

AZIONE A3

**MEMORANDUM INTERPRETATIVO PER MODELLI
FAUNISTICI IN OUTPUT ALL'AZIONE A3**

Indice

1	Modelli	3
1.1	Teoria alla base dei modelli e della conseguente validazione	3
	MAXENT, ovvero la ricerca della migliore distribuzione probabilistica con dati disponibili incompleti. Modelli <i>random</i>	3
	Le curve ROC, ovvero, discernere tra falso allarme e allarme vero: verifica di affidabilità . . .	4
1.2	Software e metodologie	5
1.3	Validazione, interpretazione e limiti	8
1.4	Sviluppi futuri	8
	Bibliografia	11

Coordinamento progetto LIFE+T.E.N.:

Claudio Ferrari

PROVINCIA AUTONOMA DI TRENTO

INCARICO DIRIGENZIALE PER LA VALORIZZAZIONE DELLA RETE DELLE AREE PROTETTE

claudio.ferrari@provincia.tn.it

Coordinamento Azione A3:

Paolo Pedrini / MUSE, *paolo.pedrini@muse.it*

Relazione a cura di:

Aaron Iemma, *aaron.iemma@muse.it*

Paolo Pedrini, *paolo.pedrini@muse.it*

1 Modelli

Uno dei principali obiettivi dell’Azione A3 prevede lo sviluppo e l’utilizzo di procedure statistiche che possano efficacemente *modellizzare* (ovvero, fornire validi e verificabili indizi) la distribuzione potenziale di un *set* di *specie focali* opportunamente individuate all’interno della Azione A2.

Il presente documento costituisce quindi una linea guida interpretativa dedicata ai principali *output* dell’Azione A3. Lo stesso è organizzato nelle seguenti:

- Breve teoria alla base dei modelli;
- Software e metodologie impiegate;
- Validazione, interpretazione e limiti;
- Aggiornamenti e sviluppi futuri;
- Bibliografia.

1.1 Teoria alla base dei modelli e della conseguente validazione

MAXENT, ovvero la ricerca della migliore distribuzione probabilistica con dati disponibili incompleti. Modelli *random*

Il modello adottato per le analisi basa la sua forza su alcune importanti assunzioni:

- che i rilevamenti delle specie trattate possano essere considerati alla stregua di “*particelle indipendenti l’una dall’altra*”;
- che le stesse particelle siano *indistinguibili* l’una dall’altra;
- che le stesse specie tendano a disporsi sul territorio occupando il più grande areale *possibile*, ...
- ... dove la “*possibilità di occupare un certo areale*” è soggetta a limiti imposti da variabili ambientali disponibili per la totalità del territorio considerato alla migliore risoluzione possibile (*i.e.*, DTM, copertura arborea, presenza di centri abitati, ...), e, infine ...
- ... che l’occupazione del territorio sia un processo governato, all’interno delle limitazioni precedenti, da un processo casuale.

Considerando le precedenti assunzioni, la modellizzazione della distribuzione delle specie coincide con la ricerca della distribuzione che *massimizza l’entropia della stessa distribuzione*, ove con “*entropia*” si intende la quantità di informazione associata ad un certo messaggio.[5]

Se il nostro messaggio è rappresentato dalla presenza o meno di una certa specie sul territorio, il modello cerca quindi di massimizzare l’informazione contenuta nel messaggio, tenendo conto dei vincoli imposti dalle variabili ambientali: non essendo possibile fare alcuna assunzione *a priori* riguardo alla distribuzione della specie, si sceglie quindi di modellizzare la distribuzione specie-specifica scegliendo il processo che meglio rappresenta lo stato attuale di conoscenza delle dinamiche della specie, ovvero, quello di una variabile casuale soggetta a vincoli ambientali[2]: in questo senso, la nostra intuizione è confermata dal *principio della massima entropia*, secondo il quale:

Teorema 1.1.1 (Principio della Massima Entropia) *La distribuzione di probabilità di una variabile casuale soggetta a delle proposizioni stabilite a priori che meglio rappresenta la conoscenza attuale del problema è proprio quella con la massima entropia [1].*

Da un punto di vista matematico, il processo di modellizzazione corrisponde quindi ad un *algoritmo di Gibbs*, nel quale si cerca di minimizzare il logaritmo della probabilità media dell'entropia soggetta ad una distribuzione di probabilità che rispetta un certo numero di vincoli. Ovvero:

$$\min(H = \sum_i -p_i \ln p_i) \quad (1.1)$$

Rimandando considerazioni di maggior dettaglio alla bibliografia specifica segnalata, si evidenzia comunque che:

