



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DOTTORATO DI RICERCA IN GESTIONE SOSTENIBILE DELLE  
RISORSE AGRARIE, FORESTALI E ALIMENTARI**

*Indirizzo di Economia, Pianificazione Forestale e Scienze del Legno*

**Ciclo XXVIII**

---

**ANALISI DELL'APPLICAZIONE  
DELL'ALGORITMO DI MASSIMA ENTROPIA ALLO  
STUDIO DELLA DISTRIBUZIONE POTENZIALE  
DEL RISCHIO DI INCENDI BOSCHIVI**

**Cristiano Foderi**

Anni 2012 / 2015





UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

DOTTORATO DI RICERCA IN GESTIONE SOSTENIBILE DELLE  
RISORSE AGRARIE, FORESTALI E ALIMENTARI  
*Indirizzo di Economia, Pianificazione Forestale e Scienze del Legno*

**Ciclo XXVIII**

**Coordinatore Prof. Leonardo Casini**

---

**ANALISI DELL'APPLICAZIONE DELL'ALGORITMO  
DI MASSIMA ENTROPIA ALLO STUDIO DELLA  
DISTRIBUZIONE POTENZIALE DEL RISCHIO DI  
INCENDI BOSCHIVI**

SETTORE SCIENTIFICO DISCIPLINARE AGR/06

Dottorando

**Cristiano Foderi** \_\_\_\_\_

Tutor

**Prof. Enrico Marchi** \_\_\_\_\_

Cotutor

**Prof. Fabio Salbitano** \_\_\_\_\_

Coordinatore

**Prof. Leonardo Casini** \_\_\_\_\_

Anni 2012 / 2015



La lotta generale per l'esistenza degli  
esseri viventi non è una lotta per  
l'energia, ma è una lotta per l'entropia.

---

*Ludwig Boltzmann*



---

# Indice

---

<b>1</b>	<b>Introduzione</b>	<b>1</b>
1.1	Obiettivi della ricerca . . . . .	3
<b>2</b>	<b>Algoritmo di massima entropia</b>	<b>5</b>
2.1	Teoria di base . . . . .	6
2.2	Applicazione in SDM . . . . .	10
2.2.1	Trattamento delle covariate . . . . .	12
2.2.2	Descrizione del funzionamento dell'algoritmo . . . . .	13
2.2.3	Risoluzione del modello . . . . .	14
2.3	Revisione bibliografica dell'applicazione dell'algoritmo di massima entropia al rischio incendi . . . . .	16
<b>3</b>	<b>Applicazione dell'algoritmo di massima entropia ai casi di studio</b>	<b>19</b>
3.1	Metodologie comuni ai casi di studio . . . . .	20
3.1.1	Valutazione della accuratezza dei modelli . . . . .	21
3.1.2	Valutazione dell'importanza delle variabili . . . . .	22
3.1.3	Ricampionamento dei punti di presenza . . . . .	23
<b>4</b>	<b>Elaborazione su scala vasta: Amazzonia Boliviana</b>	<b>25</b>
4.1	Inquadramento del progetto . . . . .	25
4.2	Premesse metodologiche . . . . .	27
4.3	Scelta delle variabili . . . . .	28
4.4	Preparazione dei dati in input . . . . .	29
4.5	Elaborazione . . . . .	30
4.5.1	Selezione delle variabili . . . . .	30

4.5.2 Parametri del modello . . . . .	34
4.6 Risultati . . . . .	34
<b>5 Applicazione a scala locale in Bolivia</b>	<b>43</b>
5.1 Elaborazione a scala locale . . . . .	45
5.2 Risultati . . . . .	46
5.2.1 Dipartimento 1 - Beni . . . . .	46
5.2.2 Dipartimento 2 - Santa Cruz . . . . .	53
5.2.3 Area 1 . . . . .	60
5.2.4 Area 2 . . . . .	64
5.2.5 Area 3 . . . . .	68
<b>6 Elaborazione su scala regionale: Valle d'Aosta</b>	<b>77</b>
6.1 Premesse metodologiche . . . . .	78
6.1.1 Preparazione dei dati . . . . .	80
6.2 Scelta delle variabili: analisi delle componenti principali . . . . .	81
6.3 Elaborazione . . . . .	83
6.4 Risultati . . . . .	84
6.4.1 Incendi estivi . . . . .	84
6.4.2 Incendi boschivi invernali . . . . .	89
6.4.3 Incendi invernali da aree a prevalente copertura di prati e pascoli . . . . .	94
<b>7 Discussione</b>	<b>103</b>
7.1 Valutazione delle prestazioni dei modelli . . . . .	104
7.2 Numerosità campionaria . . . . .	105
7.3 Rapporti con la risoluzione spaziale . . . . .	106
7.4 Metodi di selezione delle variabili . . . . .	107
7.5 Influenza delle variabili sul rischio di incendio . . . . .	109
<b>8 Conclusioni</b>	<b>113</b>
<b>Bibliografia</b>	<b>127</b>

---

# CAPITOLO 1

---

## Introduzione

---

Tradizionalmente il rischio incendi boschivi è stato valutato impiegando dati e metodologie differenti che hanno condotto alla determinazione di vari indici, tutti con lo scopo di individuare e misurare i fattori predisponenti e le variabili che influenzano le dinamiche di propagazione. Gli indici di rischio di incendio boschivo sono costruiti per valutare le condizioni che favoriscono il verificarsi e il comportamento di un incendio, oppure sono stati sviluppati per la determinazione della probabilità di innesco dell'incendio e alle modalità e capacità di espansione.

Molte attività di ricerca in materia di incendi forestali sono state condotte allo scopo di ottenere una migliore comprensione dei fattori spaziali e temporali che influenzano i fenomeni di innesco, con lo scopo di sviluppare modelli che possano essere utilizzati per predire la probabilità di accensione a diverse scale geografiche e con differenti condizioni ambientali [60]. I lavori riguardanti l'applicazione di diverse metodologie, impiegate per determinare i fattori che influenzano gli incendi, sono stati molti. Gli studi sviluppati in tale ambito includono l'utilizzo di test di ipotesi statistiche tradizionali come l'analisi delle frequenze e test di correlazione [11, 45], modelli lineari [17] e di analisi di regressione logistica [11, 13, 48, 56, 57, 66], di alberi di classificazione e regressione [1] e metodi bayesiani [15, 16]. In alcuni lavori su base spaziale è stato indagato il raggruppamento dei punti di innesco e hanno utilizzato K-function e L-function per valutare il *clustering* e la *kernel density smoothing* per fornire rappresentazioni grafiche [27, 32, 61, 67, 68, 72].

Negli ultimi anni si sono moltiplicati studi volti ad applicare ed adattare la modellistica ecologica (più specificamente di ecologia del paesaggio),

sviluppati prevalentemente nell'ambito dello studio della distribuzione di specie animali e vegetali (*Species Distribution Models - SDM*), sono stati testati per la valutazione della distribuzione potenziale del rischio degli incendi boschivi [5]. Tale tendenza si è consolidata in base all'assunto della similarità concettuale ed empirica residente nel comportamento funzionale e dinamico dei due fattori di disturbo ecologico: da un lato l'incendio, dall'altro l'attività alimentare e comportamentale delle specie animali e in particolare degli erbivori.

L'applicazione di SDM per la determinazione della distribuzione del rischio, oltre a risultare concettualmente simile ai metodi applicati per la fauna selvatica<sup>1</sup>, trova ulteriore conferma nel lavoro di Bond e Keely<sup>2</sup>, dove viene evidenziato il parallelismo nel "comportamento" tra gli incendi e gli erbivori, nell'accezione di "consumatori primari", modellatori di ecosistemi e mineralizzatori di sostanza organica [7]. Questo, assieme allo stretto legame degli incendi con le variabili derivate dall'uso del suolo, il clima e l'interazione con le componenti umane del sistema, crea tutti i presupposti per l'approfondimento dell'applicazione dei metodi derivati dai SDM per lo studio della distribuzione del rischio di incendi.

La presente ricerca è stata sviluppata per verificare l'applicabilità e l'eventuale potenza di MAXENT [58] nello studio della distribuzione degli incendi in ambiente naturale. La ricerca si basa su una serie di casi di studio selezionati sulla base di diverse scale geografiche.

MAXENT è un algoritmo di apprendimento automatico (*machine - learning*), considerato, tra gli SDM, tra i più performanti in termini di affidabilità, capacità di classificazione delle variabili e determinazione della complessità computazionale dei modelli [21]; Il metodo, basato sul teorema di concentrazione che massimizza l'entropia [37], sviluppato per dataset di sola presenza, impiegato per la previsione della distribuzione degli inneschi per gli incendi forestali, ha evidenziato performance simili, se paragonato ad altri approcci modellistici come *Random Forest (RF)* o *Generalised Linear Model (GLM)*, come descritto da Bar Massada [5].

---

<sup>1</sup>"Conceptually and methodologically, ignition-distribution modelling is closely related to species-distribution modelling (SDM). The basic approach is to analyse wildfire ignition locations (analogous to locations of species occurrence) in relation to environmental variables hypothesised to influence the spatial distribution of ignitions (or species occurrences). Models estimate the response of wildfire ignitions (or species) to these environmental predictor variables." (BarMassada, 2013 - p. 175).

<sup>2</sup>"We have shown major similarities between fire and herbivory and argued for a more-inclusive view of top-down, or, in this instance, consumer control of biomes. We believe that the global extent of fire as a consumer, its many parallels with herbivory, its role in selecting for particular plant traits and in the evolution of biomes, is worthy of much wider attention from ecologists. There is an added incentive for greater understanding of fire as a globally important consumer. Climate change, habitat fragmentation, the unprecedented transport of highly flammable plants to novel settings, and the ubiquitous overlay of human impacts on fire regimes demand a new level of synthetic understanding for our peaceful coexistence with this charismatic beast." (Bond and Keely, 2005 - p. 393)

La parte finale dell'elaborato vengono discusse in modo sinottico le comparazioni fra i risultati dei diversi casi di studio in modo da evidenziare le performance scala-dipendenti della metodologia. .

Tali differenze sono state ricercate principalmente in funzione della dimensione campionaria dei punti di presenza e al rapporto con la risoluzione spaziale delle variabili in funzione della variabilità del fenomeno, sia stagionale che in relazione alle attività antropiche locali, spesso legate al tessuto socio-culturale delle aree indagate.

Infine è stata prestata particolare attenzione al problema della colinearità<sup>3</sup> in funzione della selezione delle variabili e dei fattori da modellizzare, per limitare la ridondanza di informazioni generate da alti livelli di correlazione fra covariate.

### 1.1 Obiettivi della ricerca

---

Dall'analisi delle premesse introduttive e, alla luce dell'esperienza condotta durante l'applicazione della metodologia ai casi di studio descritti nei successivi capitoli, si è ritenuto opportuno, oltre all'analisi critica dei risultati ottenuti nei singoli casi, evidenziare alcune peculiarità, relative all'applicazione dell'algoritmo di massima entropia, nello studio della distribuzione potenziale del rischio di innesco per gli incendi forestali.

Obiettivo generale della ricerca è quello di definire gli ambiti ed i criteri di applicabilità dell'algoritmo di massima entropia allo studio del rischio di innesco degli incendi forestali.

A corollario di tale obiettivo si volevano delineare i punti di forza, in termini di accuratezza e applicabilità, e le debolezze dell'algoritmo di massima entropia. Attraverso la comparazione dei risultati dell'applicazione a diversi casi di studio, si sono volute testare le prestazioni di MAXENT in relazione alle principali cause di incertezza, spesso comuni anche ad altri approcci modellistici, ancora scarsamente trattate in letteratura.

In particolare, gli obiettivi specifici posti hanno riguardato:

- 1) la comparazione delle *performance* nell'applicazione a diverse scale geografiche dell'algoritmo MAXENT;
- 2) testare le risposte dell'algoritmo in funzione della numerosità campionaria per i punti di presenza necessari all'elaborazione dei modelli;
- 3) valutare le prestazioni dei modelli in rapporto con la risoluzione spaziale delle covariate e l'estensione dell'area di indagine;

---

<sup>3</sup>“Collinearity refers to the non independence of predictor variables, usually in a regression-type analysis. It is a common feature of any descriptive ecological data set and can be a problem for parameter estimation because it inflates the variance of regression parameters and hence potentially leads to the wrong identification of relevant predictors in a statistical model.”(Dormann, 2013 - p. 27)

- 4) definire e testare una metodologia di selezione e valutazione delle variabili da utilizzare per il calcolo dei modelli;
- 5) individuare le componenti ambientali, a diverse scale geografiche, che maggiormente influiscono nella definizione della distribuzione del rischio di incendio.

---

## Algoritmo di massima entropia

---

L'algoritmo di massima entropia (MAXENT) è una metodologia generica per caratterizzare distribuzioni di probabilità a partire da informazioni di base incomplete. Nella sua applicazione alla realizzazione di modelli di distribuzione specifica, viene posto l'obiettivo di ottenere una distribuzione che sia in perfetto accordo con le informazioni disponibili, evitando di formulare (o di basarsi su) assunti che non siano direttamente deducibili dai dati disponibili.

Il Principio della Massima Entropia venne introdotto nel 1957 da Jaynes in un articolo in cui riuscì a riformulare risultati di meccanica statistica precedentemente ottenuti da Gibbs, in termini di distribuzioni di probabilità, utilizzando appunto un nuovo approccio teorico fisico-statistico e una nuova metodologia di lavoro che oltre a semplificare notevolmente la notazione e i dettagli matematici ne permise un'ulteriore estensione. Fu infatti possibile studiare problemi di meccanica statistica come particolari istanze di problemi inferenziali basati sulla parziale conoscenza dei dati reali.

In meccanica statistica, l'entropia è una coordinata generalizzata che viene interpretata come una misura del disordine presente in un sistema fisico qualsiasi, incluso, come caso limite, l'universo. In termodinamica, lo stato di equilibrio, verso il quale i sistemi reali evolvono spontaneamente e che quindi possiede caratteristiche di stabilità, è caratterizzato da un valore massimo di entropia. Per converso la diminuzione di entropia, pur non essendo spontaneamente perseguita dai sistemi reali, non è impossibile ma richiede l'imposizione di vincoli esterni. Questa è l'osservazione centrale che ha ispirato il principio di massima entropia che serve a riformulare la meccanica statistica ma che è suscettibile di una generalizzazione che ne fa il cardine della teoria dell'infe-

renza, ovvero della teoria della ricostruzione di informazione probabilistica da dati incompleti. Massimizzare l'entropia significa prendere in considerazione tutte le scelte possibili, senza escluderne alcuna. Al contrario, l'indebita limitazione del campo di scelta, in assenza di prescrizioni esplicite in questo senso nella formulazione del problema, equivale all'imposizione di un vincolo arbitrario. Il formalismo che scongiura questa eventualità indesiderata è proprio quello codificato dal principio di massima entropia.

Il Principio della Massima Entropia è enunciato nella seguente definizione:

*In presenza di dati e/o evidenze sperimentali riguardanti un ben determinato fenomeno fisico-statistico per stimarne la relativa distribuzione di probabilità è sufficiente scegliere un modello che sia consistente con i dati disponibili ma che altrove abbia la massima entropia.*

Massimizzare l'entropia di una distribuzione di probabilità, significa quindi generalizzarne il comportamento statistico su tutto ciò che non è conosciuto a priori. Si massimizza cioè l'incertezza assunta sui parametri del sistema non noti ottenendo dei modelli basati solo su dati certi.

### 2.1 Teoria di base

---

Al fine di comprendere il sistema con cui l'algoritmo funziona sono necessari alcuni richiami teorici, derivati dalla meccanica statistica e dai lavori sulla teoria dell'informazione di Shannon [65].

Di seguito si riporta l'enunciazione del teorema di concentrazione dell'entropia proposta nel 1982 da Jaynes [37].

- i) Sia dato un esperimento casuale con  $n$  esiti possibili e sia compiuto per  $N$  tentativi. Avremo allora  $n^N$  sequenze di uscite di  $N$  risultati ed ogni sequenza  $N_i$  avrà frequenza  $f_i = N_i/N$ , con  $1 \leq i \leq n$  (ad esempio in meccanica statistica  $N_i$  potrebbe essere lo stato quantistico della  $i$ -esima particella di un sistema di  $N$  particelle), in questo contesto, l'entropia del sistema può essere definita come

$$H(f_i) = - \sum_{i=1}^n (f_i \ln(f_i)) \quad (2.1)$$

- ii) Consideriamo ora una sottoclasse  $C$  delle  $n^N$  sequenze di uscite con  $m < n$  vincoli della forma

$$\sum_{i=1}^n A_{ij} f_i = d_j \quad (2.2)$$

con  $1 \leq j \leq m$

Allora, le quantità fisiche misurate  $m$  presentano una natura definita dalla matrice  $A_{ij}$  e  $d_j$  sono i dati che otteniamo dalle misure. I dati  $d_j$ , però, al contrario dell'entropia  $H$ , non forniscono informazioni riguardo alle frequenze  $f_i$ .

Se prendiamo, infatti, una frazione  $F$  della sottoclasse  $C$  di sequenze di uscite, allora l'entropia sarà compresa entro questo intervallo

$$H_{max} - \Delta H \leq H(f_i) \leq H_{max}$$

con  $H_{max}$  calcolato con l'algoritmo dei moltiplicatori di Lagrange e funzione di partizione.

- iii) Per risolvere la sommatoria dell'algoritmo MAXENT, viene definita la funzione di partizione

$$Z(\lambda_1 \dots \lambda_m) \equiv \sum_{i=1}^n \exp \left( - \sum_{j=1}^m \lambda_j A_{ji} \right) \quad (2.3)$$

Quindi

$$H_{max} = \log Z + \sum_{j=1}^m \lambda_j d_j \quad (2.4)$$

nella quale, applicando i moltiplicatori di Lagrange,  $\lambda_j$  è dato da

$$\frac{\partial}{\partial \lambda_j} \log Z + d_j = 0 \quad (2.5)$$

con  $1 \leq j \leq m$

dove  $m$  è un insieme, di dimensione non nota, di equazioni simultanee.

La distribuzione di frequenza con la massima entropia sarà quindi

$$f_i = Z^{-1} \exp \left( - \sum_{j=1}^m \lambda_j A_{ji} \right) \quad (2.6)$$

Tutte le altre distribuzioni  $f_i^1$  consentite dai vincoli definiti nella 2.2 avranno differenti entropie, tutte minori dell'entropia massima  $H_{max}$

La concentrazione delle sequenze di uscita, prossime al limite superiore, è data dalla seguente formulazione, per la cui dimostrazione si rimanda al lavoro di Jaynes del 1982 [37].

*Teorema di Concentrazione* : Asintoticamente,  $2N\Delta H$  è distribuito sulla classe  $C$  secondo  $\chi^2$  con  $k = n - m - 1$  gradi di libertà, indipendentemente dalla natura dei vincoli. Ovvero, denotando che il  $\chi^2$  critico per  $k$  gradi di libertà

## Capitolo 2. Algoritmo di massima entropia

---

al  $100P$  livello di significatività percentuale è dato da  $\chi_k^2(P)$ ,  $\Delta H$  è espresso nei termini dell'area sottesa alla coda superiore  $(1 - F)$  della distribuzione, secondo

$$2N\Delta H = \chi_k^2(1 - F) \quad (2.7)$$

Risulta a questo punto necessario chiarire che il suddetto teorema è una affermazione sul conteggio degli stati possibili e non sulle probabilità che il sistema raggiunga un certo stato. Frequenze e probabilità coincidono solo nel caso si consideri valido il principio di equiprobabilità a priori<sup>1</sup>.

Per esempio, se consideriamo il classico caso del lancio di un dado a sei facce, lanciato 1000 volte, è facile affermare che la frequenza di uscita di una faccia qualsiasi è  $f_i^* = 1/6$ ; questo, dal punto di vista del teorema di concentrazione dell'entropia può essere tradotto così:

- a) In assenza di altri vincoli oltre la normalizzazione data da  $\sum f_i = 1$ , l'entropia raggiunge il massimo valore con

$$H_{max} = \ln 6 = 1.79176$$

a causa dell'uniformità di distribuzione  $f_i^*$  (dove non sono presenti altri vincoli).

- b) Applicando il teorema di concentrazione dell'entropia, in questo caso, con  $n = 6$  e  $N = 1000$ , si hanno  $6 - 1 = 5$  gradi di libertà, che nella tavola di distribuzione del  $\chi^2$ , al livello di significatività convenzionale del 5 per cento, riporta  $\chi_5^2(0.05) = 11.07$ . Quindi, il 95 per cento di tutte le possibili risultati avrà entropia compresa in

$$2N\Delta H = 11.07$$

ovvero

$$1.786 \leq H \leq 1.792$$

- c) Si supponga, adesso, che dai dati emerga un nuova evidenza, per cui un'influenza sistematica (dovuta, ad esempio, al fatto che si reiteri l'esperimento con un dado truccato) porta per  $N = 1000$  lanci non un valore 3,5 della relazione, con matrice  $A=i$  ( $\sum_{i=1}^6 i f_i = 3,5$ ), ma un valore 4,5, spostando la distribuzione delle frequenze dalla condizione di uniformità

$$\sum_{i=1}^6 i f_i^* = 4,5$$

---

<sup>1</sup>Il postulato fondamentale della meccanica statistica, conosciuto anche come postulato dell'equiprobabilità a priori, è il seguente: *Dato un sistema isolato in equilibrio, ogni microstato ha eguale probabilità di manifestarsi.* Questo postulato è una premessa fondamentale in meccanica statistica in quanto stabilisce che un sistema in equilibrio non ha nessuna preferenza per nessuno dei suoi microstati possibili. Dati  $\Omega$  microstati a una particolare energia, la probabilità di trovare il sistema in un particolare microstato è  $p = 1/\Omega$ . Questo postulato è necessario perché permette di concludere che per un sistema in equilibrio, il macrostato che può risultare dal maggior numero di microstati è anche il macrostato più probabile del sistema.

È possibile considerare, questa situazione, come una simulazione di ciò che frequentemente accade nei fenomeni fisici e naturali, dove la stima della frequenza del verificarsi di un particolare evento non si distribuisce in maniera uniforme (ovvero, quel determinato fenomeno risulta essere favorito o sfavorito nella probabilità di verificarsi), a causa di vincoli imposti dal sistema stesso.

- d) La distribuzione  $f_i^*$ , che ha la massima entropia assoggettata al vincolo, e definita per  $n = 6$ ,  $m = 1$  e  $A = i$ , è data da

$$f_i^* = e^{-\lambda^i (\sum e^{-\lambda^i})^{-1}} \quad (2.8)$$

con  $\lambda = -0.37105$

per la cui dimostrazione si rimanda a [36].

La distribuzione, quindi ha entropia  $H_{max} = 1.61358$  che, se confrontata con i valori del range di entropia massima calcolati in assenza di vincoli ( $1.786 \leq H \leq 1.792$ ), ci fa capire quanto questa particolare influenza sistematica (il dado truccato, in questo caso) costituisca un vincolo tanto forte da restringere le possibilità ad una piccola sottoclasse di tutte le  $6^N$  combinazioni di risultati ipotizzabili *a priori*.

- e) Ancora una volta, applicando il teorema di concentrazione, si hanno  $6 - 1 - 1 = 4$  gradi di libertà, e, il 95 percento di tutti i possibili risultati, consentiti dal vincolo, saranno in un range definito da

$$\Delta H = (2N)^{-1} \chi_4^2(0.05) = 0.00474$$

o, con sufficiente approssimazione per arrotondamento

$$1.609 \leq H \leq 1.614$$

Di conseguenza, per quanto riguarda "l'ipotesi nulla", la quale prevede che nessuna ulteriore influenza sistematica, oltre quella considerata nel vincolo, agisca sull'esperimento, ci sono meno del 5 percento di probabilità che la frequenza di distribuzione abbia valori di entropia al di fuori dell'intervallo definito con l'applicazione del teorema di concentrazione.

Partendo quindi da informazioni incomplete sul sistema è possibile concludere che non solo l'entropia massima fornisce la più probabile sequenza di risultati, ma che qualsiasi valore di entropia sarà molto vicino al massimo per tutte le possibili sequenze di risultati, con  $\Delta H \sim N^{-1}$ .

- f) In un numero di ripetizioni  $N$ , in un esperimento randomizzato, l'*i*esimo risultato si verifica  $N_i = N f_i$  volte, per  $1 \leq i \leq n$

Il numero di possibili combinazioni  $W$  con le quali è possibile realizzare uno specifico campione di risultati è dato, secondo Laplace e Bernoulli, dal coefficiente multinomiale

$$W(f_1 \dots f_i) \equiv \frac{N!}{(N f_1)! \dots (N f_n)!}$$

che per  $N \rightarrow \infty$  e applicando l'approssimazione di Stirling restituisce la funzione di entropia

$$N^{-1} \log W \rightarrow H(f_1 \dots f_i)$$

Risulta quindi che distribuzioni a maggior entropia hanno maggior molteplicità di combinazioni  $W$ .

Anche quando si riducono le reiterazioni dell'esperimento a  $N = 50$ , nell'esempio dei dadi, con  $\sum N_k = 50$ ,  $N_k = \{3, 4, 6, 8, 12, 17\}$  per il caso senza vincoli e  $N_k^1 = \{0, 1, 7, 16, 18, 8\}$  per il caso in cui agisca un'influenza sistematica, il calcolo delle combinazioni che realizzano la distribuzione di massima entropia di  $N_k$  è  $W/W^1 = \{(7!16!18!)/(3!4!6!12!17!)\} = 38220$

Questo porta ad affermare che, nel caso in cui i dati a disposizione non riescano a determinare a pieno la distribuzione  $f_i$  è prudente utilizzare comunque, per fini inferenziali, la distribuzione che massimizza entropia. Questo perché le predizioni di MAXENT sono le più affidabili che possono essere fatte con le informazioni disponibili secondo Jaynes [37]. La scelta di altri stimatori, infatti, rischierebbe di prendere in considerazione una sottoclasse piccola di sequenze di risultati possibili e, di conseguenza non rappresentativa, che potrebbero accadere sì, ma con una frequenza talmente bassa che non si potrebbero ripetere, controllare e osservare nelle reiterazioni dell'esperimento.

## 2.2 Applicazione in SDM

---

In questo paragrafo, viene riportata la spiegazione statistica del metodo di modellizzazione MAXENT, con particolare riferimento sulle assunzioni della procedura modellistica. Infatti, la comprensione delle relazioni fra assunzioni di base, scelte di processo e risultati che ne derivano è ritenuta fondamentale per illustrare il campo ed i limiti di applicabilità del modello nell'esplorare le relazioni ecologiche dei fenomeni naturali con i dati ambientali.

### Il caso dei dati di sola presenza

Analogamente a ciò che succede per in molti casi nelle indagini di distribuzione specifiche, anche per gli incendi, i dati disponibili possono essere limitati

temporalmente o non essere disponibili per tutto il territorio indagato. Ad esempio, nel caso degli incendi non è possibile affermare *a priori* che l'assenza di registrazioni, in una zona, sia dovuta al fatto che lì non ci sono le condizioni per il verificarsi di un incendio, potrebbe essere avvenuto in passato e non registrato, oppure potrà verificarsi in futuro. Per questo motivo è necessario massimizzare l'utilità delle informazioni disponibili per i dati di sola presenza (*presence-only data*), non potendo fare assunzioni riguardo al motivo dell'assenza in una determinata area.

L'espansione di utilizzo di dati *presence-only* per la modellizzazione di distribuzione specifiche ha suscitato un'ampia discussione sui tipi di distribuzione che possono essere stimate con solo i dati di sola presenza a differenza dei dati di presenza-assenza (vedi, [38, 39, 43]). Come sostenuto in queste pubblicazioni, l'argomento è complesso poiché strettamente connesso con la qualità dei dati che dipende da: quantità e precisione dei dati; rilevanza ecologica di variabili predittive; disponibilità di informazioni sui disturbi; limitazioni della diffusione a causa di interazioni biotiche; metodo di modellazione; scala di analisi. Jiménez-Valverde suggerisce che i dati *presence-only*, in un certo senso, possono svincolare l'analisi dai problemi generati dalla registrazione di dati di assenza inaffidabili [39], in particolare sottolineando che le assenze, spesso, sono generate o comunque riconducibili a interazioni, vincoli di dispersione e disturbi che possono precludere la modellizzazione di eventuali distribuzioni potenziali.<sup>2</sup>

Assumendo quindi di avere a disposizione dati di sola presenza, sia definito con

- $L$ , la regione geografica in cui si osserva il determinato fenomeno da modellizzare;
- $y = 1$  definisca la condizione di presenza e  $y = 0$  la condizione di assenza;
- $z$ , il vettore delle covariate ambientali che rappresentano le condizioni ambientali e che sono disponibili su tutto  $L$ ;

---

<sup>2</sup>“The kind of absence data used for the calibration of the models and the modelling technique used condition the characteristics of the distribution of the focal species that are described by model results. Species distributions are not only constrained by abiotic (e.g. climate) factors. Rather, they are also shaped by biotic interactions; dispersal constrains; anthropogenic effects; stochastic events; and other historical, unique, and contingent factors. These effects can only be accounted for using data on the absence of the species to restrict model predictions. Hence, the realized distribution of a species cannot be estimated without data on its absence from environmentally suitable localities. In addition, predictors must not only include environmental variables; it is necessary to incorporate other factors that might be restricting the distribution of the species. If, on the contrary, the goal is to estimate the potential distribution of a species, the absences caused by nonenvironmental factors must be avoided. Here, absence data must come from environmental conditions that are known to be unsuitable for the species. If information on absence due to environmental constraints is not available, two alternatives can be taken: (i) generate absences outside the environmental domain where the species is present and use them for model parameterization; and (ii) use profile techniques in order to estimate the location of climatically suitable places.” (Jiménez-Valverde et al., 2008 - p. 886 )

- *background*, sono tutte le posizioni spaziali (o un insieme casuale di esse) contenute in  $L$

La densità di probabilità<sup>3</sup> incondizionata delle covariate in  $L$  è data da  $f(z)$ , di conseguenza la densità di probabilità della presenza del fenomeno è data da  $f_1(z)$  e la densità di probabilità dell'assenza da  $f_0(z)$ , sempre nello spazio  $L$ .

La quantità che si vuole stimare è la probabilità, condizionata dall'ambiente, di presenza del fenomeno:  $Pr(y = 1|z)$ . Con dati di sola presenza è possibile costruire un modello di  $f_1(z)$ , che però non è sufficiente per approssimare la probabilità di presenza. Utilizzando sia i dati di sola presenza, sia i dati di *background* è possibile definire  $f_1(z)$  e  $f(z)$  che, utilizzando la regola di Bayes, portano a definire

$$Pr(y = 1|z) = \frac{f_1(z)Pr(y = 1)}{f(z)} \quad (2.9)$$

Come è facile vedere l'unico termine incognito, dai presupposti definiti fino a questo punto, è  $Pr(y = 1)$ , che fornisce la misura della prevalenza del fenomeno (in termini di proporzione di siti in cui è presente) nell'area di indagine. Come dimostrato da Ward, questa prevalenza non è esattamente determinabile partendo da dati di sola presenza [69], poiché dipende fortemente dalla dimensione campionaria. D'altro canto, visti i problemi di probabilità di rilevazione che affliggono i dati di assenza [70], anche i dati di presenza-assenza non possono produrre buone stime riguardo alla prevalenza dei fenomeni.

### 2.2.1 Trattamento delle covariate

Le covariate o predittori sono le variabili indipendenti in un modello. Nei SDM, queste includono fattori ambientali (ad esempio, la piovosità, le temperature, la topografia o l'uso del suolo), che possono avere relazioni tendenzialmente complesse rispetto al fenomeno indagato. Per questo motivo, di solito, è opportuno adottare funzioni non lineari in fase regressiva [3], applicando delle trasformazioni alle covariate. I modelli complessi sono solitamente costruiti come combinazione lineare di queste funzioni di base, come accade per i modelli additivi generalizzati (*Generalised Addictive Model - GAM*) e i modelli lineari generalizzati (*Generalised Linear Model - GLM*). In *machine learning*, funzioni di base e altre trasformazioni di dati disponibili sono denominate *features*, che corrispondono ad un set esteso di trasformazioni delle covariate originali.

La funzione generata da MAXENT, che meglio descrive il fenomeno (inteso come il *fit* del modello rispetto ai dati di presenza e alle covariate) è di solito

---

<sup>3</sup>densità o le funzioni di densità di probabilità descrivono la relativa probabilità di variabili casuali nel loro range e possono essere univariate o multivariate.

definita tramite molte *features*, generate per le singole covariate e per combinazioni di più covariate, il che significa che nella maggior parte dei modelli ci saranno più *features* rispetto al numero di variabili originali. La metodologia prevede sei classi di *features*: lineare, di prodotto<sup>4</sup>, di secondo grado, *hinge*<sup>5</sup>, di soglia e categoriche. Molte *features* di soglia o *hinge* possono essere *fittate* simultaneamente rispetto ad una covariata, generando una funzione potenzialmente molto complessa.

### 2.2.2 Descrizione del funzionamento dell'algoritmo

MAXENT utilizza i dati delle covariate estratti in corrispondenza dei punti di presenza e dei punti di *background* per stimare il rapporto  $f_1(z)/f(z)$ . Di tutte le possibili distribuzioni di  $f_1(z)$ , l'algoritmo seleziona quella che maggiormente si avvicina alla distribuzione di  $f(z)$ . Minimizzare la distanza da  $f(z)$  trova la sua ragione di essere nel fatto che  $f(z)$  è un modello nullo per  $f_1(z)$ , ovvero, senza nessun punto di presenza non ci sarebbe ragione di aspettarsi una qualche "preferenza", da parte di un determinato fenomeno, al verificarsi in concomitanza dei particolari condizioni ambientali rispetto a tutte le altre. Questo equivale a dire che non ci sono alternative migliori rispetto a quella di prevedere che il fenomeno si verifica con maggiore probabilità in maniera proporzionale alla disponibilità delle condizioni ambientali stesse, e quindi come funzione di esse.

In MAXENT la distanza da  $f(z)$  è data dall'entropia relativa di  $f_1(z)$  rispetto a  $f(z)$ . Le informazioni riguardi la densità delle covariate  $f(z)$ , in  $L$ , sono estratte dai punti di *background*, provvedendo a fornire la base per la comparazione con la densità delle covariate nei punti di presenza del fenomeno, i quali vanno a generare i vincoli da imporre alla risoluzione del modello, di modo da far sì che la soluzione sia quella che meglio riflette le informazioni corrispondenti ai punti di presenza registrati. Ad esempio, se si utilizza, come covariata, la temperatura massima nel mese più caldo (*TMhm*) dell'anno, allora il vincolo imposto alla soluzione del modello assicurerà che *TMhm* media per le  $f_1(z)$  stimate, sarà prossima alla media di *TMhm* registrata nei punti di presenza. A questo punto, la distribuzione del fenomeno sarà stimata tramite la minimizzazione della distanza tra  $f_1(z)$  e  $f(z)$ , soggetta al vincolo che prevede che la *TMhm* media stimata per  $f_1$  e per il valore medio delle altre covariate, sia prossima al valore medio registrato nei punti di presenza.

Nella definizione di Phillips [58], l'obiettivo era definire  $\pi(x) = Pr(x|y = 1)$ , ovvero la distribuzione di probabilità del fenomeno su una località  $x$ , e

<sup>4</sup>I prodotti sono prodotti di tutte le possibili combinazioni a coppie delle covariate, che consentono di *fittare* interazioni semplici.

