

UNIVERSITÀ DEGLI STUDI DI GENOVA

DIPARTIMENTO DI SCIENZE DELLA TERRA,
DELL'AMBIENTE E DELLA VITA (DISTAV)

Classe delle Lauree Magistrali in Biologia LM/6

Corso di Laurea Magistrale in

Biologia ed Ecologia Marina

curriculum

Valutazione e Gestione dell'Ambiente Marino

CLASSIFICAZIONE SUPERVISIONATA OBJECT-BASED DI DATI
MULTIBEAM PER LA CARTOGRAFIA BIONOMICA AD ALTA
RISOLUZIONE DEGLI HABITAT MARINI

Laureando:

Luca Di Vita

Relatori:

Prof.ssa Monica Montefalcone

Dott.ssa Alice Oprandi

Dott. Sante Francesco Rende

Correlatore:

Prof. Francesco Massa

ANNO ACCADEMICO 2024-2025

RIASSUNTO

La cartografia bionomica degli habitat bentonici rappresenta uno strumento fondamentale sia per la ricerca, nell'ambito dell'ecologia del paesaggio marino (seascape ecology), sia per la gestione dell'ambiente marino. Le praterie di *Posidonia oceanica* rappresentano un habitat prioritario (Direttiva Habitat 92/43/CEE), per la cui tutela è imposta agli stati membri dell'Unione Europea la mappatura e il monitoraggio nel tempo. Il presente studio si inserisce nella cornice della campagna oceanografica del progetto PNRR MER (Marine Ecosystem Restoration), in cui è prevista la mappatura di dettaglio delle praterie di fanerogame marine a scala nazionale attraverso l'acquisizione di dati acustici tramite ecoscandaglio multibeam (MBES) con risoluzione nativa sub-metrica (0,2 m), integrata da altre tecniche di rilevamento dei fondali. Il presente studio adotta l'approccio GEOBIA (*Geographic Object-Based Image Analysis*) per classificare il dato acquisito con multibeam al fine di produrre una carta bionomica ad alta risoluzione. Nello specifico, si è mirato a ottimizzare un modello di classificazione supervisionata assumendo l'area di Levanto come sito pilota, con il duplice scopo di massimizzare l'accuratezza predittiva locale e di incrementare la risoluzione tematica, introducendo e isolando la classe della *matte* morta. La classificazione si è basata sui dati acustici del MBES, backscatter e batimetrica, e sulle derivate morfometriche estratte dal DTM. L'algoritmo di classificazione supervisionata è stato addestrato su un campione di punti estratti e identificati sul dato acustico tramite interpretazione esperta; il set di punti è stato suddiviso randomicamente in proporzione 70 : 30 tra un campione di addestramento ed uno di validazione. Per forzare la specializzazione del modello sulle firme acustiche locali, la configurazione canonica Random Forest è stata sostituita con un'architettura *Decision Tree Bagging*. Lo spazio delle feature è stato rigorosamente ottimizzato tramite analisi statistiche (Test Boruta, RFE, analisi di collinearità), integrando descrittori chiave come la batimetria media. La robustezza del modello è stata testata attraverso la trasferibilità spaziale in quattro diverse aree, distribuite tra il Mar Ligure e il Mar di Sardegna, e variando il livello di complessità tassonomica (da 3 a 5 classi). L'accuratezza della classificazione è stata calcolata sui campioni di validazione con pesatura areale delle frequenze nelle matrici confusionali. Infine, le classificazioni sono state validate tramite un confronto geospaziale multisorgente, impiegando dati di verità a mare a microscala (ortofoto da AUV), classificazioni satellitari e la cartografia bionomica locale di riferimento. Le carte tematiche prodotte sono state utilizzate per calcolare

il limite inferiore delle praterie di *P. oceanica*, adottando il filtro statistico del 99° percentile della distribuzione batimetrica, e confrontando il risultato con i valori ottenuti dalla cartografia bionomica di riferimento.

I risultati evidenziano l'elevata performance dell'architettura ottimizzata, che ha raggiunto valori di Overall Accuracy superiori all'85%, con picchi del 95% nell'area pilota per i modelli di base a 3 classi (roccia, sedimento, prateria). Tuttavia, le metriche non si sono mantenute stabili in tutte le aree, evidenziando andamenti di accuratezza sito-specifici. L'introduzione della classe matte morta ha innescato il regolare trade-off tra dettaglio tematico e accuratezza globale; tuttavia, il modello ha dimostrato una notevole capacità di isolare tale habitat a Levanto, pur manifestando difficoltà di generalizzazione nelle aree in cui la classe era spazialmente sottorappresentata. Il confronto multisorgente ha evidenziato il disallineamento di scala tra il dato acustico e le carte bionomiche a mesoscala, dimostrando la capacità del classificatore di delineare variazioni ad un dettaglio inedito e di estendere il range batimetrico noto della prateria, misurandone il limite inferiore.

L'approccio GEOBIA applicato al dato MBES ad altissima risoluzione si conferma uno strumento potente per la cartografia bionomica, a condizione che l'addestramento algoritmico campioni esaustivamente l'eterogeneità morfologica locale. Lo studio ha inoltre evidenziato e documentato i limiti strumentali del sensore acustico, in particolare la difficoltà nel risolvere il mascheramento morfologico nel mosaico tra *P. oceanica* e substrato roccioso e nel discriminare fanerogame pioniere a bassa biomassa come *Cymodocea nodosa*.

La presente ricerca è stata condotta presso il laboratorio di Ecologia del Paesaggio Marino nel Dipartimento di Scienze della Terra, dell'Ambiente e della Vita (DiSTAV) dell'Università di Genova.

ABSTRACT

The bionomic mapping of benthic habitats represents a fundamental tool for both research, within the framework of seascape ecology, and marine environmental management. *Posidonia oceanica* meadows represent a priority habitat (Habitats Directive 92/43/EEC), for the protection of which European Union member states are required to carry out mapping and continuous monitoring. This study falls within the framework of the oceanographic campaign of the PNRR MER (Marine Ecosystem Restoration) project, which involves the detailed mapping of marine seagrass meadows on a national scale through the acquisition of acoustic data using a multibeam echosounder (MBES) with sub-metric native resolution (0.2 m), integrated with other seafloor surveying techniques. This study adopts the GEOBIA (Geographic Object-Based Image Analysis) approach to classify the multibeam-acquired data in order to produce a high-resolution bionomic map. Specifically, it aimed to optimize a supervised classification model using the Levanto area as a pilot site, with the dual purpose of maximizing local predictive accuracy and increasing thematic resolution by introducing and isolating the dead matte class. The classification was based on MBES acoustic data (backscatter and bathymetry) and on morphometric derivatives extracted from the DTM. The supervised classification algorithm was trained on a sample of points extracted and identified on the acoustic data through expert interpretation; the point set was randomly split in a 70:30 ratio between a training and a validation sample. To force the model's specialization on local acoustic signatures, the canonical Random Forest configuration was replaced with a Decision Tree Bagging architecture. The feature space was rigorously optimized through statistical analyses (Boruta Test, RFE, collinearity analysis), integrating key descriptors such as mean bathymetry. The model's robustness was tested through spatial transferability across four different areas, distributed between the Ligurian Sea and the Sardinian Sea, and by varying the level of taxonomic complexity (from 3 to 5 classes). Classification accuracy was calculated on the validation samples using area-weighted frequencies in the confusion matrices. Finally, the classifications were validated through a multi-source geospatial comparison, employing micro-scale ground truth data (AUV orthophotos), satellite classifications, and the local reference bionomic cartography. The generated thematic maps were used to calculate the lower limit of *P. oceanica* meadows, adopting the 99th percentile statistical filter of the bathymetric

distribution, and comparing the result with values obtained from the reference bionomic cartography.

The results highlight the high performance of the optimized architecture, which achieved Overall Accuracy values exceeding 85%, with peaks of 95% in the pilot area for the basic 3-class models (rock, sediment, meadow). However, the metrics did not remain stable across all areas, revealing site-specific accuracy trends. The introduction of the dead matte class triggered the typical trade-off between thematic detail and global accuracy; nevertheless, the model demonstrated a remarkable ability to isolate this habitat in Levanto, despite showing generalization difficulties in areas where the class was spatially underrepresented. The multi-source comparison highlighted the scale misalignment between the acoustic data and mesoscale bionomic maps, demonstrating the classifier's ability to delineate variations at an unprecedented level of detail and to extend the known bathymetric range of the meadow by measuring its lower limit.

The GEOBIA approach applied to ultra-high-resolution MBES data proves to be a powerful tool for bionomic cartography, provided that the algorithmic training exhaustively samples the local morphological heterogeneity. Furthermore, the study highlighted and documented the instrumental limits of the acoustic sensor, particularly the difficulty in resolving the morphological masking in the mosaic between *P. oceanica* and rocky substrate, and in discriminating low-biomass pioneer seagrasses such as *Cymodocea nodosa*.

This research was conducted at the Marine Landscape Ecology Laboratory in the Department of Earth, Environmental and Life Sciences (DiSTAV) of the University of Genoa.

INDICE

1. INTRODUZIONE	1
1.1 La cartografia bionomica	1
1.2 Tecniche di mappatura	3
1.3 Sistemi di classificazione Machine Learning	4
1.4 Progetto MER	5
2. SCOPO DELLA TESI	7
3. MATERIALI E METODI	9
3.1 Aree di studio	9
3.2 Tecniche di rilevamento dei dati utilizzati	13
3.3 Pre-processamento dei dati multibeam	15
3.4 Estrazione delle derivate morfometriche	16
3.5 Acquisizione dei punti di addestramento	19
3.6 Flusso di analisi GEOBIA	21
3.6.1 Segmentazione e validazione degli oggetti immagine	22
3.6.2 Sviluppo del modello di classificazione supervisionata	24
3.6.3 Post-processamento della classificazione	27
3.7 Disegno sperimentale	27
3.8 Valutazione delle performance dell' algoritmo di classificazione	29
3.9 Confronto geospaziale multi-sorgente	31
3.9.1 Armonizzazione tematica	32
3.10 Stima del limite inferiore delle praterie di <i>Posidonia oceanica</i>	34
RISULTATI E DISCUSSIONE	35
4.1 Ottimizzazione dello spazio delle feature	35
4.2 Accuratezza e robustezza del processo di classificazione	39
4.2.1 Accuratezza globale, PA e UA	39
4.2.2 Varianza della segmentazione	43
4.3 Confronto geospaziale con il dato di validazione da AUV	43
4.4 Confronto geospaziale con classificazione da satellite	46
4.5 Confronto geospaziale con cartografia bionomica di riferimento	48
4.6 Analisi del limite inferiore	53
CONCLUSIONI	56
BIBLIOGRAFIA	60
RINGRAZIAMENTI	64
APPENDICE	65

1. INTRODUZIONE

Il presente studio si propone di mettere in atto la costruzione di una carta bionomica ad alta risoluzione utilizzando il dato di rilievo acustico del fondale, attraverso il processamento delle immagini secondo il paradigma Geographic Object-Based Image Analysis di classificazione automatica supervisionata, basata su algoritmi di apprendimento automatico.

1.1 La cartografia bionomica

La cartografia bionomica è la mappatura degli elementi ecologici che costituiscono il paesaggio marino. Il concetto ha origine nella scuola francese di Pérès e Picard (1964), che per primi hanno classificato i popolamenti bentonici e identificato i gradienti abiotici principali che regolano la distribuzione degli organismi nei fondali marini. Con l'affermarsi, nei decenni successivi, dell'ecologia del paesaggio, e, al seguito, il concetto di paesaggio marino, lo studio spaziale dell'ambiente marino ha portato ad una transizione dalla catalogazione descrittiva delle biocenosi ad uno strumento di misura e analisi spaziale quantitativa, nella quale vengono adottati gli strumenti teorici ed analitici dell'ecologia del paesaggio terrestre. La mappatura dell'ambiente bentonico (benthic habitat mapping) ha assunto in questo modo un ruolo imprescindibile per la comprensione delle dinamiche costiere (Brown et al., 2011). Un obiettivo primario nella cartografia bionomica contemporanea è quello di definire uno stato iniziale di riferimento (*baseline* ecologica), che permetta di identificare anomalie e traiettorie nella distribuzione spaziale degli habitat.

Le difficoltà materiali per l'osservazione dell'ambiente marino, accompagnate dal ritardo teorico nella affermazione del concetto di paesaggio marino, hanno fatto sì che ancora oggi le informazioni cartografiche sono molto limitate, spesso con estensione limitata, o datate e con ipergeneralizzazione spaziale. Questo è il caso delle cartografie storiche a mesoscala, spesso ottenute solo tramite campionamenti del substrato puntiformi e distanziati.

Il paradigma di base del mosaico, proveniente dall'ecologia del paesaggio terrestre, permette di analizzare la strutturazione del paesaggio attraverso l'analisi delle forme geometriche. Le unità bionomiche si considerano strutturate in un complesso mosaico in cui interagiscono, e dove i

pattern spaziali delle macchie degli habitat influenzano i processi ecologici e la vita degli organismi che vi abitano. La struttura spaziale in patch, corridoi e matrice determina la connettività e le dinamiche di popolazione che possono instaurarsi.

La complessità della struttura spaziale dell'ambiente marino, ed in particolare del dominio bentonico costiero, rendono necessario lo studio a livello sub-metrico per comprendere queste dinamiche.

La specie *Posidonia oceanica* (L.) Delile è una pianta superiore endemica del mediterraneo, e per il suo ruolo di ingegnere ecosistemico, nel costruire le biocenosi climax del piano infralitorale, rappresenta un indicatore ecologico cardine per l'ecosistema marino mediterraneo. Per questa ragione è stata anche adottata come "Elemento di Qualità Biologica" dalla normativa europea MSFD. Come sottolineato da Montefalcone (2009), la diagnostica dello stato di salute dell'ecosistema costiero dipende strettamente dalla capacità di misurare con precisione i descrittori strutturali della prateria. Tra questi, il limite inferiore della prateria rappresenta il parametro bionomico più sensibile: esso risponde in modo diretto e integrato nel tempo alle variazioni di trasparenza della colonna d'acqua e agli stress ambientali (es. torbidità indotta da ripascimenti o inquinamento). Pergent-Martini et al. (2005) evidenziano come la regressione del limite o l'aumento della frammentazione interna possano essere quantificati rigorosamente solo disponendo di cartografie bionomiche ripetibili e dotate di una risoluzione spaziale adeguata alla scala del disturbo.

Oggi, la produzione di cartografia bionomica di precisione non è funzionale soltanto alla ricerca ecologica di base, ma costituisce un obbligo normativo vincolante, essendo essenziale per ogni strategia di conservazione e gestione dell'ambiente marino.

In primo luogo, la Direttiva Habitat (92/43/CEE) impone agli Stati membri la mappatura e il monitoraggio periodico degli habitat di interesse comunitario, con priorità assoluta per le praterie di *Posidonia oceanica* (codice 1120*) e le scogliere (codice 1170), al fine di istituire e gestire i siti della Rete Natura 2000 (ZSC e SIC). Parallelamente, la Marine Strategy Framework Directive (MSFD, 2008/56/CE) richiede il raggiungimento del Buon Stato Ambientale (Good Environmental Status, GES), dove i Descrittori 1 (Biodiversità) e 6 (Integrità del fondo marino) richiedono esplicitamente la quantificazione dell'estensione spaziale degli habitat bentonici e la valutazione della percentuale di fondale non soggetta ad alterazioni fisiche.

Inoltre, la cartografia bionomica rappresenta lo strumento fondativo della Pianificazione Spaziale Marittima (MSP). In un mare sempre più affollato da attività antropiche (cavi sottomarini, acquacoltura, parchi eolici offshore, turismo e pesca), l'approccio ecosistemico

integrato (Ecosystem-Based Management) necessita di mappe continue e affidabili per minimizzare i conflitti d'uso e indirizzare le attività economiche lontano dalle aree di elevato pregio o vulnerabilità ecologica (Katsanevakis et al., 2011).

1.2 Tecniche di mappatura

L'evoluzione della cartografia bentonica ha richiesto il progressivo superamento dei limiti spaziali intrinseci ai tradizionali campionamenti diretti. Le ispezioni visive, condotte tramite operatori subacquei o sistemi video filoguidati (ROV, *drop-camera* e veicoli autonomi AUV), garantiscono un inestimabile dato di verità mare ad alta risoluzione (*ground truth*). Tuttavia, tali metodi restituiscono informazioni di natura puramente puntuale o strettamente limitate alla traiettoria di un transetto. L'impossibilità logistica ed economica di coprire vaste aree con continuità spaziale rende questi approcci inadeguati per la mappatura a scala di paesaggio, imponendo interpolazioni geostatistiche storicamente affette da elevata incertezza (Brown et al., 2011; Diesing et al., 2016).

Per ovviare alla frammentarietà del dato ottico, la cui operatività è inibita dall'assorbimento della radiazione luminosa e dalla torbidità della colonna d'acqua, la cartografia marina recente fa ampio uso del telerilevamento acustico (*Remote Sensing*). Il sistema d'elezione per la mappatura bionomica costiera è l'ecoscandaglio multifascio (*Multibeam Echosounder*, abbreviato in MBES). A differenza dei sistemi a singolo fascio, il MBES emette un ventaglio trasversale di impulsi sonori ad altissima frequenza (*swath*), garantendo una copertura del fondale pressoché totale. L'acquisizione sinottica generata dall'elaborazione del segnale acustico restituisce due output fondamentali. Il primo è il Modello Digitale del Terreno (DTM), derivato dalla misurazione del tempo di ritorno dell'impulso sonoro corretto per la velocità del suono in acqua. Dal DTM si estraggono i parametri morfometrici del fondale, quali pendenza, esposizione, indici di curvatura e rugosità strutturale (Lecours et al., 2016). Tali metriche quantificano la complessità topografica e rappresentano fattori ambientali determinanti per l'identificazione spaziale delle nicchie ecologiche degli organismi bentonici (Brown et al., 2011). Il secondo output, strettamente complementare, è il segnale di *backscatter* (riflettività), che misura l'intensità dell'eco di ritorno. L'ampiezza di questo segnale varia in funzione dell'angolo di incidenza, della durezza e della microrugosità del substrato, permettendo di discriminare le diverse litologie abiotiche (es. roccia vs sedimento incoerente). Inoltre, il *backscatter* si rivela fortemente sensibile alla presenza di coperture biotiche: formazioni dotate di biomassa volumetrica strutturata, come la *canopy* fogliare e l'apparato rizomatoso delle praterie di

Posidonia oceanica, generano un peculiare fenomeno di dispersione volumetrica, fornendo una firma acustica essenziale per la discriminazione algoritmica dell'habitat (Parnum, 2007; Fakiris et al., 2019).

1.3 Sistemi di classificazione Machine Learning

Per superare la digitalizzazione manuale dei dati acustici, che è molto onerosa in termini di tempo e intrinsecamente soggettiva, si rende necessaria l'adozione di rigorosi processi di estrazione automatizzata dell'informazione. L'applicazione di algoritmi convenzionali basati sull'analisi spettrale del singolo pixel (*pixel-based*), ampiamente utilizzati nel telerilevamento terrestre, risulta tuttavia inefficace nella mappatura degli ecosistemi marini. Valutando esclusivamente il dato puntuale decontestualizzato, tali sistemi ignorano la naturale eterogeneità spaziale e la sfumatura degli ecotoni bentonici, producendo mappe eccessivamente frammentate da severi artefatti statistici ("effetto sale e pepe") (Blaschke, 2010; Diesing et al., 2016). Lo standard analitico affermato per superare questo limite strutturale è l'approccio *Geographic Object-Based Image Analysis* (GEOBIA). Tale paradigma processa le immagini geografiche simulando la percezione cognitiva umana, operando su due livelli gerarchici e sequenziali. In una prima fase, definita segmentazione, il dato raster viene frazionato in poligoni spazialmente contigui ed ecologicamente omogenei (*image objects*), i quali costituiscono le nuove unità spaziali di base. In una seconda fase, la classificazione non valuta più il singolo pixel, ma l'oggetto nel suo complesso. Questo salto scalare permette di estrapolare un ampio spazio multidimensionale di variabili (*feature space*), integrando i valori acustici medi con indici di dispersione statistica (deviazioni standard), parametri geometrici di forma, metriche di *texture* e relazioni topologiche di vicinato (Blaschke, 2010). L'assegnazione dei segmenti poligonali alle rispettive classi bionomiche viene eseguita da algoritmi di *Machine Learning* supervisionati. Fornendo un set di campioni a classificazione nota (*training samples*), l'algoritmo apprende induttivamente le complesse regole non lineari di separazione delle firme acustico-morfologiche. Tra i classificatori impiegati in ambito GEOBIA, il *Random Forest* (Breiman, 2001) si distingue per le elevate performance predittive applicate all'ecologia spaziale. Esso si basa su un'architettura di apprendimento d'insieme (*ensemble learning*) che, durante la fase di addestramento su sottocampioni casuali, costruisce una moltitudine di alberi decisionali decorrelati. In fase di predizione, ogni albero valuta l'oggetto ed esprime un voto; l'assegnazione della classe tematica finale avviene per votazione maggioritaria (*majority vote*). Al fine di evitare fenomeni di iperadattamento (*overfitting*) causati dall'eccessiva dimensionalità delle

variabili in ingresso, tale architettura richiede tuttavia una rigorosa ottimizzazione (o selezione) dello spazio delle *feature*, mantenendo operativi i soli descrittori acustici e topografici strettamente necessari a garantire la stabilità e la trasferibilità spaziale del modello (Kursa & Rudnicki, 2010; Ismail et al., 2015).

1.4 Progetto MER

Il presente lavoro si innesta nella cornice operativa del progetto istituzionale PNRR MER (Marine Ecosystem Restoration). Tale intervento, coordinato dall'Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA) in qualità di stazione appaltante, rappresenta un'iniziativa oceanografica senza precedenti per estensione, finalizzata alla mappatura nazionale e alla tutela degli habitat marini, con particolare focus sulle praterie di *Posidonia oceanica* e *Cymodocea nodosa*.

L'esecuzione dei rilievi per il progetto MER prevede l'impiego sinergico di molteplici tecnologie di ultima generazione. Per l'ambiente sommerso costiero, l'utilizzo estensivo di sistemi ecoscandaglio multibeam conformi ai massimi standard idrografici ha permesso l'acquisizione di un dato acustico con risoluzione nativa submetrica (pixel fino a 20 cm). Un tale livello di dettaglio morfologico inaugura nuove potenzialità per lo studio dei processi ecologici paesaggistici, ma pone parallelamente una complessa sfida analitica: la necessità di ottimizzare e validare architetture algoritmiche in grado di tradurre l'iper-risoluzione acustica in una cartografia bionomica affidabile e riproducibile su vasta scala (Diesing et al., 2016; Rende et al., 2020).

In questo contesto tecnologico, la sfida geomatica si intreccia indissolubilmente con le urgenze ecologiche. Negli ecosistemi costieri, e con particolare evidenza nel bacino ligure, le praterie di *Posidonia oceanica* hanno storicamente subito marcati fenomeni di regressione (Montefalcone, 2009). Come ampiamente documentato in letteratura, la valutazione rigorosa dello stato di conservazione di questi habitat non può limitarsi al censimento della prateria viva, ma richiede l'identificazione spaziale della matte morta (Pergent-Martini et al., 2005; Montefalcone, 2009). Quest'ultima non rappresenta unicamente la testimonianza storica dell'estensione originaria del posidonieto, ma costituisce il substrato d'elezione per possibili interventi di ripristino ecologico, i quali rappresentano l'obiettivo centrale del progetto MER (Rende et al., 2022).

Tuttavia, l'identificazione acustica della matte morta costituisce un limite tecnico complesso: essendo priva della canopy fogliare, essa perde la firma di scattering tipica di *P. oceanica* viva, confondendosi spettralmente con i sedimenti incoerenti circostanti (Parnum, 2007; Fakiris et al., 2019). Emerge quindi la necessità di migliorare i limiti predittivi dei modelli di classificazione per rispondere all'esigenza di interesse ecologico e gestionale di isolare la matte con il segnale acustico.

2. SCOPO DELLA TESI

Il presente lavoro si inserisce nel contesto della cartografia bionomica ad altissima risoluzione, sfruttando i dati acustici multibeam acquisiti nell'ambito del progetto PNRR MER. A fronte di una iperrisoluzione spaziale nativa (pixel di 20 cm), lo studio si propone di superare i limiti predittivi dei modelli standard, sviluppando un'architettura di classificazione GEOBIA (Geographic Object-Based Image Analysis) mirata a massimizzare congiuntamente la precisione spaziale e il dettaglio ecologico.

Per conseguire tale finalità, la ricerca è stata strutturata attorno a due obiettivi operativi primari. In primo luogo, si è mirato alla specializzazione dell'algoritmo per incrementarne l'accuratezza predittiva, assumendo l'area di Levanto come sito pilota. Al fine di forzare l'adattamento del modello sulle firme acustiche locali, si è scelto di sostituire la configurazione canonica Random Forest – utilizzata come baseline metodologica nel progetto MER – con un'architettura Decision Tree Bagging. Contestualmente, si è proceduto a un'ottimizzazione mirata dello spazio delle feature, comprendente la rigida selezione delle variabili e l'integrazione della batimetria media, necessaria per la risoluzione delle ambiguità ecotonali. In secondo luogo, lo studio ha perseguito l'incremento della risoluzione tematica. Partendo da un modello di base a tre classi (roccia, sedimento, seagrass), l'obiettivo è stato quello di introdurre e isolare una classe specifica per la matre morta (classe “matte”) poiché ecologicamente importante nell'ambito dell'analisi dello stato delle praterie di *Posidonia oceanica* e per il loro restauro. L'estensione dei modelli a quattro e cinque classi ha consentito di quantificare il conseguente trade-off tra il maggiore dettaglio bionomico e la stabilità delle metriche di accuratezza globale.

A supporto e verifica di queste implementazioni analitiche, il disegno sperimentale ha previsto una rigorosa fase di validazione. Per testare la trasferibilità del metodo, l'architettura iperspecializzata su Levanto è stata applicata a quattro ulteriori aree di studio distribuite tra il Mar Ligure e il Mar di Sardegna, con lo scopo di quantificare il livello di overfitting metodologico introdotto e valutare la tenuta del classificatore al variare del contesto geomorfologico. Parallelamente, le predizioni acustiche sono state sottoposte a un confronto geospaziale multisorgente, incrociandole sia con indagini visive di microscala (ortofoto da

AUV) sia con la cartografia istituzionale a mesoscala, come il *Nuovo Atlante degli Habitat Marini della Liguria* (Coppo et al., 2020). Tale analisi comparativa è stata delineata per valutare il disallineamento spaziale (*mismatch* di scala) e per misurare la precisione del modello nell'identificare confini critici, primo fra tutti il limite batimetrico inferiore della prateria. Infine, l'indagine si è posta l'obiettivo di identificare i limiti fisici strumentali, documentando le condizioni bionomiche in cui il sensore MBES non dispone di chiavi spettrali sufficienti alla classificazione. Ciò ha permesso di circoscrivere specifiche criticità acustiche, riscontrabili in presenza di mascheramento morfologico, come nel mosaico tra *Posidonia oceanica* e substrato roccioso, o nel tentativo di discriminare fanerogame pioniere a bassa biomassa, quali *Cymodocea nodosa*.