Tale metodo è attualmente considerato una delle più efficaci tecniche per modellizzare la distribuzione di specie animali e vegetali e i modelli ottenuti utilizzando variabili ambientali di dettaglio appaiono particolarmente accurati, in grado di stimare la reale qualità ambientale per le specie, al punto da essere in grado di predire anche alcuni parametri riproduttivi e non solo la presenza/assenza delle specie

cit. Clara Tattoni, Mattia Brambilla & Paolo Pedrini - Report Azione A3

Le curve ROC, ovvero, discernere tra falso allarme e allarme vero: verifica di affidabilità

Una volta prodotto, il modello statistico ha bisogno di essere validato: è quindi necessario istituire un opportuno *test* che sia in grado di discriminare la bontà dello stesso modello.

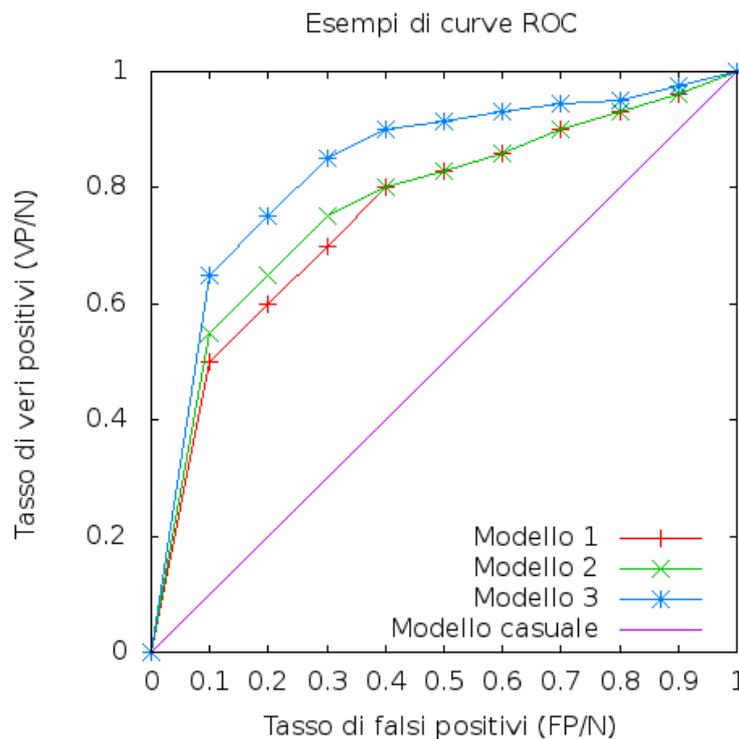


Figura 1.1: Esempi di curve ROC per vari modelli. Da notare come tutti i modelli “spiegano” meglio la variabilità (hanno un’area sottesa alla curva ROC superiore) dei dati osservati rispetto ad un modello che scelga i punti in maniera del tutto casuale (linea retta viola)

Il metodo considerato in questo lavoro è quello che considera (AUC: *Area Under Curve*) l’area totale sottesa ad una curva (la curva ROC, *Receiver Operating Characteristics*) che rappresenta in ogni suo punto il rapporto tra il numero di falsi positivi rispetto ai veri positivi che risultano dall’applicazione di un dato classificatore ad un set di punti classificabili [9]. Nel nostro caso:

Punti classificabili : sono i punti di presenza di una certa specie usati nella produzione del modello;

Classificatore : il modello della specie stessa.

Falso positivo : punto che secondo il modello è nell'areale di presenza potenziale della specie, ma che nella realtà non vi rientra.

Vero positivo : punto che secondo il modello è nell'areale potenziale della specie e che vi rientra effettivamente.

Intuitivamente inoltre, il motivo per cui modelli con **AUC** maggiore forniscono una rappresentazione più fedele della realtà si coglie facilmente osservando la figura di cui sopra: semplicemente, più area è sottesa ad una curva, più essa si avvicina alla curva del *classificatore perfetto* (la retta verticale dal punto (0,0) che restituisce solo veri positivi).

1.2 Software e metodologie

I rilievi puntiformi delle singole specie identificate nell'azione A3 (vedi [8]) e/o ritenute di particolare interesse conservazionistico sono stati estratti dal *database* costruito durante il corso dell'Azione A1 (per i dettagli sull'origine della banca dati si rimanda alla relazione specifica [6]). I criteri adottati nell'elaborazione e selezione dei modelli seguono in primo luogo tre linee:

- specie con almeno 30 osservazioni;
- incertezza del singolo rilievo inferiore a 200 m (dato preciso);
- rilievo successivo all'anno 2000 (dato recente).

