

UNIVERSITÀ DEGLI STUDI DI TRENTO

Facoltà di Ingegneria

Corso di Laurea

in Ingegneria per l'Ambiente e il Territorio

Tesi di Laurea

**METODOLOGIE GIS PER LA VALUTAZIONE
PREDITTIVA DEI MIGLIORAMENTI
AMBIENTALI SU PICCOLA SCALA**

Relatori

Prof. G. Battista Benciolini

Prof. Maria Giulia Cantiani

Correlatori

Dott. Cesare Furlanello

Dott. Marco Ciolli

Laureando

Stefano Menegon

Anno Accademico 1998/1999

DOI:

<https://dx.doi.org/10.5281/zenodo.845721>

Metodologie GIS per la valutazione predittiva dei miglioramenti ambientali su piccola scala by Stefano Menegon is licensed under a Creative Commons Attribution 4.0 License <https://creativecommons.org/licenses/by/4.0/>.

Ringraziamenti

Desidero ringraziare innanzitutto il dott. Cesare Furlanello per avermi guidato e aiutato in questo lavoro, i miei relatori: il prof. G. Battista Benciolini e la prof.ssa Maria Giulia Cantiani per la disponibilità e per l'attenzione offerta, in particolare nell'ultimo e delicato periodo, e il dott. Stefano Merler per gli importanti consigli. Desidero ricordare anche i responsabili del servizio faunistico per la fondamentale collaborazione.

Un grazie sincero va poi agli amici Roberta, Lorenzo, Beatrice, Marco, Elena e Paola per la loro grande disponibilità offerta nei mesi trascorsi all'Irst e soprattutto nelle settimane finali della tesi. Voglio poi ringraziare Marianna, Enrico e i miei genitori per la comprensione e l'aiuto concreto offerto. Grazie ai miei compagni di appartamento al dott. Bruno Caprile per i preziosi consigli e all'ing. Roberto Flor.

Ringrazio quindi l'Istituto di Ricerca Scientifica e Tecnologica di Trento per avermi ospitato e messo a disposizione le attrezzature.

Indice

Sommario	xi
1 Introduzione: miglioramenti ambientali e metodologie GIS	1
1.1 I problemi ambientali del territorio montano in Trentino	2
1.2 Gli obiettivi della gestione faunistica e i miglioramenti ambientali	2
1.3 Gli strumenti software GIS per il monitoraggio ambientale in Trentino	3
1.4 Definizioni	4
1.4.1 Sistema GIS integrato	6
1.4.2 Strutture e formati dei dati	8
1.5 Basi dati cartografiche in Trentino	11
1.5.1 DTM	11
1.5.2 Carta topografica generale 1:10000	12
1.5.3 OFD	13
1.6 Esempi dimostrativi	13
1.6.1 Esempio 1: calcolo dell'interfaccia	14
1.6.2 Esempio 2: indici di struttura del paesaggio	15
2 I miglioramenti ambientali per il fagiano di monte	19
2.1 Caratteristiche della specie	19
2.1.1 Descrizione della specie	20
2.1.2 Areale di distribuzione	22
2.1.3 Arene di canto	22
2.1.4 Dinamiche della popolazione	24
2.2 Il problema della gestione faunistica del fagiano di monte	26
2.2.1 Data base dei censimenti estivi	27

2.3	Data base dei censimenti primaverili	28
3	Il nuovo data base informatizzato	37
3.1	Strumenti software	37
3.2	Data base	40
3.2.1	Organizzazione del data base	40
3.2.2	L'informatizzazione dei censimenti estivi 1998 e 1999	43
3.3	Cartografia vettoriale delle aree campione	45
3.4	Tematismi GIS e grafici riassuntivi	46
4	Metodi di ecologia del paesaggio	52
4.1	Approccio dell'ecologia del paesaggio	52
4.1.1	Definizione di paesaggio	53
4.1.2	Struttura del paesaggio	54
4.1.3	Macchie paesistiche	55
4.1.4	Concetti di spazio, tempo e scala	56
4.2	Le variabili di ecologia del paesaggio	57
4.2.1	I dati che descrivono il paesaggio	57
4.2.2	Analisi dei patches	58
4.3	Analisi delle scale	62
4.3.1	Analisi delle distanze	64
4.3.2	Modelli frattali e configurazioni spaziali	66
4.4	Strumenti informatici di calcolo	68
5	Classificazione di ortofoto digitali	71
5.1	Classificazione supervisionata	72
5.1.1	Classificatore alla massima verosimiglianza	73
5.1.2	Classificatore SMAP	74
5.1.3	Confronto dei risultati	75
5.2	Classificazione non supervisionata	77
5.3	Applicazione alla classificazione delle aree campione	82
6	Modelli per la valutazione dei miglioramenti ambientali	86
6.1	La costruzione del data base per i modelli di miglioramento ambientale	87

6.1.1	I siti di contatto delle covate di fagiano di monte	87
6.1.2	I siti generici	88
6.1.3	Scale di analisi	90
6.1.4	Variabili di struttura del paesaggio	91
6.2	Gli alberi di classificazione binaria	97
6.2.1	Cross validation	99
6.3	Modelli di classificazione rpart	100
6.4	Discussione dei risultati	102
7	Conclusioni	109
A	Dimensione frattale	111
A.1	Dimensione di Hausdorff	111
A.2	Dimensione di Box Counting	113
B	Classificazione supervisionata	115
B.1	Introduzione	115
B.2	Statistica delle decisioni	116
B.3	Classificazione di immagini	118
B.3.1	Analisi quantitativa: modello normale multidimensionale	118
B.3.2	Classificazione alla minima distanza	119
B.4	Classificazione contestuale	120
B.4.1	Tecniche di pre elaborazione	120
B.4.2	Tecniche di post elaborazione	121
C	L'algoritmo SMAP	122
C.1	L'approccio per la segmentazione multiscalarare	122
C.2	Lo stimatore SMAP	124
C.3	L'algoritmo di segmentazione	125
C.4	Stima dei parametri	129
D	Stima delle densità di probabilità	131
D.1	Metodi parametrici	131
D.1.1	La funzione di massima verosimiglianza	133
D.2	Metodi semiparametrici	134

D.2.1	I modelli di mistura	134
E	Classificazione non supervisionata	137
E.1	Metriche e criteri di clustering	137

Elenco delle figure

1	Fagiano di monte in parata sulla neve.	xiv
1.1	Componenti e relazioni del sistema GIS integrato.	7
1.2	Esempio di struttura raster e vettoriale dei dati spaziali.	9
1.3	Esempio di calcolo della fascia di interfaccia lago-canneto (in rosso) per il biotopo Lago d'Ampola.	15
1.4	Esempio di calcolo della fascia di interfaccia lago-canneto (in rosso) per il biotopo Canneti di San Cristoforo.	16
1.5	Esempio di classificazione da foto aerea (volo Italia 1994): foto aerea originaria (a); risultato della procedura per la classificazione automatica delle categorie "lago" e "canneto" e calcolo interfaccia (b).	16
1.6	Ortofoto volo Italia 94.	18
1.7	Risultato della classificazione della categoria "bosco".	18
2.1	Maschi di fagiano di monte in parata.	20
2.2	Femmine di fagiano di monte.	21
2.3	Distribuzione esemplari adulti in funzione della quota: dato aggregato per gli anni 1998-1999 derivato da elaborazione dei censimenti estivi.	23
2.4	Distribuzione esemplari adulti in funzione della pendenza: dato aggregato per gli anni 1998-1999 derivato da elaborazione dei censimenti estivi.	24
2.5	Censimento estivo: prospetto riassuntivo.	29
2.6	Censimento estivo: cartografia dell'area campione e degli avvistamenti.	32

2.7	Censimento estivo: scheda di campagna.	33
2.8	Censimento primaverile: scheda di campagna.	34
2.9	Censimento primaverile: cartografia dell'area campione e degli avvistamenti.	35
2.10	Censimento primaverile: prospetto riassuntivo.	36
3.1	Risultato della ricerca del toponimo "malga primaluna" con <code>d.positionate</code>	39
3.2	Esempio di interfaccia per l'inserimento dei siti nel data base.	40
3.3	Schema della procedura sviluppata ed utilizzata per l'inserimento dei dati dei censimenti estivi di fagiano di monte.	41
3.4	Esempio di utilizzo del modulo GRASS <code>d.miglia.labels</code> per la visualizzazione degli attributi "n° maschi", "n° femmine", "n° pulli".	44
3.5	Cartografia dei siti di avvistamento e cartografia vettoriale dei distretti faunistici.	46
3.6	Cartografia generale delle aree campione inserite e cartografia vettoriale dei distretti faunistici.	47
3.7	OFD e area campione digitalizzata "spinale.1" (in rosso).	48
3.8	Distribuzione dell'altitudine per area campione.	49
3.9	Distribuzione della pendenza per area campione.	50
3.10	Distribuzione delle esposizioni per area campione.	51
4.1	Esempio di cartografia tematica della copertura vegetale prodotta con la classificazione da foto aerea: in bianco si è riportata l'area campione "pellizzano.17".	59
4.2	Esempio di configurazione spaziale di patches.	59
4.3	Superficie marginale e interna di un patch.	60
4.4	Transetti di analisi per l'individuazione delle scale nell'immagine.	62
4.5	Variazione distanza euclidea lungo il transetto di analisi	64
4.6	Distribuzione di probabilità dei modelli spaziali individuati in funzione della loro dimensione	65
4.7	Relazioni di distanza "bordo" - "bordo vicino" tra patches appartenenti alla stessa classe.	67

4.8	Esempio di calcolo della dimensione bi box-counting	68
5.1	Risultato procedura di classificazione gaussiana e massima verosimiglianza: foto aerea originaria (a); immagine classificata (b).	74
5.2	Risultato della procedura di classificazione mistura di gaussiane e massima verosimiglianza: foto aerea originaria (a); immagine classificata (b).	75
5.3	Risultato procedura di classificazione utilizzando l'algoritmo SMAP: foto aerea originaria (a); immagine classificata (b). . .	76
5.4	Confronto risultati procedure di classificazione: foto aerea originaria (a); risultato regressione gaussiana e massima verosimiglianza (b); risultato mistura di gaussiane e massima verosimiglianza (c); risultato mistura di gaussiane e algoritmo SMAP (d).	77
5.5	Individuazione del numero di classi e delle caratteristiche spettrali attraverso <code>i.gensigset</code>	79
5.6	Risultati procedura non supervisionata: foto aerea originaria (a); immagine classificata (b).	80
5.7	Risultato finale procedura non supervisionata.	81
5.8	Classificazione delle aree campione: superfici delle categorie per area campione.	84
5.9	Classificazione delle aree campione: percentuale di superficie boscata.	85
6.1	Numero di covate per area campione negli anni 1998 e 1999. .	87
6.2	Risultato della procedura di definizione del dominio di analisi per le aree campione "casteltesino.23" e "pievetesino": (a) aree campione; (b) area di estrazione per gli esempi generici. .	90
6.3	Esempi di siti di covate con l'indicazione delle tre scale di analisi utilizzate.	91
6.4	Precisione di calcolo del coefficiente CPA al variare della dimensione del patch.	93

6.5	Distribuzione degli esempi di training per la variabile di dimensione frattale: in verde gli esempi di classe A in rosso gli esempi di classe B	94
6.6	Distribuzione degli esempi di training per ciascuna variabile nelle tre scale di analisi: in verde gli esempi di classe A , in rosso gli esempi di classe B	95
6.7	Distribuzione degli esempi di training per ciascuna variabile nelle tre scale di analisi: in verde gli esempi di classe A , in rosso gli esempi di classe B	96
6.8	Visualizzazione di un albero di classificazione binaria.	98
6.9	Confronto degli errori in funzione del numero di split per i quattro modelli.	101
6.10	Visualizzazione dell'albero di classificazione binaria per il Modello 123.	105
6.11	Grafico di importanza delle variabili per il Modello 1.	106
6.12	Grafico di importanza delle variabili per il Modello 2.	106
6.13	Grafico di importanza delle variabili per il Modello 3.	107
6.14	Grafico di importanza delle variabili per il Modello 123.	107
6.15	Esempi di classificazione del modello: siti di presenza (classe A) (a); siti generici (classe B) (b).	108
C.1	Struttura a piramide del MSRF.	123
C.2	Struttura a <i>quadtree</i> utilizzata per il modello MSRF.	127
C.3	Struttura a grafo piramidale utilizzata nel modello MSRF.	128
C.4	Equivalente unidimensionale del grafo a struttura ibrida: in per le scale $n > 2$ si utilizza un grafo a struttura piramidale, mentre per le scale $n < 2$ si utilizza una struttura a quadtree.	129

Sommario

Lo scopo di questo lavoro è lo sviluppo di metodologie software, in ambiente GIS, da utilizzarsi per la valutazione predittiva delle operazioni di miglioramento ambientale su piccola scala. In particolare si sono realizzati degli strumenti per l'analisi sistematica e il monitoraggio, in grado di associare ad una configurazione ambientale esistente o obiettivo un opportuno indicatore della conservazione o del miglioramento dell'habitat per la fauna selvatica.

Il lavoro rientra nell'ambito degli sforzi intrapresi dal Servizio Faunistico della Provincia Autonoma di Trento (PAT), per applicare le indicazioni della L.P. n. 24/1991. Questa legge pone tra gli obiettivi prioritari dell'amministrazione provinciale, la conservazione ed il miglioramento ambientale per la fauna selvatica e individua nella gestione diretta degli ambienti naturali, attraverso interventi di carattere locale, una importante strada da seguire per la realizzazione di tali miglioramenti.

Per il perseguimento del suddetto obiettivo, si è sviluppato uno strumento informatico di analisi e valutazione che fornisca l'indispensabile supporto alle fasi di progettazione e verifica degli interventi di miglioramento. Più precisamente si è realizzato un sistema GIS integrato (capitolo 1) che soddisfa i seguenti obiettivi principali:

- la gestione e l'aggiornamento di tutti i dati, in particolare quelli di carattere spaziale (SIAT, dati faunistici, immagini SAR), disponibili per la descrizione dell'ambiente;
- il calcolo degli indicatori numerici di configurazione ambientale, utilizzando le informazioni contenute nei tematismi GIS;

-
- lo sviluppo di un modello predittivo valido per il Trentino, che associ agli indicatori numerici calcolati in maniera automatica un valore di qualità ambientale per le diverse zone esaminate.

L'ambiente GIS che costituisce la base del sistema integrato è GRASS (*Geographical Resource Analysis Support System*), il quale è stato potenziato dall'interfacciamento con il motore statistico S-PLUS e dallo sviluppo di moduli software per la gestione di basi dati specifiche. Per interfacciare i diversi ambienti che costituiscono il sistema GIS integrato, si sono prodotte apposite procedure software nei linguaggi C, Perl e Shell di Unix. Si è anche realizzato il potenziamento delle procedure ordinarie di GRASS e S-Plus, per risolvere specifiche problematiche insite nell'elaborazione dei dati spaziali a nostra disposizione.

Come caso sperimentale specifico di impiego delle metodologie, si è studiato l'habitat del fagiano di monte (*Tetrao tetrix*) per il quale si hanno a disposizione in Trentino dati precisi sul numero, sul tipo (es. maschi, femmine, pulli) e sulla localizzazione geografica degli esemplari presenti, questo grazie ad un censimento sistematico organizzato, a partire dal 1993, dal Servizio Faunistico della Provincia Autonoma di Trento (PAT) (capitolo 2).

In collaborazione con tale servizio, e all'interno di uno studio svolto presso i laboratori dell'ITC-Irst, sono stati trasferiti su supporto informatico, per la prima volta in Trentino, i dati relativi ai censimenti estivi del fagiano di monte per gli anni 1998 e 1999. Il nuovo data base prodotto, comprende, oltre alle coordinate geografiche dei punti di avvistamento dei vari esemplari, anche la cartografia in formato vettoriale di 28 aree campione.

Per risolvere i problemi di inserimento e aggiornamento dei dati specifici di interesse faunistico è stata utilizzata un'interfaccia grafica per GRASS (capitolo 3), che ha velocizzato notevolmente le operazioni di informatizzazione dei dati relativi ai censimenti. Per l'archiviazione e il trattamento dei dati si è sviluppato anche uno specifico data base relazionale.

Al fine di esprimere dei giudizi quantitativi sulle dinamiche ecologiche in corso, si sono adottati i principi e le metodologie derivanti dall'*ecologia del paesaggio* (capitolo 4). Più precisamente, si sono implementate delle

procedure in ambiente GIS che consentono di estrarre dai dati ambientali a nostra disposizione numerosi indicatori numerici di ecologia del paesaggio.

Per facilitare il calcolo degli indicatori ambientali si sono sviluppati degli strumenti di interpretazione e classificazione dei dati spaziali disponibili per il Trentino. Si sono, in particolare, messe a punto alcune procedure per la classificazione di immagini applicandole principalmente al caso delle ortofoto digitali (capitolo 5).

I principali risultati di questa tesi hanno riguardato lo sviluppo di modelli predittivi in grado di individuare le configurazioni ambientali più favorevoli alla presenza di covate di fagiano di monte (capitolo 6). I modelli sono stati addestrati sui dati di presenza effettiva di esemplari di fagiano di monte derivati dai censimenti estivi: gli indicatori di ecologia del paesaggio sono stati applicati a tre diverse scale di risoluzione ed infine combinati per ottenere un modello multiscala basato su alberi di classificazione. Il miglior modello sviluppato permette di discriminare adeguatamente (10% di errore predittivo calcolato tramite cross-validation) i siti di presenza dai siti generici utilizzati come esempi negativi. Tra le variabili automaticamente selezionate dal modello va segnalata l'importanza della misura di box-counting ed i coefficienti di forma e di caratterizzazione dei patch piccoli.



Figura 1: Fagiano di monte in parata sulla neve.

Capitolo 1

Introduzione: miglioramenti ambientali e metodologie GIS

In questo capitolo si vuole innanzitutto descrivere la situazione ambientale del territorio montano trentino, per evidenziare le problematiche esistenti e motivare di conseguenza la necessità di interventi di miglioramento ambientale. In particolare, l'attenzione viene focalizzata sui miglioramenti ambientali ai fini della gestione del patrimonio faunistico in provincia di Trento. Si presenteranno le indicazioni contenute nella legislazione provinciale, e le iniziative intraprese dal Servizio Faunistico della Provincia Autonoma di Trento, per la progettazione e realizzazione di interventi di miglioramento ambientale di tipo locale.

Successivamente, si descriverà il ruolo che assumono le metodologie software in ambiente GIS, per la valutazione dei suddetti miglioramenti a carattere locale. Si evidenzierà soprattutto l'importanza del sistema GIS integrato, sviluppato nel presente lavoro, per la gestione dei dati ad alta risoluzione e per l'estrazione automatica degli indicatori ambientali. Si esporranno infine anche alcuni esempi preliminari, i quali vogliono descrivere i passi che hanno portato all'idea attuale di metodologia GIS. Attraverso gli esempi, si è provata la fattibilità dell'estrazione di indicatori ambientali dai tematismi prodotti attraverso la classificazione di foto aeree.

1.1 I problemi ambientali del territorio montano in Trentino

In tutto l'arco alpino, a partire dal secondo dopoguerra, si è verificata una modificazione radicale nelle interazioni tra uomo e ambiente. Sono cambiate principalmente le dinamiche insediative e le tipologie di sfruttamento del territorio montano.

Queste modificazioni hanno alterato il ruolo di primaria importanza che aveva raggiunto l'uomo nel regolare i flussi energetici e di informazioni, all'interno di tale ambiente. L'abbandono dei pascoli d'alta quota e di talune pratiche agricole, lo sfruttamento del territorio montano per attività turistiche, in particolare sciistiche, e la costruzione delle necessarie infrastrutture hanno modificato gli equilibri dinamici tra uomo e natura che si erano creati nei secoli.

Le conseguenze dei mutamenti sono lontane dall'essere pienamente comprese, principalmente per i tempi rapidi con cui si sono verificati i cambiamenti. La scala temporale che li ha contenuti è infatti di alcuni ordini di grandezza più piccola rispetto a quella che aveva portato alla condizione di equilibrio precedente. D'altra parte, alcune conseguenze sono già riscontrabili; nonostante la forte espansione degli ungulati (capriolo, cervo, muflone), si osserva infatti una significativa riduzione del numero di esemplari di alcune specie quali il fagiano di monte, la coturnice, la lepre comune e il francolino di monte.

1.2 Gli obiettivi della gestione faunistica e i miglioramenti ambientali

Anche per il Trentino, una delle conseguenze delle alterazioni ambientali maggiormente riscontrabili è la diminuzione del numero di esemplari e delle aree di distribuzione di alcune specie selvatiche animali. Il calo degli effettivi, è da attribuirsi principalmente alla frammentazione e scomparsa degli habitat e all'isolamento delle popolazioni, fenomeni questi che sono riscontrabili in tutto l'arco alpino. La riduzione interessa specie che rivestono un ruolo

importante, dal punto di vista biologico e sociale, negli equilibri ecologici dell'ambiente montano.

In considerazione di ciò la L.P. n. 24/1991 pone tra gli obiettivi prioritari della Provincia Autonoma di Trento la conservazione ed il miglioramento della fauna selvatica. In tale legge si individuano due vie per perseguire questo obiettivo: la prima si realizza attraverso una gestione diretta del patrimonio faunistico provinciale, la seconda mediante una corretta ed equilibrata politica di gestione degli ambienti naturali, che costituiscono peraltro gli habitat indispensabili per le varie popolazioni faunistiche.

Gli aspetti di gestione del patrimonio faunistico, nella Provincia Autonoma di Trento, sono di competenza del Servizio Faunistico, il quale, applicando le indicazioni della legge provinciale, ha realizzato uno sforzo particolare per affiancare agli interventi ordinari di gestione del patrimonio faunistico, anche interventi diretti al miglioramento ambientale a carattere locale e finalizzati al ripristino delle condizioni favorevoli per le specie di maggior interesse biologico. A riguardo, sono state individuate un primo gruppo di specie per le quali esiste un'esigenza prioritaria di conservazione: la coturnice alpina, il fagiano di monte e la lepre comune. Già da alcuni anni, per tali specie, sono state intraprese campagne di raccolta dati e si sono avviati studi specifici finalizzati a cogliere le concrete possibilità di ripristinare, almeno localmente, le condizioni favorevoli allo svolgimento delle loro fasi biologiche più delicate.

Si vedranno nella seguente sezione e nei capitoli successivi, le importanti implicazioni che il carattere prevalentemente locale dei miglioramenti assume nello sviluppo delle metodologie software per l'analisi e il monitoraggio dell'ambiente, e nella scelta della scala di studio e di monitoraggio.

1.3 Gli strumenti software GIS per il monitoraggio ambientale in Trentino

La realizzazione dei miglioramenti ambientali per la conservazione delle specie individuate necessita dello sviluppo e dell'impiego di adeguati strumenti informatici di analisi e valutazione, che forniscano l'indispensabile supporto nelle fasi di progettazione e di verifica degli interventi.

Si è cercato pertanto di realizzare uno strumento GIS in grado di analizzare automaticamente il territorio ed associare ad ogni configurazione ambientale esistente, un opportuno indicatore della conservazione o del miglioramento dell'habitat per la fauna selvatica, tutto questo sfruttando nel miglior modo possibile i dati di carattere ambientale disponibili per il Trentino. Sono necessari strumenti software GIS poiché si deve operare con dati di tipo spaziale e ad alta risoluzione; in particolare il dato georeferenziato permette di integrare informazioni territoriali provenienti da fonti diverse, e che presentano disomogeneità anche sul tipo di struttura delle diverse basi dati. Un siffatto strumento dovrebbe permettere altresì di facilitare notevolmente le fasi di analisi sistematica su ampie porzioni di territorio, e agevolare anche il monitoraggio nel tempo, attraverso l'acquisizione ed elaborazione di nuovi dati (foto aeree).

Si analizzeranno nelle seguenti sezioni innanzitutto le caratteristiche principali dei GIS, e soprattutto del GIS GRASS, mostrando come, per i nostri scopi, sia stato necessario potenziare l'ambiente GRASS con l'integrazione del motore statistico S-PLUS [1] e con l'integrazione di strumenti specifici per la gestione di strutture di data base relazionali. Parallelamente si è sperimentato lo sviluppo di nuove procedure in GRASS che consentano di migliorare soprattutto gli aspetti di elaborazione dei dati.

1.4 Definizioni

Negli ultimi anni, la possibilità di disporre di tecnologie informatiche sempre più potenti e versatili ha visto la nascita e lo sviluppo, di efficaci applicazioni per trattare in modo digitale i dati spaziali¹. Tali applicazioni, conosciute come *sistemi informativi spaziali*, consentono in genere di lavorare con dati in diversi formati, e sono attualmente in via di sviluppo e continuo miglioramento.

In particolare, nel nostro ambito di interesse, si parlerà in maniera più appropriata di *sistemi informativi geografici* (GIS) invece che di *sistemi in-*

¹Per dati spaziali si intendono i dati referenziati per la descrizione degli oggetti in un qualche spazio geometrico di riferimento (in genere geografico) [2]

formativi spaziali, poiché il sistema di riferimento degli oggetti è appunto di tipo geografico.

Tutti i GIS hanno quattro fondamentali componenti:

1. Il sistema per l'inserimento e processamento dei dati spaziali quali cartografie tematiche, fotografie aeree, immagini satellitari ed altre fonti;
2. Il sistema per l'immagazzinamento e la gestione dei dati (*database management system*);
3. Il sistema per l'analisi, l'elaborazione e il confronto dei dati (es. processamento e classificazione delle immagini, estrazione di indicatori ambientali);
4. Il sistema per la produzione e visualizzazione di mappe e tabelle di dati.

Un GIS, in definitiva, può essere descritto, come una collezione di dati e applicazioni per studiare spazialmente i fenomeni che avvengono nel mondo che si sta osservando [2]. La nostra percezione del mondo si manifesta attraverso processi cognitivi come la selezione, la generalizzazione e la sintesi che ci consentono di ottenere informazioni su ciò che si sta osservando o sentendo, informazioni che possono essere condivise o meno e utilizzate da altre persone. I dati sono una rappresentazione fisica di tali informazioni e costituiscono la base per un modello empirico di quei fenomeni. Così l'insieme di dati che costituiscono il database sono l'immagazzinamento fisico di diverse viste sul mondo reale, e rappresentano la nostra conoscenza in un determinato momento nel tempo. Il concetto di informazione differisce da quello di dato; l'informazione è ciò che deriva dal processo interpretativo umano che si cerca di riprodurre tramite le operazioni di elaborazione e modellizzazione.

Si vedrà a riguardo, nel capitolo 5, come la simulazione del procedimento interpretativo umano sia la base per estrarre nuove informazioni dai dati. Verranno descritte, in particolare, le procedure di classificazione per estrarre dalle immagini aeree delle zone di interesse solo le informazioni relative a particolari aspetti ambientali producendo mappe riclassificate.

1.4.1 Sistema GIS integrato

L'utilizzo del GIS ha permesso in questo progetto la gestione e l'elaborazione combinata di molte informazioni territoriali. In particolare sono stati utilizzati i dati del SIAT (Sistema Informativo Ambiente e Territorio della Provincia Autonoma di Trento) e il nuovo data-base informatizzato dei censimenti del fagiano di monte, come descritto in dettaglio nel capitolo 2.

Il sistema informativo spaziale utilizzato per le nostre elaborazioni è *GRASS* (*Geographical Resource Analysis Support System*). Sviluppato inizialmente dal CERL (organismo di ricerca militare USA), GRASS è un GIS principalmente a base raster, anche se consente numerose elaborazioni dei dati in formato vettoriale², ed è quello che si può definire un open-GIS. Gli open-GIS sono caratterizzati da un insieme di moduli indipendenti, che possono essere facilmente migliorati e sviluppati, inseriti in una struttura di base che garantisce il collegamento tra i diversi moduli e la gestione di dati georeferenziati. A riguardo si può osservare come GRASS sia fornito di un gran numero di librerie C, che garantiscono il collegamento tra i nuovi moduli sviluppati ed il sistema di base. L'aggiornamento dei moduli è coordinato a livello internazionale dalle università di Baylor e Hannover. Il gruppo di lavoro ITC-Irst utilizza questo strumento in numerosi progetti di ricerca ed ha contribuito alla creazione di nuovi programmi.

Oltre a moduli sviluppati per la gestione e l'interrogazione degli oggetti spaziali, GRASS dispone anche di moduli che consentono la classificazione di immagini. Questi però, non sono in genere particolarmente flessibili, poiché sono sviluppati nell'ambito di specifiche procedure standard. Esistono anche dei moduli per l'estrazione di informazioni sulle caratteristiche spaziali delle immagini o delle carte tematiche, ma non sono supportati da adeguati strumenti di elaborazione statistica che consentano una successiva analisi multivariata. Invece, in riferimento alla gestione delle basi dati, è nata l'esigenza di disporre di particolari strutture per gestire i dati relativi ai censimenti di fagiano di monte (capitolo 3). Le caratteristiche base di GRASS non includono la gestione di data base a struttura complessa.

²I concetti di raster e vector saranno descritti in dettaglio nella successiva sezione (1.4.2).

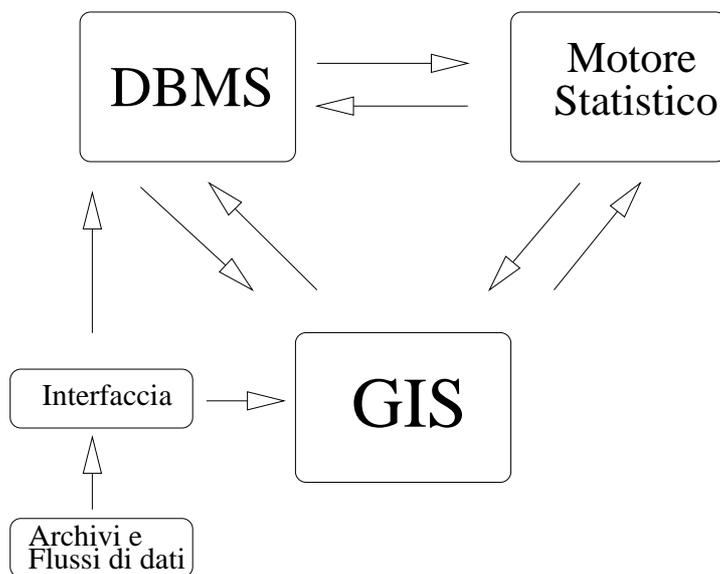


Figura 1.1: Componenti e relazioni del sistema GIS integrato.

Si è cercato quindi di potenziare le normali possibilità del GIS GRASS con l'integrazione di programmi per l'elaborazione statistica dei dati e con particolari moduli appositamente sviluppati per la gestione del data-base specifico da noi utilizzato. In particolare l'utilizzo di procedure di statistica computazionale nel programma S-PLUS ha consentito elaborazioni avanzate di analisi multivariata dei dati.

Il sistema GIS integrato così strutturato (Figura 1.1) è perciò costituito dal GIS GRASS, dal software statistico S-PLUS, e da una serie di moduli sviluppati in linguaggio Perl, Shell-Unix e C per l'inserimento dei dati, la gestione degli stessi e l'interfacciamento delle varie parti del sistema integrato.

I diversi moduli di interfacciamento saranno descritti nei capitoli successivi. Si presentano invece nella prossima sezione, in maniera più dettagliata, alcuni concetti di analisi spaziale per cogliere gli importanti aspetti concernenti l'utilizzo dei GIS e delle base dati spaziali. Alcune definizioni qui introdotte consentono di affrontare meglio gli argomenti illustrati nei successivi capitoli.

1.4.2 Strutture e formati dei dati

La struttura più comune per organizzare i dati spaziali è quella di pensarli come insiemi di mappe disposte su diversi strati o *layers*. Ogni layer o *tematismo GIS* rappresenta un diverso punto di osservazione del fenomeno cioè una sua descrizione specifica; in alcuni casi diversi layer corrispondono ad un'evoluzione temporale. Tale approccio deriva dalle tecniche di sovrapposizione di diversi tematismi utilizzate per analizzare e confrontare le mappe cartacee, prima dell'avvento della cartografia digitale.

Si è precedentemente osservato come i dati siano essenziali per la descrizione dei fenomeni che si possono osservare nel mondo reale. Perché la descrizione sia significativa, è però necessario individuare le caratteristiche elementari per descrivere univocamente un fenomeno e poterlo quindi rappresentare attraverso i dati.

Si introduce a riguardo il concetto di entità, per riferirci a fenomeni che non possono essere suddivisi in unità più semplici. Assumendo che la porzione geospaziale individui definiti fenomeni, si considerano cinque categorie di informazioni per specificare le entità base o oggetti GIS [2].

1. *l'identificatore*: è un nome o un codice oppure un numero che serve per distinguere differenti entità;
2. *la posizione*: è la localizzazione rispetto alla superficie della terra degli oggetti naturali o di origine antropica;
3. *gli attributi*: definiscono alcune caratteristiche generali delle entità, ad esempio il numero di abitanti di una casa. Questi sono necessari in quanto le entità possono avere caratteristiche che non sono usate nella loro definizione;
4. *il ruolo, il comportamento e la funzione delle entità*: sono aspetti caratteristici delle entità a comportamento dinamico;
5. *le proprietà spaziali*: descrivono le caratteristiche spaziali degli attributi, quali la dimensione, l'area, la forma, ma anche l'adiacenza e la relazione con altre entità. Sono molte volte gestite dal sistema informativo allo stesso modo degli attributi.

Aspetti più dettagliati sugli attributi e sui comportamenti dinamici delle entità saranno considerati nei capitoli 3 e 4.

Si focalizzi l'attenzione sulle proprietà spaziali di base di ciascuna entità che sono gli elementi fondamentali e non divisibili che la identificano e costituiscono i dati codificati. Si supponga, quindi, di osservare la superficie della terra sulla quale sono distribuiti i fenomeni; un primo modo intuitivo per descriverli è utilizzando le loro caratteristiche geometriche: le entità si possono allora pensare come punti, linee, aree e volumi. Si deve precisare che questa distinzione dipende dalla scala di osservazione; infatti, se si considera l'esempio di un lago alpino, questo può apparire o un punto o un poligono, a seconda della distanza da cui lo si osserva. Una rappresentazione dei fenomeni si concretizza con una particolare struttura nei dati chiamata *struttura vettoriale ad oggetti* in riferimento ai vettori che possono essere disegnati da punto a punto per ricostruire il contorno dell'oggetto (figura 1.2). Molti

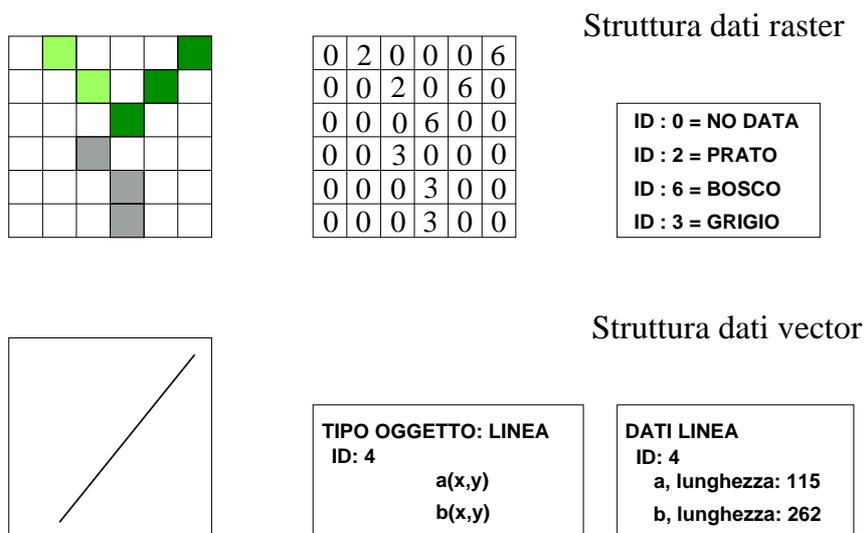


Figura 1.2: Esempio di struttura raster e vettoriale dei dati spaziali.

fenomeni si possono descrivere utilizzando particolari combinazioni di unità elementari. Si intendono per la precisione combinazioni di punti che definiscono un reticolo, o combinazioni di unità bidimensionali discrete e connesse che operano invece una tassellazione dello spazio. L'utilizzo di tali combinazioni di entità facilita soprattutto la descrizione dei fenomeni che variano con

continuità nello spazio, si pensi all'orografia del terreno, alla concentrazione di inquinanti in atmosfera, al rischio di incendi.

La tassellazione attraverso poligoni, e il reticolo di punti sono anch'essi codificati attraverso un formato vettoriale. Nel caso particolare di tassellazione regolare e di reticolo regolare, con risoluzione fissata a priori, è possibile tuttavia definire un secondo tipo di struttura per la codifica dei dati, alternativa a quella vettoriale.

Questa diversa struttura si basa sull'organizzazione dei dati in una matrice contenente righe e colonne di celle, che identificano l'entità e l'attributo associato, ed è denominata *struttura raster* (figura 1.2). Ci sono numerosi modi per codificare nel dettaglio una struttura a dati raster: la più comune è costituita da una tabella di dati che rappresentano gli attributi, i quali sono indicizzati per righe e per colonne, in funzione della distribuzione spaziale delle entità. Un importante parametro della struttura a dati raster è la risoluzione, intesa come la più piccola distanza che può essere registrata e misurata tra il centro di due celle [2].

Esempi di dati codificati con la struttura raster sono le immagini aeree o da satellite, i modelli digitali del terreno e le carte d'uso del suolo. La struttura raster è molto utile anche per la produzione di DTM (digital elevation model) e derivati (es. carta della pendenza e carta delle esposizioni) e mappe di rischio.

Seguendo queste due fondamentali strutture nella rappresentazione dei dati, anche i sistemi informativi geografici si sono sviluppati secondo due diverse filosofie: i GIS a base vettoriale e quelli a base raster. Si è visto però che le strutture dei dati vettoriale e raster si completano; la tendenza attuale è sicuramente quella di sviluppare GIS in grado di gestire adeguatamente sia l'una che l'altra forma di rappresentazione. Tuttavia, normalmente, un singolo GIS è più votato al calcolo e alla rappresentazione raster o vettoriale a seconda che sia nato come raster GIS oppure come vector GIS [3].

GRASS ad esempio è stato sviluppato principalmente per la gestione dei dati raster; sono presenti moduli che consentono svariati tipi di analisi e elaborazioni:

- analisi, elaborazione e classificazione di immagini e fotografie aeree;

- analisi di tipo morfologico, idrografico e idrologico;
- modelli di diffusione;
- analisi di ecologia del paesaggio;
- intervisibilità e analisi delle esposizioni solari.

Nei recenti miglioramenti del software è stato incluso anche un modulo per la visualizzazione tridimensionale avanzata (NVIZ).

1.5 Basi dati cartografiche in Trentino

Per completare la visione sui sistemi informativi territoriali, si presentano le principali basi dati cartografiche di importanza ambientale disponibili per il Trentino.

