



Università
Ca' Foscari
Venezia

**Dottorato di ricerca
in Scienze Ambientali
Scuola di dottorato in Scienze e tecnologia
Ciclo XXIV
(A.A. 2010 - 2011)**

***Caratterizzazione e identificazione di banchi di
specie pelagiche mediante l'uso di descrittori
strutturali e parametri ambientali***

**SETTORE SCIENTIFICO DISCIPLINARE DI AFFERENZA: BIO/07
Tesi di dottorato di Marta D'Elia, matricola 955636**

Coordinatore del Dottorato

Prof. Bruno Pavoni

Tutore del dottorando

Dott. Stefano Malavasi

Dott. Bernardo Patti

Co-tutore del dottorando

Dott. Angelo Bonanno

*Il mare non cambia mai ed il suo operare,
per quanto ne parlino gli uomini,
è avvolto nel mistero.
Joseph Conrad [1899]*

Indice

1	Introduzione	11
1.1	Survey acustici per la valutazione della biomassa e della distribuzione di stock ittici	11
1.2	Studi precedenti sulla classificazione delle eco-tracce	13
1.3	Uso di dati a multifrequenza per l'attività di classificazione delle specie . . .	15
1.4	Obiettivo e struttura della tesi	17
2	Materiali e metodi	21
2.1	Processamento di dati a singola frequenza per l'estrazione di banchi di pesce e delle aggregazioni planctoniche	21
2.2	Processamento dei dati a multifrequenza per la caratterizzazione e l'identificazione dei banchi di pesci	23
2.2.1	Raccolta dei dati acustici e biologici per la costruzione del learning sample	23
2.2.2	Processamento dei dati acustici per l'estrazione delle informazioni sui banchi di pesce da dati a multifrequenza	27
2.2.3	Metodo acustico per la classificazione delle proprietà del fondo marino	34
2.2.4	Procedura per l'estrazione delle informazioni relative al fondo	37
2.2.5	Raccolta e selezione di dati idrologici	39
2.2.6	Analisi statistiche: statistica descrittiva e classification tree	40
3	Risultati	43
3.1	Comportamento e distribuzione di banchi di pesci pelagici in relazione alla disponibilità di cibo nel Canale di Sicilia	43
3.2	Caratterizzazione dei banchi con dati a multifrequenza	50
3.2.1	Caratterizzazione batimetrica e morfologica dei banchi di pesce . . .	50
3.2.2	Proprietà energetiche di acciughe e sardine alle diverse frequenze . .	56
3.2.3	Proprietà energetiche di acciughe e sardine per range di profondità .	57
3.2.4	Modelli di classificazione ad albero con variabili predittive strutturali	58
3.2.5	Confronto dei risultati provenienti da modelli di classificazione ad albero con le stime di biomassa	67
3.2.6	Modelli di classificazione ad albero costruiti su variabili ambientali e parametri strutturali del banco.	76

4	Discussioni	81
5	Conclusioni	91
	Bibliografia	93
	Ringraziamenti	103
	Estratto per riassunto della tesi di dottorato	105

Elenco delle figure

1.1	Effetto della frequenza sull'intensità di backscatter in volume (Sv) dei pesci, dei gruppi zooplanctonici e delle strutture fisiche (Lavery <i>et al.</i> , 2007) . . .	16
1.2	Modello predittivo del backscatter strength di un'aringa Atlantica di 20-25 cm di taglia ad una profondità di 112 m. La linea continua spessa riporta il modello derivato dalla combinazione dell'effetto della vescica natatoria con l'effetto del corpo. La linea tratteggiata raffigura il modello usato per rappresentare l'effetto della vescica natatoria, mentre la linea continua sottile raffigura il modello usato per rappresentare l'effetto del corpo del pesce. (Fässler <i>et al.</i> , 2007)	17
2.1	Area di studio. Mostra I siti, lungo il tracciato acustico giornaliero del survey ANCHEVA1998, in cui banchi di pesce e plankton sono compresenti. È riportata anche la separazione dell'area di studio in tre zone, indicate come Zona 1, Zona 2, Zona 3.	22
2.2	Area di studio. In rosso sono visualizzate le catture monospecifiche selezionate e i transetti costa-largo effettuati durante il survey ANCHEVA2009. La lettera indica la specie predominante nella cala: A=acciuغه, O= altre specie pelagiche, S=sardine, D=demersali.	24
2.3	Struttura dell'algoritmo "fish school filter" creato con il modulo "Virtual Variable" del software Sonardata Echoview.	27
2.4	Ecogramma risultante dalla somma delle tre frequenze con il minimo valore soglia impostato (-210 dB).	29
2.5	Applicazione di un filtro che elimina il rumore al sopra della linea superficiale e al di sotto del fondo.	29
2.6	Applicazione di un filtro mediano con elemento strutturale 3x3.	30
2.7	Applicazione di un filtro di dilatazione con elemento strutturale 5x5.	30
2.8	Applicazione del nuovo filtro soglia a -175dB e visualizzazione delle ecotracce candidate ad essere considerate banchi.	31
2.9	Ecogramma finale con le ecotracce isolate e soglia minima di visualizzazione di -70 dB.	31

2.10	Ampiezza del segnale registrato da un echosounder in funzione del tempo. La coda del primo eco è il risultato della riflessione obliqua ed è indice di rugosità del fondo, mentre l'intero secondo eco risultante da una singola riflessione in superficie e una doppia riflessione al fondo, fornisce informazioni sulle proprietà di durezza del fondo (ICES report 2007).	35
2.11	Interazione di un ping o impulso acustico con il fondo. Il lato sinistra dell'immagine fornisce una visione laterale dell'energia riflessa da un fondo orizzontale raggiunto da un impulso acustico, mentre il lato destro mostra una sezione trasversale dell'impulso a contatto con il fondo in un determinato istante. Nella sezione centrale, la parte perpendicolare al trasduttore dell'impulso acustico raggiunge il fondo ad un tempo t_1 e insonifica un'area circolare, mentre nella sezione in basso la parte esterna dell'impulso acustico raggiunge il fondo ad un tempo t_2 , insonificando un'area ad anello	36
2.12	Geometria del segnale di ritorno dal fondo. d_{θ_a} e d_{θ_b} sono le profondità delle riflessioni oblique corrispondenti agli angoli θ_a θ_b mentre d_a e d_b sono gli intervalli costanti di integrazione del secondo eco (Siwabessy, 2001).	39
3.1	Effetto della densità e della biomassa planctonica sulla variabile risposta: (a) biomassa dei banchi di pesce vs. $\ln(\text{biomassa di plancton})$; (b) densità dei banchi di pesce vs. $\ln(\text{biomassa di plancton})$; (c) densità dei banchi di pesce vs. $\ln(\text{densità di plancton})$. La biomassa e la densità di plancton si riferiscono alle aggregazioni individuate nell'intera colonna d'acqua. La linea continua indica i valori stimati dal GAM. La linea tratteggiata rappresenta l'intervallo di confidenza al 95% stimato con il metodo bootstrap.	45
3.2	Effetto della densità e della biomassa planctonica sulla variabile risposta in Zona 1: (a) biomassa dei banchi di pesce vs. $\ln(\text{biomassa del plancton profondo})$; (b) densità dei banchi di pesce vs. $\ln(\text{biomassa del plancton profondo})$; (c) densità dei banchi di pesce vs. $\ln(\text{densità di plancton profondo})$. La biomassa e la densità di plancton si riferiscono alle aggregazioni con distanza dal fondo inferiore a 5 m. La linea continua indica i valori stimati dal GAM. La linea tratteggiata rappresenta l'intervallo di confidenza al 95% stimato con il metodo bootstrap.	47
3.3	Effetto della densità e della biomassa planctonica sulla variabile risposta in Zona 3: (a) biomassa dei banchi di pesce vs. $\ln(\text{biomassa del plancton})$; (b) densità dei banchi di pesce vs. $\ln(\text{biomassa del plancton})$; (c) densità dei banchi di pesce vs. $\ln(\text{densità di plancton})$. La biomassa e la densità di plancton si riferiscono alle aggregazioni nell'intera colonna d'acqua. La linea continua indica i valori stimati dal GAM. La linea tratteggiata rappresenta l'intervallo di confidenza al 95% stimato con il metodo bootstrap.	48
3.4	Banchi di pesce (in grigio scuro) e aggregazioni di plancton (in grigio chiaro) acusticamente individuate in Zona 3.	49

3.5	Valori medi e intervalli di confidenza al 95% dei parametri morfologici, batimetrici ed energetici estratti da banchi di acciughe, sardine, OPS e demersali: (a) Lunghezza (Corrected_length), altezza (Corrected_thickness) ed elongazione (Elong) corretti secondo Diner (2001); (b) Perimetro (Corrected_Perimeter), rettangolarità (Rectang) e dimensione frattale (Fractaldim) corretti secondo Diner (2001).	53
3.6	Valori medi e intervalli di confidenza al 95% dei parametri morfologici, batimetrici ed energetici estratti da banchi di acciughe, sardine, OPS e demersali: (a) volume del banco rispetto a un cilindro (X3D_school volume) e volume rispetto al numero di ping (Beam_volume_sum); (b) area (Corrected_area) ed area del banco rispetto a un cilindro (X3D_school area).	54
3.7	Valori medi e intervalli di confidenza al 95% dei parametri morfologici, batimetrici ed energetici estratti da banchi di acciughe, sardine, OPS e demersali: (a) variabilità della superficie di un banco (Unevenness2) e variabilità della superficie del banco normalizzata rispetto al numero di ping (Unevenness1); (b) profondità media del banco (Depth_mean) e distanza dal fondo (Mean_dist_bottom).	55
3.8	(a) Valori medi di Sv per la sardina (cerchio blu) e per l'acciuga (cerchio rosso) a 38 e 120 kHz. (b) la differenza in decibel relativa alla frequenza di 38 kHz per la sardina (cerchio blu) e per l'acciuga (cerchio rosso). I triangoli con la punta verso l'alto rappresentano il quantile 0.975, mentre i triangoli con la punta invertita rappresentano il quantile 0.025; è quindi riportata la dispersione del 95% dei dati attorno al valore medio.	56
3.9	(a) differenza in decibel relativa alla frequenza di 38 kHz per giovanili di sardina (cerchio blu) e acciuga (cerchio rosso). (b) differenza in decibel relativa alla frequenza di 38 kHz per adulti di sardina (cerchio blu) e acciuga (cerchio rosso). I triangoli con la punta verso l'alto rappresentano il quantile 0.975, mentre i triangoli con la punta invertita rappresentano il quantile 0.025; è quindi riportata la dispersione del 95% dei dati attorno al valore medio.	57
3.10	Valori medi di $dB120$ dei banchi con profondità comprese tra 40-60 m (a) e 60-80 m (b), identificati come sardine (in blu) e acciughe (in rosso). I triangoli con la punta verso l'alto rappresentano il quantile 0.975, mentre i triangoli con la punta invertita rappresentano il quantile 0.025.	58
3.11	Modello di classificazione ad albero costruito su 26 predittori.	60
3.12	Parametro di complessità (cp, in ascissa) correlato alla taglia dell'albero (size of tree) e l'errore medio di predizione (in ordinata). L'errore medio è il risultato di una cross-validation effettuata per differenti valori di cp. La barra verticale indica la variazione (dev. St) nell'errore predetto. La linea orizzontale tratteggiata si ottiene sommando all'errore predetto dall'albero di dimensioni maggiori, una volta la sua deviazione standard (regola "1-SE").	61
3.13	Modello di classificazione ottimale individuato dalla tecnica di <i>pruning</i> . . .	62

3.14	Modello ottimale costruito escludendo dai 26 predittori la profondità del banco.	63
3.15	Modello ottimale costruito sulle sole variabili energetiche.	64
3.16	Modello ottimale costruito sulle sole variabili morfologiche.	65
3.17	Distribuzione delle capacità di classificazione di trenta differenti alberi per i 4 gruppi di specie: (a) sardine, (b) acciughe, (c) OPS e (d) demersali.	67
3.18	Albero di classificazione con la migliore capacità di discriminazione per tutte le specie (MaxRate).	68
3.19	Albero di classificazione del modello con una capacità di discriminazione superiore al 75% per acciughe e sardine (MaxRateSpecies).	69
3.20	Albero di classificazione con la maggiore capacità di discriminazione per tutte le specie e per le specie target (MaxRate o MaxRateSpecies).	71
3.21	Il grafico riporta la relazione tra PRC_NASC per miglio estratto dai banchi (in ascissa) e il NASC per miglio stimato con la metodologia tradizionale (in ordinata).	72
3.22	Porzione del transetto 20-21. Le linee orizzontali rappresentano la suddivisione in miglia nautiche. Le frecce blu indicano il segnale corrispondente a pesce disperso nella colonna d'acqua, mentre la freccia rossa indica la traccia acustica prodotta da banchi di pesce.	75
3.23	Porzione del transetto 57-58. Le linee orizzontali rappresentano la suddivisione in miglia nautiche. La freccia rossa indica la traccia acustica prodotta da banchi di pesce localizzati sul fondo.	76
3.24	Modello MaxRate stimato usando come predittori variabili morfologiche, batimetriche energetiche e parametri della colonna d'acqua.	77
3.25	Distribuzione delle capacità di classificazione di trenta differenti alberi per i 4 gruppi di specie: (a) sardine, (b) acciughe, (c) demersali e (d) OPS.	78
3.26	Modello MaxRate stimato usando come predittori variabili morfologiche, batimetriche, energetiche del banco.	79
3.27	Modello MaxRate stimato utilizzando le variabili strutturali e la variabile durezza del fondo tra i parametri ambientali.	80

Elenco delle tabelle

2.1	La tabella riporta il modello di echosounder e le frequenze dei trasduttori usati in fase di acquisizione. Ad ogni trasduttore corrisponde una potenza trasmessa e un'ampiezza del cono acustico.	23
2.2	Elenco delle cale selezionate con le relative informazioni riguardanti: la campagna e l'anno di pescata (survey), il numero della cala (Trawl), la specie presente (species), la percentuale in peso con cui la specie è stata catturata (Amount %), il tempo di inizio della cala (Start Time), la durata della cala (duration), la distanza della rete dalla base al fondo (Dist ft_bot), la distanza della rete dalla testa al fondo (dist hd_bot), la profondità del fondale (bottom), il peso della cattura (Catch), la lunghezza media delle specie target (Species_Msize).	26
2.3	Parametri estratti con l'algoritmo SHAPE di Echoview o da essi derivati tramite script in R.	33
3.1	Valori medi dei descrittori di banchi di pesce e di aggregazioni planctoniche nell'area di studio (Zone 1, Zone 2 e Zone 3). Per ogni descrittore è riportato tra parentesi l'errore standard.	44
3.2	Risultati dei GAMs per l'intera area di studio.	46
3.3	Risultati dei GAM per la Zona 1.	49
3.4	Risultati dei GAM per la Zona 3.	49
3.5	Valori medi e dev. Standard (in parentesi) per specie di alcuni descrittori morfologici e batimetrici dei banchi.	52
3.6	Confusion matrix del modello predittivo finale della classification tree, stimato utilizzando tutti i predittori selezionati (v. testo).	62
3.7	Confusion matrix del modello predittivo della classification tree ottenuto escludendo la profondità dai predittori selezionati (v. testo).	64
3.8	Confusion matrix del modello costruito sulle sole variabili energetiche.	65
3.9	Confusion matrix del modello costruito sulle sole variabili morfologiche.	66
3.10	Esempio di classificazione fornito da trenta differenti modelli, con la previsione finale rappresentata dalla classe più frequente.	70
3.11	Numero e percentuale di banchi attribuiti a ogni specie dal modello MaxRate o MaxRateSpecies.	71
3.12	Numero e percentuale di banchi attribuiti a ogni specie dal modello Mode.	71

3.13	Valori e percentuale di biomassa (in m^2/nmi^2) per miglio e per specie stimati dai modelli predittivi MaxRate o MaxRateSpecies, Mode e dal metodo tradizionale, nei transetti giornalieri (1-60) del survey ANCHEVA2009, effettuati a profondità comprese fra 30 e 140 m.	73
3.14	valori e percentuali di biomassa per specie e gruppi di transetti ottenuti con il metodo tradizionale.	73
3.15	Valori e percentuali di biomassa per specie e gruppi di transetti ottenuti con il modello MaxRate o MaxRateSpecies.	74
3.16	Valori e percentuali di biomassa per specie e gruppi di transetti ottenuti con il modello Mode.	74

Capitolo 1

Introduzione

1.1 Survey acustici per la valutazione della biomassa e della distribuzione di stock ittici

Survey acustici che fanno uso di echosounder scientifici sono ormai da tempo impiegati per stimare la biomassa e mostrare la distribuzione di stock di specie ittiche di interesse commerciale o ecologico, rappresentando così strumenti sempre più importanti per la gestione delle risorse marine (Simmonds and MacLennan, 2005). Ne sono un esempio i survey acustici condotti per le aringhe del Mar del Nord (*Clupea harengus*) (Simmond, 2003); per il Blu whiting (*Micromesistius poutassou*) nel Nord Est dell'Atlantico (ICES 2004g); per i sandeel (*Ammodytes marinus*) nel Mar del Nord (Greenstreet *et al.*, 1998); per le sardine (*Sardina pilchardus*) e le acciughe (*Engraulis encrasicolus*) nella Baia di Biscaglia (Massé *et al.*, 1996), nel Canale di Sicilia (García Lafuente *et al.*, 2002; Patti *et al.*, 2004; Bonanno *et al.*, 2005) e nel Mar Egeo (Giannoulaki *et al.*, 2005). Nelle sole acque del Nord Europa ci sono più di venti stock di pesci per i quali sono effettuate valutazioni acustiche (ICES 2000). Nel resto del mondo le tecniche acustiche sono applicate a una grande varietà di altri pesci come il merluzzo del Newfoundland (*Gadus morhua*) (Rose 1995) o l'aringa del Pacifico in Alaska (*Clupea pallasii*) (Stokesbury *et al.*, 2002). Con lo sviluppo dell'idea che un approccio ecosistemico fosse necessario per una gestione sostenibile delle risorse marine (Sainsbury *et al.*, 2000) cominciarono a essere effettuate anche stime di abbondanza e distribuzione di pesci non sfruttati commercialmente e di altri componenti dell'ecosistema come il micro necton, es. krill (Hewitt and Demer 1991, Reiss *et al.*, 2008) o lo zooplancton (Lavery *et al.*, 2007; Lebourges-Dhaussy *et al.*, 2009).

Uno dei problemi principali connessi all'uso dell'acustica è quello dell'identificazione della specie. Gli strumenti acustici tipicamente utilizzati per le stime di biomassa o densità misurano l'eco di ritorno prodotto dagli esemplari della colonna d'acqua. Il risultato dell'acquisizione è una rappresentazione bidimensionale (ecogramma) di tutti quegli oggetti in grado di determinare una retrodiffusione del segnale acustico. Molti di questi oggetti, come i banchi di pesce, formano negli ecogrammi strutture coerenti definite eco-tracce contenenti una grande quantità di informazioni come posizione, morfologia ed energia, ma non la composizione per specie.

Le tradizionali stime di biomassa o densità di pesci vengono effettuate usando il metodo del filtro soglia su ecogrammi a singola frequenza (38 kHz). Tale metodo aiuta a discriminare nell'ecogramma i banchi di pesce e i pesci dispersi nella colonna d'acqua, ma per ottenere stime di densità per specie è necessario conoscere le proprietà acustiche o "Target Strength" della specie acusticamente individuata. Il processo di identificazione della specie e delle classi di taglia per specie è comunemente effettuato tramite campioni di pesce raccolti mediante catture sperimentali con rete pelagica (cala). Tuttavia è chiaramente impossibile ottenere un campione per ognuno dei banchi di pesce visualizzati durante un survey acustico. Per cui, in pratica, per identificare tutti i banchi, ci si basa sulle caratteristiche di similarità in termini di forma o profondità con i banchi su cui è stata effettuata la cala (Reid *et al.*, 2000), oppure semplicemente sull'esperienza di chi analizza gli ecogrammi rendendo così il processo di identificazione soggettivo e dispendioso in termini di tempi. MacLennan & Simmonds (2005) hanno stimato che l'errore associato alla scorretta classificazione delle tracce acustiche può essere pari anche a $\pm 80\%$.

Quindi un'identificazione obiettiva della specie, direttamente dai dati acustici, potrebbe fornire un importante contributo all'accuratezza delle stime di biomassa, specialmente nel caso di pesci con comportamento aggregativo. Inoltre l'abilità di attribuire ai banchi una particolare specie potrebbe consentire in un futuro studi più dettagliati sulla distribuzione e le caratteristiche comportamentali dei banchi. Infine l'identificazione acustica potrebbe essere usata dai pescatori per individuare banchi di specie target per la loro attività di pesca. Ciò consentirebbe di rendere la pesca molto più selettiva, massimizzando la sopravvivenza di quelle specie che non sono bersaglio della pesca su cui si esercitano inutili pressioni che hanno effetto poi sull'intero ecosistema, andando ad incidere profondamente sul suo equilibrio funzionale e la sua biodiversità. Una pesca molto più selettiva inoltre ridurrebbe il cosiddetto scarto, cioè quella parte del pescato che, benché sfruttato dal punto di vista commerciale, non viene sbarcato per motivi di mercato o a causa di limitazioni in termini di taglia minima di sbarco.

Per questo motivo una delle finalità della ricerca in tale ambito è quella di tentare di migliorare l'accuratezza, la precisione e soprattutto la capacità discriminante del dato acustico. La ricerca di una metodologia che consenta l'identificazione della specie definita da MacLennan and Holliday (1996) come "la grande sfida della pesca applicata all'acustica" o, come definito da Horne (2000), il "Santo Graal" dei ricercatori che operano in questo campo.

Nello Stretto di Sicilia l'IAMC-CNR di Capo Granitola conduce dal 1998, tipicamente nel periodo estivo, campagne di ricerca mirate alla valutazione acustica (echosurvey) dello stato delle risorse ittiche pelagiche (Patti *et al.*, 2004). Il principale risultato di ciascun survey è la stima della distribuzione ed abbondanza delle specie pelagiche maggiormente presenti nell'area di studio (Canale di Sicilia) e nel Mediterraneo, ovvero sardina (*Sardina pilchardus*) e acciuga (*Engraulis encrasicolus*). I dati dell'IREPA (2010), indicano che, nel Mediterraneo, gli sbarchi di acciughe rappresentano il 26% della produzione totale, mentre su livelli inferiori si attestano le quantità di sardine (9%). Entrambe quindi costituiscono una grossa componente della pesca marina complessiva, il che le porta ad essere annoverate in una categoria di notevole importanza nell'industria della pesca e nell'industria conserviera

nazionale (Alvares *et al.*, 2003).

Risultati presentati sulla valutazione dello stato delle risorse, mostrano come il comparto della pesca dei piccoli pelagici sia stato interessato negli ultimi anni da riduzioni delle catture (IREPA, 2009). Nel caso della sardina, dopo i gravi sintomi di crisi manifestatisi fino al 2004, si è assistiti negli ultimi anni a lievi segni di miglioramento delle abbondanze dello stock (Patti *et al.*, 2009). Nel caso dell'acciuga, al contrario, gli indicatori di livelli di biomassa ed il tasso corrente di sfruttamento mostrano chiari segni di allarme, essendosi registrati nel triennio 2006-2008 i più bassi livelli di abbondanza dello stock a partire dal 1998, con la necessità di intervenire sullo sforzo di pesca. La situazione degli stock di piccoli pelagici sembra essere condizionata principalmente dalla variabilità ambientale, sebbene anche un aumento complessivo dello sforzo di pesca può avere effetti sulla biomassa ittica. L'opportunità di proseguire la valutazione annuale della biomassa e dello stato dello stock di sardine e acciughe, mediante survey acustici, è quindi ancor più necessaria alla luce dei recenti risultati sullo stato di sfruttamento delle risorse, che evidenziano pericolo di collasso da un anno all'altro in relazione alla variabilità ambientale ed allo sforzo di pesca. Gli stock di piccoli pelagici devono quindi essere opportunamente gestiti dagli organi competenti nazionali e regionali in accordo con le indicazioni scientifiche derivate dagli studi di "stock assessment", negli ultimi anni sempre più attenti al rispetto delle linee guida dettate dal cosiddetto "Ecosystem approach to fisheries" (Fao, 2003).

Nel Canale di Sicilia tutt'oggi l'iter che porta alle stime di biomassa di acciughe e sardine richiede la ripartizione dell'energia acustica tra le due specie mediante l'interpretazione soggettiva degli ecogrammi e le informazioni biologiche provenienti dalle catture sperimentali. Un'identificazione obiettiva delle due specie comporterebbe in definitiva un aumento dell'accuratezza delle stime di biomassa di queste specie, fondamentale ai fini di una gestione sostenibile della risorsa, una maggiore comprensione delle dinamiche di queste popolazioni e un'attività di pesca più selettiva per i pescatori locali. Per questi ultimi le due specie rappresentano una risorsa storicamente importante dal punto di vista socio economico, se si pensa che da sole acciughe e sardine costituiscono più del 90% dello sbarcato di piccoli pelagici (Patti *et al.*, 2004).

1.2 Studi precedenti sulla classificazione delle eco-tracce

Il primo approccio utilizzato per l'identificazione acustica delle eco-tracce è stato quello di estrarre aggregazioni coerenti da ecogrammi a singola frequenza (38 kHz), e classificarli in relazione alla forma, alla posizione batimetrica e all'energia retrodiffusa. Rose & Leggett (1988) furono i primi a classificare banchi di pesce mediante informazioni relative alla forma e alla taglia delle singole aggregazioni. Da allora numerosi studi sono stati compiuti per caratterizzare le tracce acustiche. Massé & Rouxel (1991) provarono a discriminare i banchi usando il software MOVIES-B (Weill & al., 1993) e le informazioni derivanti da catture pelagiche. Successivamente, Scalabrin *et al.* (1996) studiarono le tracce acustiche selezionate da 11 survey (dal 1989 al 1993), estratte da cale dove una sola specie era presente. Questo studio produsse un database con più di 7000 banchi caratterizzati da 22 parametri che ne descrivevano la morfologia, la batimetria e l'energia. Mediante

l'uso di tali descrittori sembrò possibile distinguere tra acciughe, sardine, blu whiting (*Micromesistius poutassou*) abbastanza precisamente, mentre risultava difficile isolare horse mackerel (*Trachurus trachurus*). I primi studi cercarono quindi di attribuire la specie alle aggregazioni acusticamente individuate, studiando principalmente il comportamento aggregativo (in termini di forma, taglia, distribuzione spaziale). Ma applicazioni successive indicarono che il comportamento aggregativo poteva essere variabile. Per esempio una specie poteva avere un tipo di comportamento quando presente da sola, in una determinata area geografica, e un comportamento diverso in presenza di altre specie (Massé *et al.*, 1996). Oppure alcune specie potevano mostrare un comportamento diverso a seconda dell'area geografica, della stagione o dell'arco della giornata. Da allora la ricerca si è indirizzata verso l'analisi e l'integrazione di informazioni provenienti dall'ambiente con i descrittori del banco, producendo studi che riguardano, ad esempio, la distribuzione dei banchi di pesce in relazione alla profondità del termoclino (Swartzmann *et al.*, 1994), la distribuzione di pesci pelagici in funzione della temperatura e della salinità superficiale (Walsh *et al.*, 1995; Castillo *et al.*, 1996), la distribuzione dei banchi pelagici e la relazione con il tipo di fondo (D'Elia *et al.*, 2009), le caratteristiche spaziali dei banchi di pesce e la loro relazione con la densità di cibo (Nøttestad *et al.*, 1996; Patti *et al.*, 2011).

Tra il 1997 e il 1999 fu approvato un progetto noto come "CLUSTER (FAIR-CT-96.1799)" che prevedeva di analizzare il comportamento aggregativo e la distribuzione spaziale di eco tracce derivanti da serie storiche di tutti gli stock di pesce Europei e di esaminarne le variazioni interannuali in relazione a più parametri. L'obiettivo finale era quello di individuare delle regole generali che potessero essere applicate a tutti gli stock di pesci pelagici. Contestualmente in questo progetto si lavorò per individuare adeguate procedure di analisi di immagine, da applicare agli ecogrammi, che tenessero conto della risoluzione verticale e orizzontale di acquisizione, e che consentissero di isolare aggregazioni di pesce dal resto del segnale acustico. Si lavorò anche alla stesura di un protocollo per la nomenclatura e la definizione dei parametri che caratterizzano un banco (Reid *et al.*, 2000) e si mise in luce la necessità di fattori di correzione sui parametri morfologici come suggerito da Diner (2001).

Da allora l'analisi delle relazioni fra comportamento aggregativo e distribuzione spaziale è stata affrontata mediante l'uso di diverse tecniche statistiche. Clustering, analisi discriminante (Scalabrin *et al.*, 1996; Lawson *et al.*, 2001), neural network (Simmonds *et al.*, 1996), support vector machines (Robotham *et al.*, 2009) sono state utilizzate per l'identificazione della specie e lo studio della distribuzione spaziale dei banchi, mentre i modelli additivi generalizzati sono stati impiegati per analizzare le complesse relazioni tra le aggregazioni e i fattori esterni (Agenbag *et al.*, 2003; Swartzmann *et al.*, 2008). Alcune regole generali valide annualmente e in differenti stock sono state individuate al termine di tali studi: una relazione positiva tra la dimensione del banco in termini di numero di individui e la sua dimensione geometrica; una relazione inversa tra il numero di individui nel banco e il numero di banchi vicini; l'influenza dell'ora del giorno sul numero di banchi e le loro caratteristiche; una relazione tra il numero totale di banchi e l'organizzazione spaziale delle aggregazioni, indipendentemente dall'abbondanza totale di pesce: al crescere del numero dei banchi, si osservano banchi più piccoli organizzati in più cluster.

1.3 Uso di dati a multifrequenza per l'attività di classificazione delle specie

Numerosi studi sono stati condotti negli ultimi anni per classificare ed identificare ecotracce basandosi sulle informazioni derivanti da un echosounder. Questi studi vanno dall'analisi delle caratteristiche delle ecotracce acquisite ad una sola frequenza (Scalabrin *et al.*, 1996), alla combinazione di informazioni acquisite a più frequenze (Madureira *et al.*, 1993; Kloser *et al.*, 2002), fino all'uso di broad band echosounders (Simmonds *et al.*, 1996; Lundgren and Nielsen, 2008). Le attività di classificazione condotte con l'uso di una sola frequenza hanno dato buoni risultati in ambienti monospecifici ma sembra che raramente siano da sole in grado di gestire situazioni più complesse (MacLennan & Holliday, 1996). L'uso di più frequenze consente invece di ottenere informazioni biologiche più accurate e un maggior numero di informazioni acustiche da ogni eco traccia, specialmente se le proprietà energetiche della specie variano in maniera consistente con le frequenze utilizzate (Madureira *et al.*, 1993). Differenze nelle proprietà di backscatter, alle varie frequenze, sono considerate da Kang *et al.* (2002) una delle tecniche più promettenti per l'identificazione della specie. Questo tipo di approccio è stato utilizzato anche da Kloser *et al.* (2002) per discriminare tre gruppi di specie dominanti delle acque profonde. L'uso di differenze nelle proprietà di backscatter è stato anche raccomandato da Watkins & Brierley, (2002) per rendere meno soggettivo il processo di interpretazione dell'ecogramma, ma tale metodologia però ancora oggi non è stata adottata universalmente per le stime di biomassa. Anche l'attività di classificazione delle ecotracce isolate ha visto l'applicazione di diverse tecniche statistiche. Queste comprendono analisi dei componenti principali e analisi discriminante (Scalabrin *et al.*, 1996; Lawson *et al.*, 2001), artificial neural networks (Simmonds *et al.*, 1996); nearest-neighbour analysis (Richards *et al.*, 1991), k-mean clustering (Tegoswski *et al.*, 2003) o un insieme di questi modelli (Fleischman & Burwen, 2003). Molti di questi modelli risultano spesso buoni classificatori ma sono poi o difficili da applicare (es. neural networks) o richiedono l'osservanza di ipotesi statistiche che possono essere violate (es. DFA, vedi ICES 2000). Uno dei metodi più promettenti sembra essere la *classification tree* (Fernandes, 2009).

L'idea di utilizzare dati acustici a multifrequenza per la discriminazione in specie deriva dagli sforzi, compiuti negli ultimi anni, nell'identificare il segnale acustico proveniente da differenti gruppi di taglia zooplanctonici. Questi sforzi hanno portato allo sviluppo di modelli di scattering che sono oggi alla base della ricerca nel campo dell'acustica applicata alla pesca e al plancton. Secondo tali modelli, in relazione alle frequenze generalmente utilizzate nella pesca (18, 38, 120 e 200 kHz) alcuni target si trovano nella cosiddetta regione di Rayleigh (Es. plancton), cioè in quella regione in cui il backscattering varia significativamente con la frequenza. Viceversa, i pesci dotati di vescica natatoria, all'interno di questo range di frequenze, mostrano un backscattering con comportamento di tipo geometrico, vale a dire uno scattering pressoché costante al variare della frequenza. (Fig. 1.1 da Lavery *et al.*, 2007).

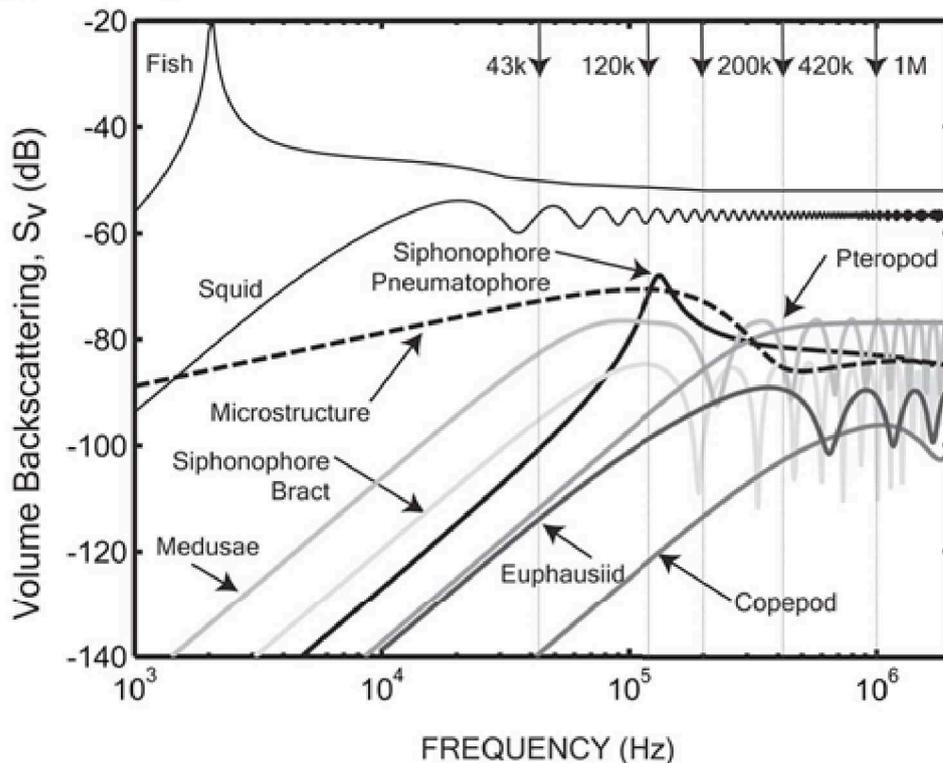


Figura 1.1: Effetto della frequenza sull'intensità di backscatter in volume (S_v) dei pesci, dei gruppi zooplanctonici e delle strutture fisiche (Lavery *et al.*, 2007)

Il modello adottato per i pesci descrive però il comportamento della sola vescica natatoria, perché considerata la componente dominante nel determinare le capacità riflettenti dei singoli individui. Tuttavia, recentemente, nuovi modelli sono stati elaborati combinando l'effetto della vescica natatoria con l'effetto del corpo, come il modello usato da Fässler *et al.* (2007) per descrivere le capacità riflettenti dell'aringa atlantica. Secondo questo modello il corpo determina nel pesce un comportamento risonante. Questa componente corporale, a certe profondità, diventa la componente dominante e le caratteristiche di risonanza alla varie frequenze assumono una forte dipendenza dalla taglia del pesce. Per un'aringa atlantica con taglia di 20-25 cm, il contributo del corpo all'energia totale di backscatter diventa predominante sulla vescica natatoria già a una profondità di 112 m, per effetto della compressione della vescica natatoria (Fig. 1.2).