Per l'elaborazione dei modelli sono state considerate alcune variabili ambientali derivate dalla cartografia digitale fornita dalla Provincia Autonoma di Trento, scaricate dal portale provinciale nel Marzo 2013. Tutta la cartografia è quindi aggiornata a quella disponibile in quella data: ove alcune variabili sono state elaborate altre sono state utilizzate così come fornite dalla PAT. Le si indica di seguito, assieme alle elaborazioni di cui sono state eventualmente fatte oggetto rispetto alla carta di uso del suolo (il cui ultimo aggiornamento ufficiale disponibile risale al 2003):

- Quota - DTM (dato originale);
- Esposizione (elaborata dal DTM, riclassificata in 8 classi);
- Pendenza (elaborata dal DTM);
- Cartografia di uso del suolo (riclassificata accorpendo alcune tipologie);
- Carta dei tipi forestali (riclassificata accorpendo alcune tipologie);
- Distanza dai fiumi (elaborata da carta di uso del suolo);
- Distanza dai laghi (elaborata da carta di uso del suolo);
- Distanza da viabilità principale (elaborata da carta di uso del suolo);
- Distanza da viabilità forestale (elaborata da carta di uso del suolo);
- Distanza da case e aree urbane (elaborata da carta di uso del suolo).

Per gli ambienti aperti sono state utilizzate, in aggiunta alle precedenti, delle variabili interamente derivate dal rilievo LiDaR provinciale:

- percentuale di alberi in aree aperte;
- superficie di un'area aperta;
- altezza di alberi e cespugli;
- lunghezza dell'ecotono.

I modelli delle singole specie sono stati ottenuti integralmente tramite elaborazioni con il software GIS GRASS v. 6.4.1, in congiunzione al *plugin Java maxent*, dell'Università di Princeton ¹ [4] [3]. Il risultato del modello è una scala continua di probabilità di presenza compresa tra zero e uno. Per facilitare la lettura dell'elaborato cartografico si è riclassificato questo risultato in 4 classi di idoneità dell'*habitat*: nulla (*nodata*), bassa (1), media (2) ed elevata (3).

Successivamente, le cartografie potenziali delle singole specie sono state sovrapposte e *sommate* tra loro, con l'obiettivo di ottenere un'unica cartografia di sintesi del numero di specie potenzialmente presenti (suddivise quindi per ogni ambiente considerato) per ogni cella quadrata della mappa, di 1 km di lato. Per quanto riguarda le aree rupestri di bassa quota, data la loro estensione relativamente modesta e la localizzazione isolata, si è utilizzato un approccio differente, che ha permesso di produrre delle carte di ricchezza *reale*, non potenziale. Per gli ambienti rupestri, incrociando la distribuzione degli habitat con quella delle specie, è stato possibile associare ai diversi contesti rupicoli/rupestri le specie presenti.

In sostanza quindi, questo *addendum* giunge di corredo alla cartografia allegata all'Azione A3:

- Modelli specie-specifici delle specie focali individuate dall'Azione A2²:

- Ululone dal ventre giallo, *Bombina variegata*
- Re di quaglie, *Crex crex*
- Gipeto, *Gypaetus barbatus*
- Averla piccola, *Lanius collurio*
- Succiacapre, *Caprimulgus europaeus*
- Civetta nana, *Glaucidium passerinum*
- Aquila reale, *Aquila chrysaetos*
- Picchio cenerino, *Picus canus*
- Gallo forcello, *Tetrao tetrix*
- Gallo cedrone, *Tetrao urogallus*
- Picchio nero, *Dryocopus martius*
- Francolino di monte, *Bonasa bonasia*
- Civetta capogrosso, *Aegolius funereus*
- Nibbio bruno, *Milvus migrans*
- Falco pecchiaiolo, *Pernis apivorus*
- Pernice bianca, *Lagopus muta*
- Orso bruno, *Ursus arctos*

- Modelli di altre specie focali rappresentative di un dato *taxa*

- Astore, *Accipiter gentilis*,
- Cannaiola, *Acrocephalus scirpaceus*,
- Coturnice, *Alectoris graeca*,
- Germano reale, *Anas platyrhynchos*,
- Airone cenerino, *Ardea cinerea*,
- Rospo comune, *Rospo comune*,
- Poiana, *Buteo buteo*,
- Usignolo di fiume, *Cettia cetti*,
- Rampichino comune, *Certhia brachydactyla*,

¹<http://www.cs.princeton.edu/~schapire/maxent/>

²con caratteristiche compatibili a quanto detto sopra: più di 30 avvistamenti nel *dataset* dell'azione A1, precisione del rilievo superiore ai 200m e dato campionato dall'anno 2000 in poi

