

DIPARTIMENTO DI  
ECONOMIA E FINANZA

**METODI E ANALISI  
STATISTICHE  
2022**





**DIPARTIMENTO DI  
ECONOMIA E FINANZA**

**METODI E ANALISI  
STATISTICHE**

**2022**



**UNIVERSITÀ  
DEGLI STUDI DI BARI  
ALDO MORO**

Tutti i diritti di traduzione, riproduzione e adattamento, totale o parziale, con qualsiasi mezzo (comprese le copie fotostatiche e i microfilm) sono riservati

Toma E., (a cura di). *Metodi e Analisi Statistiche 2022*, Dipartimento di Economia e Finanza, Università degli studi di Bari Aldo Moro.

© Copyright 2022 by Università degli Studi di Bari Aldo Moro  
[www.uniba.it](http://www.uniba.it)

Prima edizione: dicembre 2022

ISBN 978-88-6629-078-0

Gli articoli qui presentati sono stati tutti oggetto di valutazione interna.

*Editing finale:* F.D. d'Ovidio, E. Toma

# Sommario

Ernesto Toma <b>Presentazione</b> .....	pag. 5
Giovanni Armillotta, Crescenza Calculli, Alessio Pollice, Letizia Sion, Porzia Maiorano, Francesca Capezzuto, Angela Carluccio, Gianfranco D'Onghia <b>Studio della distribuzione spazio-temporale dei pesci cartilaginei nel Mar Ionio nord-occidentale</b> .....	« 7
Paola Perchinunno, Antonella Massari, Samuela L'Abbate, Lucia Mongelli <b>Sustainable Development Goals per l'analisi statistica della povertà</b> .....	« 27
Samuela L'Abbate, Paola Perchinunno <b>Un nuovo approccio al DBSCAN per raggruppare aree dense</b> .....	« 41
Angela Maria D'Uggento, Massimo Armenise, Monica Carbonara <b>A synthetic index for the analysis of the eno-gastronomic tourist attractiveness of Italian territories</b> .....	« 55
Sabrina Diomede, Giovanni Tagliatela <b>Disuguaglianze bilaterali di tipo Pearson per alcune classi di variabili aleatorie</b> .....	« 65
Francesco Domenico d'Ovidio, Angela Maria D'Uggento, Monica Cazzolle, Ernesto Toma <b>Dall'università al mercato del lavoro: un focus sulle aspettative dei laureati</b> .....	« 83
Angela Maria D'Uggento, Riccardo Roselli, Barbara Cafarelli <b>Modelli probabilistici per la gestione dell'incertezza in ambito finanziario</b> ..	« 95
Angela Maria D'Uggento, Luana Cecca, Barbara Cafarelli <b>Tecniche di text mining per l'analisi delle dichiarazioni non finanziarie delle aziende</b> .....	« 115

Dominga Capozzo, Nunziata Ribecco, Clelia Punzo, Rita Laforgia, Giustina Angellotti, Angela Pezzolla <i>Analisi statistica sulla patologia nodulare tiroidea</i> .....	« 131
Marta Schito, Nunziata Ribecco, Giustina Angellotti, Clelia Punzo <i>Analisi statistica su macro metastasi linfonodali dopo biopsia del linfonodo sentinella nel trattamento chirurgico del carcinoma mammario</i> ..	« 143
Carmine Lamanna, Barbara Cafarelli, Angela Maria D’Uggento <i>Tecniche statistiche per la previsione del rischio di insolvenza</i> .....	« 155
Laura Antonucci, Corrado Crocetta, Angelica Di Terlizzi <i>Analisi del gender gap nelle Università italiane fra il 2021 e il 2020</i> .....	« 179
Laura Antonucci, Antonia Cofano, Corrado Crocetta, Massimo Russo <i>La customer satisfaction della Sanitaservice ASL Foggia s.r.l.</i> .....	« 187
Monica Carbonara, Agata Maria Madia Carucci, Giovanni Vannella <i>Una valutazione del benessere “a misura di comune” attraverso una proposta di indicatori sintetici per il reddito e per il dominio “istruzione”</i> ...	« 203
Laura Antonucci, Giulio Biscardi, Najada Firza, Leonardo Grilli <i>Un modello ad effetti misti con dati longitudinali per il progress test dei corsi di laurea italiani in odontoiatria</i> .....	« 219
Angela Maria D’Uggento, Barbara Cafarelli, Massimo Iaquinta <i>Il fenomeno dell’abbandono universitario in Italia</i> .....	« 235

# Presentazione

*Metodi e analisi statistiche 2022* è il settimo della serie di volumi pubblicati annualmente dal 2016, ed è come sempre un momento in cui una serie di studiosi, statistici in primis ma anche di altre estrazioni scientifiche, si propongono al pubblico riportando alcuni dei contenuti delle proprie ricerche basate, come si evince dal titolo del volume, sia su metodi che su analisi statistiche.

Anche quest'anno, l'eterogeneità degli argomenti trattati la fa da padrona, ovviamente a causa (o meglio "per merito") della diversa provenienza degli autori. La contaminazione scientifica è sicuramente arricchimento culturale e, per quanto riguarda la statistica, emerge ancora una volta il ruolo centrale della stessa per tutte le analisi in ambiti tanto metodologici quanto applicati a fenomeni di ogni tipo.

Il primo contributo, di Armillotta & al., attraverso dati provenienti dal progetto internazionale MEDITS svolto sul Mar Ionio nord-occidentale, cerca di individuare, attraverso opportune metodologie multivariate (Modelli Additivi Generalizzati, Modelli Lineari Generalizzati), alcune possibili cause influenti sulla distribuzione spazio-temporale delle risorse demersali del Mar Ionio.

Con il secondo contributo, di Perchinunno & al., si passa ad affrontare un argomento di estrema attualità, visto anche il periodo di grandi difficoltà economica causato dalla pandemia. Gli autori realizzano un'analisi statistica multidimensionale (Totally Fuzzy and Relative) degli indicatori utili ai fini del raggiungimento dell'obiettivo di sviluppo sostenibile consistente nel cercare di sconfiggere la Povertà e cercano di individuare quali territori necessitino di maggiori interventi da parte delle autorità o dei policy makers.

L'Abbate e Perchinunno, si interessano invece di applicare metodi di clusterizzazione per identificare masse tumorali nelle immagini radiografiche di una mammografia. I clusters sono intesi dagli autori come delle regioni dense di spazio, separate da aree di densità inferiore. A tal fine essi utilizzano uno dei principali metodi di clustering basato sulla densità, il DBSCAN, proponendone un nuovo approccio che, attraverso curve ROC e distanza ponderata, consenta di scegliere i parametri di input.