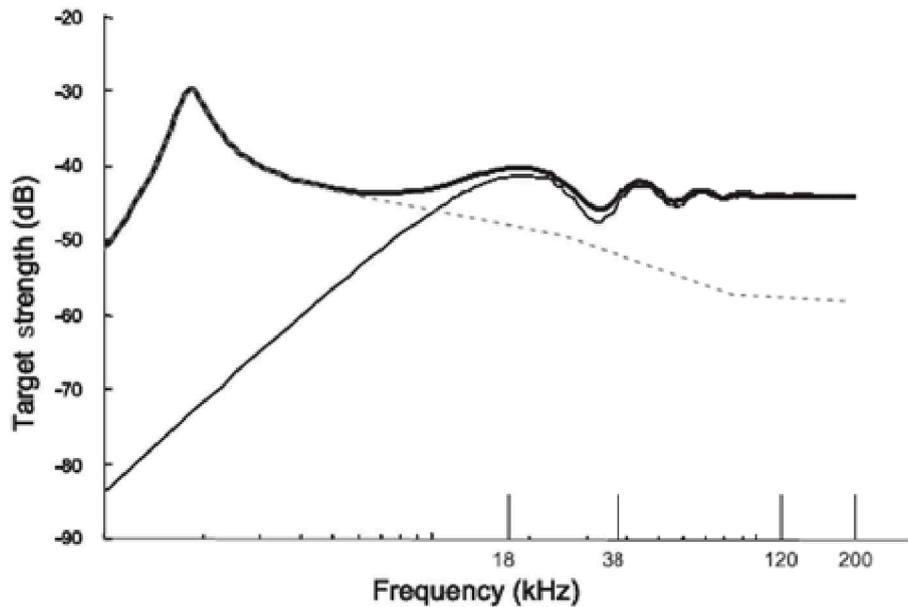


Figura 1.2: Modello predittivo del backscatter strength di un'aringa Atlantica di 20-25 cm di taglia ad una profondità di 112 m. La linea continua spessa riporta il modello derivato dalla combinazione dell'effetto della vescica natatoria con l'effetto del corpo. La linea tratteggiata raffigura il modello usato per rappresentare l'effetto della vescica natatoria, mentre la linea continua sottile raffigura il modello usato per rappresentare l'effetto del corpo del pesce. (Fässler *et al.*, 2007)

1.4 Obiettivo e struttura della tesi

L'obiettivo generale del progetto è quello di caratterizzare ed identificare banchi di specie pelagiche target (acciughe e sardine) nel Canale di Sicilia, utilizzando informazioni ambientali (provenienti dalla colonna d'acqua e dal fondo) e parametri che ne descrivono la struttura, in termini morfologici, batimetrici ed energetici a più frequenze. Le informazioni provenienti dalla colonna d'acqua sono state estratte da stazioni CTD, mentre informazioni sul fondo sono state ricavate acusticamente, utilizzando la metodologia di cui fa uso il sistema RoxAnn (SonaVision Ltd.). La tematica è stata affrontata in due fasi distinte e complementari. In una prima fase dello studio una caratterizzazione dei banchi e ipotesi sulla possibile composizione in specie sono state effettuate usando informazioni estratte da una singola frequenza (38 kHz) ed analizzando la loro relazione con la densità e la biomassa di plancton, considerata indice della disponibilità di cibo. In una seconda fase invece si è cercato di individuare quali parametri, fra i descrittori del banco e le variabili ambientali, potessero essere d'aiuto nel discriminare queste specie. Il primo punto ha richiesto lo svolgimento delle seguenti fasi:

- *Uso di un algoritmo a singola frequenza per l'estrazione delle caratteristiche strutturali di un banco e l'individuazione delle aggregazioni planctoniche:* l'algoritmo per l'isolamento dei banchi è stato sviluppato in Matlab da Patti *et al.* (2006) ed è una versione del metodo presentato da Swartzmann *et al.* (1999).

- *Analisi delle relazioni tra biomassa e densità di banchi e plancton mediante l'uso di Modelli addittivi generalizzati (GAMS):* L'analisi è stata affrontata suddividendo in aree il settore in studio (Canale di Sicilia) e separando tra plancton superficiale e plancton profondo.
- *Caratterizzazione per aree dei banchi e delle aggregazioni planctoniche:* un'analisi statistica descrittiva è stata utilizzata per caratterizzare i banchi di pesce e le patch di plancton estratte acusticamente dalla singola frequenza a 38 kHz.

Il secondo punto invece ha richiesto lo sviluppo delle seguenti fasi:

- *Creazione e applicazione di un algoritmo a multifrequenza per l'identificazione di banchi di pesce dotati di vescica natatoria ed estrazione delle caratteristiche strutturali a più frequenze.* L'algoritmo a multifrequenza noto come "Fish school filter", è stato sviluppato nel progetto SINFAMI (Fernandes *et al.*, 2006), modificato e adattato alle caratteristiche acustiche dei banchi individuati nel Canale di Sicilia e implementato come algoritmo nel modulo "Virtual Variable" del software Sonardata Echoview v.4.9. Questo algoritmo consente di rendere spazialmente comparabili i dati acustici a differenti frequenze in modo da consentire un confronto delle informazioni energetiche estratti dai banchi a differenti frequenze. Questo nuovo algoritmo consente inoltre di interpretare con maggiore accuratezza il dato acustico in termini biologici, cioè in termini di biomassa, in quanto elimina parte del segnale acustico proveniente da altri target presenti nella colonna d'acqua, cefalopodi, meduse, sedimento in sospensione, che nelle analisi rappresentano rumore, ma che possono essere erroneamente interpretati come banchi di pesce quando si fa uso di una sola frequenza. Il nuovo algoritmo risulta anche più efficiente nel mantenere la reale struttura del banco. Il metodo tradizionale che fa uso di valori soglia sulla singola frequenza a 38 kHz, sembra infatti che tenda ad erodere parte della struttura del banco (MacLennan and Simmonds, 2005).
- *Applicazione di una metodologia acustica per l'estrazione delle proprietà di durezza del fondo marino:* si descrive la tecnica utilizzata per classificare il tipo di fondo usando dati acustici raccolti nel corso dei survey acustici. L'approccio descritto è simile a quello utilizzato dal sistema commerciale RoxAnn.
- *Creazione di un learning sample da utilizzare per l'attività di classificazione:* un database di cui è nota la composizione è stato creato combinando dati biologici provenienti da cale monospecifiche e tracciati acustici effettuati in corrispondenza delle cale.
- *Caratterizzazione batimetrica e morfologica dei banchi dei gruppi di pesci mediante un'analisi statistica descrittiva:* è fornita una descrizione delle caratteristiche medie di ogni specie in termini di distribuzione batimetrica, e caratteristiche morfologiche estratte dalla singola frequenza a 38 kHz.
- *Analisi delle differenze nelle proprietà energetiche tra le due specie target:* parametri energetici estratti a differenti frequenze sono stati analizzati per valutarne il potenziale

uso nell'attività di classificazione delle due specie. I parametri energetici a più frequenze, sono infatti potenzialmente utili all'identificazione della specie, perchè connessi principalmente alle caratteristiche della specie quali presenza o meno di vescica natatoria, forma del pesce e taglia del pesce (Fernandes *et al.*,2006). Specie come acciughe e sardine sono abbastanza simili dal punto di vista acustico: entrambe possiedono vescica natatoria e formano banchi nelle ore diurne generalmente simili in termini di struttura. Per questo motivo tradizionalmente è sempre stato abbastanza difficile distinguere tra le due specie. Tuttavia le sardine hanno solitamente dimensioni maggiori in termini di taglia rispetto alle acciughe, principalmente in autunno, quando acciughe sono nel periodo del reclutamento, per cui le loro proprietà acustiche potrebbero differire.

- *Analisi delle differenze nelle proprietà energetiche per range di profondità diversi:* differenze energetiche dovute a cambiamenti di profondità sono state prese in considerazione. I clupeidiformi (come acciughe e sardine) sono infatti pesci fisiostomi e contrariamente ai fisoclisti non hanno meccanismi che gli permettono di aggiustare attivamente il volume della loro vescica natatoria, in risposta a cambiamenti di profondità e quindi di pressione. Di conseguenza questi pesci potrebbero mostrare una riduzione in volume della vescica natatoria all'aumentare della pressione, in accordo con la legge di Boyle, come osservato da Fässler *et al.* (2009) per le aringhe. Ci si potrebbe aspettare quindi che una distinzione tra le due specie sia possibile confrontando le proprietà di backscattering di queste ultime in relazione a cambiamenti di profondità.
- *Costruzione di modelli decisionali ad albero mediante l'uso di parametri strutturali del banco:* modelli decisionali ad albero sono stati costruiti con le variabili strutturali del banco per individuare quali parametri meglio discriminano tra le specie e valutarne le capacità di classificazione.
- *Costruzione di modelli decisionali ad albero mediante l'uso di parametri ambientali e strutturali:* modelli decisionali ad albero sono stati costruiti mettendo insieme le informazioni ambientali e le variabili strutturali del banco per valutarne le capacità di classificazione.

Capitolo 2

Materiali e metodi

2.1 Processamento di dati a singola frequenza per l'estrazione di banchi di pesce e delle aggregazioni planctoniche

Informazioni sui banchi di pesce e aggregazioni planctoniche sono state ricavate dai dati acustici acquisiti alle frequenze di 38 e 120 kHz mediante l'echosounder scientifico della SIMRAD EK 500 dalla campagna ANCHEVA98, effettuata nel periodo di giugno del 1998. Poiché le specie di piccoli pelagici si aggregano in banchi esclusivamente nelle ore di luce, solo i transetti diurni sono stati selezionati per il presente studio (Fig. 2.1). I dati acustici selezionati sono stati filtrati mediante una procedura elaborata con il software Echoview V.3 e convertiti in formato ASCII per le successive analisi in Matlab. Per mezzo di quest'ultimo software sono stati sviluppati degli script che applicano algoritmi di analisi di immagine agli ecogrammi di entrambe le frequenze, consentendo di individuare e delimitare oggetti attribuibili a patch di plancton o a banchi di pesce, sulla base dell'energia scatterata. La procedura specifica per l'identificazione dei banchi di pesce è quella riportata in D'Elia et al. (2009). Tale procedura, applicata esclusivamente agli ecogrammi a 38kHz, adopera la soglia dell'area minima di 4 m^2 e un filtro nell'intervallo $[-58, -36 \text{ dB}]$ per determinare quali potenziali patch acustiche possano essere considerate banchi di pesce. Le informazioni sui banchi sono state infine filtrate in modo tale da rimuovere tutte quelle aggregazioni poco dense, che si trovano nella metà superiore della colonna d'acqua, attribuite principalmente a banchi in via di formazione (D'Elia *et al.*, 2009).

Di seguito sono invece riportati gli step della procedura utilizzata per l'identificazione delle patch di plancton. Sugli ecogrammi a 38 e 120 kHz si applica un primo filtraggio, in modo tale da includere nell'analisi solo valori di backscattering all'interno dell'intervallo $[-70, -50 \text{ dB}]$, corrispondente al segnale acustico retro-diffuso dal plancton, e successivamente una maschera binaria setta a 0 i valori di ogni pixel al di fuori di questo intervallo e a 1 i valori all'interno di tale intervallo. Si esegue poi un'operazione di differenza tra gli ecogrammi binari di ambedue le frequenze. Sull'ecogramma risultante, i pixel con valore minore o uguale a 5 dB sono posti uguali a zero al fine di eliminare l'effetto di "scattering geometrico", mentre i pixel con valore superiore a 5 dB si sono posti a 1, ottenendo così una maschera binaria. A tale maschera binaria si applica un filtro morfologico, caratterizzato

da un'operazione di chiusura e successiva apertura, effettuata mediante un elemento strutturale di dimensione 3x3. La dimensione dell'elemento strutturale tiene conto della tipica distribuzione del plancton, che tende a stratificarsi orizzontalmente in relazione alle caratteristiche delle masse d'acqua che condizionano gli spostamenti. La procedura si conclude moltiplicando la versione dell'ecogramma binario risultante per l'ecogramma a 120 kHz originario (Patti *et al.*, 2006). Dalle patch individuate vengono infine estratti, dalla sola frequenza a 38 kHz, i parametri che ne descrivono la struttura. I parametri estratti riguardano la posizione (profondità della patch, distanza dal fondo), la dimensione (area, lunghezza, altezza, elongazione) e l'energia (NASC, indice di biomassa e NASC/AREA, indice di densità).

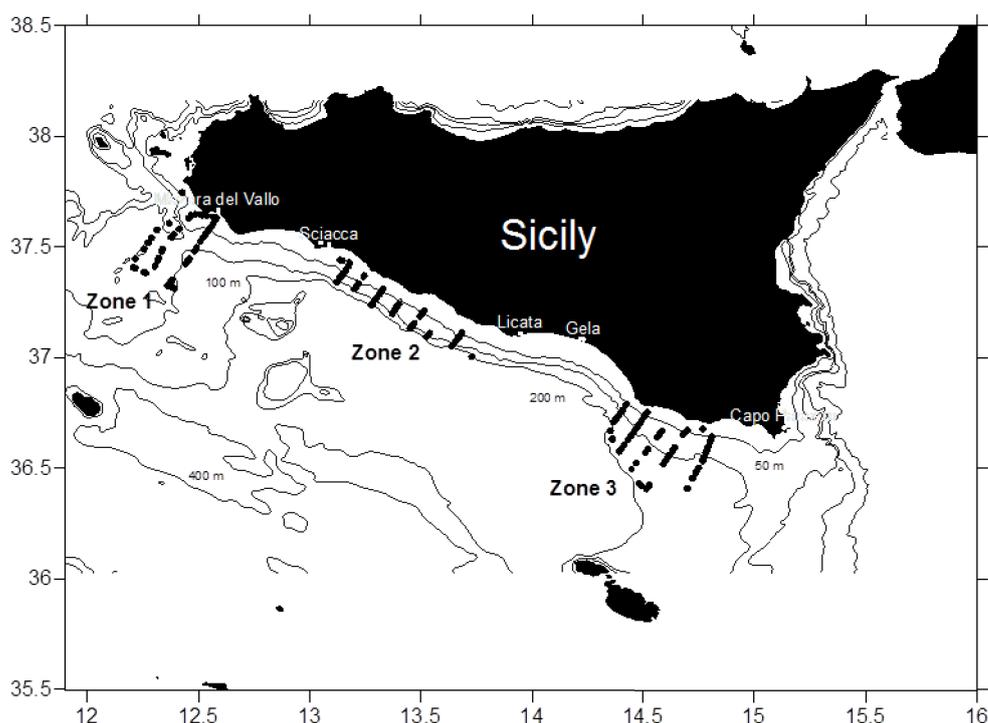


Figura 2.1: Area di studio. Mostra i siti, lungo il tracciato acustico giornaliero del survey ANCHEVA1998, in cui banchi di pesce e plancton sono compresenti. È riportata anche la separazione dell'area di studio in tre zone, indicate come Zona 1, Zona 2, Zona 3.

Con lo scopo di analizzare la relazione tra biomassa o densità dei banchi di pesce e biomassa o densità delle patch di plancton, i valori medi di NASC sono stati mediati ogni miglio nautico e dall'analisi sono state escluse tutte le aree in cui banchi e plancton non fossero compresenti. La relazione è stata investigata utilizzando Modelli Additivi Generalizzati (GAMs), al fine di individuare possibili relazioni non lineari tra le due variabili. L'analisi è stata poi approfondita separando il plancton superficiale (con profondità comprese tra 1 e 20 m) da quello più profondo (individuato a distanze dal fondo inferiore a 5 m), e analizzando le relazioni per le singole aree denominate Zona 1, Zona 2, Zona 3 (corrispondenti ai transetti giornalieri separati dal tratto di mare investigato nel corso

della notte). Per l'applicazione dei modelli è stata utilizzata una distribuzione Gaussiana dell'errore con l'identità come funzione di link. Inoltre, una trasformazione logaritmica dei predittori è stata applicata sui dati per soddisfare le condizioni di normalità nelle distribuzioni. La selezione delle funzioni di smoothing è avvenuta seguendo la metodologia proposta da Hastie & Tibshirani, 1990 usando la libreria “mgcv” del software statistico R (R. Development Core Team, 2005).

2.2 Processamento dei dati a multifrequenza per la caratterizzazione e l'identificazione dei banchi di pesci

2.2.1 Raccolta dei dati acustici e biologici per la costruzione del learning sample

I dati acustici sono stati raccolti utilizzando due differenti echosounder (EK 500 ed EK60), dotati di due o tre trasduttori split-beam, operanti alle frequenze di 38, 120 e 200 kHz. Questi ultimi sono stati calibrati usando il metodo raccomandato da Foote *et al.*, (1987). La tabella di seguito (tab. 2.1) riporta la potenza trasmessa e l'ampiezza del cono acustico di ogni trasduttore.

Tabella 2.1: La tabella riporta il modello di echosounder e le frequenze dei trasduttori usati in fase di acquisizione. Ad ogni trasduttore corrisponde una potenza trasmessa e un'ampiezza del cono acustico.

Echosounder Frequency [Khz]	Transmit power [W]	Beam-width [3dB]
EK 500/38	2000	7°
EK 500/120	1000	7°
EK 500/200	1000	7°
EK 60/38	2000	7°
EK 60/120	1000	7°
EK 60/200	1000	7°

I tre differenti trasduttori sono stati configurati per emettere ping contemporaneamente ogni secondo. La durata dell'impulso emesso nell'EK 500 è di 1.0, 1.0, e 0.6 ms rispettivamente per le frequenze 38, 120 e 200 kHz, mentre la durata dell'impulso scelta per l'EK 60 è di 1.024 ms per tutte e tre le frequenze. Il raw data utilizzato per il seguente studio consiste di un ecogramma, in cui ogni sample (unità campionata) corrisponde ad un valore di backscatter in volume (Sv, dB re 1m^{-1}). I dati acustici sono stati acquisiti, lungo la piattaforma continentale della costa meridionale siciliana, mantenendo una velocità di 10 nodi durante l'ordinaria fase di navigazione lungo transetti costa largo (vedi Fig. 2.2 per la posizione delle catture). Nella fase di cattura invece la velocità dell'imbarcazione è stata ridotta a circa 4 nodi. Per le operazioni di pesca è stata adoperata una rete pelagica con apertura orizzontale di 13-15 m, apertura verticale della bocca di 6-8 m e dimensione delle maglie nella parte terminale di 10 mm.

Per la costruzione del learning sample sono state selezionate, da survey condotti tra il 2002 e il 2009, solo eco tracce provenienti da catture sperimentali giornaliere con composizione monospecifica (Fig. 2.2).

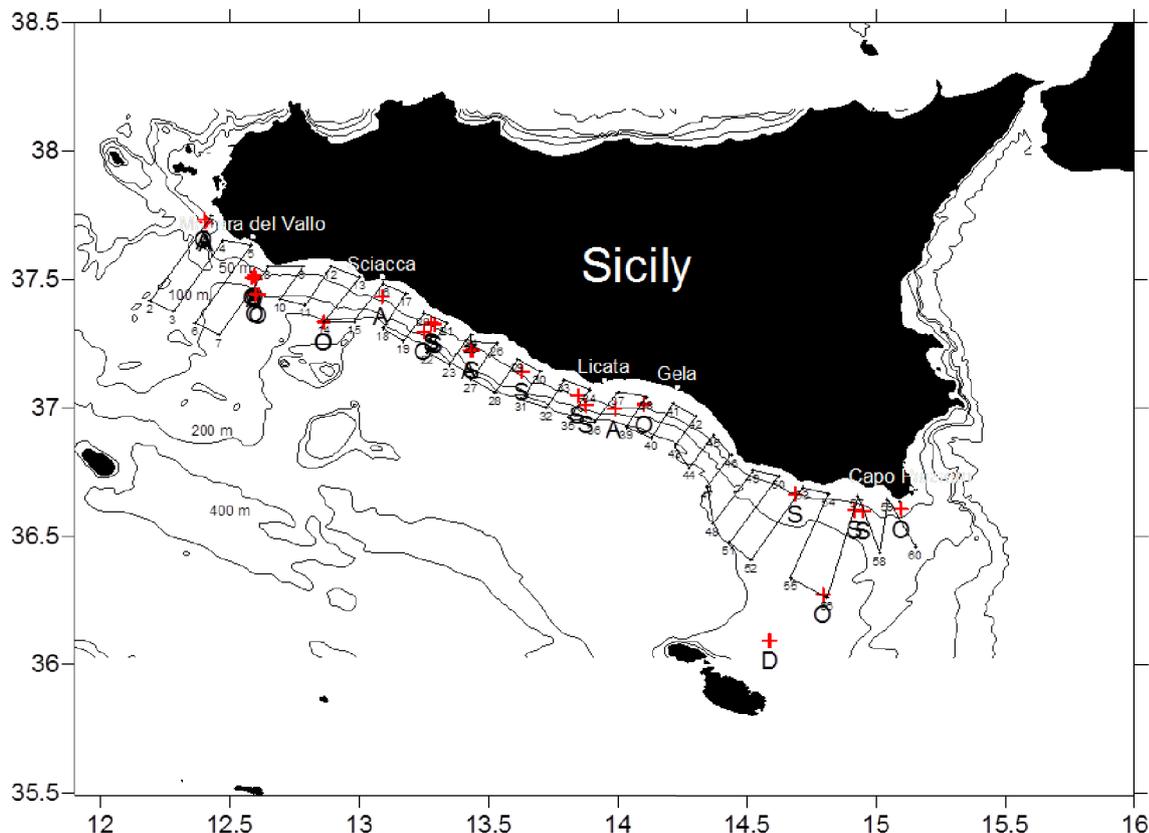


Figura 2.2: Area di studio. In rosso sono visualizzate le catture monospecifiche selezionate e i transetti costa-largo effettuati durante il survey ANCHEVA2009. La lettera indica la specie predominante nella cala: A=acciughe, O= altre specie pelagiche, S=sardine, D=demersali.

Qualsiasi attività di classificazione richiede, infatti, per prima cosa, la creazione di un dataset di cui è nota l'informazione sulla composizione per specie. Il "learning sample" è stato realizzato combinando dati biologici, derivanti da catture pelagiche "monospecifiche" di acciughe e sardine, con le corrispondenti eco tracce (dati acustici acquisiti durante la cala). La combinazione di questi dati consente di ottenere per ogni cala informazioni sulla composizione per specie (in termini di peso), la classe di taglia, la latitudine, la longitudine, il tempo di inizio e fine della cala e altri parametri connessi alla profondità e alle caratteristiche delle eco tracce individuate. Sono state considerate monospecifiche per le sardine tutte quelle cale aventi una percentuale superiore all'80% del totale in peso delle catture e monospecifiche per le acciughe quelle aventi una percentuale superiore al 70%. Tali soglie hanno permesso di selezionare un totale di 17 cale monospecifiche (11 per le sardine e 6 per le acciughe), effettuate nelle seguenti campagne acustiche condotte

tra il periodo estivo e autunnale: Ancheva2002, Juvenile2004, Ancheva2005, Juvenile2005, Ancheva2006, Ancheva2007, Ancheva2008, Ancheva2009. Di tali campagne purtroppo solo nel corso delle campagne Ancheva2002, Ancheva2008 e Ancheva2009 sono stati acquisiti e calibrati dati acustici a tre frequenze (38 kHz, 120 kHz, 200 kHz), mentre per tutte le altre campagne solo due trasduttori (38kHz e 120kHz) sono stati utilizzati e/o calibrati. Per ottenere un dataset più robusto sono state quindi prese in considerazione solo le informazioni energetiche a due frequenze ricavate da tutte le otto campagne. Insieme alle cale monospecifiche sono state selezionate nove cale con percentuale in peso superiore all' 80%, costituite da altre specie pelagiche diverse dalle specie target (Es. *Trachurus trachurus*, *Sardinella aurita*, *Boops boops*, *Spicara maena*...), denominate OPS (Other Pelagic Species) e una cala con una percentuale in peso superiore al 70%, costituita da specie appartenenti al gruppo dei demersali (in cui rientrano tutte quelle specie che vivono a stretto contatto con il fondo). La tabella (2.2) riporta un elenco delle cale selezionate con le relative informazioni registrate durante l'operazione di pesca. Il passo successivo alla raccolta dei dati biologici è stato quello di individuare dei poligoni sul tracciato acustico di ogni cala, utilizzando le informazioni riguardanti la profondità della rete e la sua apertura, al fine di delimitare l'area campionata dalla rete all'interno dell'ecogramma. Su ogni ecogramma è stata poi individuata la linea del fondo, automaticamente rilevata dallo strumento acustico (detected-sounded bottom), ed una linea a 8.5 m dalla superficie. Queste due linee (limiti) sono state impiegate per eliminare eventuali echi associati ad oggetti e bolle d'aria presenti nello strato più superficiale e ad echi provenienti dal fondo. Questa procedura è stata svolta utilizzando il software Sonardata Echoview v.4.9.

Tabella 2.2: Elenco delle cale selezionate con le relative informazioni riguardanti: la campagna e l'anno di pescata (survey), il numero della cala (Trawl), la specie presente (species), la percentuale in peso con cui la specie è stata catturata (Amount %), il tempo di inizio della cala (Start Time), la durata della cala (duration), la distanza della rete dalla base al fondo (Dist ft_bot), la distanza della rete dalla testa al fondo (dist hd_bot), la profondità del fondale (bottom), il peso della cattura (Catch), la lunghezza media delle specie target (Species_Msize).

Survey	Trawl	Species	Amount (%)	Start Time	Dur. (min)	Dist. ft_bot (m)	Dist. hd_bot (m)	Bottom (m)	Catch (g)	Species Msize (cm)	Specie Comp
Juv2004	21	Anchovy	81.58	12.04	20	0	4.8	44.8	57500	7.57	
Anch2005	1	Anchovy	98.75	19.10	32	0.3	7.4	77.9	581900	14.27	
Anch2005	3	Anchovy	82.57	7.37	30	0	5.5	80.5	236800	13.09	
Juv2005	12	Anchovy	97.27	16.35	30	0.7	7.5	67.2	461500	9.33	
Anch2009	2	Anchovy	73.70	9.00	35	0.00	6.20	80.60	23037	13.84	
Anch2009	10	Anchovy	73.00	17.37	30	0.00	4.90	62.12	172023	12.81	
Anch2002	22	Sardine	85.17	9.40	30	0.2	6.9	46.4	66400	12.02	
Juv2004	20	Sardine	82.38	7.15	30	0.2	5.4	68.4	90900	13.77	
Anch2005	7	Sardine	85.45	7.35	35	0	5.3	44.1	304000	11.48	
Juv2005	4	Sardine	80.14	17.16	34	0	4.2	63.1	43400	10.1	
Anch2006	8	Sardine	87.68	11.20	32	0	5.6	43.7	173600	12.77	
Anch2007	8	Sardine	87.61	12.11	30	0	6.1	57.8	57414	13.85	
Anch2007	16	Sardine	83.49	11.02	36	0	5.7	58.9	101440	14.05	
Anch2008	2	Sardine	89.76	7.12	30	0	6.7	77.5	43030	13.38	
Anch2008	5	Sardine	93.13	7.53	30	0	5.8	40.6	99413	12.54	
Anch2009	11	Sardine	91.41	8.03	30	0	4.8	43.67	169845	12.12	
Anch2009	20	Sardine	94.62	13.30	30	0	4.9	70.42	159369	13.9	
Anch2005	29	Demersali	73.97	12.18	32	0	6.1	143.6	14000		<i>Mullus barbatus,</i> <i>Macroronphosus scolapax,</i> <i>Lepidogatus,</i> <i>Arnoglossus lanterna</i>
Juv2004	3	OPS	87.02	10.24	31	0	5.8	75	74100		<i>Trachurus trachurus,</i> <i>Sardinella aurita</i>
Anch2005	11	OPS	81.21	11.32	30	0	5.8	34.5	106200		<i>Trachurus trachurus</i> <i>Sardinella aurita,</i> <i>Spicara spp.</i>
Anch2005	15	OPS	80.27	15.57	32	0	5.5	119.4	80600		<i>Trachurus trachurus,</i> <i>Sardinella aurita,</i> <i>Spicara spp.</i>
Juv2005	20	OPS	90.03	19.00	35	9.8	17.6	94	10000		
Anch2006	22	OPS	92.4	19.20	30	0	5.5	44.7	745600		<i>Trachurus trachurus,</i> <i>Sardinella aurita,</i> <i>Spicara maena</i>
Anch2009	9	OPS	96.79	10.05	30	0	5.3	77.07	202367		<i>Scomber japonicus,</i> <i>Trachurus trachurus,</i> <i>Spicara maena</i>
Anch2006	3	OPS	92.04	9.04	33	0	6.6	115.4	301700		<i>Trachurus trachurus</i>
Anch2007	2	OPS	100	7.45	30	0	4.9	82.1	730000		<i>Trachurus trachurus</i>
Anch2007	4	OPS	93.61	14.27	30	0	5.8	115.4	74566		<i>Trachurus trachurus</i>

2.2.2 Processamento dei dati acustici per l'estrazione delle informazioni sui banchi di pesce da dati a multifrequenza

Le eco tracce di potenziali banchi di pesce sono state isolate dall'ecogramma mediante un processo che consiste di tre fasi:

- Applicazione di un “fish school filter” agli ecogrammi.

Questo filtro è stato sviluppato nel progetto SINFAMI (Fernandes *et al.*, 2006), modificato e adattato alle caratteristiche dei banchi individuati nell'area di studio, e implementato come algoritmo nel modulo “Virtual Variable” del software Sonardata Echoview. La figura seguente (Fig. 2.3) mostra la geometria dell'algoritmo sviluppato per questo studio.

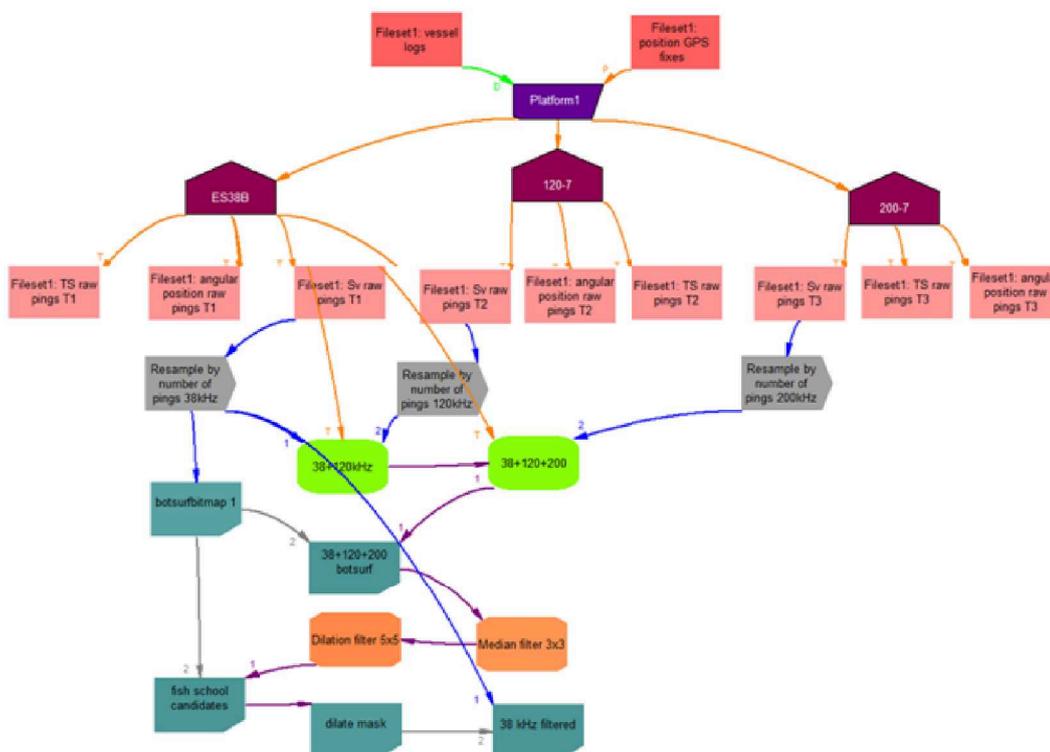


Figura 2.3: Struttura dell'algoritmo “fish school filter” creato con il modulo “Virtual Variable” del software Sonardata Echoview.

L'algoritmo applica dei filtri che consentono di isolare banchi di pesce dotati di vescica natatoria e fa uso di strumenti di analisi di immagine per rendere spazialmente comparabili i dati acquisiti a differenti frequenze. In condizioni normali infatti, fra gli ecogrammi a differenti frequenze, si verifica una parziale sovrapposizione orizzontale e verticale di ogni sample (unità campionata). La parziale sovrapposizione orizzontale è

legata alla distanza tra i diversi trasduttori, installati a scafo, e all'ampiezza del beam (se sono utilizzate ampiezze di beam differenti alle diverse frequenze). Una parziale sovrapposizione verticale può invece presentarsi quando, alle varie frequenze, vengono utilizzate larghezze di beam differenti, se è impostata una diversa durata dell'impulso, se gli impulsi emessi non sono sincronizzati, se durante la fase di acquisizione la start depth non è impostata a zero. A parità di forma e durata dell'impulso e per ampiezze di beam differenti, il vertical offset è invece attribuito alla differenza nel ritardo totale del sistema e alla risoluzione verticale (Korneliussen *et al.*, 2008). La parziale sovrapposizione orizzontale si riduce con l'incremento della profondità a causa dello sviluppo conico del beam, mentre l'offset verticale si mantiene. Di seguito è riportata una descrizione più dettagliata dell'algoritmo. Partendo da ogni singolo ecogramma l'algoritmo applica, qualora necessario, un'operazione di ricampionamento della variabile di input (valori di Sv), sulla base della risoluzione verticale con cui la variabile è stata acquisita (nel nostro caso una risoluzione verticale di 20 cm). Questa operazione è necessaria quando fra gli ecogrammi a differenti frequenze si verifica una non completa sovrapposizione verticale di ogni sample. Il ricampionamento consente così di allineare i diversi ecogrammi ai fini delle successive operazioni di somma. All'ecogramma risultante dalla somma delle tre frequenze (vedi Fig. 2.4) viene applicata una maschera binaria per limitare l'analisi allo strato d'acqua compreso tra la linea del fondo e quella superficiale, impostata a 8.5 m (Fig. 2.5). Al passo successivo l'algoritmo utilizza un filtro mediano, con elemento strutturante 3 per 3, per effettuare un'operazione di smoothing sulle eco tracce individuate, ed eliminare così tutte le strutture molto piccole (Fig. 2.6). Quindi esegue un'operazione di dilatazione (mediante elemento strutturante 5x 5), per ridurre lo shift orizzontale fra i banchi e mettere in luce le aggregazioni (Fig. 2.7); ed applica un valore soglia minimo per la visualizzazione dei nuovi valori di Sv (corrispondente a -175 dB per l'ecogramma risultante dalla somma di tre frequenze e -116 dB per gli ecogrammi risultanti dalla somma di due frequenze). Questi valori sono stati determinati empiricamente fino al raggiungimento dell'ecogramma "più pulito", ovvero di un ecogramma da poter analizzare ai fini dell'identificazione delle caratteristiche dei banchi di pesce. Le ecotracce così individuate sono quelle candidate ad essere considerate banchi (Fig. 2.8). Infine l'algoritmo trasforma l'ecogramma in una maschera binaria e la moltiplica per l'ecogramma originale alla frequenza di 38kHz, su cui viene impostata la soglia minima di -70 dB, utilizzata durante il survey per la visualizzazione di plancton e di banchi di pesce. I Valori di Sv dei candidati banchi vengono così riportati su scala originaria (Fig. 2.9).

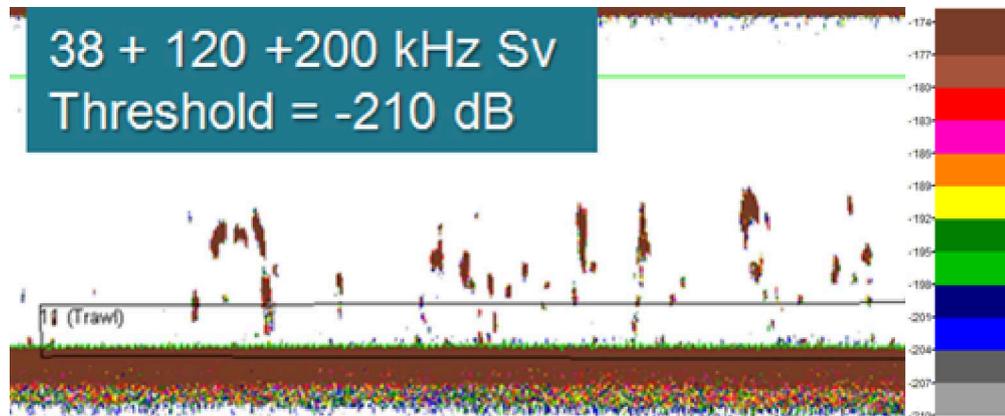


Figura 2.4: Ecogramma risultante dalla somma delle tre frequenze con il minimo valore soglia impostato (-210 dB).

