



UNIVERSITÀ DEGLI STUDI DI GENOVA

SCUOLA POLITECNICA

**DITEN – Dipartimento di Ingegneria Navale, Elettrica, Elettronica e delle
Telecomunicazioni**

CORSO DI LAUREA MAGISTRALE IN YACHT DESIGN

TESI DI LAUREA MAGISTRALE

**Analisi dei Dati AIS per la Generazione di Mappe di Rischio di
Collisione in Aree Costiere**

Relatori:

Prof. Michele Martelli

Prof. Tomaso Gaggero

Correlatore:

Dott. Filippo Ponzini

Candidato:

Paolo Pedalino

A.A. 2025/2026

Alla gente di mare

Abstract

La tesi presenta una procedura per costruire mappe di criticità degli incontri fra unità navali a partire dai dati AIS. Il lavoro è stato svolto all'interno del progetto Interreg JASON e applicato al Golfo del Tigullio, per il 2023, e allo Stretto di Messina, per il 2022, il 2023 e il 2024.

Dai messaggi AIS sono estratti gli incontri fra coppie di unità a distanza inferiore a 500 m. Sul sottoinsieme cinematico sono calcolati otto indicatori di criticità, costruiti sulla distanza osservata e sulle grandezze cinematiche. La distribuzione spaziale e temporale degli indicatori è analizzata con mappe su griglia, indici di autocorrelazione spaziale di Moran e Getis-Ord, clustering DB-SCAN e un test di riallocazione a densità invariata. Una verifica euristica valuta la conformità delle manovre al COLREG.

La scelta dell'indicatore modifica poco le mappe di criticità aggregata e molto la selezione degli incontri più critici. La criticità aggregata segue la densità di traffico, mentre la criticità per incontro presenta una struttura spaziale non riproducibile dalla sola densità. Le due aree mostrano regimi di incontro diversi. Nel Tigullio la criticità si concentra presso Portofino e nella stagione estiva, con un profilo compatibile con un traffico prevalentemente diportistico. Nello Stretto oltre l'80 % degli incontri coinvolge un'unità ferma, le zone principali ricadono nelle aree portuali e la criticità è distribuita sull'intero anno; la posizione delle zone e le principali statistiche variano poco nei tre anni.

Gli indicatori descrivono la criticità degli incontri e non stimano il rischio, inteso come combinazione di frequenza e conseguenze, né una probabilità di collisione. Le zone individuate non sono state confrontate con dati di incidenti.

Indice

Abstract	2
1 Introduzione	15
1.1 Obiettivi della tesi	15
1.2 Motivazioni	16
1.3 Struttura del documento	17
2 Background e stato dell'arte	18
2.1 L'AIS come base dati	18
2.1.1 Definizione e quadro normativo	18
2.1.2 Principi di funzionamento	19
2.1.3 Contenuto informativo	20
2.1.4 L'AIS come sorgente dati per lo studio del traffico	20
2.1.5 Limiti e criticità	21
2.2 Il quadro normativo per la prevenzione delle collisioni	22
2.2.1 Il COLREG nel sistema internazionale della sicurezza marittima	22
2.2.2 Le regole rilevanti per l'interpretazione del rischio	22
2.2.3 Le due dimensioni del rischio di collisione	23
2.2.4 Dalle regole alle zone operative	24
2.3 Strumenti concettuali e metodologici per la valutazione del rischio	25
2.3.1 Indicatori geometrici: CPA, DCPA, TCPA	25
2.3.2 Ship domain e rilevazione delle near collision	27
2.3.3 Indici di rischio compositi	29
2.3.4 Aggregazione spaziale: Kernel Density Estimation pesata	30
2.3.5 Clustering density-based per l'identificazione degli hotspot	32
2.4 Analisi critica dello stato dell'arte	34
2.4.1 Introduzione: frammentazione terminologica e safety pyramid	34
2.4.2 Approcci geometrici basati su CPA, DCPA e TCPA	35
2.4.3 Indici di rischio compositi in letteratura	36
2.4.4 Approcci basati su densità: KDE e clustering density-based in letteratura	38
2.4.5 Modelli probabilistici	39

2.4.6	Approcci alternativi e spatial statistics	42
2.4.7	Approcci data-driven e deep learning	43
2.4.8	Eterogeneità della letteratura	43
2.5	Limiti della letteratura e research gap	46
3	Presentazione e Pre-elaborazione dei dati	47
3.1	La fonte grezza: dati AIS 2023 dell'area ligure	47
3.2	Il pre-processing dei dati grezzi	49
3.2.1	Consolidamento dell'archivio: dai CSV alla struttura unificata	49
3.2.2	Estrazione degli encounter sincroni	50
3.3	Il dataset di lavoro	51
3.3.1	Dimensione e struttura	51
3.3.2	Copertura spaziale e temporale	51
3.4	Raffinamento cinematico del dataset	52
3.4.1	Stabilità del dataset cinematico rispetto alle soglie del filtro	54
3.5	Analisi preliminare	54
3.5.1	Statistiche di sintesi e distribuzione spaziale	54
3.5.2	Composizione del traffico per tipologia di unità	56
3.6	Limiti e trasparenza metodologica	59
3.6.1	Eterogeneità nella ricchezza temporale degli encounter	59
3.6.2	Anomalie nei campi cinematici	60
3.6.3	Sintesi dei sottoinsiemi	60
3.6.4	Enrichment dei dati statici e copertura del join	60
3.6.5	Risoluzione del dato e ricostruzione del singolo incontro	61
3.7	Estensione a un secondo caso di studio: lo Stretto di Messina	61
3.7.1	Fonte e pre-processing	62
3.7.2	Dimensione e raffinamento cinematico	64
3.7.3	Composizione del traffico e copertura statica	67
4	Metodologia	71
4.1	Struttura della pipeline	71
4.2	Indicatori elementari di prossimità	72
4.3	Indicatori cinematici	72
4.3.1	Velocità relativa e approach rate	72
4.3.2	Indicatori basati su DCPA e TCPA stimati	73
4.3.3	Indicatori velocity-aware	74
4.4	Procedura di confronto fra indicatori	76
4.4.1	Correlazione di rango globale (Spearman)	76
4.4.2	Concordanza top-N (Jaccard)	77
4.4.3	Mappe pesate globali	78
4.4.4	Mappe disaggregate per macro-classe	79

4.5	Classificazione degli encounter per tipo di incontro	79
4.6	Autocorrelazione spaziale della criticità	80
4.7	Profilo temporale della criticità	81
4.8	Zone di criticità per densità (DBSCAN)	81
4.9	Verifica di conformità COLREG	84
5	Risultati	86
5.1	Tigullio: mappe globali e rapporto criticità-densità	86
5.2	Tigullio: composizione per macro-classe	92
5.3	Tigullio: tipologia degli encounter	93
5.4	Tigullio: struttura spaziale e hotspot	94
5.5	Tigullio: zone di criticità DBSCAN	97
5.6	Tigullio: profilo temporale	100
5.7	Tigullio: verifica euristica di conformità COLREG	102
5.8	Stretto di Messina: risultati dei tre anni	103
5.8.1	Mappe globali e rapporto criticità-densità	103
5.8.2	Composizione per macro-classe	109
5.8.3	Tipologia degli encounter	110
5.8.4	Struttura spaziale e hotspot	112
5.8.5	Zone di criticità DBSCAN	116
5.8.6	Profilo temporale	120
5.8.7	Verifica euristica di conformità COLREG	124
5.8.8	Casi illustrativi per tipo di incontro	125
5.9	Sintesi dei risultati	129
6	Conclusioni	131
6.1	Scelta dell'indicatore	131
6.2	Criticità e densità di traffico	132
6.3	Differenze fra le due aree	132
6.4	Stabilità sui tre anni dello Stretto	132
6.5	Limiti	133
6.6	Sviluppi	133
A	Stretto di Messina: figure del 2022 e del 2024	134
A.1	Dati	135
A.2	Risultati	138
B	Casi illustrativi della verifica COLREG	143
B.1	Caso 1: sorpasso, 2023, esito non conforme	143
B.2	Caso 3: sorpasso, 2023, esito non conforme	144
B.3	Caso 4: incrocio, 2022, esito non conforme	145
B.4	Caso 5: incrocio, 2024, esito conforme	146

Bibliografia	148
Ringraziamenti	151

Elenco delle figure

2.1	Schema di funzionamento del sistema AIS: canali VHF dedicati, accesso al mezzo SOTDMA, flussi di trasmissione tra navi, stazioni costiere e ricevitori satellitari, integrazione dei dati nei centri VTS.	19
2.2	Situazioni di incontro definite dal COLREG (Parte B): head-on (Regola 14), crossing (Regola 15) e overtaking (Regola 13).	23
2.3	Zone operative concentriche attorno alla nave: zona di collisione (raggio r_3), zona di conformità COLREG (r_2) e zona di navigazione libera (r_1). I raggi sono parametrizzati in funzione di dimensioni, velocità e manovrabilità.	24
2.4	Geometria di CPA, DCPA e TCPA nel sistema di riferimento relativo. OWN nell'origine, TARGET in posizione p_{ij} con velocità relativa v_R ; il punto CPA è la proiezione ortogonale dell'origine sulla traiettoria relativa, la DCPA è la distanza minima prevista e il TCPA il tempo di percorrenza fino al CPA.	26
2.5	Le tre formulazioni principali di ship domain a confronto. Fujii (ellisse fissa $4L \times 1,6L$), Quaternion Ship Domain (dominio asimmetrico con quattro raggi distinti) e safety domain adattivo (dimensioni variabili con velocità e manovrabilità). . . .	28
2.6	Schema esplicativo del funzionamento di DBSCAN. I core point (verdi) hanno almeno minPts vicini entro il raggio ϵ ; i border point (arancioni) ricadono nel vicinato ϵ di un core point ma non soddisfano la soglia minPts; i noise point (X rosse) restano isolati e non sono assegnati ad alcun cluster.	33
2.7	Esempio di Bayesian Network per il rischio di collisione marittima. Tre livelli di variabili: input (caratteristiche della nave, condizioni meteo-marine, densità del traffico, fattore umano), intermedie (tipo di incontro, geometria critica, tempo di reazione) e target (probabilità di collisione). Gli archi rappresentano le dipendenze probabilistiche tra le variabili.	41
3.1	Area di studio del Tigullio, coperta dal dataset AIS 2023, con il riquadro della vista di dettaglio del promontorio di Portofino ($44,29-44,37^\circ$ N; $9,13-9,26^\circ$ E). .	48
3.2	Distribuzione delle 3 006 unità navali tracciate nel 2023 per macro-tipologia. Sailing e pleasure craft dominano la popolazione; cargo e passenger sono minoritari.	50

3.3	Distribuzione mensile degli encounter nel 2023. Confronto tra l'intero dataset e quello clean. Si osserva una concentrazione del 98 % degli eventi nei mesi tra aprile e ottobre.	52
3.4	Analisi di sensitività del dataset cinematico alle soglie del filtro. La variazione percentuale dei cinematici e la quota di encounter migrati restano sotto il 2 % per perturbazioni fino al $\pm 20\%$	54
3.5	Mappa di densità degli encounter cinematici ($N = 8\,573$) su base cartografica, in scala logaritmica. Rispetto alla densità del dataset complessivo, l'hotspot di Portofino risulta attenuato, mentre i corridoi di transito e gli incroci di rotta restano visibili. La densità è calcolata su una griglia di 160×120 celle (longitudine $\times$ latitudine), di lato pari a circa 200 m (202 m in latitudine, 195 m in longitudine). .	55
3.6	Densità degli encounter cinematici nella vista di dettaglio del promontorio di Portofino ($44,29\text{--}44,37^\circ$ N; $9,13\text{--}9,26^\circ$ E), su celle di circa 100 m, in scala logaritmica.	56
3.7	Composizione del dataset cinematico per macro-classe delle unità coinvolte (Tigullio).	58
3.8	Distribuzione delle lunghezze delle unità coinvolte negli encounter cinematici: unità i , unità j e $L_{\max}$, la maggiore delle due lunghezze, calcolata sugli encounter con entrambe le lunghezze note. La mediana di $L_{\max}$ è pari a 23 m, coerente con la dominanza del naviglio da diporto.	59
3.9	Area di studio dello Stretto di Messina, comune ai tre anni analizzati, con il riquadro della vista di dettaglio dello Stretto ($38,08\text{--}38,30^\circ$ N; $15,50\text{--}15,72^\circ$ E). .	63
3.10	Schema di separazione del traffico dello Stretto di Messina (settori Nord, Centro e Sud), con zone di precauzione Zp1–Zp4 e zone di traffico costiero A–D. Tratta da Italian Coast Guard Headquarters (2015), pag. MES 04.	64
3.11	Mappa di densità degli encounter cinematici dello Stretto di Messina nel 2023 ($N = 83\,494$) su base cartografica, in scala logaritmica. La densità è calcolata su una griglia di 254×287 celle (longitudine $\times$ latitudine), di lato pari a circa 250 m (249 m in latitudine, 250 m in longitudine).	66
3.12	Densità degli encounter cinematici del 2023 nella vista di dettaglio dello Stretto ($38,08\text{--}38,30^\circ$ N; $15,50\text{--}15,72^\circ$ E), su celle di circa 100 m, in scala logaritmica. .	67
3.13	Distribuzione delle lunghezze delle unità coinvolte negli encounter cinematici dello Stretto di Messina nel 2023, calcolata per encounter sulle sole lunghezze dichiarate: unità i , unità j e $L_{\max}$, la maggiore delle due lunghezze. Asse limitato a 300 m.	69
3.14	Composizione del dataset cinematico per macro-classe delle unità coinvolte (Stretto di Messina, 2023).	70
4.1	Matrici di correlazione di Spearman fra gli otto indicatori sul dataset cinematico: (a) Tigullio, 2023; (b) Stretto di Messina, 2023.	77
4.2	Concordanza top-100 (Jaccard) fra gli otto indicatori sul dataset cinematico: (a) Tigullio; (b) Stretto di Messina.	78

4.3	Curva k-distance ($k = 8$) del 25 % degli encounter con l'indicatore principale più elevato. La linea continua indica il raggio adottato per DBSCAN (300 m), le linee punteggiate i valori di 200 e 400 m dell'analisi di sensitività. L'asse verticale è limitato a 600 m; i valori massimi sono 12 836 m nel Tigullio e 23 325 m nello Stretto di Messina.	83
5.1	Tigullio: mappe di criticità aggregata per gli indicatori R_1 – R_4 . Somma per cella dell'indicatore normalizzato, in scala logaritmica.	87
5.2	Tigullio: mappe di criticità aggregata per gli indicatori velocity-aware R_5 – R_8 . Somma per cella dell'indicatore normalizzato, in scala logaritmica.	88
5.3	Tigullio: criticità aggregata di R_7 nella vista di dettaglio del promontorio di Portofino (44,29–44,37° N; 9,13–9,26° E), su celle di circa 100 m, in scala logaritmica.	89
5.4	Tigullio: densità degli encounter cinematici per cella.	90
5.5	Tigullio: criticità aggregata (somma di R_7 per cella), con la correlazione di Spearman con la densità calcolata sulle celle con almeno 3 encounter.	91
5.6	Tigullio: criticità per incontro (media di R_7 per cella), con la correlazione di Spearman con la densità calcolata sulle celle con almeno 3 encounter.	92
5.7	Tigullio: criticità aggregata di R_7 per le cinque coppie di macro-classi più numerose, sugli encounter con entrambe le unità identificate, in scala logaritmica.	93
5.8	Tigullio: mappe di criticità aggregata di R_7 per tipo di incontro (incrocio, frontale, sorpasso), con classificazione basata su $ \Delta COG $, in scala logaritmica.	94
5.9	Tigullio: celle con concentrazione locale significativa della criticità aggregata di R_7 secondo la statistica di Getis-Ord G_i^* , al 95 % e al 99 %.	95
5.10	Tigullio: criticità per incontro (media di R_7 per cella) sulle celle con almeno 3 encounter.	96
5.11	Tigullio: frazione di encounter di ciascuna cella con R_7 superiore alla soglia del 95° percentile (0,219).	97
5.12	Tigullio: indice di Moran della criticità per incontro nelle 199 riallocazioni casuali di R_7 fra gli encounter a densità invariata, e valore osservato (0,487).	97
5.13	Tigullio: zone DBSCAN del 25 % degli encounter con R_7 più elevato ($\epsilon = 300$ m, $minPts = 8$).	99
5.14	Tigullio: zone DBSCAN nella vista di dettaglio del promontorio di Portofino (44,29–44,37° N; 9,13–9,26° E).	99
5.15	Tigullio: profilo temporale della criticità aggregata, somma di R_7	101
5.16	Tigullio: somma di R_7 per mese e ora del giorno (UTC).	102
5.17	Tigullio: tassi di conformità stimati con la verifica euristica COLREG, calcolati sui soli casi con esito determinato.	103
5.18	Stretto di Messina, 2023: mappe di criticità aggregata per gli indicatori R_1 – R_4 . Somma per cella dell'indicatore normalizzato, in scala logaritmica.	104
5.19	Stretto di Messina, 2023: mappe di criticità aggregata per gli indicatori velocity-aware R_5 – R_8 . Somma per cella dell'indicatore normalizzato, in scala logaritmica.	105

5.20	Stretto di Messina, 2023: criticità aggregata di R_8 nella vista di dettaglio dello Stretto (38,08–38,30° N; 15,50–15,72° E), su celle di circa 100 m, in scala logaritmica.	106
5.21	Stretto di Messina, 2023: densità degli encounter cinematici per cella.	107
5.22	Stretto di Messina, 2023: criticità aggregata (somma di R_8 per cella), con la correlazione di Spearman con la densità calcolata sulle celle con almeno 3 encounter. 108	
5.23	Stretto di Messina, 2023: criticità per incontro (media di R_8 per cella), con la correlazione di Spearman con la densità calcolata sulle celle con almeno 3 encounter. 109	
5.24	Stretto di Messina, 2023: criticità aggregata di R_8 per le cinque coppie di macroclassi più numerose, in scala logaritmica.	110
5.25	Stretto di Messina, 2023: mappe di criticità aggregata di R_8 per tipo di incontro (incrocio, frontale, sorpasso), con classificazione basata su $ \Delta COG $, in scala logaritmica.	111
5.26	Stretto di Messina, 2023: celle con concentrazione locale significativa della criticità aggregata di R_8 secondo la statistica di Getis-Ord G_i^* , al 95 % e al 99 %. . . .	113
5.27	Stretto di Messina, 2023: criticità per incontro (media di R_8 per cella) sulle celle con almeno 3 encounter.	114
5.28	Stretto di Messina, 2023: frazione di encounter di ciascuna cella con R_8 superiore alla soglia del 95° percentile (0,165).	115
5.29	Stretto di Messina, 2023: indice di Moran della criticità per incontro nelle 199 riallocazioni casuali di R_8 fra gli encounter a densità invariata, e valore osservato (0,565).	116
5.30	Stretto di Messina, 2023: zone DBSCAN del 25 % degli encounter con R_8 più elevato ($\epsilon = 300$ m, $minPts = 8$).	118
5.31	Stretto di Messina, 2023: zone DBSCAN nella vista di dettaglio dello Stretto (38,08–38,30° N; 15,50–15,72° E).	119
5.32	Stretto di Messina: zone DBSCAN del 2022, del 2023 e del 2024 sulla stessa mappa. 120	
5.33	Stretto di Messina, 2023: profilo temporale della criticità aggregata, somma di R_8 . 122	
5.34	Stretto di Messina, 2023: somma di R_8 per mese e ora del giorno (UTC).	123
5.35	Stretto di Messina: profilo temporale della criticità aggregata (R_8) nel 2022, nel 2023 e nel 2024.	123
5.36	Stretto di Messina, 2023: tassi di conformità stimati con la verifica euristica COLREG, calcolati sui soli casi con esito determinato.	124
5.37	Stretto di Messina: posizione dei casi illustrativi all'istante di minima distanza osservata.	126
5.38	Caso 2: sorpasso (Regola 13), Stretto di Messina, 2024, esito conforme. (a) Tracce ricostruite dalle posizioni trasmesse, con i minuti rispetto all'istante di minima distanza osservata; (b) distanza osservata e DCPA stimata; (c) variazione di rotta rispetto alla fase di approccio; (d) SOG. Bande grigie: fase di approccio e passaggio. 127	

5.39	Caso 6: incrocio (Regola 15), Stretto di Messina, 2023, esito non conforme. (a) Tracce ricostruite dalle posizioni trasmesse, con i minuti rispetto all'istante di minima distanza osservata; (b) distanza osservata e DCPA stimata; (c) variazione di rotta rispetto alla fase di approccio; (d) SOG. Bande grigie: fase di approccio e passaggio.	128
A.1	Stretto di Messina: densità degli encounter cinematici nel 2022 e nel 2024, in scala logaritmica, sulla stessa griglia del 2023.	135
A.2	Stretto di Messina: distribuzione delle lunghezze delle unità coinvolte negli encounter cinematici nel 2022 e nel 2024, calcolata per encounter sulle sole lunghezze dichiarate: unità i , unità j e $L_{\max}$. Asse limitato a 300 m.	136
A.3	Composizione del dataset cinematico per macro-classe delle unità coinvolte (Stretto di Messina, 2022 e 2024).	137
A.4	Stretto di Messina: criticità aggregata (somma di R_8 per cella) nel 2022 e nel 2024, con la correlazione di Spearman con la densità calcolata sulle celle con almeno 3 encounter.	138
A.5	Stretto di Messina: criticità per incontro (media di R_8 per cella) nel 2022 e nel 2024, con la correlazione di Spearman con la densità calcolata sulle celle con almeno 3 encounter.	139
A.6	Stretto di Messina: celle con concentrazione locale significativa della criticità aggregata di R_8 secondo la statistica di Getis-Ord G_i^* , al 95 % e al 99 %, nel 2022 e nel 2024.	140
A.7	Stretto di Messina: zone DBSCAN del 25 % degli encounter con R_8 più elevato ($\epsilon = 300$ m, $minPts = 8$) nel 2022 e nel 2024.	141
A.8	Stretto di Messina: somma di R_8 per mese e ora del giorno (UTC) nel 2022 e nel 2024.	142
B.1	Caso 1: sorpasso (Regola 13), Stretto di Messina, 2023, esito non conforme. (a) Tracce ricostruite dalle posizioni trasmesse, con i minuti rispetto all'istante di minima distanza osservata; (b) distanza osservata e DCPA stimata; (c) variazione di rotta rispetto alla fase di approccio; (d) SOG. Bande grigie: fase di approccio e passaggio.	144
B.2	Caso 3: sorpasso (Regola 13), Stretto di Messina, 2023, esito non conforme. (a) Tracce ricostruite dalle posizioni trasmesse, con i minuti rispetto all'istante di minima distanza osservata; (b) distanza osservata e DCPA stimata; (c) variazione di rotta rispetto alla fase di approccio; (d) SOG. Bande grigie: fase di approccio e passaggio.	145

B.3	Caso 4: incrocio (Regola 15), Stretto di Messina, 2022, esito non conforme. (a) Tracce ricostruite dalle posizioni trasmesse, con i minuti rispetto all'istante di minima distanza osservata; (b) distanza osservata e DCPA stimata; (c) variazione di rotta rispetto alla fase di approccio; (d) SOG. Bande grigie: fase di approccio e passaggio.	146
B.4	Caso 5: incrocio (Regola 15), Stretto di Messina, 2024, esito conforme. (a) Tracce ricostruite dalle posizioni trasmesse, con i minuti rispetto all'istante di minima distanza osservata; (b) distanza osservata e DCPA stimata; (c) variazione di rotta rispetto alla fase di approccio; (d) SOG. Bande grigie: fase di approccio e passaggio.	147

Elenco delle tabelle

2.1	Messaggi AIS ripartiti per categoria.	20
2.2	Confronto tra le formulazioni principali di ship domain.	28
2.3	Sintesi dei principali lavori citati nello stato dell'arte.	45
3.1	Schema dei campi del dataset <i>encounter-level</i> (formato JSON).	51
3.2	Distribuzione degli encounter del dataset clean nelle quattro categorie di scarto cinematico e nella selezione finale del dataset cinematografico (Tigullio).	53
3.3	Statistiche di sintesi del dataset di encounter cinematografici (Tigullio).	56
3.4	Distribuzione degli encounter cinematografici per coppia di macro-classi (soli encounter con identificazione completa di entrambe le navi, "full match", cfr. Sezione 3.6.4).	57
3.5	Composizione dei sottoinsiemi del dataset definiti dai successivi passaggi di filtraggio (Tigullio).	60
3.6	Dai messaggi AIS grezzi agli encounter clean (Stretto di Messina).	62
3.7	Distribuzione degli encounter del dataset clean nelle categorie di scarto cinematico e nella selezione finale (Stretto di Messina). Fra parentesi la quota sul clean.	65
3.8	Archivio statico e disponibilità delle lunghezze (Stretto di Messina).	68
3.9	Distribuzione degli encounter cinematografici per coppia di macro-classi (Stretto di Messina). Fra parentesi la quota sul totale. A differenza del Tigullio, nello Stretto tutte le unità coinvolte sono presenti nell'archivio statico in tutti e tre gli anni: la categoria <i>unknown-involved</i> è vuota.	68
5.1	Tigullio: zone DBSCAN del 25 % degli encounter con R_7 più elevato ($\epsilon = 300$ m, $minPts = 8$).	98
5.2	Stretto di Messina: rapporto con la densità e struttura spaziale della criticità (R_8) nei tre anni.	112
5.3	Stretto di Messina, 2023: zone DBSCAN del 25 % degli encounter con R_8 più elevato ($\epsilon = 300$ m, $minPts = 8$).	116
5.4	Stretto di Messina: zone DBSCAN del 25 % degli encounter con R_8 più elevato nei tre anni.	117

5.5	Stretto di Messina: tipologia degli encounter, profilo temporale della criticità (R_8) e conformità COLREG nei tre anni. Estate: giugno–settembre; notte: 21:00–06:00 UTC.	125
5.6	Casi illustrativi nello Stretto di Messina: unità coinvolte ed esito. L'unità A è quella tenuta a manovrare.	125
5.7	Casi illustrativi: grandezze misurate sulle posizioni trasmesse. Le variazioni di rotta (positive a dritta) e di velocità sono calcolate fra la fase di approccio (da -12 a -10 min) e il passaggio (da -1 a $+1$ min); la DCPA stimata è la minima prima del passaggio.	125
5.8	Confronto fra i quattro dataset con l'indicatore R_7 . Estate: giugno–settembre; notte: 21:00–06:00 UTC. L'ultima riga si riferisce all'indicatore principale (R_7 nel Tigullio, R_8 nello Stretto).	130

Capitolo 1

Introduzione

Il trasporto marittimo è parte essenziale del sistema dei trasporti e degli scambi internazionali. La sicurezza della navigazione riguarda la prevenzione di eventi con conseguenze sulle persone, sulle unità coinvolte e sull'ambiente marino.

Nelle aree costiere la presenza di rotte convergenti, vincoli geografici, collegamenti portuali e traffico eterogeneo può determinare una forte sovrapposizione di traiettorie in condizioni cinematiche diverse. Navi mercantili, traghetti, unità da pesca e imbarcazioni da diporto operano nello stesso spazio con dimensioni, velocità e modalità di navigazione differenti. L'analisi sistematica delle traiettorie osservate permette di caratterizzare queste interazioni e la loro distribuzione nello spazio.

1.1 Obiettivi della tesi

Il presente lavoro sviluppa un approccio, basato su dati AIS (Automatic Identification System) reali, alla valutazione della criticità degli incontri fra unità navali in aree costiere. Le mappe di criticità sono costruite aggregando spazialmente gli indicatori cinematici calcolati sugli incontri osservati. Il valore associato a ciascuna cella rappresenta la criticità degli incontri presenti nella cella e non una probabilità di collisione: il rischio, nel senso ingegneristico di combinazione fra frequenza e conseguenze, non è stimato.

La tesi si inserisce nel progetto Interreg Marittimo JASON, dedicato alla sicurezza della navigazione. La metodologia è applicata a due casi di studio con configurazioni di traffico differenti: il Golfo del Tigullio, con traffico costiero misto, prevalentemente diportistico, e lo Stretto di Messina, interessato da uno schema di separazione del traffico e da un servizio intenso di traghetti. Per il Tigullio sono analizzati i dati del 2023; per lo Stretto di Messina tre anni (2022, 2023 e 2024), elaborati con la stessa procedura per verificare che dia risultati coerenti su dataset diversi.

Il lavoro affronta due questioni metodologiche. La prima riguarda la scelta dell'indicatore di criticità: indicatori costruiti secondo formulazioni differenti, applicati agli stessi *encounter*, possono

attribuire livelli di criticità diversi ai medesimi eventi e, di conseguenza, individuare insiemi differenti di eventi critici. Nel seguito, per *encounter* si intende un evento di prossimità tra due unità navali, ricostruito a partire dalle rispettive tracce AIS e caratterizzato dalla presenza sincrona delle due unità entro una determinata soglia di prossimità. La seconda riguarda il rapporto tra criticità e traffico: una mappa ottenuta sommando la criticità dei singoli *encounter* all'interno di ciascuna cella cresce anche in funzione del numero di *encounter* presenti e deve pertanto essere distinta dalla criticità media del singolo *encounter*.

Il lavoro si articola nei seguenti passaggi:

- acquisizione e pre-elaborazione dei dati AIS disponibili per le aree di studio;
- ricostruzione delle traiettorie e identificazione degli encounter tra coppie di unità;
- selezione del sottoinsieme cinematico mediante criteri di qualità e coerenza dei dati;
- calcolo e confronto di indicatori geometrici e cinematici di prossimità e di criticità;
- costruzione di mappe di criticità, aggregando sulla griglia di analisi gli encounter pesati con l'indicatore;
- analisi della struttura spaziale della criticità mediante statistiche di autocorrelazione e identificazione degli hotspot;
- individuazione delle concentrazioni spaziali degli encounter con criticità più elevata mediante clustering density-based;
- analisi della distribuzione temporale e della tipologia degli encounter;
- verifica euristica della conformità alle regole di manovra del COLREG (International Regulations for Preventing Collisions at Sea), noto anche in italiano come RIPA o RIPAM (Regolamento internazionale per prevenire gli abbordi in mare).

1.2 Motivazioni

L'interesse del lavoro nasce dall'esigenza di descrivere quantitativamente le interazioni tra unità navali dove il traffico è eterogeneo e le traiettorie sono concentrate nello spazio. Le collisioni hanno un peso rilevante fra gli incidenti marittimi: nella rassegna di Marino et al. (2023) la collisione fra due unità è il tipo di incidente più frequente nelle aree portuali, con il 31 % dei casi a livello mondiale e il 37 % nel Mediterraneo.

Il COLREG stabilisce i comportamenti da adottare nelle principali situazioni di incontro, mentre i sistemi di bordo e i servizi di monitoraggio del traffico forniscono informazioni per valutare la situazione in tempo reale. L'analisi storica dei dati AIS offre una prospettiva complementare: permette di studiare le interazioni effettivamente avvenute su un intervallo temporale esteso e di descriverne la distribuzione spaziale, temporale e cinematica.

L'analisi caratterizza la criticità cinematica degli incontri e il suo rapporto con la distribuzione del traffico; non fornisce una stima della probabilità di collisione. Il confronto fra due aree con caratteristiche differenti permette di verificare il comportamento della procedura al variare del contesto operativo.

1.3 Struttura del documento

Il Capitolo 2 definisce il quadro tecnico e normativo di riferimento: l'AIS come sorgente dati, il COLREG come riferimento per le situazioni di incontro e gli strumenti impiegati in letteratura per valutare il rischio di collisione. Il capitolo comprende una rassegna critica degli approcci esistenti e l'individuazione del research gap.

Il Capitolo 3 descrive i dataset AIS dei due casi di studio e le procedure di consolidamento, estrazione degli encounter, filtraggio cinematico ed enrichment dei dati statici, insieme ai limiti del dato e alla struttura dei sottoinsiemi usati nelle analisi.

Il Capitolo 4 presenta la metodologia: indicatori di prossimità e di criticità, confronto fra indicatori, costruzione delle mappe, analisi della struttura spaziale e temporale, clustering density-based, e verifica euristica di conformità al COLREG.

Il Capitolo 5 riporta i risultati per il Tigullio e per i tre anni dello Stretto di Messina: distribuzione spaziale della criticità e relazione con la densità di traffico, composizione per tipologia di unità e di incontro, hotspot, zone DBSCAN, profilo temporale, tassi di conformità e casi illustrativi della verifica COLREG nello Stretto.

Il Capitolo 6 raccoglie le conclusioni, i limiti dell'approccio e le possibili estensioni della metodologia. L'Appendice A riporta le figure del 2022 e del 2024 dello Stretto di Messina, l'Appendice B i casi illustrativi della verifica COLREG non riportati nel testo.

Capitolo 2

Background e stato dell'arte

Questo capitolo pone le basi concettuali, tecniche e bibliografiche su cui si sviluppa la tesi. Il capitolo parte dal quadro tecnico e normativo di riferimento e sviluppa successivamente la rassegna critica della letteratura, fino all'identificazione dei principali limiti degli approcci esaminati.

Il capitolo si articola in cinque sezioni. La Sezione 2.1 introduce l'*Automatic Identification System* (AIS) come sorgente dati, descrivendone principi di funzionamento, contenuto informativo, punti di forza e limiti per l'analisi del traffico marittimo. Segue, nella Sezione 2.2, il quadro normativo per la prevenzione delle collisioni, con particolare riferimento al *Regolamento Internazionale per Prevenire gli Abbordi in Mare* (COLREG), di cui viene evidenziata la natura duale del rischio, geometrica e comportamentale. Nella Sezione 2.3 sono richiamati gli strumenti concettuali e metodologici di cui la letteratura AIS-based si serve per valutare il rischio di collisione (indicatori geometrici, ship domain, indici compositi, KDE, clustering), presentati come riferimento metodologico per la successiva analisi della letteratura. Su questa base, la Sezione 2.4 sviluppa l'analisi critica dello stato dell'arte, organizzata per famiglie metodologiche. Il capitolo si chiude con la Sezione 2.5, che raccoglie i limiti emersi dalla letteratura e definisce il research gap che la tesi si propone di colmare.

2.1 L'AIS come base dati

2.1.1 Definizione e quadro normativo

L'*Automatic Identification System* (AIS) è un sistema di identificazione automatica basato su transponder *Very High Frequency* (VHF), concepito per incrementare la sicurezza della navigazione marittima e per supportare la prevenzione delle collisioni, le operazioni di ricerca e soccorso e la gestione del traffico da parte dei *Vessel Traffic Services* (VTS). L'AIS è stato reso obbligatorio dall'International Maritime Organization (IMO) attraverso la Regola 19 del Capitolo V della Convenzione internazionale per la salvaguardia della vita umana in mare (SOLAS (Int, 1974)), che prescrive un transponder AIS per tutte le navi di stazza lorda pari o superiore a 300 GT impegnate in viaggi internazionali, per le navi da carico di 500 GT o superiori non impegnate in viaggi in-

ternazionali e per tutte le navi passeggeri, indipendentemente dalle dimensioni. L'obbligo è stato introdotto gradualmente fra il 1° luglio 2002 e il 1° luglio 2008. Restano in linea di principio escluse alcune categorie di unità, quali navi militari, imbarcazioni da diporto e pescherecci di piccola taglia, che possono tuttavia montare dispositivi di Classe B su base volontaria.

2.1.2 Principi di funzionamento

Dal punto di vista tecnico, l'AIS opera su due canali dedicati della banda VHF marittima (canali AIS 1 e AIS 2, centrati rispettivamente a 161,975 MHz e 162,025 MHz) e si basa su un meccanismo di accesso al mezzo di tipo Self-Organised Time Division Multiple Access (SOTDMA) per i transponder di Classe A: ciascuna stazione suddivide autonomamente il tempo in slot e seleziona quelli liberi per trasmettere, consentendo la gestione simultanea di un elevato numero di trasmissioni sullo stesso canale, limitando le interferenze reciproche. La portata tipica è dell'ordine di 20-30 miglia nautiche in mare aperto, funzione soprattutto dell'altezza delle antenne e della conformazione costiera. I messaggi vengono captati sia da altre navi nelle vicinanze, sia da stazioni costiere, sia, attraverso il servizio noto come Satellite-AIS (S-AIS), da costellazioni satellitari dedicate, indispensabili per coprire le aree non raggiunte dalla rete a terra.

I dispositivi si distinguono in due classi principali:

- Transponder di Classe A: obbligatori per le navi SOLAS, trasmettono con cadenza variabile da pochi secondi ad alcuni minuti in funzione dello stato di moto della nave (in navigazione, in manovra, all'ancora).
- Transponder di Classe B: destinati a unità non SOLAS, operano con priorità inferiore e intervalli di aggiornamento più lunghi.

La Fig. 2.1 sintetizza l'architettura complessiva del sistema, mostrando i flussi di trasmissione nave-nave, nave-costa e nave-satellite, con l'integrazione finale verso il centro VTS.

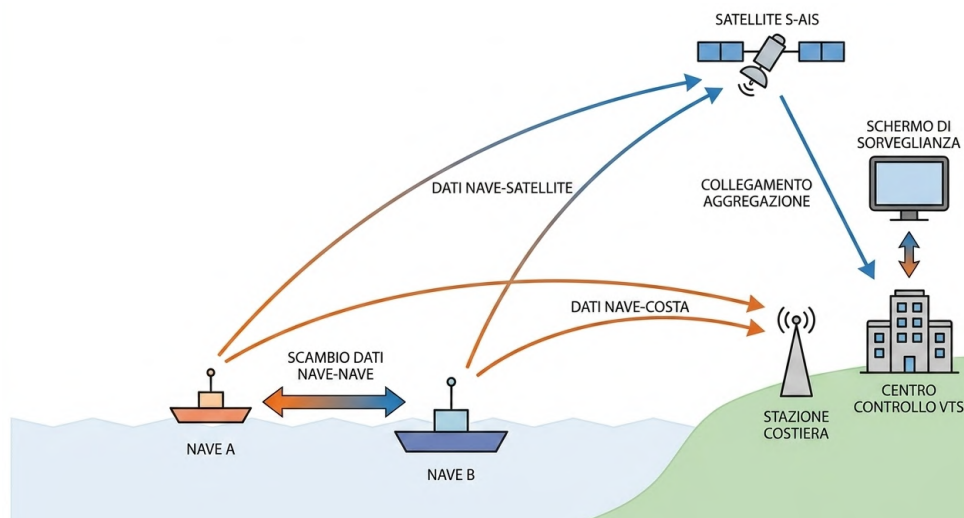


Figura 2.1: Schema di funzionamento del sistema AIS: canali VHF dedicati, accesso al mezzo SOTDMA, flussi di trasmissione tra navi, stazioni costiere e ricevitori satellitari, integrazione dei dati nei centri VTS.

2.1.3 Contenuto informativo

Le informazioni trasmesse sono convenzionalmente ripartite in tre categorie. I dati statici (Maritime Mobile Service Identity (MMSI), numero IMO, nominativo radio o call sign, nome e tipo di nave, dimensioni principali, posizione dell'antenna del sistema di posizionamento) sono inseriti al momento dell'installazione e aggiornati solo raramente. I dati dinamici (posizione in latitudine e longitudine, Speed Over Ground (SOG), Course Over Ground (COG), heading, rate of turn, stato di navigazione) sono acquisiti automaticamente dai sensori di bordo (ricevitore Global Navigation Satellite System (GNSS), girobussola) e trasmessi con frequenza elevata. I dati di viaggio (destinazione, Estimated Time of Arrival (ETA), pescaggio effettivo, eventuale presenza di merci pericolose) sono invece inseriti manualmente dall'equipaggio e aggiornati a discrezione del comandante. A queste tre categorie si affiancano brevi messaggi di sicurezza (safety-related messages) a testo libero, complementari alla fonia VHF. La Tab. 2.1 riassume in forma schematica questa ripartizione.

Tabella 2.1: Messaggi AIS ripartiti per categoria.

Categoria	Campi principali	Modalità di aggiornamento
Dati statici	MMSI; numero IMO; nominativo radio (call sign); nome della nave; tipo di nave; dimensioni principali; posizione dell'antenna del sistema di posizionamento	Inseriti al momento dell'installazione e aggiornati solo raramente (campi manuali di configurazione)
Dati dinamici	Posizione in latitudine e longitudine; SOG; COG; heading; rate of turn; stato di navigazione	Acquisiti automaticamente dai sensori di bordo (GNSS, girobussola) e trasmessi con frequenza elevata
Dati di viaggio	Destinazione; ETA; pescaggio effettivo; eventuale presenza di merci pericolose	Inseriti manualmente dall'equipaggio e aggiornati a discrezione del comandante
Messaggi di sicurezza	Safety-related messages (testo libero)	A discrezione dell'operatore, complementari alla fonia VHF

Sul piano della codifica, i messaggi AIS sono trasmessi in formato binario e veicolati a bordo dei sistemi di navigazione attraverso lo standard NMEA 0183, secondo cui ciascuna trama (con header !AIVDM per messaggi ricevuti da altre navi o !AIVDO per messaggi propri della stazione) contiene il payload codificato in 6-bit ASCII. La decodifica di un archivio AIS grezzo richiede pertanto una fase preliminare di parsing che dal flusso testuale NMEA ricostruisca i campi binari dei singoli messaggi.

2.1.4 L'AIS come sorgente dati per lo studio del traffico

La disponibilità continua e strutturata di queste informazioni ha progressivamente trasformato l'AIS, nel corso dell'ultimo decennio, da puro strumento operativo di supporto alla navigazione a

una delle principali sorgenti di dati per lo studio quantitativo del traffico marittimo. L'aggregazione di archivi storici AIS, provenienti da reti costiere e da ricevitori satellitari, consente di ricostruire a posteriori il movimento delle navi in aree di interesse, di individuare rotte ricorrenti, fasce orarie di maggiore intensità, corridoi di attraversamento e concentrazioni di traffico, nonché di isolare interazioni nave-nave potenzialmente critiche.

Nella prospettiva del presente lavoro, dedicato alla generazione di mappe di criticità degli incontri in area costiera a partire da traffico reale, l'AIS presenta alcuni punti di forza particolarmente rilevanti. Ogni messaggio è associato a un identificativo univoco (MMSI) e a un riferimento temporale esplicito, il che rende immediata la ricostruzione della traiettoria di ciascuna unità. I campi dinamici (SOG, COG, heading) permettono di descrivere non soltanto la posizione, ma anche l'assetto cinematico dell'imbarcazione in ogni istante, requisito indispensabile per qualsiasi successiva analisi di prossimità e di interazione fra navi. La combinazione di dati statici (tipo di nave, dimensioni) e dati dinamici consente inoltre di caratterizzare il traffico in modo disaggregato, distinguendo classi di unità con profili operativi e livelli di rischio differenti. Infine, la copertura ormai capillare delle aree costiere europee, sostenuta sia dalla rete di stazioni riceventi sia dai servizi satellitari, offre basi dati sufficientemente estese da consentire analisi statisticamente significative su periodi temporali ampi.

2.1.5 Limiti e criticità

L'uso dell'AIS per analisi data-driven richiede tuttavia una chiara consapevolezza dei suoi limiti. La copertura non è uniforme: in prossimità delle stazioni costiere la densità dei messaggi è elevata, mentre in zone più distanti o schermate da ostacoli orografici possono manifestarsi interruzioni di ricezione e finestre temporali prive di aggiornamenti. Non tutte le unità in mare sono rappresentate: imbarcazioni non SOLAS prive di transponder, pescherecci di piccola taglia e navi militari possono sfuggire al sistema, di conseguenza il quadro del traffico ricostruito resta per definizione parziale. La qualità delle informazioni dipende inoltre fortemente dalla corretta installazione e manutenzione del dispositivo di bordo e dall'accuratezza con cui l'equipaggio compila i campi manuali: destinazione, ETA, pescaggio e talvolta il tipo di nave possono risultare incompleti, non aggiornati o palesemente errati. Occorre inoltre considerare possibili interruzioni volontarie della trasmissione, previste dalla normativa in casi eccezionali, e, più raramente, fenomeni di spoofing o di dati dinamici inconsistenti (posizioni che "saltano", valori di SOG o COG incompatibili con la traiettoria).

Un fenomeno particolarmente attuale in materia di interruzioni intenzionali è quello della cosiddetta *shadow fleet* (o *dark fleet*), espressione con cui si designa l'insieme di navi commerciali, spesso impiegate in attività elusive di sanzioni internazionali, che disattivano deliberatamente il transponder AIS per periodi prolungati o ne falsificano i dati. Il fenomeno si è intensificato dal 2022 nel contesto delle sanzioni successive al conflitto in Ucraina e rappresenta una criticità riconosciuta a livello regolatorio (IMO, European Maritime Safety Agency (EMSA)), con implicazioni concrete per qualsiasi analisi data-driven del traffico marittimo basata sull'AIS.

Infine, gli archivi AIS grezzi presentano tipicamente messaggi duplicati, timestamp non monotoni e record parzialmente troncati, elementi che rendono necessaria una fase preliminare di pulizia e validazione dei dati prima di qualsiasi elaborazione.

2.2 Il quadro normativo per la prevenzione delle collisioni

2.2.1 Il COLREG nel sistema internazionale della sicurezza marittima

La normativa di riferimento per la prevenzione delle collisioni tra unità navali è costituita dalla *Convention on the International Regulations for Preventing Collisions at Sea* (COLREG), adottata nel 1972 sotto tutela dell'IMO. Il COLREG si affianca ad altri strumenti del quadro internazionale della sicurezza marittima, in particolare al Capitolo V della SOLAS sulla *Safety of Navigation*, che disciplina dotazioni e procedure di bordo (radar, Electronic Chart Display and Information System (ECDIS), AIS, Voyage Data Recorder (VDR)), e ai Vessel Traffic Services (VTS), che costituiscono la componente di monitoraggio da terra, definendo però un livello logico distinto. Laddove SOLAS prescrive gli strumenti e i VTS forniscono la cornice operativa di supervisione, il COLREG stabilisce il comportamento che le navi devono tenere le une rispetto alle altre. La sua funzione è quindi quella di fornire un linguaggio comune e un insieme di obblighi reciproci che rendano prevedibili le manovre in situazione di incontro: una prevedibilità che è, essa stessa, un fattore primario di sicurezza.

Nel testo consolidato (International Maritime Organization, 2018) le regole sono 41, suddivise in sei parti: Parte A, Generalità; Parte B, Regole di governo e di manovra; Parte C, Fanali e segnali visivi; Parte D, Segnali sonori e luminosi; Parte E, esenzioni; Parte F, verifica dell'applicazione della Convenzione, che comprende le Regole 39-41 aggiunte con emendamento. Il testo comprende inoltre quattro allegati tecnici. Ai fini del presente lavoro, l'area di maggiore interesse è la Parte B, che disciplina la condotta delle navi in navigazione e contiene gli articoli direttamente connessi alla valutazione e alla prevenzione delle collisioni.

2.2.2 Le regole rilevanti per l'interpretazione del rischio

Ai fini della tesi è utile richiamare in modo selettivo le regole che costituiscono il riferimento concettuale per l'analisi del rischio. La Regola 5 impone l'obbligo di un *proper look-out* (appropriato servizio di vedetta) con ogni mezzo disponibile, includendo quindi sia l'osservazione diretta sia il ricorso agli strumenti di bordo. La Regola 6 introduce il principio di *safe speed*, una velocità che deve essere tale da permettere un'azione efficace per evitare la collisione in relazione alle condizioni contingenti. La Regola 7, dedicata specificamente al *risk of collision*, stabilisce che il rischio debba essere valutato con tutti i mezzi a disposizione e, in particolare, individua nel mancato apprezzabile cambiamento del rilevamento (*compass bearing*) della nave in avvicinamento un indicatore fondamentale del rischio; precisa inoltre che il rischio possa sussistere anche in presenza di variazione del rilevamento, soprattutto a brevi distanze o con unità di grandi dimensioni. La Regola 8 disciplina l'azione di evitamento, che deve essere positiva, intrapresa con ampio margine

temporale e nel rispetto dei principi della *good seamanship* (buon senso nautico), con variazioni di rotta o di velocità sufficientemente ampie da risultare percepibili all'altra unità.

A queste si affiancano le regole che codificano le tre situazioni di incontro tipiche tra unità a propulsione meccanica in vista reciproca: la Regola 13 per l'*overtaking*, la Regola 14 per il *head-on* e la Regola 15 per il *crossing*. Tali regole individuano, per ciascuno scenario, la nave che deve mantenere la rotta (*stand-on vessel*) e quella che deve cedere il passo (*give-way vessel*), fornendo un'interpretazione deterministica di ciò che, altrimenti, sarebbe una situazione geometrica ambigua. La Regola 10, infine, disciplina la navigazione all'interno dei *Traffic Separation Schemes*, istituti particolarmente rilevanti in aree ad alta densità costiera. La Fig. 2.2 rappresenta sinteticamente le tre situazioni di incontro e le manovre di evitamento previste.

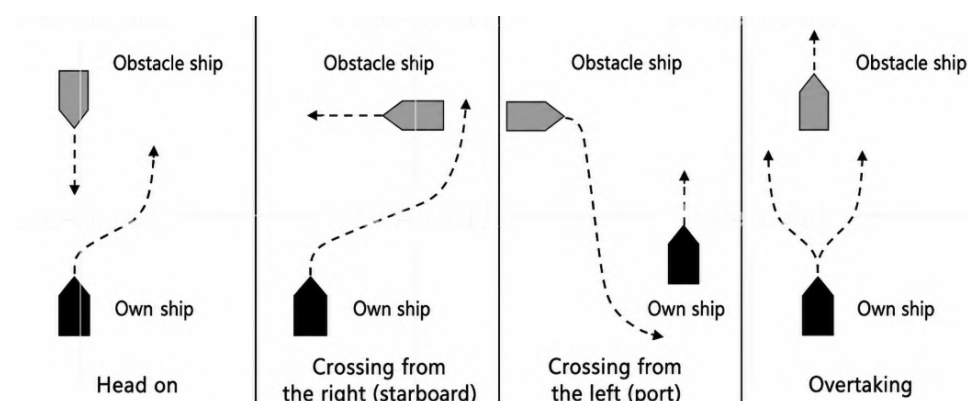


Figura 2.2: Situazioni di incontro definite dal COLREG (Parte B): head-on (Regola 14), crossing (Regola 15) e overtaking (Regola 13).

2.2.3 Le due dimensioni del rischio di collisione

Un aspetto che il COLREG rende particolarmente evidente, e che è centrale per la prospettiva metodologica della tesi, è la natura duale del rischio di collisione. Se letto solo in chiave geometrica, il rischio si riduce alla prossimità attuale o futura tra due navi, cioè a una questione di distanze e tempi di avvicinamento. La normativa, tuttavia, integra tale lettura con un livello comportamentale: il rischio non è solo funzione della traiettoria, ma anche della congruenza delle manovre rispetto a ciò che le regole prescrivono nella specifica situazione di incontro. In questa logica, due navi che si avvicinano con rilevamento costante sono in condizione di rischio anche se la distanza istantanea è ancora ampia; al contrario, un avvicinamento geometricamente serrato ma gestito con manovre conformi e tempestive può essere considerato sotto controllo. Il COLREG, di fatto, trasforma il rischio di collisione da mera variabile cinematica in un concetto che integra traiettoria, prevedibilità e conformità normativa. Quando si costruiscono mappe di rischio a partire da dati osservativi, ignorare la dimensione regolativa significherebbe trattare come equivalenti situazioni che, nella pratica marittima, hanno implicazioni molto diverse.

2.2.4 Dalle regole alle zone operative

L'applicazione operativa del COLREG al problema della valutazione del rischio ha condotto a una rappresentazione per zone concentriche attorno a ciascuna nave (Ballotta, 2025), funzionale a discriminare i differenti livelli di attenzione richiesti. Si individua in genere una *zona di navigazione libera*, in cui le distanze tra le unità sono tali da non richiedere verifiche specifiche; una *zona di conformità COLREG*, all'interno della quale le regole della Parte B entrano in vigore in modo sostanziale e la condotta delle navi deve essere valutata alla luce della situazione di incontro configurata; e una *zona di collisione*, di raggio molto più ridotto, in cui la prossimità è tale da rendere critica qualsiasi manovra e da richiedere eventualmente azioni di emergenza. L'ampiezza di queste zone dipende da parametri operativi (dimensioni della nave, velocità, manovrabilità) e può essere quindi parametrizzata in funzione dei dati disponibili. Questa struttura, pur essendo una semplificazione, costituisce uno schema concettuale utile per collegare la prescrizione normativa alle grandezze quantitative con cui il rischio viene poi misurato. La Fig. 2.3 rappresenta schematicamente le tre zone concentriche e i raggi associati.

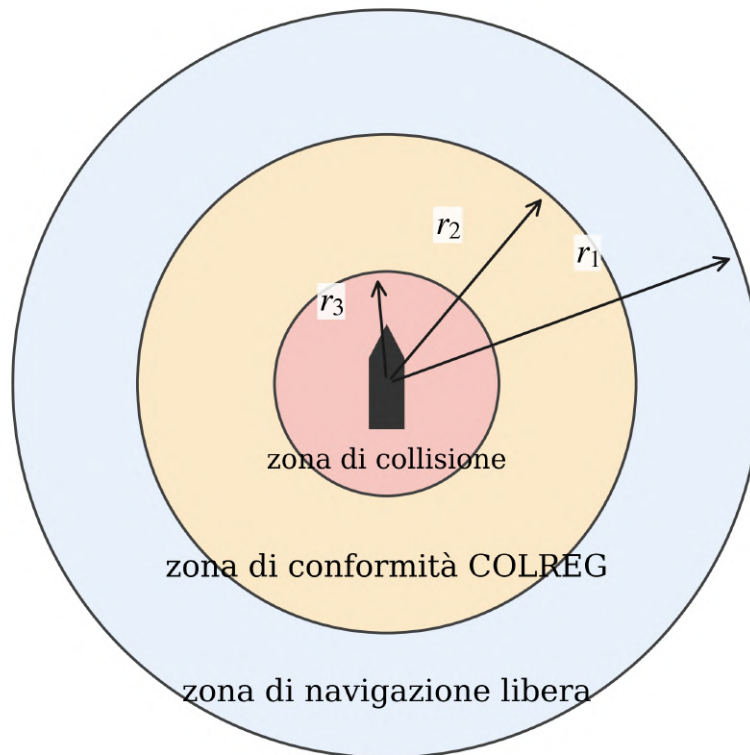


Figura 2.3: Zone operative concentriche attorno alla nave: zona di collisione (raggio r_3), zona di conformità COLREG (r_2) e zona di navigazione libera (r_1). I raggi sono parametrizzati in funzione di dimensioni, velocità e manovrabilità.

2.3 Strumenti concettuali e metodologici per la valutazione del rischio

Questa sezione presenta gli strumenti quantitativi di cui la letteratura AIS-based si serve per valutare il rischio di collisione. Tali strumenti costituiscono il vocabolario tecnico necessario per leggere la rassegna dei paper nella Sezione 2.4 e collocare la metodologia adottata rispetto agli approcci presenti in letteratura. Il percorso si articola in cinque passaggi: indicatori geometrici di prossimità (2.3.1), ship domain come criterio di rilevazione delle near collision (2.3.2), indici compositi per lo scoring del rischio (2.3.3), Kernel Density Estimation per la mappa continua (2.3.4), clustering density-based per gli hotspot discreti (2.3.5).

2.3.1 Indicatori geometrici: CPA, DCPA, TCPA

Quasi tutti i metodi di valutazione del rischio di collisione nave-nave si fondano, direttamente o indirettamente, sul Closest Point of Approach (CPA) e sulle due grandezze derivate DCPA e TCPA. Questi indicatori costituiscono il linguaggio comune con cui si descrive la prossimità tra due navi e sono il fondamento operativo dei sistemi ARPA (Automatic Radar Plotting Aid), ECDIS e dei sistemi di supporto alle decisioni a bordo. La formulazione classica risale a Lenart (1983) ed è ripresa, con notazioni equivalenti, in numerosi contributi successivi (Rong et al., 2019; Liu et al., 2022).

Geometria della velocità relativa e definizione di CPA. Si considerino due navi, denominate *own ship* (O) e *target ship* (T). Nell'istante corrente, la own si trova in (x_o, y_o) , si muove alla velocità v_o con rotta φ_o ; la target è in (x_t, y_t) , con velocità v_t e rotta φ_t . Il problema del CPA consiste nel determinare, assumendo che entrambe le navi mantengano velocità e rotta, il punto futuro di minima distanza tra le due unità e l'istante in cui tale minimo si verifica.

Il passo chiave è l'introduzione di un sistema di riferimento relativo, solidale alla own: in questo sistema la own è ferma nell'origine e la target si muove con vettore velocità relativa $\vec{v}_R = \vec{v}_t - \vec{v}_o$, il cui modulo è:

$$|v_R| = \sqrt{v_o^2 + v_t^2 - 2v_o v_t \cos(\varphi_o - \varphi_t)} \quad (2.1)$$

Se $\vec{p}_{ij} = (x_t - x_o, y_t - y_o)$ è il vettore posizione dalla own alla target, la distanza relativa al tempo τ è $|D(\tau)| = |\vec{p}_{ij} + \vec{v}_R \tau|$. Il CPA è il punto in cui $|D(\tau)|$ raggiunge il minimo: imponendo la condizione di stazionarietà si ottengono le due formule utilizzate da quasi tutta la letteratura AIS-based:

$$TCPA = -\frac{\vec{p}_{ij} \cdot \vec{v}_R}{|\vec{v}_R|^2} \quad (2.2)$$

$$DCPA = |\vec{p}_{ij}| \sqrt{1 - \left(\frac{\vec{p}_{ij} \cdot \vec{v}_R}{|\vec{p}_{ij}| |\vec{v}_R|} \right)^2} \quad (2.3)$$

I tre indicatori assumono significati semplici: il CPA è il punto previsto di minima distanza tra le due navi nella geometria relativa; la DCPA (*Distance to CPA*) è la distanza minima prevista; il TCPA (*Time to CPA*) è il tempo previsto per arrivare al CPA. Un TCPA positivo associato a una DCPA piccola identifica una condizione di potenziale rischio; un TCPA negativo indica che il CPA è già stato superato. Nei sistemi operativi si definiscono quindi soglie di allarme:

- una soglia di prossimità sulla distanza, del tipo $DCPA < DCPA_{soglia}$;
- una soglia temporale di anticipo, del tipo $0 < TCPA < TCPA_{soglia}$.

La Fig. 2.4 rappresenta schematicamente la geometria di CPA, DCPA e TCPA nel sistema di riferimento relativo, evidenziando la traiettoria della target rispetto alla own e la proiezione che identifica il punto di minimo avvicinamento.

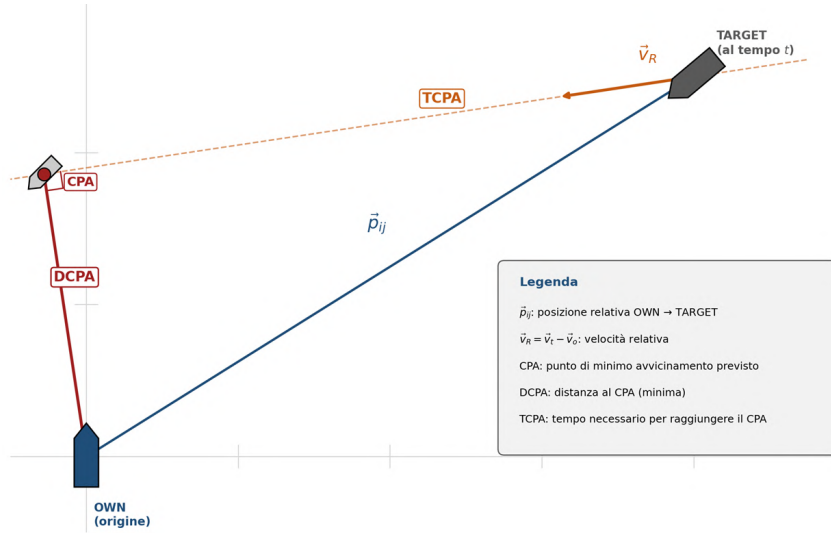


Figura 2.4: Geometria di CPA, DCPA e TCPA nel sistema di riferimento relativo. OWN nell'origine, TARGET in posizione p_{ij} con velocità relativa v_R ; il punto CPA è la proiezione ortogonale dell'origine sulla traiettoria relativa, la DCPA è la distanza minima prevista e il TCPA il tempo di percorrenza fino al CPA.

