

Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Scienze ambientali: tutela e gestione delle risorse naturali

Ciclo XXV

Settore Concorsuale di afferenza: 05/C1

Settore Scientifico disciplinare: BIO/07

TITOLO TESI

**Sviluppo di un approccio congiunto fuzzy –
Bayesiano per l’analisi e la modellizzazione
degli ecosistemi: applicazione ad ecosistemi
marini costieri**

Presentata da: Dott. Fabio Bozzeda

Coordinatore Dottorato

Prof. Enrico Dinelli

Tutore

Dott. Andrea Pasteris

Relatore

Prof. Victor Ugo Ceccherelli

Co-Relatore

Dott.ssa Marina A. Colangelo

Esame finale anno 2013

Sommario

| | |
|---|----|
| 1. INTRODUZIONE..... | 6 |
| 1.1 Gli ecosistemi come oggetto della ricerca..... | 6 |
| 1.2 Modellistica ecologica | 8 |
| 1.2.1 Modelli come strumento per la gestione | 10 |
| 1.2.2 Modelli come strumento per la ricerca..... | 11 |
| 1.2.3 Elementi di un modello | 14 |
| 1.2.4 Differenti tipologie di modelli ecosistemici..... | 15 |
| 1.3 Ecosistemi : complessità | 17 |
| 1.4 Ecosistemi : Incertezza..... | 21 |
| 2. FUZZY BAYESIAN ECOLOGICAL MODEL (FBEM) : un approccio generale per la modellistica ecosistemica..... | 25 |
| 2.1 Fuzzy Logic..... | 25 |
| 2.3 Compilatore naive Bayes | 31 |
| 2.4 Approccio fuzzy naive Bayes..... | 34 |
| 2.4.1 Considerazioni sul metodo..... | 36 |
| 3. APPLICAZIONE DELL'APPROCCIO F.B.E.M. | 37 |
| 3.1 Il progetto "Theseus" | 37 |
| 3.1.1 Decision Support System | 40 |
| 3.2.1 Habitat recettore: intertidale di spiagge sabbiose..... | 43 |
| 3.3 Le spiagge | 43 |
| 3.3.1 Descrittori degli stati ecosistemici | 44 |
| 3.4. Caratterizzazione delle aree di studio..... | 45 |
| 3.4.1 Cesenatico | 46 |
| 3.4.2 Lido di Spina – Bellocchio..... | 46 |
| 3.4.3 Inquadramento generale dei siti | 47 |

| | | |
|-----------------------------|--|-----|
| 3.5 | Scenari di flooding..... | 53 |
| 3.6 | Cesenatico : applicazione FBEM..... | 53 |
| 3.7 | Validazione dei modelli generativi (Figg. 10 – 15) | 58 |
| 3.8 | Risultati: Cesenatico..... | 62 |
| 3.8.1 | Macrofauna (Figg. 16 a-c)..... | 62 |
| 3.8.2 | Meiofauna (Figg. 17 a-c)..... | 65 |
| 3.9 | BELLOCCHIO: applicazione FBEM | 66 |
| Impostazione del FBEM | | 67 |
| 3.10 | Validazione dei modelli generativi | 69 |
| 3.11 | Risultati: Lido di Spina | 75 |
| 3.11.1 | Macrofauna (Figg. 31 a-c)..... | 75 |
| 3.11.2 | Meiofauna (Figg. 32 a-c)..... | 77 |
| 3.12 | Risultati: Bellocchio..... | 79 |
| 3.12.1 | Macrofauna (Figg. 33 a-c)..... | 79 |
| 3.12.2 | Meiofauna (Figg. 34 a-c)..... | 81 |
| 3.13 | Funzioni di danno..... | 83 |
| 3.14 | Implementazione dei modelli generativi FBEM nel DSS | 85 |
| 4. | SCENARIO ECOLOGICO SUL LUNGO PERIODO..... | 88 |
| 4.1 | Risultati | 89 |
| 4.2 | Discussione | 90 |
| 5. | ULTERIORI APPLICAZIONI DELL'APPROCCIO FUZZY NAIVE BAYES..... | 95 |
| 5.1 | Non linear power law | 95 |
| 5.2 | .Classificazione di banchi di pesce mediante descrittori strutturali e parametri ambientali in relazione a dati acustici | 98 |
| 6. | CONSIDERAZIONI CONCLUSIVE..... | 100 |
| | BIBLIOGRAFIA | 103 |

1. INTRODUZIONE

L'applicazione di approcci modellistici alle problematiche ecosistemiche nasce in ritardo rispetto ad altre discipline scientifiche, risalgono, infatti, alla seconda metà del secolo scorso i primi modelli numerici costruiti per rappresentare dinamiche ecosistemiche. Il settore ha avuto un grande sviluppo nel corso degli ultimi 3 decenni in relazione ad evidenze sperimentali e concettuali relative alla complessità degli ecosistemi ed all'impossibilità di trattare tale complessità con le metodologie classiche dell'ecologia. L'applicazione di approcci modellistici ha trovato ulteriore impulso nella necessità di dotarsi di metodologie in grado di descrivere le risposte ecosistemiche a forzanti naturali ed antropiche.

In relazione al paradigma dominante della complessità si assiste nelle scienze ecologiche ad un incremento dell'utilizzo di modelli matematici come strumento per descrivere e comprendere le dinamiche complesse degli ecosistemi e per contestualizzare i risultati derivanti da approcci classici meno complessi.

1.1 Gli ecosistemi come oggetto della ricerca

Occorre fare chiarezza circa la definizione di ecosistema. Generalmente gli ecosistemi vengono riconosciuti come livelli specifici di organizzazione senza specifiche riguardanti il tempo o la scala spaziale.

Nel contesto della modellizzazione ecologica la definizione di ecosistema nasce dalla seguente affermazione fornita da Morowitz (1968) :

“Le condizioni con cui un ecosistema sostiene la vita nello stato attuale sono da considerare proprietà dell'ecosistema piuttosto che di un singolo organismo o specie”
in cui vengono sottolineati due concetti chiave come l'elevata dipendenza degli ecosistemi dalle condizioni ambientali del contesto specifico in cui l'ecosistema si trova e lo stato di “semplice componente” del sistema che viene assegnato alle componenti biotiche che fino a quel momento avevano mantenuto un ruolo centrale negli studi effettuati con approcci classici dell'ecologia.

Alcune definizioni di ecosistema come ad esempio quella di Hutchinson (1970, 1978) riportano la componente biotica al centro della problematica definendo un ecosistema

come la porzione di spazio adatta al sostentamento di un determinato organismo o specie per cui per una forma di vita microbica l'ecosistema sarà rappresentato da pochi metri quadrati mentre per un grande predatore la scala passerà ai km quadrati ed oltre.

Una definizione maggiormente integrata è quella fornita da Lindeman (1942) per cui "Un ecosistema è costituito da processi chimici, fisici e biologici operanti in una unità spazio temporale".

A partire dalla precedente definizione (Lindeman, 1942) si è sviluppato l'approccio funzionale di processo che ha dominato gli ultimi 50 anni di ecologia degli ecosistemi. Il medesimo approccio è stato personalmente interpretato da diversi studiosi circa la componente da considerare come principale nello studio delle dinamiche ecosistemiche. In particolare l'attenzione è stata secondo i casi focalizzata sui cicli bio-geo-chimici (Odum, 1953, 1959, 1969, 1971), su bilanci funzionali di input output di biomassa (Bormann and Likens, 1967), sull'energia di cattura e nutrienti (O'Neill, 1976) oppure enfatizzando aspetti termodinamici per spiegare trasferimenti di energia fra le componenti ecosistemiche (Prigogine, 1947; Mauersberger, 1983; Jorgensen, 1981, 1982, 1986).

Considerando le oggettive difficoltà di una descrizione dettagliata degli ecosistemi, nell'interessa delle componenti di sistema, un ecosistema non può essere descritto secondo Morowitz (1968) senza aver prima definito i limiti e gli obiettivi dello studio specifico o, in altre parole, non prima di aver definito il punto di vista dell'osservatore. Pertanto la definizione di ecosistema comunemente accettata nell'ambito dell'ecologia degli ecosistemi e della modellistica ecosistemica è la seguente:

" Un ecosistema è un sistema o unità biotica e funzionale in grado di sostenere la vita includendo in sé componenti (variabili) biotiche ed abiotiche. La scala spazio-temporale non è definita a priori ma dipende dagli obiettivi dello specifico studio eco sistemico".

Conseguentemente alla precedente definizione "aperta" di ecosistema esistono diversi approcci allo studio degli ecosistemi raggruppabili in 4 filoni principali (Likens, 1985):

1. Studi empirici – Vengono raccolti frammenti d'informazione seguiti da tentativi di aggregazione ed integrazione per ottenere una rappresentazione completa del sistema in esame.
2. Studi comparativi – Componenti strutturali e funzionali sono comparate all'interno di un determinato range di tipologie ecosistemiche.

3. Studi manipolativi – Interi ecosistemi o loro parti vengono manipolate allo scopo di identificare o chiarirne particolari meccanismi.
4. Studi di modellistica o simulazione numerica

Dalla definizione stessa di ecosistema si evince come la trasversalità dei sistemi ecologici alle problematiche ambientali dia non pochi problemi agli studiosi che si avvicinano alla loro comprensione. Appare chiaro come la mancanza di una scala di definizione e di osservazione inserisca negli studi ecosistemici una componente di relativismo e, di conseguenza, di indeterminismo che è dovuta a quella parte dell'informazione che pur rimanendo fuori dalle condizioni al contorno determinate dalla scala utilizzata e dalla scelta della problematica dello studio ne condiziona comunque i risultati emergendo, in apparenza, come una componente casuale, non lineare se non caotica. Nonostante la definizione stringente gli ecosistemi sono infatti sistemi dinamici aperti e, soprattutto nella componente biotica, estremamente sensibili ad eventuali perturbazioni provenienti dall'esterno del sistema considerato.

1.2 Modellistica ecologica

I modelli rappresentano, in linea di principio, strumenti adatti ad analizzare le problematiche scientifiche tenendo conto di tutte le caratteristiche del sistema reale in cui la problematica considerata va contestualizzata.

I modelli ecologici, in particolare, possono essere comparati ad una mappa (che è a sua volta un modello) cioè ad una descrizione nello spazio, nel tempo od in entrambi di una successione di stati discreti di un sistema dinamico continuo.