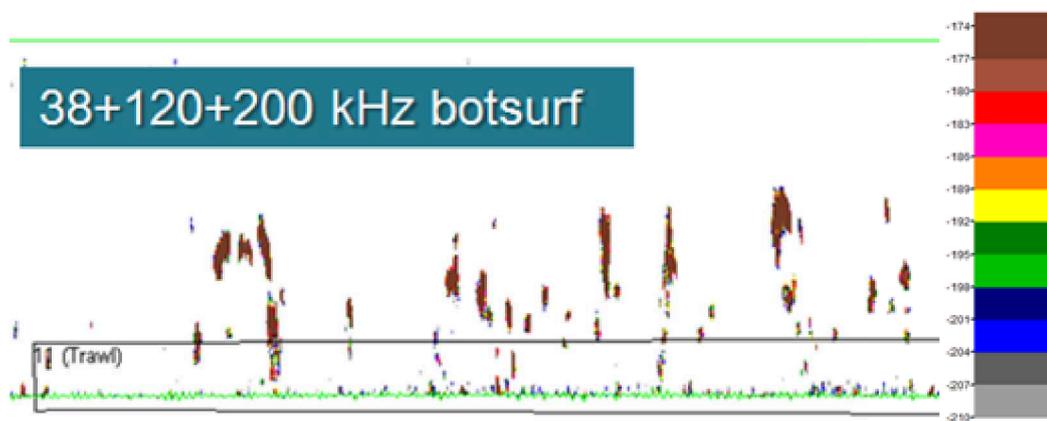


Figura 2.5: Applicazione di un filtro che elimina il rumore al sopra della linea superficiale e al di sotto del fondo.

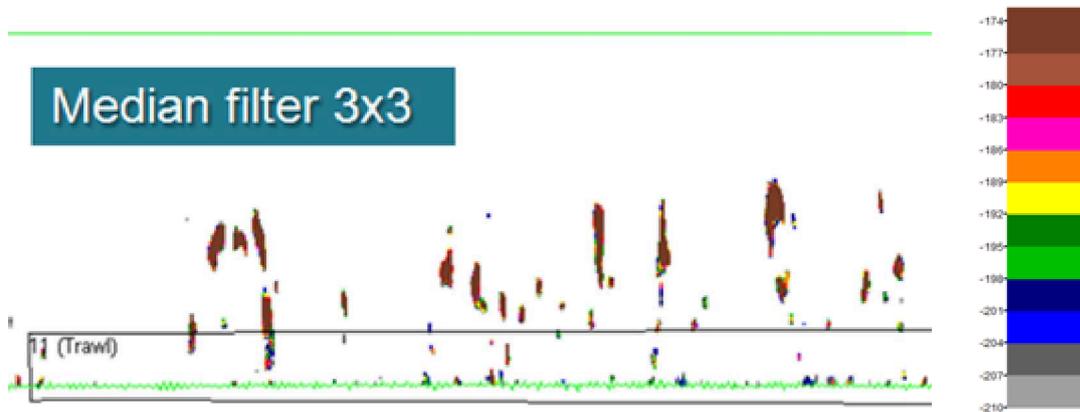


Figura 2.6: Applicazione di un filtro mediano con elemento strutturale 3x3.

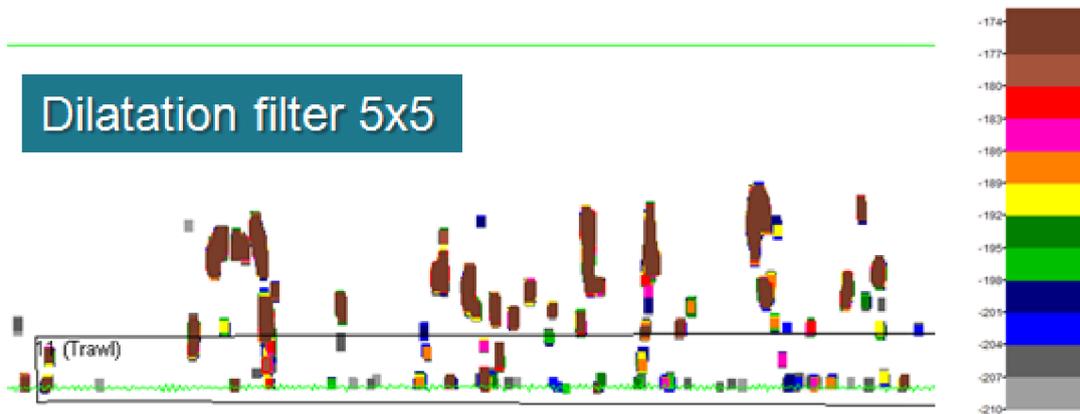


Figura 2.7: Applicazione di un filtro di dilatazione con elemento strutturale 5x5.

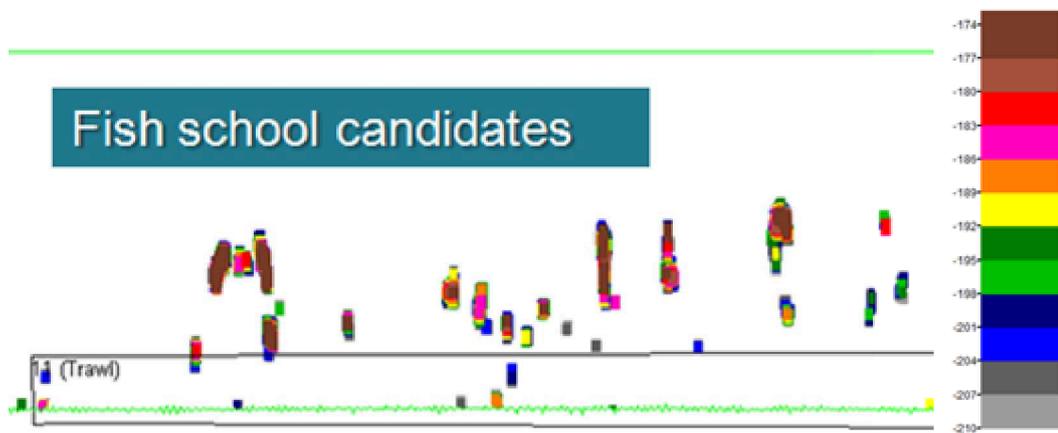


Figura 2.8: Applicazione del nuovo filtro soglia a -175dB e visualizzazione delle ecotracce candidate ad essere considerate banchi.

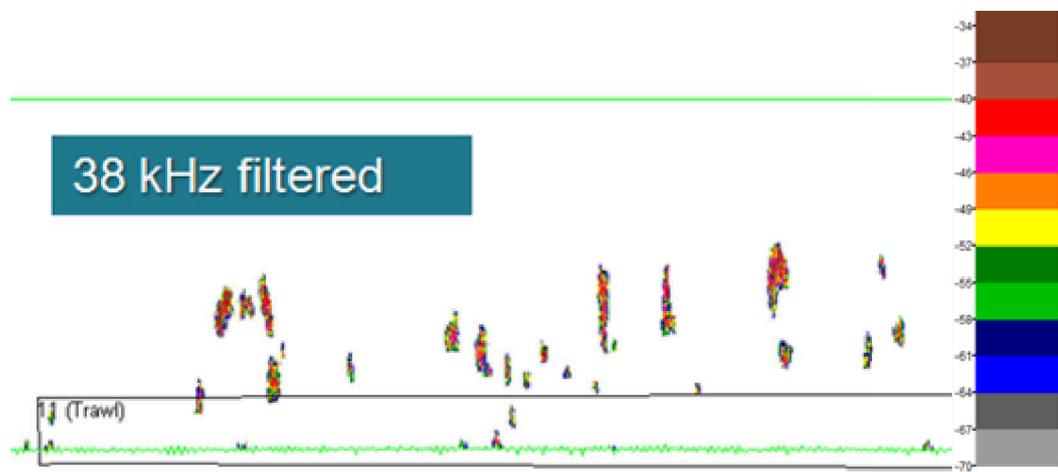


Figura 2.9: Ecogramma finale con le ecotracce isolate e soglia minima di visualizzazione di -70 dB .

- Individuazione e isolamento delle eco tracce corrispondenti a potenziali banchi di pesce.

Questa seconda fase viene eseguita sull'ecogramma a 38kHz risultante dalla precedente operazione di filtraggio e utilizza l'algoritmo SHAPE presente in Echoview (Barange, 1994). I parametri che l'algoritmo adopera per l'identificazione dei banchi sono: lunghezza minima totale 10 m, altezza minima 1 m, lunghezza minima di un probabile banco 5 m, altezza minima di 1 m, distanza verticale di collegamento 2 m, massima distanza orizzontale di collegamento di 15 m. Tutte le eco tracce isolate con questa procedura e che ricadono all'interno dell'area campionata dalla rete vengono così selezionate come oggetti per le successive analisi di classificazione.

- Estrazione dei parametri energetici, morfologici e batimetrici.

I parametri energetici, sono stati ricavati dagli originali ecogrammi a differenti frequenze, ed estratti utilizzando la procedura di export in Echoview. Mentre i parametri batimetrici e morfologici (che non cambiano al variare della frequenza) sono stati estratti dalla singola frequenza a 38 kHz. Utilizzando poi uno script elaborato in ambiente R, dalle variabili morfologiche esportate, sono stati derivati i seguenti parametri relativi, calcolati tramite opportune formule ricavate da letteratura (vedi Tab. 2.3): Dimensione frattale (indice della complessità della forma del banco, *Fractal dimension*), Elongazione (rapporto lunghezza altezza, *Elongation*), Rettangolarità (indice della relazione tra l'area del banco stimata e l'area del rettangolo che il banco, avente una certa lunghezza e altezza, dovrebbe formare, *Rectangularity*), Variabilità della superficie del banco (*Unevenness2*), indice della relazione tra il perimetro del banco e il perimetro del rettangolo che il banco, avente una certa altezza e lunghezza, dovrebbe formare), Variabilità della superficie del banco normalizzata per il numero di ping (*Unevenness1*). Alle variabili energetiche è stata invece aggiunta la variabile ddB120 (differenza tra valori medi di Sv a 38 kHz e a 120 kHz).

Tabella 2.3: Parametri estratti con l'algoritmo SHAPE di Echoview o da essi derivati tramite script in R.

Variabile	Descrizione
A.morfologica	
Corrected length [m]	Lunghezza della traccia acustica associata ad un banco corretta come richiesto da Diner (2001)
Corrected thickness [m]	Altezza della traccia acustica associata ad un banco corretta come richiesto da Diner (2001)
Corrected Area [m ²]	Area della traccia acustica associata ad un banco corretta come richiesto da Diner (2001)
Corrected Perimeter [m]	Perimetro della traccia acustica associata ad un banco corretto come richiesto da Diner (2001)
Image compactness	Stima della compattezza dell'immagine calcolata come rapporto tra il raggio di un cerchio avente lo stesso perimetro del banco osservato e il raggio di un cerchio avente l'area del banco osservato
3D_School_volume [m ³]	Volume calcolato assumendo che la forma del banco sia cilindrica
3D_School_area [m ²]	Area calcolata assumendo che la forma del banco sia cilindrica
Beam_volume_sum [m ³]	Volume di un banco calcolato come volume dei ping che compongono l'ecotraccia
Unevenness1 [m] (Weill <i>et al.</i> , 1993)	Indice della variabilità della forma del banco normalizzata per il numero di ping
Unevenness2 [m] (Weill <i>et al.</i> , 1993)	Indice della variabilità della forma del banco non normalizzata per il numero di ping
Rectangularity [m]	Indice della forma del banco espresso come rapporto tra il prodotto della lunghezza e dell'altezza del banco osservato e l' area del banco
Fractal dimension [m] (Nero and Magnuson, 1989)	Indice della complessità della forma di un banco
Elongazione [m] (Weill <i>et al.</i> , 1993)	Indice della forma del banco espresso come rapporto tra valori di lunghezza e altezza del banco corretti come richiesto da Diner (2001)
B.energetica	
Sv_mean [dB re m^{-1}]	Energia retrodiffusa in volume registrata dal software Echoview per ogni ecotraccia
Skewness	Skewness dell'energia acustica
Kurtosis	Kurtosi dell'energia acustica
Horizontal roughness	Misura statistica della dispersione dell'energia di un banco nella direzione orizzontale
Vertical roughness	Misura statistica della dispersione dell'energia di un banco nella direzione verticale
ddB120 [dB re m^{-1}]	Differenza tra l'energia retro diffusa a 38 kHz e l'energia a 120 kHz
C.batimetrica	
Depth_mean [m]	Profondità media del banco
Mean_distance_bottom [m]	Distanza media del banco dal fondo calcolata come differenza tra la profondità del fondo rilevata dallo strumento di acquisizione e la profondità media del banco

2.2.3 Metodo acustico per la classificazione delle proprietà del fondo marino

Per identificare e classificare i diversi tipi di fondo vengono impiegate oggi, principalmente, due strumentazioni che fanno uso di echosounder a single-beam a trasmissione verticale: QTC-view and RoXAnn. Queste analizzano e misurano la forma e le caratteristiche energetiche contenute nel segnale proveniente dal fondo (Prager *et al.*, 1995; Greenstreet *et al.*, 1997).

Echosounders come EK 500 e EK 60 sono anch'essi strumenti a single-beam come QTC e Rox-Ann ma sono comunemente utilizzati per l'analisi di banchi di specie pelagiche e la valutazione delle stime di biomassa. Entrambi i modelli (EK500 ed EK60) sono in grado di acquisire contemporaneamente informazioni dalla colonna d'acqua e dal fondo marino. Tuttavia, negli studi che hanno come obiettivo le stime di biomassa pelagica, il segnale acustico dal fondo è solitamente scartato. L'uso di una singola strumentazione, per l'acquisizione contemporanea di misure sulla colonna d'acqua e sul fondo, risulterebbe estremamente utile al fine di studiare i principi di funzionamento dell'ecosistema pelagico mediante approccio interdisciplinare. Inoltre avere la possibilità di effettuare misure con la stessa strumentazione su domini differenti permetterebbe di ridurre notevolmente i costi, viste le ingenti risorse umane e finanziarie che richiedono le campagne di ricerca oceanografiche, siano esse con obiettivi scientifici o indirizzi applicativi.

Numerosi approcci sono stati utilizzati da strumentazioni a single beam o multi beam per il riconoscimento delle caratteristiche del fondo marino. Uno di questi è quello oggi usato dal sistema RoxAnn, che consiste nel stimare le proprietà energetiche del primo e del secondo eco proveniente dal fondo. Orłowski (1984) e Chivers *et al.* (1990) sono stati i primi ad aver usato le caratteristiche energetiche contenute nei due echi dal fondo, come descrittori delle caratteristiche del substrato marino. Heald and Pace (1996) invece hanno fornito per primi una descrizione teorica della relazione tra proprietà energetiche dei due echi e i parametri del fondo. Recentemente le tecniche acustiche sono state utilizzate anche per mappare le comunità bentoniche. Ne sono un esempio gli studi di Magorrian *et al.* (1995), Greenstreet *et al.* (1997), Kaiser *et al.* (1998), Sorensen *et al.* (1998), Siwabessy *et al.* (1999), che hanno usato il sistema commercial RoxAnn, e infine Prager *et al.* (1995), che hanno invece fatto uso del sistema QTC-view. Più recentemente il sistema a single beam RoxAnn e il sistema a multibeam QTC sono stati adoperati contemporaneamente per la discriminazione del fondo marino e la classificazione degli habitat nell'area di Stoneheaven in Scozia (Serpetti *et al.*, 2011). La tecnica utilizzata dal sistema Rox-Ann, a cui si fa riferimento in questa tesi, consiste nell'utilizzare la coda del primo eco per caratterizzare le proprietà di rugosità del fondo e l'intero secondo eco (l'eco che subisce due riflessioni al fondo e un singolo scattering all'interfaccia aria-acqua) per ottenere informazioni sulla durezza (Fig. 2.10). In generale, laddove l'epibenthos non produce uno scattering significativo, il primo eco dal fondo è composto da almeno tre componenti: 1) la riflessione prodotta dalla superficie perpendicolare all'asse del trasduttore (porzione del segnale che impiega meno tempo per raggiungere il ricevitore). Questa parte del primo eco detta componente coerente costituisce il picco del primo segnale dal fondo ed è molto sensibile ai movimenti di rollio e beccheggio; 2) la riflessione obliqua prodotta dalla superficie esterna insonificata dal trasduttore (componente incoerente); 3) possibili riflessioni dal sub-bottom (riverbero

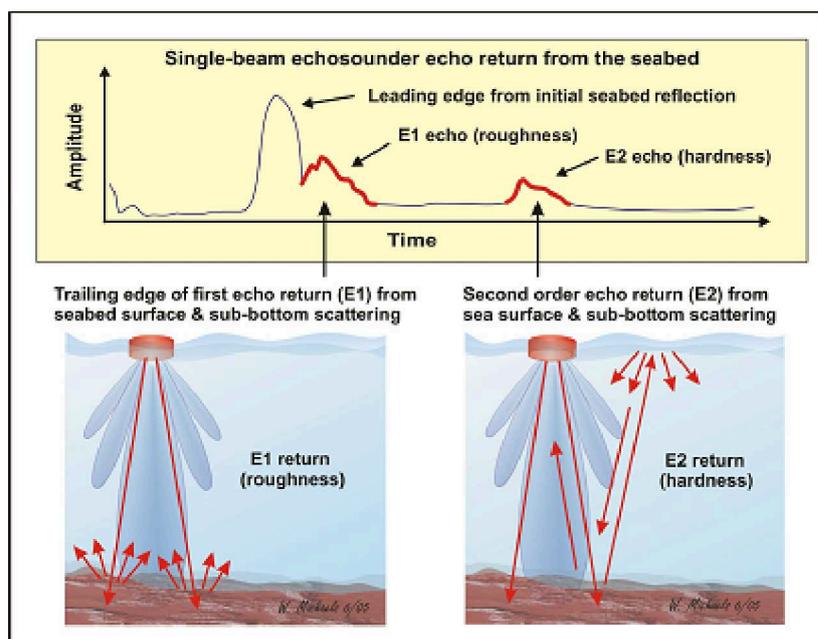


Figura 2.10: Ampiezza del segnale registrato da un echosounder in funzione del tempo. La coda del primo eco è il risultato della riflessione obliqua ed è indice di rugosità del fondo, mentre l'intero secondo eco risultante da una singola riflessione in superficie e una doppia riflessione al fondo, fornisce informazioni sulle proprietà di durezza del fondo (ICES report 2007).

del sub-bottom).

Le prime due componenti del primo eco possono essere comprese meglio analizzando l'interazione di un singolo impulso con il fondo (Fig. 2.11).

A causa della naturale curvatura del fronte d'onda emesso da un echosounder con ampio beam-angle, l'impulso insonifica sul fondo prima un'area corrispondente a un cerchio e poi progressivamente anelli di raggio crescente. La prima parte del segnale che raggiunge il ricevitore è un picco dominato principalmente dalla riflessione speculare e prodotta dall'area corrispondente al cerchio. La seconda parte, o coda del primo eco, è la riflessione prodotta dall'area ad anello e rappresenta il contributo della riflessione obliqua. La lunghezza e l'energia contenuta nel primo eco sono i parametri che forniscono informazioni sulla rugosità del fondo. Se il fondo è perfettamente smussato, l'energia trasmessa verticalmente tornerà al ricevitore con una maggiore o minore intensità in dipendenza del grado di penetrazione nel sedimento e l'energia ricevuta da altri angoli verrà riflessa via. In questo caso la coda del primo eco sarà stretta o assente. Se invece il fondo non è totalmente piatto una parte dell'energia riflessa da altre angolazioni ritornerà al trasduttore. Più è rugoso il fondo e più lunga ed energetica apparirà la coda del primo eco.

A destra della Fig. 2.10 è invece riportato il secondo eco, il quale si assume che abbia un tempo di arrivo al ricevitore che è approssimativamente 2 volte il tempo di ritorno del primo. Il secondo eco è quindi una copia ritardata del primo, in cui però gli effetti della rugosità diventano secondari.

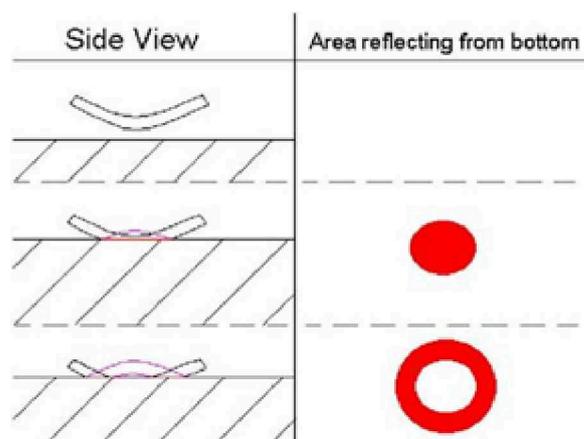


Figura 2.11: Interazione di un ping o impulso acustico con il fondo. Il lato sinistra dell'immagine fornisce una visione laterale dell'energia riflessa da un fondo orizzontale raggiunto da un impulso acustico, mentre il lato destro mostra una sezione trasversale dell'impulso a contatto con il fondo in un determinato istante. Nella sezione centrale, la parte perpendicolare al trasduttore dell'impulso acustico raggiunge il fondo ad un tempo t_1 e insonifica un'area circolare, mentre nella sezione in basso la parte esterna dell'impulso acustico raggiunge il fondo ad un tempo t_2 , insonificando un'area ad anello

L'ammontare di energia contenuta nel secondo eco è legata esclusivamente alla differenza di impedenza acustica tra la superficie del mare e il fondo. Più il fondo è duro maggiore sarà l'energia trasmessa alla superficie e da questa al trasduttore. In genere sedimenti ad alta porosità come silt e argilla hanno una bassa impedenza acustica, cioè riflettono e disperdono poca energia rispetto a materiali ad alta densità come roccia e ghiaia, che hanno coefficienti di riflessione più alti (Hamilton, 2001). Heald & Pace (1996) affermano che il percorso compiuto dal secondo eco sarebbe equivalente ad una configurazione bi-statica, cioè vuol dire con trasduttore e ricevitore disposti in modo tale che la distanza verticale sia due volte la profondità della colonna d'acqua. La conseguenza di questa configurazione è che il ricevitore si trovi nel campo vicino del segnale emesso dal fondo e che l'energia ricevuta sia quindi costituita principalmente dalla riflessione. Le riflessioni da superfici smussate sono sempre per definizione nel campo vicino. L'energia del secondo eco risulta quindi direttamente proporzionale al coefficiente di riflessione del materiale e non contiene un'informazione significativa sulla rugosità.

L'interpretazione dei risultati però è tutt'altro che semplice visto la dipendenza del backscatter del fondo da altri parametri come la pendenza o le caratteristiche della strumentazione acustica utilizzata. Kloser et al. (2001) e Schlagintweit (1993) hanno osservato una dipendenza della classificazione del fondo dalla frequenza. Per fondi con le stesse caratteristiche, differenti indici di rugosità sono stati osservati a frequenze diverse. Schlagintweit (1993) trovò che le differenze da lui osservate tra la frequenza a 40 kHz e la frequenza a 208 kHz erano legate a differenze nella profondità di penetrazione di queste ultime. Ma la frequenza influenza anche la risoluzione del segnale. A basse frequenze dove la lunghezza d'onda è maggiore rispetto alla scala di rugosità del fondo, la superficie del fondo apparirà acusticamente smussata. In questo caso la riflessione dominerà sullo scattering.

Dall'altro lato con alte frequenze e lunghezze d'onda più piccole della scala di rugosità del fondo lo scattering può essere dominante e il fondo può essere considerato acusticamente rugoso. Inoltre il backscatter da altri angoli e la riflessione del sub-bottom a basse frequenze potrebbero arrivare contemporaneamente, rendendo più difficile l'interpretazione dell'informazione contenuta nella coda del secondo eco. Hamilton et al. (1999) e Kloser et al. (2001) hanno inoltre evidenziato un errore in corrispondenza di cambiamenti di pendenza del fondo o di improvvisi cambiamenti di profondità con i dati di RoXAnn. Alte pendenze o improvvisi cambiamenti di profondità del fondo producono lunghe code nel primo eco stimando quindi alti indici di rugosità. Kloser et al. (2001) inoltre hanno trovato che ampiezze del beam più piccole sono più sensibili ai cambiamenti di profondità. Von Szalay & McConnaughey (2002) hanno riscontrato un effetto della pendenza ed anche della velocità dell'imbarcazione sul sistema di classificazione. I risultati infatti indicano che velocità comprese fra 3 e 12 nodi non hanno un effetto significativo sulle performance di classificazione, ma pendenze che eccedono i $5-6^\circ$ determinano una riduzione nell'accuratezza della classificazione. Ampiezze del beam maggiori forniscono migliori risultati in termini di rugosità perché minimizzano il potenziale contributo del riverbero dal sub-bottom. Anche il grazing-angle o angolo di incidenza può fortemente influenzare il livello di backscatter. Ad angoli di incidenza perpendicolari o vicini alla verticale il backscatter è dominato dalla riflessione speculare e dalla rugosità a grande scala, mentre per angoli di incidenza maggiori il backscatter è dominato dalla rugosità a piccola scala e dall'eterogeneità del sedimento.

A parità di caratteristiche acustiche dello strumento utilizzato (beam angle, frequenza, pulse length, grazing angle), i fattori che possono interferire con il primo eco sono:

- Pendenza del fondo;
- Movimenti della piattaforma (nel nostro caso la stessa imbarcazione, essendo l'echo-sounder ancorato alla chiglia della nave);
- L'effetto delle bolle d'aria in superficie.

I fattori che possono interferire con il secondo eco sono:

- Rugosità della superficie del mare;
- Velocità dell'imbarcazione;
- Dimensione, forma e profondità dello scafo;
- Background noise (Essendo il segnale del secondo eco di ridotta intensità (circa 10-20 dB sotto il primo eco) è più sensibile al rumore dell'ambiente).

2.2.4 Procedura per l'estrazione delle informazioni relative al fondo

Per estrarre le informazioni di rugosità si assume che la riflessione obliqua sia il solo contributo alla coda del primo eco. In realtà la riflessione obliqua compete con il riverbero dal sub-bottom (Chivers *et al.*, 1990). L'ampiezza del beam e la naturale caratteristica del fondo (compreso l'assorbimento acustico) determina, nel primo eco, la relativa intensità

della riflessione obliqua e del riverbero dal sub-bottom. È quindi necessario utilizzare ampiezze del beam opportune e isolare quanto più possibile la componente incoerente, legata esclusivamente alla riflessione obliqua per ottenere un'accurata misura di rugosità. Questo può essere ottenuto individuando i due intervalli di integrazione che corrispondono all'inizio e alla fine della coda del primo eco.

Questi due intervalli (indicati come d_{θ_a} e d_{θ_b} in Fig. 2.12) nel software Sonar Data Echoview vengono calcolati sulla base di due differenti angoli di incidenza della riflessione obliqua. L'intervallo d_{θ_a} (in Fig. 2.12) più la lunghezza dell'impulso (o un multiplo di questo) corrisponde alla profondità di partenza, al di sotto della quale si stima che il riverbero dal sub-bottom sia trascurabile. L'intervallo d_{θ_b} corrisponde alla profondità sotto la quale il segnale di backscatter si confonde con il rumore di fondo. La scelta di un opportuno θ_a ha l'effetto di rimuovere il picco del primo eco, il quale è generalmente diversi decibel sopra il livello della coda. La procedura per l'individuazione dei due angoli, ripresa da Siwabessy (2001), richiede sperimentalmente l'identificazione della profondità alla quale il backscatter (o valore di Sv) si è ridotto di 45 dB rispetto al picco massimo, ad ognuna delle frequenze a cui si opera. 45 dB sotto il picco massimo è il valore stimato da Siwabessy (2001) al di sotto del quale il riverbero risulta trascurabile. I valori di profondità individuati vengono messi in grafico in relazione alla profondità del fondo e ad essi si applica una semplice regressione lineare mediante la seguente relazione:

$$d_{\theta_a} = mR_0 + c\tau \quad (2.1)$$

dove $m = \sec(\theta_a)$, R_0 = profondità del fondo, $c\tau$ = l'offset, ovvero la lunghezza dell'impulso o un multiplo di essa. La regressione lineare viene quindi forzata a intercettare un valore che dipende dalla lunghezza dell'impulso o un multiplo di essa (in genere 3) e un nuovo angolo corrispondente a θ_a sarà individuato dalla pendenza della retta.

La stessa procedura deve essere applicata per ricavare θ_a , anche se la scelta dell'angolo necessario a individuare il termine della coda non è così critico come l'individuazione dell'angolo di partenza.

Utilizzando questa procedura Siwabessy (2001) ha stimato un intervallo di integrazione compreso tra 14° e 40° con un offset pari a 3 m per ampiezze del beam di 12° . Le profondità corrispondenti a questi due angoli variano in relazione alla profondità della colonna d'acqua; aumentando infatti la profondità aumenta l'area insonificata. Echoview possiede un algoritmo in grado di stimare automaticamente la variazione delle profondità corrispondenti a θ_a e θ_b al variare della profondità della colonna d'acqua.

L'informazione sulla durezza è invece ottenuta dall'integrazione completa del secondo eco. Il secondo eco contiene entrambe le componenti coerente e incoerente, tuttavia la componente coerente è molto più grande rispetto alla componente incoerente che viene quindi considerata trascurabile. La procedura per l'estrazione del secondo eco prevede intervalli di integrazione costanti. L'intervallo di partenza è definito come 2 volte la profondità della colonna d'acqua (d_a), mentre l'intervallo finale è definito come 2 volte la profondità della colonna d'acqua più 30 m (d_b); vedi figura 2.12 per la geometria del segnale. Nei tempi della stesura della presente tesi non è stato tuttavia possibile estrarre informazioni sulla rugosità del fondo.

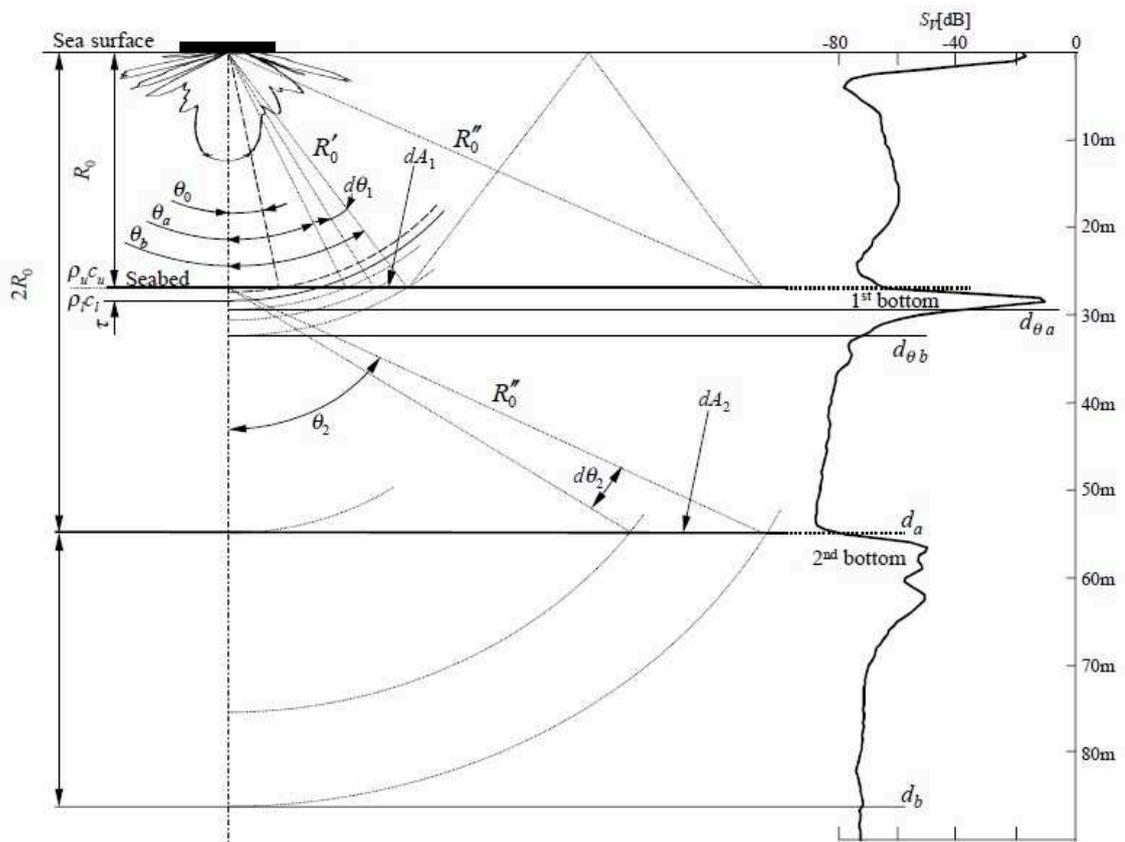


Figura 2.12: Geometria del segnale di ritorno dal fondo. d_{θ_a} e d_{θ_b} sono le profondità delle riflessioni oblique corrispondenti agli angoli θ_a θ_b mentre d_a e d_b sono gli intervalli costanti di integrazione del secondo eco (Siwabessy, 2001).

Gli intervalli di integrazione riportati in letteratura fanno riferimento ad ampiezze del beam maggiori di 8° (in genere da 10° in su). Una procedura sperimentale, che non è stato possibile mettere a punto nei tempi del progetto, sarebbe stata necessaria per individuare gli intervalli che consentono di isolare la coda del primo eco. La procedura per l'estrazione delle proprietà di durezza dal secondo eco è stata invece effettuata mediante il software Echowiew sui tracciati acustici delle cale monospecifiche e sull'intero survey Ancheva2009. I valori di durezza del fondo sono stati espressi come valori medi di NASC calcolati per intervalli di 0.05 nmi. Analisi granulometriche su campioni di sedimento, raccolti in prossimità dei transetti costa-largo di Ancheva2009, sono state utilizzate per verificare la corrispondenza tra segnale acustico e durezza del fondo.

2.2.5 Raccolta e selezione di dati idrologici

Al fine di analizzare l'effetto dei parametri ambientali sulla discriminazione della specie sono stati selezionati i profili CTD più vicini ai transetti acustici delle cale monospecifiche. Da

ogni profilo sono stati estratti i seguenti parametri: profondità del termoclino, temperatura e salinità media sopra il termoclino, temperatura e salinità media sotto il termoclino. Tuttavia solo i parametri sopra il termoclino sono stati presi in considerazione essendo alta la correlazione tra la profondità del banco e i parametri ambientali medi misurati sotto il termoclino.

2.2.6 Analisi statistiche: statistica descrittiva e classification tree

La statistica descrittiva è stata utilizzata nel secondo approccio per analizzare le caratteristiche morfologiche e batimetriche medie dei banchi per specie che compongono il learning sample costituito da un totale di 484 banchi di cui 155 banchi di sardine, 111 di acciughe, 21 di demersali e 197 OPS. La significatività delle differenze fra tutte le possibili coppie di specie o i gruppi di specie è stata valutata mediante test di Kruskal-Wallis.

Per le sole specie target, è stata verificata, mediante il test U di Mann-Whitney, la significatività delle differenze tra i valori medi di Sv per le due frequenze e la ddb120. Successivamente, per analizzare l'effetto della taglia degli individui sui livelli di backscatter, le cale con individui giovanili per le due specie target sono state separate dalle cale con gli individui adulti. Mentre l'effetto della profondità sui livelli di backscatter è stato valutato suddividendo, le eco tracce del dataset (155 e 111 rispettivamente per sardine e acciughe), in due gruppi ricadenti nei seguenti range di profondità: 40-60 m e 60-80 m, individuati sulla base della profondità media del banco. Le differenze in decibel delle eco tracce, corrispondenti a questi tre range, sono stati esaminati e confrontati mediante il test U di Mann-Whitney.