Dal 1990 la Provincia Autonoma di Trento, ha intrapreso lo sviluppo del Sistema Informativo Ambiente e Territorio (SIAT), con l'obiettivo di attrezzare la Provincia di migliori e nuove risorse informative riferite al territorio, a sostegno dell'attività di programmazione, gestione, monitoraggio e controllo [4].

Nell'ambito di tale progetto, sono stati fino ad ora prodotti una notevole mole di dati e di tematismi, alcuni dei quali, in particolare quelli di interesse primario ed intersettoriale, sono stati utilizzati ampiamente nel presente lavoro.

I dati forniti dal SIAT sono per il Trentino la principale fonte di informazioni territoriali, e si presentano omogenei per referenziamento geografico e per formato degli archivi. Si descrivono di seguito le caratteristiche principali dei tematismi utilizzati, compresi anche quelli provenienti da fonti diverse dal SIAT.

1.5.1 DTM

Il modello digitale del terreno (DTM) fa parte della cartografia di base del SIAT e rappresenta la morfologia del territorio, mediante una struttura matriciale a celle quadrate con passo di campionamento 10 m. Ad ogni cella

di tale reticolo è associata la quota al suolo, espressa in metri, riferita al centro della cella stessa.

Il DTM è inquadrato nella rappresentazione conforme Gauss-Boaga nel sistema geodetico nazionale (ellissoide internazionale con orientamento a Roma Monte Mario 1940). Tale riferimento è utilizzato per tutto il SIAT.

Dall'elaborazione del DTM sono stati prodotti per tutto il Trentino anche altri due tematismi:

- la carta delle pendenze in gradi, con risoluzione di 10 metri, calcolata per ogni cella del DTM utilizzando un intorno 3×3 .
- la carta delle esposizioni in gradi est, con risoluzione di 10 metri, calcolata per ogni cella del DTM utilizzando un intorno 3×3 .

1.5.2 Carta topografica generale 1:10000

La cartografia numerica in formato raster (CTP) è l'immagine digitale georeferenziata della Carta Topografica Generale alla scala 1:10000 (CTG) della PAT. La CTG fornisce una rappresentazione generale della morfologia, delle acque, della vegetazione, delle opere dell'uomo, riportando tutto ciò che può essere utile come riferimento topografico e rappresentato in relazione alla scala nominale. L'altimetria è resa mediante curve di livello, con equidistanza di 10 m, e punti quotati isolati.

La CTG è stata realizzata su supporto cartaceo mediante fotorestituzione, a partire dai fotogrammi di voli effettuati in due diversi periodi: 1980 (Trentino meridionale) e 1983 (Trentino settentrionale). Le tolleranze massime ammesse per la CTG sono le seguenti:

- planimetrica: $\pm 2,0$ m
- altimetrica relativa alle curve di livello in terreno scoperto: $\pm 2,0$ m
- altimetrica relativa ai punti isolati: $\pm 0,9$ m

Anche la cartografia numerica è parte del SIAT e quindi ha le stesse caratteristiche di referenziazione spaziale del DTM.

Nel maggio del 1999 è stata prodotta una nuova cartografia numerica generale per la provincia di Trento, per ora ad esclusivo uso interno dei servizi provinciali. Poichè l'ITC-Irst partecipa alla validazione di questo tematismo, è stato possibile utilizzarlo anche nel presente lavoro.

1.5.3 OFD

L'ortofoto digitale (OFD) è anch'essa estratta dal SIAT, ed è il risultato della procedura di raddrizzamento differenziale della ripresa aerofotogrammetrica ad alta risoluzione denominata *Volo Italia 1994*. La procedura per la produzione dell'ortofoto digitale è stata eseguita con metodologia digitale utilizzando come base il modello digitale dell'orografia, con maglia regolare (DTM). Le principali caratteristiche del volo sono:

- quota relativa di volo: 11.500 m
- scala media fotogrammi: 1:75.000
- camera da presa: Wild RC 20 (FMC) $f=152.82$ mm
- risoluzione immagine: tra 0,5 e 1,0 metri

La digitalizzazione dei fotogrammi è stata eseguita con scansione a risoluzione di 1016 dpi a 256 livelli di grigio. L'ortonormalizzazione è stata invece effettuata mediante il modello digitale del terreno passo 40 m e utilizzando come punti di appoggio a terra quelli rilevati per la formazione della Carta Topografica Generale 1:10000. La risoluzione finale dell'Ortofoto è di 1 m² in coordinate terreno.

Il processo di ortorettificazione calcola i punti immagine sul piano cartografico con una tolleranza di 3 metri.

1.6 Esempi dimostrativi

La possibilità di utilizzare delle metodologie GIS per lo studio dei miglioramenti ambientali a carattere locale è stata sperimentata svolgendo alcuni studi preliminari. Si è provata soprattutto la potenzialità dei sistemi

informatici nell'estrazione di indicatori di qualità ambientale da tematismi GIS.

1.6.1 Esempio 1: calcolo dell'interfaccia

Come primo esempio si è eseguito il calcolo dell'interfaccia tra lago e canneto di due biotopi di interesse provinciale, utilizzando la cartografia relativa all'uso del territorio prodotta dall'Ufficio Biotopi del Servizio Parchi e Foreste Demaniali. Sia per il biotopo "Lago d'Ampola" (Figura 1.3) che per il biotopo "Canneti di San Cristoforo" (Figura 1.4), l'interfaccia lago-canneto è stata calcolata in ambiente GRASS e sovrapposta in rosso alla mappa raster dei biotopi. Per il biotopo "Canneti di San Cristoforo" si è inoltre confrontata la variazione di interfaccia tra lago e canneto calcolata in due diversi periodi dapprima utilizzando i tematismi GIS disponibili (Figura 1.4), e quindi i risultati della classificazione automatica da ortofoto digitale del volo Italia 1994 (Figura 1.5). I risultati del confronto, che evidenzia una variazione significativa, sono riportati nella seguente tabella; in entrambi i casi la zona di margine utilizzata per il calcolo aveva larghezza 3 m ed i dati sono relativi all'area nord-est del biotopo visualizzata in Figura 1.5.

<i>mappa</i>	<i>area interfaccia (m²)</i>
cartografia PAT	1381
class. da OFD 94	3200

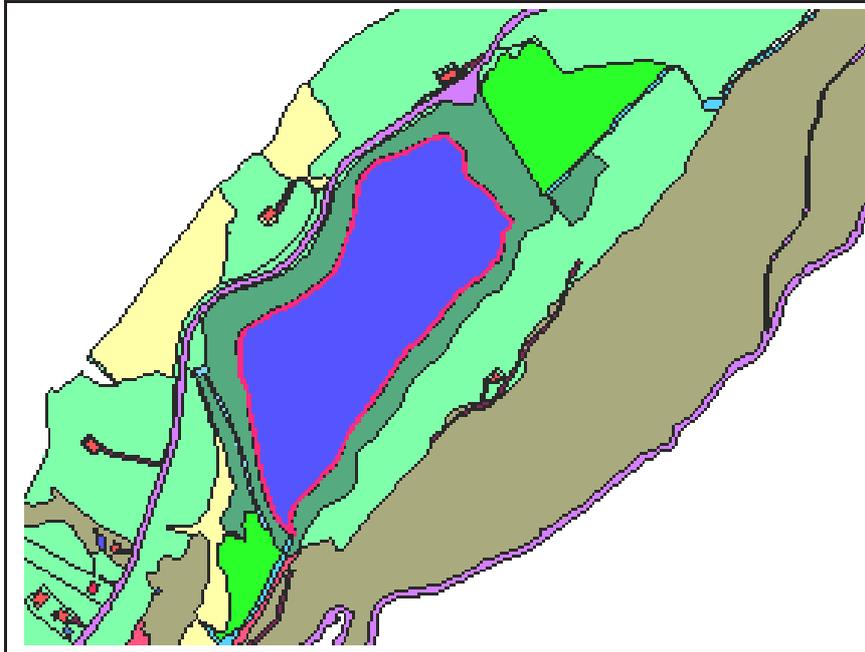


Figura 1.3: Esempio di calcolo della fascia di interfaccia lago-canneto (in rosso) per il biotopo Lago d'Ampola.

1.6.2 Esempio 2: indici di struttura del paesaggio

Si sono studiate la conformazione e distribuzione di zone boscate isolate (*patch*) nella zona in prossimità del CEA (Centro di Ecologia Alpina) in località Viote del monte Bondone vedi (figura 1.6).

Tutti i dati sono stati estratti dall'ortofoto (volo Italia '94) della zona di interesse. L'ortofoto è inizialmente stata riclassificata utilizzando l'algoritmo *smap*, nelle categorie "*bosco*", "*prati e pascoli*" e

"*altro*". I risultati della classificazione per la sola categoria "*bosco*" sono riportati in figura 1.7.

In figura 1.7 sono pure evidenziate le 9 aree (*sampling areas*) soggette ad analisi, aventi estensione di 458×385 m ciascuna. In ognuna di queste aree i *patch* di bosco presenti sono stati classificati automaticamente ed analizzati tramite opportuni indici. Questi ultimi misurano le dimensioni dei *patch*, la loro forma, le relazioni tra *patch* vicini e la loro distribuzione spaziale.

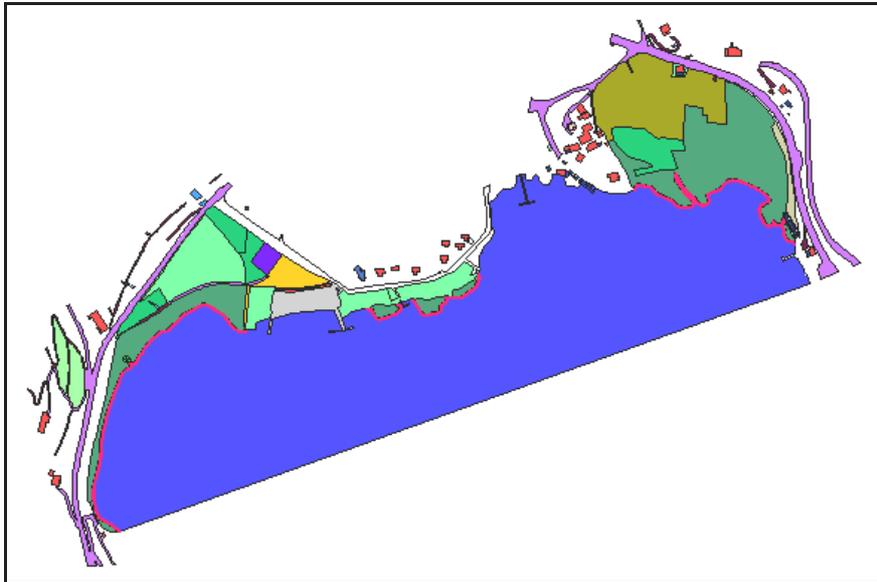


Figura 1.4: Esempio di calcolo della fascia di interfaccia lago-canneto (in rosso) per il biotopo Canneti di San Cristoforo.

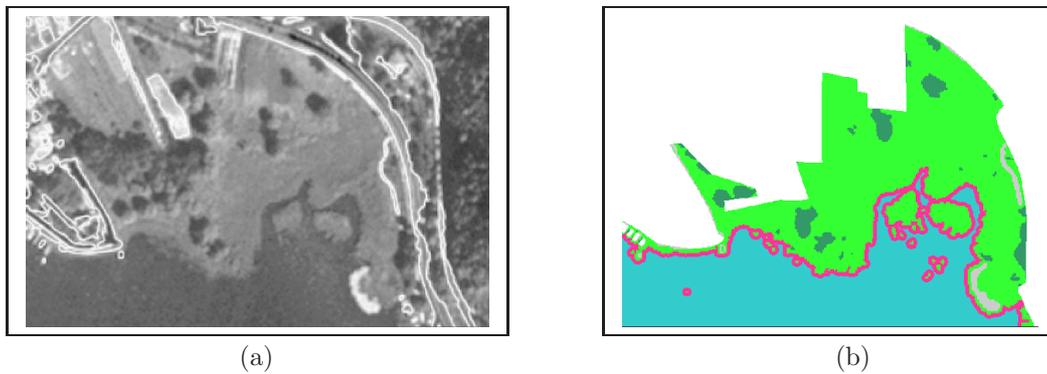


Figura 1.5: Esempio di classificazione da foto aerea (volo Italia 1994): foto aerea originaria (a); risultato della procedura per la classificazione automatica delle categorie “lago” e “canneto” e calcolo interfaccia (b).

Si riportano di seguito i risultati delle elaborazioni eseguite; i risultati sono espressi in valori medi per le 9 *sampling area*.

	Sampling area								
	1	2	3	4	5	6	7	8	9
$n^{\circ} patch$	176	192	39	73	132	1	4	34	22
% Cop.	0.49	0.48	0.41	0.036	0.090	0.98	0.96	0.84	0.83
P_{tot} (m)	16284	17116	4780	2950	5804	2135	3179	9698	10363
$\overline{P/A(m^{-1})}$	1.30	1.29	1.11	0.92	1.08	0.02	0.90	1.30	1.43
$\overline{PC/A}$	1.53	1.51	1.62	1.42	1.49	2.52	1.86	1.57	1.80
\overline{RCC}	0.70	0.71	0.66	0.72	0.70	0.78	0.78	0.74	0.70
$\overline{d_n(m)}$	5.8	4.9	10.6	18.1	6.6	-	4.5	3.2	3.4

$\overline{P/A} = \frac{Perimetro}{Area}$: media, tra tutti i *patch* di una *sampling area*, del rapporto dato dalla lunghezza totale del perimetro e l'area di ogni *patch*. È indicativo della superficie specifica di interfaccia tra bosco e prato.

$\overline{PC/A} = \frac{0.282 * Perimetro}{Area^{1/2}}$: anche per questo indice viene calcolata la media di tutti i *patch* in ogni *sampling area*. Questo indice è indicativo della forma media dei *patch*; varia da 1.0 per forme circolari, a ∞ per forme infinitamente allungate.

$\overline{RCC} = \frac{2 * (area/\pi)^{1/2}}{asse\ maggiore}$: rappresenta il rapporto tra l'area del singolo *patch* ed il più piccolo cerchio circoscritto. Questo è più propriamente un indice della compattezza dei *patch* e varia tra 0.0 per forme poco compatte, a 1.0 per forme circolari. È anche questo mediato tra tutti i *patch* in una *sampling area*.

$\overline{d_n}$: valore medio della distanza di un *patch* con il suo *patch* più vicino, misurata tra i bordi degli stessi. È indice di porosità e presenza di corridoi tra le aree boscate.

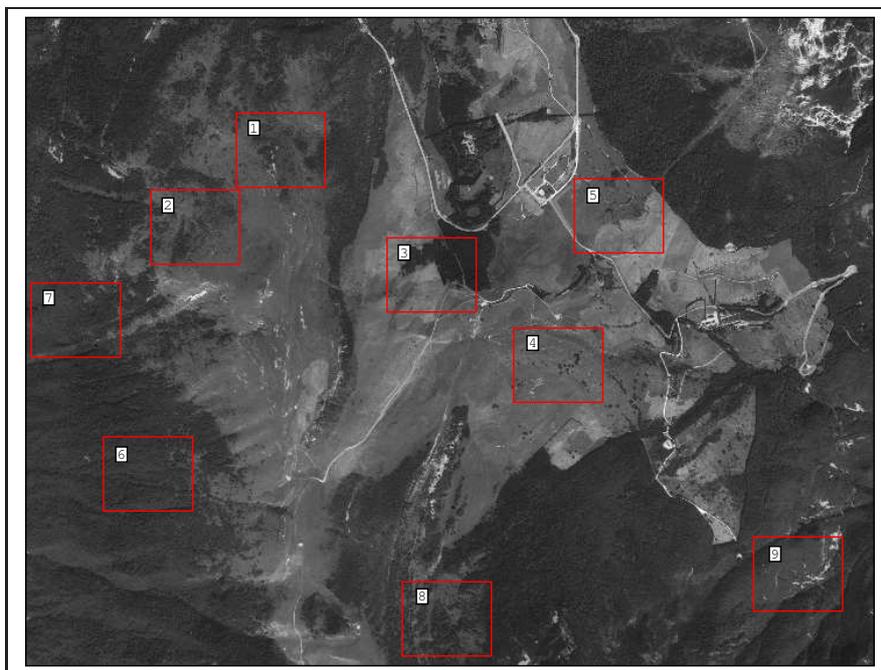


Figura 1.6: Ortofoto volo Italia 94.

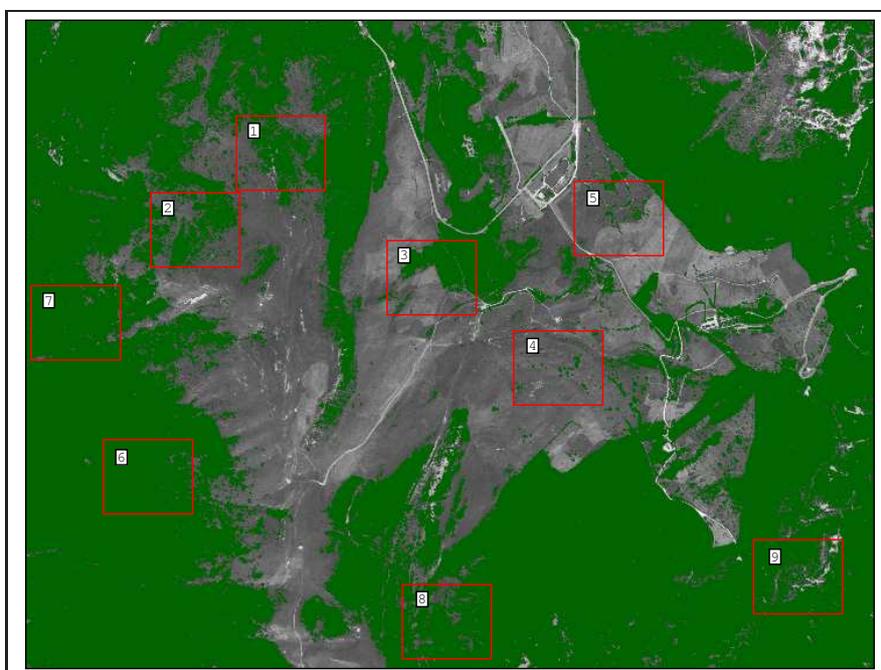


Figura 1.7: Risultato della classificazione della categoria "bosco".

Capitolo 2

I miglioramenti ambientali per il fagiano di monte

Come caso sperimentale di impiego delle metodologie GIS si è studiato l'habitat del fagiano di monte (*Tetrao tetrix*) o gallo forcello. Obiettivo di questo capitolo è la presentazione di tutte le informazioni disponibili per la specie in esame, sia quelle desunte dalla letteratura, sia quelle ottenute dall'elaborazione dei censimenti estivi e primaverili.

A riguardo si esporranno inizialmente le caratteristiche principali della specie, evidenziando gli aspetti relativi all'habitat e alle conseguenze dei disturbi antropici sulle dinamiche delle popolazioni. Si proseguirà poi esponendo gli aspetti fondamentali da conoscere e le informazioni essenziali di cui disporre per una corretta gestione faunistica del fagiano di monte e, di conseguenza, si esporranno le tipologie di censimento applicate nella provincia di Trento.

Si presenteranno anche alcuni risultati derivati dall'elaborazione del data base informatizzato, relativo ai censimenti del fagiano di monte, data base che sarà descritto nel capitolo 3. Si vogliono infatti confrontare i dati relativi alla realtà trentina con quelli desunti dalla letteratura specifica.

2.1 Caratteristiche della specie

Si descrivono inizialmente le principali caratteristiche fisiche e biologiche della specie, e poi gli aspetti legati all'habitat e alle alterazioni avvenute negli

ultimi anni. Si cercherà infine di evidenziare gli aspetti che interesseranno gli sviluppi successivi dal lavoro.

2.1.1 Descrizione della specie

Il maschio adulto di fagiano di monte ha una lunghezza totale di circa 50 cm, apertura alare di circa 80 cm, peso variabile stagionalmente tra i 1100 ed i 1450 g, colorazione sul dorso bluastra con sottoala e sottocoda bianchi e tipica coda a forma di lira. La femmina adulta invece, ha una lunghezza totale di circa 45 cm, apertura alare di circa 70 cm, peso variabile stagionalmente tra i 750 e gli 950 g e presenta piumaggio bruno con piccole barrature bianche.



Figura 2.1: Maschi di fagiano di monte in parata.

In riferimento al nutrimento esiste una netta differenza tra periodo invernale ed estivo. I rametti di mirtillo nero, le gemme e le foglie di rododendro, rappresentano nella Alpi il nutrimento principale. Gli arbusti nani, se mancano, possono venir sostituiti da piante erbacee e, in particolare, dai loro semi. Quando la fruttificazione è buona, le bacche di mirtillo nero e rosso, in



Figura 2.2: Femmine di fagiano di monte.

autunno, sostituiscono quasi totalmente i rametti e le gemme. Se accessibili, le bacche del mirtillo rosso e dell'erica a bacche (empetro nero), che restano sulla pianta in inverno, vengono mangiate fino in primavera. Gli arbusti nani offrono nutrimento anche in inverno; quando però non sono più accessibili a causa della neve, vengono sostituiti dal nutrimento proveniente dagli alberi. In inverni particolarmente nevosi il larice può infatti costituire, seppur per breve tempo, l'unica fonte di nutrimento. In mancanza del larice, questo può essere sostituito dal sorbo degli uccellatori e dal pino mugo strisciante [14].

Interessanti sono le caratteristiche fisiche degli uccelli, per il loro adattamento al clima invernale. Innanzitutto in posizione di riposo sono completamente mimetizzati; questo aspetto vale sia per la femmina nel suo piumaggio a fasce nero-rossicce, giallobruno o nerobruno, ma anche per il maschio, che con il suo piumaggio a contrasti neri e bianchi, riesce a mimetizzarsi molto bene nel paesaggio invernale. Entrambi i sessi presentano poi, un fitto piu-

maggio che permette la protezione dal freddo; nei mesi invernali, peraltro, non solo il piumino è più fitto rispetto ad altri uccelli, ma le piume presentano anche un vessillo secondario. A riguardo, il fagiano di monte presenta anche una particolare protezione per le zampe, che in inverno sono coperte da piume sulla parte anteriore e sui lati, in estate solo sulla parte anteriore.

Oltre alla protezione contro il freddo, troviamo altri adattamenti che permettono agli uccelli di muoversi senza fatica nella neve polverosa, e cioè la superficie delle dita anteriori che viene ingrandita da ambo i lati da due file di scagliette cornee sulla pianta della zampa. Inoltre si trovano delle fini scagliette plantari lunghe circa 2-4 mm, per mezzo delle quali la superficie della zampa viene ingrandita ulteriormente di 1-2 mm per parte.

2.1.2 Areale di distribuzione

Per le nostre zone alpine, l'areale di distribuzione del fagiano di monte è compreso tra i 1400 e i 2300 m, in boschi radi di resinose misti a latifoglie, ricchi di sottobosco. L'ambiente più caratteristico è però quello compreso tra i 1800 ed i 2200 m al limite della foresta, in boschi radi di conifere, dominanti di mirtillo, rododendro, ontano verde e betulla. In estate predilige i pendii freschi esposti a settentrione [14].

Per il territorio del Trentino, i dati relativi all'altitudine, alla pendenza e all'esposizione sono facilmente ricavabili dall'elaborazione del data base informatizzato, prodotto a partire dai censimenti estivi del 1998 e 1999. Nelle figure 2.3 e 2.4 sono riportate le distribuzioni del numero di esemplari adulti, stimati dai censimenti estivi, in riferimento alla quota. I dati sono espressi in forma aggregata per i due anni di censimento. Si possono così confrontare i valori proposti in letteratura, con quelli risultati dai censimenti delle popolazioni presenti nel territorio provinciale.

2.1.3 Arene di canto

Un aspetto caratteristico della specie sono indubbiamente le parate collettive che precedono l'accoppiamento, che avvengono nelle cosiddette arene di canto. Il periodo è compreso tra la fine di aprile e la fine di maggio, in funzione delle condizioni meteorologiche e dell'esposizione del luogo di parata.

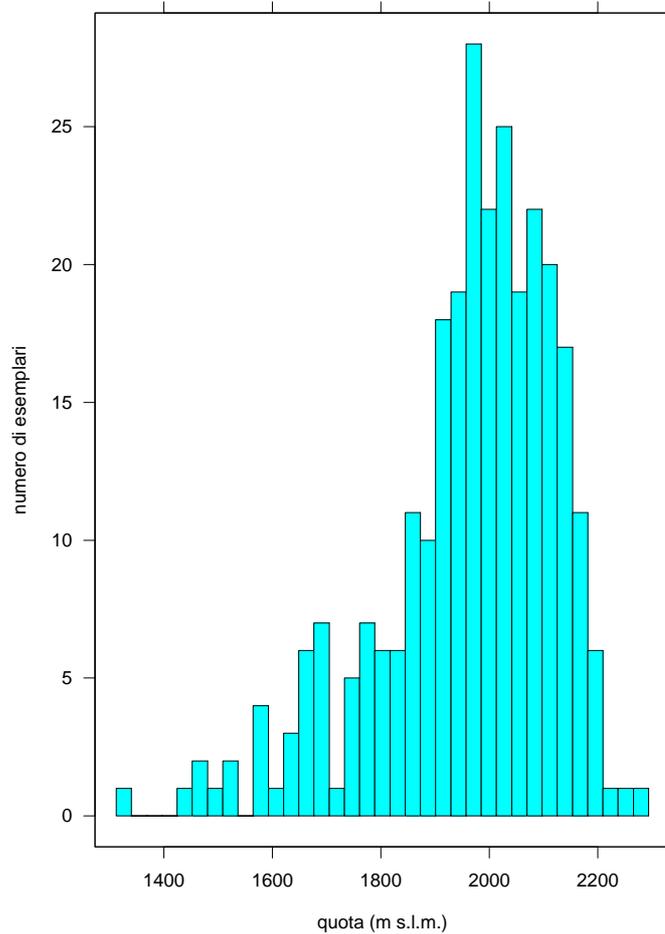


Figura 2.3: Distribuzione esemplari adulti in funzione della quota: dato aggregato per gli anni 1998-1999 derivato da elaborazione dei censimenti estivi.

Le arene di canto sono localizzate in genere su dossi o terrazzi ancora coperti almeno in parte da neve e posti in genere in prossimità del limite superiore del bosco. Sono scelti questi luoghi perché valorizzano al massimo l'effetto dei segnali ottici (dovuti al piumaggio) ed acustici emessi al momento della parata.

Le arene di canto sono delle aree importanti per gli equilibri dell'intera popolazione. I disturbi antropici che si manifestano in questi luoghi, soprat-

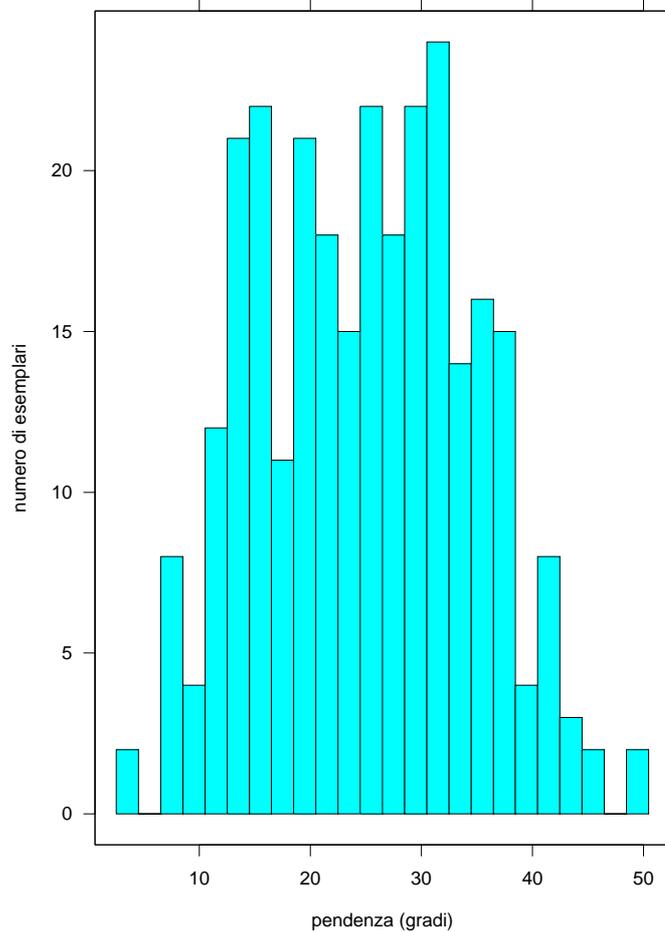


Figura 2.4: Distribuzione esemplari adulti in funzione della pendenza: dato aggregato per gli anni 1998-1999 derivato da elaborazione dei censimenti estivi.

tutto durante l'attività di canto, hanno in genere conseguenze molto negative sulle dinamiche della popolazione.

2.1.4 Dinamiche della popolazione

Da alcuni studi generalizzabili all'intero arco alpino [14], si è evidenziato che il numero di esemplari ha subito un grosso calo, a cominciare dall'ultimo optimum climatico che terminò nei primi anni cinquanta. Importanti

relazioni si sono infatti verificate, tra le dinamiche delle popolazioni di fagiano di monte e le condizioni climatiche. In particolare, fondamentali sono le temperature e le precipitazioni che si hanno nelle prime tre settimane di svezzamento dei pulcini, quando questi non sono in grado di conservare da soli la temperatura corporea. In questo periodo, le precipitazioni, soprattutto di tipo nevoso, causano la perdita di intere covate. Le aree di allevamento costituiscono, al pari delle arene di canto, delle zone di importante rilevanza biologica per la specie. Anche le temperature e le precipitazioni invernali sono particolarmente importanti perché, per evitare eccessive perdite di energia, i fagiani di monte sono portati a passare gran parte del giorno e della notte in gallerie scavate nella neve. Sono quindi sfavorevoli inverni con poca neve o con neve gelata.

Il calo che si è avuto negli ultimi anni non è comunque legato solamente alla variazione delle condizioni climatiche. Esistono infatti numerose altre influenze, soprattutto di tipo antropico, da porsi in relazione sia all'aumento del carico turistico (primaverile e invernale), sia il diverso sfruttamento del territorio montano. La pratica dello sci fuori pista, ad esempio, ha conseguenze paragonabili alla mancanza di neve nel periodo invernale [14]. I pendii all'ombra, ideali per lo sciatore data la loro neve polverosa, vengono scelti anche dal fagiano di monte per scavarvi le proprie gallerie. Gli uccelli, disturbati, perdono perciò più energia di quella che possono acquistare nutrendosi e perdono così peso. Compromettono in questo modo, oltre alla sopravvivenza invernale, anche la loro produttività nella stagione successiva [14]. La pratica degli sport invernali ha anche comportato la costruzione di una serie di infrastrutture, seggiovie, piste, skilift che arrecano disturbo, inoltre le stazioni di tali impianti sono poste negli stessi luoghi della arene di canto (dossi, terrazzi e creste). Anche le arene che si trovano lungo il tracciato di una pista vengono in genere abbandonate, perché i territori di parata conquistati in autunno non riescono ad essere mantenuti durante l'inverno. L'abbandono delle arene di canto ha conseguenze molto negative su popolazioni di vaste zone [14].

2.2 Il problema della gestione faunistica del fagiano di monte

Con il termine *gestione faunistica* si intende, l'insieme delle regole poste da enti competenti al fine di migliorare o mantenere lo stato di una popolazione in base a determinati obiettivi, che nel caso del fagiano di monte riguardano principalmente lo stato di salute degli esemplari, la loro presenza sul territorio e la garanzia di un sufficiente grado di riproduzione.

Tralasciando gli aspetti venatori della gestione delle popolazioni alpine di fagiano di monte, un corretto approccio dovrebbe stabilire il valore faunistico delle aree frequentate dagli esemplari della specie durante le stagioni dell'anno più critiche e delicate:

- aree di canto (punti o arene)
- aree di allevamento di covata
- territori individuati dagli adulti al di fuori del periodo degli amori (aree di svernamento in particolare)

Per questo, nella valutazione del valore faunistico si dovrebbero esaminare diversi fattori estrinseci alla specie, con attenzione soprattutto ai seguenti aspetti: [15]

- estensione e stato delle aree pascolate;
- utilizzazioni forestali e problemi selvicolturali;
- vie di penetrazione, turismo, frequentazione antropica;
- densità di predatori e parassiti in genere;
- disturbi nei luoghi di parata e nelle aree di svernamento.

Elementi fondamentali per reperire tali informazioni sono sicuramente i censimenti che, per essere validi e realmente utilizzabili, devono essere attuati in modo standardizzato rispetto ai tempi, e con protocolli ben precisi. Per questo l'amministrazione provinciale ha intrapreso una campagna di censimento rivolta alla conoscenza delle specie ritenute più importanti.

Considerando l'estensione delle aree in cui la specie vive e la necessità di ridurre i costi di queste operazioni, i censimenti sono effettuati in aree campione omogenee. Le aree campione sono individuate nell'ambito di ciascuna riserva faunistica, in modo da rappresentare la situazione attuale e quindi garantire l'acquisizione di dati affidabili.

In riferimento alla gestione del fagiano di monte, sono in corso in provincia di Trento due tipi di censimenti. Il primo è il censimento primaverile con il quale si intende rielvare la frequentazione delle arene di canto, ed è compiuto durante il periodo degli amori, a partire dal 1995. Il secondo è invece il censimento estivo e si riferisce al rilevamento delle covate e degli esemplari adulti presenti sul territorio ed è eseguito in estate a partire dal 1993.

Per il Trentino le aree campione rimangono pressoché invariate di anno in anno ed hanno ampiezza variabile dai 40 ettari a valori maggiori di 700.

2.2.1 Data base dei censimenti estivi

Il censimento estivo del fagiano di monte è finalizzato al rilevamento del numero e del tipo di esemplari presenti all'interno di ciascuna delle aree campione. Attraverso il censimento estivo si vogliono evidenziare tre importanti aspetti:

1. La consistenza faunistica nelle aree campione distinguendo tra:
 - numero maschi adulti;
 - numero femmine adulte senza covate;
 - numero femmine con covata;
 - numero covate.
2. La localizzazione sul territorio dei punti di contatto di ciascun esemplare.
3. L'indagine sulle aree di allevamento prole:
 - numero e sesso dei pulli e valutazione del loro stato di sviluppo;
 - analisi vegetazionale per la zona di covata;

- condizione atmosferiche al momento del censimento.

Dal punto di vista pratico, il censimento è svolto da operatori che, con l'ausilio di cani da ferma, coprono l'intera area campione. Il periodo di censimento è compreso tra la fine di agosto e i primi giorni di settembre.

Tutte le informazioni rilevate attraverso i censimenti sono riportate nelle schede e nella carta compilate per ciascuna area campione. Gli aspetti relativi alla consistenza faunistica per area campione sono riportati direttamente nel *prospetto riassuntivo* (figura 2.5). Parallelamente, è riportata su cartografia 1:10000 la posizione del punto di contatto degli esemplari adulti e delle covate di fagiano di monte (figura 2.6). Le informazioni dettagliate sulle zone di presenza di covate sono invece riportate nella scheda di *indagine sulle aree di allevamento prole* (figura 2.7).

In tabella 2.2.1 si sono riportate tutte le aree campione censite negli anni 1998 e 1999. Con il simbolo *S* si sono evidenziate le aree campione per le quali si dispone di informazioni spaziali, cioè per le quali nei censimenti si è allegata la cartografia.

Da tale tabella si evince che, meno del 50% delle aree campione censite ha riportata in allegato la relativa cartografia dei punti di contatto. Nel presente lavoro si è invece interessati in modo particolare alle caratteristiche spaziali dei dati di censimento, poiché queste devono essere confrontate con i tematismi GIS a nostra disposizione. In riferimento a questo, l'informatizzazione dei dati e le successive elaborazioni sono state eseguite solo per queste aree campione.

2.3 Data base dei censimenti primaverili

Il censimento primaverile del fagiano di monte, è essenzialmente finalizzato al rilevamento del numero e tipo di esemplari che frequentano le arene di canto nel periodo degli amori. In particolare, le informazioni rilevate si possono suddividere nei seguenti punti:

1. Il numero di esemplari presenti nell'arena di canto distinguendo tra:
 - maschi adulti;

Riserva	1998	1999	Riserva	1998	1999
Ala	N	-	P.n. Stelvio	A	N
Amblar-Don	N	N	Panchià - Ziano	AS	S
Avio 1	S	N	Peio Pns	N	AS
Avio 2	S	N	Peio Val Comasine	N	N
Bleggio Superiore	-	N	Peio-Pellizzano	AS	S
Brentonico Lastè	A	N	Pieve Tesino	S	AS
Caderzone	S	S	Pieve di Ledro	N	N
Canal S. Bovo Paradisi	AS	S	Pinzolo	-	N
Canal S. Bovo Valsorda	AS	S	Pozza di Fassa	N	-
Castel Tesino	-	AS	Primiero Calaita	N	AS
Castelfondo	N	-	Primiero Tognola	N	AS
Castello di Fiemme	N	N	Rabbi Pns	S	AS
Cles	N	-	Roncegno	-	N
Commezzadura	-	AS	Rovereto	N	-
Coredo	N	N	Rumo	N	N
Daone	N	N	Sfruz-Smarano	S	N
Demanio Bondone	S	N	Soraga	AS	N
Demanio S. Martino	N	N	Spinale-Manez	AS	S
Dimaro	S	-	Strigno	AS	N
Fiavé	-	N	Terlago	AS	S
Fierozzo	AS	S	Tesero	-	S
Folgaria	N	N	Tesero La Porta	S	AS
Folgaria 2	N	-	Tesero Slavin	S	AS
Frassilongo	N	N	Tesero Varena	-	N
Malé	-	AS	Trambileno	N	-
Malga Campo	N	N	Val Campelle	N	N
Malga Valli	S	N	Valfloriana	A	S
Mezzana	S	-	Vallarsa	-	N
Mezzano	S	AS	Vezzano	N	AS
Moena	AS	S			

A: Riserva con cartografia delle aree campione.

S: Riserva censita con cartografia dei punti di contatto.

N: Riserva censita senza cartografia dei punti di contatto o con cartografia incompleta.

–: Riserva non censita.

Tabella 2.1: Riserve censite e tipo di materiale cartografico disponibile per gli anni 1998 e 1999

- maschi giovani;

- femmine.

2. La localizzazione sul territorio delle arene di canto.

3. Le condizioni atmosferiche al momento del censimento e l'entità della copertura nevosa al suolo.

I censimenti primaverili sono svolti per aree campione omogenee durante il mese di maggio, da parte di operatori che osservano le arene di canto presenti. In questo periodo sono previste almeno tre o quattro uscite per garantire una maggiore uniformità nei dati.

In figura 2.8 si è riportato un esempio di scheda di campagna per i censimenti primaverili che viene compilata per ciascuna arena di canto. Come per il censimento estivo, anche in quello primaverile è prevista da parte degli operatori l'individuazione dei punti di localizzazione delle arene di canto sulla carta tecnica provinciale 1:10000; un esempio di tale cartografia è riportato in figura 2.9.