Ipotesi di moto rettilineo uniforme. La costruzione precedente poggia sulla seguente ipotesi: le due navi mantengono velocità e rotta costanti fino al CPA. Questa assunzione di moto rettilineo uniforme (MRU) è necessaria per ottenere una soluzione chiusa e per mantenere il calcolo computazionalmente trattabile su grandi volumi di dati AIS. È ragionevole su orizzonti brevi (pochi minuti) in acque aperte, ma perde validità in presenza di manovre, riduzioni di velocità in avvicinamento a porti, aree ristrette e condizioni meteo-marine avverse. Il CPA così calcolato deve quindi essere letto come previsione condizionata al mantenimento della cinematica corrente, non come distanza di minima assoluta.

Limiti strutturali. Al di là dell'MRU, l'uso diretto di DCPA e TCPA come stima del rischio di collisione presenta limiti strutturali ben noti in letteratura:

- **Trattamento puntuale della nave.** DCPA e TCPA ignorano dimensioni e geometria della nave, che è trattata come un punto materiale. Una DCPA di 100 m ha un significato diverso per una portacontainer di 300 m e per una piccola unità da diporto. Questo limite è affrontato, in parte, dall'introduzione dello ship domain (Sezione 2.3.2).
- **Indistinguibilità dei tipi di incontro.** Il solo CPA non distingue scenari di *head-on*, *crossing* e *overtaking*: scenari diversi possono produrre valori identici di DCPA e TCPA pur avendo implicazioni operative molto diverse in termini di priorità COLREG e manovre disponibili. Un indicatore puramente geometrico non è quindi sufficiente a caratterizzare il rischio sul piano comportamentale, come già evidenziato nella discussione del COLREG in 2.2.3.
- **Assenza di variabili contestuali.** Condizioni meteo-marine, densità di traffico, tipologia di nave, manovrabilità, esperienza dell'equipaggio restano tutte fuori dal perimetro geometrico del CPA.

Nonostante questi limiti, DCPA e TCPA rimangono il nucleo comune degli approcci AIS-based. Le estensioni trattate nelle sezioni successive (ship domain, indici compositi, aggregazione spaziale) sono costruite su questo nucleo.

2.3.2 Ship domain e rilevazione delle near collision

Il concetto di *ship domain* risale a Fujii and Tanaka (1971), che lo definiscono come l'area bidimensionale attorno a una nave che il pilota desidera mantenere libera da altre unità e da ostacoli. La violazione di tale dominio, ovvero l'ingresso di un'altra nave nella sua area, viene interpretata come un evento di prossimità critica, o *near collision*. Lo ship domain svolge dunque un duplice ruolo: fornisce un criterio fisico per la rilevazione dell'evento critico, più robusto di una soglia arbitraria su DCPA, e integra implicitamente le dimensioni della nave nell'analisi.

Formulazioni principali. La letteratura ha sviluppato tre formulazioni principali:

- **Dominio fisso ellittico (Fujii).** È la formulazione più diffusa negli studi AIS-based: un'ellisse orientata secondo la prua, con semiasse maggiore $4L$ e semiasse minore $1,6L$, dove L è la lunghezza della nave. La scelta è motivata da semplicità e replicabilità: richiede solo la lunghezza, direttamente estraibile dai messaggi AIS statici, e non necessita di calibrazione operativa. Il limite è la staticità: il dominio non varia con velocità, manovrabilità o densità di traffico.
- **Quaternion Ship Domain (QSD).** È un dominio asimmetrico che distingue quattro raggi: prua, poppa, dritta, sinistra. L'asimmetria riflette il fatto che il rischio percepito è più alto a prua che a poppa, e differisce tra dritta e sinistra per via delle regole COLREG. Il QSD è più

coerente con la pratica della navigazione ma introduce più parametri e maggiore sensibilità alla calibrazione.

- **Safety domain adattivo.** È un dominio le cui dimensioni variano con velocità e manovrabilità: una nave veloce richiede più spazio di sicurezza di una nave in manovra portuale. È l'approccio concettualmente più fisico, ma richiede informazioni (caratteristiche manovriere) non sempre disponibili nei dati AIS, tipicamente approssimate tramite tabelle IMO per classi di navi.

La ricognizione delle scelte specifiche effettuate in letteratura (quali autori abbiano adottato Fujii, quali QSD, quali formulazioni adattive, e con quali parametri) è sviluppata nella Sezione 2.4. In questa sede interessa il principio operativo comune ai tre approcci. La Fig. 2.5 confronta visivamente le tre formulazioni e la Tab. 2.2 ne riassume in forma comparativa proprietà, parametri richiesti e principali vantaggi/limiti.

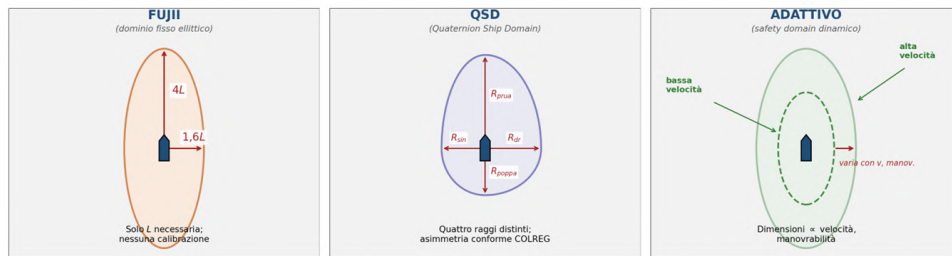


Figura 2.5: Le tre formulazioni principali di ship domain a confronto. Fujii (ellisse fissa $4L \times 1,6L$), Quaternion Ship Domain (dominio asimmetrico con quattro raggi distinti) e safety domain adattivo (dimensioni variabili con velocità e manovrabilità).

Tabella 2.2: Confronto tra le formulazioni principali di ship domain.

Formulazione	Forma	Parametri richiesti	Pro	Contro
Fujii (ellittico fisso)	Ellisse orientata secondo la prua; semiasse maggiore $4L$, minore $1,6L$	Solo la lunghezza della nave (da dati AIS statici)	Estrema semplicità; replicabilità; nessuna calibrazione operativa	Staticità; sovrastima delle near collision in aree portuali e a bassa velocità
Quaternion Ship Domain (QSD)	Dominio asimmetrico con quattro raggi distinti (prua, poppa, dritta, sinistra)	Quattro raggi calibrati; dati sulla percezione del rischio per classe di nave	Asimmetria coerente con la pratica operativa e con le regole COLREG	Calibrazione più onerosa; maggiore sensibilità ai parametri
Safety domain adattivo	Dimensioni variabili con velocità, manovrabilità e condizioni operative	Caratteristiche manovriere (spesso da tabelle IMO per classi)	Concettualmente più fisico; meno falsi positivi in area portuale	Richiede dati manovrieri non sempre disponibili; calibrazione complessa

Rilevazione dell'evento near collision. Dal punto di vista operativo, la violazione dello ship domain è il criterio di detection. Per ciascun istante temporale e per ciascuna coppia di navi simultaneamente presenti nell'area di studio, si verifica se la posizione della target cade all'interno del dominio della own. In caso affermativo, l'evento è registrato come near collision e i suoi

attributi (DCPA, TCPA, distanza relativa D_R , angolo di incontro θ_{ij} , tipo di incontro, posizione geografica, istante temporale, identificativi e caratteristiche delle due navi) vengono salvati. L'insieme degli eventi near collision così rilevati costituisce il dataset di eventi critici su cui operano i passaggi successivi: lo scoring del rischio e l'aggregazione spaziale.

Si anticipano qui due osservazioni. La prima riguarda il bilanciamento tra semplicità e accuratezza: un dominio fisso come quello di Fujii sovrastima leggermente la near collision rate in aree ad alta densità e bassa velocità (tipicamente acque portuali, ancoraggi e aree di manovra), in quanto il raggio di dominio non si riduce proporzionalmente alla bassa velocità operativa. Un dominio adattivo corregge in parte questo effetto, al prezzo di una maggiore richiesta di dati input e di un costo computazionale elevato, particolarmente rilevante in caso di processing in tempo reale. La seconda osservazione è che la near collision, così definita, è un evento non-incidentale: non si tratta di una collisione reale, ma di una prossimità che viola le distanze di sicurezza implicite nel dominio. La sua frequenza e distribuzione spaziale fungono da proxy del rischio, secondo l'impostazione della *safety pyramid* richiamata nella Sezione 2.4.1.

2.3.3 Indici di rischio compositi

Una volta rilevato l'evento critico, è necessario quantificarne la gravità. Gli indicatori geometrici presi singolarmente forniscono un'informazione binaria (violazione del dominio sì/no) oppure unidimensionale (un valore di DCPA o di TCPA), insufficiente per graduare il rischio e, soprattutto, per pesare gli eventi in una successiva mappa di densità. La letteratura ha sviluppato a questo scopo una famiglia di indici compositi, che combinano più variabili in un unico scalare di rischio.

Un indice composito di rischio di collisione è una funzione scalare

$$CR = f(d_{CPA}, t_{CPA}, D_R, L_o, L_t, \theta_{ij}, \dots) \quad (2.4)$$

che associa a ciascuna near collision un valore di rischio crescente al decrescere della prossimità spazio-temporale e, in alcune formulazioni, modulato dal tipo di incontro. Le scelte di design di un indice composito ricadono tipicamente in tre categorie:

- Lineare, in cui il rischio è una somma pesata delle variabili normalizzate.
- Esponenziale, in cui il contributo di ciascuna variabile decade esponenzialmente al suo crescere, amplificando la sensibilità agli eventi a bassa prossimità.
- Geometrico-direzionale, in cui all'indice è aggiunta una modulazione trigonometrica sull'angolo di incontro, per assegnare pesi diversi a *head-on*, *crossing* e *overtaking*.

Le formulazioni esponenziali hanno vantaggi rilevanti per il *risk mapping*: penalizzano in modo marcato gli eventi a bassissime DCPA e TCPA, coerentemente con l'idea della piramide della sicurezza; integrano esplicitamente distanza relativa e dimensioni delle navi; producono un valo-

re continuo e strettamente positivo utilizzabile come peso nella Kernel Density Estimation della Sezione 2.3.4.

Gli indici compositi di questa famiglia condividono alcune criticità. La scelta dei coefficienti è empirica, tipicamente calibrata su un dataset specifico, il che pone questioni di trasferibilità quando l'indice viene applicato a contesti geografici o a regimi di traffico diversi da quello di calibrazione. L'assenza di variabili contestuali (condizioni meteo-marine, stato del traffico, tipo di nave oltre la lunghezza) rappresenta un limite intrinseco delle formulazioni puramente cinematiche. La validazione incrociata è in generale scarsa in letteratura, per ragioni che hanno a che fare con la rarità degli incidenti reali e con la difficoltà di accedere a dataset di riferimento indipendenti.

Le applicazioni specifiche (CR di Rong et al. (2019), VCRO e sue evoluzioni, indici multi-componente) e la discussione critica delle loro calibrazioni sono sviluppate nella Sezione 2.4.3.

2.3.4 Aggregazione spaziale: Kernel Density Estimation pesata

L'indice composito fornisce un valore scalare di rischio per ciascuna near collision, localizzato in un punto geografico. Per passare da un insieme discreto di eventi a una rappresentazione continua sul territorio, cioè a una mappa di rischio, serve un operatore di aggregazione spaziale. Lo strumento standard è la Kernel Density Estimation (KDE), formalizzata da Rosenblatt (1956) e Parzen (1962), nella sua versione pesata con il valore di rischio.

Principio del metodo. La KDE è un metodo non-parametrico per stimare la densità di una distribuzione a partire da un campione di osservazioni. Nel caso unidimensionale, dato un campione $\{x_1, \dots, x_n\}$, la KDE stima la densità nel punto x come:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (2.5)$$

dove K è una funzione kernel (simmetrica, non negativa, con integrale unitario) e h è il *bandwidth*. Intuitivamente, la KDE dispone attorno a ogni osservazione una “campana” di forma K e larghezza h e somma i contributi: dove i punti sono densi, le campane si sovrappongono e la densità stimata è alta; dove sono radi, è bassa.

Nel caso bidimensionale, rilevante per le coordinate geografiche delle near collision:

$$\hat{f}_h(x, y) = \frac{1}{nh_1 h_2} \sum_{i=1}^n K\left(\frac{x-x_i}{h_1}, \frac{y-y_i}{h_2}\right) \quad (2.6)$$

con h_1 e h_2 bandwidth nelle due direzioni e K kernel bidimensionale.

Scelta del kernel e del bandwidth. La scelta del kernel ha un impatto modesto sul risultato finale. Il kernel gaussiano bidimensionale

$$K(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right) \quad (2.7)$$

è quello più diffuso, preferito per regolarità e semplicità computazionale.

La scelta del bandwidth è invece determinante. Un h troppo piccolo produce una stima rumorosa, con picchi isolati attorno a ciascun punto; un h troppo grande appiattisce le strutture locali, perdendo informazione sulla distribuzione. La regola di riferimento è quella di Silverman (1986):

$$h_{Silverman} \propto \hat{\sigma} n^{-1/(d+4)} \quad (2.8)$$

(con $d = 2$ per la KDE bidimensionale). La regola è ottimale sotto l'ipotesi di distribuzione gaussiana, notoriamente disattesa nel traffico marittimo, che si concentra lungo corridoi e nodi: produce quindi un certo *oversmoothing*. Alternative più sofisticate, come la cross-validation o l'uso di un bandwidth adattivo, esistono ma sono raramente impiegate per ragioni computazionali.

KDE pesata per il risk mapping. Nella formulazione standard ogni osservazione contribuisce alla stima della densità con peso unitario. Per il risk mapping è naturale che eventi più gravi pesino più di eventi meno gravi: si adotta allora la KDE pesata

$$\hat{f}_h(x, y) = \frac{1}{h_1 h_2 \sum_i w_i} \sum_{i=1}^n w_i K\left(\frac{x - x_i}{h_1}, \frac{y - y_i}{h_2}\right) \quad (2.9)$$

dove w_i è il peso del singolo evento, tipicamente il suo valore di indice composito di rischio. Il risultato non è più una densità di frequenza ma una *densità di rischio pesata*: una superficie in cui i valori alti indicano congiuntamente alta frequenza e alta severità degli eventi near collision.

Il prodotto della KDE pesata è una mappa continua, definita su una griglia regolare dell'area di studio, che può essere sovrapposta a carte nautiche per la comunicazione visiva e utilizzata come base per confronti spaziali qualitativi.

Limiti e motivazione per lo strumento complementare. La KDE pesata si è imposta come standard per il risk mapping marittimo: è non-parametrica e non richiede ipotesi forti sulla distribuzione degli eventi; è continua e interpretabile e si presta alla sovrapposizione con carte nautiche; è computazionalmente trattabile su grandi dataset grazie a implementazioni basate su FFT.

Presenta tuttavia limiti che motivano l'introduzione di uno strumento complementare. È uno strumento descrittivo, non inferenziale: un valore alto segnala una zona densa di eventi pesati, ma non afferma nulla sulla significatività statistica del cluster. È sensibile al bandwidth: scelte diverse producono mappe visivamente diverse, e la regola di Silverman non è ottimale per le distribuzioni tipiche del traffico marittimo. Infine, non identifica gli hotspot come oggetti discreti: restituisce una superficie liscia, non un insieme perimetrato di eventi "appartenenti a un hotspot" che si

possano contare, caratterizzare e confrontare con altre entità (incidenti storici, infrastrutture, aree protette).

Quest'ultimo limite motiva l'introduzione del secondo strumento di aggregazione spaziale: il clustering density-based.

2.3.5 Clustering density-based per l'identificazione degli hotspot

Dove la KDE pesata produce una mappa continua di rischio, il clustering density-based produce un insieme discreto di hotspot. Per ciascun evento near collision, risponde alla domanda “questo evento appartiene a un hotspot, e se sì a quale?”. I due strumenti sono complementari: la KDE comunica la distribuzione spaziale del rischio; il clustering identifica esplicitamente gli oggetti hotspot su cui ragionare quantitativamente.

Il clustering density-based è particolarmente indicato per l'analisi di dati AIS per due ragioni. In primo luogo, non richiede di specificare a priori il numero di cluster, che nel contesto del traffico marittimo non è noto. In secondo luogo, identifica cluster di forma arbitraria, coerentemente con il fatto che gli hotspot di near collision possono estendersi lungo rotte, concentrarsi ai nodi di incrocio o distribuirsi in configurazioni geometricamente irregolari. Questa caratteristica distingue gli algoritmi density-based dai metodi di clustering basati su distanze centroidali (come k-means), che impongono cluster di forma convessa.

DBSCAN. Il metodo Density-Based Spatial Clustering of Applications with Noise (DBSCAN) costituisce uno degli algoritmi di riferimento per il clustering basato sulla densità. Si basa su due parametri: ϵ (raggio di vicinato) e minPts (numero minimo di punti nel raggio ϵ per considerare una zona “densa”). Ogni punto del dataset è classificato in una delle tre categorie:

- *core point*: ha almeno minPts punti (incluso sé stesso) entro ϵ ;
- *border point*: non è core, ma ricade nel vicinato ϵ di un core point;
- *noise point*: né core né border, un evento isolato.

Un cluster DBSCAN è l'insieme massimale di core point mutuamente raggiungibili per densità, più i border point adiacenti. I noise point restano non assegnati. La Fig. 2.6 illustra schematicamente i tre tipi di punto e il ruolo del raggio ϵ e del parametro minPts.

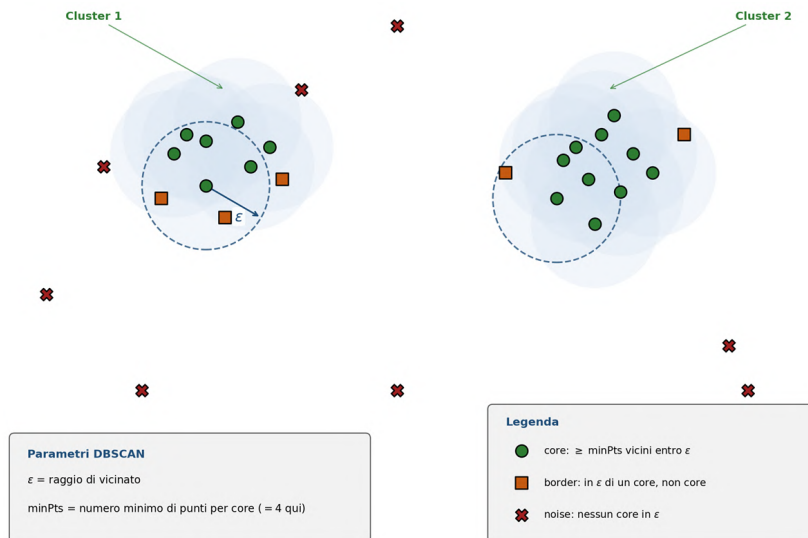


Figura 2.6: Schema esplicativo del funzionamento di DBSCAN. I core point (verdi) hanno almeno minPts vicini entro il raggio ϵ ; i border point (arancioni) ricadono nel vicinato ϵ di un core point ma non soddisfano la soglia minPts; i noise point (X rosse) restano isolati e non sono assegnati ad alcun cluster.

I vantaggi di DBSCAN sono sostanziali: nessun numero di cluster da specificare a priori, identificazione di cluster di forma arbitraria inclusi allungamenti lungo rotte, gestione nativa del rumore. I limiti, altrettanto noti, riguardano principalmente l'ipotesi di densità omogenea: DBSCAN è inadeguato per aree in cui coesistono zone ad alta e bassa densità, come tipicamente le coste a traffico misto, in cui tende a identificare come rumore molti eventi in zona rada che pure presentano rischio non trascurabile. È inoltre sensibile alla scelta dei parametri: ϵ è tipicamente determinato tramite il *k-distance plot*; minPts viene posto pari o superiore alla dimensione dello spazio più uno, spesso $2d$ in ambienti rumorosi.

HDBSCAN. Hierarchical DBSCAN (HDBSCAN) supera il limite principale di DBSCAN (la densità omogenea) attraverso una costruzione gerarchica. Abbandona la nozione di soglia fissa ϵ e lavora invece su un'intera gerarchia di densità, selezionando come cluster quelli più persistenti al variare della soglia. Senza entrare nel dettaglio algoritmico, l'idea è quella di rendere la nozione di cluster robusta rispetto a densità locali diverse: due punti in zona rada non vengono considerati vicini con lo stesso criterio di due punti in zona densa.

Il parametro principale è minClusterSize, dimensione minima di un cluster valido, concettualmente più intuitivo dei parametri di DBSCAN. Il costo computazionale è superiore a DBSCAN ma resta trattabile. Nel contesto di un'area costiera a traffico misto (in cui corridoi commerciali e zone di diporto coesistono con densità molto diverse) HDBSCAN è indicato quando si vogliono individuare cluster sia nelle zone dense sia in quelle rade senza sovra-segmentare le prime o sotto-rilevare le seconde.

ST-DBSCAN: estensione alla dimensione temporale. DBSCAN e HDBSCAN operano nativamente nello spazio. I dati AIS portano tuttavia con sé anche il tempo: una zona può essere un

hotspot in certe ore o in certi periodi dell'anno e non in altri (hotspot diurni o notturni, stagionali, legati a campagne di pesca o a rotte turistiche estive). Mescolare spazio e tempo in un DBSCAN standard non è corretto, per la diversa unità di misura delle due dimensioni.

Spatio-Temporal DBSCAN (ST-DBSCAN) risolve il problema introducendo due soglie separate: ε_1 (raggio di vicinato spaziale) e ε_2 (raggio di vicinato temporale). Due eventi sono vicini se distano meno di ε_1 nello spazio e meno di ε_2 nel tempo. La definizione di core point, density-reachability e cluster segue la stessa logica di DBSCAN ma in uno spazio ibrido spazio-tempo con metriche distinte.

L'adozione di ST-DBSCAN nel dominio AIS-based è ancora limitata e rappresenta una linea di ricerca attiva. Nella prospettiva del lavoro, la disponibilità della dimensione temporale nei dati AIS rende ST-DBSCAN interessante come strumento esplorativo per caratterizzare la stagionalità e la ciclicità degli hotspot, oltre la loro mera distribuzione spaziale. Nel presente lavoro non è applicato: la dimensione temporale è trattata separatamente con il profilo temporale della criticità (Sezione 4.7).

Altri algoritmi e scelta della famiglia. Per completezza si richiama l'esistenza di *Ordering Points To Identify the Clustering Structure* (OPTICS), che non produce direttamente un partizionamento ma un *reachability plot* ordinato, da cui è possibile estrarre cluster a diversi livelli di densità: è utile come strumento esplorativo preliminare. Una menzione specifica merita anche *Mean Shift*, algoritmo di clustering basato sulla stima della densità tramite kernel mobile, che converge ai modi della distribuzione e può essere considerato un'alternativa density-based di natura più simile alla KDE. Mean Shift non richiede di specificare il numero di cluster e identifica cluster di forma arbitraria, ma è sensibile alla scelta della bandwidth e ha un costo computazionale maggiore di DBSCAN per dataset estesi. Esistono inoltre algoritmi non density-based (k-means, clustering gerarchico, spectral clustering) con assunzioni diverse: k-means richiede il numero di cluster a priori e assume cluster convessi; lo spectral clustering cattura strutture globali ma è computazionalmente oneroso. Nessuno di questi risulta particolarmente adatto alla natura dei dati trattati. Nel presente lavoro è impiegato DBSCAN (Sezione 4.8).

2.4 Analisi critica dello stato dell'arte

2.4.1 Introduzione: frammentazione terminologica e safety pyramid

La diffusione globale dei transponder AIS e la progressiva disponibilità di grandi archivi storici hanno trasformato la base dati per lo studio quantitativo del rischio di collisione marittima. L'evoluzione letteraria del ventennio recente ha prodotto un ampio spettro metodologico. Tale ricerca spazia dal micro-monitoraggio dei singoli eventi di interazione (encounter-level) alla macro-caratterizzazione del rischio regionale, culminando nella modellazione spaziale basata su dati reali di navigazione.

Il presupposto teorico trasversale agli approcci data-driven è la *safety pyramid* di Heinrich, richiamata da Du et al. (2020) come fondamento dell'analisi basata su eventi critici non-incidentali: la frequenza di eventi di bassa gravità (le *near collision*) può essere usata come proxy della frequenza di incidenti reali, sfruttando la relazione statistica tra i livelli della piramide. L'ipotesi è tanto più attrattiva quanto più gli incidenti reali sono rari e scarsamente documentati, mentre gli eventi non-incidentali sono identificabili in grandi quantità da dati AIS. Marino et al. (2023) confermano la rilevanza del fenomeno per il contesto mediterraneo, riportando che le collisioni rappresentano oltre il 30 % degli incidenti nelle aree portuali.

La rassegna di Du et al. (2020) segnala tuttavia una criticità trasversale della letteratura: esistono numerose terminologie diverse per designare lo stesso tipo di evento critico (*near miss*, *near collision*, *collision candidate*, *ship conflict*, *traffic conflict*, *non-accident critical event*), senza che esista una definizione condivisa. Questa frammentazione riflette scelte metodologiche sostanziali su cosa costituisca un evento rilevante e su come ne venga quantificata la severità. Una lettura organica della letteratura richiede quindi di distinguere le famiglie metodologiche e di valutarne separatamente i contributi.

Gli strumenti matematici e algoritmici richiamati nella Sezione 2.3 costituiscono il vocabolario di riferimento. Qui l'attenzione si sposta sulle loro applicazioni: come sono stati impiegati, da quali autori, in quali aree geografiche, con quali parametri e con quali risultati.

2.4.2 Approcci geometrici basati su CPA, DCPA e TCPA

L'uso degli indicatori geometrici CPA, DCPA e TCPA come base per la valutazione del rischio di collisione è la linea più consolidata della letteratura AIS-based. Costituisce il punto di partenza operativo per la quasi totalità dei lavori successivi ed è la base su cui si sono innestati gli sviluppi verso indici compositi, domini di manovra e mappe di rischio.

I lavori che impiegano direttamente CPA/DCPA/TCPA come strumento di valutazione del rischio adottano in genere tre strategie. La prima è l'uso di soglie binarie su DCPA e TCPA per identificare eventi critici: un evento è considerato rilevante quando la DCPA prevista scende sotto una certa soglia (tipicamente qualche centinaio di metri in area costiera) entro un TCPA inferiore a una finestra predefinita (tipicamente 5–10 minuti). Questa strategia è alla base della maggior parte dei sistemi operativi di bordo (ARPA, ECDIS) e della loro traduzione in algoritmi di analisi offline su dati AIS. La seconda strategia consiste nel trattare DCPA e TCPA come variabili continue e nel costruire distribuzioni, quantili o statistiche descrittive per caratterizzare il livello di rischio di un'area o di un periodo. La terza, più diffusa negli sviluppi successivi, è quella di usare DCPA e TCPA come input per un indice composito che ne ricompona il significato in un unico valore scalare (Sezione 2.4.3).

Formulazioni geometriche equivalenti a quelle richiamate nella Sezione 2.3.1 sono state proposte in diverse varianti. Rong et al. (2019) ne adottano la versione classica in forma vettoriale. Lin et al. (2024) utilizzano una formulazione alternativa basata esplicitamente su distanza relativa

D_R , rilevamento della target α_t e rotte, operativamente equivalente. Formulazioni analoghe sono presenti in Liu et al. (2022) e in Chen et al. (2021) nel contesto delle rispettive pipeline.

Un contributo metodologico importante è stato il riconoscimento formale dei limiti strutturali del CPA come stima del rischio. Liu et al. (2023) dimostrano attraverso il concetto di *Closest Points of Collision* (CPC) sulle frontiere di collisione dinamica che scenari di head-on, crossing e overtaking possono produrre valori identici di DCPA e TCPA pur avendo implicazioni operative molto diverse in termini di priorità COLREG e manovre disponibili. Questa ambiguità rappresenta la motivazione principale degli sviluppi verso indici che integrino esplicitamente la geometria dell'incontro (Sezione 2.4.3).

La rassegna di Du et al. (2020) identifica inoltre l'ipotesi di moto rettilineo uniforme come uno degli elementi più trasversali e al tempo stesso più limitanti della letteratura AIS-based: il CPA calcolato su dati reali è una previsione condizionata al mantenimento della cinematica corrente, e la sua affidabilità decresce rapidamente in aree in cui le manovre sono frequenti, come acque ristrette, avvicinamenti portuali e zone di pesca.

2.4.3 Indici di rischio compositi in letteratura

Il passaggio dal CPA come indicatore puntuale a un indice composito di rischio risponde a una duplice esigenza: graduare la severità degli eventi near collision in modo continuo e produrre un peso utilizzabile per l'aggregazione spaziale. La letteratura ha sviluppato in questa direzione diverse famiglie di indici, che si possono raggruppare in tre linee principali.

Dall'indice lineare all'indice esponenziale. Nella linea di studi basata su DCPA e TCPA, Deb-nath and Chin (2016) misurano la severità dei conflitti fra unità con un modello fondato sulle due grandezze. Nella forma più semplice, un indice di questo tipo combina DCPA e TCPA con coefficienti di peso:

$$C_{ij} = w_1 \cdot DCPA + w_2 \cdot TCPA \quad (2.10)$$

in cui il rischio è una combinazione pesata delle due variabili, con pesi negativi, così che l'indice cresca al diminuire di DCPA e TCPA. La semplicità del modello è un vantaggio per l'interpretazione e la calibrazione, ma presenta un limite sostanziale: il contributo di DCPA e TCPA scala linearmente, con la conseguenza che un evento a DCPA molto bassa non viene penalizzato in modo sproporzionatamente maggiore di un evento a DCPA moderata, contrariamente a quanto la logica della piramide della sicurezza suggerirebbe.

Rong et al. (2019) superano questo limite proponendo l'indice *Collision Risk* (CR) in forma esponenziale:

$$CR = \alpha_{DCPA} e^{\beta_{DCPA} d_{CPA}} + \alpha_{TCPA} e^{\beta_{TCPA} t_{CPA}} + \alpha_{DR} \frac{L_o + L_t}{D_R} \quad (2.11)$$

dove d_{CPA} e t_{CPA} sono DCPA e TCPA dell'evento, D_R è la distanza relativa istantanea, L_o e L_t sono le lunghezze delle due navi e α_k , β_k sono coefficienti calibrati empiricamente (con $\beta_k < 0$, così da ottenere decadimento esponenziale al crescere di DCPA e TCPA). L'indice è stato applicato ai dati della costa portoghese e utilizzato, nello stesso lavoro, alla produzione di mappe KDE pesate distinte per tipo di incontro. Nel risk scoring pairwise il CR presenta alcune caratteristiche utili: la formulazione esponenziale è coerente con l'intuizione della safety pyramid; l'integrazione della distanza relativa normalizzata sulla somma delle lunghezze porta esplicitamente la dimensione delle navi nell'indice; il valore continuo e strettamente positivo si presta naturalmente all'uso come peso nella KDE. Il limite principale, condiviso con tutti gli indici calibrati empiricamente, è la *trasferibilità*: i coefficienti α_k , β_k sono stati stimati su un dataset specifico e non sono stati validati su contesti geografici diversi.

La linea VCRO e le sue evoluzioni. In parallelo alla linea CR si è sviluppata una famiglia di indici fondati sul *Vessel Conflict Ranking Operator* (VCRO), introdotto da Zhang et al. (2015, 2016). Nella formulazione base:

$$VCRO = k d_{ij}^{-1} v_{ij} [m \sin(\theta_{ij}) + n \sin(2\theta_{ij})] \quad (2.12)$$

dove d_{ij} è la distanza tra le due navi, v_{ij} la velocità relativa, θ_{ij} l'angolo di incontro e k , m , n sono coefficienti da calibrare. L'elemento distintivo del VCRO è la modulazione trigonometrica $\sin \theta + \sin 2\theta$, che assegna pesi differenti agli incontri di tipo head-on, crossing e overtaking. Il VCRO cattura quindi la geometria dell'incontro in modo più esplicito del CR, ma tratta la nave come punto materiale e non include esplicitamente DCPA e TCPA.

La linea VCRO ha dato luogo a due evoluzioni rilevanti. Liu et al. (2022) propongono l'*Enhanced VCRO* (EVCRO), che integra il VCRO con il Quaternion Ship Domain, con coefficienti stimati ai minimi quadrati; l'EVCRO è impiegato nella stessa pipeline come peso per una KDE su griglia spazio-temporale. Liu et al. (2023) propongono il *Kinematics Feature-based VCRO* (KF-VCRO), che introduce i *Closest Points of Collision* per superare l'ambiguità dimostrata tra head-on e overtaking con identici DCPA/TCPA.

Rispetto al CR, la linea VCRO presenta un vantaggio (classificazione più esplicita del tipo di incontro) e due svantaggi (trattamento della nave come punto, necessità di calibrazione più articolata). Le due linee restano distinte nella formulazione; nelle evoluzioni più recenti la linea VCRO integra tuttavia le dimensioni delle unità attraverso lo ship domain, elemento già presente nel CR (Liu et al., 2022).

Indici multi-componente. Un ulteriore sviluppo è rappresentato dagli indici multi-componente, che integrano il rischio cinematico con il rischio di aggregazione. Lin et al. (2024) propongono un indice regionale in cui il rischio pairwise è espresso come:

$$RI_{ij} = \omega_1 \cdot CRI_{DCPA} + \omega_2 \cdot CRI_{TCPA} + \omega_3 \cdot ADRI \quad (2.13)$$

con un termine ADRI (*Aggregated Density Risk Indicator*) legato alla densità locale di navi. Il rischio di sistema per l'area è poi ricomposto come:

$$RI_i = 1 - \prod_j (1 - RI_{ij}) \quad (2.14)$$

riconoscendo che il rischio in un'area densa non è la semplice somma dei rischi pairwise. La formulazione è interessante sul piano concettuale (integra due scale di analisi in un unico indice) ma richiede la scelta di pesi ω_k che la letteratura non ha ancora consolidato su dataset indipendenti.

Valutazione trasversale. Le tre linee (CR esponenziale, VCRO e sue evoluzioni, indici multi-componente) condividono criticità ricorrenti: scelta dei pesi sempre empirica su dataset specifico, assenza di variabili contestuali (condizioni meteo-marine, stato del traffico, tipo di nave oltre la lunghezza), validazione incrociata scarsa, limitata all'area di origine o a benchmark interni al singolo lavoro. Il CR di Rong et al. (2019) ha una forma direttamente utilizzabile come peso in una KDE, per il valore continuo e strettamente positivo.

2.4.4 Approcci basati su densità: KDE e clustering density-based in letteratura

La produzione di mappe di rischio a partire da eventi near collision richiede due strumenti complementari: un operatore di aggregazione spaziale continua (KDE) e un meccanismo di identificazione di hotspot discreti (clustering density-based). La letteratura AIS-based ha consolidato entrambi, con applicazioni significative in diverse aree geografiche.

Applicazioni della Kernel Density Estimation pesata. L'applicazione della KDE pesata con indice di rischio al risk mapping marittimo è descritta in dettaglio da Rong et al. (2019), che la impiegano sulla costa portoghese con kernel gaussiano, una griglia spaziale di 600×500 celle e bandwidth stimato con la regola di Silverman, ottenendo mappe distinte per i tre tipi di incontro. I risultati mostrano che le near collision di tipo *overtaking* si concentrano lungo le rotte principali, mentre *crossing* e *head-on* si localizzano ai nodi di intersezione, una distinzione che non emerge da una KDE non pesata o da un semplice conteggio degli eventi.

Liu et al. (2022) estendono l'approccio a una dimensione spazio-temporale, applicando la KDE pesata con EVCRO su una griglia di $1 \text{ nm} \times 1 \text{ h}$ nelle acque di Chengshantou, in Cina. La scelta di una griglia spazio-temporale anziché puramente spaziale risponde alla necessità di cogliere la variabilità diurna/notturna e stagionale del traffico, ma introduce una sensibilità elevata alla scelta della risoluzione temporale. Altri lavori adottano la KDE con varianti nella scelta del bandwidth (cross-validation invece di Silverman, bandwidth adattivo) senza tuttavia convergere su una pratica consolidata.

Due considerazioni trasversali emergono dalla letteratura KDE-based. La prima riguarda la sensibilità al bandwidth: mappe prodotte con bandwidth diverso sullo stesso dataset possono condurre a interpretazioni visivamente difformi, e la regola di Silverman produce un certo *oversmoothing* in presenza di distribuzioni concentrate lungo corridoi, come tipicamente il traffico marittimo. La seconda è la natura descrittiva dello strumento: la KDE produce una mappa di densità di rischio ma non afferma nulla sulla significatività statistica dei cluster osservati, limite che motiva il ricorso a strumenti complementari come le *spatial statistics* inferenziali (Sezione 2.4.6) o il clustering density-based.

Applicazioni del clustering density-based. L'applicazione del clustering density-based ai dati AIS è documentata in diversi lavori centrali per il risk mapping. Chen et al. (2021) adottano DBSCAN all'interno della loro pipeline *Time-varying Collision Risk* (TCR) per identificare cluster spaziali di navi a rischio, combinando il clustering con il concetto di velocity obstacle e con il Quaternion Ship Domain. Lin et al. (2024) impiegano HDBSCAN, nella variante HDBSCAN ϵ , per il clustering adattivo in aree a densità variabile, integrandolo con il loro indice multi-componente e con un'architettura encoder-decoder Long Short-Term Memory (LSTM) per la previsione regionale. Marino et al. (2023), nella loro rassegna dedicata al Mediterraneo, discutono DBSCAN come uno dei metodi più efficaci per gestire le distribuzioni irregolari del traffico navale, elencandone pregi e limiti.

I parametri adottati dai diversi autori variano in modo significativo. DBSCAN viene tipicamente configurato con ϵ stimato tramite *k-distance plot* e minPts pari a $2d$ o poco superiore; HDBSCAN con minClusterSize calibrato empiricamente sulla scala degli hotspot attesi. La trasferibilità di questi parametri tra contesti geografici diversi non è consolidata, il che rende necessaria una calibrazione specifica quando il metodo viene applicato a una nuova area di studio.

Il contributo principale del clustering density-based rispetto alla KDE è l'identificazione degli hotspot come oggetti discreti, che permette analisi statistiche condizionate (distribuzione dei tipi di incontro dentro ciascun hotspot, estensione spaziale, durata temporale se si usa ST-DBSCAN) e il confronto quantitativo con entità esterne: incidenti storici, infrastrutture, aree protette, schemi di separazione del traffico.

2.4.5 Modelli probabilistici

Accanto agli approcci geometrici e density-based, una parte importante della letteratura classica sulla valutazione del rischio di collisione marittima si fonda su modelli probabilistici. A differenza degli approcci AIS-based, che costruiscono il rischio a partire da eventi osservati, i modelli probabilistici stimano frequenze attese di incidente tramite combinazioni di probabilità di accadimento e probabilità di esito, tipicamente su orizzonti temporali lunghi e su aree estese.

Modelli analitici: IWRAP e COWI. Il riferimento storico più consolidato è il modello IWRAP (*IALA Waterway Risk Assessment Program*), sviluppato dall'International Association of Marine

Aids to Navigation and Lighthouse Authorities (IALA, 2009). L'IWRAP stima la frequenza attesa di collisioni in una via navigabile come prodotto tra il numero atteso di incontri geometrici (calcolato a partire dalla distribuzione del traffico lungo rotte discretizzate) e la probabilità causale di collisione condizionata all'incontro, quest'ultima stimata a partire da dati storici o da giudizio esperto. Il modello è tipicamente applicato su griglie di rotte definite a priori in un'area di studio, e ammette la distinzione tra incontri head-on, crossing e overtaking, ciascuno con probabilità causali distinte.

Il modello COWI (COWI, 2008), sviluppato in ambito consulenziale danese, segue un'impostazione analoga ma introduce una maggiore granularità nella modellazione della distribuzione laterale del traffico attorno alla rotta nominale, elemento particolarmente rilevante in aree come lo stretto del Grande Belt o i terminal portuali nord-europei per cui il modello è stato calibrato.

Entrambi i modelli condividono una logica comune: il rischio è la composizione di una componente geometrica (incontri) e di una componente causale (probabilità di collisione dato un incontro). La calibrazione avviene su dati di traffico aggregati (distribuzioni di rotta, volumi annui, tipologie di nave) e su statistiche di incidenti storici. Sono modelli pensati per la pianificazione infrastrutturale: valutazione di schemi di separazione del traffico, dimensionamento di schemi VTS, analisi di impatto per nuovi insediamenti portuali o campi eolici offshore.

Reti bayesiane. Un secondo filone probabilistico è quello delle *Bayesian Networks* (BN), che modellano le relazioni causali tra fattori di rischio e probabilità di incidente attraverso un grafo diretto aciclico. I nodi rappresentano variabili (caratteristiche della nave, condizioni ambientali, affaticamento dell'equipaggio, tipologia di incontro), gli archi le dipendenze probabilistiche tra esse, e l'inferenza consente di propagare evidenza da variabili osservate a variabili non osservate. La Fig. 2.7 propone un esempio prototipale di rete bayesiana per il problema del rischio di collisione, articolata su tre livelli (variabili di input, variabili intermedie, variabile target).

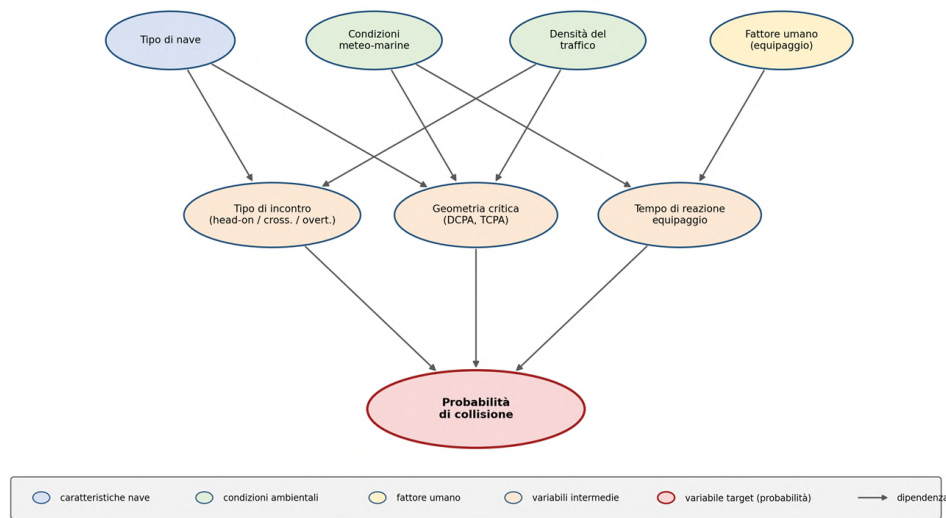


Figura 2.7: Esempio di Bayesian Network per il rischio di collisione marittima. Tre livelli di variabili: input (caratteristiche della nave, condizioni meteo-marine, densità del traffico, fattore umano), intermedie (tipo di incontro, geometria critica, tempo di reazione) e target (probabilità di collisione). Gli archi rappresentano le dipendenze probabilistiche tra le variabili.

Sotiralis et al. (2016) sviluppano una BN che integra elementi della tecnica TRACER per stimare la probabilità di collisione dovuta all'errore umano, e la applicano a una nave feeder in navigazione nello Stretto di Dover. Sul fattore umano si concentrano anche Chauvin et al. (2013), che analizzano con una versione dell'HFACS adattata al contesto marittimo 39 collisioni indagate dagli organismi investigativi britannico (MAIB) e canadese (TSB), senza costruire una rete bayesiana. Marino et al. (2023), nella loro rassegna, richiamano i modelli bayesiani come classe rilevante per l'analisi causale degli incidenti, ma ne evidenziano i limiti di applicazione al risk mapping ad alta risoluzione.

Confronto con gli approcci AIS-based. Nella rassegna di Marino et al. (2023), che classifica i lavori esaminati anche per famiglia di metodo, gli approcci a indicatori sintetici del rischio (*Synthetic Indicators*) rappresentano il 50 % dei lavori; gli approcci di tipo statistico o probabilistico (analisi statistica, *Fault Tree Analysis* e reti bayesiane) ne rappresentano complessivamente il 34 %.

Approcci probabilistici e approcci AIS-based hanno finalità diverse. I modelli probabilistici operano a un livello di aggregazione superiore, producono stime di frequenza attesa di incidente su orizzonti temporali estesi, e sono lo strumento naturale per valutazioni a priori di configurazioni infrastrutturali. Gli approcci AIS-based operano a posteriori sul traffico effettivamente osservato, identificano eventi critici a scala di singolo incontro e producono mappe di rischio ad alta risoluzione spaziale. I due filoni sono complementari: una valutazione completa del rischio marittimo in un'area tipicamente richiede entrambi.

Per gli obiettivi della presente tesi il riferimento operativo è l'approccio AIS-based. Il modello geometrico dei candidati a collisione di Pedersen (1995), alla base dei modelli di tipo IWRAP,

non è stato implementato.

2.4.6 Approcci alternativi e spatial statistics

Accanto alle linee principali discusse finora, la letteratura ha sviluppato alcuni approcci alternativi che è opportuno richiamare per completezza.

Velocity Obstacle. Chen et al. (2021) integrano il concetto di *velocity obstacle* (VO) con un *Time-varying Collision Risk* (TCR) basato su Quaternion Ship Domain e con clustering density-based. Il VO sposta il ragionamento da “qual è il punto di minimo avvicinamento?” a “quante velocità evasive sono disponibili senza collisione?”, cogliendo la difficoltà di evitamento come dimensione del rischio. È un’alternativa concettualmente interessante al CPA, ma nelle sue formulazioni lineari più diffuse richiede assunzioni sul comportamento futuro della target e non cattura traiettorie curve.

Entropia informativa. Feng et al. (2022) propongono un approccio fondato sull’entropia di Shannon applicata alla distribuzione dei Course Over Ground:

$$H = -\sum_i p_i \log_2 p_i \quad (2.15)$$

L’idea è che un’area ordinata (navi che si muovono prevalentemente in una direzione) abbia entropia bassa, mentre un’area disordinata abbia entropia alta. L’approccio è semplice e computazionalmente leggero, ma si tratta di una proxy indiretta del rischio: un’alta entropia può riflettere semplicemente la geometria del sito (incroci, ancoraggi) senza corrispondere a un effettivo pericolo di collisione.

Autocorrelazione spaziale e hotspot analysis inferenziale. Una categoria a parte è rappresentata dalle *spatial statistics* inferenziali, che aggiungono un livello di validazione statistica alle mappe prodotte con KDE o con clustering. Rong et al. (2021) applicano al dominio marittimo l’indice globale di Moran e la statistica locale di Getis-Ord G_i^* . L’indice di Moran misura l’autocorrelazione spaziale globale di una variabile distribuita su celle contigue e consente, attraverso uno z-score sotto ipotesi di casualità, di confermare o rigettare la presenza di clustering. La statistica G_i^* localizza invece gli hotspot, fornendo per ciascuna cella uno z-score che ne segnala la significatività come cluster di valori alti. Rong et al. (2021) applicano questi strumenti alla stessa area della loro pipeline KDE del 2019 e identificano quattro hotspot statisticamente significativi.

Dal punto di vista metodologico, queste tecniche aggiungono rigore inferenziale alla pipeline descrittiva (KDE) ed esplorativa (clustering). La loro applicazione a grandi dataset AIS presenta tuttavia costi computazionali non trascurabili e richiede scelte di modellazione (definizione del vicinato, standardizzazione dei pesi, correzione per test multipli) che ne rendono la calibrazione

delicata. Nel presente lavoro l'indice di Moran e la statistica G_i^* sono applicati alle superfici di criticità di entrambe le aree (Sezione 4.6).

2.4.7 Approcci data-driven e deep learning

Lo sviluppo di architetture di machine learning e deep learning per dati spazio-temporali ha aperto una linea di ricerca recente nel collision risk assessment. Liu et al. (2022) integrano il loro framework KDE con una *Convolutional Long Short-Term Memory* (ConvLSTM) per la previsione spazio-temporale del rischio; Lin et al. (2024) propongono un'architettura encoder-decoder LSTM per la previsione regionale, accoppiata con HDBSCAN; Zhang et al. (2020) applicano *Convolutional Neural Networks* per la classificazione del rischio.

Il vantaggio di questi approcci è la capacità di catturare pattern non lineari e dipendenze spazio-temporali che sfuggono ai modelli analitici. I limiti sono ben noti: i modelli sono largamente *black-box* e scarsamente interpretabili, richiedono dataset estesi non sempre disponibili in aree nuove, apprendono tipicamente la dinamica di una proxy di rischio calcolata a monte (un VCRO o un CR) e generalizzano male fuori dominio.

Nella prospettiva della tesi, gli approcci data-driven sono richiamati per completezza ma non rappresentano l'oggetto del lavoro. Una pipeline analitica basata su indicatori di rischio espliciti resta un prerequisito per qualunque estensione predittiva: un modello di deep learning addestrato su una proxy di rischio ne apprende la dinamica, compresi i suoi limiti.

2.4.8 Eterogeneità della letteratura

La letteratura sulla valutazione del rischio di collisione AIS-based, pur ricca, è profondamente eterogenea lungo cinque dimensioni. Questa eterogeneità riflette scelte metodologiche sostanziali e condiziona la confrontabilità dei risultati tra contesti diversi.

Definizione dell'evento critico. Du et al. (2020) documentano numerose terminologie diverse per l'evento non-incidentale. Alcuni lavori usano la violazione dello ship domain, altri soglie fisse su DCPA/TCPA, altri ancora operatori di conflitto superiori a soglia. Il “numero di near collision” in studi diversi non è, in generale, confrontabile.

Costruzione dell'indicatore di rischio. Le scelte vanno dai singoli DCPA/TCPA alle combinazioni di DCPA e TCPA (Debnath and Chin, 2016), alle formulazioni esponenziali (Rong et al., 2019), agli indici VCRO-based (Zhang et al., 2015; Liu et al., 2022, 2023), fino alle proxy indirette come l'entropia del COG (Feng et al., 2022) e agli indici multi-componente (Lin et al., 2024).

Dati e contesto applicativo. La letteratura oscilla tra AIS reale, simulazioni e studi ibridi. Le finestre temporali variano da pochi mesi ad anni; le aree geografiche spaziano tra acque ristrette asiatiche, coste atlantiche europee, ZEE nord-europee e porti del Sud-Est asiatico. La trasferibilità di un indice calibrato su Ningbo-Zhoushan o sulla costa portoghese al Mar Ligure non è scontata.

Scala di analisi. Esistono tre scale spesso confuse ma non intercambiabili: encounter-level (coppia nave-nave), locale/hotspot (distribuzione spaziale in un'area ristretta) e regionale (rischio complessivo di un'area estesa). Un indice ottimale a una scala non lo è necessariamente a un'altra.

Validazione. Du et al. (2020) rilevano che la maggior parte dei lavori presenta validazione insufficiente: gli indici sono giustificati per costruzione teorica e illustrati su un caso di studio, raramente confrontati con dati indipendenti di incidenti reali. Il limite è in parte intrinseco alla rarità degli incidenti, ma resta un punto debole riconosciuto del campo.

A queste cinque dimensioni si aggiunge la coesistenza di classificazioni costruite su criteri diversi: Goerlandt and Montewka (2015) distinguono i lavori per prospettiva sul concetto di rischio (*Strong Realist*, *Scientific Proceduralist*, *Moderately Constructivist*, fra le altre), mentre altre classificazioni li raggruppano per famiglia di metodo; Marino et al. (2023) le applicano entrambe alla stessa raccolta di lavori.

A chiusura della rassegna, la Tab. 2.3 riassume in forma sinottica i principali lavori discussi nella sezione, organizzati per famiglia metodologica, area di studio, indicatore di rischio adottato, strategia di aggregazione spaziale e contributo principale.

Tabella 2.3: Sintesi dei principali lavori citati nello stato dell'arte.

Autori (anno)	Famiglia	Area	Indicatore di rischio	Aggregazione	Contributo principale
IALA, IWRAP (2009)	Modello probabilistico.	Applicaz. internaz.	Frequenza attesa di collisione (incontri $\times$ prob. causale)	Griglie di rotte a priori	Modello analitico di riferimento per pianificazione e TSS
Chauvin et al. (2013)	Analisi dei fattori umani (HFACS)	Collisioni indagate da MAIB (UK) e TSB (Canada)	Fattori contributivi classificati con HFACS	Corrispondenze multiple e clustering gerarchico	Tre profili di fattori umani e organizzativi
Zhang et al. (2015, 2016)	Risk index VCRO	Mar Baltico settentrionale	VCRO con modulazione trigonometrica dell'angolo di incontro	Ranking pairwise	Classificazione esplicita del tipo di incontro
Sotiralis et al. (2016)	Rete bayesiana	Stretto di Dover	Probabilità di collisione per errore umano (BN con TRACER)	Albero degli eventi per le conseguenze	Integrazione dei fattori umani nel rischio di collisione
Debnath & Chin (2016)	Risk index lineare	Porto di Singapore	Indice di conflitto basato su DCPA e TCPA	Analisi statistica pairwise	Indice di conflitto su DCPA e TCPA (tecnica NTCT)
Rong et al. (2019)	Risk index + densità	Costa portoghese	CR esponenziale (DCPA, TCPA, distanza rel., lunghezze)	KDE pesata con CR, kernel gaussiano, griglia 600×500	Risk mapping AIS-based con KDE pesata per tipo di incontro
Du et al. (2020)	Rassegna critica	n.a.	n.a.	n.a.	Rassegna di riferimento; documenta la frammentazione terminologica
Zhang et al. (2020)	Deep learning	Mar Baltico settentrionale	Classificazione di rischio via CNN su dati AIS	Predizione spaziale	CNN per classificazione del rischio
Chen et al. (2021)	Velocity Obstacle + clustering	Acque portuali cinesi	Time-varying Collision Risk (TCR) basato su QSD	DBSCAN su navi a rischio	Integrazione di VO, QSD e clustering density-based
Rong et al. (2021)	Spatial statistics	Costa portoghese	CR + Moran's I e Getis-Ord Gi*	Validazione inferenziale degli hotspot	Applicazione di Moran's I e Getis-Ord Gi* agli hotspot di near collision
Feng et al. (2022)	Entropia informativa	Area portuale asiatica	Entropia di Shannon sulla distribuzione dei COG	Mappa spaziale di entropia	Proxy indiretta del rischio basata sulla dispersione direzionale
Liu R.W. et al. (2022)	Risk index + densità + DL	Acque di Chengshantou (Cina)	EVCRO con QSD, coefficienti calibrati LSQ	KDE pesata spazio-temp. (1 nm $\times$ 1 h); ConvLSTM	Pipeline spazio-temporale con KDE pesata e ConvLSTM
Liu Z. et al. (2023)	Risk index VCRO	Acque cinesi	KF-VCRO con Closest Points of Collision (CPC)	Ranking encounter-level	Dimostrazione dell'ambiguità del CPA tra head-on e overtaking
Marino et al. (2023)	Rassegna	Mediterraneo	n.a.	n.a.	Rassegna dedicata al Mediterraneo; collisioni > 30% nelle aree portuali
Lin et al. (2024)	Indice multi-comp. + DL	Regione costiera	Indice RI con termine di densità (ADRI); ricomposizione di sistema	HDBSCAN adattivo; LSTM	Integrazione rischio cinematico e di aggregazione

2.5 Limiti della letteratura e research gap

Dall'analisi emergono quattro direzioni lungo cui la letteratura presenta limiti o aree meno esplorate. Esse definiscono, nel loro insieme, lo spazio entro cui si colloca il presente lavoro.

Gap geografico. Nonostante il Mediterraneo sia un bacino ad alta densità di traffico e nonostante Marino et al. (2023) mostrino che le collisioni rappresentano il 37 % degli incidenti nelle aree portuali del Mediterraneo, la letteratura AIS-based è dominata da studi su acque asiatiche, nord-europee e atlantiche. Il Mediterraneo settentrionale e l'area costiera ligure, in particolare, non sono stati finora oggetto di una pipeline completa di risk mapping basata su dati AIS reali. Il progetto Interreg Marittimo JASON, nel cui ambito si inserisce la tesi, rende disponibili dati AIS qualificati e definisce un contesto operativo specifico.

Gap metodologico. La letteratura AIS-based ha consolidato separatamente diversi strumenti per la valutazione del rischio di collisione (rilevazione degli eventi critici, scoring composito, aggregazione spaziale continua, identificazione di hotspot discreti), ma la loro integrazione in un'unica pipeline applicata allo stesso dataset di eventi pesati resta poco esplorata. Le combinazioni esistenti tendono a utilizzare questi strumenti in modo parziale o in contesti disgiunti. Questa osservazione definisce una direzione di contributo, la cui concretizzazione in scelte metodologiche specifiche è rimandata al capitolo di metodologia.

Gap sul traffico misto costiero. Gli studi AIS-based si concentrano prevalentemente su aree a traffico omogeneo: corridoi commerciali, acque portuali, canali ristretti. Le aree costiere a *traffico misto* (commerciale, passeggeri, pesca, servizi e diporto, con rotte incrociate e forte variabilità temporale) sono meno studiate. L'area costiera ligure è un esempio paradigmatico di questa categoria, e l'assenza di letteratura dedicata rende non banale la taratura degli indici e la scelta degli algoritmi.

Gap applicativo. La letteratura accademica è raramente collegata a progetti operativi di gestione della sicurezza marittima. Il progetto JASON rappresenta un contesto in cui la produzione analitica ha un destinatario operativo definito e obiettivi applicativi concreti. Il metodo proposto mantiene quindi un collegamento diretto tra l'analisi dei dati AIS e le possibili applicazioni nel monitoraggio del traffico marittimo.

Capitolo 3

Presentazione e Pre-elaborazione dei dati

Il presente capitolo descrive i dataset utilizzati come base empirica del lavoro, ne ricostruisce la genesi a partire dalle sorgenti AIS grezze e ne discute la struttura, la copertura e i limiti. Il capitolo si articola come segue: la Sezione 3.1 introduce la fonte primaria, costituita dai dati AIS dell'anno 2023 nell'ambito del progetto Interreg Marittimo JASON; la Sezione 3.2 illustra le fasi di pre-processing che hanno trasformato i record grezzi in un dataset di interazioni nave-nave; la Sezione 3.3 presenta il dataset di lavoro, organizzato a livello di *encounter* sincroni e geograficamente prossimi; la Sezione 3.4 documenta il passaggio di raffinamento cinematico, volto a isolare gli encounter cinematici dalle anomalie; la Sezione 3.5 fornisce una caratterizzazione esplorativa preliminare del dataset; la Sezione 3.6 raccoglie i limiti residui e le scelte di trasparenza metodologica; la Sezione 3.7, infine, introduce un secondo caso di studio, lo Stretto di Messina, analizzato su tre anni (2022, 2023 e 2024) e utilizzato per verificare la generalizzabilità della pipeline a un regime di traffico differente.

L'elaborazione è organizzata in una pipeline modulare di script MATLAB, dedicati ai diversi passaggi di importazione, pre-processing, ricostruzione dell'evoluzione degli encounter, raffinamento cinematico e preparazione dei dati per le analisi successive. La stessa struttura è applicata alle due aree di studio, Tigullio e Stretto di Messina, tenendo conto delle differenze nelle rispettive fonti dati.