In secondo tempo modelli di classificazione ad albero, sviluppati in ambiente R, tramite la libreria "rpart", sono stati adoperati per comprendere quali variabili strutturali del banco e/o parametri ambientali meglio caratterizzano e discriminano tra le due specie. Tali modelli sono stati applicati al dataset completo di 484 banchi. Per analizzare invece l'effetto del fondo è stato necessario ridurre il dataset completo. Sono state utilizzate infatti esclusivamente cale monospecifiche dai survey, Ancheva2005, Ancheva2007, Ancheva2008, e Ancheva2009 per cui erano disponibili le informazioni sul secondo eco. I modelli decisionali che includono classification o regression tree sono considerati una delle tecniche statistiche più promettenti per la discriminazione della specie. I campi di applicazione delle metodologie ad albero comprendono i settori più disparati, che spaziano da quello socio-economico (es., marketing, qualità, etc.) a quello bio-sanitario (es. biometria, oncologia, immunologia, genetica, etc.) e botanico (classificazione di specie vegetali). In ambiente marino questa tecnica è stata applicata per la prima volta ad un problema analogo a quello dell'identificazione della specie, che consisteva nel riconoscimento di classi di imbarcazioni mediante lo studio dei profili radar (Hooper & Lucero, 1976). Alberi decisionali sono stati applicati anche in campo ecologico da Franklin (1995), Guisande & Zimmermann (2000). Le classificazioni ad albero si basano sulla ripartizione dell'insieme di osservazioni da esaminare in sottoinsiemi che raccolgono osservazioni, le quali presentano valori uguali o simili per qualche attributo considerato importante e significativo. Normalmente un albero di decisione viene costruito utilizzando un dataset di cui è nota la composizione (learning sample), il quale è diviso in due sottoinsiemi: il training set sulla base del quale

si crea la struttura dell'albero e il test set che viene utilizzato per testare l'accuratezza del modello di classificazione. I modelli decisionali possono essere applicati a variabili predittive continue e categoriche, anche contemporaneamente, mentre la variabile risposta è una variabile continua nel caso dell'applicazione di modelli di regressione ad albero (o regression tree) e categorica nell'applicazione di modelli di classificazione ad albero (o classification tree). Entrambi i modelli sono tecniche non parametriche, per cui non è necessario testare la normalità o altre ipotesi che riguardano la distribuzione statistica dei dati e che generalmente richiedono a priori una conoscenza del dataset da analizzare. Questa tecnica consente di trattare strutture di dati molto complesse producendo in uscita dei grafici facilmente interpretabili. I risultati sono infatti presentati sotto forma di diagramma ad albero rovesciato costituito da parent node o nodi genitori i quali hanno due discendenti diretti o child node e nodi terminali che non subiscono ulteriori bipartizioni. In una classification tree ogni nodo terminale avrà come variabile risposta la variabile categorica più frequente, mentre in una regression tree al nodo terminale sarà assegnata la media delle variabile risposta continua. La suddivisione in sottogruppi viene effettuata sulla base del predittore più rilevante. Quest'ultimo è rappresentato dalla quella variabile e dal suo valore soglia in grado di produrre due sottogruppi molto omogenei al loro interno (o con basso grado di impurità) ma tanto più differenti tra loro. Diversi indici possono essere utilizzati per misurare il grado di impurità o omogeneità che serve a individuare il predittore candidato allo sviluppo dell'albero. Alcuni software utilizzano il calcolo della devianza fra gruppi e all'interno dei gruppi altri invece preferiscono utilizzare l'indice di Gini o l'errore nella misclassificazione (cioè la proporzione di osservazioni misclassificate) (Breiman *et al.*, 1984). Lo sviluppo dell'albero viene poi bloccato o quando i risultanti nodi terminali hanno un numero minimo di osservazioni (5) oppure quando un'ulteriore suddivisione in sottogruppi non comporta più una diminuzione significativa dell'errore. In ambiente R, il pacchetto "R-part", di cui si fa uso in questa tesi per l'applicazione di modelli decisionali, impiega di default un parametro di complessità pari a 0.01 per arrestare lo sviluppo del modello. Il valore del parametro di complessità indica la minima diminuzione dell'errore nella classificazione che deve produrre una suddivisione per essere presa in considerazione. Quindi se il predittore scelto produce una riduzione dell'errore inferiore a 0.01 lo sviluppo dell'albero viene arrestato.

Come in una regressione lineare bisogna però trovare il modello decisionale ottimale. Alberi troppo grandi sono poco comprensibili perché presentano molti nodi terminali con poche osservazioni al loro interno e molte informazioni da interpretazione. Inoltre modelli di grande taglia soffrono spesso di over-fitting ossia di un adattamento eccessivo dell'albero ai dati in esame, che rende poi difficilmente generalizzabili i risultati ottenuti. Allo stesso modo un albero di piccole dimensioni produce pochi gruppi, con poche osservazioni al loro interno. Risultando quindi facilmente interpretabile ma con un grado di adattamento scarso. Esistono quindi strumenti, come L'AIC (Akaike's Information Criterion) per la regressione lineare, che consentono di capire qual è la taglia dell'albero ottimale in termini di semplicità nell'interpretazione e di fit. Lo strumento migliore per la scelta del modello ottimale è far uso una di tecnica di pruning. Il principio della tecnica consiste nell'utilizzo del metodo della cross-validation per individuare il parametro di complessità (cp) che

consente di ottenere l'albero ottimale in termini di taglia, e utilizzare il cp trovato (correlato alla taglia dell'albero) per eliminare i nodi che non contribuiscono ad un incremento significativo dell'accuratezza (vale a dire un aumento del numero di osservazioni classificate correttamente rispetto al numero totale di osservazioni). Nel dettaglio la tecnica di pruning consiste nell'utilizzare il 90% delle osservazioni presenti nel training set per stimare l'albero e il restante 10% per testarlo. Il processo è applicato dieci volte, su dieci training set e test set differenti, fino a stimare dieci valori di errore su cui è possibile calcolare un valore medio di errore nella predizione e la sua deviazione standard. Questa operazione è poi ripetuta per differenti taglie di albero e valori di cp. Infine la regola nota come "1-deviazione standard (1-SE)" è adoperata per selezionare la "giusta taglia" del modello (Breiman *et al.*, 1984).

Capitolo 3

Risultati

3.1 Comportamento e distribuzione di banchi di pesci pelagici in relazione alla disponibilità di cibo nel Canale di Sicilia

In questo paragrafo l'attribuzione della specie alle aggregazioni acusticamente individuate è stata affrontata studiando, per zone, il comportamento giornaliero di banchi di pesce in relazione alla distribuzione del plancton. Questa analisi preliminare ha riguardato i dati giornalieri raccolti nel corso della campagna ANCHEVA98, condotta nel mese di Luglio 1998 (v. paragrafo 2.2).

Una caratterizzazione delle aggregazioni di plancton e dei banchi di pesce, individuati con una sola frequenza nelle tre differenti zone, è riportata in Tab. 3.1. I parametri che ne descrivono la struttura sono quelli estratti da ecogrammi a singola frequenza. I banchi di pesce appaiono più energetici (NASC e NASC/AREA) e più profondi delle aggregazioni planctoniche. Di contro le aggregazioni planctoniche mostrano maggiori valori di lunghezza, elongazione e distanza dal fondo rispetto ai banchi di pesce.

La relazione preda-predatore è stata esaminata mediante Modelli Additivi Generalizzati (GAMS). In Fig. 3.1 sono riportati i risultati dei GAM che mostrano un'interazione significativa tra le due variabili. Il primo modello (Fig. 3.1a) riporta l'effetto della biomassa planctonica (in logaritmo) dell'intera colonna d'acqua sulla biomassa media dei banchi di pesce. Per bassi valori di biomassa planctonica il modello mostra un'interazione negativa tra le due variabili, cioè a un incremento della biomassa media del predatore si osserva un decremento della biomassa della preda. Per alti valori di biomassa planctonica il trend risulta opposto, ad un aumento della biomassa del predatore corrisponde un incremento della biomassa della preda. Una situazione analoga è stata riscontrata analizzando l'interazione tra biomassa planctonica e densità media dei banchi di pesce, che è risultata negativa per bassi valori di biomassa e positiva per valori di biomassa più elevati (Fig. 3.1b). Il grafico che mostra la densità delle due variabili indica, invece, un'interazione positiva per alti valori di densità planctonica e l'assenza di interazione per bassi valori di densità (Fig. 3.1c). Un riepilogo dei modelli sopra descritti è riportato in Tab. 3.2.

Tabella 3.1: Valori medi dei descrittori di banchi di pesce e di aggregazioni planctoniche nell'area di studio (Zone 1, Zone 2 e Zone 3). Per ogni descrittore è riportato tra parentesi l'errore standard.

Variable	Fish schools				Plankton patches			
	1	2	3	TOT	1	2	3	TOT
<i>N</i>	131	188	159	478	1683	281	860	2824
<i>N/nmi</i>	2.91	4.70	2.74	3.34	37.40	7.03	14.83	19.75
Bottom depth (m)	78.86 (19.55)	68.76 (12.56)	107.52 (15.22)	84.42 (8.91)	90.78 (20.31)	98.72 (21.94)	140.29 (28.00)	106.65 (14.97)
Distance from the bottom (m)	5.61 (1.45)	7.31 (1.50)	6.19 (1.07)	6.47 (0.80)	36.75 (8.89)	39.58 (10.44)	91.17 (18.47)	53.60 (7.81)
Area (m ²)	75.39 (18.81)	78.19 (16.99)	77.10 (13.21)	77.06 (9.52)	538.98 (136.40)	316.71 (123.02)	272.24 (59.34)	435.63 (84.20)
Length (m)	16.09 (3.16)	19.19 (3.82)	21.59 (3.32)	19.14 (2.06)	51.26 (10.32)	44.35 (11.91)	48.97 (9.75)	49.88 (6.93)
Height (m)	4.28 (0.86)	4.24 (0.86)	3.64 (0.54)	4.05 (0.45)	6.98 (1.30)	5.46 (1.13)	5.64 (1.14)	6.42 (0.86)
Elongation (m)	4.14 (0.78)	5.24 (0.99)	6.49 (0.96)	5.35 (0.55)	5.21 (0.98)	6.64 (1.35)	7.78 (1.39)	6.14 (0.73)
Depth (m)	71.11 (18.25)	59.33 (10.91)	99.51 (14.40)	75.92 (8.19)	50.55 (11.20)	56.40 (12.90)	46.30 (9.91)	49.83 (7.44)
NASC (m ² nmi ⁻²)	3885.52 (909.79)	4595.18 (1084.53)	2029.53 (423.54)	3547.27 (516.27)	52.21 (10.28)	164.57 (35.88)	63.18 (14.49)	66.92 (8.37)
NASC/AREA (nmi ⁻²)	105.94 (30.83)	92.75 (23.52)	38.93 (10.80)	78.46 (13.10)	0.53 (0.12)	1.75 (0.41)	0.60 (0.14)	0.67 (0.09)
Mean Sv (dB)	-50.71 (10.14)	-47.81 (8.92)	-50.64 (7.28)	-29.40 (3.89)	-72.54 (13.10)	-68.07 (13.24)	-73.18 (13.71)	-72.29 (8.95)

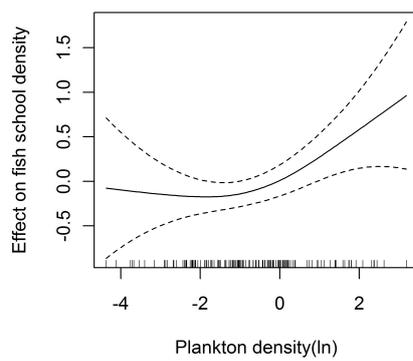
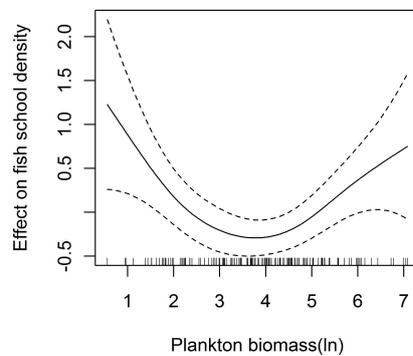
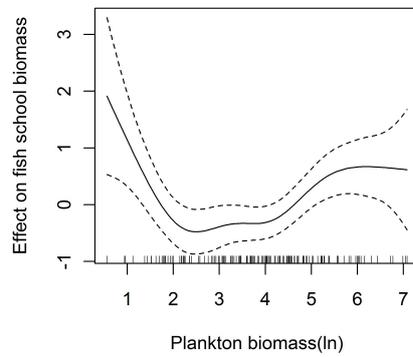


Figura 3.1: Effetto della densità e della biomassa planctonica sulla variabile risposta: (a) biomassa dei banchi di pesce vs. $\ln(\text{biomassa di plancton})$; (b) densità dei banchi di pesce vs. $\ln(\text{biomassa di plancton})$; (c) densità dei banchi di pesce vs. $\ln(\text{densità di plancton})$. La biomassa e la densità di plancton si riferiscono alle aggregazioni individuate nell'intera colonna d'acqua. La linea continua indica i valori stimati dal GAM. La linea tratteggiata rappresenta l'intervallo di confidenza al 95% stimato con il metodo bootstrap.

Tabella 3.2: Risultati dei GAMs per l'intera area di studio.

Model	p-value	Explained deviance (%)	Degrees of freedom
Nasc school = s(nasc plankton)	0.00099	15.9	4.83
Nasc/Area school = s(nasc plankton)	0.00777	10.2	2.77
Nasc/Area school = s(nasc/area plankton)	0.04010	5.8	1.90

Per comprendere l'effetto della distribuzione spaziale su questi modelli le relazioni sono state analizzate, inizialmente, separando il plancton superficiale dal plancton profondo nell'intera area di studio, senza ottenere però nessun risultato significativo. Quindi, successivamente, l'area di studio è stata suddivisa in tre zone (vedi Fig. 2.1), tenendo in considerazione la differente distribuzione del plancton lungo la colonna d'acqua. Risultati significativi sono stati ottenuti in Zona 1, usando come predittore la biomassa del plancton profondo, e in Zona 3 usando il plancton dell'intera colonna d'acqua. In Fig. 3.2 sono riportate le interazione osservate in Zona 1 tra la biomassa dei banchi di pesce e la biomassa del plancton profondo, la densità dei banchi di pesce e la biomassa del plancton profondo e per ultima la densità dei banchi e la densità del plancton profondo (vedi Tab. 3.3 per le caratteristiche dei modelli).

In Fig. 3.3 sono invece riportate le interazioni tra le stesse variabili osservate in Zona 3 (le caratteristiche dei modelli stimati sono mostrate in Tab. 3.4). I risultati ottenuti dalla suddivisione per zone suggeriscono che l'interazione negativa tra biomassa (o densità) di banchi e biomassa di plancton nell'intera colonna d'acqua è legata principalmente alla relazione osservata in Zona 1. Invece, l'interazione positiva osservata per l'intera area di studio (Fig. 3.2 a,b) per valori più alti di biomassa /densità planctonica, è associata principalmente alla relazione diretta osservata in Zona 3 e dovuta principalmente al plancton dell'intera colonna d'acqua. Un esempio della distribuzione dei banchi di pesce e delle aggregazioni di plancton in Zona 3 è fornita in (Fig. 3.4).

L'approccio esplorativo utilizzato ha permesso di studiare le aggregazioni ittiche pelagiche in specifiche aree del Canale di Sicilia. Ciò ha consentito di formulare delle ipotesi sulla possibile composizione in specie dei banchi localizzati nelle zone in cui era stata riscontrata una relazione significativa con la disponibilità di cibo. Tale approccio ha però evidenziato notevoli limiti legati principalmente alla carenza di informazioni sulla "verità a mare". Di conseguenza, si è scelto di analizzare i dati acustici e biologici raccolti nel corso di campagne sperimentali monospecifiche e di individuare la metodologia più appropriata per attribuire la specie ai banchi.

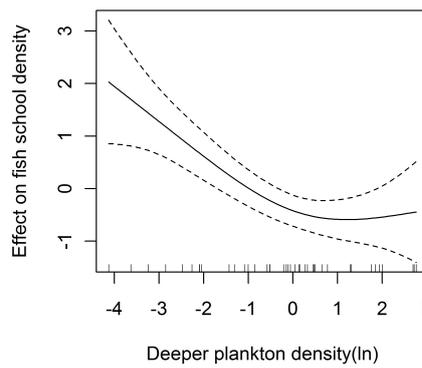
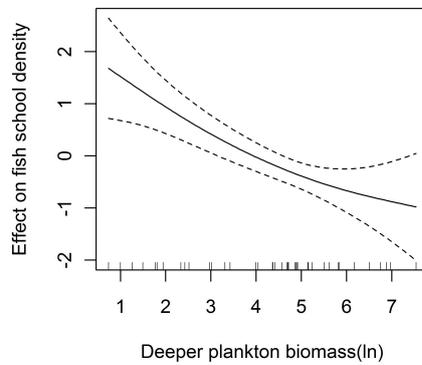
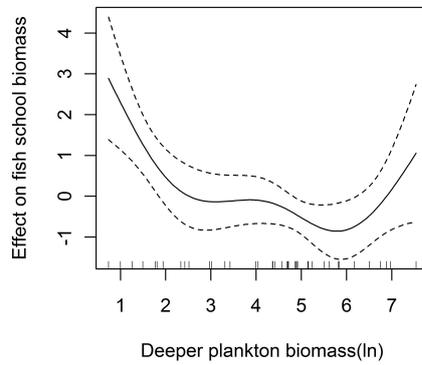


Figura 3.2: Effetto della densità e della biomassa planctonica sulla variabile risposta in Zona 1: (a) biomassa dei banchi di pesce vs. $\ln(\text{biomassa del plancton profondo})$; (b) densità dei banchi di pesce vs. $\ln(\text{biomassa del plancton profondo})$; (c) densità dei banchi di pesce vs. $\ln(\text{densità di plancton profondo})$. La biomassa e la densità di plancton si riferiscono alle aggregazioni con distanza dal fondo inferiore a 5 m. La linea continua indica i valori stimati dal GAM. La linea tratteggiata rappresenta l'intervallo di confidenza al 95% stimato con il metodo bootstrap.

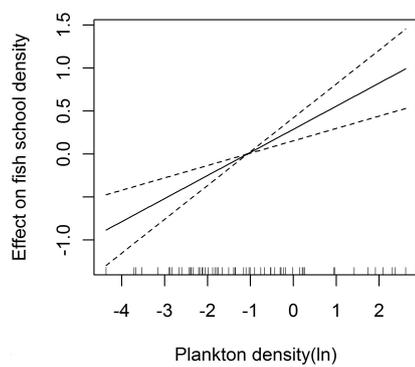
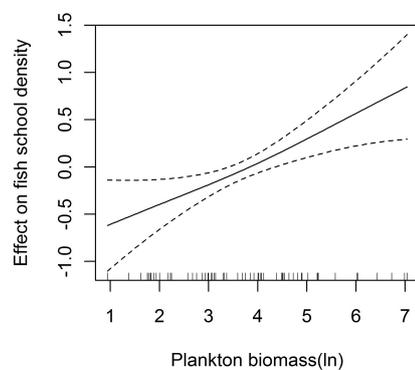
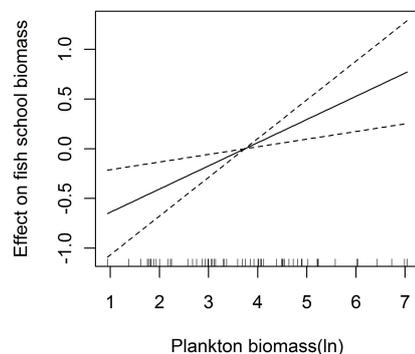


Figura 3.3: Effetto della densità e della biomassa planctonica sulla variabile risposta in Zona 3: (a) biomassa dei banchi di pesce vs. $\ln(\text{biomassa del plancton})$; (b) densità dei banchi di pesce vs. $\ln(\text{biomassa del plancton})$; (c) densità dei banchi di pesce vs. $\ln(\text{densità di plancton})$. La biomassa e la densità di plancton si riferiscono alle aggregazioni nell'intera colonna d'acqua. La linea continua indica i valori stimati dal GAM. La linea tratteggiata rappresenta l'intervallo di confidenza al 95% stimato con il metodo bootstrap.

Tabella 3.3: Risultati dei GAM per la Zona 1.

Model	p-value	Explained deviance (%)	Degrees of freedom
Nasc school = s(nasc plankton)	0.00300	35.9	3.64
Nasc/Area school = s(nasc plankton)	0.00214	28.0	1.57
Nasc/Area school = s(nasc/area plankton)	0.00163	32.0	2.04

Tabella 3.4: Risultati dei GAM per la Zona 3.

Model	p-value	Explained deviance (%)	Degrees of freedom
Nasc school = s(nasc plankton)	0.00898	13.7	1
Nasc/Area school = s(nasc plankton)	0.00492	16.8	1.24
Nasc/Area school = s(nasc/area plankton)	0.00017	24.8	1

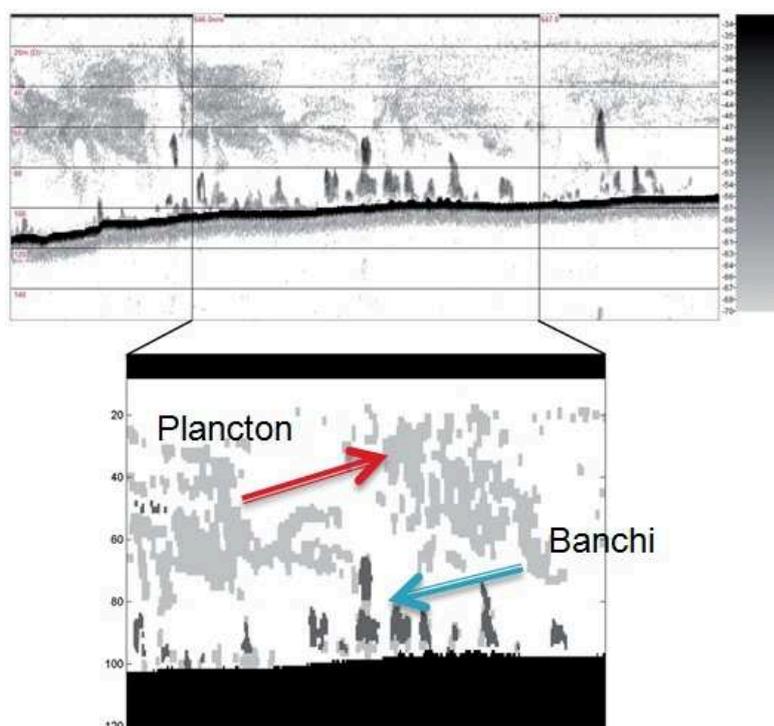


Figura 3.4: Banchi di pesce (in grigio scuro) e aggregazioni di plancton (in grigio chiaro) acusticamente individuate in Zona 3.

3.2 Caratterizzazione dei banchi con dati a multifrequenza

3.2.1 Caratterizzazione batimetrica e morfologica dei banchi di pesce

In tabella 3.5 sono riportati i valori medi e le rispettive deviazioni standard di alcuni dei parametri stimati attraverso l'algoritmo SHAPE del software Echoview, o da essi derivati, calcolati sul learning sample (dati relativi a cale monospecifiche raccolti nel corso di otto campagne acustiche condotte dal 2002 al 2009; v. paragrafo 2.2). L'analisi statistica descrittiva riportata è stata effettuata al fine di ottenere una preliminare caratterizzazione batimetrica e morfologica dei banchi di acciughe e sardine e del gruppo di OPS e demersali. I valori medi e gli intervalli di confidenza dei parametri batimetrici e morfologici dei quattro gruppi di specie sono stati rappresentati mediante Box-plot (Fig. 3.7 a, b, c, d, f).

I banchi di sardine, acciughe OPS e demersali non presentano differenze in termini di valori medi di altezza (`corrected_thickness`). Valori di lunghezza maggiori si osservano invece per acciughe e OPS rispetto a banchi di demersali e sardine, sebbene differenze significative siano state evidenziate solo tra sardine e OPS (Kruskal-Wallis $p=0.003$). Il rapporto lunghezza altezza (`Elong`) indica valori più alti per le sardine e più bassi per le acciughe rispetto alle altre specie, sebbene in questo caso la variabilità del rapporto sia elevata e tra nessuno dei gruppi di specie le differenze risultino significative.

In relazione ai parametri Area (`Corrected_area`) e Perimetro (`Corrected_perimeter`) i banchi di acciughe e di OPS hanno valori maggiori rispetto a sardine e demersali, con differenze significative tra OPS e sardine (Kruskal-Wallis $p=0.008$) e acciughe e sardine (Kruskal-Wallis $p=0.002$) per quanto riguarda l'area, e tra OPS e sardine per quanto riguarda il perimetro (Kruskal-Wallis $p=0.005$). Non si osservano invece differenze significative in termini di rettangolarità.

La dimensione frattale o `Fractaldim`, è un indice della complessità nella forma del banco. Un valore prossimo a 1 di `Fractaldim` indica margini del banco smussati, mentre valori prossimi a 2 denotano margini più complessi. OPS, sardine e demersali possiedono valori prossimi a 2, indicando per essi margini più complessi, mentre banchi di acciughe hanno invece margini più smussati, essendo questo indice più vicino ad 1. Tuttavia le differenze non risultano significative.

`3D_school volume` e `3D_school area` rappresentano una stima rispettivamente dell'area e del volume di un banco assumendo che essi siano cilindrici. Acciughe e OPS mostrano valori medi più alti rispetto a demersali e sardine, sia in termini di area che di volume. In termini di area le differenze risultano significative tra acciughe e sardine (Kruskal-Wallis $p=0.002$) e tra OPS e sardine (Kruskal-Wallis $p=0.04$), mentre per il volume differenze significative sono state riscontrate solo tra acciughe e sardine (Kruskal-Wallis $p=0.004$). Anche la variabile `Beam_volume_sum`, un indice di volume del banco, calcolato però come somma del volume dei ping che lo compongono, ci consente di ottenere differenze significative tra tutte le coppie di specie (Kruskal-Wallis $p<0.000$), eccetto tra banchi di demersali e OPS. Valori medi e variabilità più alta di `Beam_volume_sum` si osservano per demersali e OPS rispetto ad acciughe e sardine. Tra le due specie target, la `Beam_volum_sum` risulta più elevata per le acciughe.

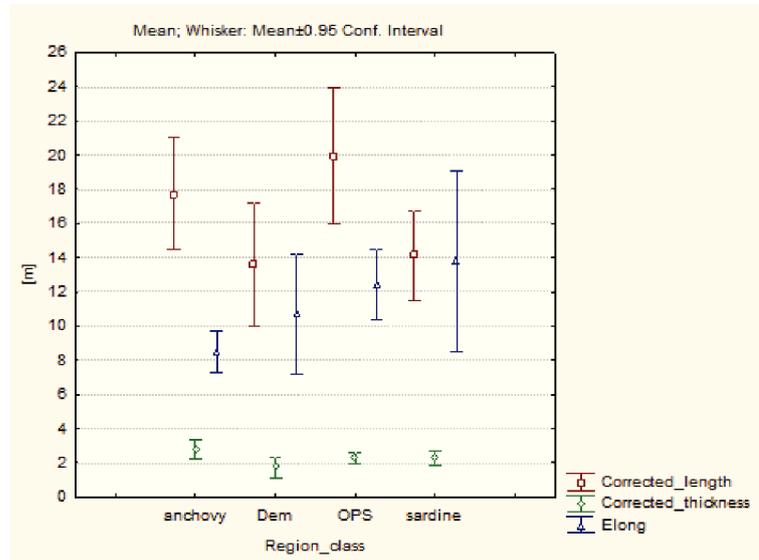
Il parametro Unevenness rappresenta un indice della variabilità nella forma del banco. Rispetto al valore non normalizzato per il numero di ping (Unevenness2) i quattro gruppi di specie non presentano differenze significative, che invece si individuano in termini di unevenness1 ($p=0.003$) tra sardine e OPS. In particolare, i banchi di OPS risultano avere una maggiore variabilità nella forma rispetto a banchi di sardine.

I valori di Depth_mean indicano invece la distribuzione batimetrica dei gruppi di specie. Le due specie target, acciughe e sardine, mostrano preferenze per batimetrie differenti, corrispondenti a profondità medie di 65 m per le acciughe, e a profondità di 45 m per le sardine. Pur distribuendosi su fasce batimetriche differenti, queste due specie presentano distanze dal fondo simili con valori di poco più alti per le acciughe. A maggiori profondità si individuano banchi di demersali, i quali risultano anche molto più vicini al fondo rispetto alle specie target. Il gruppo degli OPS ha invece una distribuzione batimetrica più ampia, con profondità medie di 74 m e distanze dal fondo più elevate. Le differenze relative alla profondità sono significative per tutte le coppie di specie, mentre la significatività per la distanza media dal fondo si osserva solo tra OPS e demersali (Kruskal-Wallis $p<0.002$) e tra OPS e sardine (Kruskal-Wallis $p<0.0001$).

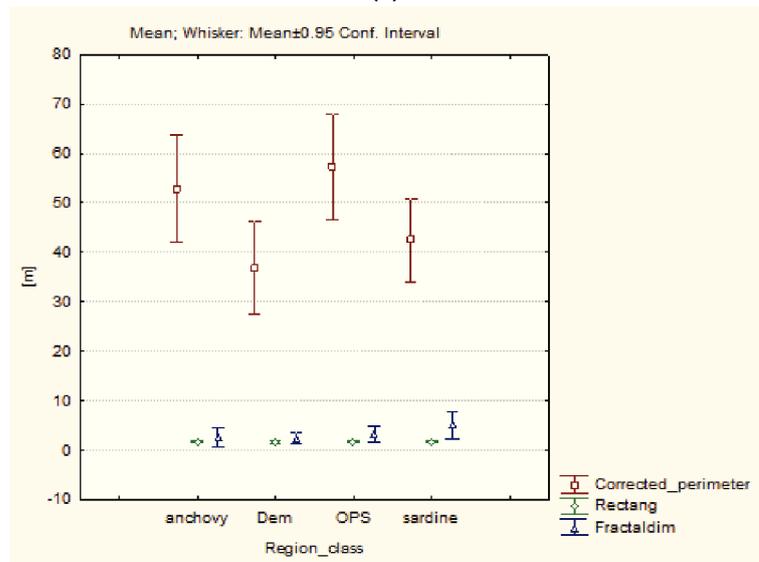
La profondità del banco è l'unico parametro strutturale per il quale si individuano differenze significative tra tutte le coppie di specie e che quindi potrebbe essere d'aiuto in un processo di identificazione. Per tutti gli altri parametri morfologici la significatività si osserva solo tra alcune coppie di specie. Acciughe e sardine differiscono principalmente in termini di area e volume.

Tabella 3.5: Valori medi e dev. Standard (in parentesi) per specie di alcuni descrittori morfologici e batimetrici dei banchi.

	Anchovy	Sardine	Demersal	OPS
Depth_mean [m]	65.77 (11.25)	44.93 (13.44)	138.13 (1.58)	74.37 (28.37)
Bottom [m]	69.50 (9.89)	48.41 (12.09)	140.11 (0.72)	82.14 (28.05)
Corrected_length [m]	17.74 (17.28)	14.12 (16.53)	13.61 (7.86)	19.96 (28.42)
Corrected_thickness [m]	2.79 (2.90)	2.30 (2.73)	1.74 (1.36)	2.29 (2.17)
Corrected_perimeter [m]	52.84 (57.34)	42.48 (52.61)	36.83 (20.77)	6.32 (720.44)
Corrected_area [m ²]	37.41 (83.19)	21.70 (43.56)	16.62 (17.00)	30.54 (52.68)
X3D_school_area [m ²]	350.55 (878.80)	229.58 (714.20)	105.29 (116.87)	269.82 (555.39)
X3D_school_volume [m ³]	132.67 (677.01)	64.66 (305.17)	22.77 (47.10)	63.14 (149.51)
Beam_volume_sum [m ³]	1523.30 (1904.20)	545.15 (701.70)	4283.76 (3994.96)	3515.53 (7102.48)
Elong [m]	8.50 (6.42)	13.80 (33.40)	10.69 (7.70)	12.39 (14.73)
Rectang [m]	1.68 (0.54)	1.77 (0.73)	1.62 (0.49)	1.69 (0.58)
Unevennes2 [m]	1.24 (0.29)	1.24 (0.29)	1.23 (0.40)	1.30 (0.42)
Unevennes1 [m]	50.19 (57.17)	39.84 (52.23)	34.54 (20.12)	52.83 (71.20)
Fractaldim [m]	1.51 (1.79)	1.88 (1.74)	1.85 (1.67)	1.87 (1.33)
Mean_dist_bottom [m]	3.73 (3.06)	3.48 (3.61)	1.98 (1.35)	7.78 (11.88)

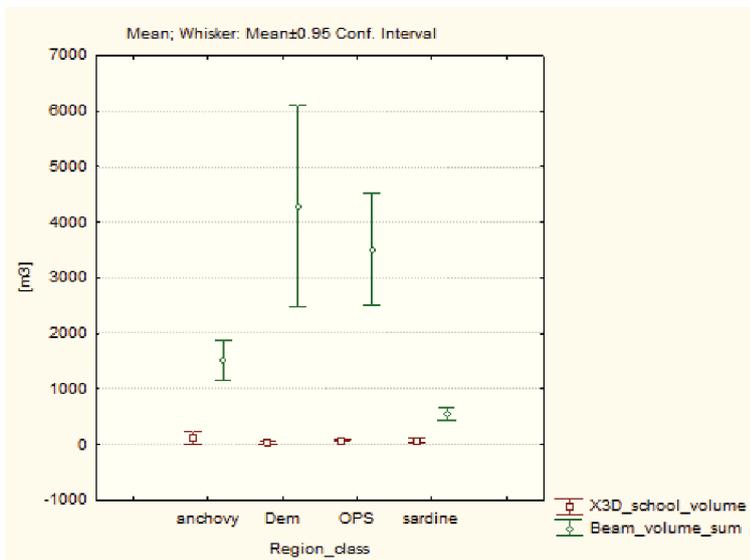


(a)

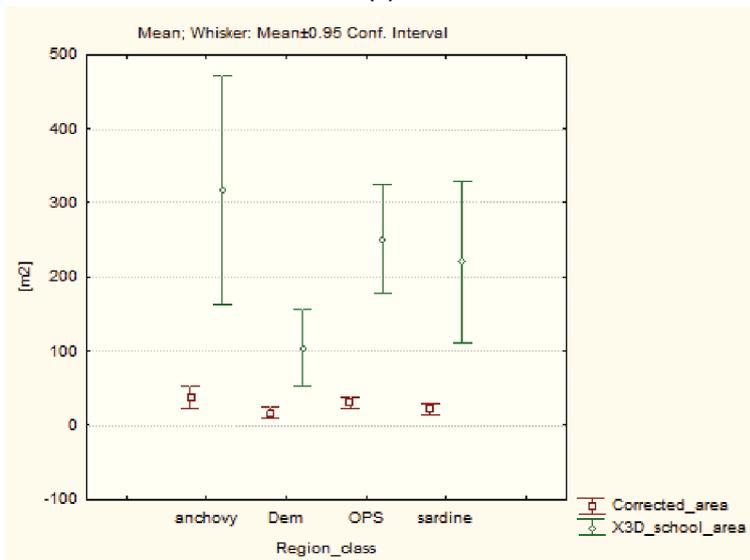


(b)

Figura 3.5: Valori medi e intervalli di confidenza al 95% dei parametri morfologici, batimetrici ed energetici estratti da banchi di acciughe, sardine, OPS e demersali: (a) Lunghezza (Corrected_length), altezza (Corrected_thickness) ed elongazione (Elong) corretti secondo Diner (2001); (b) Perimetro (Corrected_Perimeter), rettangolarità (Rectang) e dimensione frattale (Fractaldim) corretti secondo Diner (2001).