3. MATERIALI E METODI

Questo studio è stato svolto utilizzando i rilevamenti provenienti dal progetto MER (*Marine Ecosystem Restoration*) – ricompreso nel PNRR - Missione M2 - Componente C4 - Investimento 3.5 “Ripristino e tutela dei fondali e degli habitat marini” – Intervento A16 – A18 “Esecuzione dei rilievi per la mappatura nazionale delle praterie di *Posidonia oceanica* e *Cymodocea nodosa* mediante rilievi LiDAR topografici, batimetrici e con indagini geofisiche di completamento”, coordinato dall’Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA) in qualità di Stazione Appaltante.

L’intervento A16 – A18, in corso d’opera al momento della stesura della presente tesi, mette in atto molteplici campagne di rilevamento con molte tecnologie diverse. I dati acquisiti che prendiamo in esame nel presente studio sono: il rilevamento con ecoscandaglio multifascio (multibeam), il telerilevamento mediante satellite di immagini multispettrali ad altissima risoluzione, ed il rilevamento videofotografico ortorettificato, realizzato con veicoli subacquei autonomi (AUV, *autonomous underwater vehicle*). Questi tre rilievi risultano completati nelle aree di studio selezionate, ad eccezione del rilevamento con AUV nella regione della Sardegna.

3.1 Aree di studio

Le aree prese in esame sono state selezionate tra le aree mappate di cui i dati sono risultati disponibili all’inizio di questo studio. La campagna di rilevamento con *multibeam* nella regione ligure ricopre quattordici aree di fondale marino, estese da decine a centinaia di ettari e distanziate di circa 1–60 km lungo la costa. Tra queste, sono state selezionate quattro aree in modo da rappresentare al meglio la varietà di ambienti infralitorali vegetati da praterie di *Posidonia oceanica* nel Mar Ligure. In aggiunta, è stata studiata un’area ulteriore nella costa occidentale della Sardegna, caratterizzata da condizioni oceanografiche molto diverse da quelle presenti nelle coste liguri. Di seguito si riporta l’elenco delle cinque aree di studio con le denominazioni adoperate nel presente lavoro (Figura 1.A), provenienti dalle denominazioni di progetto:

- Levanto
- Genova_03 – Secca di Santo Stefano

- Genova_19 – Punta Baffe
- Genova_20 – Punta Moneglia
- Sar_W_62 – Scoglio del Catalano

L'area Levanto (SP) è stata impiegata come area pilota per lo sviluppo del sistema di classificazione, esteso in una seconda fase sulle altre quattro aree. Questa area dispone di un dataset che comprende il rilievo MBES, la classificazione derivata da immagine satellitare multispettrale, ed il transetto di ortofoto da AUV. Il rilievo acustico con multibeam ricopre una superficie di 28 ha, con estensione rettangolare di 490 m per 550 m e si trova nel lato settentrionale della baia di Levanto (coordinate geografiche WGS84: Lat. 44.169081 °N, Long. 9.596009 °E), a circa 40 m dalla linea di costa. L'area è caratterizzata da fondali di sedimento sabbioso che declinano dal margine nord, corrispondente alla batimetrica di 10 m, verso il punto più profondo nel margine sud-est, che raggiunge i 30,8 m di profondità (Figura 1.b). Nella zona centrale dell'area mappata è presente un rilievo topografico costituito da una formazione rocciosa che raggiunge i 3,13 m dal livello medio mare, noto come Secca di Perea. In prossimità di questo affioramento, aumenta la pendenza del fondale, formando una lieve scarpata tra le batimetriche di 19 e 25 m. A nord e ad ovest dell'area rocciosa, nel *Nuovo Atlante degli Habitat Marini della Liguria* (Coppo et al. 2020), risulta la presenza di *matte*, rappresentata nella classe “mosaico di *Posidonia oceanica* viva e matte morta”.

Il rilievo con multibeam su questa area è stato realizzato il 23/08/2024. L'acquisizione dell'immagine satellitare multispettrale è avvenuta in data 14/04/2024, mentre per il dato di

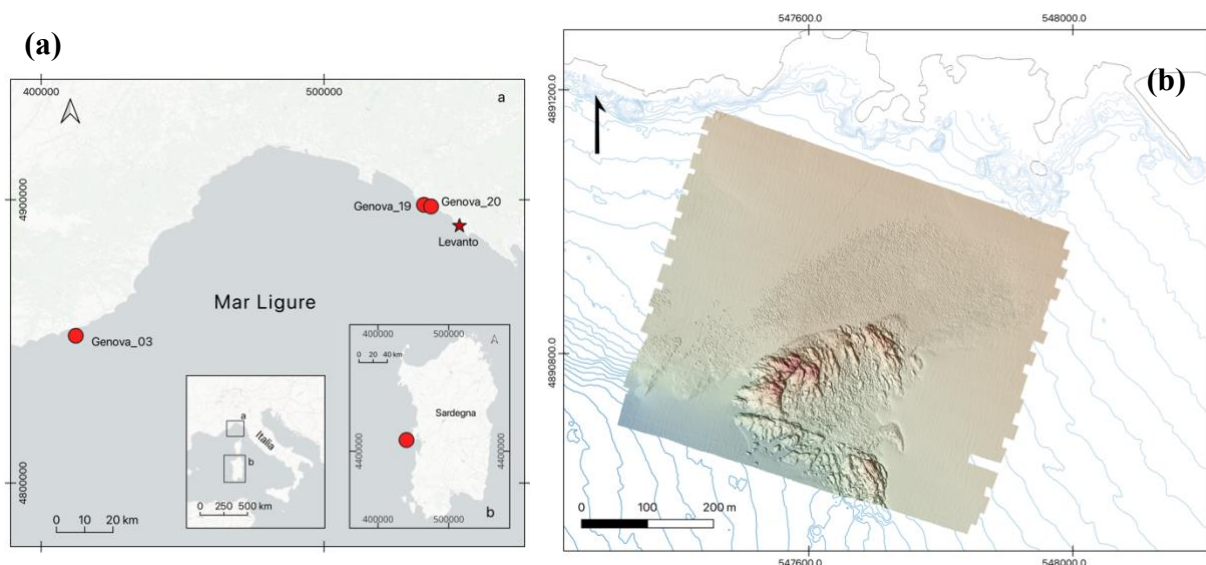


Figura 1. A) Inquadramento geografico delle aree di studio. a) Aree del Mar ligure; b) Sar_W_62. **B)** Modello digitale del terreno (DTM) dell'area di studio di Levanto, in scala a falsi colori con ombreggiatura. In nero, linea di costa; in gradazione di blu, curve isobatimetriche (equidistanza 1 m). Sistema di riferimento: WGS 84 / UTM 32N

verità a mare (in letteratura *ground truth data*) è avvenuta il 17/12/2024 per il transetto con Seacat ed il 09/10/2025 per le acquisizioni con AUV Quadroin.

Il secondo dataset ricopre l'area mappata nel rilievo presso Punta Baffe (Sestri Levante, GE), situato alle coordinate 44.235449 °N, 9.443005 °E (WGS84). Questa area, denominata "Genova 019" secondo la nomenclatura della campagna di rilevamento con multibeam, ricopre una superficie di 22,5 ha, con estensione batimetrica da 17 a 50 m (Figura 2.a). Sono presenti numerosi rilievi positivi, dovuti all'affioramento dell'assetto geologico locale, con sviluppo in direzione nord-sud, in continuità con la topografia della punta limitrofa.

L'area Genova_20 (Figura 2.b) si trova in prossimità di Punta Moneglia, coordinate 44.2311160 °N, 9.4698745 °E (Moneglia, GE). Il rilievo MBES ricopre un'area di 23,3 ha dalla profondità di 20 m a 47 m. Come nell'area prospiciente a Punta Baffe, la topografia è caratterizzata da affioramenti rocciosi che conformano un declivio ripido in prossimità del capo retrostante.

Nelle aree di Punta Moneglia e di Punta Baffe, i rilevamenti con ecoscandaglio multibeam sono stati acquisiti nella seconda metà di settembre 2024 (26/09/24), mentre le immagini satellitari risalgono al 29/07/2024 ed al 23/08/2024, rispettivamente. I transetti fotografici, infine, sono stati campionati il 17/12/2024.

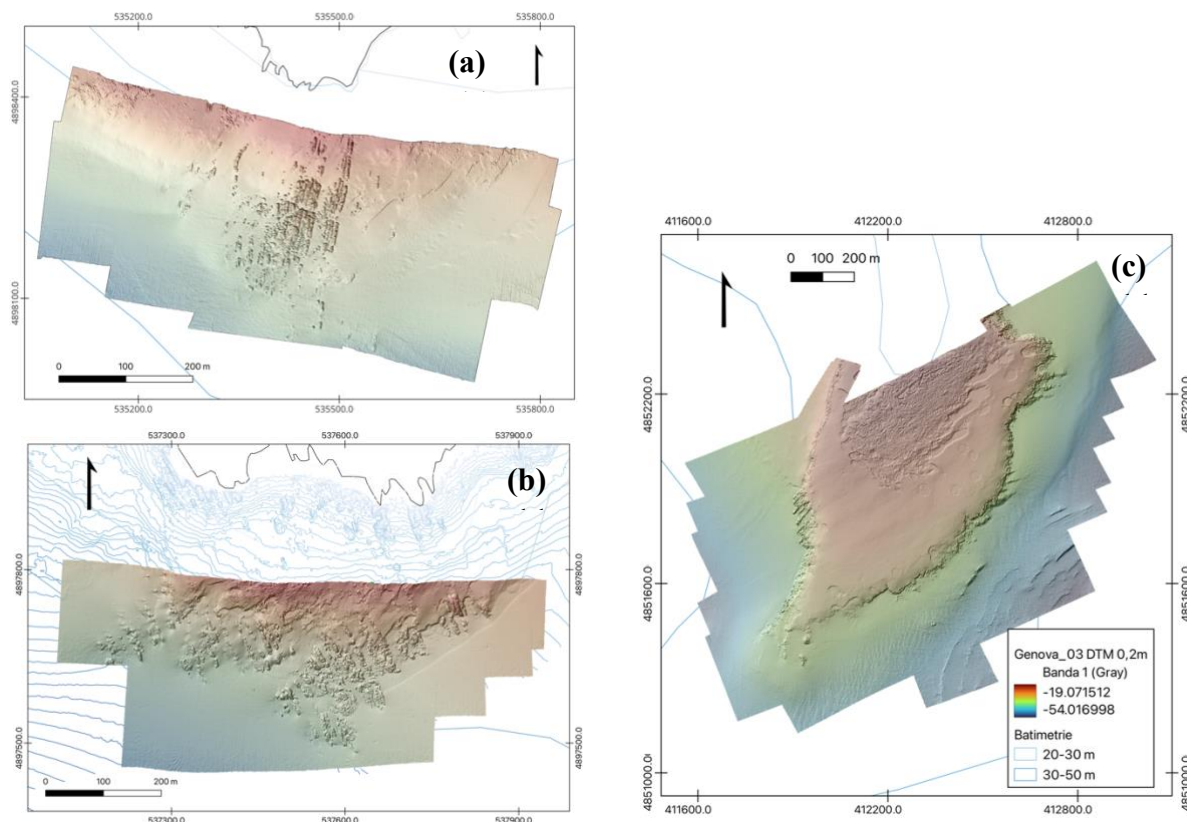


Figura 2. Modello digitale del terreno (DTM) delle aree di studio Genova_19 (a), Genova_20 (b), Genova_03 (c) in scala a falsi colori con ombreggiatura. In nero, linea di costa; in gradazione di blu, curve isobatimetriche. Sistema di riferimento: WGS 84 / UTM 32N

La quarta area, denominata Genova 003, ricopre la Secca di Santo Stefano, al largo di S. Stefano al Mare (IM) (coordinate geografiche WGS84 43.814273 °N, 7.909191 °E). Questo dataset presenta una estensione di 127 ha, che ricoprono la porzione più meridionale della prateria di *P. oceanica*, che si estende dal margine nord, profondo 19 m, fino a circa 30 m (Figura 2.c). La secca è contornata da una scarpata rocciosa di 3–6 m, alla base della quale degrada il fondale caratterizzato da sedimento fine (sabbie e fanghi) fino a 54 m (Coppo et al., 2020). Nell'area Genova 003, il rilevamento con ecoscandaglio multibeam è stato acquisito il 7/10/2024, mentre l'immagine satellitare il 26/07/2024 ed i transetti con AUV il 05/12/2024.

La quinta area è situata all'interno dell'Area Marina Protetta Isola di Mal di Ventre e Penisola del Sinis, nella costa occidentale della Sardegna (OR), tra Capo San Marco e lo Scoglio del Catalano, in zona di protezione C, alle coordinate 39.891745 °N, 8.302246 °E (WGS 84). Questa area, denominata SAR_W_062, si estende per 652,5 ha su fondali pianeggianti, ampiamente ricoperti da praterie di *P. oceanica*, che si estende oltre il margine nord-est nell'ampio posidonieto della Secca di Mezzo, e con affioramenti rocciosi circoscritti all'area più prossima allo Scoglio del Catalano (Figura 3). La profondità in quest'area è compresa tra 20 e 45 m. Nell'area Sar_W_062, il rilevamento con multibeam è stato acquisito il 21 e 22 febbraio 2025, mentre l'immagine satellitare risale al 21/08/2023. Al momento del presente studio, non è stata svolta la campagna di rilevamento dei transetti di verità a mare con AUV nel sito Sar_W_062.

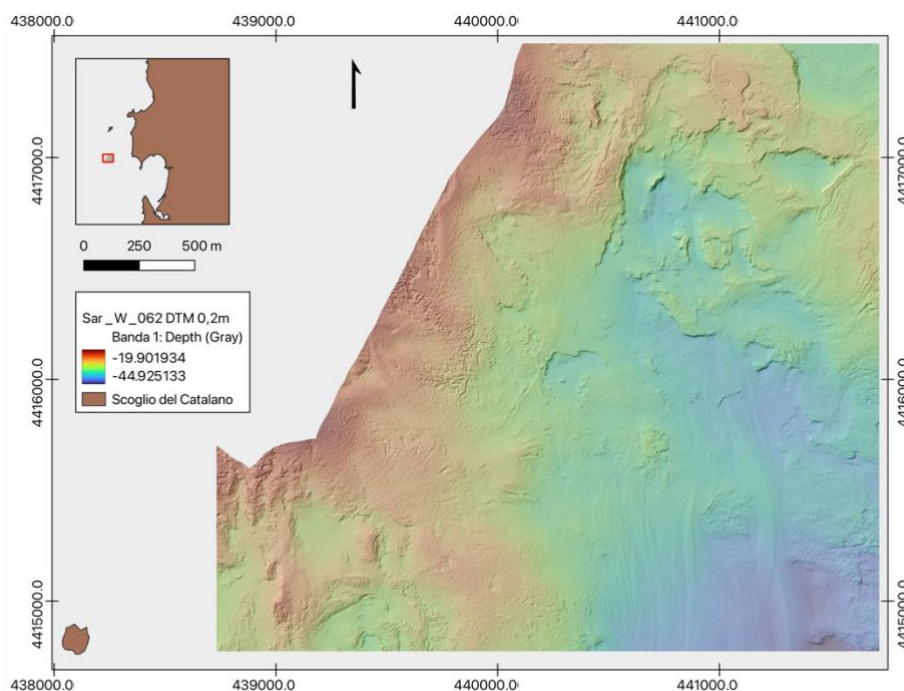


Figura 3. Modello digitale del terreno (DTM) dell'area di studio Sar_W_62, in scala a falsi colori con ombreggiatura. Sistema di riferimento: WGS 84 / UTM 32N

3.2 Tecniche di rilevamento dei dati utilizzati

L'ecoscandaglio multifascio è un sensore batimorfologico acustico attivo, che emette e riceve onde ad altissima frequenza (oltre i 12 KHz). Ogni sensore (trasduttore di ricezione) misura riflettività e tempo di ritorno della eco, riflessa sul fondale, del fascio di onde acustiche generato dai trasduttori di emissione. L'intensità del segnale di ritorno (riflettività o *backscatter*) è determinata da molteplici caratteristiche, tra cui le principali risultano la durezza e compattezza della superficie riflettente e l'angolo di incidenza dell'onda acustica. Ne risulta che l'immagine bidimensionale e georeferenziata del *backscatter* rappresenta l'informazione sul materiale e la rugosità del fondale marino. Il tempo di ritorno costituisce una misura indiretta della distanza della superficie riflettente dal sensore. Per cui questo segnale permette di investigare la topografia del fondale, con risoluzione pari alla distanza tra i singoli segnali acustici registrati dal sensore.

In questo progetto è stato utilizzato un sistema multibeam Kongsberg EM2040 MKII conforme allo standard IHO S44, con l'ecoscandaglio montato sullo scafo dell'imbarcazione, frequenza non inferiore a 400 kHz, impronta del fascio di $0,4^\circ \times 0,7^\circ$ e frequenza di campionamento (*ping rate*) non inferiore a 40 Hz (lunghezza dell'impulso nell'intervallo 25 μ s – 100 μ s). Al *multibeam* è stata associata una sonda SVP per la misurazione in continuo della velocità del suono lungo l'intera colonna d'acqua. Inoltre, sono stati impiegati i sensori di assetto (*inertial motion unit*, IMU) per la calibrazione delle oscillazioni, rollio, beccheggio imbardata (inglese *roll*, *pitch* e *yaw*), e delle accelerazioni lineari (*surge*, *sway*, *heave*). Infine, si è impiegato un sistema di geoposizionamento di tipo GNSS, con correzioni differenziali basate sulle stazioni afferenti a reti permanenti inquadrature nella RDN 2008, e calibrazione degli offset strumentali nell'imbarcazione.

I dati batimetrici e di riflettività sono, quindi, georeferenziati (tramite sistema GNSS-IMU), corretti con la velocità del suono (dalla sonda SVP) ed il livello di marea, ripuliti da spikes e outliers e normalizzati (con correzione lineare e lambertiana del *backscatter*). La restituzione dei dati processati è in formato raster con risoluzione di 0,2 m, analoga alla dimensione dei singoli *soundings* misurati, con eccezione per il dato batimetrico dell'area di Levanto, che ha risoluzione di 0,5 m. Il dato batimetrico è restituito come DTM, in formato raster a banda singola, in scala negativa di profondità in metri. Il *backscatter* è restituito in raster a banda singola, che riporta il segnale in decibel con range da -60 dB (scuro) a -10 dB (chiaro).

La classificazione bionomica a mosaico derivata da immagine satellitare è stata acquisita con un sistema di telerilevamento multispettrale VHR a 8 bande, alla risoluzione di 2 m. La classificazione delle immagini, ortorettificate e calibrate radiometricamente con correzione atmosferica e della colonna d'acqua, presenta accuratezza planimetrica di 2 pixel. Su queste immagini è stata elaborata la classificazione del fondale marino in quattro categorie:

1. associazioni a *Cymodocea nodosa* su sabbie fini ben classate,
2. Biocenosi di *Posidonia oceanica*
3. Accumuli di detrito fogliare di *Posidonia oceanica*
4. Altri habitat bentonici

La terza classe fa riferimento agli ammassi di foglie morte galleggianti, che possono concentrarsi in prossimità della battigia, specialmente nella stagione estiva e autunnale, raggiungendo una densità che impedisce il telerilevamento del fondale sottostante. Tuttavia, le uniche classi riscontrate in corrispondenza delle aree di studio sono la seconda e la quarta.

Con un veicolo autonomo subacqueo (AUV) è stato ottenuto un dato di verità a mare, acquisendo immagini ottiche del fondale marino lungo transetti larghi circa 2 m. L'immagine ottica, georeferenziata e ortorettificata, è stata acquisita con videocamera a risoluzione 4K montata su AUV. Sono stati impiegati due mezzi diversi: un modello SeaCat ed un Quadroin di EvoLogics. L'ortomosaico presenta una risoluzione dell'immagine subcentimetrica, e accuratezza posizionale di 2 m.

Questi transetti sono stati classificati con sistema *Machine Learning* supervisionato, restituendo un grigliato vettoriale di punti a distanza di 1 m, classificati in 6 tematismi:

- sabbia fine (0),
- sabbia grossolana (1),
- sabbia con componente organica (2),
- sedimento grossolano (3),
- roccia o massi (4),
- posidonia (5).

Si evidenzia che la classificazione adottata nel presente dato non comprende le classi specifiche per le praterie di *C. nodosa* né la matte morta delle praterie di *P. oceanica*. Di conseguenza, questa classificazione delle ortofoto da AUV risulta, allo stato attuale, incompleta per la validazione della classificazione elaborata nel presente lavoro.

3.3 Pre-processamento dei dati multibeam

I dati raster di batimetria e backscatter sono stati ulteriormente processati con il software open-source Quantum GIS (versione long-term-release QGIS 3.34.12-Prizren), poiché i raster presentano dei pixel vuoti. Questi possono essere dovuti a porzioni di fondale dalle quali non è pervenuto il segnale di ritorno, per effetto di ombreggiamento dalla topografia circostante, o sono generate dal processo di ortorettificazione nel caso in cui la distanza sul fondale dei segnali registrati sia risultata superiore alla risoluzione del raster prodotto. Queste aree in cui manca l'informazione sono costituite in massima parte da singoli pixel o da gruppi di meno di 10 pixel adiacenti, e rappresentano meno dello 0,5% dell'estensione totale del raster (massimo 0,495% in Genova_3, minimo di 0,006% in Sar_W_062). Le porzioni marginali con densità di vuoti maggiore del 20 per cento sono state escluse dall'area di classificazione nei dataset Genova_20 e Genova_19. I pixel vuoti restanti sono stati riempiti per interpolazione, adoperando la funzione *r.fillnulls* di GRASS implementata su Quantum GIS. Per le aree di Levanto, Punta Baffe e Punta Moneglia è stato applicato il metodo RST (regularized spline interpolation with tension), con dimensione del bordo (edge) pari a 3 pixel e tensione pari a 40 (si veda Mitsova and Mitsova, 1993, e GRASS Development Team, 2025). Nelle aree di Genova 3 (Secca di Santo Stefano) e Sar_W_062, per praticità computazionale dovuta alla maggiore estensione di questi dataset, è stata applicata l'interpolazione spline bilineare, con bordo di 8 pixel e $\lambda = 0,1$ (si veda Getreuer, 2011, e GRASS Development Team, 2024). L'interpolazione con *r.fillnulls* è stata applicata ai raster DTM e di backscatter; tuttavia, i secondi presentano meno pixel privi di informazione come nel caso Punta Baffe e Sar_W_62, o nessun pixel vuoto come nel dato di Genova_20 e di Genova_03.

La classificazione di dati raster basata su oggetti in eCognition richiede un vettoriale che definisca l'area di lavoro per la analisi di immagine. Questo poligono, denominato *cover*, non può contenere pixel *no-data* in nessuno dei layer in ingresso. Pur avendo interpolato i dati raster per rimuovere i vuoti, si è deciso di adoperare un filtro spaziale per circoscrivere l'espansione dell'area di lavoro ad un limite massimo di 3 pixel, rispetto all'estensione nativa. La creazione delle maschere è stata elaborata in QGIS Prizren tramite vettorializzazione dei file raster non interpolati (strumento *gdal_polygonize.py*), fusione di tutte le feature in un singolo poligono (strumento *dissolve*), e infine ripulitura dei buchi attraverso l'applicazione di un doppio buffer: prima di +0,6 m, e successivamente di -0,8 m. Nelle maschere risultanti sono cancellati solo i buchi di larghezza non superiore a 6 pixel, e si è ottenuto un poligono di copertura con margine

rientrato di circa 1 pixel rispetto all'estensione iniziale del raster. Il poligono *cover* è stato quindi ottenuto dall'intersezione di queste maschere del DTM e del backscatter, in ogni area di studio. Nell'area di Levanto, essendo presente una condotta nel fondale, è stato modificato manualmente il vettoriale *cover* ritagliando la condotta, in modo da escluderla dall'area di classificazione.

3.4 Estrazione delle derivative morfometriche

Per la classificazione del fondale, i dati diretti di batimetria e riflettività, ottenuti dal multibeam, vengono affiancati da ulteriori dati spaziali secondari, denominati *derivative morfometriche*.

Queste sono dei parametri spaziali del terreno (Lecours et al., 2016) calcolati dalla morfologia del fondale, rappresentata dal DTM, che ne descrivono le caratteristiche geometriche. La caratterizzazione della superficie del fondale rappresenta uno strumento fondamentale nell'identificazione degli habitat [*sensu* nicchie ecologiche], a molteplici scale spaziali (Wilson et al., 2007). L'uso della morfometria nella ecologia spaziale è stato sviluppato in ambiente terrestre, per la classificazione degli habitat, la modellizzazione predittiva delle tipologie di suoli, copertura vegetale e distribuzione delle specie (Wilson et al., 2007).

L'analisi della topografia risulta particolarmente informativa in ambito ecologico perché sono in grado di rappresentare le caratteristiche topografiche che identificano la nicchia ecologica degli organismi bentonici. Secondo Brown et al. (2011) e Lecours et al. (2016), i parametri morfometrici del fondale marino influenzano l'idrodinamica, la dinamica sedimentaria e la penetrazione della luce, che rappresentano fattori ambientali determinanti per la distribuzione della fauna. Tra questi, si includono la dinamica locale delle correnti, il tasso di sedimentazione e la tessitura del sedimento, la diversità strutturale, che a loro volta influenzano la disponibilità di nutrienti e cibo per gli organismi filtratori e detritivori e la disponibilità di rifugi e ripari dai predatori.

Oltre a rappresentare dei fattori predittivi della distribuzione degli organismi bentonici, i parametri morfometrici permettono di identificare la copertura del fondale. La presenza di praterie di fanerogame marine è evidenziata da un aumento della rugosità della superficie, che le distingue dalle aree di sedimento prive di vegetazione (Micallef et al., 2012).

L'impiego dei dati morfometrici per la mappatura predittiva degli habitat bentonici (*habitat suitability*) e per la classificazione del fondale marino trova ampio riscontro in letteratura (Wilson et al., 2007, Micallef et al., 2012, Ierodicanou et al., 2007, Ierodicanou et al., 2018,

Janowski et al., 2018a, Janowski et al., 2018b, Wicaksono et al., 2019, Rende et al., 2020, si veda in particolare l'articolo di revisione di Diesing et al., 2016).

Le derivate morfometriche si distinguono in quattro categorie: di orientamento, pendenza, curvatura e rugosità (Wilson et al., 2007). In questo studio sono stati utilizzati sette parametri:

1. Pendenza (Slope): angolo di inclinazione massima della superficie misurato in gradi (0–90).
2. Orientamento (Aspect): misura in gradi (0–360), in senso orario, l'angolo dalla direzione nord. Rappresenta l'esposizione, o orientazione, del versante, ovvero la direzione lungo la massima pendenza.
3. inclinazione nord (Northness): conversione trigonometrica dell'orientamento in variabile lineare continua (da –1 a 1). Rappresenta la pendenza lungo la direzione nord.
4. inclinazione est (Eastness): conversione trigonometrica dell'orientamento in variabile lineare continua (da –1 a 1). Rappresenta la pendenza lungo la direzione est.
5. curvatura generale (General Curvature), descrive la curvatura planare totale, valori positivi indicano concavità, mentre valori negativi corrispondono ad una forma convessa.
6. curvatura totale (Total Curvature): misura scalare dell'intensità della curvatura complessiva in tutte le direzioni (valore minimo di 0 indica una superficie piatta).
7. TRI (terrain ruggedness index) è un indice di rugosità del terreno, basato sul dislivello complessivo tra una cella e le otto circostanti.

Questi parametri sono stati derivati in formato raster a partire dai singoli multibeam di ogni area, utilizzando il software SAGA (System for Automated Geoscientific Analyses, versione 9.11.1) (Conrad et al., 2015), alla risoluzione nativa del modello digitale del terreno su cui sono calcolate.