3.1 La fonte grezza: dati AIS 2023 dell'area ligure

Il dataset di partenza è costituito dall'archivio storico AIS dell'anno 2023 relativo all'area costiera ligure compresa, in prima approssimazione, tra Bogliasco e Sestri Levante (cfr. Fig. 3.1).

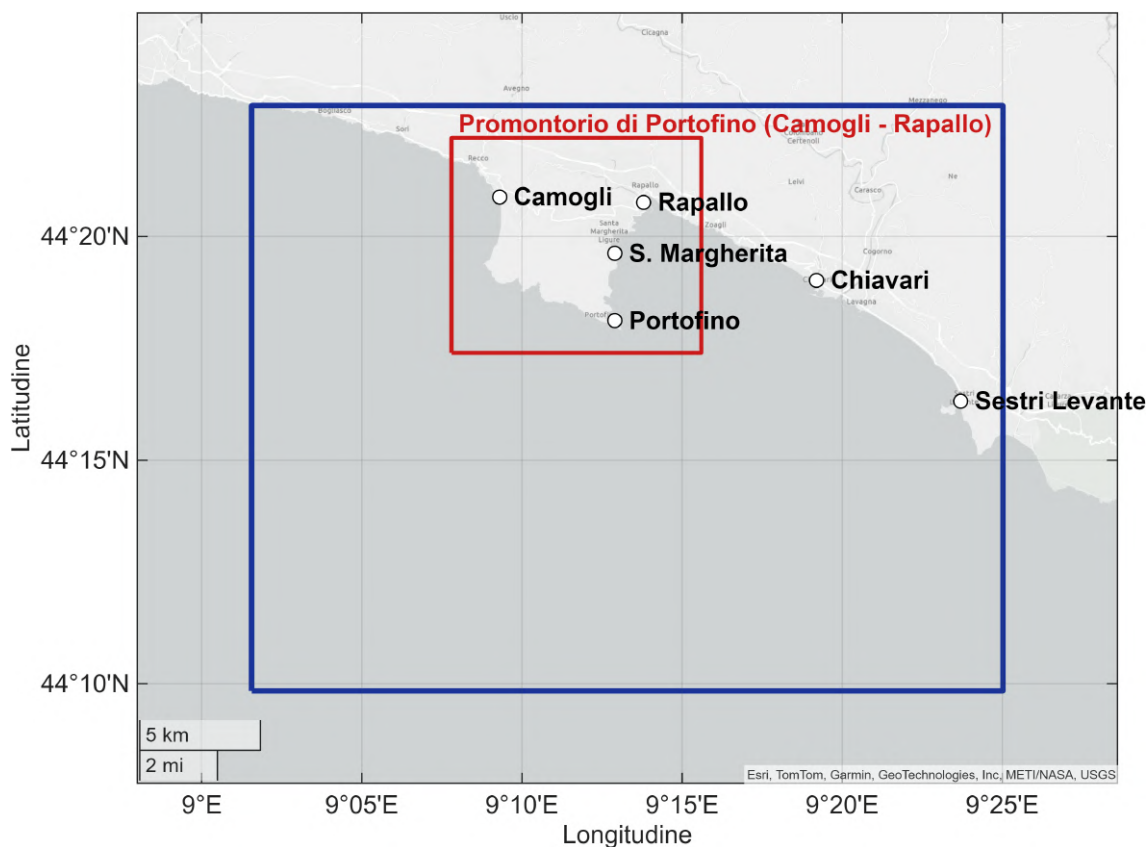


Figura 3.1: Area di studio del Tigullio, coperta dal dataset AIS 2023, con il riquadro della vista di dettaglio del promontorio di Portofino (44,29–44,37° N; 9,13–9,26° E).

L'archivio si presenta in formato *Comma-Separated Values* (CSV), organizzato su base mensile, denominato secondo la convenzione `ais-a-raw_BBOX.2023-MM.csv`, dove MM indica il mese e il suffisso BBOX rimanda al ritaglio spaziale operato in fase di acquisizione. Ciascun record corrisponde a un singolo messaggio AIS ricevuto da una stazione costiera e include i campi tipici descritti nella Sezione 2.1.3 (cfr. Tab. 2.1): identificativo Maritime Mobile Service Identity (MMSI), nominativo radio (*call sign*), nome della nave (*target name*), numero IMO, codice di tipologia operativa (*ship cargo type*) secondo la classificazione ITU-R M.1371 (International Telecommunication Union, 2014), posizione geografica (latitudine, longitudine), velocità rispetto al fondo (*Speed Over Ground*, SOG), rotta rispetto al fondo (*Course Over Ground*, COG), prua vera (*true heading*) e quattro campi geometrici (*ref_pos_bow*, *ref_pos_stern*, *ref_pos_beam*, *ref_pos_starbeam*) che indicano la posizione dell'antenna GNSS rispetto agli estremi della nave e da cui sono ricavabili lunghezza e larghezza del mezzo.

L'insieme include sia messaggi provenienti da transponder di Classe A (obbligatori per le navi soggette a SOLAS) sia, per quanto disponibili, da transponder di Classe B installati su unità non SOLAS, prevalentemente da diporto. Nell'area in esame le unità da diporto sono la componente più numerosa del traffico durante la stagione estiva.

A titolo esemplificativo, il solo file relativo a gennaio 2023, classe A, contiene 61 218 record

corrispondenti a 186 MMSI distinti, di cui 183 dotati di numero IMO. Tali conteggi si riferiscono ai dati raw, prima dell'interpolazione temporale. La copertura mensile è continua: il primo record ricevuto è datato 2023-01-01 00:01:18, l'ultimo 2023-01-31 23:56:33. L'estensione all'intero anno e l'inclusione dei dispositivi di Classe B portano la popolazione navale tracciata a 3 006 unità distinte (Sezione 3.2.1).

3.2 Il pre-processing dei dati grezzi

La trasformazione dei messaggi AIS grezzi nel dataset di lavoro analizzato è avvenuta in due fasi successive, descritte nelle sottosezioni seguenti. La prima fase ha consolidato i CSV mensili in una struttura arricchita di informazioni statiche per nave; la seconda ha individuato e isolato gli eventi di prossimità tra coppie di unità, producendo il dataset *encounter-level* utilizzato nei capitoli successivi.

3.2.1 Consolidamento dell'archivio: dai CSV alla struttura unificata

I dodici file CSV mensili sono stati convertiti in archivi *SQLite* separati, uno per ciascun mese, per facilitare l'interrogazione e la gestione dei dati. Il campo temporale *position_dt*, presente nei CSV in due formati, con e senza millisecondi, è stato uniformato al formato *yyyy-MM-dd HH:mm:ss*, con arrotondamento al secondo. Per ciascun mese è stato quindi ottenuto un database relazionale contenente i messaggi AIS originari, organizzati in sedici colonne e con un formato uniforme per il campo temporale.

La conversione in *SQLite* è stata utilizzata nella fase preliminare di interrogazione ed esplorazione dei dati. Le analisi successive sono state invece eseguite in ambiente MATLAB.

In una seconda passata, l'insieme dei CSV mensili è stato consolidato in una struttura MATLAB gerarchica, organizzata in tre livelli. Al primo livello le unità navali sono partizionate secondo il *ship cargo type*: ogni codice di tipologia (es. 36 per *sailing*, 37 per *pleasure craft*, 70 per *cargo*, 60 per *passenger*) costituisce un campo della struttura. All'interno di ciascun gruppo di tipologia, le unità sono ulteriormente identificate dal numero IMO laddove disponibile o, in alternativa, dal solo MMSI per le unità non SOLAS che ne sono prive.

A ciascuna unità sono associati la serie temporale completa dei suoi record, gli identificativi statici (*mmsi*, *target_name*, *imo_number*) e le caratteristiche geometriche come la lunghezza fuori tutto e la larghezza, calcolate come somma dei campi *ref_pos*. In fase di consolidamento è stato applicato un primo filtro sulla velocità (*SOG* > 1 nodo) volto a escludere i record relativi a unità ferme o in ormeggio.

Al termine del consolidamento, l'archivio annuale 2023 contiene 3 006 MMSI distinti, distribuiti su 56 codici di *ship cargo type*. La distribuzione, riportata in Fig. 3.2, è disomogenea: le tipologie *sailing* (codice 36) e *pleasure craft* (codice 37) sommano insieme oltre il 67 % del totale; seguono le navi da carico e le unità passeggeri, con quote inferiori.

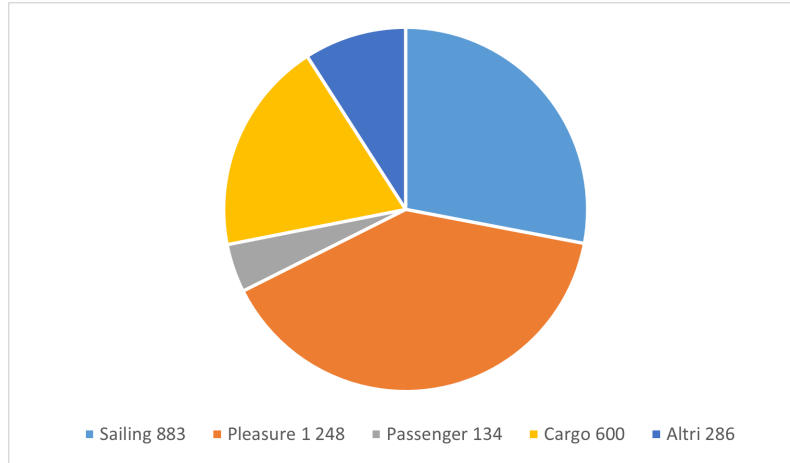


Figura 3.2: Distribuzione delle 3 006 unità navali tracciate nel 2023 per macro-tipologia. Sailing e pleasure craft dominano la popolazione; cargo e passenger sono minoritari.

3.2.2 Estrazione degli encounter sincroni

Il dataset degli *encounter* è stato estratto a monte del presente lavoro, nell’ambito del progetto JASON, e fornito in formato JSON. Il filtro sulla velocità descritto nella Sezione 3.2.1 non è applicato a questo dataset, che comprende pertanto anche record con SOG inferiore a 1 nodo. L’archivio consolidato è utilizzato esclusivamente per il recupero dei dati statici delle unità, secondo la procedura descritta nella Sezione 3.6.4.

La definizione di *encounter* è quella introdotta nella Sezione 1.1. Ai fini dell’estrazione, un *encounter* è costituito dalla sequenza di osservazioni sincrone associate a una coppia di unità, sulla quale vengono successivamente applicati i criteri di pulizia e di affinamento cinematico descritti nelle sezioni successive.

Le serie temporali delle unità sono state risincronizzate mediante interpolazione lineare su un passo temporale regolare di un minuto. Per ogni coppia di navi compresenti nella stessa finestra temporale è stata quindi calcolata, a ogni istante, la distanza reciproca. La coppia è stata identificata come *encounter* quando la distanza è risultata inferiore a 500 m; sono stati quindi mantenuti i record temporalmente contigui per i quali la condizione di prossimità è rimasta verificata.

La soglia di 500 m è stata fissata nella fase di estrazione. Trattandosi di una soglia fissa, il criterio non tiene conto della dimensione delle unità: risulta relativamente ampio per le piccole unità da diporto e relativamente restrittivo per le navi di maggiori dimensioni.

Per ciascun *encounter* sono stati inoltre registrati il valore minimo della distanza osservata lungo la sequenza temporale, indicato come *min_dist*, e il relativo istante, indicato come *cpa_time*.

Per ciascun *encounter* sono stati conservati sia i metadati di sintesi (istante di inizio, istante di fine, distanza minima, posizione e cinematica delle due unità all’istante di minima distanza osservata), sia la sequenza completa dei record minuto per minuto. Quest’ultima struttura costituisce la base per le analisi cinematiche descritte nei capitoli successivi.

3.3 Il dataset di lavoro

3.3.1 Dimensione e struttura

Il dataset finale, esportato in formato *JavaScript Object Notation* (JSON) e successivamente importato nell'ambiente MATLAB come *struct array*, si compone di 23 769 encounter riferiti al 2023, per un totale di 110 395 record sincroni a passo di un minuto. Lo schema dei campi è uniforme e verificato su tutti gli encounter: ciascuno contiene quindici campi a livello di evento e un sotto-array *records* di lunghezza variabile, con sedici campi per ogni record sincrono. La struttura completa dei campi è riportata in Tab. 3.1.

Sono coinvolte 1 650 unità navali distinte (MMSI univoci), che danno luogo a 8 331 coppie distinte. Il numero di record per encounter ha distribuzione asimmetrica: la mediana è pari a 2 record temporali, mentre il massimo osservato è di 216 record, corrispondenti a un'interazione protratta per oltre tre ore e mezza.

Tabella 3.1: Schema dei campi del dataset *encounter-level* (formato JSON).

Livello	Campo	Tipo	Descrizione
encounter	mmsi_i, mmsi_j	intero	MMSI delle due unità coinvolte
encounter	start, end	datetime	istanti di inizio e fine dell'encounter
encounter	min_dist	reale [m]	distanza minima osservata
encounter	cpa_time	datetime	istante di distanza minima osservata
encounter	lat_*_cpa, lon_*_cpa	reale [°]	posizioni geografiche all'istante di minima distanza osservata
encounter	sog_*_cpa, cog_*_cpa	reale	velocità e rotta in tale istante (kn, deg)
encounter	records	array	sotto-array dei record sincroni
record	t	datetime	timestamp del record (passo 1 min)
record	lat_i/j, lon_i/j	reale [°]	posizioni geografiche istantanee
record	x_i/j, y_i/j	reale [m]	posizioni cartografiche (ausiliarie)
record	sog_i/j, cog_i/j	reale	velocità e rotta istantanee
record	distance	reale [m]	distanza istantanea

3.3.2 Copertura spaziale e temporale

La finestra temporale del dataset si estende dal 2023-01-02 12:51:00 al 2023-12-30 16:33:00, con copertura annuale. La distribuzione mensile, mostrata in Fig. 3.3, non è uniforme: il 98 % degli encounter si concentra nei mesi da aprile a ottobre, con picchi a giugno (4 323 encounter), agosto (4 719) e settembre (4 618). I mesi invernali contribuiscono complessivamente per meno di 500 eventi. Tale stagionalità è coerente con la natura prevalentemente turistica del traffico nell'area in esame.

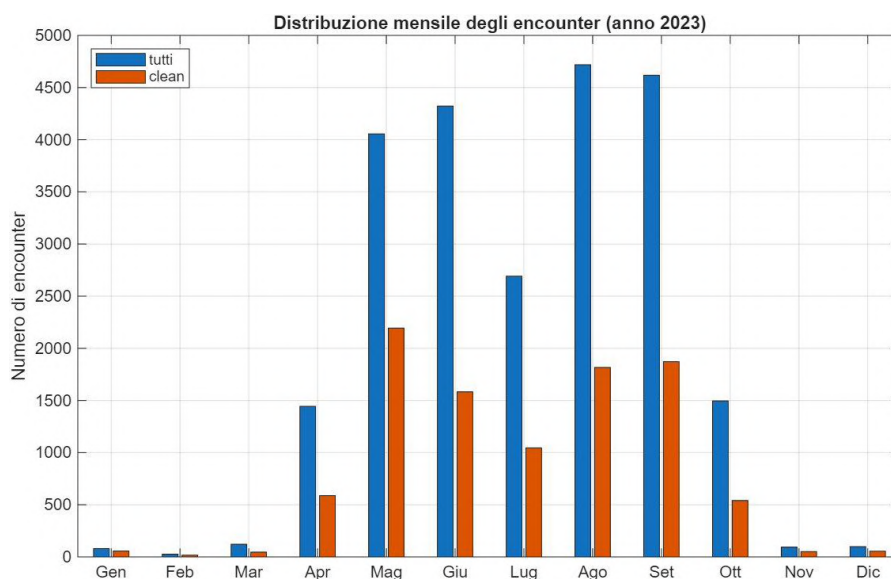


Figura 3.3: Distribuzione mensile degli encounter nel 2023. Confronto tra l'intero dataset e quello clean. Si osserva una concentrazione del 98 % degli eventi nei mesi tra aprile e ottobre.

La copertura spaziale è delimitata in un riquadro compreso tra $44,164^{\circ}$ N e $44,382^{\circ}$ N in latitudine e tra $9,026^{\circ}$ E e $9,417^{\circ}$ E in longitudine. L'area include il promontorio di Portofino e si estende a est fino a Sestri Levante e a nord-ovest fino al litorale tra Camogli e Bogliasco. La distribuzione mostra una forte disomogeneità: alcune zone come l'imboccatura del porto di Portofino e il tratto antistante Santa Margherita Ligure–Rapallo registrano densità di encounter di ordini di grandezza superiori al resto dell'area. Nell'area coesistono rotte commerciali di transito, traffico passeggeri locale, pesca e traffico da diporto, che è la componente più numerosa.

3.4 Raffinamento cinematico del dataset

Il dataset depurato dagli errori di tracciamento (*clean dataset*: encounter con almeno tre record, con SOG e COG validi per entrambe le unità, privi di valori SOG anomali e di durata non inferiore a un minuto) viene sottoposto a una procedura di filtraggio cinematico. Il filtro esclude gli encounter associati ad affiancamenti, stazionamento reciproco e anomalie cinematiche compatibili con errori di trasmissione o con oscillazioni della traccia a bassa velocità.

La logica del filtro cinematico è stata articolata in quattro macro-criteri sequenziali:

- **Rimozione degli artefatti AIS:** vengono identificati e scartati gli encounter che presentano valori cinematici fisicamente non plausibili. Il criterio adottato isola i picchi di velocità superiori a 60 nodi. Tali valori vengono rimossi a monte del filtraggio.
- **Rimozione degli affiancamenti:** vengono esclusi gli encounter in cui le due unità procedono su rotte sostanzialmente parallele e a velocità simile. Operativamente, l'affiancamento è riconosciuto quando, lungo l'evoluzione, la differenza di rotta (COG) e la differenza di velocità (SOG) fra le due unità si mantengono entrambe entro le soglie definite nel filtro.

- **Definizione dello stazionamento reciproco:** un'interazione viene classificata come stazionamento e, di conseguenza, esclusa, se entrambe le unità coinvolte presentano una velocità mediana inferiore a 1,0 nodo. Si è adottata la mediana anziché la media per la sua maggiore robustezza rispetto alle anomalie e ai picchi di velocità presenti nei dati AIS. Il criterio mantiene gli encounter cinematicamente attivi nei quali un'unità in movimento si avvicina a un'unità ferma, all'ancora o alla boa, come avviene presso i campi boe dell'Area Marina Protetta di Portofino. Gli encounter in cui una sola unità ha velocità mediana inferiore a 1,0 nodo restano quindi nel dataset cinematico: nel Tigullio sono il 10,7 % degli encounter cinematici.
- **Identificazione dei comportamenti caotici:** per non penalizzare la navigazione a bassa velocità, tipica del naviglio da diporto in transito sottocosta, la soglia di velocità per l'identificazione dei profili caotici è fissata a 1,5 nodi. Un evento è classificato come caotico se la velocità mediana dell'unità più lenta si colloca al di sotto di tale soglia e, contestualmente, il suo indice di dispersione supera il 50 %. L'indice è definito come rapporto fra deviazione standard e mediana della velocità dell'unità più lenta, valutata record per record. Il denominatore è limitato inferiormente a 0,5 nodi, per evitare che velocità mediane prossime a zero producano valori di dispersione artificialmente elevati. Per mediane superiori a 0,5 nodi, la soglia del 50 % corrisponde a una deviazione standard pari alla metà della velocità mediana; al di sotto di 0,5 nodi, il confronto è effettuato rispetto al denominatore fissato a tale valore. Il criterio è rivolto ai casi di oscillazione della posizione e della velocità di unità lente in condizioni nelle quali la traccia AIS non descrive un movimento regolare.

Non viene applicato un filtro geometrico fisso di co-localizzazione, ad esempio una soglia di distanza mediana inferiore a 30 metri. Un criterio puramente spaziale potrebbe infatti scartare anche close encounter reali nei quali due unità transitano legittimamente a distanza ridotta. La classificazione dei casi limite è quindi affidata alle componenti cinematiche di velocità (SOG) e rotta (COG).

I quattro criteri isolano il dataset degli *encounter cinematici*. La distribuzione quantitativa delle categorie di scarto e della selezione finale è riportata in Tab. 3.2.

Tabella 3.2: Distribuzione degli encounter del dataset clean nelle quattro categorie di scarto cinematico e nella selezione finale del dataset cinematico (Tigullio).

Categoria di scarto / Selezione	N. encounter	% sul clean
Artefatto AIS ($SOG > 60$ kn)	0	0,0 %
Affiancamento (rotte parallele, velocità simili)	251	2,5 %
Stazionamento (entrambe le unità ferme)	304	3,1 %
Caotico (navigazione erratica sotto 1,5 kn)	816	8,2 %
Dataset cinematico (selezionato)	8 573	86,2 %
<i>Totale clean dataset (input filtro)</i>	<i>9 944</i>	<i>100,0 %</i>

Il filtraggio evita che encounter caratterizzati da stazionamento o da anomalie cinematiche contribuiscano agli indicatori di criticità. Un'unità ferma in prossimità di un'altra può infatti produrre valori elevati di $1/d_{\min}$ senza rappresentare un'interazione dinamica; analogamente, anomalie nella velocità relativa possono compromettere il calcolo di DCPA e TCPA.

3.4.1 Stabilità del dataset cinematico rispetto alle soglie del filtro

Poiché due soglie del filtro, quella di affiancamento sulla differenza di SOG e quella di stazionamento sulla SOG mediana, sono fissate su base operativa, è stata verificata la sensibilità del dataset cinematico al loro valore. Le due soglie sono state perturbate del $\pm 10\%$ e del $\pm 20\%$ rispetto ai valori nominali, misurando quanti encounter migrano tra le categorie cinematica e non cinematica.

La migrazione resta contenuta in ogni configurazione (Fig. 3.4): l'1,63 % degli encounter nel caso peggiore sulla griglia congiunta delle due perturbazioni, con il numero di cinematici che oscilla tra 8 457 e 8 713 attorno al valore nominale di 8 573. La variazione contenuta del numero di encounter cinematici indica una sensibilità limitata del dataset rispetto alle soglie considerate.

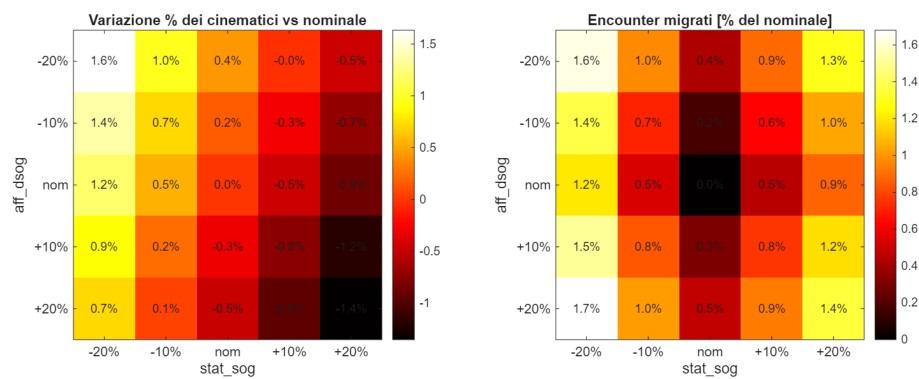


Figura 3.4: Analisi di sensibilità del dataset cinematico alle soglie del filtro. La variazione percentuale dei cinematici e la quota di encounter migrati restano sotto il 2 % per perturbazioni fino al $\pm 20\%$.

3.5 Analisi preliminare

Prima del calcolo degli indicatori di criticità, il dataset cinematico è stato sottoposto a un'analisi esplorativa delle principali caratteristiche del traffico osservato e degli encounter.

3.5.1 Statistiche di sintesi e distribuzione spaziale

Le caratteristiche aggregate del sottoinsieme degli encounter cinematici sono riportate in Tab. 3.3. La distribuzione spaziale dopo il filtro cinematico, mostrata in Fig. 3.5, conserva i principali corridoi di traffico e riduce la densità delle aree dominate da artefatti di stazionamento; in particolare l'effetto è massimo all'imboccatura del porto di Portofino: la cella più densa passa da 1 472 encounter nel dataset complessivo a 917 in quello cinematico.

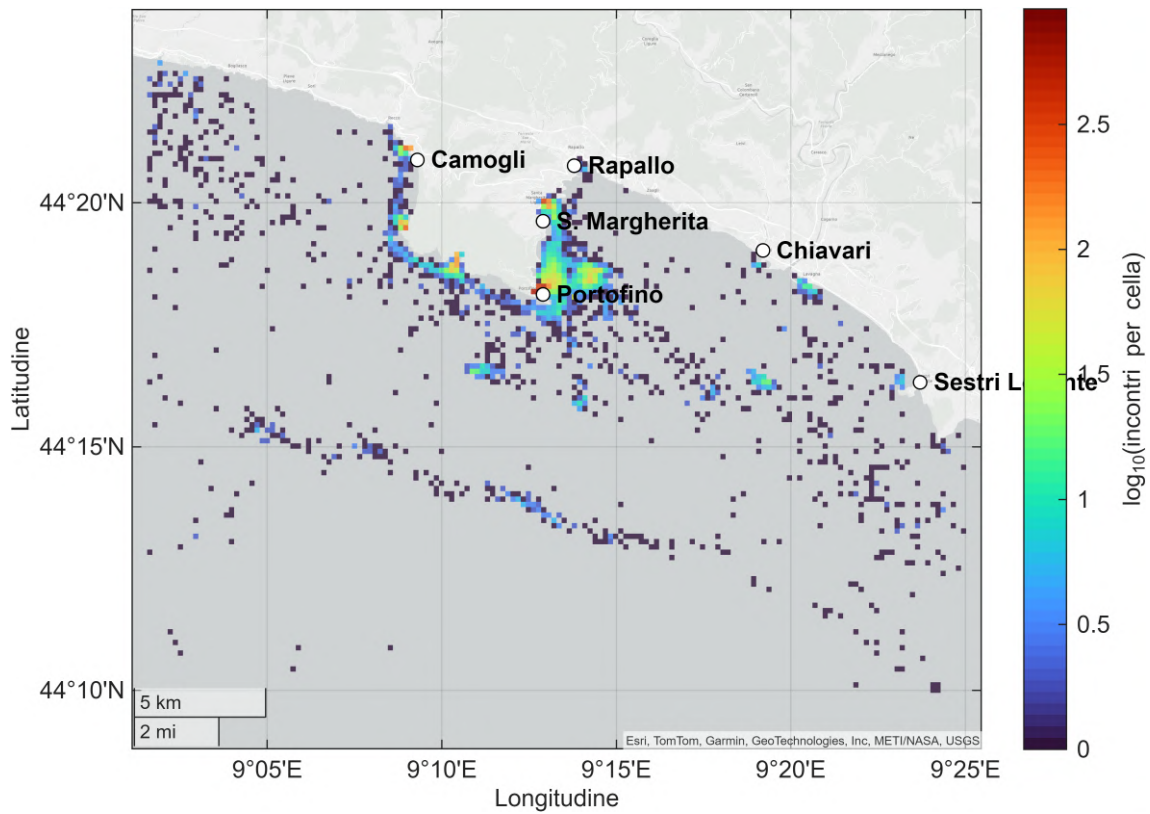


Figura 3.5: Mappa di densità degli encounter cinematici ($N = 8573$) su base cartografica, in scala logaritmica. Rispetto alla densità del dataset complessivo, l'hotspot di Portofino risulta attenuato, mentre i corridoi di transito e gli incroci di rotta restano visibili. La densità è calcolata su una griglia di 160×120 celle (longitudine $\times$ latitudine), di lato pari a circa 200 m (202 m in latitudine, 195 m in longitudine).

La vista di dettaglio del promontorio di Portofino ($44,29\text{--}44,37^\circ$ N; $9,13\text{--}9,26^\circ$ E, celle di circa 100 m) contiene 7 359 degli 8 573 encounter cinematici (85,8 %); la cella più densa si trova all'imboccatura del porto di Portofino (Fig. 3.6).

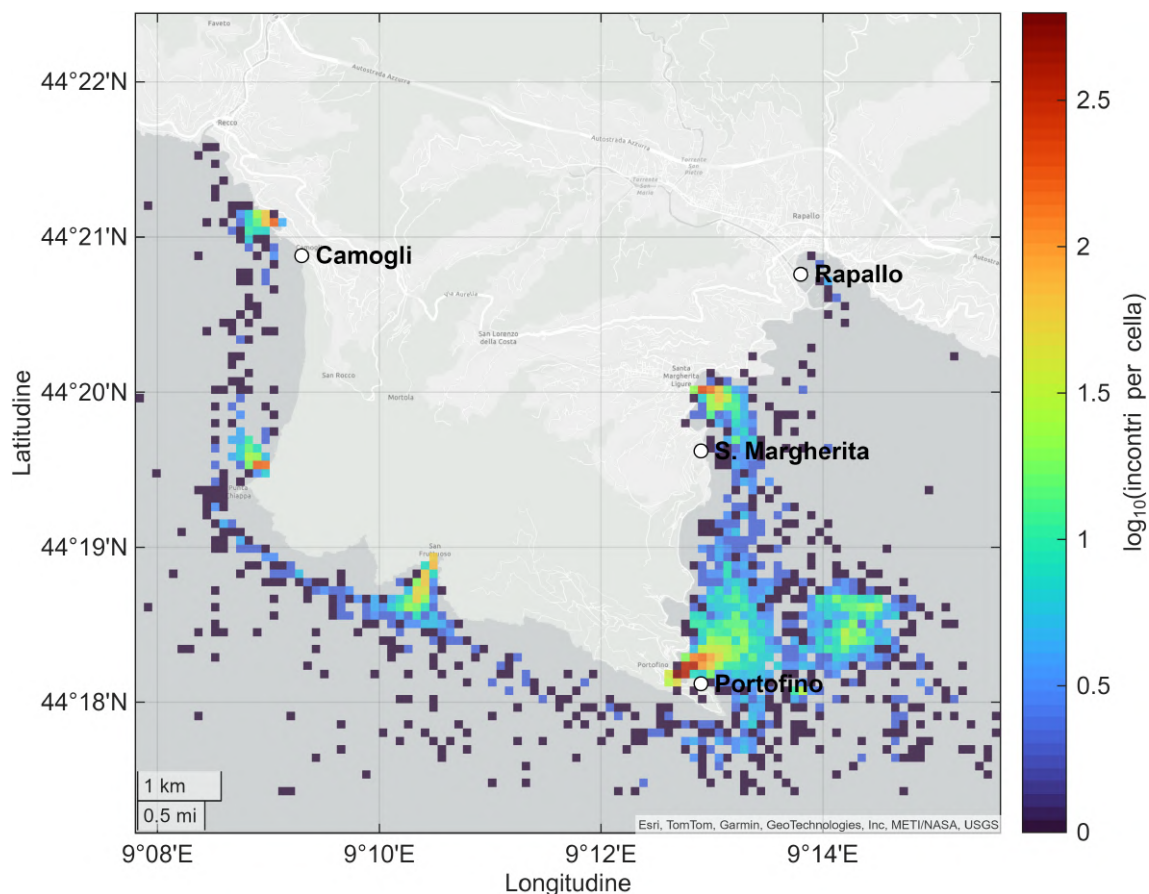


Figura 3.6: Densità degli encounter cinematici nella vista di dettaglio del promontorio di Portofino (44,29–44,37° N; 9,13–9,26° E), su celle di circa 100 m, in scala logaritmica.

Tabella 3.3: Statistiche di sintesi del dataset di encounter cinematici (Tigullio).

Grandezza	Valore
N. encounter cinematici	8 573
Finestra temporale	2 gen 2023 – 30 dic 2023
MMSI distinti coinvolti	1 053

3.5.2 Composizione del traffico per tipologia di unità

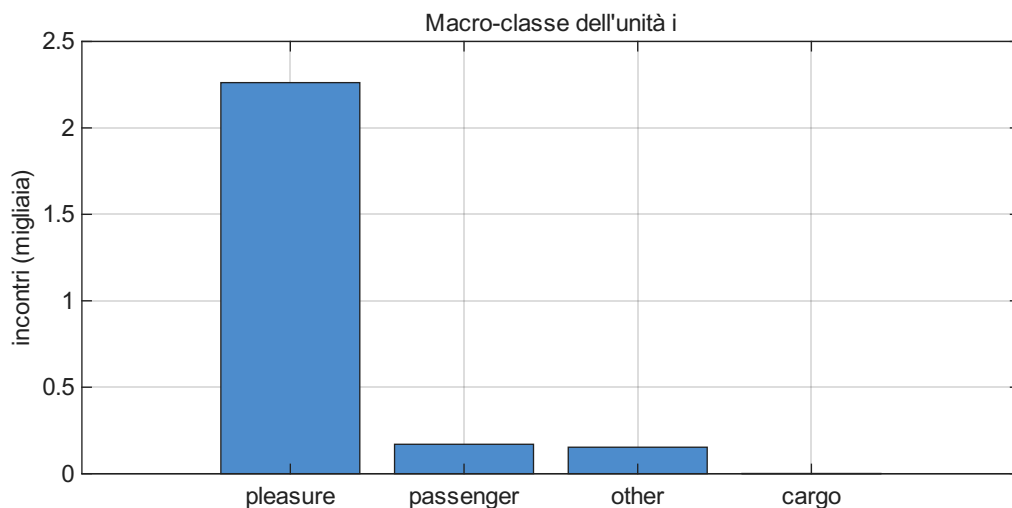
Una caratterizzazione complementare riguarda la composizione del dataset cinematico per *tipologia di unità coinvolta*. Tale informazione, non presente direttamente nel dataset *encounter-level* ma recuperabile tramite *enrichment* sull'archivio annuale dei dati statici (cfr. Sezione 3.6.4), viene utilizzata per interpretare le analisi successive e per confrontare la composizione del traffico nelle due aree di studio.

La procedura di *enrichment* ha associato a ciascun MMSI coinvolto in encounter cinematici il rispettivo *ship cargo type* secondo la classificazione ITU-R M.1371 (International Telecommunication Union, 2014). Aggregando i 56 codici originari in cinque macro-classi (*cargo* per i codici

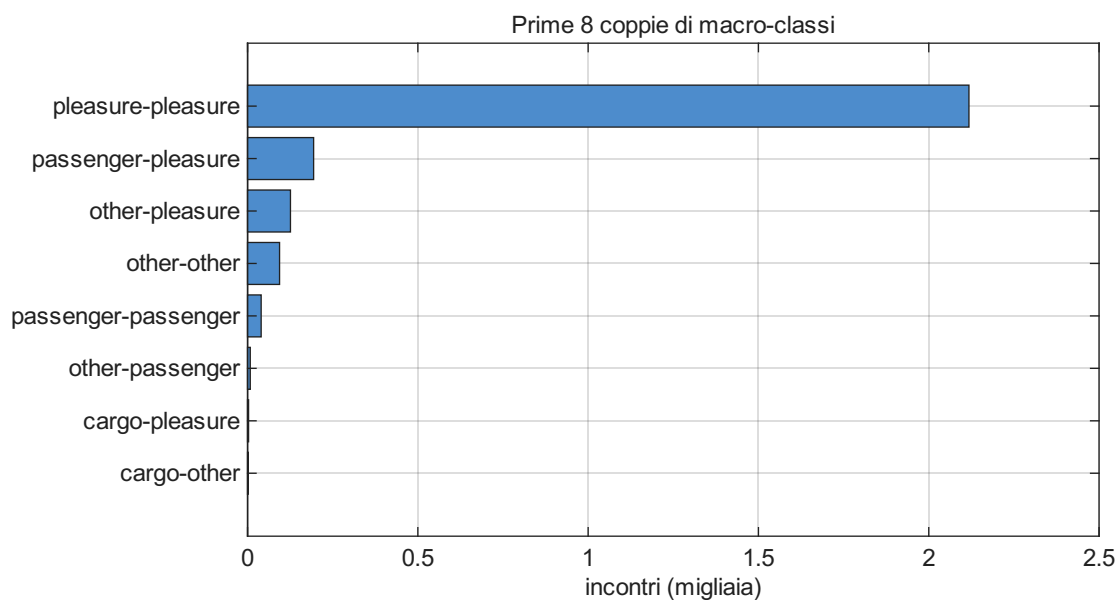
70–79; *tanker* per 80–89; *passenger* per 60–69; *pleasure/sailing* per i codici 36 e 37; *other* per le tipologie residue), è stata ricostruita la matrice delle coppie. La distribuzione, riportata in Tab. 3.4 e illustrata in Fig. 3.7, mostra una marcata presenza del traffico da diporto.

Tabella 3.4: Distribuzione degli encounter cinematici per coppia di macro-classi (soli encounter con identificazione completa di entrambe le navi, “full match”, cfr. Sezione 3.6.4).

Coppia di macro-classi	N. encounter	% del full match
pleasure–pleasure	2 118	82,3 %
passenger–pleasure	194	7,5 %
other–pleasure	126	4,9 %
other–other	94	3,7 %
passenger–passenger	40	1,6 %
altre coppie (cargo, tanker)	0	0,0 %
Totale full match	2 572	100 %



(a) Macro-classe dell'unità i .



(b) Prime otto coppie di macro-classi, sugli encounter con entrambe le unità identificate (2 572 su 8 573).

Figura 3.7: Composizione del dataset cinematico per macro-classe delle unità coinvolte (Tigullio).

Oltre l'82 % degli encounter cinematici con identificazione completa coinvolge due unità da diporto; considerando le coppie miste con almeno un'unità da diporto si raggiunge il 94,8 % del totale. Le unità cargo e tanker non compaiono nel sottoinsieme in full match. Il traffico passeggeri puro (passenger-passenger) rappresenta l'1,6 % del totale.

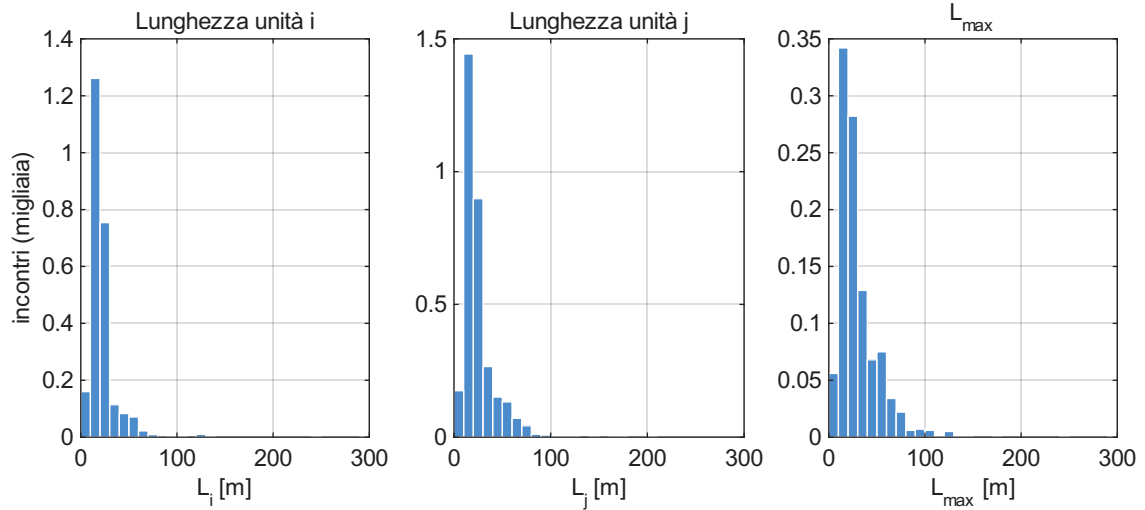


Figura 3.8: Distribuzione delle lunghezze delle unità coinvolte negli encounter cinematici: unità i , unità j e $L_{\max}$, la maggiore delle due lunghezze, calcolata sugli encounter con entrambe le lunghezze note. La mediana di $L_{\max}$ è pari a 23 m, coerente con la dominanza del naviglio da diporto.

Questa composizione deve essere considerata nella scelta degli indicatori di criticità adottati nel Capitolo 4. In particolare, gli indici compositi della letteratura (fra cui l'indice CR di Rong et al. (2019), applicato alle near collision rilevate con lo *ship domain* di Fujii) sono stati sviluppati e calibrati in contesti a traffico commerciale, nei quali la lunghezza tipica delle unità è dell'ordine delle centinaia di metri. La loro applicazione a un dataset dominato da unità con lunghezze dell'ordine delle decine di metri richiede pertanto una verifica dei parametri di taratura. Nel presente lavoro questi indici non sono applicati con i coefficienti di letteratura; gli indicatori adottati sono descritti nel Capitolo 4.

3.6 Limiti e trasparenza metodologica

Il dataset così costruito presenta alcuni limiti che devono essere considerati nell'interpretazione dei risultati. I principali riguardano la diversa ricchezza temporale degli encounter, la presenza di anomalie nei campi cinematici, la copertura dei dati statici e la risoluzione della ricostruzione delle tracce.

3.6.1 Eterogeneità nella ricchezza temporale degli encounter

La distribuzione del numero di record per encounter è asimmetrica. Gli encounter rappresentati da un solo record sincrono sono 7743, il 32,6% del totale. Per tali eventi non è possibile ricostruire una traiettoria di evoluzione temporale, il che limita la calcolabilità degli indicatori cinematici introdotti nei capitoli successivi, tra cui DCPA e TCPA istantanei e gli indici compositi sull'evoluzione.

Nelle analisi che richiedono una storia temporale è stato pertanto introdotto un filtro minimo di tre record per encounter. Il minimo di tre record è una scelta operativa del presente lavoro, motivata

dalla necessità di disporre di almeno un intervallo centrale per la stima a differenze finite del tasso di avvicinamento.

3.6.2 Anomalie nei campi cinematici

Una piccola frazione di record raw presenta valori di SOG superiori a 60 nodi, fra cui il valore 102,3, che nella codifica AIS indica una velocità non disponibile (International Telecommunication Union, 2014). Gli encounter con tali anomalie sono stati segnalati con un apposito flag, in modo da poter essere esclusi dalle analisi cinematiche senza modificare il dataset utilizzato per le analisi descrittive.

3.6.3 Sintesi dei sottoinsiemi

I diversi passaggi di filtraggio descritti nelle sezioni precedenti definiscono una gerarchia di sottoinsiemi del dataset, la cui composizione numerica è riassunta in Tab. 3.5.

Tabella 3.5: Composizione dei sottoinsiemi del dataset definiti dai successivi passaggi di filtraggio (Tigulio).

Sottoinsieme	Definizione	N. encounter
Dataset complessivo	encounter con $dist \leq 500$ m	23 769
Encounter clean	≥ 3 record (≥ 3 con SOG e COG validi), no SOG anomale, durata ≥ 60 s	9 944
Encounter cinematici	clean $\cap$ no artefatti cinematici (Sezione 3.4)	8 573

Le analisi descritte nei capitoli successivi specificano di volta in volta su quale sottoinsieme sono condotte. La caratterizzazione esplorativa più ampia utilizza l'intero dataset complessivo e il sottoinsieme clean, mentre le analisi cinematiche, inclusi DCPA e TCPA lungo l'evoluzione, gli indici compositi e le mappe di criticità pesate, utilizzano esclusivamente il sottoinsieme degli encounter cinematici.

3.6.4 Enrichment dei dati statici e copertura del join

Il dataset *encounter-level* non include i campi statici delle unità coinvolte: tipo di nave, lunghezza, larghezza e nominativo radio. Tali informazioni sono disponibili nell'archivio annuale consolidato descritto nella Sezione 3.2.1, dove ciascuna unità è caratterizzata dal proprio *ship cargo type* e dalle dimensioni geometriche derivate dai campi *ref_pos*. Il recupero dei dati statici è stato effettuato tramite *join* sull'identificativo MMSI fra il dataset di encounter e l'archivio annuale, che contiene 3 006 unità navali distinte con i propri metadati.

La copertura del *join* è stata valutata su due livelli. A livello di MMSI distinti, la maggior parte degli identificativi coinvolti negli encounter cinematici è stata rintracciata nell'archivio annuale. A livello di encounter la copertura è inferiore, perché è sufficiente che una delle due unità non sia

identificata per impedire il *full match*: gli encounter con tipologia nota per entrambe le unità sono 2 572; i restanti ricadono in *partial match* o *no match*.

La divergenza fra la copertura a livello di MMSI e quella a livello di encounter è riconducibile a un numero esiguo di unità non identificate ma coinvolte in un numero sproporzionato di encounter. Non viene utilizzata un'ulteriore attribuzione della tipologia o delle dimensioni per le unità prive di corrispondenza nell'archivio statico.

Il campo *length* entra soltanto nell'indicatore R_8 (Sezione 4.3.3); gli altri indicatori, compreso R_7 , non usano le dimensioni delle unità. Le lunghezze mancanti non sono stimate da altre informazioni. La disaggregazione per macro-classe opera sui 2 572 encounter in *full match*; la distribuzione di $L_{\max}$, la maggiore delle due lunghezze, opera sui 1 039 encounter in cui entrambe le lunghezze sono note, pari al 12,1 % degli encounter cinematici (Fig. 3.8). R_8 è calcolato su tutti gli encounter cinematici: una lunghezza non nota è sostituita dalla lunghezza di riferimento dell'indicatore.

3.6.5 Risoluzione del dato e ricostruzione del singolo incontro

La ricostruzione geometrica del singolo incontro è limitata dalla risoluzione del dato AIS. Nel dataset complessivo il numero mediano di record per encounter è pari a due, e solo il 15,5 % degli incontri ne raggiunge 8. Nel Tigullio solo il 17,5 % dei record sincroni corrisponde a messaggi effettivamente trasmessi, riconoscibili dalla risoluzione di un decimo prevista dallo standard per SOG e COG (International Telecommunication Union, 2014); i restanti derivano dall'interpolazione a passo di un minuto. Il confronto fra la velocità ricavata dallo spostamento delle posizioni e quella dichiarata nello stesso messaggio mostra coerenza per le unità sopra i 5 nodi, mentre al di sotto di tale soglia il rapporto raggiunge valori superiori a 6 nel 10 % dei casi. A bassa velocità lo spostamento fra record consecutivi è confrontabile con l'incertezza di posizione, per la quale lo standard AIS distingue un'accuratezza alta, non superiore a 10 m, da una bassa (International Telecommunication Union, 2014).

Poiché il dataset è costituito da incontri entro 500 m e in larga parte fra unità da diporto dotate di transponder di Classe B, l'incertezza della ricostruzione può essere dello stesso ordine della separazione fra le unità. Per questa ragione le analisi operano sull'insieme degli encounter cinematici; i singoli incontri sono ricostruiti solo nei casi illustrativi della verifica COLREG, sulle posizioni trasmesse dalle unità in Classe A dello Stretto di Messina (Sezione 5.8.8).

3.7 Estensione a un secondo caso di studio: lo Stretto di Messina

Il secondo caso di studio è costituito dagli archivi AIS 2022, 2023 e 2024 relativi allo Stretto di Messina. I tre anni sono elaborati separatamente, con la stessa procedura, per verificare che la metodologia dia risultati coerenti su dataset diversi. Lo Stretto presenta condizioni di traffico differenti da quelle del Tigullio: l'area è attraversata da un corridoio di transito Nord-Sud ad alta intensità e da un servizio di traghetti con traiettorie trasversali. Il confronto fra i due domini

consente quindi di applicare la pipeline a una configurazione nella quale le rotte e le tipologie di incontro presentano una distribuzione diversa rispetto all'area ligure.

La navigazione nello Stretto è regolata da uno schema di separazione del traffico di tipo rotatorio (Italian Coast Guard Headquarters, 2015), articolato in tre settori (Nord, Centro, Sud) con corsie a senso unico nei tratti settentrionale e meridionale e una rotonda nel settore centrale, tra Messina e Villa San Giovanni (Fig. 3.10). Questo sistema contribuisce, insieme alla morfologia del passaggio e alla posizione dei porti, all'organizzazione dei flussi osservati. Nel presente lavoro è considerato solo come elemento di contesto: le traiettorie AIS non sono state confrontate con la geometria dello schema.

3.7.1 Fonte e pre-processing

Per ciascun anno la sorgente è costituita da 24 file CSV mensili di messaggi AIS dell'area dello Stretto, separati per transponder di Classe A e di Classe B. Sono state scartate le righe con istante non standard, cioè con frazioni di secondo o con formato diverso da `aaaa-mm-gg hh:mm:ss`, e le righe prive di posizione o di MMSI.

Gli encounter sono estratti con la logica descritta nella Sezione 3.2.2. Le tracce sono ricampionate a passo di un minuto per interpolazione lineare. Pulizia, ricostruzione dell'evoluzione e filtro cinematico sono eseguiti con lo stesso codice MATLAB applicato al Tigullio e con gli stessi valori di soglia. I tre criteri di esclusione della pulizia non sono mutuamente esclusivi: un encounter può soddisfarne più d'uno. L'area di analisi, comune ai tre anni, è mostrata in Fig. 3.9.

Tabella 3.6: Dai messaggi AIS grezzi agli encounter clean (Stretto di Messina).

	2022	2023	2024
Righe dei CSV	10 657 812	23 102 794	31 519 093
Scartate: istante non standard	33 019 (0,31 %)	48 532 (0,21 %)	43 327 (0,14 %)
Scartate: posizione o MMSI mancanti	25	16	85
Encounter estratti	569 001	591 638	553 616
Esclusi: meno di 3 record	141 983	154 298	148 396
Esclusi: durata inferiore a 60 s	69 987	77 734	79 934
Esclusi: SOG superiore a 60 kn	2 458	4 325	8 833
Esclusi: meno di 3 record con SOG e COG validi	11	0	0
Encounter clean	424 758	433 358	397 005

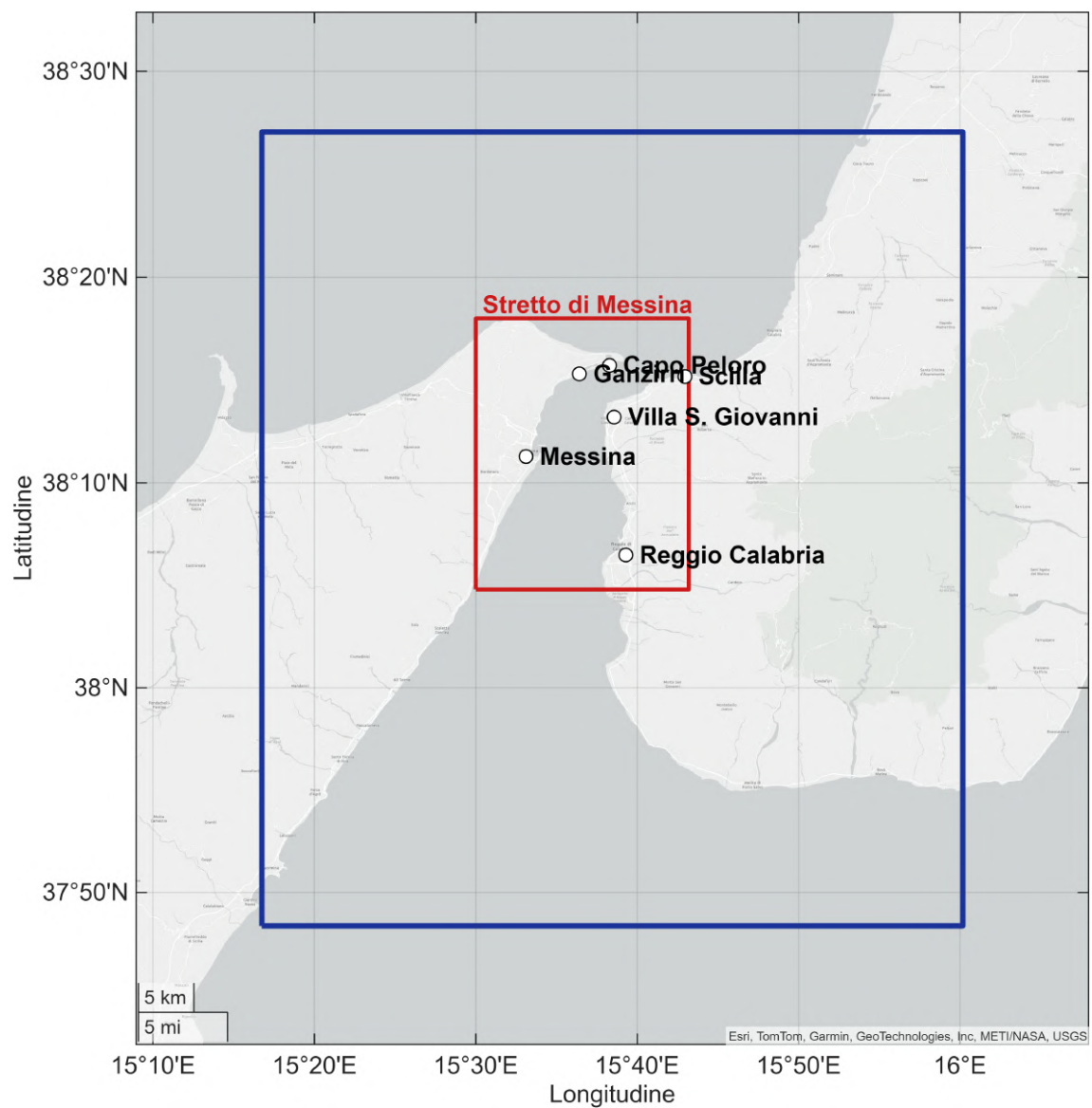


Figura 3.9: Area di studio dello Stretto di Messina, comune ai tre anni analizzati, con il riquadro della vista di dettaglio dello Stretto (38,08–38,30° N; 15,50–15,72° E).

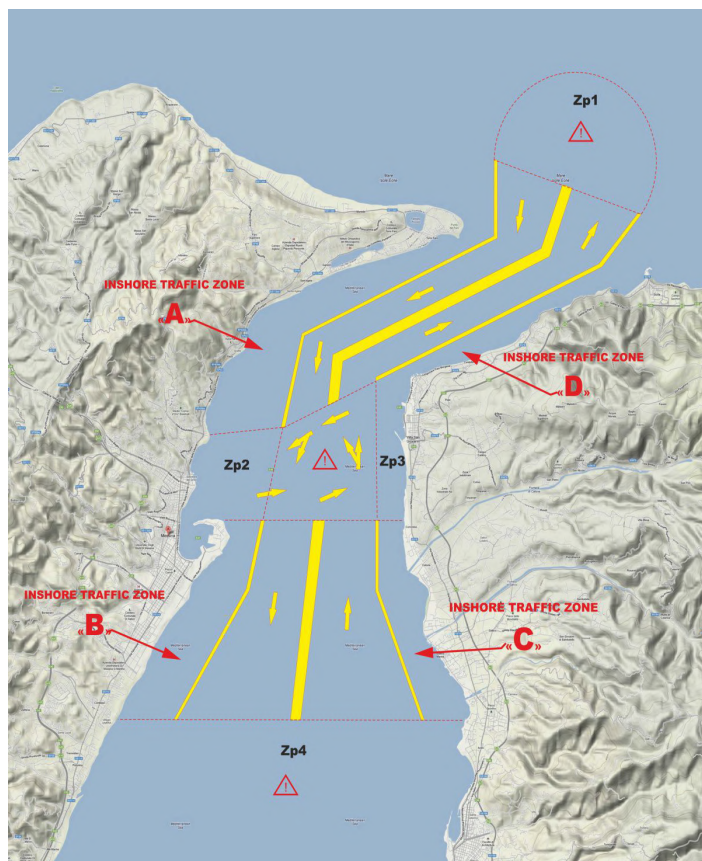


Figura 3.10: Schema di separazione del traffico dello Stretto di Messina (settori Nord, Centro e Sud), con zone di precauzione Zp1–Zp4 e zone di traffico costiero A–D. Tratta da Italian Coast Guard Headquarters (2015), pag. MES 04.

3.7.2 Dimensione e raffinamento cinematico

La maggiore intensità di traffico dello Stretto si riflette in un numero di encounter cinematici da 8 a 10 volte superiore a quello del Tigullio. Il filtro cinematico, riassunto in Tab. 3.7, isola 85 018 encounter cinematici nel 2022, 83 494 nel 2023 e 69 241 nel 2024.

Rispetto al Tigullio lo stazionamento è la categoria più numerosa: 66,6, 67,4 e 68,6 % del clean nei tre anni, contro il 3,1 %. La frazione cinematica è pari al 20,0, 19,3 e 17,4 % del clean. La quota di stazionamento è compatibile con la presenza di aree di attesa e ancoraggio in prossimità dell'imbocco dello Stretto. Anche la composizione degli encounter cinematici è diversa: quelli in cui una sola unità ha velocità mediana inferiore a 1,0 nodo sono l'84,8, l'82,9 e l'81,6 % nei tre anni, contro il 10,7 % del Tigullio.

Tabella 3.7: Distribuzione degli encounter del dataset clean nelle categorie di scarto cinematico e nella selezione finale (Stretto di Messina). Fra parentesi la quota sul clean.

Categoria di scarto / Selezione	2022	2023	2024
Artefatto AIS (<i>SOG</i> > 60 kn)	0 (0,0 %)	0 (0,0 %)	0 (0,0 %)
Affiancamento	44 274 (10,4 %)	45 657 (10,5 %)	44 506 (11,2 %)
Stazionamento	282 717 (66,6 %)	292 082 (67,4 %)	272 160 (68,6 %)
Caotico	12 749 (3,0 %)	12 125 (2,8 %)	11 098 (2,8 %)
Dataset cinematico	85 018 (20,0 %)	83 494 (19,3 %)	69 241 (17,4 %)
<i>Totale clean (input filtro)</i>	<i>424 758</i>	<i>433 358</i>	<i>397 005</i>

La distribuzione spaziale degli encounter cinematici del 2023 è riportata in Fig. 3.11. Tutti gli 83 494 encounter ricadono nella griglia e occupano 2 244 celle. La densità si concentra nel tratto settentrionale dello Stretto, fra Messina, Ganzirri e Villa San Giovanni, e lungo la costa calabra presso Reggio Calabria. La cella più densa, nella rada di Messina, contiene 6 174 encounter. Nuclei minori compaiono ai margini dell'area, a Giardini Naxos e sulla costa tirrenica calabra; nel tratto meridionale dello Stretto gli encounter sono sparsi. Le celle occupate sono 1 996 nel 2022 e 2 277 nel 2024, con un massimo per cella di 11 586 e 4 647 encounter; le mappe dei due anni sono riportate in Appendice A.1.

La vista di dettaglio dello Stretto (38,08–38,30° N; 15,50–15,72° E, celle di circa 100 m; Fig. 3.12) contiene 59 994 encounter del 2023 (71,9 %), con un massimo di 3 709 per cella; nel 2022 e nel 2024 vi ricadono 70 506 (82,9 %) e 48 929 (70,7 %) encounter.