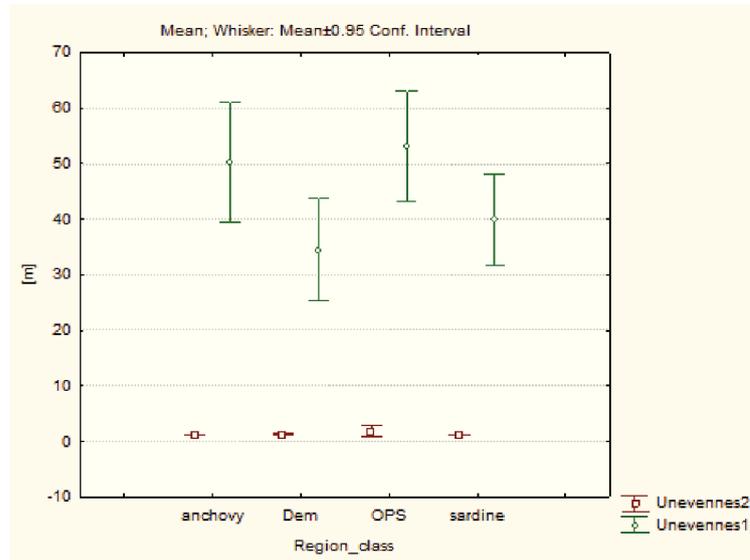


(a)

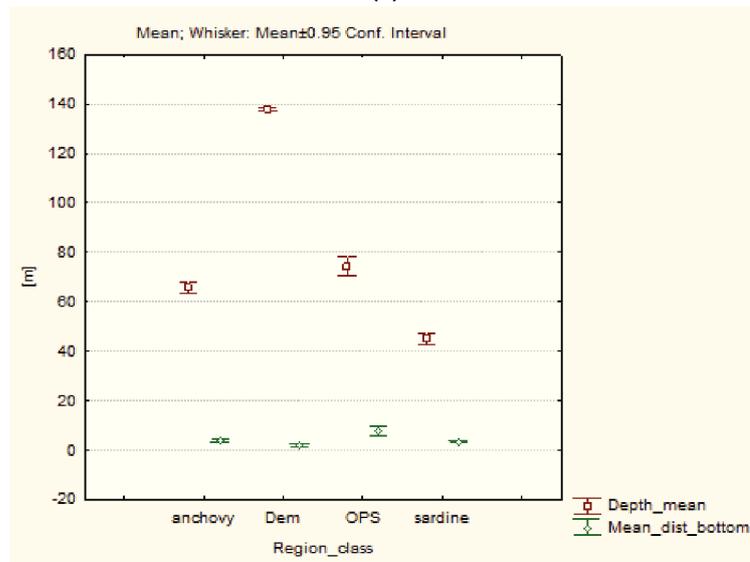


(b)

Figura 3.6: Valori medi e intervalli di confidenza al 95% dei parametri morfologici, batimetrici ed energetici estratti da banchi di acciughe, sardine, OPS e demersali: (a) volume del banco rispetto a un cilindro (X3D_school volume) e volume rispetto al numero di ping (Beam_volume_sum); (b) area (Corrected_area) ed area del banco rispetto a un cilindro (X3D_school area).



(a)



(b)

Figura 3.7: Valori medi e intervalli di confidenza al 95% dei parametri morfologici, batimetrici ed energetici estratti da banchi di acciughe, sardine, OPS e demersali: (a) variabilità della superficie di un banco (Unevenness2) e variabilità della superficie del banco normalizzata rispetto al numero di ping (Unevenness1); (b) profondità media del banco (Depth_mean) e distanza dal fondo (Mean_dist_bottom).

3.2.2 Proprietà energetiche di acciughe e sardine alle diverse frequenze

In questo paragrafo si è cercato di comprendere se un'a distinzione tra le due specie target, acciughe e sardine, fosse possibile confrontando l'intensità di backscatter dei banchi in relazione alla frequenza. L'uso di ecogrammi a più frequenza per l'individuazione di banchi di pesce, ci ha infatti consentito di estrarre valori medi di intensità di backscatter (Sv) per le frequenze 38 e 120 kHz. Per ogni banco è stato poi identificato un indice, chiamato ddB , per esprimere la differenza nei valori di intensità relativamente alla frequenza di 38 kHz:

$$ddB120 = Sv38 - Sv120 \quad (3.1)$$

Il confronto tra i valori medi di Sv (Fig. 3.8a) indica che acciughe e sardine presentano livelli di backscatter e variabilità, nei valori medi, molto simili alla frequenza di 38 kHz, mentre le differenze, sono molto più evidenti a 120 kHz dove risultano peraltro significative (Mann-Whitney test: $p_{120kHz}=0.016$). Una più chiara distinzione tra le due specie in termini di proprietà acustiche è messa in evidenza dai valori di ddB (Fig. 3.8b). La $ddB120$ assume valori pari a +1.55 e +4.18 dB rispettivamente per sardine e acciughe (con valori quindi di circa 2.5 dB in più per le acciughe), e sebbene la dispersione dei valori intorno alla media sia molto simile per le due specie, le differenze nei valori di $ddB120$ risultano significative (Mann-Whitney test: $p_{ddb120}=0.0002546$).

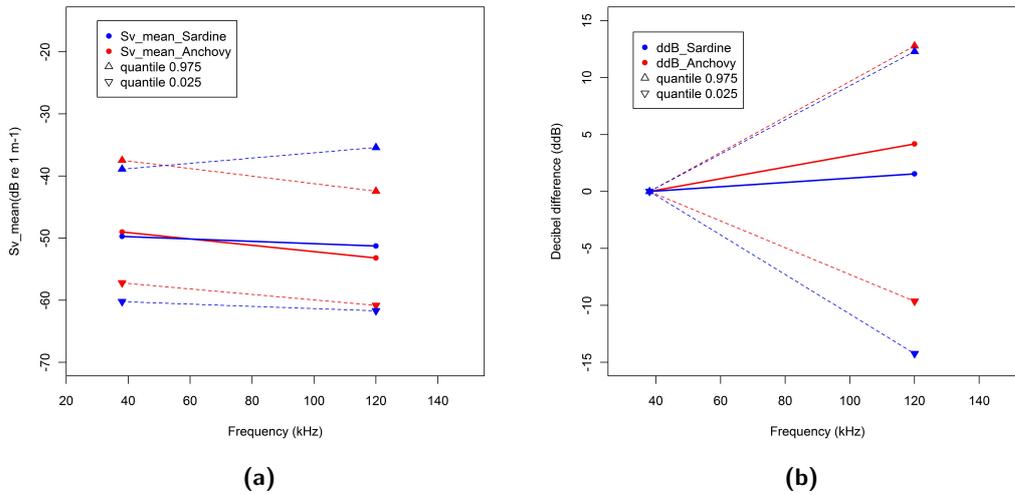


Figura 3.8: (a) Valori medi di Sv per la sardina (cerchio blu) e per l'acciuga (cerchio rosso) a 38 e 120 kHz. (b) la differenza in decibel relativa alla frequenza di 38 kHz per la sardina (cerchio blu) e per l'acciuga (cerchio rosso). I triangoli con la punta verso l'alto rappresentano il quantile 0.975, mentre i triangoli con la punta invertita rappresentano il quantile 0.025; è quindi riportata la dispersione del 95% dei dati attorno al valore medio.

Per comprendere se differenze significative tra le due specie fossero dovute alla presenza di differenti classi di taglia, il dataset è stato suddiviso in banchi di giovanili di acciughe e sardine (indicando come giovanili individui di classi di taglia media compresa tra 7 e 10 cm)

e banchi di adulti delle due specie target (comprendenti individui con classe di taglia media compresa tra 12 e 14 cm). L'analisi dei valori di ddB_{120} indica che differenze significative tra le due specie (Mann-Whitney: $p=0.0007$) si individuano fra banchi di giovanili, con valori medi che risultano positivi e pari a 5.7 dB e 2.6 dB rispettivamente per banchi di acciughe e sardine (Fig. 3.9a). Un valore medio più alto di ddB_{120} (pari a 3.3 dB) si riscontra anche per banchi di individui adulti di acciughe rispetto agli adulti di sardine (il cui valore medio è invece pari a 1.4 dB) (Fig. 3.9b), sebbene le differenze tra questi ultimi siano al limite della significatività (Mann-Whitney: $p=0.05$).

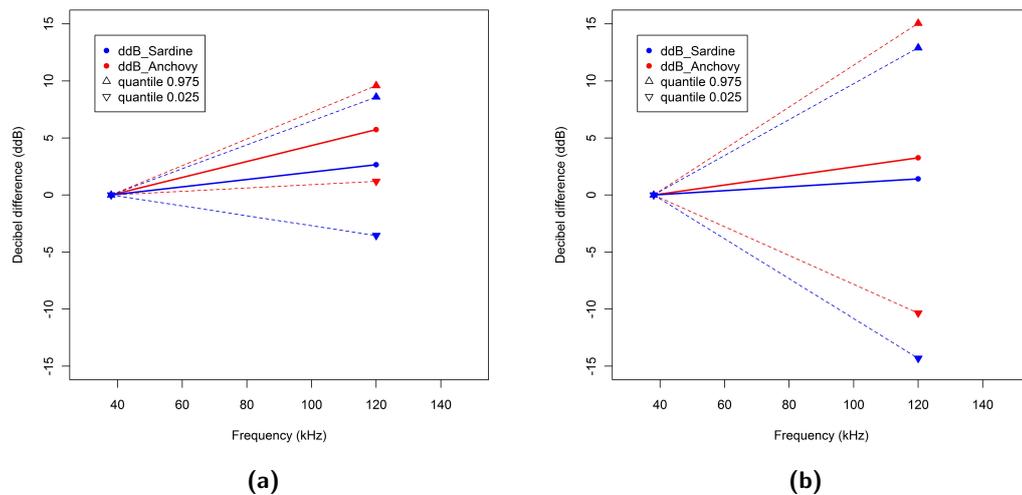


Figura 3.9: (a) differenza in decibel relativa alla frequenza di 38 kHz per giovanili di sardina (cerchio blu) e acciuga (cerchio rosso). (b) differenza in decibel relativa alla frequenza di 38 kHz per adulti di sardina (cerchio blu) e acciuga (cerchio rosso). I triangoli con la punta verso l'alto rappresentano il quantile 0.975, mentre i triangoli con la punta invertita rappresentano il quantile 0.025; è quindi riportata la dispersione del 95% dei dati attorno al valore medio.

3.2.3 Proprietà energetiche di acciughe e sardine per range di profondità

L'analisi dei valori di backscatter per range di profondità indica differenze significative nei valori di ddB tra le due specie sia a profondità comprese tra 40-60 m (Mann-Whitney: $p_{ddB_{120}}=0.004$) (Fig. 3.10a) che a profondità comprese tra 60-80 m (Mann-Whitney: $p_{ddB_{120}}=0.003$) (Fig. 3.10b). A profondità minori (tra 40 e 60 m) i banchi di acciughe presentano valori medi di ddB più bassi rispetto ai banchi di sardine. Mentre gli stessi assumono valori di ddB più alti rispetto ai banchi di sardine a profondità maggiori (tra 60 e 80 m).

Una distinzione dei banchi per range di profondità ci consente quindi di osservare delle differenze significative tra i valori di energia retro diffusa dalle due specie. Tuttavia la variabilità attorno ai valori medi misurati è sempre abbastanza alta e tende a sovrapporsi.

Non è quindi possibile pensare di utilizzare la sola variabile energetica per discriminare tra acciughe e sardine.

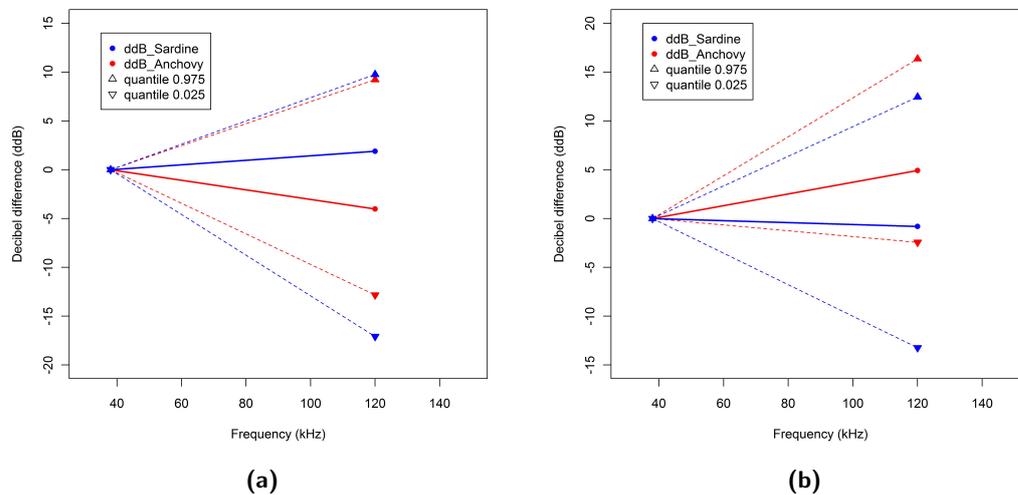


Figura 3.10: Valori medi di ddB_{120} dei banchi con profondità comprese tra 40-60 m (a) e 60-80 m (b), identificati come sardine (in blu) e acciughe (in rosso). I triangoli con la punta verso l'alto rappresentano il quantile 0.975, mentre i triangoli con la punta invertita rappresentano il quantile 0.025.

3.2.4 Modelli di classificazione ad albero con variabili predittive strutturali

In questo paragrafo modelli di classificazione ad albero sono stati usati per classificare banchi di pesci pelagici in gruppi di specie, adoperando come predittori i parametri morfologici (lunghezza, altezza e misure derivate), i parametri batimetrici (profondità del banco e distanza dal fondo) e i parametri energetici (intensità di backscatter e misure derivate alle varie frequenze) estratti, da ogni banco, con l'algoritmo SHAPE di Echoview o calcolati mediante script in R. L'algoritmo estrae numerose variabili spesso tra loro altamente correlate o non corrette come richiesto da Diner (2001). Per ridurre il numero di variabili da utilizzare per la costruzione del modello le variabili morfologiche non corrette sono state eliminate e delle variabili altamente correlate tra loro solo una è stata scelta per l'analisi. Es. la profondità del fondo automaticamente rilevata dallo strumento è stata esclusa dall'analisi perchè altamente correlata alla profondità del banco ($r=0.96$). Adottando questo criterio è stato selezionato un totale di 26 predittori: Sv_mean 38, Sv_min 38, Depth_mean, Corrected length, Corrected thickness, Corrected perimeter, Corrected area, Image compactness, Horizontal_roughness 38, Horizontal_roughness 120, Vertical_roughness 38, Vertical_roughness 120, Sv_mean 120, Sv_min 120, Beam_volume_sum, 3D_school_volume, 3D_school_area, Skewness_120, Skewness_38, Kurtosis_120, Kurtosis_38, ddB_{120} , Rectangularity, Unevenness1, Fractaldim, Mean_dist_bottom. In Tabella (2.2.1) è riportato il significato di ogni variabile.

La procedura di correzione suggerita da Diner (2001) si applica ai parametri morfologici quando si usa un echosounder verticale perchè le dimensioni orizzontali e verticali dei banchi soffrono di alcune distorsioni legate alla lunghezza dell'impulso e alla larghezza del beam. In particolare l'altezza deve essere corretta per la durata dell'impulso e la lunghezza per la larghezza del beam. La correzione di Diner (2001) sulla lunghezza tuttavia è applicabile solo quando il rapporto tra la lunghezza stimata (uncorrected length) e la lunghezza reale è superiore a 1.5. Laddove la correzione non era applicabile sono stati utilizzati i valori non corretti. Per alcune delle variabili selezionate come predittori si è riscontrata la presenza di valori mancanti (sostituiti con la dicitura "NA"), ma poiché questi ultimi non impediscono l'utilizzo delle relative variabili per l'applicazione dei modelli decisionali ad albero, esse sono state ugualmente considerate nell'analisi. La classification tree ha infatti il vantaggio di trattare il dato mancante come categoria speciale. Quando, infatti, non è possibile determinare una suddivisione su una certa variabile esplicativa perchè il valore di una determinata unità è mancante, si effettua una suddivisione su un'altra variabile detta surrogata, scelta in modo tale che le due suddivisioni abbiano un indice di similarità massimo in quel determinato nodo. In questo modo si evita la perdita d'informazione che deriverebbe dall'eliminazione delle variabili contenenti dati mancanti o dall'eliminazione dei banchi che possiedono valori mancanti per determinate variabili.

I modelli decisionali sono stati applicati a un dataset composto da 155 banchi di sardine e 111 banchi di acciughe. A questo sono stati poi aggiunti banchi di altre specie, presenti nell'area di studio, ed etichettati come OPS (Other Pelagic Species, v. sottoparagrafo 2.2.6) e demersali. Il numero totale dei banchi etichettati come OPS è di 197, mentre solo 21 è il numero dei banchi attribuiti a specie demersali. L'inserimento di nuove specie o gruppi di specie risulta fondamentale ai fini di un futuro utilizzo dei modelli di classificazione come strumenti classificativi o operativi. Il dataset iniziale (con tot di n. 484 banchi) (v.paragrafo è stato suddiviso in un training set (di n. 324 banchi) e in un test set (di n. 160 banchi), prelevati in maniera random selezionando rispettivamente 2/3 e 1/3 dal dataset iniziale.

Il primo albero (in Fig. 3.11), sviluppato inserendo tutte le variabili esplicative (in totale 26), classifica le osservazioni del nostro training set in 15 nodi terminali utilizzando in totale 7 variabili: Profondità media del banco (Depth_mean), Distanza media dal fondo (Mean_dist_bottom), variabilità energetica lungo l'asse verticale a 120 kHz (vertical_roughness 120), volume del banco (Beam_volume_sum), energia retrodiffusa a 120 kHz (Sv_mean 120), valori minimi di energia retrodiffusa a 120 (Sv_min 120) e valori minimi di energia retrodiffusa a 38 kHz (Sv_min 38). Di queste 7 variabili, 2 sono batimetriche, 1 morfologica e 4 energetiche. Le osservazioni per cui la risposta alla condizione sulla variabile selezionata è verificata sono raggruppate a sinistra, altrimenti le osservazioni ricadono nel ramo di destra. Ogni nodo terminale rappresenta le variabili risposta nel seguente ordine: anchovy, demersali, OPS e sardine. Il gruppo maggiormente rappresentato nel nodo terminale è quello da cui deriva il nome. Secondo il modello la profondità rappresenta la variabile esplicativa più importante nella discriminazione tra gruppi di specie. Questa stessa variabile appare più volte durante lo sviluppo dell'albero, indicando l'esistenza di una relazione non lineare con la variabile risposta.

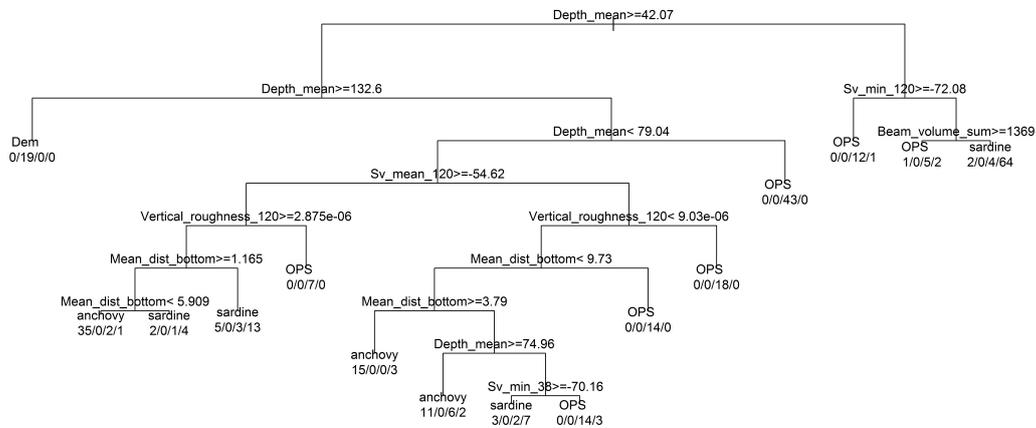


Figura 3.11: Modello di classificazione ad albero costruito su 26 predittori.

Il modello classifica le acciughe in tre gruppi terminali, mentre le sardine sono classificate principalmente in due gruppi terminali, sebbene solo uno sia più numeroso. I banchi di sardine (64/100) si trovano principalmente a profondità inferiori a 42 m, hanno valori minimi di Sv a 120 kHz più bassi di -72 dB e beam_volume.sum inferiori a 1369 m³. Il volume del banco (indicato come beam_volume.sum) classifica correttamente un numero molto basso di osservazioni. La lunghezza dei rami dell'albero è infatti proporzionale al numero di osservazioni correttamente classificate. Un gruppo più piccolo banchi di sardine (13/100) si trova su fondali compresi tra 42 m e 79 m, ha valori medi di Sv a 120 kHz superiori o uguali a -54.6 dB, è caratterizzato da una variabilità energetica lungo l'asse verticale nei valori di Sv superiori a 2.87×10^{-6} e distanze dal fondo minori di 1.16 m. Le acciughe rientrano in tre gruppi terminali. Gran parte dei banchi (35/74 banchi) si trova a profondità comprese tra 42 e 79 m, ha valori medi di Sv a 120 kHz superiori o uguali a -54.6 dB, è caratterizzato da una variabilità energetica lungo l'asse verticale nei valori di Sv a 120 kHz (vertical_roughness_120) superiori a 2.87×10^{-6} e distanze dal fondo comprese tra 1.16 m e 5.9 m. Un secondo gruppo di 15 banchi di acciughe si trova nella stessa fascia batimetrica del precedente ma presenta valori medi di Sv a 120 kHz inferiori e -54.6 dB, una variabilità energetica lungo l'asse verticale a 120 kHz più bassa di 9.03×10^{-6} e distanze dal fondo comprese tra 3.8 m e 9.7 m. Il terzo gruppo di acciughe (11/74 acciughe) ha profondità comprese tra 75 e 79 m, le stesse caratteristiche del precedente il termini di vertical_roughness_120 ma distanze medie dal fondo inferiori a 3.79 m. I demersali (19/21) sono invece tutti distribuiti in un gruppo a profondità superiori a 133 m, mentre gli OPS sono suddivisi in più sottogruppi con un gruppo predominante (43/131 banchi) caratterizzato da profondità superiori a 79 m.

Una volta analizzate le caratteristiche del modello è stato necessario capire se la taglia

dell'albero fosse ottimale e potesse essere quindi adattato anche a nuove osservazioni. La tecnica migliore per la scelta dell'albero ottimale è applicare la tecnica di pruning, che impiega la cross-validation. Quest'ultima consente di trovare il parametro di complessità (cp) per individuare la dimensione ottimale dell'albero in termini di numero di nodi terminali. Il risultato della cross-validation è riportato nel grafico di seguito (Fig. 3.12).

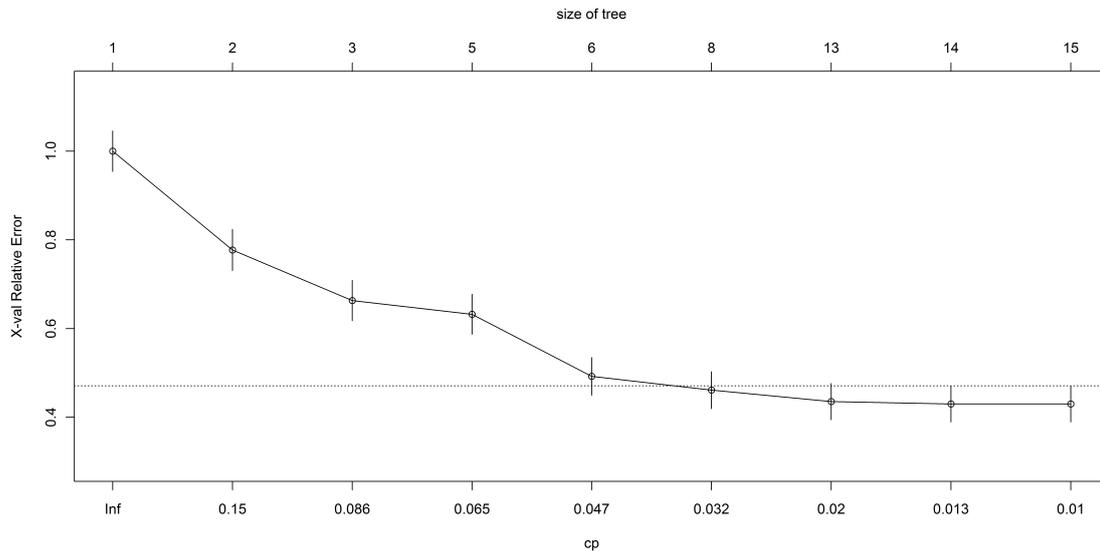


Figura 3.12: Parametro di complessità (cp, in ascissa) correlato alla taglia dell'albero (size of tree) e l'errore medio di predizione (in ordinata). L'errore medio è il risultato di una cross-validation effettuata per differenti valori di cp. La barra verticale indica la variazione (dev. St) nell'errore predetto. La linea orizzontale tratteggiata si ottiene sommando all'errore predetto dall'albero di dimensioni maggiori, una volta la sua deviazione standard (regola "1-SE").

La regola "1-SE" stabilisce che il valore ottimale di cp (a cui corrisponde una taglia ottimale) è quel valore per il quale l'errore medio di predizione della cross validation ricade sotto la linea orizzontale. Il valori di cp ottimale è pari quindi a 0.032 a cui corrisponde una taglia dell'albero pari a 8. In Fig. 3.13 è riportato l'albero ottimale selezionato dalla tecnica di pruning. La tecnica di pruning raggruppa le specie target in due soli più grandi gruppi terminali. Per valutare l'accuratezza del modello, l'albero è stato testato sul test set e i risultati sono stati confrontati con la classificazione a priori mediante una confusion matrix, riportante proprio il numero e la percentuale di banchi per specie classificati correttamente e non correttamente dal modello con tutte le variabili predittive. I risultati della confusion matrix indicano che la probabilità di classificare correttamente i gruppi di specie, utilizzando le variabili predittive selezionate dal modello, è dell'81% con capacità di discriminazione per specie pari a 82% per OPS , 100% per demersali, 89% per sardine, 65% per acciughe (Tab. 3.6). La capacità di classificazione totale è stata calcolata come rapporto tra la somma delle osservazione correttamente classificate e il totale delle osservazioni.

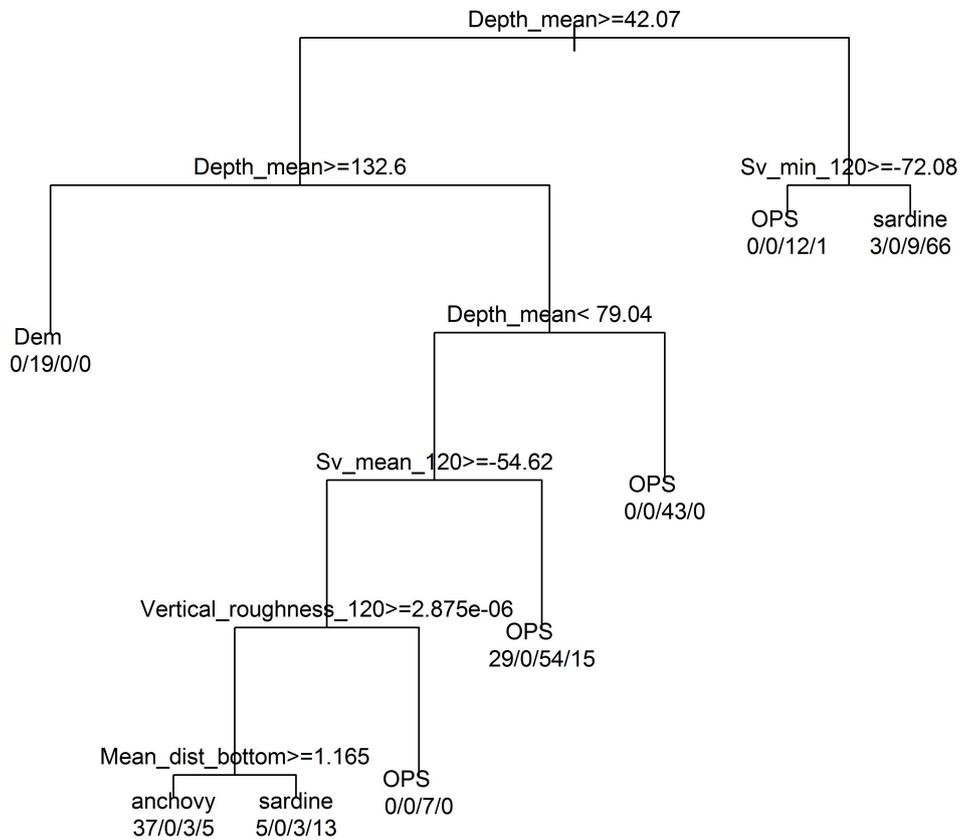


Figura 3.13: Modello di classificazione ottimale individuato dalla tecnica di *pruning*.

Tabella 3.6: Confusion matrix del modello predittivo finale della classification tree, stimato utilizzando tutti i predittori selezionati (v. testo).

	Anchovy	Dem	OPS	Sardine	Specie Rate
Anchovy	24	0	4	9	65%
Demersal	0	2	0	0	100%
OPS	4	0	55	7	82%
Sardine	5	0	1	49	89%

Per testare quale fosse l'effetto della profondità sulla capacità di discriminazione per specie è stato costruito un nuovo modello, escludendo dalle variabili predittive la profondità del banco, che risulta fortemente correlata alla profondità del fondo ($r^2=0.96$). Il modello ottimale è riportato in Fig. 3.14. In assenza della profondità il predittore più importante nella suddivisione in sottogruppi è la variabile Sv_mean 120, vale dire il valore medio di backscatter misurato alla frequenza di 120 kHz. Altre variabili energetiche entrano in gioco nella classificazione in specie sono: *ddB120*, *horizontal_roughness 120*, *vertical_roughness 120*, *Sv_min120*. Tra le variabili morfologiche rimane invece la *beam_volume_sum*.

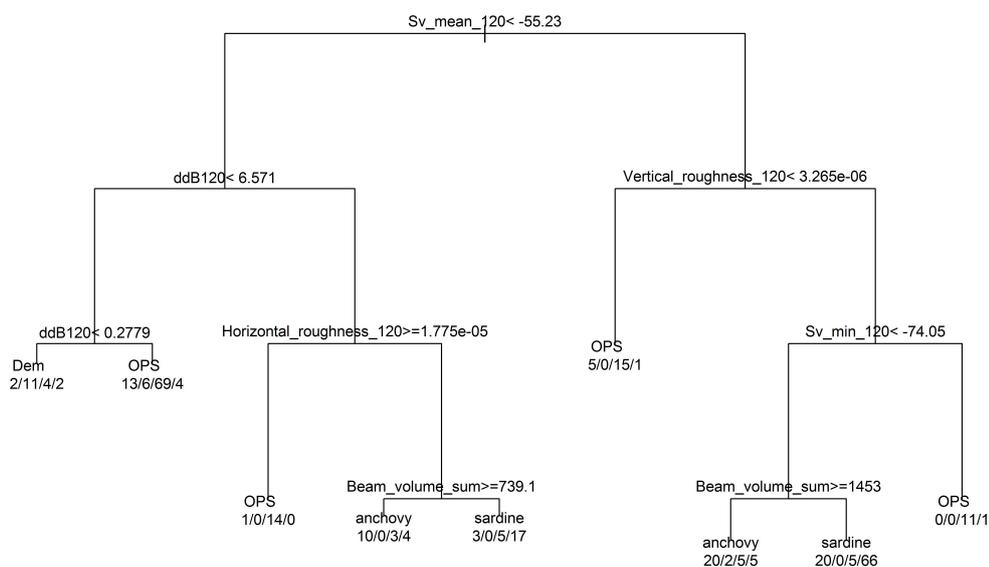


Figura 3.14: Modello ottimale costruito escludendo dai 26 predittori la profondità del banco.

Questo modello ha una capacità totale di classificazione del 60% con valori per specie riportati in Tab. 3.7. Rispetto al modello precedente il tasso di classificazione si riduce di circa il 20%, abbassando in maniera considerevole la capacità di classificazione delle sardine e soprattutto delle acciughe. Non viene invece modificata di molto la capacità di classificazione degli OPS, che passa da un 82% nel modello precedente al 79%. Ciò induce a pensare che la profondità del banco non fornisca un grande contributo nella discriminazione dei banchi di OPS da altre specie, mentre risulta di notevole importanza nella classificazione di acciughe e sardine.

Il passo successivo è stato quello di verificare il contributo alla discriminazione delle specie fornito dalle sole variabili morfologiche e dalle sole variabili energetiche.

Tabella 3.7: Confusion matrix del modello predittivo della classification tree ottenuto escludendo la profondità dai predittori selezionati (v. testo).

	Anchovy	Dem	OPS	Sardine	Specie Rate
Anchovy	8	0	11	18	22%
Demersal	0	2	0	0	100%
OPS	4	1	52	9	79%
Sardine	10	0	11	34	62%

Due nuovi modelli sono stati quindi costruiti e testati per analizzarne l'accuratezza. In Fig. 3.15 è riportato il modello costruito sulle sole variabili energetiche e in Tab. 3.8 la capacità di classificazione per specie. Le variabili energetiche classificano correttamente pochi banchi di acciughe (16/74), mentre si conferma la presenza di un gruppo numeroso e prevalente per gli OPS e per le sardine. Ciò indica che le sole variabili energetiche hanno una scarsa capacità di discriminare le acciughe dalle altre specie.

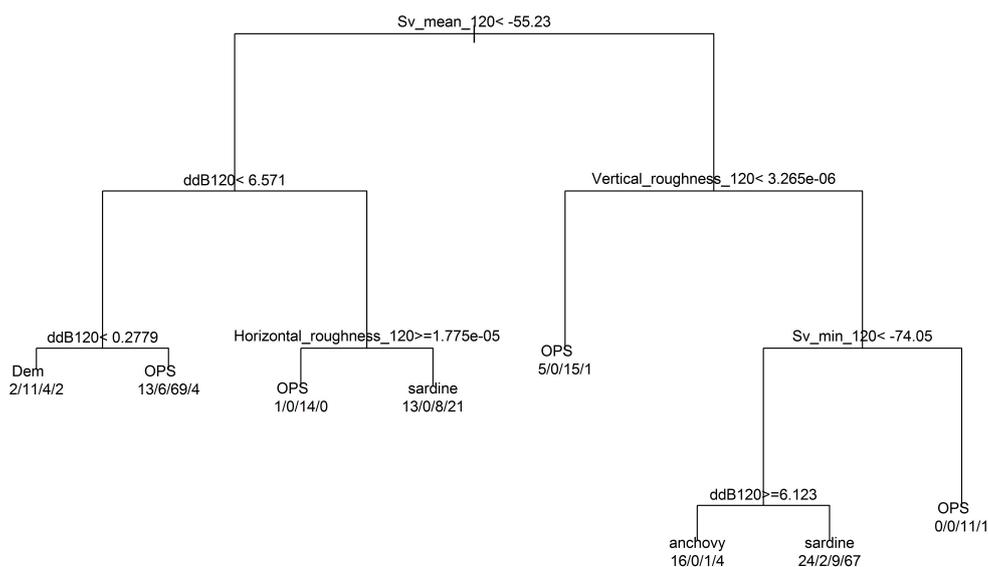


Figura 3.15: Modello ottimale costruito sulle sole variabili energetiche.

In effetti il tasso di classificazione totale è del 59% con un tasso del 79% per OPS, del 100% per Demersali, del 69% per sardine, e di solo 5% per acciughe (Tab. 3.8).

Tabella 3.8: Confusion matrix del modello costruito sulle sole variabili energetiche.

	Anchovy	Dem	OPS	Sardine	Specie Rate
Anchovy	2	0	11	24	5%
Demersal	0	2	0	0	100%
OPS	2	1	52	11	79%
Sardine	6	0	11	38	69%

In Fig. 3.16 è riportato il modello stimato sulle sole variabili morfologiche.

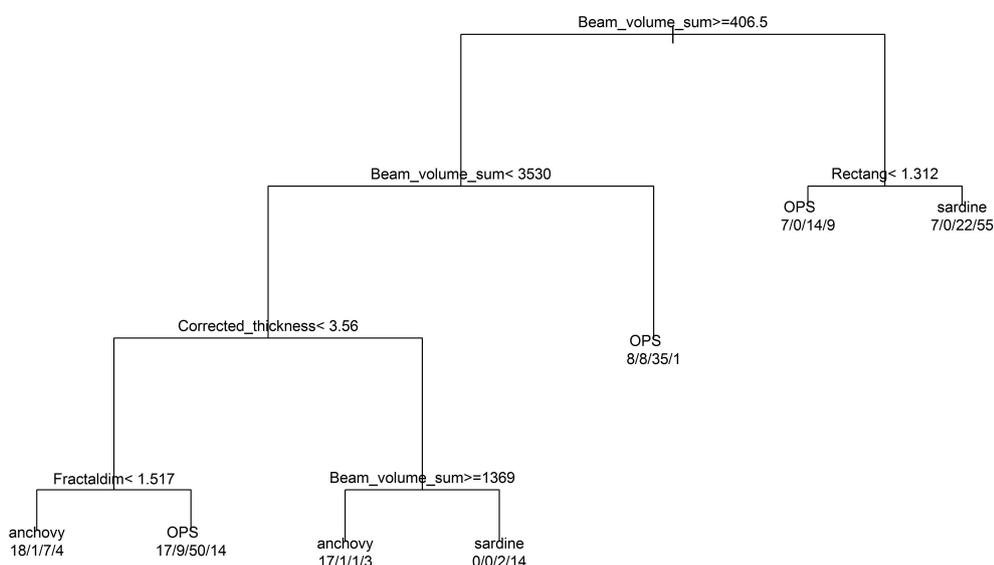


Figura 3.16: Modello ottimale costruito sulle sole variabili morfologiche.