Nel quarto articolo, D'Uggento, Armenise e Carbonara utilizzano dati ISTAT a livello comunale per studiare la distribuzione territoriale del turismo enogastronomico italiano ed è quasi superfluo sottolineare l'importanza di un simile studio, visto che il settore turistico è ormai a pieno titolo tra quelli maggiormente trainanti dell'economia del nostro paese.

Dopo la nota in cui Diomede e Tagliatela affrontano il problema metodologico delle disuguaglianze bilaterali di tipo Pearson per alcune classi di variabili aleatorie, d'Ovidio et al. presentano taluni dati estrapolati dalla ricerca PETERE (Preferences for Employment and Training as Elected by recent Graduates) che coinvolge, oltre all'Università degli studi di Bari, molte altre sedi universitarie di ogni parte d'Italia. In tale contributo, gli autori cercano di individuare le difficoltà che i laureati affrontano nel mercato del lavoro e di far emergere le loro capacità di far fronte alle difficoltà di tale

mercato, grazie, talvolta, al supporto dell'università stessa. Dal punto di vista metodologico si segnala l'applicazione di tecniche di post-stratificazione per sopperire alla purtroppo ridotta dimensione campionaria.

Nel settimo contributo, di D'Uggento, Roselli e Cafarelli, si riscontra un connubio tra il mondo della probabilità e quello del mercato finanziario. Gli autori, a tal fine, realizzano una interessante disamina dei principali modelli probabilistici in campo finanziario.

Sempre D'Uggento e Cafarelli, con Cecca, nel successivo articolo trattano ancora una volta un fenomeno economico, in questo caso le dichiarazioni non finanziarie delle aziende, usando delle tecniche di text mining.

I due contributi successivi affrontano invece, sempre con metodo statistico, delle tematiche mediche. Capozzo et al. svolgono una interessante analisi sulla patologia nodulare tiroidea cercando di individuare quali variabili influiscono maggiormente sulla diagnosi e sulla variabilità del fenomeno. Schito et al., invece, studiano le variabili potenzialmente influenti nella linfadenectomia ascellare dopo biopsia positiva del linfonodo nel trattamento del carcinoma mammario. Gli autori, attraverso un campione di donne trattate chirurgicamente dal 2008 al 2021 presso l'Azienda Ospedaliera Universitaria Policlinico di Bari, hanno voluto verificare se nelle pazienti fosse possibile individuare le caratteristiche predittive dell'eventuale negatività dei restanti linfonodi rimossi attraverso l'asportazione ascellare.

Il contributo di Lamanna, Cafarelli e D'Uggento si sofferma sulle tecniche statistiche per la previsione del rischio di insolvenza, tema di fondamentale importanza, specie in questo periodo di difficoltà a livello mondiale, in quanto una previsione errata della probabilità di insolvenza può produrre rating inadeguati nel supportare per esempio le banche nella concessione di prestiti.

Dopo l'intervento di Antonucci, Crocetta e Di Terlizzi, che in una breve nota fanno il punto sul gender gap nelle Università italiane negli anni 2020 e 2021, Antonucci et al. riportano i principali risultati di un'analisi di customer satisfaction nell'ambito della ASL Foggia con dati rilevati da Sanitaservice.

La commistione tra metodologia statistica e ambito socio-economico, viene invece affrontata da Carbonara, Carucci e Vannella che propongono degli indicatori sintetici per il reddito e il dominio "istruzione", utilizzando dati ISTAT a livello comunale.

Infine, dopo un contributo di Antonucci et al. che si occupano dei Corsi di Laurea in Odontoiatria mediante i risultati dei Progress Test, utilizzando da un punto di vista metodologico un modello ad effetti misti con dati longitudinali, il volume si conclude con un'analisi di D'Uggento, Cafarelli e Iaquina sul fenomeno dell'abbandono in Italia. In quest'ultimo lavoro, gli autori forniscono un quadro dimensionale aggiornato del fenomeno a livello nazionale attraverso dati di origine amministrativa.

Ringraziando tutti gli autori che hanno contribuito alla realizzazione del presente volume, speriamo che esso potrà riscontrare il favore dei lettori tutti.

# Studio della distribuzione spazio-temporale dei pesci cartilaginei nel Mar Ionio nord-occidentale

Giovanni Armillotta<sup>1</sup>, Crescenza Calculli<sup>1\*</sup>, Alessio Pollice<sup>1</sup>,  
Letizia Sion<sup>2</sup>, Porzia Maiorano<sup>2</sup>, Francesca Capezzuto<sup>2</sup>, Angela  
Carluccio<sup>2</sup>, Gianfranco D'Onghia<sup>2</sup>

<sup>1</sup>Dipartimento di Economia e Finanza, Università degli Studi di Bari Aldo Moro,

<sup>2</sup>Dipartimento di Biologia, Università degli Studi di Bari Aldo Moro

**Riassunto:** Le caratteristiche ambientali e le attività antropiche legate alla pesca a strascico sono fattori che possono determinare cambiamenti nelle comunità marine. Questo lavoro propone l'analisi di alcune possibili cause influenti sulla distribuzione spazio-temporale delle risorse demersali del Mar Ionio. Il lavoro si basa su dati raccolti durante il progetto internazionale MEDITS relativamente alla Geographic Sub-Area 19 (Mar Ionio nord-occidentale). L'analisi si focalizza sulla presenza e su alcune misure di abbondanza quali, densità e biomassa, dell'intera classe dei pesci cartilaginei. Tramite l'utilizzo dei Modelli Additivi Generalizzati sono state ottenute mappe continue sull'area di studio per i descrittori della dinamica della popolazione di interesse. Inoltre, è stata analizzata la significatività degli effetti di due variabili collegate allo sforzo di pesca e di due variabili ambientali sulla densità, biomassa e presenza dei pesci cartilaginei. Infine, è stata analizzata la dipendenza del numero di specie di pesci cartilaginei dalle stesse quattro variabili esplicative. L'uso di Modelli Lineari Generalizzati per variabili di conteggio ha permesso di evidenziare la significatività degli effetti delle quattro variabili. Quest'ultima analisi è stata condotta sia sull'intera GSA19 sia su una ripartizione del dominio nelle due sub-aree pugliese e calabrese.

**Keywords:** Modelli additivi generalizzati (GAM); Modelli lineari generalizzati (GLM); Modelli spazio-temporali; zero-inflation; Modelli per dati di conteggio.

---

\* Autore corrispondente: [crescenza.calculli@uniba.it](mailto:crescenza.calculli@uniba.it)

Il lavoro qui descritto è frutto di un progetto comune, ma G. Armillotta ha provveduto alla redazione dei paragrafi 1, 2 e 2.1, mentre C. Calculli ha redatto il paragrafo 2.2, 3 e 4. La supervisione del lavoro di redazione e di elaborazione dei dati è da attribuirsi ad A. Pollice; la concettualizzazione e il disegno dello studio è frutto del lavoro di L. Sion e G. D'Onghia mentre l'esecuzione e la raccolta dei dati di L. Sion, P. Maiorano, F. Capezzuto e A. Carluccio.