Infine è anche prevista la compilazione del prospetto riassuntivo (figura 2.10) che raccoglie i risultati delle osservazioni per le arene di canto presenti in un'area campione. Nel prospetto riassuntivo sono riportati contemporaneamente i rilevamenti effettuati nelle diverse uscite.

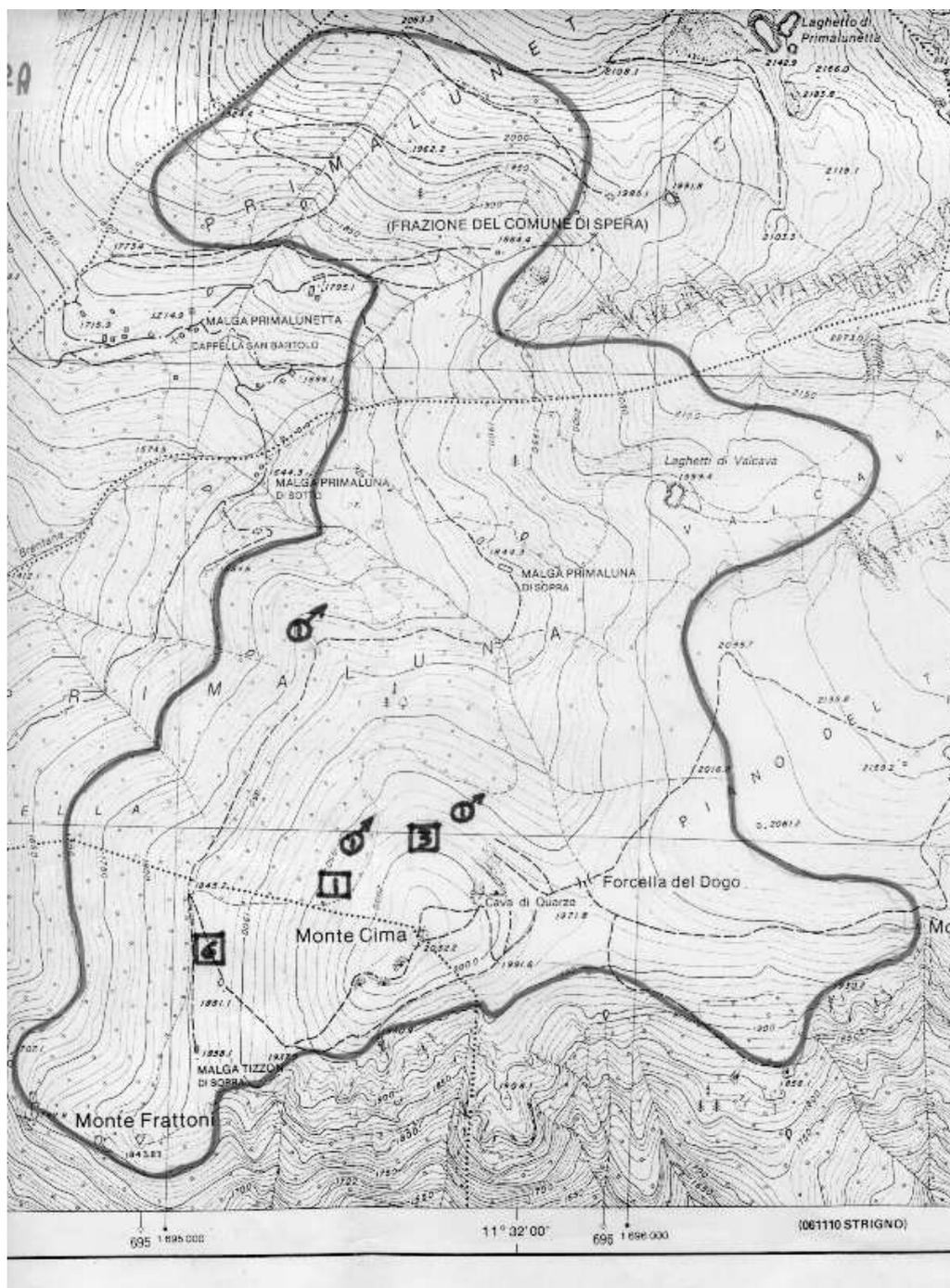


Figura 2.6: Censimento estivo: cartografia dell'area campione e degli avvistamenti.



PROVINCIA AUTONOMA DI TRENTO
Servizi Forestali

SCHEDA DI
CAMPAGNA

CENSIMENTO ESTIVO FAGIANO DI MONTE - ANNO 1998

INDAGINE SULLE AREE DI ALLEVAMENTO PROLE

Rilevatore n. 3

Area campione n. _____ Riserva AVIO Località PRA' ALPESINA

Ente propr. Com. BREZZONICO Cod. piano 322 N° sez. p. assest. 157

Altitudine 1600 - 1800 Esposizione EST Pendenza % 40

AVVISTAMENTO ADULTI

Sesso M F ind. Localizzazione: VEDI CARTINA

N° 1

SEGNALAZIONE COVATA

Femmina madre N° giovani 4 maschi 2 femmine 2

Stadio di sviluppo DISCRETO

ANALISI VEGETAZIONALE ZONA COVATA

Specie arboree - antano verde - pino mugo

- altezza maggiore di 2 m: grado % di copertura specie _____

- altezza minore di 2 m: grado % di copertura specie _____

totale 1

Specie arbustive ed erbacee

rododendro - ginepro - salice: grado % di copertura 30 altezza media in cm 100

mirtilli e ericacee: grado % di copertura altezza media in cm _____

graminacee: grado % di copertura 60 altezza media in cm 10

altre: (rovi _____ grado % di copertura 10 altezza media in cm 30

e lamponi - _____

felci - lonicera - rose ecc.)

totale 100 1

100 1 - 2

Zona soggetta a pascolo si no

Note: CIELO SERENO - TEMPERATURA ELEVATA

Rilevatore ROSSI DOMENICO - Rossi Domènica Qualifica AGENTE

FIRMA DEL RILEVATORE - CENSIMENTO ATTRIBUITO IN MECCANO
AD OGNI FOLIA COSTE

Data rilievo 22.08.98

Sig. CERADINI FLAVIO, CONDUTTORE N° 2 CANI

Figura 2.7: Censimento estivo: scheda di campagna.



PROVINCIA AUTONOMA DI TRENTO
Servizi Forestali

SCHEDA
CAMPAGNA AREE
CAMPIONE

CENSIMENTO PRIMAVERILE - ANNO 98

FAGIANO DI MONTE

 GALLO CEDRONE

AREA CAMPIONE N. _____ ARENA N. _____
 RISERVA AVIO LOCALITA' ARTICOLONE
 CAPO SQUADRA _____ COLLABORATORE/I _____
 DATA RILIEVO 10-05-98

fredda <input type="checkbox"/>	sereno <input checked="" type="checkbox"/>	nebbia <input type="checkbox"/>	vento <input type="checkbox"/>	intens. _____
mita <input checked="" type="checkbox"/>	variabile <input type="checkbox"/>	foschia <input type="checkbox"/>	poggia <input type="checkbox"/>	intens. _____
caldo <input type="checkbox"/>	coperto <input type="checkbox"/>	ragiada <input type="checkbox"/>	neve <input type="checkbox"/>	intens. _____
Condizioni del mano nevoso:		copertura continua <input type="checkbox"/>	a chiazze <input type="checkbox"/>	assenza <input type="checkbox"/>

Maschi adulti (+ di 1 anno) n° <u>3</u>	maschi di età indet. n° <input type="checkbox"/>	sogg. indet. per sesso n° <input type="checkbox"/>
Maschi giovani (1 anno) n° <u>1</u>	femmine n° <u>2</u>	accoppiamenti n° <input type="checkbox"/>
Disturbo antropico durante il rilievo <input checked="" type="checkbox"/> <u>NO</u>		

OSSERVAZIONI:

Data rilievo 10-5-98 RILEVATORI GR. S. M. MARIZIO

Figura 2.8: Censimento primaverile: scheda di campagna.

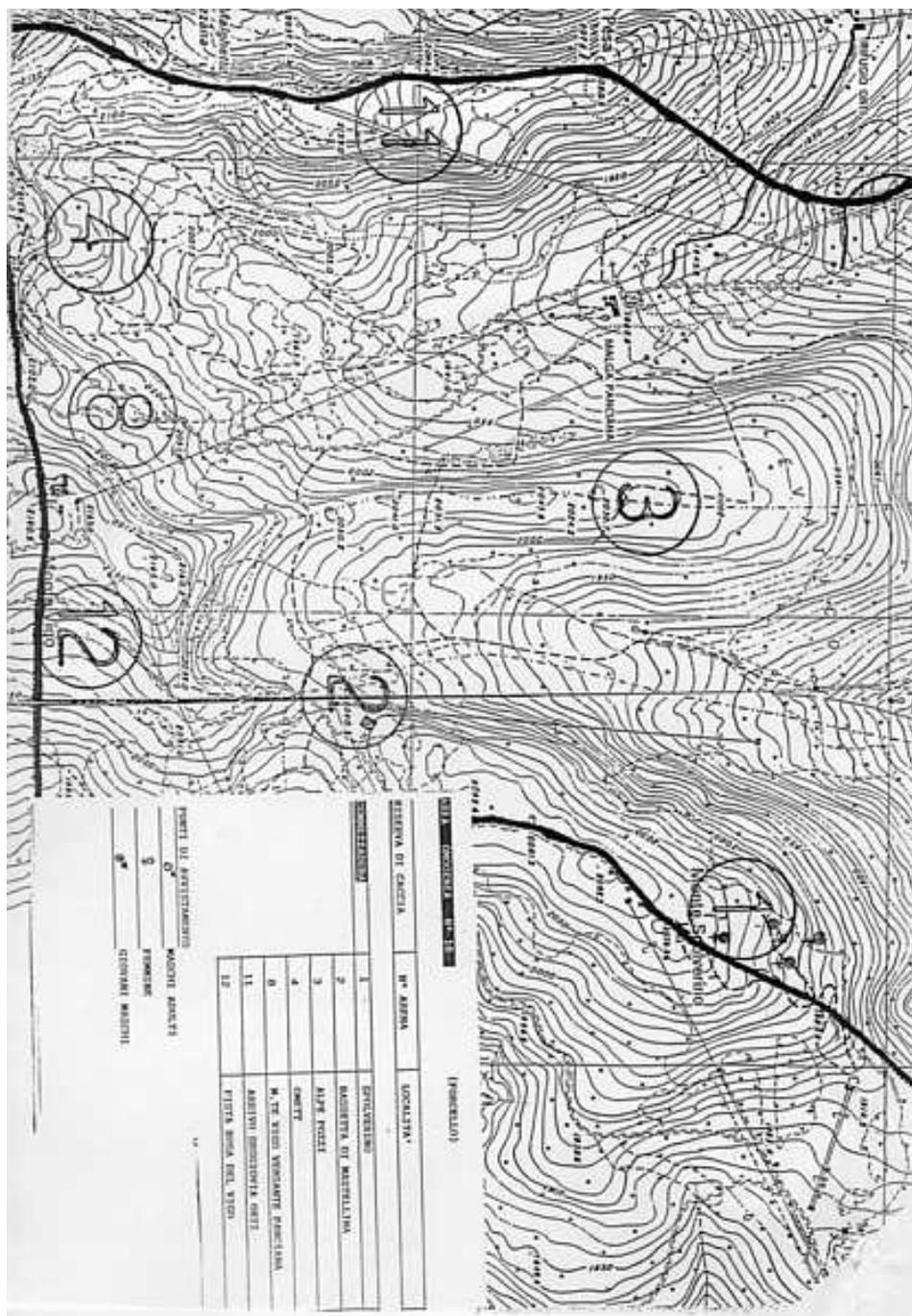


Figura 2.9: Censimento primaverile: cartografia dell'area campione e degli avvistamenti.



 Provincia Autonoma di Trento

 Servizi Forestali

CENSIMENTO FAGIANO DI MONTE ANNO 1998

PROSPETTO RIASSUNTIVO

Area Omogenea n./nome:
 Riserva/le:
 Superficie totale ha:

DATA	Osserv. n. 1		Osserv. n. 2		Osserv. n. 3		Osserv. n. 4		Osserv. n. 5		Osserv. n. 6		N. max. ind. osserv.	
	m	f	m	f	m	f	m	f	m	f	m	f	maschi	F
	a	g	a	g	a	g	a	g	a	g	a	g	a	g
05/05/98	1												1	1
10/05/98	3	1	2										3	1
12/05/98	3		1										3	1
22/05/98	1												1	1
N. uccise predat.													3	1
													4	2

adulti # giovani # addebiatati
 F. femmine di ritorno ac. accoppiamenti

Figura 2.10: Censimento primaverile: prospetto riassuntivo.

Capitolo 3

Il nuovo data base informatizzato

Si presentano in questo capitolo, gli aspetti relativi all'informatizzazione dei dati di carattere faunistico. L'obiettivo è quello di creare un data base correttamente strutturato per consentirne l'utilizzo nell'ambito del sistema informativo territoriale.

Si vedranno le scelte operate in fase di progettazione del data base relazionale, e gli strumenti software appositamente sviluppati per l'inserimento dei dati nel nostro sistema GIS integrato. Le procedure sono state utilizzate per l'inserimento dei dati dei censimenti estivi del fagiano di monte, relativi al biennio 1998-1999. Si esporranno le caratteristiche di questo data base prodotto ed alcune elaborazioni spaziali eseguite sullo stesso.

3.1 Strumenti software

L'inserimento dei dati di censimento ha presentato alcuni problemi che hanno richiesto l'utilizzo di strumenti software specifici.

Le problematiche, principalmente di carattere pratico, e in parte legate alle scarse caratteristiche grafiche dell'ambiente GRASS, possono essere così evidenziate:

1. Difficoltà di individuare, dalla cartografia in allegato ai censimenti l'inquadramento geografico delle zone di interesse. Tale cartografia, è molte volte un estratto parziale dei fogli della carta tecnica 1:10000 per

territori montani privi di riferimenti significativi. A riguardo, è risultata complicata anche la ricerca partendo dalle informazioni sulle riserve.

2. Difficoltà delle normali interfacce GRASS di gestire, per un dato tematismo, la visualizzazione di insieme e di dettaglio attraverso zoom dinamici. Questo è invece un aspetto importante che facilita il corretto e rapido inserimento dei dati spaziali.
3. Difficoltà di collegare in maniera automatica l'inserimento della localizzazione spaziale dei dati e gli attributi ad essi associati.

Per risolvere il primo dei problemi appena descritti si sono utilizzate le informazioni particolari contenute nel data base sul corredo toponomastico allegato alla CTP digitale. Questo data base, parzialmente rivisto dal servizio beni librari e archivistici e dalla commissione provinciale per la toponomastica, contiene essenzialmente i seguenti campi:

- scritta del corredo toponomastico;
- classe del toponimo;
- coordinate Gauss Boaga di inquadramento della scritta nella CTP.

La disponibilità di tali informazioni ha consentito lo sviluppo di una procedura automatica che, utilizzando i nomi dei toponimi presenti nella cartografia dei censimenti identifi chi e visualizzi la stessa porzione di territorio per la CTP digitale. La procedura è stata implementata in ambiente GRASS tramite un nuovo modulo denominato `d.positionate`. Un esempio di applicazione della procedura è riportato in figura 3.1 ricercando nel data base del corredo toponomastico la localizzazione per la scritta “malga primaluna” e visualizzando la regione quadrata di 1300 m di lato nell’intorno del sito.

Si consideri ora l’esigenza di utilizzare un’adeguata interfaccia grafica per l’inserimento nel data base digitale delle informazioni sulla localizzazione spaziale dei punti di contatto e sugli attributi associati a questi. A riguardo si è sviluppata in collaborazione diretta con l’Ite-Irst una interfaccia grafica per GRASS che consente di gestire la visualizzazione complessa di un qualunque

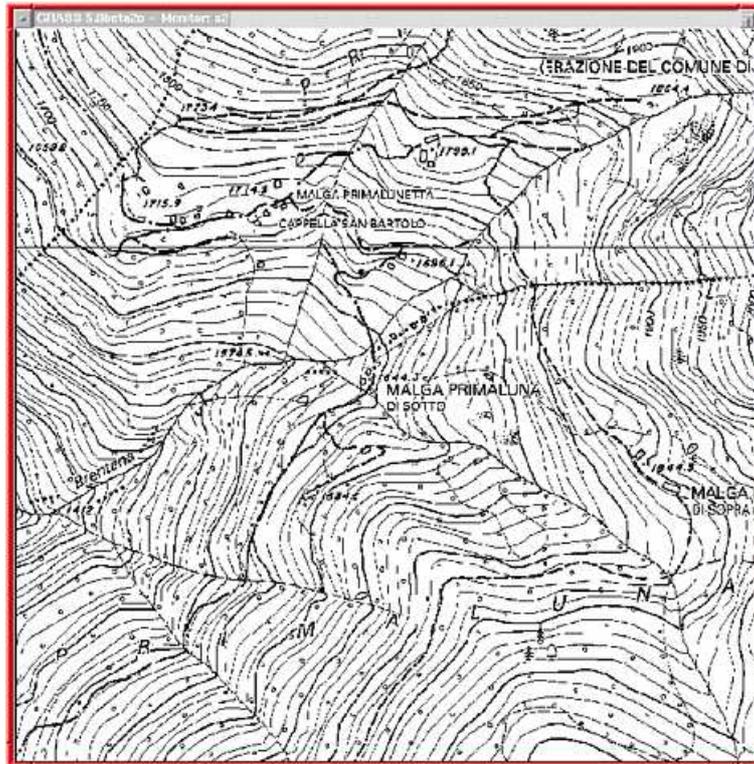


Figura 3.1: Risultato della ricerca del toponimo “malga primaluna” con `d.positionate`.

tematismo GIS a disposizione. L’interfaccia permette la visualizzazione contemporanea della zona di interesse a diverse scale e dopo l’inserimento dei siti consente di inserire direttamente nel data base digitale le informazioni ad essi associate desunte dai censimenti. L’interfaccia grafico utilizzato per l’inserimento dei siti di censimento per il fagiano di monte è implementato nel modulo GRASS `s.miglia` ed un esempio di utilizzo è riportato in figura 3.2. Sono visibili nella figura le sotto finestre con i diversi “zoom” e il terminale a caratteri per l’inserimento degli attributi. La zona è la stessa visualizzata precedentemente utilizzando `d.positionate` (figura 3.1).

La procedura generale sviluppata per l’inserimento dei punti di contatto dei censimenti deriva dalla combinazione dei due moduli appena descritti e nella sua globalità ed è descritta nelle sue parti in figura 3.3.

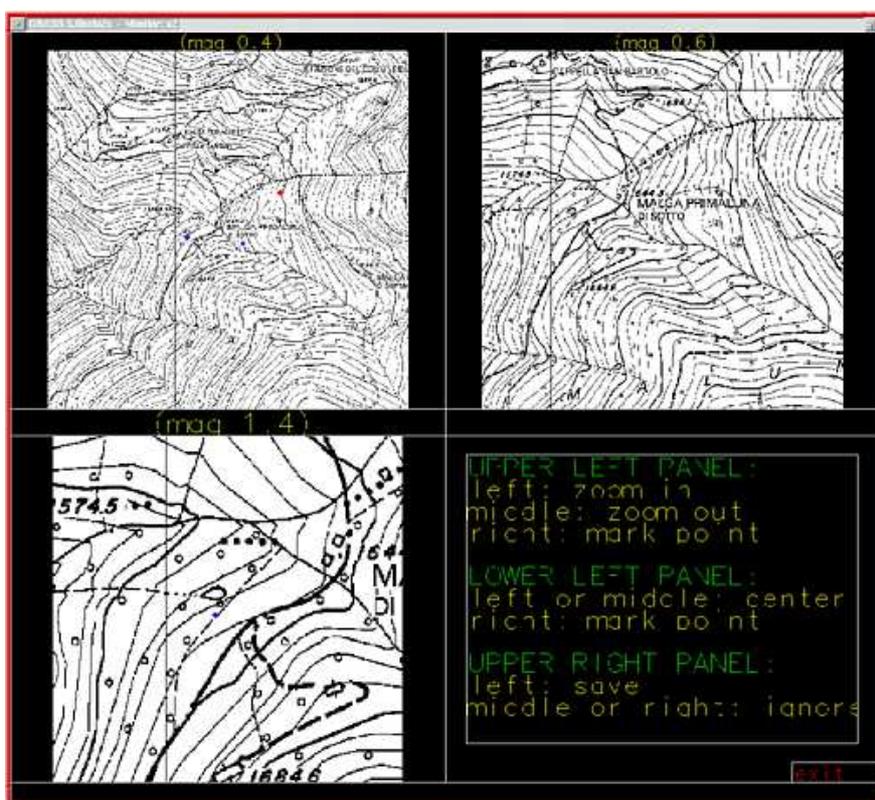


Figura 3.2: Esempio di interfaccia per l'inserimento dei siti nel data base.

3.2 Data base

3.2.1 Organizzazione del data base

L'organizzazione di un data base è strettamente legata allo scopo del suo utilizzo e al tipo di informazioni contenute. Nel nostro caso per lo sviluppo del data base ci siamo posti i seguenti obiettivi:

1. Immagazzinare tutte le informazioni contenute nelle schede ed allegati dei censimenti estivi.
2. Garantire un rapido accesso ai dati dai diversi ambienti che formano il sistema integrato:
 - l'ambiente GRASS;
 - l'ambiente S-PLUS.

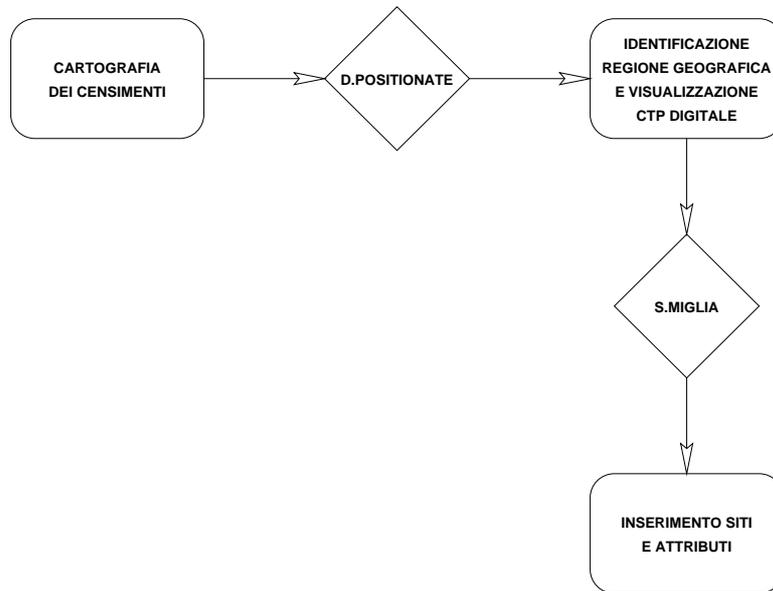


Figura 3.3: Schema della procedura sviluppata ed utilizzata per l'inserimento dei dati dei censimenti estivi di fagiano di monte.

3. Garantire un rapido inserimento di nuovi dati.
4. Evitare l'eccessiva ridondanza e garantire l'integrità del data base.

Le singole tabelle che compongono il data base sono:

CONTATTI: contiene come chiave primaria il numero identificativo del contatto e negli altri campi le coordinate Gauss Boaga del sito di avvistamento e la codifica degli esemplari avvistati;

CONTATTI					
<u>ID</u>	coord EAST	coord NORD	n maschi	n femmine	n pulli
1	1647731	5073950	0	1	6
2	1648401	5074082	0	1	7
3	1646202	5073977	0	1	0
4	1646092	5073771	0	1	7
5	1645958	5073421	0	1	3
6	1646048	5073451	1	0	0

COVATE: la chiave primaria è ancora una volta il numero identificativo del contatto e contiene le informazioni desunte dalla scheda di campagna per le aree di allevamento della prole del censimento estivo;

COVATE									
ID	>2m	%>2m	<2m	%<2m	%arb1	%arb2	%arb3	%arb4	pascolo
1	-	-	-	-	50	30	20	0	n
2	1	20	1	10	10	30	30	0	y
4	1	20	1	10	10	30	30	0	y
5	1	20	1	10	10	30	30	0	y

I campi presenti in questa tabella descrivono gli aspetti vegetazionali delle in cui si è localizzata la covata ed hanno i seguenti significati:

>2m: Tipo di specie arboree con altezza maggiore di 2 m.

%>2m: Grado percentuale di copertura delle specie arboree con altezza maggiore di 2m.

<2m: Tipo di specie arboree con altezza minore di 2 m.

%<2m: Grado percentuale di copertura delle specie arboree con altezza minore di 2 m.

%arb1: Grado percentuale di copertura per le specie arbustive ed erbacee di primo tipo: rododendro, ginepro, salice.

%arb2: Grado percentuale di copertura per le specie arbustive ed erbacee di secondo tipo: mirtilli, ericacee.

%arb3: Grado percentuale di copertura per le specie arbustive ed erbacee di terzo tipo: graminacee.

%arb4: Grado percentuale di copertura per le specie arbustive ed erbacee di quarto tipo: altre (es. rovi, lamponi, felci, lonicera, rose).

pascolo Identifica la presenza di pascolo.

AREE CAMPIONE : contiene per ciascun punto di contatto il codice identificativo dell'area campione di appartenenza;

AREE CAMPIONE	
ID	codice aree
3	AC6
4	AC6
5	AC3
6	AC3

CODICI VEGETAZIONE: interpretazione del codice identificativo per la vegetazione utilizzati nella tabella covate;

CODICI VEGETAZIONE	
codice vegetazione	descrizione
1	abete rosso, larice
2	larice
3	larice, abete rosso, ontano
4	larice, pino cembro
5	pascolo, pino mugo, faggio

CODICI AREE CAMPIONE: interpretazione del codice identificativo per le aree campione.

CODICI AREE CAMPIONE		
codice area	area campione	anno ins.
AC3	pellizzano.17	1998
AC4	terlago.9	1998
AC5	strigno	1998
AC6	moena.1	1998

Si sono infine sviluppati dei moduli in linguaggio Perl e Shell di Unix per operare interrogazioni al data base e anche per visualizzare graficamente tali risultati. Un esempio di questo è il modulo per GRASS `d.miglia.labels` che consente di visualizzare a video la localizzazione dei siti di contatto con il codice identificativo del sito e il valore di alcuni suoi attributi. In figura 3.4 si è riportato il risultato dell'utilizzo del modulo `d.miglia.labels` per visualizzazione dei punti di contatto e degli attributi relativi al numero di maschi, di femmine e di pulli presenti nel 1998.

3.2.2 L'informatizzazione dei censimenti estivi 1998 e 1999

Si sono inseriti nel data base le informazioni contenute nei censimenti estivi del fagiano di monte per gli anni 1998 e 1999. Più precisamente, poiché si è interessati a cogliere le relazioni spaziali dei dati, sono stati inseriti i



Figura 3.4: Esempio di utilizzo del modulo GRASS `d.miglia.labels` per la visualizzazione degli attributi “*n*° maschi”, “*n*° femmine”, “*n*° pulli”.

punti di contatto con i relativi attributi soltanto per le aree campione i cui censimenti riportavano in allegato la cartografia degli avvistamenti.

Nella tabella 2.2.1 descritta nel precedente capitolo si sono evidenziate con il simbolo *S* le riserve per le quali è disponibile la cartografia e che quindi rappresentano le sole aree campione inserite nel data base informatizzato.

In tabella 3.2.2 è presentato il prospetto riassuntivo con la quantità di dati inseriti per ciascuno dei due anni.

	1998	1999
aree campione censite	49	50
aree campione inserite	23	23
siti inseriti	135	122
esemplari inseriti		
maschi	65	55
femmine	80	84
covate	58	52
pulli	208	176
totale esemplari	353	315

Tabella 3.1: Riassunto dei dati informatizzati.

In figura 3.5 si è mostrata la visualizzazione generale per il Trentino di tutti i siti inseriti per gli anni 1998 e 1999.

3.3 Cartografia vettoriale delle aree campione

Il data base informatizzato dei censimenti estivi è stato completato con la digitalizzazione dei limiti geografici delle aree campione. Poiché l'estensione e la forma delle aree rimangono invariate negli anni, e visto che per uno stesso anno non sono riportate nella cartografia allegata tutte le aree campione, si è proceduto a digitalizzare prima tutte quelle disponibili con il censimento del 1999 e per quelle mancanti si è utilizzata la cartografia presente nel censimento del 1998.

In tabella 2.2.1, con il simbolo *A* sono evidenziate le aree campione digitalizzate e l'anno di censimento a cui si è fatto riferimento.

Si sono quindi digitalizzate 25 aree campione utilizzando in sequenza i due moduli GRASS:

- `d.positionate`
- `v.digspline`

Del modulo `d.positionate` si è già parlato nella sezione precedente; mentre il modulo `v.digspline`¹ consente di digitalizzare e gestire dati in formato vettoriale, utilizzando un'interfaccia grafica.

Dalla cartografia vettoriale delle aree campione si è anche prodotta la cartografia in formato raster delle stesse, per facilitare alcune elaborazioni successive. Il modulo utilizzato per questo è `v.to.rast`.

In figura 3.6 sono riportate tutte le aree digitalizzate per il Trentino.

¹`v.digspline` è disponibile con la versione 5 di GRASS.

3.4 Tematismi GIS e grafici riassuntivi

Si riportano in questa sezione alcuni esempi delle nuove cartografie digitali prodotte (figure 3.5, 3.6, 3.7), e gli istogrammi dell'altitudine, della pendenza e dell'esposizione per ciascuna area campione (figure 3.8, 3.9, 3.10).

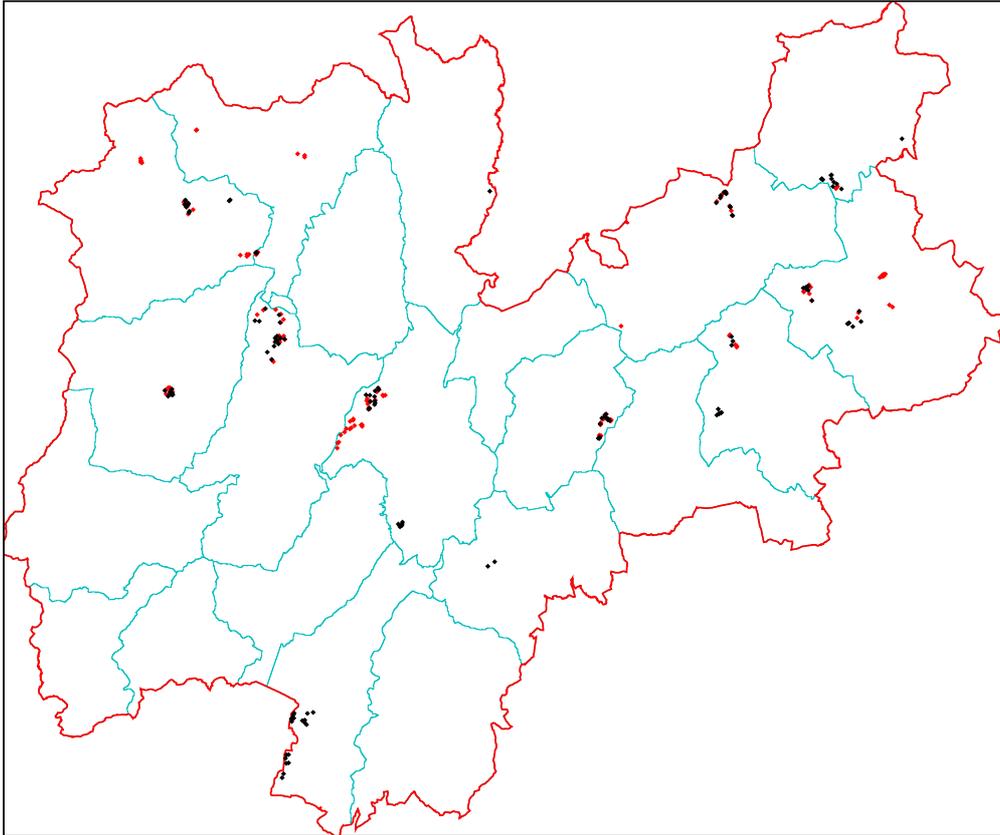


Figura 3.5: Cartografia dei siti di avvistamento e cartografia vettoriale dei distretti faunistici.

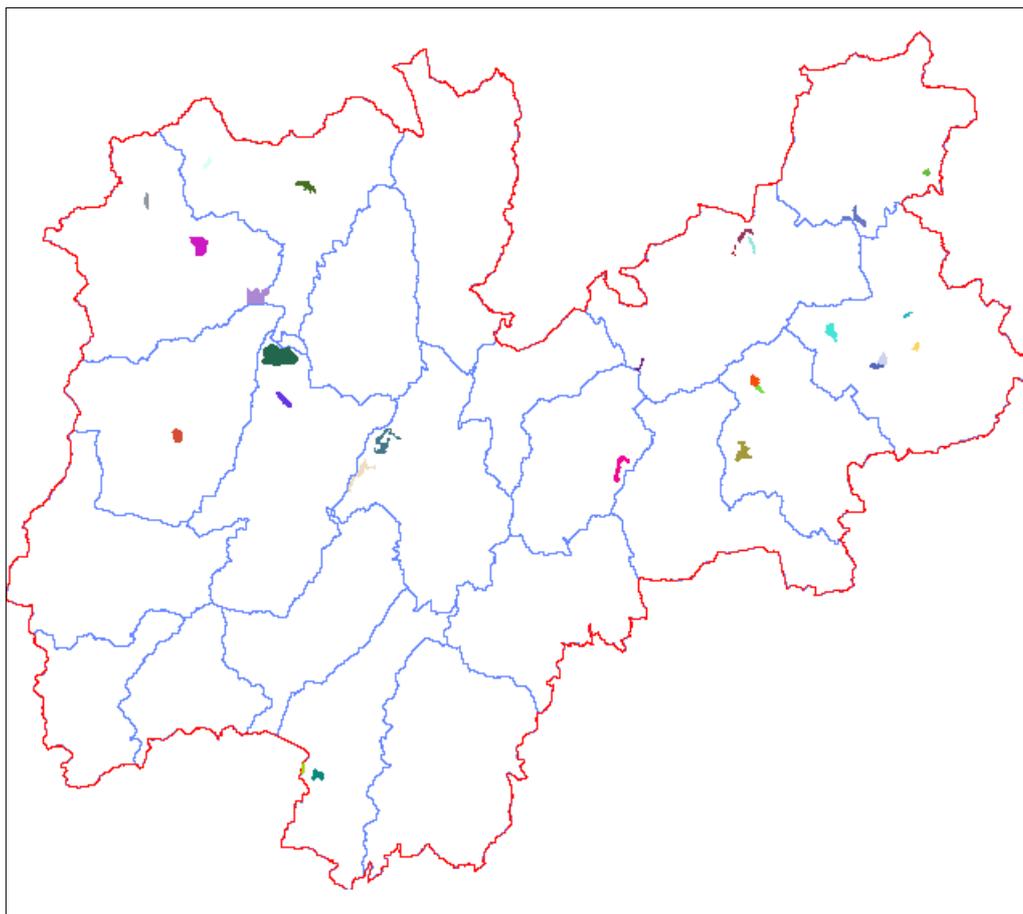


Figura 3.6: Cartografia generale delle aree campione inserite e cartografia vettoriale dei distretti faunistici.

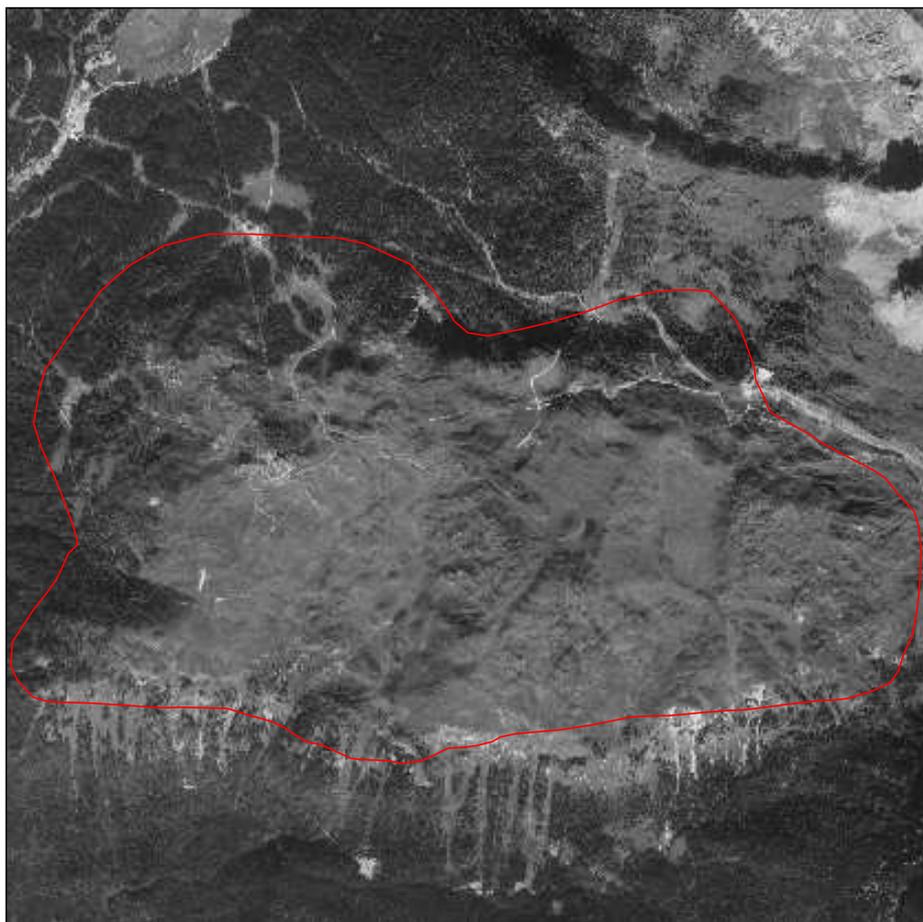


Figura 3.7: OFD e area campione digitalizzata “spinale.1” (in rosso).