Le prime sei derivate sono state generate con il modulo di analisi del terreno *Slope*, *Aspect*, *Curvature* del software SAGA, applicando la curva polinomiale di secondo ordine a 9 parametri, che approssima la superficie topografica locale su ogni pixel determinando i parametri su una finestra di 3×3 celle (Zevenbergen e Thorne, 1987). Le derivate di orientamento e pendenza si ottengono dalla derivazione di primo grado della curva, in accordo con le definizioni di Olaya (2006):

$$\text{aspect} = \theta = [90 + \text{atan2}(\partial z/\partial y, \partial z/\partial x)] \times 180/\pi \quad [\text{eq. 1}]$$

$$\text{slope} = \arctan \sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2} \quad [\text{eq. 2}]$$

dove x, y, z rappresentano le coordinate di ogni cella negli assi nord, est ed elevazione rispettivamente. Il calcolo delle derivate di inclinazione nord ed est (Figura 4) corrisponde alla conversione trigonometrica descritta da Wilson et al. (2007):

$$\text{northness} = \cos \theta \quad [\text{eq. 3}]$$

$$\text{eastness} = \sin \theta \quad [\text{eq. 4}]$$

Le derivate di curvatura sono invece calcolate dalla derivata di secondo grado della curva polinomiale. Le formule implementate nel modulo *Slope, Aspect, Curvature* per il calcolo della curvatura generale e curvatura totale sono state verificate direttamente nel codice sorgente del software (modulo *ta_morphometry*) (SAGA GIS Source Code Repository), e si discostano da altre formule proposte in letteratura:

$$\text{General Curvature} = -2 \times \left[\left(\frac{\partial^2 z}{\partial x^2} \right) + \left(\frac{\partial^2 z}{\partial y^2} \right) \right] \quad [\text{eq. 5}]$$

$$\text{Total Curvature} = \left(\frac{\partial^2 z}{\partial x^2} \right)^2 + 2 \times \left(\frac{\partial^2 z}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 z}{\partial y^2} \right)^2$$

[eq. 6]

La curvatura generale rappresenta una versione della curvatura planare totale basata solo sulla variazione di pendenza nelle direzioni orizzontali x , y . È una metrica sensibile alle forme del terreno, che distingue una forma concava (valori positivi) da una forma convessa (valori negativi), e nominalmente può assumere qualsiasi valore reale da $-\infty$ a $+\infty$.

La curvatura totale (equazione 6) è una combinazione quadratica delle tre derivate seconde della curva polinomiale. Questa formula considera la curvatura in tutte le direzioni, con un fattore di peso doppio per la derivata mista, che definisce la curvatura di torsione diagonale. I valori di questa derivativa aumentano dal minimo di 0, per una superficie piatta, a $+\infty$ per curvature crescenti, indipendentemente che si tratti di forme concave o convesse.

Il calcolo dell'indice TRI è basato sulla formula originale di Riley et al. (1999) (equazione 7). Il modulo specifico in SAGA integra una funzione di peso, per bilanciare contributo delle celle in funzione della loro distanza, e permette di modificare l'estensione della finestra di calcolo (equazione 8). In questo studio è stata adoperata una finestra circolare con raggio di 2, che corrisponde alla finestra convenzionale di 3×3 , ed una funzione di peso gaussiana con sigma pari a 75, per ottenere un risultato comparabile alla formula originale. Si evidenzia che la formula utilizzata in SAGA include la cella centrale, sbilanciando il peso a favore del denominatore.

$$TRI = \sqrt{\frac{\sum_{i=1}^8 (z_i - z_0)^2}{8}} \quad [\text{eq. 7}]$$

$$TRI_{SAGA} = \sqrt{\frac{\sum_{i \in N} w(d_i) \cdot (z_i - z_0)^2}{\sum_{i \in N} w(d_i)}} \quad [\text{eq. 8}]$$

dove z_0 è la quota della cella centrale, d_i è la distanza tra la cella i -esima e la cella centrale, in una finestra di N celle, con funzione di peso $w(d_i)$.

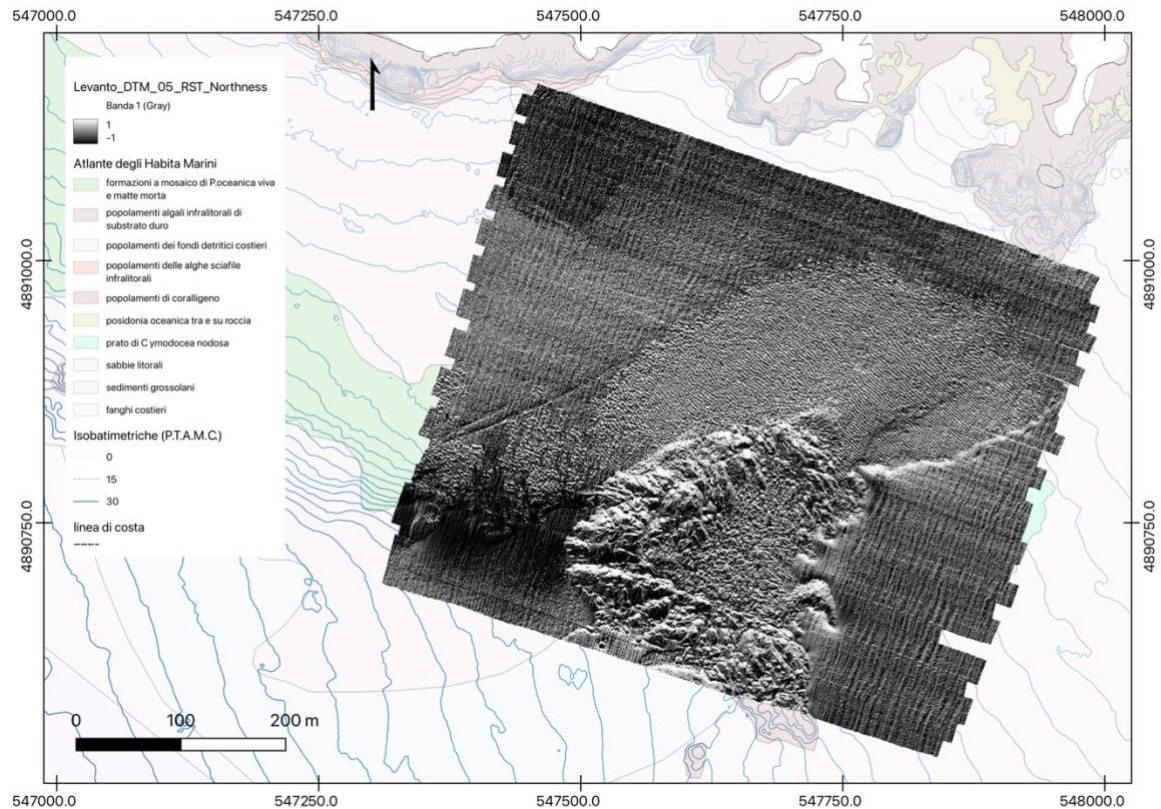


Figura 4. Derivativa morfometrica *northness* calcolata sul modello digitale del terreno dell'area di Levanto. Nello sfondo: atlante degli habitat (Coppo et al. 2020), curve isobatimetriche e linea di costa (dalla cartografia P.T.A.M.C.). Coordinate WGS84 / UTM N32.

3.5 Acquisizione dei punti di addestramento

La classificazione supervisionata necessita di un set di dati di addestramento (*training set*). In questo studio non utilizziamo un dato campionato *in situ*, indipendente rispetto dato acustico che intendiamo classificare, ma utilizziamo un campione di punti di addestramento estratti tramite interpretazione visiva esperta sul dato acustico stesso, del backscatter affiancato al modello digitale del terreno. In questo modo, utilizzando il programma Quantum GIS, per ogni area è stato prodotto un file vettoriale contenente i punti di addestramento (training samples). Le classi utilizzate sono: “seagrass”, “rock”, “sediment”, “sediment F” e “matte”. Le prime tre ricalcano le classi adoperate nel progetto MER, nella mappatura tramite multibeam. La classe *seagrass* rappresenta le praterie di fanerogame marine, che nelle aree del presente studio sono costituite da *P. oceanica*; la classe *rock* racchiude le zone a substrato duro (roccioso); la classe *sediment* contiene le aree a substrato molle. La classe *sediment F* identifica il sedimento fine, introdotta come suddivisione aggiuntiva per il sedimento, utilizzata per distinguere le zone con segnale di backscatter marcatamente più scuro (riflettività molto bassa). La classe *matte*

rappresenta le zone dove si estende la struttura biogenica priva del ricoprimento della prateria di *P. oceanica*, nella condizione comunemente definita “matte morta”. La classe *matte* è stata identificata con il supporto delle ortofoto acquisite con Quadroin, della cartografia pregressa (Coppo et al., 2020) e di una valutazione morfologica della topografia in prossimità delle praterie di posidonia. Il segnale di backscatter delle zone identificate come *matte* non è risultato distinguibile dal segnale del sedimento, per cui non è stato sufficiente per campionare i punti di addestramento per questa classe.

I punti classificati sono stati presi in numero proporzionale all'estensione dell'area e della singola classe nell'area. Il numero di punti presi varia da 176 per l'area di Genova 19 a 569 per l'area Sar_W_62. Nelle aree di Genova 03, Genova 20 e Sar_W_62 non è stata riscontrata la presenza della classe di sedimento grossolano.

Dai vettoriali con i punti classificati sono stati estratti due campioni, separati con suddivisione casuale stratificata per classe, nella proporzione 70:30 proposta da Gholamy et al. (2018). Il set di punti più numeroso è utilizzato come campione di addestramento (*training set*) per la classificazione supervisionata, mentre il secondo campione è utilizzato per la validazione della classificazione (*validation set*). La suddivisione dei punti campione è stata eseguita in ambiente QGIS Prizren tramite script Python che utilizza le classi dell'API PyQGIS “QgsVectorLayer”. Per garantire la riproducibilità dell'estrazione casuale dei punti, è stato impostato un valore fisso al seme casuale: `random.seed(42)`.

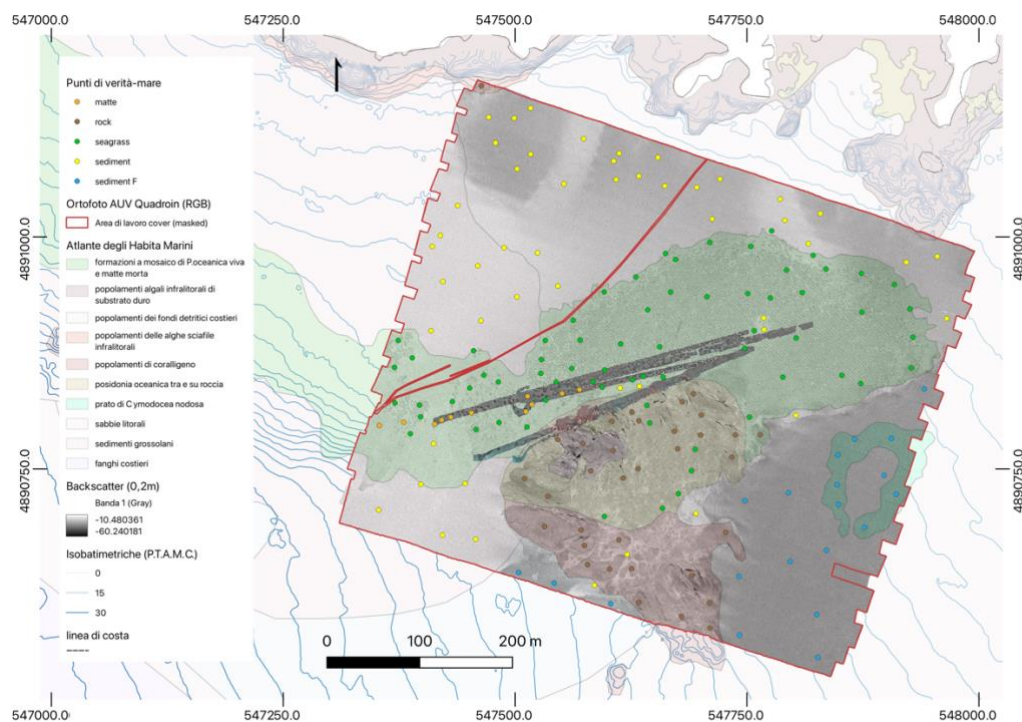


Figura 5. Punti di addestramento per il classificatore dell'area di Levanto. Nello sfondo: backscatter del multibeam (dB), atlante degli habitat (Coppo et al. 2020), ortofoto acquisite con Quadroin. Coordinate WGS84 / UTM N32.

3.6 Flusso di analisi GEOBIA

La mappatura degli habitat è stata ottenuta mediante l'identificazione delle classi con un sistema di apprendimento automatico (*Machine Learning*, siglato ML). In particolare, è stato adoperato un processo di classificazione supervisionata basato su algoritmi di Geographic Object-Based Image Analysis (analisi delle immagini geografiche basata sulla creazione di oggetti, GEOBIA). Questo sistema di classificazione automatizzato permette di superare il limite delle analisi delle firme acustiche basate sui valori puntuali (i singoli pixel), includendo la variabilità spaziale nella distinzione topologica delle classi tematiche analizzate (Rende et al., 2020). Questo passaggio incrementa esponenzialmente lo spazio dimensionale delle variabili che descrivono la 'firma' acustica degli habitat indagati. L'utilizzo di modelli statistici di apprendimento automatico per permettere di identificare queste firme complesse non generalizzabili, superando i limiti delle tecniche di analisi classiche, basate su logiche deterministiche.

La classificazione supervisionata è stata eseguita utilizzando esclusivamente i dati di batimetria e *backscatter* del multibeam. La supervisione del classificatore è derivata dall'impiego di punti di addestramento e validazione, individuati e classificati da operatore esperto dall'interpretazione del dato multibeam affiancato dalla cartografia bionomica di riferimento (si veda il paragrafo 3.5).

L'utilizzo di punti non preclude i dati di addestramento del contenuto informativo spaziale, poiché, nel processo di analisi GEOBIA, l'informazione sulla variabilità locale viene elaborata dall'algoritmo di classificazione sull'oggetto in cui ricade il punto di addestramento. Tale oggetto corrisponde all'intorno topologico del punto di addestramento, e si assume essere rappresentativo della variabilità spaziale dell'habitat che ricopre. Tale assunzione si sorregge sul metodo di segmentazione, fondato nella analisi dell'eterogeneità spaziale (si veda il paragrafo 3.6.1).

L'analisi GEOBIA è stata eseguita con il software eCognition Developer, versione 10.5. In ragione alla terminologia specifica del programma utilizzato, definiamo *rule set* l'insieme degli algoritmi che costituiscono il flusso completo di analisi, dai dati di ingresso alla carta classificata. I tre processi principali sono la creazione degli oggetti, la classificazione supervisionata e la rifinitura del risultato (Figura 6). Dal *rule set* iniziale sono state sviluppate molteplici versioni. Queste mantengono gli stessi algoritmi per i processi principali, e differiscono solamente nei parametri utilizzati per la classificazione.

In tutti i progetti, corrispondenti ciascuno ad una area di studio, si è scelto di mantenere la stessa denominazione dei layer, per garantire il trasferimento diretto degli algoritmi nell'analisi delle diverse aree (Tabella 1). In Allegato.1 sono riportati tutti gli algoritmi ed i relativi parametri della versione Originale. Tra i molteplici algoritmi integrati, la segmentazione (*multiresolution segmentation*) e la classificazione (*supervised classification*) costituiscono i due processi chiave della analisi GEOBIA.

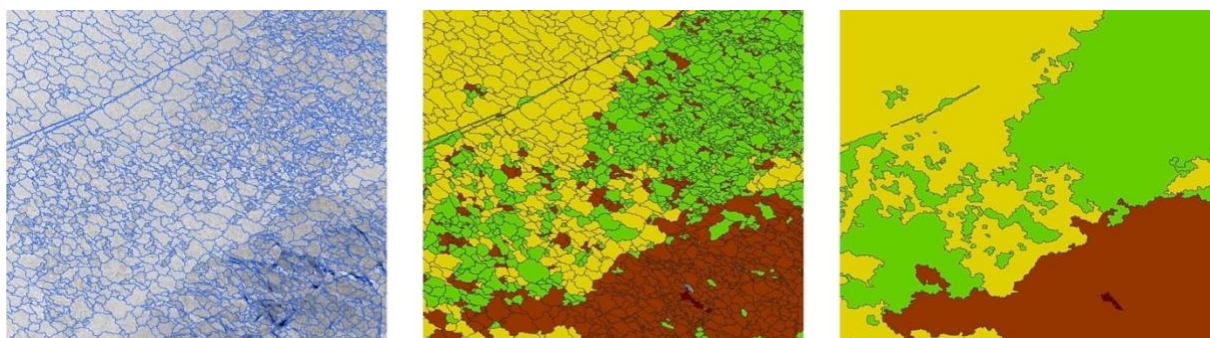


Figura 6. Le tre fasi del processo di analisi basato su oggetti delle immagini geografiche (GEOBIA), particolare dell'area di Levanto. In a) segmentazione dell'immagine in oggetti poligonali (nello sfondo il layer di backscatter); b) classificazione degli oggetti (sediment in giallo, seagrass in verde, rock in marrone); c) rifinitura di post-processamento.

Tabella 1. Dati di ingresso per la classificazione supervisionata in eCognition.

Dato raster	Nome del layer	Dato vettoriale	Nome del layer
Batimetria (DTM)	BATHY	Area di lavoro	cover
Backscatter	BS	punti di addestramento	samp_train
Aspect	A	punti di validazione	samp_val
Slope	S		
General Curvature	GC		
Total Curvature	TC		
TRI	TRI		
Eastness	E		
Northness	N		

3.6.1 Segmentazione e validazione degli oggetti immagine

L'applicazione del paradigma GEOBIA su dati ad alta risoluzione richiede la definizione di un'unità spaziale di base che non corrisponda necessariamente all'estensione complessiva della unità bionomica indagata (la patch ecologica), bensì a un oggetto primitivo (*superpixel*) in grado di catturare la variabilità intrinseca (texture) dell'habitat (Diesing et al., 2016). Affinché la classificazione sia efficace, la segmentazione deve generare oggetti che costituiscano tasselli

elementari rappresentativi della texture propria della classe in cui ricadono: sufficientemente ampi da ricoprire l'eterogeneità fisiologica del segnale, ma confinati all'interno dei limiti dell'habitat per evitare l'inclusione di ecotoni, che comporterebbe un incremento esponenziale della varianza intra-oggetto (Ismail et al., 2015; Drăguț et al., 2014).

La creazione degli oggetti in analisi GEOBIA viene definita *segmentazione*. Nel presente studio, è stato utilizzato l'algoritmo di segmentazione multiscala (*multiresolution segmentation*), che crea gli oggetti attraverso un processo iterativo di accorpamento dei pixel e oggetti adiacenti reciprocamente più simili, in base al criterio di omogeneità. Il processo di "crescita" degli oggetti per accorpamento si interrompe al raggiungimento della soglia massima per il criterio di omogeneità, definita dal parametro di scala. Il criterio di omogeneità è calcolato in funzione di tre componenti: eterogeneità del colore (*color*), *smoothness* (levigatezza) e compattezza (*compactness*) degli oggetti. L'eterogeneità del colore è definita come sommatoria delle deviazioni standard dei valori di ogni layer considerato, pesate con i parametri *image layer weights*. La *smoothness* è definita come rapporto tra lunghezza del perimetro dell'oggetto e lunghezza del perimetro del rettangolo circoscritto (*bounding box*). La compattezza è determinata dal rapporto tra perimetro e area dell'oggetto. Il parametro *shape parameter* regola il contributo percentuale delle componenti di forma (*smoothness* e compattezza) nel criterio di omogeneità, rispetto al contributo residuo attribuito al colore. Il parametro *compactness* regola il contributo percentuale della compattezza in rapporto alla *smoothness*. Il criterio di omogeneità, adottato dall'algoritmo nella regolazione del processo iterativo di crescita degli oggetti, è sintetizzato nella equazione 9 (rielaborazione dalla documentazione ufficiale del software eCognition, Trimble).

$$(1 - a) \left[\sum_{i \in N} (\sigma_{ij} w_i) \right] + a \left\{ b \left[\frac{\text{perimetro}_j}{\text{area}_j} \right] + (1 - b) \left[\frac{\text{perimetro}_j}{\text{perimetro}_{B(j)}} \right] \right\} \leq c \quad [\text{eq. 9}]$$

dove:

a = shape

b = compactness

c = scale parameter

σ_{ij} = deviazione standard del valore di banda del layer i nei pixel dell'oggetto j

w_i = peso del layer i

N = numero di layer

$B(j)$ = *bounding box* dell'oggetto j

Nel presente lavoro, sono stati utilizzati i valori riportati in Allegato 1 per l'algoritmo *multiresolution segmentation*, in tutte le versioni del *rule set* ad eccezione del parametro di scala,

che differisce nella versione Rev10 (*scale parameter* = 10) rispetto a tutte le altre versioni (*scale parameter* = 12).

Al fine di validare la segmentazione prodotta a seguito dell'incremento del parametro di scala, ed escludere fenomeni di sovradimensionamento o sottodimensionamento spaziale degli oggetti immagine, è stata estratta la varianza ponderata spazialmente (Area-Weighted Variance, WV). Questa metrica viene calcolata in ambiente R sull'intero mosaico generato in eCognition, per ciascuna area. Essa fornisce la misura globale dell'eterogeneità intra-oggetto delle feature, pesata per l'estensione areale degli oggetti, secondo la formula definita da Espindola et al. (2006) che qui viene applicata ai due layer impiegati nella segmentazione (equazione 10).

$$WV_{\text{combinato}} = \frac{\sum_{i=1}^n [a_i \cdot (\sigma_{BS,i}^2 + \sigma_{TRI,i}^2)]}{\sum_{i=1}^n a_i} \quad [\text{eq. 10}]$$

dove a_i rappresenta l'area dell'oggetto i -esimo e σ_i^2 la varianza dei pixel al suo interno.

Poiché la minimizzazione assoluta della WV collaserebbe inevitabilmente alla scala del singolo pixel (Espindola et al., 2006), il suo utilizzo operativo non è mirato alla ricerca del valore minimo, bensì alla verifica della stabilità topologica del modello. Fissata l'architettura algoritmica, la stazionarietà dei valori di WV tra il sito pilota e le aree di trasferimento dimostra che il parametro di scala non subisce derive morfologiche. Il calcolo certifica oggettivamente che gli oggetti generati si mantengono costantemente al di sotto della soglia spaziale di miscelazione degli habitat in contesti ambientali differenti.

3.6.2 Sviluppo del modello di classificazione supervisionata

Il processo di classificazione supervisionata con software eCognition è stato messo a punto attraverso uno sviluppo iterativo volto ad ottenere una ottimizzazione progressiva del risultato nell'area pilota di Levanto. Attuando una specializzazione del processo di classificazione sito-specifica sul dato dell'area di Levanto, si intende ricercare il massimo dettaglio morfologico interpretabile dai soli dati multibeam ad alta risoluzione adoperati.

Il modello random forest messo a punto nella versione Rev10 del *rule set*, nell'ambito del progetto MER per la classificazione del dato multibeam, è stato adottato come sistema di partenza. Si tratta di un algoritmo random forest con configurazione canonica. Lo spazio delle feature è costituito da un set di valori di media e deviazione standard estratti da sei dei nove layer di ingresso, per un totale di otto feature: le medie BATHY, BS, TRI, TC, S e le deviazioni standard di TRI, GC ed S. Questo modello di classificazione è stato impiegato per l'estrazione delle tre classi di base utilizzate nel progetto (*seagrass*, *rock*, *sediment*). Tuttavia, questo modello genera classificazioni con livelli di accuratezza subottimali dovuti a parziale

misclassificazione nelle zone di transizione ecotonale, corrispondenti particolarmente alle aree di mosaico di prateria di *P. oceanica* viva e matte morta e della zona rocciosa con *P. oceanica* tra e su roccia (informazioni desunte dall'atlante degli habitat, Coppo et al., 2020). Al fine, dunque, di ampliare gli obiettivi bionomici, integrando una classe specifica per la matte morta, si è attuata una ridefinizione del modello di classificazione volta a massimizzarne accuratezza e sensibilità nel dato dell'area di Levanto.

3.6.2.1 Costruzione dello spazio delle feature

La selezione dello spazio delle feature ottimale è stata eseguita in ambiente RStudio, attraverso una sequenza di analisi statistiche sulla capacità di predizione di ciascuna feature rispetto alle tre classi principali. Il dataset di allenamento è costituito dai valori delle feature negli oggetti contenenti i punti di addestramento e validazione classificati come seagrass, rock o sediment, con i valori incorporati sulla segmentazione propria del *rule set* Rev10. Lo spazio di partenza scelto include tutte le feature di media e deviazione standard per batimetria, backscatter e ciascuna derivativa, con altre tre variabili della geometria degli oggetti: area (pixel), compattezza, lunghezza (pixel). Su questo dataset è stata eseguita una conversione preventiva delle variabili circolari in indici di dispersione cartesiani (con trasformazione trigonometrica seno e coseno). Al fine di garantire efficienza computazionale e interpretabilità del modello, si è scelto di mantenere un numero limitato di feature come condizione di base.

Il primo approccio esplorativo adottato è il test Boruta, una analisi statistica dell'informazione delle variabili predittive basata su algoritmo random forest (Kursa e Rudnicki 2010). Questo algoritmo permette di isolare i descrittori veicolanti un segnale statisticamente superiore al rumore casuale di distribuzione corrispondente (shadow features).

Per minimizzare la collinearità tra i predittori, si applica una analisi di correlazione al set di variabili risultate informative con il test Boruta. Si è stabilita la soglia di correlazione $r < 0.9$, oltre la quale è stata eliminata una feature per coppia di correlate, secondo il criterio di preferenza di minima importanza, estratto dall'algoritmo Boruta, fino ad ottenere un gruppo di variabili con correlazione ≥ 0.9 totalmente disaccoppiate.

Si applica in successione la Recursive Feature Elimination (RFE) per l'identificazione del sottoinsieme ottimale di feature. Questo algoritmo opera nuovamente innestato su un modello random forest, ma con criterio di selezione di tipo minimal-optimal, con numerosità determinata a priori. Per mantenere lo spazio delle feature con numero di dimensioni comparabile al set della versione Rev10, è stata fissata l'indagine della curva nell'intorno di $N=8$.

Infine, si esegue anche una analisi per la risoluzione della multicollinearità con metodo LASSO (*Least Absolute Shrinkage and Selection Operator*). La penalizzazione di tipo L1 comporta una esclusione dei predittori collineari, secondo il valore dei coefficienti beta di regressione logistica multinomiale.