<sup>5</sup>Le *hinge features*, se usate singolarmente, restituiscono modelli analoghi ai modelli additivi generalizzati (GAM), ovvero, modelli additivi, con funzioni non lineari dotati di complessità variabile ma senza i passaggi repentini generati dalle *feature* di soglia.

che viene definita "*raw distribution*". Come dimostrato da Elith, massimizzare l'entropia della *raw distribution* è equivalente a minimizzare l'entropia relativa di  $f_1(z)$  rispetto  $f(z)$ , per cui le due formulazioni sono equivalenti [23].

Ricapitolando, il modello nullo per la *raw distribution*, come già detto, corrisponde alla distribuzione uniforme sul paesaggio e in spazio ambientale discreto, il modello nullo equivalente per  $z$  è  $f(z)$ . I vincoli sono stati definiti sulla base delle covariate, ma come spiegato nella sezione precedente, MAXENT adegua (*fit*) il modello sulla base di caratteristiche (*features*) costruite in base alla trasformazione delle covariate, in tale modo la procedura di modellizzazione consente di valutare relazioni complesse rispetto al semplice accoppiamento spaziale delle covariate con l'ambiente di riferimento.

I vincoli che limitano al distribuzione alla massima entropia, imposti dalle condizioni ambientali, quindi, sono estesi in modo da passare da essere vincoli delle medie delle covariate ad essere vincoli delle medie delle *features*.

Se viene definito con  $h(z)$  il vettore delle *features* e con  $\beta$  il vettore dei coefficienti, la minimizzazione dell'entropia relativa risulta appartenere alla famiglia dei modelli esponenziali secondo

$$f_1(z) = f(z)e^{\eta(z)} \quad (2.10)$$

dove  $\eta(z) = \alpha + \beta h(z)$

e  $\alpha$  è la costante di normalizzazione che garantisce che  $f_1(z)$  sommi a 1.

Risulta chiaro, quindi, che il *target* del modello è  $e^{\eta(z)}$ , che è lo stimatore del rapporto  $f_1(z)/f(z)$ . Similarmente ai GLM, è un modello log-lineare che, in questo caso, dipende contemporaneamente dai punti di presenza e dai campioni estratti dal *background*, in modo che le caratteristiche dello spazio ambientale siano intimamente legate con la soluzione risultante dal modello [23].

### 2.2.3 Risoluzione del modello

Come già evidenziato nell'equazione 2.10, per la risoluzione del modello, è necessario individuare i coefficienti  $\beta$  che faranno sì che le limitazioni imposte dai vincoli siano soddisfatte, ma non in modo troppo restrittivo per evitare il rischio di produrre un modello con limitata capacità di generalizzazione a causa del fenomeno di *overfit*<sup>6</sup>.

---

<sup>6</sup>*overfit* o sovradeterminazione è il fenomeno secondo cui, un modello, per arrivare a spiegare "tutto" di una certa configurazione di dati, considera come informazioni rilevanti anche ciò che è imputabile alle inevitabili fluttuazioni sperimentali, che sono tipiche del particolare campione utilizzato per la costruzione del modello e che, di conseguenza, non sono generalizzabili ad altre osservazioni. In altri termini, oltre ad un certo grado di dettaglio si inizia a descrivere il "rumore" e non più le regolarità statistiche del fenomeno studiato, generando una rapida degenerazione della capacità predittiva dei modelli. Questo fenomeno contraddice l'idea intuitiva che un modello accurato sia anche il migliore, poiché, per definizione, un modello è una rappresentazione parziale e approssimata di alcuni aspetti rilevanti di un fenomeno, non la sua descrizione completa (Giuliani and Zbilut, 2009) [29].

MAXENT affronta la questione imponendo un limite di errore ammissibile, ovvero una massima deviazione consentita rispetto alle medie delle *features* calcolate nei punti di presenza. Questo viene fatto, in primo luogo, attraverso un processo di *rescale* di tutte le *features* in modo che abbiano un intervallo 0-1. Successivamente il limite di errore ( $\lambda_j$ ) è calcolato per ogni *feature* come

$$\lambda_j = \lambda \sqrt{\frac{s^2(h_j)}{m}} \quad (2.11)$$

dove  $\lambda_j$  è il parametro di regolarizzazione della *feature*  $h_j$  e  $s^2(h_j)$  è la varianza di  $h_j$  sugli  $m$  punti di presenza. Concettualmente,  $\lambda_j$  corrisponde all'ampiezza dell'intervallo di confidenza, e in ogni caso, questo prende la forma dell'errore standard (l'espressione sotto radice quadra) moltiplicato per il parametro  $\lambda$  in base all'intervallo di confidenza desiderato [23]. Il lambda dell'equazione 2.11 opera la regolarizzazione applicando una sorta di *smoothing* alla distribuzione e rendendola più regolare. In altre parole, è possibile immaginare questa regolarizzazione come un sistema di contrazione dei coefficienti  $\beta$  che porta ad avere valori che riescano a bilanciare la relazione tra la complessità del fenomeno e la determinazione (il *fit*) del modello, garantendo contemporaneamente sia una buona accuratezza predittiva, sia la capacità di generalizzazione.

In MAXENT la determinazione del modello è misurata, nei punti di presenza, con la funzione di log-verosomiglianza (*log-likelihood*)<sup>7</sup>, per cui un modello ad elevata complessità registrerà il più elevato valore di log-verosomiglianza. Questa condizione, da sola, potrebbe risultare limitante o carente per quanto riguarda l'aspetto di generalizzabilità, per questo motivo l'obiettivo della regolarizzazione operata dall'algorithm è quello di "mediare" tra la determinazione del modello e la sua complessità. Per fare ciò MAXENT determina il modello secondo la distribuzione della funzione di massima verosomiglianza penalizzata [59]

$$\max_{\alpha, \beta} \frac{1}{m} \sum_{i=1}^m \ln(f(z_i) e^{\eta(z_i)}) - \sum_{j=1}^n \lambda_j |\beta_j| \quad (2.12)$$

per  $j = 1 \dots n$  *features*

e dove  $\int_L f(z) e^{\eta(z)} dz = 1$

con  $z$  che è il vettore delle *features* per i punti di presenza  $i$  nelle  $m$  località.

Il primo termine dell'equazione 2.12 è

<sup>7</sup>In statistica, la log-verosomiglianza (*log-likelihood*) descrive il logaritmo della probabilità di un risultato osservato, in un intervallo che va da 0 per la probabilità uguale ad 1 e tende a meno infinito per la probabilità uguale a 0. Se i risultati esistono in uno spazio continuo, allora ciò che verrà misurato sarà la densità della probabilità, invece della probabilità stessa.

Per cui, massimizzare la log-verosomiglianza penalizzata è equivalente a minimizzare l'entropia relativa soggetta ai vincoli dei limiti di errore ammesso.

### 2.3 Revisione bibliografica dell'applicazione dell'algoritmo di massima entropia al rischio incendi

---

Sono riportati i risultati di alcuni dei più recenti studi, basati su modelli che utilizzano l'algoritmo di massima entropia, effettuati per indagare la distribuzione potenziale degli incendi boschivi e il conseguente rischio. Questi studi sono stati utilizzati come punto di partenza, assieme agli altri lavori precedentemente citati, oltre che per definire e mettere a punto la metodologia applicata ai casi di studio, descritti nei successivi capitoli, anche per individuare le principali variabili indipendenti da considerare per la costruzione dei modelli.

Bar Massada *et al.* [5] hanno messo a confronto diversi approcci di modellizzazione per determinare quali variabili sono più fortemente correlate con l'innescò degli incendi evidenziando che il modello migliore, tra quelli usati, per determinare la distribuzione potenziale degli inneschi di incendi è MAXENT. La ricerca ha dimostrato che il miglior indicatore previsionale della distribuzione del fenomeno di innesco è la distanza dalla strada più vicina; gli altri tre principali fattori che influiscono su dove un incendio si potrebbe generare sono la distanza da case, la densità abitativa nella zona circostante, e la quota altimetrica.

Gli incendi boschivi sono un problema di gestione ricorrente anche nelle zone montuose occidentali dell'India. Sebbene la maggior parte degli incendi si verificano durante la stagione secca, le informazioni sulla distribuzione spaziale degli incendi è necessaria per migliorare la prevenzione di questi eventi [63]. Per questo studio, Ranesh *et al.* hanno utilizzato il database delle anomalie termiche (*Hotspots*) del satellite MODIS e l'algoritmo MAXENT per fornire una comprensione quantitativa delle variabili ambientali e per analizzare la distribuzione spaziale degli incendi boschivi nel periodo dal 2003 al 2007. In tutta l'area montuosa dell'India occidentale e in due sottoregioni con caratteristiche contrastanti è stato utilizzato il partizionamento gerarchico per valutare i contributi indipendenti di clima, topografia e vegetazione all'accuratezza dei modelli e per evidenziare la suscettibilità al fuoco più rappresentativa per ciascuna area di studio. I risultati hanno dimostrato che, sebbene le aree ipotizzate come altamente soggette agli incendi boschivi sono state localizzate principalmente sulle pendici orientali dei monti, le previsioni spaziali e l'accuratezza del modello differivano significativamente tra le aree di studio. Gli autori suggeriscono quindi un approccio in due fasi: una prima analisi con

### 2.3. Revisione bibliografica dell'applicazione dell'algoritmo di massima entropia al rischio incendi

---

particolare attenzione alle condizioni climatiche, ovvero alle condizioni della stagione dei monsoni prima della stagione secca, che determina il tasso di umidità del combustibile, per individuare le aree soggette a incendi agli incendi; una seconda analisi all'interno delle aree soggette a incendi utilizzando i modelli locali principalmente basati sul tipo di vegetazione.

Parisien e Moritz, in uno studio condotto in Nord America [53], hanno valutato quello che definiscono lo "Spazio ambientale biofisico" degli incendi da scala regionale fino a scala sub-continentale, con metodi ampiamente utilizzati per la modellazione di distribuzioni di specie (SDM). Questo approccio, spiegano, è particolarmente adatto per lo studio biogeografico degli incendi in quanto considera, contemporaneamente e "a macchia d'olio", più fattori che controllano l'idoneità o la predisposizione su grandi aree. Gli autori hanno usato l'algoritmo di massima entropia e potenziato gli algoritmi ad albero di regressione (*Classification and Regression Trees - CART*) per valutare l'interazione incendi-ambiente per tre livelli di complessità, in termini di inclusione delle variabili, e a tre scale spaziali: l'intera estensione degli Stati Uniti, lo stato della California, e cinque ecoregioni della California maggiormente soggette a incendi. Le risultanti dei modelli sono state proiettate geograficamente per ottenere previsioni spaziali e di idoneità ambientale degli incendi, inoltre, per valutare la trasferibilità spaziale il modello è stato applicato anche ad altre regioni. I modelli hanno identificato una serie di variabili chiave che non erano imputate importanti, su larga scala, per il fenomeno degli incendi, ma che potrebbero indirettamente influire sulla presenza di vegetazione infiammabile. Lo studio dimostra che i modelli di distribuzione di incendio e degli habitat sono concetti correlati e possono essere utilizzati per caratterizzare i fattori ambientali in un processo disturbo naturale.

Lo scopo di un altro studio condotto da Parisien *et al.* è stato quello di utilizzare le osservazioni degli incendi verificati durante 25 anni e una vasta gamma di variabili ambientali, in un quadro statistico, per produrre stime di probabilità ad alta risoluzione spaziale [54]. Utilizzando MAXENT per la modellizzazione sono stati messi in relazione i punti di innesco degli incendi, campionati nelle zone bruciate durante il periodo 1984-2008, con le variabili maggiormente rappresentative del fenomeno come la vegetazione infiammabile, il clima e la topografia. Per valutare l'effetto antropico sulla distribuzione spaziale di rischio incendio, gli autori hanno costruito un modello alternativo che escludeva tutte le variabili che hanno una forte impronta antropica, dimostrando che la distribuzione di probabilità nella parte occidentale degli USA è ben lungi dall'essere uniforme; l'effetto di fattori antropici sulla probabilità di incendi varia da regione a regione ma, nel complesso, la presenza di attività umane sembra inibire l'attività del fuoco nell'area indagata. I risultati, non solo forniscono quelle che sembrano essere robuste previsioni di rischio incendi,

ma migliorano la comprensione a lungo termine dei fattori che influenzano gli incendi a supporto della pianificazione strategica del territorio.

Utilizzando dati climatici e degli incendi raccolti in 25 anni nel bacino del fiume Colorado, negli Stati Uniti occidentali, Arnold *et al.*, hanno studiato le condizioni climatiche e ambientali antecedenti agli eventi di incendio [2]. Nell'area indagata, l'algoritmo di massima entropia è stato applicato per stimare l'estensione territoriale alla quale la robustezza della relazione tra incendi e particolari condizioni predisponenti si verifica, determinando una quantificazione del rischio. Dai risultati emerge la chiara distinzione spaziale tra i regimi di incendio limitati dalle condizioni climatiche, dove la componente combustibile è abbondante, e i regimi in cui è la tipologia di combustibile a determinare le limitazioni di diffusione, nonostante le condizioni climatiche siano spesso ottimali per gli incendi.

Nel lavoro condotto da Duane *et al.*, di recente pubblicazione [19], l'algoritmo di massima entropia è stato utilizzato per determinare il contributo relativo dei fattori predisponenti per la determinazione della probabilità di incendio in Catalonia, nel Nord Est della Spagna. Introducendo una nuova classificazione basata sui modelli di diffusione dominante del fuoco, al fine di esaminare le variabilità spatio-temporale dei regimi di incendio, lo studio ha evidenziato una grande accuratezza nella determinazione della distribuzione spaziale dei diversi modelli di diffusione dominante del fuoco, validando i risultati su scala temporale. Inoltre, è stato evidenziato come i fattori legati alla tipologia di combustibile, in ambiente forestale, siano determinanti per gli incendi a prevalente diffusione convettiva, mentre, le condizioni topografiche giocano un ruolo fondamentale per gli incendi la cui diffusione è fortemente influenzata dal vento.

---

## Applicazione dell'algoritmo di massima entropia ai casi di studio

---

Al fine di valutare le prestazioni in termini di accuratezza e la capacità predittiva dei modelli generati con MAXENT nello studio delle distribuzioni di rischio per gli incendi forestali, nei successivi capitoli sono riportate le esperienze di applicazione dell'algoritmo in tre casi di studio in cui sono state testate diverse combinazioni di covariate applicate ad ambienti e scale geografiche differenti.

Il primo caso di studio riguarda la regione il bacino idrografico afferente al Rio delle Amazzoni nell'area amministrativa di pertinenza boliviana (*Cuenca Amazzonica* in Bolivia). In questo caso la procedura metodologica è stata costruita per un area geografica decisamente vasta (circa 725000 km<sup>2</sup> de la Cuenca Amazonica de Bolivia) con un diversità ambientale e fisiografica estremamente elevata. Si ricorda, tra l'altro, che l'area fa parte del principale ambito di biodiversità del pianeta<sup>1</sup>.

---

<sup>1</sup>“Bolivia tiene una superficie de 1.098.581 km<sup>2</sup> con una división político administrativa que comprende 9 departamentos y 327 municipios. En el país están presentes 3 cuencas hidrográficas: Cuenca del Amazonas, Cuenca del Plata, Cuenca del Altiplano. De las tres, la cuenca del Amazonas es la más importante por su extensión. La cuenca amazónica de Bolivia abarca el 66% territorio nacional e involucra a 6 departamentos del país, 236 municipios y a 32 pueblos indígenas de los 36 existentes en Bolivia. Los departamentos que forman parte de la cuenca amazónica en Bolivia son: Pando, La Paz, Beni, Oruro, Cochabamba, Santa Cruz, Potosí y Chuquisaca. La divisoria de la Cuenca Amazonas se encuentra delimitada por la línea topográfica de los accidentes orográficos: de oeste a este por las cordilleras de Apolobamba, La Paz, Tres Cruces, Los Frailes (cordillera de Levichuco); sigue por los cerros Wila Kollu, Saketa, Kellu, Canatari, Ravelo, Tarabuco, Cerro Jatun, Matancería, Azurduy; para luego continuar por la planicie de la provincia Cordillera del departamento de Santa Cruz, hasta encontrar las nacientes del río Iténez, en la serranía de Aguapehí, en la República Federal del Brasil. La diferencia de relieve está determinada por las siguientes costas: altitud 6.421 m s.n.m., correspondiente al punto más alto (nevado Illampu) y 90 m s.n.m., referido al punto más bajo (población de Mnoa). La cuenca amazónica boliviana presenta dieciséis ecorregiones.” (Rodríguez, J.P.,

### **Capitolo 3. Applicazione dell'algoritmo di massima entropia ai casi di studio**

---

E', di conseguenza, estremamente elevata la variabilità dei predittori sia della componente naturale (fisica e biotica) sia antropica (da aree densamente urbanizzate a zone praticamente prive di insediamenti umani permanenti). Dal punto di vista biogeografico, infatti, è presente una vasta gamma di ecosistemi compresi, secondo gradienti di umidità e, subordinatamente, di temperatura, fra gli ambienti tropicali caldi ed umidi della foresta amazzonica (Dipartimenti di Pando e Beni) e condizioni climatiche caratterizzate da prolungata stagione secca e pronunciata stagionalità termica. In quest'ultimo caso si fa riferimento alla vegetazione xerica di savana arborea e arbustiva (savannah-like ecosystems) nell'area periferica meridionale del Bosque Chiquitano (Dipartimento Chiquitano) di Santa Cruz de la Sierra)

Nel secondo caso di studio riportato, l'indagine è stata ristretta ad aree dell'Amazonia boliviana con estensione geografica progressivamente più limitata, prima a livello di dipartimento e successivamente a livello di territori comunali (singole o per piccoli gruppi) al fine di limitare la variabilità ambientale. Sono stati utilizzati gli stessi parametri e le medesime covariate del primo caso di studio, senza variare la risoluzione spaziale, al fine di valutare le risposte dei predittori utilizzati e le *performance* dei modelli in situazioni più omogenee e uniformi rispetto all'intera *Cuenca* amazzonica boliviana.

Il terzo caso di studio, ha interessato la modellizzazione della distribuzione del rischio di incendio in ambiente alpino. L'algoritmo di massima entropia è stato applicato per indagare il fenomeno degli incendi nella Regione Autonoma Valle d'Aosta (3.263 km<sup>2</sup>) in funzione delle stagioni e in relazione ai fattori antropici legati alle attività zootecniche. In questo caso è stata prestata particolare attenzione all'approccio statistico di selezione delle covariate, con il preciso obiettivo di limitare la colinearità tra esse.

#### **3.1 Metodologie comuni ai casi di studio**

---

Alla luce dell'obiettivo generale della ricerca ed in relazione agli obiettivi specifici volti a testare la metodologia a diversi livelli di scala, alcuni aspetti e procedure sono state comuni per tutti i casi di studio, prevalentemente in riferimento alle procedure di valutazione della accuratezza dei modelli, delle covariate e del sistema di ricampionamento dei punti di presenza per l'estrazione dei dataset di *train* e *test* nelle reiterazioni di calcolo.

M.A. Oliveira- Miranda, R. Lazo, S. Zambrano-Martínez, E. Tapiquén, A. Ruiz, E.E.Gutiérrez, M. Armas, L. Gerik, L.A. Solórzano y F. Rojas-Suárez. 2005. Geografía de la Conservación de los Andes Tropicales: Bolivia, Colombia, Ecuador, Perú y Venezuela.)

### 3.1.1 Valutazione della accuratezza dei modelli

Il potere diagnostico di un modello è di per sé un concetto multidimensionale, in quanto include la sensibilità, la specificità<sup>2</sup>, il potere predittivo positivo, il potere predittivo negativo e l'accuratezza. Sottoponendo il modello ad un campione misto di veri positivi (TP) e veri negativi (TN), in questo caso rispettivamente le presenze e le assenze, il modello risponderà restituendo:

- una frazione di veri positivi (*true positive rate* - TPR) o potere predittivo positivo;
- una frazione di falsi positivi (*false positive rate* - FPR);
- una frazione di veri negativi (*true negative rate* - TNR) o potere predittivo negativo;
- una frazione di falsi negativi (*false negative rate* - FNR).

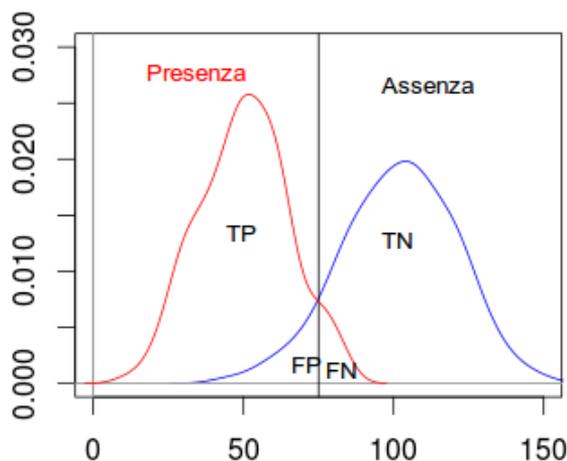
La presenza di falsi positivi (FP) e falsi negativi (FN) è dovuta alla natura inferenziale del modello, nel quale la distribuzione di frequenza delle presenze e quella delle assenze, possono non essere nettamente distinte, sovrapponendosi come mostrato in figura 3.1

Nei test di tipo dicotomico (sì/no, assenza/presenza) l'uso della coppia di valori di sensibilità e specificità, rispetto ad un determinato valore di soglia, è in genere sufficiente a descrivere l'accuratezza del modello. Nei test in cui, invece, è possibile operare un *rating*, come per i modelli di distribuzione potenziale, è necessario utilizzare coppie di valori di sensibilità e specificità per tutti i possibili valori di soglia del modello. La rappresentazione grafica di queste coppie di valori ( $SE, 1 - SP$ ), prendono il nome di curva ROC (*Receiver Operating Characteristic*)<sup>3</sup>.

La curva ROC è una raffigurazione bidimensionale della capacità di classificazione ottenibile da un modello predittivo [24] della presenza di un fenomeno, che però male si presta alle indagini comparative di modelli diversi, per questo motivo la valutazione delle *performance*, nei casi di studio riportati, è stata misurata attraverso il valore di AUC (*Area Under the Curve*), ovvero l'area sottesa alla curva ROC. Questo è un metodo che è largamente impiegato [25] e che ha il pregio di permettere confronti che non sono influenzati

<sup>2</sup>La sensibilità (SE) di un modello è definita dal *true positive rate* ed è dato da  $SE = TPR = TP/(TP + FN)$ , mentre la specificità (SP) è data da  $SP = 1 - FPR$ , dove  $FPR = FP/(FP + TN)$  quindi  $SP = TN/(FP + TN)$

<sup>3</sup>“Each threshold value produces a different point in ROC space. Conceptually, we may imagine varying a threshold from  $-\infty$  to  $+\infty$  and tracing a curve through ROC space. Computationally, this is a poor way of generating an ROC curve, and the next section describes a more efficient and careful method. [...] A threshold of  $+\infty$  produces the point (0, 0). As we lower the threshold to 0.9 the first positive instance is classified positive, yielding (0, 0.1). As the threshold is further reduced, the curve climbs up and to the right, ending up at (1, 1) with a threshold of 0.1. Note that lowering this threshold corresponds to moving from the "conservative" to the "liberal" areas of the graph.”(Fawcett, 2006 - p. 863)



**Figura 3.1:** Esempio di distribuzione di frequenze per i risultati di assenza e di presenza di un generico modello

né dalle dimensioni del campione, né dal numero di covariate coinvolte per la costruzione del modello<sup>4</sup>. In particolare, il valore di AUC, che può variare da 0 a 1, indica la qualità del modello stesso. Valori pari a 0.5 indicano un comportamento assolutamente casuale. Valori superiori a 0.6 indicano un modello che nel 40 per cento dei casi effettua previsioni errate, e così via.

L'analisi di questi valori, e la scelta di soglie efficienti, pur ponendo numerose difficoltà [44], permette di considerare efficienti i modelli il cui valore di AUC supera 0,8 [46].

#### 3.1.2 Valutazione dell'importanza delle variabili

L'esame dei risultati del processo di creazione del modello ha consentito di ordinare le covariate utilizzate in base alla loro importanza relativa percentuale nel determinare la probabilità di innesco per gli incendi. Inoltre, le covariate prese in considerazione hanno subito un ricampionamento secondo il metodo *jackknife* [51, 62] (*Jackknife Repeated Replication - JRR*), che consiste nel ricalcolare le stime escludendo dal calcolo una osservazione alla volta

<sup>4</sup>“The AUC has an important statistical property: the AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. This is equivalent to the Wilcoxon test of ranks.” (Fawcett, 2006 - p. 868)

dal campione. In questo caso è stato utilizzato per misurare l'apporto delle singole variabili al modello: escludendo di volta in volta una covariata e ricalcolando il modello si evidenzia quanto l'assenza di questa incida sull'attendibilità del modello e sulla sua capacità predittiva, quindi, per estensione sul comportamento del fenomeno indagato.

L'idea generale su cui si basa il metodo JRR è quella di formare diversi sottocampioni di un generico campione estratto eliminando di volta in volta una singola osservazione o, nel caso di disegni complessi e che prevedono un'elevata numerosità, una piccola parte del campione originario. Ciascun sottocampione misurerà così il contributo alla variabilità totale fornita da tale piccola parte eliminata dalla stima: la procedura JRR calcola infatti le stime dei parametri di interesse su ciascuna replica ed utilizza la variabilità tra tali valori per ottenere una stima della variabilità campionaria totale.

Le variabili indipendenti considerate sono state valutate anche sulla base di due parametri calcolati sul modello e di seguito descritte:

- i) il contributo percentuale, che è espressione della variazione marginale di una singola covariata. Questo prevede che venga calcolato un modello per ogni singola covariata, considerando le altre *features* bloccate al valore medio registrato nei punti di presenza;
- ii) l'importanza di permutazione, in cui il contributo di ciascuna covariata è determinato dalla permutazione casuale dei valori di questa tra i punti di presenza del fenomeno e quelli di *background* dell'area indagata, misurando la conseguente variazione (solitamente una diminuzione) nei valori di AUC. Una diminuzione elevata (espressa in termini di perdita percentuale) del valore di AUC indica che il modello dipende fortemente dalle informazioni contenute in quella covariata, o, in altri termini, che le informazioni apportate da quella covariata non sono ricalcolabili a partire dalle altre covariate.

#### 3.1.3 Ricampionamento dei punti di presenza

I punti di presenza sono stati ricampionati secondo la tecnica *bootstrap* [20] per estrarre i dataset di *train* e *test*. Il *bootstrap* è un metodo generale per ottenere informazioni circa la variabilità e la distribuzione di statistiche campionarie rispetto ad un parametro, e quindi da esse, stimare i limiti di confidenza dello stesso parametro in quella popolazione, quando non si possiedono informazioni sulla sua distribuzione. Questo metodo ha il grande vantaggio di fornire le stime statistiche necessarie all'inferenza anche per funzioni molto complesse, come il calcolo dei momenti, dei coefficienti di variazione, dei rapporti tra valori medi e fra varianze, dei coefficienti di correlazione, degli autovalori delle matrici di varianze e covarianze. Usi frequenti nella ricerca

### **Capitolo 3. Applicazione dell'algoritmo di massima entropia ai casi di studio**

---

ambientale riguardano l'inferenza su indici di similarità o distanza calcolati mediante la presenza-assenza di specie o l'intensità di alcuni parametri fisici e chimici. L'idea di base è, per un campione formato da  $k$  osservazioni, simulare il processo di selezione di molti campioni di dimensione  $k$ , allo scopo di determinare la probabilità che i valori cadano all'interno di intervalli predefiniti. Il campione *bootstrap* è nient'altro che il campione originario nel quale, per effetto dell'estrazione con ripetizione, alcuni dati sono ripetuti ed altri, per mantenere lo stesso numero di osservazioni, sono assenti. E' proprio la modalità di estrazione, fondata sulla ripetizione, a generare la variabilità nelle stime, questo in conseguenza al fatto che il metodo richiede che i campioni abbiano tutti lo stesso numero di osservazioni e se si estraesse senza ripetizione sarebbero tutti identici. La metodologia è fondata sulla generazione casuale di un numero molto grande di copie (di norma fino a un milione), del campione di  $k$  osservazioni. Tutte le copie vengono casualmente mescolate, e grazie alla generazione di numeri casuali, vengono estratti milioni di stringhe della stessa dimensione del campione originale ( $k$ ). Ognuna di queste stringhe può contenere due o più valori identici, con l'ovvia esclusione di altri valori che sono contenuti nel campione originale. Nei casi di studio riportati *bootstrap* è stato usato per costruire diversi campioni, tutti ugualmente rappresentativi del campione originale dei punti di presenza disponibili, per reiterare il modello con combinazioni differenti di dataset di *train* e di *test*.

---

## Elaborazione su scala vasta: Amazzonia Boliviana

---

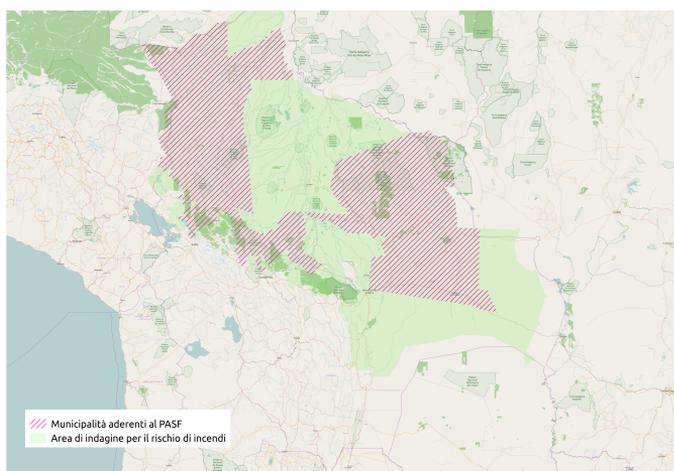
In questo caso di studio è stata applicata la metodologia basata sull'algoritmo di massima entropia per la determinazione della distribuzione del rischio di incendio all'interno della regione amazzonica della Bolivia. In quest'area, che si estende per una superficie di oltre 720000 km<sup>2</sup>, il maggior numero degli incendi forestali è riconducibile, o comunque indirettamente correlato, alle pratiche agronomiche che prevedono l'utilizzo del fuoco (debbio), in condizioni teoricamente controllate, per liberare una superficie dalla vegetazione naturale al fine di rendere possibile la coltivazione o per rinnovare il cotico erboso e renderlo disponibile per il pascolo del bestiame. In molti casi queste pratiche tradizionali sfuggono al controllo generando incendi, spesso di grandi dimensioni.

### 4.1 Inquadramento del progetto

---

Le attività che vengono di seguito descritte sono state sviluppate all'interno del Programma Amazonía Sin Fuego (PASF), eseguito in Bolivia dall'Autorità Plurinazionale della Madre Terra, organo autarchico dello Stato boliviano dipendente dal Ministero dell'Ambiente e Acqua, grazie al finanziamento del Ministero degli Affari Esteri italiano e della Cooperazione Italiana allo Sviluppo, ed ai contributi del Governo Brasiliano e della Banca di Sviluppo dell'America Latina (CAF).

L'obiettivo generale del Programma è quello di ridurre l'incidenza degli incendi nella regione amazzonica della Bolivia, attraverso la promozione e



**Figura 4.1:** *Inquadramento del Programma PASF e area di indagine per il rischio di incendi forestali.*

diffusione di pratiche alternative all'uso del fuoco nelle attività agricole, contribuendo a proteggere l'ambiente ed a migliorare le condizioni di vita delle comunità indigene e rurali. All'interno di questo contesto è sorta parallelamente la necessità di individuare uno strumento per incrementare la capacità operativa nelle attività di lotta agli incendi su una superficie tanto vasta, quanto poco densamente abitata. Vista la grande disponibilità di dati satellitari e la frequenza di aggiornamento con cui questi dati sono restituiti l'obiettivo specifico è stato quello di individuare una metodologia per la discriminazione degli incendi forestali rilevati a seguito di una lettura critica dei dati dalle registrazioni di anomalie termiche rilevati da sensori satellitari (*hotspots*).

La metodologia sviluppata ha previsto l'utilizzo di MAXENT per la creazione di carte della distribuzione del rischio di incendio, sulla base dei dati storici degli incendi forestali, per creare uno strumento di valutazione degli *hotspot*. Misurando il valore di rischio in corrispondenza dell'anomalia termica registrata da satellite è possibile stimare con quale probabilità essa sarà stata generata da un incendio forestale o da un'altra fonte di calore. Con l'applicazione dell'algoritmo di massima entropia, in questo caso di studio, si è voluto indagare la capacità della metodologia di individuare i fattori che più contribuiscono a caratterizzare i fenomeni di incendio boschivo in un territorio tanto esteso quanto variabile dal punto di vista ambientale, testando un metodo di selezione delle variabili basato sulla capacità classificatrice dell'algoritmo.

### 4.2 Premesse metodologiche

---

I modelli interpretativi di dati satellitari disponibili (ad esempio MODIS o IN-PE) riportano gli *hotspots* quali alterazioni termiche della superficie terrestre (pixel “caldi”), una volta filtrate le discrepanze più evidenti come i falsi positivi dovuti alla riflettanza dei corpi idrici in particolari situazioni stagionali e di incidenza dei raggi solari, è necessario determinare il tipo di evento di fuoco. In questo senso si può fare riferimento, in via preliminare, a due grandi classi:

- a) incendio libero di coperture vegetali (incendio d’ora in avanti);
- b) fuoco controllato condotto con obiettivi agronomici (localmente dette *chaqueo* o *quema*).

Gli algoritmi progettati per le discriminazioni tra *hotspots* e incendi si basano principalmente su criteri legati alle proprietà di risoluzione delle immagini, attraverso l’analisi delle bande spettrali o basati sulle caratteristiche geofisiche di acquisizione delle immagini e sul tempo di ritorno del satellite sullo stesso punto della terra. Ciò è riportato in letteratura [10, 28, 42, 47] sia a livello internazionale sia in riferimento all’esperienza specifica della Bolivia.

Gli algoritmi di discriminazione degli *hotspots* possono infatti utilizzare:

- i) dati biometeorologici derivati da satelliti geostazionari di seconda generazione (ad esempio Meteosat II della Agenzia Spaziale Europea);
- ii) Dati Lidar che restituiscano il modello digitale della superficie e della struttura della vegetazione volti a generare modelli realistici di combustibile;
- iii) immagini satellitari ad alta risoluzione con tempi di rivisitazione entro le 24 ore.

L’utilizzo di questi dati, in aree del mondo in cui la frequenza di fuoco controllato come pratica agronomica è trascurabile, portano ad una probabilità massima di coincidenza tra *hotspots* e incendio pari al 90%.

In Bolivia, tali dati non sono attualmente disponibili e i fuochi controllati (localmente *quemaz* o *chaqueo*) legati alle pratiche agricole sono molto diffusi. I dati reperibili e le metodologie menzionate (i,ii,iii) consentono una stima di affidabilità dei dati *hotspots* come indicatori diagnostici degli incendi forestali non supera il 50%. È quindi necessario (ed è tema dibattuto e richiesto a livello nazionale) operare con metodologie differenti che meglio interpretino i dati territoriali e ambientali effettivamente disponibili per l’area vasta del bacino amazzonico boliviano. Si tratta, di conseguenza, di un ambito estremamente rilevante per testare un approccio metodologico che prevedeva la valutazione delle covariate da includere nel modello e l’applicazione di MAXENT per la

determinazione della distribuzione di probabilità di incendio forestale. Tale approccio si è fondato sui dati di eventi di fuoco registrati nella banca dati governativa di ABT (*Autoridad de Fiscalización y Control Social de Bosques y Tierra*).