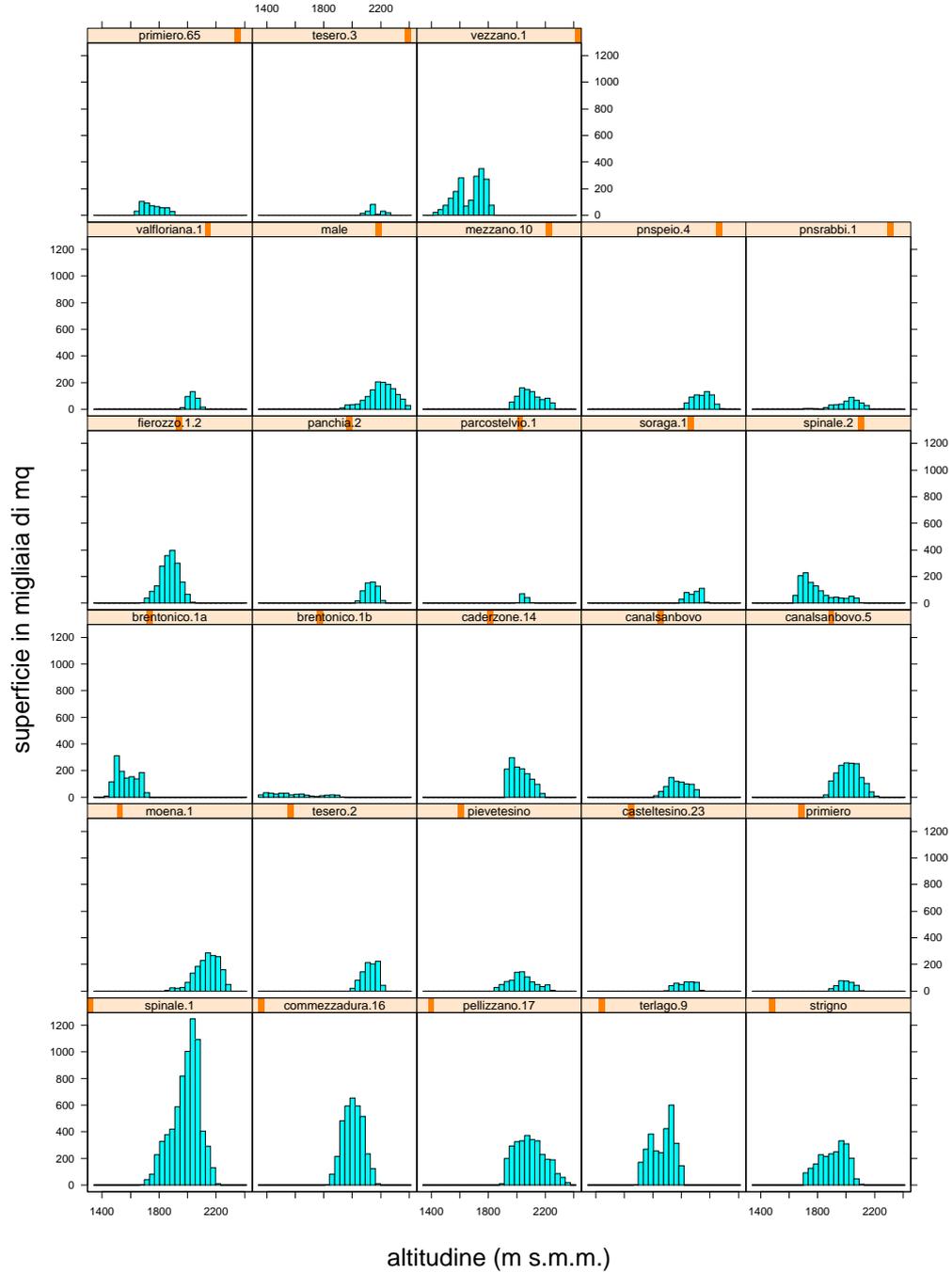


Figura 3.8: Distribuzione dell'altitudine per area campione.

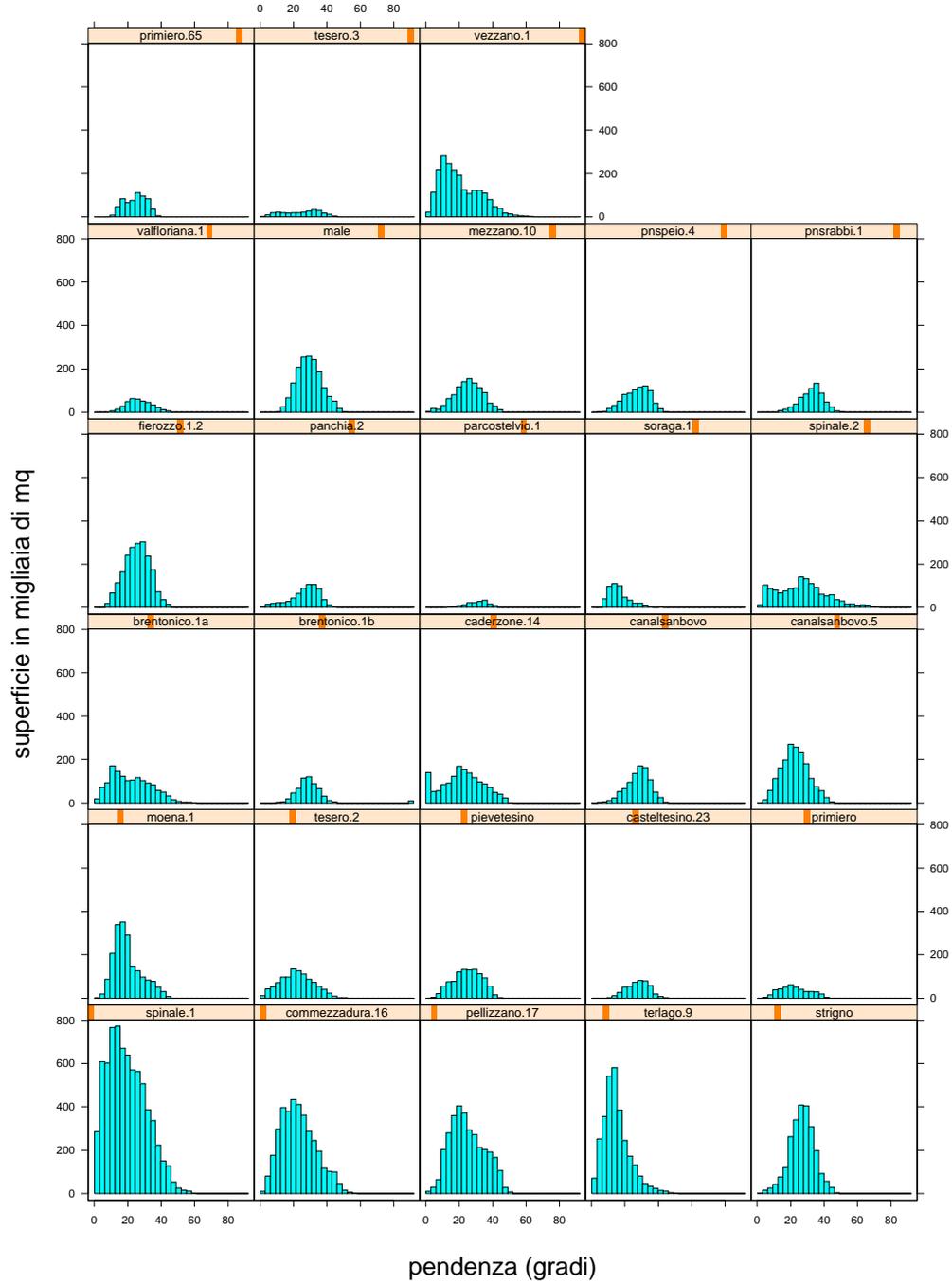


Figura 3.9: Distribuzione della pendenza per area campione.

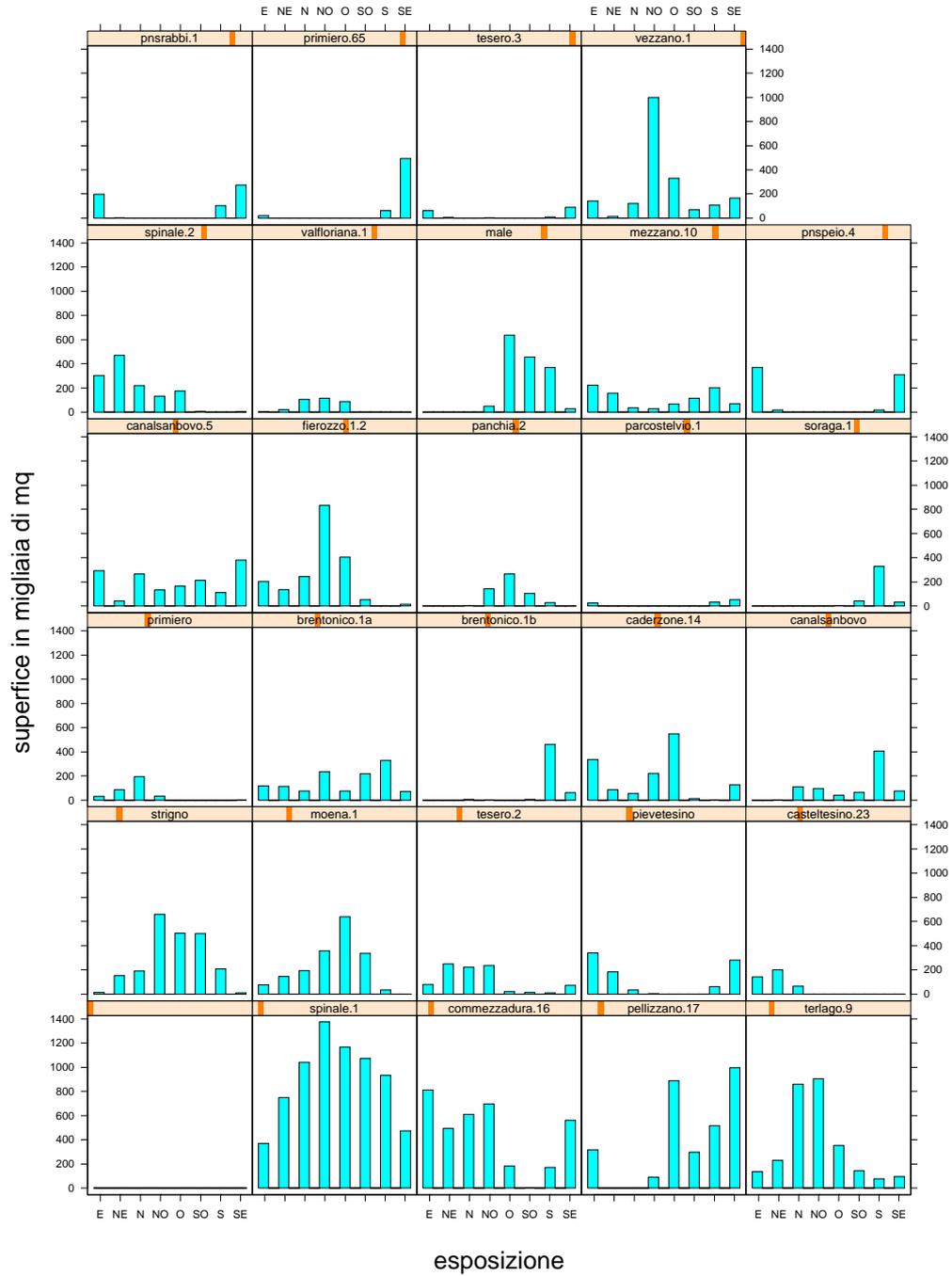


Figura 3.10: Distribuzione delle esposizioni per area campione.

Capitolo 4

Metodi di ecologia del paesaggio

Nel presente capitolo vengono illustrati i fondamenti teorici dell'ecologia del paesaggio e le metodologie di analisi proprie di questa disciplina a cui si è fatto ricorso per lo sviluppo del modello predittivo.

4.1 Approccio dell'ecologia del paesaggio

L'ecologia del paesaggio può essere pensata allo stesso tempo “sia come una teoria innovatrice dell'ecologia generale, sia come un ramo particolare di questa, occupandosi del paesaggio come specifico livello di organizzazione biologica” [12].

L'innovazione forse più importante, introdotta con l'ecologia del paesaggio è l'importanza che viene attribuita agli schemi spaziali per capire i processi ecologici in corso. Questo aspetto oltre ad essere valido a livello paesistico lo è anche agli altri livelli di organizzazione biologica; “l'eterogeneità e la variabilità del mosaico ambientale, sono in molti casi una componente anche evolutamente più importante del determinismo della distribuzione degli organismi e delle loro interazioni” [7].

Nell'ambito degli studi di ecologia del paesaggio, si sono fatti numerosi sforzi per inquadrare i mosaici spaziali in modelli descrittivi sempre più aderenti alla realtà delle cose. Si sono utilizzate tecniche ed indicatori derivati

da altre discipline, ma si sono anche sviluppate metodologie specifiche per lo studio dei paesaggi.

Con l'ecologia del paesaggio si è anche capita l'importanza di utilizzare molteplici scale spaziali e temporali per cogliere fenomeni che si manifestano a scale diverse; di conseguenza sono state sviluppate metodologie per integrare le informazioni tra tali scale.

In riferimento agli aspetti particolari di questo lavoro, si cercheranno di formalizzare le metodologie di ecologia del paesaggio per caratterizzare, attraverso un approccio multiscalare e adottando opportuni indicatori, gli aspetti strutturali dell'ambiente oggetto di studio.

L'obiettivo ultimo è, comunque, quello di mettere in relazione una siffatta descrizione dell'ambiente con i processi ecologici in corso. Le zone biologicamente rilevanti sono infatti individuate in base alla presenza e qualità di particolari specie animali. L'ecologia del paesaggio, peraltro, ha operato anche nell'ambito dell'ecologia animale fondamentali integrazioni, legate anche in questo caso alle analisi della disposizione spaziale degli ambienti. Un considerevole contributo alla comprensione del comportamento animale in un ambiente eterogeneo è venuto di recente sia dalla *teoria della lacunarità* (*percolation theory*), che consente di qualificare la condizione strutturale di un mosaico ambientale, sia dall'impiego della matematica dei frattali.

Si vedrà successivamente come si cercato di caratterizzare il paesaggio sono localmente, senza considerare i concetti di metapopolazione e di frammentazione dell'habitat.

4.1.1 Definizione di paesaggio

In Italia l'uso comune della parola paesaggio, non sembra comprendere il suo senso funzionale e interattivo, limitandosi invece a quello percettivo e visuale. È pertanto opportuno formalizzare in maniera più rigorosa ciò che si intende per paesaggio.

La prima definizione di paesaggio è quella proposta a Forman e Godron: "il paesaggio è descritto come porzione di territorio eterogenea composta da un insieme di ecosistemi interagenti che si ripete con struttura riconoscibile" [16].

Il paesaggio può essere descritto in maniera più esaustiva come sistema di unità spaziali ecologicamente diverse ma fra loro interagenti; in definitiva può essere visto come *sistema di ecosistemi* o *metaecosistema*. Questo metaecosistema è caratterizzato da molteplici domini gerarchici di scale spazio-temporali e rappresenta inoltre un livello di organizzazione della vita superiore all'ecosistema [12].

Da qui deriva la corretta definizione di ecologia del paesaggio, cioè la disciplina che studia i sistemi di ecosistemi, considerandoli però come specifico livello di organizzazione biologica. In tale senso, il paesaggio presenta caratteristiche proprie che non sono derivabili da una semplice combinazione degli attributi del livello ecosistemico. Tale visione olistica quindi, “supera la pur vera affermazione che l'ecologia del paesaggio si occupa dello studio dei fenomeni ecologici in rapporto ai cambiamenti di scala e di configurazione multidimensionale e gerarchica degli ecosistemi” [12].

In maniera schematica si possono distinguere tre fondamentali caratteristiche del paesaggio: *la struttura*, *la funzione* e *la dinamica*. La struttura è costituita dall'insieme delle relazioni spaziali tra distinti ecosistemi che si manifestano attraverso la distribuzione di energia, materiali e specie, in relazione alla dimensione, forma, numero, tipo e configurazione dei componenti. La funzione si riferisce invece alle interazioni tra elementi spaziali, che sono il flusso di energia, materia e organismi attraverso gli *ecotopi*. I cambiamenti o le dinamiche del paesaggio si riferiscono all'alterazione nella struttura e funzione del mosaico ecologico nel tempo [11].

4.1.2 Struttura del paesaggio

Dalla definizione di paesaggio come sistema di ecosistemi, segue che ogni tipo di paesaggio si può riferire ad una configurazione di ecosistemi interagenti, riconducibile a uno specifico modello strutturale [7]. La difficoltà nel definire correttamente i modelli strutturali di riferimento è comunque indice dell'estremo disagio di cogliere in modo formale le complessità organizzate.

Secondo Forman e Gordon si considerano come elementi base della struttura le *macchie di ecosistemi* (*patches*) [16]. Essi descrivono le macchie in rapporto alla dimensione, alla forma, al tipo biotico, al numero, alla confi-

gurazione. Insieme di elementi di base formano mosaici caratteristici o ecomosaici, dove si possono distinguere aggregazioni di macchie configurabili ancora in macchie, oppure in corridoi, e in cui si possono distinguere le *matrici paesistiche*. Le matrici sono formate da quegli elementi che determinano il carattere prevalente dell'ecomosaico per estensione, connessione e funzione.

Il precedente modello deve comunque essere integrato da nuovi concetti [7], primo fra tutti quello di *ecotopo* inteso come elemento minimo che forma il paesaggio. L'ecotopo è l'elemento spaziale che presenta uniformità di composizione locale (es. caratteristiche fisiche, chimiche, morfologiche). Altro concetto che deve essere introdotto, deriva dall'osservazione che gli organismi non percepiscono la struttura del paesaggio nello stesso modo, quindi sembra più corretto riconoscere configurazioni complesse di elementi, cioè un ecomosaico pluridimensionale denominato *ecotessuto paesistico*. In genere quest'ultimo può essere pensato come costituito da un ecomosaico di base, quasi sempre individuabile dalle biocenosi¹ vegetali, e da una gerarchia di mosaici correlati.

4.1.3 Macchie paesistiche

Si cercherà ora di approfondire il concetto di *macchia* (patch) per la notevole importanza che assume nel presente lavoro.

In un paesaggio si possono individuare diversi tipi di macchie classificabili in rapporto alla loro origine [7]:

1. disturbo (cronico, ciclico, isolato)
2. reliquato (associato o no a disturbo)
3. risorsa ambientale (biotica o abiotica)
4. antropizzazione (diretta o indiretta)
5. colonizzazione (naturale o antropica)
6. evento transitorio (naturale o antropico)

¹Per biocenosi si intende la consociazione di diverse specie animali o vegetali in ambienti caratterizzati da determinate condizioni.

Le macchie nascono da trasformazioni nell'ambiente e a loro volta si trasformano nel tempo. In genere poi, le macchie che hanno diversa origine presentano anche diverse caratteristiche fisiche e biologiche.

In una macchia si può individuare una zona marginale e una zona interna: le due zone presentano in genere caratteristiche diverse (es. diversità nella biomassa e nella ricchezza di specie).

Aspetti importanti che caratterizzano una macchia possono essere l'estensione, la forma e la disposizione spaziale rispetto ad altre macchie. Saranno queste le caratteristiche particolarmente utilizzate in questo lavoro.

4.1.4 Concetti di spazio, tempo e scala

Si presentano ora alcuni chiarimenti e precisazioni sui concetti di scale spaziali e temporali.

Anche se gli ecotopi, che costituiscono le parti fondamentali del paesaggio, possono essere ritenuti invarianti con la scala, non lo sono la struttura, la funzione e i cambiamenti. È stato infatti dimostrato [17], che le misure sul numero, dimensione, forma delle *macchie* dipendono dall'estensione della finestra con cui si osserva il paesaggio. La variazione della funzione con la scala è evidente invece se si pensa al fatto che la percezione di paesaggio è diversa per i diversi organismi. Un classico esempio di questo è l'individuazione delle scale critiche nella studio della frammentazione del paesaggio [18], dove si evidenzia la diversa importanza e sensibilità che assumono i *patches* al variare della scala spaziale di indagine.

È poi importante precisare le caratteristiche dei dati attraverso i quali si vogliono estrarre informazioni per le diverse scale. Si possono distinguere la *grana* e l'*estensione*. La grana si riferisce al più piccolo livello di risoluzione spaziale o temporale dell'insieme dei dati. L'estensione si riferisce invece all'estensione dell'area di studio e alla durata dello studio. Ultimo aspetto da considerare in un approccio multiscalare è sicuramente l'integrazione delle informazioni relative a diverse scale, sia spaziali che temporali. Dal punto di vista teorico possiamo distinguere tre tipi di approcci: somma, indici e modelli. Nel nostro caso si cercherà di integrare le informazioni con lo sviluppo di un *modello* specifico dell'ambiente in esame che, oltre a garantire

la corretta integrazione tra scale spazio-temporali, consenta anche una più corretta descrizione dei fenomeni in gioco e che sia infine predittivo.

4.2 Le variabili di ecologia del paesaggio

Si esporranno ora i criteri per applicare le considerazioni teoriche espresse nelle precedenti sezioni, riconoscendo nei dati spaziali a disposizione le macchie che formano la struttura del paesaggio e poi analizzando forma, dimensione e disposizione spaziale di tali macchie.

Si vedranno anche le tecniche di analisi multiscalare e di analisi frattali che consentono di evidenziare in maniera più precisa i fenomeni che si manifestano alle diverse scale.

4.2.1 I dati che descrivono il paesaggio

Le informazioni utilizzate per caratterizzare le configurazioni ambientali del territorio sono sia quelle contenute nel SIAT, ma in particolare le informazioni contenute nella cartografia della vegetazione derivata dalla classificazione delle OFD tramite le procedure descritte in capitolo 5.

Si dispone quindi, oltre che delle informazioni sulle caratteristiche geomorfologiche del territorio ad una risoluzione di 10 metri, anche e soprattutto delle informazioni sulla copertura vegetale ad elevata risoluzione (1 m). Le classi evidenziate in quest'ultimo tematismo sono:

Classe 1 “Boschi”: è comprensiva di tutta la vegetazione ad alto fusto (abeti, pini, larice, faggio), e di particolari tipi di vegetazione a comportamento arboreo, tra quest'ultimi in particolare il pino mugo.

Classe 2 “Prati”: sono compresi tutti i tipi di vegetazioni erbacee.

Classe 3 “Altro”: ricadono in questa categoria il suolo nudo

(rocce affioranti, ghiaioni), le zone coperte da neve, strade asfaltate, case ed edifici.

Le informazioni contenute in questo tematismo non consentono una rappresentazione dettagliata delle diverse specie vegetali presenti sul territorio,

ma costituiscono ugualmente una fonte preziosa di informazioni sulla struttura del paesaggio. Come si vedrà infatti dettagliatamente nel capitolo 5, l'elevata risoluzione delle OFD ha consentito di evidenziare, attraverso la classificazione alcune informazioni indirette sul tipo di vegetazione presente. Tali informazioni sono legate principalmente alla diversa densità di piante evidenziabile dai tematismi prodotti. Un esempio della cartografia prodotta è riportato in figura 4.1 dove sono evidenziati i limiti dell'area campione "pellizzano.17".

4.2.2 Analisi dei patches

Per poter caratterizzare la struttura del paesaggio è innanzitutto necessario individuare dai dati a disposizione i diversi *patches* presenti. Si è già visto precedentemente cosa si intende dal punto di vista teorico per patches. Da tale concetto teorico si deve però passare alla definizione operativa per l'individuazione dei patches nel tematismo GIS prodotto dalla riclassificazione delle OFD.

Intenderemo in questo caso per patch tutte le celle adiacenti appartenenti ad una stessa classe. In figura 4.2 si è riportato un esempio del concetto; si devono pensare le celle che appartengono ad un patch come ad un unico oggetto. Esisteranno quindi patch costituiti da "bosco", da "prato" e da "altro".

Una volta individuati sull'immagine i diversi patches, è stato possibile definire su di essi un insieme di misure.

Le misure utilizzate si possono suddividere in due categorie principali:

1. misure per caratterizzare la forma del patch.
2. misure per caratterizzare le relazioni tra patches.

Le misure per descrivere la forma dei patches vengono presentate di seguito in questa sezione, mentre quelle per caratterizzare le relazioni tra patches si esporranno nella sezione successiva. Per alcune variabili si evidenzieranno anche degli aspetti pratici ed i limiti applicativi legati all'utilizzo di immagini raster.

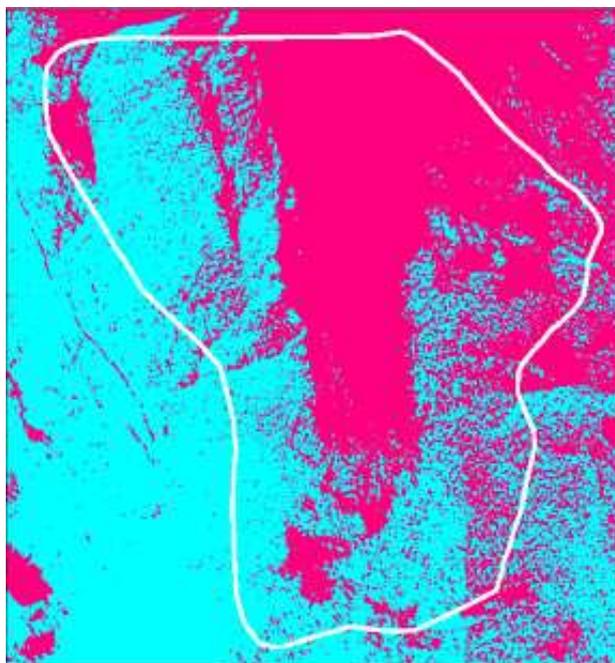


Figura 4.1: Esempio di cartografia tematica della copertura vegetale prodotta con la classificazione da foto aerea: in bianco si è riportata l'area campione "pellizzano.17".

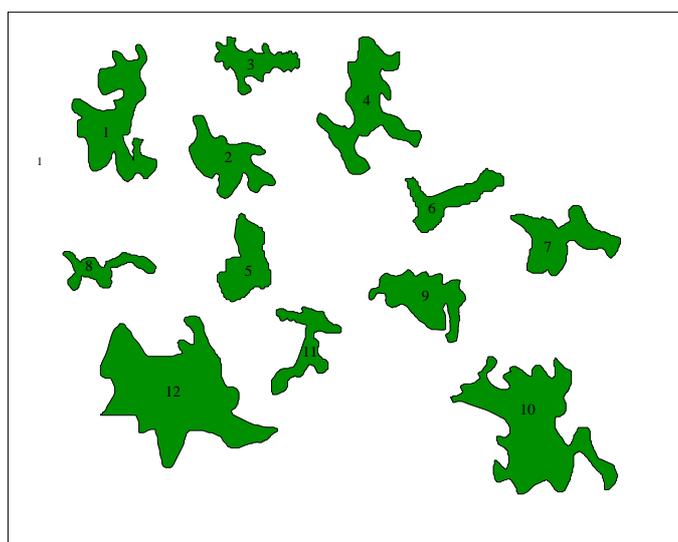


Figura 4.2: Esempio di configurazione spaziale di patches.

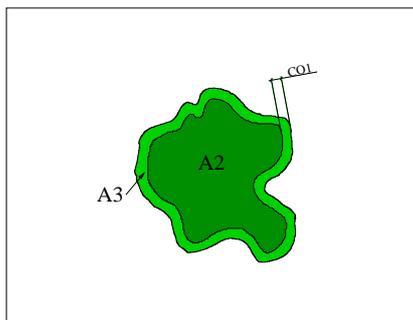


Figura 4.3: Superficie marginale e interna di un patch.

A1: superficie del patch in m^2 . Nel caso di immagini raster è sempre un multiplo della superficie s della singola cella.

A2 superficie interna del patch in m^2 . È la superficie residua del patch una volta tolta la fascia marginale di spessore col (figura 4.3).

A3 superficie di marginale del patch in m^2 . È equivalente a $A1 - A2$.

P: perimetro totale del patch in m . Equivale alla somma tra il perimetro esterno e l'eventuale perimetro interno del patch. Nei casi applicativi è multiplo di \sqrt{s} .

PA: rapporto tra perimetro (P) e superficie del patch:

$$PA = \frac{P}{A1}$$

È misurato in m^{-1} e a parità di forma varia con la dimensione del patch. I limiti teorici sono:

$+\infty$ per patch infinitamente allungati;

$2\sqrt{\frac{\pi}{A1}}$ per patch di forma circolare.

Dal punto di vista pratico i limiti sono invece diversi e legati alla risoluzione dell'immagine raster. In questo caso, oltre ad essere funzione della dimensione del patch, PA è anche funzione della dimensione delle celle che formano l'immagine. Il limite superiore è infatti $\frac{4}{\sqrt{s}}m^{-1}$ ed è

valido per celle disposte in diagonale, mentre il limite inferiore è minimizzato dal rapporto $\frac{4}{\sqrt{A}}m^{-1}$ e vale per patch di forma quadrata. In realtà anche il limite inferiore dipende da s .

PCA: è un coefficiente di forma ed è definito dall'espressione:

$$PCA = \frac{1}{2\sqrt{\pi}} \frac{P}{\sqrt{A}}$$

Dal punto di vista teorico non dipende dalla dimensione dei patches e gli estremi di variazione sono:

$$1.0 \leq PCA \leq \infty$$

dove l'estremo inferiore è valido per patches di forma circolare e l'estremo superiore per patches infinitamente allungati.

Come nel caso di PA , anche in questo i limiti pratici differiscono da quelli teorici. Per immagini raster infatti, l'intervallo di variazione è inferiormente limitato dal valore $\frac{4}{2\sqrt{\pi}} \approx 1.129$ e superiormente dal valore $\frac{2}{\sqrt{\pi}} \sqrt{\frac{A}{s}}$.

RCC: è un ulteriore coefficiente di forma e descrive la compattezza; è definito dalla relazione:

$$RCC = \frac{2(\frac{A}{\pi})^{1/2}}{d_{max}}$$

dove d_{max} equivale al diametro del cerchio minimo che circoscrive il patch.

Non dipende, almeno teoricamente, dalla dimensione del patch e varia tra 0.0 per forme infinitamente lunghe a 1.0 per forme circolari. Nelle applicazioni il limite superiore è $\sqrt{2} \frac{A}{\sqrt{s}}$ per pixel disposti lungo la diagonale.

Le variabili finora descritte servono semplicemente per caratterizzare i singoli patches che compongono il paesaggio. Si vedranno invece nel capitolo 6 le metodologie utilizzate per la combinazione delle informazioni dei singoli patches così da ottenere una descrizione completa del paesaggio alla scala di analisi.

4.3 Analisi delle scale

Si vogliono ora descrivere ed applicare delle tecniche che, utilizzando le informazioni contenute nella copertura vegetale di ciascuna area campione, consentono di individuare tutte le scale necessarie per descrivere correttamente il paesaggio.

Si è utilizzata in particolare, una tecnica di analisi multivariata per l'individuazione dei bordi (*edge detection methods*) [11]. La tecnica consiste nell'individuare nell'immagine alcuni transetti di analisi e suddividerli in sottofinestre adiacenti di lunghezza costante 4.4. La lunghezza della sottofine-

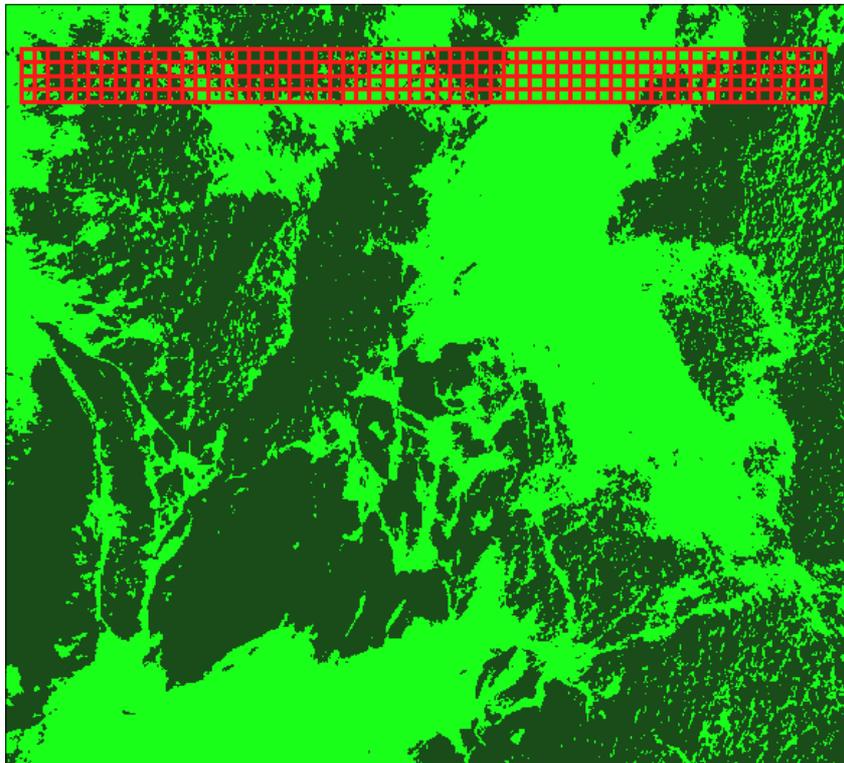


Figura 4.4: Transetti di analisi per l'individuazione delle scale nell'immagine.

stra equivale alla risoluzione con cui si individueranno i bordi. Per ciascuna sottofinestra si eseguiranno una serie di misurazioni; nel nostro caso le misure sono state:

1. Superficie boscata.

2. Perimetro totale dei patches di bosco.
3. Numero di patches bosco.
4. Numero dei patch di prato.

I risultati di tali misurazioni sono rappresentabili in uno spazio a quattro dimensioni. Per non attribuire importanza diversa tra le variabili, si è eseguita la standardizzazione delle stesse:

$$z_i = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}$$

dove $x_i, i = 1, \dots, 4$ sono le 4 variabili misurate con media campionaria \bar{x}_i e varianza $\sigma_{x_i}^2$.

Scegliendo una direzione di analisi del transetto, si è calcolata la distanza euclidea tra i vettori nello spazio delle misure per ciascuna sottofinestra e la sottofinestra adiacente successiva; i risultati sono riportati in figura 4.5. È rappresentata in tale figura la distanza euclidea in funzione della posizione nel transetto. Più il picco in questo grafico risulta elevato, maggiore è la discontinuità presente in quel punto, cioè maggiore è la differenza in termini delle variabili misurate tra due sottofinestre contigue. I picchi più elevati identificano perciò i bordi presenti nell'immagine, mentre i picchi più piccoli rappresentano il rumore.

Un grafico come quello di figura 4.5 non è facilmente interpretabile. Non è infatti possibile riconoscere direttamente le scale dominanti. Si è quindi proceduto ad una rielaborazione dei risultati individuando le distanze ds che intercorrono lungo il transetto tra due picchi successivi. Per non considerare in questa analisi anche il rumore si sono calcolate le distanze solo tra i picchi superiori ad una certa soglia. Se ora si calcola la densità di probabilità di ds si ottiene il grafico in figura 4.6, dove le dimensioni dominanti dei patches (50, 250, 350 m) sono bene evidenziate. Le scale di analisi che permettono di isolare i tre diverse tipologie di patches evidenziate sono invece: 150, 300, 400m. Il grafico in figura 4.6 si è ottenuto analizzando contemporaneamente le cartografie della vegetazione per 4 aree campione. Si utilizzeranno tali risultati per impostare correttamente le analisi multiscalare di struttura del paesaggio descritte nel capitolo 6.

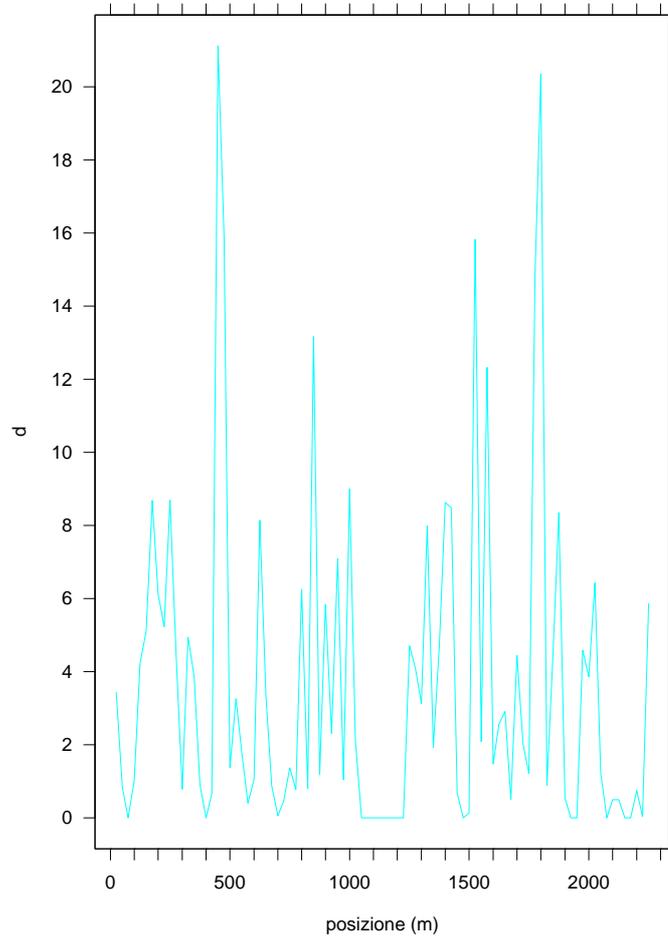


Figura 4.5: Variazione distanza euclidea lungo il transetto di analisi

4.3.1 Analisi delle distanze

Per poter caratterizzare anche gli aspetti relativi alle relazioni tra i diversi patches si sono utilizzate due tipi di tecniche:

- misurazioni delle distanze tra i patches.
- calcoli di dimensione frattale.

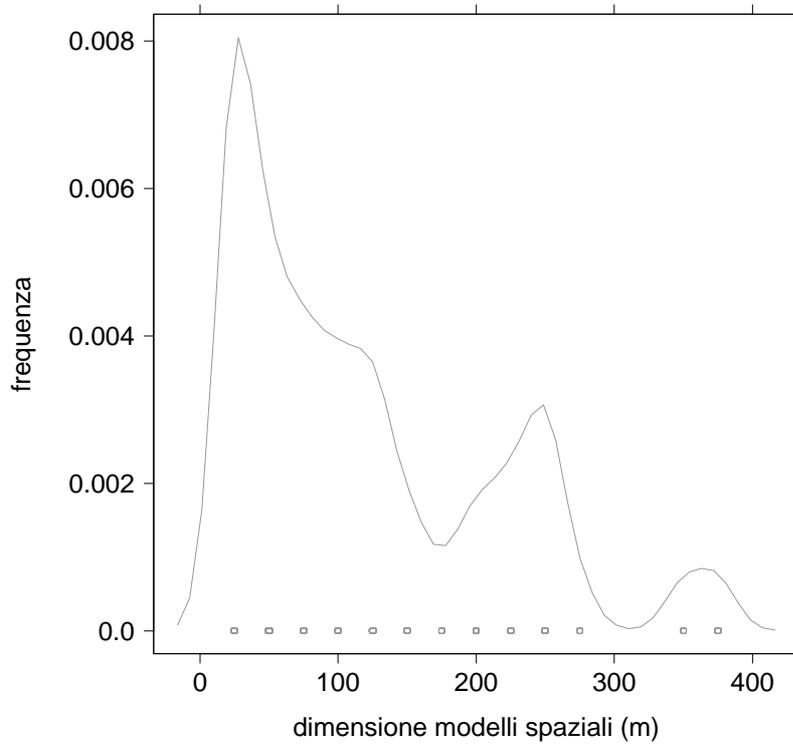


Figura 4.6: Distribuzione di probabilità dei modelli spaziali individuati in funzione della loro dimensione

Le problematiche relative all'utilizzo delle dimensioni frattali saranno viste successivamente, mentre ora si presentano gli aspetti sul calcolo delle distanze.