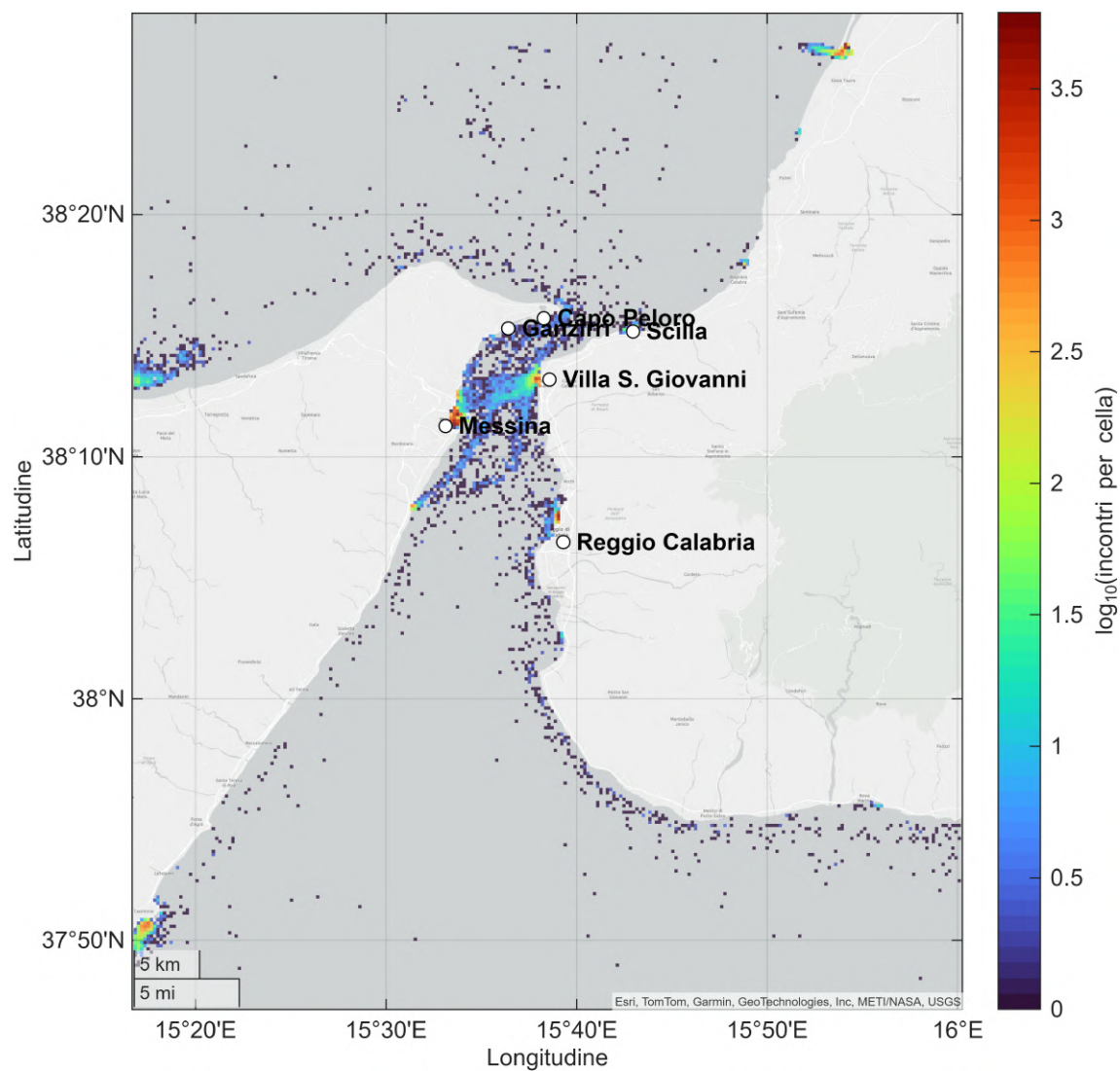


Figura 3.11: Mappa di densità degli encounter cinematici dello Stretto di Messina nel 2023 ($N = 83\,494$) su base cartografica, in scala logaritmica. La densità è calcolata su una griglia di 254×287 celle (longitudine $\times$ latitudine), di lato pari a circa 250 m (249 m in latitudine, 250 m in longitudine).

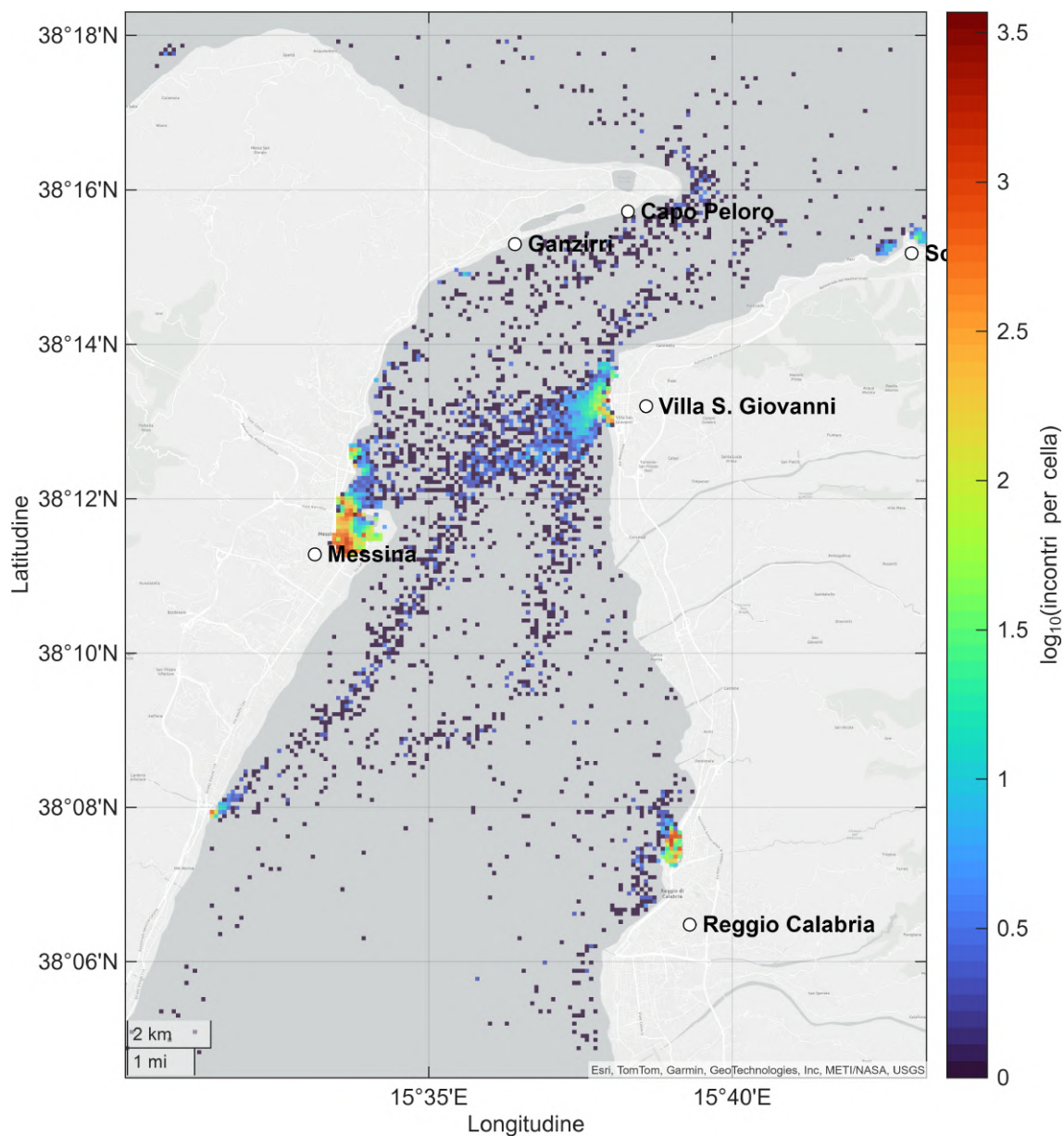


Figura 3.12: Densità degli encounter cinematici del 2023 nella vista di dettaglio dello Stretto (38,08–38,30° N; 15,50–15,72° E), su celle di circa 100 m, in scala logaritmica.

3.7.3 Composizione del traffico e copertura statica

L'archivio statico di ciascun anno è ricavato dagli stessi CSV. Per tipo, lunghezza e larghezza di ogni MMSI si adotta il valore positivo dichiarato nel maggior numero di messaggi dell'anno, così che un singolo messaggio errato non modifichi il valore attribuito all'unità. Le unità con più di una lunghezza dichiarata nell'anno sono 125 nel 2022, 105 nel 2023 e 99 nel 2024; per 34, 28 e 38 di esse il valore massimo dichiarato supera di oltre il 10 % quello adottato. La consistenza dell'archivio e la disponibilità delle lunghezze sono riportate in Tab. 3.8.

Tabella 3.8: Archivio statico e disponibilità delle lunghezze (Stretto di Messina).

	2022	2023	2024
MMSI nell'archivio statico	7 178	7 765	7 824
MMSI con lunghezza valida	92,8 %	92,3 %	92,2 %
Encounter cinematici con entrambe le lunghezze note	89,1 %	93,2 %	95,2 %

La composizione per macro-classe, illustrata in Fig. 3.14 e riportata in Tab. 3.9, mostra un peso maggiore delle unità *passenger* e *cargo* rispetto al Tigullio. Sulle coppie con entrambe le lunghezze dichiarate (77 812 nel 2023) la distribuzione di $L_{\max}$ è multimodale, con picchi fra 40 e 60 m, fra 90 e 120 m e fra 140 e 150 m, la classe più numerosa (Fig. 3.13). I picchi sono compatibili con le lunghezze delle unità di linea che operano nello Stretto. Le coppie con $L_{\max}$ superiore a 300 m, fuori dall'asse della figura, sono 3 621 nel 2022, 6 334 nel 2023 e 6 863 nel 2024. Le figure del 2022 e del 2024 sono riportate in Appendice A.1.

Tabella 3.9: Distribuzione degli encounter cinematici per coppia di macro-classi (Stretto di Messina). Fra parentesi la quota sul totale. A differenza del Tigullio, nello Stretto tutte le unità coinvolte sono presenti nell'archivio statico in tutti e tre gli anni: la categoria *unknown-involved* è vuota.

Coppia di macro-classi	2022	2023	2024
other-passenger	30 236 (35,6 %)	22 629 (27,1 %)	15 406 (22,2 %)
other-other	15 474 (18,2 %)	14 701 (17,6 %)	11 501 (16,6 %)
passenger-passenger	16 501 (19,4 %)	13 329 (16,0 %)	11 201 (16,2 %)
cargo-other	6 544 (7,7 %)	9 551 (11,4 %)	9 703 (14,0 %)
pleasure-pleasure	4 847 (5,7 %)	9 038 (10,8 %)	8 295 (12,0 %)
passenger-pleasure	5 257 (6,2 %)	5 856 (7,0 %)	4 918 (7,1 %)
other-pleasure	2 536 (3,0 %)	3 798 (4,5 %)	4 382 (6,3 %)
cargo-cargo	1 187 (1,4 %)	1 806 (2,2 %)	1 842 (2,7 %)
cargo-passenger	1 395 (1,6 %)	1 688 (2,0 %)	1 031 (1,5 %)
other-tanker	655 (0,8 %)	646 (0,8 %)	545 (0,8 %)
altre coppie	386 (0,5 %)	452 (0,5 %)	417 (0,6 %)
Totale	85 018	83 494	69 241

Le macro-classi più rappresentate sono *other* e *passenger*, in relazione alla presenza del servizio di traghetti dell'area.

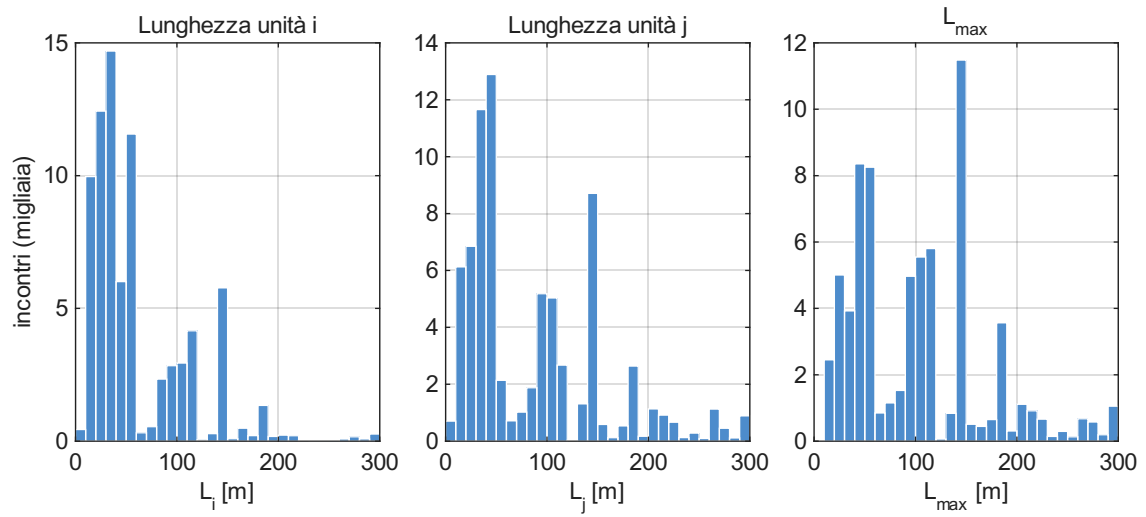
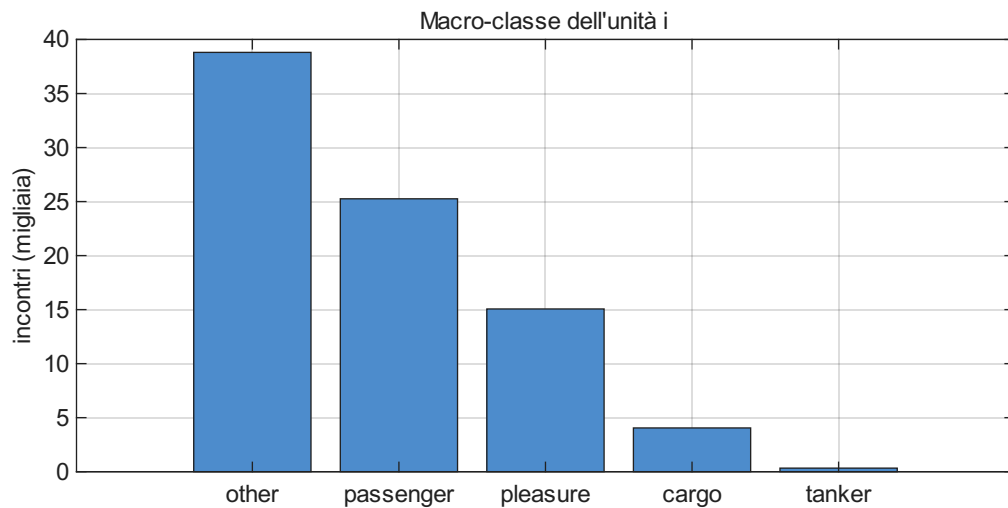
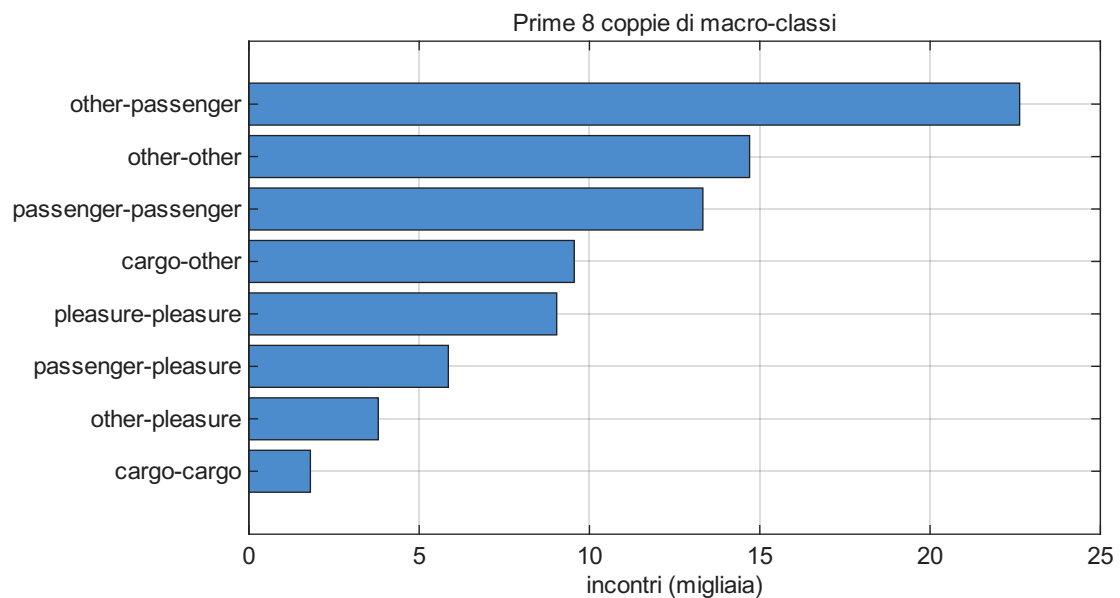


Figura 3.13: Distribuzione delle lunghezze delle unità coinvolte negli encounter cinematici dello Stretto di Messina nel 2023, calcolata per encounter sulle sole lunghezze dichiarate: unità i , unità j e L_{max} , la maggiore delle due lunghezze. Asse limitato a 300 m.



(a) Macro-classe dell'unità i .



(b) Prime otto coppie di macro-classi.

Figura 3.14: Composizione del dataset cinematico per macro-classe delle unità coinvolte (Stretto di Messina, 2023).

La differenza tra i due dataset riguarda sia la composizione del traffico sia la sua distribuzione spaziale: nel Tigullio prevale un traffico costiero diffuso, con forte componente da diporto, mentre nello Stretto gli encounter sono maggiormente associati a flussi di traffico organizzati e alla presenza di traffico passeggeri e di transito. Questa differenza costituisce la base del confronto fra le due aree sviluppato nel Capitolo 5.

Capitolo 4

Metodologia

Il presente capitolo descrive la metodologia adottata per analizzare i dataset introdotti nel Capitolo 3 e per produrre, a partire da essi, indicatori e mappe di criticità. La metodologia considera progressivamente indicatori di diversa complessità, dagli indicatori elementari di prossimità alle formulazioni cinematiche e composite. I risultati ottenuti con i diversi indicatori vengono quindi confrontati sullo stesso dataset, così da evidenziare le differenze nella classificazione degli encounter prodotte dalle diverse formulazioni.

Il capitolo si articola come segue. La Sezione 4.1 espone la sequenza complessiva degli indicatori. La Sezione 4.2 descrive gli indicatori elementari di prossimità. La Sezione 4.3 introduce gli indicatori cinematici, distinguendo fra grandezze ricavate dalla distanza osservata, grandezze ricavate da DCPA e TCPA stimati e indicatori velocity-aware. La Sezione 4.4 sviluppa la procedura di confronto fra indicatori (correlazione di Spearman e concordanza top-N). Le Sezioni 4.5–4.9 descrivono le analisi complementari: classificazione per tipo di incontro, autocorrelazione spaziale, profilo temporale e verifica di conformità al COLREG. La metodologia è la stessa per i due casi di studio, il Tigullio e lo Stretto di Messina, e per quest’ultimo ai tre anni 2022, 2023 e 2024; differisce soltanto l’indicatore principale delle analisi spaziali e temporali (Sezione 4.4). Le figure e le mappe che ne derivano sono raccolte nel Capitolo 5.

4.1 Struttura della pipeline

La pipeline si articola in quattro livelli. Il *primo livello* è costituito dagli indicatori elementari di prossimità, tra cui il più semplice è $1/d_{\min}$, dove $d_{\min}$ è la minima distanza osservata di ciascun encounter. Questo livello considera la sola prossimità e non la cinematica. Il *secondo livello* introduce due grandezze cinematiche: la velocità relativa all’istante di minima distanza osservata, cioè il modulo della velocità di una nave rispetto all’altra, e l’approach rate, cioè la componente della velocità relativa lungo la congiungente delle due unità, che quantifica la rapidità con cui la distanza si riduce. Il *terzo livello* è costituito dagli indicatori cinematici lungo l’evoluzione: per ogni record di ogni encounter cinematico, DCPA e TCPA sono ricalcolati come se quell’istante

fosse il presente, estrapolando i vettori velocità correnti sotto l'ipotesi di moto rettilineo uniforme istantaneo. Il *quarto livello* introduce indicatori cinematici velocity-aware, che integrano in un'unica formulazione la prossimità (DCPA stimata o distanza istantanea) e l'intensità della cinematica. Questo livello è motivato dalla composizione del traffico (cfr. Sezione 3.5.2): in un'area dominata dal naviglio da diporto a bassa velocità occorre distinguere gli encounter “vicini ma lenti” da quelli “distanti ma veloci”. Ogni livello aggiunge un elemento rispetto al precedente, e gli indicatori dei quattro livelli sono confrontati sullo stesso dataset (Sezione 4.4).

4.2 Indicatori elementari di prossimità

Il primo livello di analisi assume come unica grandezza di interesse la distanza minima $d_{\min}$ raggiunta dalle due unità nel corso dell'encounter, ricavata dalla traiettoria di record sincroni descritta nella Sezione 3.3. L'indicatore elementare è definito come:

$$R_1 = \frac{1}{d_{\min}} \quad (4.1)$$

inversamente proporzionale alla distanza minima: a distanze più piccole corrisponde un valore più alto dell'indicatore. Per stabilità numerica, $d_{\min}$ è limitata inferiormente a 1 m. R_1 assegna valori più alti alle coppie più vicine, ma non usa l'informazione cinematica: due encounter con la stessa $d_{\min}$ ricevono lo stesso punteggio anche se uno corrisponde a un incrocio di traghetti a 20 nodi e l'altro a un sorpasso lento in rada.

La formulazione di R_1 come inverso della distanza minima è la più elementare delle famiglie di indicatori AIS-based discusse nella Sezione 2.4.2. Formulazioni equivalenti o lievemente modificate sono adottate, fra gli altri, da Rong et al. (2019), il cui indice CR include esplicitamente un termine proporzionale a $1/D_R$, mentre la rassegna di Du et al. (2020) identifica la distanza relativa fra le variabili più ricorrenti negli indicatori AIS-based censiti; nel presente lavoro essa è mantenuta come riferimento di confronto rispetto ai livelli successivi.

4.3 Indicatori cinematici

Per superare i limiti di R_1 si introducono grandezze cinematiche derivate da SOG e COG delle due unità. Le grandezze della Sezione 4.3.1 derivano dalla distanza osservata; quelle della Sezione 4.3.2 da DCPA e TCPA stimati per ogni record dell'encounter (formulazioni *time-resolved*).

4.3.1 Velocità relativa e approach rate

All'istante di minima distanza osservata, la velocità relativa v_{rel} è definita come il modulo della differenza vettoriale delle velocità delle due unità in quell'istante:

$$v_{rel} = |\vec{v}_i - \vec{v}_j| \quad (4.2)$$

dove le velocità vettoriali sono ricostruite dalle coppie (SOG, COG) registrate in tale istante, scomponendo ciascuna nelle componenti Est e Nord secondo $\vec{v} = (\text{SOG} \cdot \sin \text{COG}, \text{SOG} \cdot \cos \text{COG})$. La seconda grandezza è il *max approach rate*: l'approach rate istantaneo è il tasso di variazione della distanza reciproca fra le due unità, da non confondere con la velocità SOG delle singole unità, ed è ottenuto come derivata negativa di tale distanza rispetto al tempo, a differenze finite centrate sui record interni dell'encounter:

$$a(t_k) = -\frac{d(t_{k+1}) - d(t_{k-1}))}{t_{k+1} - t_{k-1}} \quad (4.3)$$

con valore positivo associato a una fase di avvicinamento. L'indicatore costruito sull'approach rate è il massimo dei valori di $a(t)$ lungo la durata dell'encounter:

$$R_3 = \max_t a(t) \quad (4.4)$$

R_3 è espresso in m/s e misura la massima velocità con cui la distanza si riduce. L'uso dell'approach rate come descrittore cinematico di severità è consolidato nella letteratura AIS-based: la grandezza è adoperata, in forme analoghe, da Bakdi et al. (2020) all'interno del safety domain adattivo, ed è citata da Du et al. (2020) fra le variabili dinamiche più frequenti. L'adattamento qui adottato differisce per il calcolo a differenze finite centrate sui record sincroni, scelta motivata dalla griglia temporale a passo regolare di un minuto.

4.3.2 Indicatori basati su DCPA e TCPA stimati

La distanza osservata e le grandezze che ne derivano dipendono dalla discretizzazione temporale dei dati. Per superare questa limitazione si introduce una formulazione *time-resolved*, basata su DCPA e TCPA stimati: per ogni record t_k , DCPA e TCPA sono ricalcolati come se t_k fosse il presente, sotto l'ipotesi di moto rettilineo uniforme istantaneo:

$$\text{TCPA}(t) = -\frac{\vec{p}(t) \cdot \vec{v}(t)}{|\vec{v}(t)|^2} \quad (4.5)$$

$$\text{DCPA}(t) = |\vec{p}(t) + \text{TCPA}(t) \cdot \vec{v}(t)| \quad (4.6)$$

dove $\vec{p}(t)$ è il vettore posizione relativa e $\vec{v}(t)$ il vettore velocità relativa. Per ricondurre le curve a uno scalare si definisce il valore minimo della curva DCPA, minDCPA , e l'indicatore cinematico lungo l'evoluzione:

$$R_2 = \frac{1}{\text{minDCPA}} \quad (4.7)$$

con la stessa convenzione di stabilità numerica di R_1 (limite inferiore 1 m). Diversamente da R_1 , R_2 può essere più elevato del proprio omologo geometrico in encounter in cui la cinematica indicava un avvicinamento estremo, anche se la distanza minima osservata è stata moderata: situazione tipica di una manovra evasiva in prossimità del CPA stimato. La versione *time-resolved*, che ricostruisce la DCPA istante per istante, è alla base dei modelli *Enhanced VCRO* di Liu et al. (2022) e *Kinematics Feature-based VCRO* di Liu et al. (2023).

Una formulazione composita più articolata, anch'essa basata sulle curve *time-resolved*, è il *max risk geometrico*:

$$R_4(t) = \frac{1}{\max(\text{DCPA}(t), 1)} \cdot \frac{1}{1 + |\text{TCPA}(t)|/60 \text{ s}}, \quad R_4 = \max_t R_4(t) \quad (4.8)$$

che combina prossimità spaziale e prossimità temporale. Il fattore $1/(1 + |\text{TCPA}|/60 \text{ s})$ penalizza i CPA temporalmente distanti. R_4 si colloca nella famiglia degli indici compositi DCPA–TCPA, di cui l'indice di conflitto di Debnath and Chin (2016), basato su DCPA e TCPA, e il *Collision Risk* esponenziale di Rong et al. (2019) costituiscono i riferimenti canonici. La costante di tempo di 60 s è scelta come ordine di grandezza tipico delle fasi di approach-and-evade in acque costiere e non ha pretesa di taratura predittiva.

4.3.3 Indicatori velocity-aware

Gli indicatori dei primi tre livelli trattano la prossimità (R_1 , R_2) e la cinematica (R_3) come dimensioni separate; R_4 le combina in un prodotto geometrico, senza integrare direttamente la velocità relativa. In un contesto di traffico costiero misto questa separazione è limitante: due unità a 2 nodi con distanza minima osservata di 30 m non presentano la stessa criticità di due unità a 15 nodi con distanza minima osservata di 50 m, sebbene il primo encounter riceva un valore di R_1 più alto. Il quarto livello introduce quattro indicatori *velocity-aware*, valutati lungo l'evoluzione e ridotti a uno scalare col massimo. Per stabilità numerica, nei denominatori di R_5 e R_6 si introduce una distanza di smorzamento $d_0 = 10 \text{ m}$.

Il primo è il *Time-To-Collision istantaneo* ($1/\text{TTC}$):

$$R_5 = \max_t \frac{a(t)}{d(t) + d_0}, \quad a(t) > 0 \quad (4.9)$$

con unità $[1/\text{s}]$: è il tasso di chiusura riferito alla distanza corrente. Il vincolo $a(t) > 0$ esclude gli istanti di allontanamento; encounter senza fase di avvicinamento netto sono identificati con un flag ed esclusi dal ranking di R_5 . Il TTC, rapporto fra distanza istantanea e tasso di chiusura, è alla base del *Time-varying Collision Risk* di Chen et al. (2021), basato su velocity obstacle.

Il secondo è una *kinetic energy density proxy* (KEDP):

$$R_6 = \max_t \frac{v_{rel}(t)^2}{d(t) + d_0} \quad (4.10)$$

con unità [m/s²] e dimensionalità di un'accelerazione, proporzionale all'energia cinetica relativa per unità di distanza. La dipendenza quadratica da v_{rel} rende R_6 sensibile agli encounter ad alta velocità anche a distanze moderate. La logica di un descrittore proporzionale all'energia cinetica relativa è propria dei modelli probabilistici di tipo IWRAP (IALA, 2009) e della spatial risk function di Bakdi et al. (2020); la formulazione qui adottata applica il principio lungo l'evoluzione dell'encounter e non dipende dalla massa delle unità.

Il terzo è il *Velocity-weighted CPA* (V-CPA):

$$R_7 = \max_t \frac{v_{rel}(t)}{DCPA(t) + d_c} \quad (4.11)$$

formalmente analogo a R_5 , con la velocità relativa al posto dell'approach rate e la DCPA stimata al posto della distanza osservata. R_7 ha unità [1/s] e misura la rapidità con cui le due unità si avvicinerebbero al punto di minimo avvicinamento stimato, rispetto alla prossimità che vi raggiungerebbero. Il termine $d_c = 25$ m impedisce la divergenza dell'indicatore per valori di DCPA stimata prossimi a zero: per $DCPA(t) \rightarrow 0$ il rapporto tende a $v_{rel}(t)/d_c$. Il valore di d_c è una scelta operativa del presente lavoro. R_7 dipende dalla sola cinematica della coppia ed è calcolato in tutte le aree. Indicatori che pongono il rapporto fra velocità relativa e DCPA come misura sintetica della criticità dell'incontro sono discussi nel filone VCRO: la famiglia introdotta da Zhang et al. (2015) contiene il termine $d_{ij}^{-1} \cdot v_{ij}$, preservato nelle evoluzioni di Liu et al. (2022, 2023).

Il quarto indicatore tiene conto delle dimensioni delle unità misurando la DCPA scalata rispetto alla lunghezza media della coppia, con L_{ref} come riferimento dimensionale:

$$R_8 = \max_t \frac{v_{rel}(t)}{d_c + DCPA(t) L_{ref}/L_p}, \quad L_p = \frac{L_i + L_j}{2} \quad (4.12)$$

con $d_c = 25$ m e $L_{ref} = 10$ m. A parità di velocità relativa e di DCPA stimata, R_8 cresce con la lunghezza media della coppia: una DCPA stimata di 100 m vale 10 lunghezze della coppia per due unità di 10 m e una sola per due unità di 100 m.. Le lunghezze L_i e L_j sono quelle dell'archivio statico (Sezioni 3.6.4 e 3.7.3); una lunghezza non nota o non positiva è posta pari a L_{ref} , unità per unità, senza altre imputazioni. Con entrambe le lunghezze non note, R_8 coincide con R_7 .

La lunghezza di entrambe le unità è nota per il 12,1 % degli encounter cinematici nel Tigullio (1 039 su 8 573) e per l'89,1, il 93,2 e il 95,2 % nello Stretto di Messina nel 2022, 2023 e 2024. Sulle coppie con entrambe le lunghezze note la mediana di L_p è 18,0 m nel Tigullio e 74,0, 70,5 e 74,0 m nello Stretto. Nello Stretto la correlazione di Spearman con L_p è 0,34, 0,40 e 0,46 per R_8 e $-0,12$, $-0,12$ e $-0,06$ per R_7 : R_8 attribuisce valori più alti agli incontri fra unità più lunghe, mentre R_7 non dipende dalle dimensioni. Fra il 25 % degli encounter del 2023 con R_8 più elevato

la lunghezza di entrambe le unità è nota per il 94,2 %, contro il 93,2 % sul totale, e la selezione non dipende quindi dalla disponibilità del dato statico. Nel Tigullio la stessa quota scende al 5,1 % fra il 25 % degli encounter con R_7 più elevato: gli incontri più critici coinvolgono soprattutto unità di lunghezza non nota, per le quali R_8 non aggiunge informazione.

R_8 è quindi adottato come indicatore principale nello Stretto di Messina; nel Tigullio l'indicatore principale è R_7 e R_8 è riportato solo come confronto. Il confronto fra le aree usa R_7 , calcolato ovunque con la stessa formula. Né R_7 né R_8 stimano una probabilità di collisione: sono indicatori di criticità dell'incontro.

R_7 e R_8 combinano componenti consolidati nella letteratura AIS-based. La struttura a rapporto fra velocità relativa e distanza al CPA appartiene alla famiglia dei *Vessel Conflict Ranking Operator* introdotta da Zhang et al. (2015) e ripresa nelle sue evoluzioni cinematiche da Liu et al. (2022, 2023); il quadro di prossimità fondato sulla DCPA è quello classico degli indici DCPA–TCPA (Debnath and Chin, 2016; Rong et al., 2019), censito fra le formulazioni più ricorrenti dalla rassegna di Du et al. (2020); la pesatura per velocità relativa recepisce il principio velocity-aware degli approcci a Velocity Obstacle (Chen et al., 2021) e dei modelli energetici di tipo IWRAP (IALA, 2009; Bakdi et al., 2020). Il contributo metodologico del presente lavoro consiste nell'integrare tali componenti in un'unica pipeline time-resolved, applicata a tutti gli encounter cinematici e confrontata con indicatori di diversa natura.

4.4 Procedura di confronto fra indicatori

Gli otto indicatori ($R_1, \dots, R_8$) sono calcolati sullo stesso dataset cinematico. Il confronto misura quanto indicatori di livelli diversi producano classificazioni di criticità diverse sugli stessi eventi. La procedura si articola in una fase globale (Spearman) e una mirata alla parte alta del ranking (concordanza top-N di tipo Jaccard).

4.4.1 Correlazione di rango globale (Spearman)

Per ciascun indicatore gli encounter sono ordinati dal più al meno critico, ottenendo un rango; ai valori uguali è assegnato il rango medio. La correlazione di Spearman $\rho(R_i, R_j)$ è la correlazione di Pearson tra i vettori di ranghi e consente di confrontare la coerenza delle classificazioni prodotte da indicatori non comparabili nelle unità di misura.

Nel Tigullio le matrici evidenziano una forte concordanza fra indicatori che condividono la stessa componente cinematica. In particolare, R_7 e R_8 mostrano la correlazione più elevata ($\rho = 0,97$), mentre R_2 e R_4 raggiungono $\rho = 0,96$ e R_3 e R_5 $\rho = 0,92$. Le correlazioni di R_7 e R_8 con gli indicatori basati sulla velocità relativa e sulla DCPA stimata confermano la presenza di strutture comuni nel ranking. R_1 presenta invece correlazioni generalmente moderate con gli altri indicatori.

Nello Stretto di Messina la struttura delle correlazioni risulta in parte diversa. Rimangono elevate le associazioni fra R_3 e R_5 e fra R_2 e R_4 , mentre R_8 mostra una minore concordanza con gli

indicatori R_1 – R_4 . Tale comportamento è stabile nei tre anni considerati: le correlazioni di R_8 con R_7 , R_1 e R_6 rimangono infatti su valori prossimi fra loro nel 2022, 2023 e 2024. La differenza rispetto al Tigullio è riconducibile alla dipendenza di R_8 dalla lunghezza della coppia, che ne modifica parzialmente il ranking rispetto agli indicatori basati sulle sole grandezze cinematiche.

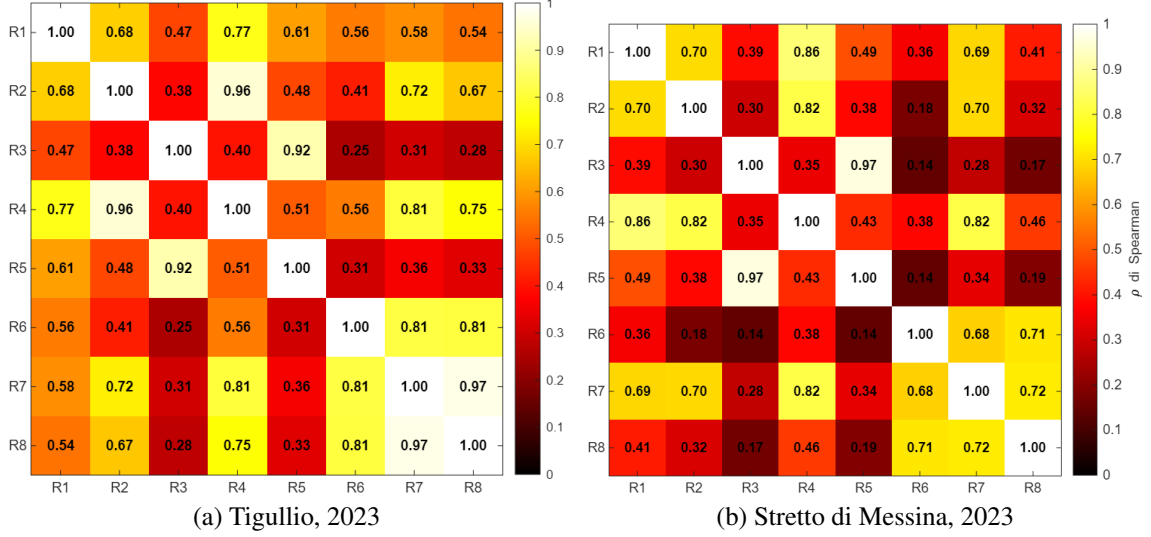


Figura 4.1: Matrici di correlazione di Spearman fra gli otto indicatori sul dataset cinematografico: (a) Tigullio, 2023; (b) Stretto di Messina, 2023.

4.4.2 Concordanza top- N (Jaccard)

La correlazione di Spearman valuta l'accordo globale fra gli ordinamenti, ma non quantifica direttamente la sovrapposizione degli encounter collocati nella parte alta dei ranking. Per questo confronto si considera la concordanza top- N mediante indice di Jaccard, definita come

$$J(R_i, R_j; N) = \frac{|T_N(R_i) \cap T_N(R_j)|}{|T_N(R_i) \cup T_N(R_j)|}, \quad (4.13)$$

dove $T_N(R_k)$ rappresenta l'insieme dei primi N encounter secondo l'indicatore R_k .

La matrice top-100 evidenzia una sovrapposizione molto ridotta fra R_1 e gli indicatori velocity-aware adottati successivamente, nonostante la correlazione globale di Spearman non sia nulla. Nel Tigullio, ad esempio, la concordanza fra R_1 e R_7 è $J = 0,01$, mentre quella fra R_7 e R_8 raggiunge $J = 0,85$. Nello Stretto di Messina la sovrapposizione fra R_1 e R_8 è altrettanto ridotta. Il confronto mostra quindi che la parte alta del ranking non è determinata dalla sola prossimità, ma risente della componente cinematica introdotta negli indicatori velocity-aware. Le matrici complete sono riportate in Fig. 4.2.

Per R_2 il confronto top-100 è invece poco informativo, poiché numerosi encounter assumono lo stesso valore massimo in presenza di minDCPA stimata non superiore a 1, m. Nel Tigullio tali encounter sono 422; la selezione dei primi cento dipende pertanto dall'ordinamento dei dati più che da una distinzione effettiva di criticità.

La concordanza è stata inoltre verificata sul 25,% degli encounter con valore più elevato dell'indicatore principale. L'aumento dell'insieme considerato determina una maggiore sovrapposizione fra gli indicatori, ma non modifica la distinzione osservata nella parte più alta dei ranking. Il confronto top- N viene pertanto utilizzato come verifica complementare alla correlazione globale e non come criterio autonomo di selezione.

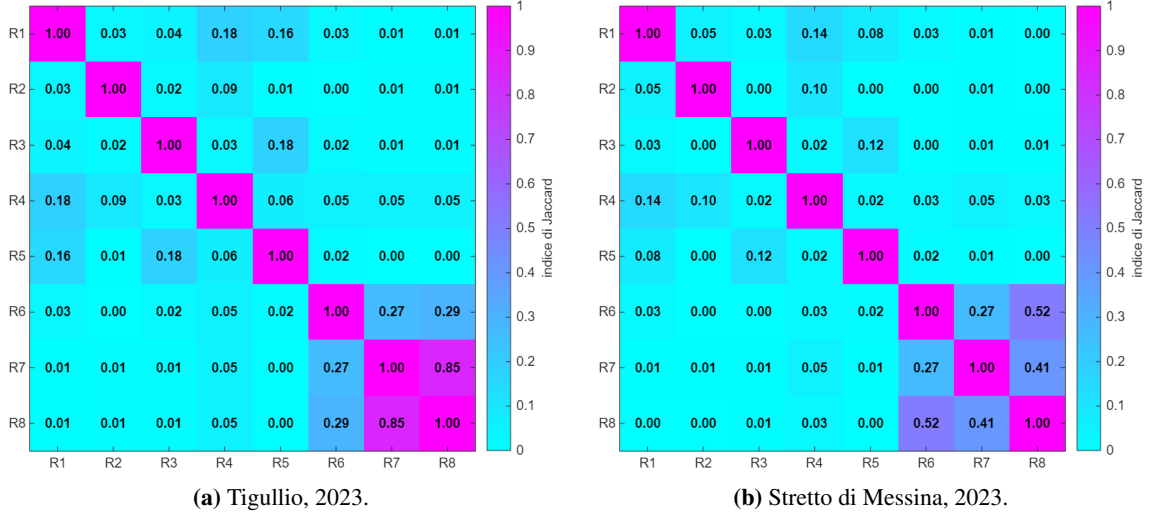


Figura 4.2: Concordanza top-100 (Jaccard) fra gli otto indicatori sul dataset cinematografico: (a) Tigullio; (b) Stretto di Messina.

Sulla base del confronto complessivo fra gli indicatori, per le analisi successive viene adottato un indicatore principale distinto per area: R_7 nel Tigullio e R_8 nello Stretto di Messina (Sezione 4.3.3). La scelta consente di mantenere costante la metrica all'interno di ciascuna area nelle successive analisi spaziali, temporali e per tipologia di encounter. Entrambi gli indicatori combinano la DCPA stimata con una componente legata alla velocità relativa e si distinguono quindi dagli indicatori basati esclusivamente sulla prossimità o sull'approach rate. La selezione dell'indicatore principale è successivamente mantenuta nelle mappe di criticità e nelle analisi spaziali e temporali, mentre il confronto con gli altri indicatori rimane alla base della valutazione della sensibilità della classificazione.

4.4.3 Mappe pesate globali

Per ciascun indicatore R_k i valori sono normalizzati nell'intervallo $[0, 1]$ dividendo per il massimo osservato, normalizzazione che non altera l'informazione ordinale e rende confrontabili le mappe. Il valore aggregato nella cella (i, j) è la somma dei pesi normalizzati degli encounter la cui posizione all'istante di minima distanza osservata cade nella cella:

$$M_k(i, j) = \sum_{e \in \text{cella}(i, j)} \tilde{R}_k(e), \quad \tilde{R}_k(e) = \frac{R_k(e)}{\max_e R_k(e)} \quad (4.14)$$

I valori $M_k(i, j)$ sono rappresentati in scala logaritmica ($\log_{10}$), con le celle vuote rese trasparenti

per distinguere le aree prive di encounter da quelle a bassa criticità aggregata. Il confronto fra le mappe globali ottenute con i diversi indicatori, per entrambe le aree di studio, è presentato nel Capitolo 5.

4.4.4 Mappe disaggregate per macro-classe

Le mappe globali mascherano la variazione spaziale della composizione delle macro-classi. Per esplicitare i contributi differenziali, gli encounter sono raggruppati per coppia di macro-classi (Sezione 3.5.2) e per ciascun raggruppamento è costruita una mappa pesata con l'indicatore principale dell'area. Il peso unico tiene costante l'indicatore al variare della classe, in modo da isolare l'effetto della tipologia di unità da quello della scelta dell'indicatore. Le mappe disaggregate per coppia di macro-classi, per entrambe le aree, sono presentate nel Capitolo 5.

4.5 Classificazione degli encounter per tipo di incontro

Le mappe per macro-classe distinguono gli encounter in base alla tipologia di unità; una classificazione complementare distingue invece la geometria dell'incontro. Ogni encounter cinematico è assegnato a una di tre classi geometriche in base alla differenza di rotta $|\Delta COG|$ nel record in cui la DCPA è minima. Le classi sono denominate frontale (*head-on*), incrocio (*crossing*) e sorpasso (*overtaking*) per analogia con le situazioni del COLREG (International Maritime Organization, 2018), ma non coincidono con esse:

- frontale: $|\Delta COG| \geq 150^\circ$;
- sorpasso: $|\Delta COG| \leq 45^\circ$;
- incrocio: $45^\circ < |\Delta COG| < 150^\circ$.

La classificazione ha scopo descrittivo. Le soglie sono una scelta operativa del presente lavoro: il Regolamento definisce le situazioni attraverso il rilevamento relativo e l'orientamento reciproco delle rotte, senza fissare soglie di $|\Delta COG|$. La verifica di conformità (Sezione 4.9) adotta un criterio diverso, basato sul rilevamento relativo nella fase di approccio, e i due criteri non producono gli stessi totali. Le mappe per tipo usano la classificazione su $|\Delta COG|$ perché si applica a tutti gli encounter cinematici; il criterio della verifica richiede almeno 4 record e si applica al 69,0 % degli encounter nel Tigullio e al 54,1 % nello Stretto di Messina nel 2023. Sugli encounter classificati da entrambi i criteri la classe coincide nel 42,0 % dei casi nel Tigullio, contro il 34,3 % atteso per caso, e nel 55,9, 55,3 e 57,1 % nello Stretto nel 2022, 2023 e 2024, contro il 40,5, 40,9 e 40,7 % attesi. La classificazione su $|\Delta COG|$ descrive quindi la geometria delle rotte e non individua la situazione regolamentare. Per ciascuna classe è costruita una mappa di criticità aggregata pesata con l'indicatore principale; le mappe, per entrambe le aree, sono riportate nel Capitolo 5.

4.6 Autocorrelazione spaziale della criticità

Le mappe di criticità mostrano dove la criticità si concentra, ma non se la concentrazione sia statisticamente significativa. Per valutarlo si adottano due statistiche di autocorrelazione spaziale, in linea con l'analisi degli hotspot di near collision di Rong et al. (2021).

La prima è l'indice di Moran globale

$$I = \frac{n}{W} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (4.15)$$

dove x_i è il valore della cella i , $\bar{x}$ la media sulla griglia, w_{ij} vale 1 per le celle adiacenti secondo la contiguità queen e 0 altrimenti, e W è la somma dei pesi. L'indice è calcolato sulla superficie della criticità aggregata, cioè sulla somma per cella dell'indicatore normalizzato, per ciascuno degli otto indicatori. In questo calcolo ciascun encounter contribuisce sia alla cella dell'unità i sia a quella dell'unità j , le cui posizioni all'istante di minima distanza osservata distano meno di 500 m. Il calcolo comprende tutte le celle della griglia, incluse quelle prive di encounter, e la significatività è valutata con 199 permutazioni casuali dei valori di cella. Poiché le celle vuote sono incluse, l'indice riflette anche la concentrazione degli incontri nella porzione di area effettivamente frequentata.

La seconda è la statistica locale di Getis-Ord

$$G_i^* = \frac{\sum_j w_{ij} x_j - \bar{x} \sum_j w_{ij}}{S \sqrt{\frac{n \sum_j w_{ij}^2 - (\sum_j w_{ij})^2}{n-1}}} \quad (4.16)$$

calcolata sulla superficie della criticità aggregata dell'indicatore principale, costruita con la sola posizione dell'unità i , dove S è la deviazione standard sulla griglia e w_{ij} vale 1 per le celle di una finestra di 5×5 celle comprendente la cella stessa, pari a circa 1 km nel Tigullio e a circa 1,25 km nello Stretto. Sono considerate hotspot le celle con $G_i^* > 1,96$ (significatività al 95 %) e, fra queste, quelle con $G_i^* > 2,58$ (99 %). Le soglie sono applicate cella per cella, senza correzione per confronti multipli. Le mappe degli hotspot, per entrambe le aree, sono riportate nel Capitolo 5.

L'indice di Moran è usato anche per un secondo scopo: verificare se la struttura spaziale della criticità per incontro possa essere ricondotta alla distribuzione del traffico. In questo caso il calcolo è limitato alle celle con almeno 3 encounter, con contiguità queen fra le sole celle selezionate, ed è applicato a tre superfici: densità di encounter, criticità aggregata (somma dell'indicatore principale) e criticità per incontro (media dell'indicatore principale). I valori dell'indicatore vengono poi riallocati casualmente fra gli encounter, lasciando invariata la cella di ciascuno: la superficie di densità resta identica per costruzione, mentre la criticità per incontro viene ridistribuita. Ricostruita la superficie della media per cella, se ne calcola l'indice di Moran; la procedura è ripetuta 199 volte, con seme fisso del generatore casuale per rendere il risultato riproducibile. La distribuzione ottenuta descrive l'autocorrelazione attesa quando la criticità per incontro non dipende dalla

posizione a parità di traffico. Uno scarto significativo del valore osservato indica che la struttura spaziale della criticità per incontro non è riproducibile dalla sola distribuzione del traffico; il test non ne individua la causa.

Come grandezza alternativa alla media si considera la frazione di encounter di ciascuna cella con l'indicatore principale superiore alla soglia che delimita il 5 % degli encounter dell'area con il valore più elevato. La grandezza stima la probabilità che un encounter della cella superi tale soglia e non è una probabilità di collisione.

4.7 Profilo temporale della criticità

Le analisi precedenti sono di natura puramente spaziale. Per introdurre la dimensione temporale, la criticità aggregata (somma dell'indicatore principale) è profilata rispetto al mese, all'ora del giorno e al giorno della settimana, e sintetizzata in una matrice mese $\times$ ora. Si tratta di un'elaborazione descrittiva, motivata dal fatto che l'esposizione non è stazionaria: la firma temporale del traffico da diporto (stagionale e diurna) è diversa da quella di un corridoio di traffico ordinato. Gli orari sono espressi in UTC. Il periodo estivo comprende i mesi da giugno a settembre; il periodo notturno le ore fra le 21:00 e le 06:00 UTC. La somma dell'indicatore cresce con il numero di encounter: ogni quota è quindi riportata anche sul numero di encounter, e la differenza fra le due quote indica se nel periodo considerato il valore medio dell'indicatore è superiore o inferiore alla media annua. I profili temporali, per entrambe le aree, sono riportati nel Capitolo 5.

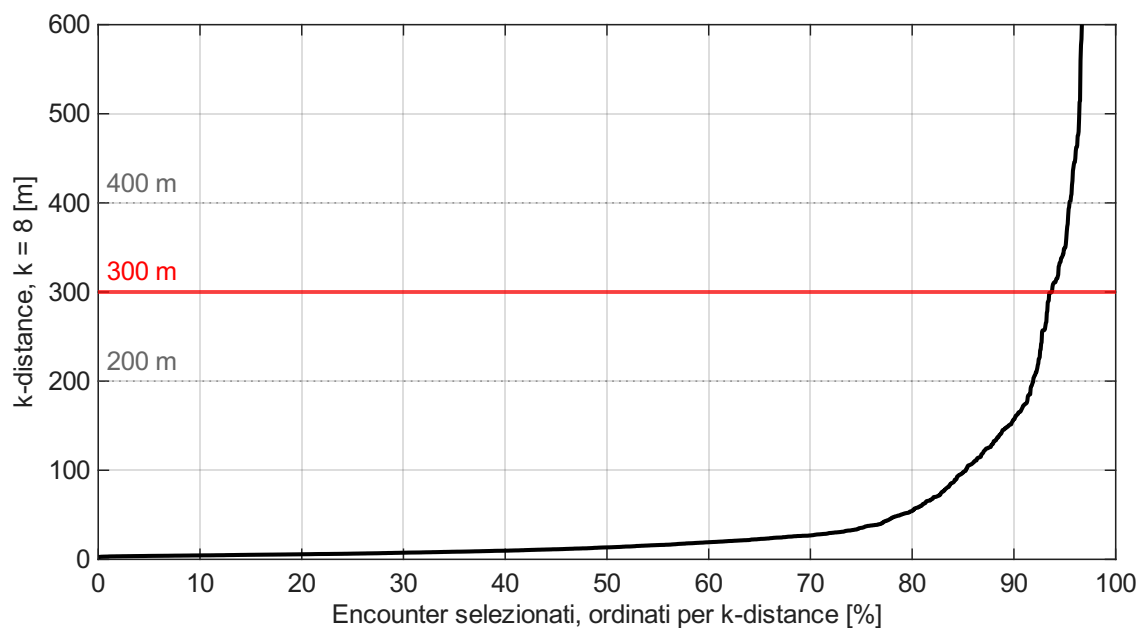
4.8 Zone di criticità per densità (DBSCAN)

Accanto agli hotspot su griglia (Sezione 4.6), sono state individuate concentrazioni spaziali discrete degli incontri con valori elevati dell'indicatore principale. È stato considerato il 25 % degli encounter con l'indicatore principale più elevato nell'area (R_7 nel Tigullio, R_8 nello Stretto di Messina). Agli encounter selezionati è stato applicato l'algoritmo DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) di Ester et al. (1996), sulle coordinate della posizione dell'unità i all'istante di minima distanza osservata, convertite in metri su un piano locale. L'indicatore interviene soltanto nella selezione e non è impiegato come peso. I parametri sono un raggio di vicinato $\varepsilon = 300$ m e un numero minimo di punti $minPts = 8$, identici nelle due aree. L'algoritmo non richiede il numero di cluster a priori e classifica come rumore i punti isolati. Per ciascuna zona sono calcolati numerosità, centroide, valore medio dell'indicatore ed estensione, espressa come distanza quadratica media dal centroide.

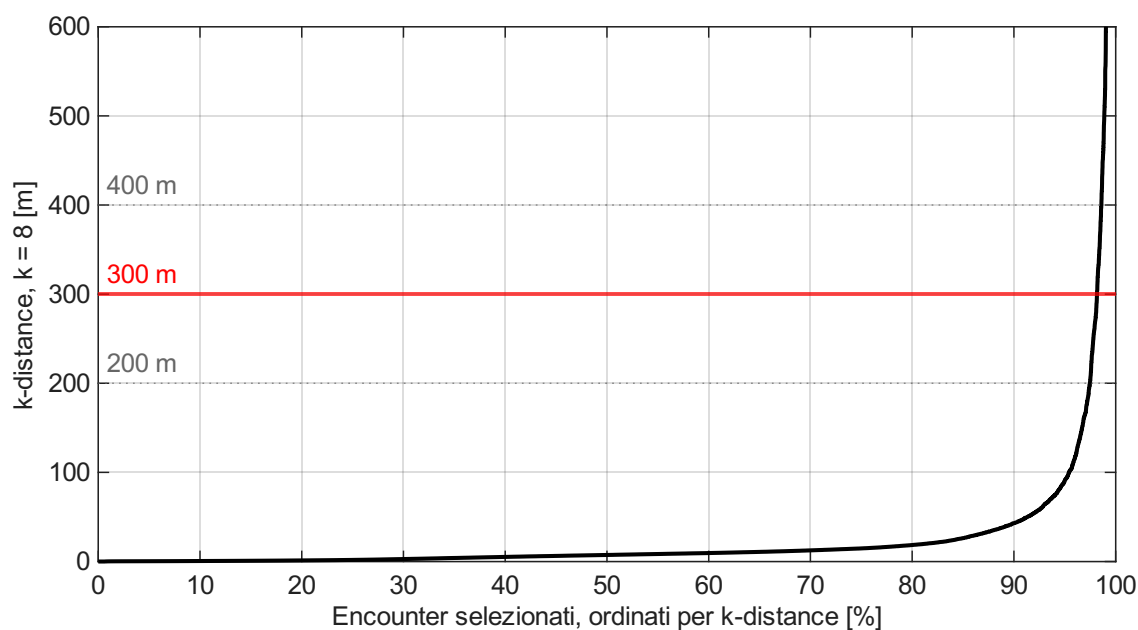
Il valore di ε è stato confrontato con la curva k-distance (Ester et al., 1996): per ciascun encounter selezionato si calcola la distanza dall'ottavo punto più vicino, contando il punto stesso, e i valori sono ordinati in modo crescente (Fig. 4.3). Con questa definizione la quota di encounter con k-distance non superiore a ε coincide con la quota di punti di nucleo di DBSCAN con $minPts$ pari a 8. La k-distance ha mediana di 13 m nel Tigullio e di 7 m nello Stretto di Messina nel 2023, e supera 350 e 91 m nel 5 % degli encounter. Nel Tigullio il raggio di 300 m include il 93,7 % degli

encounter e cade nel tratto in cui la curva cresce rapidamente: fra 200 e 400 m la quota di punti di nucleo passa dal 91,9 al 95,4 %. Nello Stretto la quota è già del 97,5 % a 200 m e sale al 98,2 e al 98,6 % a 300 e 400 m. Il ginocchio della curva, individuato come il punto più distante dalla retta che ne congiunge gli estremi dopo averla normalizzata e limitata al 99 % dei valori più piccoli, cade a circa 185 m nel Tigullio, al 92 % dei punti, e a circa 48 m nello Stretto, al 91 % (34 e 46 m nel 2022 e nel 2024); il raggio adottato è superiore a entrambi.

Per verificare la dipendenza delle zone dai parametri, il clustering è stato ripetuto sulla stessa selezione con ϵ pari a 200, 300 e 400 m e minPts pari a 5, 8 e 12, e con $\epsilon = 300$ m e minPts pari a 8 sul 30 % e sul 20 % degli encounter con l'indicatore principale più elevato. Ogni partizione è stata confrontata con quella di riferimento mediante l'indice di Rand corretto, calcolato sugli encounter comuni alle due selezioni con il rumore trattato come una classe, e mediante lo stesso indice limitato agli encounter assegnati a una zona in entrambe le partizioni.



(a) Golfo del Tigullio (R_7)



(b) Stretto di Messina, 2023 (R_8)

Figura 4.3: Curva k-distance ($k = 8$) del 25 % degli encounter con l'indicatore principale più elevato. La linea continua indica il raggio adottato per DBSCAN (300 m), le linee punteggiate i valori di 200 e 400 m dell'analisi di sensitività. L'asse verticale è limitato a 600 m; i valori massimi sono 12 836 m nel Tigullio e 23 325 m nello Stretto di Messina.

Poiché gli encounter selezionati sono un sottoinsieme della flotta complessiva, la loro densità spaziale dipende anche da quella del traffico. Le zone individuano quindi concentrazioni di incontri con valori elevati dell'indicatore principale, e non aree in cui il singolo incontro è mediamente più severo. Le zone, per entrambe le aree, sono riportate nel Capitolo 5.

4.9 Verifica di conformità COLREG

Le analisi precedenti descrivono gli incontri in termini geometrici e cinematici, ma non indicano come l'incontro sia stato gestito. La verifica di conformità al COLREG (International Maritime Organization, 2018) aggiunge questa dimensione comportamentale: distingue, entro i limiti del dato, gli incontri in cui una manovra evasiva è rilevabile nella traccia da quelli in cui non lo è.

La verifica di conformità si articola in tre passi: individuazione della situazione regolamentare, attribuzione del ruolo, ricerca della manovra prescritta. La classificazione non si fonda sulla sola differenza di rotta, ma sul rilevamento relativo, come nelle definizioni del Regolamento. Per ciascuna unità si determina la rotta in approccio dalle posizioni successive dei primi record dell'incontro, impiegando la rotta trasmessa solo come valore di ripiego, e si calcola il rilevamento con cui ciascuna unità vede l'altra all'inizio dell'incontro.

Si riconosce una situazione di sorpasso ai sensi della Regola 13 quando una delle due unità vede l'altra con un rilevamento relativo superiore a $112,5^\circ$ dalla propria prua, cioè oltre $22,5^\circ$ a poppavia del proprio traverso, e l'altra è la più veloce. L'unità più veloce è la nave raggiungente ed è tenuta a manovrare. Si riconosce una situazione di incontro frontale ai sensi della Regola 14 quando le rotte in approccio differiscono di almeno 150° e ciascuna unità vede l'altra entro 45° dalla propria prua, condizione che obbliga la manovra ad entrambe. Le restanti configurazioni sono trattate come incrocio ai sensi della Regola 15, attribuendo l'obbligo di manovra all'unità che vede l'altra nel proprio settore prodiero di dritta; nei casi in cui entrambe soddisfino tale condizione si assume tenuta a manovrare quella con rilevamento relativo maggiore.