Esistono differenti tipologie di mappe in relazione alle caratteristiche dell'informazione che si intende rappresentare, le mappe geografiche ad esempio differiscono secondo la scala di rappresentazione della realtà riportata sulla mappa stessa. Una mappa non contiene mai tutte le caratteristiche di una zona geografica e la sua attendibilità è legata alla variabilità degli oggetti rappresentati, se una mappa ad esempio contenesse la posizione di tutte le automobili in una data area la sua validità decadrebbe nel momento stesso in cui una sola delle automobili dovesse cambiare posizione, la mappa così costruita avrebbe validità solo nel momento della sua implementazione e nelle

condizioni in cui sono state rilevate le osservazioni. La medesima difficoltà di applicazione è stata riscontrata nel tentativo di includere livelli di complessità sempre maggiori nella modellistica degli ecosistemi per cui, per le componenti più complesse, ci si è ritrovati a dover processare variabilità troppo elevate non spiegabili a partire da dati sperimentali o osservazioni di campo.

Per questo motivo, almeno inizialmente, sono stati concepiti modelli descrittivi di problematiche ben delimitate o comunque parziali rispetto alla reale complessità riscontrata nei sistemi naturali. In risposta alle nuove esigenze di conoscenza e gestione degli ecosistemi si è assistito ad un deciso incremento, come numero e tipologie, delle applicazioni modellistiche in ecologia .

La ragione del proliferare di nuovi approcci modellistici è imputabile essenzialmente a tre fattori:

1. Lo sviluppo delle tecnologie di calcolo informatico adatte al trattamento di sistemi matematici altamente complessi.
2. Un generale aumento dell'attenzione dedicata alle problematiche ambientali dovute all'aumento dell'inquinamento e degli impatti antropici. La consapevolezza dell'impossibilità di eliminare completamente inquinamento ed impatti dovuti all'attività umana ha visto emergere la necessità di dotarsi di mezzi adatti alla gestione ottimale degli ecosistemi.
3. L'aumento significativo delle conoscenze quantitative e qualitative circa le dinamiche ecosistemiche ha imposto all'attenzione dei ricercatori la rilevanza delle interazioni fra proprietà puramente ecologiche e fattori ambientali da inquadrare in un ottica complessa.

I modelli possono essere considerati come una sintesi di quanto è conosciuto su un ecosistema in riferimento ad un determinato problema i cui responsi possono anche essere in contrasto con l'analisi statistica dei dati che rivela invece le relazioni statistiche esistenti fra osservazioni.

Attraverso l'utilizzo di un approccio modellistico è possibile includere conoscenze complesse circa un ecosistema come:

1. Le modalità secondo cui alcune componenti ecosistemiche interagiscono con altre (ad esempio come lo zooplancton incide sul phytoplankton).
2. Inclusione di processi fisici sottoforma di funzioni matematiche considerate generali per il processo considerato.
3. L'importanza (numericamente il peso) di uno o più processi in relazione ad uno specifico problema.

Ovviamente i due approcci (modellistico e statistico) non sono reciprocamente esclusivi ma risultano complementari in quanto in base ad analisi statistiche effettuate su osservazioni è possibile calibrare al meglio le funzioni dei modelli i quali hanno la caratteristica aggiuntiva di contestualizzare l'osservazione statistica che altrimenti rimarrebbe circoscritta al caso di studio sia nel tempo che nello spazio.

Tornando all'esempio delle vetture la cui posizione deve essere rappresentata in una mappa, attraverso l'analisi statistica vengono studiate le posizioni reciproche delle vetture in un dato momento; l'analisi di diversi momenti (e quindi diverse posizioni) potrebbe permettere di estrarre una funzione descrittiva del succedersi di diverse posizioni reciproche adatta alla costruzione di un modello di posizionamento reciproco delle vetture in un dato quadro temporale e spaziale.

1.2.1 Modelli come strumento per la gestione

Urbanizzazione e sviluppo tecnologico hanno provocato un incremento degli impatti ambientali. Incrementi dei flussi energetici ed inquinanti hanno provocato negli ecosistemi variazioni a livello di singole componenti come crescita di alghe, batteri e specie dannose nonché alterazioni dell'intera struttura ecologica. Un ecosistema è un sistema estremamente complesso per cui predire gli effetti ambientali di alcuni impatti anche puntiformi diventa una sfida ambiziosa in cui si inserisce l'approccio modellistico.

Attraverso approcci modellistici si stima di riuscire a ricostruire le modalità di propagazione di perturbazioni esterne all'interno degli ecosistemi in modo da

indirizzarne la gestione e la scelta delle tecnologie più adatte alla minimizzazione di possibili impatti ecosistemici.

L'utilizzo gestionale di approcci modellistici acquista ulteriore valenza nel contesto degli studi sul "Global Climate Change" in relazione ai quali appare fondamentale predire quali potranno essere gli effetti sugli ecosistemi dei cambiamenti climatici (Clark, 2001).

1.2.2 Modelli come strumento per la ricerca

I modelli possono anche essere utilizzati per generalizzare quanto emerso da osservazioni in situ o da esperimenti di laboratorio e per ricontestualizzarne i risultati in presenza di eventuali influenze provenienti da processi considerati irrilevanti in un'indagine sperimentale specifica. Ad esempio le camere termostatiche sono utilizzate per misurare la crescita algale in funzione solamente della concentrazione di nutrienti, le carote di sedimento sono trattate in laboratorio per indagare l'interazione acqua-sedimento senza disturbi provenienti da altre componenti ecosistemiche e così via. I risultati provenienti da esperimenti possono essere utilizzati per stimare dei parametri o delle funzioni che una volta confluiti in un modello interagiscono in modo da rendere responsi che tengano conto, almeno in parte, della complessità del sistema reale rispetto alla semplicità del sistema sperimentale.

L'applicazione dei modelli in ecologia è praticamente obbligatoria nel momento in cui l'obiettivo diventa la comprensione del maggior numero possibile di interazioni all'interno del sistema. D'altro canto l'aumento delle interazioni da descrivere complica la descrizione modellistica in quanto provoca la comparsa di fenomeni non lineari per cui spesso il comportamento complessivo degli ecosistemi non è dato dalla somma di dinamiche individuali. Questo implica che le proprietà globali di un ecosistema non possono essere rilevate se non con approcci adeguatamente complessi. L'applicazione di modelli alla comprensione di fenomeni di sistema ha messo in evidenza alcuni vantaggi derivanti dall'utilizzo in ecologia di approcci modellistici:

1. I modelli sono buoni strumenti da utilizzare nel monitoraggio degli ecosistemi.

2. I modelli possono rivelare proprietà degli ecosistemi che non emergerebbero da studi sperimentali concentrati su problematiche particolari.
3. I modelli possono rivelare punti deboli nella conoscenza e possono in questo modo indirizzare l'evoluzione della ricerca.
4. I modelli possono essere utili nel testare ipotesi difficili da replicare sperimentalmente attraverso il confronto fra dati reali e risultati modellistici.

L'ultimo punto sopra affrontato (4) comporta problemi di opzioni alternative. Infatti un modello rappresenta di per sé un'ipotesi sul funzionamento dell'ecosistema in esame per cui si hanno 4 casi annidati in 2 (accettazione o rifiuto del modello):

1. Il modello è corretto nel contesto del problema e l'ipotesi testata è corretta
2. Il modello non è corretto ma l'ipotesi è corretta
3. Il modello è corretto ma l'ipotesi testata non è corretta
4. Il modello non è corretto e non lo è nemmeno l'ipotesi testata

Partendo dal fatto che un modello non validato nella specificità dei casi possibili non può essere utilizzato per testare ipotesi sperimentali su proprietà del sistema è possibile liquidare velocemente i punti 2 e 4. Nascono, al contrario, diverse difficoltà circa i punti 1 e 3 in cui ci si trova ad esaminare l'incertezza portata dal sistema in esame in congiunzione con l'incertezza insita nell'approccio modellistico. La sovrapposizione delle incertezze da esaminare ne confonde gli effetti per cui è possibile solo in maniera limitata capire se l'accettazione/rifiuto di un'ipotesi attraverso test modellistici sia dovuta alla correttezza/inesattezza dell'ipotesi o ad un errore interno alla struttura matematica del modello utilizzato. Ciò suggerisce, come già in altre discipline scientifiche, la fisica in primis, come questi strumenti vadano utilizzati in maniera integrata all'interno di una struttura di feedback in cui i risultati sperimentali indirizzano la costruzione dei modelli mentre questi ultimi possono a loro volta aiutare nella

comprensione di alcuni risultati e nella messa a punto di impianti sperimentali sempre più adatti all'indagine delle problematiche considerate.

Ora, dopo aver definito le possibili implicazioni/applicazioni degli approcci modellistici nelle scienze ecologiche occorre definire, in maniera generale, le considerazioni che hanno portato allo sviluppo di queste metodologie ricordando che la seguente lista ha validità puramente indicativa e va adattata caso per caso:

1. Il progredire delle scienze ha evidenziato come il mondo reale sia molto più complesso rispetto a quanto inizialmente previsto. In particolare in ecologia si scoprono continuamente nuove interazioni e nuove problematiche. La successione negli studi e nei casi studio ha mostrato come risultati basati su assunzioni e approssimazioni eccessivamente semplificative siano non realistici.
2. Lo sviluppo dell'ecologia degli ecosistemi ha evidenziato il bisogno di metodologie adatte alla considerazione della complessità ecosistemica e, considerando la trasversalità degli ecosistemi, che prendessero in considerazione risultati, interpretazioni e dinamiche provenienti da altre discipline scientifiche (fisica, chimica, economia, sociologia, etc.).
3. La necessità di valutare l'elevato grado di incertezza dovuto all'elevata complessità degli ecosistemi.
4. La necessità di dotarsi di approcci elastici in considerazione dell'elevata variabilità degli ecosistemi e delle loro componenti. L'utilizzo di approcci fra loro coerenti favorisce il confronto fra diversi ecosistemi e fra diverse contestualizzazioni spazio temporali degli stessi.
5. Le necessità gestionali circa le problematiche ecologiche hanno favorito lo sviluppo di approcci modellistici gestionali a complemento degli approcci tradizionali (esperimenti, studi d'impatto, calcolo di indici) non adatti, da soli, a fornire responsi dinamici ed elastici.
6. L'evidenza dell'impossibilità di formalizzare per intero le dinamiche che intercorrono all'interno di un sistema complesso ambientale tenendo in conto delle non linearità e delle dinamiche caotiche che da essi possono emergere.