Per ciascun patches si è calcolata la distanza dal più vicino patch della stessa categoria, questo vuole rappresentare in qualche modo il grado di isolamento in cui si trova il patch rispetto agli altri. I modi possibili per calcolare le distanze tra patches sono:

1. distanza centro-centro;
2. distanza bordo-bordo;
3. distanza bordo-centro;

4. distanza centro-bordo.

Ciascuna di queste distanze descrive meglio particolari aspetti della struttura. A seconda di ciò che si vuole misurare si utilizzerà la misura più opportuna. Se si pensa per esempio che le zone interne del patch costituiscono il carattere più importante da studiare è in generale più corretto utilizzare la distanza centro-centro.

Nel nostro caso invece, si è interessati a caratterizzare le zone di margine, verrà a riguardo utilizzata la misura bordo-bordo tra un patches ed il suo più vicino appartenente alla stessa classe. In figura 4.7 è evidenziata una configurazione spaziale di patches e sono mostrate le distanze bordo-bordo tra ciascun patches e il suo più vicino. Si osservi che la relazione di vicinanza non è biunivoca quindi se il patch A è il più vicino a B , non è in generale vero il contrario.

4.3.2 Modelli frattali e configurazioni spaziali

Una grosso contributo allo studio delle configurazioni spaziali del paesaggio si è avuto con la teoria dei frattali.

I fenomeni che si verificano nel paesaggio, come moltissimi fenomeni naturali presentano dipendenza di scala e necessitano, per essere descritti, di strumenti adeguati quale appunto la teoria dei frattali. Nel caratterizzare la struttura del paesaggio si ricorrerà quindi al calcolo della dimensione frattale di box counting della superficie boscata. La definizione di dimensione di box counting è riportata in appendice A.1. Mentre in figura 4.8 è riportato lo schema utilizzato per osservare il paesaggio alle diverse scale e necessario per calcolare la dimensione di box-counting.

Per il calcolo della dimensione frattale si devono escludere le scale confrontabili con la risoluzione dell'immagine e le scale paragonabili alla dimensione della finestra di calcolo.

L'utilizzo della dimensione frattale dovrebbe portare ad evidenziare particolari fenomeni che non possono essere descritti da altri indicatori ambientali. In particolare dovrebbe descrivere la complessità della superficie boscata, il grado di dispersione dei patches e la presenza di strutture che si ripetono a scale diverse.

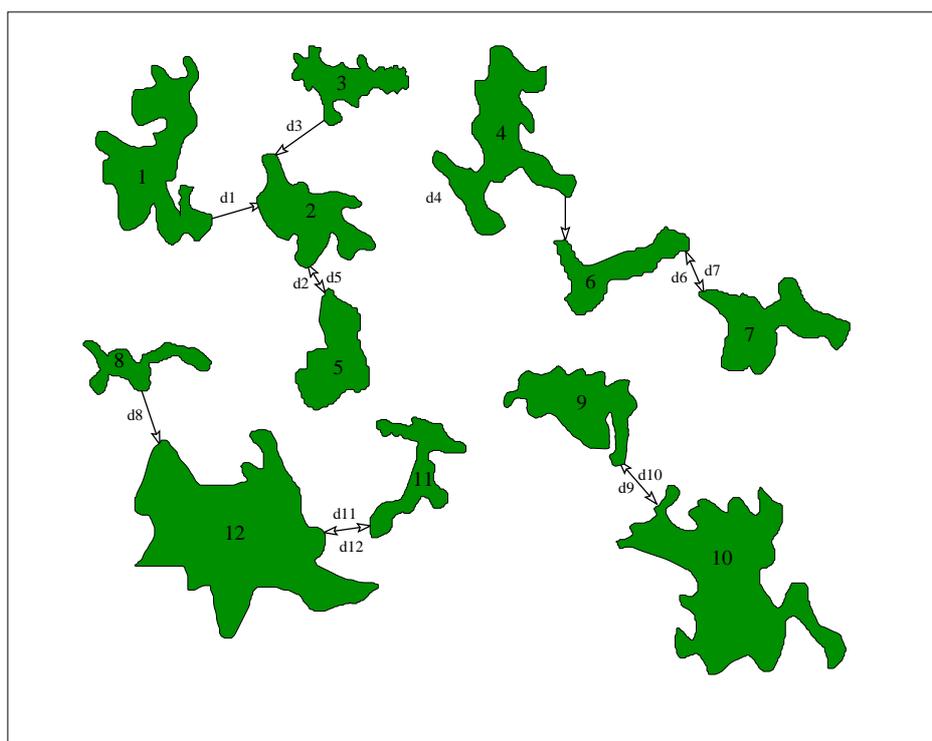


Figura 4.7: Relazioni di distanza “bordo” - “bordo vicino” tra patches appartenenti alla stessa classe.

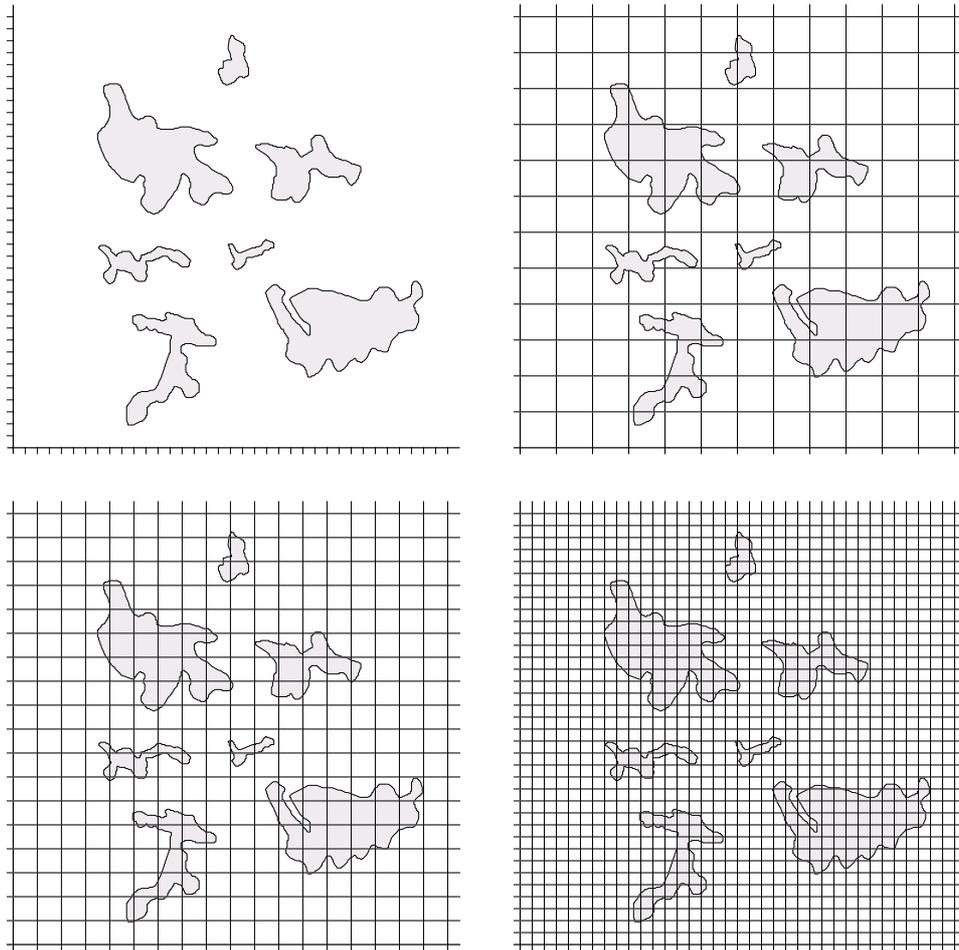


Figura 4.8: Esempio di calcolo della dimensione bi box-counting

4.4 Strumenti informatici di calcolo

Per l'individuazione dei diversi patches nel tematismo raster ed il calcolo degli indici relativi si sono utilizzati i moduli `r.le programs` disponibili dalla versione 4.1 di GRASS. Sono un insieme di programmi per l'analisi quantitativa della struttura del paesaggio [19]. Le informazioni sulla struttura del paesaggio sono ottenute sovrapponendo all'immagine un insieme di finestre di analisi (*sampling areas*) e le misure di struttura sono calcolate separatamente per ciascuna di queste aree. La possibilità di controllare la forma, il numero e la distribuzione delle aree di analisi all'interno dell'immagine consente anche

analisi di struttura di tipo multiscalare

Dell'intero gruppo di moduli si sono in particolare utilizzati i seguenti programmi:

r.le.setup: permette di selezionare i vari parametri, in particolare le *sampling areas*, che vengono utilizzati dagli altri programmi **r.le**.

r.le.patch: individua i *patches* presenti in ciascuna finestra di analisi e ne calcola gli attributi: area, perimetro, coefficienti di forma, area interna e marginale. I tipi di output del programma sono:

- misure individuali per ciascun *patch* di tutte le *sampling areas*;
- indicatori aggregati per *sampling areas* (media e varianza);
- tematismo raster con evidenziati i diversi *patches* individuati nell'immagine.

r.le.dist: individua nell'immagine i diversi *patches* presenti e opera diversi tipi di misure di distanza tra *patches*. I risultati possono essere riportate in diversi modi similmente a quanto visto per il programma **r.le.patch**

Per il calcolo della dimensione frattale si è invece utilizzato il programma **r.boxcount**. È un modulo per GRASS che permette di calcolare la *dimensione di box counting* delle celle con categoria diversa da 0 in regioni quadrate. La definizione di dimensione di box counting è riportata in appendice A.1.

Per potenziare, velocizzare e automatizzare le operazioni di analisi del paesaggio si sono sviluppati appositamente dei moduli in linguaggio Perl. Questi moduli hanno permesso da un lato, di combinare i diversi programmi per l'analisi del paesaggio visti in precedenza, e dall'altro di collegare il flusso di dati tra ambiente GRASS e ambiente statistico S-PLUS. L'utilizzo di S-PLUS consente l'analisi più complessa della struttura alle diverse scale, e permette anche di eseguire un'analisi multiscalare.

I moduli sviluppati in linguaggio Perl sono riportati di seguito.

landscape.pl: è il programma principale che consente di analizzare la struttura del paesaggio nell'intorno di particolari siti di interesse. Come da-

ti di input utilizzano il file contenente i siti da analizzare e la mappa tematica sulla quale fare l'analisi; consente anche di selezionare le dimensioni delle scale spaziali di analisi. Il programma svolge essenzialmente la seguente procedura:

1. individuare all'interno della mappa il numero e la posizione dei siti contenuti nel file di input (i siti che cadono fuori dalla mappa non vengono analizzati).
2. generare direttamente i file contenenti i parametri per i programmi `r.le.patch` ed `r.le.dist` senza utilizzare in modulo specifico `r.le.setup`².
3. analizzare il paesaggio per le diverse scale in prossimità di ogni sito utilizzando in sequenza i moduli GRASS `r.le.patch`, `r.le.dist` e `r.boxcount`.
4. generare dei file riassuntivi sulle operazioni eseguite.

`dist.convert.pl`: converte i file generati dal programma `r.le.dist` eseguito all'interno di `landscape.pl` per renderli compatibile per S-PLUS.

`patch.convert.pl`: è analogo al programma `dist.convert.pl` per convertire i file di output di `r.le.patch`.

`fractal.convert.pl`: converte invece file di output del programma `r.boxcount` sempre per renderli compatibili per S-PLUS.

Si vedrà nel capitolo 6 l'utilizzo di questi moduli per l'analisi della struttura del paesaggio nelle zone di presenza di covate di fagiano di monte.

²Il modulo `r.le.setup` è eseguibile solo in modo interattivo e quindi non è utilizzabile nell'ambito di una procedura automatica

Capitolo 5

Classificazione di ortofoto digitali

Il presente capitolo è dedicato allo sviluppo e all'applicazione di procedure automatiche per la classificazione delle Ortofoto Digitali del SIAT, descritte in sezione 1.5.

Si vuole garantire con tali procedure il conseguimento dei seguenti obiettivi principali:

- disporre di una procedura automatica applicabile in maniera sistematica, senza eccessivo controllo da parte dell'operatore;
- generare per le aree campione una cartografia contenente le informazioni di carattere ambientale e utilizzabile per il calcolo degli indicatori di ecologia del paesaggio;
- possibilità di utilizzare la procedura anche per la classificazione di immagini di tipo diverso, quali OFD a colori ed immagini SAR o per altri dati multidimensionali;
- garantire la facilità di aggiornamento della cartografia prodotta.

In appendice si descrivono i riferimenti teorici degli algoritmi e delle procedure utilizzate per la classificazione di immagini.

5.1 Classificazione supervisionata

Si presentano innanzitutto i risultati e le applicazioni di classificazione supervisionata (*supervised classification*). I riferimenti teorici sono riportati in appendice B.

Si descrivono innanzitutto i passi essenziali per una classificazione supervisionata applicata alle immagini sono i seguenti:

1. Decidere i tipi di copertura del territorio nei quali l'immagine sarà classificata. Queste sono le classi di informazione e nel nostro caso, vista la caratteristica delle immagini, si sono scelte le categorie "bosco", "prato" ed "altro".
2. Scegliere degli esempi rappresentativi per ognuna delle classi desiderate. Questi formeranno i *training data* (dati di addestramento). L'insieme di training per ogni classe è stato ottenuto da fotointerpretazione non automatica di alcune zone delle immagini da classificare.
3. Utilizzare i *training data* per stimare i parametri del particolare algoritmo di classificazione utilizzato; questi parametri possono essere le proprietà del modello probabilistico usato, o le equazioni di ripartizione che definiscono appunto la partizione nello spazio delle misure. L'insieme di parametri per una data classe è a volte denominato *signature* della classe.
4. Utilizzare gli algoritmi di classificazione "addestrati" con l'insieme di *training*, per classificare ogni pixel dell'immagine in una delle coperture del terreno scelte.
5. Produrre quindi le cartografie e le mappe che riassumono il risultato della classificazione.

Dal punto di vista statistico le immagini si presentano come un reticolo regolare di celle con associata una variabile casuale discreta, cioè i livelli di luminosità. In particolare queste variabili casuali sono spazialmente correlate.

In riferimento a questa osservazione sono presenti nella classificazione di immagini due approcci fondamentali:

- classificare ciascun pixel dell'immagine utilizzando soltanto le informazioni sugli attributi (appendice B.3);
- segmentare l'immagine individuando le zone a comportamento spettrale uniforme (appendice B.4).

5.1.1 Classificatore alla massima verosimiglianza

Si sono innanzitutto provati dei classificatori che non considerano il contesto spaziale dei pixel nell'immagine, ma operano la decisione soltanto in base ai livelli di luminosità. Per le OFD che sono immagini in bianco e nero quindi lo spazio delle misure è monodimensionale ed è il livello di grigio che assume il pixel.

Il classificatore più semplice da utilizzare è quello della *massima verosimiglianza* sotto l'ipotesi che le distribuzioni di probabilità delle classi nello spazio delle misure siano gaussiane multidimensionali (equazione B.10) (le distribuzioni di probabilità delle immagini in bianco e nero sono un caso particolare). Per la stima dei parametri a partire dai dati di training si è a sua volta applicata la regola della massima verosimiglianza (equazioni D.1 D.7).

All'interno di GRASS è implementata, a riguardo, una procedura a due passi che utilizza i seguenti moduli:

i. **gensig**: algoritmo che calcola, dai dati di addestramento, i parametri per le gaussiane multidimensionali delle diverse classi, scrivendoli poi in un file (*signature file*). I dati di training sono passati all'algoritmo attraverso un mappa raster

denominata *training map* che contiene alcuni pixel o regioni di pixel già classificati. Si osservi che per creare la mappa di training sono stati selezionati manualmente nell'immagine da classificare esempi di zone di diverse classi.

`i.maxlik`: algoritmo che opera la classificazione, utilizzando il classificatore della massima verosimiglianza e le informazioni contenute nel *signature file* precedentemente generato.

In figura 5.1 è riportato il risultato della classificazione utilizzando la procedura appena descritta.

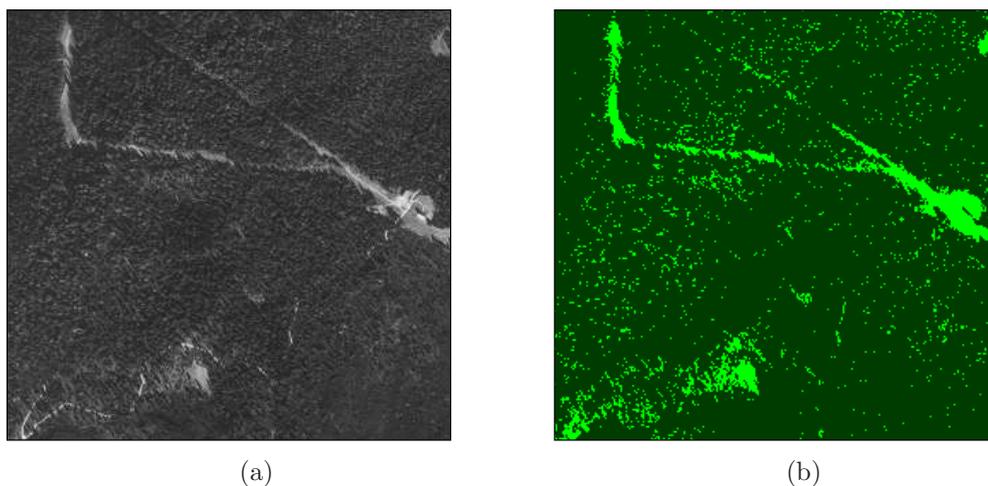


Figura 5.1: Risultato procedura di classificazione gaussiana e massima verosimiglianza: foto aerea originaria (a); immagine classificata (b).

Un miglioramento di tale procedura di classificazione è possibile utilizzando una diversa regressione per i dati di addestramento nel calcolo delle probabilità $p(x | C_i)$. Invece della distribuzione gaussiana, si è applicato il metodo semiparametrico a mistura di gaussiane. A riguardo in GRASS sono implementati, nel modulo `i.gensigset`, gli algoritmi *MDL* e *EM* ([20],[5]) per il calcolo dei

parametri della mistura di gaussiane per ciascuna delle classi con le quali sono etichettati i dati di addestramento. In figura 5.2 è riportato il risultato della classificazione di OFD.

5.1.2 Classificatore SMAP

Per cercare di migliorare i risultati ottenuti con il classificatore alla massima verosimiglianza si sono eseguite delle prove utilizzando il *massimo sequenziale a posteriori* (SMAP). Il modello di classificazione SMAP, descritto

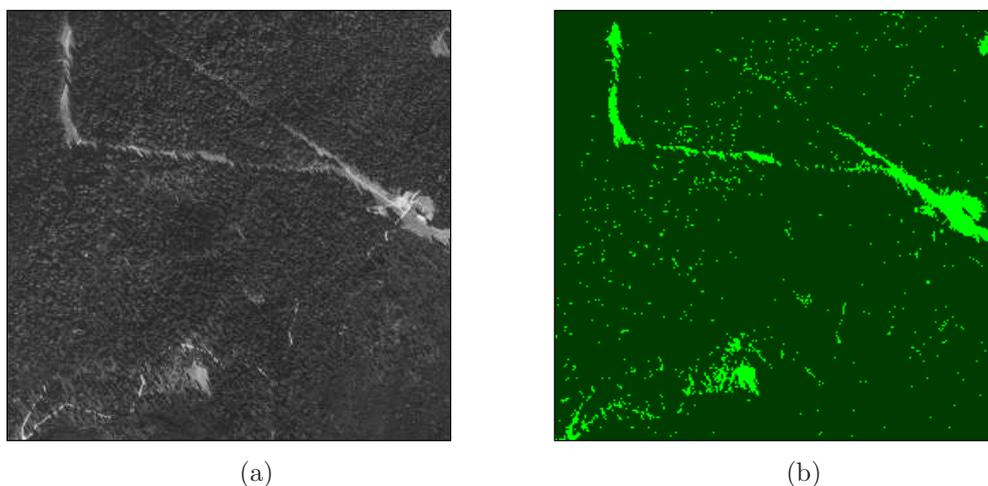


Figura 5.2: Risultato della procedura di classificazione mistura di gaussiane e massima verosimiglianza: foto aerea originaria (a); immagine classificata (b).

in appendice C), consente di segmentare l'immagine tenendo conto del contesto spaziale dei singoli pixel. La dipendenza è ottenuta attraverso un modello a campi casuali multiscalarari (MSRF Multi Scale Random Field).

Il classificatore *SMAP* è implementato in GRASS nell'algoritmo `i.smap`. In questo caso, quindi, la procedura a due passi utilizzata è stata:

1. `i.gensigset`: modulo che calcola, dai dati di addestramento, i parametri per le misture di gaussiane multidimensionali delle diverse classi, scrivendoli in un file (signature file).
2. `i.smap`: modulo che opera la classificazione secondo l'algoritmo *SMAP*, utilizzando il *signature file* precedentemente generato.

In figura 5.3 è riportato il risultato dell'utilizzo della procedura `i.gensigset`, `i.smap` applicata alla OFD.

5.1.3 Confronto dei risultati

È possibile confrontare in figura 5.4 i risultati delle tre procedure di classificazione descritte nelle sezioni precedenti. Si osservi che con l'utilizzo del

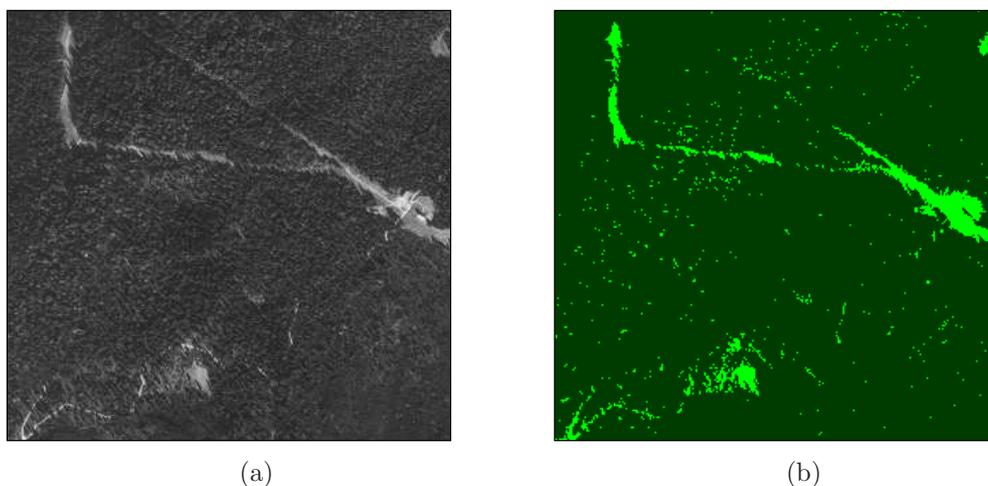


Figura 5.3: Risultato procedura di classificazione utilizzando l’algoritmo SMAP: foto aerea originaria (a); immagine classificata (b).

classificatore SMAP (figura 5.4c) vengono ridotte notevolmente le singole radure presenti all’interno della superficie boscata e anche che l’interfaccia tra le due classi “bosco” e “prato” si presentano più lisciate. Questo risultato è legato al fatto che con l’algoritmo SMAP si introduce una correlazione spaziale tra i diversi pixel. Si osservi inoltre che la stima dei parametri per le funzioni di trasferimento tra le diverse scale (appendice C è fatta localmente, quindi le eventuali classi che localmente presentano un’elevata probabilità a priori rispetto alle altre saranno predilette dal classificatore.

Confrontando invece tra loro i risultati delle altre due procedure di classificazione (figura 5.4b e figura 5.4c) emerge che l’utilizzo di un algoritmo migliore per la regressione dei dati di training (`i.gensigset` e mistura di gaussiane) porta a risultati peggiori rispetto all’algoritmo `i.gensig` (funzione di regressione gaussiana), questo a parità di classificatore utilizzato (`i.maxlik`). La spiegazione di questo non può che risiedere in una scelta non appropriata dei dati di training. Infatti quello che si vuole ottenere è il riconoscimento del prato, distinguendo anche le piccole radure che si trovano all’interno della superficie boscata, per poter riconoscere anche regioni con densità di piante arboree differenti. Sotto queste ipotesi è quindi tecnicamente impossibile individuare correttamente dei dati di training.

Vedremo come questo problema può essere risolto utilizzando una

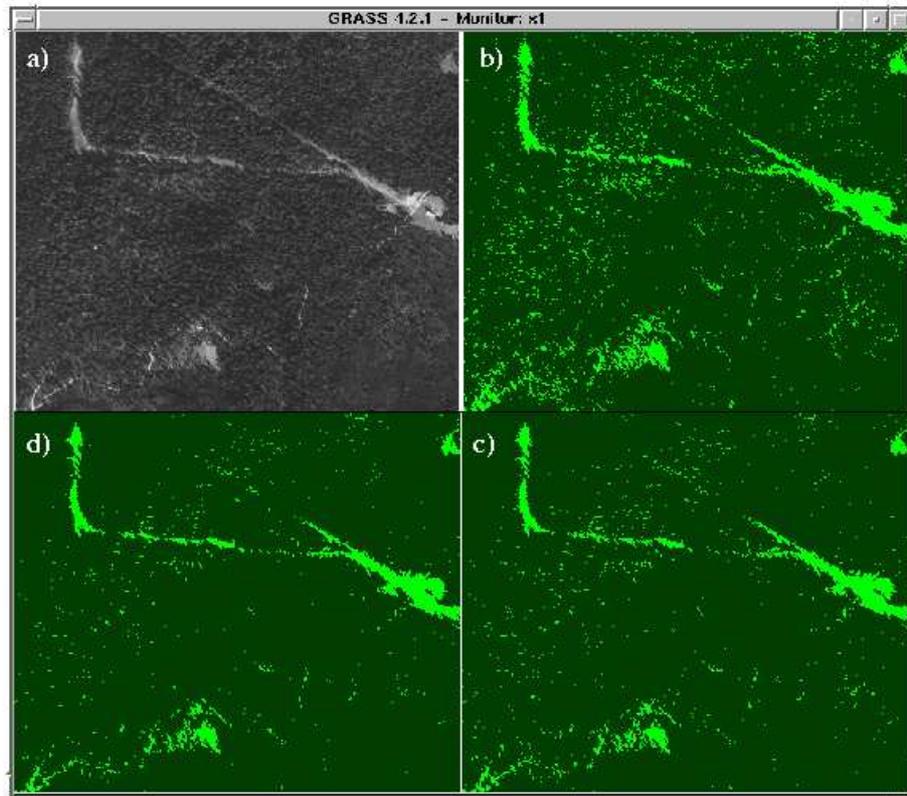


Figura 5.4: Confronto risultati procedure di classificazione: foto aerea originaria (a); risultato regressione gaussiana e massima verosimiglianza (b); risultato mistura di gaussiane e massima verosimiglianza (c); risultato mistura di gaussiane e algoritmo SMAP (d).

procedura non supervisionata per la selezione automatica delle diverse classi.

5.2 Classificazione non supervisionata

Viste le difficoltà di estrarre correttamente i dati di training si è pensato di utilizzare una procedura non supervisionata che consenta di individuare il probabile numero di classi presenti e le caratteristiche spettrali delle classi soltanto in base ai dati. Come descritto in appendice E, gli algoritmi di clustering richiedono un numero notevole di operazioni e di tempo, quindi non è sempre possibile applicarli direttamente all'intera immagine da classificare.

Si tenga presente che il tipo di immagini del data base OFD sono ad alta risoluzione (1 m) e per analizzare la vegetazione di alcune aree campione si sono utilizzate sotto immagini di 3000×3000 pixel.

Si deve invece calcolare il clustering su un sotto insieme dei dati (training set). La scelta del training set costituisce un punto importante per una corretta riuscita della procedura. Si dovranno comprendere all'interno del training set dati che appartengono verosimilmente a tutte le classi, ed in particolare dati localizzati in zone di discontinuità dell'immagine (es. bordo tra bosco e prato). Questo perché, le zone di interfaccia hanno molte volte comportamento spettrale distinto dalle altre classi. È in queste zone che la classificazione presenta le maggiori difficoltà ed è importante che tali dati siano utilizzati nell'algoritmo di clustering.

Per l'individuazione del numero di classi spettralmente distinte e per determinare le loro caratteristiche spettrali, si è utilizzato il modello a mistura di gaussiane .

Si è già visto l'utilizzo di un modello a mistura di gaussiane nell'ambito di una procedura supervisionata che permette di avere una migliore regressione dei dati di training D.2.1. Ora invece lo stesso modello a mistura di gaussiane verrà utilizzato nell'ambito di una procedura non supervisionata per estrarre dal training set il numero e le caratteristiche delle classi spettralmente distinte.

Si è scelto il modello a mistura di gaussiane invece di un algoritmo di clustering classico¹ poiché non è necessario stabilire a priori il numero di classi, ma questo viene determinato utilizzando l'algoritmo MDL [20]. L'applicazione del programma GRASS `i.gensigset` in figura 5.5 esemplifica il suo utilizzo nell'ambito della procedura non supervisionata evidenziando le gaussiane della mistura che individuano anche le diverse classi. Successivamente all'individuazione delle diverse classi si è applicato il classificatore SMAP. In figura 5.6 è riportato un esempio del risultato di questa procedura per le classi non ancora etichettate.

Poiché i moduli di GRASS non prevedevano l'utilizzo combinato di questi algoritmi, è stato sviluppato il programma in linguaggio Perl `classifier.pl`

¹In GRASS è presente il modulo `i.cluster` che implementa l'algoritmo di clustering *migrating means*.

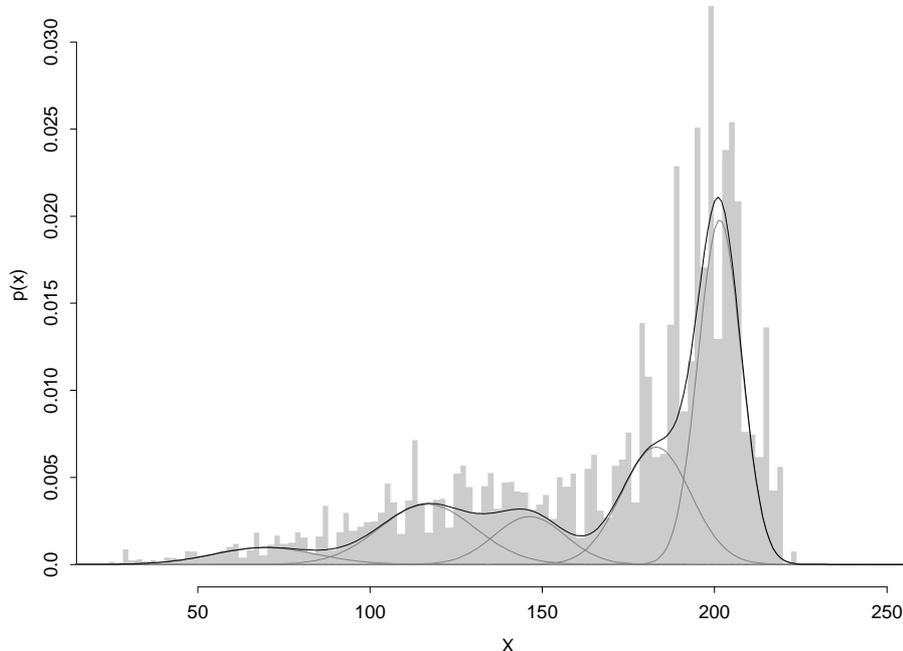


Figura 5.5: Individuazione del numero di classi e delle caratteristiche spettrali attraverso `i.gensigset`.

per interfacciare correttamente i diversi moduli GRASS creando così una procedura di classificazione automatica.

Si osservi che il classificatore SMAP in questo caso non ha gli inconvenienti riscontrati in precedenza, anzi garantisce una migliore classificazione rispetto all'utilizzo della massima verosimiglianza, apportando un miglioramento effettivo all'immagine classificata dovuto ad una migliore definizione dei particolari 5.6. Il risultato sembra dovuto al fatto che il modello considera ora le diverse classi separatamente e non come sottoclassi, questo aspetto influisce direttamente sulla stima dei parametri evitando i problemi di lisciamento tra le superfici di separazione delle diverse classi.

Per concludere la procedura di classificazione si devono etichettare le diverse classi della nuova immagine prodotta. Con l'utilizzo di `classifier.pl` alcune classi saranno aggregate assieme per formare l'immagine finale com-

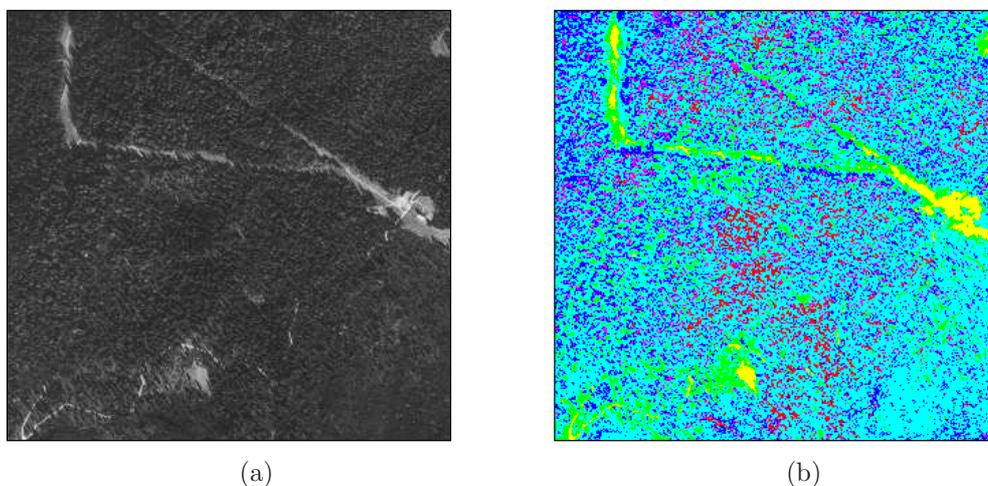


Figura 5.6: Risultati procedura non supervisionata: foto aerea originaria (a); immagine classificata (b).

posta dalle tre classi definitive “bosco”, “prato” e “altro”. In figura 5.7 è riportato il risultato finale di questa procedura di classificazione ed è possibile notare come in questo caso, rispetto agli esempi esposti in figura 5.4, siano state classificate più correttamente le zone boscate, riuscendo ad evidenziare quelle a diversa densità.

Per tentare di estrarre ulteriori informazioni dalle immagini, con lo scopo di distinguere tra i diversi tipi di copertura vegetale, si sono provate anche tecniche diverse di classificazione contestuale, introducendo in particolare nuovi canali contenenti informazioni sulla tessitura dell’immagine (varianza e diversità). L’utilizzo di questa tecnica ha portato ad individuare parzialmente le zone a mugheto², ma la procedura è risultata instabile e non generalizzabile ad immagini diverse. Non è quindi stata applicata nelle successive elaborazioni.

In riferimento a quest’ultimo problema, si deve anche ricordare come l’obiettivo sia l’analisi spaziale tramite indicatori di ecologia del paesaggio. È preferibile non utilizzare classificatori troppo complessi per tentare di distinguere i diversi tipi di copertura dal loro comportamento spaziale, poiché della struttura spaziale si terrà conto nel modello finale.

²Per lo studio dell’habitat del fagiano di la presenza di pino mugo assume grande importanza biologica.

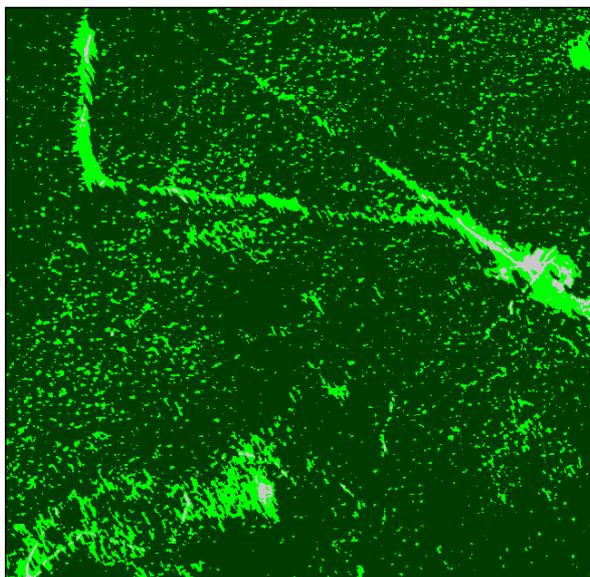


Figura 5.7: Risultato finale procedura non supervisionata.

In conclusione è stata costruita una procedura implementata nel modulo software `classifier.pl` in grado di rilevare in maniera sistematica per le OFD della provincia di Trento le seguenti categorie:

Cat.1 “Boschi”: è comprensiva di tutta la vegetazione ad alto fusto (abeti, pini, larice, faggio), e di particolari tipi di vegetazione a comportamento arbustivi, tra quest’ultimi in particolare il *pino mugo*.

Cat.2 “Prati”: sono compresi tutti i tipi di vegetazioni erbacee.

Cat.3 “Altro”: ricadono in questa categoria il suolo nudo (rocce affioranti, ghiaioni), le zone coperte da neve, strade asfaltate, case ed edifici.

Le procedure sono sicuramente applicabili anche alle ortofoto digitali a colori, per le quali peraltro sono stati eseguiti degli esempi di prova. Per la classificazione di immagini da satellite, in particolare di immagini SAR, le procedure saranno ancora valide, ma non si esclude siano necessari ulteriori adattamenti.

5.3 Applicazione alla classificazione delle aree campione

La procedura di classificazione descritta nella sezione precedente è stata utilizzata per la produzione della mappa della vegetazione per le zone interessate dalla presenza del fagiano di monte.

Non sono state analizzate tutte le aree campione, ma soltanto un sotto campione di dieci, che comunque, come risulterà evidente nel capitolo dedicato all'elaborazione statistica dei dati, sono significative del totale.

La tabella 5.1 riporta i codici utilizzati per indicare le dieci aree campione analizzate; la terza colonna indica l'anno di rilevamento della data area campione.

codice	denominazione	anno
AC1	spinale.1	1998
AC2	commezzadura.16	1999
AC3	pellizzano.17	1998
AC4	terlago.9	1998
AC5	strigno	1998
AC6	moena.1	1998
AC7	tesero.2	1999
AC8	pievetesino	1999
AC9	casteltesino.23	1999
AC10	primiero	1999

Tabella 5.1: Codici aree campione

In tabella 5.2 sono invece riportati, in modo sintetico, i risultati della classificazione; vengono indicate in ettari (ha) le superfici occupate dalle varie categorie per ciascuna area campione. Gli stessi risultati si sono riportati anche in forma di grafico 5.8 ordinando le aree campione a seconda della dimensione.

Un secondo grafico che riassume i risultati di classificazione è mostrato in figura 5.9 dove si mostra per ciascuna area campione la percentuale di superficie boscata.