## 1. Introduzione

Il Mar Mediterraneo è un complesso sistema biogeografico che comprende una grande varietà di aree specifiche. Mentre la maggior parte degli stock ittici è definita a livello nazionale, il loro sfruttamento è frazionato in un vasto numero di porzioni specifiche (definite dalla tipologia di navi, dalle tecniche di pesca, dai porti di sbarco, dalle specie catturate, ecc.). A tal riguardo, le informazioni globali disponibili sullo stato delle risorse della pesca (risorse alieutiche) risultano essere poche e poco omogenee. Il progetto di ricerca MEDITS (MEDiterranean International Trawl Surveys) è stato avviato nel 1994 con l'obiettivo di condurre un monitoraggio sulle risorse alieutiche demersali (di fondale) con metodologie, attrezzi e protocolli di campionamento standardizzati per tutti i partecipanti (Bertrand, 2002). Attualmente i Paesi coinvolti sono 10: Spagna, Francia, Italia, Slovenia, Croazia, Montenegro, Albania, Grecia, Malta e Cipro.

Al largo delle coste di questi Paesi, sono realizzate campagne sperimentali di pesca a strascico tra 10 e 800 m di profondità. Il progetto MEDITS considera una suddivisione del Mar Mediterraneo in 30 sub-aree denominate GSA (Fig. 1). Questo lavoro riguarda l'analisi della sola GSA 19 (Mar Ionio nord-occidentale) che si estende per circa 16.500 km<sup>2</sup> interessando più di 1.000 km di costa delle regioni Puglia, Basilicata, Calabria e Sicilia, da Capo d'Otranto (Lecce) sino a Capo Passero (Siracusa). Il bacino nord-occidentale del Mar Ionio è ulteriormente suddiviso dal canyon di Taranto in due sub-aree, approssimativamente identificabili con le coste di Puglia ad est e Basilicata, Calabria e Sicilia ad ovest, differenti fra loro per caratteristiche geomorfologiche e idrografiche. Nella GSA 19 sono state realizzate annualmente 74 stazioni sperimentali di pesca a strascico (4 di esse presenti solo dal 1995 al 2001). Le stazioni sono state individuate applicando uno schema di campionamento stratificato con scelta casuale all'interno di ogni strato. Il parametro di stratificazione adottato è la *profondità*, con i seguenti limiti batimetrici: 10, 50, 100, 200, 500 e 800 m. La posizione di ciascuna cala varia solo per piccoli scostamenti accidentali da un anno all'altro ed è stata selezionata casualmente all'interno di piccole sotto aree definite per ottenere un compromesso tra campionamento casuale e rappresentatività degli strati.

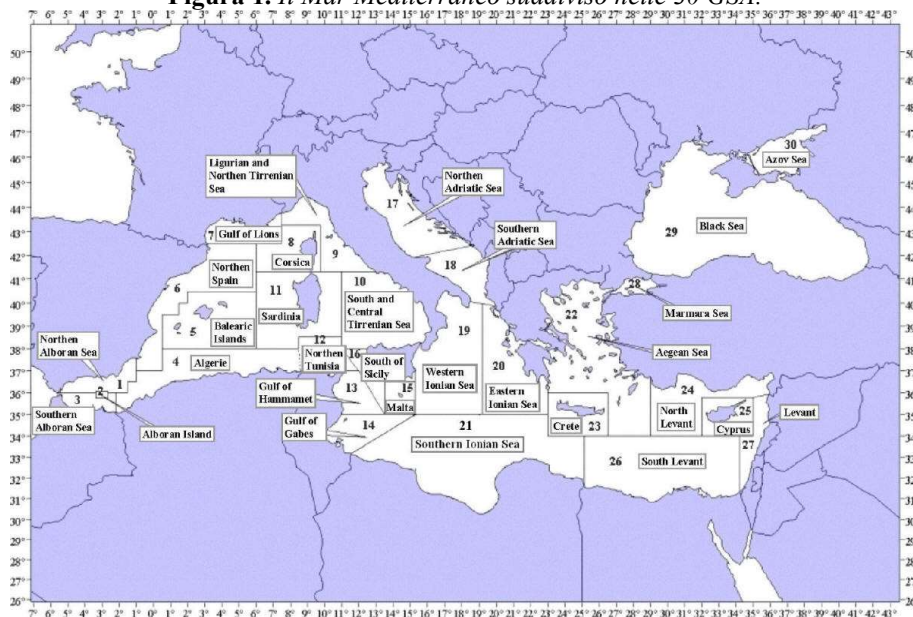
L'obiettivo di questo lavoro è quello di indagare la distribuzione spazio-temporale dei pesci cartilaginei (*Chondrichthyes* o Condritti) presenti nell'area di studio in relazione ad alcune variabili ambientali e a fattori antropici legati allo sforzo di pesca. Le variabili considerate per descrivere la dinamica spazio-temporale della distribuzione della classe dei Condritti sono state rilevate puntualmente per ciascuna

stazione e fanno riferimento a misure di abbondanza quali la *densità* e la *biomassa*, alla *presenza* e al *numero* di specie (taxa) nell'area. Le variabili *densità* e *biomassa* sono caratterizzate da un eccesso di valori nulli che può essere rappresentato, nella stima dei modelli GAM (Wood, 2017), da specifiche assunzioni distributive (Jørgensen, 1997). Per la presenza delle specie, invece, l'approccio modellistico GAM prevede un modello con risposta dicotomica.

Per tutti i modelli, la relazione di dipendenza tra la variabile risposta e le variabili esplicative viene descritta da funzioni *smooth* non lineari per la modellazione della componente spazio-temporale e da effetti lineari delle variabili esplicative considerate, con l'obiettivo di pervenire a mappe di previsione riferite all'intervallo temporale 1995 - 2020.

Inoltre, è stata investigata la serie storica del numero di specie della classe dei pesci cartilaginei per l'intera GSA 19 in relazione sia a variabili ambientali che relative allo sforzo di pesca. In questo caso sono stati utilizzati dei modelli lineari generalizzati per variabili di conteggio con assunzione quasi-Poisson (Hastie e Pregibon, 1992). La stessa analisi è proposta tenendo conto anche della distinzione tra due sub-aree: versante ionico salentino (sal) e versante ionico calabro-siculo (cal).