### 4.3 Scelta delle variabili

---

L'elaborazione dei modelli calcolati per la *Cuenca* amazzonica boliviana è stata condotta attraverso una prima fase esplorativa volta alla discriminazione delle variabili indipendenti da utilizzare. L'elenco completo delle variabili considerate è riportato nella tabella 4.1.

Prendendo in considerazione quanto suggerito in letteratura [12, 54], sono state inizialmente incluse tutte le covariate disponibili dal dataset *Worldclim*<sup>1</sup>, il rilievo terrestre ricavato dal modello digitale di elevazione (DEM) contenuto nel dataset GMTED2010 (*Global Multi-resolution Terrain Elevation Data 2010*) distribuito da USGS<sup>2</sup>, che ha fornito anche il *Normalized Difference Vegetation Index* (NDVI) elaborato a partire dai dati registrati dal satellite MODIS (*Moderate Resolution Imaging Spectroradiometer*).

La componente antropica, secondo Hawbaker *et al.*, può essere efficacemente rappresentata dalle informazioni geografiche relative alla densità abitativa e dalla distanza dalle infrastrutture stradali [30]. In questo caso di studio, vista la scarsa disponibilità di dati censuari ad adeguata scala di dettaglio riguardanti la densità abitativa,<sup>3</sup> è stata utilizzata solo la variabile di distanza dalle infrastrutture delle celle raster, calcolata sulla distanza euclidea dagli elementi stradali alla risoluzione di 250 metri. I dati sono stati ricavati dall'archivio geografico nazionale boliviano ed integrati con i dati delle infrastrutture forniti dalle più di 40 municipalità interessate dallo studio.

La componente di copertura del suolo è stata inclusa attraverso la rasterizzazione dello strato informativo nazionale denominato COBUSO (acronimo di *COBertura y USO de la tierra*), aggiornata all'anno 2010, fornita da ABT.

Per integrare le informazioni sulla tipologia e sulla struttura delle vegetazione sono stati considerati, inizialmente, sia la carta delle ecoregioni elabo-

---

<sup>1</sup>In formato raster, alla risoluzione di 30 arco-secondi (corrispondente a circa un chilometro), riferite ai dati climatici legati a temperatura e precipitazioni, riportati con la nomenclatura "BIO" seguita dal numero della covariata, come descritto in tabella 4.1; per ulteriori dettagli si rimanda al sito web worldclim e per la metodologia di elaborazione dei dati climatici a [34]

<sup>2</sup>U.S. Geological Survey e National Geospatial-Intelligence Agency (NGA) hanno sviluppato il modello globale di elevazione GMTED2010 aumentando la precisione rispetto ai precedenti prodotti come il GTOPO30. Sono state incluse nuove fonti di dati come il Digital Terrain Elevation dati (DTED®) dalla Shuttle Radar Topography Mission (SRTM), dati Spot 5 Reference 3D, e dati dal Land Elevation Satellite (ICESat).

<sup>3</sup>I dati censuari presenti in Bolivia sono registrati in forma aggregata per municipalità e, di conseguenza poco informativi ai fini modellistici; inoltre, dal confronto con le autorità locali è emerso che nella maggior parte dei casi le piccole comunità, a dimensione poco più che familiare e molto frequenti nell'area amazzonica, non sono registrate.

rata da Ibisch [35], che la carta della vegetazione di Navarro [52]. Quest'ultima nasce come carta della vegetazione potenziale della Bolivia e classifica le strutture vegetali in oltre 1900 tipi. Per l'utilizzo ai fini della modellistica di rischio incendi è stato necessario riclassificare le tipologie originali sulla base delle componenti strutturali principali fino a ridurre la classificazione a 177 tipi. L'elevato dettaglio dalla carta originale di Navarro, su una scala geografica così elevata, se da un lato fornirebbe una maggiore quantità di informazioni, aumentando presumibilmente la predittività dei modelli per questa covariata, dall'altro renderebbe più difficilmente interpretabile la relazione tra tipo (sia specifico che strutturale) di vegetazione e il fenomeno indagato. Nella fase esplorativa, di selezione delle covariate da utilizzare nei modelli, è stata posta particolare attenzione a comparare il possibile uso alternativo di questi ultimi due strati informativi: se risulta infatti chiaro che le due variabili siano espressione di informazioni comuni, derivate principalmente dalle caratteristiche dalla vegetazione, non è da escludere *a priori* che le due differenti scale di dettaglio (elevata per la vegetazione di Navarro e grossolana per la carta delle ecoregioni) influiscano diversamente sul rischio di incendio forestale.

Infine, come ulteriore fonte di informazione, è stata inclusa nei modelli la carta dei suoli secondo la classificazione FAO, in quanto considerata espressione della variabilità topografica e delle pendenze a scala di paesaggio [31].

#### 4.4 Preparazione dei dati in input

---

I punti di presenza degli incendi, ricavati dalla banca dati governativa di ABT sono stati registrati nel periodo di 7 anni (dal 2005 al 2011), per un totale di oltre 1800 eventi confermati,<sup>4</sup> forniti in formato vettoriale nel sistema di riferimento WGS84, in coordinate geografiche non proiettate (latitudine e longitudine).

Le variabili indipendenti sono state ricampionate per omogeneizzare le risoluzioni spaziali dei dati provenienti da fonti diverse, inoltre quando necessario, è stata applicata la conversione del sistema di riferimento delle coordinate, in modo da farlo coincidere con il sistema di riferimento dei punti di incendio. La risoluzione dei raster è stata impostata a 7.5 secondi di arco ( 0.002075 gradi decimali) che corrispondono, a quella latitudine, a circa 250 metri.

---

<sup>4</sup>Il sistema di registrazione degli incendi, adottato da ABT, include soltanto gli eventi sui quali è stato necessario l'intervento del personale addetto alla lotta attiva (squadre antincendio dell'esercito), e non considera tutti gli altri incendi, per questo motivo il numero è così esiguo, se paragonato all'estensione del territorio investigato

## Capitolo 4. Elaborazione su scala vasta: Amazzonia Boliviana

---

**Tabella 4.1:** *Elenco delle variabili considerate nella fase esplorativa per il modello di distribuzione dei rischi di incendio nell'Amazzonia boliviana*

Variabile	Descrizione
BIO1	Temperatura media annuale
BIO2	Range medio delle temperature giornaliere
BIO3	Isotermalità (BIO2/BIO7) (* 100)
BIO4	Stagionalità delle temperature (deviazione standard *100)
BIO5	Massima temperatura nel mese più caldo
BIO6	Minima temperatura nel mese più freddo
BIO7	Range annuale delle temperature (BIO5-BIO6)
BIO8	Temperatura media del quarto più piovoso
BIO9	Temperatura media del quarto più siccitoso
BIO10	Temperatura media del quarto più caldo
BIO11	Temperatura media del quarto più freddo
BIO12	Precipitazioni annuali
BIO13	Precipitazioni nel mese più piovoso
BIO14	Precipitazioni nel mese più siccitoso
BIO15	Stagionalità delle precipitazioni (coefficiente di variabilità)
BIO16	Precipitazioni del quarto più piovoso
BIO17	Precipitazioni del quarto più siccitoso
BIO18	Precipitazioni del quarto più caldo
BIO19	Precipitazioni del quarto più freddo
DEM	Modello digitale del terreno da dataset GMTED2010
DIST INFR	Raster di distanza dalle strutture e infrastrutture antropiche
ECOREG	Ecoregioni di Ibisch
NDVI	MODIS Normalized Difference Vegetation Index (NDVI)
SOIL	Carta dei suoli FAO
USO.TIERRA	Uso e copertura della terra (ABT)
VEGET.NAVARRO	Carta della vegetazione di Navarro (generalizzata a 177 categorie)

### 4.5 Elaborazione

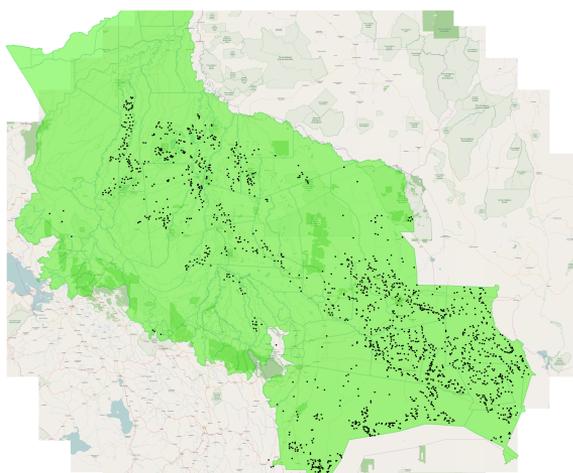
---

L'elaborazione dei modelli per la valutazione della distribuzione del rischio di incendi è stata strutturata in due fasi: un primo calcolo del modello, per individuare le covariate maggiormente informative; una seconda elaborazione, con le covariate selezionate per il calcolo del modello per valutare la distribuzione del rischio e l'accuratezza.

#### 4.5.1 Selezione delle variabili

Il primo modello calcolato, come precedentemente spiegato, ha incluso tutte le variabili descritte per avere la possibilità di applicare la procedura di ricampionamento secondo il metodo *jackknife* e valutare il guadagno, in termini di *performance*, del modello alla sottrazione di una covariata alla volta. Per il processo di selezione sono stati valutati, inoltre, i valori di importanza di permutazione e di contributo relativo (espressi in forma percentuale) all'informatività del modello di distribuzione del rischio.

I punti di presenza degli incendi sono stati suddivisi in maniera casuale per creare due dataset distinti e generare un campione per il calcolo del modello,



**Figura 4.2:** I punti in nero rappresentano la posizione degli incendi confermati da ABT per la Cuenca amazonica boliviana.

composto dal 70% dei punti originali, e un campione di verifica (*test*), con il restante 30% dei punti.

I parametri utilizzati per il modello esplorativo sono riportati in tabella 4.2.

**Tabella 4.2:** Parametri del modello esplorativo utilizzato per la selezione delle variabili da utilizzare per la valutazione del rischio incendio in Amazzonia boliviana.

Parametro	Valore / modalità
Punti training	1270
Punti test	544
Totale dei punti utilizzati per determinare la distribuzione	11270 (punti di training + punti di background)
Valore di regolarizzazione feature lineare	0.05
Valore di regolarizzazione feature quadratica	0.05
Valore di regolarizzazione feature prodotto	0.05
Valore di regolarizzazione feature categorica	0.25
Valore di regolarizzazione feature soglia	1
Valore di regolarizzazione feature hinge	0.5
Combinazione di feature usata	prodotto - lineare - quadratica - hinge - soglia

Dall'analisi del ricampionamento col metodo JRR è emersa chiaramente la ridondanza di informazioni fornite dalle variabili. Quest'ultima considerazione nasce dal fatto che non risultano evidenti differenze tra i valori di AUC dei modelli alla sottrazione di una singola variabile (in turchese nella figura 4.3); contemporaneamente, si può notare che le variabili che, considerate singolarmente, generano modelli con maggiore accuratezza in termini di AUC, sono la carta dei suoli con il valore di 0.754, la carta di uso e copertura del

## Capitolo 4. Elaborazione su scala vasta: Amazzonia Boliviana

**Tabella 4.3:** Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili del modello esplorativo.

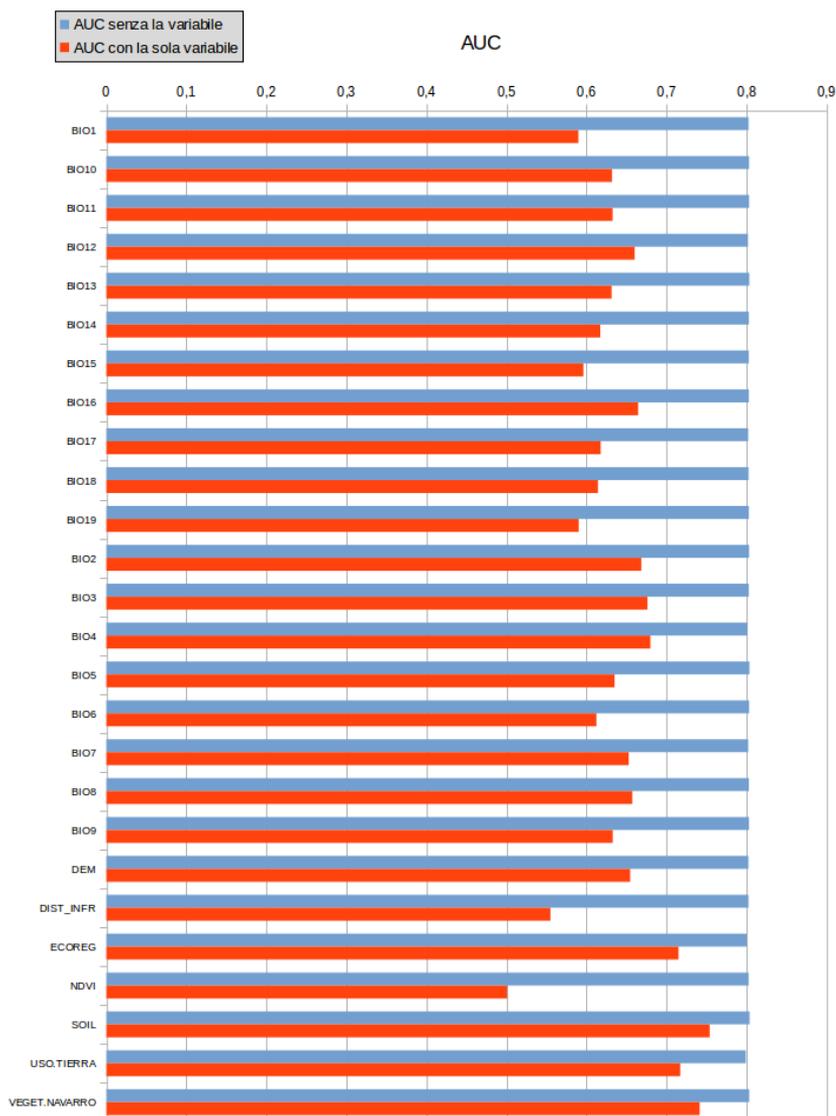
Variabile	CP (%)	IPP (%)	Variabile	CP (%)	IPP (%)
BIO1	0,55	1,72	BIO14	0,055	0,16
<b>BIO2</b>	25,85	11,48	BIO15	0,09	0,34
BIO3	0,099	0,19	<b>BIO16</b>	2,75	6,51
<b>BIO4</b>	18,33	38,19	<b>BIO17</b>	1,45	1,71
<b>BIO5</b>	2,47	3,73	BIO18	0,098	0,029
BIO6	0,46	0,87	BIO19	0,68	1,93
<b>BIO7</b>	1,42	1,94	<b>DEM</b>	6,22	0,31
BIO8	0,095	0,25	<b>DIST INFR</b>	2,92	1,38
BIO9	0,11	0,09	<b>ECOREG</b>	6,31	0,33
BIO10	0,81	0,03	NDVI	0,17	0,16
<b>BIO11</b>	1,41	1,21	<b>SOIL</b>	7,45	8,35
<b>BIO12</b>	2,52	5,95	<b>USO.TIERRA</b>	11,074	6,91
<b>BIO13</b>	1,57	2,03	<b>VEGET.NAVARRO</b>	5,01	4,18

suolo (valore di 0.717) e le due variabili considerate per le caratteristiche della vegetazione: la carta delle ecoregioni con  $AUC = 0.714$  e Carta della vegetazione potenziale di Navarro con  $AUC = 0.741$  (graficamente riportati in rosso, sempre nella figura 4.3).

Dai valori registrati per gli altri due parametri utilizzati per la valutazione dell'informatività delle variabili, riportati in tabella 4.3, è stato determinato il valore di soglia per la scelta delle variabili, eliminando quelle con valore di contributo percentuale (CP) inferiore a 1.

Prima di elaborare il modello con le variabili selezionate, sono state valutate singolarmente le due alternative di variabili legate alla tipologia di vegetazione, ricalcolando i modelli esplorativi includendo o l'una o l'altra fonte di informazione. Questo, per limitare la ridondanza di informazione che rischierebbe di portare ad una attribuzione eccessiva del "peso" della componente vegetale del modello, e per rendere più comprensibili le relazioni tra covariate e il fenomeno degli incendi forestali. Dalla valutazione delle *performance* delle elaborazioni fatte con i due differenti strati informativi per la tipologia e la struttura della vegetazione, non sono emerse differenze sostanziali in termini di AUC (in entrambi i modelli il valore registrato è stato prossimo a 0,81), inoltre, anche le covariate, con valori di contributo percentuale e di importanza di permutazione, superiori alla soglia dell'1% sono risultate praticamente le stesse per entrambi i casi, come mostrato in tabella 4.4.

Questi risultati confermano che a questa scala geografica, le due variabili - ecoregioni e vegetazione potenziale - sono equivalenti in termini di quantità e qualità di informazioni che riescono a conferire al modello; alla luce di questa considerazione è stata scelta la combinazione di variabili che ha incluso la carta della vegetazione di Navarro, riclassificata a 177 tipologie, poiché,



**Figura 4.3:** Rappresentazione grafica di JRR applicato al modello esplorativo per la selezione delle variabili in Amazonia boliviana: in rosso sono riportati i valori di AUC registrati per modelli che considerano una sola variabile alla volta (in ordinata), in turchese sono riportati i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata.

## Capitolo 4. Elaborazione su scala vasta: Amazzonia Boliviana

**Tabella 4.4:** Confronto tra le variabili con valori di CP e IPP superiori alla soglia (1 %) per la valutazione dei modelli, elaborato per la scelta dello stato informativo relativo alle caratteristiche della vegetazione. Sulla sinistra i risultati per l'elaborazione con la carta della vegetazione potenziale di Navarro (V) e sulla destra rualli per l'elaborazione con la carta delle ecoregioni di Ibisch (E).

Variabile V (%)	PC V(%)	IPP V(%)	Variabile E(%)	PC E(%)	IPP E(%)
VEGET.NAVARRO	23,1	14,6	BIO2	22,6	11,4
BIO2	21,9	17,2	USO.TIERRA	18,7	5,9
USO.TIERRA	14,8	7	BIO4	16,4	40,9
SOIL	10,3	8,5	SOIL	11	12,5
DEM	9,4	2,4	ECOREG	8,9	3,1
BIO4	7	22,3	DEM	7,7	0,1
BIO7	4,1	5	BIO16	2,8	6,6
BIO16	2,5	6,2	BIO7	2,6	1,1
BIO19	1,6	5,2	BIO19	1,6	1,3
BIO1	1,3	2,6	BIO10	1	0
BIO18	0,9	3,2	BIO5	0,6	1,3
BIO5	0,9	1,8	BIO1	0,4	7,1
BIO17	0,3	1,2	BIO12	0,2	2,1
			DIST INFR	3,1	1,4

pur essendo "equivalente" per la determinazione della distribuzione del rischio incendi, questa comprende un maggiore dettaglio e rende più immediata l'interpretazione della struttura e della tipologia di vegetazione in relazione al rischio di incendio.

### 4.5.2 Parametri del modello

Considerando la grande estensione dell'area indagata per l'elaborazione del modello di distribuzione del rischio di incendi sono state considerate un numero superiore di punti di *background*. Sono state effettuate cinque repliche di calcolo, dove, per ogni replica i punti di presenza che compongono i dataset di *train* e di *test* sono stati ricampionati con il metodo *bootstrap*. Questo ha permesso di evitare che la stima della distribuzione del rischio potesse essere influenzata da l'estrazione casuale di un dataset parzialmente rappresentativo.

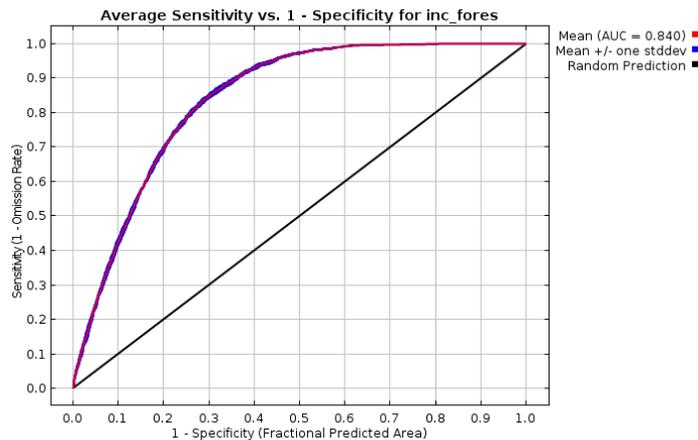
Il dettaglio dei parametri utilizzati per il modello elaborato con le variabili selezionate è riportato in tabella 4.5.

## 4.6 Risultati

L'elaborazione condotta con le variabili selezionate ha prodotto, su cinque ripetizioni di calcolo, una stima della distribuzione del rischio per gli incendi forestali con un'accuratezza, misurata in termini di area sottesa alla curva ROC (in figura 4.4), in media pari a  $AUC = 0,84$  e con una deviazione standard media pari a  $s = 0,002$ .

**Tabella 4.5:** Parametri del modello elaborato con le variabili selezionate per la valutazione del rischio incendio in Amazzonia boliviana.

Parametro	Valore / modalità
Punti training	1270
Punti test	544
Totale dei punti utilizzati per determinare la distribuzione	31267 (punti di training + punti di background)
Valore di regolarizzazione feature lineare	0.05
Valore di regolarizzazione feature quadratica	0.05
Valore di regolarizzazione feature prodotto	0.05
Valore di regolarizzazione feature categorica	0.25
Valore di regolarizzazione feature soglia	1
Valore di regolarizzazione feature hinge	0.5
Combinazione di feature usata	prodotto - lineare - quadratica - hinge - soglia
Coeff. di regolarizzazione	2.0
Repliche	5
Ricampionamento	Bootstrap

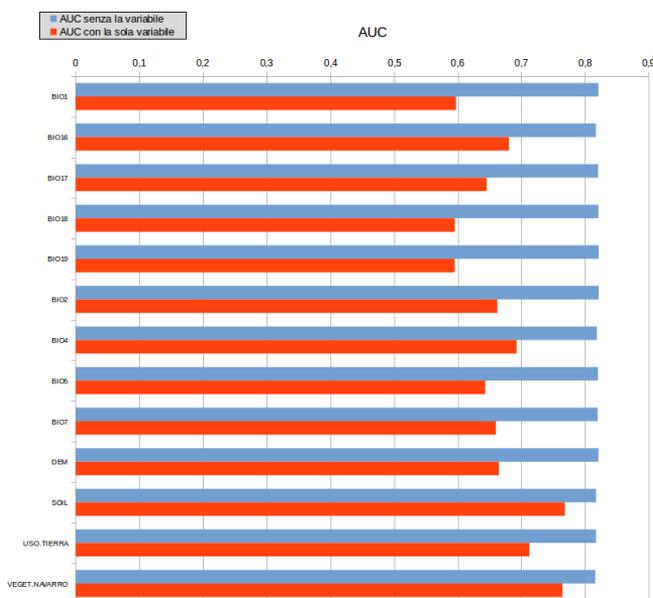


**Figura 4.4:** Curva ROC delle performance media delle cinque reiterazioni di calcolo del modello per gli incendi forestali in Amazzonia boliviana.

Dalla stima delle prestazioni dei modelli calcolati a seguito del ricampionamento JRR, applicato alle covariate selezionate, e riportati nel grafico a barre in figura 4.5 è possibile notare che la sottrazione di nessuna variabile (in turchese) restituisca una diminuzione sensibile delle prestazioni dei modelli, non abbassandosi mai sotto la soglia del valore 0,81 di AUC. Parallelamente si nota come le migliori prestazioni si registrino nei modelli (in rosso nel grafico) in cui si considerino solo le informazioni contenute nelle variabili di uso del suolo, o con la carta della vegetazione di Navarro, o con la carta dei suoli, rispettivamente con valori di AUC di 0,71, 0,76 e 0,77.

Dai valori stimati per il contributo percentuale e per l'importanza di permutazione percentuale, riportate in tabella 4.6, emergono, oltre alle variabili

## Capitolo 4. Elaborazione su scala vasta: Amazzonia Boliviana



**Figura 4.5:** *Rappresentazione grafica di JRR applicato al modello per gli incendi forestali in Amazzonia boliviana: in rosso sono riportati i valori di AUC registrati per modelli che considerano una sola variabile alla volta (in ordinata), in turchese sono riportati i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata.*

evidenziate dalla procedura JRR, anche gli alti valori delle componenti climatiche legate alla temperatura. In particolare l'escursione termica media giornaliera<sup>5</sup> (BIO2 con  $CP = 19,9\%$ ) e la stagionalità delle temperature<sup>6</sup> (BIO4 con  $IPP = 31\%$ ) risultano essere particolarmente informative per i modelli di distribuzione del rischio incendi in quest'area. Da sottolineare, inoltre anche il contributo, in termini di  $IPP$  delle precipitazioni dei tre mesi più piovosi (BIO16, con  $9,9\%$ ).

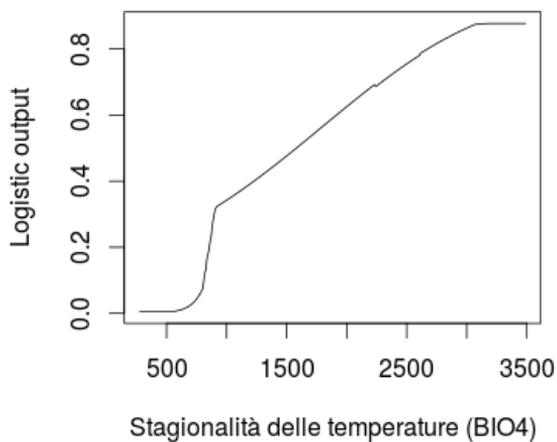
<sup>5</sup>L'escursione termica media giornaliera, espressa come *range* medio giornaliero delle temperature, è calcolata come media mensile della differenza tra le temperature massime e le temperature minime.

<sup>6</sup>La stagionalità è riportata come la deviazione standard delle temperature annuali.

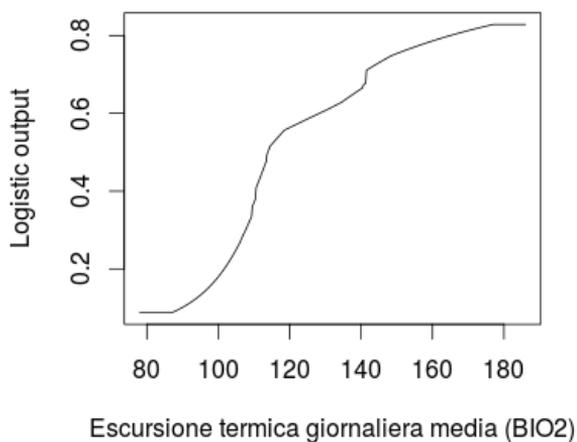
**Tabella 4.6:** Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili utilizzate per l'elaborazione del modello di distribuzione del rischio incendio in Amazzonia boliviana.

Variabile	CP(%)	IPP(%)
BIO2	19,94	11,47
VEGET.NAVARRO	18,92	13,69
USO.TIERRA	14,82	6
BIO4	12,46	31
SOIL	12	11,18
DEM	6,87	2,95
BIO16	4,62	9,96
BIO5	3,83	3,57
BIO7	3,27	2,22
BIO19	1,44	2,86
BIO17	1,28	1,21
BIO1	0,63	4
BIO18	0,39	0,67

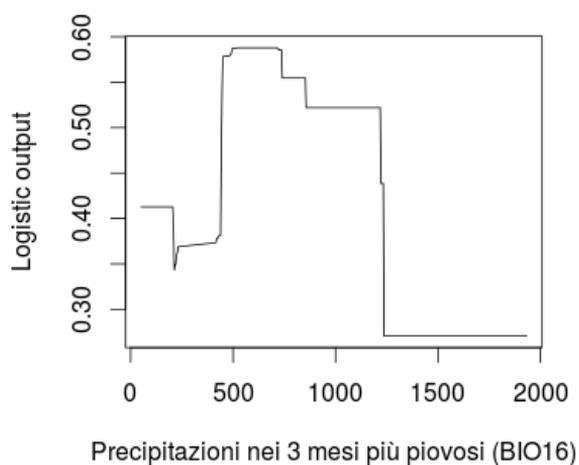
Di seguito sono riportate le restituzioni grafiche della variazione di probabilità di presenza, o più correttamente di variazione del rischio, media per gli incendi, al variare delle covariate maggiormente informative precedentemente menzionate. Il rischio è riportato in ordinata in formato logistico (con *range* 0 – 1) che, come suggerito da Baldwin [4], facilita e rende potenzialmente più accurata l'interpretazione rispetto ai valori della covariata in ascisse (figure 4.6 - 4.7 - 4.8). Per le covariate a variazione non continua sono riportate le categorie con valore di rischio, sempre in formato logistico, superiore a 0, 6.



**Figura 4.6:** *Rischio di incendio in relazione alla stagionalità delle temperature per l'area di studio (deviazione standard moltiplicata per 100).*

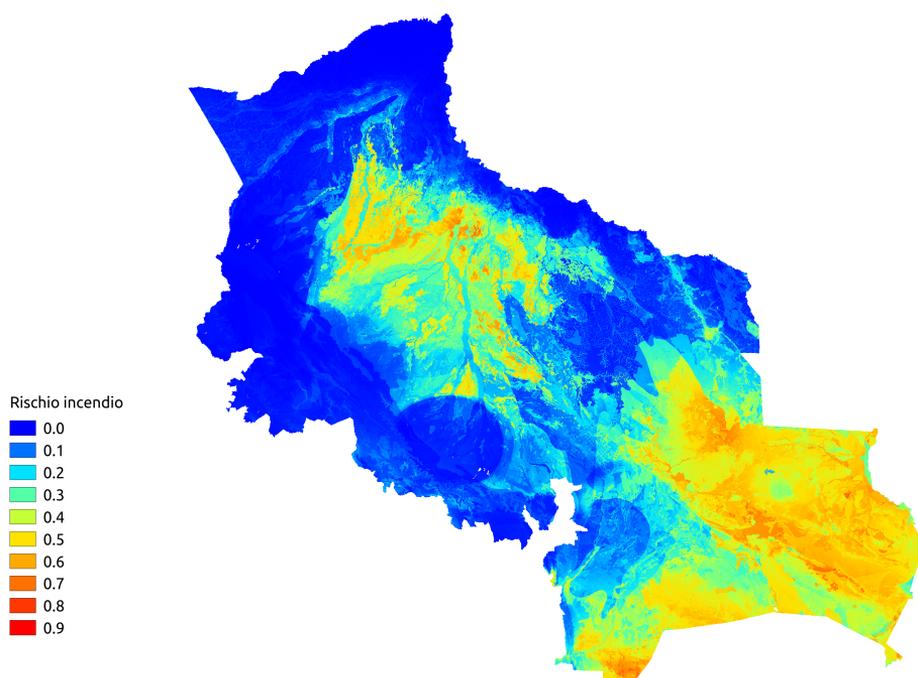


**Figura 4.7:** *Rischio di incendio in relazione all'escursione termica media giornaliera (le temperature sono espresse in gradi Celsius e moltiplicati per 10).*



**Figura 4.8:** *Rischio di incendio in relazione alle precipitazioni cumulate nei tre mesi più piovosi dell'anno.*

Infine, in figura 4.9 è riportata la restituzione grafica della distribuzione spaziale media del rischio di incendio. I valori di rischio, in formato logistico, sono stati riportati su gradiente cromatico per facilitare la lettura della carta.



**Figura 4.9:** *Distribuzione media del rischio di incendi nella Cuenca Amazonica boliviana.*

**Tabella 4.7:** Valori in formato logistico del rischio incendio delle categorie più significative (valore  $\geq 0,6$ ) nelle variabili non continue per il modello elaborato per l'Amazzonia boliviana

Variabile	Categoria	Valore di rischio
USO.TIERRA	Pascoli estensivi	0,67
	Aree silvopastorali a bassa densità di pascolo	0,61
VEGET.NAVARRO	Bosco basso chiquitano su suolo sabbioso o roccioso	0,60
	Foreste di Varzea e vegetazione ripariale del Mamoré-Ibare	0,60
	Boschi semidecidui idrofilo	0,61
	Paludi del basso Río Grande	0,61
	Boschi ripari del Nord del Beni	0,61
	Boschi di transizione del Nord del Chaco	0,61
	Foreste di Varzea e vegetazione ripariale del Río Beni	0,62
	Savana aperta igrofila seminaturale	0,62
	Savana alberata e arbustiva dell'alta Amazoniz	0,62
	Savana period. allagata su suoli non alcalini del Rogagua	0,63
	Savana period. allagata del Mamoré-Ibare	0,64
	Savana arborata del Beni	0,64
	Boschi di Varzea e liane del Beni	0,64
	Savana erbacea mesotrofica del Beni	0,65
	Savana period. allagata del Omi-Tapado	0,66
	Bosco paludoso di palme di pianura alluvionale del Sud Amazonia	0,67
	Boschi ripari del Omi-Tapado	0,69
Boschi di Varzea del Maniqui-Cuberene	0,70	
Macchia di sclerofille della Chiquitania	0,72	
Paludi del Beni Est	0,73	
Foresta planiziale occidentale	0,73	
SOIL	Dystric Gleysols	0,61
	Mollic Gleysols	0,61
	Orthic Acrisols	0,61
	Acric Ferralsols	0,64
	Dystric Cambisols	0,65
	Plinthic Acrisols	0,74



---

## Applicazione a scala locale in Bolivia

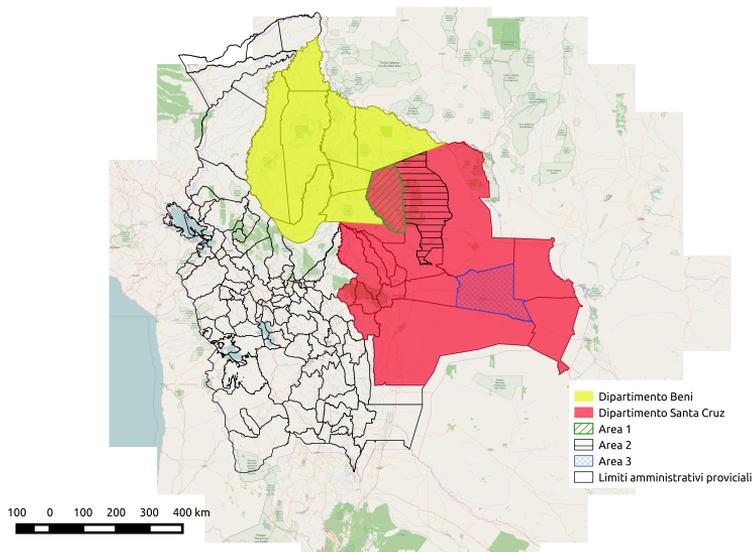
---

Per testare le prestazioni dell'algoritmo di massima entropia in relazione al rapporto tra l'estensione dell'area di indagine e la risoluzione spaziale delle covariate utilizzate, sono stati calcolati modelli locali di distribuzione del rischio di incendio. In queste elaborazioni, non potendo aumentare il dettaglio delle variabili, il rapporto tra la risoluzione spaziale delle informazioni e l'estensione dell'area sulla quale si applica l'algoritmo è stato indagato mantenendo invariata la risoluzione rispetto al modello generale del capitolo 4 (Raster a celle quadrate di 250 metri di lato) e variando la superficie di indagine.

L'elaborazione ha previsto l'analisi dei due dipartimenti di Beni e Santa Cruz; all'interno di quest'ultimo sono state individuate ulteriori tre aree, di estensione più limitata, rappresentate in figura 5.1.