Tra le variabili morfologiche selezionate dal modello vi sono: il volume, che rimane la variabile morfologica più importante nella discriminazione tra specie, la rettangolarità (Rectang) che discrimina tra un gruppo OPS e Sardine, l'altezza del banco (Corrected thickness) e la dimensione frattale (Fractaldim).

L'albero costruito sulle sole variabili morfologiche ha una capacità totale di classificazione del 58% e non consente di classificare il gruppo di demersali. La percentuale di specie classificate correttamente è pari al 74% per OPS, al 62% per sardine e al 27% per acciughe (Tab. 3.9). Un confronto tra questi ultimi due modelli indica che le sole variabili morfologiche discriminano meglio le acciughe dalle altre specie rispetto alle sole variabili energetiche. I modelli hanno invece capacità di classificazione comparabili nei confronti delle sardine.

Tabella 3.9: Confusion matrix del modello costruito sulle sole variabili morfologiche.

	Anchovy	OPS	Sardine	Specie Rate
Anchovy	10	17	10	27%
Demersal	0	2	0	0%
OPS	3	49	14	74%
Sardine	8	13	34	62%

Ad ogni modo l'albero costruito con l'insieme delle variabili morfologiche, batimetriche ed energetiche, risulta comunque il modello con la maggiore capacità di discriminazione totale e per specie. Si è cercato di capire, a questo punto, se la struttura e la dimensione, nonché la capacità di discriminazione del modello con tutte le variabili strutturali, subisse delle modifiche in relazione al training set scelto. Si è deciso quindi di costruire trenta alberi differenti, usando 30 differenti training set e test set e di analizzare le capacità di discriminazione per specie dei trenta diversi alberi. Il risultato di questa ulteriore analisi è riprodotto in 4 istogrammi (Fig. 3.17), ognuno dei quali mostra la distribuzione delle capacità di classificazione di trenta differenti alberi per ogni gruppo di specie.

In media la capacità totale di classificazione è pari al 76%, ma risulta piuttosto variabile selezionando training set differenti. La selezione di un training set diverso modifica infatti le regole di classificazione e con esse la capacità classificativa di ogni specie o gruppo di specie. Ciò riflette probabilmente la grande variabilità dei parametri descrittivi del banco, usati nei modelli come variabili esplicative. Tuttavia gran parte degli alberi ha buone capacità di classificazione per le sardine e gli OPS, per lo più comprese tra 70 e 80% per le sardine e tra 75 e 85% per OPS. La capacità dei modelli di classificare correttamente le acciughe risulta, invece, piuttosto variabile. Un numero più alto di modelli ha capacità di discriminazione comprese tra 50 e 55% per le acciughe. Per i demersali si ha sempre una buona capacità di discriminazione, ma di questo risultato si è tenuto poco conto essendo il learning sample per questo gruppo estremamente ridotto.

Ci si è chiesto a questo punto quali di questi 30 modelli fosse più opportuno utilizzare per identificare banchi di specie diverse da un qualunque survey. Sono stati quindi selezionati tre differenti modelli predittivi, da applicare a tutto il survey 2009 e ed è stato scelto come criterio di valutazione della bontà del modello il grado di accordo dello stesso con le informazioni di biomassa per specie ricavate con la tradizionale metodologia per le stime di biomassa. Il paragrafo successivo riporta una breve descrizione dei tre modelli decisionali selezionati per associare la specie ai banchi estratti dal survey ANCHEVA2009, una descrizione della metodologia tradizionale per le stime di biomassa e infine un confronto tra i risultati dei tre modelli predittivi e i valori di biomassa stimati con la metodologia tradizionale. Il confronto è stato effettuato prima sui dati stimati per l'intero survey e successivamente in 6 gruppi di dieci transetti.

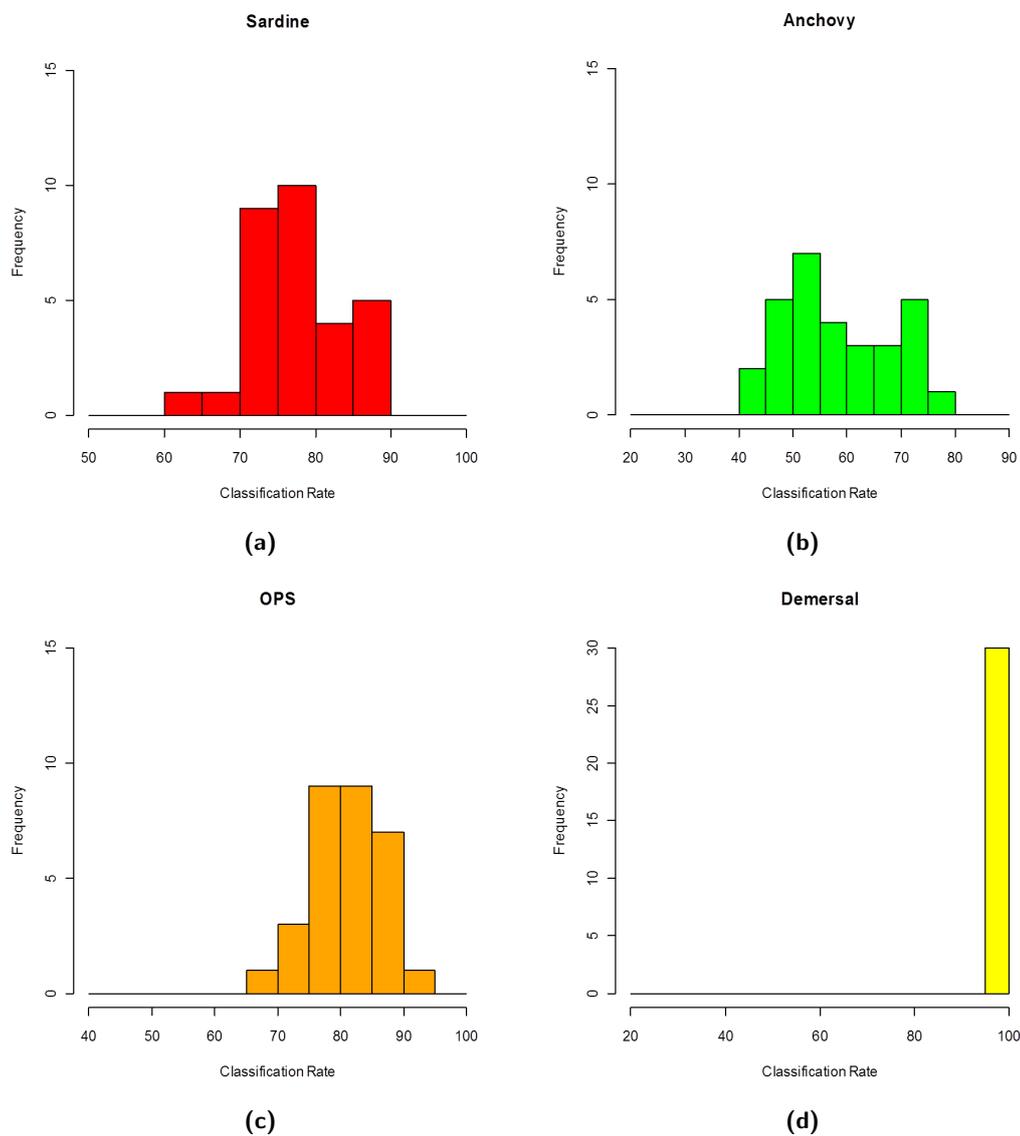


Figura 3.17: Distribuzione delle capacità di classificazione di trenta differenti alberi per i 4 gruppi di specie: (a) sardine, (b) acciughe, (c) OPS e (d) demersali.

3.2.5 Confronto dei risultati provenienti da modelli di classificazione ad albero con le stime di biomassa

Per verificare le capacità predittive dei modelli decisionali l'algoritmo a multifrequenza è stato applicato ai transetti diurni del survey ANCHEVA2009. Ogni transetto è stato suddiviso in miglia nautiche, in modo tale da ricavare, oltre alle già citate informazioni batimetriche, morfologiche ed energetiche, un parametro denominato PRC_NASC (estratto

dalla sola frequenza a 38 kHz). Il PRC_NASC misura il contributo fornito da ogni banco al NASC all'interno di ogni miglio nautico. Nel calcolo della biomassa totale fornita dai banchi per miglio nautico sono state escluse tutte quelle eco tracce a profondità inferiori a 30 m e superiori a 140 m, in quanto non avendo informazioni biologiche provenienti dalle cale sperimentali, non era noto quali specie contribuiscano alla biomassa a tali batimetrie. In questo modo è stato individuato un totale di 1065 banchi su un percorso totale di 232 miglia nautiche.

La classificazione per specie è stata effettuata mediante l'uso di tre differenti modelli:

- L'albero con la migliore capacità di discriminazione per tutte le specie fra i trenta differenti alberi Fig 3.18. L'albero (denominato MaxRate) ha una capacità totale di classificazione dell'82% con tassi dell' 87% per le sardine, del 72% per le acciughe, del 100% per i demersali, dell'83% per gli OPS. Questo modello utilizza solo due variabili batimetriche (depth_mean e mean_dist_bottom), tre variabili morfologiche (Beam_volum_sum, corrected length e Fractaldim) e due variabili energetiche (Sv_mean_38, Sv_min_120). Il modello distribuisce gli OPS in vari gruppi principalmente sulla base della profondità e individua un gruppo predominante di sardine e due gruppi principali di acciughe. L'analisi del training set mostra che i due principali gruppi di acciughe corrispondono ad una suddivisione in adulti (gruppo di acciughe con 34/75 banchi) e in giovanili (gruppo di acciughe) con 18/75 banchi. Una tale suddivisione consente di raggiungere maggiori capacità di discriminazione per questa specie. Ciò induce a ipotizzare che individui di taglia diversa abbiano caratteristiche strutturali differenti, sebbene variabili.

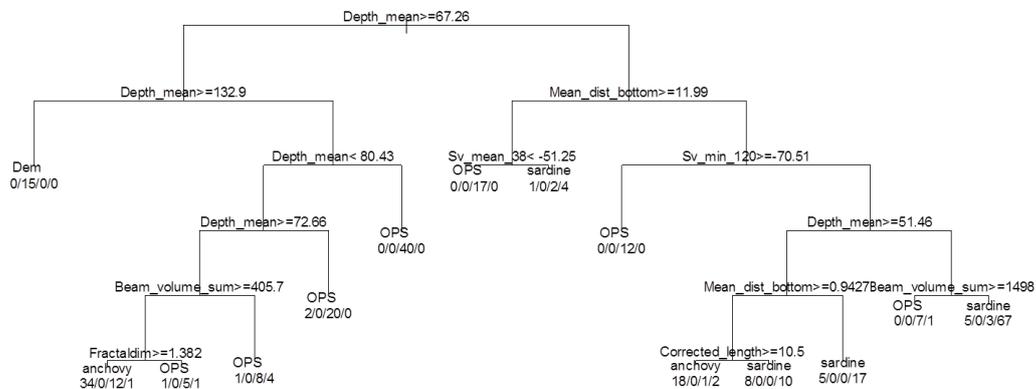


Figura 3.18: Albero di classificazione con la migliore capacità di discriminazione per tutte le specie (MaxRate).

- L'albero con un tasso di classificazione superiore al 75% per acciughe e sardine (denominato MaxRateSpecies). Essendo queste le specie target e anche le specie più

abbondanti nell'area di studio si decide di scegliere il modello che fornisca più alti tassi di classificazione per esse Fig. 3.19. Esso ha una capacità totale di classificazione del 79% con valori dell'81% per le sardine, del 77% per acciughe, del 100% per demersali e del 78% per OPS. Si mantiene un gruppo predominante di sardine e la distinzione dei banchi di acciughe in giovanili (sul ramo destro dell'albero) e adulti (sul ramo sinistro). In entrambi i casi giovanili di acciughe risultano più costieri rispetto agli individui adulti e vengono discriminati da banchi di sardine sulla base di valori medi di lunghezza e distanza dal fondo.

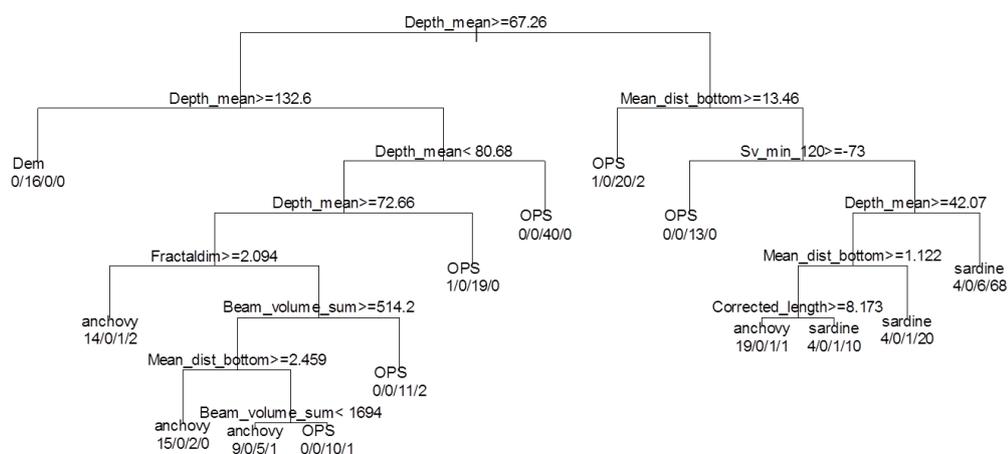


Figura 3.19: Albero di classificazione del modello con una capacità di discriminazione superiore al 75% per acciughe e sardine (MaxRateSpecies).

- Il terzo modello adottato è stato denominato “Mode” e classifica ogni banco utilizzando la classe più frequente risultante da trenta differenti alberi. In altre parole tutti i modelli sono utilizzati per effettuare la classificazione predizioni, ma la classificazione finale è rappresentata dalla classe più frequente. Questo modello in presenza di più gruppi, con lo stesso numero di banchi per specie associa l’informazione “non determinato”.

Un esempio è rappresentato nella tabella di seguito (Tab. 3.10).

Tabella 3.10: Esempio di classificazione fornito da trenta differenti modelli, con la previsione finale rappresentata dalla classe più frequente.

Anchovy	Sardine	Dem	OPS	Previsione
1	10	5	14	OPS
5	10	10	5	notDet

La distinzione tra banchi di acciughe giovanili e banchi di adulti in due distinti gruppi terminali ha indotto a pensare che individui di taglia diversa potessero avere caratteristiche strutturali differenti. È stato quindi necessario verificare che il learning sample creato rispecchiasse la composizione per specie e taglia della campagna scelta per l'attività di classificazione dei banchi. Il survey Ancheva2009, selezionato per questo scopo, è stato condotto nel periodo estivo, dove banchi di giovanili di acciughe non erano presenti. Ciò ha richiesto una riduzione del learning sample (a un totale di 443 banchi), per eliminazione delle cale con individui giovanili per questa specie. Sono state invece mantenute le cale con giovanili di sardine, che per effetto del reclutamento erano presenti nel survey estivo del 2009 insieme ad individui adulti. Con il learning sample ridotto sono stati simulati trenta diversi modelli selezionando per scopi classificativi un nuovo modello MaxRate (coincidente in questo caso anche con il modello MaxRateSpecies) e un nuovo modello Mode. I trenta nuovi alberi hanno mostrato una capacità totale di classificazione del 78% con valori dell'84% per le sardine, del 47% per le acciughe, del 100% per demersali e dell'81% per OPS. In generale l'assenza di banchi di giovanili non ha determinato un miglioramento nelle capacità predittive rispetto ai trenta alberi stimati sul dataset completo. Di seguito sono riportate le caratteristiche dei due modelli selezionati per la classificazione dei banchi estratti dal survey Ancheva2009:

- Il Modello MaxRate o MaxRateSpecies: il modello con la massima capacità di discriminazione totale e per le due specie target. Esso ha una capacità di classificazione totale dell'83% con tassi del 90% per le sardine, del 67% per le acciughe, del 100% per demersali e dell'80% per gli OPS. In Fig. 3.20 sono riportate le caratteristiche del modello. Esso classifica i banchi di sardine in un gruppo terminale principale (93/104 banchi), sulla base della profondità (inferiore a 67.22 m), della distanza minima dal fondo e dell'Sv minimo a 120. Le acciughe sono invece classificate principalmente in un unico nodo terminale (31/52 banchi). Sono localizzate a profondità comprese tra 79 e 73 m e differiscono da banchi di OPS, presenti alla stessa profondità, per il volume del banco (indicato come beam_volume_sum). Questo modello classifica i banchi del survey Ancheva2009 come riportato in Tab. 3.11.
- Il modello "Mode", che utilizza tutti e trenta gli alberi per fare delle predizioni classifica come riportato in Tab. 3.14 i banchi del survey Ancheva2009.

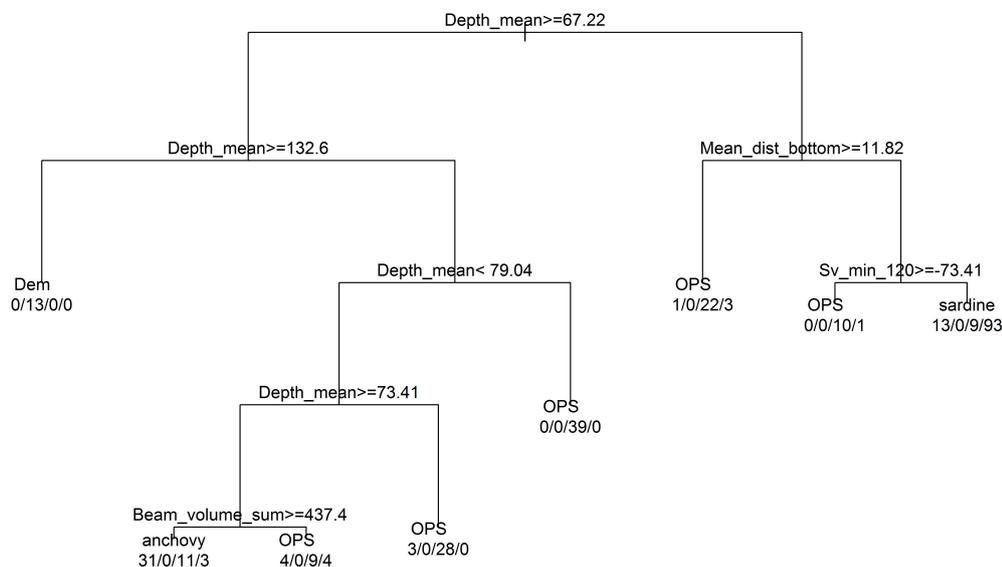


Figura 3.20: Albero di classificazione con la maggiore capacità di discriminazione per tutte le specie e per le specie target (MaxRate o MaxRateSpecies).

Tabella 3.11: Numero e percentuale di banchi attribuiti a ogni specie dal modello MaxRate o MaxRateSpecies.

Anchovy	Dem	OPS	Sardine
20	21	703	321
2%	2%	66%	30%

Tabella 3.12: Numero e percentuale di banchi attribuiti a ogni specie dal modello Mode.

Anchovy	Dem	OPS	Sardine	notDet
21	21	670	352	1
2%	2%	62%	33%	0%

Un opportuno script elaborato in ambiente R ha permesso di stimare i valori di PRC_NASC (indice di biomassa) per miglio associati ai banchi di ogni specie. Con un semplice scatterplot è stato poi verificato che esistesse una correlazione tra il PRC_NASC per miglio estratto dai banchi e il NASC per miglio calcolato con la tecnica tradizionale. La tecnica tradizionale infatti non lavora esclusivamente sui banchi, ma integra l'intera colonna

d'acqua ad intervalli di 1 nmi con un filtro soglia, dopo aver eliminato manualmente le aggregazioni che, secondo l'esperienza acquisita dall'operatore, corrispondono a plancton (la tecnica quindi risulta alquanto soggettiva). Un semplice filtro soglia consente di visualizzare non solo i banchi di pesce ma anche i singoli individui dispersi nella colonna d'acqua. Con la tecnica tradizionale i valori di NASC per miglio possono quindi essere estratti sia da transetti notturni, quando il pesce è disperso nella colonna d'acqua, sia in transetti diurni, quando i singoli individui risultano principalmente aggregati in banchi. Inoltre, essendo solo le specie pelagiche i target delle stime di biomassa, una linea impostata sull'ecogramma a 0.5 m dal fondo elimina parte delle aggregazioni molto vicine ed esso è attribuite a demersali. Infine, ai valori di NASC per miglio viene attribuita la specie utilizzando la tecnica della cala più vicina e l'esperienza dell'operatore. L'algoritmo a multifrequenza invece isola dall'ecogramma i banchi (la tecnica funziona quindi solo su transetti acquisiti di giorno) e li discrimina automaticamente dal plancton, escludendo quella fase di soggettività che caratterizza la metodologia tradizionale. L'attribuzione per specie è invece effettuata tramite i modelli di classificazione (MaxRate o MaxRateSpecies e Mode).

Sebbene il valore di PRC_NASC estratto da i soli banchi a profondità comprese tra 30 e 140 sia nel totale superiore ($34091m^2/nmi^2$) al valore stimato dalla metodologia tradizionale ($32175 m^2/nmi^2$), il grafico in fig. 3.21 riporta comunque una forte correlazione tra di essi ($R^2=0.77$).

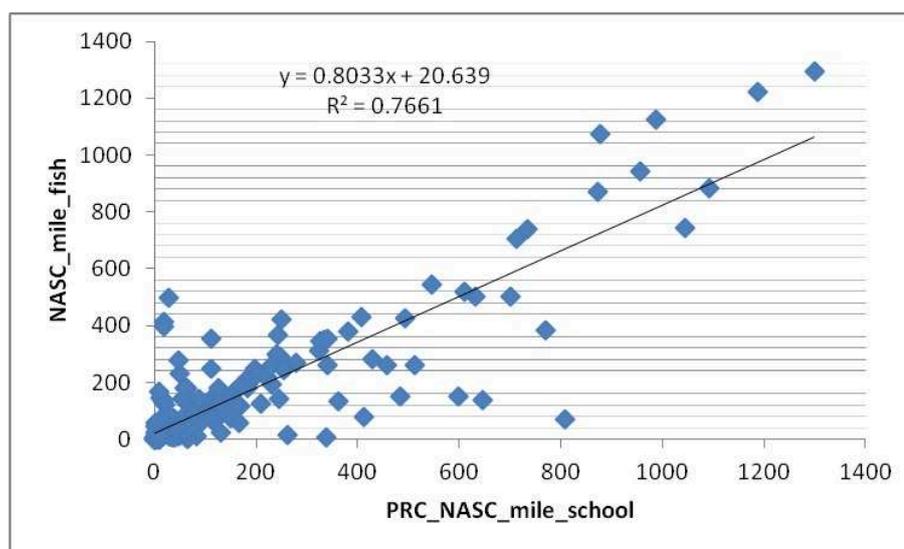


Figura 3.21: Il grafico riporta la relazione tra PRC_NASC per miglio estratto dai banchi (in ascissa) e il NASC per miglio stimato con la metodologia tradizionale (in ordinata).

Ciò indica che sebbene le due metodologie lavorino in maniera diversa risultano comunque tra loro confrontabili. La tabella 3.15 confronta le percentuali di PRC_NASC per specie ottenute dai tre modelli predittivi con i valori stimati dalla metodologia tradizionale per l'intero survey. Nei modelli predittivi i valori di PRC_NASC di demersali e OPS sono stati sommati, al fine di semplificare il confronto con le stime tradizionali dove una

valutazione dei demersali non è prevista (si ipotizza infatti di escluderli completamente impostando una linea a 0.5 m dal fondo).

Tabella 3.13: Valori e percentuale di biomassa (in m^2/nmi^2) per miglio e per specie stimati dai modelli predittivi MaxRate o MaxRateSpecies, Mode e dal metodo tradizionale, nei transetti giornalieri (1-60) del survey ANCHEVA2009, effettuati a profondità comprese fra 30 e 140 m.

tr	Biomassa OPS+Dem	Biomassa acciughe	Biomassa sardine	Biomassa notDet	Biomassa totale	Modello
1-60	19620	2722	9833	0	32175	Traditional
	61%	8%	31%	0%	100%	
1-60	16670	3632	13710	0	34091	MaxRate o MaxRateSpecies
	49%	11%	40%	0%	100%	
1-60	16508	3626	13934	24	34091	Mode
	48%	11%	41%	0%	100%	

In generale entrambi i due modelli (MaxRate o MaxRateSpecie e Mode) forniscono risultati consistenti con le stime tradizionali. Essi indicano infatti valori di biomassa a favore di specie diverse dalle specie target e consentono di ottenere percentuali più elevate per le sardine rispetto alle acciughe, in accordo con le stime tradizionali. Le capacità di classificazione dei tre modelli sono state successivamente analizzate per gruppi di 10 transetti, per capire se le differenze o le analogie osservate con il metodo tradizionale siano attribuibili a determinati settori dell'area di studio. Di seguito la tabella 3.14 visualizza i risultati ottenuti dalla metodologia tradizionale mentre i risultati ottenuti dai due modelli utilizzati per la classificazione sono riportati nelle Tabelle 3.15 e 3.16.

Tabella 3.14: valori e percentuali di biomassa per specie e gruppi di transetti ottenuti con il metodo tradizionale.

Biomassa tot metodo tradizionale	Biomassa OPS	Biomassa acciughe	Biomassa sardine	Tr
7953	5077	801	2076	1-10
	64%	10%	26%	
6414	4199	546	1669	10-20
	65%	9%	26%	
8969	3992	1030	3947	20-30
	44%	12%	44%	
2401	1447	204	750	30-40
	60%	8%	31%	
5177	3685	142	1351	40-50
	71%	2%	8%	
1260	1220	0	40	50-60
	97%	0%	3%	

Nei transetti 10-20 e 20-30 (corrispondenti al settore di piattaforma tra Sciacca e Licata, vedi Fig.2.2) si osservano le maggiori differenze tra i modelli decisionali utilizzati

Tabella 3.15: Valori e percentuali di biomassa per specie e gruppi di transetti ottenuti con il modello MaxRate o MaxRateSpecies.

Biomassa tot modello MaxRate	Biomassa OPS+dem	Biomassa acciughe	Biomassa sardine	Tr
8075	3044 38%	2165 27%	2866 35%	1-10
7718	3195 41%	342 4%	4181 54%	10-20
1492	288 19%	0 0%	1204 81%	20-30
5055	2002 40%	417 8%	2635 52%	30-40
2038	1602 79%	32 2%	404 20%	40-50
9713	6618 68%	675 7%	2420 25%	50-60

Tabella 3.16: Valori e percentuali di biomassa per specie e gruppi di transetti ottenuti con il modello Mode.

Biomassa tot metodo Mode	Biomassa OPS+dem	Biomassa acciughe	Biomassa sardine	Biomassa notDet	Tr
8075	2891 36%	2161 27%	3022 37%	0 0%	1-10
7716	3153 41%	342 4%	4221 55%	2 0%	10-20
1492	260 17%	19 1%	1213 81%	0 0%	20-30
5055	1989 39%	396 8%	2649 52%	22 0%	30-40
2038	1597 78%	32 2%	409 20%	0 0%	40-50
9713	6618 68%	675 7%	2420 25%	0 0%	50-60

e il metodo tradizionale per le stime di biomassa. Nei transetti 10-20 infatti i modelli attribuiscono gran parte della biomassa stimata alle sardine piuttosto che al gruppo di OPS. Nei transetti 20-30, dove le percentuali di biomassa appaiono equamente distribuite tra OPS e sardine con la metodologia tradizionale (vedi Tab. 3.14), i modelli sono invece a favore esclusivamente delle sardine per entrambi i modelli classificativi (vedi Tab. 3.15 e 3.16). La maggiore corrispondenza con le stime di biomassa si osserva invece principalmente nel gruppo di transetti 40-50, ricadenti nel settore di piattaforma che segue Gela e precede Capo Passero.

Differenze evidenti tra il metodo tradizionale e i modelli decisionali si osservano invece in termini di biomassa totale per gruppi di transetti. Nei transetti 10-20 e 20-30 si osservano le maggiori differenze. Nel gruppo di transetti 20-30, ad esempio, i valori di biomassa totali stimati con la metodologia tradizionale (Tab. 3.14) risultano molto più alti rispetto ai valori per miglio estratti esclusivamente dai banchi (Tab. 3.15 e 3.16). Maggiori valori di biomassa in questo settore sono probabilmente da attribuire ad un maggior contributo del pesce disperso nella colonna d'acqua alla biomassa per miglio rispetto a quello fornito dai banchi di pesce (Fig. 3.22). Laddove invece i valori di biomassa stimati con la metodologia tradizionale risultano molto più bassi rispetto ai valori di biomassa per miglio associati ai banchi (es. transetti 50-60), è stato osservato un maggior numero di banchi localizzati al fondo ed esclusi con la metodologia tradizionale (Fig. 3.23).

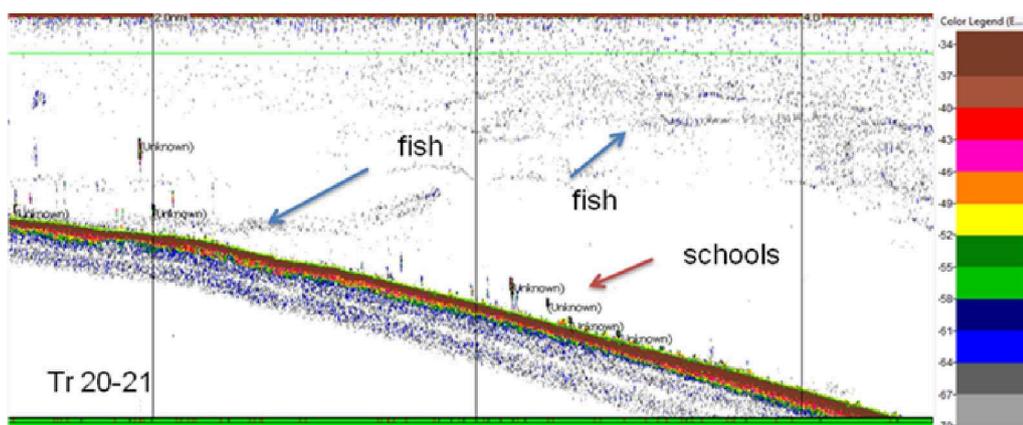


Figura 3.22: Porzione del transetto 20-21. Le linee orizzontali rappresentano la suddivisione in miglia nautiche. Le frecce blu indicano il segnale corrispondente a pesce disperso nella colonna d'acqua, mentre la freccia rossa indica la traccia acustica prodotta da banchi di pesce.

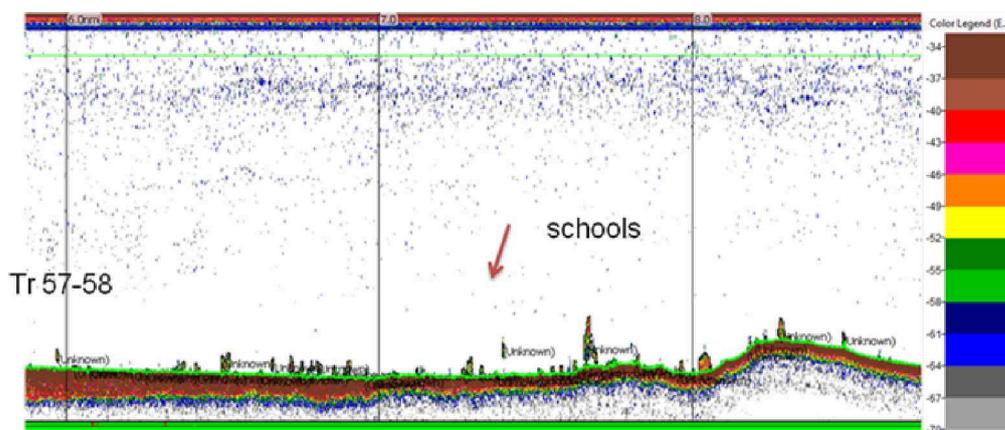


Figura 3.23: Porzione del transetto 57-58. Le linee orizzontali rappresentano la suddivisione in miglia nautiche. La freccia rossa indica la traccia acustica prodotta da banchi di pesce localizzati sul fondo.

3.2.6 Modelli di classificazione ad albero costruiti su variabili ambientali e parametri strutturali del banco.

In questo paragrafo si analizzano i risultati ottenuti nella discriminazione di banchi di specie diverse aggiungendo le variabili ambientali ai modelli di classificazione ad albero. Le variabili ambientali scelte come variabili predittive comprendono i parametri della colonna d'acqua, ricavati da CTD, e i valori di intensità di backscatter, estratti integrando l'intero secondo eco dal fondo e usati come indicatori delle proprietà di durezza del substrato marino. Tra le variabili estratte dal CTD sono state selezionate esclusivamente temperatura e salinità sopra il termoclino e profondità del termoclino (indicati come Temp_up_thermo, e salinity_up_thermo, Depth_thermo), mentre sono state escluse le variabili dipendenti dal fondo (temperatura, salinità media al fondo). Essendo però le condizioni ambientali differenti tra il periodo estivo e quello autunnale è stato utilizzato il learning sample ridotto alle sole cale estive. Quindi un dataset con totale di 396 banchi è stato adottato. Avendo verificato nei paragrafi precedenti che l'albero può avere una struttura diversa e incorporare anche variabili differenti a seconda del training set scelto, decidiamo di tenere in considerazione la variabilità del nostro learning sample stimando trenta alberi differenti ed estraendo da questa procedura i modelli MaxRate e MaxRateSpecies, che nel paragrafo precedente sono stati utilizzati a scopo classificativo.

In questo caso tutti e trenta gli alberi stimati hanno tassi di classificazione molto alti per le specie (in media 92% sardine, 81% acciughe, 100% demersali e 98% OPS), e i modelli MaxRate e MaxRateSpecies coincidono. La capacità di classificazione del MaxRate è in media del 100% per le sardine, 95% per le acciughe, 100% per demersali, 96% per OPS. La struttura del Modello è riportata in Fig. 3.24.

Dai modelli stimati di evidenza immediatamente che i tassi di classificazione per specie sono veramente molto alti e tra l'altro nessuna delle variabili morfologiche o energetiche, riscontrate nei modelli precedenti ad eccezione del valore minimo di Sv a 120 kHz (Sv_min_120), viene più selezionata per la discriminazione della specie. È ipotizzabile

quindi che l'aggiunta delle variabili ambientali possa non essere d'aiuto nel processo di identificazione per scopi classificativi. Tutti i banchi associati ad una cala hanno infatti lo stesso valore medio relativamente al parametro ambientale e quindi la distribuzione delle specie prodotta dal modello rispecchia zone del canale completamente diverse dal punto di vista ambientale. Tra l'altro però non ci si aspetta nemmeno che tra banchi vicini all'interno di una stessa zona ci siano differenze significative in termini di temperatura o salinità media sopra il termocline. Il modello quindi risulta alla fine solo indicativo di habitat differenti o meglio di condizioni ambientali diverse tra settori del Canale di Sicilia. I risultati ci inducono quindi a pensare che il problema vada studiato utilizzando un dataset più ampio, dotato di maggiore contrasto.