1.2.3 Elementi di un modello

In generale la formulazione matematica di un modello ecologico è composta da 5 componenti:

1. **Funzioni forzanti o variabili esterne** : funzioni o variabili esterne al target del modello che influenzano o possono influenzare lo stato dell'ecosistema. Si suddividono in funzioni di controllo se riferite ad attività antropiche e funzioni naturali se indipendenti da esse.
2. **Variabili di stato**: sono le variabili che descrivono gli stati dell'ecosistema, la scelta delle variabili di stato dipende dalla problematica che si intende indagare e dal livello di complessità che si considera caratterizzante l'ecosistema. All'interno della struttura gerarchica del sistema è possibile scegliere diverse variabili di stato secondo il loro livello di annidamento nel sistema ed alla complessità delle loro dinamiche.
3. **Equazioni matematiche**: sono utilizzate per descrivere i processi fisici, chimici e biologici esistenti all'interno del sistema. Le equazioni costituiscono la rappresentazione delle relazioni che intercorrono fra funzioni forzanti e variabili di stato.
4. **Parametri**: coefficienti utilizzati nella rappresentazione matematica dei processi. La stima dei coefficienti è un passaggio obbligato nella costruzione di un modello ed è realizzata in due modi o attraverso osservazioni in situ o attraverso i risultati di esperimenti di laboratorio.
5. **Costanti universali**: costanti fisse come la costante dei gas o il peso atomico che possono essere utilizzate in alcune applicazioni modellistiche.

1.2.4 Differenti tipologie di modelli ecosistemici

Esistono diverse tipologie di approccio modellistico agli ecosistemi ognuna adatta allo studio di problematiche specifiche e a differenti tipologie di informazione proveniente dagli ecosistemi stessi.

L'alto livello di complessità presente negli ecosistemi permette l'applicazione di varie tipologie d'approccio funzione della complessità stessa indagabile da differenti punti di vista ed a differenti livelli di profondità all'interno della struttura complessa dei sistemi naturali.

Gli approcci modellistici possono essere classificati come segue:

- Modelli biogeochimici dinamici
- Modelli biogeochimici stazionari
- Modelli di dinamica di popolazione
- Modelli strutturali dinamici
- Modelli fuzzy
- Reti neurali
- Modelli spaziali
- Modelli "Individual based"
- Modelli eco tossicologici
- Modelli stocastici
- Modelli ibridi

L'applicabilità dei diversi approcci proposti va valutata attraverso una comparazione fra gli svantaggi ed i vantaggi che ogni approccio porta con sé in relazione alla tipologia di dato disponibile, alla problematica indagata ed al livello di complessità che si intende raggiungere nella rappresentazione dell'ecosistema (Tab. 1).

| Approccio modellistico | Vantaggi | Svantaggi |
|------------------------------------|---|---|
| Modelli biogeochimici dinamici | Basati su rapporti causali. Agevoli nella comprensione e nello sviluppo. | Non utilizzabili con dati eterogenei. Difficoltà nella calibrazione in presenza di relazioni complesse. Non adattativi. |
| Modelli biogeochimici discreti | Eccellenti per grandi impatti o situazioni nella media. | Non utilizzabili come strumento previsionale. |
| Modelli di dinamica di popolazione | Basati su rapporti causali. Agevoli nella comprensione e nello sviluppo. | Modelli auto ecologici. Limiti di applicabilità. Bassa complessità. |
| Modelli strutturali dinamici | Non necessitano di calibrazione di parametri. Relativamente agevoli da implementare. | Tempi di computazione elevati. Difficoltà di validazione. Non generativi. |
| Modelli fuzzy | Applicabili a dati incerti. Applicabili ad informazioni incomplete. | Difficoltà di applicazione a dinamiche complesse. |
| Reti neurali | Utilizzabili dove altri approcci falliscono. Utilizzabili con dataset eterogenei. | Non riportano rapporti di causalità. Limitazioni nell'accuratezza delle previsioni. |
| Modelli spaziali | Rappresentativi nello spazio. | Statici. |

| | | |
|----------------------------------|---|---|
| | Possono contenere diverse informazioni (G.I.S.) | Non applicabili a dinamiche complesse. |
| Modelli “Individual agent based” | Adatti a descrivere relazioni anche complesse di autocorrelazione. Descrizioni spaziali. | Tolleranti di poche variabili. |
| Modelli ecotossicologici | Specificità. Semplicità di utilizzo. | Specificità. Alta incertezza. |
| Modelli stocastici | Considerano la casualità. Robusti rispetto all’incertezza. | Aprioristicità. Richiedono grandi dataset. |
| Modelli ibridi | A seconda del tipo di ibridazione | A seconda del tipo di ibridazione |

Tabella 1: Tipologie di approcci modellistici con riportati i vantaggi e gli svantaggi rispetto alla problematica ecosistemica.

1.3 Ecosistemi: complessità

Gli ecosistemi portano all'uomo una serie di servizi e funzioni fondamentali nel determinare a livello globale le tessiture sociali economiche e politiche. Appare dunque fondamentale dotarsi di mezzi adatti alla comprensione delle dinamiche ecosistemiche e di metodologie che siano in grado di comprenderne l'evoluzione in risposta a forzanti naturali ed antropiche.

La comprensione delle dinamiche che portano a cambiamenti di stato permanenti o momentanei negli ecosistemi rappresenta una sfida, resa ancora più attuale dalla necessità di comprendere le conseguenze a livello ambientale dei fenomeni correlati ai cambiamenti climatici globali (Schlacher et al., 2008; Planque et al., 2011; Migliavacca et al., 2012).

Ogni ecosistema è, per numero di componenti e per grado di connettività, un sistema complesso le cui dinamiche dipendono dalle singole componenti e dalla struttura sistemica in cui sono inserite.

I sistemi complessi, e gli ecosistemi non fanno eccezione, godono di alcune proprietà che ne rendono difficile se non impossibile la descrizione deterministica classica.

Una di queste è la presenza di comportamenti emergenti che agiscono sulle variabili di sistema e ne condizionano i valori in modo apparentemente indipendente dall'interazione lineare delle singole componenti (Grimm et al., 2005; Svirezhev, 2008).

I comportamenti emergenti ricadono sulle singole componenti senza poter essere spiegati da altre componenti del sistema, la loro tempistica ed intensità appare casuale e li ricollega al concetto più ampio di auto-organizzazione dei sistemi (Sibani et al., 1993; Levin, 1998, 2005).

E' molto importante definire la presenza di comportamenti emergenti in relazione alla definizione di ecosistema per la quale, come spiegato precedentemente, non esiste una definizione specifica di scala né spaziale né temporale. Nel momento in cui viene deciso il livello di complessità che si intende raggiungere nella modellizzazione di un ecosistema ci si pongono automaticamente dei limiti di comprensione delle dinamiche ecosistemiche in quanto non è possibile superare il livello di complessità proprio della variabile target del modello. La scelta stessa delle variabili di controllo in numero limitato a quanto si pensa causalmente correlato alla variabile target impedisce la comprensione dei comportamenti emergenti i cui effetti sulla variabile target vengono spesso considerati delle anomalie.

In funzione del punto di osservazione, un comportamento emergente apparirà come una dinamica non lineare non descrivibile attraverso le leggi statistiche dedotte dai pattern di osservazioni raccolte.

Tali difficoltà di comprensione si riflettono nelle difficoltà nel definire la topologia degli ecosistemi vale a dire la definizione e lo studio di alcune proprietà invarianti fra gli spazi (scala spaziale) e le funzioni (scala temporale) che ne caratterizzano struttura e connessioni fra le componenti. La definizione di una topologia corretta è di fondamentale importanza nella formalizzazione di una rappresentazione modellistica in quanto permette aprioristicamente di definire il numero di dimensioni (numero di variabili) e di interazioni da considerare. Topologicamente un ecosistema potrebbe essere definito come una varietà n-dimensionale in cui n indica il numero delle

componenti. Tale varietà, può essere rappresentata, tramite un'unione di complessi di celle di dimensioni inferiori ad n chiamata n -scheletro. L' n -scheletro rappresenta funzionalmente uno spazio in cui sono comprese tutte le n -celle della sua partizione con proprietà differenti rispetto alle n -celle considerate singolarmente. Una proprietà interessante dell' n -scheletro è quella per cui le proprietà delle singole n -celle vengono modificate nel momento in cui aumentando le dimensioni dello spazio si viene a formare l' n -scheletro che rappresenta una realtà funzionale indipendente ed autonoma. Ora, appare intuitivamente chiaro come ipotesi eccessivamente semplificative allo studio di sistemi complessi possano risultare fuorvianti nella comprensione degli stessi a causa della riduzione di dimensioni e conseguentemente di funzionalità osservabili. Proviamo ad immaginare un ecosistema come un'ipersfera n -dimensionale (un n -scheletro con una struttura spaziale ben definita), l'ipersfera ecosistemica avrà tante dimensioni quante sono le componenti del sistema: sarà quindi suddividibile in ipersfere a dimensionalità minori, sfere, circonferenze, punti. Il numero di dimensioni di ogni sottoinsieme indica la complessità intrinseca del sottosistema scelto che non risente delle deformazioni indotte dalle dimensioni non considerate dando così luogo a dinamiche non completamente corrispondenti alle dinamiche che invece caratterizzano la struttura dell'ipersfera n -dimensionale complessiva. L'osservazione dei sottoinsiemi isolabili dall'ipersfera renderà quindi necessariamente un'informazione parziale circa le dinamiche del sistema nel suo complesso difficilmente generalizzabile alle dinamiche proprie dell'ipersfera principale. Nell'ipersfera, così come negli ecosistemi ciò che varia è quindi la percezione esterna delle dinamiche di sistema che risulta condizionata dal numero di dimensioni in cui si trova l'osservatore. Il responso dell'osservatore sarà quindi accompagnato da una quantità di rumore statistico dovuto alle dimensionalità non considerate e, soprattutto, non formalizzabile all'interno del sottospazio preso in considerazione.

Un'altra importante caratteristica dei sistemi complessi è la struttura di dipendenze che interseca su ogni livello le componenti di sistema dando luogo ad un insieme gerarchizzato di variabili (Sibani et al., 1993; Clauset et al., 2008). La gerarchizzazione sistemica ha come primo effetto quello di rendere difficoltosa la comprensione di intensità e direzione di un disturbo all'interno del sistema complesso.

Per descrivere gli stati di un ecosistema è necessario individuare all'interno della struttura gerarchizzata quelle variabili maggiormente dipendenti i cui valori sono quindi

determinati dalle variazioni reciproche delle altre secondo la struttura di dipendenze del sistema in esame. Fra tutte le variabili di sistema quelle che, per gli ecosistemi, maggiormente rispondono ai requisiti di dipendenza sopra descritti sono le variabili biologiche considerate le migliori indicatrici degli stati eco sistemici (Water Framework Directive, 2000/60/EC).

All'interno del sistema possiamo classificare le diverse variabili secondo :

- grado di dipendenza
- livello di complessità
- variabilità nello spazio
- variabilità nel tempo

L'analisi di queste 4 caratteristiche descrive una struttura all'interno della quale le diverse componenti scambiano informazione determinando così le reciproche variazioni spazio/temporali (Woodard et al.2007).