	AC1	AC2	AC3	AC4	AC5	AC6	AC7	AC8	AC9	AC10
sup. bosco (ha)	250	262	166	217	156	73	19	63	39	21
sup. prato (ha)	479	106	157	77	76	119	84	42	10	23
sup. altro (ha)	6	-	-	3	7	1	3	-	-	-
tot	734	368	323	297	240	193	106	105	49	44

Tabella 5.2: Classificazione delle aree campione: superfici in ettari (ha) occupate dalle diverse categorie.

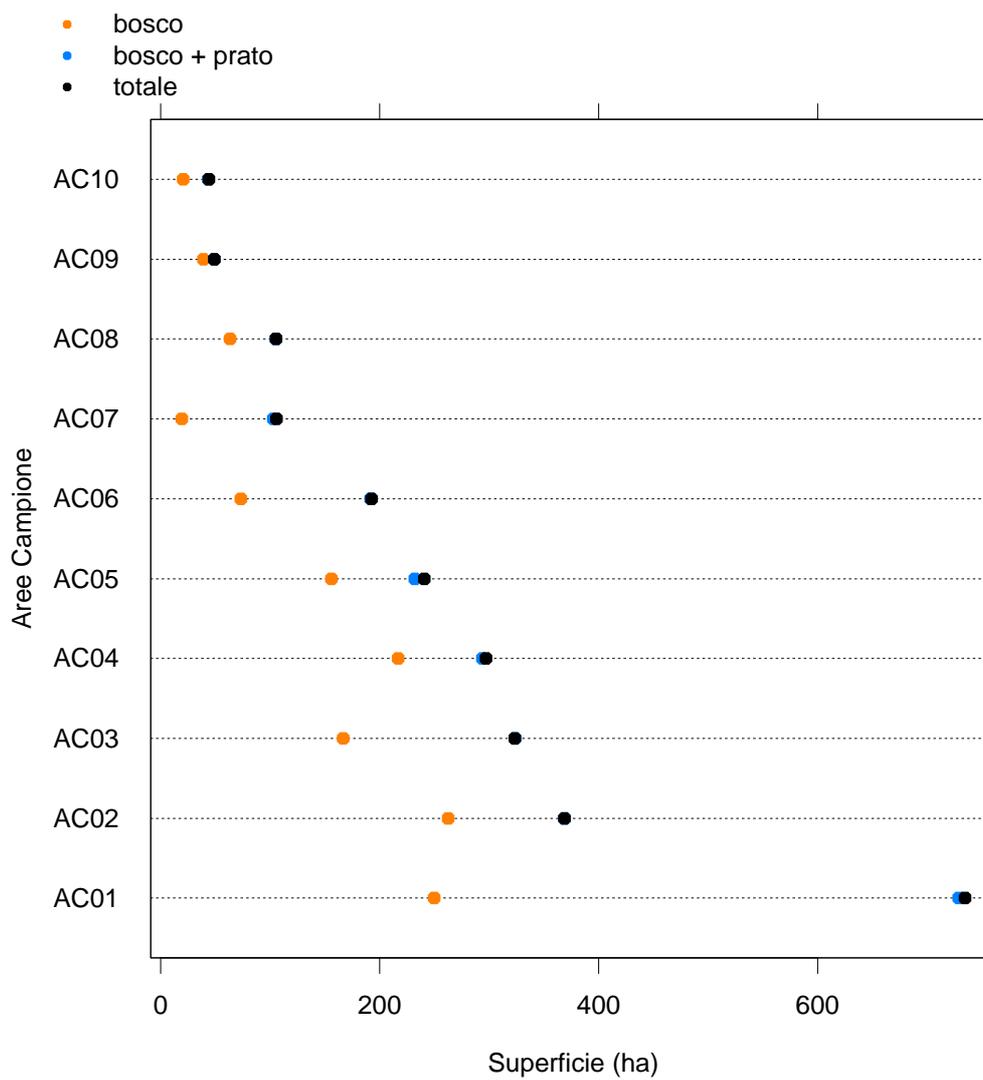


Figura 5.8: Classificazione delle aree campione: superfici delle categorie per area campione.

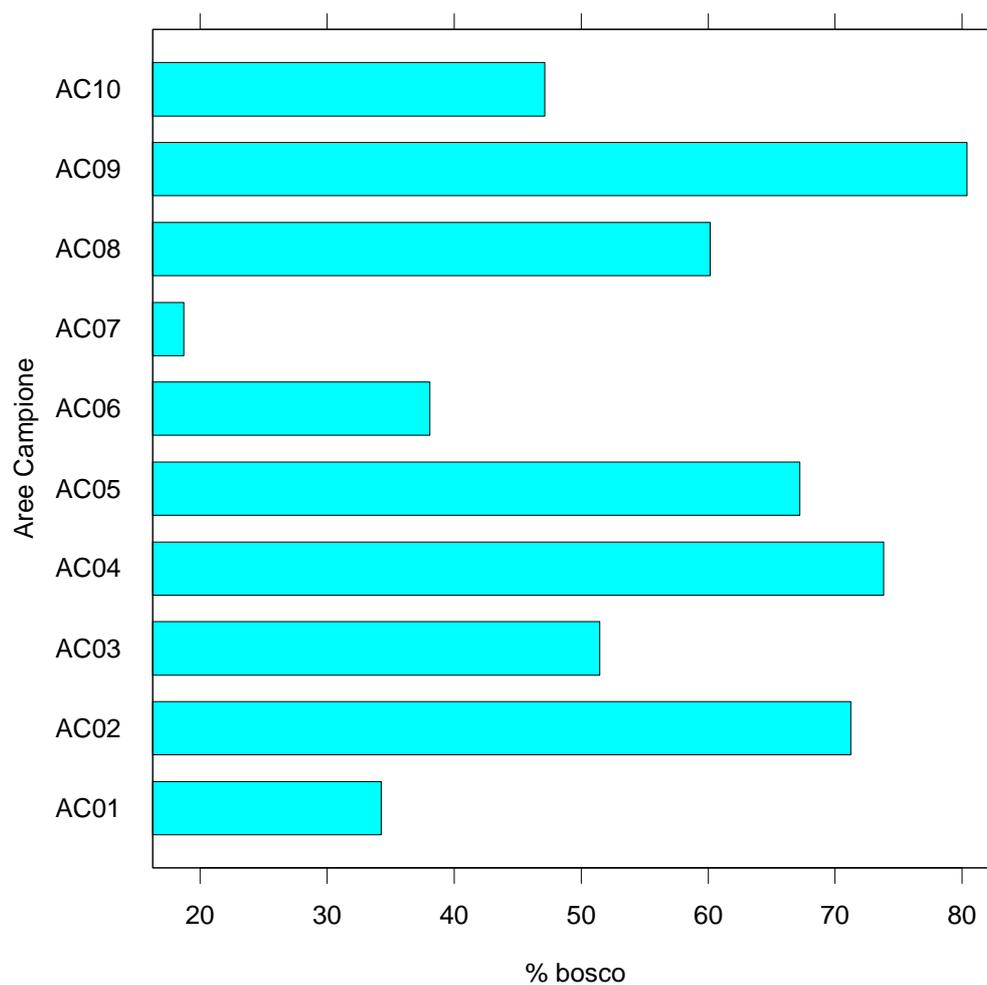


Figura 5.9: Classificazione delle aree campione: percentuale di superficie boscata.

Capitolo 6

Modelli per la valutazione dei miglioramenti ambientali

Gli obiettivi fondamentali di questo capitolo sono:

- individuare le variabili ambientali più importanti che influenzano lo svolgimento delle fasi biologiche più delicate del fagiano di monte.
- sviluppo di un modello predittivo finalizzato allo studio dell'habitat del fagiano di monte.

Si sono trattati in particolare i seguenti aspetti:

- individuazione dei fenomeni spaziali da studiare;
- descrizione dei fenomeni attraverso variabili di ecologia del paesaggio e costruzione del data base per i modelli di miglioramento;
- sviluppo di modelli ad albero binario di classificazione (recursive partitioning) sui dati di training;
- selezione del modello migliore ed analisi dei risultati.

6.1 La costruzione del data base per i modelli di miglioramento ambientale

6.1.1 I siti di contatto delle covate di fagiano di monte

Ci si è concentrati sulla caratterizzazione delle zone interessate dalla presenza di covate di fagiano di monte poichè, come descritto nel capitolo 2, le aree di allevamento dei pulli costituiscono delle zone ad elevata importanza biologica per la specie.

La distribuzione del numero di covate per le aree campione inserite nel data base informatizzato nei de anni 1998 e 1999 è visualizzata in figura 6.1

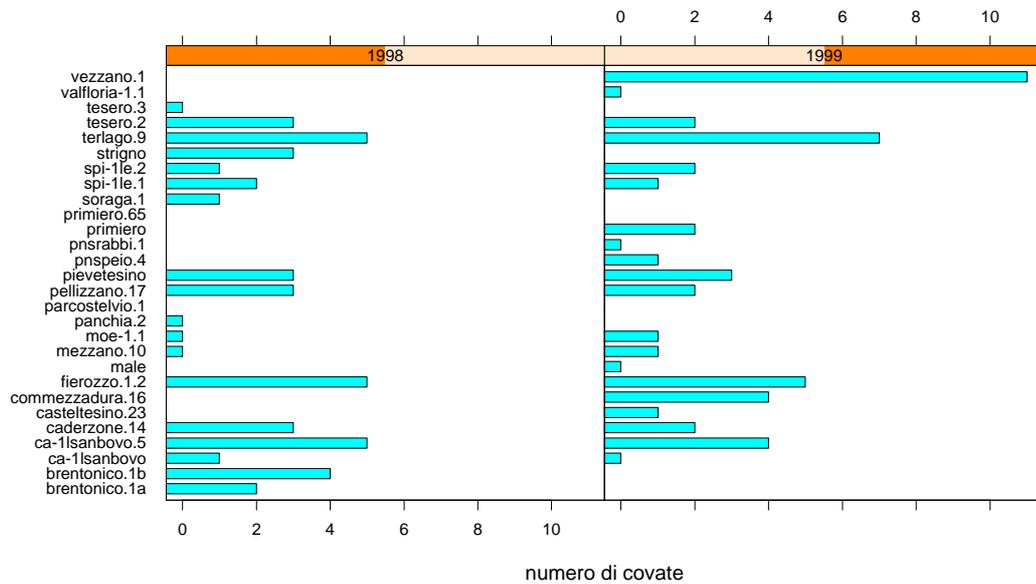


Figura 6.1: Numero di covate per area campione negli anni 1998 e 1999.

Come descritto nella sezione 5.3, la cartografia digitale della vegetazione è stata prodotta per le seguenti 10 aree campione:

codice	denominazione
AC1	spinale.1
AC2	commezzadura.16
AC3	pellizzano.17
AC4	terlago.9
AC5	strigno
AC6	moena.1
AC7	tesero.2
AC8	pievetesino
AC9	casteltesino.23
AC10	primiero

Nella successiva analisi sono state considerate le covate presenti in tali 10 aree campione negli anni 1998, 1999. Il numero di covate complessivamente analizzate, per anno e in totale è:

anno	n° punti
1998	20
1999	24
totale	44

6.1.2 I siti generici

Da un punto di vista formale, per poter stabilire quali variabili di ecologia del paesaggio siano determinanti per la caratterizzazione delle zone più vocate all'allevamento dei pulli, devono essere costruiti degli esempi di configurazioni ambientali sfavorevoli o almeno non registrate come favorevoli. Si tratta di una problematica caratteristica nella costruzione di modelli di preferenza in ambito territoriale. Ad esempio, nella creazione di un modello per archeologia preventiva, è in generale difficile presumere che un sito abbia caratteristiche negative per la frequentazione antropica. Allo stesso modo, il mancato incontro di una covata nel corso di un censimento può dipendere dalle condizioni meteorologiche, da precedenti disturbi, da inefficienze non sistematiche dei rilevatori o più semplicemente da fattori casuali dovuti alla distribuzione spaziale della covata al momento del censimento.

Una metodologia per risolvere il problema dell'assegnazione di etichette negative a siti ove non è stata riscontrata la presenza di una covata potrebbe prevedere l'analisi di tutte le configurazioni di struttura del paesaggio possibili all'interno dell'area campione. Se la distribuzione dei siti con presenza di covate differisce significativamente rispetto a quella dei siti generici nello spazio multidimensionale delle variabili utilizzate, è in via di principio possibile costruire un modello statistico in grado di discriminare le due classi. Tuttavia, in questi problemi il numero di siti di presenza è di diversi ordini di grandezza più piccolo di quelli potenzialmente interessati al fenomeno. La disponibilità in Trentino di un modello digitale del terreno a celle di 10 m, e di un sistema di ortofoto alla risoluzione di 1 m rendono elevatissima la scala di analisi dei problemi ambientali. È perciò rischioso confrontare campioni di poche decine o centinaia di esempi positivi con alcuni milioni di esempi non positivi. Per tale motivo, si preferisce costruire tramite un processo di campionamento casuale un campione di esempi di presenza non rilevata, detti anche siti generici (o, in senso lato, siti di assenza).

Nel caso studiato, si è deciso allora di estrarre un campione per tutte le aree campione individuando in modo casuale un numero di siti pari allo 0.002% del numero di celle dell'area campione totale, da cui sono state preventivamente esclusi opportuni intorno dei siti di presenza. Più precisamente, nell'estrazione casuale dei siti si sono escluse le zone troppo vicine ai punti di contatto per le covate e definite da cerchi di raggio 100 m centrati nei siti di contatto. Si osservi che la dimensione della zona di esclusione è inferiore alla più piccola scala utilizzata per studiare la struttura del paesaggio. La misura di 100 metri è peraltro inferiore all'home range stimato per le covate in circa 300 metri.

In riferimento a questa definizione delle zone di esclusione, si è sviluppata in ambiente GRASS una procedura automatica in linguaggio Perl che consente di:

1. generare un maschera per limitare il dominio di estrazione dei siti solo all'interno delle aree campione;
2. generare un ulteriore maschera per escludere l'estrazione di non siti in prossimità dei punti di localizzazione effettiva delle covate. Si sono

utilizzate a riguardo delle maschere circolari di raggio 100 m centrate nei siti;

3. estrarre in maniera casuale i siti di tipo generico all'interno del dominio ristretto individuato dalle maschere in ciascuna delle aree campione. Per tale procedura è stato utilizzato il modulo GRASS `r.random`.

In figura 6.2 è riportato un esempio per l'individuazione automatica del dominio di estrazione degli esempi generici. Sono evidenti i buffer circolari di esclusione attorno ai siti di incontro di covate.

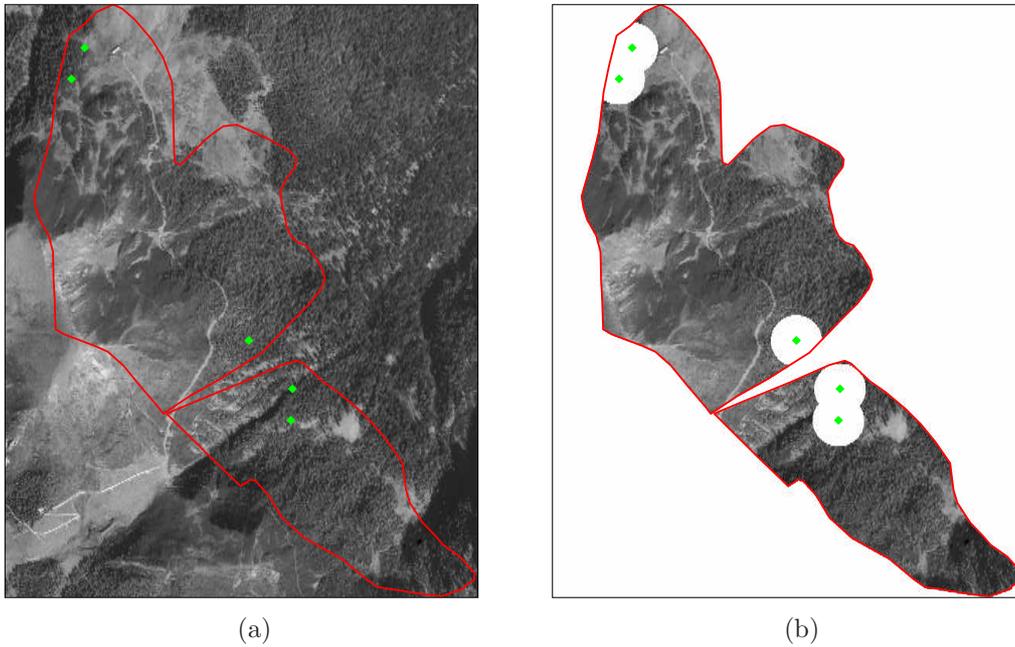


Figura 6.2: Risultato della procedura di definizione del dominio di analisi per le aree campione “casteltesino.23” e “pievetesino”: (a) aree campione; (b) area di estrazione per gli esempi generici.

Il numero totale di siti che formano il campione generico estratto dalle aree campioni è formato da 417 esempi.

6.1.3 Scale di analisi

Si è analizzata la struttura del paesaggio nell'intorno di ciascun sito di presenza (classi **A**) e sito generico (classe **B**).

Nel capitolo 4 si è descritta ed applicata una metodologia per individuare le scale più idonee alla descrizione della struttura del paesaggio. I risultati hanno portato all'individuazione di tre scale significative di analisi: 400 m, 300 m, 150 m.

Questi valori sono ora utilizzati per definire le dimensioni delle finestre per la caratterizzazione della struttura del paesaggio nell'intorno di ciascun sito. In figura 6.3 è riportato un esempio delle scale di analisi utilizzate.

L'utilizzo di tali scale è compatibile con quanto noto riguardo l'*home range* del fagiano di monte come segnalato dal Servizio Faunistico PAT, ed in particolare per quanto relativo alle covate di fagiano di monte; da recenti studi non pubblicati si è infatti osservato che l'*home range* varia dai 4 ai 15 ettari.

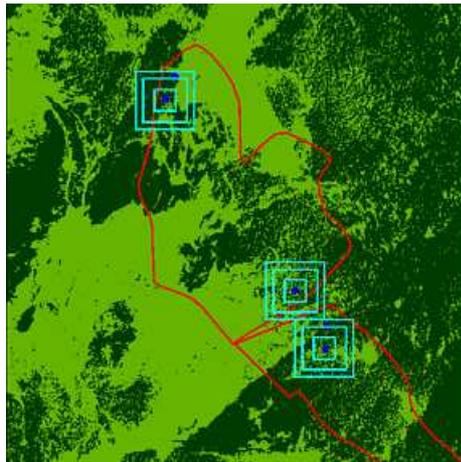


Figura 6.3: Esempi di siti di covate con l'indicazione delle tre scale di analisi utilizzate.

6.1.4 Variabili di struttura del paesaggio

Si definiscono ora le variabili che descrivono la struttura del paesaggio nell'intorno di ogni sito e per le tre differenti scale.

Si utilizzano degli indicatori globali che riassumono le informazioni sui singoli patch calcolate per ciascun sito e per ciascuna scala con le modalità descritte nella sezione 4.2. Gli indicatori globali rappresentano il più delle

volte i momenti statistici delle variabili elementari su tutti i patch individuati nella scala o su sotto insiemi ben precisi. Le variabili globali utilizzate sono quindi:

npatch numero dei patches di bosco.

npatchmin numero dei patches di bosco di dimensione inferiore a $100 m^2$.

sup1tot somma della superficie dei patches di bosco di area maggiore di $100 m^2$.

sup2tot somma delle superfici interne dei patches di bosco di area maggiore di $100 m^2$.

sup3tot somma delle superfici marginali dei patches di bosco di area maggiore di $100 m^2$.

supmin somma delle superfici dei patches di bosco di area inferiore a $100 m^2$.

pertot somma dei perimetri totali dei patches di bosco di area superiore ai $100 m^2$.

msize1 superficie media dei patches di bosco di area superiore a $100 m^2$.

msize2 superficie interna media dei patches di bosco con area maggiore di $100 m^2$.

msize3 superficie marginale media dei patches di bosco con area maggiore di $100 m^2$.

mper perimetro medio dei patches di bosco con area maggiore di $100 m^2$.

mPA coefficiente di forma PA medio dei patches di bosco con dimensione maggiore di $100 m^2$.

mCPA coefficiente di forma CPA medio dei patches di bosco con dimensione maggiore di $100 m^2$.

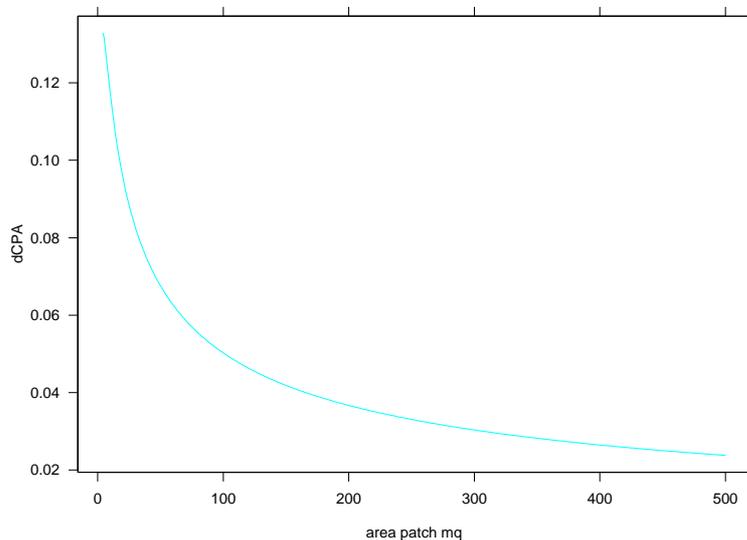


Figura 6.4: Precisione di calcolo del coefficiente CPA al variare della dimensione del patch.

mRCC coefficiente di forma RCC medio dei patches di bosco con dimensione maggiore di $100 m^2$.

dist distanza media tra bordo e bordo dei patches di bosco.

box dimensione di box counting calcolata su una finestra 400×400 .

log coefficiente angolare della regressione lineare tra $\frac{\log(per)}{\log(size)}$ per tutti i patches.

La scelta del valore di soglia di $100 m^2$ nella definizione di molte variabili è dovuta a due motivazioni principali. La prima, di carattere prevalentemente pratico, è legata all'utilizzo di coefficienti di forma che non hanno significato per patch con dimensioni confrontabili con la risoluzione dell'immagine. In figura 6.4 è riportata una misura della precisione del calcolo del coefficiente CPA (dCPA) in funzione della dimensione del patch.

La seconda motivazione è legata ad un'ipotesi sull'importanza che i patches di piccola dimensione possono avere per l'ecosistema del fagiano di monte: si noti che nonostante sia i patches di bosco e cespuglio che quelli di prato

dovrebbero essere di interesse per il fagiano di monte, in questo lavoro di tesi, l'analisi è stata applicata ai soli patch classificati bosco.

Prima di procedere con lo sviluppo del modello multivariato, conviene analizzare separatamente la capacità discriminativa delle singole variabili (analisi monovariata); inoltre è istruttivo eseguire l'analisi alle tre diverse scale. Nelle figure 6.6, 6.7 e 6.5 sono esposte per tutti i siti di tipo **A** e **B** le distribuzioni di probabilità delle variabili. Per ciascuna variabile e per ogni scala è possibile evidenziare la capacità di separazione per le due classi di esempi. Poiché tale discriminazione verrà complessivamente utilizzata dai modelli sviluppati più avanti in questo capitolo, non sono qui inclusi test di significatività specifici per le variabili.

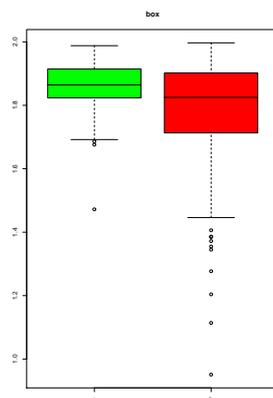


Figura 6.5: Distribuzione degli esempi di training per la variabile di dimensione frattale: in verde gli esempi di classe **A** in rosso gli esempi di classe **B**.

Per una più corretta interpretazione dei dati e per valutare meglio l'importanza degli indicatori, si è quindi proceduto ad un'analisi multivariata che consideri contemporaneamente le variabili descrittive considerate. Nelle sezioni successive si descriverà la metodologia utilizzata per lo sviluppo del modello predittivo per la classificazione dei siti di tipo **A** e **B** utilizzando come spazio delle misure gli indicatori globali appena descritti. Il modello di classificazione utilizzato è il recursive partitioning (rpart), o albero a classificazione binaria.

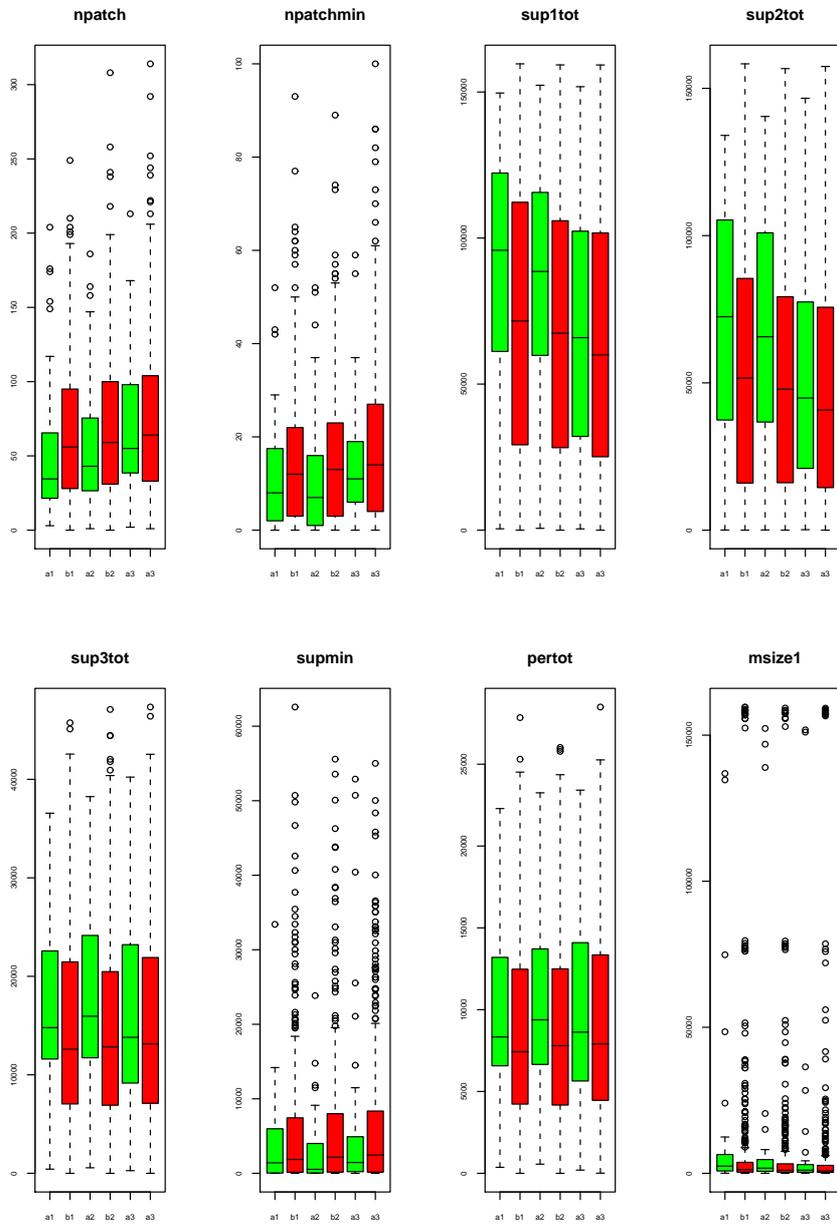


Figura 6.6: Distribuzione degli esempi di training per ciascuna variabile nelle tre scale di analisi: in verde gli esempi di classe A, in rosso gli esempi di classe B.

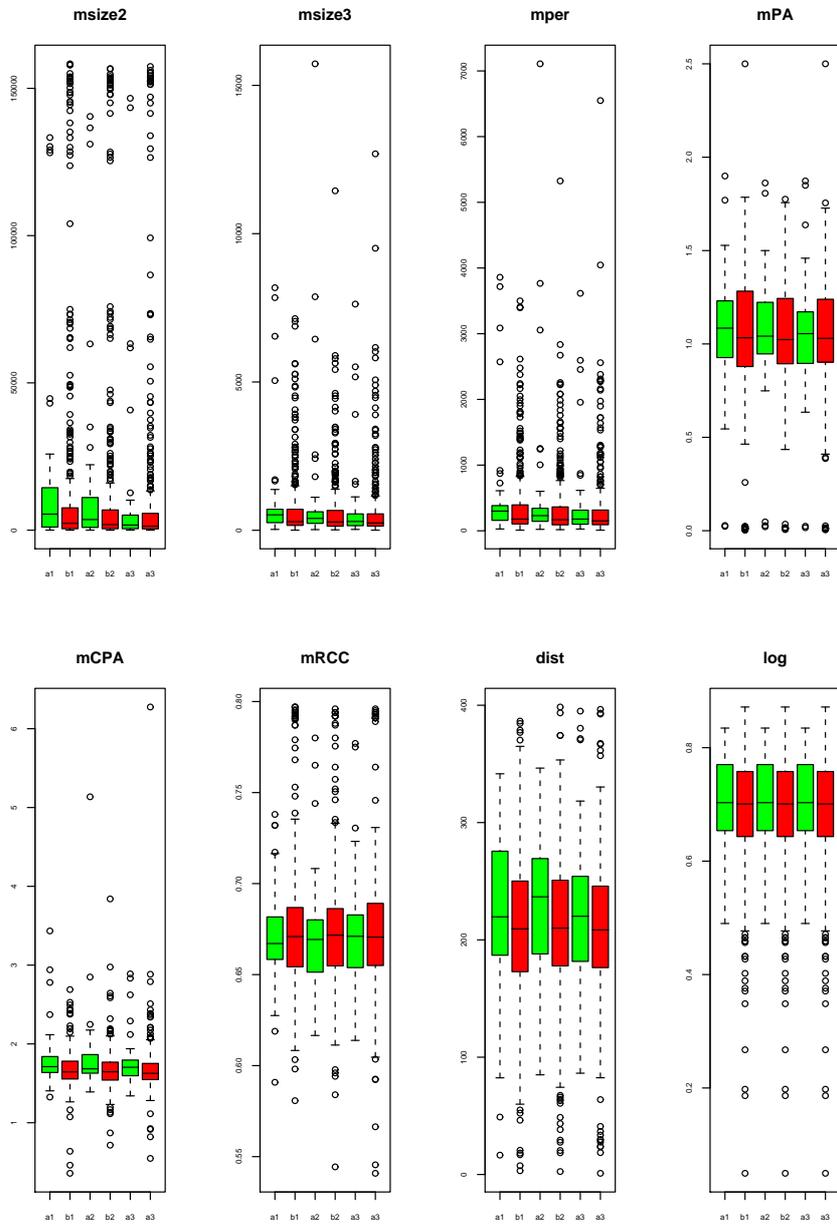


Figura 6.7: Distribuzione degli esempi di training per ciascuna variabile nelle tre scale di analisi: in verde gli esempi di classe A, in rosso gli esempi di classe B.

6.2 Gli alberi di classificazione binaria

Gli *alberi di classificazione binaria* sono stati introdotti da Breiman et al. (1984) [21] in alternativa alle analisi statistiche classiche. Un albero binario è essenzialmente una procedura ricorsiva non parametrica di classificazione che permette di suddividere tramite test numerici o discreti lo spazio delle variabili. Le variabili gestite possono essere sia di tipo quantitativo che di tipo qualitativo.

La rappresentazione grafica della procedura di classificazione attraverso un albero permette la visualizzazione dei risultati in una forma facilmente interpretabile che consente di comprendere in maniera immediata molti aspetti della struttura dati analizzata.

Da un punto di vista formale un classificatore ad albero opera una partizione dello spazio degli input $x = (x_j)$ in sottoregioni. La partizione è eseguita ricorsivamente per le singole variabili x_j dei dati disponibili per l'addestramento (insieme di training): $\mathcal{L} = \{(x_i, y_i)\}_{i=1\dots n}$. Nell'esempio che segue in figura 6.8 si riporta un diagramma relativo ad un classificatore ad albero associato ad un problema di classificazione da dati GIS. Per una descrizione delle convenzioni di rappresentazione di tali diagrammi, si veda Venables and Ripley 1999.

Le regole per la partizione dello spazio (*splitting*) dipendono dal tipo di variabili utilizzate e sono:

- $x_j < t$ per variabili numeriche o categoriche ordinate;
- $x_j \in S$ per variabili categoriche a L livelli, dove S è un possibile sottoinsieme di $\{1 \dots L\}$, Il numero di split possibili è perciò $2^{L-1} - 1$.

In ogni nodo la procedura ricorsiva sceglie la variabile x_j che rende più omogenei i nodi successivi.

La regola di partizione e l'idea di omogeneità si formalizzano introducendo il concetto di impurità del nodo t . Si definisce con $p(k | t)$, con $k = 1, \dots, N$, la frazione dei casi $x_n \in t$ appartenenti alla classe k , in modo tale che $\sum_{k=1}^N p(k|t) = 1$. Si definisce invece misura d'impurità la funzione:

$$i(t) = f(p(1 | t), p(2|t), \dots, p(N | t))$$

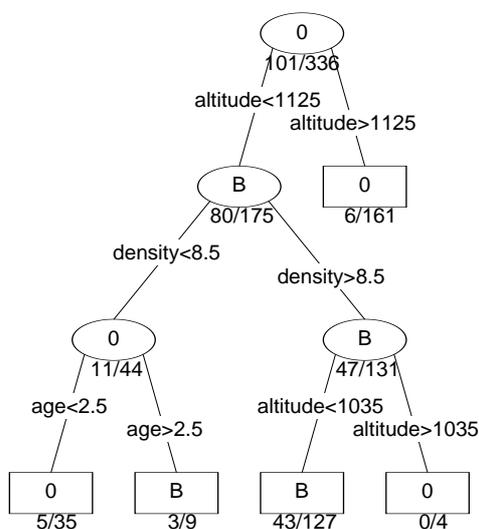


Figura 6.8: Visualizzazione di un albero di classificazione binaria.

tale che $i(t)$ è massima quando tutte le classi sono presenti in proporzione uguale, e minima quando il nodo contiene dati di un'unica classe. Le misure di impurità più note sono:

$$i(t) = \sum_{i \neq j} p(i|t)p(j|t) \tag{6.1}$$

che prende il nome di indice di Gini, oppure la misura di entropia che è definita come:

$$i(t) = \sum_{j=1}^N p(j|t) \log p(j|t). \tag{6.2}$$

La partizione avviene massimizzando tra tutti i possibili split di tutte le variabili la diminuzione di impurità $i(t)$. Il procedimento di partizione ricorsiva termina nei nodi dove non è possibile un'ulteriore decremento di impurità, tali nodi sono chiamati terminali. Per ogni nodo terminale è assegnata la classe

$$k = \arg \max_{1 \leq j \leq K} p(j | t)$$

con $k \in \{1, \dots, k\}$ insieme delle classi.

Si ottiene così un albero *massimale*, che il più delle volte tende a descrivere troppo bene i dati (overfitting). Si deve quindi scegliere un modello ridotto e

per fare ciò solitamente si considera una funzione di costo regolarizzata che penalizza gli alberi troppo dettagliati:

$$R_\alpha(T) = R(T) + \alpha \text{size}(T)$$

dove T è l'albero binario, $R(T)$ l'errore di classificazione e $\alpha \text{size}(T)$ è un termine di complessità. Si può dimostrare che esiste solo una sequenza finita di sottoalberi $T(\alpha_i)$, ognuno ottimo in $[\alpha_i, \alpha_{i+1}]$. In definitiva, viene implementata una procedura di regolarizzazione che bilancia la diminuzione dell'errore di modello $R(T)$ con la penalizzazione di modelli ad elevata complessità. In questo lavoro di tesi, per la scelta del $T(\alpha)$ migliore si è utilizzata la procedura di *cross validation* [22] secondo l'implementazione di febbraio 2000 di *rpart* per i linguaggi *S - PLUS* e *R* di T. Thernau (Mayo Foundation).

6.2.1 Cross validation

Considerato T^b l'albero ottimale per l'intervallo $[\alpha_b, \alpha_{b+1}]$, la tecnica di *cross validation* consiste nel suddividere l'insieme di training in k sottoinsiemi disgiunti e di stessa cardinalità $I_{k=1, \dots, n}$. Per ciascun insieme di training

$$\mathcal{L}_i = \cup \{I_k, \forall j \neq i\} \forall i \in 1, \dots, k$$

si costruiscono dei nuovi alberi binari

$$T_i^b$$

. È quindi possibile calcolare l'errore di predizione e_i^b per ciascuno degli alberi testandoli sull'insieme $I_i = \mathcal{L} \setminus \mathcal{L}_i$.

Per ogni intervallo $[\alpha_b, \alpha_{b+1}]$ è definito l'errore predittivo complessivo di cross validation come

$$E_{cv}^b = \frac{1}{k} \sum_{k=1}^k e_i^b$$

La stima di cross-validation individua una misura ottimale di errore predittivo che viene utilizzato per selezionare l'albero ottimale di base T^b , definito come

$$T = T_b \mid b = \arg \min_b E_{cv}^b$$

6.3 Modelli di classificazione rpart

Per l'applicazione di un modello di classificazione ad albero si sono utilizzate le procedure implementate nel programma statistico S-PLUS e in particolare i moduli `rpart`. Le stesse procedure sono disponibili anche per il linguaggio statistico R, utilizzato nella fase finale di analisi e modellizzazione.

A partire dalla base dati di esempi di tipo A e B per il fagiano di monte, sono stati sviluppati 3 diversi modelli in corrispondenza a ciascuna scala di analisi:

Modello 1 scala $a_1 = 400$ m.

Modello 2 scala $a_2 = 300$ m.

Modello 3 scala $a_3 = 150$ m.

Per ogni modello sono stati utilizzati gli indicatori di ecologia del paesaggio calcolati alla scala corrispondente, ottenendo configurazioni abbastanza differenti. Una descrizione dei modelli in termini dei loro parametri fondamentali (n° nodi, n° variabili utilizzate (N var), n° di variabili totali, coefficiente Cp^1) è data in tabella 6.3

	Cp	N nodi	Xerror	N var.	N var. tot
Modello 1	0.0000	43	0.142	13	16
Modello 2	0.0014	42	0.119	14	16
Modello 3	0.0014	46	0.137	14	16
Modello 123	0.0029	30	0.096	20	46

Tabella 6.1: Errori di cross validation per i diversi modelli costruiti.

Osservato che il nostro interesse è per modelli a miglior capacità predittiva, qui modellizzata dal valore della stima di cross-validation *xerror*, dalla 6.3 si ricava che il miglior modello monoscala è il Modello 2 che descrive la struttura del paesaggio per finestre quadrate di lato 300 m.