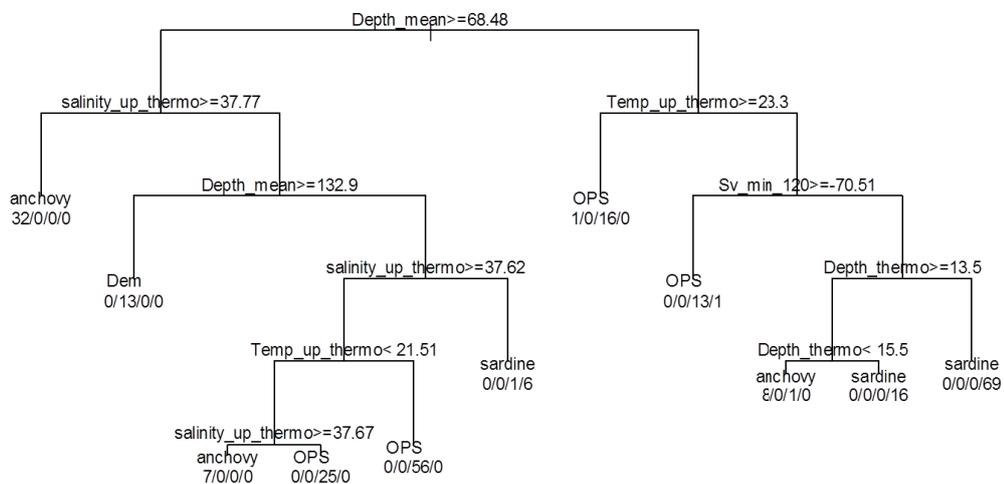


Figura 3.24: Modello MaxRate stimato usando come predittori variabili morfologiche, batimetriche energetiche e parametri della colonna d'acqua.

La variabile durezza del fondo, estratta acusticamente utilizzando il secondo eco, è stata inserita nei modelli predittivi separatamente rispetto alle altre variabili ambientali. Questo perché il secondo eco dal fondo non è stato registrato in corrispondenza di alcune cale con rete pelagica; va sottolineato, infatti, che in anni precedenti il segnale successivo al primo eco del fondo non veniva registrato perché l'analisi degli ecogrammi era esclusivamente finalizzata alle stime di biomassa ittica. Il learning sample è stato quindi ridotto a 370 banchi, di cui 70 acciughe, 21 demersali, 152 OPS e 127 sardine. Per comprendere quale fosse l'effetto della durezza del fondo sulle capacità di classificazione di banchi di specie diverse è stato necessario prima stimare nuovamente, su un dataset ristretto, il modello MaxRate, risultante dalla scelta di trenta training set e test set differenti, sulle sole variabili strutturali: batimetriche, morfologiche ed energetiche. I trenta differenti alberi consentono

buone capacità di discriminazione di banchi di sardine e OPS (Fig. 3.25 a,d) ma le capacità di classificare correttamente le acciughe sono molto variabili e in media intorno al 50% (Fig. 3.25 b).

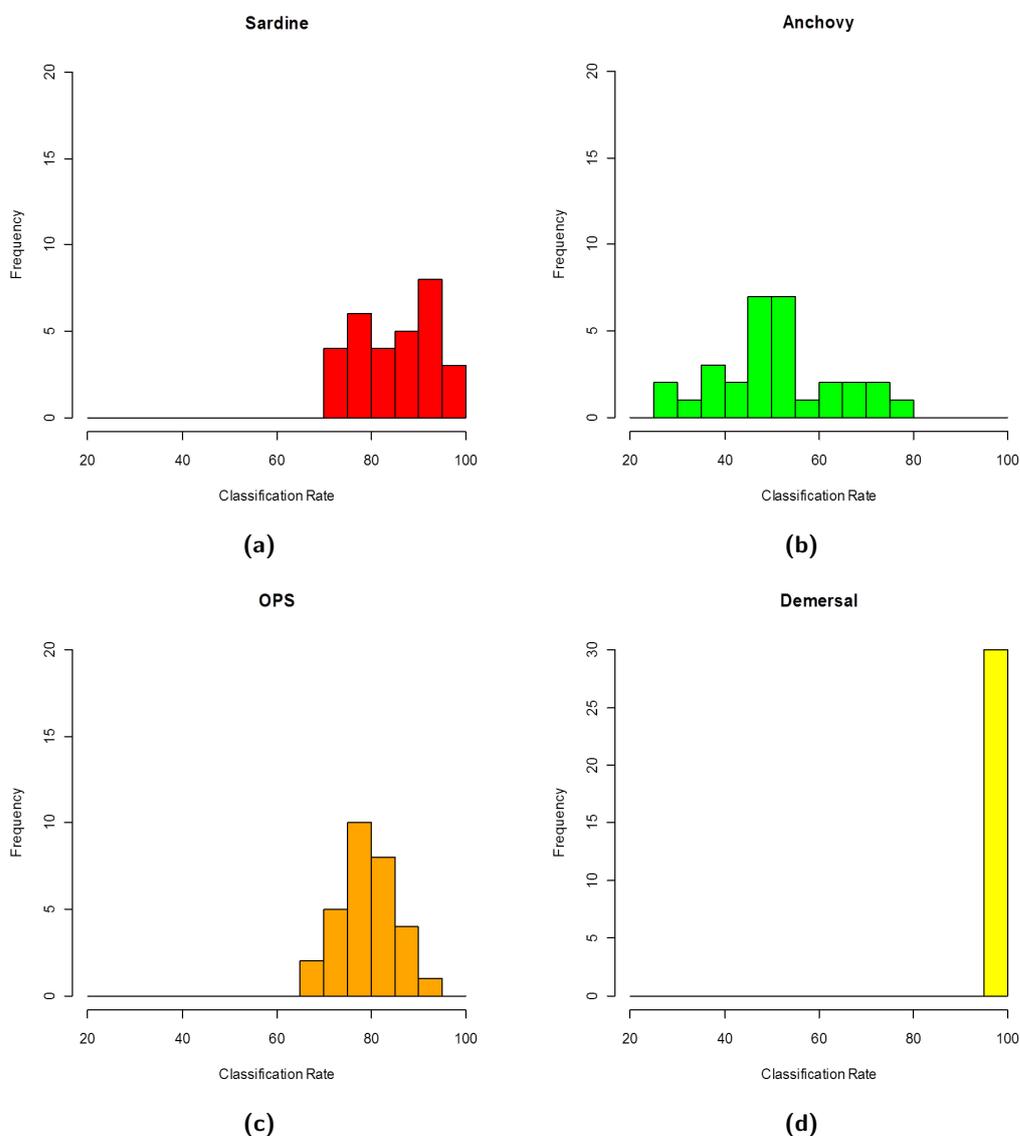


Figura 3.25: Distribuzione delle capacità di classificazione di trenta differenti alberi per i 4 gruppi di specie: (a) sardine, (b) acciughe, (c) demersali e (d) OPS.

Il modello MaxRate estratto dai trenta alberi è mostrato in Fig. 3.26. Il modello MaxRate classifica correttamente l'100% di banchi di sardine, il 83% di demersali, l'100% di OPS e l'56% di acciughe. Profondità del banco, distanza dal fondo Sv_min_120 e la dimensione frattale sono i predittori scelti per la suddivisione in sottogruppi. A profondità

inferiori a 68 m il modello individua banchi di OPS e sardine che sono tra loro discriminate sulla base del valore minimo di Sv misurato nel banco a 120 kHz e della distanza dal fondo. A profondità intermedie (comprese tra 74 e 79 m) la dimensione frattale e la distanza dal fondo a discriminare tra banchi di acciughe, sardine e OPS.

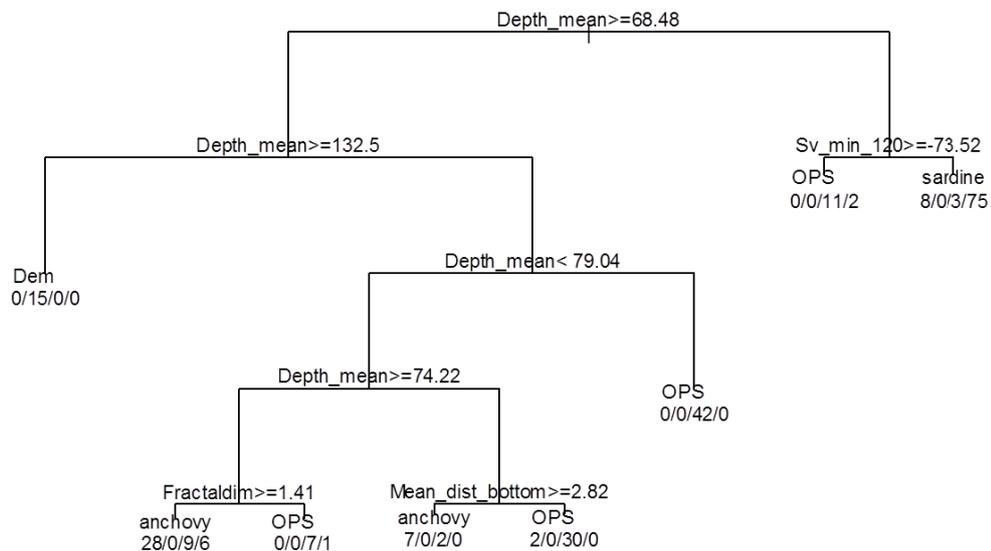


Figura 3.26: Modello MaxRate stimato usando come predittori variabili morfologiche, batimetriche, energetiche del banco.

Alle variabili predittive è stato poi aggiunto l'indice di durezza del fondo (NASC_E2) e trenta nuovi alberi sono stati stimati estraendo infine il MaxRate. Il nuovo modello (Fig. 3.27) con l'aggiunta della durezza del fondo ci consente di ottenere maggiori capacità di classificare correttamente sardine e OPS (89% e 100% rispettivamente per sardine e OPS) e acciughe (68%). La durezza del fondo interviene nella discriminazione di sardine e acciughe localizzate a profondità inferiori a 68 m, sebbene i due gruppi terminali relativi a queste specie siano poco numerosi.

Purtroppo non avendo campioni di sedimento in corrispondenza delle catture sperimentali non è stato possibile verificare a quale composizione granulometrica corrispondessero queste differenze in termini di distribuzione delle specie. Anche il tema del fondo deve essere tuttavia trattato più accuratamente e su un dataset più ampio prima di potere avanzare delle ipotesi relativamente alle capacità della durezza del fondo a discriminare tra specie diverse.

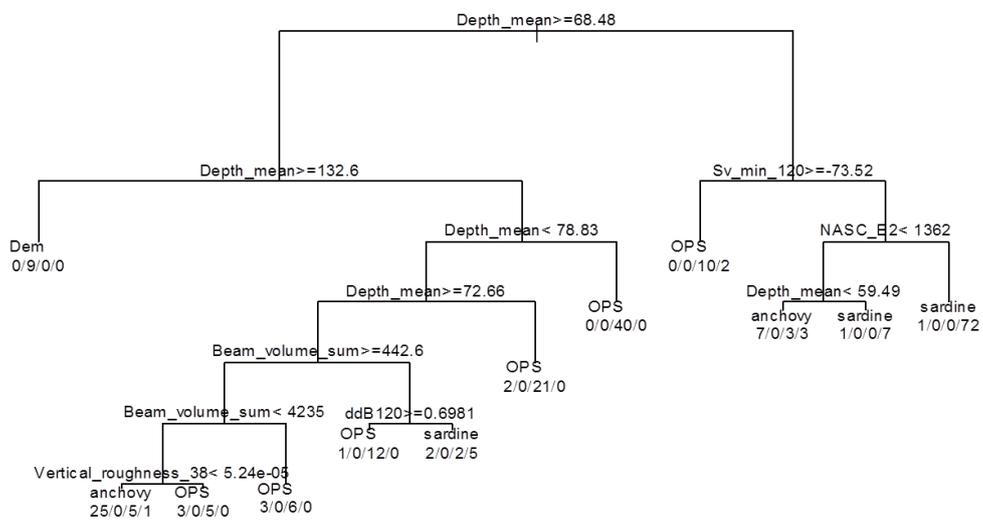


Figura 3.27: Modello MaxRate stimato utilizzando le variabili strutturali e la variabile durezza del fondo tra i parametri ambientali.

Capitolo 4

Discussioni

Le tecniche acustiche sono oggi strumenti importanti per le stime di biomassa e lo studio della distribuzione spaziale degli organismi marini che vivono nella colonna d'acqua. Questo è particolarmente vero per specie come acciughe, sardine e altre specie pelagiche che si riuniscono in banchi durante le ore diurne formando delle eco tracce facilmente individuabili dagli echosounder. Al fine di stimare l'abbondanza di una risorsa, i survey acustici sono molto efficienti in quanto forniscono, per ampie superfici, dati continui su oggetti presenti nella colonna d'acqua ad una risoluzione molto alta (tipicamente 20 cm lungo l'asse verticale). Tuttavia gli oggetti acusticamente individuati richiedono ancora un processo di identificazione, noto come "processo di interpretazione dell'ecogramma", che viene solitamente effettuato usando le catture sperimentali e l'esperienza dell'operatore a cui è affidata l'interpretazione, introducendo una certa soggettività nella procedura di stima. La capacità di identificare banchi di pesce, direttamente da dati acustici, associandoli ad una particolare specie, consentirebbe studi più dettagliati sulla distribuzione e il comportamento aggregativo dei banchi e fornirebbe un importante contributo all'accuratezza delle stime di biomassa, specialmente quando le informazioni derivanti dalle catture sperimentali non riescono a dare sufficiente copertura spaziale o dettaglio in relazione alla distribuzione della biomassa complessiva di pesce (NASC fish).

In questo lavoro di tesi diverse procedure per identificare le specie più abbondanti (*Engraulis encrasicolus* e *Sardina pilchardus*) nel Canale di Sicilia sono state applicate ai dati acustici acquisiti durante i survey.

In una prima fase dello studio una semplice caratterizzazione e identificazione dei banchi (senza distinzione per specie) è stata eseguita isolando le aggregazioni di pesci con un algoritmo in Matlab sviluppato da Patti et al. (2006). Le informazioni così estratte da una singola frequenza (38 kHz) sono state analizzate in relazione alla densità e alla biomassa di plancton, considerato indice della disponibilità di cibo. L'analisi dei banchi di pesce e delle aggregazioni planctoniche acusticamente individuate nel survey "Ancheva98" ha evidenziato due tipi di interazioni: una relazione inversa, che si osserva in Zona 1 (parte nord-occidentale dell'area di studio, ricadente sul Banco Avventura) per bassi valori di biomassa e densità planctonica, attribuita principalmente alle aggregazioni di plancton profondo; e una relazione diretta, localizzata in Zona 3 (parte sud-orientale dell'area

di studio, ricadente sulla piattaforma Maltese) che coinvolge aggregazioni di plancton dell'intera colonna. La relazione inversa osservata potrebbe essere indicativa di un consumo della preda (plancton) da parte dei predatori (pesci). Si potrebbe ipotizzare che in Zona 1 si trovino banchi di pesce in grado di ridurre la biomassa del plancton a basse densità. Effetti simili sono stati osservati sulla distribuzione e l'abbondanza di organismi marini in studi precedenti. Per esempio, un'alta abbondanza di jack mackerel (*Trachurus murphii*) è stata osservata a largo della costa del Cile in relazione a una bassa abbondanza della comunità mesopelagica (Bertrand *et al.*, 2004). Un effetto di diminuzione del biovolume di zooplancton in presenza di alte densità di acciughe è stato osservato anche a largo delle coste del Perù (Ayon *et al.*, 2008). Ancora Swartzmann *et al.* (2002) e Winter & Swartzmann (2006) hanno evidenziato una relazione inversa tra i giovanili di pollock d'Alaska (*Theragra Chalcogramma*) e lo zooplancton vicino le isole Pribilof. L'interazione positiva osservata in Zona 3 potrebbe indicare invece che, a più alte densità di plancton, banchi di piccoli pelagici abbiano una preferenza per aree caratterizzate da un'alta disponibilità di cibo. Lo stesso effetto positivo è stato riscontrato da Maravelias *et al.* (2000), analizzando l'effetto della biomassa zooplanctonica sull'abbondanza di aringhe.

Il differente comportamento tra Zona 1 e Zona 3 potrebbe essere attribuito a specie dominanti differenti all'interno delle due aree. Un primo suggerimento proviene dalla differente distribuzione batimetrica dei banchi nelle due aree e dalla distanza dal fondo più bassa in Zona 1 rispetto alla Zona 3. Altre informazioni utili provengono dall'analisi della composizione in specie delle catture sperimentali, effettuate durante il survey acustico "Ancheva98". Le due cale effettuate in Zona 1 mostrano una dominanza di demersali e altre specie pelagiche non target come tracuri (*Trachurus spp.*) e sgombri (*Scomber spp.*) in Zona 1, mentre in Zona 3 gran parte delle catture sperimentali (due su tre) sono composte da sardina (*Sardina pilchardus*), la quale risulta la specie più abbondante nel survey del 1998. Sebbene durante il survey gran parte della biomassa di sardine era concentrata in Zona 2 (Patti *et al.*, 2004), la relazione diretta e significativa osservata in Zona 3 potrebbe essere relativa a più bassi valori di biomassa e densità planctonica nella colonna d'acqua rispetto alla Zona 2. In particolare si può pensare che in Zona 3 la proporzionalità diretta tra biomassa media dei banchi e aggregazioni di plancton testimoni una più bassa disponibilità di cibo rispetto alla Zona 2, che spingerebbe i banchi più energetici alla ricerca attiva di aree relativamente più ricche di plancton. Peraltro questo tipo di interazione è presente anche in Zona 2 (sebbene non sia risultata statisticamente significativa usando i GAM) dove la relazione diretta è parzialmente mascherata da valori locali più alti, in confronto alla Zona 3. (vedi Tab.3.1). Sfortunatamente nessun campionamento di plancton è stato effettuato durante il survey Ancheva98, quindi non è stato possibile verificare il risultato acustico con misure dirette sull'abbondanza e la composizione di plancton. Tuttavia dati oceanografici ed ittioplanctonici raccolti approssimativamente nello stesso periodo del survey Ancheva98 che evidenziano più alti tassi di produzione primaria e di densità di larve di pesci concentrato in Zona 2 (e Zona 3) rispetto alla Zona 1 (García Lafuente *et al.*, 2002). Per quanto riguarda la composizione del meso-zooplancton (che comprende organismi di taglia compresa nel range 0.2 -20 mm), essa è stata investigata durante due survey planctonici estivi effettuati tra il 1999 e il 2000, mostrando una struttura

della comunità dominata da specie di copepodi, appartenenti ai generi *Clausocalanus*, *Centropages* e *Oncaea*, i quali sono stati indicati come cibo preferenziale per le acciughe (Aguzzi *et al.*, 2003). Sebbene la taglia massima di questi organismi sia solitamente inferiore rispetto al range individuabile a 120 kHz, organismi zooplanctonici più piccoli sono ancora visibili ad alte concentrazioni. Di conseguenza, le informazioni disponibili sulla distribuzione, l'abbondanza e la composizione di plancton sebbene indirette, indicano che i risultati derivanti dall'analisi acustica possono testimoniare condizioni locali più favorevoli per la sopravvivenza e la crescita di differenti specie di pesci. Questo primo approccio ha consentito di formulare delle ipotesi sulla possibile composizione in specie dei banchi localizzati nelle due zone in cui era stata riscontrata una relazione significativa con la disponibilità di cibo. Ma tale approccio ha però evidenziato notevoli limiti legati principalmente alla carenza di informazioni sulla reale composizione delle specie a mare.

In una seconda fase della tesi si è cercato di individuare quali parametri, fra i descrittori del banco e le variabili ambientali, potessero essere utili nel discriminare queste specie. Per il raggiungimento di questo obiettivo è stato realizzato un algoritmo che potesse rendere i dati acustici a differenti frequenze spazialmente comparabili, in modo da poter estrarre le informazioni energetiche a più frequenze ed è stato creato un *learning sample* di banchi estratti acusticamente da tracciati di cale con composizione monospecifica, effettuate nel corso di echosurvey estivi e autunnali dal 2002 al 2009. In un primo momento sono stati analizzati esclusivamente i parametri morfologici e batimetrici, provenienti dall'analisi dei dati acquisiti alla frequenza di 38 kHz, per caratterizzare e discriminare le due specie target e altri gruppi di specie presenti.

L'analisi di questi parametri mostra che banchi di acciughe e sardine differiscono tra loro principalmente in termini di area e volume con valori maggiori per le acciughe rispetto alle sardine. Differenze significative tra sardine e Other Pelagic Species (OPS) si hanno invece in termini di area, lunghezza, perimetro e variabilità nella forma (Unevenness1): i banchi di OPS risultano più lunghi, hanno area e perimetro maggiore ma anche maggiore variabilità nella forma rispetto a banchi di sardine. La maggiore variabilità nella forma riflette probabilmente la presenza di specie diverse all'interno del gruppo OPS. Tutti i gruppi di specie, in ogni caso, mantengono durante il giorno una forma "oblunga" (il rapporto medio tra lunghezza e altezza è infatti superiore a 8). Lo stesso tipo di rapporto è stato riscontrato da Zwolinski *et al.* (2007) per banchi di *Sardina pilchardus* individuati lungo la costa portoghese.

I quattro gruppi di specie (sardine, acciughe, OPS e demersali) mostrano invece preferenze per batimetrie differenti. Banchi di demersali si individuano a profondità più elevate e hanno distanze dal fondo inferiori rispetto alle altre specie; banchi di acciughe hanno profondità medie maggiori delle sardine, ma distanze medie dal fondo molto simili ad esse; contrariamente a OPS per i quali si individua una distribuzione batimetrica più ampia e distanze dal fondo maggiori. Differenze nella distribuzione spaziale tra le due specie target sono già state osservate in studi precedenti. Scalabrin & Massé (1993), ad esempio, osservano nella Baia di Biscaglia una segregazione batimetrica tra acciughe (*Engraulis encrasilolus*), sardine (*Sardine pilchardus*), tracuri (*Trachurus trachurus*) e sgombri (*Scomber scombrus*) in due anni investigati (1990, 1991), con le sardine più vicine

alla costa, le acciughe nella zona intermedia (tra 50 e 100 m), e sgombri e tracuri più a largo. Più di recente Zwolinski et al. (2010) individuano, a largo della costa occidentale Portoghese, banchi di sardine preferenzialmente a profondità inferiori a 100 m e tipicamente in acque più costiere nel periodo estivo (tra 40 e 45 m). Una segregazione spaziale è stata osservata anche nel Nord del Mar Egeo e nel Golfo del Benguela tra *Engraulis encrasicolus* e *Sardina pilchardus* da Giannoulaki et al. (2005) e tra *Sardinops sagax* ed *Engraulis encrasicolus* da Barange et al. (2005). Questi ultimi evidenziano che la segregazione spaziale tra le due specie persiste quando i livelli di biomassa risultano bassi. In presenza di alti livelli di biomassa invece le due specie mostrano un'alta sovrapposizione spaziale. Questa evidenza non sembra verificarsi nel Mediterraneo, molto probabilmente perché i livelli di biomassa presenti sono sempre molto più bassi rispetto a quelli riportati da Barange et al. (2005). Tugores et al. (2010) riportano infatti, anche per il Mediterraneo, una segregazione spaziale persistente tra le due specie target, sebbene quest'ultima analisi sia stata effettuata esclusivamente in tardo autunno, quando le acciughe sono in fase di reclutamento e le sardine nel periodo riproduttivo. Queste differenze nella distribuzione spaziale osservate in Mediterraneo sono probabilmente relative ad habitat preferenziali. Sardine e acciughe occupano infatti differenti nicchie ecologiche in termini di masse d'acqua, tolleranza ai fattori abiotici, capacità migratorie e nutrizione (Bertrand et al., 2004; Gutierrez et al., 2007).

Gli altri descrittori strutturali dei banchi estratti dalla singola frequenza a 38 kHz sono invece molto variabili. Le caratteristiche morfologiche, ad esempio, possono variare anche all'interno di una stessa specie in relazione alle condizioni ambientali e quindi all'area geografica in cui si trovano (Soria et al., 1993; Swartzmann 1997) o ancora in relazione alla presenza di altre specie (Massé et al., 1996; Scalabrin et al., 1996, Misund et al., 1998). Essi sono quindi buoni descrittori del comportamento aggregativo di una specie all'interno di una certa area geografica ma da soli difficilmente possono essere usati come strumenti di classificazione. L'uso invece del solo parametro batimetrico (profondità del banco), le cui differenze tra le specie sembrano persistenti nel tempo, escluderebbe la possibilità di discriminare i diversi banchi a quelle fasce batimetriche dove specie come acciughe e sardine possono essere compresenti.

Si è scelto quindi di impiegare un secondo approccio per l'identificazione della specie, consistente nell'uso di dati energetici a più frequenze e nell'analisi delle differenze energetiche tra le due specie target. Questo studio è il primo a classificare in Mediterraneo specie simili come acciughe e sardine usando informazioni a più frequenze. Fernandes et al. (2006) sostengono che le proprietà energetiche del banco a più frequenze siano attribuibili principalmente alle caratteristiche di una specie in termini di forma, presenza o meno di vescica natatoria, o taglia degli individui. Sebbene acciughe e sardine abbiano entrambe vescica natatoria e caratteristiche simili in termini di forma, stagionalmente presentano in media taglia differente, per cui differenze significative tra di esse potrebbero osservarsi in termini energetici. L'analisi delle proprietà acustiche dei banchi di acciuga e sardina rivela simili valori di backscatter alle frequenze di 38 e 120 kHz ed elevata variabilità attorno ai valori medi. La differenza tra l'energia retrodiffusa a 38 kHz e l'energia retrodiffusa a 120 kHz (ddB120) risulta essere un parametro importante per distinguere tra individui

giovanili di acciughe e sardine, ma una discriminazione tra individui di taglia diversa non appare comunque possibile sulla base della sola caratteristica energetica. Questo si afferma in ragione dell'elevata sovrapposizione nella distribuzione dei valori di ddB attorno alla media delle due specie. Le stesse conclusioni sono state ottenute da Fässler et al. (2007) dopo il confronto delle variabili energetiche tra Norway pout (*Trisopterus esmarki*) e aringhe (*Clupea harengus*). L'uso della ddB e dei valori di backscatter alle varie frequenze appare invece promettente nella discriminazione di specie prive di vescica natatoria come gli sgombri (*Scomber scombrus*) da altre specie (Korneliussen, 2010).

Differenze significative nelle proprietà energetiche si osservano anche tra banchi di adulti delle due specie target per range di profondità. Per profondità del banco inferiori (tra 40 e 60 m) la $ddb120$ assume valori negativi per le sardine e positivi per le acciughe, mentre per profondità del banco maggiori la differenza in decibel tra le due frequenze risulta positiva per le sardine e negativa per le acciughe. Incrementi nella profondità tendono probabilmente a modificare il contributo della vescica natatoria sul totale dell'energia, determinando un cambiamento nei valori di ddB al variare della profondità all'interno della stessa specie. Generalmente la vescica natatoria contribuisce per il 90% al totale dell'energia retrodiffusa rendendo trascurabile il contributo proveniente dal corpo. Quando una specie, che non è in grado di immettere gas nella propria vescica natatoria scende in profondità, il volume della vescica diminuisce per l'incremento della pressione, in accordo con la legge di Boyle (Nero et al., 2004, Fässler et al., 2009, Ona 1990). Questo tipo di comportamento è tipico di pesci fisostomi, come i Clupeidiformi, che non possiedono organi come la ghiandola del gas, in grado di aggiustare attivamente il volume della vescica natatoria al cambiare della pressione e quindi della profondità. La profondità sembra quindi avere un effetto sui valori di backscatter delle due specie target. Tuttavia, nemmeno considerando l'effetto della profondità sui banchi di individui adulti, appare possibile discriminare tra acciughe e sardine, data la notevole sovrapposizione nella variabilità. Quindi né l'approccio che fa uso dei soli parametri strutturali ad una frequenza, né l'approccio che fa uso dei soli parametri energetici a più frequenze sembra promettente nella soluzione del problema dell'identificazione della specie per banchi di acciughe e sardine, sebbene dal punto di vista statistico si osservino differenze significative tra le due specie target per alcuni dei parametri morfologici, batimetrici ed energetici a più frequenze. Non si può escludere inoltre che differenze significative nei valori di $ddb120$ o dei parametri morfologici siano da attribuire all'erronea assegnazione della specie ai banchi visualizzati nell'ecogramma. In alcuni casi infatti le informazioni ricavate dall'ecogramma potrebbero non corrispondere esattamente alle informazioni sulla specie ricavate dalle catture sperimentali.

Il terzo approccio usato per risolvere il problema dell'identificazione della specie è stato quello di estrarre dai banchi tutte le possibili informazioni (energetiche a più frequenze, morfologiche e batimetriche) e testarne le capacità di discriminazione usando modelli di classificazione ad albero. Per costruire un modello che potesse avere anche capacità classificative, oltre che descrittive, sono stati inseriti nel learning sample anche banchi di altri gruppi di specie (OPS e demersali). È necessario infatti creare un learning sample che risulti rappresentativo della composizione in specie dell'area di studio perché questi modelli possano essere utilizzati per classificare le tracce acustiche corrispondenti a banchi

estratti da un qualunque survey.

Un training set random è stato selezionato per costruire il modello con tutte le variabili morfologiche, batimetriche ed energetiche. Il modello indica buone capacità di discriminare tra OPS, sardine e demersali, con valori rispettivamente dell' 82%, 89% e del 100%; per quest'ultimo gruppo di specie però il dataset è veramente ridotto per cui è difficile fare delle considerazioni su di esso. Il modello ha invece capacità più contenute (pari al 65%) nel classificare banchi di acciughe. La profondità del banco, fortemente correlata alla profondità del fondo, risulta la variabile predittiva più importante; escludendola dal modello infatti le capacità predittive totali si riducono dall'81% al 60%, con valori per specie del 22% per le acciughe e del 62% per le sardine.

Il contributo alla classificazione fornito dalle variabili energetiche e morfologiche è molto simile per le sardine, mentre le sole variabili morfologiche classificano meglio le acciughe, rispetto alle variabili energetiche. Le variabili energetiche sembrano quindi più importanti nella discriminazione di banchi di sardine piuttosto che di acciughe. Questo potrebbe essere una conseguenza delle differenti strategie alimentari tra le due specie. Nel nord del Benguela, ad esempio, le sardine sono principalmente filtratori mentre le acciughe sono consumatori di plancton di dimensioni maggiori e quindi predatori (Van der Linger, 1994). Van der Linger (1994), in accordo con Eggers (1976), sostiene che i predatori usino strategie di attacco quando localizzano la loro preda, richiedendo un incremento nelle distanze tra gli individui di un banco per evitare la competizione per il cibo. Banchi di acciughe quindi potrebbero trarre beneficio nel rompere la loro struttura in presenza di cibo, contrariamente alle sardine le quali, essendo principalmente filtratori, non traggono alcun beneficio energetico nel rompere la struttura del proprio banco. In mediterraneo e soprattutto nel Canale di Sicilia in realtà si conosce ancora ben poco sulle strategie alimentari di acciuga e sardina. Le uniche informazioni dal Canale di Sicilia provengono da Aguzzi et al. (2003) che riportano le specie di copepodi appartenenti ai generi *Clausocalanus*, *Centropages*, and *Oncaea*, come elemento principale della dieta di *Engraulis encrasicolus*.

L'importanza invece della profondità del banco nella discriminazione di specie come acciughe e sardine è stata ampiamente documentata. Lawson et al. (2001), ad esempio facendo uso di un'analisi discriminante, individuano una riduzione della capacità totale di classificazione dall' 88.3% (in presenza di profondità), al 64.5% (escludendo la profondità) di *Engraulis encrasicolus*, *Sardinops sagax*, ed *Etrumeus whitehead* (specie pelagiche della costa meridionale africana), con riduzioni maggiori delle capacità di classificazione delle acciughe rispetto alle sardine, in accordo quindi con i risultati ottenuti in questo studio. Risultati simili sono stati riscontrati da Cabreira et al. (2009) nella discriminazione di specie demersali e pelagiche lungo la piattaforma continentale dell'Argentina. Essi riscontrano una capacità totale di classificazione delle specie dell' 84.1% considerando tutti insieme i descrittori batimetrici, morfologici ed energetici, ed una riduzione delle capacità di classificazione al 67.9% in assenza dei parametri batimetrici. Più di recente anche Robotham et al. (2010) indicano i parametri batimetrici come descrittori più importanti nella discriminazione delle specie pelagiche più comuni lungo la costa centro-meridionale del Cile, *Engraulis ringens* (acciuga comune), *Strangomera bentincki* (sardina comune) e

Trachurus Murphy

La simulazione di trenta differenti modelli ha consentito di capire che la variabilità dei parametri morfologici, batimetrici ed energetici, anche all'interno della stessa specie, possa influenzare la struttura, la dimensione e quindi le capacità di classificazione dei modelli decisionali. Il comportamento di un banco varia infatti nello spazio e nel tempo determinando una variazione dei parametri che ne descrivono la struttura (Freon and Misund, 1999). Dalla simulazione è stato possibile osservare che gran parte dei modelli costruiti su tutte le variabili strutturali ha buone capacità (tra 70 e 90%) di classificare correttamente sardine e OPS da altre specie, mentre la capacità di discriminare acciughe è piuttosto variabile (dal 40% all' 85%). Miglioramenti nelle capacità di classificazione di questa specie potrebbero essere ottenute con l'aggiunta di nuove osservazioni al dataset iniziale o da una separazione per specie e classi di taglia del gruppo di OPS.

I risultati della simulazione di trenta differenti modelli hanno posto il problema di comprendere quale training set fosse più opportuno scegliere per classificare i banchi di un qualunque survey. Per tenere in considerazione la variabilità del learning sample, la scelta è stata infine orientata verso tre modelli denominati per questo studio come "MaxRate", "MaxRateSpecies" e "Mode". Il primo (MaxRate) utilizza fra trenta training set differenti, quello che fornisce il più alto tasso di classificazione totale; il secondo "MaxRateSpecies" utilizza il training set che fornisce tassi di classificazione superiori al 75% per le due specie target (acciughe e sardine), e infine il modello "Mode" che adopera tutti e trenta i training set differenti per costruire i modelli e seleziona la classe più frequente per fare previsioni su un singolo banco. I primi due modelli possono essere utilizzati sia come strumenti descrittivi che come modelli classificativi, contrariamente al modello "Mode" che è stato generato esclusivamente per scopi classificativi. I modelli MaxRate e MaxRateSpecies, estratti dalla simulazione, forniscono alte capacità di classificazione di acciughe e sardine e determinano entrambi una separazione di banchi di adulti e giovanili di acciughe. Questa separazione si basa principalmente sulla differente distribuzione batimetrica di individui di taglia diversa; i giovanili di acciuga risultano più costieri rispetto agli adulti. Un comportamento più costiero di giovanili di acciughe è stato evidenziato anche da Lawson et al. (2001). La discriminazione di adulti di acciughe e giovanili da altre specie avviene poi sulla base di descrittori morfologici (Lunghezza, volume e dimensione frattale) e posizionali (distanza dal fondo). Ciò suggerisce che l'insieme dei parametri strutturali possa discriminare anche tra individui di taglia differente e che giovanili di una specie possano avere un comportamento differente da quello degli adulti. Quindi per applicare modelli classificativi ad un qualunque survey bisogna tener conto delle classi di taglia per specie presenti stagionalmente. Sulla base di questa considerazione per classificare i banchi del survey Ancheva2009, condotto nel periodo estivo, dove i banchi di acciughe erano esclusivamente costituiti da adulti in fase di deposizione, è stato necessario ridurre il learning sample eliminando dal dataset banchi di giovanili di acciughe che non erano presenti nello stesso periodo. Con il learning sample ridotto sono stati simulati nuovamente trenta diversi modelli selezionando per scopi classificativi un nuovo modello MaxRate (concidente in questo caso anche con il modello MaxRateSpecies) e un modello Mode.

Per valutare la bontà dei modelli classificativi è stato analizzato il grado di accordo degli

stessi con le informazioni di biomassa per specie, ricavate con la metodologia tradizionale. Un primo risultato è stato ottenuto confrontando il PRC_NASC dell'intero survey (ottenuto dalla somma dei PRC_NASC per miglio nautico estratti dai banchi) con il NASC ricavato dalla metodologia tradizionale. Il contributo dei banchi al NASC è infatti consistente con le stime di biomassa che non sono esclusivamente il risultato dei banchi ma anche di tutto ciò (principalmente esemplari singoli) che è disperso nella colonna d'acqua. Ciò significa che i valori di biomassa ricavati esclusivamente dai banchi risultano confrontabili con le stime tradizionali. Stime di biomassa annuali potrebbero quindi basarsi esclusivamente sul PRC_NASC estratto da banchi, isolati acusticamente utilizzando l'algoritmo a multi-frequenza adottato per questo studio. La distribuzione della biomassa per le varie specie (prodotta dai tre modelli predittivi) sembra consistente con il NASC, stimato per acciughe, OPS e sardine con la metodologia tradizionale solo in un settore dell'area del survey, in cui le stime tradizionali indicano una minore abbondanza di acciughe in termini di biomassa. Un'analisi più approfondita è però necessaria per spiegare questo risultato e individuare quale modello (dei due adoperati) abbia maggiori capacità di discriminare correttamente tra le specie.