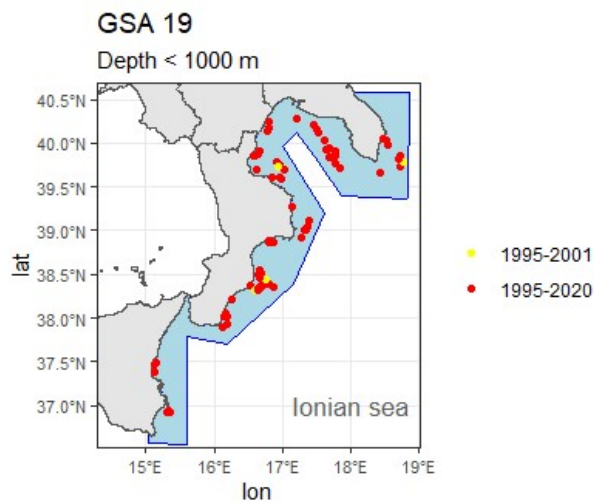
**Figura 1. Il Mar Mediterraneo suddiviso nelle 30 GSA.**



## 2. Materiali e metodi

I dati analizzati si riferiscono all'intervallo temporale 1995 - 2020 e consistono in 1848 record corrispondenti a 74 cale per i primi 7 anni e a 70 cale per i successivi 19 anni. La distribuzione delle cale nella GSA19 è riportata in Fig. 2. Le variabili di interesse sono state sottoposte preliminarmente ad un'approfondita analisi esplorativa i cui risultati, qui solo parzialmente riportati per brevità, sono stati determinanti nella scelta dei modelli di analisi utilizzati successivamente. In particolare, la specificazione delle assunzioni distributive e delle variabili esplicative dei GAM e dei GLM ha risentito fortemente delle deduzioni effettuate in questa fase.

**Figura 2.** GSA19: allocazione delle cale relative al progetto MEDITS



### 2.1 Fonti dei dati, struttura del database

I record del database rappresentano le cale/stazioni e sono identificati tramite un codice alfanumerico che contiene i riferimenti all'anno di campionamento, alla stazione di monitoraggio e alla profondità del punto in cui è avvenuto il campionamento.

Per ciascuna cala, le variabili o i campi più rilevanti, ai fini della nostra analisi, sono le seguenti:

- coordinate (longitudine e latitudine) della cala;
- stazione (identificativo della cala);
- anno della rilevazione;

- *presenza* della classe dei Condritti;
- *densità* della classe dei Condritti (numero di pesci nella cala diviso la superficie dell'area strascicata);
- *biomassa* della classe dei Condritti (peso dei pesci diviso la superficie dell'area strascicata);
- *numero di specie* della classe dei Condritti;
- potenza imbarcazioni (in hp);
- numero imbarcazioni;
- temperatura al fondo;
- salinità.

Le prime tre variabili identificano le singole rilevazioni, le successive quattro variabili sono state scelte per descrivere la dinamica della popolazione in esame, potenza e numero di imbarcazioni sono proxy dell'impatto antropico dovuto allo sforzo di pesca, infine, i fattori abiotici salinità e temperatura al fondo, sono stati considerati per spiegare eventuali alterazioni ambientali derivanti dal cambiamento climatico. I dati riferiti alle variabili ambientali sono stati ottenuti dalla piattaforma *Copernicus* (Escudier et al., 2020) mentre l'informazione sullo sforzo di pesca deriva dalla consultazione dei registri di *Fishing fleet capacities* dell'EU consultabili all'indirizzo [https://oceans-and-fisheries.ec.europa.eu/fisheries/rules/fishing-fleet-capacities\\_en](https://oceans-and-fisheries.ec.europa.eu/fisheries/rules/fishing-fleet-capacities_en). Per le ultime cinque variabili in elenco, sono state considerate sia le misure aggregate per l'intera GSA19 che distinte nelle due sub-aree: versante ionico salentino (sal) e versante ionico calabro-siculo (cal).

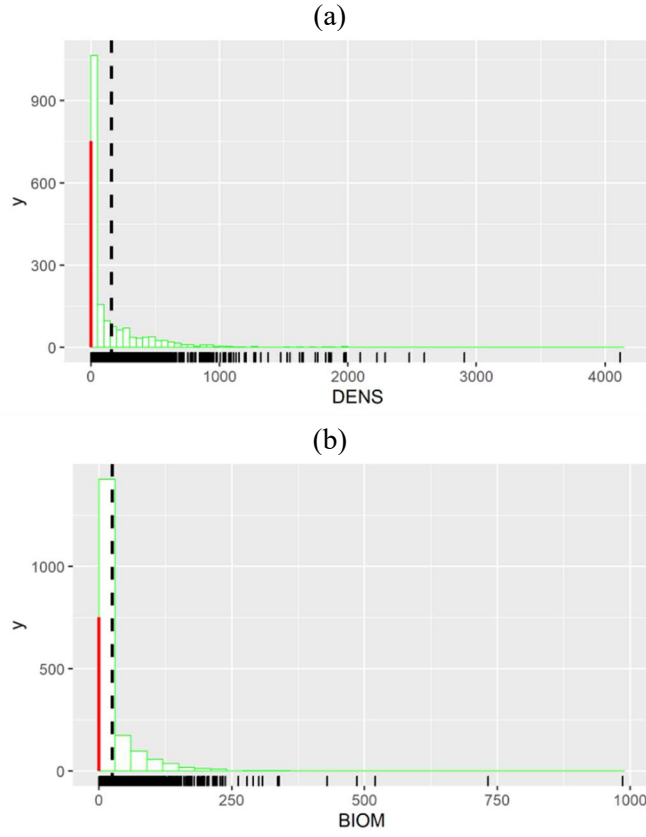
## **2.2 Metodi di analisi utilizzati**

Al fine di investigare i fattori che influenzano la distribuzione spazio-temporale della comunità di pesci cartilaginei nel Mar Ionio, sono stati considerati quattro potenziali fattori influenti. In particolare, fattori abiotici come la *temperatura* al fondo e la *salinità* nonché fattori antropici legati allo sforzo di pesca (il *numero di imbarcazioni* e la *potenza delle imbarcazioni*), sono stati investigati rispetto a tre principali descrittori della dinamica della popolazione in oggetto, quali *presenza*, *densità* e *biomassa* riferiti alla classe faunistica dei *Chondrichthyes*.