La motivazione della scelta di queste aree è legata ai risultati della precedente elaborazione. I due dipartimenti individuati in questa fase sono quelli dove si sono registrati i valori medi più elevati di rischio per gli incendi. Inoltre, le tre sotto-aree sono state scelte, oltre per la disponibilità dei punti di presenza, anche perché hanno un'estensione di circa un decimo della superficie del singolo dipartimento. In questo modo è stato possibile indagare le prestazioni dei modelli su differenti estensioni, il cui dettaglio è riportato in tabella 5.1, e confrontarne i risultati.

Al fine di garantire la comparabilità dei risultati, oltre a mantenere invariata la risoluzione spaziale delle covariate utilizzate, queste non hanno subito un processo di selezione a scala locale, ma sono stati impiegati gli stessi strati informativi del modello generale, ad unica eccezione della carta dei suoli ri-



**Figura 5.1:** *Inquadramento delle aree di indagine per la valutazione delle prestazioni dell’algoritmo di massima entropia in relazione al rapporto tra risoluzione spaziale delle variabili ed estensione territoriale.*

**Tabella 5.1:** *Estensione delle superfici e percentuale rispetto all’area del progetto generale (PASF) delle aree considerate per la valutazione delle performance dei modelli a diversa scala geografica.*

Area di indagine	Superficie (km <sup>2</sup> )	Superficie Percentuale (%)
Area totale PASF	732448	100
Dip. Beni	202859	28
Dip. Santa Cruz	345501	47
Area 1	19027	3
Area 2	34887	5
Area 3	26251	4

portata secondo la classificazione FAO. Sebbene questa sia stata inclusa nel modello generale come espressione della variabilità topografica e delle pendenze a scala di paesaggio [31] ed abbia registrato valori significativi per alcune categorie (tabella 4.7), con la riduzione dell’estensione della superficie di indagine, si riduce progressivamente anche la capacità rappresentativa dei tipi di suolo per quanto riguarda la variabilità topografica e le pendenze. Se si considera infatti il numero delle categorie di suolo nelle diverse aree di indagine, si può notare la progressiva diminuzione al diminuire della superficie, come riportato in tabella 5.2.

Per ovviare alla verosimile perdita di informazioni, causata dall’impiego di una covariata idonea a scala di paesaggio come la carta dei suoli, per l’applica-

**Tabella 5.2:** *Categorie di suoli (FAO) in relazione alla superficie delle aree considerate per la valutazione delle performance dei modelli a diversa scala geografica.*

Area di indagine	Superficie (km <sup>2</sup> )	Numero di categorie di suolo FAO
Area totale PASF	732448	92
Dip. Santa Cruz	345501	54
Dip. Beni	202859	26
Area 2	34887	12
Area 3	26251	6
Area 1	19027	5

zione a scala locale, la variabilità topografica è stata rappresentata includendo la carta delle pendenze. Le informazioni sulle pendenze sono state calcolate sulla base delle quote altimetriche riportate dal modello digitale del terreno GMTED2010.

### 5.1 Elaborazione a scala locale

Come precedentemente accennato, per favorire la comparabilità dei risultati all'applicazione dell'algoritmo di massima entropia nella stima della distribuzione spaziale del rischio di incendi a diverse scale geografiche, tutte le covariate considerate per il modello generale sono state incluse<sup>1</sup>. Inoltre, anche i parametri in input usati per l'elaborazione del modello, riepilogati in tabella 5.3, sono rimasti invariati ad eccezione della dimensione campionaria (i punti di presenza variano per le singole aree), per la quale è stato comunque mantenuto il rapporto tra punti di *training* e punti per il *test*, rispettivamente di 70% e 30%. Sia il numero di punti utilizzati, che la combinazione di *feature* utilizzate dall'algoritmo sono riportate, per ogni area di indagine, al rispettivo paragrafo.

**Tabella 5.3:** *Parametri comuni ai modelli elaborati a scala locale per la comparazione delle performance di MAXENT nella distribuzione del rischio incendio in Amazzonia boliviana.*

Parametro	Valore / modalità
Valore di regolarizzazione feature lineare	0.05
Valore di regolarizzazione feature quadratica	0.05
Valore di regolarizzazione feature prodotto	0.05
Valore di regolarizzazione feature categorica	0.25
Valore di regolarizzazione feature soglia	1
Valore di regolarizzazione feature hinge	0.5
Coeff. di regolarizzazione	2.0
Repliche	5
Ricampionamento	Bootstrap

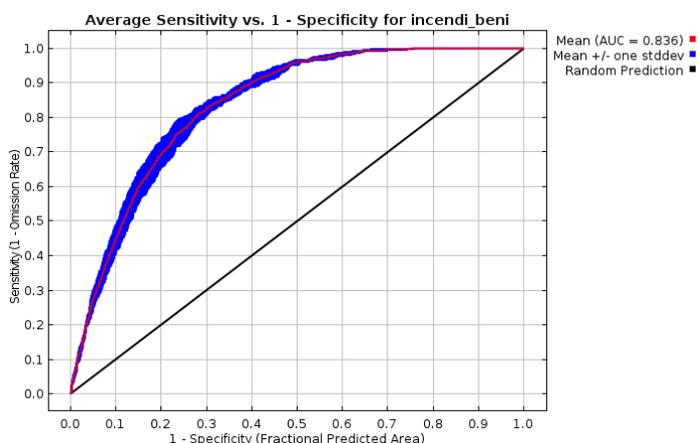
<sup>1</sup> ad eccezione della carta dei suoli FAO sostituita, per le motivazioni descritte all'inizio del capitolo, con la carta delle pendenze.

## 5.2 Risultati

### 5.2.1 Dipartimento 1 - Beni

L'elaborazione è stata condotta a partire da 338 punti di presenza per il *training*, e 114 per il *test*, per un totale di 10338 punti per determinare la distribuzione (*training + background*). Le combinazioni di *feature* selezionate dall'algoritmo sono state: prodotto, lineare, quadratica, *hinge* e di soglia.

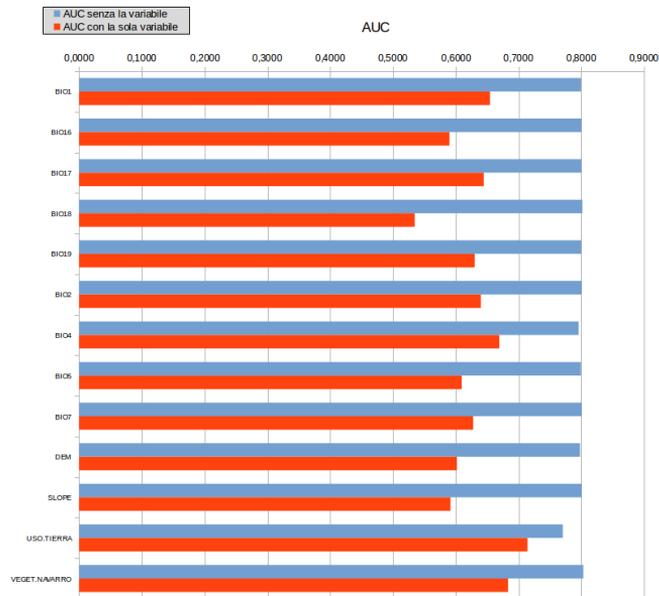
Su cinque repliche di calcolo, per il dipartimento del Beni è stata stimata la distribuzione del rischio di incendi forestali con un valore medio di  $AUC = 0,836$ , con deviazione standard pari a  $s = 0,011$ , rappresentata in figura 5.2.



**Figura 5.2:** Curva ROC delle performance medie delle cinque reiterazioni di calcolo del modello per gli incendi forestali nel dipartimento del Beni.

La stima delle prestazioni dei modelli calcolati a seguito del ricampionamento JRR, riportati nel grafico in figura 5.3, evidenzia come la componente delle categorie di uso del suolo (USO.TIERRA nel grafico) risulti apportare il maggiore contributo alla accuratezza del modello, facendo registrare il valore più basso (0,771) in termini di AUC, nel caso la si escluda dal modello. Parallelamente, se la si considera singolarmente (le barre in rosso del grafico) escludendo tutte le altre covariate, l'uso del suolo registra il punteggio più elevato con  $AUC = 0,714$ .

In tabella 5.4 sono riportati i valori di contributo percentuale e importanza di permutazione per le variabili utilizzate, in ordine decrescente rispetto al valore di contributo percentuale. Analogamente a quanto evidenziato dalla procedura di ricampionamento JRR, anche per questi stimatori risulta evidente l'informatività legata alle categorie di uso del suolo ( $CP = 42,33$  e



**Figura 5.3:** Rappresentazione grafica dei risultati di ricampionamento JRR delle covariate nel Beni: in rosso sono riportati i valori di AUC registrati per modelli che considerano una sola variabile alla volta (in ordinata), in turchese sono riportati i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata.

$IPP = 16,59$ ). Anche la componente di tipologia e struttura della vegetazione e la stagionalità delle temperature (rispettivamente VEGET.NAVARRO e BIO4 nella tabella) risultano essere componenti importanti, con valori di CP superiori al 15%. Delle altre variabili, il cui valore di contributo percentuale supera la soglia dell'1%, è da notare che sono tutte riferite al regime termico dell'area di indagine, con le sole eccezioni di BIO17 (precipitazioni nei tre mesi più secchi con  $IPP = 8,63$ ) e delle quote altimetriche (DEM con  $IPP = 9,50$ ).

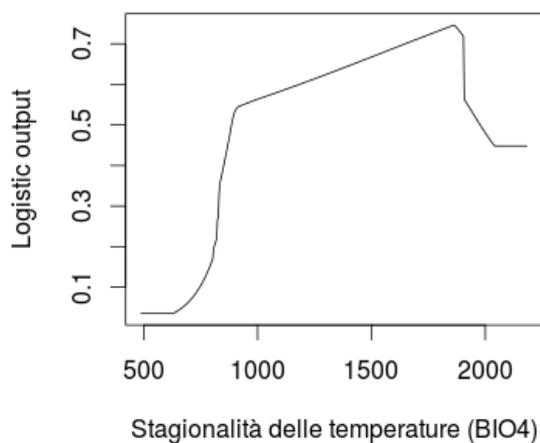
## Capitolo 5. Applicazione a scala locale in Bolivia

---

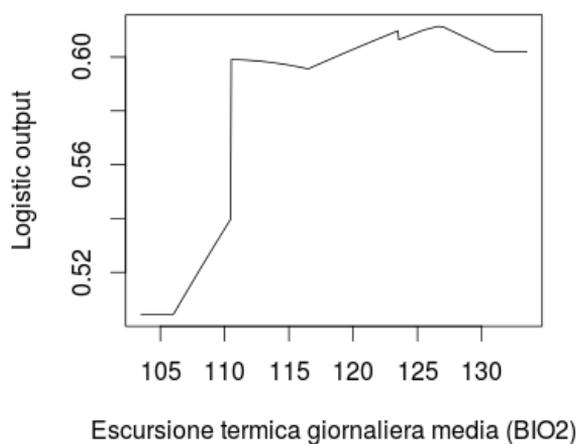
**Tabella 5.4:** *Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili utilizzate per l'elaborazione del modello di distribuzione del rischio incendio nel dipartimento del Beni.*

Variabile	CP(%)	IPP(%)
USO.TIERRA	42,33	16,59
VEGET.NAVARRO	17,15	11,30
BIO4	15,61	24,76
DEM	7,86	9,50
BIO1	7,35	0,97
BIO5	3,44	19,01
BIO2	2,56	1,48
BIO17	1,58	8,63
BIO19	0,54	2,38
SLOPE	0,54	0,60
BIO7	0,44	1,31
BIO18	0,42	2,58
BIO16	0,18	0,89

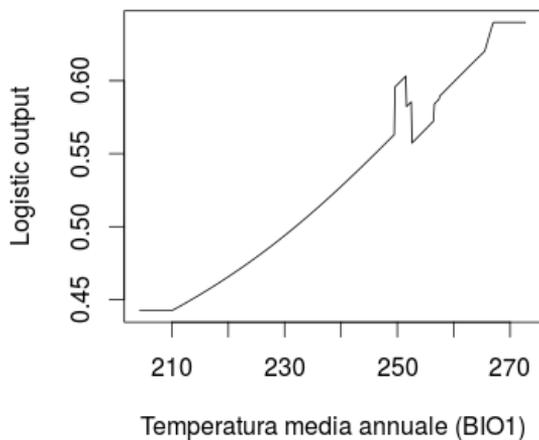
Come per il caso di studio generale, trattato nel capitolo 4, di seguito sono riportate le restituzioni grafiche della variazione media del rischio per gli incendi, in formato logistico, al variare delle covariate maggiormente informative appena descritte (da figura 5.4 a figura 5.9). Per le covariate a variazione non continua sono riportate in tabella 5.5 le categorie con valore di rischio maggiormente significativo (superiore a 0,6), in formato logistico.



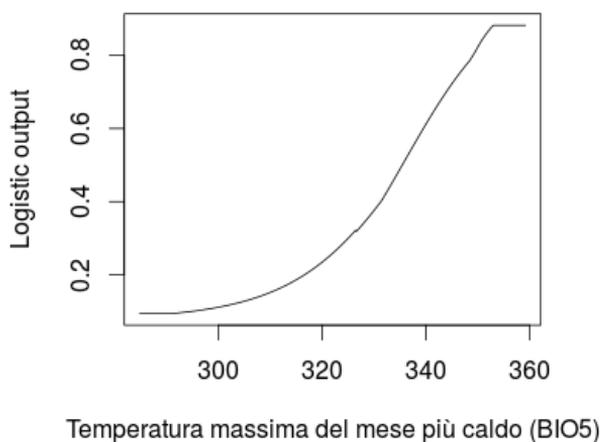
**Figura 5.4:** *Rischio di incendio in relazione alla stagionalità delle temperature per il dipartimento del Beni (deviazione standard delle temperature annuali in gradi Celsius moltiplicata per 100).*



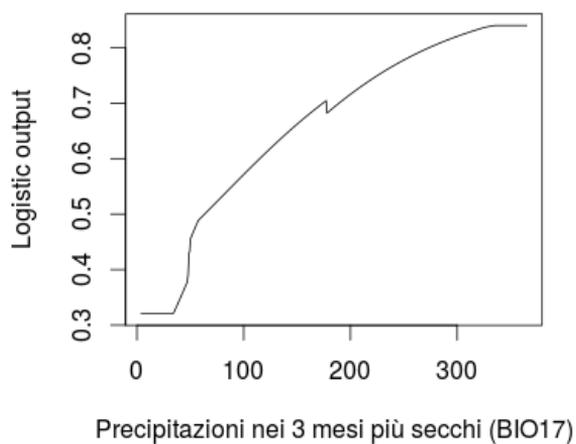
**Figura 5.5:** *Rischio di incendio in relazione all'escursione termica media giornaliera nel dipartimento del Beni (le temperature sono espresse in gradi Celsius e moltiplicate per 10).*



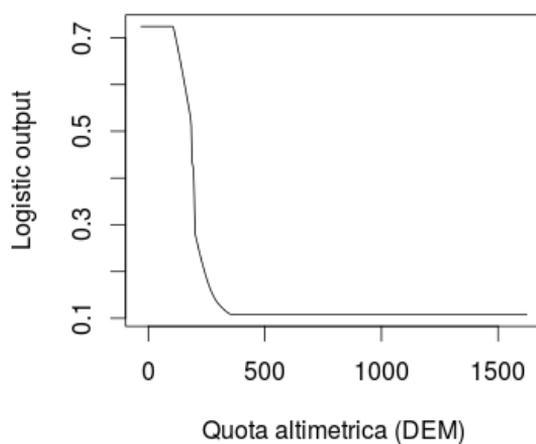
**Figura 5.6:** *Rischio di incendio in relazione alle temperature medie annuali nel dipartimento del Beni (le temperature sono espresse in gradi Celsius e moltiplicate per 10).*



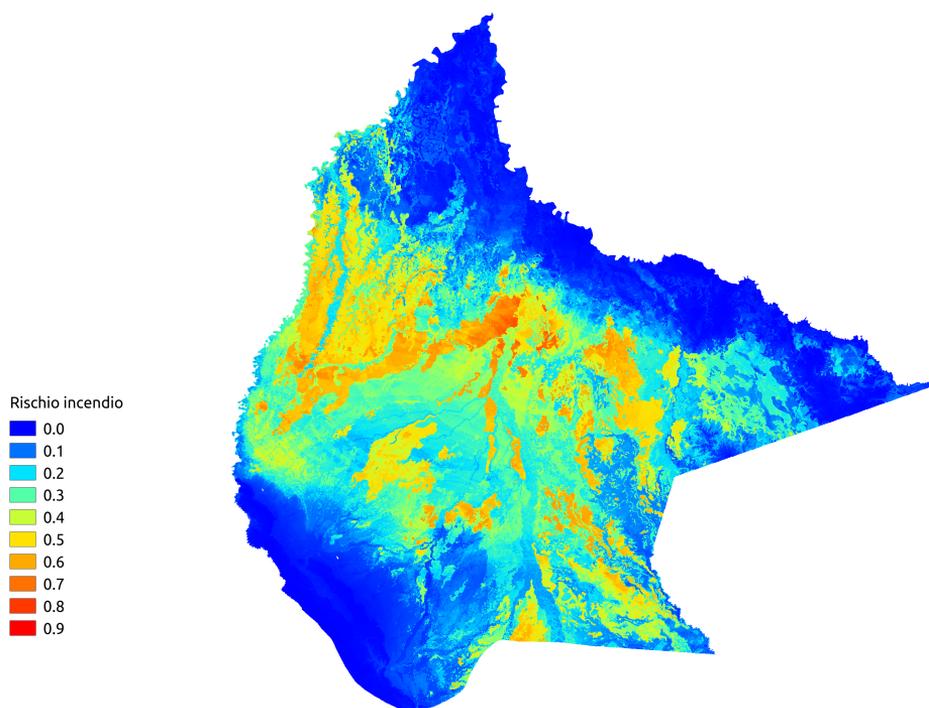
**Figura 5.7:** *Rischio di incendio in relazione alle temperature massime, registrate nel mese più caldo, nel dipartimento del Beni (le temperature sono espresse in gradi Celsius e moltiplicate per 10).*



**Figura 5.8:** *Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) dei tre mesi più siccitosi nel dipartimento del Beni.*



**Figura 5.9:** *Rischio di incendio in relazione alla quota altimetrica nel dipartimento del Beni (m s.l.m.).*



**Figura 5.10:** *Distribuzione media del rischio di incendi nel Dipartimento del Beni.*

**Tabella 5.5:** Valori in formato logistico del rischio incendio delle categorie più significative (valore  $\geq 0,6$ ) nelle variabili non continue per il modello elaborato per il dipartimento del Beni

Variabile	Categoria	Valore di rischio
USO.TIERRA	Pascoli estensivi	0,6
VEGET.NAVARRO	Boschi di Varzea e liane del Beni	0,61
	Savana arborata del Beni	0,62
	Paludi del Beni Est	0,62
	Savana periodicamente allagati del Mamoré-Ibare	0,62
	Boschi ripari del Nord del Beni	0,63
	Savana period. allagati su suoli non alcalini del Secure	0,63
	Savana period. allagati su suoli non alcalini del Omi-Tapado	0,64
	Paludi del basso Rio Grande	0,64
	Boschi di Varzea e vegetazione riparia del Rio Beni	0,65
	Vegetazione acquatica e palustre neotropica del Beni	0,65
	Boschi di Varzea e vegetazione riparia del Secure	0,66
	Savana aperta igrofila delle alture del Beni	0,66
	Boschi ripari del Omi-Tapado	0,67
	Savana period. allagata su suoli non alcalini del Rogagua	0,67
	Paludi del Beni Ovest	0,69
	Boschi di Varzea del Maniqui-Cuberene	0,69
	Boschi semidecidui su suoli mediamente ben drenati del Beni	0,69
	Bosco paludoso di palme della pianura alluvionale del Sud Amazonia	0,7
	Boschi amazzonici period. allagati con acqua stagnante	0,74
	Boschi period. allagati di pianura alluvion. del Sud-Ovest Amazzonia	0,87

### 5.2.2 Dipartimento 2 - Santa Cruz

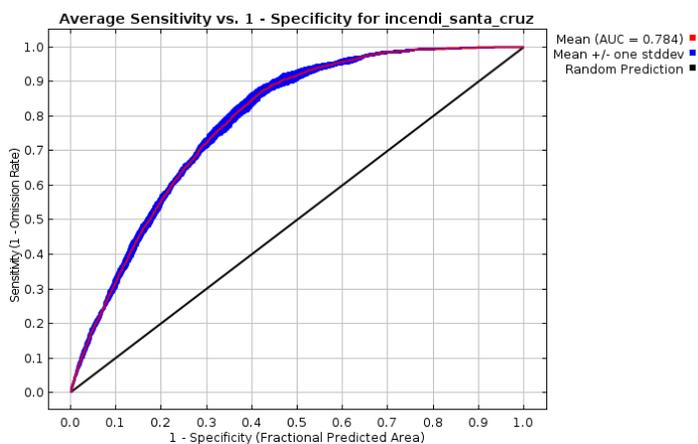
Per il secondo dipartimento sono stati impiegati 929 punti di presenza per il *training*, e 398 per il *test*, per un totale di 10929 punti per determinare la distribuzione (*training + background*). Anche per quest'area le combinazioni di *feature* selezionate dall'algoritmo sono state: prodotto, lineare, quadratica, *hinge* e di soglia.

La stima della distribuzione del rischio per gli incendi forestali ha registrato un valore medio di  $AUC = 0,784$ , con deviazione standard pari a  $s = 0,008$ , rappresentata in figura 5.11.

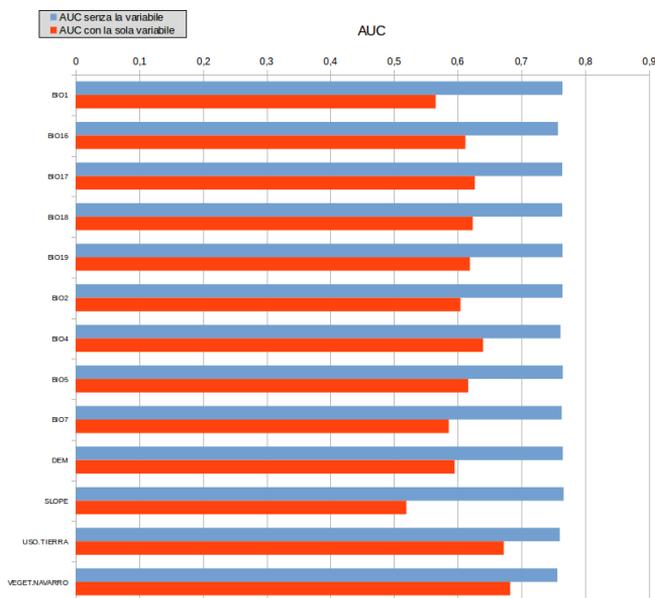
Dal ricampionamento JRR si evidenziano risultati simili al dipartimento del Beni, anche se meno marcati (grafico in figura 5.12), con la componente delle categorie di uso del suolo che apporta il maggiore contributo alla accuratezza del modello, ovvero facendo registrare il valore più basso (0,761) in termini di AUC, nel caso la si escluda dal modello e il valore maggiore nel caso in cui si calcoli un modello con solo questa variabile ( $AUV = 0,672$ ).

Dall'esame dei valori stimati di contributo percentuale e di importanza di permutazione percentuale, riportati in tabella 5.6, si può notare come in questo caso, il risultato della procedura JRR non corrisponda con i risultati di questi due stimatori. La variabile a maggiore informatività risulta essere BIO4

## Capitolo 5. Applicazione a scala locale in Bolivia



**Figura 5.11:** Curva ROC delle performance medie delle cinque reiterazioni di calcolo del modello per gli incendi forestali nel dipartimento di Santa Cruz.



**Figura 5.12:** Rappresentazione grafica dei risultati di ricampionamento JRR delle covariate nel dipartimento di Santa Cruz: in rosso sono riportati i valori di AUC registrati per modelli che considerano una sola variabile alla volta, in turchese, i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata.

(la stagionalità delle temperature con  $CP = 27,6$  e  $IPP = 34,2$ ), invece delle categorie di uso del suolo. L'elevato valore di importanza di permutazione evidenzia il fatto che, per il dipartimento di Santa Cruz le informazioni

apportate da questa variabile, oltre ad essere determinanti nella stima della distribuzione degli incendi forestali, non possono essere ricostruite utilizzando le informazioni derivate dagli altri strati informativi. Da notare inoltre come per quest'area di indagine, oltre alle componenti legate alla tipologia di vegetazione e uso del suolo, e alle covariate termiche (BIO4, BIO7, BIO5 e BIO2), notevole peso hanno le variabili pluviometriche. In particolare il valore di IPP registrato dalle precipitazioni nei tre mesi più piovosi (BIO16), prossimo al 19%, rende questa covariata particolarmente informativa.

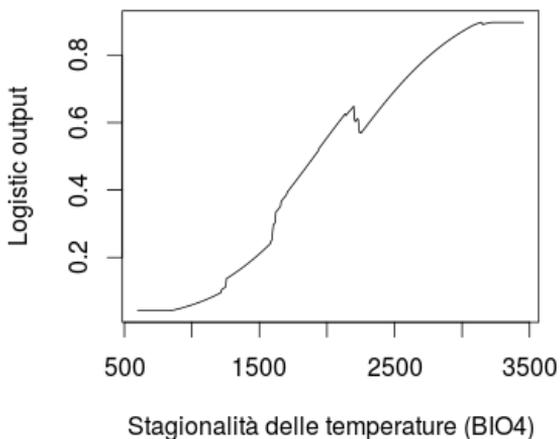
**Tabella 5.6:** *Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili utilizzate per il modello di distribuzione del rischio incendio nel dipartimento di Santa Cruz.*

Variabile	CP(%)	IPP(%)
BIO4	27,62	34,23
USO.TIERRA	21,54	7,41
VEGET.NAVARRO	16,24	8,54
BIO7	15,74	9,59
BIO16	7,83	18,77
BIO17	2,28	2,79
BIO5	2,26	1,84
BIO2	2,25	7,23
BIO19	1,36	4,37
DEM	1,31	0,96
SLOPE	0,71	1,28
BIO18	0,46	2,12
BIO1	0,41	0,88

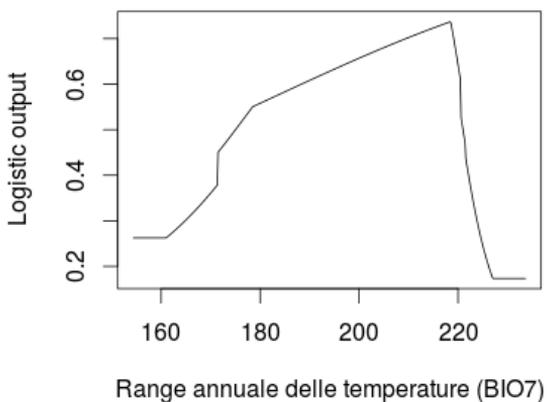
Di seguito sono riportati i grafici della variazione media del rischio per gli incendi in formato logistico, al variare delle covariate maggiormente informative (da figura 5.13 a figura 5.20).

Per quanto riguarda le covariate categoriche di uso del suolo e di tipologia di vegetazione sono da sottolineare i valori relativamente bassi di IPP, rispettivamente di 7,4% e 8,5%. Questo è riconducibile alla bassa variabilità negli usi del suolo e da tipi di vegetazione simili dal punto di vista della suscettibilità al passaggio degli incendi, come confermato dalle categorie maggiormente significative in termini di rischio.

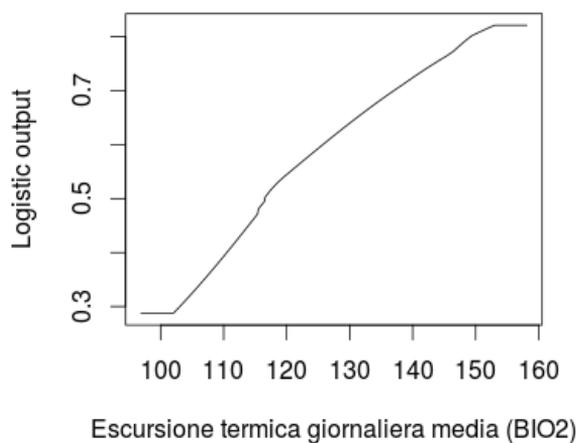
Per quanto riguarda l'uso del suolo, infatti, più del 65% delle categorie (15 su un totale di 23) registra un valore di rischio incendio superiore a 0,6. Tutte queste categorie risultano legate all'attività agricola (massimo valore per le aree ad agricoltura intensiva di 0,69) e di pascolo (pascolo estensivo con 0,66 e aree silvopastorali con 0,63). Similarmente, anche le categorie di tipi vegetazionali significative, dal punto di vista del rischio, sono preponderanti, più dell'88% (75 su 85 totali per Santa Cruz), con i valori massimi per la vegetazione di savana del centro Sud dell'Amazzonia (0,89) e le aree a vegetazione palustre (0,91).



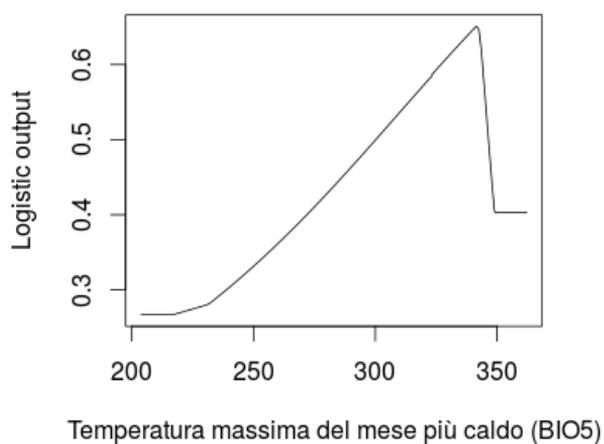
**Figura 5.13:** *Rischio di incendio in relazione alla stagionalità delle temperature per il dipartimento di Santa Cruz (deviazione standard delle temperature annuali in gradi Celsius moltiplicata per 100).*



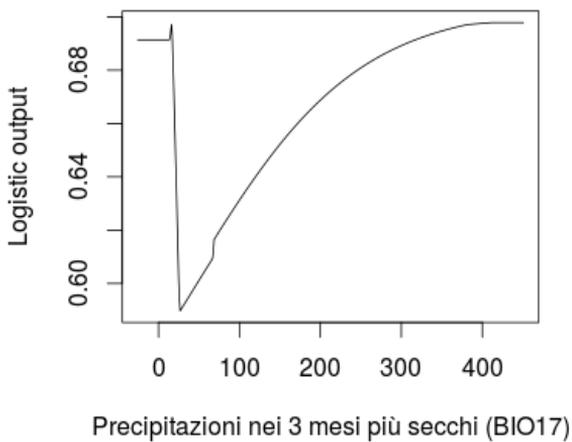
**Figura 5.14:** *Rischio di incendio in relazione al range annuale di temperatura nel dipartimento di Santa Cruz (le temperature sono espresse in gradi Celsius e moltiplicate per 10).*



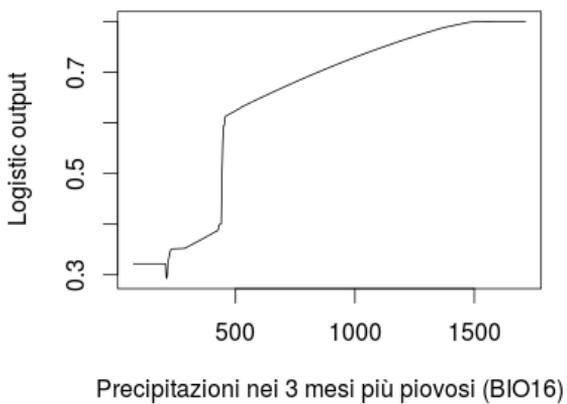
**Figura 5.15:** *Rischio di incendio in relazione all'escursione termica media giornaliera nel dipartimento di Santa Cruz (le temperature sono espresse in gradi Celsius e moltiplicate per 10).*



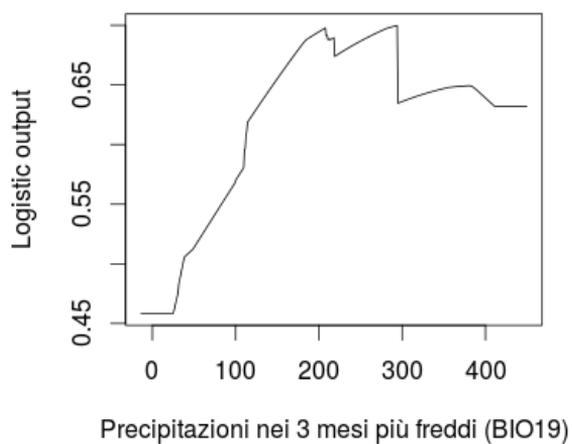
**Figura 5.16:** *Rischio di incendio in relazione alle temperature massime, registrate nel mese più caldo, nel dipartimento di Santa Cruz (le temperature sono espresse in gradi Celsius e moltiplicate per 10).*



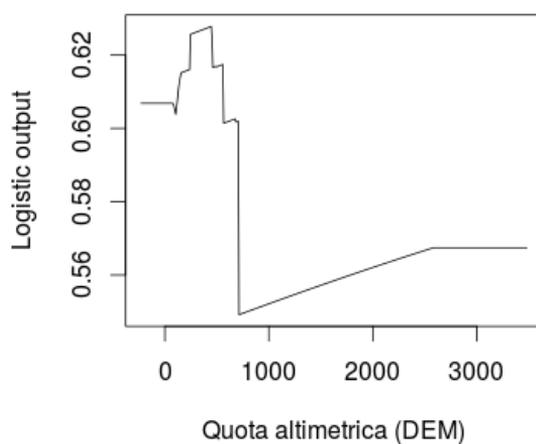
**Figura 5.17:** *Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) dei tre mesi più siccitosi nel dipartimento di Santa Cruz.*



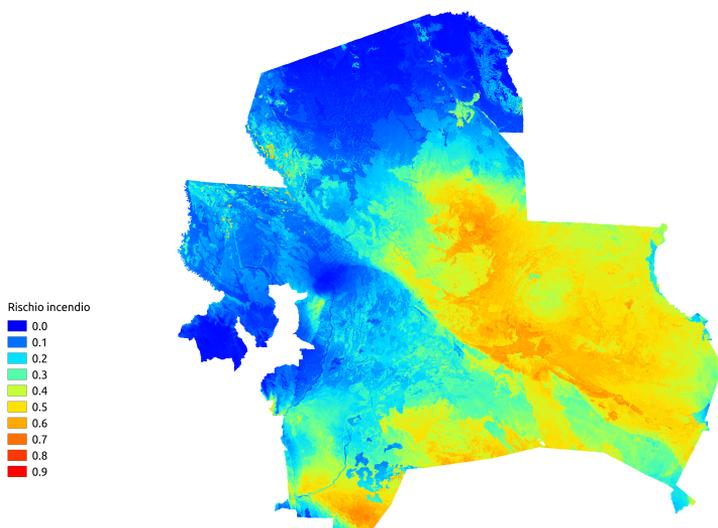
**Figura 5.18:** *Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) dei tre mesi più piovosi nel dipartimento di Santa Cruz.*



**Figura 5.19:** *Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) dei tre mesi più freddi nel dipartimento di Santa Cruz.*



**Figura 5.20:** *Rischio di incendio in relazione alla quota altimetrica nel dipartimento di Santa Cruz (m s.l.m.).*



**Figura 5.21:** *Distribuzione media del rischio di incendi nel Dipartimento di Santa Cruz.*

### 5.2.3 Area 1

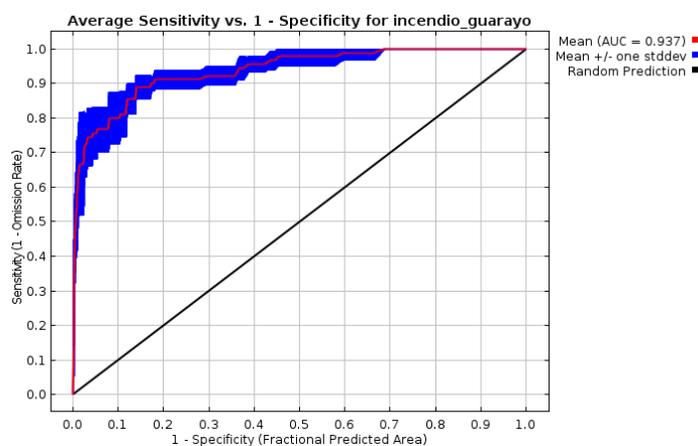
La prima area di indagine, all'interno del dipartimento di Santa Cruz, è situata nella parte Nord, al confine col dipartimento del Beni, in corrispondenza delle municipalità di Urubichà e di Ascención de Guarayos.