Per migliorare le prestazioni del modello si è infine considerato un approccio multiscalare analizzando simultaneamente le variabili per le tre diverse

¹ Cp è il coefficiente di complessità dell'albero paragonabile ad α

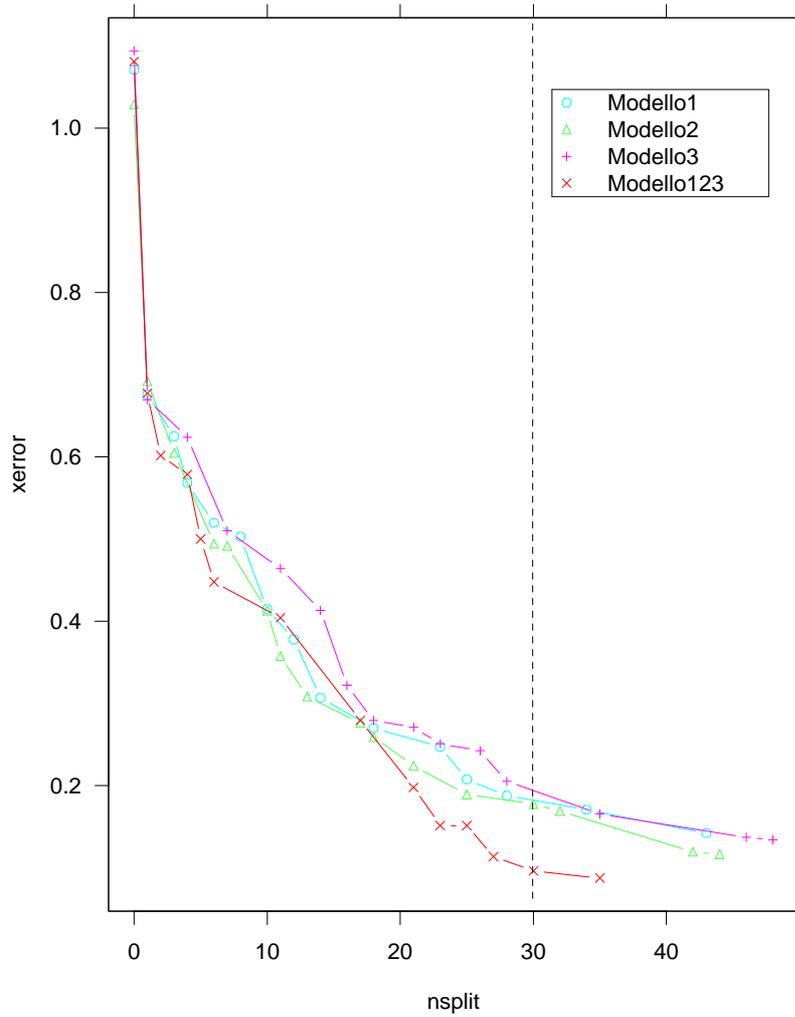


Figura 6.9: Confronto degli errori in funzione del numero di split per i quattro modelli.

scale: il modello così ottenuto è stato definito come Modello 123. In figura 6.9 è riportato il confronto i termini di errore di cross validation dei 4 modelli ad albero prodotti (alle tre scale e alla multiscala). L'errore di cross

validation è calcolato in funzione del numero di split e quindi della complessità dell'albero. È particolarmente interessante osservare come il modello migliore risulti proprio il Modello 123 che presenta prestazioni superiori sia in termine di errore predittivo, sia di inferiore complessità. I valori riportati in tabella per i diversi modelli non si riferiscono all'albero massimale, ma a quello ottimizzato utilizzando la stima di cross validation.

Il risultato finale del Modello 123 è descritto dalla sua rappresentazione grafica ad albero 6.10. Per un'ulteriore analisi di questo risultato, sono stati realizzati quattro grafici in cui è confrontata l'importanza delle diverse variabili a livello della decisione al nodo radice (grafici 6.11, 6.12, 6.13, 6.14). Inoltre, il differente uso delle variabili nei quattro migliori modelli è riassunta in tabella (6.2). Si noti come il modello multiscala sia composto da variabili di tutte e tre le risoluzioni.

6.4 Discussione dei risultati

La struttura gerarchica dei modelli ad albero, caratterizzati da split binari ricorsivi, permette sia una facile interpretazione del modello in termini ecologici che la determinazione delle variabili, e dei loro valori di soglia con buona capacità discriminante. Osservando il modello ad albero rappresentato in figura 6.10, risulta evidente come la variabile *box* sia la più importante per la caratterizzazione delle aree di presenza di covate. In particolare, le aree di presenza sono state rilevate principalmente in corrispondenza dei patches che presentano una dimensione frattale di box counting elevata ($box > 1.8$). Da un punto di vista strutturale, ciò significa che le aree di presenza sono caratterizzate da una netta separazione tra le zone di bosco e pascolo.

Inoltre, la forma media dei patch di bosco delle aree di presenza è caratterizzata da valori bassi della variabile *mPA* ($mPA < 1.7$). Questo significa, ad esempio, che le covate di fagiano di monte sono state osservate in zone di pascolo interrotte da zone di bosco a simmetria radiale. In termini pratici, ciò significa che le zone di bosco, utilizzate come zone di riparo, devono essere maggiormente simili ad una circonferenza che ad una ellisse caratterizzata da alta eccentricità.

	Modello1	Modello 2	Modello 3	Modello 123
1	box	box	box	box
2	msize3	msize3	supmin	msize3.2
3	mCPA	npatchmin	npatchmin	npatchmin.2
4	npatch	msize1	mCPA	msize1.2
5	sup2tot	supmin	msize3	sup2tot
6	msize 1	msize2	dist	sup1tot.2
7	sup3tot	mCPA	log	dist.1
8	npatchmin	npatch	sup1tot	dist.2
9	mPA	dist	mPA	mCPA.2
10	mRCC	sup2tot	mRCC	mCPA.3
11	pertot	mPA	msize1	mPA.1
12	dist	pertot	sup2tot	mPA.2
13	supmin	sup3tot	npatch	mRCC.2
14	-	mRCC	sup3tot	msize1.1
15	-	-	-	sup1tot.1
16	-	-	-	msize2.2
17	-	-	-	sup1tot.3
18	-	-	-	sup3tot.1
19	-	-	-	sup3tot.2
20	-	-	-	supmin.3

Tabella 6.2: Variabili utilizzate nei modelli: in grassetto sono riportate le variabili utilizzate nel modello migliore (Modello 123)

Un altro gruppo di variabili significative sono risultate essere gli indicatori *supmin.3* e *npatchmin.2* che sono stati utilizzati per calcolare la superficie ed il numero di patches di bosco di dimensione ridotta (inferiori ai 100 mq). In particolare, si può osservare che le zone di presenza sono caratterizzate da un numero ridotto di patches di bosco piccoli e tali che la loro superficie sia ancora mediamente piccola ($npatchmin.2 < 10.5$, $supmin.3 < 3148$). Questo risultato, al pari di quell ottenuto con l'indice di dimensione frattale, può indicare che non sono favorevoli alla presenza di covate le zone caratterizzate dall'espansione del bosco verso il pascolo. Questo risultato potrebbe avere forte rilevanza dal punto di vista applicativo dopo le opportune verifiche sul campo.

Se si osservano i risultati esposti in tabella 6.2, si può notare il diverso

contributo delle singole scale al modello finale (Modello 123); in tale tabella sono indicate in grassetto le variabili utilizzate nel Modello 123. La scala che apporta le maggiori informazioni è la numero 2 (300 m) che contribuisce con 10 variabili, mentre le variabili utilizzate della scala numero 3 (150 m) sono 3, la scala 1 invece contribuisce con 6. Si osservi poi dal grafico in figura 6.14 che le variabili della scala 2 oltre ad essere le più numerose sono anche le più importanti. Questo risultato può indicare la scala 2 come alla base della descrizione dei fenomeni.

In figura 6.15 si è infine riportata la visualizzazione della struttura del paesaggio nell'intorno di 4 siti classificati dal modello come presenza **A** (6.15a) e nell'intorno di 4 siti classificati come generici **B** (6.15b). Nell'intorno di ciascun sito si è rappresentata la finestra più grande di analisi di lato 400×400 m e l'intorno circolare di 100 metri utilizzato nella procedura di esclusione per la determinazione degli esempi generici. Si noti come tra gli esempi di tipo B siano compresi una zona particolarmente scoperta, mentre i 4 siti di tipo A siano abbastanza simili fra loro.

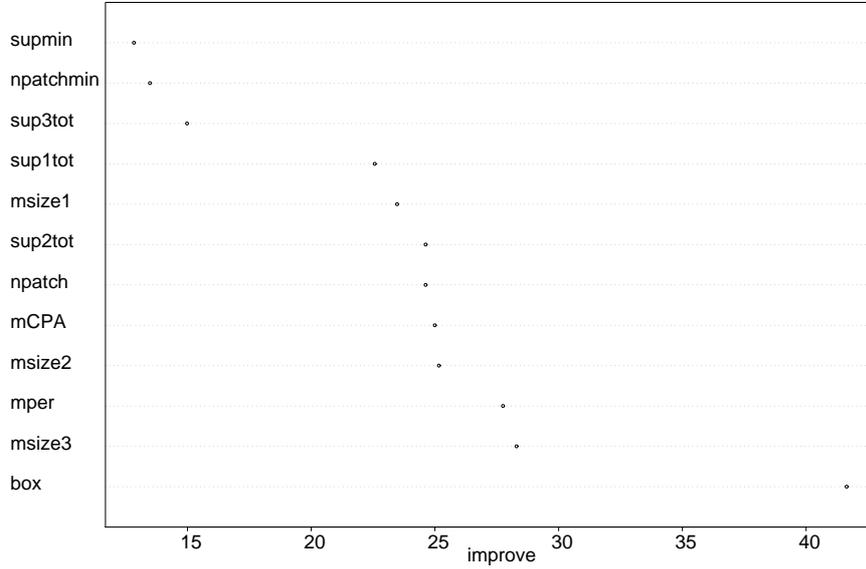


Figura 6.11: Grafico di importanza delle variabili per il Modello 1.

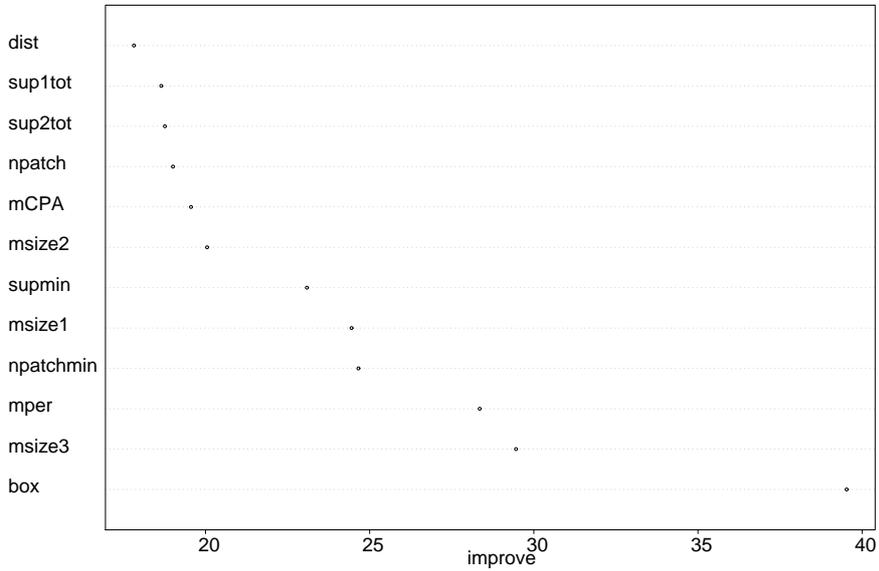


Figura 6.12: Grafico di importanza delle variabili per il Modello 2.

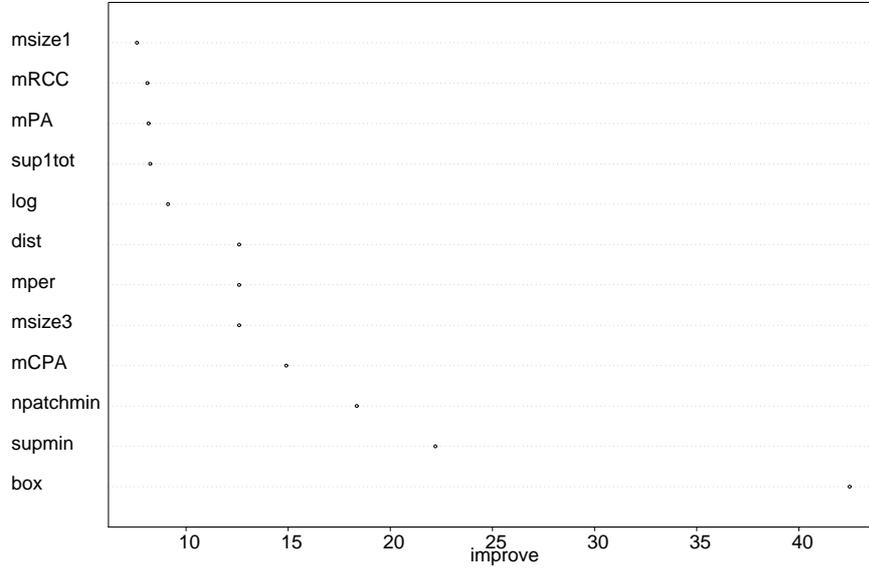


Figura 6.13: Grafico di importanza delle variabili per il Modello 3.

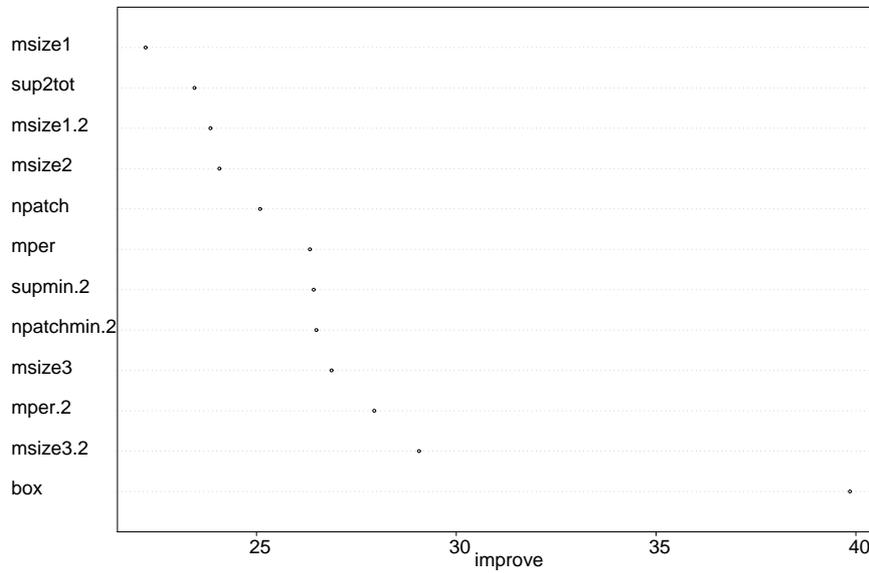


Figura 6.14: Grafico di importanza delle variabili per il Modello 123.

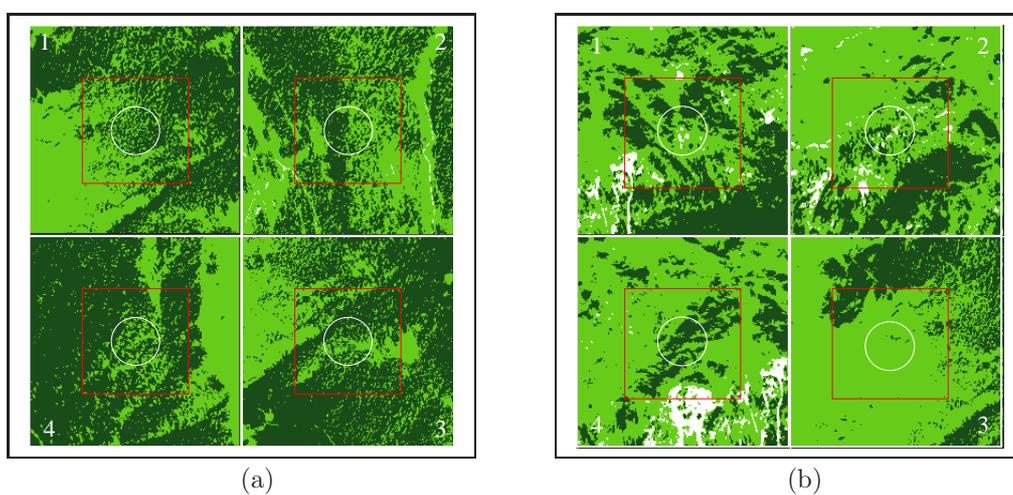


Figura 6.15: Esempi di classificazione del modello: siti di presenza (classe **A**) (a); siti generici (classe **B**) (b).

Capitolo 7

Conclusioni

Il modello predittivo sviluppato ha fornito alcune importanti indicazioni sulle esigenze di habitat delle covate di fagiano di monte. Pur utilizzando una base OFD a livelli di grigio che non permette di distinguere compiutamente il dettaglio di caratteristiche vegetazionali interessanti per la descrizione completa dell'habitat (ad es. i mugheti), gli indicatori di ecologia del paesaggio calcolati hanno mostrato di poter discriminare adeguatamente (10% errore predittivo) i siti di presenza dai siti generici utilizzati come esempi negativi.

In questo studio sono stati utilizzati un numero di punti di contatto limitato. I modelli realizzati vanno quindi considerati con cautela: tuttavia va segnalata l'importanza della misura box-counting e le indicazioni relative alle variabili di forma e di caratterizzazione dei patch piccoli. Questi risultati sono interessanti da un punto di vista applicativo, poiché non è difficile ipotizzare la costruzione di un sistema software per simulazione di intervento che ne tenga adeguatamente conto.

In questo senso si può ritenere che gli obiettivi principali di questo lavoro di tesi siano stati raggiunti. Si è infatti provata la fattibilità dell'applicazione delle metodologie GIS come supporto predittivo ai miglioramenti ambientali di carattere locale. Si è confermata l'esperienza recente di alcuni gruppi di ricerca per quanto riguarda l'integrazione delle strutture dati e software GIS con procedure modellistiche e sistemi di gestione dei dati. Questo lavoro estende in particolare l'approccio seguito presso l'ITC-irst per la costruzione di modelli predittivi a partire da variabili ambientali. È stato istruttivo verificare non solo come la combinazione in maniera automatica di un GIS con

software statistici e di gestione di basi dati permetta di sviluppare un sistema integrato in grado di gestire dati ambientali provenienti da fonti diverse, ma anche come tali fonti possano essere basate sulla classificazione automatica di foto aeree anche alla scala necessaria ad un intervento territoriale locale.

I principali risultati di questa tesi hanno infine riguardato la fattibilità di basare su indicatori numerici della struttura del paesaggio la modellizzazione della qualità dell'ambiente per una specie a rischio quale il fagiano di monte. Lo sviluppo di nuove metodologie software in ambiente GIS per il fagiano di monte ha avuto caratteristiche sufficientemente generali da ritenere possibile l'estensione ad altre specie in ambiente alpino, specialmente in relazione all'utilizzo di descrizioni spaziali più sofisticate (ortofoto a colori e dati SAR).

Appendice A

Dimensione frattale

A.1 Dimensione di Hausdorff

Si definirà in questa sezione la *dimensione di Hausdorff* per insiemi A che appartengono allo spazio

$$\mathbb{R}^n = \{x \mid x = (x_1, \dots, x_n), x_i \in \mathbb{R}\}$$

per un numero naturale [8].

Sono comunque necessarie alcune notazioni matematiche per arrivare alla definizione. La distanza su \mathbb{R}^n sarà la distanza euclidea definita da: $d(x, y)$, in \mathbb{R}^n ,

$$d(x, y) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2}$$

Usando queste notazioni possiamo definire il diametro di un sottoinsieme U di \mathbb{R}^n :

$$diam(U) = \sup\{d(x, y) \mid x, y \in U\}$$

L'ultima nozione di cui abbiamo bisogno è di ricoprimento aperto di un sottoinsieme A di \mathbb{R}^n . Un sottoinsieme U di \mathbb{R}^n è detto aperto se per ogni $x \in U$ esiste una palla, $B_\epsilon(x) = \{y \in \mathbb{R}^n \mid d(x, y) < \epsilon\}$ di raggio $\epsilon > 0$ centrata in x , che è contenuta in U . Sia $\{U_i\}_{i \in I}$ una famiglia di aperti di

\mathbb{R}^n diremo che questa famiglia è un ricoprimento aperto di A se soddisfa la seguente condizione:

$$A \subseteq \bigcup_{i \in I} U_i$$

A questo punto siamo pronti per definire la *dimensione di Hausdorff* di A . Dati s e ϵ numeri reali positivi si definisce:

$$h_\epsilon^s(A) = \inf \left\{ \sum_{i=0}^{\infty} \text{diam}(U_i)^s \mid \{U_1, U_2, \dots\} \text{ ricoprimento aperto di } A \text{ con } \text{diam}(U_i) < \epsilon \right\}$$

La somma della serie può essere finita o infinita. Al decrescere di ϵ la classe dei possibili ricoprimenti di A diminuisce, perciò l'estremo inferiore aumenta o rimane uguale avvicinandosi quindi ad un valore limite per $\epsilon = 0$ che può essere infinito od un numero reale. Dunque

$$\exists h^s(A) = \lim_{\epsilon \rightarrow 0} h_\epsilon^s(A)$$

Il limite $h^s(A)$ è chiamato *misura s-dimensionale di Hausdorff* di A . Si può vedere che se A è l'insieme vuoto la misura s-dimensionale di Hausdorff è 0 $\forall s \in \mathbb{R}$ e che $h^s(A) \leq h^s(B)$ se $A \subset B$. Se A è il supporto di una curva regolare $a : I \subseteq \mathbb{R} \rightarrow \mathbb{R}^n$ allora si può vedere che $h^1(A)$ è la lunghezza della curva $|a|$ e che $h^2(A)$ è l'area di una superficie regolare a meno di un fattore $\pi/4$ e che $h^3(A)$ è il volume di un corpo regolare tridimensionale a meno di un fattore $4\pi/3$.

Un'altra importante proprietà è che se $f : A \rightarrow \mathbb{R}^n$ è α -Hölderiana, cioè esiste una costante c reale tale che per tutte le coppie $x, y \in A$ si ha che

$$d(f(x), f(y)) \leq c(d(x, y))^\alpha$$

allora si può vedere che

$$h^{s/\alpha}(f(A)) \leq c^{s/\alpha} h^s(a).$$

Per esempio, se f è una omotetia con fattore di contrazione $0 \leq c < 1$ allora f soddisfa la condizione di Holder con $\alpha = 1$ e quindi $h^s(f(a)) \leq c^s h^s(A)$.

Si può mostrare che esiste un numero $D_H(A)$ tale che

$$h^s(A) = \begin{cases} \infty & \text{se } s < D_H(A) \\ 0 & \text{se } s > D_H(A) \end{cases}$$

Il numero $D_H(A)$ è definito *dimensione di Hausdorff* di A e si può vedere che vale la seguente relazione

$$D_H(A) = \inf\{s \mid h^s(A) = 0\} = \sup\{s \mid h^s(A) = \infty\}.$$

Se $s = D_H(A)$ allora $h^s(A)$ può essere zero, infinito o qualsiasi numero reale positivo.

Riportiamo di seguito alcune proprietà fondamentali della dimensione di Hausdorff:

1. Se $A \subset \mathbb{R}^n$ allora $D_H(A) \leq n$
2. Se $A \subset B$ allora $D_H(A) \leq D_H(B)$
3. Se A è un insieme numerabile allora $D_H(A) = 0$
4. Se $A \subset \mathbb{R}^n$ e $D_H(A) < 1$ allora A è totalmente sconnesso.

A.2 Dimensione di Box Counting

Ci sono numerose difficoltà nel valutare la dimensione di Hausdorff in un caso concreto. La dimensione di box-counting in questo senso vuole superare queste difficoltà.

La difficoltà fondamentale nel calcolare la dimensione di Hausdorff è data dal termine $\sum_{i=0}^{\infty} \text{diam}(U_i)^s$. La dimensione di box-counting semplifica questo problema sostituendo il termine $\text{diam}(U_i)^s$ con un termine δ^s che dipende da un ricoprimento in pluritangoli di A .

Una definizione formale della dimensione di box-counting D_b di qualsiasi insieme limitato A di \mathbb{R}^n è la seguente: sia $N_\delta(A)$ il più piccolo numero di insiemi di diametro al più δ che coprono A . Allora

$$D_b(A) = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(A)}{\log 1/\delta}$$

se esiste il limite. Ci sono diverse definizioni equivalenti di $D_b(A)$. Per esempio, consideriamo una suddivisione di \mathbb{R}^n in plurirettangoli con dimensione della maglia δ . Sia ora $N'_\delta(A)$ il numero dei plurirettangoli che ricoprono \mathbb{R}^n , che contengono A . Allora si dimostra che

$$D_b(A) = \lim_{\delta \rightarrow 0} \frac{\log N'_\delta(A)}{\log 1/\delta}$$

a condizione che il limite esista. Dunque $N_\delta(A) \propto \delta^{-s}$ per δ piccoli, dove $s = D_b(A)$, più precisamente

$$N_\delta(A)\delta^s \rightarrow \begin{cases} \infty & \text{se } s < D_b(A) \\ 0 & \text{se } s > D_b(A) \end{cases}$$

Ma

$$N_\delta(A)\delta^s = \inf \left\{ \sum_{i=0}^{\infty} \delta^s \mid \left. \begin{array}{l} \{U_1, U_2, \dots\} \text{ ricoprimento aperto} \\ \text{di } A \text{ con } \textit{diam}(U_i) < \epsilon \end{array} \right\} \right\}$$

Questa espressione si può confrontare con la definizione di dimensione di Hausdorff per osservare che la sola differenza è nel termine $\textit{diam}(U_i)$ contro il termine δ^s .

Tuttavia la dimensione di box-counting e la dimensione di Hausdorff non sempre coincidono. La dimensione di box-counting dei razionali di $[0,1]$ è 1, la dimensione di Hausdorff dello stesso insieme numerabile è 0 [8]. Un altro esempio è l'insieme $A = \{0.1/2, 1/3, 1/4, \dots\}$. che ha dimensione di box-counting frazionaria, infatti $D_b(A) = 1/2$. Il fatto che la dimensione di box-counting $D_b(A)$ non sia un intero non è sufficiente per stabilire che l'insieme A abbia proprietà frattali. È però vero che la dimensione di Hausdorff e la dimensione di Box-counting coincidono per una larga classe di insiemi che includono i classici frattali come l'insieme di Cantor [8].

Appendice B

Classificazione supervisionata

B.1 Introduzione

In questa appendice ci occuperemo di problemi di classificazione supervisionata (*Supervised Classification*). Data un insieme X di oggetti e una famiglia di classi C_i si vuole risolvere il problema di assegnare ogni $x \in X$ ad una classe C_i .

Nella classificazione supervisionata si suppone di conoscere a priori un insieme di k classi e di disporre di un sufficiente numero di dati d'esempio già etichettati (*training set*) per ciascuna classe. Allora si può definire un sistema di *classificazione supervisionata* come una particolare procedura che, utilizzando l'informazione contenuta nei dati di training, valuta un nuovo dato non ancora etichettato, attribuendolo ad una delle k classi. In genere, per la descrizione dei dati si utilizza un insieme di attributi (*features*).

Più formalmente una regola di classificazione è una funzione:

$$F : \mathbf{X} \rightarrow \{C_1, \dots, C_k\} \quad (\text{B.1})$$

Il problema della classificazione è quindi ricondotto alla stima della funzione F sulla base degli esempi del training set D .

Dal punto di vista statistico si possono perciò individuare due fasi fondamentali nella classificazione: la prima consiste nell'utilizzare un modello parametrico o non parametrico per determinare la distribuzione degli esempi di ciascuna classe; la seconda invece applica una statistica delle decisioni per trovare la classe ottimale per i dati da classificare.

Nelle sezioni seguenti, sempre in riferimento a problemi di classificazione, supervisionata verranno trattati gli aspetti della statistica delle decisioni, in particolare il teorema di Bayes e l'algoritmo SMAP, affrontando infine un metodo parametrico di inferenza e si discuterà l'uso di misture di distribuzioni gaussiane applicate alla classificazione di immagini.

B.2 Statistica delle decisioni

In questa sezione supponiamo di conoscere già le probabilità a priori per le diverse classi. Sia cioè $P(C_i)$ la probabilità a priori della classe C_i nell'insieme dei dati. Si suppone anche nota la distribuzione di probabilità nello spazio delle misure per i dati che appartengono ad una determinata classe $p(\mathbf{x} | C_i)$.

Un primo problema che si pone è quello di utilizzare un valido e ragionevole criterio per confrontare tra di loro diverse procedure di classificazione F , come definite in (B.1). Tra i diversi criteri che possono essere introdotti, quello sicuramente più utilizzato consiste nel minimizzare la probabilità di commettere errori di classificazione:

$$p_{mc}(k) = P\{C(\mathbf{x}) \neq C_k | C = C_k\} \quad (\text{B.2})$$

Si può dimostrare che una forma equivalente alla minimizzazione della probabilità di misclassificazione $p_{mc}(k)$ può essere espressa nel seguente modo [5]:

$$F(\mathbf{x}) = C_i \quad \text{se } p(C_i | \mathbf{x}) > p(C_j | \mathbf{x}) \forall j \neq i \quad (\text{B.3})$$

Si osserva però che nei casi pratici le $P(C_i | \mathbf{x})$ in (B.3), non sono in generale note. È però possibile risolvere il problema attraverso il teorema di Bayes [5]:

$$P(C_i | \mathbf{x}) = p(\mathbf{x} | C_i)P(C_i)/p(\mathbf{x}) \quad (\text{B.4})$$

dove $P(C_i)$ è la probabilità a priori per la classe C_i ; mentre $p(\mathbf{x})$ si può scrivere:

$$p(\mathbf{x}) = \sum_{i=1}^M p(\mathbf{x} | C_i)p(C_i) \quad (\text{B.5})$$

Sostituendo a $P(C_i | \mathbf{x})$ in (B.3) l'espressione ottenuta attraverso il teorema di Bayes (B.4) si ottiene una nuova regola di classificazione:

$$F(\mathbf{x}) = C_i \quad \text{se} \quad \frac{p(\mathbf{x} | C_i)P(C_i)}{p(\mathbf{x})} > \frac{p(\mathbf{x} | C_j)P(C_j)}{p(\mathbf{x})} \forall j \neq i \quad (\text{B.6})$$

e poiché $p(\mathbf{x})$ non è funzione delle classi C_i l'equazione diventa:

$$F(\mathbf{x}) = C_i \quad \text{se} \quad p(\mathbf{x} | C_i)P(C_i) > p(\mathbf{x} | C_j)P(C_j) \forall j \neq i \quad (\text{B.7})$$

La regola di classificazione espressa in (B.7) dipende dalle probabilità $p(\mathbf{x} | C_i)$ che possono essere stimate direttamente dai *training data* così come le $p(C_i)$.

Se si introduce poi la *funzione discriminante* g definita come:

$$\begin{aligned} g_i(\mathbf{x}) &= \ln(p(\mathbf{x} | C_i)p(C_i)) \\ &= \ln p(\mathbf{x} | C_i) + \ln p(C_i) \end{aligned} \quad (\text{B.8})$$

allora la regola di decisione può essere espressa in una forma ancora più compatta [6]:

$$\mathbf{x} \in C_i \quad \text{se} \quad g_i(\mathbf{x}) > g_j(\mathbf{x}) \forall j \neq i \quad (\text{B.9})$$

Le regole di classificazione che si sono descritte possono avere una importante interpretazione; possiamo infatti pensare lo spazio delle misure suddiviso in k *regioni di decisione* $\{R_1, \dots, R_k\}$, tali che, se un dato ricade nella regione R_i , gli sarà assegnata la classe C_i . Il bordo tra queste regioni è chiamato *superficie di separazione* ed è definito dall'intersezione delle funzioni discriminanti.

Nella sezione successiva viene presentato il problema della classificazione di immagini, mentre alcuni metodi (parametrici e non parametrici) per la stima delle distribuzioni di probabilità delle varie classi saranno discussi più avanti.

B.3 Classificazione di immagini

Sviluppiamo in questa sezione alcune metodologie di base per l'interpretazione di immagini digitali.

Utilizzando lo stimatore della massima verosimiglianza che definiremo più in dettaglio in appendice D, per la classificazione di immagini. Classificare un'immagine equivale a classificare gli attributi che la definiscono e quindi considerarne i pixel. Questo può essere fatto seguendo due approcci distinti: possiamo classificare direttamente i pixel oppure cercare relazioni spaziali fra gruppi di pixel. In questa appendice ci concentreremo sul primo approccio.

B.3.1 Analisi quantitativa: modello normale multidimensionale

In questo caso, dal punto di vista statistico, i diversi pixel non sono considerati tra loro correlati; sono quindi applicabili direttamente le procedure di stima viste in B.2, classificando i pixel come singoli dati. Le misure sui dati sono in questo caso i livelli di luminosità nelle diverse bande. Nel caso di immagini in bianco e nero lo spazio delle misure sarà monodimensionale.

La regola base per la classificazione è quindi ancora lo stimatore di Bayes (B.9). In genere viene però adottata in questo caso una precisa forma funzionale per le distribuzioni di probabilità delle varie classi, e precisamente si assume che siano gaussiane multidimensionali. Questa è appunto un'assunzione piuttosto che una proprietà dimostrabile degli spettri naturali delle classi di informazioni. Così facendo però si ottiene una semplificazione matematica rilevante.

Si può vedere in [6] che sotto queste ipotesi la regola di classificazione si implementa introducendo nell'equazione (B.9) la seguente espressione per $g_i(\mathbf{x})$:

$$g_i(\mathbf{x}) = \ln p(C_i) - \frac{1}{2} \ln |\sum_i| - \frac{1}{2}(\mathbf{x} - \mathbf{m}_i)^t \sum_i^{-1} (\mathbf{x} - \mathbf{m}_i) \quad (\text{B.10})$$

Dove \mathbf{m}_i e \sum_i sono il vettore media e la matrice covarianza dei dati nella classe C_i . Essendo il termine $(\mathbf{x} - \mathbf{m}_i)^t \sum_i^{-1} (\mathbf{x} - \mathbf{m}_i)$ in (B.10) una funzio-

ne quadratica di \mathbf{x} , anche le superfici di separazione sono in questo caso quadriche.

Alcune volte, come vedremo, nelle analisi non si hanno informazioni circa le $p(C_i)$, nel qual caso si assume che le diverse classi abbiano probabilità a priori uguale. In tal caso l'equazione precedente si semplifica assumendo la forma:

$$g_i(\mathbf{x}) = -\ln \left| \sum_i \right| -(\mathbf{x} - \mathbf{m}_i)^t \sum_i^{-1} (\mathbf{x} - \mathbf{m}_i) \quad (\text{B.11})$$

B.3.2 Classificazione alla minima distanza

L'efficienza della classificazione con massima verosimiglianza dipende da un'accurata stima per il vettore media \mathbf{m} e per la matrice di covarianza Σ , per cui è necessaria una quantità di dati di training sufficiente, altrimenti la stima sarà inaccurata.

Quando il numero dei dati di training è piccolo, è più efficace utilizzare una classificazione che non fa uso di stime dei momenti secondi (varianza e covarianza), poiché troppo instabili, ma solo della media. Questa viene chiamata *classificazione alla minima distanza* [6]. Si utilizza quindi l'insieme di training per la stima della media m_i di ciascuna classe e la classificazione è fatta attribuendo al dato \mathbf{x} la classe che ha media m_i più vicina; questo è equivalente ad utilizzare per la funzione discriminante $g_i(\mathbf{x})$ la seguente espressione:

$$g_i(\mathbf{x}) = 2\mathbf{m}_i \cdot \mathbf{x} - \mathbf{m}_i \cdot \mathbf{m}_i \quad (\text{B.12})$$

dove $m_i, i = 1, \dots, K$ sono le medie delle K calssi calcolate dai dati di training. Le superfici di separazione sono definite da equazioni lineari in x .

B.4 Classificazione contestuale

Si presentano ora alcune procedure di classificazione¹ che utilizzano, oltre agli attributi dei singoli pixel, anche le informazioni spaziali contenute nell'immagine.

Si tratta di evidenziare le correlazioni esistenti tra i pixel dell'immagine e utilizzarle nella fase di decisione: si parla, in questo caso, di *classificazione contestuale*.

Verranno innanzitutto descritte le procedure base e successivamente l'algoritmo SMAP implementato in GRASS.

B.4.1 Tecniche di pre elaborazione

Un primo metodo molto semplice per considerare il contesto spaziale è di elaborare l'immagine prima di classificarla, così da modificare o migliorare le sue proprietà spaziali. Gli algoritmi di pre elaborazioni sono rivolti in generale al miglioramento dei contrasti o alla riduzione del rumore presente nell'immagine. .

Una procedura alternativa è quella di produrre una nuova banda che contenga proprietà spaziali, da utilizzarsi assieme alle bande dell'immagine multispettrale.

Sempre su questo approccio, ma con caratteristiche diverse, si basa la metodologia di classificazione ECHO (Extraction and Classification of Homogeneous Object) [6]. In ECHO le regioni che hanno proprietà spettrali simili sono individuate prima che venga eseguita la classificazione. Le regioni uniformi sono individuate e quindi trattate come un singolo oggetto per poi essere classificate; si può poi utilizzare la massima verosimiglianza per trattare i pixel singoli che non cadono in alcuna regione.

¹Per le immagini si parla anche di segmentazione per enfatizzare il contesto spaziale nel quale sono inseriti i dati. Segmentare significa infatti riconoscere all'interno dell'immagine regioni a comportamento statistico uniforme.

B.4.2 Tecniche di post elaborazione

Una volta che le mappe tematiche sono state generate utilizzando un classificatore per punti, è possibile rendere conto del contesto spaziale utilizzando il filtraggio della mappa. In questo ambito rientrano anche le procedure ricorsive, che utilizzano la funzione di vicinanza per correlare alla classe di ciascun pixel dell'immagine anche le classi dei pixel presenti in un certo intorno.

Appendice C

L'algoritmo SMAP

Si descrive l'approccio proposto da Bouman e Shapiro [10] per la segmentazione di immagini. L'algoritmo di segmentazione è implementato in GRASS nel modulo `i.smap`.

Si introducono i campi casuali multiscalarari di Markof (MSRF) e un criterio di stima chiamato massimo sequenziale a posteriori (SMAP). L'algoritmo non è iterativo ed ha un costo computazionale proporzionale ad MN , dove M è il numero di classi e N il numero di pixel che formano l'immagine.

Il MSRF è composto da una serie di campi casuali progressivi dalla scala più grossolana alla più fine. Ogni campo è assunto dipendente solamente dal campo precedente più grossolano, quindi i diversi campi formano una catena di Markov in scala, che consente di stimare i parametri in modo ricorsivo dalla scala più grossolana a quelle più fine.