La manovra è ricercata confrontando rotta e velocità della fase di approccio con quelle della fase finale dell'incontro, ricavate dai primi e dagli ultimi record; la rotta è stimata dalle posizioni successive. È riconosciuta un'accostata quando la rotta varia di oltre 15° , a dritta se la variazione è in senso orario, e una riduzione di velocità quando la velocità diminuisce di oltre 1 nodo e di oltre il 20 % del valore iniziale. Nell'incrocio e nel sorpasso l'esito è conforme quando l'unità tenuta a manovrare riduce la velocità oppure accosta a dritta, non conforme quando accosta a sinistra senza ridurre la velocità oppure non manovra. Nel frontale l'esito è conforme solo quando entrambe le unità accostano a dritta, come prescritto dalla Regola 14; la sola riduzione di velocità non è sufficiente e ogni altro caso è non conforme. Il criterio sul verso dell'accostata è applicato a tutti i tipi di incontro: nel sorpasso la Regola 13 non prescrive il lato di passaggio, e il criterio è quindi più restrittivo del Regolamento. L'esito è indeterminato quando, nell'incrocio, nessuna delle due unità ha l'altra nel proprio settore prodiero di dritta, oppure quando la traccia è troppo breve. La procedura usa due soglie sul numero di record: almeno 4 per classificare l'incontro e almeno 6 per valutare la manovra; gli incontri con 4 o 5 record ricevono quindi una classificazione ma un esito indeterminato.

La procedura eredita i limiti del dato AIS: non coglie l'intenzione, non distingue le precedenze della Regola 18 (in particolare la precedenza delle unità a vela sulle unità a propulsione meccanica) e risente dell'intermittenza dei transponder di Classe B; i verdetti sono pertanto indicativi e, in

particolare, la classificazione *non conforme* segnala l'assenza di una manovra rilevabile e non un illecito accertato. La procedura non verifica inoltre se all'inizio dell'incontro esisteva rischio di collisione ai sensi della Regola 7: l'unità tenuta a manovrare che transita già a distanza adeguata e non modifica rotta né velocità riceve un esito non conforme. Anche per questa ragione i tassi di conformità sono stime per difetto. I tassi di conformità per tipo di incontro sono riportati nel Capitolo 5.

La verifica è illustrata su singoli incontri dello Stretto di Messina dei tre anni (Sezione 5.8.8). Con il passo di un minuto, un incontro fra unità a velocità di servizio resta entro la soglia di estrazione per 2–3 minuti e non raggiunge i 6 record richiesti per valutare la manovra: la situazione è classificata, ma l'esito è indeterminato. I casi illustrativi sono quindi valutati sulle posizioni trasmesse dalle unità in Classe A, più frequenti dei record sincroni, con le stesse soglie di accostata e di riduzione di velocità.

I candidati sono incontri cinematici fra unità in Classe A dichiarate come unità da diporto, unità veloci, navi passeggeri, da carico o cisterne, con velocità mediana di almeno 6 kn e velocità minima di almeno 4 kn nell'incontro e distanza osservata minima non superiore a 400 m. La geometria iniziale, valutata sui record al minuto, deve essere netta: nel sorpasso rotte che differiscono di non più di 20° e velocità che differiscono di almeno 3 kn, nel frontale rotte che differiscono di almeno 150°, nell'incrocio rotte che differiscono fra 60° e 120°. Per ciascun tipo di incontro sono estratti fino a 15 candidati in ordine casuale, con seme fisso; fra i casi validi la scelta privilegia tipi di nave, unità e anni diversi ed esiti diversi.

Per ogni incontro le posizioni trasmesse fra 12 minuti prima e 4 minuti dopo l'istante di minima distanza osservata sono ricampionate a passo di 5 s. Il caso è valido se l'intervallo fra due posizioni non supera 150 s, se entrambe le unità procedono ad almeno 6 kn nella fase di approccio e ad almeno 2 kn nell'intera finestra, e se la situazione, riclassificata nella fase di approccio con i criteri della verifica, coincide con quella di partenza. La rotta, ricavata dalle posizioni, e la velocità mediana della fase di approccio, da 12 a 10 minuti prima dell'istante di minima distanza osservata, sono confrontate con quelle del passaggio, da un minuto prima a un minuto dopo. La fase di approccio precede in genere l'ingresso nella soglia di estrazione e la manovra anticipata che la Regola 16 richiede all'unità tenuta a manovrare. Poiché le finestre di confronto sono diverse, gli esiti dei casi illustrativi non sono confrontabili con i tassi di conformità calcolati sui record sincroni.

Capitolo 5

Risultati

Il capitolo presenta i risultati ottenuti applicando ai dataset del Capitolo 3 la metodologia del Capitolo 4. Per il Golfo del Tigullio ciascuna analisi è descritta per esteso. Per lo Stretto di Messina (Sezione 5.8) sono riportate le figure del 2023, con un testo limitato alle differenze rispetto al Tigullio; i valori dei tre anni sono riuniti in tabelle e le figure del 2022 e del 2024 sono riportate in Appendice A.2.

L'indicatore principale è R_7 nel Tigullio e R_8 nello Stretto di Messina (Sezione 4.4).

5.1 Tigullio: mappe globali e rapporto criticità-densità

Le Fig. 5.1 e 5.2 riportano le mappe di criticità aggregata per gli otto indicatori. Ciascun encounter contribuisce alla cella che contiene la posizione dell'unità i all'istante di minima distanza osservata, con peso pari all'indicatore normalizzato; il valore della cella è la somma dei pesi, rappresentata in scala logaritmica comune a tutte le mappe. Le mappe sono costruite sulle stesse celle, poiché gli encounter considerati sono gli stessi, e per tutti gli indicatori i valori massimi si concentrano negli stessi due settori, presso Camogli e fra Portofino e Santa Margherita Ligure. La cella con la criticità aggregata massima, all'imboccatura del porto di Portofino ($44,303^\circ$ N; $9,213^\circ$ E), coincide con la cella di densità massima. Le differenze fra gli indicatori riguardano soprattutto l'intensità relativa dei valori nelle celle, più che la posizione dei massimi. La Fig. 5.3 riporta la criticità aggregata di R_7 nella vista di dettaglio del promontorio di Portofino.

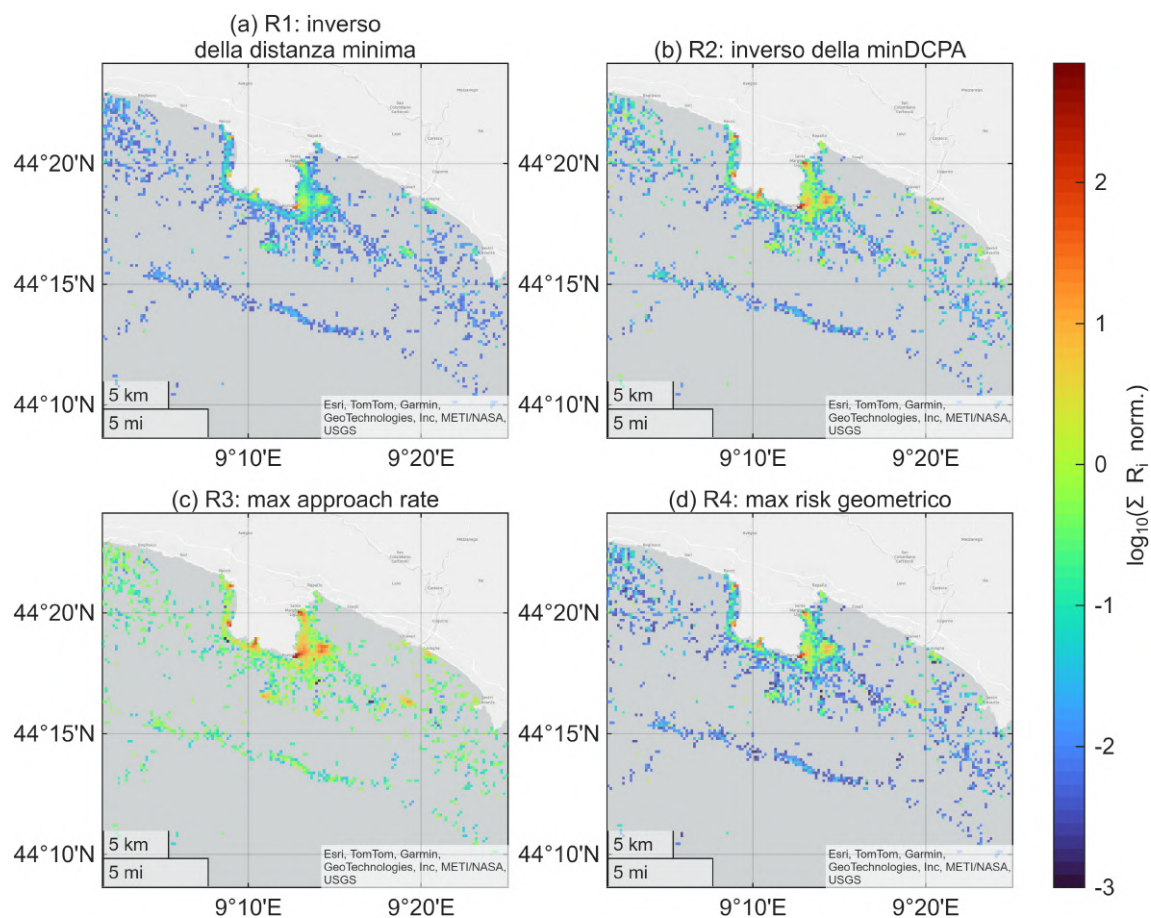


Figura 5.1: Tigullio: mappe di criticità aggregata per gli indicatori R_1 – R_4 . Somma per cella dell'indicatore normalizzato, in scala logaritmica.

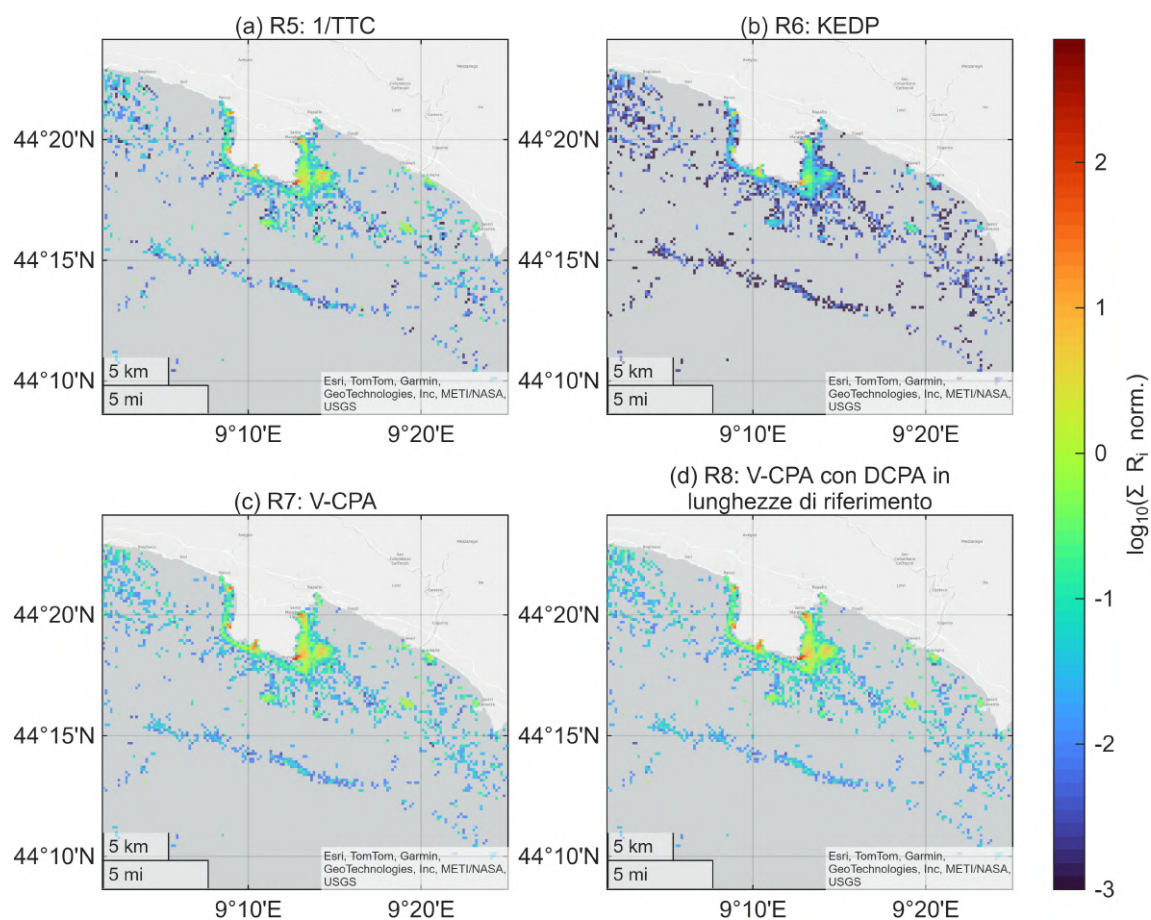


Figura 5.2: Tigullio: mappe di criticità aggregata per gli indicatori velocity-aware R_5 – R_8 . Somma per cella dell'indicatore normalizzato, in scala logaritmica.

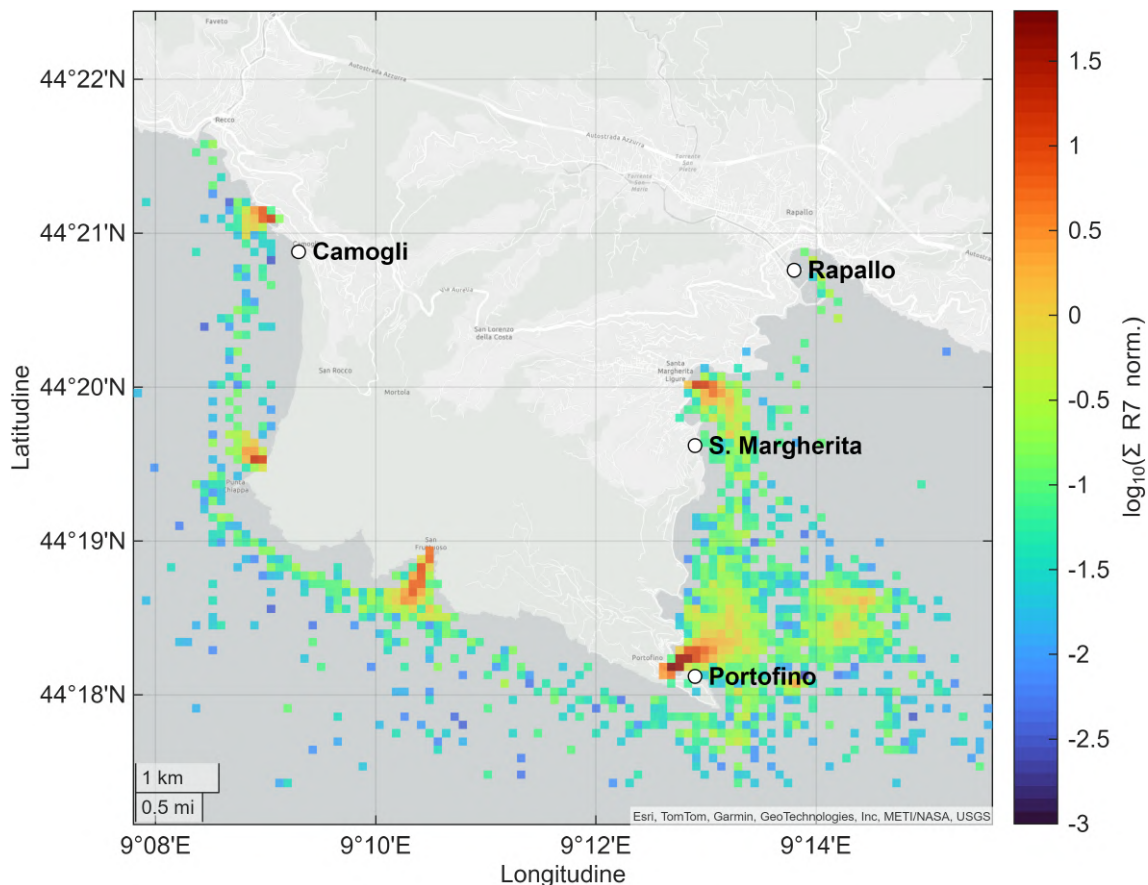


Figura 5.3: Tigullio: criticità aggregata di R_7 nella vista di dettaglio del promontorio di Portofino (44,29–44,37° N; 9,13–9,26° E), su celle di circa 100 m, in scala logaritmica.

Il rapporto con la densità di traffico è stato misurato sulle 297 celle con almeno 3 encounter (Fig. 5.4–5.6). La correlazione di rango fra criticità aggregata (somma di R_7 per cella) e numero di encounter è $\rho = 0,890$; per la criticità per incontro (media di R_7 per cella) scende a $\rho = 0,317$. L'ordinamento delle celle secondo la criticità aggregata coincide quindi in larga parte con quello della densità, e la mappa aggregata descrive soprattutto dove si concentrano gli incontri. La media per cella riduce la dipendenza dal volume di traffico e descrive la criticità del singolo incontro. Nella mappa della criticità per incontro i valori più alti non si concentrano soltanto nelle celle a densità massima, presso Portofino, ma compaiono anche in celle periferiche: presso Camogli, in un settore circa 3 km a ovest di Portofino e, nel settore orientale, presso Chiavari. La fascia al largo, a sud del promontorio, presenta invece valori bassi.

La relazione con la densità non si annulla: con $\rho = 0,317$ le celle con maggior densità tendono ad avere valori medi di R_7 più alti. La media calcolata su poche osservazioni è inoltre più variabile; la soglia di 3 encounter per cella limita questo effetto senza eliminarlo.

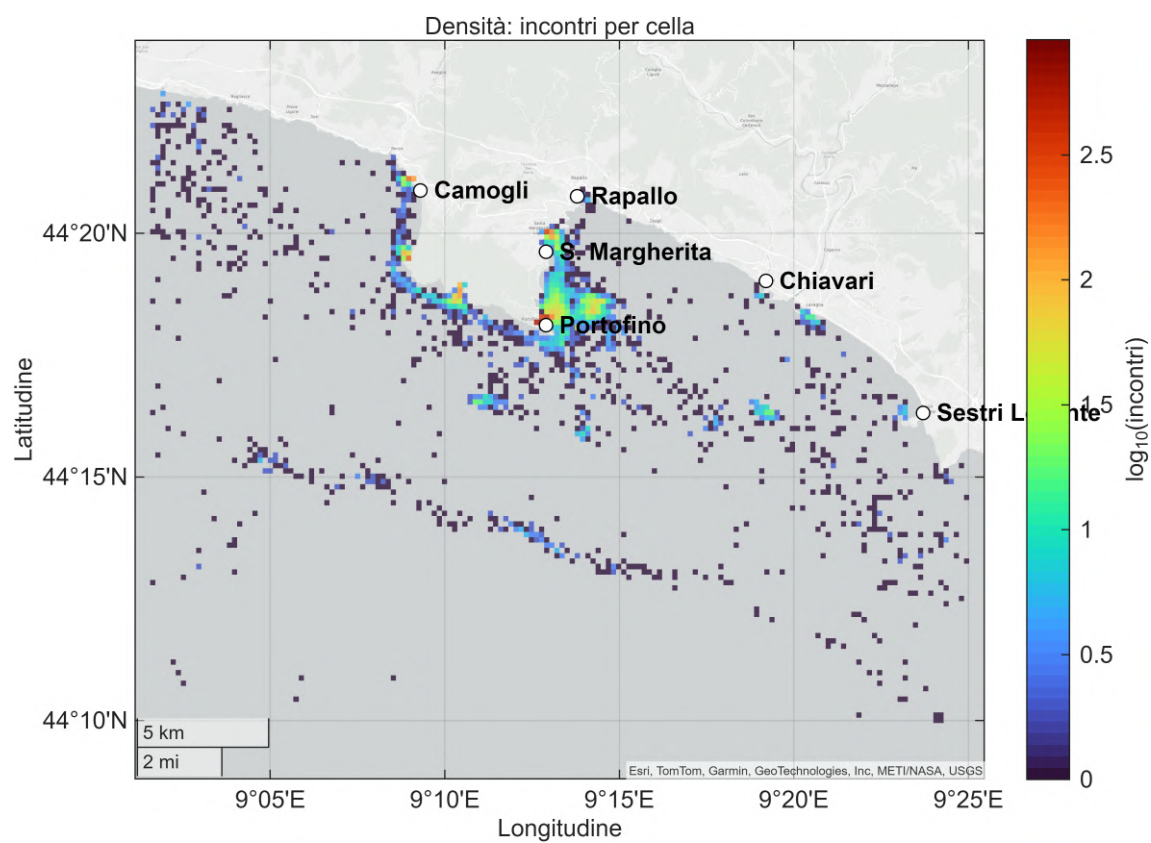


Figura 5.4: Tigullio: densità degli encounter cinematici per cella.

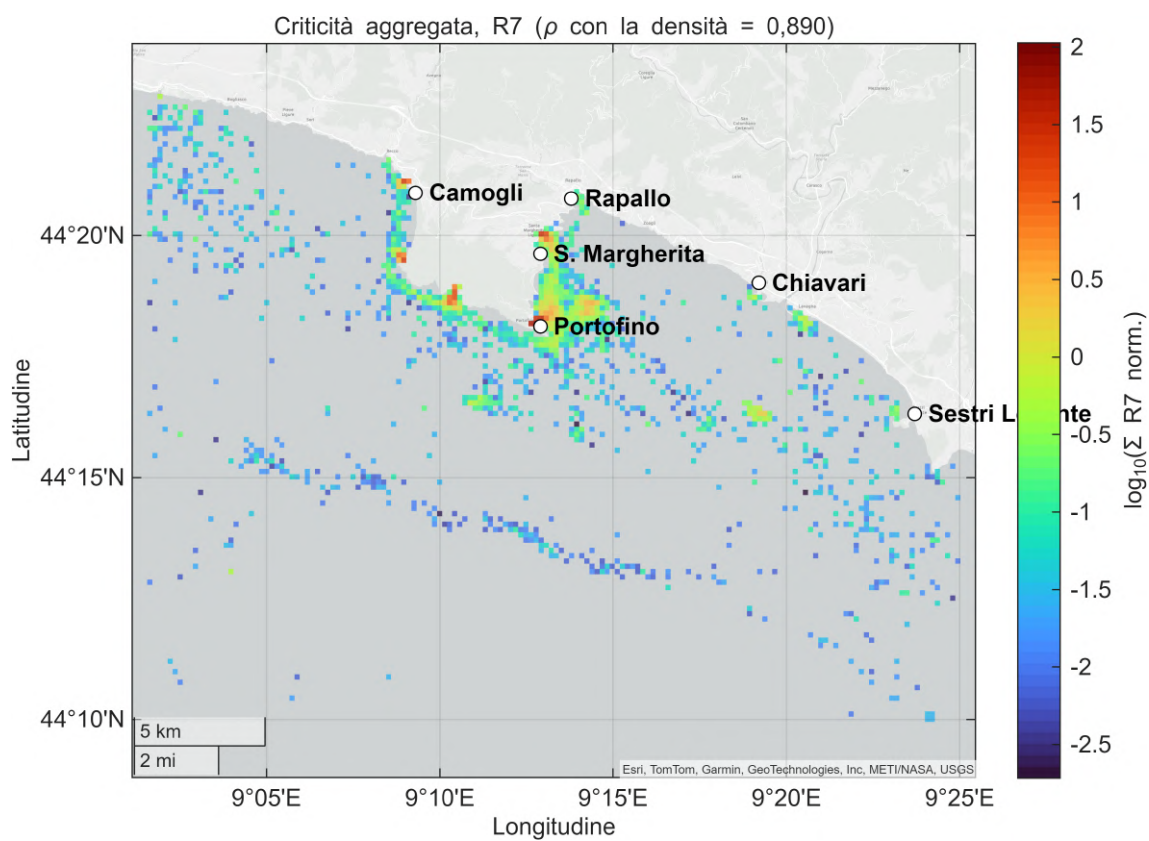


Figura 5.5: Tigullio: criticit   aggregata (somma di R_7 per cella), con la correlazione di Spearman con la densit   calcolata sulle celle con almeno 3 encounter.

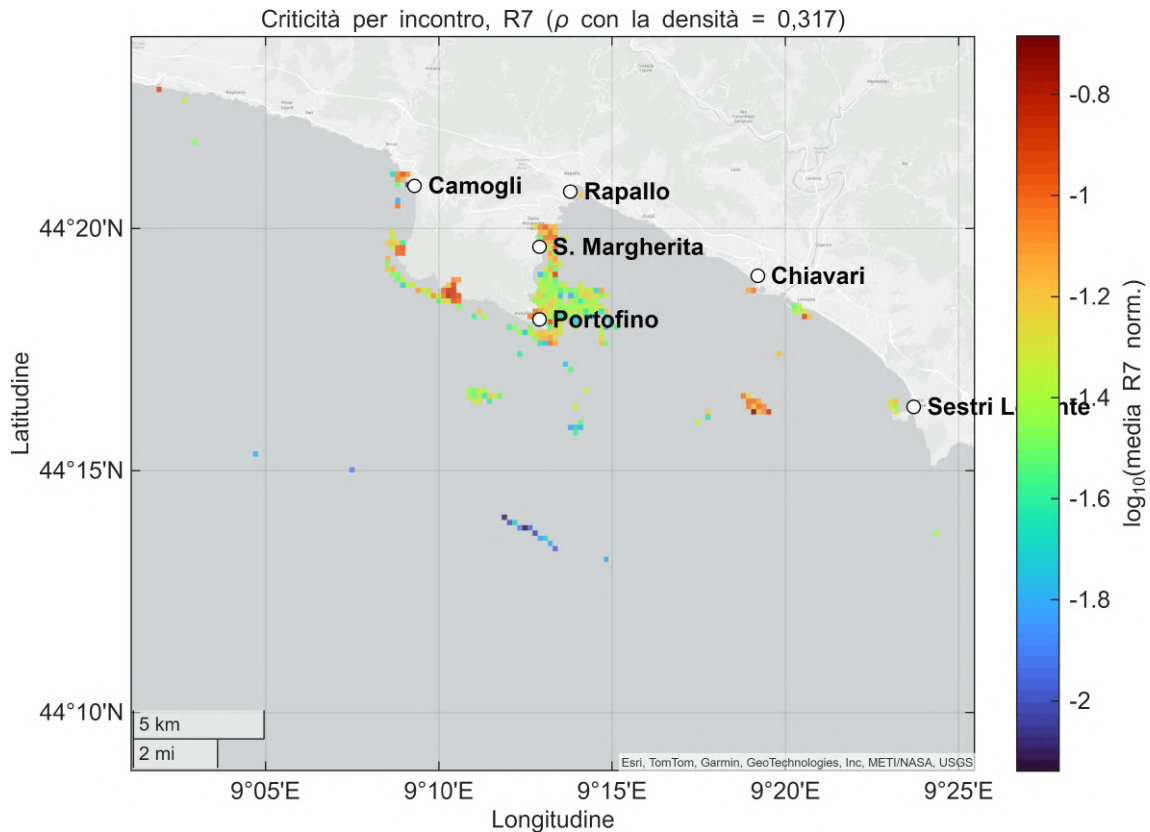


Figura 5.6: Tigullio: criticità per incontro (media di R_7 per cella), con la correlazione di Spearman con la densità calcolata sulle celle con almeno 3 encounter.

5.2 Tigullio: composizione per macro-classe

La disaggregazione per coppia di macro-classi richiede che entrambe le unità siano identificate nell'archivio statico ed è quindi limitata ai 2 572 encounter in full match (Sezione 3.6.4), circa il 30 % degli 8 573 encounter cinematici. La Fig. 5.7 riporta le mappe di criticità aggregata di R_7 per le cinque coppie più numerose: pleasure–pleasure (2 118 encounter), passenger–pleasure (194), other–pleasure (126), other–other (94) e passenger–passenger (40).

Le coppie fra unità da diporto costituiscono l'82 % del sottoinsieme, come nella composizione del traffico dell'area (Sezione 3.5.2). Trattandosi di mappe aggregate, ciascuna riflette anzitutto la distribuzione spaziale degli incontri della coppia; per le coppie con meno di 200 encounter le mappe hanno valore qualitativo. Il sottoinsieme in full match non è necessariamente rappresentativo dell'intero dataset: gli encounter esclusi coinvolgono almeno un'unità non identificata nell'archivio statico.

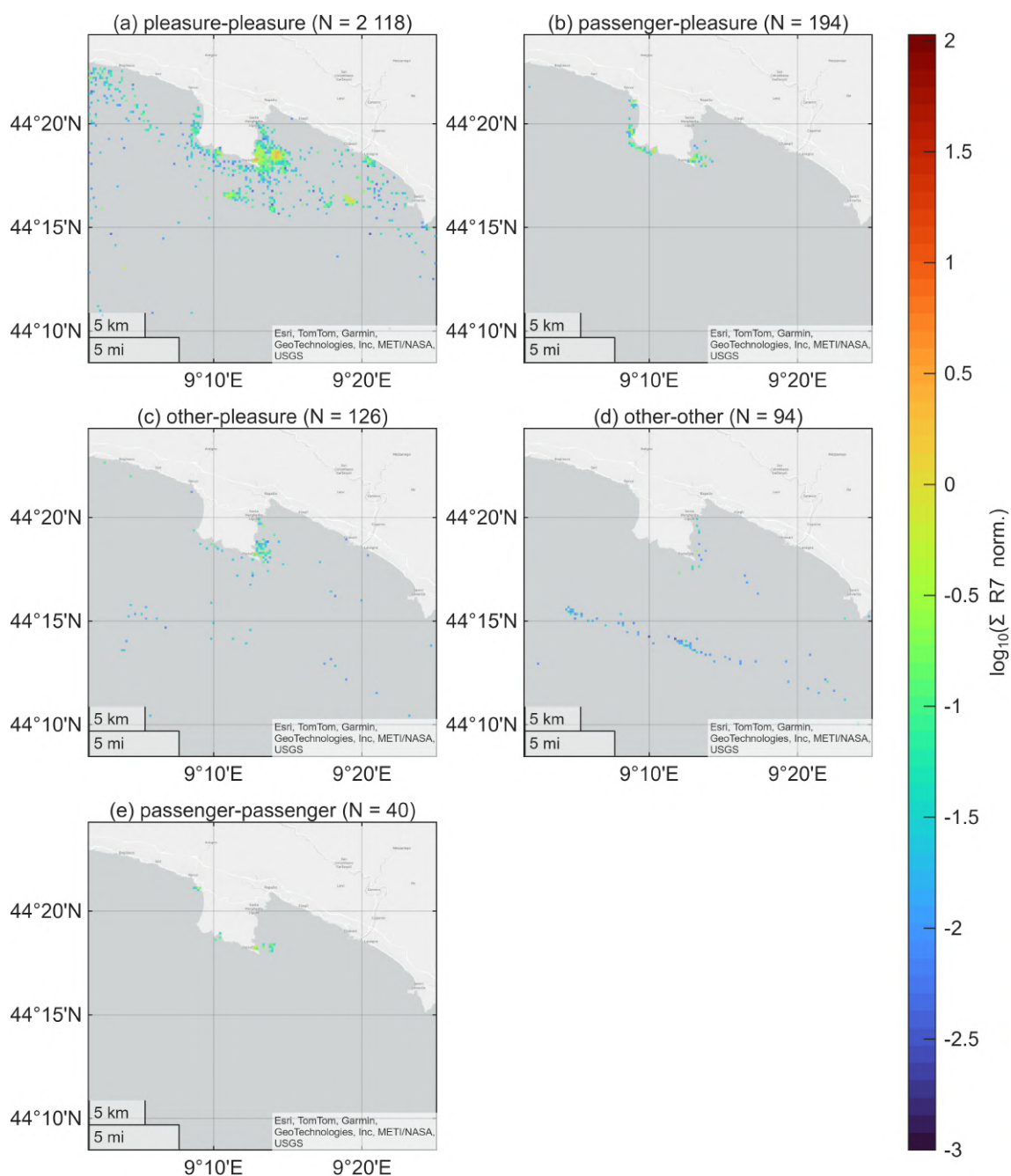


Figura 5.7: Tigullio: criticità aggregata di R_7 per le cinque coppie di macro-classi più numerose, sugli encounter con entrambe le unità identificate, in scala logaritmica.

5.3 Tigullio: tipologia degli encounter

La classificazione della Sezione 4.5, basata su $|\Delta COG|$ nel record in cui la DCPA stimata è minima, assegna gli 8 573 encounter cinematici all'incrocio (3 508; 40,9 %), al sorpasso (2 811; 32,8 %) e al frontale (2 254; 26,3 %). La Fig. 5.8 riporta le corrispondenti mappe di criticità aggregata di R_7 .

Le tre mappe presentano i valori più alti nello stesso settore, fra Portofino e Santa Margherita Ligure. La mappa del frontale presenta inoltre un allineamento di celle al largo, a sud del promon-

torio di Portofino, meno evidente nelle altre due. Poiché le mappe sono aggregate, le differenze riflettono in primo luogo dove ciascun tipo di incontro è più frequente. La classificazione è distinta da quella della verifica di conformità (Sezione 5.7).

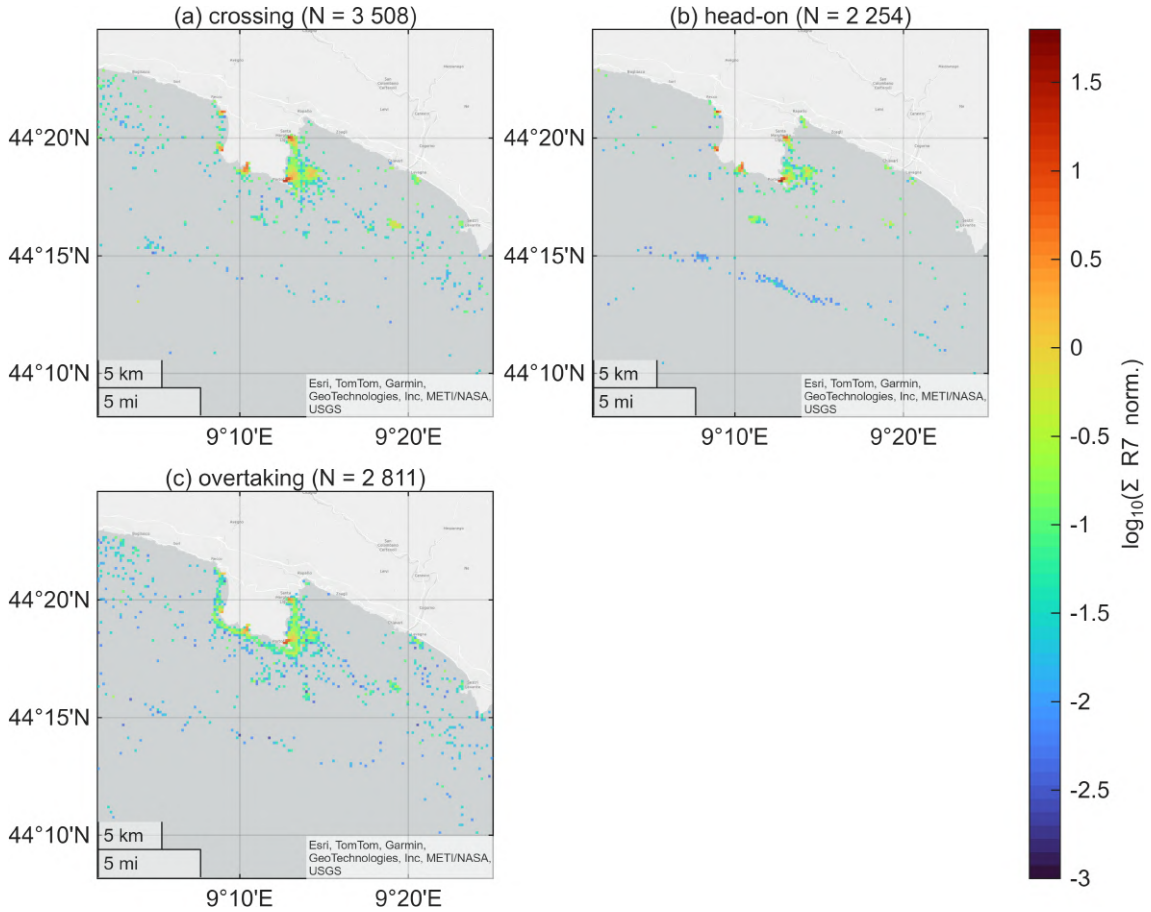


Figura 5.8: Tigullio: mappe di criticità aggregata di R_7 per tipo di incontro (incrocio, frontale, sorpasso), con classificazione basata su $|\Delta COG|$, in scala logaritmica.

5.4 Tigullio: struttura spaziale e hotspot

L'indice di Moran è usato in due modi distinti (Sezione 4.6). Il primo riguarda la superficie della criticità aggregata, calcolata sull'intera griglia, comprese le celle prive di encounter. Su questa superficie l'indice è positivo per tutti gli otto indicatori, con valori compresi fra 0,21 (R_1) e 0,28 (R_3); per R_7 vale $I = 0,267$, con $p = 0,005$ su 199 permutazioni, il valore minimo ottenibile con quel numero di permutazioni, e per R_8 vale 0,277. La criticità aggregata non è quindi distribuita a caso sulla griglia; poiché le celle vuote sono incluse, il valore riflette anche la concentrazione degli incontri nella parte di area effettivamente frequentata.

Sulla superficie di R_7 costruita con la sola unità i , la statistica di Getis-Ord individua 250 celle con $G_i^* > 1,96$ (95 %), di cui 218 con $G_i^* > 2,58$ (99 %) (Fig. 5.9). Le soglie sono applicate cella per cella, senza correzione per confronti multipli. Le celle significative formano pochi nuclei compatti fra le longitudini 9,15° e 9,25° E, presso Camogli e fra Portofino e Santa Margherita Ligure. Gli

hotspot riflettono l'associazione della criticità aggregata con la densità (Sezione 5.1) e indicano aree di concentrazione degli incontri e dei valori di R_7 , non luoghi di collisione.

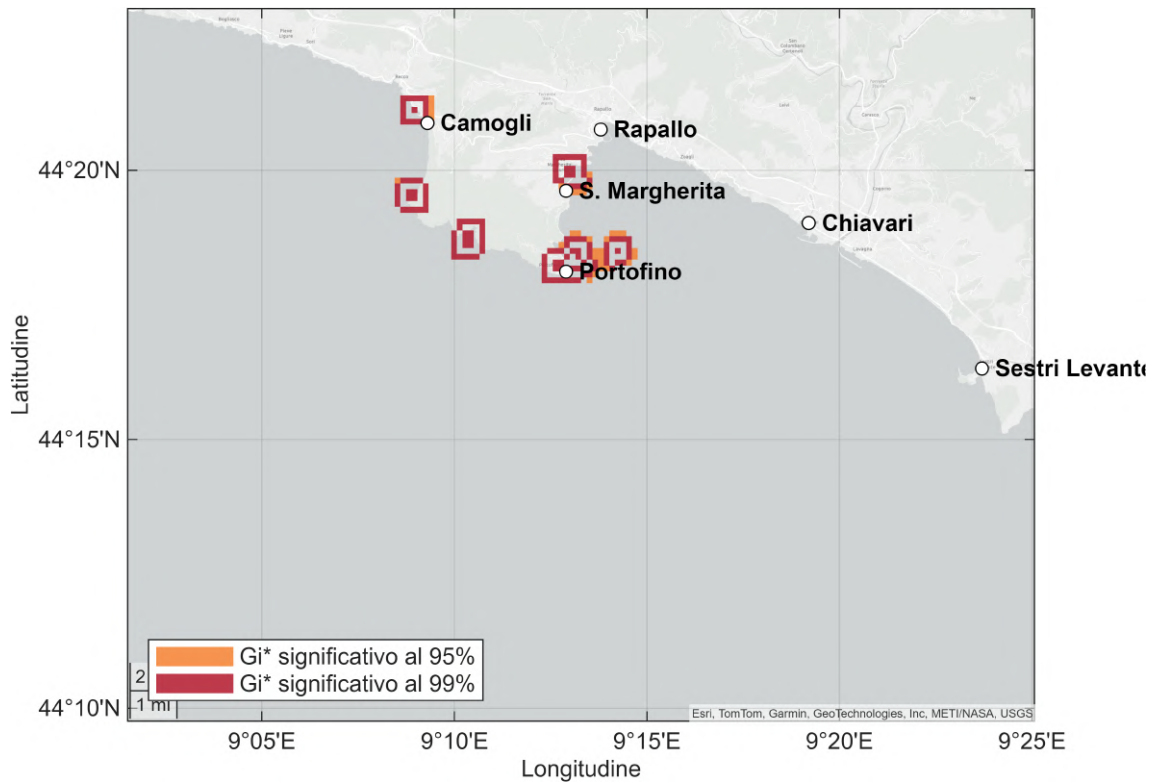


Figura 5.9: Tigullio: celle con concentrazione locale significativa della criticità aggregata di R_7 secondo la statistica di Getis-Ord G_i^* , al 95 % e al 99 %.

Il secondo uso dell'indice riguarda le 297 celle con almeno 3 encounter, sulle quali sono confrontate tre superfici: densità degli encounter ($I = 0,481$), criticità aggregata ($I = 0,419$) e criticità per incontro ($I = 0,487$). Per verificare se la struttura spaziale della criticità per incontro sia riconducibile alla sola distribuzione del traffico, i valori di R_7 sono stati riallocati casualmente fra gli encounter, mantenendo invariata la cella di appartenenza e quindi la superficie di densità. Nelle 199 riallocazioni l'indice della criticità per incontro ha media $-0,0012$ e deviazione standard $0,0328$; il valore osservato se ne discosta di $14,9$ deviazioni standard, con $p = 0,005$ (Fig. 5.10–5.12). La criticità per incontro presenta quindi una struttura spaziale che la sola densità degli encounter non spiega; il test non ne individua la causa. La sua autocorrelazione è di poco superiore a quella della densità ($0,487$ contro $0,481$).

La frazione di encounter di ciascuna cella con $R_7 > 0,219$, soglia che delimita il 5 % degli encounter con R_7 più elevato nell'area, ha con la densità una correlazione di rango $\rho = 0,470$, intermedia fra quelle della criticità aggregata e della criticità per incontro.

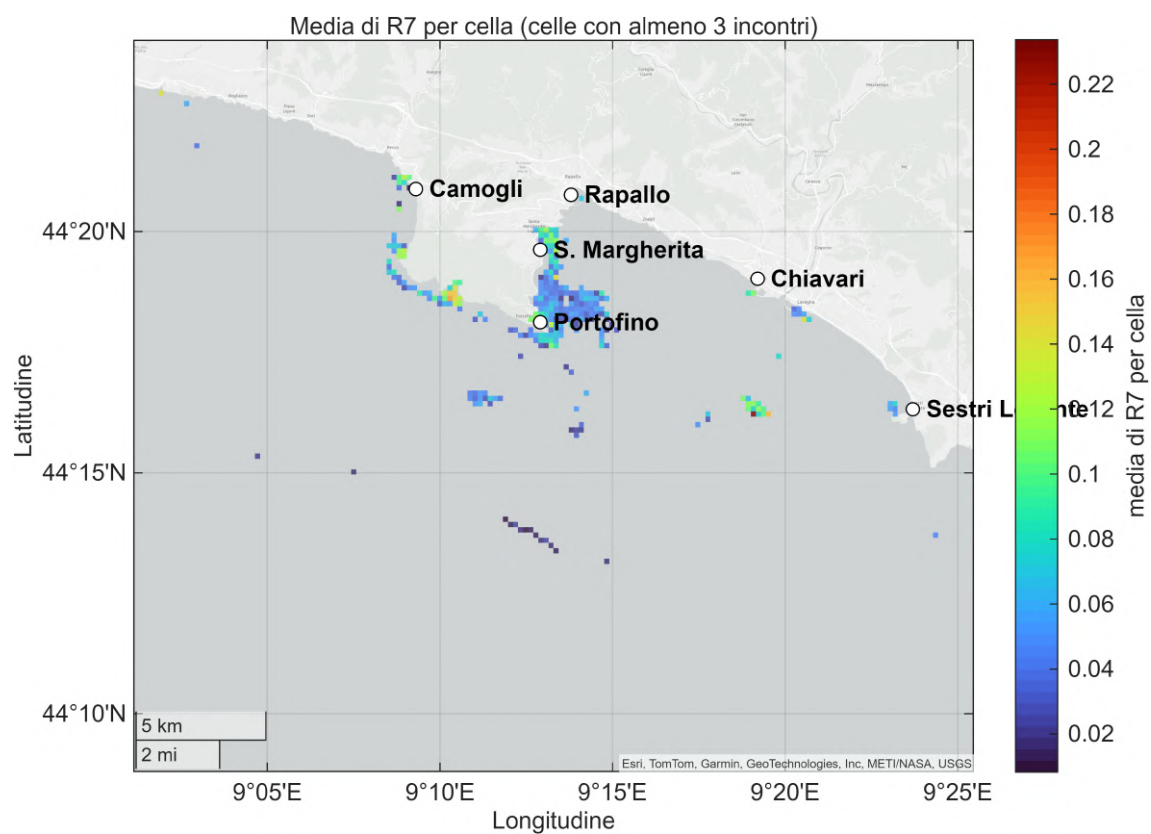


Figura 5.10: Tigullio: criticità per incontro (media di R_7 per cella) sulle celle con almeno 3 encounter.

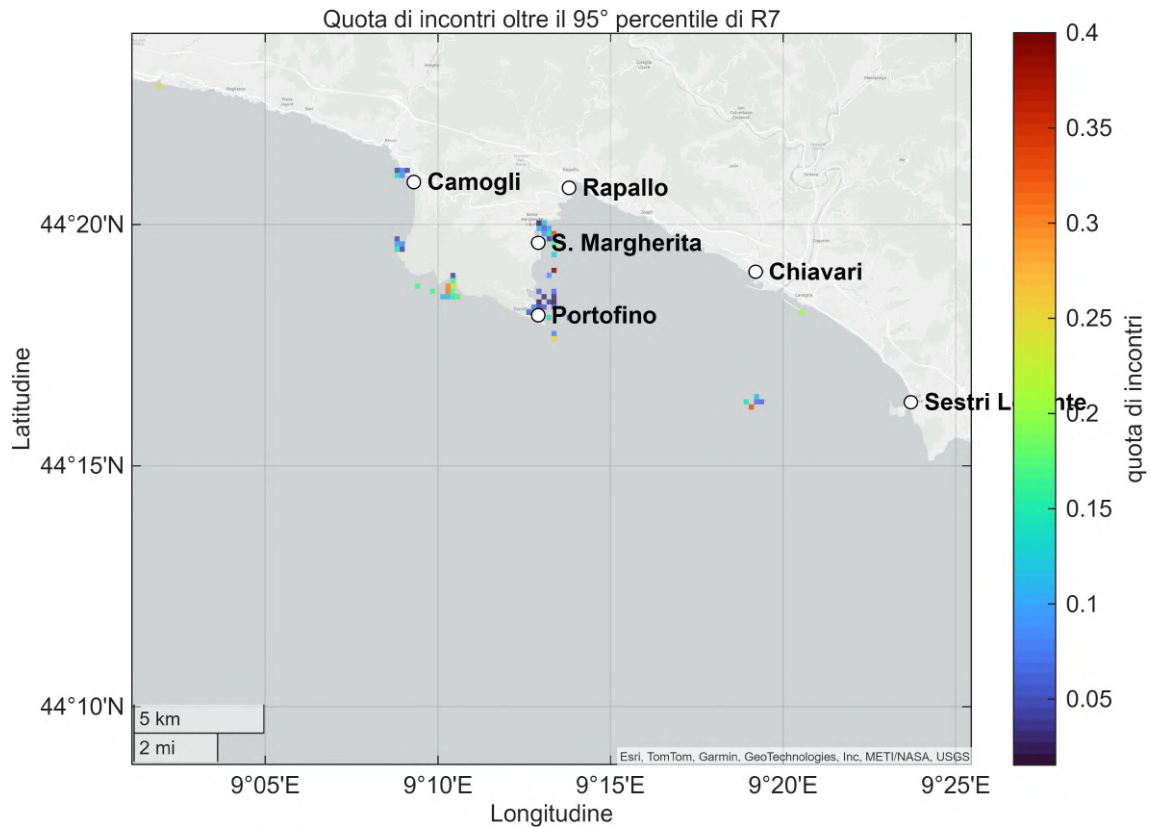


Figura 5.11: Tigullio: frazione di encounter di ciascuna cella con R_7 superiore alla soglia del 95° percentile (0,219).

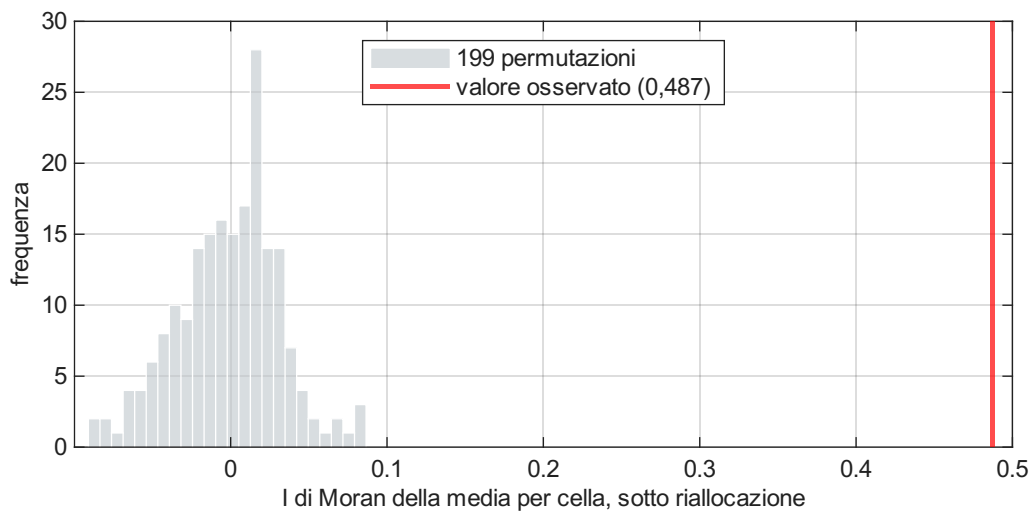


Figura 5.12: Tigullio: indice di Moran della criticità per incontro nelle 199 riallocazioni casuali di R_7 fra gli encounter a densità invariata, e valore osservato (0,487).

5.5 Tigullio: zone di criticità DBSCAN

È stato considerato il 25 % degli encounter con R_7 più elevato ($R_7 \geq 0,1168$), pari a 2 143 encounter. DBSCAN, applicato alle posizioni di minima distanza osservata con $\epsilon = 300$ m e $minPts = 8$

(Sezione 4.8), individua 9 zone e classifica 107 encounter come rumore, il 5,0 % (Fig. 5.13 e 5.14, Tab. 5.1). La zona più numerosa, di 1 053 encounter, il 49,1 % di quelli selezionati, è centrata presso Portofino. Le altre zone si trovano attorno al promontorio, fra Camogli e Santa Margherita Ligure, tranne una, di 35 encounter, al largo di Chiavari.

Tabella 5.1: Tigullio: zone DBSCAN del 25 % degli encounter con R_7 più elevato ($\varepsilon = 300$ m, $\minPts = 8$).

Encounter	% dei selezionati	Lat. [°N]	Lon. [°E]	Posizione
1 053	49,1	44,304	9,213	Portofino
298	13,9	44,312	9,173	circa 3 km a ovest di Portofino
235	11,0	44,333	9,217	Santa Margherita Ligure
190	8,9	44,326	9,149	a sud di Camogli
168	7,8	44,352	9,150	Camogli
35	1,6	44,272	9,320	al largo di Chiavari
31	1,4	44,310	9,239	a est di Portofino
15	0,7	44,301	9,230	a est di Portofino
11	0,5	44,318	9,221	fra Portofino e Santa Margherita Ligure
107	5,0			rumore

Le zone ricadono nei settori in cui è più alta anche la densità complessiva degli encounter. Il risultato indica dove si concentrano gli incontri con R_7 elevato e non separa l'effetto del volume di traffico da quello della criticità per incontro, distinzione affidata alla superficie della media per cella (Sezioni 5.1 e 5.4). Numero ed estensione delle zone dipendono inoltre dai valori di ε e $\minPts$.

Con ε fra 200 e 300 m e $\minPts$ fra 5 e 12 la zona di Portofino comprende fra 1 036 e 1 114 encounter e l'indice di Rand corretto rispetto alla partizione di riferimento è almeno 0,926; le zone sono 8 o 9, tranne con $\varepsilon = 200$ m e $\minPts$ pari a 5, che dà 14 zone. Con $\varepsilon = 400$ m e $\minPts$ pari a 5 o 8 la zona di Portofino si unisce a quella di Santa Margherita Ligure e alle tre zone minori di 31, 15 e 11 encounter: la zona maggiore sale a 1 368 e 1 365 encounter, il suo centroide si sposta di circa 595 m e l'indice scende a 0,673 e 0,677. Con $\minPts$ pari a 12 la fusione non avviene (1 076 encounter, indice 0,970). Con la selezione estesa al 30 % degli encounter con R_7 più elevato ($R_7 \geq 0,104$) le zone sono 7, la zona maggiore comprende 1 343 encounter e l'indice è 0,925; con la selezione ridotta al 20 % ($R_7 \geq 0,1322$) le zone sono 9, la zona maggiore comprende 849 encounter e l'indice è 0,999.

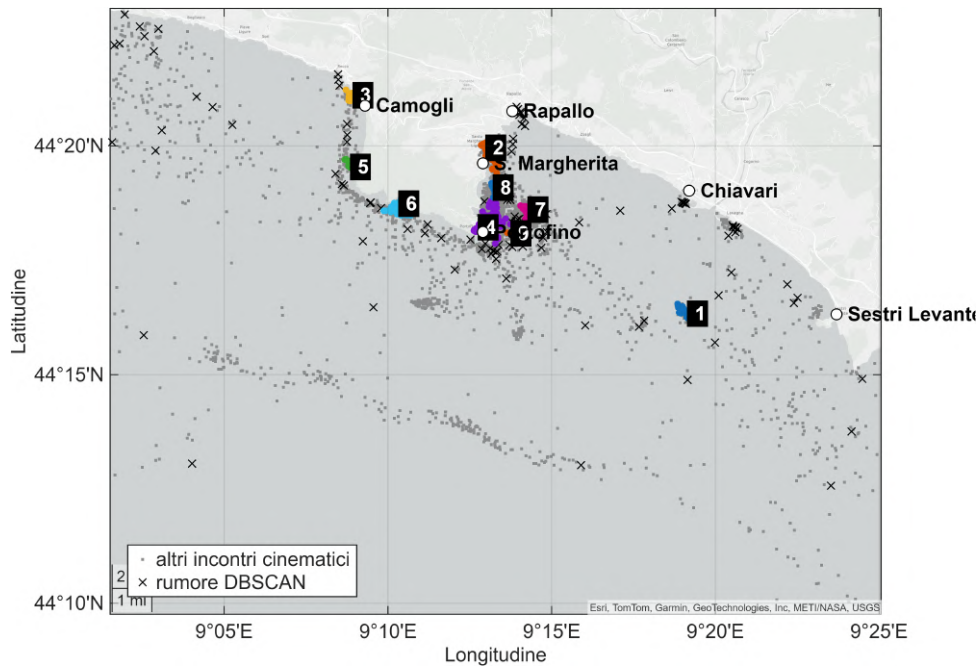


Figura 5.13: Tigullio: zone DBSCAN del 25 % degli encounter con R_7 più elevato ($\varepsilon = 300$ m, $minPts = 8$).

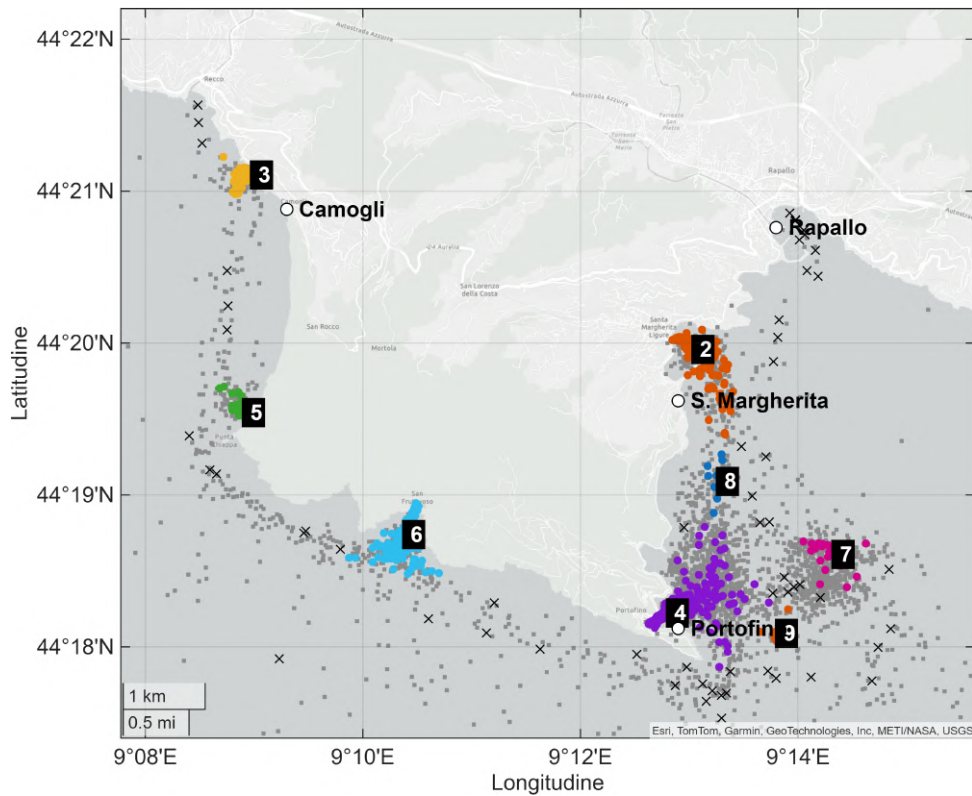
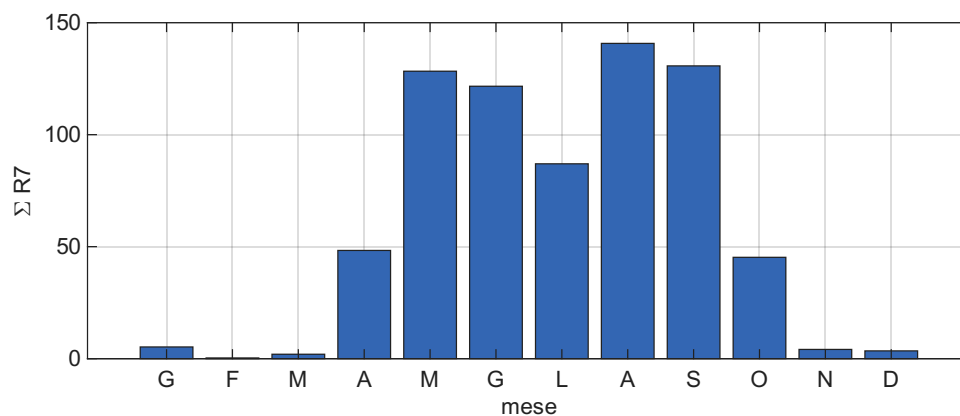


Figura 5.14: Tigullio: zone DBSCAN nella vista di dettaglio del promontorio di Portofino (44,29–44,37° N; 9,13–9,26° E).