L'analisi esplorativa ha evidenziato un chiaro comportamento evolutivo delle quattro variabili esplicative (Fig. 5) insieme ad una forte correlazione tra le stesse (Fig. 6). Per superare pertanto il problema della collinearità e degli effetti confondenti dei regressori nell'ambito di un modello di regressione multipla, sono stati sti-

mati diversi modelli additivi generalizzati o GAM (Wood, 2017) nei quali la dipendenza di ciascuna variabile risposta dai fattori abiotici e antropici è stata valutata considerando individualmente l'effetto lineare di ciascun fattore. Sulla base delle assunzioni distributive riferite alle variabili risposta considerate sono stati utilizzati modelli GAM spazio-temporali non separabili con risposta binomiale per modellare la presenza dei pesci cartilaginei mentre, per le misure di abbondanza (*densità* e *biomassa*), sono stati stimati modelli GAM con risposta Tweedie, assunzione distributiva che tiene conto dell'eccesso di zeri che caratterizza le distribuzioni delle due variabili semi-continue. Gli istogrammi in Fig. 3 mostrano, per le variabili *densità* e *biomassa*, la frequenza elevata di zeri nelle distribuzioni delle due risposte, suggerendo una trattazione appropriata al caso di dati *zero-inflated*. Il problema noto in letteratura come *zero-inflation* (Zuur *et al.*, 2009) descrive la situazione in cui la frequenza dei valori pari a zero osservati è più alta rispetto alla probabilità attesa sotto le più comuni assunzioni distributive (*e.g.* log-normale, Poisson, ecc.). L'elevata proporzione di zeri richiede che questi dati siano modellati considerando due tipi di approccio: quello dei modelli Zero-Inflated (ZI, Lambert, 1992) e quello dei modelli Hurdle (Mullahy 1986; Heilbron 1994). I modelli ZI e Hurdle differiscono in base alla loro concettualizzazione degli zeri. Se nel primo caso, infatti, gli zeri sono generati da una mistura di due distribuzioni, una che modella gli zeri "strutturali" e l'altra che modella gli zeri "veri", il secondo approccio presuppone che tutti gli zeri provengano da uno stesso processo con una parte binaria che modella la probabilità di zero ed un'altra parte che utilizza un modello troncato, come un Poisson troncato (per dati di conteggio) o una distribuzione Gamma per dati continui. In questo lavoro è stato utilizzato il primo approccio. Più precisamente, si è assunta alla base dei GAM la famiglia di distribuzioni di probabilità Tweedie che fa parte dei modelli esponenziali di dispersione (EDM, Jørgensen, 1997) e comprende la classe delle distribuzioni composte Poisson-Gamma. In generale, nell'ambito degli EDM, queste distribuzioni sono caratterizzate da una relazione media-varianza governata da due parametri,  $\mu$  e  $\phi$  (media e parametro di dispersione, rispettivamente) e da un terzo parametro  $p$  per cui, data una generica v. c.  $Y$ , vale  $E(Y) = \mu$ ,  $\text{Var}(Y) = \phi\mu^p$ . Nel caso particolare di  $1 < p < 2$ , la distribuzione Tweedie può essere specificata come una mistura di Poisson di distribuzioni Gamma particolarmente idonea a rappresentare dati caratterizzati da una massa positiva a zero e valori continui non-negativi.

**Figura 3.** Distribuzione della (a) densità e (b) biomassa dei *Chondrichthyes*. Il segmento in rosso mostra la frequenza degli zeri. Il segmento nero tratteggiato indica il valore medio.



Dato quindi il numero di cale,  $i = 1, \dots, 74$  e gli anni,  $j = 1, \dots, 26$ , per ogni variabile risposta  $Y_{ij}$  (*presenza, densità, biomassa*), i GAM sono stati specificati nel modo seguente:

$$Y_{ij} \sim \text{Bin} \text{ o } Y_{ij} \sim \text{Tw}$$

$$g(\mu_{ij}) = \beta_0 + \text{te}(\text{lon}_{ij}, \text{lat}_{ij}, \text{anno}_j) + \beta_1 x_{ij} = \eta(x_{ij})$$

dove  $g()$  è una funzione link appropriata che lega il predittore  $\eta$ , alla media della risposta,  $\text{te}()$  è il prodotto tensoriale di funzioni smooth, mentre il parametro  $\beta_1$  rappresenta il coefficiente di regressione associato a ciascuna covariata inserita a turno nel modello. La funzione semiparametrica  $\text{te}()$  considerata è una funzione non isotropica utile per rappresentare gli effetti simultanei di predittori multipli misurati in unità di misura diverse o con grado di lisciamento diverso (come per il caso delle

coordinate spaziali e dei punti temporali). Le spline del prodotto tensoriale possono essere espresse come combinazioni lineari di *funzioni base* che permettono di cogliere la struttura non lineare presente nei dati. In particolare, la stima di queste funzioni dipende dalla scelta della base e dal numero di nodi che condiziona lo smoothing di ciascuna funzione. Nel caso della stima di un elevato numero di funzioni smooth non lineari il problema dell'overfitting può essere superato utilizzando una penalità per controllare il grado di lisciamiento dell'adattamento. Per la stima dei modelli GAM tale approccio consiste nel massimizzare la verosimiglianza modificata aggiungendo una penalità per ogni funzione smooth utilizzando il metodo denominato Penalized Iteratively Reweighted Least Squares (P-IRLS, Wood, 2006). Per quanto riguarda il caso di studio, le basi utilizzate per il prodotto tensoriale sono state scelte considerando la natura delle covariate. In particolare, l'interpolazione spaziale è ottenuta utilizzando uno smoother di tipo *soap-film* (Wood, 2008) che tiene conto della complessità del dominio introducendo un vincolo sul contorno inteso come linea di costa e linea batimetrica (isobata), mentre per l'effetto temporale si è utilizzata una interpolazione polinomiale cubica. La dimensione delle basi è stata scelta inoltre facendo un compromesso tra il numero di nodi e l'efficienza della stima.