Sono stati impiegati 30 punti di presenza per il *training*, e 9 per il *test*, per un totale di 10030 punti per determinare la distribuzione (*training + background*). Per quest'area le combinazioni di *feature* selezionate dall'algoritmo sono state: lineare, quadratica e *hinge*.

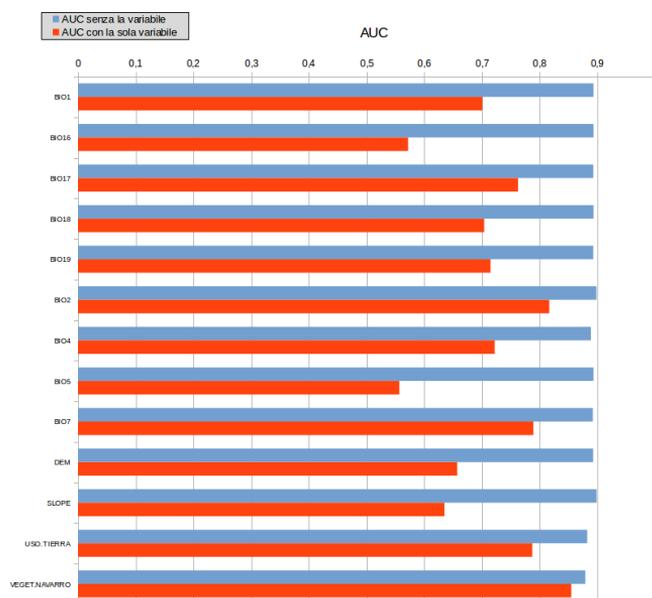
La stima della distribuzione del rischio per gli incendi forestali ha registrato un valore medio di  $AUC = 0,937$ , con deviazione standard pari a  $s = 0,008$ , rappresentata in figura 5.22

In maniera simile a quanto evidenziato per i due dipartimenti, in quest'area i risultati dei modelli calcolati a seguito del ricampionamento JRR mostrano un maggiore peso alle informazioni apportate dalle tipologie di vegetazione e dalle categorie di uso del suolo per l'accuratezza del modello (grafico in figura 5.23). Si nota infatti come, nel caso in cui si escludano l'una o l'altra variabile categorica, i relativi modelli registrino la maggiore diminuzione del valore di AUC (in turchese, nel grafico).

Il risultato trova conferma nelle stime di contributo percentuale, in tabella 5.7, dove l'uso del suolo (USO.TIERRA) registra il massimo valore, seguito dall'escursione termica media giornaliera e dalla tipologia di vegetazione.



**Figura 5.22:** Curva ROC delle performance media delle cinque reiterazioni di calcolo del modello per gli incendi forestali nell'area 1.



**Figura 5.23:** Rappresentazione grafica dei risultati di ricampionamento JRR delle covariate nell'area 1: in rosso sono riportati i valori di AUC registrati per modelli che considerano una sola variabile alla volta, in turchese, i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata.

Come per le precedenti applicazioni dell'algoritmo di massima entropia, sono riportati i grafici della variazione media del rischio per gli incendi in formato logistico, al variare delle covariate maggiormente informative, in figura

## Capitolo 5. Applicazione a scala locale in Bolivia

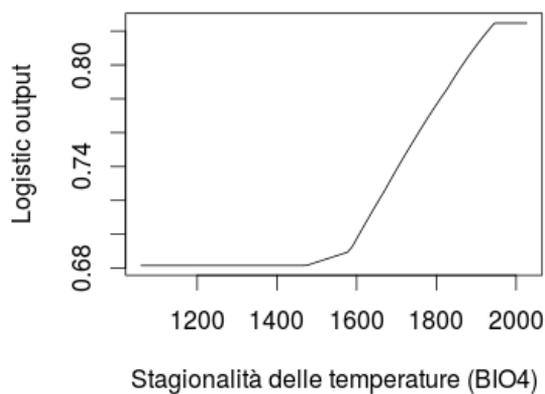
**Tabella 5.7:** *Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili utilizzate per il modello di distribuzione del rischio incendio nell'area 1.*

Variabile	CP(%)	IPP(%)
USO.TIERRA	57,29	28,92
BIO2	18,93	15,39
VEGET.NAVARRO	17,14	22,43
BIO4	1,78	18,07
BIO7	1,73	9,21
SLOPE	0,97	2,48
BIO18	0,88	0,69
BIO19	0,65	0,55
BIO16	0,29	2,08
BIO1	0,16	0,2
BIO17	0,11	0
DEM	0,07	0
BIO5	0	0

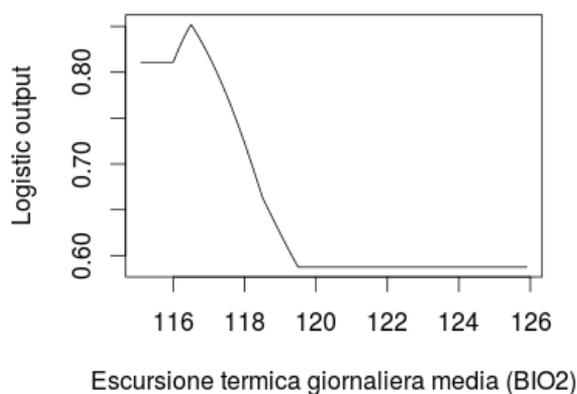
5.24 e figura 5.25, e le categorie di uso del suolo e tipologia di vegetazione più significative in termini di rischio (tabella 5.8).

**Tabella 5.8:** *Valori in formato logistico del rischio incendio delle categorie più significative (valore  $\geq 0,6$ ) nelle variabili non continue per il modello elaborato per l'area 1.*

Variabile	Categoria	Valore di rischio
USO.TIERRA	Pascoli estensivi	0,76
VEGET.NAVARRO	Paludi del Beni Est	0,81
	Savana erbacea mesotrofica	0,75
	Vegetazione acquatica e palustre neotropica della Chiquitania e del Beni	0,65
	Bosco sub-umido semideciduo della Chiquitania	0,63
	Boschi period. allagati del Sud-Est e del centro sud dell'Amazonia	0,61



**Figura 5.24:** *Rischio di incendio in relazione alla stagionalità delle temperature per l'area 1 (deviazione standard delle temperature annuali in gradi Celsius moltiplicata per 100).*

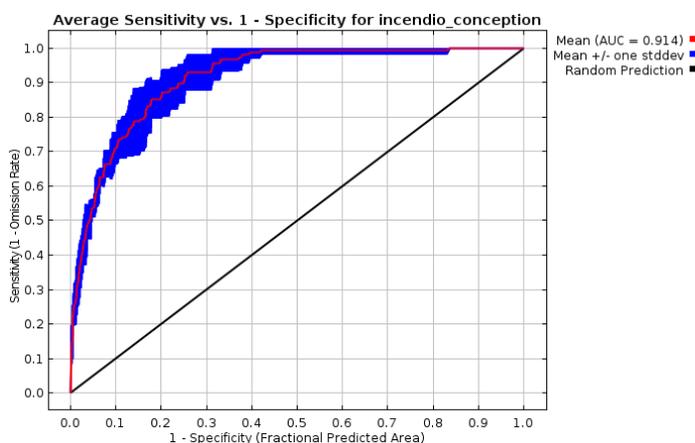


**Figura 5.25:** *Rischio di incendio in relazione all'escursione termica media giornaliera in area 1 (le temperature sono espresse in gradi Celsius e moltiplicati per 10).*

### 5.2.4 Area 2

La seconda area è situata sempre nella parte Nord del dipartimento di Santa Cruz, in corrispondenza della municipalità di Concepción. Sono stati impiegati 31 punti di presenza per il *training*, e 13 per il *test*, per un totale di 10031 punti per determinare la distribuzione (*training* + *background*). Per quest'area le combinazioni di *feature* selezionate dall'algoritmo sono state: lineare, quadratica e *hinge*.

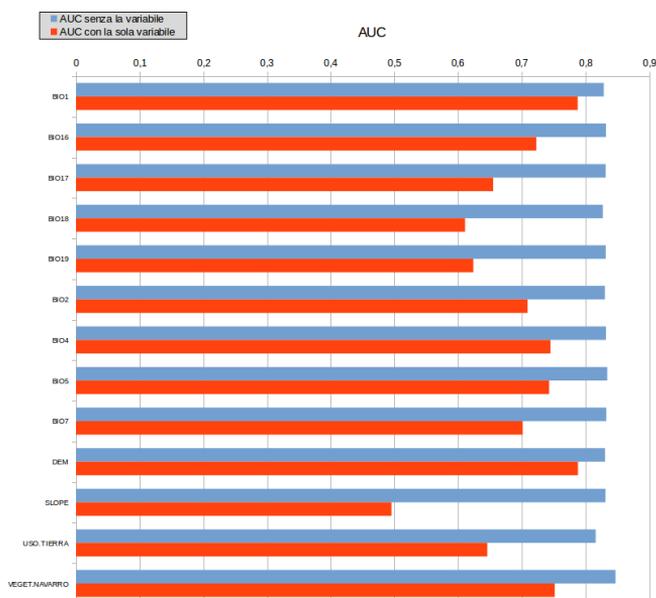
La stima della distribuzione dei rischio per gli incendi forestali ha registrato un valore medio di  $AUC = 0,914$ , con deviazione standard pari a  $s = 0,016$ , rappresentata in figura 5.26.



**Figura 5.26:** Curva ROC delle performance medie delle cinque reiterazioni di calcolo del modello per gli incendi forestali nell'area 2.

Dal grafico in figura 5.27 si può notare come, anche in questo caso, l'esclusione della componente relativa alle categorie di uso del suolo riduca l'accuratezza del modello, registrando il valore più basso di AUC. Analizzando, invece, le prestazioni dei modelli calcolati sulle singole variabili, si nota come la maggiore accuratezza si ottenga con il DEM, che registra un AUC di 0,788.

Risultati simili sono evidenziati dai valori stimati per il contributo e l'importanza di permutazione percentuali, dove le categorie di uso del suolo rappresentano la variabile che contribuisce maggiormente ( $CP = 31,64$  e  $IPP = 21,85$ ), seguita dal tipo di vegetazione, le temperature medie annuali, la quota altimetrica e la stagionalità delle temperature, i cui valori sono riepilogati in tabella 5.9.

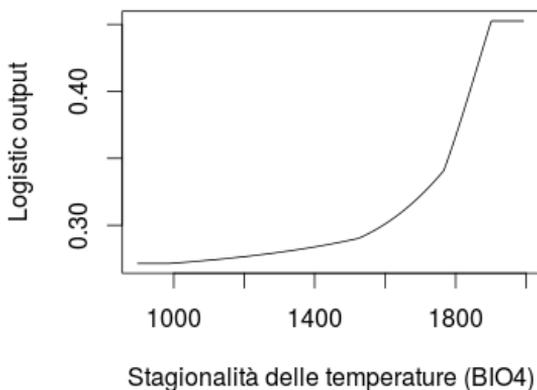


**Figura 5.27:** Rappresentazione grafica dei risultati di ricampionamento JRR delle covariate nell'area 2: in rosso sono riportati i valori di AUC registrati per modelli che considerano una sola variabile alla volta, in turchese, i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata.

**Tabella 5.9:** Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili utilizzate per il modello di distribuzione del rischio incendio nell'area 2.

Variabile	CP(%)	IPP(%)
USO.TIERRA	31,65	21,85
VEGETA.NAVARRO	22,05	18,05
BIO1	14,25	21,79
DEM	12,96	10,93
BIO4	9,73	10,15
BIO18	2,73	2,17
BIO5	2,44	1,27
SLOPE	2,04	5,78
BIO2	0,86	4,73
BIO16	0,61	1,11
BIO17	0,54	0,87
BIO7	0,12	0
BIO19	0,03	1,31

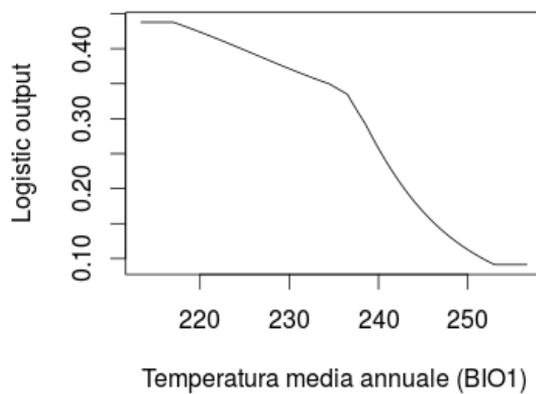
Sono di seguito riportati i grafici della variazione media del rischio per gli incendi in formato logistico, al variare delle covariate maggiormente informative (da figura 5.28 a figura 5.30), e le categorie di uso del suolo e tipologia di vegetazione più significative in termini di rischio (tabella 5.10).



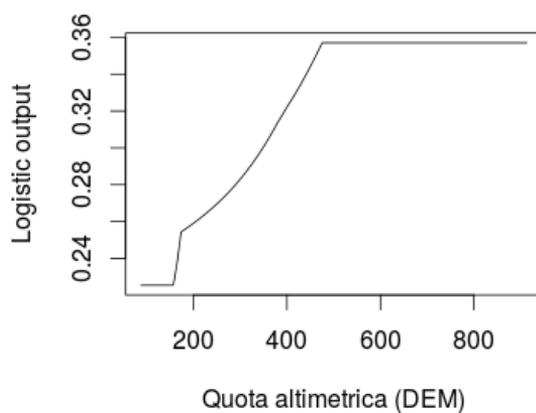
**Figura 5.28:** *Rischio di incendio in relazione alla stagionalità delle temperature per l'area 2 (deviazione standard delle temperature annuali in gradi Celsius moltiplicata per 100).*

**Tabella 5.10:** *Valori in formato logistico del rischio incendio delle categorie più significative (valore  $\geq 0,6$ ) nelle variabili non continue per il modello elaborato per l'area 2.*

Variabile	Categoria	Valore di rischio
USO.TIERRA	Pascoli estensivi	0,77
	Agropastorale estensivo	0,67
VEGET.NAVARRO	praterie paldose delle pianure alluvionali dell'alta Amazzonia	0,84
	Boschi di transizione del Nord dal Chaco alla Chiquitania	0,67



**Figura 5.29:** *Rischio di incendio in relazione alle temperature medie annuali per l'area 2 (le temperature sono espresse in gradi Celsius e moltiplicate per 10).*



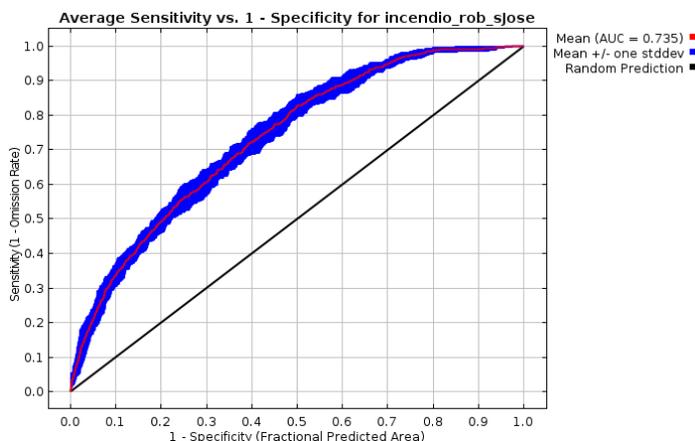
**Figura 5.30:** *Rischio di incendio in relazione alla quota altimetrica per l'area 2 (m s.l.m.).*

### 5.2.5 Area 3

L'ultima area, utilizzata per testare le prestazioni dell'algoritmo di massima entropia in funzione della scala geografica, è situata nella parte centrale del dipartimento di Santa Cruz, in corrispondenza delle due municipalità di Roborè e di San Jòsè de Chiquitos.

Sono stati impiegati 174 punti di presenza per il *training*, e 74 per il *test*, per un totale di 10174 punti per determinare la distribuzione (*training + background*). Per quest'area le combinazioni di *feature* selezionate dall'algoritmo sono state: prodotto, lineare, quadratica, *hinge* e di soglia.

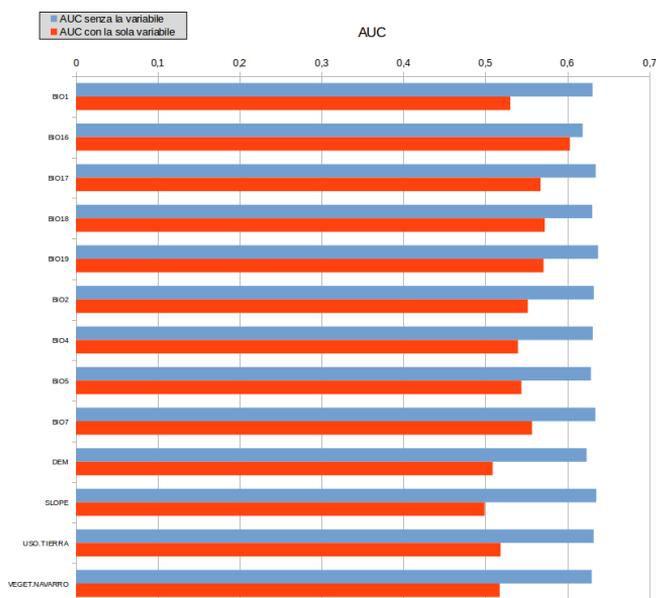
La stima della distribuzione del rischio per gli incendi forestali risulta meno accurata rispetto alle altre due aree considerate, registrando un valore medio di  $AUC = 0,735$ , con un deviazione standard pari a  $s = 0,015$ , rappresentati in figura 5.31.



**Figura 5.31:** Curva ROC delle performance media delle cinque reiterazioni di calcolo del modello per gli incendi forestali nell'area 3.

Contrariamente ai risultati emersi per le precedenti due aree, in questa zona, le variabili pluviometriche appaiono maggiormente determinanti per la definizione della distribuzione del rischio degli incendi. Dalla procedura di ricampionamento JRR riportata graficamente in figura 5.32, infatti, le precipitazioni dei tre mesi più piovosi (BIO16) registrano sia il maggiore valore di AUC (0,6), se considerata come unica covariata del modello, sia il valore più basso (0,62) nel modello in cui si considerano tutte le variabili tranne questa.

Anche le stime del CP e di IPP, riportate in tabella 5.11, evidenziano la marcata informatività delle precipitazioni nei tre mesi più piovosi e nei tre mesi più freddi (BIO19) in quest'area. A queste seguono, in termini di valore percentuale di contributo decrescente, le categorie dei tipi vegetazionali,



**Figura 5.32:** Rappresentazione grafica dei risultati di ricampionamento JRR delle covariate nell'area 2: in rosso sono riportati i valori di AUC registrati per modelli che considerano una sola variabile alla volta, in turchese, i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata.

l'escursione termica annuale, la quota altimetrica (con  $IPP = 15,78$ ) e le categorie di uso del suolo.

**Tabella 5.11:** Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili utilizzate per il modello di distribuzione del rischio incendio nell'area 3.

Variabile	CP(%)	IPP(%)
BIO16	32,18	29,02
BIO19	13,71	8,55
VEGET.NAVARRO	9,03	5,89
BIO7	8,22	6,48
DEM	7,54	15,78
USO.TIERRA	6,01	2,89
BIO5	4,88	3,52
BIO1	4,64	5,62
BIO2	3,79	1,86
BIO18	3,52	8,09
SLOPE	2,86	1,35
BIO4	2,44	3,09
BIO17	1,18	7,86

Sono infine riportati i grafici della variazione media del rischio, in formato logistico, per gli incendi al variare delle covariate maggiormente informative

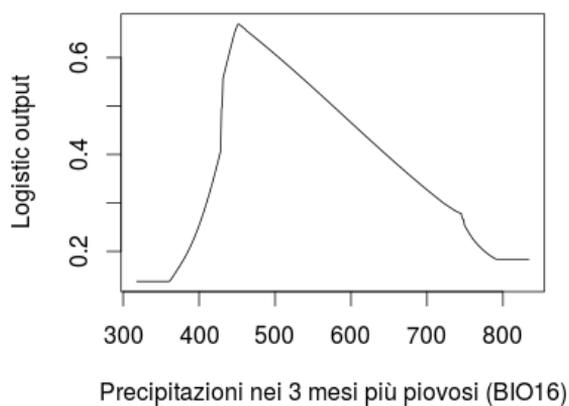
## Capitolo 5. Applicazione a scala locale in Bolivia

---

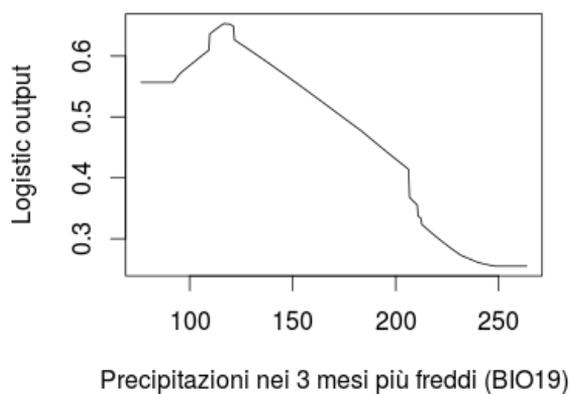
(da figura 5.33 a figura 5.36), e le categorie di uso del suolo e tipologia di vegetazione più significative in termini di rischio (tabella 5.12).

**Tabella 5.12:** Valori in formato logistico del rischio incendio delle categorie più significative (valore  $\geq 0,6$ ) nelle variabili non continue per il modello elaborato per l'area 3.

Variabile	Categoria	Valore di rischio
USO.TIERRA	Silvopastorali a bassa densità di capi	0,62
VEGET.NAVARRO	Savana idrofittica	0,66



**Figura 5.33:** *Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) dei tre mesi più piovosi nell'area 3.*



**Figura 5.34:** *Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) dei tre mesi più freddi per l'area 3.*

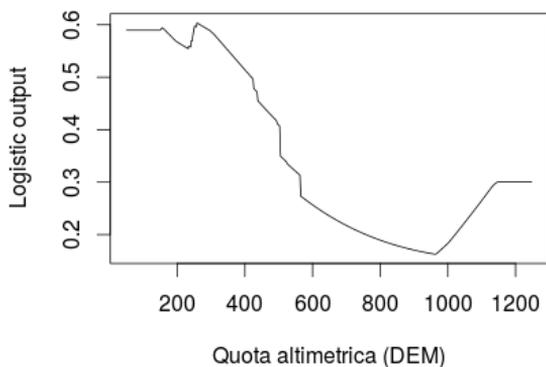


Figura 5.35: Rischio di incendio in relazione alla quota altimetrica nell'area 3 (m s.l.m.).

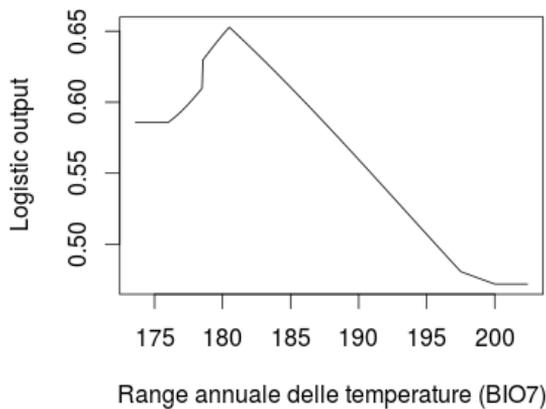


Figura 5.36: Rischio di incendio in relazione al range annuale di temperatura per l'area 3 (le temperature sono espresse in gradi Celsius e moltiplicate per 10).

Le relazioni tra la piovosità e il rischio di incendio, nei modelli appena descritti, sembrano avere andamenti contrastanti rispetto a ciò che tendenzialmente si è portati a pensare, ovvero che con l'aumentare delle precipitazioni il rischio sia più basso. Nelle aree dell'Amazonia boliviana indagate, l'incremento del rischio di incendio all'aumentare delle precipitazioni in diversi periodi dell'anno è rappresentativo delle particolari condizioni che caratterizzano la struttura della vegetazione nelle savane tropicali. Le savane tropicali e subtropicali sono determinate principalmente dalla scarsità e dalla marcata stagionalità delle precipitazioni. Nel periodo delle piogge il terreno rimane allagato fino all'inizio della stagione secca, quando asciugandosi permette la rapida crescita di specie erbacee e arbusti. Durante il successivo periodo di siccità, il manto erboso secco e continuo crea le condizioni ideali per lo sviluppo degli incendi. In questo modo, dove le precipitazioni sono più elevate si creerà uno strato di combustibile fine continuo e più abbondante, che nel periodo secco fa aumentare il rischio di incendio. Anche le precipitazioni nei mesi più secchi, sebbene molto irregolari, agiscono in maniera analoga, favorendo lo sviluppo degli strati erbacei che seccano rapidamente aumentando la quantità di combustibile.

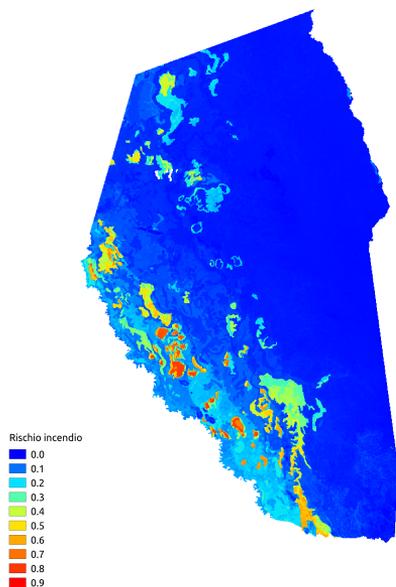


Figura 5.37: Distribuzione media del rischio di incendi nell'Area 1.

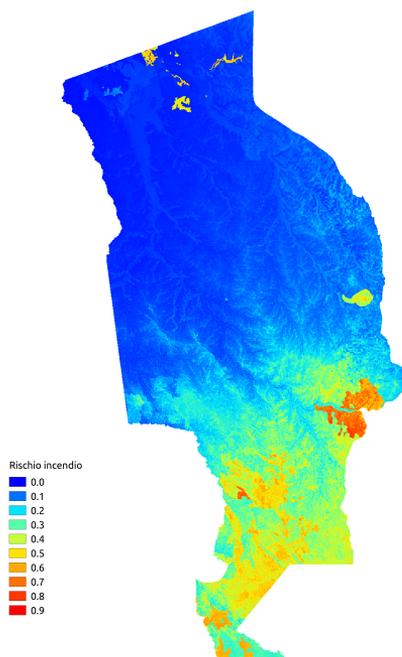
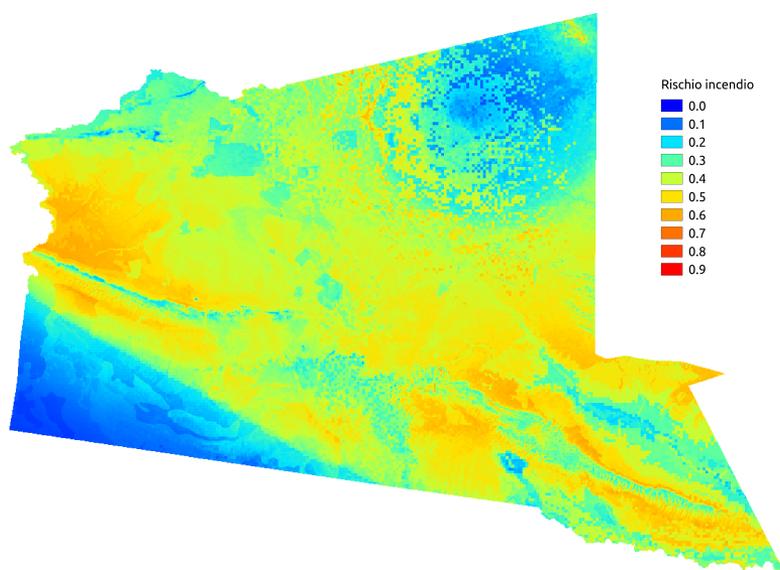


Figura 5.38: Distribuzione media del rischio di incendi nell'Area 2.



**Figura 5.39:** *Distribuzione media del rischio di incendi nell'Area 3.*



---

### Elaborazione su scala regionale: Valle d'Aosta

---

Gli incendi boschivi hanno sempre rappresentato un pericolo per le foreste della Valle d'Aosta ed è ancora vivo in molte persone il ricordo dei disastrosi eventi del 1965, del 1973 e del 1981. Negli anni '70 e '80 si è registrata una tendenza all'aumento del numero di incendi, tanto che nel 1989 se ne sono contati circa 370. Nel 1990 sono stati registrati 362 eventi con una superficie bruciata che superava i 1200 ettari di territorio e una distruzione di soprassuolo boschivo di circa 150 ettari. I fattori predisponenti e determinanti che concorrono all'innescio degli incendi boschivi sono molteplici, dipendendo dalla morfologia del territorio, dal clima, dalle caratteristiche della vegetazione e dalle attività umane.

La dipendenza da tutte queste variabili, spesso in relazione l'una con le altre, rende il processo di modellizzazione degli incendi suscettibile ai problemi derivati dalla colinearità. Come evidenziato da Dormann [18], la colinearità descrive la situazione in cui due o più covariate<sup>1</sup>, in un modello, hanno una relazione lineare. Questo, legato al fatto che molti approcci statistici, utilizzati in ecologia, sono spesso sensibili alla colinearità [6], può condurre a stime dei parametri instabili, con errori standard delle stime eccessivi, che inducono di conseguenza a statistiche inferenziali deviate. Anche per i metodi meno sensibili, sorgono in ogni caso due problemi in concomitanza con la colinearità: gli effetti delle variabili non possono essere separati e l'estrapolazione che ne deriva sarà, molto probabilmente, distorta [50]. In tutti i fenomeni naturali esiste

---

<sup>1</sup>nel caso in cui siano più di due variabili si parla di multi-colinearità

un certo grado di colinearità tra le variabili predittive, questo a causa del fatto che comunemente la dipendenza lineare tra covariate è intrinseca, ovvero che le variabili colineari sono differenti manifestazioni delle stesse condizioni (si pensi ad esempio, alle relazioni tra le variabili termiche, quelle pluviometriche e la quota altimetrica).

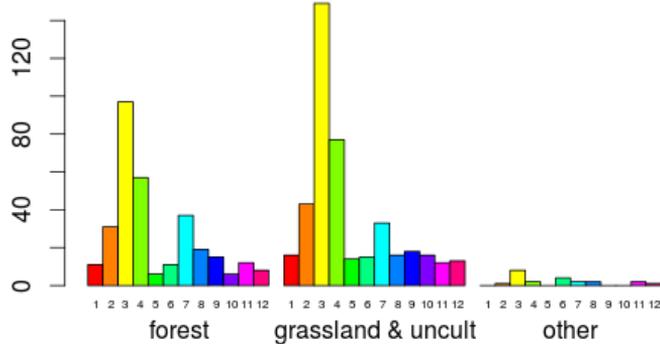
Uno dei metodi più comuni per rimuovere le correlazioni in un set di variabili prevede l'analisi delle componenti principali (PCA) [18]. Questo metodo può essere applicato alle variabili continue producendone la riproiezione su assi ortogonali e di conseguenza perfettamente non correlati. È quindi possibile sfruttare questa caratteristica per analizzare i raggruppamenti di variabili maggiormente correlate tra loro e procedere alla selezione, per ogni componente principale (PC), di una sola variabile per gruppo, eliminando, o comunque limitando fortemente, il problema della colinearità.

Il caso di studio ha interessato il territorio della Regione Autonoma Valle d'Aosta. Sono stati analizzati 15 anni di incendi e sono stati presi in considerazione gli strati informativi che meglio caratterizzano i fenomeni di innesco per gli incendi nei diversi periodi dell'anno, selezionati attraverso l'analisi delle componenti principali.

### 6.1 Premesse metodologiche

---

L'archivio degli inneschi è stato estratto dalla banca dati geografica della Regione Autonoma Valle d'Aosta, prendendo in esame gli eventi registrati dal gennaio 1995 al dicembre 2009, per un totale di 1118 incendi boschivi. Dalla analisi di frequenza degli inneschi in relazione ai mesi dell'anno e alla tipologia di vegetazione prevalente interessata dagli incendi si è ritenuto opportuno, viste le finalità dello studio, suddividere l'archivio in 3 dataset. Si può notare infatti che la maggiore frequenza dei fenomeni di incendio si registra in corrispondenza dei primi mesi dell'anno, con il massimo nel mese di marzo, riportato in giallo nella figura 6.1, sia per quanto riguarda le superfici a prevalente copertura boscata, sia per quelle a prevalenza di prati e pascoli. Secondariamente si nota come l'altro picco di frequenza si registri in corrispondenza del mese di luglio (in azzurro nel grafico).



**Figura 6.1:** *Frequenze mensili cumulate per il periodo di indagine degli incendi, suddivise per tipologia di vegetazione in corrispondenza del punto di innesco. Con "forest" per i soprasuoli boschivi; "grassland&uncult" per le aree a pascolo e gli incolti; "other" per le altre tipologie di vegetazione, prevalentemente ad uso agricolo.*

Sono stati così analizzati separatamente:

- i) 246 incendi estivi (E), registrati da maggio a settembre di ogni anno;
- ii) 316 incendi invernali in bosco (IB), registrati da ottobre ad aprile di ogni anno, il cui punto di innesco registrato nell'archivio regionale ricade sulle superfici a prevalente copertura forestale;
- iii) 556 incendi invernali su prati-pascoli ed incolti (PP) registrati da ottobre ad aprile di ogni anno, il cui punto di innesco ricade sulle superfici in cui la copertura prevalente è costituita da prati e pascoli o aree a vegetazione incolta.

I predittori considerati per la caratterizzazione della distribuzione del rischio di incendi sono riportati in tabella 6.1 e sono di seguito descritti. Le variabili climatiche sono state ricavate dagli strati informativi disponibili nel sito web worldclim, per la cui metodologia di elaborazione si rimanda a [34].

La radiazione solare incidente è stata valutata attraverso il calcolo, in ambiente GIS, del HeatLoad Index [49]. La quota altimetrica è stata rappresentata attraverso il modello digitale di elevazione (DEM), dal quale è stato possibile calcolare le pendenze. Dal DEM è stato inoltre calcolato il Southwestness index [26] per valutare la propensione alla xericità del territorio.