- Merlo acquaiolo, *Cinclus cinclus*,
 - Rampichino alpestre, *Certhia familiaris*,
 - Quaglia comune, *Coturnix coturnix*,
 - Picchio rosso maggiore, *Dendrocopos major*,
 - Zigolo giallo, *Emberiza citrinella*,
 - Folaga, *Fulica atra*,
 - Gheppio comune, *Falco tinnunculus*,
 - Torcicollo, *Jynx torquilla*,
 - Marmotta delle Alpi, *Marmota marmota*,
 - Culbianco, *Oenanthe oenanthe*,
 - Assiolo, *Otus scops*,
 - Picchio verde, *Picus viridis*,
 - Svasso maggiore, *Podiceps cristatus*,
 - Sordone, *Prunetta collaris*,
 - Porciglione eurasiatico, *Rallus aquaticus*,
 - Salamandra pezzata, *Salamandra salamandra*,
 - Stiaccino, *Saxicola rubetra*,
 - Picchio muratore, *Sitta europaea*,
 - Tuffetto comune, *Tachybaptus ruficollis*,
 - Merlo dal collare, *Turdus torquatus*,
 - Upupa, *Upupa epops*,
 - Lucertola vivipara, *Zootoca vivipara*,
 - Codirosso, *Phoenicurus phoenicurus*,
 - Cincia dal ciuffo, *Lophophanes cristatus*,
 - Fanello, *Carduelis cannabina*,
 - Gallinella d'acqua, *Gallinula chloropus*,
 - Fringuello alpino, *Montifringilla nivalis*,
 - Cinciarella, *Cyanistes caeruleus*,
 - Cincia alpestre, *Poecile montanus*,
- Ricchezza potenziale per ogni *habitat*, che mostra in ogni *pixel* di 50x50m del territorio Trentino la somma delle specie che hanno un *potenziale di insediamento* da modello favorevole o molto favorevole in quel punto;
 - Aree umide
 - Boschi perifluviali
 - Colture arboree
 - Conifere
 - Boschi latifoglie
 - Ricchezza reale (somma del numero di specie rilevate dagli anni 2000 in poi contenute nel *database* dell'Azione A1) per le sole aree rocciose.
 - Ricchezza totale risultante dalla sovrapposizione di *tutti* i modelli

Ogni cartografia giunge in formato sia *raster* (.tiff) che vettoriale (.shp), nel sistema di riferimento provinciale ETRS89 - UTM32N.

Altre mappe di corredo utili alla riproduzione completa delle Tavole in allegato alla relazione A3 sono ugualmente fornite.

1.3 Validazione, interpretazione e limiti

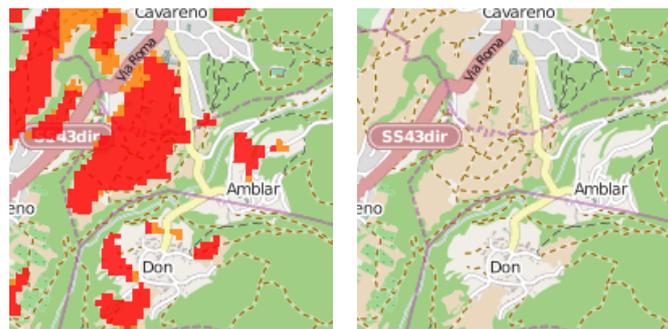


Figura 1.2: Modello di Re di quaglie *Crex crex* nel WebGIS (Azione A1), per un'area dell'Alta Val di Non.

■: alta probabilità, ■: media probabilità.

Ogni modello prodotto è stato susseguentemente validato attraverso l'analisi delle curve ROC, usando il parametro derivato AUC come discriminante tra modelli accurati e non accurati. I modelli elaborati in questo studio hanno in media una AUC del 90%, che significa che il 90% dei punti di presenza utilizzati per l'elaborazione del modello ricadono correttamente nella classi di idoneità media ed elevata del modello. Una AUC del 90% è considerata una diagnostica del modello molto soddisfacente, pur rimanendo una quota di punti di presenza della specie, circa il 10% a seconda dei casi, che non viene correttamente interpretata dal modello. Quando si sovrappone il risultato cartografico del modello con i punti di presenza reali è quindi possibile che ci sia qualche osservazione che non corrisponde con l'habitat potenziale individuato, ma questo non inficia l'attendibilità del modello.