3.6.2.2 Parametri adottati nel classificatore

Nell’algoritmo di classificazione supervisionata implementato su eCognition è stato utilizzato il metodo “Random Tree”, che si basa sulla architettura dei modelli random forest. Questo metodo, applicato sugli oggetti generati dalla segmentazione, esegue una classificazione attraverso l’apprendimento d’insieme (*ensemble learning*) basato sugli alberi decisionali (si veda il paragrafo 1.3). Per risolvere l’instabilità classificatoria, con l’obiettivo di massimizzare l’accuratezza del modello mantenendo lo spazio dimensionale il più contenuto possibile, è stata eseguita la selezione mediante analisi statistica delle feature più informative non collineari (di cui al paragrafo precedente). L’insieme delle otto feature così ottenuto (denominato “set corretto”) è stato dunque adoperato imponendo la valutazione simultanea dell’intero spazio delle feature in ogni albero decisionale. Questa modifica del modello è stata ottenuta attraverso la saturazione del parametro *active variables*, impostato a 10 in modo da eludere il sottocampionamento casuale delle feature ai nodi (*feature randomness*). Inoltre, sono stati rimossi i vincoli di pre-potatura (nessun limite di profondità degli alberi, con numero minimo di un campione per nodo). Per stabilizzare la stima e ridurre la varianza dell’ensemble, il numero massimo di alberi è stato portato a 500 (si veda l’elenco dei parametri in Allegato 1). In questo modo, si adopera un modello di apprendimento di insieme basato sulla votazione estratta da una foresta di 500 alberi eterogenei, grazie alla randomizzazione del campionamento con reimmissione (*bootstrap*), decorrelati e ad elevato contenuto informativo, grazie all’ottimizzazione dello spazio delle feature (si veda paragrafo 3.6.2.1).

La classificazione è stata sottoposta ad un duplice livello di validazione. Una prima validazione interna è stata condotta utilizzando l’insieme indipendente di punti di validazione separato dal dataset di addestramento (si veda il paragrafo 3.5), al fine di stimare l’accuratezza locale della classificazione. Una seconda validazione esterna è stata eseguita applicando la stessa procedura di classificazione alle altre aree di studio, effettuando in ciascun caso l’addestramento del modello sui rispettivi punti locali, al fine di verificare la robustezza del metodo in contesti geografici differenti.

3.6.3 Post-processamento della classificazione

La classificazione è stata infine rifinita con la aggregazione degli oggetti immagine in patch di habitat omogenee e la rimozione delle inclusioni spurie. Quest'ultima si è resa necessaria poiché le diverse versioni del classificatore sono rimaste caratterizzate dalla presenza di singoli oggetti interspersi all'interno di aree di classe diversa, generando frammentazione spaziale.

L'unione dei poligoni adiacenti assegnati alla stessa classe tematica è stata eseguita con algoritmo *merge region*, questo passaggio restituisce la generalizzazione cartografica dei singoli oggetti classificati in aree unitarie del mosaico ecologico, corrispondenti alle patch di habitat bentonici.

La rimozione della frammentazione spaziale, dovuta del rumore stocastico locale che ha generato artefatti di classificazione del dato ad alta risoluzione, si è resa necessaria per aumentare la leggibilità del risultato ad una scala coerente con l'estensione dell'indagine e con le dinamiche strutturali delle patch di habitat bentonici (Blaschke, 2010; Diesing et al., 2016; Rende et al., 2020). È stato impiegato l'algoritmo *grow region* che fonde nella matrice circostante gli oggetti circondati da una unica classe diversa dalla propria. Il riassorbimento degli oggetti interspersi è stato limitato da una soglia dimensionale degli oggetti, che rappresenta la dimensione dell'unità minima mappabile. Sono stati fissati due limiti: 500 pixel per il riassorbimento delle patches di sabbia e roccia in prateria o matte, e per il riassorbimento della sabbia in roccia; 200 pixel per il riassorbimento delle classi seagrass e matte, e per la roccia su sabbia (si vedano gli algoritmi "grow region" in Tabella 4). Questa differenziazione garantisce un risultato più conservativo per gli habitat di maggiore interesse nel presente studio, le praterie di *P. oceanica* e la matte morta.

Si sottolinea che la matrice di errore e le metriche di accuratezza del modello di classificazione sono state calcolate sul risultato diretto del classificatore, prima della rifinitura.

3.7 Disegno sperimentale

Partendo da questo modello come baseline, è stato attuato un processo di ottimizzazione del risultato, con l'obiettivo di migliorare l'accuratezza del risultato nell'area di Levanto. Al fine di garantire efficienza computazionale e interpretabilità del modello, si è scelto di mantenere un numero limitato di feature come condizione di base.

Il modello con architettura di tipo decision tree bagging è stato adottato in tre versioni di *rule set* per valutare l'efficacia delle feature. La versione Originale mantiene il set di feature identico al modello iniziale del *rule set* Rev10, con il modello di classificazione modificato. La versione

Corretto mantiene rigorosamente il set di feature ottenuto con la analisi statistica, senza il reintegro nella media della batimetria. Infine, il *rule set* Ottimizzato include quest'ultima feature, come unica differenza rispetto alla versione Corretto. In tabella 2 sono riportate le caratteristiche dei diversi *rule set* adoperati.

Il modello di classificazione costruito sull'area pilota di Levanto è stato sottoposto a verifica della robustezza mediante applicazione sulle altre aree di studio. La validazione esterna con inferenza spaziale risulta necessaria per controllare l'overfitting metodologico che può verificarsi con il modello decision tree bagging. La trasposizione dell'architettura ha previsto il riaddestramento locale, per isolare l'effetto di overfitting dell'architettura del modello. In questo modo si previene l'introduzione di errore derivante dai valori assoluti dei layer di ingresso.

Tabella 2. Modello e spazio delle feature di ciascun rule set.

Rule set	Rev10	Originale	Corretto	Ottimizzato
Modello	random forest	decision tree bagging	decision tree bagging	decision tree bagging
Feature	Mean BATHY	Mean BATHY	Mean BS	Mean BS
	Mean BS	Mean BS	Mean E	Mean E
	Mean TRI	Mean TRI	Mean GC	Mean GC
	Mean TC	Mean TC	SD BATHY	SD BATHY
	Mean S	Mean S	SD BS	SD BS
	SD GC	SD GC	SD TRI	SD TRI
	SD S	SD S	SD N	SD N
	SD TRI	SD TRI	Area	Area
				Mean BATHY

Il modello decision tree bagging, nelle versioni Originale e Ottimizzato, è stato utilizzato per ampliare il numero di classi tematiche con le aggiuntive *matte* e sediment F. Quest'ultima classe, per il sedimento grossolano con firma acustica di riflettività molto più bassa (colorazione scura del backscatter), è stata riscontrata solo nelle aree di Levanto e Genova 19.

La combinazione di prove elaborate per le diverse aree è schematizzata in Tabella 3. Abbiamo un disegno sperimentale a tre fattori (*rule set*, numero di classi, sito) sbilanciato, con il fattore classi gerarchizzato nell'interazione sito – *rule set*, i quali sono ortogonali tra loro. Questo disegno incompleto e privo di repliche indipendenti deriva dalla eterogeneità dei dati disponibili, per i quali è stato possibile adoperare la classificazione in 5 classi nelle sole aree Genova_19 e Levanto, poiché la quinta classe (sediment F) non è risultata presente nelle aree restanti. Ai fini del presente studio, avendo a disposizione i dati delle aree di studio selezionate appositamente

e senza alcuna randomizzazione spaziale, si verifica la specificità dell’algoritmo di classificazione sviluppato nell’area di Levanto, confrontando i diversi livelli del fattore sito (le cinque aree di studio), specificando la mancanza di repliche indipendenti necessarie per la eventuale generalizzazione spaziale dei risultati.

Tabella 3. Schema riepilogativo delle classificazioni elaborate.

Levanto		Genova_03		Genova_19		Genova_20		Sar_W_62	
Rule set	Classi	Rule set	Classi	Rule set	Classi	Rule set	Classi	Rule set	Classi
Rev10	3C	Rev10	3C	Rev10	3C	Rev10	3C	Rev10	3C
Originale	3C	Originale	3C	Originale	3C	Originale	3C	Originale	3C
Originale	4C	Originale	4C	Originale	5C	Originale	4C	Originale	4C
Originale	5C	Corretto	3C	Corretto	3C	Corretto	3C	Corretto	3C
Corretto	3C	Ottimizzato	3C	Ottimizzato	3C	Ottimizzato	3C	Ottimizzato	3C
Ottimizzato	3C	Ottimizzato	4C	Ottimizzato	5C	Ottimizzato	4C	Ottimizzato	4C
Ottimizzato	4C								
Ottimizzato	5C								

3.8 Valutazione delle performance dell’algoritmo di classificazione

Le prestazioni del workflow analitico sono valutate attraverso l’analisi quantitativa delle metriche di accuratezza. Nel contesto di analisi GEOBIA, vengono adottati molteplici approcci di validazione della classificazione, dipendenti dalla natura della verità a mare. In presenza di dati di validazione di tipo areale, è affermata l’applicazione di analisi spaziali, che permette di quantificare in maniera esplicita l’errore associato alla dimensione degli oggetti classificati (Whiteside et al. 2014). Tuttavia, con dati puntuali non è possibile attribuire il dato classificato alla intera estensione degli oggetti poligonali, ed è quindi comune l’utilizzo di matrici di errore tradizionali, basate sui conteggi delle corrispondenze (Congalton e Green, 2009). Nel presente studio si adotta il metodo proposto da Olofsson et al. (2014), con matrici basate sui conteggi pesate per l’estensione totale delle classi nell’area cartografata. Questo approccio permette di bilanciare la validazione asimmetrica punto–poligono, che è soggetta a rischio di sovrastima dell’errore nei casi in cui l’estensione spaziale delle classi sia fortemente eterogenea.

L’incrocio spaziale tra le topologie predittive (gli oggetti) ed il set di punti di validazione, è stato eseguito previo adattamento dell’interpretazione delle classi tematiche del dato di validazione (sempre in cinque classi) in sintonia con il processo di classificazione: per le prove con 3 e 4 classi, si è applicato il medesimo accorpamento delle classi aggiuntive all’interno di *sediment* (*sediment F e matte*). Ne consegue che la classe nominale *sediment*, rappresenta esclusivamente

il sedimento non fine solo nelle classificazioni con 5 classi. Nelle prove con 3 classi, invece, “sediment” rappresenta l’insieme di sedimento grossolano, sedimento fine e mattoni morti, in coerenza con la classificazione attualmente in corso di svolgimento nell’ambito del progetto di mappatura.

Le matrici di errore di ciascuna prova con la pesatura areale sono state generate in ambiente R, mediante il pacchetto *sf*, utilizzando per ciascuna prova le tabelle attribuite dei file vettoriali prerifinitura spazialmente incrociati con i punti di validazione indipendenti. Si è proceduto alla costruzione delle tabelle di contingenza incrociata, e al calcolo dell’accordo tra predizioni e osservazioni lungo la diagonale principale. Sono stati applicati quindi i pesi areali W_i ad ogni conteggio, per convertire la matrice di contingenza in matrice delle proporzioni di area (Olofsson et al., 2014) secondo l’equazione 11:

$$p_{ij} = W_i \cdot \frac{n_{ij}}{n_i} = \left(\frac{A_i}{A_{tot}} \right) \cdot \left(\frac{n_{ij}}{n_i} \right) \quad [\text{eq. 11}]$$

dove n_{ij} è il punteggio della cella ij , e n_i rappresenta il totale marginale di riga.

Dalla matrice delle proporzioni di area sono state estratte le metriche di accuratezza per la valutazione della classificazione, secondo le formule definite da Olofsson et al. (2014). Si ottengono quindi la metrica di Overall Accuracy (accuratezza globale), che rappresenta la proporzione stimata dell’intera area di studio che è stata classificata correttamente; e le metriche di User Accuracy (affidabilità dell’utente, complementare all’errore di commissione), e Producer Accuracy (sensibilità del modello, complementare all’errore di omissione). Quest’ultima metrica di accuratezza, diversamente dal calcolo convenzionale, subisce una severa penalizzazione in ragione inversa alla copertura percentuale della classe in oggetto, diventando una metrica sensibile all’errore di omissione per le classi rare.

In rispondenza alla metodologia di Olofsson et al. (2014), sono stati calcolati anche gli intervalli di confidenza al 95% ($\pm 1,96 \times SE$), dove gli errori standard sono calcolati sulla varianza ponderata spazialmente.

L’assunzione di distribuzione normale è sostenuta dal teorema del Limite Centrale, in presenza di un numero di punti di validazione superiore a 30 in tutti i set di dati. Per le metriche classe-specifiche, PA e UA, la possibile asimmetria degli intervalli per le classi minoritarie è controllata fissando a 1 il limite superiore.

In presenza di un disegno sperimentale sbilanciato e privo di replicazione, non si applicano analisi della varianza sulle metriche di accuratezza; tuttavia, l’errore standard calcolato permette di eseguire test inferenziali di confronti a coppie. È stato quindi adoperato il test Z per il confronto di proporzioni indipendenti. I confronti statistici a coppie sono stati eseguiti isolando

specifiche combinazioni dei fattori di variazione, al fine di valutare l'effetto del sito geografico e della risoluzione tematica (numero di classi) prendendo come modello di riferimento il *rule set* Ottimizzato.

3.9 Confronto geospaziale multi-sorgente

La validazione del risultato cartografico rappresenta un passaggio fondamentale per stimare l'effettiva corrispondenza tra l'output predittivo del modello e i pattern spaziali reali. Tale verifica si rende stringente avendo effettuato la supervisione del processo di classificazione mediante dati di addestramento derivati da interpretazione esperta del medesimo segnale acustico impiegato in fase di addestramento.

Al fine di quantificare l'affidabilità delle carte prodotte, il presente studio ha integrato i molteplici livelli informativi spaziali disponibili per le aree indagate, adottando un approccio di validazione incrociata. Ai dati puntuali di validazione da AUV, assunti come riferimento di microscala, sono state affiancate le classificazioni da immagine satellitare multispettrale (in corso di elaborazione nell'ambito del medesimo progetto ISPRA) e le cartografie di riferimento di pubblica disponibilità. Queste ultime sono rappresentate dall'atlante degli habitat (Coppo et al. 2020), per le aree nel Mar Ligure, e, per l'area Sar_W_6, dallo strato informativo *Tipologie degli ecosistemi marini* del portale EcoAtlante (ISPRA, s.d.), consultato tramite servizio web ArcGIS (SINAccloud).

Tale approccio di confronto multi-sorgente ha reso necessaria l'adozione di metodi analitici differenziati in funzione alla natura del dato. Il flusso di analisi geospaziale, eseguito in ambiente RStudio, ha previsto l'armonizzazione dei sistemi di riferimento e la delimitazione del dominio (tramite ritaglio geometrico delle aree di sovrapposizione o esclusione dei campioni puntuali ricadenti all'esterno del dato acustico). A seguito dell'intersezione spaziale, le metriche di concordanza sono state calcolate strutturando due differenti tipologie di matrici di confusione, in funzione della natura del dato di validazione.

Nel caso del grigliato puntuale AUV, la validazione si fonda su una classica matrice di confusione basata sulle frequenze assolute di intersezione esatta (point-in-polygon) tra il dato di validazione e la classe predittiva sottostante. In conformità con gli standard di valutazione dell'accuratezza tematica per la cartografia bionomica (Brown et al., 2011; Diesing et al., 2016; Rende et al., 2020), non sono state applicate tolleranze geometriche a priori per compensare l'incertezza posizionale intrinseca del mezzo di campionamento AUV (stimata in 2 m), la cui

valutazione è demandata all'analisi critica degli errori di omissione e commissione ai margini ecotonali.

Per i mosaici cartografici continui, sono state calcolate le matrici di congruenza areale (area-based error matrix), quantificando le aree di intersezione spaziale assoluta per singola classe.

3.9.1 Armonizzazione tematica

Poiché le diverse carte prese in esame contengono una eterogeneità tematica delle classi bentoniche, per poter attuare il confronto di corrispondenza spaziale e semantica è necessario definire il minimo comune denominatore che rappresenti in maniera univoca e coerente le classi confrontabili. La notevole diversità di risoluzione tematica che presentano le carte in esame è dipendente dalla natura del segnale (acustico od ottico) e dalla risoluzione nativa del sensore con cui è stato acquisito il dato. Ne consegue la necessità di attuare una degradazione semantica controllata, per le sorgenti ad alta risoluzione, fino al raggiungimento di una struttura coerente con le classi adottate nella classificazione del multibeam.

Al fine di non comprimere l'eterogeneità tematica delle diverse sorgenti di confronto con il MBES a un'unica e drastica riduzione gerarchica, si adotta un disegno fattoriale asimmetrico per la validazione incrociata in rapporto 1:1 di ciascuna carta con la classificazione multibeam (Tabella 4).

Tabella 4. Quadro del numero di corrispondenze tra la classificazione (MBES) e le classi presenti nei quattro dati di confronto (classificazione da satellite, classificazione in grigliato da AUV, EcoAtlante (ISPRA, s.d.) e l'atlante degli habitat (Coppo et al., 2020).

MBES	Satellitare	AUV	EcoAtlante	Atlante degli Habitat Marini
seagrass	1 (Biocenosis of <i>P. oceanica</i>)	1 (Posidonia)	1 (A5.535)	4 (pos-roc, mos, pos, cym)
matte	1 (Marine benthic habitats)	–	1 (A5.5353)	1 (mos)
sediment	1 (//)	3 (“0”, “1”, “2”)	2 (A5.13, A5.23)	4 (s, f, sgc, dc)
sediment F	1 (//)	1 (0 sabbia fine)	1 (A5.23)	4 (s, f, sgc, dc)
rock	1 (//)	1 (4 roccia o ghiaie)	[Non presente]	3 (c, asi, asf)

La disponibilità di versioni in 3, 4 e 5 classi dei risultati ottenuti in eCognition rappresenta uno spettro di possibilità che offre maggiore elasticità nella definizione del minimo comune denominatore tematico. Si sceglie in questa prospettiva di confrontare i diversi dati di validazione con le versioni della classificazione più direttamente confrontabili, per garantire la

minima degradazione semantica necessaria. In particolare, il confronto con il mosaico derivato dal dato satellitare si limita alla versione in 3 classi, poiché l'unica congruenza individuabile è quella della classe seagrass con 'Biocenosi di *P. oceanica*', aggregando le restanti. Diversamente, nel confronto con le carte bionomiche degli ecosistemi marini, è possibile mantenere la distinzione in 5 classi coerenti con le tipologie di habitat marini, standardizzate secondo il sistema di classificazione europeo EUNIS, presenti nell'area Sar_W_62 (Tabella 4). Riguardo all'attribuzione delle corrispondenze con la classe matte si è scelto un criterio conservativo basato su corrispondenze semantiche multiple (relazioni non biunivoche), escludendo le associazioni non supportate da corrispondenza esplicita (Tabella 5 e 6).

Tabella 5. Quadro delle conversioni multiple adottate per ogni classe dell'AUV verso le classi tematiche adottate nella classificazione del dato multibeam.

AUV	0 sand (fine ripple)	1 sand (coarse ripple)	2 sand (with organic material)	3 other sediment (gravel)	4 rocks (and coarse gravel)	5 Posidonia
3C	sediment	sediment	sediment	sediment	rock	seagrass
4C	sediment / matte	sediment / matte	sediment / matte	sediment / matte	rock	seagrass
5C	sediment / matte / sediment F	sediment / matte / sediment F	sediment / matte / sediment F	sediment / matte	rock	seagrass

Tabella 6. Quadro delle conversioni multiple adottate per il dato di confronto da satellite e la cartografia bionomica preesistente verso le classi tematiche adottate nella classificazione del dato multibeam. Atlante Habitat fa riferimento a Coppo et al. (2020).

Satellite	Biocenosis of <i>Posidonia oceanica</i>						Marine benthic habitats				
3C	seagrass						rock / sediment				
4C	seagrass						rock / sediment / matte				
5C	seagrass						rock / sediment / matte / sediment F				
Atlante Habitat	pos-roc	mos	c	asi	af	pos	cym	s	f	sgc	dc
3C	seagrass / rock	seagrass / sediment	rock	rock	rock	seagrass	seagrass	sediment	sediment	sediment	sediment
4C	seagrass / rock	seagrass / matte	rock	rock	rock	seagrass	seagrass	sediment	sediment	sediment	sediment
5C	seagrass / rock	seagrass / matte	rock	rock	rock	seagrass	seagrass	sediment / sediment F	sediment / sediment F	sediment / sediment F	sediment / sediment F
EcoAtlante	Coarse & mixed sediment			Sand		Muddy sand			Posidonia oceanica		
3C	sediment			sediment		sediment			seagrass		
4C	sediment			sediment		sediment			seagrass		

3.10 Stima del limite inferiore delle praterie di *Posidonia oceanica*

La mappatura ad altissima risoluzione dei margini della prateria tramite sistemi acustici ha dimostrato una precisione paragonabile ai rilevamenti in situ (Rende et al., 2020). Tuttavia, la derivazione cartografica di questo descrittore paesaggistico, tipicamente misurato *in situ*, espone la misurazione ad errore dovuto della risoluzione del dato cartografico, che, in presenza di margini frammentati o sfumati, può generare marcate discrepanze rispetto alla determinazione diretta. Come evidenziato dai recenti lavori di Ventura et al. (2023) sull'altissima risoluzione, e in accordo con i fondamenti dell'analisi GEOBIA descritti da Blaschke (2010), l'individuazione di un limite ecologico da un classificatore non può prescindere da filtri spaziali (MMU) e statistici (percentili).

La metodica raccomandata dagli standard cartografici RAC/SPA (UNEP/MAP - RAC/SPA, 2015) prevede l'adozione di un *minimum mapping unit* (MMU), ovvero di una unità dimensionale cartografica, che permette di semplificare l'interpretabilità del dato sulla base di un criterio dimensionale esplicito. Nel presente lavoro è stata adottata in fase di rifinitura (post-processamento) una soglia MMU di 200 pixel (pari a 8 m²) mediante l'algoritmo *grow region*. In particolare, è stata selezionata una soglia minima duplice: di 200 pixel per i poligoni classificati come seagrass, matte, o roccia su sedimento, e di 500 pixel (20 m²) per i poligoni di roccia o sedimento su seagrass, matte, rock (si veda in Allegato 1 la parametrizzazione dell'algoritmo). In questo modo, si identifica il limite inferiore con la profondità massima a cui si trova la classe seagrass, avendo oggetti immagine già filtrati ad una soglia ecologicamente significativa.

In aggiunta, si adotta l'approccio geostatistico basato sui percentili sui dati di profondità (statistiche zonali), per calcolare un limite inferiore più conservativo, escludendo potenziali outlier e falsi positivi del classificatore. Pertanto, si sovrappone il DTM ai poligoni classificati e rifiniti (con algoritmi *merge* e *grow*) e si calcola la profondità massima relativa alla classe seagrass e la profondità al il 99° percentile della distribuzione batimetrica (Steward et al., 2005).

RISULTATI E DISCUSSIONE

4.1 Ottimizzazione dello spazio delle feature

Il primo processo di selezione del set di feature è stato l'algoritmo Boruta. L'analisi dell'importanza di tutte le feature incluse nel set iniziale ha restituito un quadro conservativo, in cui quasi la totalità delle variabili risulta avere una importanza marcatamente superiore rispetto alle variabili ombra (costituite dal rimescolamento casuale dei dati delle corrispondenti feature originali). Le uniche feature risultanti prive di importanza sono compactness, ovvero il rapporto geometrico tra perimetro e area degli oggetti (si veda il paragrafo 3.6.1) e la media dell'esposizione scomposta nella componente est (Mean_A_eastness). Il risultato mostrato in Figura 7 rappresenta una condizione di relativa uniformità nella distribuzione dell'importanza nello spazio complessivo delle 21 variabili analizzate.

Questo risultato indica la presenza di uno spazio delle feature adeguatamente informativo in riferimento ad i punti di addestramento analizzati, e permette l'esclusione delle sole due feature non informative dal set.

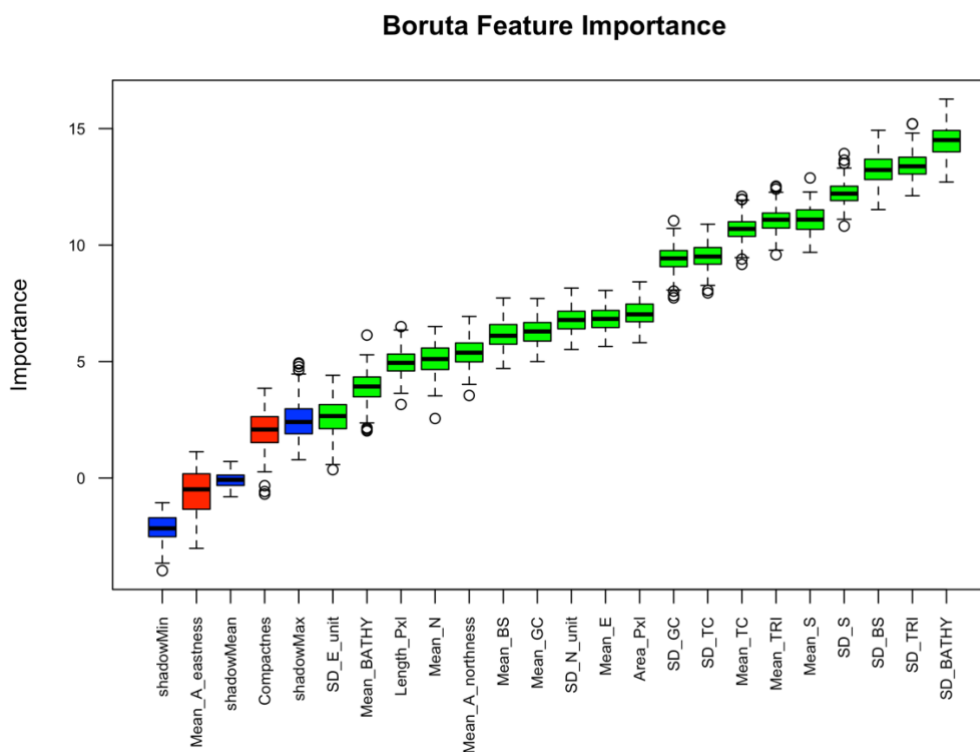


Figura 7. Diagramma a scatola e baffi dell'importanza delle feature calcolato con Boruta test. In rosso sono evidenziate le feature non importanti, in blu le variabili ombra.

Il set di feature ripulito delle due variabili ombra identificate è stato sottoposto ad analisi di correlazione con coefficiente lineare r di Pearson, che ha restituito un quadro di elevata collinearità tra le feature, con 17 coppie con correlazione $|r| > 0,9$. La correlazione di 0,993 tra la media di Slope e di TRI rappresenta una condizione elevatissima collinearità tra le due variabili. La perfetta sovrapposizione nell'informazione fornita da queste due feature rende inoltre meno stabile i valori di importanza calcolati dal test Boruta, che è innestato su modello random forest.

La risoluzione automatica della collinearità mediante ritenzione conservativa della feature con maggiore importanza (Z-score) ha portato alla rimozione di 7 feature dal set: media di S, TRI, TC, A-northness, deviazione standard di S, TC, GC (Tabella 6 e Figura 8).

Tabella 7. Modulo del coefficiente di correlazione di Pearson per i valori maggiori di 0,9 tra tutte le combinazioni del set di feature filtrato tramite Boruta test. Feature siglate in *SD* per deviazione standard e *Mean* per il valore medio. Si rimanda alla Tabella 1 per la denominazione estesa delle derivate.