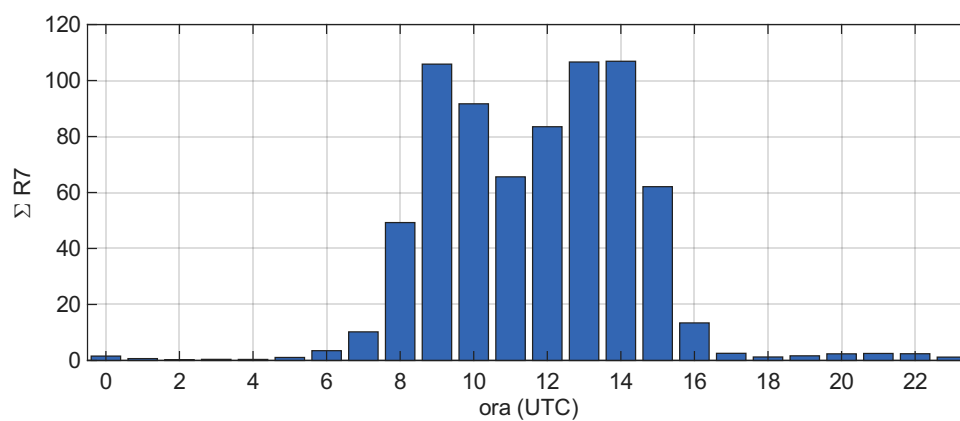
5.6 Tigullio: profilo temporale

Le Fig. 5.15 e 5.16 riportano la somma di R_7 per mese, per ora del giorno e per giorno della settimana, e la matrice mese per ora; gli orari sono in UTC. Il 67 % della somma di R_7 ricade fra giugno e settembre, con il valore mensile massimo ad agosto. Nel giorno i valori più alti cadono alle 14, alle 13 e alle 9, mentre il periodo notturno, fra le 21 e le 06, raccoglie l'1,5 % del totale. Sabato e domenica raccolgono il 46,2 % della somma di R_7 , contro il 28,6 % di una distribuzione uniforme fra i giorni, e anche il venerdì (14,2 %) supera gli altri giorni feriali.

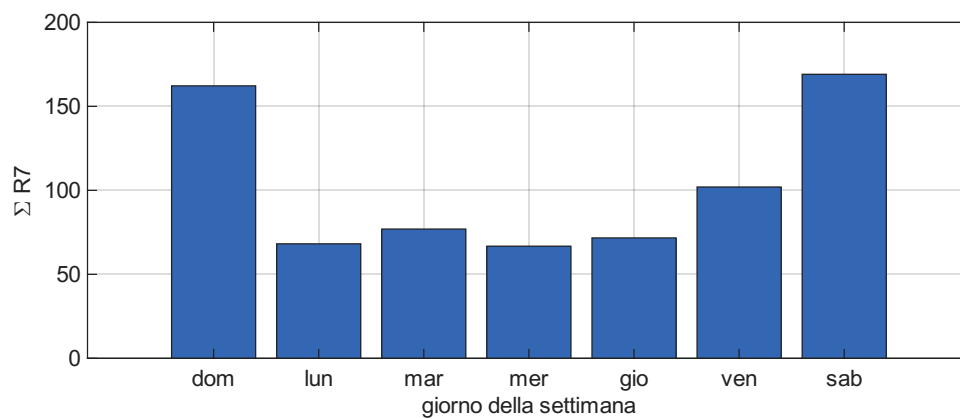
Sul numero di encounter cinematici la quota estiva è del 63 %, quella notturna del 3,4 %, e il mese con più encounter è maggio. La quota estiva di R_7 supera quindi quella degli encounter, e ad agosto la somma di R_7 è massima pur con un numero di encounter inferiore a quello di maggio: nel periodo estivo il valore medio di R_7 per encounter è più alto della media annua, e ad agosto è più alto che a maggio. Nel periodo notturno la quota di R_7 è inferiore a quella degli encounter. La concentrazione stagionale, diurna e nei giorni festivi è compatibile con un traffico prevalentemente diportistico.



(a) Per mese.



(b) Per ora del giorno (UTC).



(c) Per giorno della settimana.

Figura 5.15: Tigullio: profilo temporale della criticità aggregata, somma di R_7 .

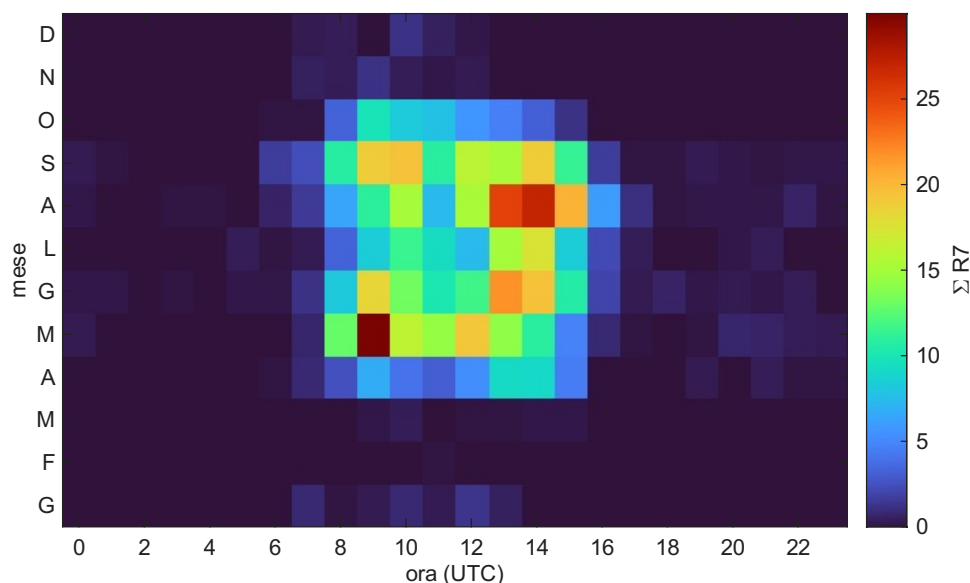


Figura 5.16: Tigullio: somma di R_7 per mese e ora del giorno (UTC).

5.7 Tigullio: verifica euristica di conformità COLREG

La verifica (Sezione 4.9) è applicabile a 5 916 degli 8 573 encounter cinematici; i restanti 2 657 hanno meno di quattro record e non ricevono una classificazione regolamentare. Fra i valutabili la procedura riconosce 1 822 frontali, 2 771 incroci e 1 323 sorpassi. I tassi di conformità, calcolati sui soli casi con esito determinato (Fig. 5.17), sono il 71,3 % nel sorpasso (652 su 915), il 60,1 % nell'incrocio (832 su 1 384) e il 20,3 % nel frontale (179 su 882). Gli esiti indeterminati sono 408 nel sorpasso, 1 387 nell'incrocio e 940 nel frontale.

Questi totali non coincidono con la composizione della Sezione 5.3. Quella classificazione assegna tutti gli 8 573 encounter in base a $|\Delta COG|$ nel record di DCPA stimata minima e ha scopo descrittivo; la verifica usa il rilevamento relativo nella fase di approccio e la sequenza temporale dei record, richiede almeno quattro record e attribuisce il ruolo previsto dalle Regole. I due criteri hanno finalità diverse e i rispettivi totali non devono coincidere.

Un esito non conforme indica che nella traccia non è rilevabile la manovra richiesta dal criterio adottato: nell'incrocio e nel sorpasso una riduzione di velocità o un'accostata a dritta dell'unità tenuta a manovrare, nel frontale un'accostata a dritta di entrambe le unità. Il criterio richiede l'accostata a dritta anche nel sorpasso, per il quale la Regola 13 non prescrive il lato di passaggio: un'accostata a sinistra senza riduzione di velocità è classificata non conforme, e il tasso del sorpasso è quindi una stima per difetto.

Il tasso del frontale non è direttamente confrontabile con quelli degli altri tipi. Nel frontale la manovra deve essere rilevata su entrambe le tracce e la sola riduzione di velocità non è considerata; l'esito è quindi più sensibile alla qualità del dato AIS e all'assenza di una verifica del rischio di collisione ai sensi della Regola 7 all'inizio dell'incontro.

Gli esiti indeterminati hanno due origini: tracce con quattro o cinque record, sufficienti per classificare l'incontro ma non per valutare la manovra, che richiede almeno sei record; e, nel solo incrocio, configurazioni in cui nessuna delle due unità ha l'altra nel proprio settore prodiero di dritta.

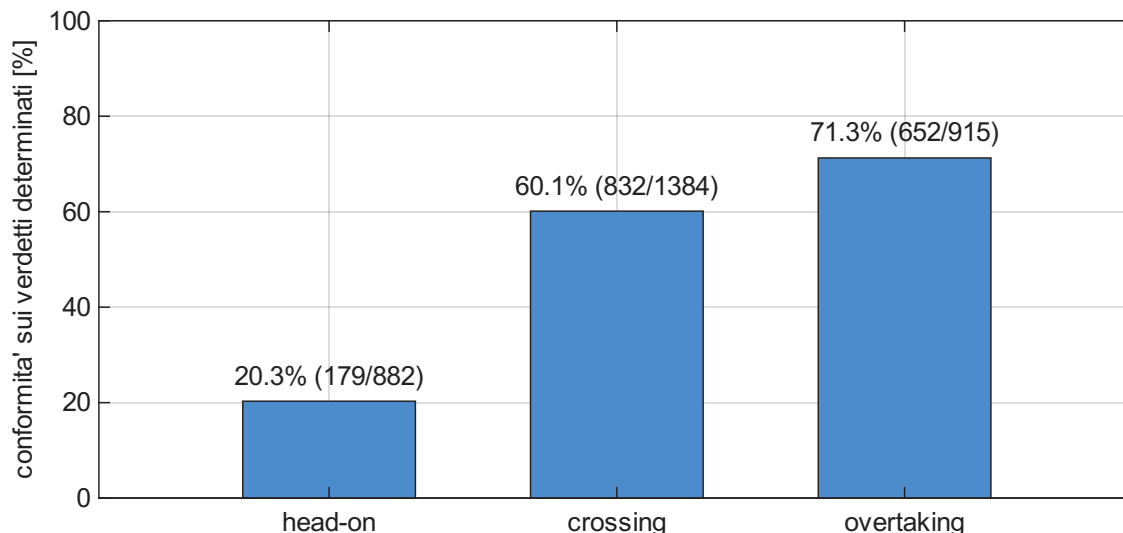


Figura 5.17: Tigullio: tassi di conformità stimati con la verifica euristica COLREG, calcolati sui soli casi con esito determinato.

5.8 Stretto di Messina: risultati dei tre anni

Lo Stretto di Messina è analizzato con la stessa procedura e con gli stessi parametri del Tigullio, separatamente per il 2022, il 2023 e il 2024, su 85 018, 83 494 e 69 241 encounter cinematici rispettivamente. L'indicatore principale è R_8 ; per il confronto con il Tigullio è riportato anche R_7 . Le figure riguardano il 2023; quelle del 2022 e del 2024 sono riportate in Appendice A.2. Le sottosezioni commentano le differenze rispetto al Tigullio e fra i tre anni.

La composizione degli encounter cinematici è diversa da quella del Tigullio (Sezione 3.7.2). Gli encounter in cui una sola unità ha velocità mediana inferiore a 1 nodo sono l'84,8, l'82,9 e l'81,6 % nei tre anni e raccolgono l'83,1, l'80,9 e l'80,4 % della somma di R_8 ; fra il 25 % degli encounter con R_8 più elevato sono il 77,6, il 75,4 e il 76,4 %. Nel Tigullio le stesse quote sono il 10,7, il 5,8 e il 2,3 %. Nello Stretto la criticità riguarda quindi in prevalenza incontri fra un'unità ferma e un'unità in moto. Un esempio è l'unità CP404, un pattugliatore della Guardia Costiera di 30 m con base a Messina¹, ferma nel 98,8 % dei suoi 8 261 encounter del 2023, con l'altra unità a circa 8 nodi: da sola raccoglie il 9,8 % della somma di R_8 dell'anno.

5.8.1 Mappe globali e rapporto criticità-densità

Le mappe di criticità aggregata del 2023 (Fig. 5.18 e 5.19) seguono il tratto centrale dello Stretto e le fasce costiere a nord e a sud. Per tutti gli indicatori i valori più alti si concentrano presso

¹Pattugliatore "Gaetano Magliano" del Corpo delle Capitanerie di porto, impiegato nella vigilanza sulla pesca.

Messina, Villa San Giovanni e Reggio Calabria. La cella con la criticità aggregata massima di R_8 si trova nella rada di Messina nel 2022 e nel 2023 e presso Reggio Calabria nel 2024. La Fig. 5.20 riporta la criticità aggregata di R_8 nella vista di dettaglio dello Stretto. Le correlazioni dell'indicatore principale con gli altri indicatori sono più basse che nel Tigullio: $\rho(R_1, R_8) = 0,41$ e $\rho(R_6, R_8) = 0,71$, contro $\rho(R_1, R_7) = 0,58$ e $\rho(R_6, R_7) = 0,81$. Con R_7 la correlazione con R_1 è invece più alta che nel Tigullio (0,69): la differenza dipende dalla lunghezza della coppia inclusa in R_8 (Sezione 4.4.1).

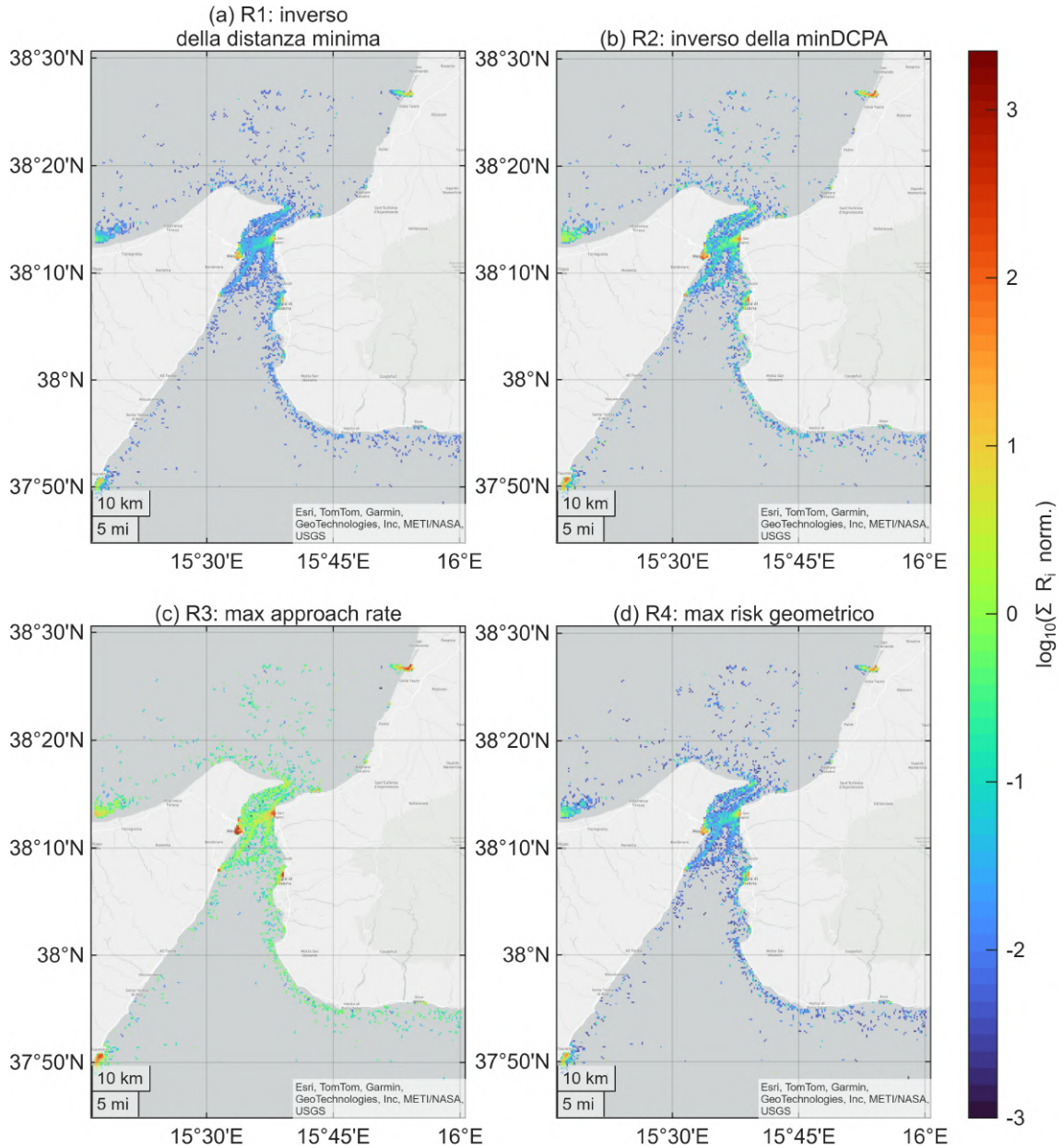


Figura 5.18: Stretto di Messina, 2023: mappe di criticità aggregata per gli indicatori R_1 – R_4 . Somma per cella dell'indicatore normalizzato, in scala logaritmica.

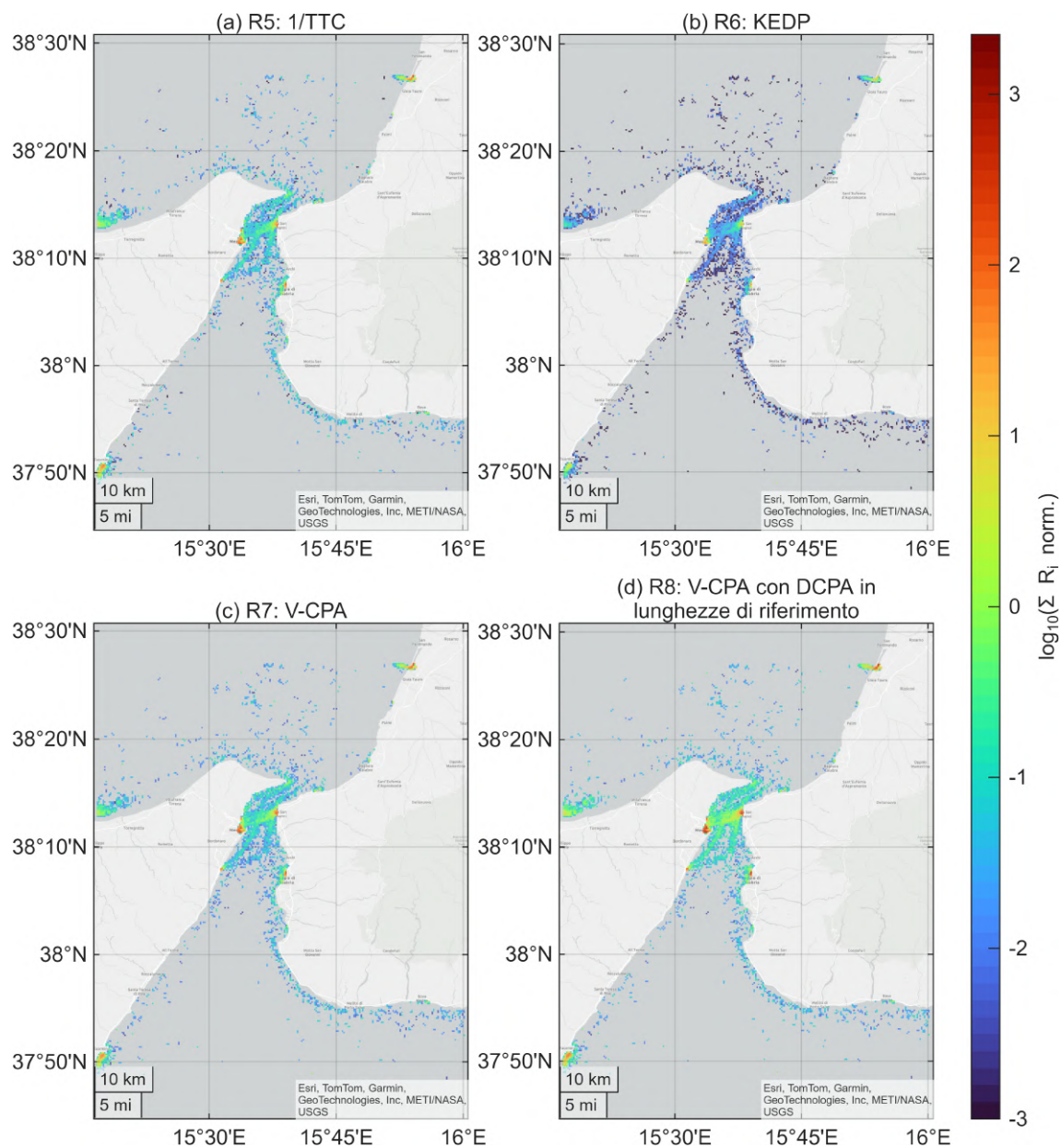


Figura 5.19: Stretto di Messina, 2023: mappe di criticità aggregata per gli indicatori velocity-aware R_5 – R_8 . Somma per cella dell'indicatore normalizzato, in scala logaritmica.

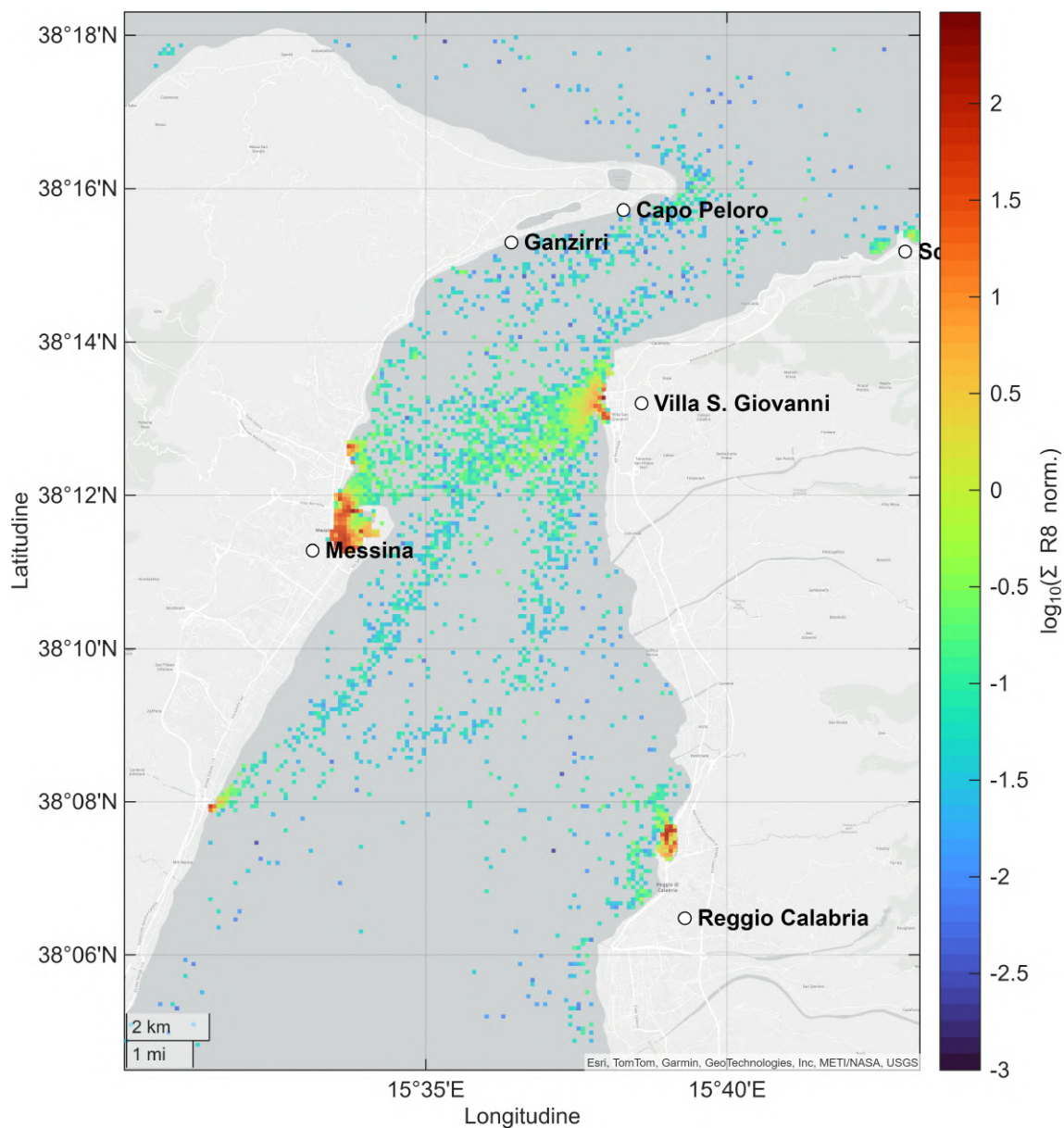


Figura 5.20: Stretto di Messina, 2023: criticità aggregata di R_8 nella vista di dettaglio dello Stretto (38,08–38,30° N; 15,50–15,72° E), su celle di circa 100 m, in scala logaritmica.

Sulle 633 celle con almeno 3 encounter del 2023 (Fig. 5.21–5.23) la correlazione di rango con la densità è $\rho = 0,902$ per la criticità aggregata e $\rho = 0,303$ per la criticità per incontro; con R_7 i valori sono 0,867 e 0,328. Nel Tigullio sono 0,890 e 0,317: in entrambe le aree la criticità aggregata segue la densità e la criticità per incontro ne dipende in misura molto minore, con valori simili. Nei tre anni la correlazione è compresa fra 0,867 e 0,902 per la criticità aggregata e fra 0,242 e 0,303 per la criticità per incontro (Tab. 5.2).

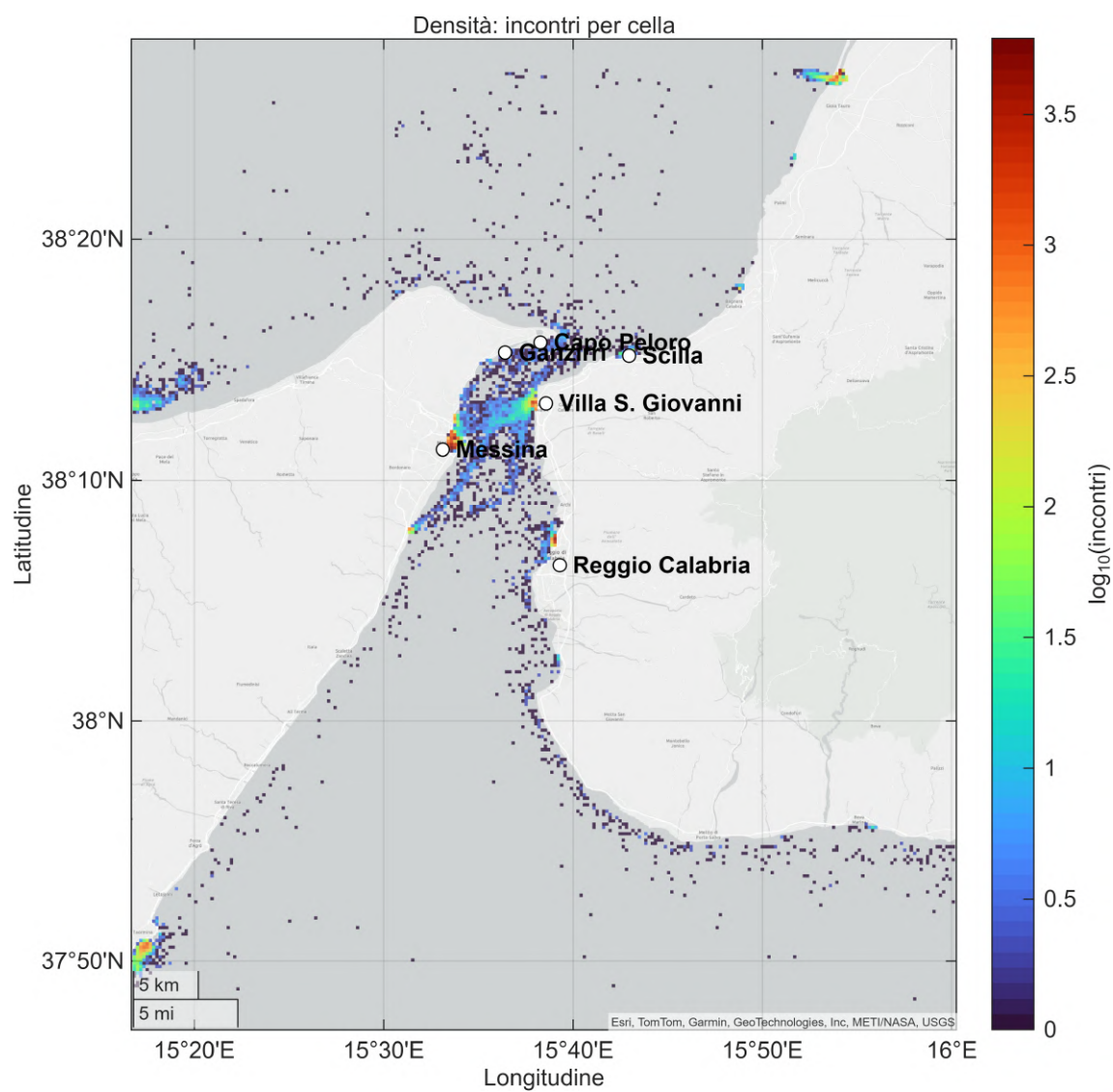


Figura 5.21: Stretto di Messina, 2023: densità degli encounter cinematici per cella.

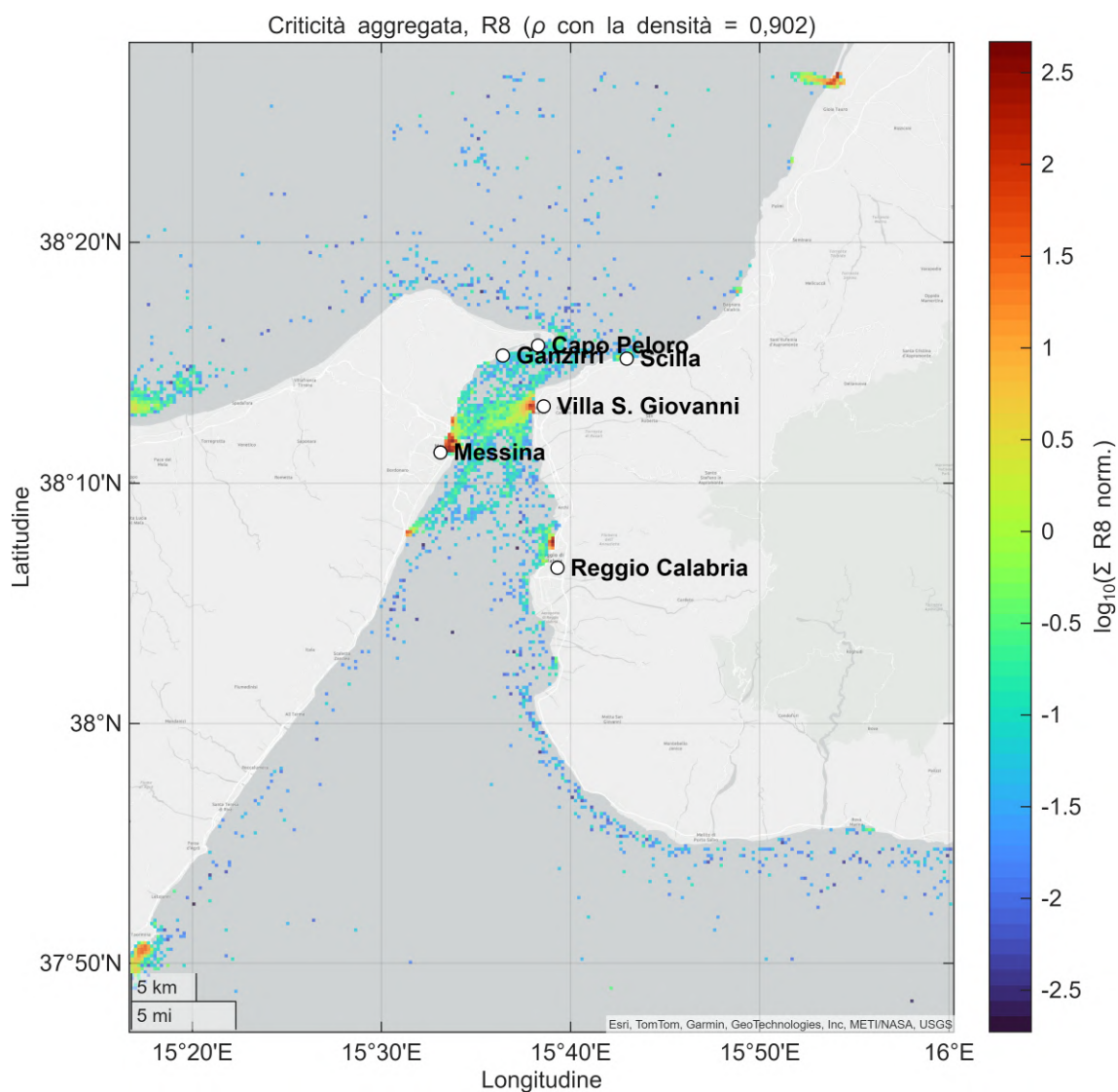


Figura 5.22: Stretto di Messina, 2023: critic   aggregata (somma di R_8 per cella), con la correlazione di Spearman con la densit   calcolata sulle celle con almeno 3 encounter.

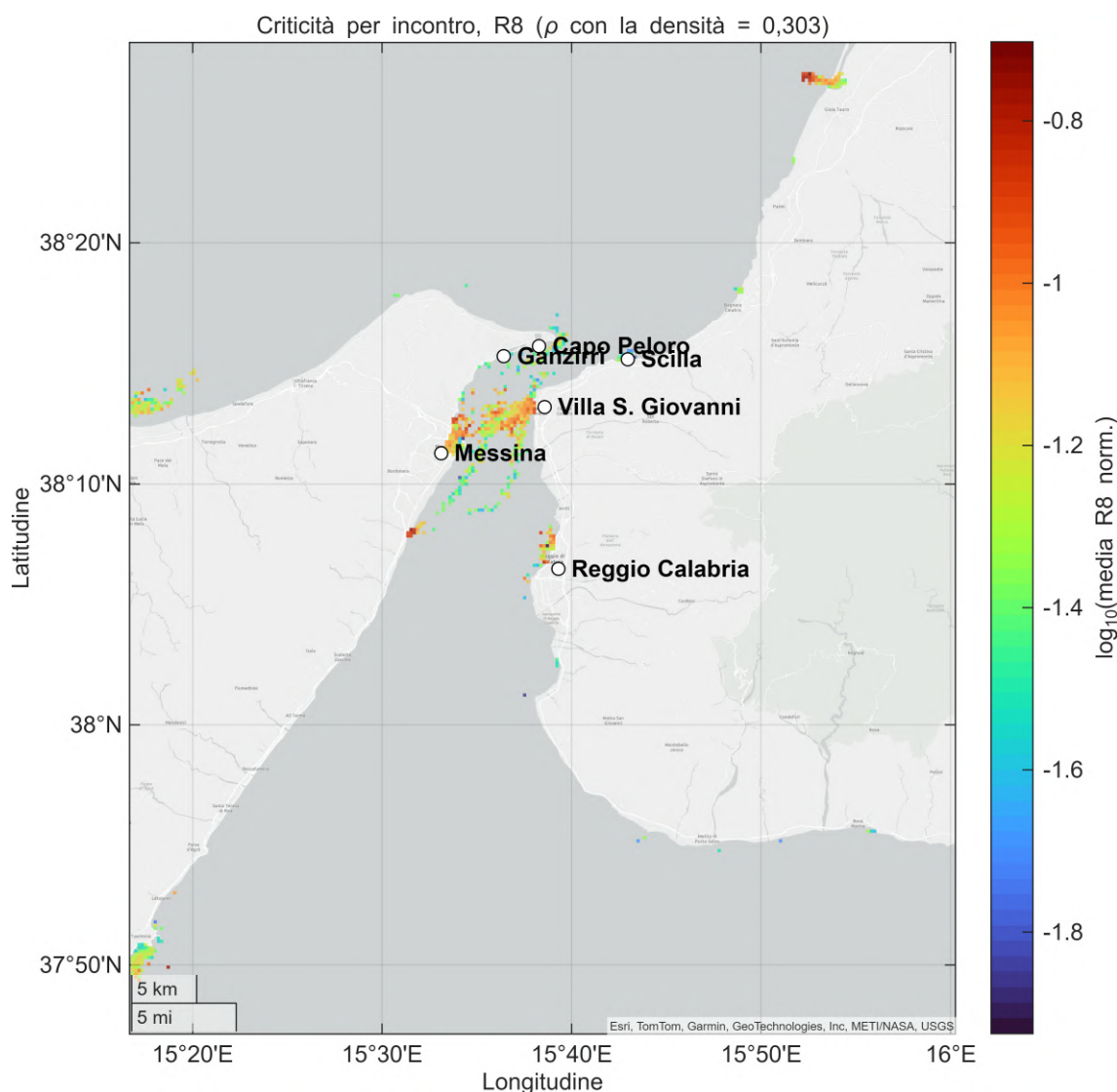


Figura 5.23: Stretto di Messina, 2023: criticità per incontro (media di R_8 per cella), con la correlazione di Spearman con la densità calcolata sulle celle con almeno 3 encounter.

5.8.2 Composizione per macro-classe

Nello Stretto tutte le unità coinvolte sono presenti nell'archivio statico (Tab. 3.9), e la disaggregazione non è limitata a un sottoinsieme come nel Tigullio. Le cinque coppie più numerose del 2023 (Fig. 5.24) sono other–passenger (22 629 encounter), other–other (14 701), passenger–passenger (13 329), cargo–other (9 551) e pleasure–pleasure (9 038), per complessivi 69 248 degli 83 494 encounter. Prevalgono le coppie con unità passeggeri e *other*, in relazione al servizio di traghetti (Sezione 3.7.3). Nei tre anni la quota delle coppie other–passenger scende dal 35,6 al 22,2 % e quella delle coppie pleasure–pleasure sale dal 5,7 al 12,0 %.

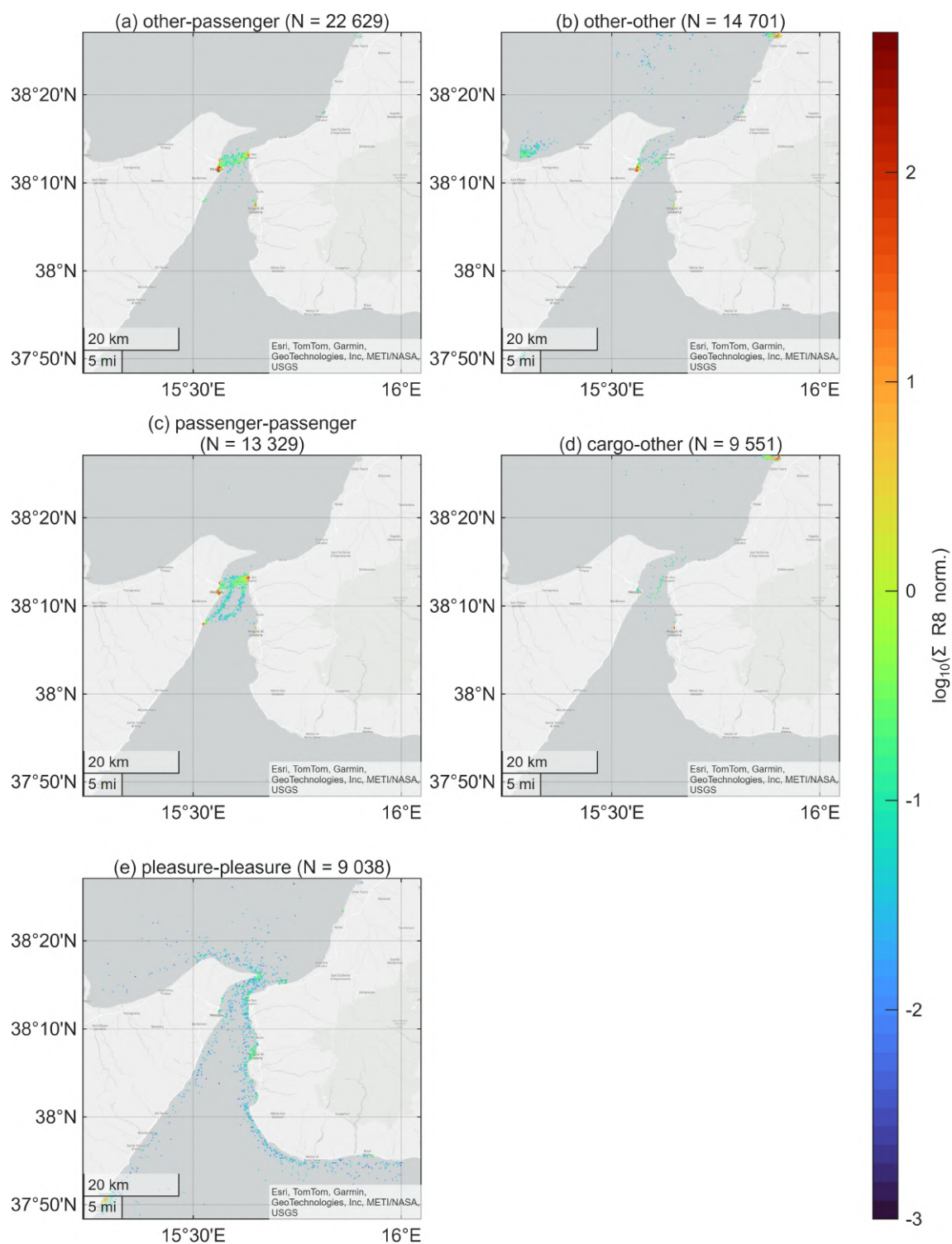


Figura 5.24: Stretto di Messina, 2023: criticità aggregata di R_8 per le cinque coppie di macro-classi più numerose, in scala logaritmica.

5.8.3 Tipologia degli encounter

Nel 2023 la classificazione per tipo assegna 42 898 encounter all'incrocio (51,4 %), 26 220 al sorpasso (31,4 %) e 14 376 al frontale (17,2 %) (Fig. 5.25). Rispetto al Tigullio (40,9, 32,8 e

26,3 %) l'ordinamento è lo stesso: aumenta la quota di incroci e si riduce quella dei frontali. La composizione è stabile nei tre anni (Tab. 5.5): l'incrocio è fra il 48,8 e il 51,6 %, il sorpasso fra il 30,0 e il 33,4 %, il frontale fra il 17,2 e il 18,4 %. Nelle mappe il frontale e l'incrocio presentano i valori più alti in pochi punti, presso Messina, Villa San Giovanni e Reggio Calabria, mentre il sorpasso si distribuisce lungo il corridoio centrale dello Stretto e lungo le fasce costiere. La quota più alta di incroci è compatibile con l'attraversamento del corridoio di transito da parte delle rotte dei traghetti.

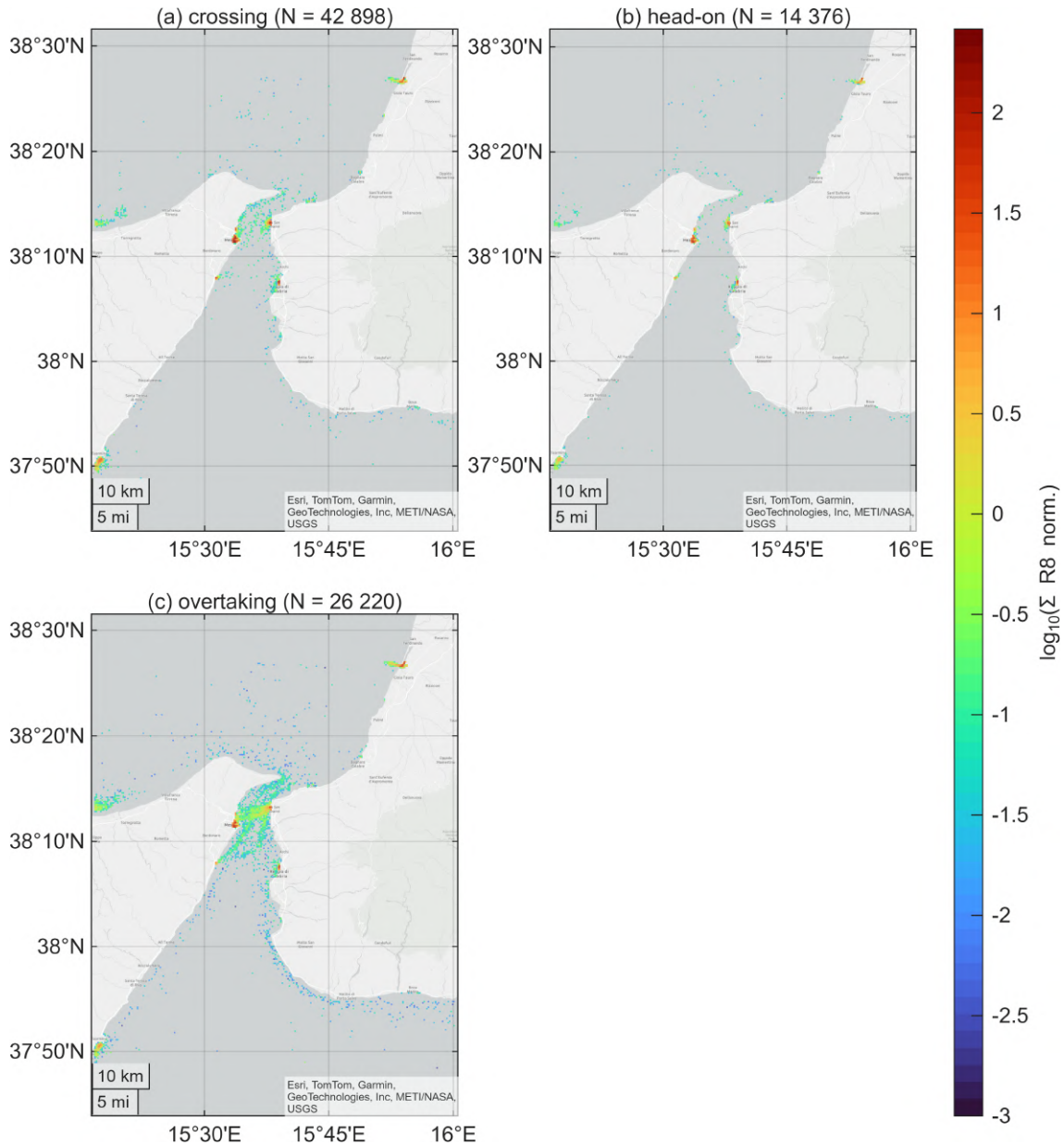


Figura 5.25: Stretto di Messina, 2023: mappe di criticità aggregata di R_8 per tipo di incontro (incrocio, frontale, sorpasso), con classificazione basata su $|\Delta COG|$, in scala logaritmica.

5.8.4 Struttura spaziale e hotspot

Nel 2023 l'indice di Moran sulla superficie della criticità aggregata è positivo per tutti gli otto indicatori, con valori fra 0,28 (R_5) e 0,36 (R_8); per R_8 vale $I = 0,363$ e per R_7 0,334, con $p = 0,005$. La statistica di Getis-Ord su R_8 individua 337 celle con $G_i^* > 1,96$, di cui 312 con $G_i^* > 2,58$ (Fig. 5.26). Le celle significative formano pochi nuclei, i principali presso Messina, Villa San Giovanni e Reggio Calabria. L'indice di Moran di R_7 è più alto che nel Tigullio (0,334 contro 0,267), ma le due griglie hanno risoluzione diversa, circa 250 m contro circa 200 m, e il confronto fra le aree ha valore indicativo. I valori dei tre anni sono riportati in Tab. 5.2.

Sulle 633 celle con almeno 3 encounter del 2023, l'indice di Moran vale 0,600 per la densità, 0,568 per la criticità aggregata e 0,565 per la criticità per incontro. Nelle 199 riallocazioni casuali dei valori di R_8 a densità invariata l'indice della criticità per incontro ha media $-0,0009$ e deviazione standard 0,0250; il valore osservato se ne discosta di 22,7 deviazioni standard, con $p = 0,005$ (Fig. 5.27–5.29). Lo scarto è di 29,5 deviazioni standard nel 2022 e di 19,7 nel 2024, sempre con $p = 0,005$: in tutti e tre gli anni la struttura spaziale della criticità per incontro non è riconducibile alla sola distribuzione del traffico. Il confronto con l'autocorrelazione della densità non dà invece un risultato stabile: l'indice della criticità per incontro è superiore a quello della densità nel 2022 (0,634 contro 0,497) e inferiore nel 2023 (0,565 contro 0,600) e nel 2024 (0,487 contro 0,572).

La frazione di encounter oltre la soglia del 95° percentile di R_8 (0,165 nel 2023) ha con la densità una correlazione di rango $\rho = 0,469$, intermedia fra quelle della criticità aggregata e della criticità per incontro, come nel Tigullio. R_7 e R_8 hanno scale diverse e le rispettive soglie non sono confrontabili; per R_7 la soglia vale 0,085 nello Stretto e 0,219 nel Tigullio.

Tabella 5.2: Stretto di Messina: rapporto con la densità e struttura spaziale della criticità (R_8) nei tre anni.

	2022	2023	2024
ρ con la densità: criticità aggregata	0,890	0,902	0,867
ρ con la densità: criticità per incontro	0,274	0,303	0,242
Moran I di R_8 (intervallo R_1 – R_8)	0,280 (0,21–0,28)	0,363 (0,28–0,36)	0,350 (0,25–0,35)
Celle G_i^* al 95 % / al 99 %	285 / 256	337 / 312	347 / 324
Moran I (celle con almeno 3 encounter):	0,497	0,600	0,572
densità			
criticità aggregata	0,464	0,568	0,542
criticità per incontro	0,634	0,565	0,487
Test di riallocazione: scarto (dev. st.)	29,5	22,7	19,7
Soglia del 95° percentile di R_8	0,163	0,165	0,174

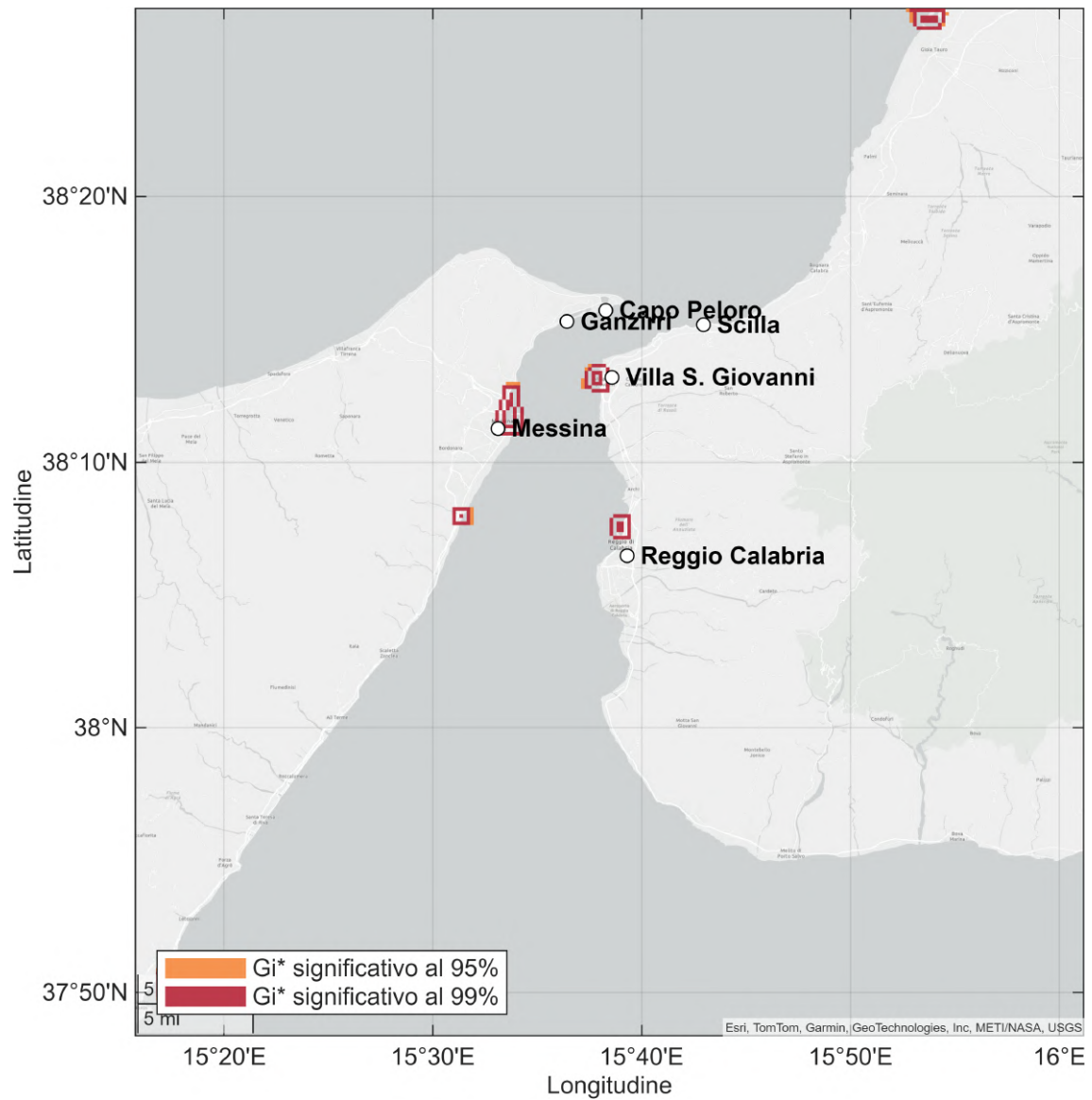


Figura 5.26: Stretto di Messina, 2023: celle con concentrazione locale significativa della criticità aggregata di R_8 secondo la statistica di Getis-Ord G_i^* , al 95 % e al 99 %.

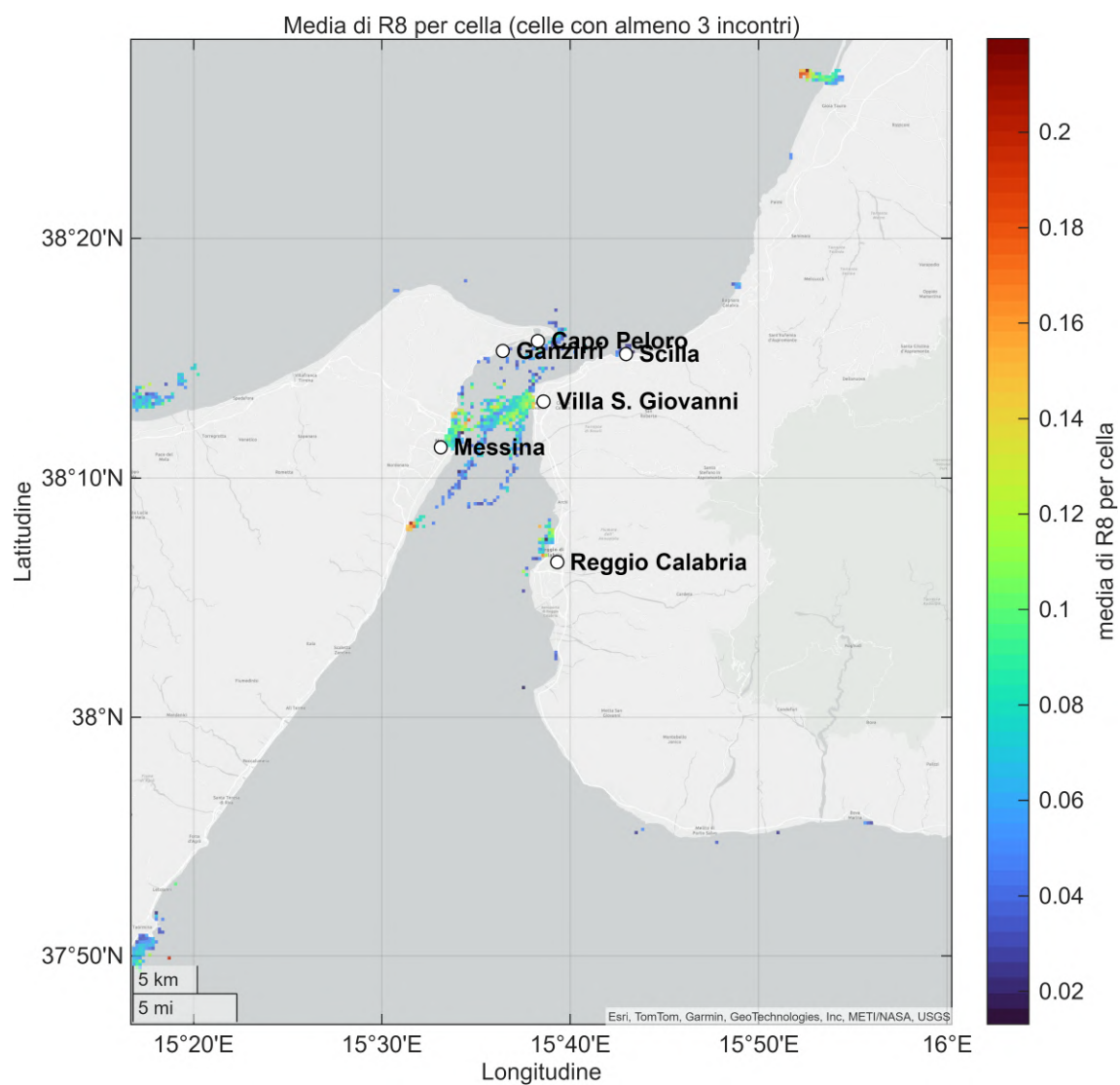


Figura 5.27: Stretto di Messina, 2023: criticità per incontro (media di R_8 per cella) sulle celle con almeno 3 encounter.