Ai modelli classificativi si è deciso infine di associare anche valori medi di temperatura e salinità estratti dalle cale stazioni CTD più vicine alle cale monospecifiche, per comprendere quale fosse l'effetto delle variabili ambientali sulla capacità di discriminare banchi di specie diverse. Ci sono infatti numerosi esempi in letteratura che associano alle acciughe in fase di spawning (durante il periodo estivo) certi range di temperatura (Richardson *et al.*, 1998; Basilone *et al.*, 2006) o alle aringhe certi gradienti di salinità (Maravelias & Reid, 1995). I risultati con le variabili della colonna d'acqua tuttavia ci suggeriscono che modelli di questo tipo potrebbero essere indicativi di habitat differenti tra le specie ma non essere d'aiuto in un processo di identificazione a scopi classificativi. La scarsa variabilità del dataset ambientale però ci spinge ad analizzare un dataset più ampio e dotato di maggiore contrasto prima di effettuare una qualunque ipotesi sulla possibilità di utilizzare i parametri della colonna d'acqua a scopi classificativi. Diverso invece potrebbe essere il contributo fornito dalla durezza del fondo, per il quale è stato possibile estrarre acusticamente un valore medio ogni 0.05 nmi. Risultati preliminari indicano che l'utilizzo di questo parametro potrebbe aumentare la capacità di discriminazione totale e per specie dei modelli classificativi. Il MaxRate estratto da trenta diversi alberi indica infatti un effetto della durezza del fondo a minori profondità (inferiori a 68 m) dove quest'ultimo potrebbe intervenire nella discriminazione di banchi di acciughe e sardine. Questi risultati sono a supporto del lavoro svolto da D'Elia *et al.* (2009) in due zone (zona 1 e 2) del settore Nord-occidentale della costa meridionale siciliana dove un comportamento differente dei banchi in relazione al tipo di fondo è stato evidenziato in zona 2, laddove catture sperimentali mostravano una dominanza di sardine e acciughe. Relazioni tra specie pelagiche e caratteristiche sedimentologiche e/o morfologiche del fondale sono state osservate anche per le aringhe del Mar del Nord (*Clupea harengus*) da Maravelias (1999) and Maravelias *et al.* (2000). Aggregazioni più consistenti di aringhe, analizzate prima della fase di spawning, sono state individuate all'interno di habitat caratterizzati da profondità comprese tra i 100 e i 150 m, su substrato di tipo ghiaioso/sabbioso con alte concentrazioni di zooplankton. Non è

quindi da escludere che nell'ipotesi di una preferenza delle specie target per un certo tipo di fondo, questa preferenza non sia connessa ad una maggiore disponibilità di cibo o alla presenza di cibo preferenziale per le specie.

Modelli decisionali ad albero proposti in questo studio come metodi di identificazione delle eco tracce si sono dimostrati strumenti flessibili e facilmente interpretabili. Le variabili predittive non richiedono alcuna trasformazione ed alcun trattamento dei valori mancanti o degli outliers, in quanto queste osservazioni vengono isolate, identificano relazioni lineari e possono essere utilizzate sia a scopi classificativi che descrittivi. Tra l'altro possono rappresentare un ottimo strumento nel processo di interpretazione di un ecogramma essendo modelli di questo tipo incorporabili all'interno di software commerciali come Echoview utilizzati nel campo dell'acustica applicata alla pesca. I modelli MaxRate o MaxRatespecies potrebbero essere costruiti in ambiente Echoview utilizzando il modulo virtual-variable al fine di ottenere una classificazione in real-time delle eco tracce. Abbiamo comunque riscontrato uno svantaggio nell'applicazione di tali modelli. Essi sono infatti molto instabili (Prasad *et al.*, 2006). Ciò significa che variando il set di osservazioni utilizzate per la loro costruzione o il set di variabili esplicative, si possono ottenere anche modelli molto differenti (in termini di predittori selezionati e regole adoperate per la suddivisione), sebbene le capacità classificative possano essere simili (Scull *et al.*, 2005). L'instabilità di alberi decisionali tuttavia può essere gestita tenendo in considerazione la variabilità del dataset fornito in input al modello. Ciò si può effettuare sviluppando ripetutamente (in genere 30-80 volte) un certo numero di modelli, basandosi su training set differenti e mediando i risultati ottenuti (come da noi effettuato con il modello "Mode").

Capitolo 5

Conclusioni

I principali risultati ottenuti in questo studio sono di seguito sintetizzati:

- Un comportamento differente è stato evidenziato in due zone del Canale di Sicilia, (zona 1 e zona 3) in relazione alla biomassa e/o alla densità planctonica. Il differente comportamento è stato attribuito a specie dominanti diverse, sulla base delle catture sperimentali effettuate nelle due zone e delle caratteristiche dei banchi (morfologiche, batimetriche ed energetiche ad una frequenza). Le relazioni con il plancton ci hanno quindi consentito di formulare delle ipotesi sulla composizione in specie in specifiche aree del survey;
- Una discriminazione tra le specie appare possibile utilizzando contemporaneamente variabili morfologiche, batimetriche ed energetiche a più frequenze in modelli decisionali ad albero. L'insieme di queste variabili consente di ottenere buone capacità di discriminare banchi di sardine e OPS, ma ancora discrete capacità di classificare banchi di acciughe;
- La profondità del banco (correlata fortemente a quella del fondo) è la variabile più importante nella discriminazione tra le specie e ciò riflette la preferenza di banchi di specie diverse per batimetrie medie differenti;
- Il contributo alla classificazione fornito dalle sole variabili energetiche e morfologiche è molto simile per le sardine, mentre le acciughe sono meglio classificate dalle sole variabili morfologiche rispetto alle energetiche. In generale comunque le variabili energetiche o morfologiche forniscono da sole una minore capacità di classificazione;
- Parametri strutturali ed energetici però risultano piuttosto variabili all'interno di una stessa specie e ciò produce instabilità nei modelli decisionali ad albero, che può essere gestita effettuando ripetutamente un certo numero di modelli su training set e test set differenti ed estraendo i modelli più congeniali;
- Il valore totale di PRC_NASC estratto dai banchi lungo tutto il survey anche-va2009 è consistente con il NASC totale stimato dalla metodologia di cui si fa uso tradizionalmente per le stime di biomassa;

- Tre modelli predittivi differenti sono stati stimati per classificare la specie in banchi estratti dai tracciati acustici del survey Ancheva2009. La biomassa attribuita ad ogni specie dai modelli risulta in accordo con le stime tradizionali principalmente nel settore che segue Gela e precede Capo Passero, laddove si osservano livelli di biomassa per le acciughe più bassi;
- L'utilizzo dei parametri della colonna d'acqua nei modelli sembra non essere d'aiuto nel processo di identificazione delle specie, contrariamente alla durezza del fondo per la quale si ritiene utile effettuare analisi più dettagliate.

Bibliografia

Agenbag J.J., Richardson A.J., Demarcq H., Fréon P., Weeks S. and Shillington F.A. 2003. Estimating environmental preferences of South African pelagic fish species using catch size- and remote sensing data. *Progress in Oceanography* 59(2-3): 275-300.

Aguzzi L., Carini V., Cuttitta A., Cavalcante C., Mazzola S., Basilone G., Buscaino G., Patti B., and Bonanno A., 2003. Primi dati sulla distribuzione e abbondanza del popolamento zooplanctonico nel Canale di Sicilia. *Biol. Mar. Medit.* 10 (2), pp. 975–979.

Alvares F., Cingolati N., Kallianiotis A., 2003. Piccoli pelagici: gestione futura delle risorse condivise in Mediterraneo. *Il Pesce* N° 5, 57-65.

Ayon P., Swartzman G, Bertrand A., Gutiérrez M., and Bertrand S. 2008. Zooplankton and foragefish species off Peru: Large-scale bottom-up forcing and local-scale depletion. *Prog. Oceanogr.* 79, pp. 208–214.

Basilone G., Guisande C., Patti B., Mazzola S., Cuttitta A., Bonanno A., Vergara A. R. & Maneiro I., 2006. Effect of habitat conditions on reproduction of the European anchovy (*Engraulis encrasicolus*) in the Strait of Sicily. *Fisheries Oceanography* 15: 271–280.

Barange M., Coetzee J. C., and Twatwa N. M. 2005. Strategies of space occupation in anchovy and sardine in the southern Benguela: the role of stock size and intra-species competition. *ICES Journal of Marine Science*, 62: 645–654.

Bertrand A., Barbieri M.A., Cordova J., Hernandez C., Gomez F. and Leiva F., 2004. Diel vertical behaviour, predator–prey relationships, and occupation of space by jack mackerel (*Trachurus murphyi*) off Chile. *ICES J. Mar. Sci.* 61 , pp. 1105–1112.

Bertrand A., Segura M., Gutiérrez M., and Vásquez, L. 2004. From small-scale habitat loopholes to decadal cycles: a habitat-based hypothesis explaining fluctuation in pelagic fish populations off Peru. *Fish and Fisheries*, 5: 296–316.

Bonanno A., Goncharov S., Mazzola S., Popov S., Cuttitta A., Patti B., Basilone G., Di Nieri A., Patti C., Aronica S., Buscaino G., 2006. Acoustic evaluation of anchovy

- larvae distribution in relation to oceanography in the Cape Passero area (Strait of Sicily). *Chemistry and Ecology*, 22 (1), S265-S273.
- Brieman F., Friedman J.H., Losen R.A., Stone C.J., 1984. *Classification and Regression Trees*. Wadsworth and Brooks/Cole, Monterrey. p. 172.
- Brierley A.S., Ward P., Watkins J.L., and Goss, C. 1998. Acoustic discrimination of Southern Ocean zooplankton. *Deep Sea Research II* 45: 1155-1173.
- Cabreira A. G., Tripode M., and Madirolas A. 2009. Artificial neural networks for fish-species identification. *ICES Journal of Marine Science*, 66: 1119–1129.
- Castillo J., Barbieri M. A., and Gonzalez A. 1996. Relationships between sea surface temperature, salinity, and pelagic fish distribution off northern Chile. – *ICES Journal of Marine Science*, 53: 139–146.
- Chivers R.C., Emerson N. and Burns D.R. 1990. “New acoustic processing for underway surveying.” *Hydro. J.* 56, 9-17.
- D’Elia M., Patti B., Sulli A., Tranchida G., Bonanno A., Basilone G., Giacalone G., Fontana I., Genovese S., Guisande C., and Mazzola S. 2009. Distribution and spatial structure of pelagic fish schools in relation to the nature of the seabed in the Sicily Channel (Central Mediterranean), *Mar. Ecol.* 30 , pp. 151–160.
- Diner N. 2001. Correction on school geometry and density: approach based on acoustic image simulation. *Aquat. Living Resour.* 14: 211-222.
- Eggers D. M. 1976. Theoretical effect of schooling by planktivorous fish predators on rate of prey consumption. *Journal of the Fisheries Research Board of Canada* 33, 1964–1971.
- Garcia S.M., Zerbi A., Aliaume C., Do Chi T., Lasserre G. The ecosystem approach to fisheries. Issues, terminology, principles, institutional foundations, implementation and outlook. *FAO Fisheries Technical Paper*. No. 443. Rome, FAO. 2003. 71 p.
- Fässler S. M. M., and Gorska N. 2009. On the target strength of Baltic clupeids. *ICES Journal of Marine Science*, 66: 1184–1190.
- Fässler S. M. M., Brierley A. S., and Fernandes P. G. 2009. A Bayesian approach to estimating target strength. *ICES Journal of Marine Science*, 66: 1197-1204.
- Fässler S.M.M., Santos R., García-Núñez N., Fernandes P.G., 2007. Multifrequency back-scattering properties of Atlantic herring (*Clupea harengus*) and Norway pout (*Trisopterus*

esmarkii). Can. J. Fish. Aquat. Sci. 64, 362–374.

Fernandes P. G. 2009. Classification trees for species identification of fish-school echotraces. ICES Journal of Marine Science, 66: 1073–1080.

Fernandes, P. G., Korneliussen, R. J., Lebourges-Dhaussy, A., Masse, J., Iglesias, M., Diner, N., Ona, E., et al. 2006. The SIMFAMI Project: Species Identification Methods from Acoustic Multifrequency Information. Final Report to the EC, Q5RS-2001–02054.

Fleischman S. J., and Burwen D. L. 2003. Mixture models for the species apportionment of hydroacoustic data, with echo-envelope length as the discriminatory variable. ICES Journal of Marine Science, 60: 592–598.

Foote K. G., Knudsen H. P., Vestnes G., MacLennan D. N., and Simmonds E. J., 1987. Calibration of acoustic instruments for fish density estimation: a practical guide. ICES cooperative research report, 144, 69 pp.

Franklin, J., 1995. Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. Prog. Phys. Geography 19, 474-499.

Fréon P., and Misund O. A. 1999. Dynamics of Pelagic Fish Distribution and Behaviour: Effects on Fisheries and Stock Assessment. Fishing News Books, London. 348 pp.

García Lafuente J., García A., Mazzola S., Quintanilla L., Delgado J., Cuttita A., Patti B. 2002. Hydrographic phenomena influencing early life stages of the Sicilian Channel anchovy. Fish. Oceanogr. 11:1, 31-44.

Giannoulaki M., Machias A., Somarakis S., and Tsimenides N. 2005. The spatial distribution of anchovy and sardine in the northern Aegean Sea in relation to hydrographic regimes. Belgian Journal of Zoology, 135: 151e156.

Greenstreet S.P.R., Tuck I.D., Grewar G.N., Armstrong E., Reid D.G. and Wright P.J.. 1997. "An assessment of the acoustic survey technique, RoxAnn, as a means of mapping seabed habitat." ICES J. Mar. Sci. 54, 939-959.

Greenstreet S.P.R., McMillan J.A. and Armstrong F. 1998. Seasonal variation in the importance of pelagic fish in the diet of piscivorous fish in the Moray Firth, NE Scotland: a response to variation in prey abundance?. ICES J. Mar. Sci. 55: 121-133.

Guisande A., Zimmermann N., 2000. On the use of static distribution models in ecology. Ecol. Model. 135, 147-186.

Gutiérrez M., Swartzman G., Bertrand A., and Bertrand S. 2007. Anchovy and sardine spatial dynamics and aggregation patterns in the Humboldt Current ecosystem, Peru, from

1983–2003. *Fisheries Oceanography*, 16: 155–168.

Hamilton L.J. 2001. "Acoustic seabed classification systems." DSTO Technical Note DSTO-TN-0401. 66pp. <http://www.dsto.defence.gov.au/corporate/reports/DSTO-TN-0401.pdf>].

Hamilton L.J., Mulhearn P.J. and Poeckert R. 1999. "Comparison of RoxAnn and QTC-View acoustic bottom classification system performance for the Cairns area, Great Barrier Reef, Australia." *Cont. Shelf. Res.* 19, 1577-1597.

Hastie T. and Tibshirani R., *Generalised Additive Models*, Chapman and Hall, London, 1990.

Heald G.J. and Pace N.G.. 1996. "An analysis of the 1st and 2nd backscatter for seabed classification." *Proc. 3rd European Conference on Underwater Acoustics*, 24-28 June 1996 vol. II, 649-654.

Hewitt R.P. and Demer D.A. 1991. Krill abundance. *Nature* 353: 310.

Hooper R., and Lucero A. 1976. Radar profile classification: a feasibility study. Technical Report Technology Service Corporation, Santa Monica, CA.

Horne, J. 2000. Acoustic approaches to remote species identification: a review. *Fisheries Oceanography*, 9: 356–371.

ICES 2000. Report of the Working Group on Fisheries Acoustics Science and Technology (WG-FAST), Ijmuiden/Haarlem, Netherlands, 10-14 April 2000. International Council for the Exploration of the Sea, Copenhagen (Denmark). ICES CM 2000/B:04, 35 pp.

ICES 2004g. Report of the planning group on Northeast Atlantic Pelagic Ecosystem Surveys (PGNAPES). ICES CM 2004g/D:07. 57 pp.

IREPA, 2010. Osservatorio economico sulle strutture produttive della pesca marittima in Italia 2008. Napoli: Edizioni Scientifiche Italiane, pp. 184.

Kaiser M.J., P.J. Armstrong P.J. Dare and R.P. Flatt. 1998. "Benthic communities associated with a heavily fished scallop ground in the English Channel." *J. Mar. Biol.* 78(4), 1045-1059.

Kang M., Furusawa M. and Miyashita K. 2002. Effective and accurate use of difference in mean volume backscattering strength to identify fish and plankton. *ICES Journal of Marine Science*, 59: 794-804.

- Kloser R., Ryan T., Sakov P., Williams A. and Koslow J.A. 2002. Species identification in deep water using multiple acoustic frequencies. *Can. J. Fish. Aquat. Sci.* 59: 1065-1077.
- Kloser R.J., Bax N.J., Ryan T., Williams A. and Baker B.A. 2001. Remote sensing of seabed types in the Australian South East Fishery – development and application of normal incident acoustic techniques and associated ground truthing. *Journal of Marine and Freshwater Research* 52: 475-489.
- Korneliussen R. J., N. Diner E. Ona L. Berger and P. G. Fernandes. 2008. Proposals for the collection of multifrequency acoustic data. *ICES Journal of Marine Science* 65:982-994.
- Korneliussen R. J. 2010. The acoustic identification of Atlantic mackerel. – *ICES Journal of Marine Science*, 67: 000–000.
- Lavery A. C., Wiebe P. H., Stanton T. K., Lawson G. L., Benfield M. C., and Copley N. 2007. Determining dominant scatterers of sound in mixed zooplankton populations. *Journal of the Acoustical Society of America*, 122: 3304–3326.
- Lawson G.L., Barange M. and Fréon 2001. Species identification of pelagic fish schools on the South African continental shelf using acoustic descriptors and ancillary information. *ICES J. Mar. Sci.* 58: 275-287.
- Lebourges-Dhaussy A., Coetzee J., Hutchings L., Roudaut G., and Nieuwenhuys C. 2009. Zooplankton spatial distribution along the South African coast studied by multifrequency acoustics, and its relationships with environmental parameters and anchovy distribution. – *ICES Journal of Marine Science*, 66: 1055–1062.
- Lundgren B., and Nielsen J. R. 2008. A method for the possible species discrimination of juvenile gadoids by broad-bandwidth backscattering spectra vs. angle of incidence. *ICES Journal of Marine Science*, 65: 581–593.
- MacLennan D.N., and Holliday D.V. 1996. Fisheries and plankton acoustics: past, present, and future. *ICES J. Mar. Sci.* 53: 513-516.
- Madureira L.S.P., Everson I. and Murphy E.J. 1993. Interpretation of acoustic data at two frequencies to discriminate between Antarctic krill (*Euphausia superba* Dana) and other scatterers. *Journal of Plankton Research* 15(7): 787-802.
- Magorrian, B.H., M. Service and W. Clarke. 1995. “An acoustic bottom classification survey of Strangford Lough, Northern Ireland.” *J. Mar. Biol. Ass. UK.* 75,987-992. *Mar. Ecol.* 30 (2009), pp. 151–160.
- Maravelias C.D. (1999) Habitat selection and clustering of pelagic fish: effects of topography and bathymetry on species dynamics. *Can J Fish Aquat Sci* 56:437–450.

Maravelias C.D., Reid D.G., and Swartzman G. 2000. Seabed substrate, water depth and zooplankton as determinants of the prespawning spatial aggregation of North Atlantic herring, *Mar. Ecol. Prog. Ser.* 195 , pp. 249–259.

Maravelias C. D. and Reid D. G. 1995. Relationship between herring (*Clupea harengus L.*) distribution and sea surface salinity and temperature in the northern North Sea. *Scientia Marina*, 59: 427–438.

Massé J. and Rouxel C. 1991. Improvement in acoustic assessments by discrimination of pelagic shoals with INES/MOVIES system. *ICES CM 1991/B:26*. 8 pp.

Massé J., Koutsikopoulos C. and Patty W. 1996. The structure and spatial distribution of pelagic fish schools in multispecies clusters: an acoustic study. *ICES Journal of Marine Science* 53(2): 155-160.

Misund O. A., A. Ferno T. Pitcher and B. Totland. 1998. Tracking herring schools with a high resolution sonar. Variations in horizontal area and relative echo intensity. *ICES J. Mar. Sci.* 55:58-66 [doi:10.1006/jmsc.1997.0228].

Nero R.W., and Magnuson J.J. 1989. Characterization of patches along transects using high-resolution 70-kHz integrated acoustic data. *Canadian Journal of Fisheries and Aquatic Sciences*, 46: 2056-2064.

Nero R.W., Thompson C.H. and Jech J.M. 2004. In situ acoustic estimates of the swim-bladder volume of Atlantic herring (*Clupea harengus*). *ICES J. Mar. Sci.* 61: 323-337.

Nottestad L., Aksland M., Beltestad A., Ferno A., Johannessen A., Misund O.A. 1996. Schooling dynamics of Norwegian spring spawning herring (*Clupea harengus L.*) in a coastal spawning area. *Sarsia* 80, 277–284.

Ona E. 1990. Physiological factors causing natural variations in acoustic target strength of fish. *J. mar. biol. Ass. U.K.* 70: 107-127.

Orlowski A. 1984. “Application of multiple echoes energy measurements for evaluation of sea bottom type.” *Oceanologia* 19, 61-78.

Patti B., A. Bonanno, S. Mazzola, G. Basilone, G. Buscaino, A. Cuttitta, S. Gontcharov, S. Aronica, G. Giacalone, I. Fontana, D. Tegolo 2006. Distribution of fish schools and plankton patches along the southern coast of Sicily in relation to hydrology. 4-8 Settembre 2006 41° European Marine Biology Symposium-EMBS, Cork.

Patti B., Bonanno A., Basilone G., Goncharov S., Mazzola S., Buscaino G., Cuttitta A., García Lafuente J., García A., Palombo G., Cosimi G., 2004. Interannual fluctuations in acoustic biomass estimates and in landings of small pelagic fish populations in relation to hydrology in the Strait of Sicily. *Chem. Ecol.*, 20, 365-375.

Patti B., Bonanno A., Basilone G., Quinci E.M., Mazzola S., 2009. Anchovy, Sardine, *Sardina pilchardus*- GSA16 (South of Sicily) presented at General Fisheries Commission for the Mediterranean, Sub-Committee On Stock Assessment (SCSA), Report of the SCSA Working Group on Stock Assessment of Small Pelagic Species, Ancona, Italy 26-30 October 2009.

Patti B. , Bonanno A. , D'Elia M. , Quinci E. , Giacalone G. , Fontana I. , Aronica S. , Basilone G. and Mazzola S. 2011. Daytime pelagic schooling behaviour and relationships with plankton patch distribution in the Sicily Strait (Mediterranean Sea)', *Advances in Oceanography and Limnology*, 2: 1, 79 — 92.

Prager, B.T., D.A. Caughey and R.H. Poeckert. 1995. "Bottom classification: operational results from QTC view." OCEANS '95 - Challenges of our changing global environment conference, October 1995, San Diego, California, USA.

Prasad, A.M., Iverson, L.R.& Liaw, A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, 9, 181–199.

R. Development Core Team, R., A Language and Environment for Statistical Computing, Vienna, Austria: R Foundation for Statistical Computing, 2005; software available at [http:// www.R-project.org](http://www.R-project.org).

Reid D., Scalabrin C., Petitgas P., Massé J., Aukland R., Carrera P., and Georgakarakos S. 2000. Standard protocols for the analysis of school based data from echo sounder surveys. *Fisheries Research*, 47: 125–136.

Reiss C. S., Cossio A. M., Loeb V., and Demer D. A. 2008. Variations in the biomass of Antarctic krill (*Euphausia superba*) around the South Shetland Islands, 1996–2006. *ICES Journal of Marine Science*, 65: 497–508.

Richards L. J., Kieser R., Mulligan T. J., and Candy, J. R. 1991. Classification of fish assemblages based on echo integration surveys. *Canadian Journal of Fisheries and Aquatic Sciences*, 48: 1264–1272.

Richardson AJ, Mitchell-Innes BA, Fowler JL, Bloomer SF, Verheye HM, Field JG, Hutchings L. and SJ Painting 1998 – The effect of sea temperature and food availability on the spawning success of Cape anchovy *Engraulis capensis* in the southern Benguela. In *Benguela Dynamics: Impacts of Variability on Shelf-Sea Environments and their Living Resources* . Pillar, SC, Moloney, CL, Payne, AIL & FA Shillington (Eds). South African

Journal of Marine Science 19 : 275-290.

Robotham H., Castillob P., Boscha J., Perez-Kallensa 2011. A comparison of multi-class support vector machine and classification tree methods for hydroacoustic classification of fish-schools in Chile. *Fisheries Research* 111(3):170-176.

Rose G.A. 1995. Cod spawning on a migration highway in the north-west Atlantic. *Nature* 366: 458-461.

Rose G.A. and Leggett W.C. 1988. Hydroacoustical signal classification of fish schools by species. *Can. J. Fish. Aquat. Sci* 45: 597-607.

Sainsbury K.J., Punt A.E. and Smith A.D.M. 2000. Design of operational management strategies for achieving fishery ecosystem objectives. *ICES J. Mar. Sci.* 57: 731-741.

Scalabrin C. and Massé J. 1993. Acoustic detection of the spatial and temporal distribution of fish shoals in the Bay of Biscay. *Aquatic Living Resources*, 6: 269–283.

Scalabrin C., Diner N., Weill A., Hillion A. and Mouchot M.C. 1996. Narrow band acoustic identification of monospecific fish shoals. *ICES J. Mar. Sci.* 53: 181-188.

Schlagintweit G.E.O. 1993. Real-time acoustic bottom classification: a field evaluation of RoxAnn. *Proceedings of Oceans '93*: 214-219. schools in relation to the nature of the seabed in the Sicily Channel (Central Mediterranean).

Scull P., Franklin J., Chadwick O.A., 2005. The application of classification tree analysis to soil type prediction in a desert landscape. *Ecol. Model.* 181:1–15.

Serpetti N., Heath M. R., Armstrong E. & Witte U., 2011. Blending single beam RoxAnn and multi-beam swathe QTC hydro-acoustic discrimination techniques for the Stonehaven area, Scotland, UK. *Journal of Sea Research* 65: 442–455.

Simmonds E. J., and MacLennan D. N. 2005. *Fisheries Acoustics: Theory and Practice*, 2nd edn. Blackwell Publishing, Oxford.

Simmonds E.J. 2003. Weighting of acoustic- and trawl-survey indices for the assessment of North Sea herring. *ICES J. Mar. Sci.* 60: 463-471.

Simmonds E.J., Armstrong F. and Copland P.J. 1996. Species identification using wideband backscatter with neural network and discriminant analysis. *ICES J. Mar. Sci.* 53: 189-195.

Siwabessy P.J.W. 2001. An investigation of the relationship between seabed type and benthic and benthopelagic biota using acoustic techniques. PhD Thesis, Curtin University

of Technology, Perth, Western Australia.

Siwabessy P.J.W., J.D. Penrose R.J. Kloser and D.R. Fox. 1999. "Seabed habitat classification." Proc. International Conference on High Resolution Surveys in Shallow Waters DSTO, 18-20 October 1999, Sydney, Australia.

Sorensen P.S., Madsen K.N., Nielsen A.A., Schultz N., Conradsen K. and Oskarsson O. 1998. "Mapping of the benthic communities common mussel and neptune grass by use of hydroacoustic measurements." Proc. 3rd European Marine Science and Technology Conference, 26 May 1998, Lisbon, Portugal.

Soria M., Gerlotto F., and Fréon P. 1993. Study of learning capabilities of tropical clupeoids using an artificial stimulus.

Stokesbury K.D.E., Kirsch J., Patrick E.V. and Norcross B.L. 2002. Natural mortality estimates of juvenile Pacific herring (*Clupea pallasii*) in Prince William Sound, Alaska. Can. J. Fish. Aquat. Sci. 59(3): 416-423.

Swartzman G., Stuetzle W., Kulman K., and Powojowski M. 1994. Relating the distribution of pollock schools in the Bering Sea to environmental factors. ICES Journal of Marine Science, 51: 481-492.

Swartzman G., 1997. Analysis of the summer distribution of fish schools in the Pacific Eastern Boundary Current. ICES J. Mar. Sci. 54, 105-116.

Swartzman G., Brodeur R., Napp J., Hunt G., Demer D., and Hewitt R. 1999. Spatial proximity of age-0 walleye pollock (*Theragra chalcogramma*) to zooplankton near the Pribilof Islands, Bering Sea, Alaska. ICES Journal of Marine Science, 56:545-560.

Swartzman G., Napp J., Brodeur R., Winter A., and Ciannelli L. 2002. Spatial patterns of pollock and zooplankton distribution in the Pribilof Islands, Alaska nursery area and their relationship to pollock recruitment, ICES J. Mar. Sci. 59 , pp. 1167-1186.

Swartzman G., Bertrand A., Gutiérrez M., Bertrand S., Vasquez L., 2008. The relationship of anchovy and sardine to water masses in the Peruvian Humboldt Current System from 1983-2005. Progress in Oceanography 79, 228-237.

Tegowski J., Gorska N., and Klusek Z. 2003. Statistical analysis of acoustic echoes from underwater meadows in the eutrophic Puck Bay (southern Baltic Sea). Aquatic Living Resources, 16: 215-221.

Tugores M. P., Iglesias M., Díaz N., Oñate D., Miquel J., and Giráldez A. 2010. Latitudinal and interannual distribution of the European anchovy (*Engraulis encrasicolus*) and sardine

(*Sardina pilchardus*) in the western Mediterranean, and sampling uncertainty in abundance estimates. ICES Journal of Marine Science, 67: 1574–1586.

Van der Lingen C. D. 1994. Effect of particle size and concentration on the feeding behaviour of adult *pilchard* *Sardinops sagax*. Marine Ecology Progress Series, 109:1–13.

Von Szalay P. G., & McConnaughey R. A. 2002. The effect of slope and vessel speed on the performance of a single beam acoustic seabed classification system. Fisheries Research, 54: 181–194.

Zwolinski J., Morais A., Marques V., Stratoudakis Y., and Fernandes P.G. 2007. Diel variation in the vertical distribution and schooling behaviour of sardine (*Sardina pilchardus*) off Portugal, ICES J. Mar. Sci. 64 , pp. 963–972.

Zwolinski J. P., Oliveira P. B., Quintino V., and Stratoudakis Y. 2010. Sardine potential habitat and environmental forcing off western Portugal. ICES Journal of Marine Science, 67: 1553–1564.

Walsh M., Reid D. G., and Turrell W. R. 1995. Understanding mackerel migration off Scotland: tracking with echosounders and commercial data, and including environmental correlates and behaviour. ICES Journal of Marine Science, 52: 925–939.

Watkins J.L. and Brierley A.S. 2002. Verification of the acoustic techniques used to identify Antarctic krill. ICES J. Mar. Sci. 59: 1326-1336.

Weill A., Scalabrin C. and Diner N. 1993. MOVIES-B: An acoustic detection description software. Application to shoal species' classification. Journal of Aquatic Living Resources 6(3): 255-267.

Winter A.G. and Swartzman G.L. 2006. Interannual changes in distribution of age-0 walleye Pollock near the Pribilof Islands, Alaska, with reference to the prediction of pollock year-class strength. ICES J. Mar. Sci. 63 , pp. 1118–1135.

Ringraziamenti

A conclusione di questo lavoro di dottorato sento di esprimere i miei più sinceri ringraziamenti a tutte le persone che ho avuto il piacere di incontrare e con cui ho avuto l'onore di lavorare in questi tre anni di esperienza nella ricerca.

Ringrazio innanzitutto i miei due tutor, il Dott. Bernardo Patti, per avermi introdotto nell'ambiente scientifico dell'IAMC-CNR e per l'opportunità datami, e il Dott. Stefano Malavasi per il suo appoggio e la sua presenza durante le varie sessioni di esami, nonché per la sua gentilezza e disponibilità.

Ringrazio inoltre enormemente il mio co-tutor, il Dott. Angelo Bonanno, per le risposte ai miei dubbi e l'aiuto materiale e morale, senza il quale questo lavoro non sarebbe mai stato portato a compimento adesso. Sono debitrice anche al "prof." Marco Barra, con il quale è stato un piacere ragionare al telefono alcuni risultati della mia ricerca.

Ringrazio ancora il Dott. Mario Sprovieri per avermi messa a conoscenza dei programmi di mobilità all'estero del CNR e il Dott. Salvatore Mazzola per avermi dato l'opportunità di partecipare al programma che mi ha consentito di effettuare l'esperienza presso il laboratorio di ricerca di Aberdeen.

E ancora mi sento di ringraziare ricercatori e amici del Marine Laboratory di Aberdeen: il Dott. Paul Fernandes innanzitutto per i suoi insegnamenti sull'acustica e la sua disponibilità e pazienza alle mie mille domande, Susan Lusseau con la quale ho avuto il piacere di confrontare l'elaborazione dei dati acustici e Andrzej Jaworski "mon amic", per i suoi grandi insegnamenti di vita, la sua gentilezza e la sua passione per R, che ha reso meno traumatico il mio approccio ad un linguaggio di programmazione, nonché per avermi spinto a partecipare alla "Ceilidh Nigh" del Marine Lab. Un grazie anche alle persone che mi hanno fatto compagnia nelle serate fredde di Aberdeen, Maria e Peter, Natalia e Douglas, Savio e Viviane e a tutti i ragazzi portoghesi e spagnoli che hanno condiviso con me l'ora di pranzo nella mensa.

Un ringraziamento sincero va anche al Dott. Heald che con pazienza e cortesia mi ha aiutato a comprendere la teoria degli echi provenienti dal fondo, fornendo risposte a miei dubbi e bibliografia su cui lavorare.

Un ringraziamento particolare a tutti i colleghi dell'IAMC, con i quali ho condiviso le lunghe ore lavorative e i viaggi sullo scudo e a tutti gli amici e colleghi dell' "Oppel Splain" per avermi aiutato a deconcentrarmi dal lavoro e a rilassare i nervi nei momenti più difficili. Un grazie infinito anche ai miei due pazienti "compagni" di lavoro, Giovanni e Ignazio.

Infine ringrazio la mia famiglia per il supporto morale ed economico e per il sostegno alle mie scelte, e a "me frate" che mi ha insegnato ed aiutato a scrivere la tesi, vista la mia assoluta mancanza di esperienza con LATEX.

A tutti gli altri amici, italiani, polacchi, greci, turchi, inglesi e americani, le persone che ho conosciuto un pò sparse per il mondo nei tre anni di ricerca e a tutti quelli con i quali ho passato dei momenti estremamente piacevoli, grazie di cuore infinitamente.

Palermo, 31-01-2012

Marta D'Elia

Estratto per riassunto della tesi di dottorato

Studente: Marta D'Elia

Matricola: 955636

Dottorato: Scienze Ambientali

Ciclo: XXIV

Titolo della tesi: Caratterizzazione e identificazione di banchi di specie pelagiche mediante l'uso di descrittori strutturali e parametri ambientali

Sommario: Le metodologie acustiche sono tecniche molto efficienti per stimare l'abbondanza delle risorse pelagiche. Tuttavia l'identificazione dei target acustici richiede ancora un processo caratterizzato da una certa soggettività perché basato sull'esperienza di chi interpreta gli ecogrammi e sulle informazioni biologiche. In questo lavoro diverse procedure sono state applicate ai dati acquisiti durante i survey acustici condotti nel Canale di Sicilia per identificare le specie più abbondanti (*E. encrasicolus*, *S. pilchardus*) e rendere la metodologia più oggettiva. I risultati indicano che una discriminazione tra banchi appartenenti a differenti specie appare possibile utilizzando nei modelli decisionali variabili morfologiche, batimetriche ed energetiche del banco, relative a più frequenze. Il metodo descritto ha consentito inoltre di acquisire informazioni sugli habitat preferenziali delle due principali specie target.

Abstract: Acoustic techniques are efficient tools for the abundance estimation of many pelagic fish species. However the technique still requires some subjective judgment when allocating the acoustic data, fish-school echotraces, to particular species. This is assisted by trawl data and operator experience in the scrutiny of echograms. In this work different procedures are applied to acoustic data, acquired during survey carried out in the Sicilian Channel, to the aim of identify the most abundant species (*E. encrasicolus*, *S. pilchardus*) and create a tool for objective species allocation. The results indicate that a discrimination between the two species seems possible using a set of morphological, bathymetric and energetic school descriptors with a classification tree approach. The method described here was also able to get important information on the preferential habitat of the two main species.

Firma dello studente