Alcune variabili del sistema appariranno pressochè costanti,altre mostreranno gradi di variabilità paragonabili e/o in rapporti di proporzionalità esprimibili tramite coefficienti e parametri numerici altre ancora subiranno variazioni più intense e repentine non descrivibili numericamente e collegati a tratti con variazioni nei valori di altre variabili. In altre parole i sistemi ambientali presentano grande disomogeneità fra le variabilità associate alle componenti del sistema.

Un'ulteriore problematica emersa dalle evidenze sperimentali e dagli scenari relativi al cambiamento climatico globale è la valutazione delle risposte ecosistemiche ad eventi climatici estremi.

Secondo l'IPCC 2001 “..un evento estremo è un evento ‘raro’ in riferimento alla sua distribuzione statistica. La definizione di ‘raro’ può variare ma normalmente è intesa come inferiore al decimo o superiore al novantesimo percentile”.

In quest'ottica il “Global Climate Change” può essere visto come una cascata di fenomeni estremi interni al sistema “ambiente” che ne deformano la dimensionalità e le caratteristiche strutturali.

Un evento estremo climatico può innescare all'interno degli ecosistemi dinamiche solo parzialmente comprensibili attraverso osservazioni ed esperimenti.

La valutazione di un effetto di un evento estremo su un sistema complesso necessita di mezzi in grado di valutare l'incertezza legata all'analisi in relazione al carattere estremo dell'evento ed alla complessità del sistema in esame.

Considerando struttura gerarchica e complessità degli ecosistemi unitamente al grado di estremismo degli eventi che possono interessare le diverse forzanti ambientali, l'analisi ecosistemica, sia descrittiva che previsionale, deve essere in grado di processare (Hughes, 2001; Wu e David, 2002; Camara, 2011; Boyd, 2012):

- l'incertezza legata al sistema in esame.
- la propagazione dell'incertezza secondo le variazioni di alcune componenti nel tempo e nello spazio.
- L'incertezza risultante dovuta alla complessità intrinseca del sistema ed al carattere estremo delle forzanti esterne.

1.4 Ecosistemi: Incertezza

Qualunque sistema complesso evolve secondo il tipo d'informazione che riceve dall'ambiente circostante in un tempo finito e con un certo grado di determinismo. Il rumore che “avanza” ,cioè l'incertezza, può essere trattato in maniera statistica. L'approccio deterministico classico della fisica descrive esclusivamente aumenti di entropia mentre gli ecosistemi alternano aumento e diminuzione dell'entropia e in quanto sistemi complessi tendono all'ordine (diminuzione di entropia) secondo modalità dipendenti dalle condizioni iniziali rispetto alla perturbazione ed alla sua intensità.

Ciò comporta in caso di applicazioni previsionali un'inevitabile perdita d'informazione nel tempo corrispondente ad un aumento dell'incertezza.

Esistono diverse forme d'incertezza, si può parlare d'incertezza in riferimento alla variabilità, all'errore di misura o all'incertezza di un modello (Cressie et al., 2009), il paradigma attualmente dominante nella descrizione dell'incertezza è la stocasticità che tratta l'incertezza in modo puramente statistico(Viertl, Hareter 2004).

In questo testo ci si riferirà all'incertezza come ad una forzante dell'informazione portata da una variabile, associata al grado di complessità (e quindi alla variabilità) della variabile stessa.

Per quanto riguarda le variabili ecologiche l'incertezza totale è legata a 2 fattori:

- il relativamente basso numero di osservazioni
- l'alta variabilità da piccola a grande scala delle variabili in esame

Per il basso numero di osservazioni nello spazio e nel tempo spesso i data set ecologici hanno la caratteristica della non asintoticità. La non asintoticità delle serie di dati non permette un'applicazione redditizia delle tecniche statistiche che classicamente si utilizzano per costruire distribuzioni di frequenza a partire da osservazioni

L'incertezza totale può essere spezzata nelle sue componenti principali che sono incertezza di stato ed incertezza funzionale. Per gli ecosistemi le due componenti principali dell'incertezza derivano dalle caratteristiche del sistema sotto esame rispettivamente: l'incertezza di stato (Cressie et al., 2009) dalla distribuzione degli organismi indicatori e dall'impossibilità di osservarli in continuum nello spazio e nel tempo mentre l'incertezza funzionale (Viertl e Hareter, 2004) è legata alla grande quantità di forzanti ambientali che, a livello sistemico, incidono sulle variabili indicatrici di stato degli ecosistemi.

Di norma i sistemi sono caratterizzati da una dominanza di una componente rispetto all'altra si hanno quindi sistemi caratterizzati da alta incertezza di stato ed altri caratterizzati da alta incertezza circa le relazioni che intercorrono fra le diverse componenti sistemiche. In diversi campi sono state sviluppate metodologie d'approccio adatte al trattamento di una o dell'altra componente. Attraverso l'applicazione di sistemi a gestione fuzzy sono stati ottenuti ottimi risultati in diversi campi in cui era dominante l'incertezza di stato mentre attraverso diversi tipi di applicazione del teorema di Bayes sono stati studiati sistemi con alto grado di incertezza funzionale (Viertl e Hareter, 2004). Sistemi più complessi quali sono gli ecosistemi che presentano spesso alti valori per entrambe le componenti creando grandi difficoltà nel valutare e processare le incertezze del sistema. L'obiettivo è di approntare un approccio generale su cui costruire modelli specifici in grado di processare entrambe le componenti dell'incertezza totale. In questo testo viene proposto un approccio congiunto di logica fuzzy e teorema

di Bayes che si presenta adatto a risolvere i problemi derivati dall'alto grado d'incertezza in entrambe le componenti principali.

2. FUZZY BAYESIAN ECOLOGICAL MODEL (FBEM) : un approccio generale per la modellistica ecosistemica

Esistono sistemi osservati con diversi gradi di incertezza alcuni con alta incertezza di stato (sistemi legati a sensori, sistemi di gestione di impianti civili, sociali, industriali, militari) e sistemi ad alta incertezza funzionale (meteorologia, oceanografia, diagnostica medica) per entrambi sono state approntate tecniche e logiche di approccio in grado di trattare l'incertezza correlata col sistema in esame. In particolare sistemi ad alta incertezza di stato sono trattati attraverso applicazioni della logica fuzzy mentre sistemi ad alta incertezza funzionale sono trattati attraverso applicazioni dirette o indirette del teorema di Bayes.

La metodologia proposta in questo lavoro di tesi nasce dalla fusione di logica fuzzy e teorema di Bayes per dar luogo ad un approccio modellistico consistente in un compilatore naïve Bayes operante in ambiente fuzzy (Tang e Xu, 2005). In sintesi, un compilatore naïve bayes le cui probabilità in input sono definite a partire dalla fuzzyficazione delle variabili di sistema ed il cui output rende un valore di probabilità associato ad un fuzzy set appartenente alla partizione fuzzy del dominio della variabile target di interesse, scelta in quanto presumibile miglior “descrittore” dello stato dell'ecosistema.

L'approccio metodologico presentato si propone di analizzare le possibilità evolutive degli ecosistemi e di individuare eventuali soglie di non linearità e tipping point in risposta al cambiamento delle condizioni ambientali al contorno dello stato dell'indicatore.

2.1 Fuzzy Logic

La fuzzy logic si basa sul concetto di fuzzy set (Zadeh, 1965) ossia di un insieme al quale un elemento appartiene (e non appartiene) con un grado di verità compreso nell'intervallo $[0,1]$. Il reale problema argomento della teoria fuzzy è la percezione o meglio la misura di grandezze reali. Assumendo di non essere in grado di ottenere misure di grandezze o fenomeni reali con un errore nullo o con l'assoluta sicurezza

della costanza dell'errore la logica fuzzy inquadra il grado d'incertezza come una qualità della grandezza misurata. Fatte queste assunzioni appare logico come un dato possa rispondere ad una determinata caratteristica secondo una scala che va dallo 0 al 100% standardizzabile in un valore compreso fra 0 e 1.

Il grado di veridicità di un certo valore rispetto ad una determinata caratteristica è chiamato coefficiente d'appartenenza ed esprime l'appartenenza di un valore ad un fuzzy set che ne definisce una caratteristica. E' dimostrato che i coefficienti di appartenenza esprimono valori di possibilità; rimane invece aperta la discussione sul considerarli o meno valori di probabilità. Formalmente i coefficienti d'appartenenza verificano i tre assiomi di Kolmogorov (Zadeh, 1968) ed il problema del complementare per cui se ne può dimostrare la loro valenza probabilistica.

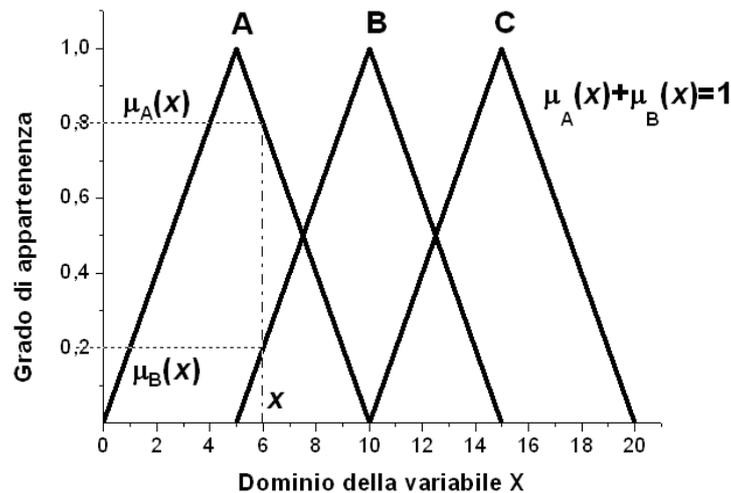


Figura 1: **fuzzy set**: partizione del dominio di una variabile X in tre fuzzy sets compenetrantesi nel punto medio e rispondenti alla funzione d'appartenenza a "triangolo isoscele". $\mu_{A/B}(x)$ =coefficiente d'appartenenza di x ad un fuzzy set; la relazione $\mu_A(x)+\mu_B(x)=1$ esprime la relazione esistente fra i coefficienti d'appartenenza di un elemento x conosciuto con incertezza nulla.

Attorno alla teoria dei fuzzy set sono state sviluppate in diversi campi applicazioni in grado di trattare l'incertezza di stato di sistemi di variabili. Ai fini dell'approccio, argomento di questa tesi di dottorato, è importante sapere che attraverso il formalismo fuzzy è possibile descrivere qualsiasi distribuzione probabilistica e soprattutto che a partire da una serie di coefficienti fuzzy è possibile stimare il valore di probabilità

marginale di un dato evento fuzzy.