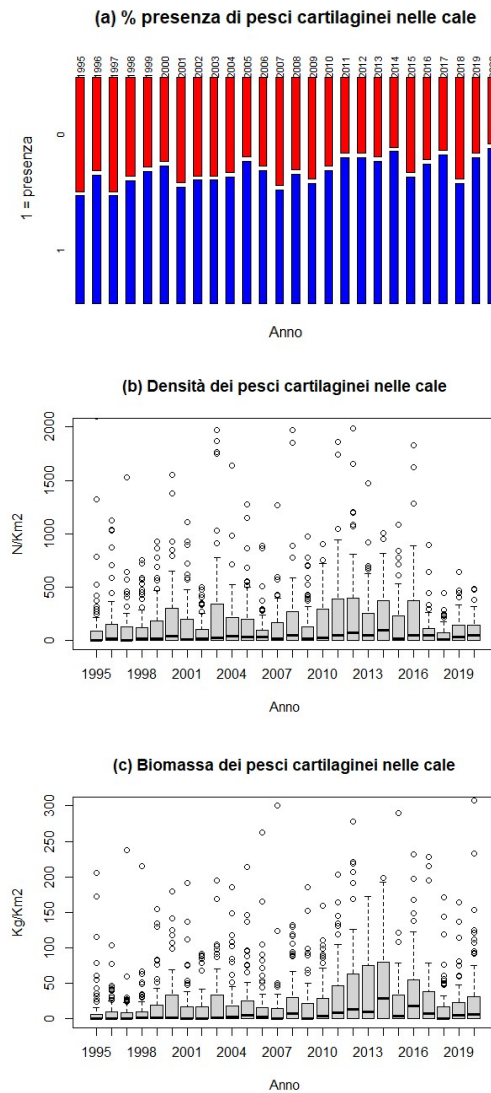
Per investigare inoltre l'influenza dello sforzo di pesca e dei fattori ambientali sul numero di specie (*numero taxa*) nei diversi anni in studio, sono stati stimati alcuni modelli lineari generalizzati (GLM) con assunzione distributiva Quasi-Poisson considerando l'intera GSA19 e la sua ripartizione in due sub-aree. Questi modelli permettono di trattare il problema, comune nell'ambito dell'analisi dei dati discreti, della *sovradisersione* che si verifica quando per una v.c. di Poisson si ha che  $Var(Y) > E(Y)$ . Poiché in genere il meccanismo di generazione della sovradisersione non è noto o non si è in grado di specificarlo precisamente, la strategia più appropriata è quella di considerare un modello distributivo sovradisperso alla base del GLM, i cui parametri vengono stimati mediante quasi-verosimiglianza (Hastie e Pregibon, 1992). Per i GLM stimati si è verificata inoltre l'assenza di autocorrelazione temporale dei residui che rende di fatto superflua la necessità di modellare l'effetto temporale poiché completamente assorbito dagli effetti delle covariate ambientali e di sforzo di pesca.

Tutte le analisi presentate sono state condotte utilizzando i pacchetti `mgcv`, `sf` e `tidyverse` del software statistico R (R Core Team, 2022).

### 3. Risultati

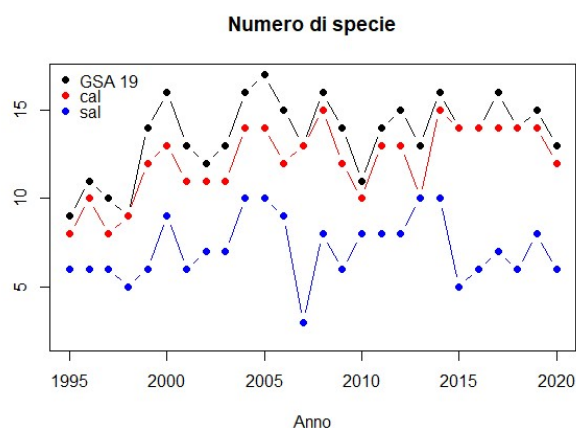
I risultati dell'analisi esplorativa riferita alle variabili investigate sono riportati nelle Figg. 4-7.

**Figura 4.** Andamento temporale della presenza (a), densità (b) e biomassa (c) dei *Chondrichthyes* catturati nella GSA 19. In (b) e (c) sono stati esclusi rispettivamente i 7 e i 10 valori più elevati.



L'evoluzione temporale di *presenza*, *densità* e *biomassa* dei pesci appartenenti alla classe dei cartilaginei in Fig. 4, suggerisce un andamento crescente delle 3 variabili risposta nell'arco temporale considerato. In particolare, mentre per la *densità* la crescita risulta meno evidente rispetto alle altre due variabili, per la *biomassa* si osserva un incremento più accentuato fino al 2014.

**Figura 5.** *Andamento temporale del numero di specie della classe dei Chondrichthyes nella GSA19 e nelle due sub-aree calabrese e salentina*

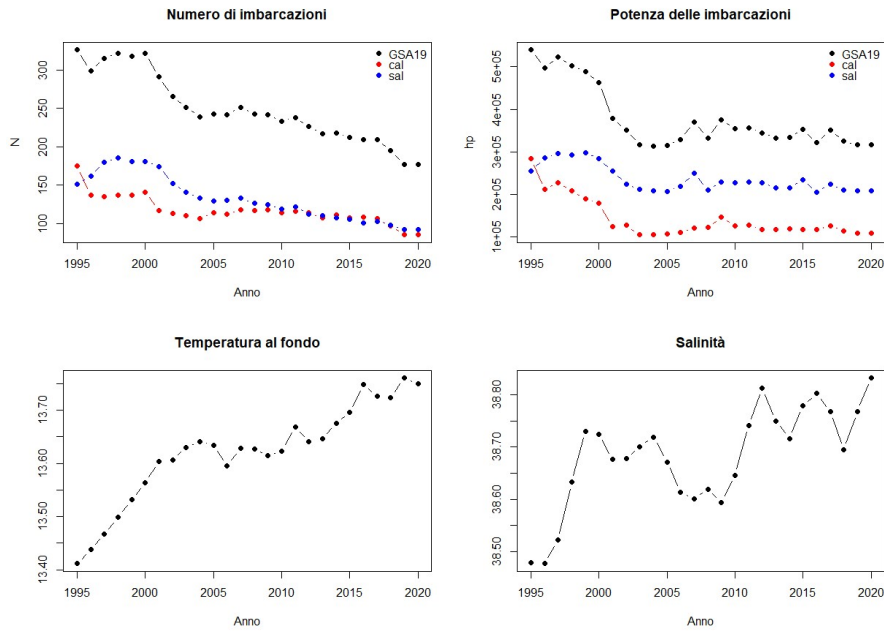


Il *numero di specie* in Fig. 5. mostra, allo stesso modo, un incremento fino al 2006 nell'intera area di studio. Si può inoltre osservare come l'area calabrese, rispetto all'area pugliese, sembri trainare l'aumento, con un picco di 15 specie rilevate nel 2008. Per gli anni successivi si evidenzia invece una certa stabilizzazione del numero di specie con riferimento sia all'intera area che alle due sub-aree individuate.