## Capitolo 6. Elaborazione su scala regionale: Valle d'Aosta

Lo strato informativo relativo all'uso del suolo e alla vegetazione è stato ricavato dai dati raccolti nel progetto Corine Land Cover all'anno 2006, distinguendo cinque macro-categorie rappresentative della tipologia di combustibile: aree urbane e naturali con vegetazione scarsa o assente; corpi idrici e ghiacciai; prati, pascoli e brughiere; aree agricole; aree forestali.

La componente antropica è stata considerata includendo, tra le variabili, la distanza del punto di innesco degli incendi dagli edifici e dalla rete viaria, disponibili nel database OpenStreetMap. Inoltre, per indagare le relazioni esistenti tra il rischio di innesco e la realtà socio-economica del territorio, sono stati inseriti i dati relativi al numero di capi al pascolo e alla consistenza numerica delle aziende zootecniche con animali al pascolo, ricavati dal più recente censimento ISTAT dell'agricoltura (2010). Il dato relativo alla consistenza numerica delle aziende, in rapporto al numero complessivo di capi, rispecchia il livello di sfruttamento del territorio in termini di concentrazione del carico di pascolo: è noto infatti che situazioni di forte carico di pascolo comportano un contenimento della biomassa combustibile [9].

**Tabella 6.1:** *Elenco delle variabili considerate per l'indagine della distribuzione del rischio incendio in Valle d'Aosta.*

Variabile	Descrizione
BIO1	Temperatura media annuale
BIO5	Massima temperatura nel mese più caldo
BIO6	Minima temperatura nel mese più freddo
BIO8	Temperatura media del quarto più piovoso
BIO9	Temperatura media del quarto più siccitoso
BIO10	Temperatura media del quarto più caldo
BIO11	Temperatura media del quarto più freddo
BIO12	Precipitazioni annuali
BIO13	Precipitazioni nel mese più piovoso
BIO14	Precipitazioni nel mese più siccitoso
BIO16	Precipitazioni del quarto più piovoso
BIO17	Precipitazioni del quarto più siccitoso
BIO18	Precipitazioni del quarto più caldo
BIO19	Precipitazioni del quarto più freddo
DEM	Modello digitale del terreno
DIST.INFR	Raster di distanza dalle strutture e infrastrutture antropiche
Heatload Index	Radiazione solare incidente( $HLI = \frac{1 - (\cos \text{esposizione} - 45)}{2}$ )
N.FARMS	Aziende con animali al pascolo dal censimento dell'agricoltura ISTAT 2010
N.GRAZ.ANIMALS	Numero di animali al pascolo dal censimento dell'agricoltura ISTAT 2010
SLOPE	Pendenza in gradi calcolata dal DEM
Southwestness Index	$SWI = \cos(\text{esposizione} - 225)$
VEGETATION	Tipologia di vegetazione da Corine Land Cover

### 6.1.1 Preparazione dei dati

Le variabili sono state riportate in formato raster alla risoluzione di 900 m<sup>2</sup>, su grid a maglia quadrata con lato di 30 m. Vista la natura aggregata, su base comunale, del dato ricavato dal censimento dell'agricoltura di ISTAT, sia il dato

## 6.2. Scelta delle variabili: analisi delle componenti principali

numerico delle aziende, sia il numero dei capi al pascolo sono stati rapportati alla superficie potenzialmente pascolabile. Per ogni comune, sono state calcolate la densità di aziende e la densità di capi sulle superfici delle categorie di uso del suolo compatibili con l'attività di pascolo; queste, infine, sono state moltiplicate per la risoluzione del raster, per riportare i valori di densità per cella, e attribuire il valore risultante alle sole celle corrispondenti ad aree a pascolo (riportate con le sigle D.GRAZ.ANIM e D.FARMS). In questo modo il dato aggregato fornito da ISTAT è stato ridistribuito sul territorio, permettendo di evidenziare la maggiore o minore concentrazione di attività legate alle sole superfici effettive sulle quali si concentrano le attività di pascolo e in cui il carico può influenzare la quantità di combustibile disponibile per gli incendi.

## 6.2 Scelta delle variabili: analisi delle componenti principali

L'utilizzo della PCA per la selezione delle variabili, in modo da ridurre il più possibile la colinearità, si basa sullo studio progressivo dei *loadings* delle variabili sui singoli assi delle componenti principali. Secondo quanto sostenuto da Booth [8], le variabili con *loadings* assoluti superiori a 0,32 formano i raggruppamenti "*proxy*" o gruppi di interesse all'interno dei quali le variabili sono correlate tra loro, come mostrato in figura 6.2, dove in ordinata sono riportati i valori di loading delle variabili, la soglia di 0,32 è rappresentata dalla retta parallela all'asse delle ascisse al di sopra della quale si può notare il gruppo delle 4 variabili correlate tra loro.

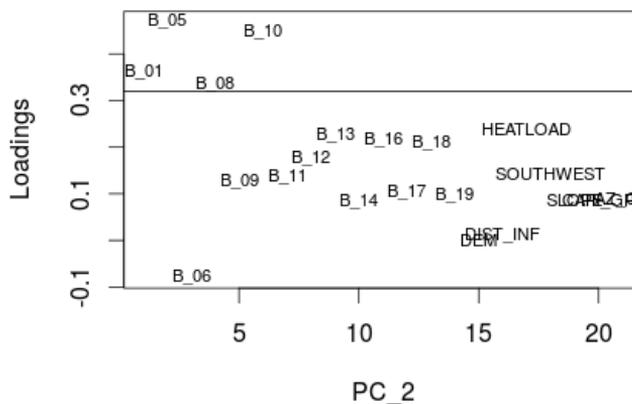


Figura 6.2: Esempio di rappresentazione grafica di un gruppo proxy per la PC 2.

## Capitolo 6. Elaborazione su scala regionale: Valle d'Aosta

La soglia del valore di 0,32 è determinata dal fatto che questa rappresenta il 10% della varianza della variabile spiegata dall'asse della PC. Dai gruppi *proxy* individuati per ogni PC, riportati in tabella 6.2, sono state selezionate le variabili con *loadings* assoluti maggiori e sono state eliminate tutte le altre.

**Tabella 6.2:** *Gruppi proxy identificati dalla analisi delle componenti principali in Valle d'Aosta.*

PC	Variabile	Loading
2	BIO8	0,335
	BIO1	0,359
	BIO10	0,447
	BIO5	0,469
3	D.GRAZ.ANIM	0,691
	D.FARMS	0,691
4	Heatload Index	-0,637
	Southwestness Index	-0,690
5	SLOPE	-0,582
	DIST.INFR	-0,770
11	BIO17	-0,362
	BIO11	-0,409
	BIO19	-0,448
	BIO6	-0,475
13	BIO17	-0,327
	BIO18	-0,725

Ad ulteriore conferma del grado di correlazione delle variabili evidenziate nei gruppi *proxy* sono riportati i valori dalle matrici di correlazione<sup>2</sup> calcolate per alcune variabili individuate nei gruppi delle componenti principali riportate nelle tabelle da 6.3 a 6.6.

**Tabella 6.3:** *Matrice di correlazione della componente principale PC 2 in Valle d'Aosta.*

	BIO1	BIO5	BIO8	BIO10
BIO1	1.0000000	0.9429162	0.8714225	0.9688531
BIO5	0.9429162	1.0000000	0.8403732	0.9944274
BIO8	0.8714225	0.8403732	1.0000000	0.8480551
BIO10	0.9688531	0.9944274	0.8480551	1.0000000

**Tabella 6.4:** *Matrice di correlazione della componente principale PC 3 in Valle d'Aosta.*

	D.GRAZ.ANIM	D.FARMS
D.GRAZ.ANIM	1.0000000	0.9925604
D.FARMS	0.9925604	1.0000000

<sup>2</sup>Per il calcolo delle matrici di correlazione, visto che le grandezze delle variabili, in alcuni casi, sono profondamente differenti (ad esempio, temperature e precipitazioni), sono stati utilizzati i valori standardizzati delle covariate. La standardizzazione è la trasformazione dei valori  $x_{ij}$  nei valori  $z_{ij}$  definiti da  $z_{ij} = \frac{x_{ij} - (\bar{x})_j}{\sigma_j}$  dove  $(\bar{x})_j$  è la media della variabile  $x_j$  e  $\sigma_j$  la deviazione standard campionaria della stessa variabile.

**Tabella 6.5:** Matrice di correlazione della componente principale PC 4 in Valle d'Aosta.

	HEATLOAD	SOUTHWEST
HEATLOAD	1.0000000	0.9806511
SOUTHWEST	0.9806511	1.0000000

**Tabella 6.6:** Matrice di correlazione della componente principale PC 11 in Valle d'Aosta.

	BIO6	BIO11	BIO17	BIO19
BIO6	1.0000000	0.9112132	-0.9914014	-0.9899902
BIO11	0.9112132	1.0000000	-0.8922469	-0.8931962
BIO17	-0.9914014	-0.8922469	1.0000000	0.9975549
BIO19	-0.9899902	-0.8931962	0.9975549	1.0000000

## 6.3 Elaborazione

I modelli sono stati prodotti considerando il 70% dei punti dei dataset e validati utilizzando il restante 30% come test indipendente, selezionato tramite ricampionamento *bootstrap* su 5 reiterazioni di calcolo per ciascun dataset. Le variabili selezionate attraverso la valutazione dei gruppi *proxy* sugli assi delle componenti principali sono riepilogate in tabella 6.7.

**Tabella 6.7:** Elenco delle variabili selezionate con PCA per l'indagine della distribuzione del rischio incendio in Valle d'Aosta.

Variabile	Descrizione
BIO5	Massima temperatura nel mese più caldo
BIO6	Minima temperatura nel mese più freddo
BIO9	Temperatura media del quarto più siccitoso
BIO12	Precipitazioni annuali
BIO13	Precipitazioni nel mese più piovoso
BIO14	Precipitazioni nel mese più siccitoso
BIO16	Precipitazioni del quarto più piovoso
BIO18	Precipitazioni del quarto più caldo
DEM	Modello digitale del terreno
DIST.INFR	Raster di distanza dalle strutture e infrastrutture antropiche
D.GRAZ.ANIMALS	Densità di animali al pascolo per cella raster dal censimento dell'agricoltura ISTAT 2010
Southwestness Index	$SWI = \cos(\text{esposizione} - 225)$
VEGETATION	Tipologia di vegetazione da Corine Land Cover

I parametri utilizzati per la stima della distribuzione del rischio incendio, nei tre dataset considerati sono riportati, in dettaglio, nella tabella 6.8. Come si può notare, tali parametri sono volutamente comuni a tutti i modelli precedentemente descritti, questo per favorirne la comparabilità. In questo caso di studio, l'unica eccezione è stata fatta per il coefficiente di regolarizzazione, impostato a 1,5 invece che a 2. Il coefficiente di regolarizzazione interviene sulla determinazione dei limiti di errore ammissibili dei vicoli imposti dalle combinazioni di *feature*  $\lambda_j$ , descritti nel secondo capitolo (equazione 2.11), come moltiplicatore di  $\lambda$ . In questo caso, il valore più basso rispetto agli altri

## Capitolo 6. Elaborazione su scala regionale: Valle d'Aosta

---

modelli è stato scelto a causa della maggiore risoluzione spaziale delle covariate, e della conseguente necessità di limitare la generalizzazione dei modelli, evidenziando le differenze per i tre dataset.

**Tabella 6.8:** Parametri impostati per i modelli di distribuzione del rischio incendio nei tre dataset della Valle d'Aosta: "dataset E" per gli incendi estivi, "dataset IB" per dli incendi invernali in aree boscate e "dataset PP" per gli incendi invernali su aree a prevalente copertura di prati, pascoli e incolti.

Parametro	Valore / modalità
Punti training dataset E	180
Punti test dataset E	76
Punti training dataset IB	222
Punti test dataset IB	94
Punti training dataset PP	390
Punti test dataset PP	166
Valore di regolarizzazione feature lineare	0.05
Valore di regolarizzazione feature quadratica	0.05
Valore di regolarizzazione feature prodotto	0.05
Valore di regolarizzazione feature categorica	0.25
Valore di regolarizzazione feature soglia	1
Valore di regolarizzazione feature hinge	0.5
Combinazione di feature usata	prodotto - lineare - quadratica - hinge - soglia
Coeff. di regolarizzazione	1,5
Repliche	5
Ricampionamento	Bootstrap

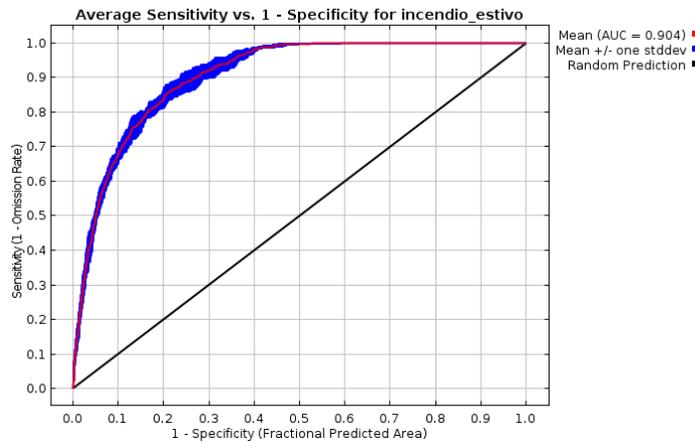
## 6.4 Risultati

---

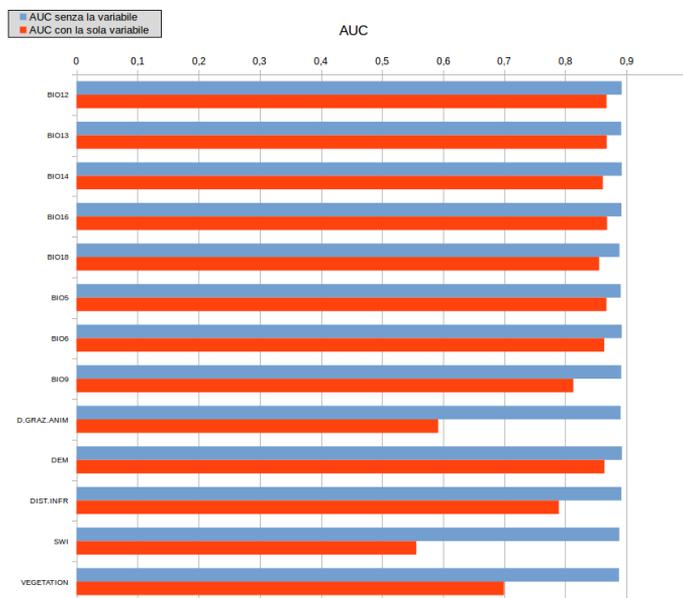
### 6.4.1 Incendi estivi

Il modello di distribuzione del rischio per gli incendi estivi ha registrato un valore di  $AUC = 0,904$ , con deviazione standard  $s = 0,008$  tra le cinque repliche. La curva ROC generata per la valutazione delle *performance* dell'elaborazione è riportata in figura 6.3. I risultati delle prestazioni dei modelli calcolati a seguito della procedura di ricampionamento JRR, graficamente riportati in figura 6.4, mostrano che la componente che maggiormente influisce sulla determinazione della distribuzione del rischio di incendi, nei mesi compresi tra maggio e settembre, è la tipologia di vegetazione, che, se esclusa, fa registrare il valore più basso di AUC (0,888). È necessario sottolineare, comunque, che il suddetto valore si scosta di poco dai valori calcolati per i modelli in cui sono alternativamente state escluse le altre covariate, per le quali si registra il massimo di 0,893 per l'esclusione della quota altimetrica.

La stima del contributo e dell'importanza di permutazione percentuali (in tabella 6.9) mostra il massimo valore di contributo della componente legata alle precipitazioni nei tre mesi più piovosi (BIO16), che però risulta essere poco informativa considerando il valore basso di IPP. Questa considerazione



**Figura 6.3:** Curva ROC delle performance medie del modello per gli incendi estivi.



**Figura 6.4:** Rappresentazione grafica di JRR applicato al modello per gli incendi estivi in Valle d'Aosta: in rosso, i valori di AUC registrati per modelli che considerano una sola variabile alla volta (in ordinata), in turchese, i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata.

è avvalorata dai valori diametralmente opposti della variabile di precipitazioni nel mese più piovoso (BIO13 con  $CP = 1,59$  e  $IPP = 33,85$ ). La quota altimetrica contribuisce alla determinazione della distribuzione di rischio per circa il 20%, ma sono le temperature massime del mese più caldo ad essere

## Capitolo 6. Elaborazione su scala regionale: Valle d'Aosta

---

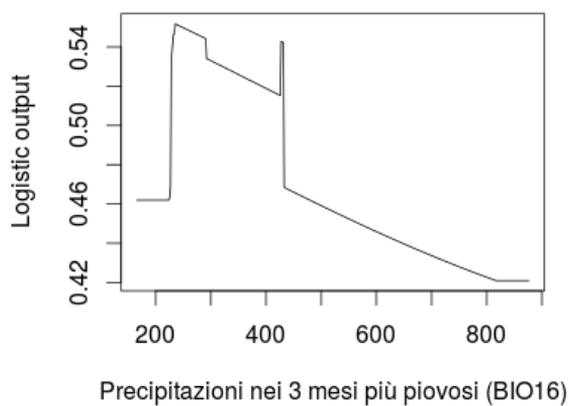
**Tabella 6.9:** *Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili selezionate nel modello di distribuzione del rischio incendio estivo.*

Variabile	CP(%)	IPP(%)
BIO16	29,41	0,76
DEM	19,99	2,78
BIO5	11,29	38,59
BIO6	9,31	1,61
BIO12	7,55	0,01
BIO14	4,48	1,12
SWI	3,47	2,05
BIO18	3,23	9,67
VEGETATION	2,96	3,18
DIST.INFR	2,77	2,58
BIO9	2,14	1,91
D.GRAZ.ANIM	1,80	1,91
BIO13	1,59	33,85

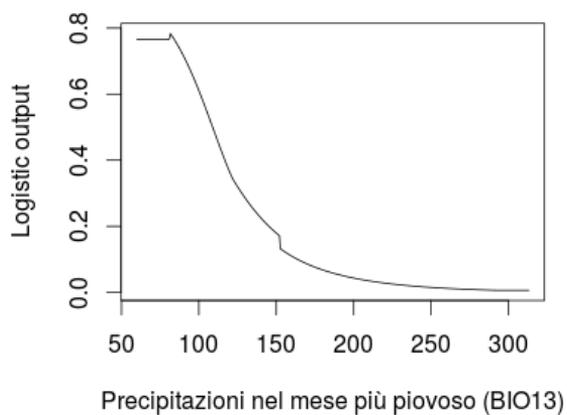
una delle variabili maggiormente informative, a causa del massimo valore di IPP stimato (38, 59).

Per quanto riguarda la tipologia di vegetazione, nonostante faccia registrare valori di CP e IPP bassi, è da sottolineare che la categoria che registra la maggiore suscettibilità agli incendi, seppur scarsamente significativa con un valore logistico di rischio di circa 0, 57, è quella che raggruppa prati, pascoli e brughiera.

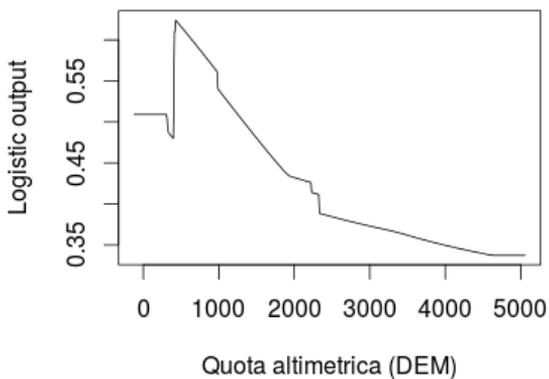
Sono riportate le restituzioni grafiche di variazione media del rischio, in formato logistico, al variare delle covariate maggiormente informative per gli incendi estivi (figure da 6.5 a 6.8).



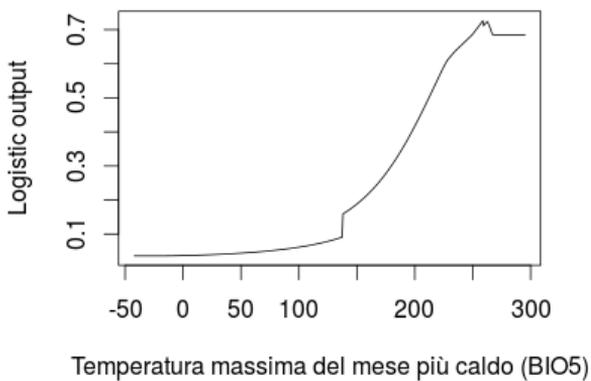
**Figura 6.5:** *Rischio di incendio estivo in relazione alle precipitazioni (in mm di pioggia) nei tre mesi più piovosi.*



**Figura 6.6:** *Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) nel mese più piovoso.*



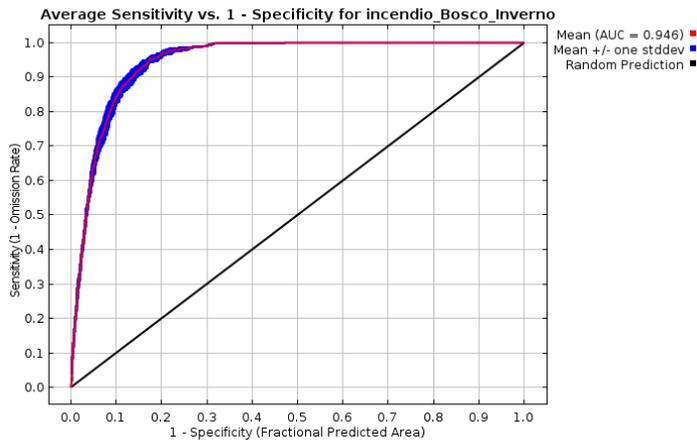
**Figura 6.7:** *Rischio di incendio in relazione alla quota altimetrica per gli incendi estivi (m s.l.m.).*



**Figura 6.8:** *Rischio di incendio in relazione alle temperature massime del mese più caldo (le temperature sono espresse in gradi Celsius e moltiplicate per 10).*

### 6.4.2 Incendi boschivi invernali

Dall'elaborazione del modello di rischio incendio sul dataset relativo agli incendi registrati da ottobre ad aprile di ogni anno, emerge un alto livello di accuratezza, espresso dal valore di  $AUC = 0,946$ , con deviazione standard  $s = 0,003$  tra le cinque repliche (in figura 6.9).

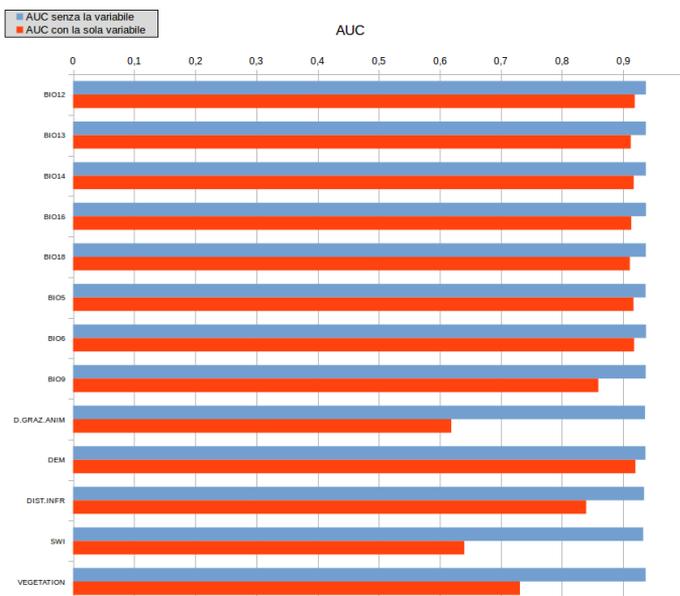


**Figura 6.9:** Curva ROC delle performance medie del modello per gli incendi boschivi invernali.

Dalla stima delle prestazioni dei modelli calcolati a seguito del ricampionamento JRR, riportati nel grafico a barre in figura 6.10 è possibile notare che il valore più basso in termini di AUC (0,932) si registra alla sottrazione delle informazioni contenute nella variabile dell'indice di xericità (*southwestness index*, SWI nel grafico).

Dai valori stimati per il contributo e per l'importanza di permutazione percentuali, riportate in tabella 6.10, emerge l'elevato valore registrato dalla quota altimetrica (DEM), seguito dalle temperature minime per il mese più freddo (BIO6), che ha, però, un basso valore di importanza di permutazione. Valori opposti si registrano per la variabile relativa alle precipitazioni nel mese più siccitoso (BIO14 con  $CP = 5,4$  e  $IPP = 30,6$ ). Le altre due covariate che risultano interessanti dal punto di vista dell'importanza di permutazione sono le temperature dei tre mesi meno piovosi (BIO9) e lo strati informarivo di distanza dalle strutture e infrastrutture antropiche (DIST.INFR), anche se non riportano valori di contributo percentuale rilevanti.

## Capitolo 6. Elaborazione su scala regionale: Valle d'Aosta

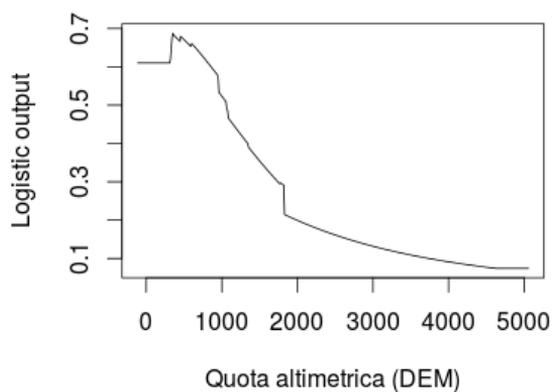


**Figura 6.10:** Rappresentazione grafica di JRR applicato al modello per gli incendi forestali invernali in Valle d'Aosta: in rosso, i valori di AUC registrati per modelli che considerano una sola variabile alla volta (in ordinata), in turchese, i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata.

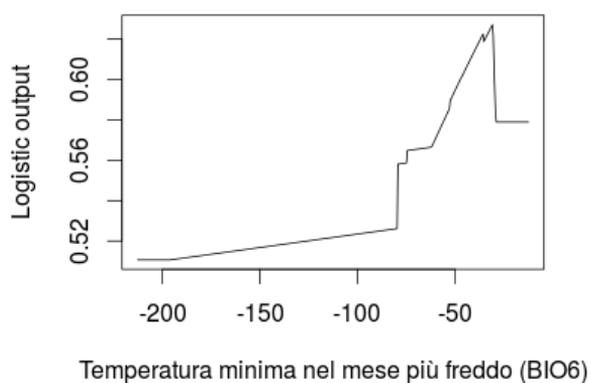
**Tabella 6.10:** Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili selezionate nel modello di distribuzione del rischio incendio boschivo invernale.

Variabile	CP(%)	IPP(%)
DEM	42,99	20,06
BIO6	11,86	0,44
BIO12	11,85	0,39
BIO5	8,42	5,69
BIO14	5,41	30,56
SWI	4,96	6,97
DIST.INFR	4,86	10,59
BIO18	4,49	3,36
BIO13	1,56	5,29
VEGETATION	1,27	2,70
BIO9	0,99	12,59
BIO16	0,91	0,22
D.GRAZ.ANIM	0,44	1,12

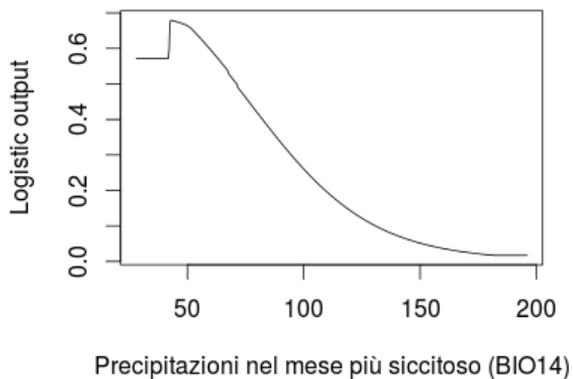
Di seguito sono riportate le restituzioni grafiche di variazione del rischio medio per gli incendi invernali, al variare delle covariate maggiormente informative precedentemente menzionate. Il rischio è riportato in ordinata, in formato logistico (figure da 6.11 a 6.16).



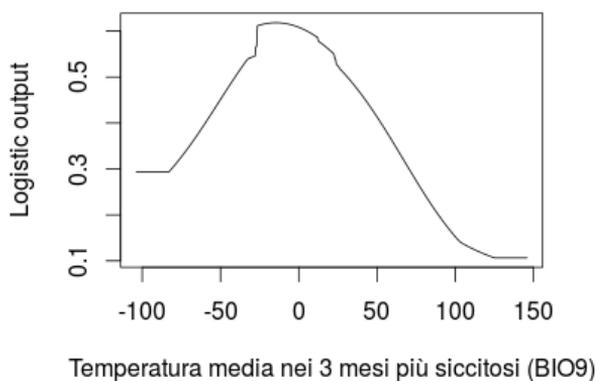
**Figura 6.11:** *Rischio di incendio in relazione alla quota altimetrica per gli incendi invernali (m s.l.m.).*



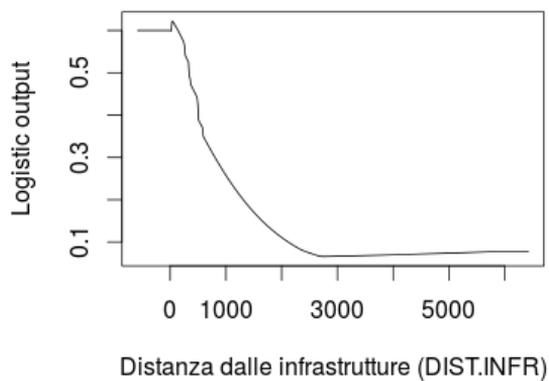
**Figura 6.12:** *Rischio di incendio in relazione alle temperature minime del mese più freddo (le temperature sono espresse in gradi Celsius e moltiplicate per 10).*



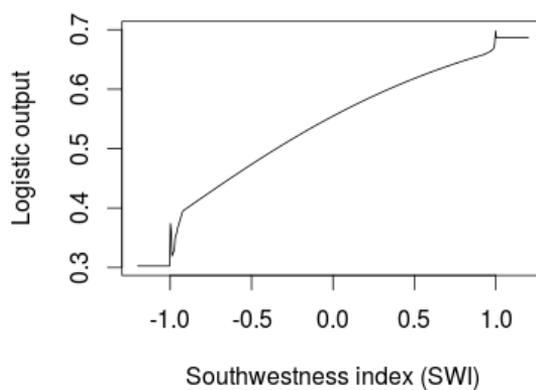
**Figura 6.13:** *Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) nel mese più siccitoso.*



**Figura 6.14:** *Rischio di incendio in relazione alle temperature medie dei tre mesi meno piovosi (le temperature sono espresse in gradi Celsius e moltiplicate per 10).*



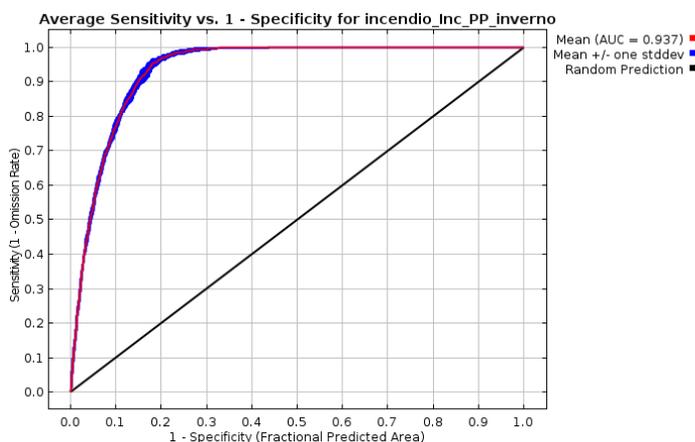
**Figura 6.15:** *Rischio di incendio in relazione alla distanza dalle infrastrutture antropiche (m).*



**Figura 6.16:** *Rischio di incendio in relazione al southwestness index.*

### 6.4.3 Incendi invernali da aree a prevalente copertura di prati e pascoli

Il modello elaborato per il terzo dataset ha fatto registrare una *performance*, in termini di valore di AUC di 0,937, con deviazione standard  $s = 0,002$  tra le reiterazioni di calcolo, riportato in figura 6.17 .



**Figura 6.17:** Curva ROC delle performance medie del modello per gli incendi invernali su prati e pascoli.

Le prestazioni dei modelli calcolati a seguito del ricampionamento JRR, riportati nel grafico a barre in figura 6.18 evidenziano l'importanza della componente antropica: il valore più basso in termini di AUC (0,925) si registra alla sottrazione delle informazioni contenute nella variabile di distanza dalle strutture e infrastrutture antropiche (DIST.INFR nel grafico).

Le stime di contributo e importanza di permutazione percentuali, riportate in tabella 6.11, mostrano gli elevati valori registrati dalla quota altimetrica (DEM), secondariamente, a sostegno di quanto emerso dalle prestazioni dei modelli calcolati nella procedura JRR, i maggiori risultati sono da riferire alla variabile di distanza dalle infrastrutture. In ordine decrescente, in termini di contributo percentuale, poi, si trovano le precipitazioni nel mese più siccitoso (BIO14) e le precipitazioni annuali (BIO12).

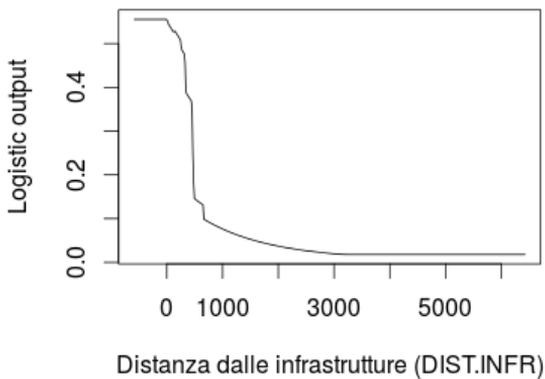


**Figura 6.18:** *Rappresentazione grafica di JRR applicato al modello per gli incendi forestali invernali in Valle d'Aosta: in rosso, i valori di AUC registrati per modelli che considerano una sola variabile alla volta (in ordinata), in turchese, i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata.*

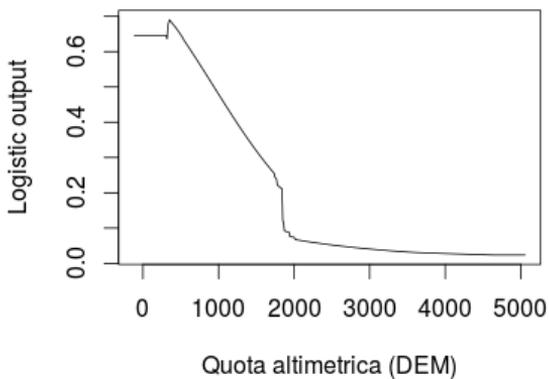
**Tabella 6.11:** *Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili selezionate nel modello di distribuzione del rischio incendio invernale su prati e pascoli.*

Variabile	CP(%)	IPP(%)
DEM	35,51	41,20
DIST.INFR	20,60	28,13
BIO14	10,54	6,35
BIO12	9,66	1,52
BIO16	5,87	1,19
VEGETATION	4,54	4,17
BIO6	3,37	0,23
BIO18	2,46	0,95
SWI	2,28	4,21
BIO5	1,94	2,76
D.GRAZ.ANIM	1,21	1,55
BIO9	1,10	5,95
BIO13	0,93	1,79

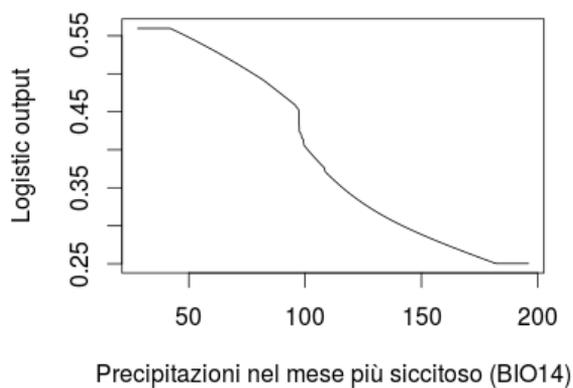
Infine sono riportate le rappresentazioni grafiche (da figura 6.11 a figura 6.16) della variazione media del rischio in formato logistico per gli incendi invernali su prati e pascoli, al variare delle covariate maggiormente informative.