Ritornando al significato formale dei coefficienti di appartenenza essi indicano un valore di possibilità. Per definire il significato di “valore di possibilità” si può ricorrere ad un esempio utilizzato da Zadeh, il padre della logica fuzzy. Si consideri l’enunciato : “è possibile che Giacomo questa mattina si mangi 3 uova, ma non l’ha mai fatto, è quindi improbabile che succeda”.

L’esempio è emblematico e fa intendere come il termine possibilità convogli in sé un significato di fattibilità/veridicità ed un significato di consistenza con l’informazione disponibile. Un evento è oggettivamente possibile se esiste una fattibilità fisica per il suo realizzarsi. Dal punto di vista “frequentista” una possibilità può essere considerata come limite superiore di una distribuzione e quindi un’estensione del concetto più restrittivo (ma più robusto in virtù del numero più elevato di implicazioni) di probabilità. A questo punto appare opportuno definire i rapporti possibilità – probabilità in modo da comprendere come, basandosi su un campionamento, sia possibile stimare valori di probabilità a partire da valori di possibilità.

Il rapporto fra possibilità e probabilità è definito attraverso il principio di consistenza probabilità/possibilità:

Definizione (DEF): sia P una probabilità su X e sia Π una possibilità su X. Si dirà che P è consistente con Π se

$$\forall A \in D(X), P(A) \leq \Pi(A)$$

Dall’enunciato si evince il rapporto dimensionale fra i valori di possibilità sempre maggiori o uguali dei valori di probabilità corrispondenti e la “subordinazione” formale dei valori di probabilità rispetto ai valori di possibilità il cui concetto di misura gioca un ruolo fondamentale nel trattamento dei dati “soft”, così chiamati in ragione della “morbidezza” dei loro contorni fuzzy.

Si consideri una variabile x con un valore nell’universo di discorso U e sia π_x la distribuzione di possibilità indotta dalla proposizione $P \triangleq x \in G$, in cui G è un sottoinsieme fuzzy di U caratterizzato dalla funzione di appartenenza μ_G . In conseguenza al postulato della possibilità si può dire

$$\pi_x = G$$

che implica

$$\pi_x(u) = \mu_G(u); u \in U$$

In cui π_x è la funzione di distribuzione di possibilità di x.

Ora, se F è un sottoinsieme fuzzy di U la misura di possibilità di F è definita dall'espressione

$$\pi(F) = \sup_U (F \cap G)$$

Oppure in maniera equivalente

$$\pi(F) = \sup_U (\mu_F(u) \wedge \mu_G(u))$$

Il numero $\pi(F)$, compreso in un range di valori tra 0 e 1 , può essere interpretato come la possibilità che x risponda alle caratteristiche sia di F che di G secondo la relazione

$$\pi(F) = Poss\{x = F/x = G\} = \sup (F \cap G)$$

In particolare se F è un insieme non-fuzzy A allora vale la relazione

$$\mu_A(u) = 1 \text{ se } u \in A$$

$$= 0 \text{ se } u \notin A$$

e quindi

$$\pi(A) = Poss\{x = A/x = G\} = \sup_A (G)$$

$$= \sup_A [\mu_G(u)]; u \in U$$

Sia $x \triangleq (x_1, \dots, x_n)$ una variabile n-aria con valori nell'universo di discorso

$$U = U_1 \times \dots \times U_n$$

con $x_i; i=1, \dots, n$ valori in U_i . Sia F una n-aria relazione fuzzy in U caratterizzata dalla funzione d'appartenenza μ_F ; la proposizione $P \triangleq x = F$ induce una n-aria distribuzione congiunta di possibilità :

$$\pi_x = \pi_{(x_1, \dots, x_n)} = F$$

La funzione di distribuzione di possibilità di x è espressa da

$$\pi_{(x_1, \dots, x_n)}(u_1, \dots, u_n) = \mu_F(u_1, \dots, u_n), \quad u \triangleq (u_1, \dots, u_n) \in U$$

$$= Poss\{x_1 = u_1, \dots, x_n = u_n\}$$

Come per la probabilità si possono definire i valori di possibilità marginale e condizionata.

Sia $s \triangleq (i_1, \dots, i_k)$ una sub sequenza della sequenza di indici $(1, \dots, n)$ e sia s' la subsequenza complementare $s' \triangleq (j_1, \dots, j_m)$.

Sia $A_{(s)} = (A_{j_1}, \dots, A_{j_m})$ una combinazione di sequenze, in particolare la variabile $x_{(s)} = (x_{i_1}, \dots, x_{i_k})$ si riferisce alla k-esima sub variabile di $x \triangleq (x_1, \dots, x_n)$ con $x_{(s')} = (x_{j_1}, \dots, x_{j_m})$ sub varabile complementare di $x_{(s)}$.

La proiezione di $\pi(x_1, \dots, x_n)$ su $U_{(s)} \triangleq U_{i_1} \times \dots \times U_{i_k}$ è la funzione di k possibilità :

$$\Pi_{x_{(s)}} \triangleq \text{proj}_{U_{(s)}} \Pi_{(x_1, \dots, x_n)}$$

definita da

$$\pi_{x_{(s)}}(u_{(s)}) \triangleq \sup_{u_{(s)'}} \pi_x(u_1, \dots, u_n)$$

in cui $\pi_{x_{(s)}}$ è la funzione di distribuzione di probabilità di $\Pi_{x_{(s)}}$.

Per analogia col concetto di probabilità marginale, $\Pi_{x_{(s)}}$ è detta possibilità marginale.

A partire da questa, coerentemente con quanto conosciuto per la probabilità, è possibile ricavare i valori di possibilità congiunta e condizionata.

La corrispondenza fra valori di possibilità e di probabilità è definibile a partire da alcune proprietà della misura di possibilità che replicano coerentemente le loro analoghe della misura di probabilità.

Si consideri la zona di studio, all'interno della quale vengono prelevati i campioni, come un insieme che chiameremo Ω . Si considerino i campioni prelevati come una famiglia F di sottoinsiemi di Ω . Si nota come la famiglia di sottoinsiemi F abbia delle proprietà di stabilità rispetto ad alcune operazioni in Ω , in particolare rispetto all'unione numerabile e al passaggio al complementare.

Siano:

Ω un insieme non vuoto

F una famiglia di sottoinsiemi di Ω

Si dice che F è una σ -algebra su Ω se:

$$\emptyset \in F$$

$$A \in F \Rightarrow F, A^c \in F$$

$$A_i \in F, \forall i \in N \Rightarrow \bigcup_{i \in N} A_i \in F$$

Una σ -algebra F su Ω è non vuota ed ha fra i suoi elementi Ω ; essa è inoltre stabile per l'intersezione numerabile :

$$A_i \in F \forall i \in N \Rightarrow \bigcap_{i \in N} A_i \in F \text{ dal momento che } \bigcap_{i \in N} A_i = \left(\bigcup_{i \in N} A_i^c \right)^c$$

Si può quindi definire un ordinamento di σ -algre su Ω all'interno del quale la σ -algebra più piccola contenente sia \emptyset che Ω è detta σ -algebra di Borel.

Nel caso in cui si abbia a che fare con funzioni a variabili reali in cui la retta reale sia equipaggiata con la sua σ -algebra Boreliana si può definire uno spazio misurabile:

(Ω, F) è uno spazio misurabile e μ è una misura su $(\Omega, F) \Rightarrow (\Omega, F, \mu)$ è detto spazio di misura.

A partire dalla definizione di spazio di misura è possibile ricavare una definizione di spazio di probabilità:

DEF: una coppia (Ω, F) in cui Ω è un insieme e F una σ -algebra di sottoinsiemi di Ω è detta spazio numerabile.

DEF: una funzione m definita sulla σ -algebra F con $m = F \rightarrow [0, \infty]$ è una misura se:

$$\forall A_i \in F, k \in [0, \infty]; A_i \cap A_j = \emptyset, \forall i \neq j$$

e vale

$$m\left(\bigcup_k A_k\right) = \sum_k m(A_k)$$

In cui A_k è una collezione di insiemi

DEF: dato uno spazio misurabile la probabilità P è una misura sulla σ -algebra F tale che $P(\Omega) = 1$ da cui ne consegue che

DEF : Lo spazio dotato di misura (Ω, F, P) è detto spazio di probabilità.

Lo spazio di probabilità gode di alcune proprietà :

$$B \subseteq A \rightarrow P(B) \leq P(A)$$

$$P(\bar{A}) = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cap B) \leq \min(P(A), P(B))$$

La funzione di probabilità è continua

In particolare, sia Ω limitato e l'osservatore in stato d'ignoranza circa la distribuzione degli eventi in Ω , è corretto scegliere una misura invariante cioè la misura di Lebesgue; assunto $\Sigma = M(\sigma$ -algebra di un sottoinsieme L -misurabile Ω) e posto $\forall E \in M$

la probabilità dell'evento è data da :

$$P(E) = \frac{|E|}{|\Omega|}$$

nel caso di un evento fuzzy la probabilità associata sarà :

$$P(E) = \frac{\sum_{E \in \Omega} \mu_E}{|\Omega|}$$

In alternativa la costruzione di una probabilità a partire da valori di possibilità può essere formalizzata a partire dalla definizione frequentista di probabilità secondo cui la probabilità di un evento è rappresentabile da un integrale :

$$P(E) = \int_E f(x) dx$$

Ridefinibile in un solo passaggio sostituendo la funzione caratteristica con la funzione di appartenenza ottenendo così la probabilità di un evento fuzzy E come :

$$P(E) = \int_{\Omega} \mu_E(x) f(x) dx$$

È interessante notare come, nonostante la seconda definizione di probabilità sia meno robusta della prima nell'utilizzo con dati incompleti o incerti tipici degli studi ecologici, la sua definizione formale nel calcolo numerico sia coerente alla definizione di probabilità di Lebesgue prima descritta.

2.3 Compilatore naive Bayes

Un compilatore bayesiano è una metodologia di tipo probabilistico che, avvalendosi di una serie di dati etichettati, stima i parametri del modello generativo e classifica le nuove istanze osservate utilizzando il teorema di Bayes (John e Langley, 1995) e selezionando la classe o categoria o intervallo di valori in un dominio continuo che ha la probabilità maggiore di aver generato l'istanza. Detto anche *idiot's Bayes e simple Bayes* questo metodo di classificazione è considerato importante per diversi motivi fra cui la semplicità di implementazione e di calcolo dei parametri che non necessita di complicati schemi di calcolo.