Feature 1	Feature 2	Correlazione	Importance_F1	Importance_F2	Importance minore
SD_S	SD_TRI	0.930277241	12.24552247	13.3991275	SD_S
Mean_S	Mean_TRI	0.993412137	11.09208351	11.06271924	Mean_TRI
Mean_S	SD_S	0.937899449	11.09208351	12.24552247	Mean_S
Mean_S	SD_TRI	0.931804481	11.09208351	13.3991275	Mean_S
Mean_S	SD_GC	0.908992298	11.09208351	9.405760175	SD_GC
Mean_S	Mean_TC	0.9086814	11.09208351	10.70587503	Mean_TC
Mean_TRI	SD_TRI	0.944877801	11.06271924	13.3991275	Mean_TRI
Mean_TRI	SD_S	0.926323551	11.06271924	12.24552247	Mean_TRI
Mean_TRI	SD_GC	0.924709593	11.06271924	9.405760175	SD_GC
Mean_TC	SD_TC	0.952932218	10.70587503	9.498741158	SD_TC
Mean_TC	SD_TRI	0.939103477	10.70587503	13.3991275	Mean_TC
Mean_TC	Mean_TRI	0.93627081	10.70587503	11.06271924	Mean_TC
Mean_TC	SD_GC	0.923465328	10.70587503	9.405760175	SD_GC
SD_TC	SD_TRI	0.90555725	9.498741158	13.3991275	SD_TC
SD_GC	SD_S	0.947569934	9.405760175	12.24552247	SD_GC
SD_GC	SD_TRI	0.92550431	9.405760175	13.3991275	SD_GC
Mean_E	Mean_A_northness	0.941587112	6.851734288	5.390737573	Mean_A_northness

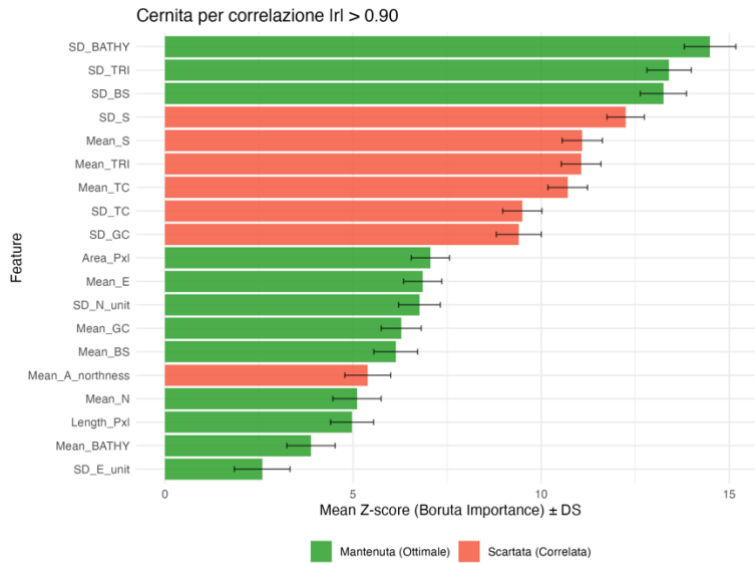


Figura 8. Grafico a barre del valore medio di importanza calcolato con Boruta test delle feature escluse per l'elevata correlazione (in rosso) e le feature mantenute a seguito della

Il test di selezione RFE per la riduzione dello spazio dimensionale del set di feature con approccio minimal-optimal è stato quindi applicato al set epurato dalle due variabili escluse con test Boruta e dalle 7 variabili ad elevata collinearità. Avendo uno spazio ridotto a 12 dimensioni, si è analizzata la performance del test RFE per l'intero intervallo di dimensioni possibili (da 2 a 12). Come si evince dal grafico della performance RFE (Figura 9.a), la dimensione ottimale minima si attesta a 6, con il valore di accuratezza che si mantiene stabile negli insiemi a 7 ed 8 feature.

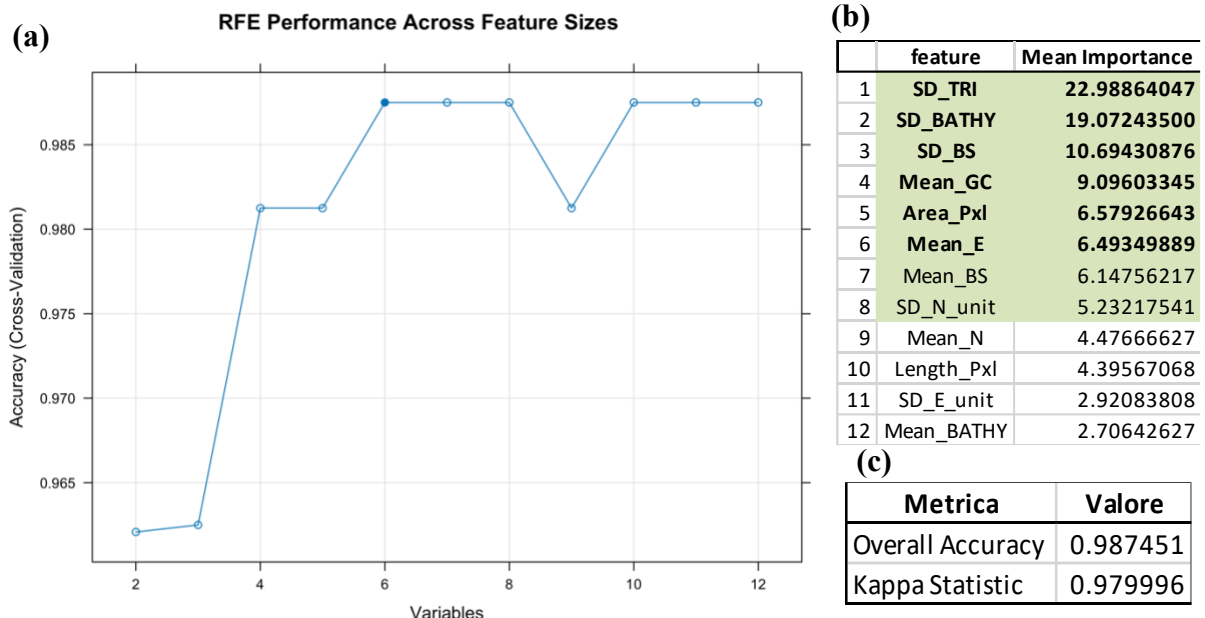


Figura 9. A) Grafico di andamento della performance, misurata in campionamento iterativi della Overall Accuracy in funzione del numero di variabili (feature). B) Valori di importanza media calcolati con modello random forest nella analisi RFE. In sfondo verde è evidenziato il set di feature mantenuto; in grassetto (prime sei) sono evidenziate le feature confermate dalla analisi. C) valori di accuratezza globale (Overall Accuracy) e indice Kappa per il set di feature ottimale.

Per non incorrere nella perdita di feature secondarie eventualmente informative per la risoluzione del dettaglio cartografico, a fronte del rumore associato ai descrittori principali, si è scelto di mantenere uno spazio di 8 feature, corrispondente all'estremo superiore del plateau di performance ottimale identificato dalla analisi RFE (Figura 9.b). Questo approccio precauzionale è supportato dal valore di accuratezza Kappa, pari a 0,980, stimato con ricampionamento iterativo a 10 partizioni (10-fold CV) su algoritmo random forest (Figura 9.c).

L'analisi di regressione logistica multinomiale LASSO è stata adottata come test diagnostico lineare. La penalizzazione L1 per le variabili che non rappresentano un vantaggio lineare netto porta all'esclusione di sei feature (Figura 10), corrispondenti a quelle identificate come meno informative in base alla importanza calcolata dall'algoritmo RFE. Si conclude quindi che il nucleo delle restanti sei feature, confermate sia dai metodi basati su alberi decisionali (il random forest su cui si innestano i test Boruta e RFE) sia dal modello lineare (LASSO), rappresenta uno spazio predittivo indipendente dall'architettura algoritmica.

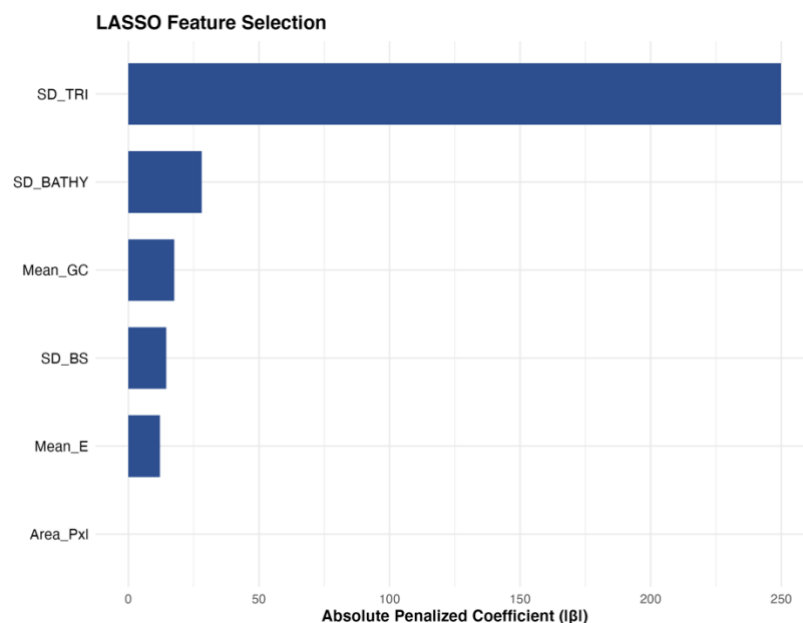


Figura 10. Grafico a barre del modulo del coefficiente beta penalizzato, determinato con analisi di regressione logistica multinomiale (LASSO).

Si nota che la forte predominanza del coefficiente relativo alla deviazione standard di TRI è derivata dal processo di segmentazione, che ha adoperato l'indice TRI ed il backscatter come variabili del criterio di omogeneità, producendo un artefatto algoritmico della *multiresolution segmentation* dato dalla circolarità informativa tra creazione dell'oggetto e classificazione.

Poiché le variabili topografiche ed acustiche utilizzate costituiscono dei predittori ad alto potenziale di interazione non lineare, si esclude la possibilità di adoperare l'analisi di regressione logistica per filtrare lo spazio delle feature in base ai coefficienti penalizzati. Si è quindi confermato lo spazio delle feature ad otto variabili, selezionato attraverso il test Boruta, l'analisi di correlazione ed il test RFE, come set ottimale per la classificazione degli oggetti.

La presente selezione ha comportato l'esclusione delle feature connesse alle derivate morfometriche di pendenza (S) e curvatura totale (TC), e l'introduzione delle derivate di direzione, *northness* (in deviazione standard) ed *eastness* (media).

4.2 Accuratezza e robustezza del processo di classificazione

La valutazione delle prestazioni del processo GEOBIA implementato si fonda sulla verifica della stabilità dell'intero processo metodologico attraverso l'analisi delle metriche di accuratezza calcolate secondo il metodo di Olofsson et al. (2014).

Coerentemente con il disegno sperimentale descritto, i risultati ottenuti dall'incrocio spaziale con i punti di validazione indipendenti sono presentati in due fasi consequenziali: in primo luogo, la quantificazione dell'accuratezza ottenuta nell'area di Levanto con il *rule set* Ottimizzato; in secondo luogo, l'analisi inferenziale della robustezza del workflow alla trasferibilità spaziale (nelle altre aree di studio), e tematica (con l'introduzione della classe *matte*).

4.2.1 Accuratezza globale, PA e UA

Le metriche di accuratezza globale, calcolate con pesatura areale, restituiscono valori compresi tra 85% e 95%, con l'accuratezza più alta ottenuta nello scenario target (Levanto, *rule set* Ottimizzato, 3 classi), sul quale è stato sviluppato il modello di classificazione. Questo risultato evidenzia la capacità del classificatore con modello decision tree bagging di interpretare le firme bionomiche delle classi indagate, a partire dalle derivate morfometriche del MBES.

La differenza in valori assoluti tra le versioni di *rule set* suggerisce un miglioramento marginale nel modello ottimizzato, ottenuto dalla stabilizzazione delle metriche. Il limite inferiore dell'intervallo di confidenza della versione Ottimizzato è superiore di quattro punti percentuali rispetto alla versione di partenza (Figura 11.a). Si conferma quindi un incremento dei valori assoluti di accuratezza, seppur non validato come statisticamente significativo dal test Z ($p > 0,05$)

L'analisi delle metriche classe-specifiche (User Accuracy e Producer Accuracy) rivela un quadro affidabile per l'utente finale (UA elevate), e bilanciato tra le diverse classi. Mentre il *rule set* iniziale presenta uno sbilanciamento per la classe rock verso l'errore di commissione, con UA inferiore a 75% e PA pari al 100%, la versione Ottimizzato restituisce valori di accuratezza più alti e bilanciati in ogni classe (Figura 11.b). Questo risultato conferma la migliore prestazione del *rule set* versione Ottimizzato nei confronti delle altre tre versioni adoperate.

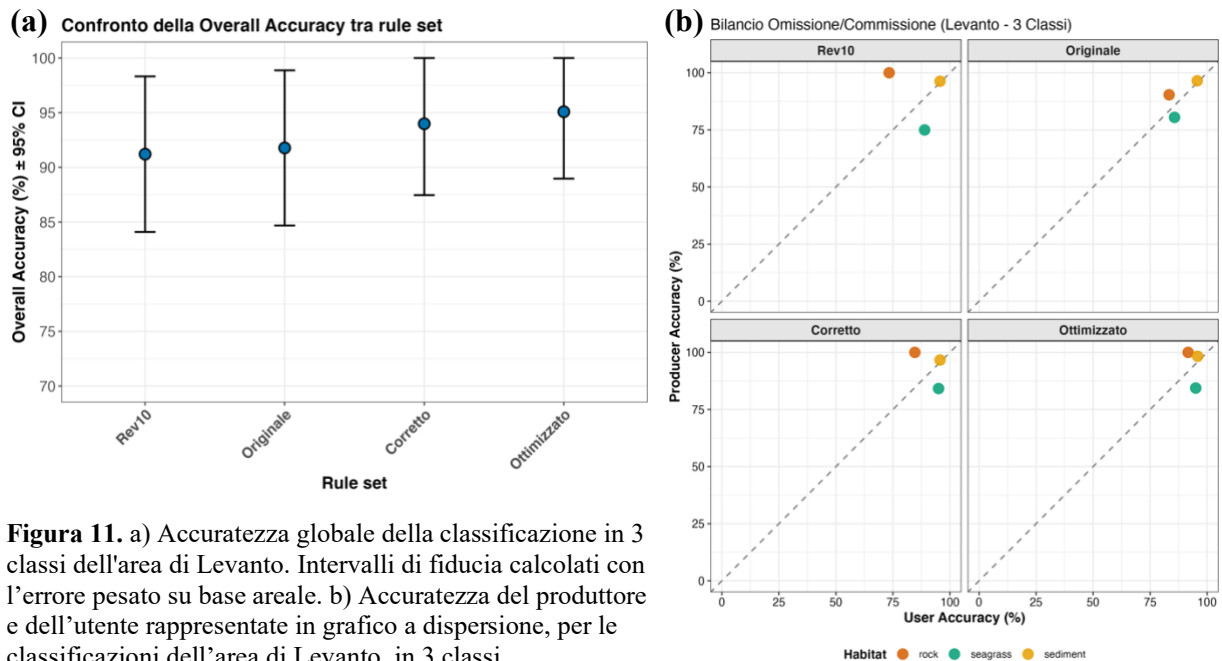


Figura 11. a) Accuratezza globale della classificazione in 3 classi dell'area di Levanto. Intervalli di fiducia calcolati con l'errore pesato su base areale. b) Accuratezza del produttore e dell'utente rappresentate in grafico a dispersione, per le classificazioni dell'area di Levanto, in 3 classi.

Osservando l'andamento delle metriche di accuratezza alla variazione del numero di classi (Figura 12.a), si conferma una OA stimata superiore a 90% in tutti i livelli. L'incremento di errore, associato alla introduzione di ogni classe ulteriore, contribuisce alla diminuzione della accuratezza globale. Tuttavia, il decremento delle metriche non risulta rilevabile come statisticamente significativo (Tabella 7). Si riscontra, dunque, una sostanziale tenuta del modello ($p > 0,05$) a fronte dell'aumento di complessità tassonomica.

La classificazione con il *rule set* Ottimizzato nelle altre aree di studio restituisce un quadro più eterogeneo (Figura 12.b), in cui l'accuratezza globale risulta marcatamente inferiore rispetto al sito pilota di Levanto, con l'eccezione dell'area Genova_20 (Punta Moneglia), che presenta OA pari a 96,5% ed errore standard di 2,1% (Allegato 6).

Le metriche di accuratezza del produttore e dell'utente analizzate per singola classe (Figura 12.c) denotano un risultato più eterogeneo, sia tra le diverse aree di studio, sia all'aumentare del numero di classi (con l'introduzione di matte e sediment F).

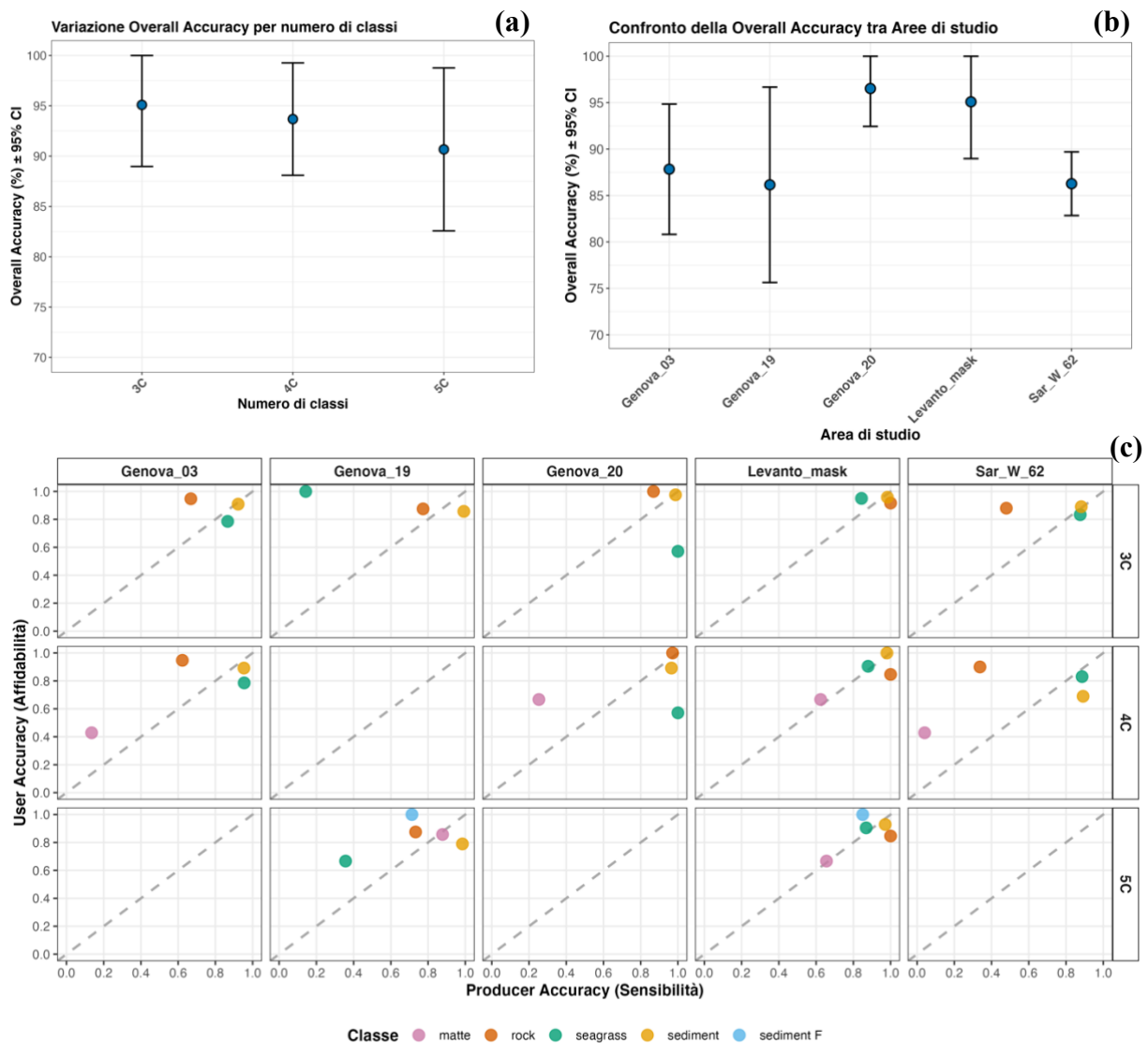


Figura 12. a) Accuratezza globale delle classificazioni nell'area Levanto con *rule set* Ottimizzato. b) Accuratezza globale delle classificazioni con *rule set* Ottimizzato e 3 classi; c) Dispersione UA – PA per ogni classe, con *rule set* Ottimizzato.

L'area Sar_W_62 ha restituito bassa sensibilità alla classe rock, e sensibilità quasi nulla per la matte morta (Allegato 7). Tuttavia, questo risultato è da ricondursi alla marcata diversità di questa area relativamente alla morfologia locale, in cui la presenza di substrato roccioso è risultata di difficile interpretazione, in fase di campionamento dei punti di addestramento.

Nelle aree di Levanto, Genova_03 (Secca di Santo Stefano) e Genova_20 si è ottenuta una accuratezza elevata e stabile per le tre classi principali (rock, seagrass, sediment) e per sediment F, mentre la classe matte presenta bassa accuratezza dell'utente (da 0,42 a 0,67) e molto bassa sensibilità (accuratezza del produttore compresa tra 0,13 e 0,66). Questo risultato conferma la difficoltà di classificare accuratamente i fondali ricoperti da matte morta di *P. oceanica* utilizzando solamente il dato acustico del multibeam e le derivate morfometriche.

Nell'area Genova_19 (Punta Baffe) si riscontra, invece, elevata accuratezza per la classe matte, con PA di 0,87 e UA di 0,86 ma bassa sensibilità per la classe seagrass (Allegato 5). Questo decadimento statisticamente confermato ($p < 0,001$) della Producer Accuracy costituisce un significativo errore di omissione sito-specifico, ed è strettamente dipendente dalla presenza molto marginale e limitata dell'habitat di prateria di *P. oceanica*, che ha comportato una sottorappresentazione spaziale della classe seagrass in fase di addestramento.

Si riscontra quindi che l'area di Levanto risulta l'unica in cui l'accuratezza dell'utente e del produttore risultano superiori alla soglia di 0,8 per le classi di sedimento, roccia e prateria, e in cui la matte presenta valori di accuratezza superiori a 0,6 (si vedano le tavole agli Allegati 2-4). Mentre, le altre aree di studio risultano affette da fonti di errore sito-specifiche, che riducono l'affidabilità della classificazione risultante per alcune classi.

L'analisi inferenziale con test Z di confronto a coppie, calcolato con gli errori standard pesati spazialmente (Olofsson et al., 2014), per determinare la significatività delle differenze nelle metriche di accuratezza, ha restituito un p-value inferiore a 0,05 solo per le differenze in OA tra Levanto e Sar_W_62, e tra Genova_20 con Genova_03 e Sar_W_62.

Secondo l'analisi delle metriche per singole classi, risultano differenze significative per un livello di fiducia di 0,05 nei confronti a coppie con Genova_03 (per rock), Genova_19 (per seagrass) e Sar_W_62 (in tutte le classi) rispetto a Levanto (in ottimizzato, 3 classi). Queste differenze ($p < 0,05$) si riconducono alla dipendenza dell'accuratezza dalla rappresentatività spaziale dell'habitat. In Sar_W_62, gli errori standard risultano più bassi in ragione del maggior numero di punti di validazione, rendendo il test Z più sensibile alle differenze. I limiti di potenza del test inferenziale, dipendenti dalla stima della varianza spaziale sulle classi minoritarie, causano un potenziale mascheramento statistico di divergenze cartograficamente rilevanti.

Le discrepanze tra i valori di User Accuracy e Producer Accuracy, laddove la seconda presenta drastici decrementi, denotano la capacità del classificatore di limitare gli errori di commissione pur perdendo capacità di generalizzazione spaziale. Tale limite, tuttavia, non dipende solo dalla sensibilità del modello di classificazione adottato, ma anche dalla ambiguità morfometrica tra matte morta e dalla sottorappresentazione della variabilità strutturale dell'habitat nei siti in cui esso risulta ecologicamente marginale.

Tabella 8. Confronti a coppie con statistica Z delle metriche di accuratezza. Le metriche per classe sono riferite tutte al confronto con 3 classi e rule set Ottimizzato.

Livelli	Confronto	OA_1	OA_2	Statistica Z	p-value
Levanto - 3C	Ottimizzato vs Rev10	0.9509	0.9121	0.811	0.4176
Ottimizzato - 3C	Levanto vs Genova_20	0.9509	0.9652	0.381	0.7029
Ottimizzato - 3C	Levanto vs Genova_19	0.9509	0.8615	1.439	0.1500
Ottimizzato - 3C	Levanto vs Genova_03	0.9509	0.8783	1.529	0.1263
Ottimizzato - 3C	Levanto vs Sar_W_62	0.9509	0.8627	2.466	0.0137
Ottimizzato - 3C	Genova_20 vs Genova_19	0.9652	0.8615	1.801	0.0717
Ottimizzato - 3C	Genova_20 vs Genova_03	0.9652	0.8783	2.099	0.0358
Ottimizzato - 3C	Genova_20 vs Sar_W_62	0.9652	0.8627	3.771	0.0002
Levanto - Ottimizzato	3C vs 4C	0.9509	0.9368	0.334	0.7386
Levanto - Ottimizzato	3C vs 5C	0.9509	0.9067	0.854	0.3931
Classe – metrica	Confronto	Acc_1	Acc_2	Statistica Z	p-value
rock – PA	Levanto vs Sar_W_62	1	0.4794	5.882	< 0.0001
seagrass – PA	Levanto vs Genova_19	0.8437	0.1429	6.074	<0.0001
sediment – PA	Levanto vs Sar_W_62	0.9828	0.8819	3.743	0.0002
rock – PA	Levanto vs Genova_03	1	0.6681	2.569	0.0102
seagrass – UA	Levanto vs Sar_W_62	0.95	0.8333	2.038	0.0415

4.2.2 Varianza della segmentazione

La valutazione dell'architettura topologica degli oggetti attraverso gradienti spaziali dissimili rappresenta un requisito critico per garantire l'assenza di overfitting morfometrico locale (Diesing et al., 2016; Ismail et al., 2015). L'estrazione della WV su tutte le aree di studio risponde alla necessità di validare la trasferibilità del *rule set*: fissata la parametrizzazione algoritmica, il mantenimento di una WV stabile tra le aree certifica che l'algoritmo preserva la medesima capacità di campionamento della varianza al di sotto degli ecotoni, indipendentemente dalla frammentazione locale.

Il calcolo della varianza pesata spazialmente ha riportato i valori di 4,9 per Levanto, 3,5 nelle aree Genova_03, Genova_20 e Genova_20 e 4,34 in Sar_W_62 confermando la stabilità della varianza adoperata dall'algoritmo di segmentazione.

4.3 Confronto geospaziale con il dato di validazione da AUV

Il confronto spaziale della classificazione da multibeam con il dato derivato da AUV si è potuto verificare solamente nelle aree di Genova_19, Genova_20 e Levanto, poiché non è disponibile un dato acquisito con AUV nelle due aree restanti. La verifica dell'accuratezza della carta

prodotta è, inoltre, limitata dalla copertura incompleta degli habitat presenti nelle aree di studio. Il grigliato vettoriale conta 598 punti nell'area di Levante, a fronte di 152 in Genova_19 e 230 nell'area Genova_20.

Analizzando le matrici di confusione sulla base delle corrispondenze multiple, si riscontra un valore di accuratezza globale che oscilla tra 0,43 e 0,92, dimostrando un andamento eterogeneo nelle diverse aree di studio.