La definizione della soglia che discrimina le quattro classi di idoneità dell'habitat è stata stabilita attraverso l'analisi ROC, cercando di minimizzare gli inevitabili falsi positivi. Trattandosi del risultato di un modello (affetto da un errore statistico intrinseco, per quanto minimo sempre presente), è importante che sia ridotta al minimo la possibilità di errore *considerata più grave* dal punto di vista gestionale. Applicando quindi, come buona norma, il principio di precauzione (per il quale in questo caso l'errore più grave sarebbe affermare che un habitat non sia idoneo quando invece lo è, ovvero avere dei falsi negativi). Il contrario dell'errore di primo tipo, in questo caso, consiste nell'avere tanti falsi positivi, ovvero affermare che un habitat è idoneo o molto idoneo ma la specie non vi è stata osservata. In conclusione:

Non essendo possibile minimizzare sia i falsi negativi che i falsi positivi, si è preferito optare per un risultato che può sovrastimare leggermente l'area adatta ad una specie, piuttosto che prediligere un risultato che possa predire un areale per la specie minore di quello reale.

Per la corretta interpretazione del risultato cartografico è quindi importante conoscere le scelte che sono alla base della suddivisione in quattro classi, così come la scala di interpretazione, che nel caso del presente studio è 1:25000. Una visualizzazione a scala più dettagliata (1:10000 o superiore) può portare a delle interpretazioni non sempre corrette.

Infine si ricorda che l'applicazione dei modelli per le specie al limite del proprio areale distributivo, come ad esempio il Re di quaglie *Crex crex*, è generalmente affetta da un errore maggiore.

1.4 Sviluppi futuri

L'approccio qui studiato si presta molto bene ad essere automatizzato in molti punti, dalla creazione dei modelli fino alla loro validazione: è quindi ferma intenzione della Sezione predisporre un adeguato *framework* di aggiornamento semiautomatico dei modelli, in modo da poter produrre mappe da modello tanto più affidabili ed aggiornate quanto più i dati specie-specifici e relativi alle covariate ambientali sono disponibili.

Inutile dire che una simile prospettiva ben si aggancia ai possibili sviluppi dei risultati dell’Azione A1 del progetto LIFE+T.E.N: il *database* risultante dell’Azione, costruito ed attualmente gestito dalla Sezione di Zoologia dei Vertebrati, si presta ad essere una fonte tenuto in costante aggiornamento, tramite l’adozione di adeguati protocolli standardizzati di campionamento e trasmissione dei dati biologici, eventualmente studiati *ad hoc* per ogni singolo ente.

Un *output* altrettanto collegato, anche se meno specialistico ma ugualmente importante, è la capacità di strumenti di visualizzazione cartografica *online* (come ad esempio lo stesso WebGIS prodotto dall’Azione A1) di restituire all’utente comune le cartografie di distribuzione potenziale costantemente aggiornate e liberamente consultabili, in una prospettiva a lungo termine che miri a sensibilizzare la cittadinanza attiva riguardo la grande varietà e interconnessione dell’ambiente trentino.

Bibliografia

- [1] Jaynes, E.T. (1957). *Information theory and statistical mechanics* (PDF). Physical Review 106 (4): 620–630. doi:10.1103/PhysRev.106.620
- [2] Elith, J. *et al.* (2011) *A statistical explanation of MaxEnt for ecologists* (PDF)
- [3] Phillips, S. J., R. P. Anderson, and R.E. Schapire (2006) *Maximum entropy modeling of species geographic distributions* (PDF). Ecological Modelling, 190:231-259.
- [4] Phillips, S. J., Dudík, M., Schapire, R. E. *A maximum entropy approach to species distribution modeling.* (PDF), Proceedings of the Twenty-First International Conference on Machine Learning, pp. 655-662, 2004.
- [5] Shannon, E. C. *A Mathematical Theory of Communication* (PDF) The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, Ottobre, 1948.
- [6] Iemma, A., Tattoni, C., Pedrini, P., (2013) *Realizzazione di una banca dati su specie e Habitat delle direttive "Uccelli" e "Habitat"* (PDF) MUSE - Museo delle Scienze, sez. Zoologia dei Vertebrati.
- [7] Brambilla, M., Pedrini, P., (2013) *Individuazione delle priorità di conservazione per specie e habitat delle direttive "Uccelli" e "Habitat"* (PDF) MUSE - Museo delle Scienze, sez. Zoologia dei Vertebrati.
- [8] Tattoni, C., Brambilla, M., Pedrini, P. (2013) *Individuazione della connettività e della frammentazione ecologica a livello provinciale e verso i territori limitrofi* (PDF) MUSE - Museo delle Scienze, sez. Zoologia dei Vertebrati.
- [9] Fawcett, T. (2005) *An introduction to ROC analysis* (PDF) Pattern Recognition Letters, Vol.27, pp. 861–874, 2006.