Teorema di Bayes :

$$P(A_i|E) = \frac{P(E|A_i)P(A_i)}{P(E)} = \frac{P(E|A_i)P(A_i)}{\sum_{j=1}^n P(E|A_j)P(A_j)}$$

Fra i compilatori bayesiani, il naive Bayes è uno dei più semplici e versatili essendo basato sull'unica assunzione semplificativa che tutti gli attributi che descrivono un'istanza siano fra loro condizionatamente indipendenti (Pazzani, 1995; Kohavi, 1996). Nonostante la sua semplicità e le restrizioni dell'assunzione semplificativa, in pratica mai rispettata per la maggior parte di pattern complessi, da letteratura emerge come il compilatore funzioni sorprendentemente bene e come le sue prestazioni complessive siano le migliori fra i classificatori bayesiani conosciuti. L'assunzione di indipendenza condizionale viene detta assunzione naive Bayes ed in base a questa il classificatore esegue una classificazione di tipo MAP (maximum a posteriori probability). L'ambito applicativo del naive Bayes riguarda la classificazione di istanze descrivibili attraverso un insieme di attributi di cardinalità anche molto elevate. A differenza di altri altri algoritmi di apprendimento il naive Bayes non ricerca la soluzione nello spazio delle possibili soluzioni (aprioristiche) ma ne stima la veridicità sulla base delle frequenze osservate nelle istanze di apprendimento.

Sia $X = \{x_1, x_2, \dots, x_n\}$ un insieme di variabili che descrivono gli attributi delle istanze da classificare e sia $Y = \{y_1, y_2, \dots, y_m\}$ una variabile i cui stati rappresentano le categorie di appartenenza possibili delle istanze (Fig. 2). Il compilatore stima i parametri di una funzione che fornisce la classe di appartenenza dell'istanza stessa.

.

.

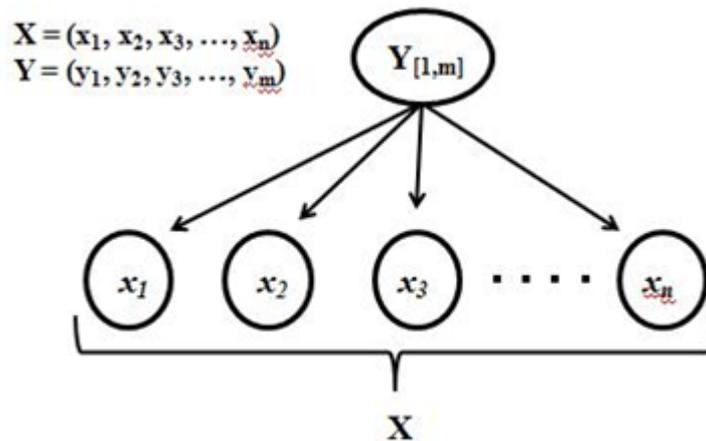


Figura 2: Schema concettuale del funzionamento di un classificatore naive Bayes in cui i parametri sono equivalenti alle probabilità condizionate di ogni n-esima “features”(x) data ogni classe m-esima(y).

L’assunzione d’indipendenza condizionale per cui le componenti delle istanze sono indipendenti all’interno di ciascuna classe fa in modo che

$$f(x|y) = \prod_{j=1}^n f(x_j|y)$$

da cui si desume che le distribuzioni marginali di ogni attributo delle classi sono stimabili separatamente. In conseguenza di questo, un problema n-multivariato risulta risolvibile come un problema n-univariato semplificando la parte computazionale della classificazione.

La relazione naive Bayes fra attributi delle istanze e classe di attribuzione è una relazione non lineare che non richiede monotonicità in x_j per cui risulta potenzialmente in grado di replicare qualunque dinamica di classificazione. L’applicazione del naive Bayes proposta in questo lavoro si basa principalmente su queste due ultime caratteristiche del classificatore che permettono rispettivamente il trattamento di data set multivariati ad elevata dimensionalità e l’individuazione nel sistema delle variabili di punti soglia che il modello identifica come punti di “treshold” che descrivono il passaggio da una classe ad un’altra.

2.4 Approccio fuzzy naive Bayes

L'alta complessità e l'alta incertezza associate ai sistemi ecologici possono essere trattate attraverso l'utilizzo congiunto di logica fuzzy e compilatore naïve Bayes. Il sistema derivante; denominato Fuzzy Bayes Ecological Model (F.B.E.M.); è un compilatore fuzzy naïve Bayes consistente in un compilatore naïve Bayes operante in ambiente fuzzy. Il compilatore come un normale naïve bayes elabora i dati per la stima dei parametri sotto forma di istanze.

L'algoritmo del compilatore è composto di 5 step successivi (Widyantoro e Yen, 2000) :

1. fuzzyficazione delle variabili
2. fase di apprendimento
3. inferenza fuzzy naïve Bayes sulle nuove istanze osservate
4. definizione del risultato utilizzando la regola del centroide all'interno dell'intervallo di incertezza risultante dal compilatore
5. defuzzyficazione dell'output

1 - Fuzzyficazione delle variabili : consiste nel descrivere le variabili considerate secondo formalismo fuzzy. Viene creata una partizione fuzzy all'interno del dominio di valori di ciascuna variabile per cui ogni dominio di sistema viene descritto da una successione di fuzzyset secondo la quale verranno fuzzyficate le osservazioni che compongono il set di training in base al quale il compilatore effettuerà la fase di learning.

2 - Fase di learning : durante la fase di learning il sistema stima i parametri del classificatore dati dalle probabilità marginali di ogni fuzzy set

$$P(x_i) = \frac{\left(\sum_{e \in L} \mu_{x_i}^e \right) + 1}{|L| + |dom(x_i)|}$$

In cui L è il set di apprendimento, $dom(x_i)$ indica il dominio della i-esimo fuzzy set della variabile x , e indica una generica istanza del set di apprendimento L,

e dalle probabilità condizionate di ogni fuzzy set dato il valore della variabile target

$$P(x_i|y) = \frac{\left(\sum_{e \in L} \mu_{x_i}^e \mu_y^e \right) + 1}{\left(\sum_{e \in L} \mu_y^e \right) + |dom(x_i)|}$$

Nella fase di learning I valori di probabilità congiunta e condizionata sono calcolati secondo la definizione di spazio di probabilità di Lebesgue e sono riferiti all'evento fuzzy cioè alla possibilità che una data variabile assuma valori inclusi in un particolare fuzzy set (Zadeh 1968).

3 - Inferenza fuzzy naïve Bayes: data una nuova istanza osservata, prevista o simulata il compilatore calcola per ogni fuzzy set della partizione della variabile target la probabilità condizionata del fuzzy set target rispetto alla nuova istanza osservata.

$$P(y|e) = P(y) \left(\sum_{x \in X_i} \frac{P(x_i|y)}{P(x_i)} \mu_{x_i}^e \right) \dots \left(\sum_{x_n \in X_n} \frac{P(x_n|y)}{P(x_n)} \mu_{x_n}^e \right)$$

L'output finale (della variabile target) è composto da una serie di probabilità condizionate associate ai fuzzy set di cui si compone la partizione della variabile target.

4 - I valori previsti sono ottenuti secondo criterio MAP (Maximum a posteriori probability) attraverso il metodo del centroide per cui si avranno uno o più valori secondo il carattere estremo delle nuove istanze a cui il compilatore associa un campo di incertezza calcolato in base alla somiglianza statistica con le istanze di training.

L'utilizzo congiunto di logica fuzzy e compilatore naïve bayes rende l'output di classificazione simile per accuratezza all'output di un metodo predittivo. Il dominio della variabile output (variabile target) sottoposto a fuzzyficazione non è infatti più rappresentato da classi discrete ma da classi sfumate.

Il dominio risulta quindi continuo rispetto ai possibili valori della variabile descritta.

La classificazione finale non è quindi assoluta ma definisce un range d'incertezza

all'interno del dominio continuo della variabile target. Nel range d'incertezza attraverso la regola del centroide viene definita una previsione puntiforme analoga per precisione ad una previsione deterministica.

L'ampiezza del range d'incertezza in cui ricade la classificazione è funzione della partizione fuzzy iniziale e della somiglianza statistica fra i dati di training e le nuove istanze.

2.4.1 Considerazioni sul metodo

L'approccio fuzzy naïve Bayes si presenta robusto rispetto al rumore inevitabilmente presente in un set complesso di dati . I set di output dovuti ad una o più particolari forzanti mostrano una diminuzione generale della variabilità in quanto la classificazione è influenzata dall'intero set di training. Una misura di variabilità reale degli output è calcolabile in riferimento ai diversi range d'incertezza dei risultati.

Il rumore presente nei dati viene “splittato” nelle varianze associate alle diverse variabili per cui in fase di elaborazione si osserva una diminuzione del rumore totale in ragione, appunto, della distribuzione del rumore. Quest'ultima caratteristica rende l'approccio potenzialmente adatto all'analisi di risposte ecologiche a eventi estremi climatici o di origine antropica. La robustezza dell'approccio rispetto pattern d'apprendimento ad elevata variabilità è legata all'elevato numero di parametri che vengono stimati durante la fase di apprendimento. Se si considera n come il numero di variabili considerate e k il numero di fuzzy set della partizione, durante la fase di apprendimento vengono stimati $k^2(n-1) + k(n-1) + 1$ parametri .

Approcci analoghi sono stati utilizzati per la descrizione, l'analisi e la modellizzazione di pattern di dati complessi in diversi ambiti, esistono infatti applicazioni fuzzy naive Bayes in oceanografia (Randon et al., 2008), robotica (Bustamante et al., 2007) e classificazione del testo (Storr, 2002)

Ogni nuova istanza osservata può essere utilizzata in un procedimento di autocalibrazione del modello aggiornando la fase di learning.

.3. APPLICAZIONE DELL'APPROCCIO F.B.E.M.

3.1 Il progetto "Theseus"

L'approccio fuzzy bayesiano ha trovato applicazione all'interno del progetto di ricerca europeo denominato "Theseus"(www.theseusproject.eu).

Il progetto Theseus fa parte dei "large cooperative project" del settimo programma quadro promossi dall'unione europea vi partecipano 31 istituti di ricerca appartenenti a 12 paesi dell'unione europea e altri 7 istituti di paesi extracomunitari.

La macroarea di ricerca all'interno della quale il progetto prende vita è quella relativa allo studi e la preservazione delle zone costiere che si estendono per circa 17000 Km lungo 20 dei 27 stati membri dell'EU occupando un'area di 2 milioni di Km quadrati in cui vivono più di 70 milioni di persone.

Il progetto si sviluppa attorno alle 4 evidenze chiave :

1. Grandi tratti delle coste europee, densamente popolate e di grande rilevanza economica, sono già ora sotto la minaccia di fenomeni di erosione e di flooding costieri.
2. Il "Global Climate Change" e l'innalzamento del livello medio marino potranno determinare un incremento nella frequenza e nell'intensità di eventi erosivi e di flooding.
3. Gli approcci gestionali fin qui utilizzati vanno reinquadrati alla luce della direttiva Habitat in modo da integrare la sicurezza dell'uomo con quella degli habitat costieri.