C.1 L'approccio per la segmentazione multiscalare

Sia il campo casuale Y l'immagine che deve essere segmentata in distinte regioni che presentano comportamenti statistici differenti. I singoli pixel in Y siano denominati Y_s , dove s è un membro del reticolo bidimensionale di aree S , con cui è stata tassellata l'immagine.

Si ipotizzi che il comportamento spettrale di ciascun pixel osservato, dipenda dal rispettivo valore in X , che è l'immagine classificata. Si ipotizzi anche che le possibili classi siano K , ognuna con il proprio comportamento

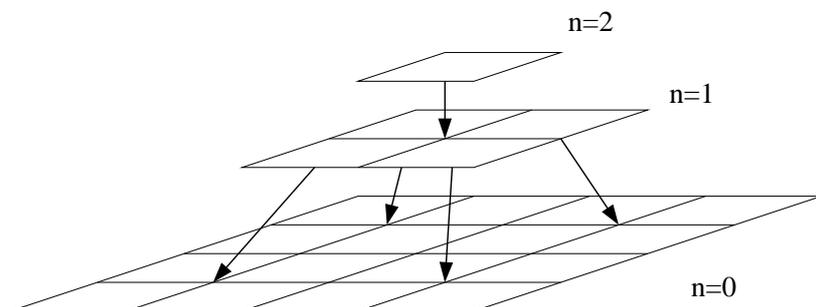


Figura C.1: Struttura a piramide del MSRF.

statistico. La dipendenza dei pixel osservati e le loro etichette è specificata attraverso $p_{y|x}(y | x)$ che è la probabilità condizionata di Y dato X .

Il modello *multi scale random field* (MSRF) si presenta allora come una sequenza di campi casuali dalla scala più grossolana alla più fine, che formano una catena di Markov (figura C.1). Per ciascuna scala n la segmentazione è denotata dal campo casuale $X^{(n)}$ ed i punti del reticolo sono denotati come $S^{(n)}$. In particolare, sia $X^{(0)}$ il campo casuale alla scala più fine con ogni punto corrispondente al singolo pixel dell'immagine. Ogni etichetta nella successiva scala più grossolana corrisponderà ad un gruppo di 4 punti nell'immagine originale. L'ipotesi che la sequenza dei campi casuali formi una catena di Markov implica che la distribuzione di $X^{(n)}$ date tutte le scale più grossolane dipenda solo da $X^{(n+1)}$.

Sotto queste ipotesi si può dimostrare [10] che la distribuzione congiunta di Y e X può essere espressa come:

$$P(Y \in dy, X = x) = p(y | x) \left\{ \prod_{n=0}^{L-1} p(x^{(n)} | x^{(n+1)}) \right\} p(x^{(L)})$$

dove L è la scala più grossolana in X . Questa struttura in scala ha il comportamento isotropico associato con i MRF, ma, grazie alla dipendenza in scala risulta un algoritmo di segmentazione non iterativo e un metodo diretto di stima dei parametri.

C.2 Lo stimatore SMAP

Il problema della minimizzazione del costo medio di misclassificazione consiste nella soluzione della:

$$\hat{x} = \underset{x}{\operatorname{arg\,min}} E[C(X, x) | Y = y] \quad (\text{C.1})$$

dove $E[.]$ è il valore atteso, $C(X, x)$ è il costo per stimare X attraverso x . La scelta del funzionale $C(., .)$ è il punto critico poiché esso determina l'importanza relativa dei diversi errori. Un primo stimatore che si può introdurre è il *MAP* (maximum a posteriori) ed è la soluzione dell'equazione (C.1) quando il funzionale di costo risulta essere:

$$C_{MAP}(X, x) = 1 - \delta(X - x)$$

dove $\delta(X - x)$ è 1 quando $(X = x)$ e 0 altrimenti.

Sviluppando l'equazione (C.1), lo stimatore si può riscrivere:

$$\hat{x} = \underset{x}{\operatorname{arg\,max}} p(X = x | Y = y)$$

Osserviamo che questa forma è del tutto analoga a quella della massima verosimiglianza. Ed anche in questo caso si può esprimere in una forma alternativa utilizzando il teorema di Bayes (B.4) e i logaritmi:

$$\hat{x} = \underset{x}{\operatorname{arg\,max}} (\log p(Y = y | X = x) + \log p(X = x)) \quad (\text{C.2})$$

La generalizzazione dello stimatore MAP nell'ambito dei MSRF si implementa utilizzando in C.1 la seguente espressione per la funzione di costo:

$$C_{MAP}(X, x) = 1 - \delta(X - x) = 1 - \prod_{n=0}^L \delta(X^{(n)} - x^{(n)})$$

Si osservi che una tale funzione attribuisce costo uguale ad 1 per l'erronea classificazione, a qualsiasi scala avvenga.

Un approccio più corretto consiste nell'utilizzare la seguente funzione di costo che assegna costi maggiori all'aumentare della regione misclassificata:

$$C_{SMAP}(X, x) = \frac{1}{2} + \sum_{n=0}^L 2^{n-1} C_n(X, x) \quad (\text{C.3})$$

dove:

$$C_n(X, x) = 1 - \prod_{i=n}^L \delta(X^{(i)} - x^{(i)})$$

Per avere subito un'interpretazione della seguente funzione di costo, si assuma che K sia l'unica scala alla quale si ha che $X^{(K)} \neq x^{(k)}$, ma con $X^{(i)} \neq x^{(i)} \quad \forall i > K$; allora il costo totale risulta $C_{SMAP}(X, x) = 2^K$ e quindi la funzione di costo C_{SMAP} rappresenta la larghezza della più larga regione misclassificata.

Utilizzando l'equazione definita in (C.3) all'interno dell'espressione (C.1) ed utilizzando anche l'ipotesi di catena di Markov tra i campi casuali alle diverse scale, è possibile dimostrare il seguente approccio ricorsivo [10] per la soluzione approssimata del problema:

$$\begin{aligned} \hat{x}^{(L)} &= \arg \max_{x^{(L)}} \log p(y | x^{(L)}) \\ \hat{x}^{(n)} &= \arg \max_{x^{(n)}} \left\{ \log p(y | x^{(n)}) + \log p(x^{(n)} | \hat{x}^{(n+1)}, y) \right\} \end{aligned} \quad (C.4)$$

Si può osservare come la ricorsione sviluppata consiste nello stimare innanzitutto la segmentazione dell'immagine alla scala più grossolana L , applicando lo stimatore MAP ai dati osservati $Y = y$. E in successione la segmentazione delle scale via via più fini si ottiene applicando ancora lo stimatore MAP data la segmentazione alla scala precedente $\hat{x}^{(n+1)}$ e l'immagine y . Ci si riferisce alla struttura ricorsiva nel suo insieme come stimatore MAP sequenziale (SMAP).

I due termini nella seconda espressione in (C.4) giocano lo stesso ruolo della funzione di verosimiglianza e della probabilità a priori nella stima convenzionale di Bayes.

C.3 L'algoritmo di segmentazione

Si presentano di seguito i modelli utilizzati per X e Y , specificando la distribuzione di probabilità condizionata $p(y | x^{(0)})$ e la densità di probabilità per la transizione tra due scale successive, dalla più fine alla più grossolana. L'algoritmo, nell'ambito della segmentazione, ipotizza che i diversi pixel

siano tra loro condizionatamente indipendenti, data la loro classe. Questo implica che la tessitura spaziale delle regioni non viene usata come attributo discriminante. La discriminazione tra classi sarà realizzata unicamente sulle caratteristiche multispettrali. Utilizzando quindi le precedenti assunzioni, la densità di probabilità condizionata per l'immagine assume la forma:

$$p(y | x^{(0)}) = \prod_{s \in S^{(0)}} p(y_s | y_s^{(0)})$$

dove $p(\cdot, k)$ è la densità di probabilità multivariata per un singolo pixel, data la classe k .

In riferimento alla scelta dei modelli per X , deve essere limitata a quelli che presentano due importanti proprietà. La prima è che i pixel in $X^{(n)}$ siano condizionalmente indipendenti ai pixel in $X^{(n+1)}$. La seconda è che ogni pixel $X_s^{(n)}$ dipenda solamente da un intorno di pixel della scala grossolana successiva. Questo insieme di elementi che formano l'intorno di s viene denotato come ∂_s . Quindi la distribuzione di probabilità dalla scala più grossolana alla più fine assume la forma:

$$p(x^{(n)} | x^{(n+1)}) = \prod_{s \in S^{(n)}} p(x_s^{(n)} | x_{\partial_s}^{(n+1)})$$

Anche se i pixel ad una certa scala sono considerati tra loro indipendenti, si può vedere come sono in effetti tra loro molto legati in quanto condividono lo stesso intorno alla scala grossolana successiva.

Si focalizzi ora l'attenzione sulla scelta dell'intorno ∂_s poiché così si definisce la struttura della piramide multiscalare. È stata proposta una soluzione ibrida composta da due tipi di transizioni di scala. La prima è una semplice struttura a *quadtree* che consente l'esatto calcolo della segmentazione SMAP. La seconda è invece una struttura più generale che non consente però il calcolo esatto della stimatore SMAP.

Per il primo modello, visibile in figura C.2, si è proposta la seguente funzione di transizione che rappresenta la probabilità che $X^{(n)}_s$ sia di classe m quando il suo genitore¹ è di classe k .

¹per genitore di s si intende l'intorno alla scala grossolana successiva ∂_s e si denota con $d(s)$. In modo simile si denota l' n -esimo genitore di s con $d^n(s)$, e così $d^{-n}(s)$ sarà l'insieme di tutti i punti che hanno come n -esimo genitore s .

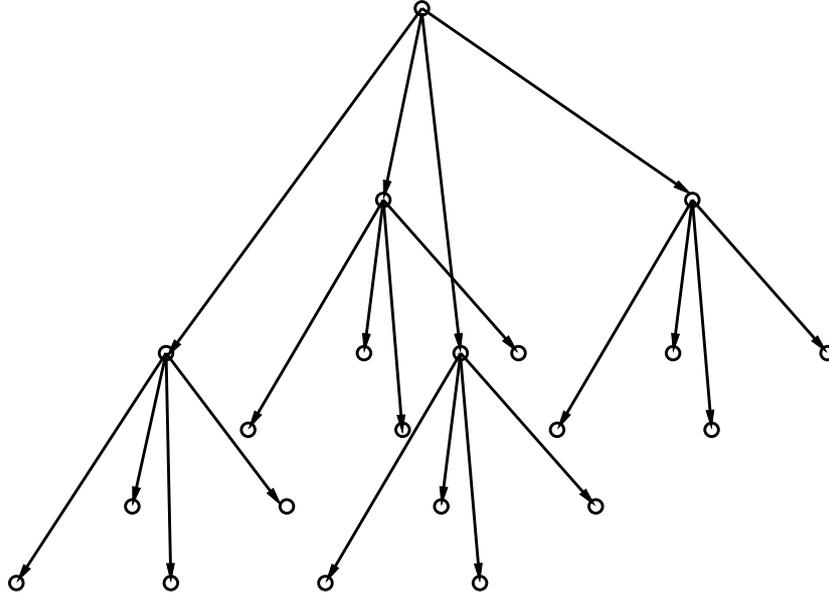


Figura C.2: Struttura a *quadtree* utilizzata per il modello MSRF.

$$p(m | k) = \theta_{n,0} \delta_{m,k} + \frac{1 - \delta_{n,0}}{M} \quad (C.5)$$

Dove il parametro $\theta_{n,0} \in [0, 1]$ che l'etichetta rimanga la stessa dalla scala $n+1$ alla n . Si può poi notare che la probabilità che avvenga un cambiamento di classe è uguale per ogni classe. In genere alle risoluzioni più fini i pixel vicini sono più portati ad avere la stessa classe. Quindi $\theta_{n,0}$ è una funzione che aumenta con la risoluzione. Si osserva infine che la distribuzione dipende dalla particolare scala n attraverso $\theta_{n,0}$, ma non dipende dal particolare pixel s .

Con queste definizioni è stato dimostrato che lo stimatore SMAP rappresentato dall'espressione (C.4) può essere riscritto nella forma seguente:

$$\hat{x}_s^{(n)} = \arg \max_{1 \leq k \leq M} \left\{ l_s^{(n)}(k) + \log p(k | \hat{x}_{\partial s}^{(n+1)}) \right\} \quad (C.6)$$

dove la stima viene eseguita per ciascun pixel e la $l_s^{(n)}$ è la funzione di log-verosimiglianza che deve essere stimata in precedenza attraverso una nuova procedura ricorsiva che va dalle scale più fini alle più grossolane:

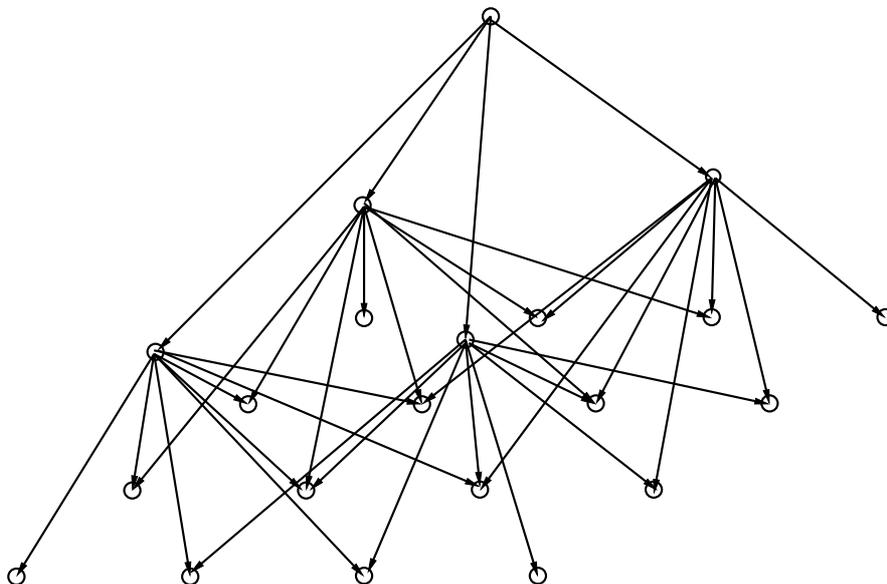


Figura C.3: Struttura a grafo piramidale utilizzata nel modello MSRF.

$$l_s^{(0)}(k) = \log p(y_s | k)$$

$$l_s^{(n+1)}(k) = \sum_{r \in d^{-1}(s)} \log \left\{ \theta_{n,0} \exp\{l_r^{(n)}(k)\} + \frac{1 - \theta_{n,0}}{M} \sum_{m=1}^M \exp\{l_r^{(n)}(m)\} \right\} \quad (C.7)$$

Inconveniente rilevante di questa struttura ad albero è che molte volte i pixel che sono spazialmente adiacenti possono non appartenere allo stesso intorno nella scala grossolana successiva. Per risolvere questa debolezza si utilizza una struttura piramidale (figura C.3).

La funzione di transizione per la struttura piramidale assume la forma:

$$\tilde{p}_{x_s^{(n)} | x_{\partial s}^{(n+1)}}(m | i, j, k) = \frac{\theta_{n,1}}{7} (3\delta_{m,i} + 2\delta_{m,j} + 2\delta_{m,k}) + \frac{1 - \theta_{n,1}}{M} \quad (C.8)$$

Non esiste però, nel caso di struttura piramidale una soluzione semplice per la funzione di verosimiglianza, come per (C.7), quindi non esiste una forma ricorsiva per computare la verosimiglianza dell'immagine, y , data l'etichetta

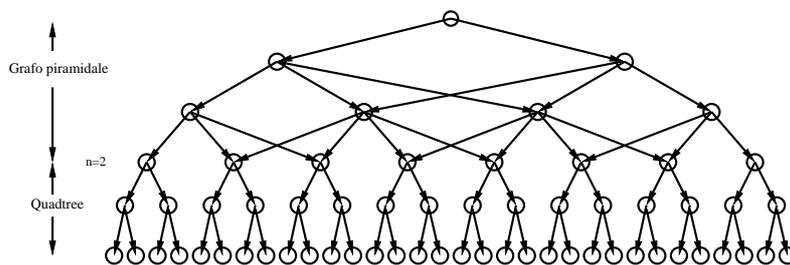


Figura C.4: Equivalente unidimensionale del grafo a struttura ibrida: in per le scale $n > 2$ si utilizza un grafo a struttura piramidale, mentre per le scale $n < 2$ si utilizza una struttura a quadtree.

$s(n)$. Il problema si può risolvere introducendo una struttura ibrida che per computare ad una singola scala, n , utilizzi la struttura ad albero (quadtree) per le scale più fine di n , ed una struttura a grafo per le scale più grossolane C.4. La procedura ricorsiva per lo stimatore SMAP, utilizzando la struttura a grafo ibrida è quindi esprimibile in questa forma approssimata:

$$\begin{aligned} \hat{x}_s^{(L)} &= \arg \max_{1 \leq k \leq M} l_s^{(L)}(k) \\ \hat{x}_s^{(n)} &= \arg \max_{1 \leq k \leq M} \left\{ l_s^{(n)}(k) + \log \tilde{p}(k \mid \hat{x}_{\partial s}^{(n+1)}) \right\} \end{aligned} \quad (C.9)$$

dove \tilde{p} è la funzione di transizione (C.8), e $l_s^{(n)}(k)$ sono calcolati utilizzando la ricorsione (C.9).

C.4 Stima dei parametri

Nelle applicazioni pratiche non si hanno in genere informazioni a priori sull'esatto comportamento della segmentazione. È però possibile determinare direttamente tali informazioni dall'immagine durante la fase stessa di segmentazione. Da calcolare sono i parametri $\theta_{n,0}$ richiesti per il calcolo della funzione di verosimiglianza nella ricorsione dalla scala più fine alla più grossolana e i parametri $\theta_{n,1}$ che sono invece richiesti per implementare lo stimatore SMAP nella ricorsione dalla scala più grossolana alla più fine. Il modello prevede comunque che tutti i parametri vengano stimati durante la

ricorsione dalle scale più grandi alle più fini. Il processo di segmentazione avviene in due fasi complete composte dai seguenti passi:

1. Eseguire la ricorsione dalle scale fini a quelle grossolane utilizzando come valore dei parametri $\theta_{n,0} = 1$
2. Stimare i parametri $\theta_{n,0}$ e $\theta_{n,1}$ durante la ricorsione inversa.
3. Eseguire nuovamente la ricorsione diretta utilizzando i parametri precedentemente stimati.
4. Ristimare i parametri $\theta_{n,1}$ durante la ricorsione finale inversa.

La stima nei passi 2 e 4 è fatta sequenzialmente per ogni scala, e non viene eseguita sull'intera immagine, ma su sottoimmagini. Questo, in particolare per le immagini ad alta risoluzione, consente di ridurre il costo computazionale dell'algoritmo; in tali immagini il numero dei pixel è elevato e perciò il sottocampionamento non determina un calo di accuratezza. È poi possibile agire sulla dimensione di questa sottoimmagine per modificare leggermente il risultato della segmentazione.

I parametri sono stimati calcolando il massimo della funzione di verosimiglianza dell'immagine y data la precedente segmentazione alla scala più grossolana, $\hat{x}^{(n+1)}$. La funzione è massimizzata usando l'algoritmo di *aspettazione massimizzazione* EM.

Appendice D

Stima delle densità di probabilità

Consideriamo in questa sezione il problema di modellizzare una funzione densità di probabilità a partire da un numero finito di dati. Nell'ambito di una procedura di classificazione supervisionata si deve procedere alla stima della probabilità $p(x | C_i)$ utilizzando l'insieme di *training* $D = \{(x_i, c(x_i))\}$

Si distinguono in genere tre diversi tipi di modelli per la stima della funzione di densità. I primi sono i *modelli parametrici* che adottano una precisa forma funzionale e un numero fissato di parametri. I secondi sono invece i *metodi non parametrici* che non utilizzano a priori nessuna forma funzionale ma la ricavano dai dati. Infine la terza classe di modelli sono quelli *semi parametrici* che sfruttano gli aspetti positivi dei primi due metodi.

D.1 Metodi parametrici

La particolarità dei metodi parametrici per la stima di $p(x | C_i)$ è quella di supporre nota la forma funzionale della distribuzione di probabilità, mentre l'incertezza riguarda l'entità di uno o più parametri caratteristici della distribuzione. L'utilizzo di una funzione a priori garantisce che la complessità del modello non dipenda dal numero di dati, ma è anche il punto debole dell'approccio parametrico, in quanto non potrebbe adattarsi correttamente ai dati specifici.

Una prima forma funzionale utilizzabile è la distribuzione *gaussiana* o *normale*. In uno spazio d -dimensionale, la funzione di densità normale multivariata può essere scritta nella forma

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}, \quad (\text{D.1})$$

dove $\boldsymbol{\mu}$ è un vettore d -dimensionale, Σ è la matrice di covarianza di dimensione $d \times d$, e $|\Sigma|$ è il determinante di Σ .

Il fattore $1/((2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}})$ assicura che $\int_{\mathbb{R}^d} p(\mathbf{x}) = 1$. La funzione densità qui definita contiene i parametri $\boldsymbol{\mu}$ e Σ che sono rispettivamente la *media* di \mathbf{x} ,

$$\boldsymbol{\mu} = E[\mathbf{x}] \quad (\text{D.2})$$

e la *covarianza*,

$$\Sigma = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T]. \quad (\text{D.3})$$

Visto che ci sono d elementi indipendenti in $\boldsymbol{\mu}$, la funzione di densità è completamente specificata una volta che i $d(d+3)/2$ parametri sono stati determinati. La quantità

$$\Delta^2 = (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}), \quad (\text{D.4})$$

che appare nell'argomento dell'esponenziale della (D.1), è chiamata *Distanza di Mahalanobis* di \mathbf{x} da $\boldsymbol{\mu}$. Le superfici per cui è costante il valore della densità di probabilità, espressa secondo la (D.1), sono le iperellissoidi su cui Δ^2 è costante. Gli assi principali di questi iperellissoidi sono gli autovettori u_i di Σ che, come abbiamo visto, esprimono la varianza lungo le direzioni principali.

È spesso conveniente considerare una forma semplificata della distribuzione gaussiana, in cui la matrice di covarianza è diagonale,

$$(\Sigma)_{ij} = \delta_{ij} \sigma_j^2. \quad (\text{D.5})$$

Si riducono così il numero di parametri indipendenti a $2d$ e si allineano gli assi principali degli ellissoidi, che rappresentano il contorno a densità costante, agli assi delle coordinate.

D.1.1 La funzione di massima verosimiglianza

Dopo aver scelto la forma funzionale $p(C_i x; \theta)$ per descrivere la distribuzione di probabilità dei dati, è necessario stimare i valori dei parametri $\theta = (\theta_1, \dots, \theta_M)^T$ a partire dai dati stessi. A riguardo viene descritta ora la tecnica di *massima verosimiglianza* (maximum log-likelihood), che consente di ricavare i valori dei parametri dalla massimizzazione di una funzione detta appunto *funzione di massima verosimiglianza*, calcolata sui dati di training.

Definiamo l'insieme di training come un insieme finito $\chi = \{x^1, \dots, x^N\}$ Si definisce allora funzione di massima verosimiglianza la densità di probabilità congiunta dei vettori dell'insieme χ ; ipotizzando che le diverse osservazioni $\{x^1, \dots, x^N\}$ siano tra loro indipendenti, assume la forma:

$$p(\chi; \theta) = \prod_{n=1}^N p(x^n; \theta) := \mathcal{L}(\chi, \theta) \quad (\text{D.6})$$

I parametri si stimano massimizzando la funzione $\mathcal{L}(\chi, \theta)$:

$$\hat{\theta}(\chi) = \arg \max_{\theta} \mathcal{L}(\theta; \chi)$$

Si suppone cioè che i dati osservati abbiano la più alta probabilità di accadimento.

In pratica è più conveniente considerare il logaritmo negativo della funzione di massima verosimiglianza (*negative log-likelihood*), sia per ragioni analitiche, sia perché può essere visto direttamente come funzione d'errore da minimizzare:

$$E = -\log \mathcal{L}(\chi, \theta) = -\sum_{n=1}^N \log p(x^n; \theta) \quad (\text{D.7})$$

Infatti, essendo il logaritmo negativo una funzione monotona decrescente, trovare il minimo di E equivale a massimizzare \mathcal{L} .

Per evitare che la matrice di covarianza sia singolare, sono richiesti almeno $N+1$ esempi, dove N è la dimensione dell'immagine multispettrale. Questo consente di calcolare l'inversa utilizzata nella funzione discriminante (B.10). Si è però osservato [6] che per ottenere una valida stima è necessario utilizzare almeno $10N$ dati di esempio per classe con valore consigliato di $100N$.

D.2 Metodi semiparametrici

I metodi parametrici pur avendo il difetto di servirsi di funzioni di densità della forma specifica, che spesso descrivono con difficoltà le distribuzioni dei dati, hanno però il vantaggio di essere efficienti e veloci [5]. I metodi non parametrici utilizzano invece funzioni appropriate, ma i loro modelli contengono un numero di variabili che aumenta con l'aumentare dei dati di *training*. I modelli di tipo *semi parametrico* si propongono di combinare i vantaggi di entrambi i metodi sopracitati: non si limitano a considerare una forma funzionale specifica e mantengono l'efficienza di un modello il cui numero di parametri aumenta con l'aumentare della complessità del modello e non con il numero di dati. Esempio di metodo semi parametrico, del quale ci occuperemo in questo lavoro è il *modello di mistura* ed in particolare il modello di mistura di gaussiane.

D.2.1 I modelli di mistura

Nella classe dei modelli di mistura si considerano modelli in cui la funzione di densità è rappresentata da una combinazione lineare di funzioni chiamate *basis functions*¹ [9], che sono generalmente funzioni di densità parametrizzate e dove il numero M delle funzioni, componenti della mistura, è trattato come un parametro del modello ed è generalmente inferiore al numero N dei dati da modellizzare. La stima delle densità di probabilità $p(\mathbf{x} | C_i)$ utilizzando un modello di mistura si può scrivere:

$$p(\mathbf{x} | C_k) = \sum_{j=1}^{M_k} p_k(\mathbf{x}|j)P_k(j), \quad (\text{D.8})$$

con

$$\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbb{R}^n. \quad (\text{D.9})$$

¹Le *basis functions* entrano in gioco nei problemi di classificazione a proposito delle *funzioni discriminanti*. È grazie a queste funzioni discriminanti $y_1(\mathbf{x}), \dots, y_c(\mathbf{x})$ che viene fatta la classificazione. In poche parole il vettore di input \mathbf{x} è assegnato alla classe C_k se $y_k(\mathbf{x}) > y_j(\mathbf{x})$ per ogni $j \neq k$. Al fine di minimizzare la probabilità di misclassificazione, si sceglierà $y_k(\mathbf{x}) = P(C_k|\mathbf{x}) = p(\mathbf{x}|C_k)P(C_k)$. Le funzioni basis generalizzano le funzioni discriminanti permettendo di considerare un numero maggiore di gruppi decisionali. La nuova funzione discriminante appare nella forma $y_k(\mathbf{x}) = \sum_{j=1}^M w_{kj}\phi_j(\mathbf{x}) + w_{k0}$, dove le $\phi_j(\mathbf{x})$ fissate sono indipendenti dai dati e sono proprio le funzioni *basis*.

Una rappresentazione di questo tipo è chiamata *mistura di distribuzioni* della classe k , i coefficienti $P_k(j)$ sono i *parametri della mistura* e $p_k(\mathbf{x}|j)$, che sono funzioni di densità, sono dette componenti per la mistura per la classe k . Si può notare che c'è una certa somiglianza tra la D.8 e la probabilità non-condizionata B.5. $P_k(j)$ rappresenta infatti la probabilità a priori che il dato x appartenente alla classe k sia stato generato dalla componente j -esima della mistura. Le probabilità a priori soddisfano $\sum_{j=1}^{M_k} P_k(j) = 1$, con $0 \leq P_k(j) \leq 1$. Allo stesso modo le componenti della funzione densità $p_k(\mathbf{x}|j)$ sono state normalizzate in modo tale che $\int p_k(\mathbf{x}|j) d\mathbf{x} = 1$, e possono così essere trattate come *densità di probabilità condizionata*.

Ad ogni dato della classe k viene associata la probabilità di essere generato dalla componente j -esima e questo vale per ogni j che varia all'interno dell'intervallo $\{1, \dots, M_k\}$, dove M_k è il numero delle componenti della mistura per la classe k . Questa probabilità è proprio la $p_k(\mathbf{x}|j)$, che quindi soddisfa $\int p_k(\mathbf{x}|j) d\mathbf{x} = 1$.

Una volta visto il collegamento tra le *probabilità a priori* e le *densità di probabilità condizionate*, si può introdurre il concetto di *probabilità a posteriori* che, utilizzando il *Teorema di Bayes* (B.4), può essere espressa nella seguente forma:

$$P_k(j|\mathbf{x}) = \frac{p_k(\mathbf{x}|j)P_k(j)}{p_k(\mathbf{x})}, \quad (\text{D.10})$$

dove $p_k(\mathbf{x})$ è data dalla (D.8).

La probabilità a posteriori inoltre soddisfa

$$\sum_{j=1}^{M_k} P_k(j|\mathbf{x}) = 1 \quad (\text{D.11})$$

e rappresenta la probabilità che la componente j -esima della mistura sia responsabile della generazione del punto \mathbf{x} .

Consideriamo un esempio di modello di mistura per ciascuna classe k in cui le componenti di densità sono funzioni di distribuzione gaussiane, le cui matrici di covarianza sono:

$$\Sigma_j^k = (\sigma_j^k)^2 I \quad (\text{D.12})$$

con I che rappresenta la matrice identità.

La densità di probabilità condizionata per un modello di questo tipo è

$$p_k(\mathbf{x}_k|j) = \frac{1}{(2\pi)^{d/2} |(\sigma_j^k)^2 I|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}_k - \mu_j^k)^T ((\sigma_j^k)^2 I)^{-1} (\mathbf{x}_k - \mu_j^k)\right\} \quad (\text{D.13})$$

e poiché

$$|(\sigma_j^k)^2 I|^{1/2} = ((\sigma_j^k)^{2d})^{1/2} = ((\sigma_j^k)^2)^{d/2} \quad (\text{D.14})$$

possiamo scrivere

$$p_k(\mathbf{x}|j) = \frac{1}{(2\pi(\sigma_j^k)^2)^{d/2}} \exp\left\{-\frac{\|\mathbf{x} - \mu_j^k\|^2}{2(\sigma_j^k)^2}\right\}. \quad (\text{D.15})$$

I parametri di questo modello sono: $P_k(j)$, μ_j^k , σ_j^k , con $j \in \{1, \dots, M_k\}$. Si può dimostrare che utilizzando la regola della massima verosimiglianza si ottiene la seguente stima per i parametri del modello a mistura di gaussiane:

$$\hat{\mu}_j^k = \frac{\sum_{n=1}^N P_k(j|\mathbf{x}^n) \mathbf{x}^n}{\sum_{n=1}^N P_k(j|\mathbf{x}^n)}, \quad (\text{D.16})$$

$$\hat{\sigma}_j^{k2} = \frac{\sum_{n=1}^N P_k(j|\mathbf{x}^n) \|\mathbf{x}^n - \hat{\mu}_j^k\|^2}{\sum_{n=1}^N P_k(j|\mathbf{x}^n)}. \quad (\text{D.17})$$

$$\hat{P}_k(j) = \frac{1}{N} \sum_{n=1}^N p_k(j|\mathbf{x}^n)$$

Appendice E

Classificazione non supervisionata

La *classificazione non supervisionata* è la tecnica che porta ad assegnare ad ogni dato dell'insieme di *training* una classe non predefinita. Le classificazioni trovate attraverso metodi non supervisionati sono detti tecniche di clustering [5].

Un vantaggio può essere il fatto che è possibile etichettare solo le classi di interesse, mentre le altre servono solo per avere migliori prestazioni nell'algoritmo di classificazione rimanendo etichettate in modo simbolico. In genere l'identificazione delle classi di interesse è più facile quando si ha già la separazione spaziale delle varie classi.

Nel problema della classificazione di immagini questa tecnica è conveniente per classi allungate, quali fiumi e strade per le quali è difficile operativamente definire un insieme di training. Un grosso svantaggio della procedura di clustering è il notevole consumo di tempo, legato al fatto che generalmente si deve elaborare l'intera immagine. Un altro aspetto da considerare è la difficoltà nell'inserire la probabilità a priori dei membri delle classi.

E.1 Metriche e criteri di clustering

Il *Clustering* comporta il raggruppamento dei pixel in uno spazio multispettrale. I pixel che appartengono ad un particolare *cluster* sono perciò spettralmente simili. Per quantificare questa relazione sarà necessario dispor-

re di misure di similarità. La similarità si definisce sulla base di metriche che molto frequentemente si rifanno alla misura di semplici distanze nello spazio multispettrale. Fra queste le più utilizzate sono le distanze euclidee:

$$\begin{aligned}
 d(x, y) &= \sqrt{\sum_{i=1}^n (x_i - y_i)^2} & (E.1) \\
 d(x, y) &= \|x_i - y_i\| \\
 d(x, y) &= \max_{1 \leq i \leq n} \{|x_i - y_i|\}
 \end{aligned}$$

Usando la misura di distanza dovrebbe essere possibile determinare i cluster nei dati. Comunque spesso ci saranno diversi tipi di clusterizzazione possibili, e così è desiderabile avere misure che stabiliscano quale delle clusterizzazioni è la migliore. Un comune indicatore della qualità di clusterizzazione è la somma degli scarti quadratici (SSE sum of squared error).

$$SSE = \sum_{c_i} \sum_{x \in C_i} ((\mathbf{x} - \mathbf{m}_i)^t (\mathbf{x} - \mathbf{m}_i)) \quad (E.2)$$

dove \mathbf{m}_i è la media del i -esimo cluster e $\mathbf{x} \in \mathbf{C}_i$ è il pixel appartenente al cluster i -esimo, mentre si osservi che la sommatoria esterna è fatta sopra tutti i cluster.

È interessante notare che SSE ha un minimo teorico in zero, quando tutti i cluster contengono un solo dato. Se viene usato un metodo iterativo per cercare i cluster di una classe spettrale, allora esiste, almeno dal punto di vista teorico, l'esistenza di una soluzione.

Come semplice esempio di questo citiamo l'algoritmo *migrating means*. È un algoritmo iterativo che consiste nello scegliere a priori il numero massimo di cluster da individuare nell'immagine. L'algoritmo sceglierà a caso nello spazio i punti più probabili per i centri dei singoli cluster e classificando i successivi nel cluster meno distante. Dopo che tutti i pixel sono classificati ci sarà un ricalcolo dei nuovi centri dei cluster, quindi la procedura dovrà essere iterata.

Una volta che la clusterizzazione è completata, oppure quando lo si ritenga più adatto, i cluster possono essere esaminati per accertare se:

- alcuni cluster contengono così pochi punti da essere insignificanti.

- alcuni cluster sono talmente vicini tra loro che non ha senso considerarli separati.

Per queste ultime due scelte si dovranno stabilire a priori i valori di soglia per il numero minimo di pixel e per la distanza minima tra due cluster.

Per consentire all'utente la scelta del numero di cluster e di decidere quali cluster considerare è possibile adottare la *clusterizzazione agglomerativa gerarchica*. La procedura comincia assumendo che tutti i pixel formino dei cluster con un solo elemento, questi poi saranno fusi con i cluster più vicini utilizzando misure di distanze. La procedura avanza fino a quando tutti i pixel cadono in un singolo cluster. Un importante aspetto del metodo è che tutto il cammino di fusione è visualizzato in un dendrogramma che evidenzia a quale distanza tra i centri i diversi cluster sono fusi assieme. L'asse verticale si può ritenere una buona misura della stabilità di un cluster. In genere questi tipi di alberi non vengono però usati per immagini con grosse quantità di dati.

Bibliografia

- [1] W.N. Venables and B.D. Ripley. *Modern Applied Statistics with S-Plus*. Springer-Verlag, 2nd edition, 1997.
- [2] R. Laurini and D. Thompson. *Fundamentals of Spatial Information System*. The Apic Series. Academic Press, 1994.
- [3] M. Ciolli. Gis: Stato dell'arte e prospettive di utilizzo in campo applicativo e scientifico. *Dendronatura*, (2):7–16, 1999.
- [4] E. Arnoldi. Il sistema informativo ambiente e territorio della provincia autonoma di trento con particolare riferimento al settore agricolo. *Dendronatura*, (2):85–89, 1999.
- [5] B.D. Ripley. *Pattern Recognition and Neural Networks*. Cambridge University Press, 1996.
- [6] John A. Richards. *Remote Sensing Digital Image Analysis: an introduction*. Springer-Verlag, second, revised and enlarged edition edition, 1993.
- [7] V. Ingegnoli. *Fondamenti di Ecologia del Paesaggio*. Città Studi, 1993.
- [8] H. Jurgens O. Peitgen and D. Saupe. *Chaos and Fractals*. Springer-Verlag, 1992.
- [9] C.M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, New York, 1995.
- [10] C.A. Bouman and M. Shapiro. A multiscale random field model for bayesian image segmentation. *IEEE Transactions on Image Processing*, 3(2):162–177, March 1994.

- [11] M.G. Turner and R.H. Gardner, editors. *Quantitative Methods in Landscape Ecology*. Springer-Verlag, 1991.
- [12] V. Ingegnoli and S. Pignatti. *L'Ecologia del Paesaggio in Italia*. Città Studi Edizioni, 1996.
- [13] R.R. Sokal and F.J. Rohlf. *Biometry*. W.H. Freeman and Company, New York, 1995.
- [14] U.N.G. von Blotzheim. *Tetraonidi*. Sempach, 1985. Stazione Ornitologica Svizzera.
- [15] P.F. De Franceschi. Gestione venatoria del fagiano di monte (o gallo forcello). *Gestione venatoria di Tetraonidi, Coturnice e Lepre Alpina*, 1997. Atti del convegno del Comitato di Gestione del Comprensorio Alpino di Caccia di Sondrio.
- [16] R.T.T. Forman and M. Gordon. *Landscape Ecology*. J. Wiley & Sons, 1986.
- [17] M.G. Turner R.H. Gardner, B.T. Milne and R.V. O'Neill. Neutral models for the analysis of broad-scale landscape pattern. *Landscape Ecology*, (1):19–28, 1987.
- [18] D.L. Urban T.H. Keitt and B.T. Milne. Detecting critical scales in fragmented landscape. *Conservation Ecology [online]*, 1(1), 1997. Available from Internet. URL: <http://www.consecol.org/vol1/iss1/art4>.
- [19] W. L. Baker. *The r.le programs*. University of Wyoming, August 1994.
- [20] J. Riassen. A universal prior for integers and estimation by minimum description length. *The Annals of Statistics*, 11(2):416–431, June 1983.
- [21] Olshen Brieman, Friedman and Stone. *Classification and Regression Trees*. Wadsworth, California, 1984.
- [22] B. Efron and R.J. Tibshirani. Cross-validation and the bootstrap: Estimating the error rate of a prediction rule. Technical Report 176, Stanford Univ., MAY 1995.