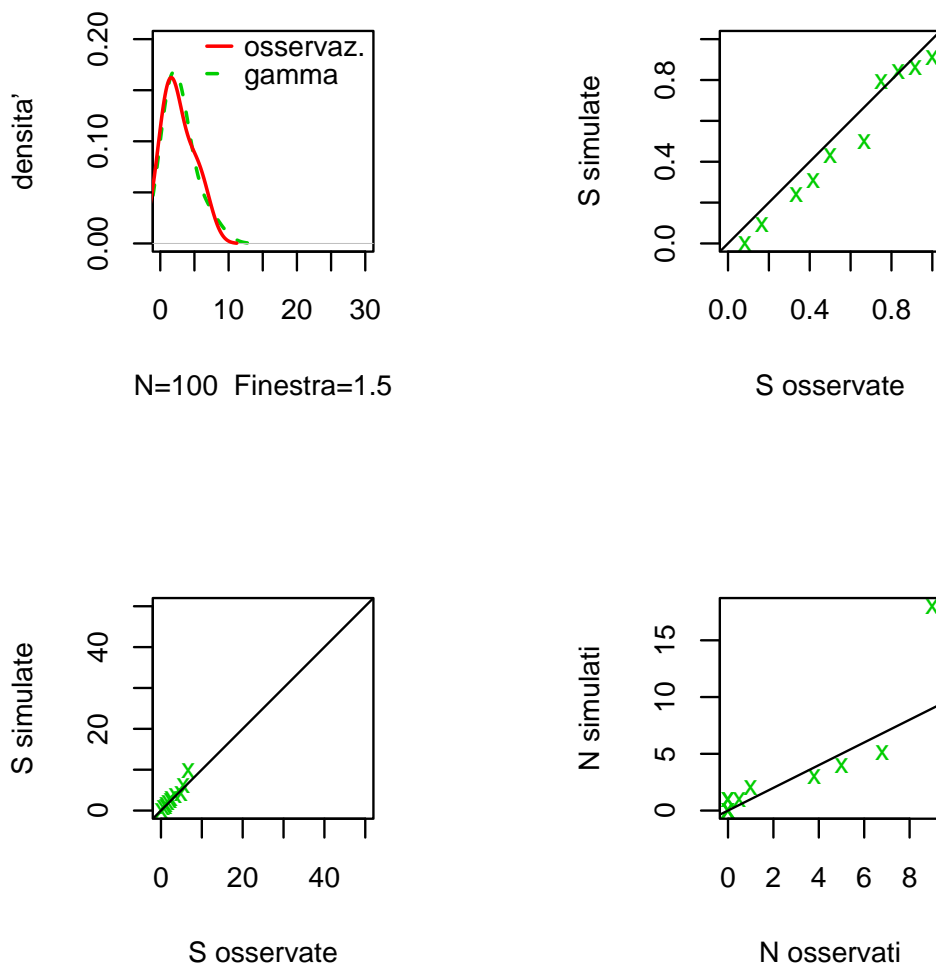
Figura 31: Rappresentazione della serie storica del sito 33, dall'anno 1952 al 2000. Per ogni valore mancante è stata utilizzata la configurazione scelta più di frequente dall'algorithm: M2 per il primo valore mancante, M1 per il secondo e M2 per il terzo.



Per il sito analizzato si è preferito affrontare il problema della possibile distorsione tra il processo di erosione, che avviene ai piedi delle rupi, e la sua misurazione, che avviene nella parte alta, tramite l'utilizzo di una soglia.

In relazione alla funzione che lega il valor medio dell'intensità delle frane a un modello lineare in funzione dei danni cumulati è stata scelta la funzione identità piuttosto dell'esponenziale. Esse si comportano allo stesso modo nell'interpolazione dei dati a disposizione ma danno luogo a differenti code nella distribuzione delle frane. La funzione esponenziale, sebbene non dia alcun problema negli algoritmi usati, non rispecchia la natura del fenomeno erosivo dando del credito a dei valori troppo alti. La funzione identità, invece, sembra rispecchiare molto di più il fenomeno, in quanto a un aumento unitario del danno cumulato corrisponde un aumento lineare nel valore atteso della distribuzione dell'intensità della frana. Con l'adozione della

Figura 32: Studio dell'adattamento del modello (funzione identità) tramite simulazioni, ai dati del Sito 33 con l'uso di una soglia di 10cm. Le stime dei valori attesi sono relative all'output del RJMCMC. Le frane del Sito 33 con cui si confrontano le simulazioni, sono quelle dal 1961 al 2000. In senso orario da in alto a sn: confronto delle densità stimate tra frane simulate ed osservate, Probability plot e Quantile plot; infine, Quantile plot per gli anni di pausa.



funzione identità occorre fare attenzione all'utilizzo degli algoritmi MCMC perché le proposte dei parametri della distribuzione delle frane, devono sempre soddisfare i necessari vincoli di positività, creando delle difficoltà nella convergenza delle catene in presenza di pochi dati. L'effetto dei dati mancanti, infatti, è quello di impedire

di avere a disposizione lunghi spezzoni completi delle serie storiche da utilizzare. Di conseguenza l'articolazione del modello può risultare un po' complessa nella fase di identificazione di tutti i parametri. Nel Sito 33, con l'uso della funzione identità, si hanno avuto delle difficoltà nell'identificazione del parametro di forma della distribuzione delle frane. Tuttavia, il modello è riuscito a spiegare globalmente le dinamiche dell'erosione tramite una compensazione tra i parametri. L'informazione censurata di quegli spezzoni di serie storiche con dati mancanti, può essere recuperata con l'utilizzo dell'algoritmo Reversible Jump MCMC: questo risulta essere molto utile perché, come si è visto nel Sito 33, bastano pochi dati in più per avere un'agevole identificazione di tutti i parametri.

Riferimenti bibliografici

- Coles S. (2001) *An introduction to statistical modeling of extreme values*, Springer, London.
- Gilks W.R., Richardson S. and Spiegelhalter D.J. (Eds.) (1996) *Markov chain Monte Carlo in practice*, Chapman & Hall, London.
- Green P. (1995) Reversible jump markov chain monte carlo computation and bayesian model determination, *Biometrika*, 82, 711–732.
- Hall J.W., Meadowcroft I.C., Lee E.M. and van Gelder P. (2002) Stochastic simulation of episodic soft coastal cliff recession, *Coastal Engineering*, 46, 159–174.
- Kamphuis J.W. (1987) Recession rate of glacial till bluffs, *ASCE Journal of Waterway, Port, Coastal, and Ocean Engineering*, 113, 1, 60–73.
- Pethick J. (1996) Coastal slope development: temporal and spatial periodicity in the Holderness Cliff Recession, in: *Advances in Hillslope Processes*, Anderson M.G. and Brooks S.M., eds., Wiley & Sons, volume 2, 897–917.
- Pethick J.S. and Legget D. (1993) The geomorphology of the Anglian Coast, in: *Coastlines of the Southern North Sea*, Hillen R. and Verhagen H.J., eds., American Society of Civil Engineers, 52–64.

Acknowledgements

This work was supported by the University of Padova (Italy) grant CPDA037217: “Methods for the analysis of extreme sea levels and for coastal erosion”.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it
Most of the working papers can also be found at the following url: <http://wp.stat.unipd.it>