Nell'area di Levante, l'accuratezza globale calcolata sul dato da AUV si attesta a 0,70 per la classificazione in 3 classi con *rule set* Ottimizzato. Valore che si mantiene stabile con l'introduzione della quarta classe (matte), e che si discosta di soli 5 punti percentuali dall'accuratezza ottenuta con la versione di *rule set* Rev10 (OA = 0,75). L'accuratezza della classe seagrass si attesta al di sopra della media, con un valore di $0.8 \pm 0,03$ in tutte le combinazioni di rule set e numero di classi, sia in accuratezza dell'utente sia del produttore. Per la classe rock, non è invece disponibile informazione sufficiente per valutare l'accuratezza della classificazione, poiché nel transetto AUV si sono riscontrati solamente 2 punti classificati come "rocca o massi" (Tabella 9).

Tabella 9. Matrice di confusione e metriche di accuratezza calcolate sul dato derivato da AUV nell'area di Levante, rule set Ottimizzato, 3 classi. Su sfondo verde i campi dei veri positivi.

Classe AUV	seagrass	sediment	rock	Totale	PA
0	55	61	0	116	0.53
3	41	13	0	54	0.24
4	2	0	0	2	0.00
5	345	81	0	426	0.81
UA	0.78	0.48	NA	OA	0.70

Nell'area Genova_20, a fronte di un marcato abbassamento della accuratezza globale (attestata tra 0,43 e 0,53), la versione di rule set Ottimizzato ha conseguito un miglioramento di 10 punti percentuali nella sensibilità per le classi rock e seagrass (con valori di PA pari a 0,41 e 0,21 rispettivamente) rispetto alla classificazione con Rev10.

Nell'area Genova_19, si riscontra invece l'accuratezza globale più elevata, con valore di 0,93 nella classificazione Ottimizzato – 3 classi (Tabella 10). Tuttavia, l'accuratezza del produttore per la posidonia (classe 5), che non supera il valore di 0,62 e risulta, inoltre, non valutabile l'accuratezza per la classe rock, la quale non ricade all'interno del transetto di verità a mare effettuato (Figura 13).

Tabella 10. Matrice di confusione e metriche di accuratezza calcolate sul dato derivato da AUV. Su sfondo verde i campi dei veri positivi.

Genova_20 – Ottimizzato – 3 classi						Genova_19 – Ottimizzato – 3 classi					
Classe	seagrass	sediment	rock	Totale	PA	Classe	seagrass	sediment	rock	Totale	PA
0	1	3	1	5	0.60	0	0	7	0	7	1.00
1	0	4	0	4	1.00	1	0	7	0	7	1.00
2	0	1	0	1	1.00	2	0	0	0		NA
3	25	71	19	115	0.62	3	1	111	0	112	0.99
4	0	11	8	19	0.42	4	0	0	0		NA
5	18	47	21	86	0.21	5	16	10	0	26	0.62
UA	0.41	0.58	0.16	OA	0.46	UA	0.94	0.93	NA	OA	0.93

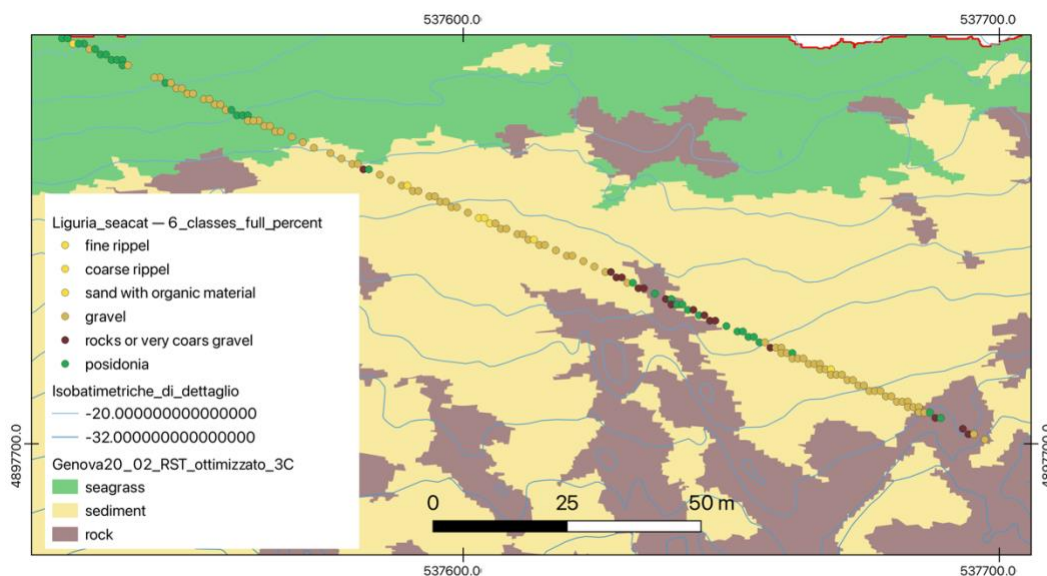


Figura 13. Sovrapposizione dei punti di validazione classificati da ortofoto AUV con classificazione del dato multibeam, area Genova_20 (dettaglio), rule set Ottimizzato, 5 classi.

Analizzando la distribuzione spaziale dei punti di AUV in discordanza con la classificazione, si evince che parte di essi si concentra lungo i margini roccia-sedimento e prateria-sedimento (Figura 14). Data l'alta frammentazione di questi margini, è ragionevole imputare una quota significativa degli errori di omissione e commissione non a un reale deficit predittivo del modello acustico, bensì al *misregistration error* intrinseco del mezzo AUV (2 m di incertezza sul posizionamento).

Tuttavia, l'analisi spaziale dei residui non esaurisce la quota di errore registrata, specialmente laddove le discordanze si manifestano nei nuclei omogenei delle patch bentoniche e non ai loro margini. La flessione delle metriche di accuratezza deve essere ricondotta ai limiti intrinseci del processo di classificazione basato su oggetti. Da un lato, la sovrapposizione delle firme

acustiche riduce la capacità discriminativa dell' algoritmo di classificazione. Dall' altro, è da considerare la componente di incertezza introdotta dall' interprete durante la fase di creazione dei punti di addestramento. Il dato AUV, derivato da evidenza ottica diretta, fa emergere la discrepanza tematica con il risultato della classificazione alla scala metrica, inferiore quindi alla dimensione degli oggetti generati con la segmentazione.

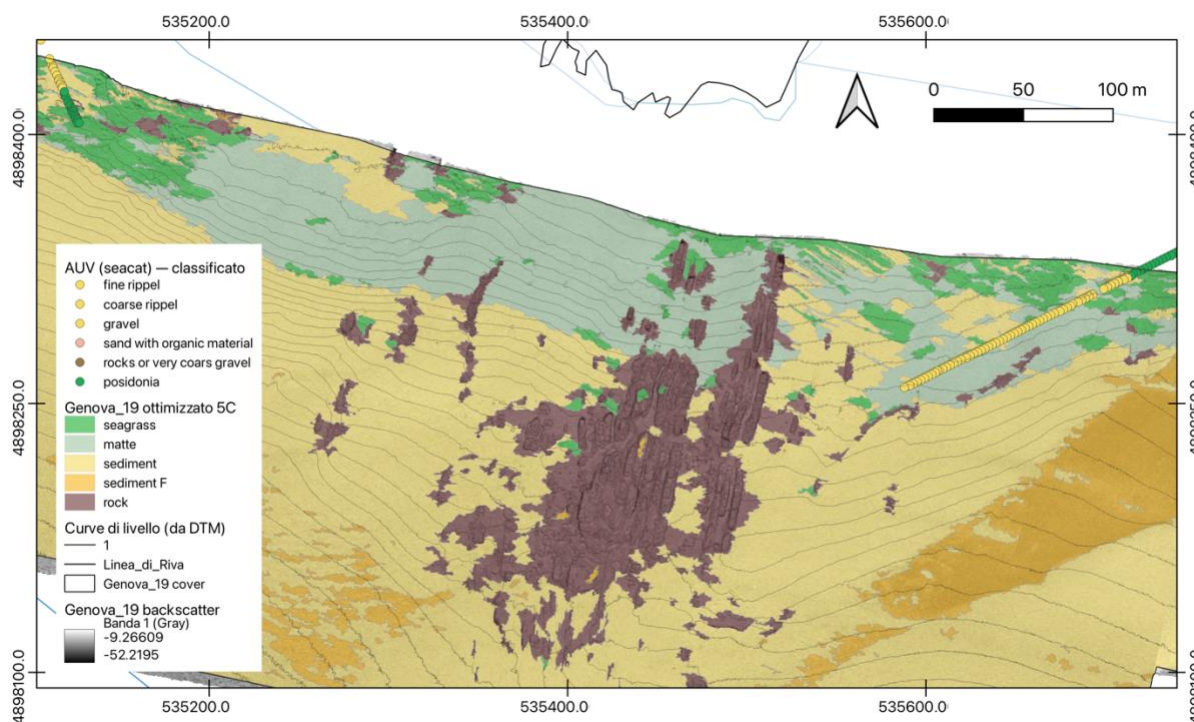


Figura 14. Sovrapposizione dei punti di validazione classificati da ortofoto AUV con classificazione del dato multibeam, area Genova_19 (dettaglio), rule set Ottimizzato, 3 classi.

4.4 Confronto geospaziale con classificazione da satellite

La sovrapposizione delle classificazioni nelle cinque aree di studio con il mosaico derivato da immagini satellitari ha restituito risultati fortemente eterogenei tra le diverse aree, ma stabili tra i diversi rule set. Nell' area di Levante, l' accuratezza globale calcolata su base areale si attesta a 0,88 sia nella classificazione con versione Rev10 (3 classi) sia nelle classificazioni con *rule set* Ottimizzato in 3 e 4 classi, presentando un decremento di un solo punto percentuale per la classificazione in 5 classi. L' accuratezza del produttore per la classe 'Biocenosi di *P. oceanica*' risulta compresa tra 74% e 78%, posizionandosi nello stesso intervallo dell' accuratezza dell' utente per la classe seagrass.

Nell' area Genova_19, la Producer Accuracy per la prateria di *P. oceanica* è compresa tra 28% e 56% e la User Accuracy oscilla tra 39% e 48%. Analogamente, nell' area Genova_20 si riscontrano valori di accuratezza PA e UA compresi tra 27% e 56% (Tabella 11), che denotano

una marcata discordanza spaziale tra la carta classificata da dato acustico ed il mosaico satellitare.

Tabella 11. Accuratezza del produttore (PA) e dell'utente (UA), del confronto tra le carte classificate ed il mosaico satellitare (classe, Biocenosi della posidonia).

Area	Rule set	Classi	PA_Seagrass	UA_Seagrass
Genova_03	Rev10	3C	70%	100%
Genova_03	Ottimizzato	3C	78%	100%
Genova_03	Ottimizzato	4C	78%	100%
Genova_19	Rev10	3C	44%	48%
Genova_19	Ottimizzato	3C	29%	46%
Genova_19	Ottimizzato	5C	60%	39%
Genova_20	Rev10	3C	27%	57%
Genova_20	Ottimizzato	3C	40%	42%
Genova_20	Ottimizzato	4C	38%	40%
Levanto	Rev10	3C	75%	77%
Levanto	Ottimizzato	3C	74%	76%
Levanto	Ottimizzato	4C	78%	75%
Levanto	Ottimizzato	5C	74%	74%
Sar_W_62	Rev10	3C	54%	94%
Sar_W_62	Ottimizzato	3C	53%	94%
Sar_W_62	Ottimizzato	4C	53%	94%

Diversamente, nell'area di Genova_03 si riscontra una accuratezza molto elevata, con PA tra il 70% e il 78%; la UA del 100% rappresenta tuttavia un valore influenzato dalla mancanza di eterogeneità tematica del dato satellitare, che ricopre solo una porzione limitata dell'area di studio, prossima al limite della penetrazione batimetrica per il sensore ottico. In Sar_W_62, infine, si riscontra un forte sbilanciamento tra UA molto elevata (94%) e PA moderata (53%). Tale condizione indica che i poligoni classificati come seagrass dal multibeam trovano quasi totale riscontro nel dato ottico (alta affidabilità), ma che quest'ultimo individua estensioni areali maggiori. Questo divario non invalida la predizione, ma riflette la tendenza del sensore satellitare ad accorpate morfologie contigue, laddove l'algoritmo acustico discrimina con maggiore selettività le discontinuità ad elevata risoluzione.

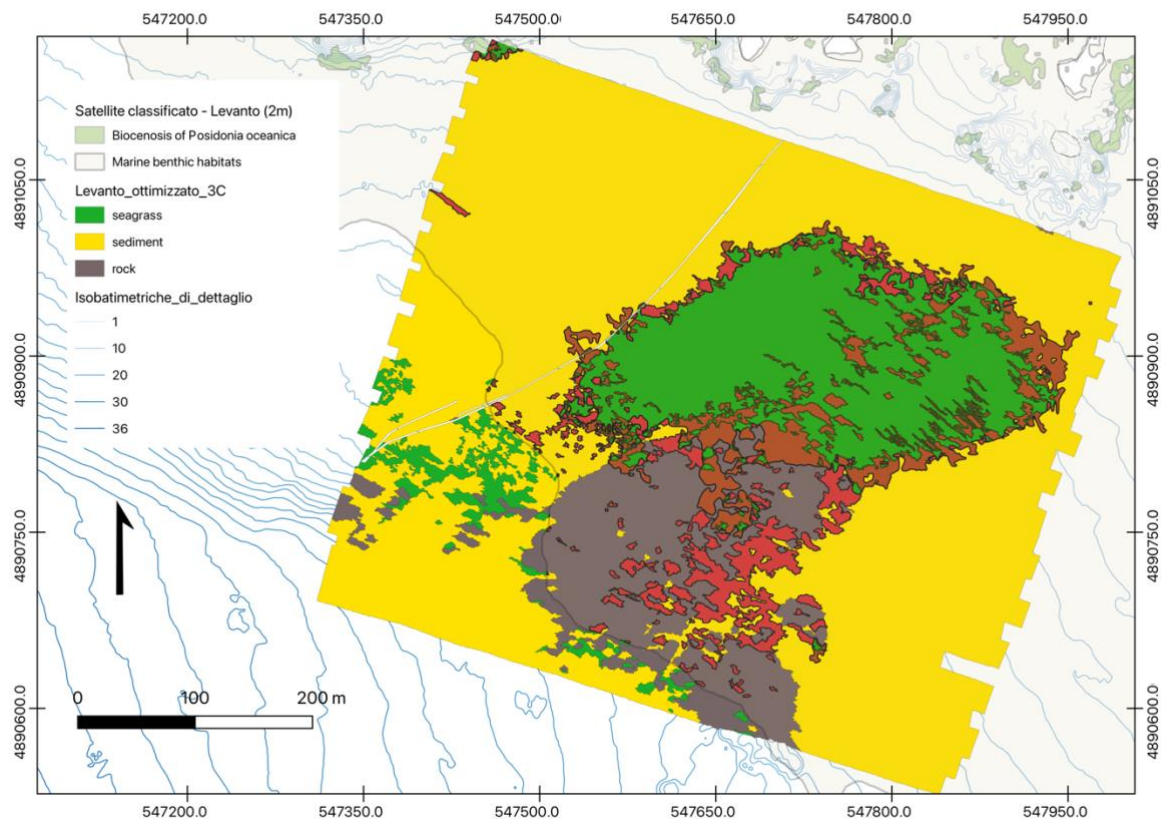


Figura 15. Quadro di sovrapposizione tra la classificazione in 3 classi con rule set Ottimizzato e dato derivato da satellite, nell'area di Levanto. In rosso le aree di discordanza.

Tuttavia, come si evince dal quadro di discordanza della sovrapposizione tra la classificazione di Levanto ottimizzato (3 classi) con il dato satellitare (Figura 15), le differenze si concentrano lungo i margini delle patches. Questa condizione, associata alla maggiore tortuosità dei margini della prateria nella classificazione da multibeam, evidenzia che le discrepanze tra le due carte sono largamente imputabili al diverso grado di risoluzione con cui sono state delineate le patches. Inoltre, la limitata capacità di penetrazione del telerilevamento satellitare alle maggiori profondità, rende il confronto tra le carte intrinsecamente limitato alle sole batimetrie più superficiali, fornendo un quadro di validazione spazialmente parziale rispetto al dominio acustico locale.

4.5 Confronto geospaziale con cartografia bionomica di riferimento

Il confronto spaziale tra la classificazione acustica e le cartografie di riferimento a mesoscala ha richiesto la mediazione delle corrispondenze multiple rappresentate in Tabella 6.

La corrispondenza della classificazione con l'atlante degli habitat (Coppo et al., 2020) risulta elevata in tutte le aree liguri, dove l'accuratezza globale si attesta tra l'85% ed il 94% nella

classificazione con 3 classi, con i valori più alti confermati per la classificazione di Levanto (Tabella 12). Questi valori evidenziano una solida coerenza topologica per le categorie bionomiche aggregate.

Tabella 12. Metriche di accuratezza calcolate sulla corrispondenza multipla della classificazione da multibeam sulle classi bionomiche dell’atlante degli habitat (Coppo et al., 2020).

Area	Rule set	Classi	OA	PA mos	PA pos-roc	PA pos	UA seagrass	UA matte
Genova_03	Rev10	3C	85%	99%		75%	82%	
Genova_03	Ottimizzato	3C	86%	99%		81%	83%	
Genova_03	Ottimizzato	4C	85%	97%		81%	81%	21%
Genova_19	Rev10	3C	88%	96%		20%	27%	
Genova_19	Ottimizzato	3C	90%	94%		18%	31%	
Genova_19	Ottimizzato	5C	79%	73%		39%	37%	8%
Genova_20	Rev10	3C	87%	86%		21%	65%	
Genova_20	Ottimizzato	3C	86%	83%		74%	80%	
Genova_20	Ottimizzato	4C	84%	59%		73%	79%	75%
Levanto	Rev10	3C	94%	98%	97%		91%	
Levanto	Ottimizzato	3C	93%	97%	96%		91%	
Levanto	Ottimizzato	4C	87%	81%	97%		89%	81%
Levanto	Ottimizzato	5C	87%	80%	97%		89%	80%
Sar_W_62	Rev10	3C	78%	100%		71%	91%	
Sar_W_62	Ottimizzato	3C	76%	100%		68%	90%	
Sar_W_62	Ottimizzato	4C	62%	2%		69%	85%	23%

Analizzando il dettaglio classe-specifico, si osserva che la Producer Accuracy (PA) calcolata sulla classe del mosaico di prateria di *P. oceanica* e matte morta (corrispondente alle classi seagrass e matte della classificazione) raggiunge valori prossimi alla totalità (> 95%) nelle configurazioni a 3 classi, e mantiene valori elevati nelle classificazioni che introducono la classe matte (versioni in 4 e 5 classi), con la sola eccezione di Sar_W_62 in cui il classificatore non ha isolato con successo questa classe. Tale risultato rappresenta una misura della sensibilità della classificazione all’estensione tematica con la classe matte, la cui User Accuracy si attesta tra il 75% e l’81% nelle aree di Levanto e Genova_20.

Si evidenzia, inoltre, come nelle aree di Genova_19 e Genova_20, che hanno valori di accuratezza generalmente più bassi, la user accuracy è risultata marcatamente più alta per la classe seagrass nel rule set Ottimizzato rispetto alla versione Rev10, grazie all’introduzione della media della batimetria nello spazio delle feature. La localizzazione delle patch della classe seagrass nella classificazione del dato multibeam è risultata fortemente polarizzata verso le

profondità minori (Figura 16), grazie all'introduzione di questa feature, restituendo un mosaico più coerente con la reale distribuzione batimetrica delle praterie di *P. oceanica*, confermata dalla cartografia bionomica (si veda il paragrafo 4.6).

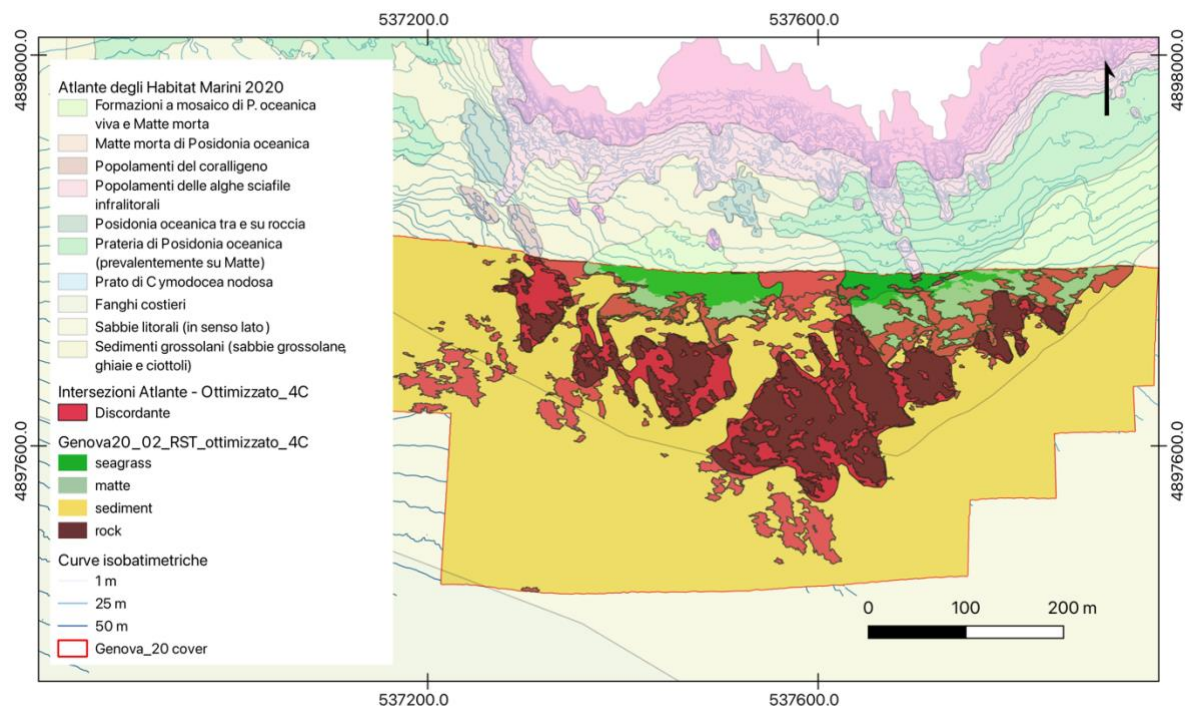


Figura 16. Sovrapposizione tra la classificazione del dato multibeam con rule set ottimizzato, 4 classi, nell'area Genova_20, e la carta Nuovo Atlante degli Habitat Marini (Coppo et al., 2020).

Nell'area di Levante si riscontrano valori di PA elevati per le classi maggioritarie (Tabella 13). La bassa sensibilità attribuita alle classi di dettaglio (asi, dc) presenti nell'atlante degli habitat (Coppo et al., 2020) dimostra la tendenza del classificatore a distinguere meglio gli habitat principali rappresentativi della classe, dimostrando la suscettibilità del modello di classificazione all'eterogeneità intra-classe dovuta all'inclusione di habitat distinti nella cartografia bionomica. Si evidenzia, inoltre, la testimonianza nell'atlante degli habitat (Coppo et al., 2020) di una piccola patch di prateria di *Cymodocea nodosa*, localizzata all'interno dell'area a sedimento fine, che non è stata identificata all'interno della classe seagrass. Tuttavia, questa discordanza non proviene da un errore del processo di classificazione, ma dalla effettiva omogeneità del segnale acustico e delle derivate morfometriche tra l'area mappata come prateria di *C. nodosa* e la circostante area classificata come sedimento costiero (s). Tale omogeneità è stata valutata effettuando una prova di addestramento e classificazione inserendo nel set di punti di addestramento una classe specifica per questo habitat, presa appositamente in corrispondenza del poligono classificato "cym" nell'atlante degli habitat (Coppo et al., 2020); tentativo che ha

restituito la stessa classificazione priva della classe cym, confermando l'impossibilità da parte del classificatore di distinguere la presente patch nello spazio delle feature adottato.

Tabella 13. Matrice di confusione e metriche di accuratezza del confronto della classificazione dell'area di Levanto con *rule set* Ottimizzato, 4 classi, rispetto alle classi bionomiche dell'atlante degli habitat (Coppo et al., 2020). Su sfondo verde i campi dei veri positivi.

MATRICE DI CONFUSIONE (Aree in mq)					Totale	PA
Classe	rock	sediment	matte	seagrass		
af	3201	98	0	0	3299	97%
asi	0	28	28	211	268	0%
c	13407	640	1140	3721	18908	71%
cym	0	6141	0	0	6141	0%
dc	443	2645	226	691	4004	66%
mos	3807	12713	12645	55641	84806	81%
pos-roc	19630	140	590	5045	25405	97%
s	365	100910	392	2499	104166	97%
sgc	287	33741	688	555	35271	96%
USER ACCURACY (UA)				OVERALL ACCURACY		
Classe	Totale	UA				
matte	15708	81%				
rock	41140	88%				
seagrass	68364	89%				
sediment	157055	87%				

Un'ulteriore criticità emerge in corrispondenza dell'habitat del coralligeno (c), la cui Producer Accuracy risulta penalizzata da un evidente errore di misclassificazione: porzioni di tale habitat sono state identificate come seagrass, estendendo la predizione al di sotto del reale limite batimetrico inferiore della prateria (Figura 17). Questa sovrapposizione testimonia l'intrinseca difficoltà dell'algorithmo nell'isolare le patches di *P. oceanica* su roccia: alla mancanza di un netto contrasto nel segnale di backscatter si somma l'inefficacia delle derivate topografiche, in quanto la prateria ricalca fedelmente la morfologia scoscesa del substrato duro sottostante. Questo limite del classificatore non è stato compensato nell'area di Levanto neppure dall'integrazione della media della batimetria nello spazio delle feature (nel *rule set* Ottimizzato).

Ciononostante, l'elevata accuratezza sia di PA sia di UA per la prateria di *P. oceanica* e matte morta attesta il successo della classificazione per le classi target del presente studio.