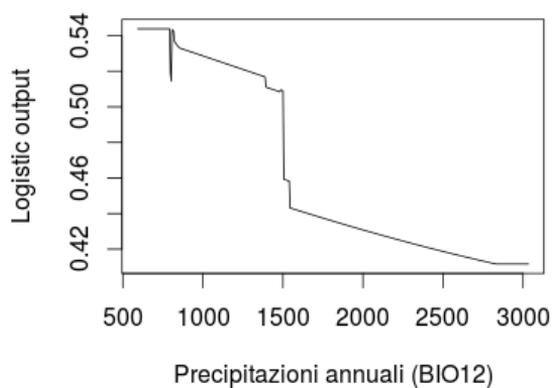
**Figura 6.19:** *Rischio di incendio in relazione alla distanza dalle infrastrutture antropiche (m).*



**Figura 6.20:** *Rischio di incendio in relazione alla quota altimetrica (m s.l.m.).*

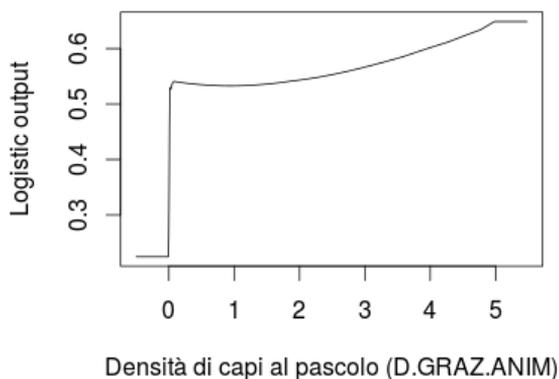


**Figura 6.21:** *Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) nel mese più siccitoso .*

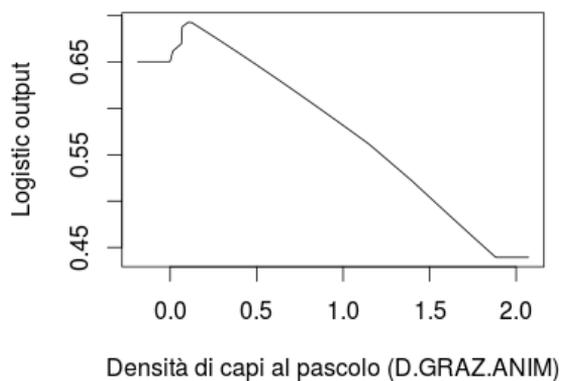


**Figura 6.22:** *Rischio di incendio in relazione alle precipitazioni annuali (in mm di pioggia).*

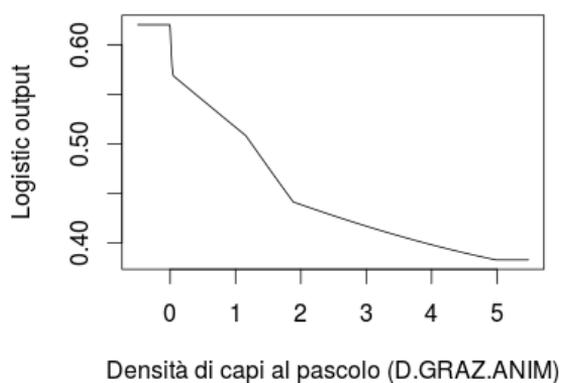
Nonostante che in tutti e tre i dataset la componente socio-economica legata alle attività di pascolo, espressa dalla densità di capi al pascolo per cella raster, risulti limitatamente informativa (come dimostrato nelle tabelle 6.9, 6.10 e 6.11), risulta interessante analizzare i trend di variazione del rischio di incendio in funzione di questa covariata, riportate nelle seguenti rappresentazioni grafiche. Come si può vedere in figura 6.24 e in figura 6.25, per quanto riguarda il rischio in inverno, l'andamento delle curve sembra confermare quanto sostenuto da Bovio [9], mostrando una progressiva diminuzione del rischio all'aumentare del carico di pascolo, influenzando sulla quantità di biomassa combustibile. Un andamento opposto si riscontra nel rischio incendio estivo, in figura 6.23, dove si nota un leggero aumento all'aumentare della densità di capi al pascolo.



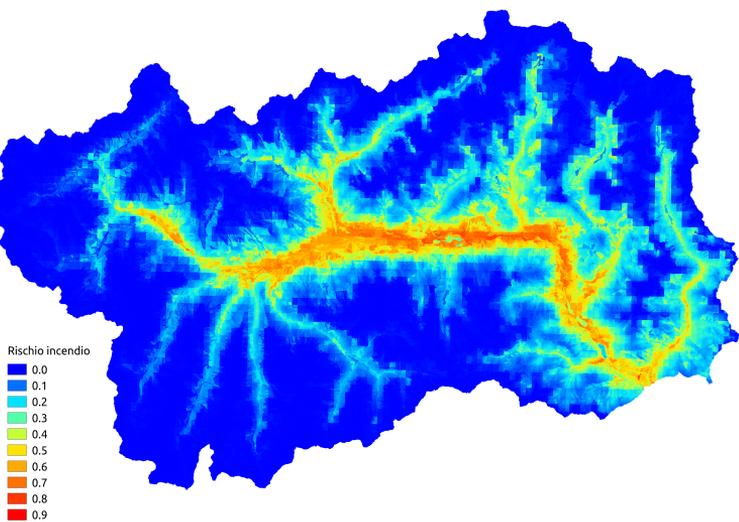
**Figura 6.23:** *Rischio di incendio estivo in relazione alla densità di animali al pascolo (numero di animali per cella raster).*



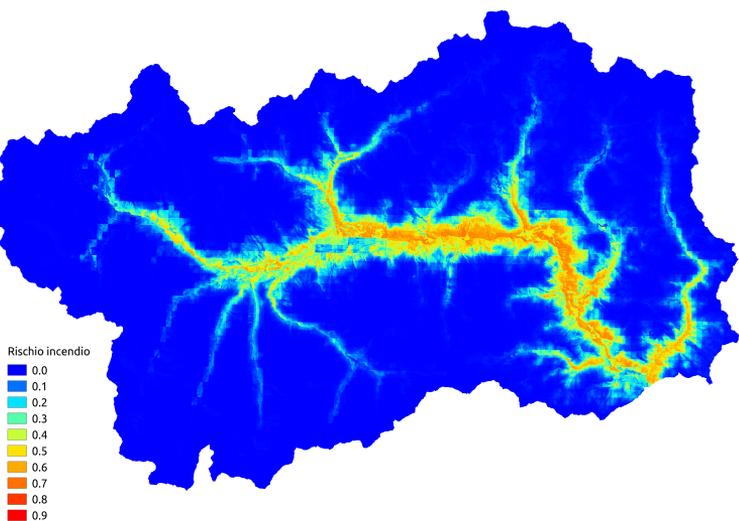
**Figura 6.24:** *Rischio di incendio invernale nelle aree a prevalente copertura boschiva in relazione alla densità di animali al pascolo (numero di animali per cella raster).*



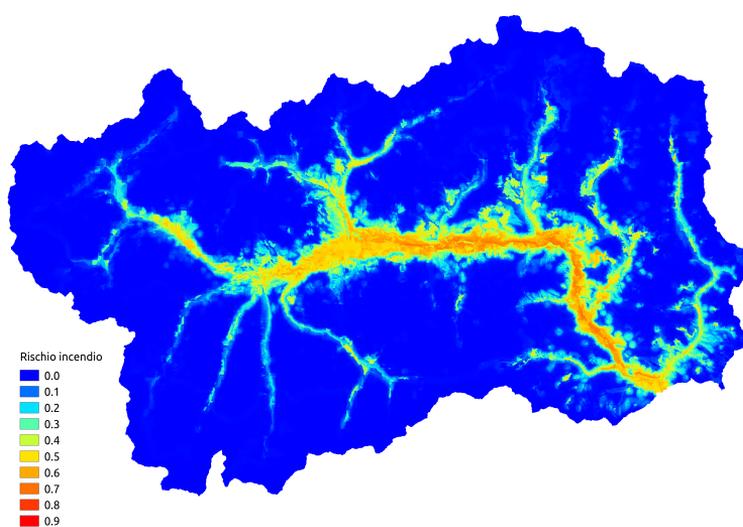
**Figura 6.25:** *Rischio di incendio invernale nelle aree a prevalente copertura di prati e pascoli in relazione alla densità di animali al pascolo (numero di animali per cella raster).*



**Figura 6.26:** *Distribuzione media del rischio in Valle d'Aosta per gli incendi estivi .*



**Figura 6.27:** *Distribuzione media del rischio in Valle d'Aosta per gli incendi boschivi invernali.*



**Figura 6.28:** *Distribuzione media del rischio in Valle d'Aosta per gli incendi invernali generati da aree a prevalente copertura di prati e pascoli.*



---

## Discussione

---

La comparazione delle prestazioni di MAXENT è stata condotta prendendo in considerazione il valore di AUC medio, registrato tra le repliche per ogni modello, definito dalla curva ROC. Come descritto nel paragrafo 3.1, la curva ROC è una raffigurazione bidimensionale della capacità di classificazione ottenibile da un modello predittivo della presenza di un fenomeno, che però male si presta alle indagini comparative di modelli diversi, per questo motivo la valutazione delle *performance* è stata misurata attraverso il valore di AUC (*Area Under the Curve*), ovvero l'area sottesa alla curva ROC. Questo è un metodo che è largamente impiegato [25] e che ha il pregio di permettere confronti che non sono influenzati né dalle dimensioni del campione, né dal numero di covariate coinvolte per la costruzione del modello.

Sulla base di quanto evidenziato da Lobo *et al.* riguardo il fatto che le previsioni dei modelli sono dipendenti dalle condizioni in cui il modello è stato effettuato [44] sono state comparate le *performance* dei modelli in relazione al numero dei punti di presenza degli incendi considerati, alla risoluzione spaziale delle covariate in funzione dell'estensione dell'area di indagine ed infine, al sistema di selezione delle variabili.

Nell'ultima parte del capitolo sono commentati i risultati dei modelli dal punto di vista delle variabili maggiormente significative e della loro influenza sul rischio di incendio. L'intento è quello di fornire una trattazione, su base ecologica, di come queste variabili influiscono nel determinare le condizioni favorevoli agli incendi, evidenziando le tendenze comuni.

## 7.1 Valutazione delle prestazioni dei modelli

La comparazione dei modelli, elaborati nei casi di studio considerati, evidenzia in generale un notevole livello di accuratezza e capacità di classificazione. Le prestazioni valutate in termini di AUC, riepilogate in tabella 7.1, mostrano per la maggior parte dei modelli, valori superiori a 0,8, considerato come valore limite al di sopra del quale si possono considerare prestazioni efficienti [46].

**Tabella 7.1:** *Comparazione delle prestazioni dei modelli elaborati nei casi di studio ( $\sigma$  indica il valore di deviazione standard di AUC nelle reiterazioni di calcolo dell'algoritmo).*

Modello	AUC	$\sigma$
Amazonia Bol con tutte le variabili	0,810	0,019
Amazonia Bol con variabili selezionate	0,840	0,002
Dipartimento 1 – Beni	0,836	0,011
Dipartimento 2 – Santa Cruz	0,784	0,008
Area 1	0,937	0,008
Area 2	0,914	0,016
Area 3	0,735	0,015
VdA Estivi	0,904	0,008
VdA Boschivi invernali	0,946	0,003
VdA Prati e Pascoli invernali	0,937	0,002

Le migliori prestazioni si riscontrano per i modelli calcolati per i tre dataset del caso di studio della Valle d'Aosta (VdA in tabella 7.1) e per l'area 1 e 2 in Bolivia (corrispondenti rispettivamente alle municipalit  di Urubich  e di Ascenci n de Guarayos e alla municipalit  di Concepci n), dove l'AUC, sempre superiore a 0,9, indica l'ottima capacit  di discriminazione del rischio di incendio e un'elevata accuratezza. Tali valori risultano superiori in confronto con quanto descritto in letteratura [53], dove sono riportati valori che al massimo raggiungono  $AUC = 0,890$  per modelli calcolati con metodologia paragonabile a quella applicata.

Gli unici due casi in cui si registrano valori di AUC inferiori a 0,8 riguardano elaborazioni condotte nello studio della distribuzione del rischio di incendi in Amazzonia boliviana:

- 1) nel Dipartimento di Santa Cruz ( $AUC = 0,784$ ) il valore, sebbene inferiore, non discosta molto dalla soglia indicata in letteratura per la valutazione dell'efficienza nella capacit  discriminante, e comunque comparabile con i risultati descritti in studi simili [54];
- 2) nell'Area 3, corrispondente alle due municipalit  di Robor  e di San J s  de Chiquitos, il basso valore ( $AUC = 0,735$ ) potrebbe essere imputabile alla relativamente bassa variabilit  delle covariate che maggiormente influiscono sul rischio di incendio: l'area   prevalentemente interessata da usi del suolo legati alle attivit  di pascolo del bestiame in ambienti con

fisionomia di vegetazione a savana idrofita, tranne le esigue aree forestali nelle zone ad altitudine maggiore. Questo assieme alla pratica tradizionale di uso del fuoco per rinnovare i pascoli, che spesso si evolvono in incendi, rende tutta l'area particolarmente suscettibile al rischio di incendio, senza però evidenziare particolari prevalenze nella distribuzione.

Inoltre, se si prende in considerazione l'approccio meno restrittivo relativo ai valori di AUC, riportato nel lavoro di Duane *et al.* [19], dove viene indicato che solo i modelli con valori inferiori a 0,7 risultano non informativi, anche i modelli elaborati per le due suddette aree risultano utili per spiegare la distribuzione del rischio degli incendi in relazione alle variabili considerate, anche se meno accurati.

## 7.2 Numerosità campionaria

---

Le prestazioni dei modelli, in relazione al numero di incendi considerati, non hanno evidenziato particolari tendenze, a conferma di quanto riportato da da Wisz *et al.* [71] e successivamente da Baldwin *et al.* [4] in merito alla ridotta sensibilità di MAXENT alla dimensione campionaria.<sup>1</sup>

Allo stesso modo, i risultati ottenuti con l'applicazione dell'algoritmo di massima entropia nel caso di studio della Valle d'Aosta sembrano confermare quanto evidenziato da Hernandez *et al.* [33], ovvero che oltre una certa quantità (tra 50 e 75 per le osservazioni su specie animali rare di Hernandez), la dimensione campionaria non comporta un miglioramento significativo delle prestazioni del modello in termini di AUC. Come si può notare in tabella 7.2, la comparazione delle prestazioni registrate nei due dataset invernali della Valle d'Aosta mostra un valore di AUC maggiore con un numero più limitato di punti di innesco in aree boschive (222 punti di innesco) rispetto ai punti considerati per la valutazione della distribuzione degli incendi generati su aree a prevalente copertura di prati e pascoli (390 inneschi).

Sebbene la distinzione dei due dataset, fatta sulla base della tipologia di vegetazione corrispondente al punto di innesco (prevalenza di bosco o prevalenza di aree pascolabili), possa far pensare ad una limitata comparabilità, è necessario tenere presente che in entrambi i casi l'analisi è stata condotta sugli incendi forestali e che la distinzione ha riguardato i punti di innesco e non la vegetazione prevalentemente bruciata.

---

<sup>1</sup>“Other algorithms were much less sensitive to sample size, including an algorithm based on maximum entropy (MAXENT) that had among the best predictive power across all sample sizes. Relative to other algorithms, a distance metric algorithm (DOMAIN) and a genetic algorithm (OM-GARP) had intermediate performance at the largest sample size and among the best performance at the lowest sample size. No algorithm predicted consistently well with small sample size ( $n < 30$ ) and this should encourage highly conservative use of predictions based on small sample size and restrict their use to exploratory modelling.”(Wisz, 2008 - p. 763)

## Capitolo 7. Discussione

---

**Tabella 7.2:** *Comparazione delle prestazioni dei modelli elaborati nei casi di studio in relazione al numero di incendi.*

Modello	AUC	N punti incendio
Amazonia Bol con tutte le variabili	0,810	1270
Amazonia Bol con variabili selezionate	0,840	1270
Dipartimento 1 – Beni	0,836	338
Dipartimento 2 – Santa Cruz	0,784	929
Area 1	0,937	30
Area 2	0,914	31
Area 3	0,735	174
VdA Estivi	0,904	180
VdA Boschivi invernali	0,946	222
VdA Prati e Pascoli invernali	0,937	390

### 7.3 Rapporti con la risoluzione spaziale

---

In merito alla scelta della risoluzione spaziale delle covariate da utilizzare per l'elaborazione dei modelli con MAXENT, in letteratura non si trovano particolari indicazioni. L'unico riferimento, fornito da Ranner e Warton [64], sottolinea che la distribuzione di probabilità, e di conseguenza l'accuratezza, generata dall'algoritmo di massima entropia dipende dalla scelta della risoluzione spaziale, definendola scala-dipendente.

Concettualmente non esiste una singola scala naturale alla quale i modelli ecologici dovrebbero essere studiati [41]. In generale, la risoluzione dovrebbe essere coerente con il contenuto informativo dei dati anche se in pratica questo non è sempre possibile [22], a causa della disponibilità, talvolta scarsa, di dati adeguatamente precisi.

In conseguenza a quanto appena detto, le prestazioni dei modelli sono state in funzione dell'estensione del territorio investigato. Le differenze di scala sono infatti fondamentali per mettere in evidenza la variabilità di rappresentazione del fenomeno laddove non siano disponibili covariate con risoluzioni spaziali di maggior dettaglio. La comparazione fra risoluzioni viene invece evidenziata (in tabella 7.3) fra contesti territoriali che hanno comunque dimensioni difficilmente comparabili fra loro. Il rapporto fra superficie (S), quindi "gradino" di scala, e risoluzione (R) è stato utilizzato come descrittore del valore di performance del modello.

Il valori delle *performance* dei modelli in relazione alla risoluzione sono riepilogati nella tabella 7.3 in ordine di AUC decrescente, nella colonna denominata "S/R" è riportato il rapporto tra la superficie coperta dall'area di studio e la risoluzione spaziale delle covariate utilizzate nei modelli.

Dal confronto dei risultati si nota come le migliori prestazioni si registrino nei casi in cui il rapporto tra estensione dell'area di indagine e risoluzione spaziale delle variabili è più basso. Indipendentemente dalla superficie, infatti

## 7.4. Metodi di selezione delle variabili

**Tabella 7.3:** Comparazione delle prestazioni dei modelli elaborati nei casi di studio in relazione alla risoluzione spaziale delle variabili e dell'estensione dell'area di indagine.

Modello	AUC	Superficie (Km <sup>2</sup> )	Risoluzione (Km <sup>2</sup> )	S/R
VdA Boschivi invernali	0,946	3262	0,0009	36, 24E + 05
Area 1	0,937	19027	0,00625	30, 44E + 05
VdA Prati e Pascoli invernali	0,937	3262	0,0009	36, 24E + 05
Area 2	0,914	34887	0,00625	55, 82E + 05
VdA Estivi	0,904	3262	0,0009	36, 24E + 05
Amazzonia Bol con variabili selezionate	0,840	732448	0,00625	1171, 92E + 05
Dipartimento 1 – Beni	0,836	202859	0,00625	324, 57E + 05
Amazzonia Bol con tutte le variabili	0,810	732448	0,00625	1171, 92E + 05
Dipartimento 2 – Santa cruz	0,784	345501	0,00625	552, 80E + 05
Area 3	0,735	25251	0,00625	40, 40E + 05

si ritrovano valori di AUC superiori a 0,9 sia per i modelli elaborati per il caso di studio in Valle d'Aosta ( $S/R = 36, 24E + 05$ ), sia per due aree (Area 1 e Area 2) esaminate in Amazzonia Boliviana, all'interno del Dipartimento di Santa Cruz.

Tale risultato fa presupporre che le prestazioni dei modelli di distribuzione del rischio degli incendi elaborati con MAXENT, in relazione alle variabili considerate, non solo siano scala-dipendenti, ma dipendano anche dal rapporto tra scala e risoluzione spaziale.

Il basso valore di AUC registrato per l'Area 3 in Bolivia, accennato anche nel paragrafo 7.1, risultando in disaccordo con quanto appena sostenuto, rafforza l'ipotesi che la bassa variabilità nell'uso del suolo (fattore molto importante per il modello elaborato in quest'area) porti l'algoritmo a generalizzare eccessivamente la distribuzione degli incendi, producendo quindi valori di AUC più bassi rispetto ad aree dove la variabilità è maggiore, con rapporto tra estensione e risoluzione spaziale comparabile.

## 7.4 Metodi di selezione delle variabili

I fattori predisponenti condizioni favorevoli all'aumento del rischio degli incendi boschivi sono molteplici, e dipendendo, come già dimostrato, dalla morfologia del territorio, dal clima, dalle caratteristiche della vegetazione e dalle attività umane. La dipendenza da molteplici variabili, spesso coagenti e tra loro interrelate, rende complessa l'interpretazione dei modelli di rischio per gli incendi, nell'individuare i fattori maggiormente coinvolti.

Questo, sommato al fatto che in tutti i fenomeni naturali esiste un certo grado di colinearità tra le variabili predittive [18] a causa della intrinseca dipendenza lineare tra covariate, e che molti approcci statistici sono sensibili alla colinearità [6], può condurre all'instabilità delle stime dei parametri, con errori standard delle stime eccessivi, che inducono di conseguenza a statistiche inferenziali deviate. Secondo quanto riportato da Kuemmerle *et al.*, le prestazioni

di MAXENT non sono influenzate dalla colinearità delle covariate, ma questa condizione ostacola l'interpretazione dei modelli [40], permanendo comunque i problemi legati all'impossibilità di distinguere gli effetti delle variabili e la conseguente probabile distorsione nell'estrapolazione di informazioni [50].

Nei casi di studio riportati sono stati utilizzati due metodi di selezione delle variabili permettendo di rendere più comprensibili le dinamiche di interazione dei predittori che contribuiscono a generare le condizioni secondo le quali il rischio di incendi si distribuisce.

Il primo metodo, utilizzato nelle aree indagate in Bolivia, ha previsto l'elaborazione di modelli esplorativi, comprendendo tutti gli strati informativi disponibili e analizzandoli in base ai risultati della procedura di ricampionamento *Jackknife Repeated Replication* (JRR) e alla stima del valore del contributo e dell'importanza di permutazione. Dal confronto delle prestazioni registrate per i modelli generali calcolati per la *Cuenca* amazzonica boliviana si può notare come il metodo di selezione tramite modelli esplorativi, dimezzando il numero delle covariate (da 26 considerate ne sono state scelte 13) porti ad una maggiore accuratezza: il valore di AUC sale da 0,81 a 0,84. In questo caso, oltre a rendere più semplice l'interpretazione delle relazioni tra le variabili il rischio di incendio, si registra anche un miglioramento della *performance*, imputabile al fatto che a questo livello di scala, la colinearità delle variabili del dataset WoldClim tende a distorcere leggermente la capacità di classificazione dell'algoritmo di massima entropia, producendo una distribuzione più generalizzata.

Il secondo metodo è stato applicato al caso di studio della Valle d'Aosta, impiegando l'analisi delle componenti principali per identificare i gruppi *proxy* di variabili linearmente correlate per ogni componente, all'interno dei quali selezionare la covariata con *loading* maggiore. Secondo quanto sostenuto da Cruz-Cárdenas *et al.* [14], l'utilizzo della PCA per la selezione delle variabili, sulla base dei più elevati valori di *loadings*, garantisce la diminuzione dell'autocorrelazione tra le variabili prima di elaborare modelli di distribuzione. Anche in questo caso la riduzione del numero delle variabili ha reso più facile la comprensione dei rapporti tra predittori e rischio di incendio, ottenendo le migliori prestazioni registrate tra tutti i casi di studio e assicurando di limitare al massimo le distorsioni legate alla colinearità, sia sulla capacità di classificazione, sia in merito all'estrapolazione delle informazioni legate al fenomeno.

Riepilogando, con entrambi i metodi utilizzati per selezionare le variabili si riesce a ridurre notevolmente il numero di predittori, facilitando la comprensione del fenomeno e semplificando l'interpretazione delle relazioni tra

## 7.5. Influenza delle variabili sul rischio di incendio

---

le covariate e la distribuzione del rischio. Il metodo che si basa su modelli esplorativi e utilizza il contributo percentuale come stimatore dell'informatività della variabile, però, appare essere meno robusto rispetto al metodo che utilizza i *loadings* delle componenti principali. Questo a causa del fatto che il calcolo operato da MAXENT per il contributo percentuale si basa su un approccio euristico all'importanza del modello, nel quale i valori di contributo sono determinati dal guadagno in termini di aumento di AUC, apportato da ogni variabile. Di conseguenza è necessario prestare particolare attenzione nell'impiego di questo stimatore che risulta essere particolarmente suscettibile a forti gradi di colinearità, imputando maggiore importanza a una di due o più variabili altamente correlate [4]. Risulta quindi importante controllare che le valutazioni fatte attraverso l'analisi dei valori di contributo percentuale siano confermate dai risultati del ricampionamento JRR, per evitare di eliminare variabili che, a causa di elevata correlazione registrino valori di contributo più bassi rispetto alle colineari, ma che sarebbero ecologicamente più informative per la spiegazione del fenomeno, in quel caso di studio.

## 7.5 Influenza delle variabili sul rischio di incendio

---

Nella maggior parte dei casi analizzati il principale fattore discriminante per le condizioni che influiscono sulla distribuzione del rischio di incendio è rappresentato dalle categorie di uso del suolo. Sia per quanto riguarda i casi di studio analizzati in Bolivia che per quanto riguarda le elaborazioni condotte in Valle d'Aosta, emerge come le tipologie di uso del suolo legate alle attività di pascolo del bestiame giochino un ruolo determinante, rappresentando le condizioni a maggiore rischio. Questo è dovuto, probabilmente, alla compresenza di due condizioni essenziali per gli incendi: da un lato l'abbondanza di combustibile fine potenzialmente molto infiammabile in ambiente naturale o seminaturale prossimo ad aree boscate, dall'altro la presenza di attività antropiche che prevedono l'uso del fuoco, in maniera volontaria come accade in Bolivia, o accidentalmente come ipotizzabile per la Valle d'Aosta. Le elaborazioni condotte in ambiente alpino, inoltre, fanno presupporre l'efficacia dell'attività di pascolo quale fattore di abbassamento potenziale del rischio di incendio visto il comportamento da "controllore" di biomassa, agendo direttamente sulla quantità di combustibile.

Altro fattore ad elevata informatività, indagato nei i modelli boliviani, è il tipo di vegetazione che, come evidenziato in studi condotti in condizioni simili [55], risulta essere un consistente predittore della probabilità di incendio nelle regioni in cui la vegetazione passa gradualmente da boschi e foreste di igrofite a savane xeriche di specie a portamento arbustivo. Sia nelle savane e nelle praterie che nelle aree periodicamente allagate, la probabilità di incendio

è maggiore rispetto ad aree con altri tipi di vegetazione. In queste condizioni, durante la stagione delle piogge o nei periodi di allagamento, le specie erbacee si sviluppano in abbondanza accumulando una grande quantità di combustibile fine, che, nella stagione secca, perde umidità molto velocemente aumentando l'inflammabilità. Le aree interessate da questi tipi di vegetazione sono tradizionalmente trattate con i fuochi controllati per favorire il pascolo e le attività agricole, con il rischio di generare incendi su superfici maggiori rispetto al previsto e in condizioni tali in cui il controllo del fuoco non è più attuabile.

I risultati dei rapporti tra le variabili utilizzate e il rischio di incendio hanno messo in evidenza la complessità delle relazioni con il clima. Le variabili legate alle precipitazioni e alle temperature, infatti, risultano sempre essere tra i principali fattori che maggiormente contribuiscono alla definizione del rischio per gli incendi. Nonostante siano state considerate inizialmente tutte le 19 variabili presenti nel dataset *WorldClim*, e successivamente circa la metà di queste, non una singola di esse è risultata essere predominante nella definizione del rischio di incendio anche se alcune combinazioni di fattori sono ricorrenti.

In Amazzonia, la combinazione di stagionalità delle temperature ed escursione termica media giornaliera, con andamenti molto simili, ricorre tra i predittori a maggiore contributo sia per il modello generale che a scala di Dipartimento. In questi casi il rischio di incendio aumenta al crescere delle due covariate, evidenziando la maggiore suscettibilità agli incendi in aree dove è più marcata la stagionalità dei regimi termici.

Diminuendo l'estensione delle superfici indagate, nelle tre aree del Dipartimento di Santa Cruz si nota come questa dipendenza dalla stagionalità e dall'escursione termica giornaliera vari fino ad essere sostituita dai fattori legati alla pioggia, passando da Nord-Ovest a Sud-Est, secondo un gradiente che ricomprende la variazione di vegetazione dalle zone tropicali alle savane xeriche del *Chaco*. Nell'area 1 si ritrova ancora una forte dipendenza dalla stagionalità delle temperature, ma l'escursione termica influenza in maniera opposta il rischio di incendio, che diminuisce al crescere della differenza tra temperatura minima e massima giornaliera.

Nell'area 2 assieme alla stagionalità delle temperature che ha un comportamento simile a quello precedentemente descritto per gli altri modelli, si nota l'importanza della temperatura media annuale come predittore della distribuzione del rischio per gli incendi.

Infine, nell'area 3 sono le precipitazioni a influire maggiormente sulle condizioni per l'aumento del rischio di incendio. Sia le precipitazioni dei tre mesi più piovosi, che le precipitazioni nei mesi più freddi costituiscono i maggiori fattori predisponenti. In una zona dove la vegetazione prevalente è rappre-

sentata dalle specie di savana relativamente più umida, il fattore limitante all'accumulo del combustibile è rappresentato dalla disponibilità idrica garantita dalle piogge. In quest'area, essendo il pascolo estensivo l'attività prevalente, è verosimile pensare che la struttura della savana sia mantenuta, se non estesa, attraverso l'utilizzo del fuoco da parte dell'uomo, con il risultato di una scarsa differenziazione dei tipi fisionomici di vegetazione (in questo caso ecosistemi simili alla savana). L'azione ripetuta di disturbo operata dal passaggio del fuoco, infatti, non consente la naturale evoluzione di strutture vegetazionali più complesse rispetto alla savana, anche là dove le condizioni naturali lo permetterebbero.

Questo suggerisce che le relazioni con il clima operino influenze diverse in regimi di incendio differenti, agendo sulla struttura e sulle condizioni dei combustibili vincolando la distribuzione geografica della vegetazione.

Le precipitazioni sono un importante predittore anche in ambiente alpino, essendo presenti tra i principali fattori che determinano la distribuzione degli incendi nei tre dataset della Valle d'Aosta, dove il regime pluviometrico è caratterizzato da due massimi nelle stagioni intermedie, e da due minimi, in estate ed in inverno. Nel settore orientale prevale il massimo primaverile, in quello occidentale il massimo autunnale, nella zona intermedia la differenza tra i due massimi è ridotta.

Per gli incendi indagati nel periodo estivo, i fattori che riducono per il rischio sono l'aumento delle precipitazioni nei mesi più piovosi e la diminuzione delle temperature del mese più caldo. In merito alle precipitazioni si può ipotizzare che la scarsità di pioggia in primavera favorisca l'accumularsi di sostanza secca, soprattutto negli starti erbacei e arbustivi, che si traduce un aumento del rischio estivo.

In tutti e due i dataset invernali, invece, le precipitazioni che influiscono maggiormente sono quelle del mese meno piovoso, influenzando direttamente sull'umidità dei combustibili. Il modello elaborato per gli incendi boschivi in inverno evidenzia, inoltre, la forte relazione della distribuzione del rischio di incendio con l'indice di xericità (*southwestness index*), rilevando la maggiore suscettibilità delle aree ad esposizione Sud-Ovest, tipicamente più soleggiate e calde.

L'analisi dei risultati del modello invernale per gli inneschi su aree a prevalente copertura di prati e pascoli, invece mostra un maggiore valore di rischio in prossimità delle infrastrutture e della viabilità, evidenziando ulteriormente l'importanza della componente antropica, che contribuisce, seppur in modo più limitato, anche nella distribuzione del rischio per gli incendi generati da inneschi in aree boschive nella stessa stagione.

L'altra componente antropica indagata in Valle d'Aosta, espressa dalla densità di capi al pascolo, è risultata limitatamente informativa probabilmente a

## Capitolo 7. Discussione

---

causa della natura aggregata del dato di partenza. Nonostante ciò, le tendenze espresse dal modello sembrano confermare come l'aumento del carico di animali influisca sulla diminuzione del combustibile riducendo il rischio di incendio. È verosimile ipotizzare che dati di maggiore dettaglio potrebbero portare ad una più precisa comprensione delle relazioni tra pascolo e rischio di incendio in Valle d'Aosta.

---

## Conclusioni

---

I risultati ottenuti dall'applicazione della metodologia, basata sull'algoritmo di massima entropia, allo studio della distribuzione del rischio di incendio hanno evidenziato buone prestazioni a tutte le scale geografiche esaminate nei casi di studio, con *performance* spesso superiori rispetto a esperienze paragonabili condotte da altri autori, riportate in letteratura. La comparazione delle prestazioni dei modelli elaborati, rispetto alle cause di incertezza indagate, ha permesso di mettere in luce i punti di forza e le criticità, riepilogate in forma di matrice di analisi SWOT<sup>1</sup> in tabella 8.1, della metodologia.

Le elevate prestazioni discriminanti nella stima della distribuzione del rischio di incendio in ambienti e condizioni profondamente differenti, unite alle buone capacità di gerarchizzazione delle variabili che contribuiscono a determinare le condizioni favorevoli agli incendi, a diverse scale geografiche e indipendentemente dalla numerosità campionaria degli eventi utilizzati per l'elaborazione dei modelli, costituiscono i principali punti di forza della metodologia, evidenziati dalle comparazioni nei casi di studio riportati.

---

<sup>1</sup>L'analisi SWOT (Strengths - Weaknesses - Opportunities - Threats) è uno strumento di pianificazione strategica semplice ed efficace che serve ad evidenziare le caratteristiche di un progetto, di un programma, di una metodologia e le conseguenti relazioni con l'ambiente operativo nel quale si colloca, offrendo un quadro di riferimento per la definizione di orientamenti strategici finalizzati al raggiungimento di un obiettivo. Essa consente la valutazione simultanea, rispetto all'obiettivo che si vuole raggiungere, delle variabili sia interne che esterne. Le variabili interne sono quelle che fanno parte del sistema e sulle quali è possibile intervenire; quelle esterne invece possono solo essere tenute sotto controllo, in modo di sfruttare i fattori positivi e limitare i fattori che invece rischiano di compromettere il raggiungimento degli obiettivi prefissati.