4. L'Unione Europea non ha ancora sviluppato un approccio integrato per la valutazione e la gestione dei rischi costieri che comprendesse le esigenze umane ed ambientali.

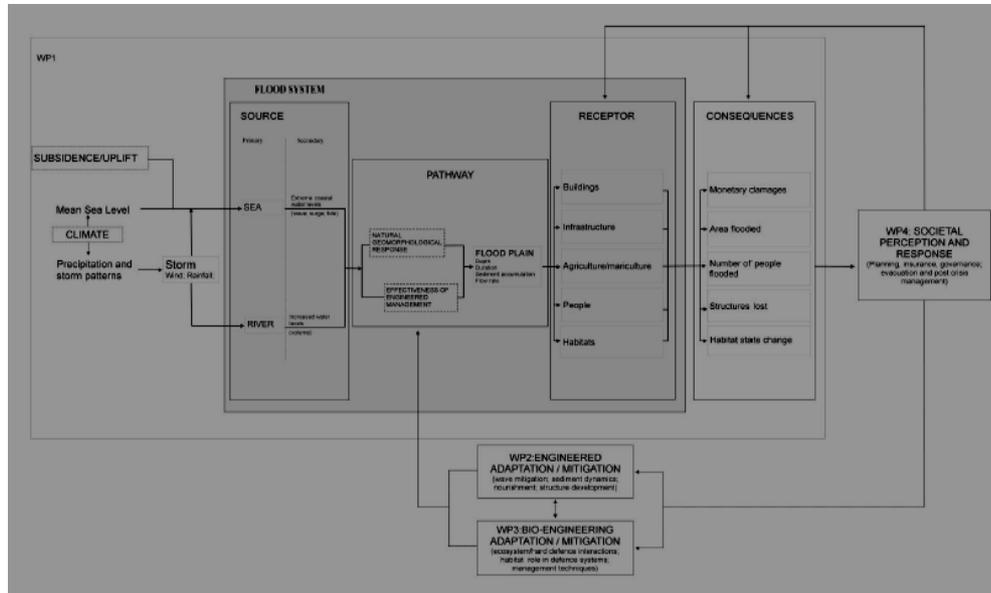


Figura 3 : Schema del progetto "Theseus" e schema di riferimento del Decision Support System basato sul principio Source Pathway Receptors Consequences (SPRC)

Alla luce di questi punti chiave e degli ormai riconosciuti limiti delle strategie di difesa costiera tradizionali, il consorzio "Theseus" si propone il non semplice obiettivo di sviluppare metodologie di valutazione e difesa dei rischi connessi ad eventi di flooding che tengano conto di implicazioni ambientali, sociali, economiche ed ingegneristiche il tutto in ragione dell'incertezza derivante dal contesto previsionale del "Global Climate Change".

Il progetto (Fig. 3) è strutturato secondo lo schema "Sources-Pathway-Receptors-Consequences" (SPRC), sorgenti-percorso-recettori-conseguenze, in cui l'informazione viene via via integrata e portata a livelli superiori di complessità fino ad una valutazione del rischio che tenga conto di componenti sociali, economiche, ambientali ed ingegneristiche. Il cammino SPRC sarà iterato inserendo ai vari livelli opzioni di mitigazione provenienti sia dalle singole aree disciplinari sia dalla loro interazione con l'obiettivo di affiancare allo studio integrato degli effetti del flooding una proposta integrata delle migliori opzioni di mitigazione secondo le specifiche dei casi.

Lo schema SPRC è stato applicato ad 8 siti (Tab. 2) individuati fra gli stati facenti parte del progetto, i siti sono stati scelti in modo da ricoprire le tipologie geografiche, ambientali, economiche e sociali presenti nell'Unione Europea cui vanno riferiti i risultati del progetto. Infatti i risultati ottenuti dal progetto Theseus verranno validati su scala continentale mentre le loro implicazioni finali dovranno essere applicabili da scala locale a scala regionale.

| Site name and number | Location | Geomorphological setting | Relevant urban centre | Habitats | Important activities | |
|----------------------|---|--------------------------------|--|---|---------------------------------|---|
| 1 | Vistula plain and environs, Poland | Baltic Sea | Spit | Gdansk, Sopot, Gdynia | Beach dunes | Harbour, fishery, chemical industry, airport |
| 2 | Elbe estuary, Germany | North Sea | Large estuary | Hamburg | Inter-tidal flats | Harbour, fisheries, nuclear power plant, (petro)chemical industry |
| 3 | Santander Bay, Spain | Gulf of Biscay, Atlantic Ocean | Open coast and spit enclosed bay/estuary | Santander | Beach, dunes, inter-tidal flats | Harbour, fisheries, airport, tourism |
| 4 | Scheldt estuary, Belgium and the Netherlands | North Sea | Large estuary | Antwerp | Inter-tidal flats | Harbour, fisheries, nuclear power plant, (petro)chemical industry |
| 5 | South Devon, UK (Plymouth Sound to the Exe estuary) | English Channel | Open coast with small estuaries | Teignmouth, Plymouth, Exeter, Newtown Abbot | Beach, dunes, inter-tidal flats | Harbour, fisheries, railway, port and naval base |
| 6 | Gironde estuary, France | Atlantic Ocean | Large estuary | Bordeaux | Beach, dunes, inter-tidal flats | Harbour, marina, fisheries, refinery, nuclear power station |
| 7 | Po delta and adjoining coast | Mediterranean Sea | Delta plain | Cesenatico, Bellocchio | Beach, dunes, inter-tidal flats | Harbour, (petro)chemical industry, tourism |
| 8 | Varna, Bulgaria | Black Sea | Open coast with small inlets | Varna | Beach, dunes | Harbour, chemical industry, power plants |

Tabella 2: Siti di studio del progetto "Theseus" con riportati il bacino di riferimento, la tipologia ecosistemica, i centri abitati rilevanti, gli habitat e le attività prevalenti.

Scendendo nel particolare del contesto ecologico, questo si inserisce nella struttura della ricerca al terzo livello corrispondente ai recettori degli eventi di flooding, gli obiettivi principali del gruppo di ricerca di ecologia (work package 3; WP3) sono :

- Valutare la vulnerabilità dei differenti habitat ad eventi di flooding costiero.
- Valutare i possibili impatti ecologici dovuti all'attuazione di opere di difesa da flooding ed erosione.
- Studiare e proporre possibili azioni di mitigazione "ecologiche" valorizzando le capacità di difesa intrinseche ai vari habitat.

- Stimare la vulnerabilità ecologica agli eventi di flooding previsti per il futuro.
- Stimare l'incidenza della sitospecificità nelle diverse possibili combinazioni di multiple stressors naturali ed antropici.

Per quanto riguarda i siti di ricerca italiani è stata individuata un'area lungo le coste del nord Adriatico all'interno della quale la scelta è ricaduta sui siti di Cesenatico e Bellocchio differenti per storia evolutiva tipologia di habitat e livello di impatto antropico. All'interno dei due siti sono riconoscibili diversi habitat costieri potenzialmente soggetti ad impatto da flooding quali : zona intertidale, spiagge, lagune costiere, pinete, sistemi di dune.

L'approccio fuzzy-bayesiano è stato utilizzato per l'analisi descrittiva e previsionale delle zone intertidali di spiagge sabbiose individuate come un importante recettore delle conseguenze di eventi di flooding in considerazione del loro posizionamento, della loro abbondanza e della sensibilità alle variazioni ambientali propria degli organismi scelti come indicatori degli stati ecosistemici.

I risultati provenienti dalle simulazioni dei modelli generativi FBEM sono quindi stati utilizzati per completare gli scenari climatici col dato ecologico, in modo da ottenere degli scenari ambientali più completi, per costruire curve evolutive della zona intertidale e curve di danno da implementare all'interno di un Decision Support System (DSS) che rappresenta il prodotto finale del progetto stesso e per stimare una matrice di parametri da implementare nel DSS assieme all'equazione generale del modello per valutare l'incidenza delle condizioni ambientali specifiche sull'evoluzione degli habitat e sulla loro risposta agli eventi di flooding previsti.

3.1.1 Decision Support System

Gli studi settoriali svolti nell'ambito del progetto "Theseus" confluiranno all'interno di sistema di supporto alle decisioni, un decision support system (DSS), che dovrà fornire ai gestori costieri le informazioni necessarie per una gestione ottimale delle zone costiere riguardo alle problematiche di flooding ed erosione. Un DSS è un sistema informativo (software) che converte dati in informazioni da fornire ai gestori, in caso di

particolari problematiche, al fine di permettere loro di prendere decisioni rapide ed efficaci.

In via teorica un DSS dovrebbe permettere all'utente/decisore di pianificare, gestire e controllare le attività delle quali è responsabile. Un DSS si appoggia a dati di un database o di una base di conoscenza (insieme di regole), che aiutano l'utilizzatore ad indirizzare in maniera ottimale scelte strategiche; non è solo un'applicazione informatica, perché può contenere anche strumenti di "Management Intelligence" e di tecnologie dei Sistemi Esperti, quali modelli di supporto decisionale.

Le decisioni prese dal *decision support system* possono essere di vario tipo :

- oggettiva/soggettiva
- routine/straordinaria
- semplice/complessa
- operativa
- tattica/strategica

L'architettura del sistema DSS del progetto Theseus è stata concepita in modo da ricalcare la struttura del progetto secondo lo schema precedentemente descritto SPRC mentre il software prodotto è operativo su base G.I.S.

Esulando da discorsi tecnici, che rimangono al di fuori delle mie competenze e dell'argomento di questa tesi di dottorato, è opportuno spiegare alcune peculiarità del DSS "Theseus" che risulta innovativo per varie ragioni.

Innanzitutto è importante delineare la complessità dell'architettura del DSS in quanto, a differenza della maggior parte dei DSS ambientali attualmente in circolazione, essa riprende la complessità dei sistemi reali nel tentativo di replicare le connessioni che intercorrono nei sistemi naturali fra componenti biotiche ed abiotiche e fra componenti naturali ed antropiche. La necessità di rendere il DSS utilizzabile in differenti siti su scala continentale ha imposto alla struttura del DSS stesso un'elevata elasticità sia nell'interazione con il decisore finale sia nell'analisi dei dati di applicazioni specifiche. Di norma infatti, vista la complessità e l'incertezza delle problematiche affrontate, i

DSS ambientali sono sistemi di regole e parametri prestimati che confluiscono in base a quanto desiderato dall'utente finale in un risultato aggregato-cumulativo, derivante dal calcolo di indici o da un'analisi multicriteriale che sono poco adatti ad un contesto di studio complesso. Inoltre non sono in grado di considerare l'incertezza legata ad alcuni sistemi naturali e comunque presente sempre, in qualche misura, nel momento in cui l'obiettivo dello studio è previsionale.