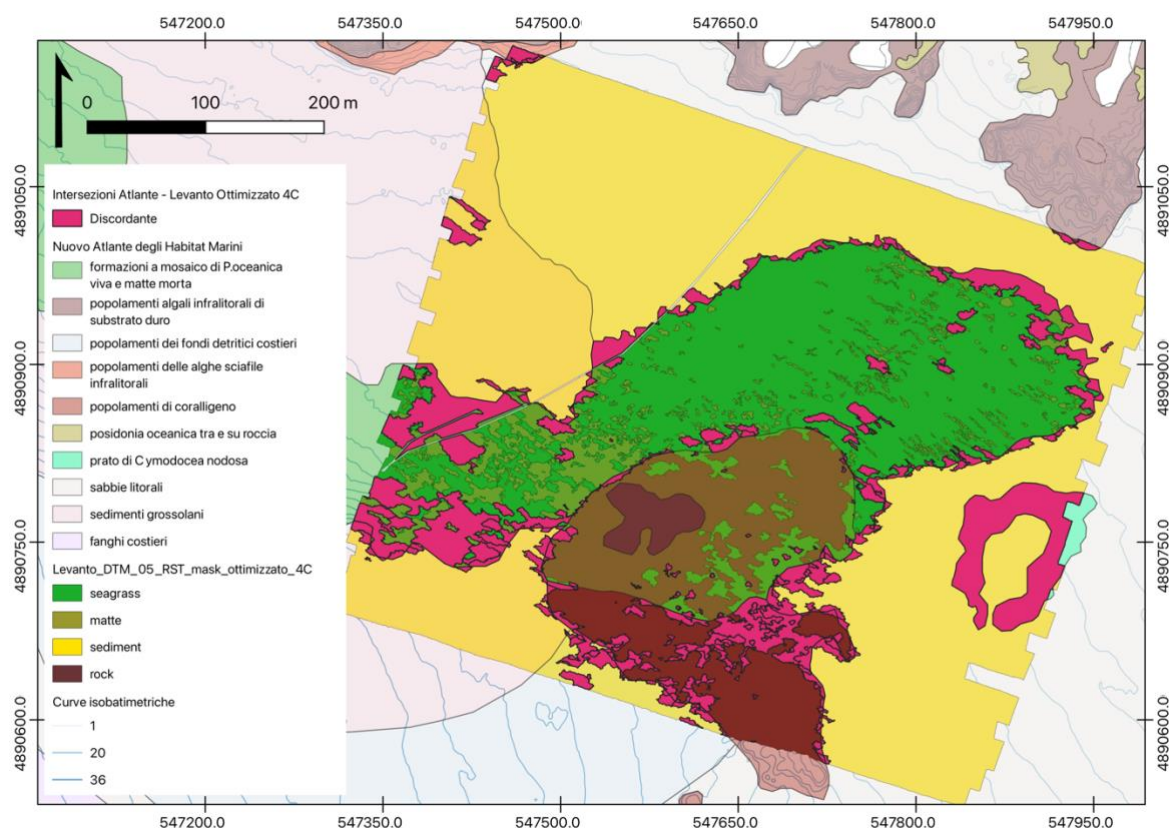


Figura 17. Sovrapposizione tra la classificazione del dato multibeam con rule set ottimizzato, 4 classi, nell'area Levanto, e la carta Nuovo Atlante degli Habitat Marini (Coppo et al., 2020).

Nelle aree di Genova_19 e Sar_W_62, la marcata flessione delle metriche di accuratezza con l'introduzione della classe matte (versioni con 4 e 5 classi) evidenzia un limite nella robustezza del processo di classificazione al trasferimento in queste aree. Questa debolezza è imputabile non solo alla sensibilità algoritmica rispetto alla classe della matte morta, ma anche alla efficacia dell'addestramento del modello di classificazione, dipendente dalla estensione e complessità del dato analizzato e dalla disponibilità di dati di addestramento in numero sufficiente per delineare la firma acustica nello spazio delle feature.

La discrepanza tra la classificazione del dato multibeam e le carte bionomiche di riferimento deve ricondursi in tutti i casi alla differente scala di acquisizione e rappresentazione del dato, sia per le classi ad elevata accuratezza sia per le classi con valori bassi, poiché l'alta risoluzione del classificatore acustico permette di catturare l'eterogeneità di microscala, individuando discontinuità reali che non trovano riscontro nel dato vettoriale di confronto. Tale dinamica è pienamente supportata dalle metriche di commissione calcolate sull'area di Levanto: a fronte di bassi valori di concordanza sulle classi frammentate dell'atlante degli habitat (Coppo et al., 2020), il modello acustico (in configurazione in 4 e 5 classi) ha mantenuto una User Accuracy

per la classe predittiva matte superiore allo 0,80. Ciò evidenzia un'elevata affidabilità spaziale del classificatore, capace di superare i limiti di risoluzione nativa della cartografia di inquadramento e di mappare variazioni bionomiche non altrimenti censite.

Nelle figure 16 e 17 si può riscontrare come le discordanze si distribuiscano in larga parte lungo i margini dei poligoni della carta bionomica, i quali presentano una tortuosità nettamente minore rispetto alla classificazione del dato multibeam, in ragione della minore risoluzione nativa del dato.

4.6 Analisi del limite inferiore

L'individuazione spaziale del limite inferiore della prateria a *Posidonia oceanica* (classe seagrass) rappresenta uno dei descrittori ecologici e bionomici più sensibili per la valutazione dello stato di conservazione dell'habitat (Pergent-Martini et al., 2005; Montefalcone, 2009). Il calcolo duplice dell'indicatore – attraverso la misura della profondità massima alla MMU e della profondità al 99° percentile – ha confermato in tutte le aree di studio la necessità di applicare un filtro geostatistico per la rimozione degli artefatti acustici (Tabella 14).

Tabella 14. Confronto delle profondità del limite inferiore (espresse in metri) calcolate tramite estrazione del valore massimo assoluto e del 99° percentile. Coordinate dei punti di massima profondità del limite inferiore calcolati sul 99° percentile; sistema di riferimento metrico RDN2008 / UTM zone 32N

Area	rule set / sorgente	Limite assoluto	Limite 99° percentile	Coord. X	Coord. Y
Genova_03	Atlante degli Habitat	34.83	32.35	412049.4	4851534.3
Genova_03	Ottimizzato – 3C	53.25	33.36	411920.0	4851908.6
Genova_03	Ottimizzato – 4C	53.25	33.35	411918.0	4851484.0
Genova_03	Rev10	54.07	36.38	411756.2	4852048.8
Genova_19	Atlante degli Habitat	29.48	28.10	535726.4	4898271.8
Genova_19	Ottimizzato – 3C	34.22	33.49	535404.4	4898227.4
Genova_19	Ottimizzato – 5C	41.42	32.25	535286.0	4898292.6
Genova_19	Rev10	46.49	39.85	535384.8	4898139.0
Genova_20	Atlante degli Habitat	30.26	28.37	537747.0	4897706.0
Genova_20	Ottimizzato – 3C	24.37	23.48	537626.2	4897746.0
Genova_20	Ottimizzato – 4C	24.37	23.53	537433.2	4897753.0
Genova_20	Rev10	41.73	40.85	537192.0	4897641.2
Levanto	Atlante degli Habitat	27.88	24.86	547371.0	4890751.5
Levanto	Originale – 3C	28.63	25.57	547623.0	4890610.0
Levanto	Ottimizzato – 3C	27.40	25.71	547610.0	4890607.5
Levanto	Ottimizzato – 4C	27.20	26.48	547568.0	4890656.0
Levanto	Ottimizzato – 5C	27.38	26.26	547546.0	4890664.5
Levanto	Rev10	27.35	24.09	547573.5	4890644

Per il modello Ottimizzato a 3 classi nell'area di Levanto, il limite inferiore è stato individuato alla quota batimetrica di 25,7 m. L'estrazione del valore massimo assoluto, vulnerabile agli outlier, posiziona il limite a 27,4 m, con uno scostamento pari a 1,7 m. Tali valori risultano invariati nelle classificazioni con diverso *rule set* e con diverso numero di classi, indicando una stabilità del risultato della classificazione, e risultano anche vicini al valore stimato sulla carta bionomica dell'atlante degli habitat (Coppo et al., 2020).

Nell'area Genova_03, si evidenzia un salto anomalo di 20 m nella profondità assoluta calcolata sulla classificazione del multibeam (53,3–54,1 m di profondità) rispetto al limite assoluto identificato nella carta bionomica di riferimento (pari a 34,8 m). Questa discrepanza marcata rappresenta dei valori non realistici rispetto alle profondità medie del limite inferiore delle praterie di *P. oceanica* nel Mar Ligure noti in letteratura. Tuttavia, l'applicazione del filtro statistico del 99° percentile risulta in questi casi efficace per rientrare entro il valore stimato con l'atlante degli habitat (Coppo et al., 2020).

Le aree di Genova_19 e Genova_20 presentano dei valori di profondità per il limite inferiore marcatamente anomali nella classificazione con *rule set* Rev10 (41,7 m a Genova_20 e 46,5 a Genova_19), mentre il risultato prodotto con il *rule set* Ottimizzato restituisce dei limiti inferiori che si discostano di circa 5 m dai valori stimati sulla carta bionomica (Tabella 14). Nell'area Sar_W_62 non è stato possibile determinare il limite inferiore della prateria poiché è risultata estendersi, seppur con andamento discontinuo, oltre il margine più profondo dell'area mappata, rendendo priva di significato ecologico la determinazione del limite più profondo presente nella carta.

Dal confronto dei risultati emergono due dinamiche non dipendenti dalla configurazione del *rule set* adottato. In primo luogo, si osserva come il modello acustico tenda generalmente a espandere il range batimetrico noto dell'habitat rispetto alla cartografia di riferimento. L'introduzione della profondità media nello spazio delle feature ha restituito una classificazione più realistica, in relazione alla stratificazione verticale, nelle aree di Genova_19 e Genova_20 in cui si è riscontrata bassa accuratezza della classificazione. Eppure, anche il *rule set* Ottimizzato ha restituito degli scostamenti ecologicamente rilevanti sul limite inferiore, rispetto ai confini regolari rappresentati nella cartografia di mesoscala.

Tale discrepanza è da attribuirsi al diverso grado di dettaglio ed alla maggiore risoluzione con cui è stata elaborata la classificazione MBES. Tuttavia, con i dati acquisiti nel presente studio,

non può escludersi un degrado dell'accuratezza della classificazione nella zona di margine del limite inferiore delle praterie, il quale è frequentemente sfumato e quindi complesso da delineare gli algoritmi di apprendimento automatico. D'altra parte, la discordanza nella resa cartografica, che si osserva tra le cinque aree di studio, può rappresentare l'effetto risultante dell'errore sistematico nel campionamento dei punti di addestramento, della diversificata complessità morfologica e dei margini paesaggistici nelle differenti aree. Tuttavia, per determinare sperimentalmente il contributo relativo di queste sorgenti di variabilità nel dato ecologico stimato, sarebbe necessaria l'acquisizione di verità a mare *in situ*, che permettano di confermare l'accuratezza della classificazione ad una scala coerente con quella del dato utilizzato.

CONCLUSIONI

Il presente studio ha valutato l'efficacia e la robustezza dell'approccio Geographic Object-Based Image Analysis (GEOBIA) applicato a dati acustici ad alta risoluzione (MBES) per la cartografia bionomica degli habitat costieri, con particolare focus sulle praterie di *Posidonia oceanica*. L'analisi condotta su cinque aree di studio del Mar Ligure e del Mar di Sardegna ha permesso di valutare la robustezza dell'architettura algoritmica messa a punto, evidenziandone le potenzialità analitiche in termini di confrontabilità del risultato rispetto a dati geospaziali di provenienza da fonti diversificate ed in termini di potere risolutivo della mappatura.

In primo luogo, l'ottimizzazione del modello di classificazione ha dimostrato come la selezione delle feature acustiche e morfometriche sia un parametro critico. La costruzione dello spazio delle feature ottimale, definito tramite le analisi statistiche dell'informazione e della collinearità tra le variabili, ha permesso di incrementare la specializzazione del modello di classificazione nell'area pilota di Levanto. L'integrazione, secondo un criterio ecologico, della media batimetrica nello spazio delle feature nel *rule set* Ottimizzato ha risolto le criticità di misclassificazione registrate dal modello di base (Rev10), correggendo la polarizzazione anomala della classe seagrass verso profondità incompatibili con la nicchia ecologica della fanerogama marina. Questo affinamento ha garantito una contestualizzazione spaziale coerente, restituendo mosaici bionomici aderenti alla distribuzione verticale degli habitat.

I risultati cartografici prodotti hanno dimostrato una resa efficace a fronte dei metodi stringenti adottati, quali l'addestramento dell'algoritmo di classificazione tramite interpretazione esperta dello stesso dato acustico utilizzato per la classificazione, e la dimensione modesta dello spazio delle feature adottato. Le configurazioni di base a 3 classi (roccia, sedimento, seagrass) hanno dimostrato elevata robustezza e consistente trasferibilità spaziale, raggiungendo valori di accuratezza globale (OA) costantemente superiori all'85% e picchi del 95%. Tuttavia, l'incremento del dettaglio semantico – attraverso l'introduzione della classe matite morta e la suddivisione dei sedimenti (modelli a 4 e 5 classi) – ha innescato un decremento strutturale delle metriche di concordanza globale. Tale flessione non rappresenta un fallimento della architettura algoritmica, ma evidenzia la riduzione strutturale dell'accuratezza globale a fronte della

maggiore risoluzione tematica, laddove ogni classe aggiuntiva richiede la determinazione di ulteriori soglie di distinzione, con conseguente incremento dell'errore associato (Tulloch et al., 2017).

Nonostante il calo di accuratezza rispetto alla classificazione in tre classi, nella classificazione con l'introduzione della classe matte si è superata la soglia standard del 85% in Overall Accuracy in 3 siti su 5, con valori maggiori del 90% nelle prove elaborate dell'area pilota di Levanto.

Il confronto geospaziale con la cartografia bionomica di riferimento (Coppo et al., 2020; EcoAtlante ISPRA, s.d.) ha restituito un quadro di elevata concordanza nelle classificazioni caratterizzate da alta accuratezza. Se da un lato è emersa una solida coerenza topologica alla mesoscala, dall'altro le apparenti discordanze registrate (in particolare per la classe matte) hanno rivelato il fenomeno del *mismatch* di scala. Il classificatore MBES si è dimostrato capace di superare la generalizzazione delle carte bionomiche di riferimento, individuando patch e discontinuità all'interno di aree precedentemente mappate come omogenee, come confermato dagli elevati valori di User Accuracy.

Il potere risolutivo del dato acustico è stato ulteriormente verificata con l'analisi del limite inferiore della prateria a *P. oceanica*. Il modello GEOBIA ha frequentemente esteso il range batimetrico dell'habitat rispetto alla cartografia istituzionale, mettendo in evidenza porzioni di prateria profonda che potrebbero essere state omesse dai censimenti precedenti. Ciononostante, la marcata variabilità di tale limite tra le diverse aree di studio evidenzia l'incertezza associata all'interpretazione ecologica del dato acustico. Particolare cautela deve essere tenuta in riferimento all'accuratezza della classificazione in presenza di margini sfumati, dove le soglie costruite dagli algoritmi di apprendimento automatico sui dati acustici rischiano di non coincidere perfettamente con i reali confini ecologici delle associazioni bentoniche.

Inoltre, la classificazione ha riscontrato difficoltà nel separare la canopy fogliare di *P. oceanica* su roccia dalla topografia del substrato, come emerso nel mosaico di *P. oceanica* tra e su roccia dell'area di Levanto, dove la classificazione delle patch di prateria insediate sul rilievo roccioso ha registrato le performance più basse. La morfologia scoscesa e irregolare del substrato roccioso prevale nelle derivate morfometriche, mascherando la microtopografia generata dalla chioma della pianta. In assenza di un forte contrasto di backscatter, l'algoritmo non dispone di chiavi spettrali sufficienti per separare la componente biotica da quella abiotica.

Parimenti, l'impossibilità dell'algoritmo di discriminare la patch di *Cymodocea nodosa*, rappresentata sul Nuovo Atlante degli Habitat Marini all'interno dell'area di Levanto, dal

sedimento sabbioso circostante conferma il limite fisico dell'acustica marina già dibattuto in letteratura: l'esigua biomassa fogliare e l'assenza di un apparato rizomatoso massiccio e lignificato determinano per *C. nodosa* un'impedenza acustica e un volume di scattering insufficienti per generare un contrasto spettrale rilevabile alle frequenze MBES standard (Fakiris et al., 2019; Monnier et al., 2021).

Alla luce di tali evidenze, nonostante il conseguito raggiungimento di un elevato contenuto informativo nelle carte prodotte con addestramento basato su interpretazione esperta del dato, si rimarca come l'adozione di dati di verità a mare in fase di addestramento possa magnificare l'accuratezza ed il potere risolutivo a piccola scala della classificazione, migliorando la prestazione della classificazione rispetto alle criticità riscontrate, avendo un dato acustico di altissima risoluzione come nel presente studio. Particolare attenzione, tuttavia, deve essere adottata della rappresentatività esaustiva dell'eterogeneità tematica presente nell'area di indagine. Il dato di addestramento per il modello di classificazione deve campionare come minimo l'eterogeneità al livello delle classi oggetto di studio; ma, nel caso di classi aggregate (come nel presente studio per la classificazione in sediment, seagrass e rock), è preferibile garantire una copertura delle firme per ogni sottocategoria bionomico-morfologica esistente nell'ambiente studiato, quale possono essere tutte le classi di habitat tematizzate nel Nuovo Atlante degli Habitat Marini (Coppo et al., 2020). In assenza di un addestramento realmente esaustivo dell'eterogeneità ambientale a cui può essere sensibile lo scandaglio acustico, la classificazione derivata tende inevitabilmente ad avere dei crolli in accuratezza del produttore relativamente alle sottocategorie non campionate.

In conclusione, l'altissima risoluzione spaziale oggi raggiunta dai più moderni sensori dei sistemi MBES, con pixel di 20 cm, impiegati nel presente studio nell'ambito del progetto MER per la mappatura delle praterie di fanerogame marine in tutte le coste nazionali, restituisce un dettaglio morfologico senza precedenti, spingendosi fino alla delineazione dei singoli ciuffi di fasci fogliari, alla scala metrica. Questo livello di dettaglio costituisce la premessa per permettere di incrementare ulteriormente il dettaglio della classificazione ed arrivare a mappare con accuratezza le singole macchie vegetate, nelle praterie discontinue di *P. oceanica*. Questo salto di scala apre nuove potenzialità e per lo studio dei processi ecologici paesaggistici, e per il monitoraggio dello stato ecologico dei posidonieti, tutelati come Habitat prioritario (1120*) ai sensi della Direttiva Habitat (92/43/CEE).

Operare a scala submetrica, inoltre, impone di integrare l'effetto dei processi fenologici rilevabili, quali la marcata stagionalità della copertura fogliare di *P. oceanica*, che presenta massimo sviluppo in estate, e successiva riduzione nel periodo autunnale. A questa scala di mappatura, la marcata stagionalità a cui è soggetta la *P. oceanica* può generare firme acustiche differenti per lo stesso habitat, introducendo potenziali distorsioni spaziali nella definizione dei margini degli habitat e influenzando la trasferibilità temporale dei modelli di classificazione. Le prospettive di miglioramento della mappatura degli habitat per la ricerca ecologica e per la cartografia bionomica risiedono dunque nella capacità di integrare la fenologia e l'ecologia delle unità bionomiche nelle tecniche geoinformatiche e statistiche di processamento dei dati.

BIBLIOGRAFIA

Blaschke T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* (2010) 65:2-16.

Breiman L. Random forests. *Mach. Learn.* (2001) 45:5-32.

Brown CJ, Smith SJ, Lawton P, and Anderson JT. Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. *Estuar. Coast. Shelf Sci.* (2011) 92:502-520.

Congalton RG, and Green K. *Assessing the accuracy of remotely sensed data: principles and practices.* (2009), CRC Press, Boca Raton.

Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J, Wichmann V, and Böhner J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* (2015) 8:1991-2007.

Coppo S, Diviacco G, and Montepagano E. *Nuovo Atlante degli habitat marini della Liguria: Descrizione e cartografia delle praterie di Posidonia oceanica e dei principali popolamenti marini costieri.* (2020), Regione Liguria, Genova.

Diesing M, Mitchell P, and Stephens D. Image-based seabed classification: what can we learn from terrestrial remote sensing? *ICES J. Mar. Sci.* (2016) 73:2425-2441.

Drăguț L, Csillik O, Eisank C, and Tiede D. Estimating scale parameters for object-based image analysis through the local variance of image features. *ISPRS J. Photogramm. Remote Sens.* (2014) 89:93-101.

Espindola GM, Camara G, Reis IA, Bins LS, and Monteiro AM. Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. *Int. J. Remote Sens.* (2006) 27:3035-3040.

Fakiris E, Blondel P, Papatheodorou G, Christodoulou D, Dimas X, Georgiou N, Kordella S, Dimitriadis C, Rzhhanov Y, and Geraga M. Multi-frequency, multi-sonar mapping of shallow habitats—Efficacy and management implications in the National Marine Park of Zakynthos, Greece. *Remote Sens.* (2019) 11:461.

Getreuer P. Linear Methods for Image Interpolation. *Image Process. On Line.* (2011) 1:238-259.

Gholamy A, Kreinovich V, and Kosheleva O. Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation. *Dept. Comput. Sci. Tech. Rep.* (2018) 1209.

GRASS Development Team. r.fillnulls manual. In GRASS GIS 8.4 Reference Manual. (2024), Open Source Geospatial Foundation. <https://grass.osgeo.org/grass-stable/manuals/r.fillnulls.html>

Ierodionou D, Laurenson L, Leblanc M, Stagnitti F, Nesq M, Felsing M, and Venema S. Marine benthic habitat mapping using multibeam data, georeferenced video and image classification techniques in Victoria, Australia. *J. Spat. Sci.* (2007) 52:93-104.

Ierodionou D, Schimel ACG, Kennedy D, Monk J, Gaylard G, Young M, Diesing M, and Rattray A. Combining pixel and object based image analysis of ultra-high resolution multibeam bathymetry and backscatter for habitat mapping in shallow marine waters. *Mar. Geophys. Res.* (2018) 39:271-285.

Ismail K, Huvenne VA, and Masson DG. Objective automated classification technique for marine habitat mapping using remote sensing data. *Mar. Geophys. Res.* (2015) 36:227-241.

ISPRA (s.d.). EcoAtlante - Tipologie degli ecosistemi marini. Servizio cartografico FeatureServer, Portale SINACloud - Sistema Informativo Nazionale Ambientale. Disponibile all'indirizzo:
https://sinacloud.isprambiente.it/arcgisina/rest/services/sina_consumo_suolo/tipologie_ecosistemi_marini/FeatureServer (Consultato il: 3 marzo 2026)

Janowski Ł, Trzcińska K, Tęgowski J, Kruss A, Rucinska-Zjadacz M, and Porskamp P. Nearshore Benthic Habitat Mapping Based on Multi-Frequency, Multibeam Echosounder Data Using a Combined Object-Based Approach: A Case Study from the Rowy Site in the Southern Baltic Sea. *Remote Sens.* (2018) 10:1983.

Janowski Ł, Tęgowski J, and Nowak J. Seafloor mapping based on multibeam echosounder bathymetry and backscatter data using Object-Based Image Analysis: a case study from the Rewal site, the Southern Baltic. *Oceanol. Hydrobiol. Stud.* (2018) 47:248-259.

Johnson B, and Xie Z. Unsupervised image segmentation evaluation and refinement using a multi-scale approach. *ISPRS J. Photogramm. Remote Sens.* (2011) 66:473-483.

Katsanevakis S, Stelzenmüller V, South A, Sørensen TK, Jones PJS, Kerr S, Badalamenti F, Makris P, Alvarez P, Fiorentino D, Bograd A, Dokken G, Alberín I, Prinz M, Erra F, and Tericho G. Ecosystem-based marine spatial management: Review of concepts, policies, tools, and critical issues. *Ocean Coast. Manage.* (2011) 54:807-820.

Kursa MB, and Rudnicki WR. Feature selection with the Boruta package. *J. Stat. Softw.* (2010) 36:1-13.

Lamarche G, and Lurton X. Recommendations for improved and coherent acquisition and processing of backscatter data from seafloor-mapping sonars. *Mar. Geophys. Res.* (2018) 39:5-22.

Lecours V, Dolan MF, Micallef A, and Lucieer VL. A review of marine geomorphometry, the quantitative study of the seafloor. *Hydrol. Earth Syst. Sci.* (2016) 20:3207-3244.

- Micallef A, Le Bas TP, Huvenne VA, Blondel P, Hühnerbach V, and Deidun A. A multi-method approach for benthic habitat mapping of shallow coastal areas with high-resolution multibeam data. *Cont. Shelf Res.* (2012) 39:14-26.
- Mitasova H, and Mitas L. Interpolation by Regularized Spline with Tension: I. Theory and Implementation. *Math. Geol.* (1993) 25:641-655.
- Monnier B, Pergent G, Mateo MÁ, Clabaut P, and Pergent-Martini C. Sizing the carbon sink associated with *Posidonia oceanica* seagrass meadows using very high-resolution seismic reflection imaging. *Mar. Environ. Res.* (2021) 170:105415.
- Montefalcone M. Ecosystem health assessment using the Mediterranean seagrass *Posidonia oceanica*: A review. *Ecol. Indic.* (2009) 9:595-604.
- Olaya V. Basic Land-Surface Parameters, in *Geomorphometry: Concepts, Software, Applications* (Hengl T, and Reuter HI, Eds.). (2006), pp 141-169, Elsevier, Amsterdam.
- Olofsson P, Foody GM, Herold M, Stehman SV, Woodcock CE, and Wulder MA. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* (2014) 148:42-57.
- Parnum IM. Benthic habitat mapping using multibeam sonar systems. (2007), PhD Thesis, Curtin University of Technology, Perth.
- Pérès JM, and Picard J. Nouveau manuel de bionomie benthique de la mer Méditerranée. *Recl. Trav. Stn. Mar. Endoume.* (1964) 31:1-137.
- Pergent-Martini C, Rico-Raimondino V, and Pergent G. Descriptors of *Posidonia oceanica* meadows: Use and application. *Ecol. Indic.* (2005) 5:213-230.
- Rende SF, Bosman A, Di Mento R, Bruno F, Lagudi A, Irving AD, and Giacobbe S. Ultra-High-Resolution Mapping of *Posidonia oceanica* (L.) Delile Meadows through Acoustic Technology and Object-Based Image Analysis (OBIA). *J. Mar. Sci. Eng.* (2020) 8:647.
- Rende SF, Bosman A, Di Paola C, Trabucco B, and Irving AD. Assessing Seagrass Restoration Using Ultra-High-Resolution Acoustic Mapping and OBIA. *J. Mar. Sci. Eng.* (2022) 10:1989.
- Riley SJ, De Gloria SD, and Elliot R. A Terrain Ruggedness that Quantifies Topographic Heterogeneity. *Intermt. J. Sci.* (1999) 5:23-27.
- SAGA GIS Source Code Repository. Modulo `ta_morphometry` (Slope, Aspect, Curvature). (2026). SourceForge Repository. https://sourceforge.net/p/saga-gis/code/ci/master/tree/saga-gis/src/tools/terrain_analysis/ta_morphometry/Morphometry.cpp
- Steward JS, Virnstein RW, Morris LJ, and Lowe EF. Setting seagrass depth, coverage, and light targets for the Indian River Lagoon system, Florida. *Estuaries.* (2005) 28:923-935.
- Tulloch VJ, et al. Trade-offs between data resolution, accuracy, and cost when choosing information to plan reserves for coral reef ecosystems. *Biol. Conserv.* (2017) 210:10-20.

UNEP/MAP - RAC/SPA. Handbook for gathering and processing data on the habitat types and species. (2015), Regional Activity Centre for Specially Protected Areas, Tunis.

Ventura D, Bonifazi A, Gravina MF, Belluscio A, and Ardizzone G. Coastal benthic habitat mapping and monitoring by integrating aerial and water surface low-cost drones. *Front. Mar. Sci.* (2023) 9:1096594.

Whiteside TG, Boggs GS, and Maier SW. Area-based and location-based validation of classified image objects. *Int. J. Appl. Earth Obs. Geoinf.* (2014) 28:117-130.

Wicaksono P, et al. Benthic Habitat Mapping Model and Cross Validation Using Machine-Learning Classification Algorithms. *Remote Sens.* (2019) 11:1279.

Wilson JP, Gallant JC, and Hutchinson MF. Measuring terrain complexity, in *Geomorphometry: Concepts, Software, Applications* (Hengl T, and Reuter HI, Eds.). (2007), pp 141-169, Elsevier, Amsterdam.

Zevenbergen LW, and Thorne CR. Quantitative analysis of land surface topography. *Earth Surf. Process. Landf.* (1987) 12:47-56.