**Tabella 8.1:** *Matrice di analisi SWOT dell'applicazione dell'algoritmo di massima entropia allo studio della distribuzione del rischio per gli incendi forestali.*

	<b>STRENGTHS ★</b>	<b>WEAKNESSES ◆</b>
<b>INTERNAL</b>	<p>★ Buona performance ed elevate prestazioni discriminanti nell'applicazione a casi di studio che presentano caratteristiche ambientali (climatiche, biogeografiche, sociali e infrastrutturali) anche estremamente diverse fra loro</p> <p>★ Relativamente indipendente dall'esiguità di punti di presenza (punti dove realmente si verifica il fenomeno in analisi) nella stima della distribuzione</p> <p>★ Buone capacità di classificare e gerarchizzare i fattori più influenti a diverse scale geografiche</p>	<p>◆ Accuratezza dipendente dal rapporto tra superficie dell'area indagata e risoluzione spaziale delle variabili utilizzate</p> <p>◆ Sensibile alla colinearità delle variabili</p> <p>◆ Tendenza all'eccessiva generalizzazione della distribuzione se i predittori più influenti hanno bassa variabilità rispetto alle altre covariate</p>
<b>EXTERNAL</b>	<p>▲ Promuove una migliore comprensione del peso dei singoli fattori ambientali e delle relazioni fra i fattori stessi implicati in fenomeni di incendio anche a diverse scale geografiche</p> <p>▲ Può diventare un utile strumento per il supporto alla pianificazione della gestione territoriale e dell'allocazione delle risorse per la lotta agli incendi nelle zone a maggiore rischio</p>	<p>▼ Mascheramento: l'utilizzo di dati poco aggiornati, distribuiti in modo discontinuo sul territorio, sia per quanto riguarda i punti di presenza degli incendi, sia per le variabili ambientali e antropiche, può portare ad un'interpretazione distorta o delle informazioni derivabili dall'applicazione del modello</p> <p>▼ Decontestualizzazione: l'impiego di un set predeterminato di variabili, non selezionato caso per caso, può condurre a valutazioni errate riguardo all'influenza dei fattori rispetto a differenti regimi di incendio</p> <p>▼ Eccesso di informazione: l'utilizzo di variabili correlate linearmente induce ad interpretazioni potenzialmente distorte.</p>
	<b>OPPORTUNITIES ▲</b>	<b>THREATS ▼</b>

---

Sono stati evidenziati gli aspetti critici in termini di accuratezza dei modelli, ai quali sarà necessario prestare particolare attenzione in fase di elaborazione, legati alla sensibilità rispetto alle interdipendenze delle covariate e al loro dettaglio.

Sia la colinearità che la marcata differenza tra la variabilità dei fattori più influenti rispetto agli altri influiscono sull'accuratezza dei modelli e sulla capacità discriminante. L'elevato grado di correlazione tra le variabili porta a produrre modelli la cui interpretazione potrebbe essere potenzialmente difficoltosa, se non addirittura forviante, a causa della scarsa capacità da parte dell'algoritmo di massima entropia ad estrapolare l'effettiva importanza dei singoli fattori quando questi sono colineari. L'algoritmo, in presenza di colinearità delle variabili, rischia di attribuire valori di contributo maggiore a una variabile rispetto a quella ad essa correlata. Per limitare le distorsioni causate da questa caratteristica è quindi necessario applicare un metodo di selezione delle variabili, tale da evitare, o comunque limitare il più possibile, l'utilizzo di combinazioni di variabili linearmente correlate tra loro.

La bassa variabilità dei fattori preponderanti sulla distribuzione del rischio rispetto alle altre covariate tende a produrre modelli eccessivamente generalizzati. Questo, a causa del fatto che, in queste condizioni, l'algoritmo non riesce a determinare una netta prevalenza delle combinazioni di fattori che caratterizzano il fenomeno e, di conseguenza, i vincoli che vengono imposti alla distribuzione di massima entropia sono meno forti.

Oltre la dipendenza rispetto alla scala geografica, già individuata in letteratura, è stata constatata anche una dipendenza rispetto al rapporto tra superficie indagata e risoluzione delle variabili utilizzate che influenza le *performance* dei modelli.

La metodologia testata sulla stima della distribuzione del rischio per gli incendi forestali può rivelarsi estremamente utile per migliorare la comprensione dei fenomeni di incendio, costituendo uno strumento efficace nell'analisi delle relazioni tra i fattori predisponenti i livelli più elevati di rischio. Le informazioni ricavabili dall'applicazione della metodologia, inoltre, possono aumentare l'efficienza della pianificazione territoriale nella lotta agli incendi boschivi, supportando e indirizzando le strategie di intervento e di allocazione delle risorse economiche nelle zone a rischio più elevato.



---

## Elenco delle figure

---

3.1	Esempio di distribuzione di frequenze per i risultati di assenza e di presenza di un generico modello . . . . .	22
4.1	Inquadramento del Programma PASF e area di indagine per il rischio di incendi forestali. . . . .	26
4.2	I punti in nero rappresentano la posizione degli incendi confermati da ABT per la Cuenca amazonica boliviana. . . . .	31
4.3	Rappresentazione grafica di JRR applicato al modello esplorativo per la selezione delle variabili in Amazonia boliviana: in rosso sono riportati i valori di AUC registrati per modelli che considerano una sola variabile alla volta(in ordinata), in turchese sono riportati i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata. . . . .	33
4.4	Curva ROC delle performance media delle cinque reiterazioni di calcolo del modello per gli incendi forestali in Amazonia boliviana. . . . .	35
4.5	Rappresentazione grafica di JRR applicato al modello per gli incendi forestali in Amazonia boliviana: in rosso sono riportati i valori di AUC registrati per modelli che considerano una sola variabile alla volta (in ordinata), in turchese sono riportati i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata. . . . .	36
4.6	Rischio di incendio in relazione alla stagionalità delle temperature per l'area di studio (deviazione standard moltiplicata per 100). . . . .	38

## Elenco delle figure

---

4.7	Rischio di incendio in relazione all'escursione termica media giornaliera (le temperature sono espresse in gradi Celsius e moltiplicati per 10). . . . .	38
4.8	Rischio di incendio in relazione alle precipitazioni cumulate nei tre mesi più piovosi dell'anno. . . . .	39
4.9	Distribuzione media del rischio di incendi nella Cuenca Amazonica boliviana. . . . .	40
5.1	Inquadramento delle aree di indagine per la valutazione delle prestazioni dell'algoritmo di massima entropia in relazione al rapporto tra risoluzione spaziale delle variabili ed estensione territoriale. . . . .	44
5.2	Curva ROC delle performance media delle cinque reiterazioni di calcolo del modello per gli incendi forestali nel dipartimento del Beni. . . . .	46
5.3	Rappresentazione grafica dei risultati di ricampionamento JRR delle covariate nel Beni: in rosso sono riportati i valori di AUC registrati per modelli che considerano una sola variabile alla volta (in ordinata), in turchese sono riportati i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata. . . . .	47
5.4	Rischio di incendio in relazione alla stagionalità delle temperature per il dipartimento del Beni (deviazione standard delle temperature annuali in gradi Celsius moltiplicata per 100). . . . .	49
5.5	Rischio di incendio in relazione all'escursione termica media giornaliera nel dipartimento del Beni (le temperature sono espresse in gradi Celsius e moltiplicate per 10). . . . .	49
5.6	Rischio di incendio in relazione alle temperature medie annuali nel dipartimento del Beni (le temperature sono espresse in gradi Celsius e moltiplicate per 10). . . . .	50
5.7	Rischio di incendio in relazione alle temperature massime, registrate nel mese più caldo, nel dipartimento del Beni (le temperature sono espresse in gradi Celsius e moltiplicate per 10). . . . .	50
5.8	Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) dei tre mesi più siccitosi nel dipartimento del Beni. . . . .	51
5.9	Rischio di incendio in relazione alla quota altimetrica nel dipartimento del Beni (m s.l.m.). . . . .	51
5.10	Distribuzione media del rischio di incendi nel Dipartimento del Beni. . . . .	52
5.11	Curva ROC delle performance media delle cinque reiterazioni di calcolo del modello per gli incendi forestali nel dipartimento di Santa Cruz. . . . .	54

5.12 Rappresentazione grafica dei risultati di ricampionamento JRR delle covariate nel dipartimento di Santa Cruz: in rosso sono riportati i valori di AUC registrati per modelli che considerano una sola variabile alla volta, in turchese, i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata. . . . .	54
5.13 Rischio di incendio in relazione alla stagionalità delle temperature per il dipartimento di Santa Cruz (deviazione standard delle temperature annuali in gradi Celsius moltiplicata per 100).	56
5.14 Rischio di incendio in relazione al range annuale di temperatura nel dipartimento di Santa Cruz (le temperature sono espresse in gradi Celsius e moltiplicate per 10). . . . .	56
5.15 Rischio di incendio in relazione all'escursione termica media giornaliera nel dipartimento di Santa Cruz (le temperature sono espresse in gradi Celsius e moltiplicate per 10). . . . .	57
5.16 Rischio di incendio in relazione alle temperature massime, registrate nel mese più caldo, nel dipartimento di Santa Cruz (le temperature sono espresse in gradi Celsius e moltiplicate per 10).	57
5.17 Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) dei tre mesi più siccitosi nel dipartimento di Santa Cruz.	58
5.18 Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) dei tre mesi più piovosi nel dipartimento di Santa Cruz.	58
5.19 Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) dei tre mesi più freddi nel dipartimento di Santa Cruz. .	59
5.20 Rischio di incendio in relazione alla quota altimetrica nel dipartimento di Santa Cruz (m s.l.m.). . . . .	59
5.21 Distribuzione media del rischio di incendi nel Dipartimento di Santa Cruz. . . . .	60
5.22 Curva ROC delle performance media delle cinque reiterazioni di calcolo del modello per gli incendi forestali nell'area 1. . . .	61
5.23 Rappresentazione grafica dei risultati di ricampionamento JRR delle covariate nell'area 1: in rosso sono riportati i valori di AUC registrati per modelli che considerano una sola variabile alla volta, in turchese, i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata. . .	61
5.24 Rischio di incendio in relazione alla stagionalità delle temperature per l'area 1 (deviazione standard delle temperature annuali in gradi Celsius moltiplicata per 100). . . . .	63
5.25 Rischio di incendio in relazione all'escursione termica media giornaliera in area 1 (le temperature sono espresse in gradi Celsius e moltiplicati per 10). . . . .	63

## Elenco delle figure

---

5.26	Curva ROC delle performance media delle cinque reiterazioni di calcolo del modello per gli incendi forestali nell'area 2. . . . .	64
5.27	Rappresentazione grafica dei risultati di ricampionamento JRR delle covariate nell'area 2: in rosso sono riportati i valori di AUC registrati per modelli che considerano una sola variabile alla volta, in turchese, i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata. . . . .	65
5.28	Rischio di incendio in relazione alla stagionalità delle temperature per l'area 2 (deviazione standard delle temperature annuali in gradi Celsius moltiplicata per 100). . . . .	66
5.29	Rischio di incendio in relazione alle temperature medie annuali per l'area 2 (le temperature sono espresse in gradi Celsius e moltiplicate per 10). . . . .	67
5.30	Rischio di incendio in relazione alla quota altimetrica per l'area 2 (m s.l.m.). . . . .	67
5.31	Curva ROC delle performance media delle cinque reiterazioni di calcolo del modello per gli incendi forestali nell'area 3. . . . .	68
5.32	Rappresentazione grafica dei risultati di ricampionamento JRR delle covariate nell'area 2: in rosso sono riportati i valori di AUC registrati per modelli che considerano una sola variabile alla volta, in turchese, i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata. . . . .	69
5.33	Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) dei tre mesi più piovosi nell'area 3. . . . .	71
5.34	Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) dei tre mesi più freddi per l'area 3. . . . .	71
5.35	Rischio di incendio in relazione alla quota altimetrica nell'area 3 (m s.l.m.). . . . .	72
5.36	Rischio di incendio in relazione al range annuale di temperatura per l'area 3 (le temperature sono espresse in gradi Celsius e moltiplicate per 10). . . . .	72
5.37	Distribuzione media del rischio di incendi nell'Area 1. . . . .	74
5.38	Distribuzione media del rischio di incendi nell'Area 2. . . . .	74
5.39	Distribuzione media del rischio di incendi nell'Area 3. . . . .	75
6.1	Frequenze mensili cumulate per il periodo di indagine degli incendi, suddivise per tipologia di vegetazione in corrispondenza del punto di innesco. Con "forest" per i soprasuoli boschivi; "grassland&uncult" per le aree a pascolo e gli incolti; "other" per le altre tipologie di vegetazione, prevalentemente ad uso agricolo. . . . .	79

6.2	Esempio di rappresentazione grafica di un gruppo proxy per la PC 2. . . . .	81
6.3	Curva ROC delle performance medie del modello per gli incendi estivi. . . . .	85
6.4	Rappresentazione grafica di JRR applicato al modello per gli incendi estivi in Valle d'Aosta: in rosso, i valori di AUC registrati per modelli che considerano una sola variabile alla volta (in ordinata), in turchese, i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata. . . . .	85
6.5	Rischio di incendio estivo in relazione alle precipitazioni (in mm di pioggia) nei tre mesi più piovosi. . . . .	87
6.6	Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) nel mese più piovoso. . . . .	87
6.7	Rischio di incendio in relazione alla quota altimetrica per gli incendi estivi (m s.l.m.). . . . .	88
6.8	Rischio di incendio in relazione alle temperature massime del mese più caldo (le temperature sono espresse in gradi Celsius e moltiplicate per 10). . . . .	88
6.9	Curva ROC delle performance medie del modello per gli incendi boschivi invernali. . . . .	89
6.10	Rappresentazione grafica di JRR applicato al modello per gli incendi forestali invernali in Valle d'Aosta: in rosso, i valori di AUC registrati per modelli che considerano una sola variabile alla volta (in ordinata), in turchese, i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata. . . . .	90
6.11	Rischio di incendio in relazione alla quota altimetrica per gli incendi invernali (m s.l.m.). . . . .	91
6.12	Rischio di incendio in relazione alle temperature minime del mese più freddo (le temperature sono espresse in gradi Celsius e moltiplicate per 10). . . . .	91
6.13	Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) nel mese più siccitoso. . . . .	92
6.14	Rischio di incendio in relazione alle temperature medie dei tre mesi meno piovosi (le temperature sono espresse in gradi Celsius e moltiplicate per 10). . . . .	92
6.15	Rischio di incendio in relazione alla distanza dalle infrastrutture antropiche (m). . . . .	93
6.16	Rischio di incendio in relazione al southwestness index. . . . .	93
6.17	Curva ROC delle performance medie del modello per gli incendi invernali su prati e pascoli. . . . .	94

## Elenco delle figure

---

6.18 Rappresentazione grafica di JRR applicato al modello per gli incendi forestali invernali in Valle d'Aosta: in rosso, i valori di AUC registrati per modelli che considerano una sola variabile alla volta (in ordinata), in turchese, i valori di AUC per i modelli che considerano l'intero set di variabili meno quella in ordinata.	95
6.19 Rischio di incendio in relazione alla distanza dalle infrastrutture antropiche (m).	96
6.20 Rischio di incendio in relazione alla quota altimetrica (m s.l.m.).	96
6.21 Rischio di incendio in relazione alle precipitazioni (in mm di pioggia) nel mese più siccitoso.	97
6.22 Rischio di incendio in relazione alle precipitazioni annuali (in mm di pioggia).	97
6.23 Rischio di incendio estivo in relazione alla densità di animali al pascolo (numero di animali per cella raster).	98
6.24 Rischio di incendio invernale nelle aree a prevalente copertura boschiva in relazione alla densità di animali al pascolo (numero di animali per cella raster).	99
6.25 Rischio di incendio invernale nelle aree a prevalente copertura di prati e pascoli in relazione alla densità di animali al pascolo (numero di animali per cella raster).	99
6.26 Distribuzione media del rischio in Valle d'Aosta per gli incendi estivi.	100
6.27 Distribuzione media del rischio in Valle d'Aosta per gli incendi boschivi invernali.	100
6.28 Distribuzione media del rischio in Valle d'Aosta per gli incendi invernali generati da aree a prevalente copertura di prati e pascoli.	101

---

## Elenco delle tabelle

---

4.1	Elenco delle variabili considerate nella fase esplorativa per il modello di distribuzione del rischi di incendio nell'Amazzonia boliviana . . . . .	30
4.2	Parametri del modello esplorativo utilizzato per la selezione delle variabili da utilizzare per la valutazione del rischio incendio in Amazzonia boliviana. . . . .	31
4.3	Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili del modello esplorativo. . . . .	32
4.4	Confronto tra le variabili con valori di CP e IPP superiori alla soglia (1 %) per la valutazione dei modelli, elaborato per la scelta dello stato informativo relativo alle caratteristiche della vegetazione. Sulla sinistra i risultati per l'elaborazione con la carta della vegetazione potenziale di Navarro (V) e sulla destra rualli per l'elaborazione con la carta delle ecoregioni di Ibisch (E). . . . .	34
4.5	Parametri del modello elaborato con le variabili selezionate per la valutazione del rischio incendio in Amazzonia boliviana. . . . .	35
4.6	Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili utilizzate per l'elaborazione del modello di distribuzione del rischio incendio in Amazzonia boliviana. . . . .	37
4.7	Valori in formato logistico del rischio incendio delle categorie più significative ( $valore \geq 0,6$ ) nelle variabili non continue per il modello elaborato per l'Amazzonia boliviana . . . . .	41

## Elenco delle tabelle

---

5.1	Estensione delle superfici e percentuale rispetto all'area del progetto generale (PASF) delle aree considerate per la valutazione delle performance dei modelli a diversa scala geografica. . . . .	44
5.2	Categorie di suoli (FAO) in relazione alla superficie delle aree considerate per la valutazione delle performance dei modelli a diversa scala geografica. . . . .	45
5.3	Parametri comuni ai modelli elaborati a scala locale per la comparazione delle performance di MAXENT nella distribuzione del rischio incendio in Amazzonia boliviana. . . . .	45
5.4	Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili utilizzate per l'elaborazione del modello di distribuzione del rischio incendio nel dipartimento del Beni. . . . .	48
5.5	Valori in formato logistico del rischio incendio delle categorie più significative ( $valore \geq 0,6$ ) nelle variabili non continue per il modello elaborato per il dipartimento del Beni . . . . .	53
5.6	Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili utilizzate per il modello di distribuzione del rischio incendio nel dipartimento di Santa Cruz. . .	55
5.7	Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili utilizzate per il modello di distribuzione del rischio incendio nell'area 1. . . . .	62
5.8	Valori in formato logistico del rischio incendio delle categorie più significative ( $valore \geq 0,6$ ) nelle variabili non continue per il modello elaborato per l'area 1. . . . .	62
5.9	Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili utilizzate per il modello di distribuzione del rischio incendio nell'area 2. . . . .	65
5.10	Valori in formato logistico del rischio incendio delle categorie più significative ( $valore \geq 0,6$ ) nelle variabili non continue per il modello elaborato per l'area 2. . . . .	66
5.11	Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili utilizzate per il modello di distribuzione del rischio incendio nell'area 3. . . . .	69
5.12	Valori in formato logistico del rischio incendio delle categorie più significative ( $valore \geq 0,6$ ) nelle variabili non continue per il modello elaborato per l'area 3. . . . .	70
6.1	Elenco delle variabili considerate per l'indagine della distribuzione del rischio incendio in Valle d'Aosta. . . . .	80
6.2	Gruppi proxy identificati dalla analisi delle componenti principali in Valle d'Aosta. . . . .	82

6.3	Matrice di correlazione della componente principale PC 2 in Valle d'Aosta. . . . .	82
6.4	Matrice di correlazione della componente principale PC 3 in Valle d'Aosta. . . . .	82
6.5	Matrice di correlazione della componente principale PC 4 in Valle d'Aosta. . . . .	83
6.6	Matrice di correlazione della componente principale PC 11 in Valle d'Aosta. . . . .	83
6.7	Elenco delle variabili selezionate con PCA per l'indagine della distribuzione del rischio incendio in Valle d'Aosta. . . . .	83
6.8	Parametri impostati per i modelli di distribuzione del rischio incendio nei tre dataset della Valle d'Aosta: "dataset E" per gli incendi estivi, "dataset IB" per gli incendi invernali in aree boscate e "dataset PP" per gli incendi invernali su aree a prevalente copertura di prati, pascoli e incolti. . . . .	84
6.9	Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili selezionate nel modello di distribuzione del rischio incendio estivo. . . . .	86
6.10	Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili selezionate nel modello di distribuzione del rischio incendio boschivo invernale. . . . .	90
6.11	Contributo percentuale (CP) e importanza di permutazione percentuale (IPP) delle variabili selezionate nel modello di distribuzione del rischio incendio invernale su prati e pascoli. . . . .	95
7.1	Comparazione delle prestazioni dei modelli elaborati nei casi di studio ( $\sigma$ indica il valore di deviazione standard di AUC nelle reiterazioni di calcolo dell'algoritmo). . . . .	104
7.2	Comparazione delle prestazioni dei modelli elaborati nei casi di studio in relazione al numero di incendi. . . . .	106
7.3	Comparazione delle prestazioni dei modelli elaborati nei casi di studio in relazione alla risoluzione spaziale delle variabili e dell'estensione dell'area di indagine. . . . .	107
8.1	Matrice di analisi SWOT dell'applicazione dell'algoritmo di massima entropia allo studio della distribuzione del rischio per gli incendi forestali. . . . .	114



---

---

## Bibliografia

---

- [1] G. Amatulli, R. Maria João, M. Trombetti, and R. Lovreglio. Assessing long-term fire risk at local scale by means of decision tree technique. *Journal of Geophysical Research: Biogeosciences*, 111:G0405, 2006.
- [2] J.D. Arnold, S.C. Brewer, and P.E. Dennison. Modeling Climate-Fire Connections within the Great Basin and Upper Colorado River Basin, Western United States. *Fire Ecology*, 10(2):64–75, 2014.
- [3] M.P. Austin. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, 157(2-3):101–118, 2002.
- [4] R. Baldwin. Use of Maximum Entropy Modeling in Wildlife Research. *Entropy*, 11(4):854–866, nov 2009.
- [5] A. Bar Massada, A.D. Syphard, S.I. Stewart, and V.C. Radeloff. Wildfire ignition-distribution modelling: a comparative study in the Huron-Manistee National Forest, Michigan, USA. *International Journal of Wildland Fire*, 22(2):174, 2013.
- [6] D.A. Belsley. *Conditioning diagnostics*. Wiley Online Library, 1991.
- [7] W.J. Bond and J.E. Keeley. Fire as a global 'herbivore': the ecology and evolution of flammable ecosystems. *Trends in ecology & evolution*, 20(7):387–94, jul 2005.
- [8] G.D. Booth, M.J. Niccolucci, and E.G. Schuster. Identifying proxy sets in multiple linear-regression - an aid to better coefficient interpretation. *USDA Forest Service Intermountain Research Station Research Paper*, pages 1–13, 1994.
- [9] G. Bovio and E. Marchi. Selvicoltura: incendi, pascolo. *L'Italia Forestale e Montana*, 2(65):113–119, 2010.
- [10] A. Calle, F. Gonzalez-Alonso, and S.M. de Miguel. Validation of active forest fires detected by MSG-SEVIRI by means of MODIS hot spots and AWiFS images. *International Journal of Remote Sensing*, 29(12):3407–3415, 2008.
- [11] J.A. Cardille and S.J. Ventura. Occurrence of wildfire in the northern Great Lakes Region: Effects of land cover and land ownership assessed at multiple scales. *International Journal of Wildland Fire*, 10(2):145–154, 2001.
- [12] F. Chen, Y. Du, S. Niu, and J. Zhao. Modeling Forest Lightning Fire Occurrence in the Daxinganling Mountains of Northeastern China with MAXENT. *Forests*, 6(5):1422–1438, 2015.
- [13] Y.H. Chou. Management of wildfires with a geographical information system. *International Journal of Geographical Information Systems*, 6(2):123–140, 1992.

## Bibliografía

---

- [14] G. Cruz-Cárdenas, L. López-Mata, J.L. Villaseñor, and E. Ortiz. Potential species distribution modeling and the use of principal component analysis as predictor variables. *Revista Mexicana de Biodiversidad*, 85(1):189–199, 2014.
- [15] T.E. Dilts, J.S. Sibold, and F. Biondi. A Weights-of-Evidence Model for Mapping the Probability of Fire Occurrence in Lincoln County, Nevada. *Annals of the Association of American Geographers*, 99(4):712–727, 2009.
- [16] W.M. Dlamini. A Bayesian belief network analysis of factors influencing wildfire occurrence in Swaziland. *Environmental Modelling & Software*, 25(2):199–208, 2010.
- [17] L.R. Donoghue and W.A. Main. Some factors influencing wildfire occurrence and measurement of fire prevention effectiveness. *Journal of Environmental Management*, 20(1):87–96, 1985.
- [18] C.F. Dormann, J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J.R.G. Marquéz, B. Gruber, B. Lafourcade, P.J. Leitão, T. Münkemüller, C. McClean, P.E. Osborne, B. Reineking, B. Schröder, A.K. Skidmore, D. Zurell, and S. Lautenbach. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1):027–046, 2013.
- [19] A. Duane, M. Piqué, and M. Castellnou. Predictive modelling of fire occurrences from different fire spread patterns in Mediterranean landscapes. *International Journal of Wildland Fire*, 24(3):407–418, 2015.
- [20] B. Efron. *The Jackknife, the Bootstrap and Other Resampling Plans*. 1982.
- [21] J. Elith, C.H. Graham, R.P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R.J. Hijmans, F. Huettmann, J.R. Leathwick, A. Lehmann, J. Li, L.G. Lohmann, B.A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J.M.C. Overton, A.T. Peterson, S.J. Phillips, K. Richardson, R. Scachetti-Pereira, R.E. Schapire, J. Soberon, S. Williams, M.S. Wisz, and N.E. Zimmermann. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(January):129–151, 2006.
- [22] J. Elith and J.R. Leathwick. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1):677–697, 2009.
- [23] J. Elith, S.J. Phillips, T. Hastie, M. Dudík, Y. En Chee, and C.J. Yates. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17(1):43–57, jan 2011.
- [24] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.
- [25] A.H. Fielding and J.F. Bell. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, null:38–49, 3 1997.
- [26] J. Franklin, P. McCullough, and C. Gray. Terrain variables used for predictive mapping of vegetation communities in southern california. *Terrain analysis: principles and applications/edited by John P. Wilson, John C. Gallant*, 2000.
- [27] M.G. Genton, D.T. Butry, M.L. Gumpertz, and J.P. Prestemon. Spatio-temporal analysis of wildfire ignitions in the St Johns River Water Management District, Florida. *International Journal of Wildland Fire*, 15(1):87–97, 2006.
- [28] L. Giglio, T. Loboda, D.P. Roy, B. Quayle, and C.O. Justice. An active-fire based burned area mapping algorithm for the MODIS sensor. *Remote Sensing of Environment*, 113(2):408–420, 2009.
- [29] A. Giuliani and J.P. Zbilut. *L'ordine della complessità*. Di fronte e attraverso. Jaca Book, 2009.
- [30] T.J. Hawbaker, V.C. Radeloff, A.D. Syphard, Z. Zhu, and S.I. Stewart. Detection rates of the MODIS active fire product in the United States. *Remote Sensing of Environment*, 112(5):2656–2664, 2008.
- [31] S.B. Hecht, Rockefeller Foundation, and Centro Internacional de Agricultura Tropical. *Amazonia, Agriculture and Land Use Research: Proceedings of the International Conference*. CIAT series ; 03E3-82. Centro Internacional de Agricultura Tropical, 1982.

- [32] A.S. Hering, C.L. Bell, and M.G. Genton. Modeling spatio-temporal wildfire ignition point patterns. *Environmental and Ecological Statistics*, 16(2):225–250, 2009.
- [33] P.A. Hernandez, C.H. Graham, L.L. Master, D.L. Albert, and A.D.L. The. The effect of sample size and species characteristics on performance of different species distribution modeling methods. 5(June):773–785, 2006.
- [34] R.J. Hijmans, S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. WORLDCLIM - a set of global climate layers (climate grids). *International Journal of Climatology*, 25:1965–1978, 2005.
- [35] P.L. Ibisch, S.G. Beck, B. Gerkmann, A. Carretero, and G. Mérida. Ecoregiones y ecosistemas. *Biodiversidad: la riqueza de Bolivia. Estado de conocimiento y conservación. Ministerio de Desarrollo Sostenible y Planificación, Editorial FAN, Santa Cruz*, 2003.
- [36] E.T. Jaynes. Where Do We Stand on Maximum Entropy? *The Maximum Entropy Formalism*, 3:15 – 118, 1978.
- [37] E.T. Jaynes. On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 70(9):939–952, 1982.
- [38] A. Jiménez-Valverde and J.M. Lobo. Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica*, 31(3):361–369, May 2007.
- [39] A. Jiménez-Valverde, J.M. Lobo, and J. Hortal. Not as good as they seem: The importance of concepts in species distribution modelling. *Diversity and Distributions*, 14(6):885–890, 2008.
- [40] T. Kuemmerle, K. Perzanowski, O. Chaskovskyy, K. Ostapowicz, L. Halada, A.T. Bashta, I. Krullov, P. Hostert, D.M. Waller, and V.C. Radeloff. European Bison habitat in the Carpathian Mountains. *Biological Conservation*, 143(4):908–916, 2010.
- [41] Levin, S.A. The problem of pattern and scale in ecology. *Ecology*, 73(6)(1165):1943–1967, 1992.
- [42] Z. Li, Y.J. Kaufman, C. Ichoku, R. Fraser, A. Trishchenko, L. Giglio, J. Jin, and X. Yu. A review of AVHRR-based active fire detection algorithms: Principles, limitations, and recommendations. *Int. J. Remote Sensing*, 20(10):1947–1985, 1999.
- [43] J.M. Lobo, A. Jiménez-Valverde, and J. Hortal. The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, 33(1):103–114, 2010.
- [44] J.M. Lobo, A. Jiménez-valverde, and R. Real. AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17(2):145–151, 2008.
- [45] J.K. Maingi and M.C. Henry. Factors influencing wildfire occurrence and distribution in eastern Kentucky, USA. *International Journal of Wildland Fire*, 16(1):23–33, 2007.
- [46] S. Manel, H.C. Williams, and S.J. Ormerod. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, 38(5):921–931, 2001.
- [47] T. Manyangadze. Forest fire detection for near real-time monitoring using geostationary satellites. *International Institute for Geo-Information Science and Earth Observation*, M.Sc.:80, 2009.
- [48] J. Martínez, C. Vega-García, and E. Chuvieco. Human-caused wildfire risk rating for prevention planning in Spain. *Journal of environmental management*, 90(2):1241–52, 2009.
- [49] B. McCune and D. Keon. Equations for potential annual direct incident radiation and heat load. *Journal of Vegetation Science*, 13(1966):603–606, 2002.
- [50] M. Meloun, J. Militký, M. Hill, and R.G. Brereton. Crucial problems in regression modelling and their solutions, 2002.
- [51] R.G. Miller. *The Jackknife - A Review*, volume 61. 1974.
- [52] G. Navarro and W. Ferreira. Zonas de Vegetación Potencial de Bolivia: Una Base para el Analisis de Vacíos de Conservación. *Rev. Bol. Ecol.* 15, pages 1–40, 2004.
- [53] M.A. Parisien and M.A. Moritz. Environmental controls on the distribution of wildfire at multiple spatial scales. *Ecological Monographs*, 79(1):127–154, 2009.

## Bibliografia

---

- [54] M.A. Parisien, S. Snetsinger, J.A. Greenberg, C.R. Nelson, T. Schoennagel, S.Z. Dobrowski, and M.A. Moritz. Spatial variability in wildfire probability across the western United States. *International Journal of Wildland Fire*, 21(4):313–327, 2012.
- [55] J. Paritsis, A. Holz, T.T. Veblen, and T. Kitzberger. Habitat distribution modeling reveals vegetation flammability and land use as drivers of wildfire in SW Patagonia. *Ecosphere*, 4(May):1–20, 2013.
- [56] M. J. Perestrello De Vasconcelos, S. Silva, M. Tome, M. Alvim, and J.M.C. Pereira. Spatial Prediction of Fire Ignition Probabilities : Comparing Logistic Regression and Neural Networks. *Photogrammetric Engineering & Remote Sensing*, 67(1):73–81, 2001.
- [57] K.L. Pew and C.P.S. Larsen. GIS analysis of spatial and temporal patterns of human-caused wildfires in the temperate rain forest of Vancouver Island, Canada. *Forest Ecology and Management*, 140(1):1–18, 2001.
- [58] S.J. Phillips, R.P. Anderson, and R.E. Schapire. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3-4):231–259, jan 2006.
- [59] S.J. Phillips and M. Dudík. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31(2):161–175, 2008.
- [60] M.P. Plucinski. A review of wildfire occurrence research. CSIRO Report. Technical report, 2012.
- [61] J. Podur, D.L. Martell, and F. Csillag. Spatial patterns of lightning-caused forest fires in Ontario, 1976–1998. *Ecological Modelling*, 164(1):1–20, 2003.
- [62] M. H. Quenouille. Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(1):66–84, 1949.
- [63] B.R. Ramesh, Q. Renard, and N. Kodandapani. Environmental susceptibility model for predicting forest fire occurrence in the Western Ghats of India . . . *Journal of Wildland Fire*, 2012.
- [64] I.W. Renner and D.I. Warton. Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology. *Biometrics*, 69(1):274–281, 2013.
- [65] C.E. Shannon. *The mathematical theory of communication.*, volume 14. 1948.
- [66] A.D. Syphard, V.C. Radeloff, N.S. Keuler, R.S. Taylor, T.J. Hawbaker, S.I. Stewart, and M.K. Clayton. Predicting spatial patterns of fire on a southern California landscape. *International Journal of Wildland Fire*, 17(5):602–613, 2008.
- [67] R. Turner. Point patterns of forest fire locations. *Environmental and Ecological Statistics*, 16(2):197–223, 2009.
- [68] Y. Wang and K.R. Anderson. An evaluation of spatial and temporal patterns of lightning-and human-caused forest fires in Alberta, Canada, 1980–2007. *International Journal of Wildland Fire*, 19(8):1059–1072, 2010.
- [69] G. Ward, T. Hastie, S. Barry, J. Elith, and J.R. Leathwick. Presence-only data and the em algorithm. *Biometrics*, 65(2):554–563, 2009.
- [70] B.A. Wintle, M.A. McCarthy, K.M. Parris, and M. Burgman. Precision and bias of methods for estimating point survey detection probabilities. *Ecological Applications*, 14(3):703–712, 2004.
- [71] M.S. Wisz, R.J. Hijmans, J. Li, A.T. Peterson, C.H. Graham, A. Guisan, J. Elith, M. Dudík, S. Ferrier, F. Huettmann, J.R. Leathwick, A. Lehmann, L. Lohmann, B.A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M.C. Overton, S.J. Phillips, K.S. Richardson, R. Scachetti-Pereira, R.E. Schapire, J. Soberón, S.E. Williams, and N.E. Zimmermann. Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, 14(5):763–773, 2008.
- [72] J. Yang, H.S. He, S.R. Shifley, and E.J. Gustafson. Spatial patterns of modern period human-caused fire occurrence in the Missouri Ozark Highlands. *Forest Science*, 53(1):1–15, 2007.