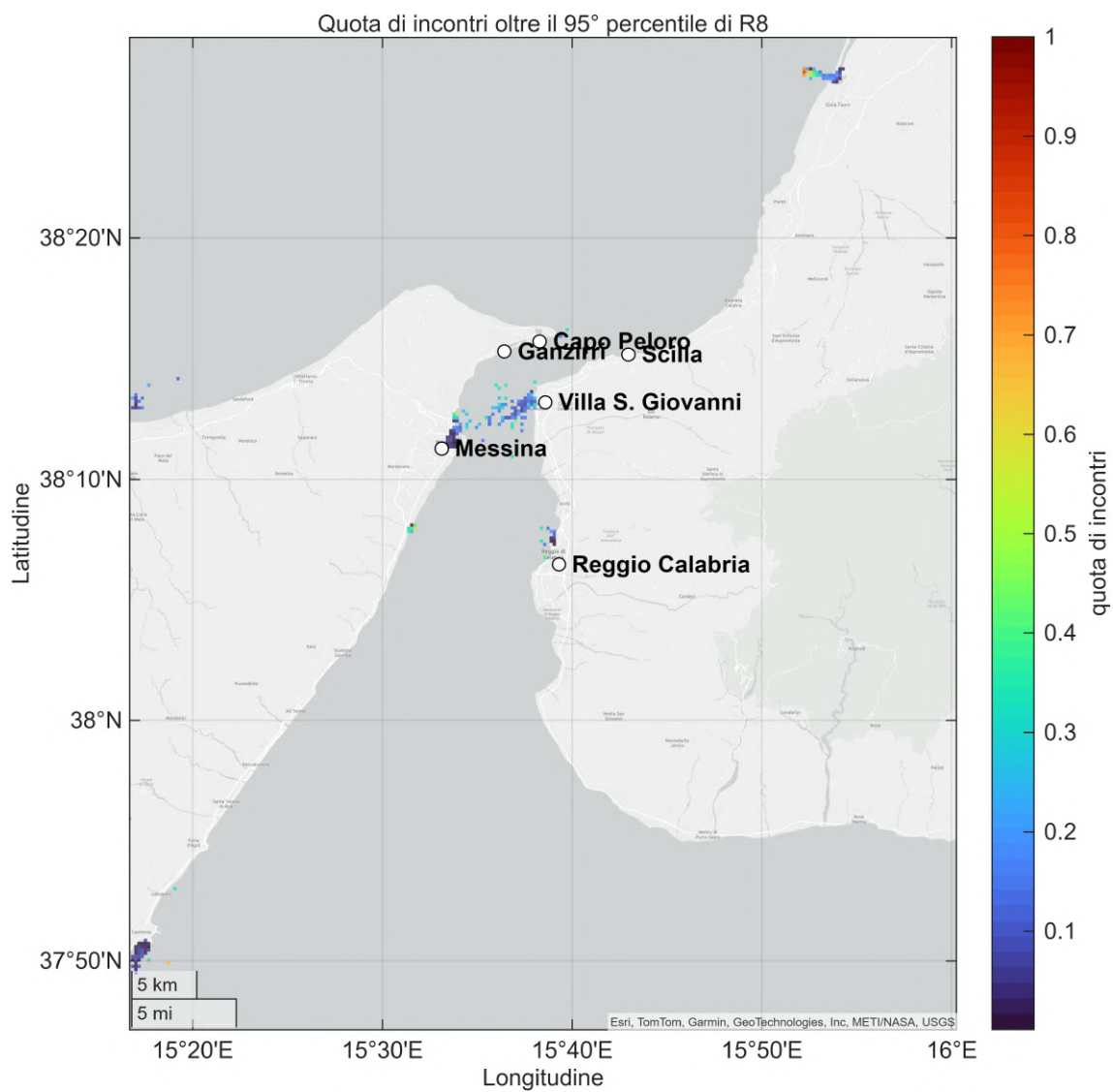


Figura 5.28: Stretto di Messina, 2023: frazione di encounter di ciascuna cella con R_8 superiore alla soglia del 95° percentile (0,165).

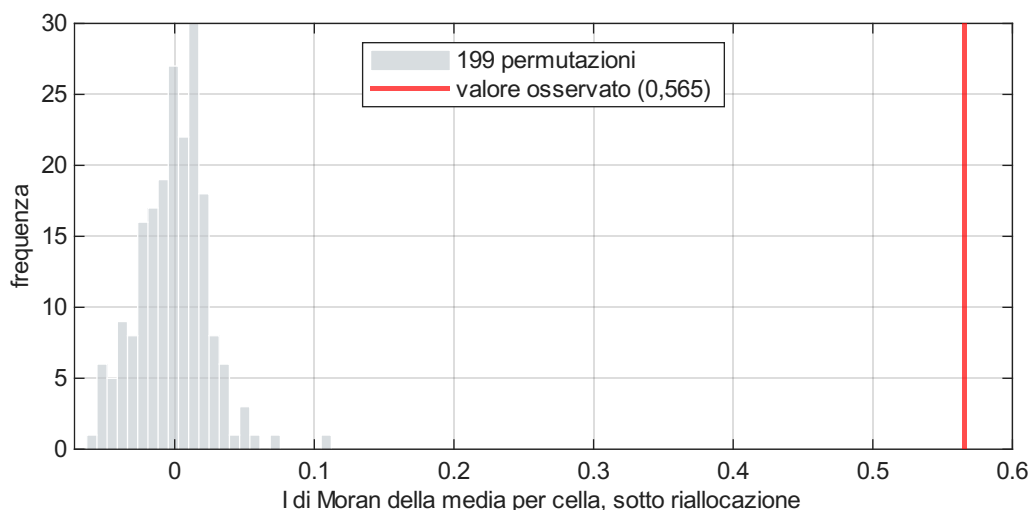


Figura 5.29: Stretto di Messina, 2023: indice di Moran della criticità per incontro nelle 199 riallocazioni casuali di R_8 fra gli encounter a densità invariata, e valore osservato (0,565).

5.8.5 Zone di criticità DBSCAN

Nel 2023 il 25 % degli encounter con R_8 più elevato ($R_8 \geq 0,1077$) comprende 20 874 encounter. Con gli stessi parametri del Tigullio DBSCAN individua 11 zone e classifica 310 encounter come rumore, l'1,5 %; i restanti 20 564 sono assegnati a una zona (Fig. 5.30, Tab. 5.3). La zona maggiore, nella rada di Messina, raccoglie il 42,1 % degli encounter selezionati e ha un'estensione di 698 m. Le zone principali si trovano in aree portuali e di approdo: rada e porto di Messina, Villa San Giovanni, Reggio Calabria, Gioia Tauro, Tremestieri e Giardini Naxos–Taormina. Sette delle 11 zone hanno il centroide nella vista di dettaglio dello Stretto (Fig. 5.31). Il numero di zone non è direttamente confrontabile con le 9 del Tigullio: con ϵ e $minPts$ fissati in valore assoluto, esso dipende anche dal numero di encounter selezionati (20 874 contro 2 143) e dall'estensione dell'area. Come nel Tigullio, le zone indicano concentrazioni di incontri con valori elevati dell'indicatore e non luoghi di collisione.

Tabella 5.3: Stretto di Messina, 2023: zone DBSCAN del 25 % degli encounter con R_8 più elevato ($\epsilon = 300$ m, $minPts = 8$).

Encounter	% dei selezionati	Lat. [°N]	Lon. [°E]	Posizione
8 798	42,1	38,194	15,561	rada di Messina
3 464	16,6	38,220	15,631	Villa San Giovanni
3 362	16,1	38,127	15,651	Reggio Calabria
3 118	14,9	38,446	15,897	porto di Gioia Tauro
1 000	4,8	38,132	15,524	porto di Tremestieri
693	3,3	37,840	15,288	Giardini Naxos–Taormina
73	0,3	38,219	15,286	a est del porto di Milazzo
56	0,3			quattro zone di 21, 14, 13 e 8 encounter
310	1,5			rumore

Con la selezione estesa al 30 % e ridotta al 20 % degli encounter con R_8 più elevato si ottengono 12 e 11 zone, con una zona maggiore di 10 900 e 6 795 encounter. In tutte le combinazioni di ε e minPts tranne una, la zona maggiore comprende fra 7 924 e 8 814 encounter e l'indice di Rand corretto rispetto alla partizione di riferimento è almeno 0,905. Con $\varepsilon = 400$ m e minPts pari a 5 la zona della rada di Messina si unisce a quella di Villa San Giovanni: la zona maggiore sale a 12 351 encounter e l'indice scende a 0,678.

Nei tre anni le zone si ripetono negli stessi punti (Fig. 5.32): i centroidi delle zone della rada di Messina, di Villa San Giovanni, di Reggio Calabria, di Gioia Tauro, di Tremestieri, di Giardini Naxos–Taormina e della piccola zona a est del porto di Milazzo differiscono di circa 100 m al massimo da un anno all'altro. È il risultato più stabile del confronto fra gli anni. Varia invece il peso relativo delle zone: la zona della rada di Messina raccoglie il 63,7, il 42,1 e il 34,6 % degli encounter selezionati nel 2022, 2023 e 2024 (Tab. 5.4), quella di Giardini Naxos–Taormina 205, 693 e 532 encounter. Il numero di righe dei CSV cresce fra gli anni più del numero di encounter (Tab. 3.6); finché non è esclusa una variazione nella frequenza di registrazione dei dati, le differenze fra gli anni non sono interpretate come variazioni del traffico.

Tabella 5.4: Stretto di Messina: zone DBSCAN del 25 % degli encounter con R_8 più elevato nei tre anni.

	2022	2023	2024
Soglia di R_8	0,1068	0,1077	0,1162
Encounter selezionati	21 255	20 874	17 310
Zone DBSCAN / rumore	12 / 1,3 %	11 / 1,5 %	11 / 1,4 %
Zona della rada di Messina	13 549 (63,7 %)	8 798 (42,1 %)	5 996 (34,6 %)

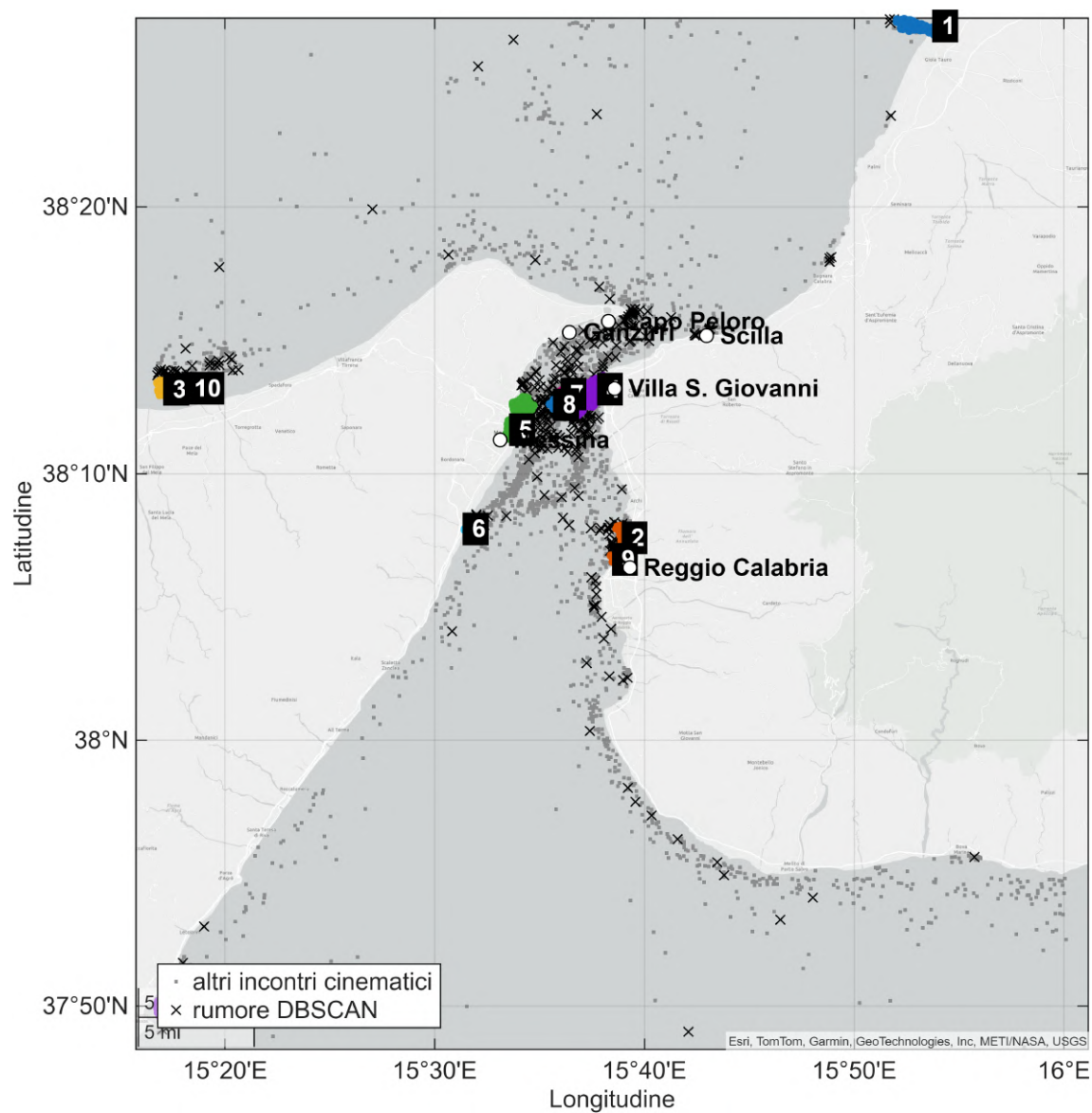


Figura 5.30: Stretto di Messina, 2023: zone DBSCAN del 25 % degli encounter con R_8 più elevato ($\epsilon = 300$ m, $minPts = 8$).

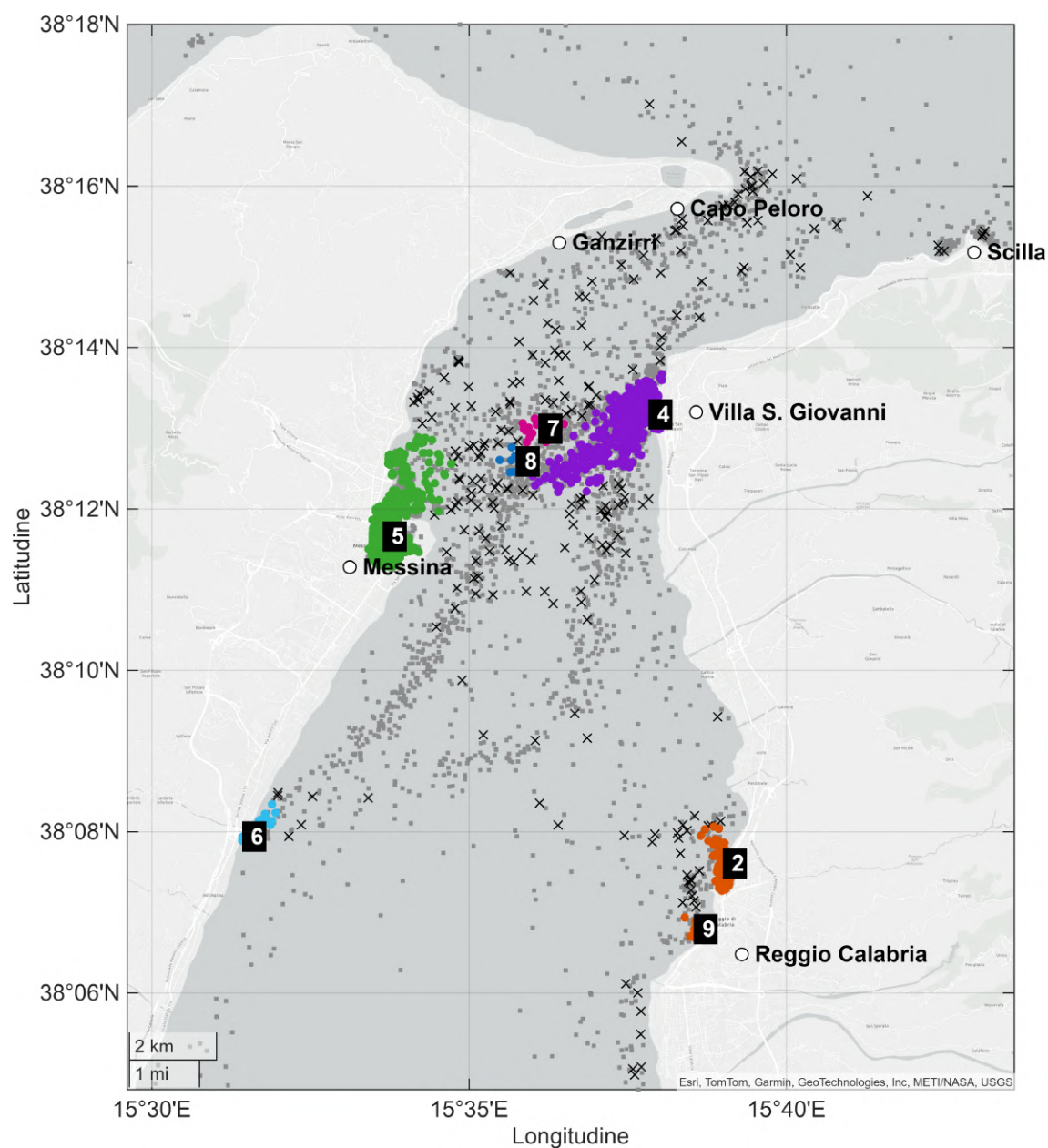


Figura 5.31: Stretto di Messina, 2023: zone DBSCAN nella vista di dettaglio dello Stretto (38,08–38,30° N; 15,50–15,72° E).

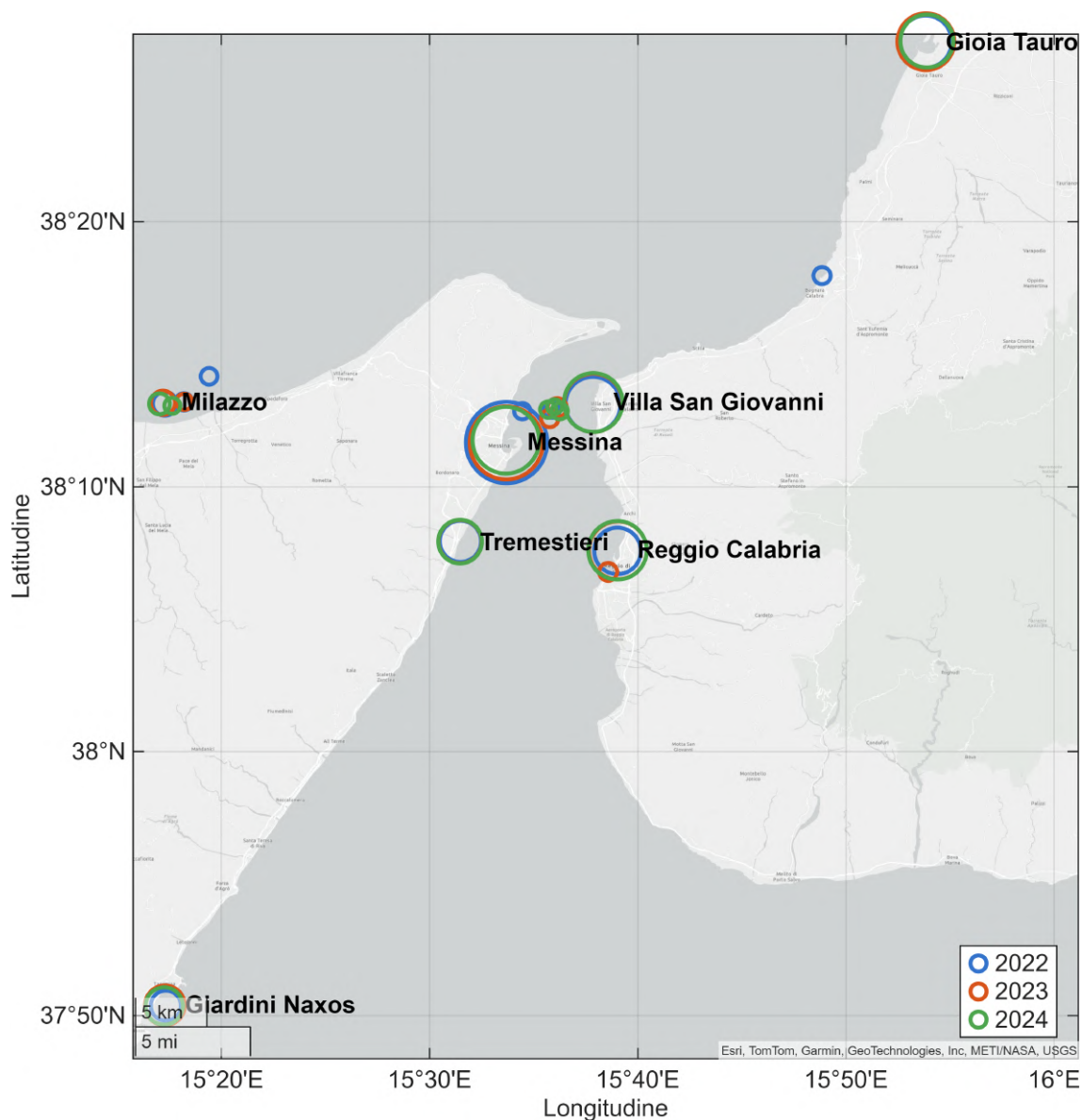


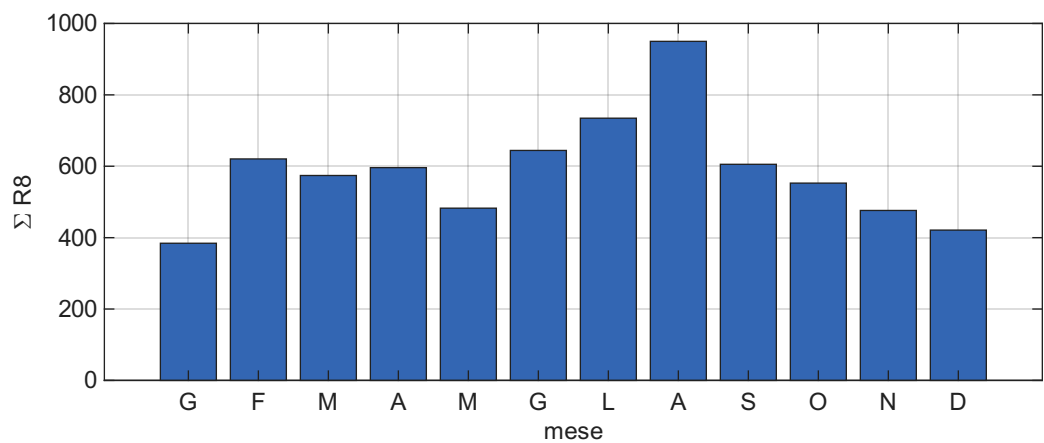
Figura 5.32: Stretto di Messina: zone DBSCAN del 2022, del 2023 e del 2024 sulla stessa mappa.

5.8.6 Profilo temporale

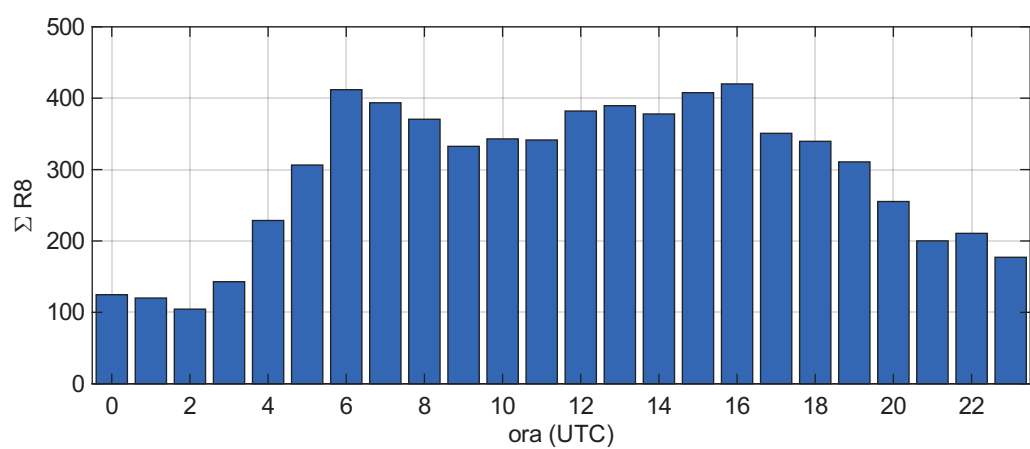
Nel 2023 il 42 % della somma di R_8 ricade fra giugno e settembre, con il valore mensile massimo ad agosto; il periodo notturno, fra le 21 e le 06 UTC, ne raccoglie il 22,9 %, e i valori orari più alti cadono alle 16, alle 6 e alle 15 UTC (Fig. 5.33 e 5.34). Negli stessi periodi ricadono il 43 % e il 22,4 % degli encounter, con il massimo mensile ad agosto. Rispetto al Tigullio il profilo è meno concentrato nella stagione estiva e nelle ore diurne, ed è compatibile con un traffico commerciale e di servizio attivo durante tutto l'anno e nell'intera giornata.

Le quote sono stabili nei tre anni (Tab. 5.5 e Fig. 5.35): la quota estiva della somma di R_8 è fra il 41,7 e il 45,9 %, quella notturna fra il 22,9 e il 24,3 %; le ore di picco cambiano da un anno all'altro. Nella settimana l'andamento è opposto a quello del Tigullio e si ripete nei tre anni:

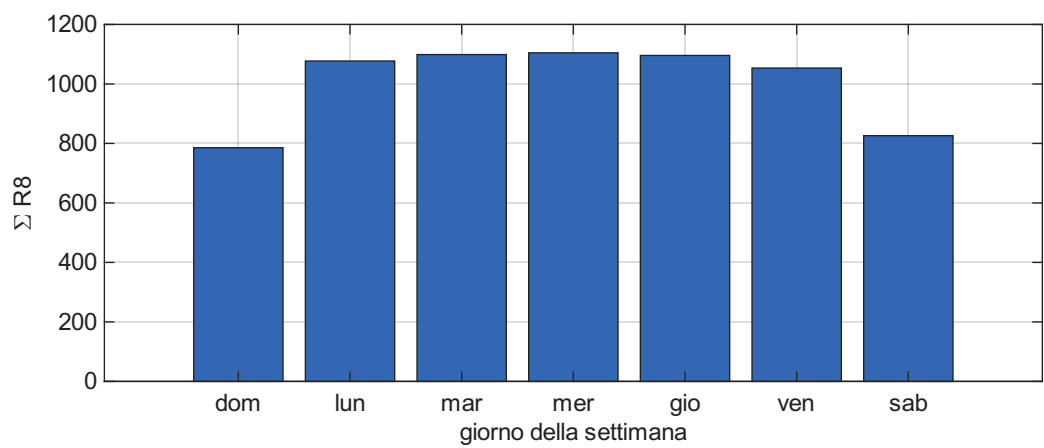
ciascun giorno feriale raccoglie fra il 15,0 e il 16,4 % della somma di R_8 , il sabato fra l'11,7 e il 12,6 % e la domenica fra il 10,4 e l'11,2 %. In tutti e tre gli anni la quota estiva di R_8 è inferiore a quella degli encounter: nel periodo estivo il valore medio di R_8 per encounter è più basso della media annua. La differenza è compatibile con l'aumento estivo degli incontri fra unità di piccole dimensioni, ai quali R_8 attribuisce valori più bassi; la composizione per lunghezza nei diversi mesi non è stata verificata.



(a) Per mese.



(b) Per ora del giorno (UTC).



(c) Per giorno della settimana.

Figura 5.33: Stretto di Messina, 2023: profilo temporale della criticità aggregata, somma di R_8 .

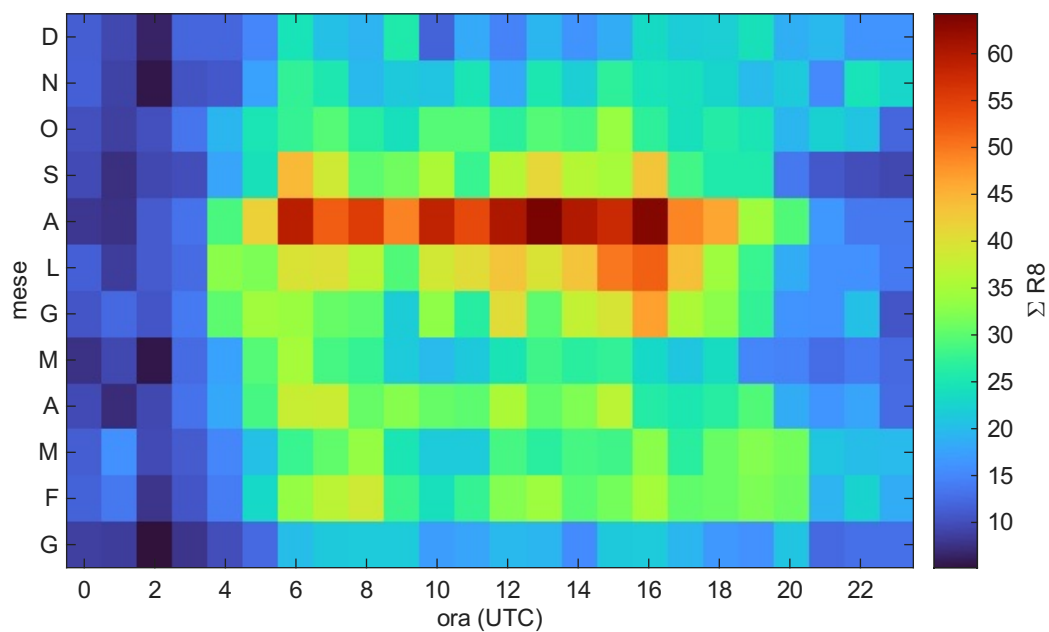


Figura 5.34: Stretto di Messina, 2023: somma di R_8 per mese e ora del giorno (UTC).

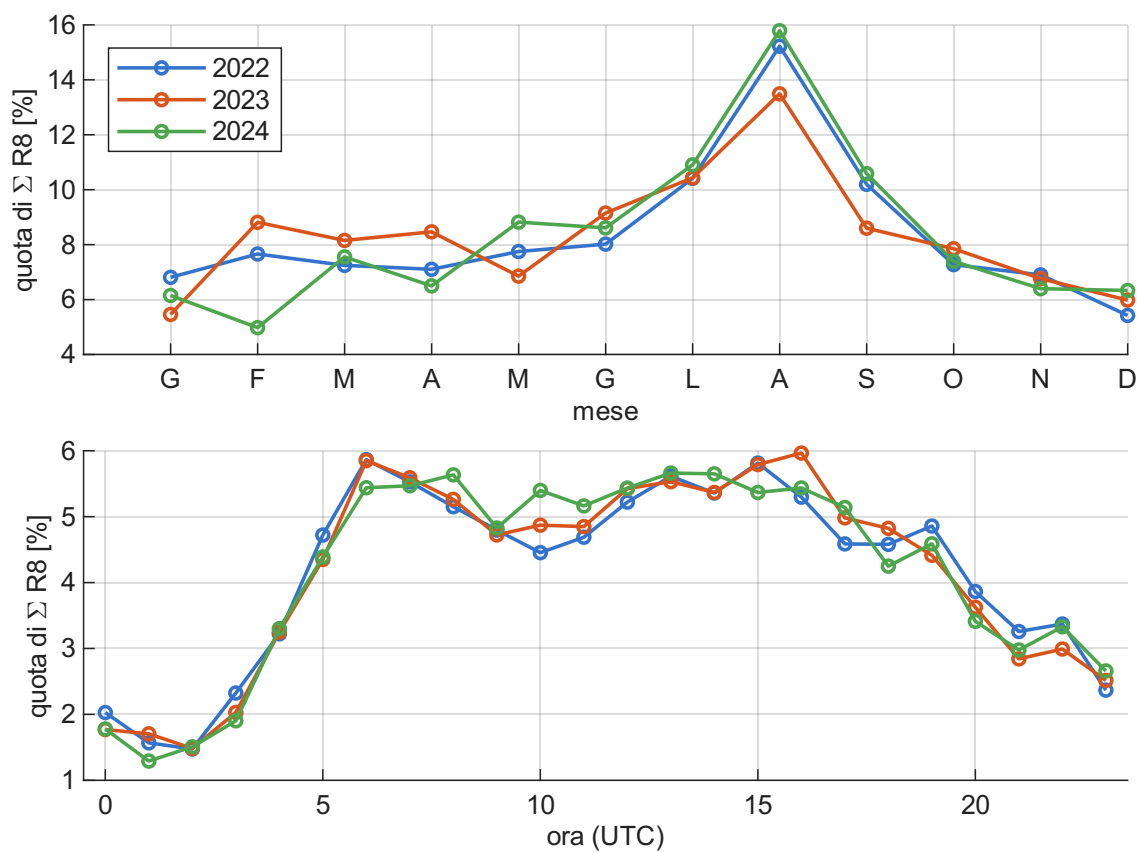


Figura 5.35: Stretto di Messina: profilo temporale della criticità aggregata (R_8) nel 2022, nel 2023 e nel 2024.

5.8.7 Verifica euristica di conformità COLREG

Nel 2023 la verifica è applicabile a 45 193 degli 83 494 encounter; i restanti 38 301, pari al 45,9 %, hanno meno di quattro record, contro il 31,0 % del Tigullio. Fra i valutabili la procedura riconosce 3 555 frontali, 22 309 incroci e 19 329 sorpassi. I tassi sui casi con esito determinato (Fig. 5.36) sono il 67,1 % nel sorpasso (6 943 su 10 343), il 49,0 % nell'incrocio (3 520 su 7 189) e l'8,1 % nel frontale (124 su 1 540). Nell'incrocio gli esiti indeterminati sono 15 120 su 22 309, e il tasso è calcolato su circa un terzo degli incroci riconosciuti. Nei tre anni l'ordinamento non cambia (Tab. 5.5): il sorpasso è fra il 64,7 e il 68,2 %, l'incrocio fra il 46,6 e il 50,8 %, il frontale fra il 6,3 e l'8,5 %. È lo stesso ordinamento del Tigullio.

Fra gli esiti determinati del 2023, gli encounter in cui almeno un'unità ha velocità mediana inferiore a 1 nodo sono il 55,9 % nel sorpasso, il 76,6 % nell'incrocio e il 92,5 % nel frontale: i tassi dello Stretto riguardano quindi in prevalenza incontri con un'unità ferma, in particolare nel frontale. Valgono le cautele esposte per il Tigullio sul significato di non conforme, sul criterio applicato al sorpasso e sul frontale, e sulla differenza fra i totali della verifica e quelli della classificazione per tipo (Sezione 5.7). La quota più alta di encounter non valutabili e di esiti indeterminati attribuisce ai tassi dello Stretto un valore indicativo.

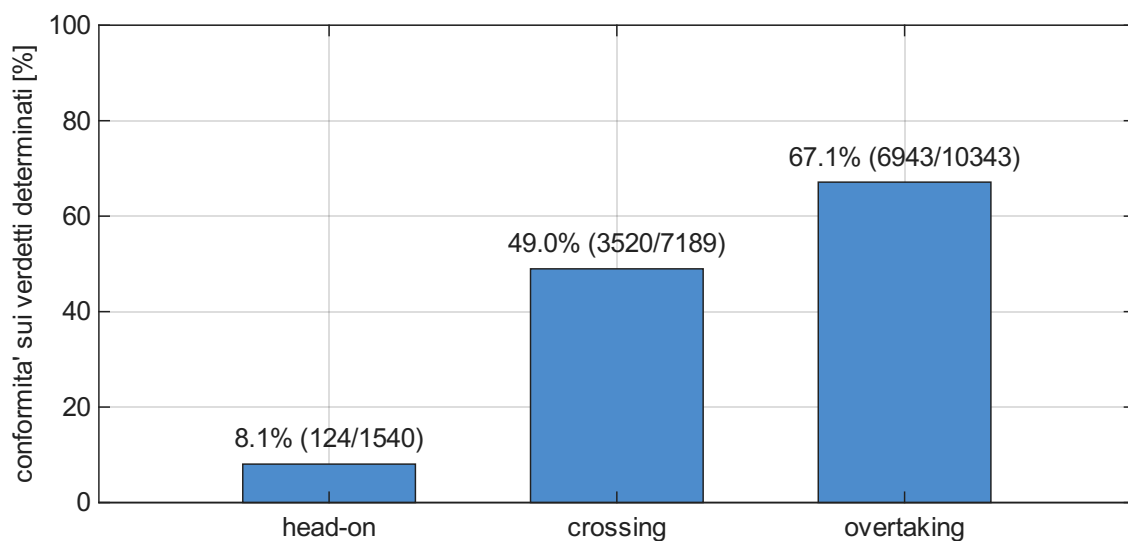


Figura 5.36: Stretto di Messina, 2023: tassi di conformità stimati con la verifica euristica COLREG, calcolati sui soli casi con esito determinato.

Tabella 5.5: Stretto di Messina: tipologia degli encounter, profilo temporale della criticità (R_8) e conformità COLREG nei tre anni. Estate: giugno–settembre; notte: 21:00–06:00 UTC.

	2022	2023	2024
Frontale / incrocio / sorpasso (%)	18,4 / 51,6 / 30,0	17,2 / 51,4 / 31,4	17,8 / 48,8 / 33,4
Quota estiva: R_8 / encounter (%)	43,8 / 46,1	41,7 / 43,2	45,9 / 48,9
Quota notturna: R_8 / encounter (%)	24,3 / 24,0	22,9 / 22,4	23,1 / 23,0
Ore di picco di R_8 (UTC)	6, 15, 13	16, 6, 15	13, 14, 8
Conformità COLREG: frontale / incrocio / sorpasso (%)	6,3 / 46,6 / 68,2	8,1 / 49,0 / 67,1	8,5 / 50,8 / 64,7

5.8.8 Casi illustrativi per tipo di incontro

Fra i candidati dei tre anni, selezionati e valutati sulle posizioni trasmesse come descritto nella Sezione 4.9, risultano validi 10 sorpassi, 4 incroci e nessun frontale. Sono stati scelti sei casi, tre sorpassi e tre incroci (Tab. 5.6 e 5.7); la Fig. 5.37 ne riporta la posizione. Le unità sono indicate con la classe e la lunghezza; l'unità A è quella tenuta a manovrare.

Tabella 5.6: Casi illustrativi nello Stretto di Messina: unità coinvolte ed esito. L'unità A è quella tenuta a manovrare.

Caso	Anno	Situazione	Esito	Unità A	Unità B
1	2023	sorpasso	non conforme	da carico, 335 m	diporto, 66 m
2	2024	sorpasso	conforme	veloce, 40 m	da carico, 90 m
3	2023	sorpasso	non conforme	da carico, 210 m	passaggeri, 97 m
4	2022	incrocio	non conforme	passaggeri, 80 m	da carico, 108 m
5	2024	incrocio	conforme	passaggeri, 97 m	da carico, 81 m
6	2023	incrocio	non conforme	passaggeri, 100 m	diporto, 80 m

Tabella 5.7: Casi illustrativi: grandezze misurate sulle posizioni trasmesse. Le variazioni di rotta (positive a dritta) e di velocità sono calcolate fra la fase di approccio (da -12 a -10 min) e il passaggio (da -1 a $+1$ min); la DCPA stimata è la minima prima del passaggio.

Caso	Dist. oss. [m]	DCPA [m]	Δ rotta A [°]	Δ SOG A [kn]	Δ rotta B [°]	Δ SOG B [kn]
1	285	0	+8	−0,2	−2	−0,1
2	145	35	+27	+0,3	+7	+1,1
3	228	38	−3	+0,8	+28	−0,2
4	163	1	−21	+0,5	+9	−1,2
5	165	12	−23	−2,6	−2	+1,5
6	309	2	−54	+2,0	−5	+0,2

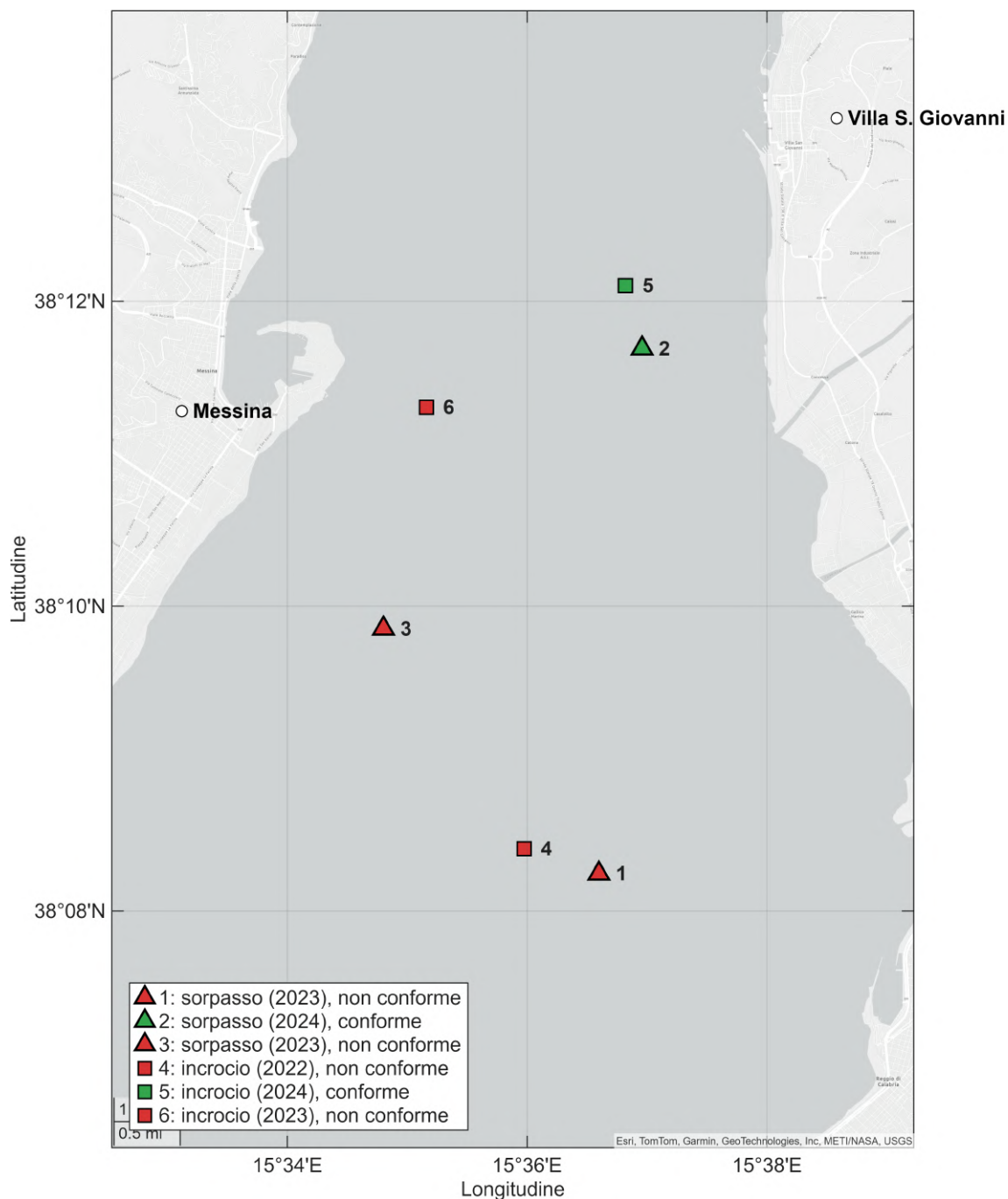


Figura 5.37: Stretto di Messina: posizione dei casi illustrativi all'istante di minima distanza osservata.

Nel caso 2 (sorpasso, 2024; Fig. 5.38) un'unità veloce di 40 m raggiunge una nave da carico di 90 m, che nella fase di approccio la rileva a 139° dalla prua: l'unità veloce è la nave raggiungente ed è tenuta a manovrare. Fra la fase di approccio e il passaggio accosta di 27° a dritta e mantiene la velocità, da 18,9 a 19,3 kn, mentre la nave da carico varia la rotta di 7° , sotto la soglia. L'accostata a dritta supera la soglia di 15° e l'esito è conforme. La distanza osservata minima è 145 m, la DCPA stimata minima prima del passaggio 35 m.

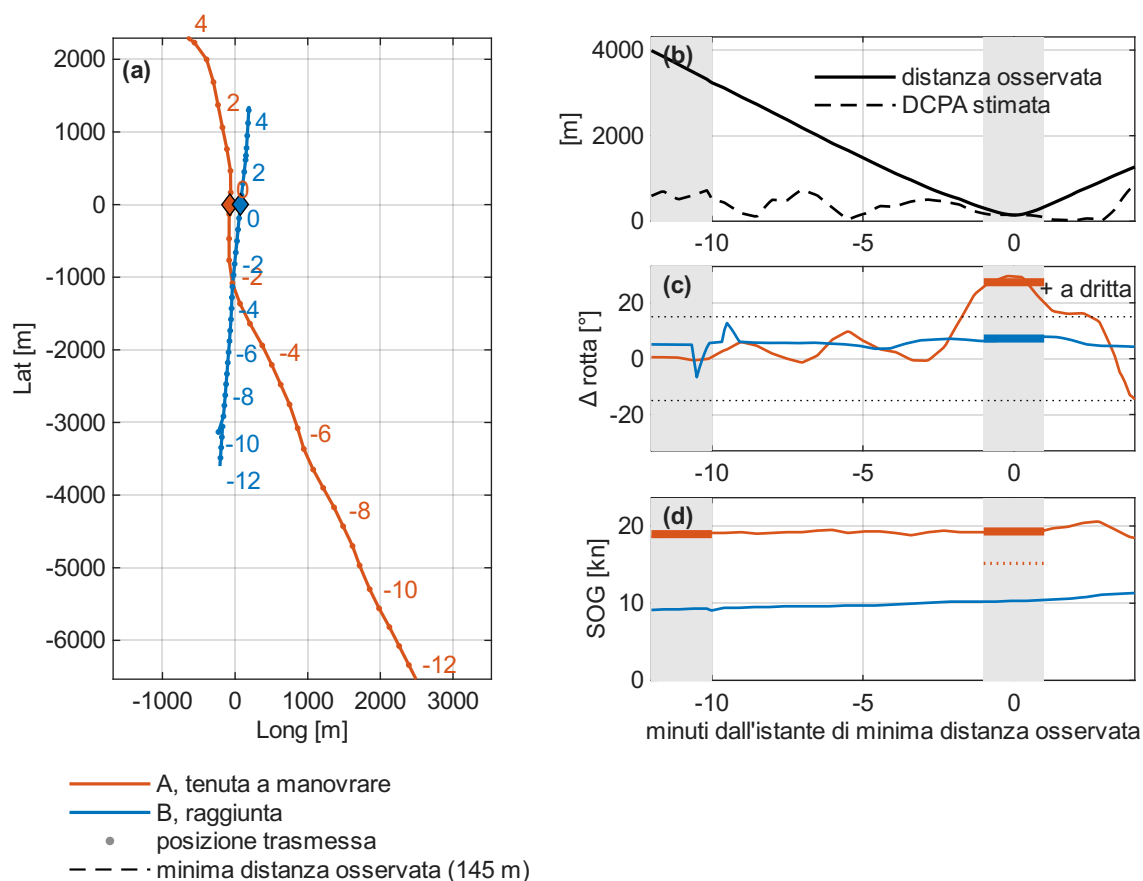


Figura 5.38: Caso 2: sorpasso (Regola 13), Stretto di Messina, 2024, esito conforme. (a) Tracce ricostruite dalle posizioni trasmesse, con i minuti rispetto all'istante di minima distanza osservata; (b) distanza osservata e DCPA stimata; (c) variazione di rotta rispetto alla fase di approccio; (d) SOG. Bande grigie: fase di approccio e passaggio.

Nel caso 6 (incrocio, 2023; Fig. 5.39) una nave passeggeri di 100 m ha sulla propria dritta, con rilevamento relativo di 54° , un'unità da diporto di 80 m ed è tenuta a manovrare; le rotte differiscono di 55° . Fra la fase di approccio e il passaggio la nave passeggeri accosta di 54° a sinistra e aumenta la velocità da 13,2 a 15,2 kn, mentre l'unità da diporto varia la rotta di 5° . L'accostata supera la soglia ma è a sinistra e non è accompagnata da una riduzione di velocità: l'esito è non conforme. La distanza osservata minima è 309 m, la DCPA stimata minima prima del passaggio 2 m.

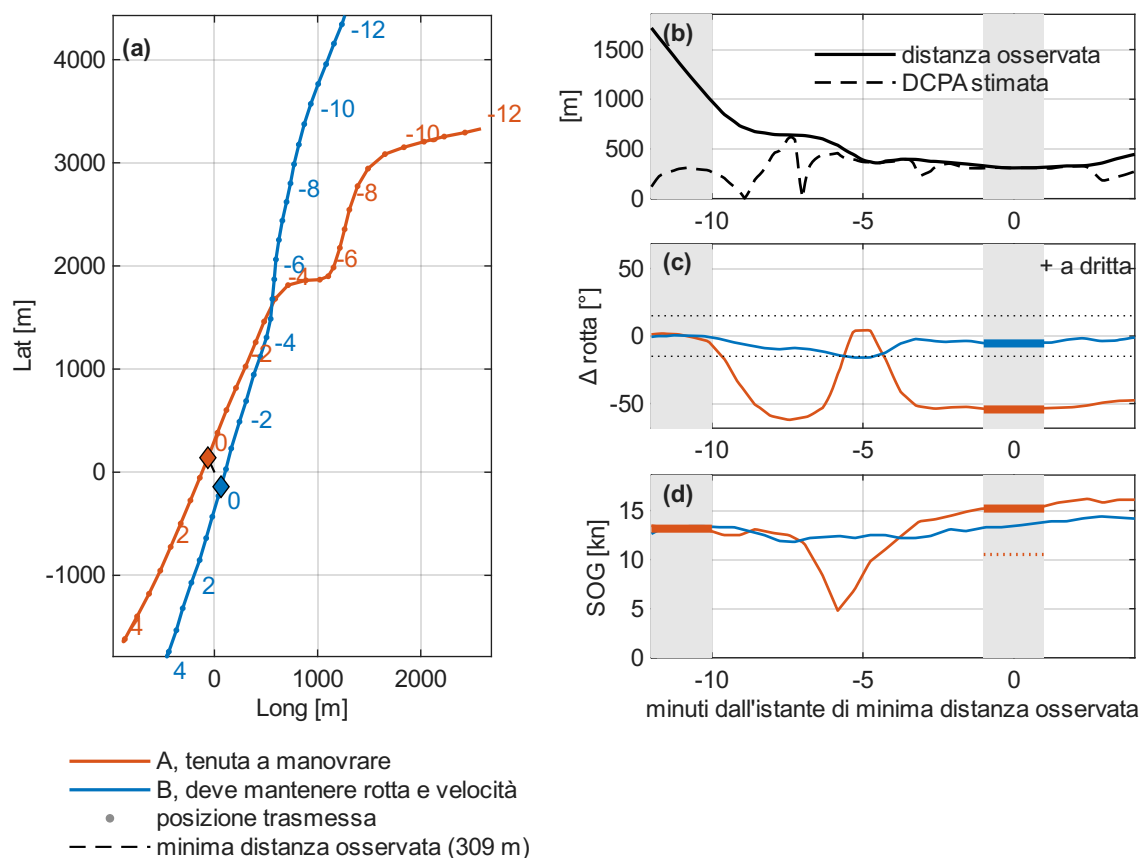


Figura 5.39: Caso 6: incrocio (Regola 15), Stretto di Messina, 2023, esito non conforme. (a) Tracce ricostruite dalle posizioni trasmesse, con i minuti rispetto all'istante di minima distanza osservata; (b) distanza osservata e DCPA stimata; (c) variazione di rotta rispetto alla fase di approccio; (d) SOG. Bande grigie: fase di approccio e passaggio.

Gli altri quattro casi sono riportati in Appendice B. Nel caso 3 a variare la rotta, di 28° a dritta, è la nave raggiunta, mentre la raggiungente, tenuta a manovrare, non varia rotta né velocità oltre le soglie: l'esito è non conforme. Il caso 5 è conforme per la riduzione di velocità, da 9,9 a 7,4 kn, accompagnata da un'accostata di 23° a sinistra. Nei casi 1, 3 e 4 le posizioni trasmesse nella finestra sono fra 8 e 11 per unità, con intervalli fino a 143 s, contro le 31–34 dei casi 2, 5 e 6.

Nessun incontro frontale supera la valutazione. Nei tre anni i candidati frontali sono 17, contro 1 179 sorpassi e 196 incroci; dei 15 valutati sulle posizioni trasmesse, 10 sono esclusi perché almeno una delle due unità procede a meno di 6 kn nella fase di approccio o a meno di 2 kn nella finestra, 5 per intervalli fra posizioni superiori a 150 s. Nel frontale la velocità relativa è prossima alla somma delle due velocità: due unità a 10 kn restano entro la soglia di estrazione di 500 m per circa un minuto e mezzo, e l'incontro è registrato con uno o due record. Gli incontri frontali con più record coinvolgono quindi in prevalenza unità lente: nel 2023 il 92,5 % dei frontali con esito determinato comprende un'unità con velocità mediana inferiore a 1 kn.

Fra i casi validi prevalgono gli esiti non conformi: 6 su 10 nel sorpasso e 3 su 4 nell'incrocio. Nei casi non conformi riportati la distanza osservata minima è compresa fra 163 e 309 m, e l'unità

tenuta a manovrare non varia rotta o velocità oltre le soglie oppure accosta a sinistra. L'esito non conforme registra l'assenza della manovra prevista fra la fase di approccio e il passaggio e non accerta una violazione. Il criterio di estrazione riduce inoltre la presenza di esiti conformi: un'unità che manovra con anticipo e passa a più di 500 m dall'altra non genera un incontro. Nello Stretto è in vigore uno schema di separazione del traffico a rotatoria (IMO SN.1/Circ.279), nelle cui corsie è raccomandato mantenere una rotta costante ed evitare accostate improvvise (Italian Coast Guard Headquarters, 2015); la verifica applica le Regole 13–15 senza modellare lo schema né le procedure del VTS di Messina.

5.9 Sintesi dei risultati

La Tab. 5.8 confronta i quattro dataset con lo stesso indicatore, R_7 , calcolato ovunque con la stessa formula.

In tutti i dataset la criticità aggregata segue la densità di traffico, con correlazioni di rango superiori a 0,85 sia con R_7 sia con l'indicatore principale. La criticità per incontro ne dipende in misura molto minore, e la sua struttura spaziale non è riproducibile dalla sola distribuzione del traffico: nel test di riallocazione a densità invariata lo scarto, calcolato con l'indicatore principale, è di 14,9 deviazioni standard nel Tigullio e fra 19,7 e 29,5 nello Stretto, sempre con $p = 0,005$. Con R_7 il test dà lo stesso esito anche nello Stretto, con uno scarto minore nel 2024. La mappa della criticità aggregata indica quindi soprattutto dove si concentrano gli incontri, quella della criticità per incontro dove il singolo incontro è in media più critico. Il confronto fra l'autocorrelazione della criticità per incontro e quella della densità non dà invece un risultato stabile fra i dataset.

Gli hotspot Getis-Ord e le zone DBSCAN si concentrano nelle aree con maggiore densità di encounter e indicano concentrazioni di incontri con valori elevati dell'indicatore, non luoghi di collisione. Nel Tigullio la zona principale è presso Portofino; nello Stretto le zone della rada di Messina, di Villa San Giovanni e di Reggio Calabria si ripetono negli stessi punti in tutti e tre gli anni.

Gli encounter con una sola unità ferma sono il 10,7 % nel Tigullio e oltre l'80 % nello Stretto, dove raccolgono una quota analoga della somma dell'indicatore principale e dove le zone principali coincidono con porti e approdi. Nello Stretto la criticità riguarda quindi soprattutto incontri fra un'unità in moto e un'unità ferma: i valori dell'indicatore nelle due aree si riferiscono a situazioni di incontro diverse e vanno confrontati tenendo conto della composizione.

Il profilo temporale distingue due regimi. Nel Tigullio la criticità è stagionale, diurna e concentrata nel fine settimana, con il 67 % della somma di R_7 in estate e l'1,5 % di notte; nello Stretto è distribuita sull'intero anno, sull'intera giornata e sui giorni feriali. Nel Tigullio la quota estiva della somma di R_7 supera quella degli encounter, per cui gli incontri estivi hanno un valore medio dell'indicatore di poco superiore alla media annua; nello Stretto, con R_8 , avviene il contrario in tutti e tre gli anni.

La verifica di conformità al COLREG dà lo stesso ordinamento in tutti i dataset, con il sorpasso al tasso più alto e il frontale al più basso. Nei casi illustrativi, valutati sulle posizioni trasmesse, nessun frontale soddisfa i criteri di validità e prevalgono gli esiti non conformi.

Nei tre anni dello Stretto la posizione delle zone, la composizione per tipo di incontro, i tassi di conformità, le quote temporali e la correlazione con la densità variano poco. Variano invece il peso relativo delle zone e il numero di righe dei dati grezzi; queste differenze non sono interpretate (Sezione 5.8.5).

Tabella 5.8: Confronto fra i quattro dataset con l'indicatore R_7 . Estate: giugno–settembre; notte: 21:00–06:00 UTC. L'ultima riga si riferisce all'indicatore principale (R_7 nel Tigullio, R_8 nello Stretto).

	Tigullio	Stretto di Messina		
	2023	2022	2023	2024
Encounter cinematici	8 573	85 018	83 494	69 241
Moran I della criticità aggregata	0,267	0,248	0,334	0,318
ρ con la densità: criticità aggregata	0,890	0,863	0,867	0,856
ρ con la densità: criticità per incontro	0,317	0,282	0,328	0,265
Test di riallocazione: scarto (dev. st.)	14,9	15,7	15,1	6,0
Soglia del 95° percentile	0,219	0,085	0,085	0,091
Quota estiva: R_7 / encounter (%)	67,0 / 63,0	45,0 / 46,1	44,4 / 43,2	47,6 / 48,9
Quota notturna: R_7 / encounter (%)	1,5 / 3,4	23,2 / 24,0	21,8 / 22,4	22,4 / 23,0
Encounter con una sola unità ferma (%)	10,7	84,8	82,9	81,6
quota dell'indicatore principale (%)	5,8	83,1	80,9	80,4

Capitolo 6

Conclusioni

La tesi ha definito una procedura che, a partire dai messaggi AIS, individua gli incontri fra coppie di unità a distanza inferiore a 500 m, ne seleziona il sottoinsieme cinematico, calcola otto indicatori di criticità e ne descrive la distribuzione nello spazio e nel tempo. La procedura è stata applicata al Tigullio per il 2023 e allo Stretto di Messina per il 2022, il 2023 e il 2024. I risultati rispondono alle due questioni poste nel Capitolo 1: quanto la scelta dell'indicatore condiziona l'identificazione degli incontri critici e quanto le mappe di criticità dipendono dalla densità di traffico. La scelta dell'indicatore pesa poco sulle mappe di criticità aggregata e molto sulla selezione degli incontri più critici; le mappe costruite sulla somma dell'indicatore seguono la densità di traffico, quelle costruite sulla media molto meno.

6.1 Scelta dell'indicatore

Indicatori di formulazione diversa, applicati agli stessi encounter, ordinano gli incontri in modo diverso soprattutto nella parte alta della classifica. Fra l'inverso della distanza minima osservata (R_1) e l'indicatore principale la correlazione di Spearman è 0,58 nel Tigullio e 0,41 nello Stretto nel 2023, ma i primi cento encounter dei due indicatori quasi non coincidono (indice di Jaccard 0,01 in entrambe le aree). Le mappe di criticità aggregata cambiano poco con l'indicatore; le analisi costruite sugli incontri più critici, come le zone DBSCAN, ne dipendono invece in modo diretto, e i loro risultati valgono per l'indicatore con cui gli incontri sono stati selezionati.

R_7 , rapporto fra velocità relativa e DCPA stimata, dipende dalla sola cinematica della coppia; R_8 scala la DCPA stimata rispetto alla lunghezza media della coppia e assegna valori più alti agli incontri fra unità più lunghe. La scelta fra i due dipende dalla disponibilità delle lunghezze nell'archivio statico: entrambe sono note per il 12,1 % degli encounter del Tigullio, dove l'indicatore principale è R_7 , e per oltre l'89 % degli encounter dello Stretto, dove è R_8 . Nessuno dei due stima una probabilità di collisione.

6.2 Criticità e densità di traffico

In tutti i dataset la somma dell'indicatore per cella segue la densità degli encounter: la mappa della criticità aggregata indica soprattutto dove si concentrano gli incontri, e gli hotspot e le zone DBSCAN ne ereditano la dipendenza dalla densità. La media dell'indicatore per cella è molto meno associata alla densità, e la sua struttura spaziale non si ottiene riallocando casualmente i valori fra gli encounter a densità invariata ($p = 0,005$ in tutti i dataset). Le due superfici descrivono quindi grandezze diverse: la prima indica dove si accumulano incontri critici, la seconda dove il singolo incontro è in media più critico. Per individuare le zone in cui gli incontri sono più critici, e non solo più numerosi, la superficie da usare è la seconda. Il test non individua la causa della struttura osservata.