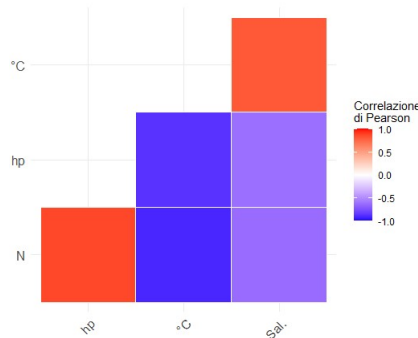
L'analisi esplorativa degli effetti dei fattori antropici sulla presenza e abbondanza dei *pesci cartilaginei* (Fig. 6) suggerisce un chiaro andamento decrescente del numero e della potenza delle imbarcazioni per la pesca a strascico in tutta la GSA19, più accentuato nell'area calabrese. Per i parametri ambientali investigati si osserva invece un comportamento evolutivo di crescita, con la *temperatura al fondo* che aumenta costantemente nei 26 anni e la *salinità* che, dopo un brusco aumento fino al 1999, continua a crescere con minore intensità ed in modo discontinuo. L'evidente comportamento evolutivo dei quattro fattori considerati suggerisce una potenziale correlazione tra le variabili ambientali e antropiche per la GSA19, investigata tramite il calcolo di coefficienti di correlazione di Pearson per le coppie di variabili. Il risultato, riportato nella heatmap in Fig. 7, mostra valori estremamente elevati e positivi tra le due variabili associate allo sforzo di pesca e tra la *salinità* e la *temperatura al*

fondo. La correlazione negativa più accentuata (spuria) si riscontra invece tra *temperatura al fondo* e il *numero di imbarcazioni da pesca a strascico*.

**Figura 6.** *Andamento temporale del numero e della potenza motore delle imbarcazioni, della temperatura al fondo e della salinità.*



**Figura 7.** *Heatmap del coefficiente di correlazione di Pearson per il numero e la potenza motore delle imbarcazioni, la temperatura al fondo e la salinità.*



Come riportato nel Paragrafo 2.2, la forte correlazione tra le variabili esplicative porta ad escludere la possibilità di considerare simultaneamente gli effetti ambientali e antropici in un unico modello di dipendenza multivariato. Pertanto, al fine di superare potenziali problemi di confondimento e multicollinearità, la verifica della significatività degli effetti delle quattro variabili (due antropiche e due ambientali) su presenza, densità e biomassa è stata ottenuta attraverso la stima di modelli GAM in cui, insieme alla componente spazio-temporale, gli effetti lineari dei fattori esplicativi vengono introdotti individualmente nel predittore. Le stime di questi modelli sono riportate in Tab. 1 dove, coerentemente con quanto atteso, si evidenzia la significatività degli effetti positivi di *salinità* e *temperatura al fondo* e, per la sola biomassa, degli effetti negativi delle due variabili legate allo sforzo di pesca.

**Tabella 1.** *Stime tramite modelli GAM degli effetti lineari del Numero e della Potenza motore delle imbarcazioni, della Temperatura al fondo e della Salinità sulla presenza, la densità e la biomassa dei Chondrichthyes.*

<i>Chondrichthyes</i>		Coef.	p-value
<b>Presenza</b>	Numero di imbarcazioni	-0.005	0.231
	Potenza delle imbarcazioni	-1.309e-06	0.459
	Temperatura al fondo	<b>2.904</b>	<b>0.015</b>
	Salinità	<b>2.717</b>	<b>&lt; 0.001</b>
<b>Densità</b>	Numero di imbarcazioni	0.003	0.246
	Potenza delle imbarcazioni	5.994e-08	0.959
	Temperatura al fondo	1.143	0.115
	Salinità	<b>2.077</b>	<b>&lt; 0.001</b>
<b>Biomassa</b>	Numero di imbarcazioni	<b>-0.012</b>	<b>&lt; 0.001</b>
	Potenza delle imbarcazioni	<b>-5.392e-06</b>	<b>&lt; 0.001</b>
	Temperatura al fondo	<b>4.477</b>	<b>&lt; 0.001</b>
	Salinità	<b>4.443</b>	<b>&lt; 0.001</b>

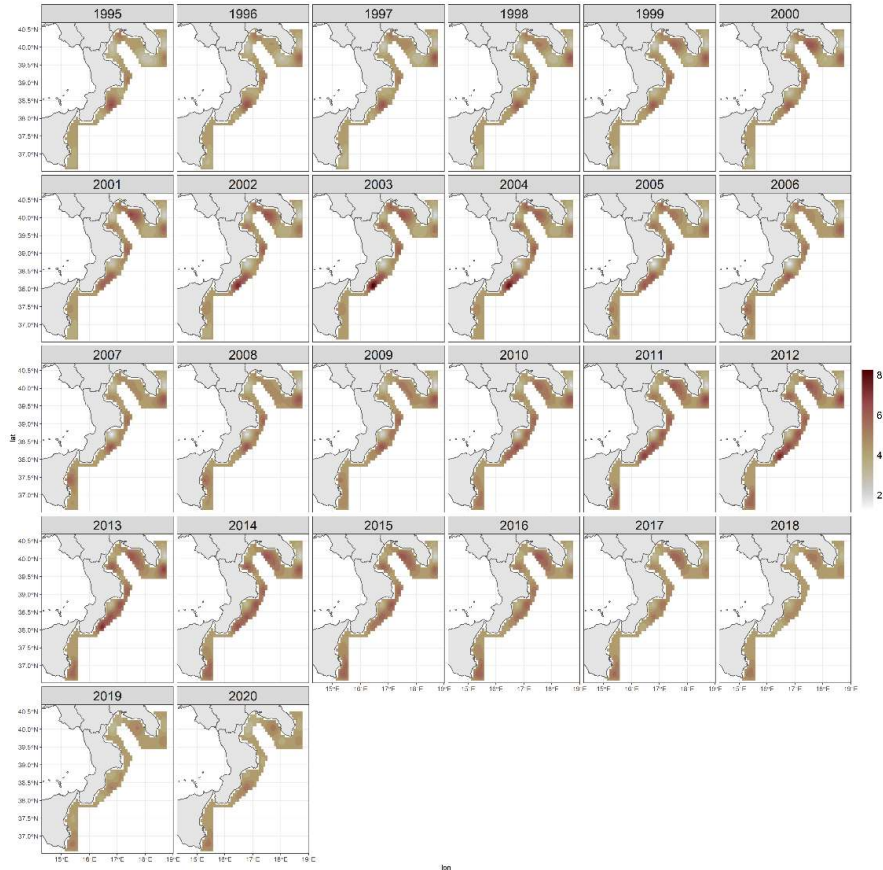
L'andamento evolutivo crescente delle prime due variabili e decrescente delle variabili di effort porta a concludere in favore di un aumento dei valori di presenza, densità e biomassa nell'arco temporale considerato, più accentuato per la biomassa, quindi per la presenza della classe, infine per la densità che risulta rispondere al solo effetto della *salinità*. Tale considerazione sembra essere confermata dalla previsione della distribuzione spazio-temporale delle tre variabili, riportata nelle mappe delle Figg. 8-10. In particolare, la presenza della classe dei pesci cartilaginei nelle cale (Fig. 8) cresce nei 26 anni in corrispondenza degli hot-spot a largo di S. M. di Leuca, nella zona del Crotonese e di Roccella Ionica e lungo la costa siciliana. Per queste aree viene stimata infatti, negli anni, una maggiore probabilità di presenza, mentre

per le misure di abbondanza, la previsione spatio-temporale riportata su scala logaritmica, suggerisce un incremento meno evidente soprattutto con riferimento alla *densità* (Fig. 9) se si considera l'intero arco temporale di studio. Dalle mappe si può tuttavia apprezzare un maggiore incremento in termini di densità e biomassa nella zona calabrese di Roccella ionica, lungo l'arco ionico della Puglia e a Largo di S.M. di Leuca se si considerano i primi anni investigati e fino al 2004.