RINGRAZIAMENTI

Desidero rivolgere un sentito ringraziamento a Michelle Wagner e Nick Rackebrandt per il prezioso e paziente supporto fornito nella complessa gestione dei dati e per avermi trasmesso le metodologie analitiche essenziali in questo studio. La loro disponibilità e dedizione sono state determinanti per la realizzazione di questo elaborato. Un sincero ringraziamento va inoltre a tutto il gruppo di lavoro di Fugro Marine Germany per il costante e cordiale affiancamento.

APPENDICE

ALLEGATO 1

Schema riepilogativo degli algoritmi e delle impostazioni nel *rule set* Ottimizzato. I parametri non rappresentati sono mantenuti al valore di default del programma.

Algoritmo	parametri	valore	Descrizione
remove object variable data	Domain	execute	rimuovi i dati di tutte le variabili da tutti gli oggetti immagine
delete image object level	Domain	execute	elimina tutti i livelli
remove classification	Domain	image object level	a New Level: rimuovi la classificazione
	Level	New Level	
delete samples	Domain	execute	elimina tutti i campioni
	Class list	all	
vector-based segmentation	level name	New Level	segmentazione vettoriale: creazione di 'New Level' usando il layer cover
	create objects in outside area	No	
	Thematic Layers	[cover]	
update region	domain	image object level	a New Level: imposta la regione corrente = COVER
	variable	COVER	
	Mode	from domain	
Multiresolution segmentation	domain	image object level	all'interno di COVER: segmentazione multiscala: 10 [shape:0.1 compct:0.5]
	map	from parent	
	region	COVER (variable)	
	Level usage	use current	
	compatibility mode	latest version	
	Image Layer weights	BS(1), TRI(1)	
	Scale parameter	12	
	shape	0,1	
	compactness	0,3	
convert thematic objects to samples	domain	image object level	converte in campioni gli oggetti tematici del layer samp_train (attributo "Class Name")
	Thematic Layer	samp_train	
	Class name attribute	Class Name	
	Match class names	[matte, sediment], [sediment F, sediment]	
supervised classification	Domain	image object level	classificazione supervisionata (train) con Random Forest utilizzando

			BS, TRI, BATHY, TC, S, GC (200x)
	Level	New Level	
	Samples only	No	
	Operation	Train	
	Configuration	RT	
	Feature Space Source	object based	
	Feature Space Features	[Mean BS, Standard deviation TRI, Mean BATHY, Mean E, Mean GC, Standard deviation BATHY, Standard deviation BS, Standard deviation N, Area]	
	Classifier Type	Random Trees	
	Classifier Depth	0	
	Classifier Min sample count	1	
	Classifier Use surrogates	No	
	Classifier Max categories	16	
	Classifier Active variables	10	
	Classifier Max tree number	500	
	Classifier Termination criteria type	Terminate learning by max tree number	
supervised classification	Domain	image object level	applica la classificazione supervisionata a New Level
	Level	New Level	
	Samples only	No	
	Operation	Apply	
	Configuration	RT	
	Feature Space Source	object based	
supervised classification	Operation	Query	interroga il modello Random Forest
	Configuration	RT	
	Query Information Type	Random Trees	
	Query Information Export Importance Table	{Workspc.OutputRoot}/results/{Project.Name}_originale_3C_Importance.{Ext}	
	Number of cycles	1	
read thematic attribute	Domain	image object level	Crea la variabile VAL con l'attributo 'Class_Name'
	Level	New Level	
	Samples only	No	
	Thematic layer	samp_val	
	Thematic layer attribute	Class_Name	
	Variable	VAL	
export confusion matrix	Domain	image object level	esporta la matrice di confusione basata sulle variabili degli oggetti
	Level	New Level	
	Samples only	No	
	Export mode	Static export item	
	Export path	{Workspc.OutputRoot}/results/{Project.Name}_originale_3C_Importance.{Ext}	

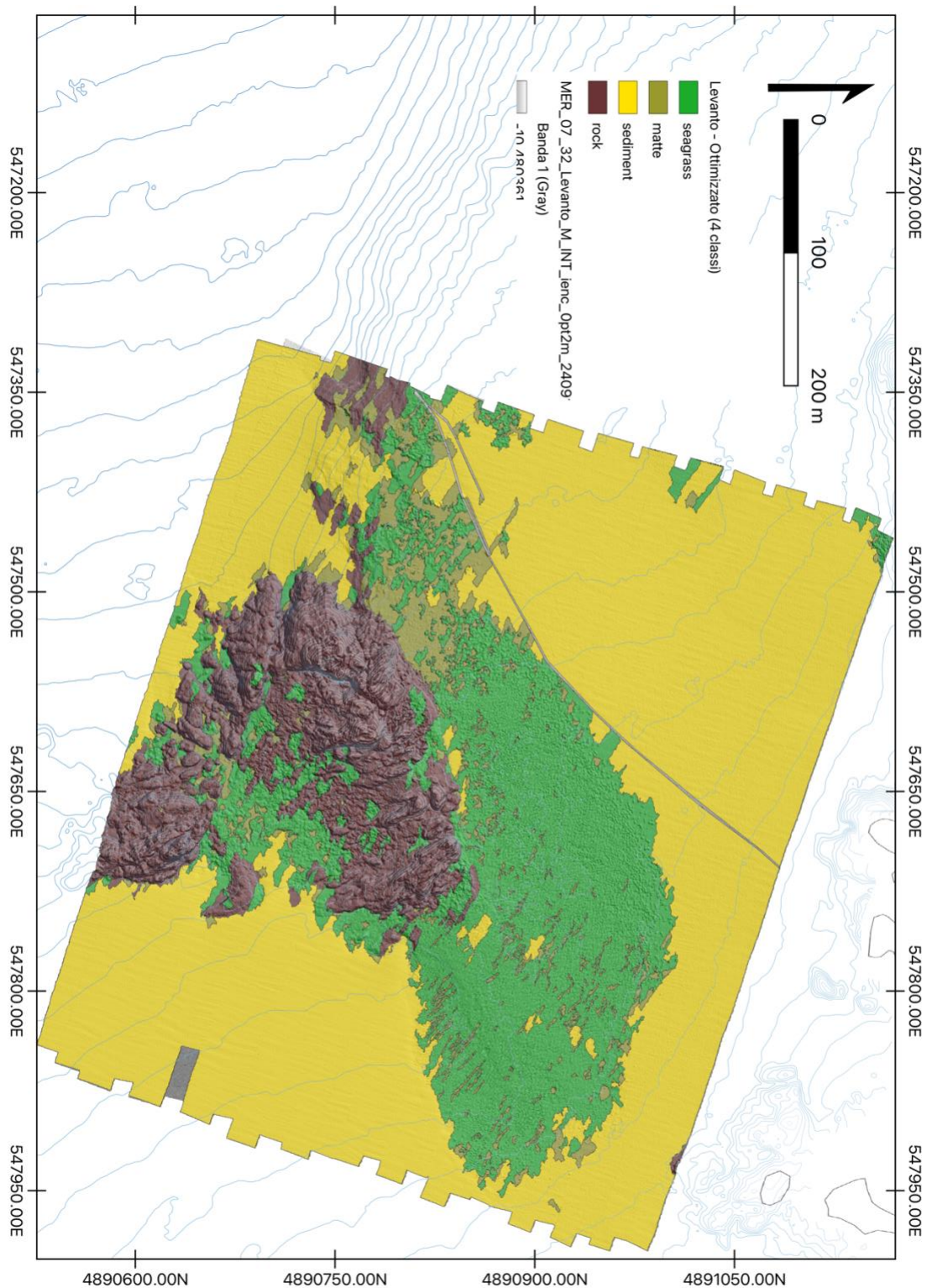
	Confusion matrix input data	image object level	
	Confusion matrix class filter	All	
	Confusion matrix ground truth	Object variable	
	Object variable	VAL	
	Add detailed statistics to export	Yes	
update feature list	Domain	execute	aggiungi le feature: [Area, Compactness, Class name(0,0), Mean A, ...]
	Variable	New Level	
	Mode	add features	
	Features	[Area, Compactness, Class name(0,0), Mean A, Mean BATHY, Mean BS, Mean E, Mean GC, Mean N, Mean S, Mean TC, Mean TRI, Standard deviation BATHY, Standard deviation BS, Standard deviation E, Standard deviation GC, Standard deviation N, Standard deviation S, Standard deviation TC, Standard deviation TRI, Main direction, Length, Sample class name]	
export vector layer	Domain	image object level	esporta le geometrie degli oggetti in pre_refinement
	Level	New Level	
	Export mode	Static export item	
	Export item name	Classification	
	Export path	{Workspc.OutputRoot}/results /{:Item.Name}/pre_refinements/{:Project.Name}_originale_3C.{:Ext}	
	Export data use fracture list	No	
	Export data shape type	Polygons	
	Export data export type	Raster	
	Export format	Shapefile (*.shp)	
merge region	Domain	image object level	unisce gli oggetti adiacenti di classe uguale
	Level	New Level	
	Class filter	none	
	Fusion super objects	No	
	Thematic Layers	[]	
	Merge	By Class	
grow region	Domain	image object level	(classe matte, seagrass): unifica gli oggetti rock, sediment F, sediment con area < 500 pixel
	Level	New Level	
	Class filter	matte, seagrass	

	Candidates classes	rock, sediment F, sediment	
	Fusion super objects	No	
	Fitting function	Area	
	Comparison	<	
	Fitting threshold	500	
	Use Thematic Layers	No	
grow region	Domain	image object level	(classe rock, sediment F, sediment): unifica gli oggetti matte, seagrass con area < 200 pixel
	Level	New Level	
	Class filter	rock, sediment F, sediment	
	Candidates classes	matte, seagrass	
	Fusion super objects	No	
	Fitting function	Area	
	Comparison	<	
	Fitting threshold	200	
	Use Thematic Layers	No	
grow region	Domain	image object level	(classe rock): unifica gli oggetti sediment F, sediment con area < 500 pixel
	Level	New Level	
	Class filter	rock	
	Candidates classes	sediment F, sediment	
	Fusion super objects	No	
	Fitting function	Area	
	Comparison	<	
	Fitting threshold	500	
	Use Thematic Layers	No	
grow region	Domain	image object level	(classe sediment F, sediment): unifica gli oggetti rock con area < 200 pixel
	Level	New Level	
	Class filter	sediment F, sediment	
	Candidates classes	rock	
	Fusion super objects	No	
	Fitting function	Area	
	Comparison	<	
	Fitting threshold	200	
	Use Thematic Layers	No	
export vector layer	Domain	image object level	esporta le geometrie degli oggetti in Classification
	Level	New Level	
	Export mode	Static export item	
	Export item name	Classification	
	Export path	{Workspc.OutputRoot}/results/{:Item.Name}/{:Project.Name}_originale_3 C. {:Ext}	
	Export data use fracture list	No	
	Export data shape type	Polygons	
	Export data export type	Raster	
	Export format	Shapefile (*.shp)	

ALLEGATO 2

Tavola 1: Area Levanto

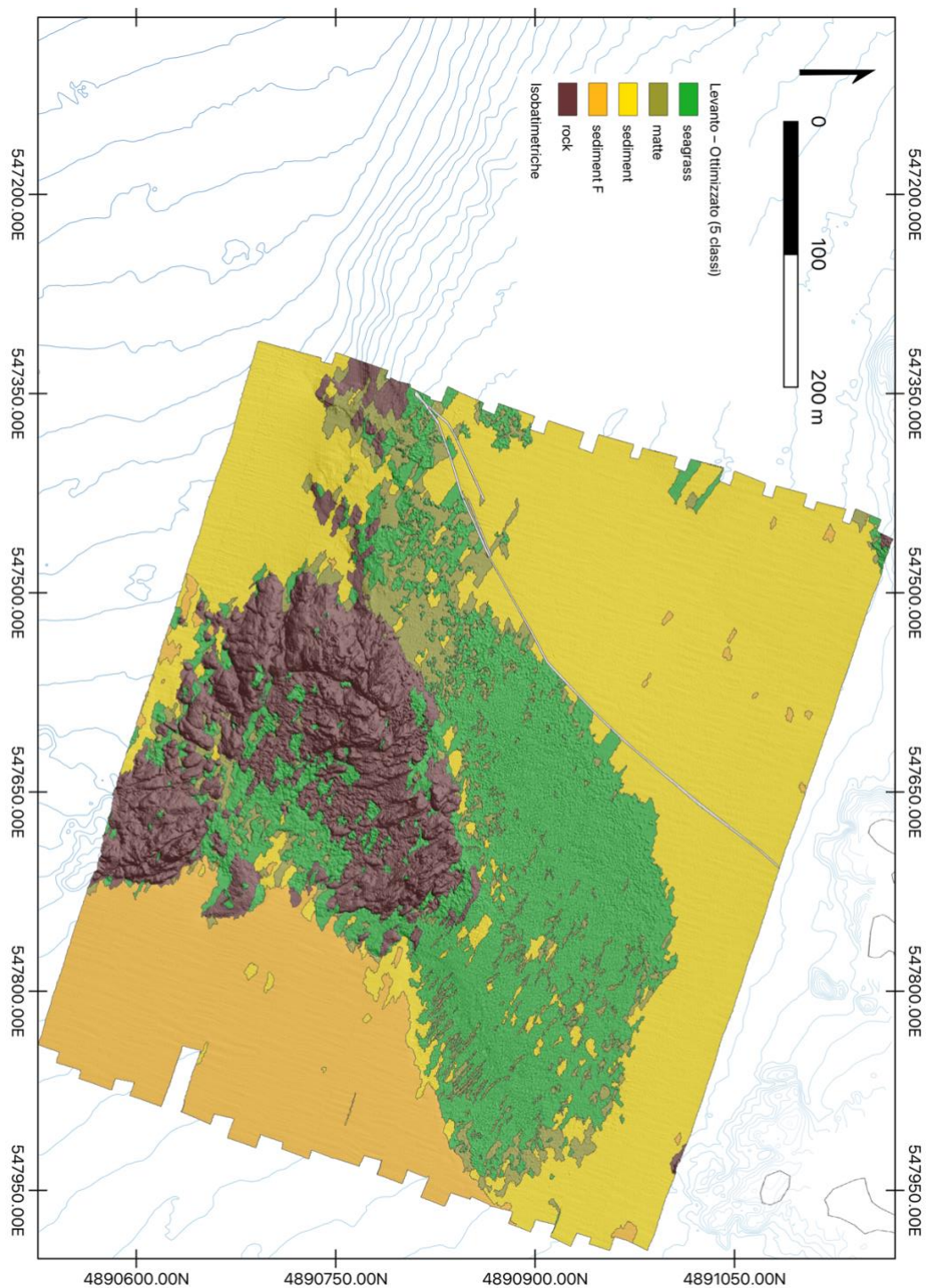
Classificazione supervisionata in 4 classi del segnale acustico acquisito tramite multibeam (MBES). Le isobate sono rappresentate con intervallo di 1 m. Sistema di Riferimento: WGS84 / UTM zone 32N.



ALLEGATO 3

Tavola 2: Area Levanto

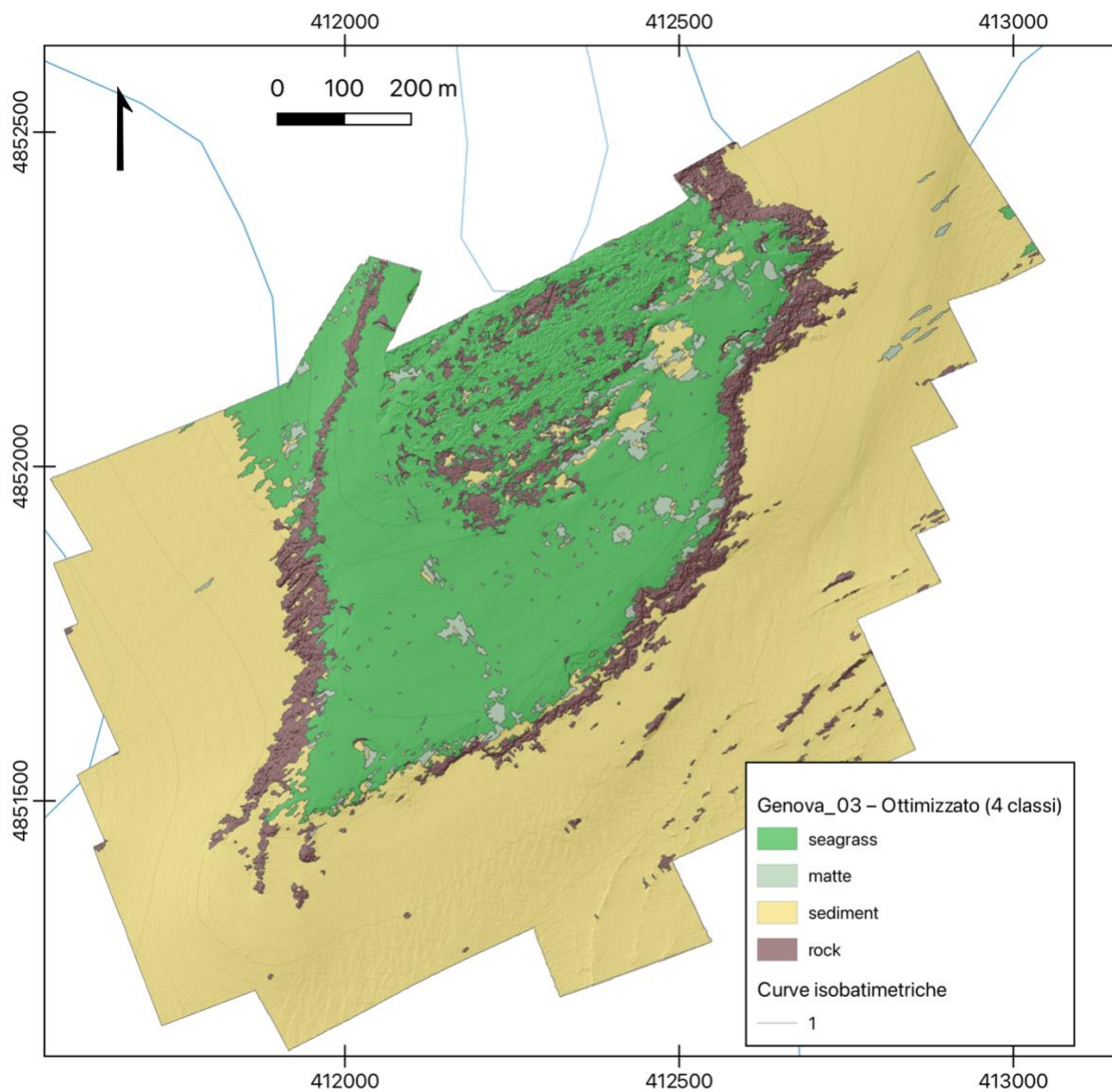
Classificazione supervisionata in 5 classi del segnale acustico acquisito tramite multibeam (MBES). Le isobate sono rappresentate con intervallo di 1 m. Sistema di Riferimento: WGS84 / UTM zone 32N.



ALLEGATO 4

Tavola 3: Area Genova_03

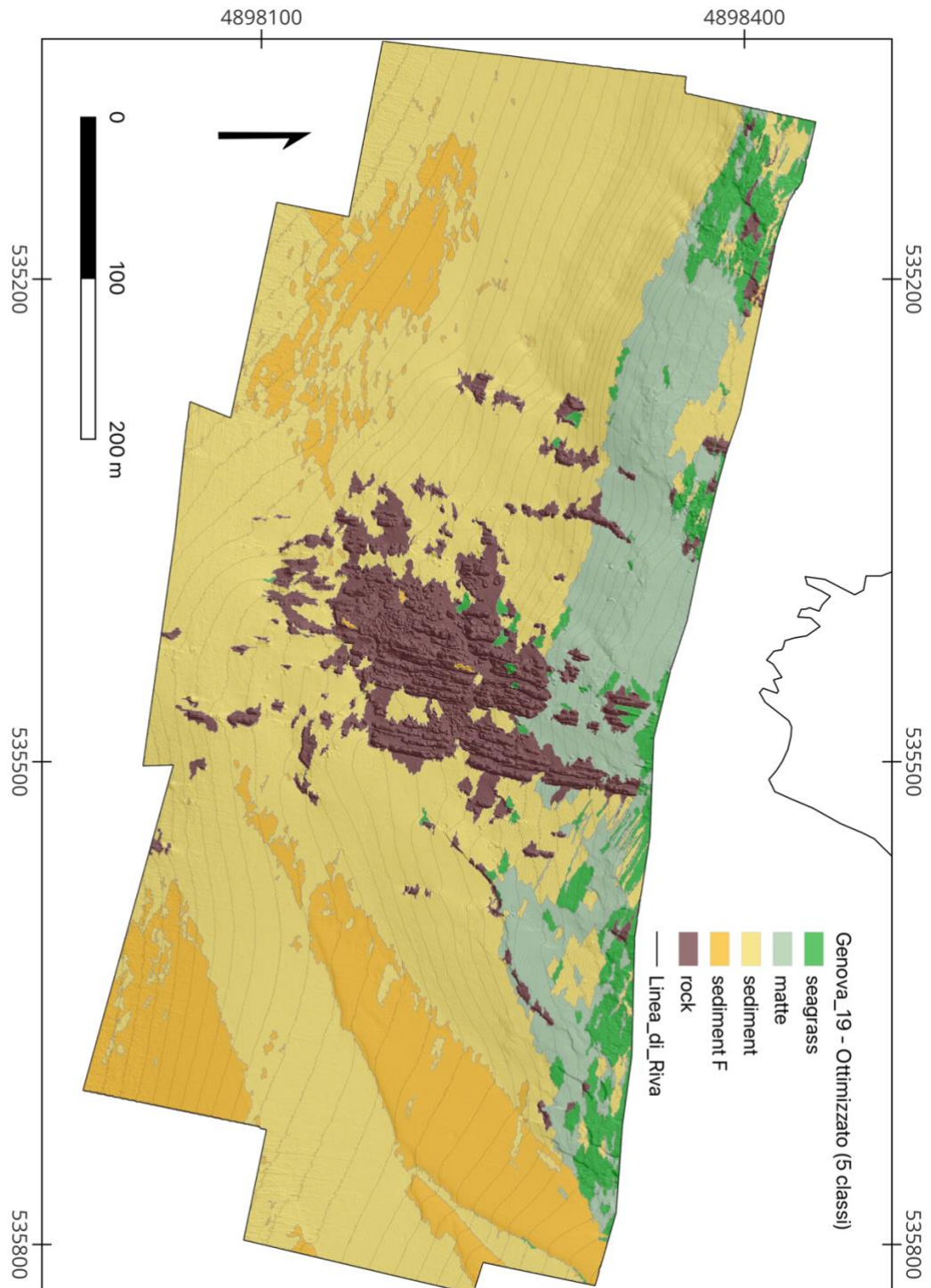
Classificazione supervisionata in 4 classi del segnale acustico acquisito tramite multibeam (MBES). Le isobate sono rappresentate con intervallo di 20 m. Sistema di Riferimento: WGS84 / UTM zone 32N.



ALLEGATO 5

Tavola 4: Area Genova_19

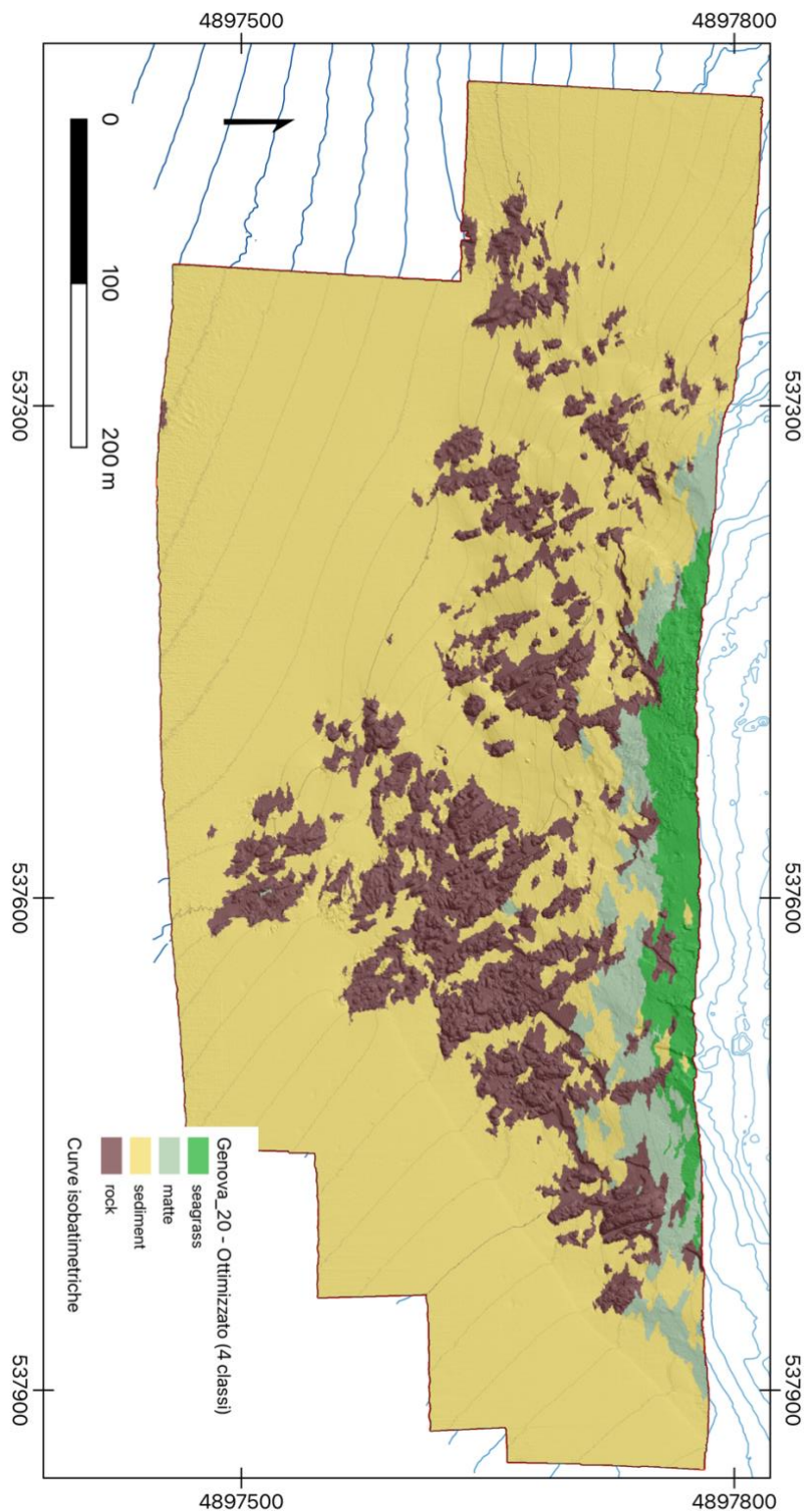
Classificazione supervisionata in 5 classi del segnale acustico acquisito tramite multibeam (MBES). Le isobate sono rappresentate con intervallo di 1 m. Sistema di Riferimento: WGS84 / UTM zone 32N.



ALLEGATO 6

Tavola 5: Area Genova_20

Classificazione supervisionata in 4 classi del segnale acustico acquisito tramite multibeam (MBES). Le isobate sono rappresentate con intervallo di 1 m. Sistema di Riferimento: WGS84 / UTM zone 32N.



ALLEGATO 7

Tavola 6: Area Sar_W_62

Classificazione supervisionata in 4 classi del segnale acustico acquisito tramite multibeam (MBES). Le isobate sono rappresentate con intervallo di 1 m. Sistema di Riferimento: WGS84 / UTM zone 32N.