6.3 Differenze fra le due aree

I due casi di studio presentano regimi di incontro diversi. Nel Tigullio la criticità si concentra attorno al promontorio di Portofino, dove la zona DBSCAN principale raccoglie circa la metà degli encounter selezionati, e segue la stagione estiva, le ore diurne e il fine settimana, con un profilo compatibile con un traffico prevalentemente diportistico. Nello Stretto le zone principali ricadono nelle aree portuali e di approdo, la criticità è distribuita sull'intero anno, sull'intera giornata e soprattutto sui giorni feriali, e oltre l'80 % degli encounter cinematici coinvolge un'unità ferma e un'unità in moto, contro il 10,7 % del Tigullio. Lo stesso indicatore descrive quindi situazioni di incontro diverse nelle due aree, e il confronto fra i valori va fatto tenendo conto della composizione degli incontri.

La verifica euristica di conformità al COLREG dà invece lo stesso ordinamento in tutti i dataset, con il sorpasso al tasso più alto e il frontale al più basso. Un esito non conforme indica che la manovra prevista non è rilevabile nella traccia, non che sia stata accertata una violazione delle Regole.

6.4 Stabilità sui tre anni dello Stretto

I tre anni dello Stretto, elaborati separatamente, danno risultati coerenti: le zone DBSCAN si ripetono negli stessi punti, con spostamenti dei centroidi entro circa 100 m, e variano poco la composizione per tipo di incontro, i tassi di conformità, le quote temporali della criticità e la correlazione con la densità. La procedura produce quindi risultati stabili su dataset diversi della stessa area. Variano invece il peso relativo delle zone e il numero di righe dei dati grezzi, che triplica fra il 2022 e il 2024 senza un aumento corrispondente del numero di encounter. Queste differenze non sono state interpretate come variazioni del traffico, perché non è stata esclusa una variazione nella frequenza di registrazione dei dati.

6.5 Limiti

Il dato AIS ha una risoluzione limitata rispetto alla scala degli incontri: nel Tigullio solo il 17,5 % dei record sincroni corrisponde a messaggi effettivamente trasmessi (Sezione 3.6.5). Le analisi operano quindi su insiemi di encounter; i singoli incontri sono ricostruiti solo nei casi illustrativi della verifica COLREG, sulle posizioni trasmesse dalle unità in Classe A dello Stretto (Sezione 5.8.8). Anche in questi casi nessun incontro frontale soddisfa i criteri di validità: con il passo di un minuto e la soglia di 500 m, un frontale fra unità a velocità di servizio è registrato con uno o due record. La soglia di estrazione di 500 m è fissa e non tiene conto delle dimensioni delle unità, e nel Tigullio la lunghezza è nota per una piccola parte degli encounter, per cui la correzione dimensionale di R_8 non è utilizzabile. Gli indicatori non considerano la manovrabilità delle unità, le condizioni meteomarine e la visibilità.

Le zone di criticità non sono state confrontate con dati di incidenti o di quasi-incidenti, che non erano disponibili, e l'influenza delle soglie della verifica di conformità sui tassi non è stata analizzata. Le griglie delle due aree hanno lato diverso, circa 200 e 250 m, e gli indici di autocorrelazione sono confrontabili fra le aree solo in modo indicativo.

6.6 Sviluppi

Gli sviluppi possibili sono di tre tipi. Il primo è la validazione: confrontare le zone di criticità con dati di incidenti e di quasi-incidenti o con le informazioni del servizio di controllo del traffico dello Stretto, e le traiettorie con lo schema di separazione del traffico (Italian Coast Guard Headquarters, 2015).

Il secondo riguarda la robustezza dei risultati: verificare la frequenza di registrazione dei messaggi nei tre anni dello Stretto, condizione per leggere le differenze fra gli anni come variazioni del traffico; estendere l'analisi su più anni al Tigullio; valutare la sensibilità dei tassi di conformità alle soglie della verifica.

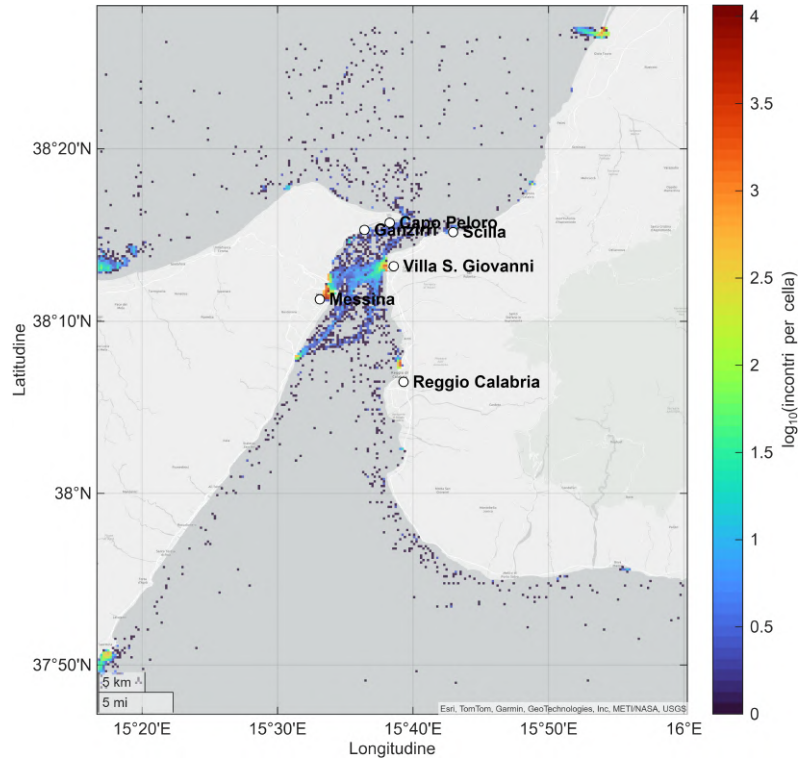
Il terzo è l'estensione del metodo: applicare ai flussi osservati il modello di Pedersen (1995), introdotto nella Sezione 2.4.5, ed esaminare gli approcci a Velocity Obstacle (Chen et al., 2021), gli incontri fra più di due unità e le acque soggette a schemi di instradamento (Feng et al., 2022). Il dominio di sicurezza adattivo di Bakdi et al. (2020) non è stato applicato: nel Tigullio la lunghezza manca per gran parte delle unità, e nello Stretto i raggi calcolati superano la distanza di estrazione degli encounter, per cui l'indicatore si ridurrebbe a una funzione della distanza osservata.

Appendice A

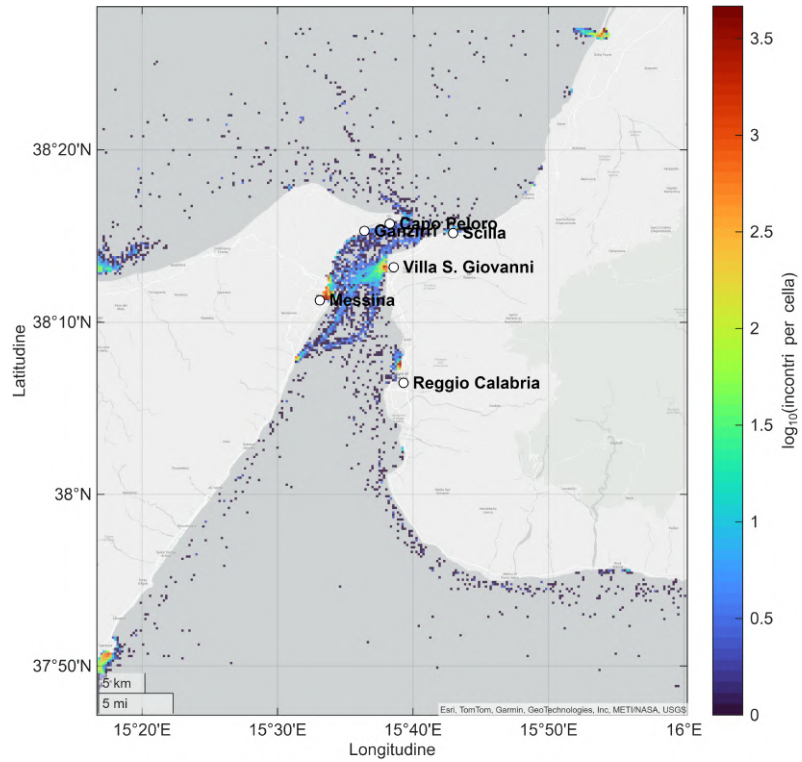
Stretto di Messina: figure del 2022 e del 2024

Le figure riportano, per il 2022 e il 2024, le elaborazioni presentate nel testo per il 2023; in ciascuna figura i due anni sono affiancati. I valori numerici dei tre anni sono riportati nelle tabelle dei Capitoli 3 e 5. Il profilo temporale per mese, ora e giorno della settimana e i tassi di conformità COLREG dei tre anni sono riportati nella Fig. 5.35 e nella Tab. 5.5 e non sono ripetuti.

A.1 Dati

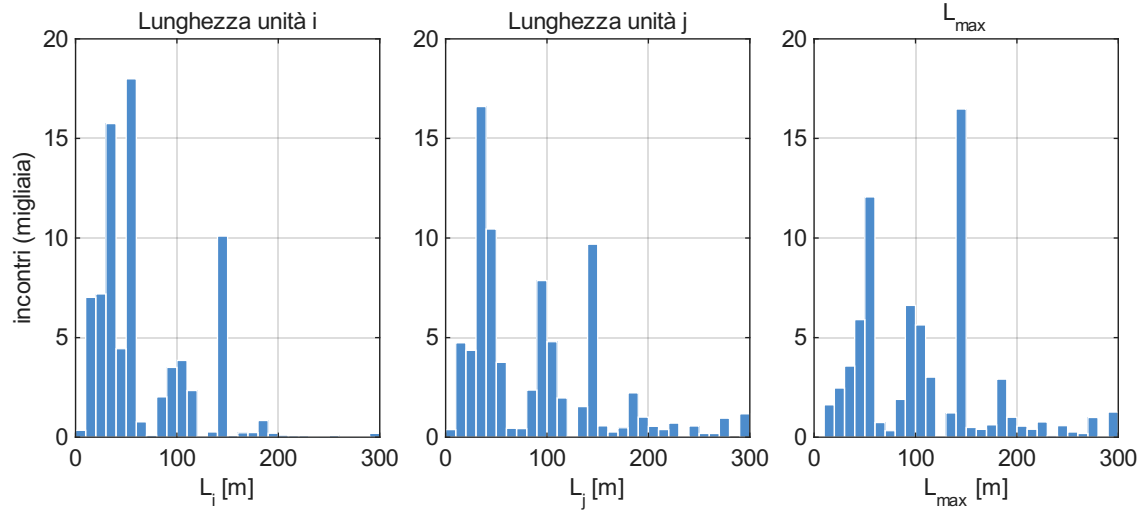


(a) 2022 (N = 85 018).

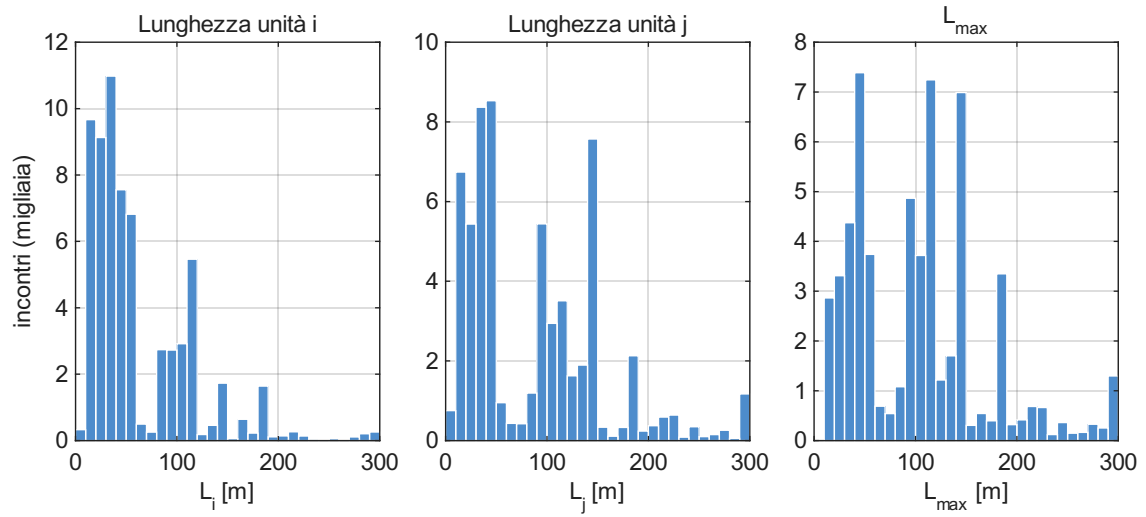


(b) 2024 (N = 69 241).

Figura A.1: Stretto di Messina: densità degli encounter cinematici nel 2022 e nel 2024, in scala logaritmica, sulla stessa griglia del 2023.

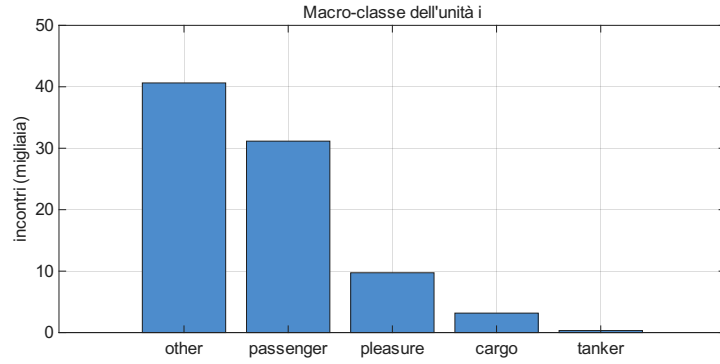


(a) 2022.

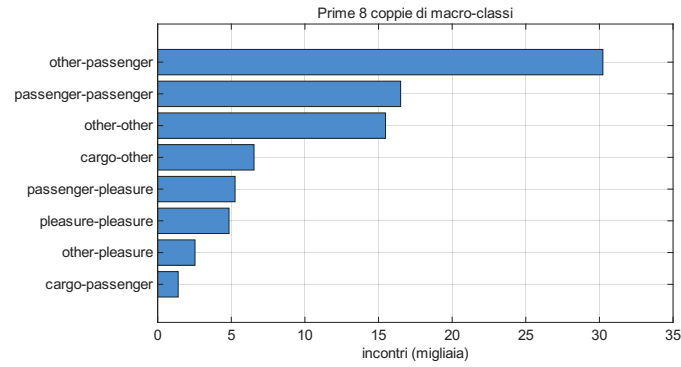


(b) 2024.

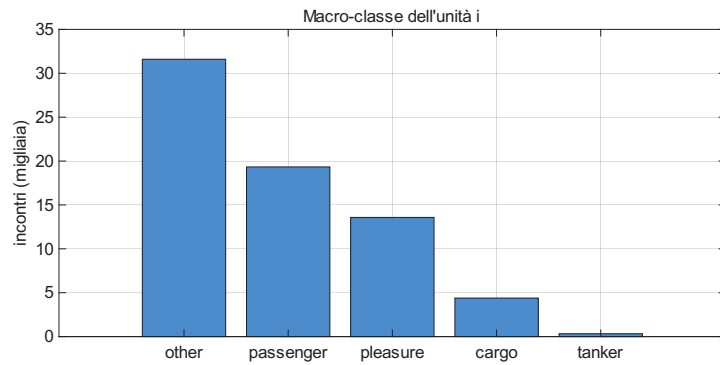
Figura A.2: Stretto di Messina: distribuzione delle lunghezze delle unità coinvolte negli encounter cinematici nel 2022 e nel 2024, calcolata per encounter sulle sole lunghezze dichiarate: unità i , unità j e $L_{\max}$. Asse limitato a 300 m.



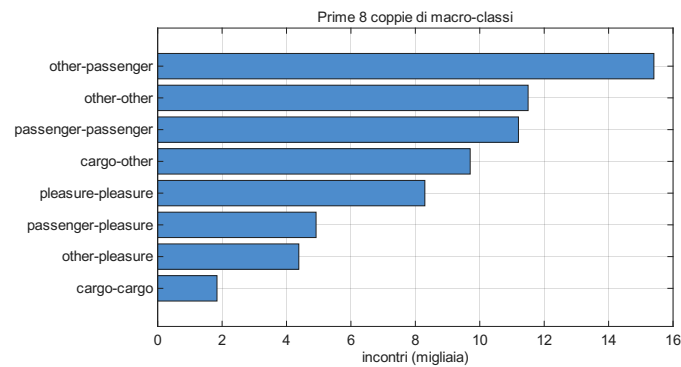
(a) 2022: macro-classe dell'unità i .



(b) 2022: prime otto coppie di macro-classi.



(c) 2024: macro-classe dell'unità i .

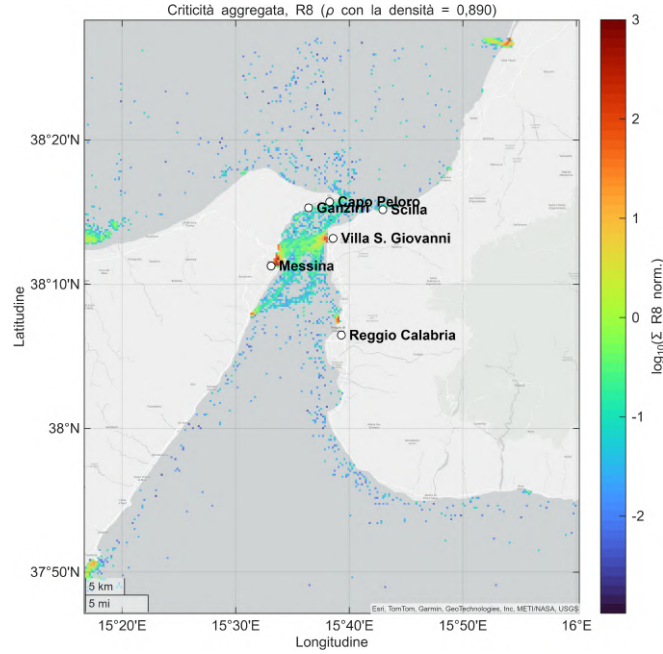


(d) 2024: prime otto coppie di macro-classi.

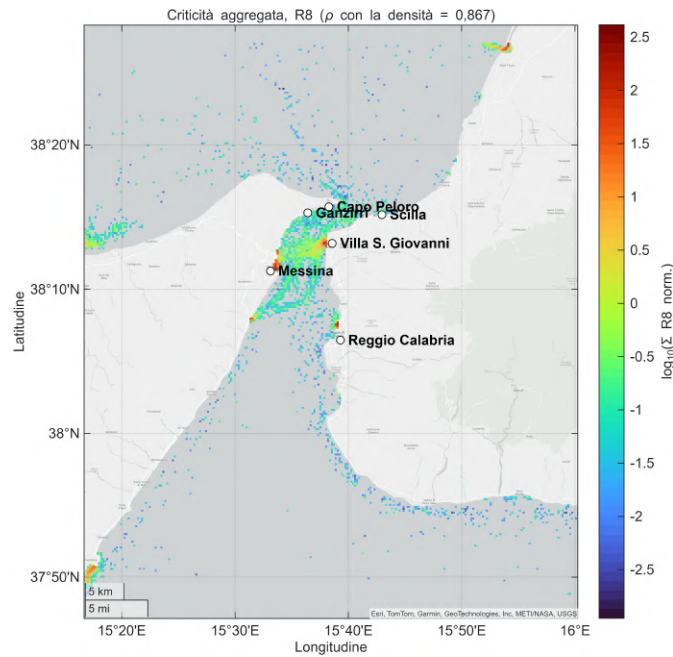
Figura A.3: Composizione del dataset cinematografico per macro-classe delle unità coinvolte (Stretto di Messina, 2022 e 2024).

A.2 Risultati

Le figure riportano, per il 2022 e il 2024, le elaborazioni della Sezione 5.8 con l'indicatore principale R_8 . I valori numerici sono riportati nelle Tab. 5.2, 5.4 e 5.5.

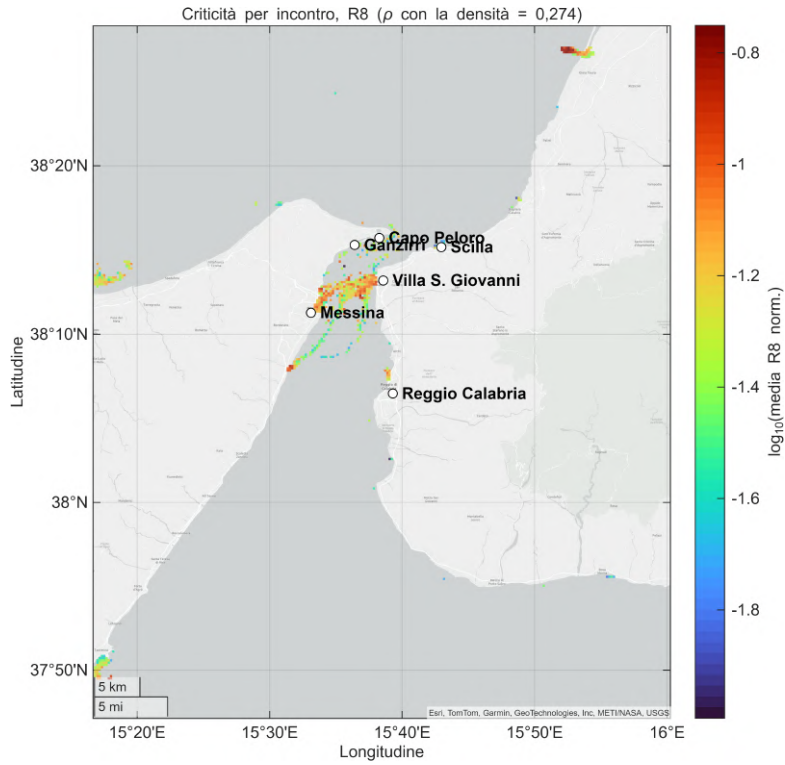


(a) 2022.

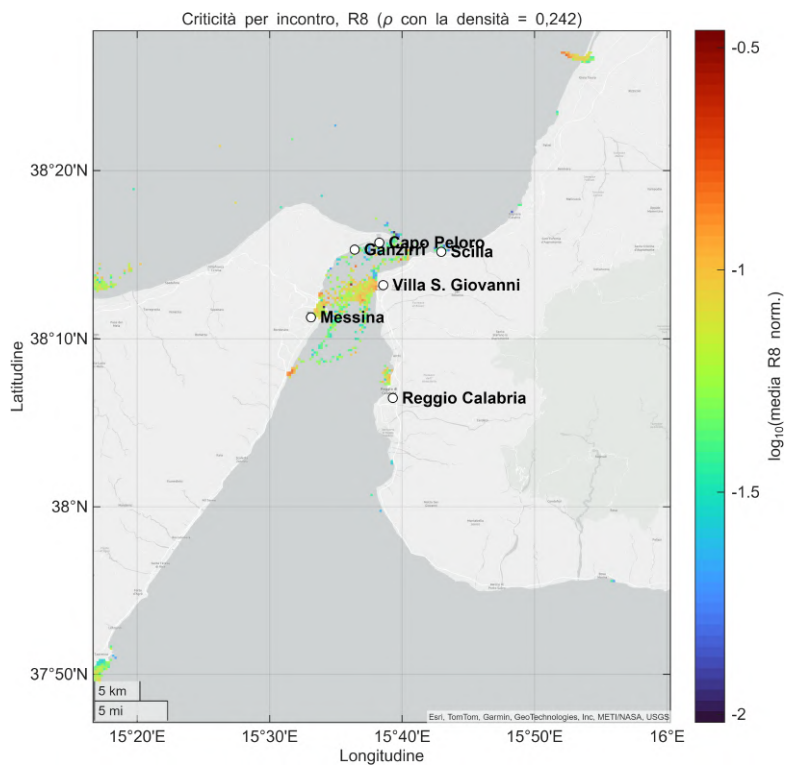


(b) 2024.

Figura A.4: Stretto di Messina: criticità aggregata (somma di R_8 per cella) nel 2022 e nel 2024, con la correlazione di Spearman con la densità calcolata sulle celle con almeno 3 encounter.

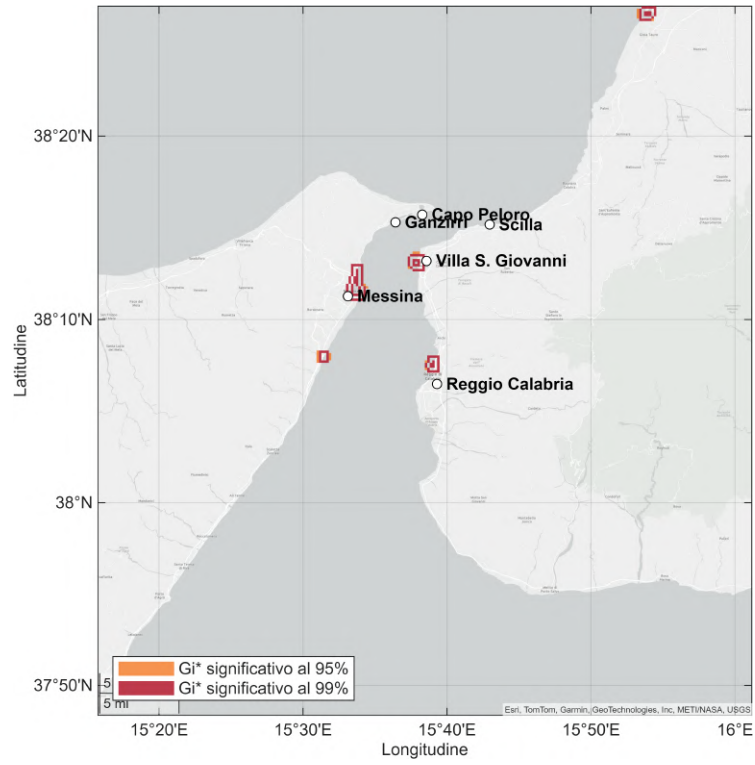


(a) 2022.

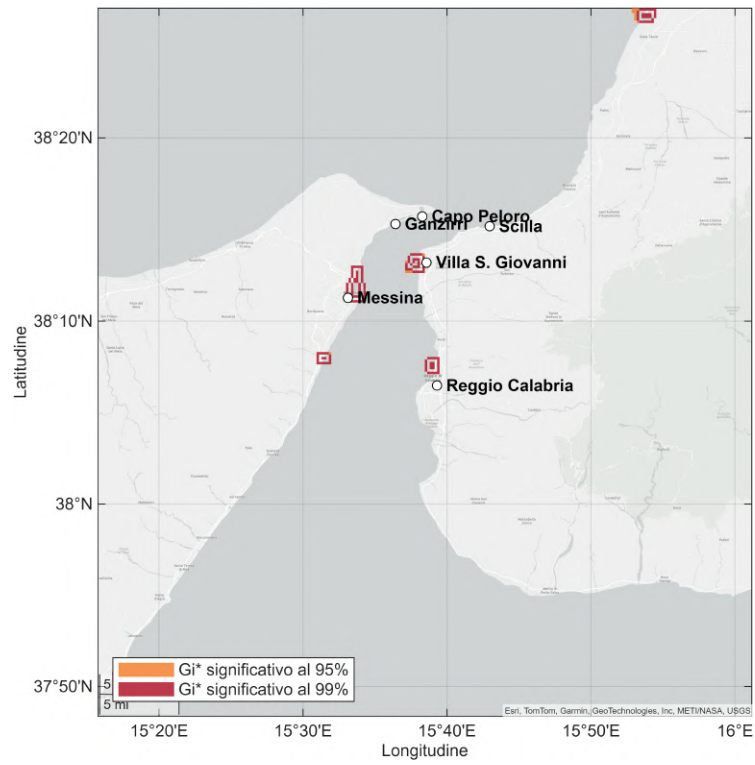


(b) 2024.

Figura A.5: Stretto di Messina: criticità per incontro (media di R_8 per cella) nel 2022 e nel 2024, con la correlazione di Spearman con la densità calcolata sulle celle con almeno 3 encounter.

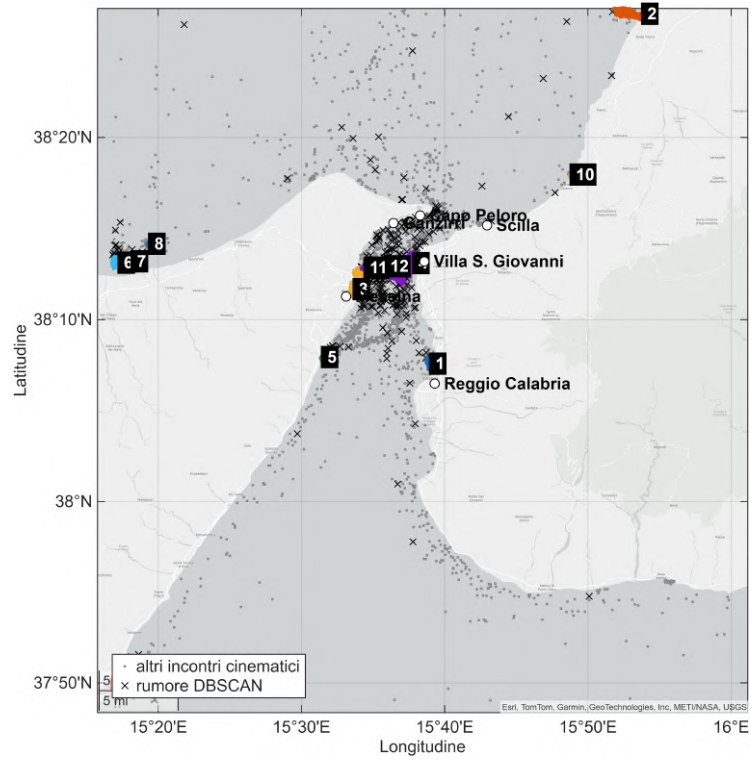


(a) 2022.

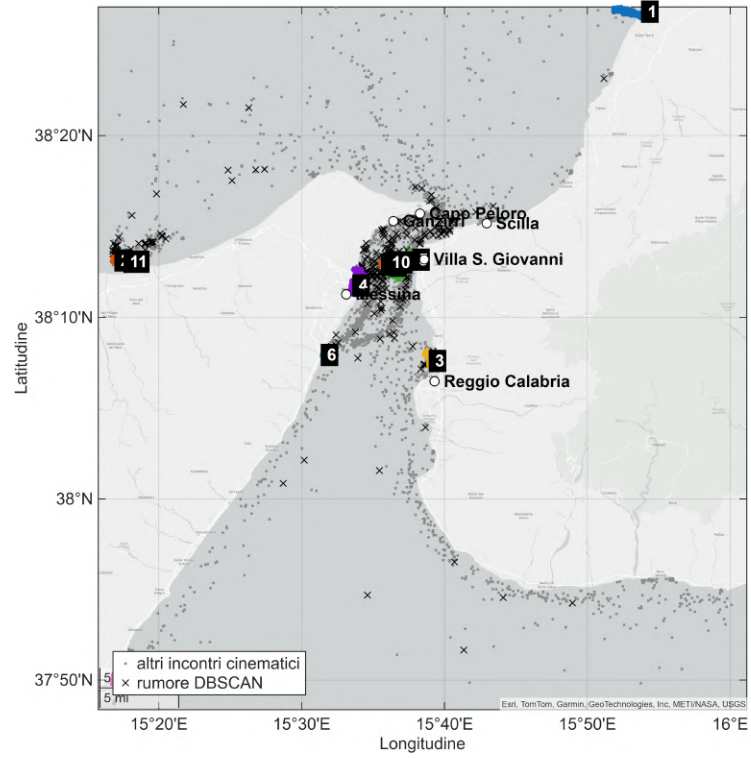


(b) 2024.

Figura A.6: Stretto di Messina: celle con concentrazione locale significativa della criticità aggregata di R_8 secondo la statistica di Getis-Ord G_i^* , al 95 % e al 99 %, nel 2022 e nel 2024.

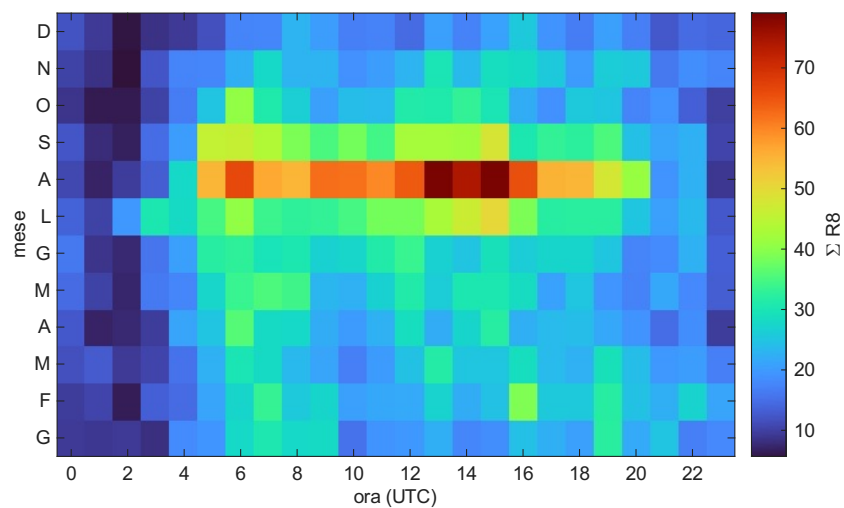


(a) 2022.

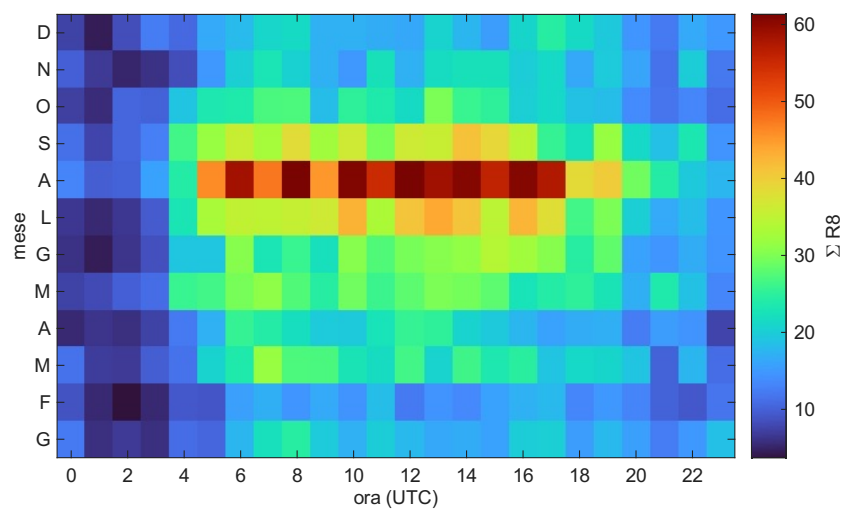


(b) 2024.

Figura A.7: Stretto di Messina: zone DBSCAN del 25 % degli encounter con R_8 più elevato ($\epsilon = 300$ m, $minPts = 8$) nel 2022 e nel 2024.



(a) 2022.



(b) 2024.

Figura A.8: Stretto di Messina: somma di R_8 per mese e ora del giorno (UTC) nel 2022 e nel 2024.

Appendice B

Casi illustrativi della verifica COLREG

Sono riportati i casi illustrativi 1, 3, 4 e 5 della Sezione 5.8.8. Le unità coinvolte e le grandezze misurate sono riassunte nelle Tab. 5.6 e 5.7; il criterio di selezione e di valutazione è descritto nella Sezione 4.9.

B.1 Caso 1: sorpasso, 2023, esito non conforme

Una nave da carico di 335 m raggiunge un'unità da diporto di 66 m, che nella fase di approccio la rileva a 170° dalla prua: la nave da carico è la raggiungente ed è tenuta a manovrare. Fra la fase di approccio e il passaggio varia la rotta di 8° a dritta e la velocità da 16,2 a 16,0 kn, entrambe sotto le soglie, e l'esito è non conforme. La distanza osservata minima è 285 m, la DCPA stimata minima prima del passaggio 0 m. Le posizioni trasmesse nella finestra sono 10 e 11, con intervalli fino a 100 s.

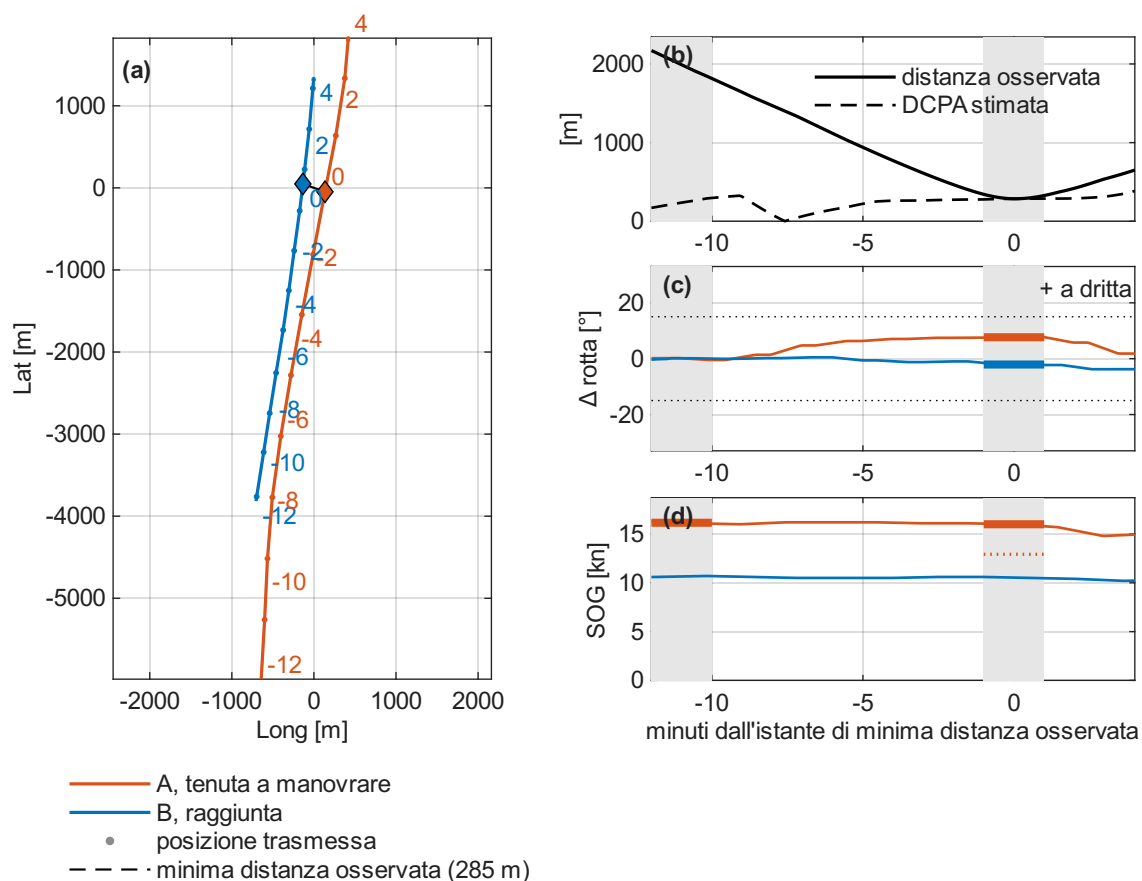


Figura B.1: Caso 1: sorpasso (Regola 13), Stretto di Messina, 2023, esito non conforme. (a) Tracce ricostruite dalle posizioni trasmesse, con i minuti rispetto all'istante di minima distanza osservata; (b) distanza osservata e DCPA stimata; (c) variazione di rotta rispetto alla fase di approccio; (d) SOG. Bande grigie: fase di approccio e passaggio.

B.2 Caso 3: sorpasso, 2023, esito non conforme

Una nave da carico di 210 m raggiunge una nave passeggeri di 97 m, che nella fase di approccio la rileva a 130° dalla prua. La nave da carico, tenuta a manovrare, varia la rotta di 3° a sinistra e porta la velocità da 12,7 a 13,5 kn; la nave passeggeri, nave raggiunta, accosta di 28° a dritta. Poiché la manovra non è rilevata sull'unità tenuta a manovrare, l'esito è non conforme. La distanza osservata minima è 228 m, la DCPA stimata minima prima del passaggio 38 m. Le posizioni trasmesse nella finestra sono 10 e 11, con intervalli fino a 100 s.

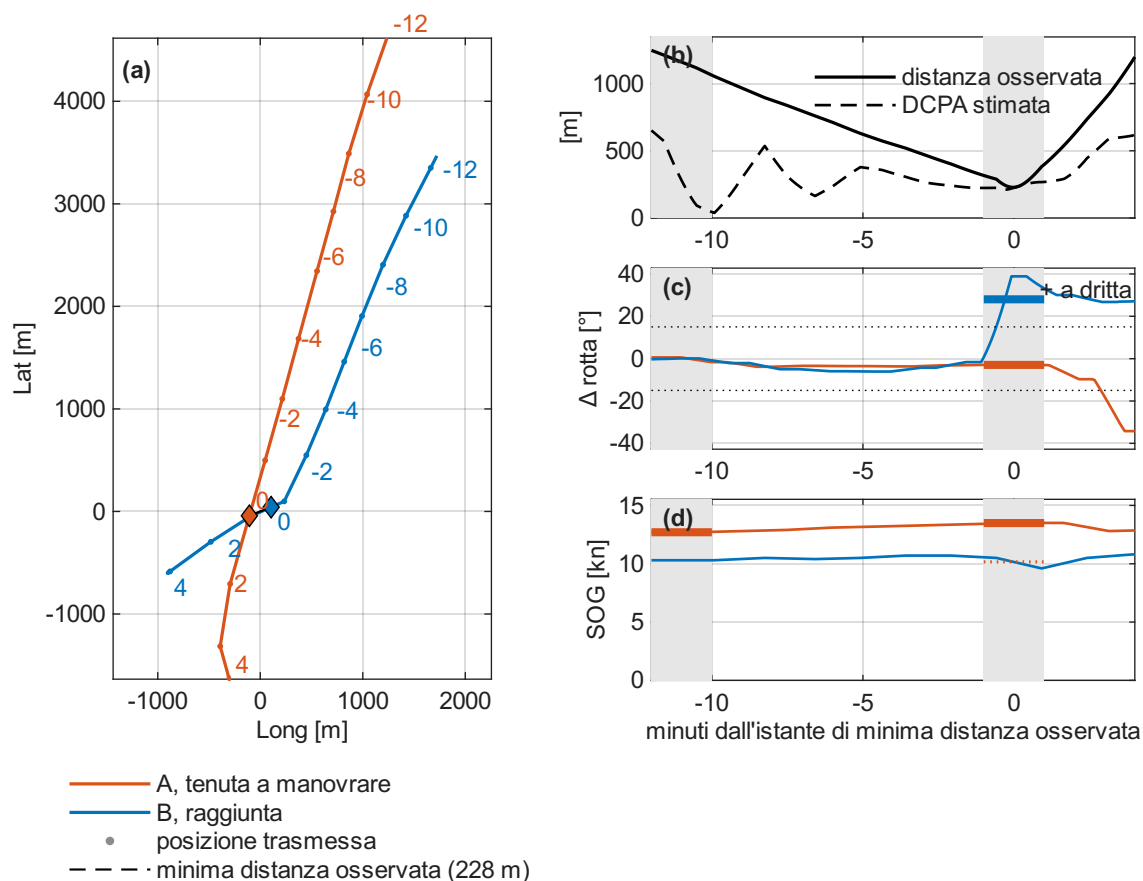


Figura B.2: Caso 3: sorpasso (Regola 13), Stretto di Messina, 2023, esito non conforme. (a) Tracce ricostruite dalle posizioni trasmesse, con i minuti rispetto all'istante di minima distanza osservata; (b) distanza osservata e DCPA stimata; (c) variazione di rotta rispetto alla fase di approccio; (d) SOG. Bande grigie: fase di approccio e passaggio.

B.3 Caso 4: incrocio, 2022, esito non conforme

Una nave passeggeri di 80 m ha sulla propria dritta, con rilevamento relativo di 46° , una nave da carico di 108 m ed è tenuta a manovrare; le rotte differiscono di 73° . Fra la fase di approccio e il passaggio la nave passeggeri accosta di 21° a sinistra e porta la velocità da 11,3 a 11,8 kn: l'accostata supera la soglia ma è a sinistra, e l'esito è non conforme. La distanza osservata minima è 163 m, la DCPA stimata minima prima del passaggio 1 m. Le posizioni trasmesse nella finestra sono 8 per ciascuna unità, con intervalli fino a 143 s.

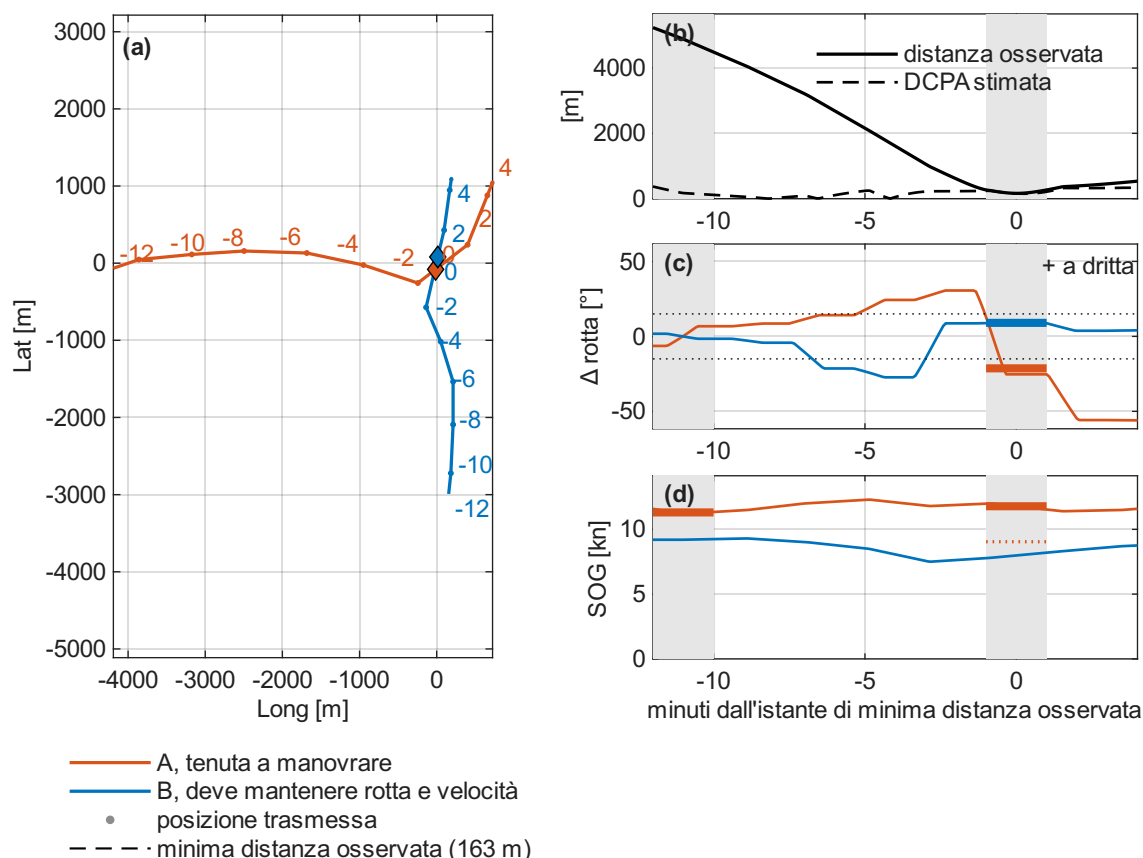


Figura B.3: Caso 4: incrocio (Regola 15), Stretto di Messina, 2022, esito non conforme. (a) Tracce ricostruite dalle posizioni trasmesse, con i minuti rispetto all'istante di minima distanza osservata; (b) distanza osservata e DCPA stimata; (c) variazione di rotta rispetto alla fase di approccio; (d) SOG. Bande grigie: fase di approccio e passaggio.

B.4 Caso 5: incrocio, 2024, esito conforme

Una nave passeggeri di 97 m ha sulla propria dritta, con rilevamento relativo di 60° , una nave da carico di 81 m ed è tenuta a manovrare; le rotte differiscono di 75° . Fra la fase di approccio e il passaggio la nave passeggeri accosta di 23° a sinistra e riduce la velocità da 9,9 a 7,4 kn. La riduzione di velocità supera la soglia, pari a 2,0 kn per questa velocità iniziale, e l'esito è conforme nonostante l'accostata a sinistra. La distanza osservata minima è 165 m, la DCPA stimata minima prima del passaggio 12 m.

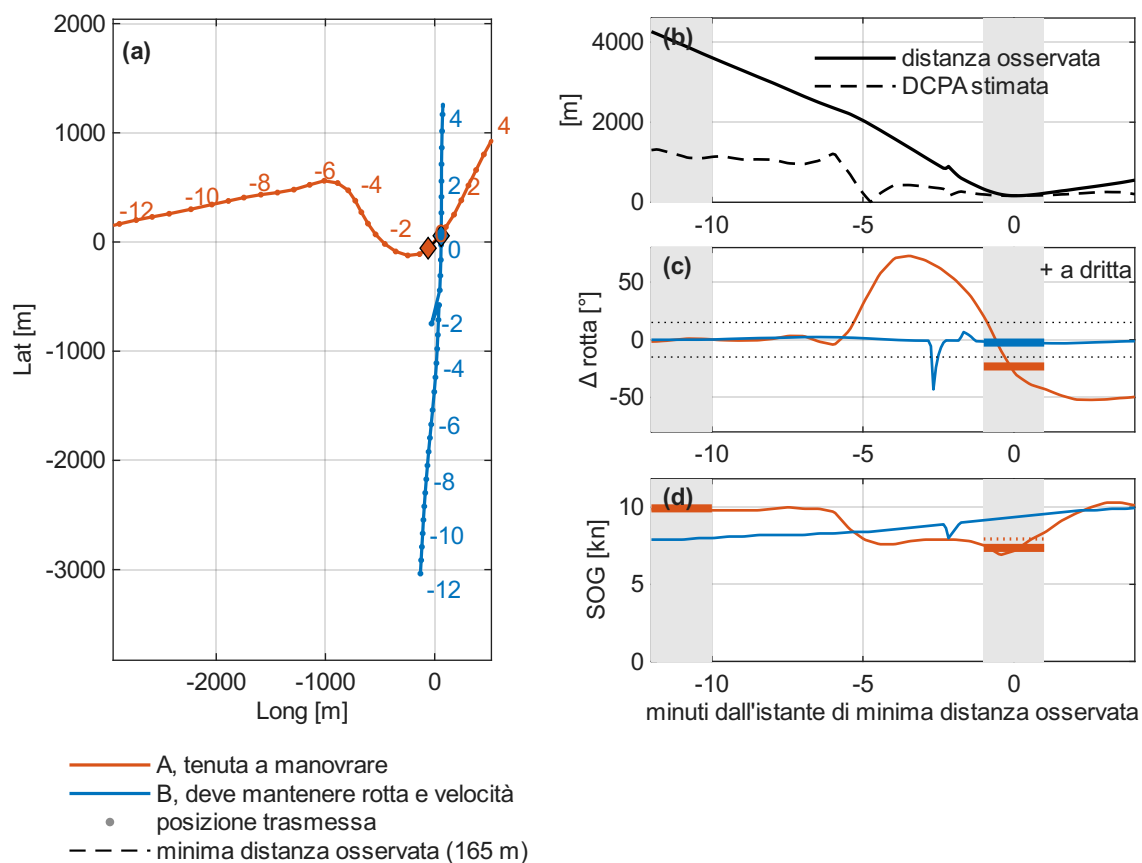


Figura B.4: Caso 5: incrocio (Regola 15), Stretto di Messina, 2024, esito conforme. (a) Tracce ricostruite dalle posizioni trasmesse, con i minuti rispetto all'istante di minima distanza osservata; (b) distanza osservata e DCPA stimata; (c) variazione di rotta rispetto alla fase di approccio; (d) SOG. Bande grigie: fase di approccio e passaggio.

Bibliografia

- Azzeddine Bakdi, Ingrid Kristine Glad, Erik Vanem, and Ørnulf Jan Økland. AIS-based multiple vessel collision and grounding risk identification based on adaptive safety domain. *Journal of Marine Science and Engineering*, 8(1):5, 2020. doi: 10.3390/jmse8010005.
- Rachel Ballotta. Analisi probabilistica della collisione tra unità navali. Tesi di laurea magistrale in yacht design, Università degli Studi di Genova, 2025.
- Christine Chauvin, Salim Lardjane, Gaël Morel, Jean-Pierre Clostermann, and Benoît Langard. Human and organisational factors in maritime accidents: Analysis of collisions at sea using the HFACS. *Accident Analysis & Prevention*, 59:26–37, 2013. doi: 10.1016/j.aap.2013.05.006.
- Pengfei Chen, Mengxia Li, and Junmin Mou. A velocity obstacle-based real-time regional ship collision risk analysis method. *Journal of Marine Science and Engineering*, 9(4):428, 2021. doi: 10.3390/jmse9040428.
- COWI. Risk analysis of sea traffic in the area around Bornholm. Technical report, COWI A/S, Kongens Lyngby, Denmark, 2008.
- Ashim Kumar Debnath and Hoong Chor Chin. Modelling collision potentials in port anchorages: Application of the navigational traffic conflict technique (NTCT). *Journal of Navigation*, 69(1): 183–196, 2016. doi: 10.1017/S0373463315000521.
- Lei Du, Floris Goerlandt, and Pentti Kujala. Review and analysis of methods for assessing maritime waterway risk based on non-accident critical events detected from AIS data. *Reliability Engineering & System Safety*, 200:106933, 2020. doi: 10.1016/j.ress.2020.106933.
- Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 226–231, Portland, Oregon, 1996. AAAI Press.
- Hongxiang Feng, Manel Grifoll, Zhongzhen Yang, and Pengjun Zheng. Collision risk assessment for ships’ routeing waters: An information entropy approach with automatic identification system (AIS) data. *Ocean and Coastal Management*, 224:106184, 2022. doi: 10.1016/j.ocecoaman.2022.106184.

- Yahei Fujii and Kenichi Tanaka. Traffic capacity. *Journal of Navigation*, 24(4):543–552, 1971. doi: 10.1017/S0373463300022384.
- Floris Goerlandt and Jakub Montewka. Maritime transportation risk analysis: Review and analysis in light of some foundational issues. *Reliability Engineering & System Safety*, 138:115–134, 2015. doi: 10.1016/j.ress.2015.01.025.
- IALA. IWRAP MkII – IALA waterway risk assessment program. Technical report, International Association of Marine Aids to Navigation and Lighthouse Authorities (IALA), Saint-Germain-en-Laye, France, 2009.
- International Maritime Organization. *Convention on the International Regulations for Preventing Collisions at Sea, 1972 (COLREGs) — Consolidated Edition*. International Maritime Organization (IMO), London, 2018.
- International Convention for the Safety of Life at Sea (SOLAS), 1974, as amended*. International Maritime Organization (IMO), London, 1974.
- International Telecommunication Union. Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile frequency band. Recommendation ITU-R M.1371-5, International Telecommunication Union, Radiocommunication Sector, Geneva, 2014.
- Italian Coast Guard Headquarters. Messina VTS — User’s Manual. Area Procedures Edition No. 007, Ministry of Infrastructures and Transport, Italian Coast Guard Headquarters, March 2015. Last updated 05.03.2015. Traffic Separation Scheme described on pages MES 04–MES 05; scheme published by IMO with SN.1/Circ.279 on 12 June 2009.
- Andrzej S. Lenart. Collision threat parameters for a new radar display and plot technique. *Journal of Navigation*, 36(3):404–410, 1983. doi: 10.1017/S0373463300039758.
- Chenyan Lin, Rong Zhen, Yanting Tong, Shenhua Yang, and Shengkai Chen. Regional ship collision risk prediction: An approach based on encoder-decoder LSTM neural network model. *Ocean Engineering*, 296:117019, 2024. doi: 10.1016/j.oceaneng.2024.117019.
- Ryan Wen Liu, Xiaojie Huo, Maohan Liang, and Kai Wang. Ship collision risk analysis: Modeling, visualization and prediction. *Ocean Engineering*, 266:112895, 2022. doi: 10.1016/j.oceaneng.2022.112895.
- Zhao Liu, Boyuan Zhang, Mingyang Zhang, Helong Wang, and Xiuju Fu. A quantitative method for the analysis of ship collision risk using AIS data. *Ocean Engineering*, 272:113906, 2023. doi: 10.1016/j.oceaneng.2023.113906.
- Massimiliano Marino, Luca Cavallaro, Elisa Castro, Rosaria Ester Musumeci, Matteo Martignoni, Federico Roman, and Enrico Foti. New frontiers in the risk assessment of ship collision. *Ocean Engineering*, 274:113999, 2023. doi: 10.1016/j.oceaneng.2023.113999.

- Emanuel Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962. doi: 10.1214/aoms/1177704472.
- Preben Terndrup Pedersen. Collision and grounding mechanics. In *Proceedings of WEMT'95, The West European Conference on Marine Technology*, pages 125–157, Copenhagen, Denmark, 1995.
- H. Rong, A. P. Teixeira, and C. Guedes Soares. Risk of ship near collision scenarios off the coast of Portugal. In Michael Beer and Enrico Zio, editors, *Proceedings of the 29th European Safety and Reliability Conference (ESREL 2019)*, pages 3660–3666, Hannover, Germany, 2019. Research Publishing. doi: 10.3850/978-981-11-2724-3_0625-cd.
- H. Rong, A. P. Teixeira, and C. Guedes Soares. Spatial correlation analysis of near ship collision hotspots with local maritime traffic characteristics. *Reliability Engineering & System Safety*, 209:107463, 2021. doi: 10.1016/j.ress.2021.107463.
- Murray Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956. doi: 10.1214/aoms/1177728190.
- B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London, 1986.
- P. Sotiralis, N. P. Ventikos, R. Hamann, P. Golyshev, and A. P. Teixeira. Incorporation of human factors into ship collision risk models focusing on human centred design aspects. *Reliability Engineering & System Safety*, 156:210–227, 2016. doi: 10.1016/j.ress.2016.08.007.
- Weibin Zhang, Floris Goerlandt, Jakub Montewka, and Pentti Kujala. A method for detecting possible near miss ship collisions from AIS data. *Ocean Engineering*, 107:60–69, 2015. doi: 10.1016/j.oceaneng.2015.07.046.
- Weibin Zhang, Floris Goerlandt, Pentti Kujala, and Yinhai Wang. An advanced method for detecting possible near miss ship collisions from AIS data. *Ocean Engineering*, 124:141–156, 2016. doi: 10.1016/j.oceaneng.2016.07.059.
- Weibin Zhang, Xinyu Feng, Floris Goerlandt, and Qing Liu. Towards a convolutional neural network model for classifying regional ship collision risk levels for waterway risk analysis. *Reliability Engineering & System Safety*, 204:107127, 2020. doi: 10.1016/j.ress.2020.107127.

Ringraziamenti

Desidero ringraziare innanzitutto i relatori, prof. Michele Martelli e prof. Tomaso Gaggero, e il correlatore, dott. Filippo Ponzini, per il tempo dedicato alle revisioni e per le osservazioni che mi hanno guidato in questo lavoro.

Ringrazio il corpo docenti del corso di laurea, per quanto mi ha insegnato in questi anni.

Grazie alla città della Spezia, che mi ha ospitato per questi cinque lunghi anni, permettendomi di trasformare i miei studi in una passione, e all'Accademia, che rendeva un pomeriggio di studio più piacevole.

Un ringraziamento speciale va alla mia famiglia, che mi ha sostenuto in tutto il percorso.

Grazie ai Five Guys e agli amici e colleghi conosciuti qui, con cui ho condiviso la vita universitaria, viaggi, esperienze di ogni tipo e forti emozioni.

E grazie agli amici di giù, che anche da lontano ci sono sempre stati e ci saranno sempre.

Grazie a tutti.

Paolo Pedalino