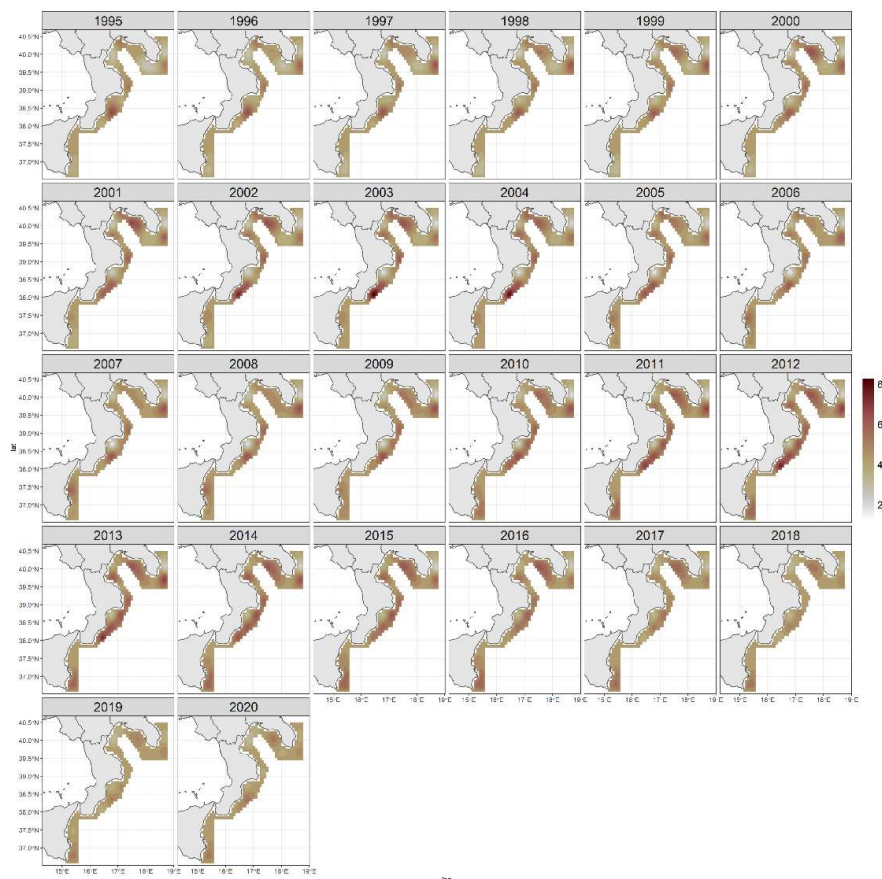
**Figura 8.** *Previsione da modello GAM della probabilità di presenza di Chondrichthyes nelle cale.*



**Figura 9.** *Previsione su scala logaritmica da modello GAM della densità dei Chondrichthyes nelle cale.*



**Figura 10.** Previsione su scala logaritmica da modello GAM della biomassa dei *Chondrichthyes* nelle cale.

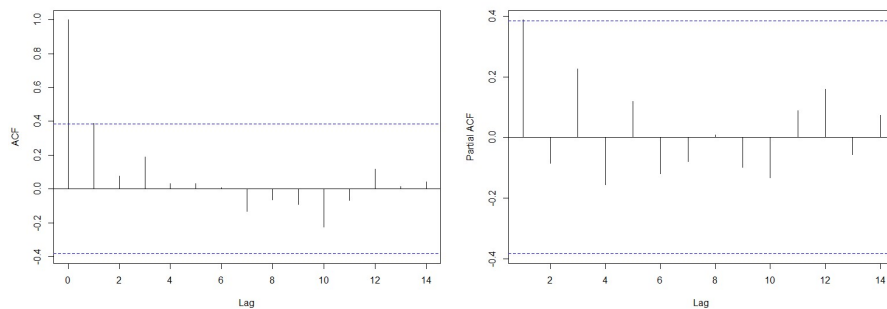


**Tabella 2.** Stime tramite modelli GLM Quasi-Poisson degli effetti lineari del Numero e la Potenza delle imbarcazioni, della Temperatura al fondo e della Salinità sul numero di taxa per la classe dei *Chondrichthyes*.

<i>Number of Taxa</i>		Coef.	p-value
<b>GSA19</b>	Numero di imbarcazioni	-0.082	0.012
	Potenza delle imbarcazioni	-0.112	0.002
	Temperatura al fondo	0.102	0.001
	Salinità	0.091	0.005
<b>area calabria</b>	Numero di imbarcazioni	-0.104	0.002
	Potenza delle imbarcazioni	-0.130	< 0.001
<b>area salentina</b>	Numero di imbarcazioni	-0.052	0.316
	Potenza delle imbarcazioni	-0.113	0.031

Per quanto riguarda invece il *numero di taxa*, nella Tab. 2 sono riportate le stime dei modelli Quasi-Poisson che tengono conto della potenziale sovradisersione nei dati di conteggio. L'autocorrelazione temporale è stata calcolata, considerando l'intera GSA19, sia per la risposta (Fig. 11) che per i residui dei modelli stimati (Fig. 12), tramite gli autocorrelogrammi delle funzioni di ACF e PACF. Il risultato conferma che l'effetto temporale presente nella risposta è completamente assorbito dalle variabili antecedenti, cosa che rende superflua la modellizzazione esplicita della componente temporale. Il risultato (non mostrato per motivi di sintesi) è del tutto analogo a quello che si ottiene se si considera la serie storica del numero di taxa per le due sub-aree calabrese e salentina. Come riportato nella Tab. 2, anche per il numero di specie gli effetti positivi delle variabili ambientali (con evoluzione temporale crescente) e gli effetti negativi delle variabili di effort (caratterizzate da evoluzione temporale decrescente) portano a concludere in favore dell'aumento del numero di specie nell'intera GSA, più accentuato nella zona calabrese che in quella salentina.

**Figura 11.** ACF e PACF della serie storica del numero di specie di pesci cartilaginei nella GSA 19.



**Figura 12.** ACF e PACF dei residui dei modelli Quasi-Poisson per la stima degli effetti lineari del numero (a) e della potenza (b) delle imbarcazioni, della temperatura al fondo (c) e della salinità (d) sul numero di specie di pesci cartilaginei nella GSA 19.