Il driver conduttore del DSS è legato alla problematica centrale del progetto, in questo caso gli eventi di flooding costiero, per cui le sorgenti sono rappresentate da un data set contenente le stime future dell'evoluzione climatica ed in particolare dei fenomeni di sea level rise (innalzamento del livello medio marino) e di aumento nella frequenza ed intensità dei fenomeni meteorologici di forte intensità.

Il DSS contiene al suo interno due modelli di propagazione delle onde sottocosta applicabili, caso per caso a seconda della morfologia costiera delle zone prese in esame, unitamente ad un modello specifico per stimare la "storm surge", il brusco innalzamento locale del livello marino dovuto a eventi tempestosi, secondo l'intensità dell'evento, l'altezza dell'onda frangente sottocosta e la morfologia dei fondali.

In funzione di sea level rise, dello storm surge, dell'altezza e ripidità dell'onda frangente e delle caratteristiche locali dei fondali, tre funzioni (modelli) stimano l'intensità, la velocità e la durata dell'eventuale conseguente evento di flooding.

I valori così stimati dei parametri fisici di flooding confluiscono nei recettori ecosistemici, sociologici ed economici in cui le conseguenze del flooding sono appropriatamente stimate in funzione di altri parametri specifici dei sopradetti contesti. Il DSS permette inoltre di scegliere fra diverse opzioni di mitigazione relative agli ambiti di studio ed integrate fra i diversi settori. In base alla scelta compiuta, il sistema stima le conseguenze in funzione della variazione dei parametri di flooding indotte dalle misure di contenimento. Ad ogni opzione di mitigazione è associata una funzione, anch'essa implementata all'interno del DSS, descrittiva dell'effetto mitigatore sulle variabili di flooding.

Al proprio interno il DSS contiene un data base riportante 4 scenari climatici corrispondenti agli anni: 2010 (considerato come la situazione attuale), 2020, 2050, 2080 che possono essere utilizzati dal decisore il quale ha comunque la possibilità di fornire al DSS valori delle variabili di flooding (intensità, velocità, durata, altezza

dell'onda frangente, ripidità dell'onda frangente) derivanti da scelte personali o specifici di situazioni particolari che si intende testare.

3.2.1 Habitat recettore: intertidale di spiagge sabbiose

Nei siti di studio italiani, fra gli habitat recettori degli effetti indotti dal sea level rise e dal flooding, è stato scelto l'habitat intertidale sabbioso in relazione ad alcune sue peculiarità quali:

- L'ampia diffusione geografica.
- La posizione intermedia fra terra e mare che ne fa un ottimo recettore degli stress provenienti da entrambe le direzioni.
- La sensibilità degli indicatori dell'intertidale.
- La complessità della risposta della zona intertidale in cui è possibile osservare effetti dovuti a tutte le componenti di flooding e per qualsiasi grado d'intensità degli eventi considerati.

3.3 Le spiagge

Considerate a lungo dei deserti marini e ampiamente trascurate dagli studiosi, le spiagge rappresentano, invece, gli habitat tra i più vitali e più produttivi del pianeta dal punto di vista ecologico, economico e sociale. Esse forniscono, infatti, numerosi servizi ecosistemici che sono essenziali per supportare l'uso che l'uomo ne fa. I servizi ecosistemici più importanti includono:

- deposito e trasporto di sedimenti;
- effetto tamponante dell'energia scaricata sulla costa dal moto ondoso e dagli eventi estremi (uragani e tsunami);
- risposta dinamica al *sea level rise* (con limiti);
- trasformazione della sostanza organica e abbattimento degli inquinanti;

- filtrazione e purificazione dell'acqua;
- mineralizzazione e riciclo dei nutrienti;
- mantenimento della biodiversità;
- aree nursery per i giovanili dei pesci;
- sedi di nido per tartarughe ed uccelli marini, e ripari per i pinnipedi;
- paesaggi panoramici, e opportunità ricreative;
- legami funzionali tra l'ambiente marino e terrestre nella zona costiera;

(Defeo *et al.*, 2009).

La fascia intertidale, caratterizzata da substrato mobile, fornisce così due habitat per le popolazioni della fauna bentonica. L'habitat tipico del macrobenthos, quello rappresentato dalla matrice sabbiosa; e l'habitat interstiziale del meio e microbenthos. Il primo forma un sistema prevalentemente a due dimensioni (la dimensione verticale in particolare è sfruttata dagli organismi "fossori", tramite però il dislocamento della sabbia), uniforme ma dinamico e instabile, il secondo più stabile, ma più complesso e con chiari gradienti verticali, forma pienamente un sistema a tre dimensioni (McLachlan, 1983).

La composizione dei popolamenti bentonici intertidali, e le abbondanze relative degli organismi che li rappresentano, sono di norma determinate da tutte quelle variabili fisiche che tipicamente caratterizzano la spiaggia e di cui si è già accennato sopra. Tra le diverse tipologie di spiagge, inoltre, quelle riflessive presentano una ricchezza in specie generalmente minore rispetto a quelle dissipative, dove le condizioni fisiche creano un ambiente più favorevole all'insediamento degli organismi e alla colonizzazione dell'habitat (McLachlan, 2006).

3.3.1 Descrittori degli stati ecosistemici

Fra i possibili descrittori degli stati ecosistemici le variabili biotiche appaiono essere quelle con il maggior grado di "memoria" delle variazioni di sistema e quindi quelle che, potenzialmente, permettono di descrivere il maggior numero possibile di stati sistemici. In particolare la componente biotica è considerata un buon descrittore per le seguenti motivazioni:

- I fattori biologici sono molto sensibili alle variazioni sia dei parametri chimici che dei parametri fisici dell'ecosistema.
- I fattori biologici sono influenzati da molti fattori ambientali
- I fattori biologici sono influenzati dalle interazioni dei fattori ambientali
- Soprattutto in caso di fenomeni ad ampia intensità e relativamente limitata durata, come i fenomeni di flooding, i fattori biologici forniscono una risposta che non ha nessuna relazione di feedback che rischi di creare problemi al momento dell'interpretazione dei dati.

Fra i possibili descrittori biologici si è scelto di utilizzare la fauna bentonica nelle sue due componenti dimensionali della fauna macrobentonica ($> 500\mu$) e meiobentonica ($63\mu < \varnothing < 500\mu$). La scelta della fauna bentonica come descrittore è motivata anche dal fatto che, oltre a godere delle proprietà sopra descritte, la fauna bentonica è onnipresente lungo tutte le coste sabbiose, il che oltre a farne un ottimo descrittore sincretico permette di usare la fauna bentonica per stabilire un continuum di stati ecosistemici possibili per le spiagge e di valutare al meglio le caratteristiche di sito specificità degli ecosistemi.

3.4. Caratterizzazione delle aree di studio

La fascia costiera Emiliano-romagnola forma un sistema litoraneo complesso che con i suoi 130 km di costa bassa e sabbiosa rappresenta, sotto il profilo turistico-balneare, una delle spiagge più importanti d'Europa. Il sistema scorre praticamente ininterrotto, se non per la presenza di canali, lagune, foci e strutture artificiali, dal margine degli Appennini a sud, fino alla foce del Po di Goro a nord. La costa è soggetta a condizioni di vento moderato e a un range di marea (circa 80 cm) abbastanza ampio se confrontato con i valori registrati nel resto del bacino del Mediterraneo. La costa è caratterizzata da spiagge generalmente definite di tipo 'dissipativo' soggette, per la maggior parte, a fenomeni erosivi.



Figura 4: Posizionamento dei siti studio sulle coste nord adriatiche occidentali da nord a sud Lido di Spina, Bellocchio, Cesenatico.

3.4.1 Cesenatico

Il sito di studio di Cesenatico è situato sulla costa nord adriatica fra Rimini e Ravenna (Fig.4). Il sito è caratterizzato da una lunga spiaggia sabbiosa “pesantemente” gestita per la presenza di stabilimenti balneari. La spiaggia di Cesenatico è composta da “spiaggia Ponente” e “spiaggia Levante” rispettivamente a nord e sud del porto canale. Le due spiagge, entrambe dissipative, sono caratterizzate da un diverso regime di trasporto per cui la spiaggia “Ponente” risulta in forte erosione mentre la spiaggia “Levante” appare stabile nel tempo nella sua lunghezza. Il campionamento di Cesenatico ha coinvolto altri due siti Cervia e Lido di Dante complementari a Cesenatico per caratteristiche morfodinamiche e gestionali.

3.4.2 Lido di Spina – Bellocchio

Lido di Spina

Il sito di studio di Lido di Spina (Fig.4), è il più a sud dei sette lidi di Comacchio e vanta di una spiaggia tra le più lunghe della riviera. La spiaggia si estende dalla foce del canale di Logonovo fino al Bagno Giamaica, dove, 5 pennelli in pali di legno costruiti nel 1999 la separano dalla Riserva Naturale delle Vene di Bellocchio. I 600 m a Nord

del primo pennello si trovano in una situazione di forte erosione, mentre il restante tratto di spiaggia è in continuo avanzamento. Per tale motivo la spiaggia è stata soggetta ogni primavera a partire dal 1995 a interventi di difesa costiera quali ripascimenti, concentrati in particolare nella zona a sud, consentendo così l'avanzamento della linea di riva e il normale svolgimento della stagione balneare turistica (Avanzi & Marzetti, 2011). Il sedimento raggiunge anche la zona a nord della spiaggia, non in erosione, grazie al trasporto operato dalla corrente lungo costa.

Bellocchio

Il sito di studio comprende la zona intertidale della spiaggia antistante alla laguna di Bellocchio (Fig. 4). Tale spiaggia ricade all'interno del Parco Naturale Regionale del Delta del Po, e nel suo complesso rappresenta un sito ad elevata diversità ambientale. L'intera area si estende per 510 ettari e al suo interno comprende: un sistema di dune sabbiose attive, la fascia marina antistante, una grande sacca salmastra, la foce del fiume Reno, zone umide d'acqua dolce e salmastra, pinete, boscaglie costiere e bosco pianiziale di temerci ed olivi. Il cordone litoraneo ha molto risentito negli ultimi decenni del marcato processo di erosione costiera che interessa la foce del Reno. Le mareggiate entrano sempre più frequentemente nell'entroterra causando la riduzione della laguna, o cosiddetta Sacca di Bellocchio, formatasi nella prima metà del 1900 e l'aumento della salinità dei terreni retrostanti. Tale fenomeno procede con ritmo molto sostenuto, sottraendo materiale al cordone di dune che la difende e determinando l'arretramento della spiaggia verso l'entroterra con una media di 10 metri l'anno (ARPA, 2007)

3.4.3 Inquadramento generale dei siti

Per mostrare un quadro generale dei valori osservati per abbondanza e numero di taxa sia per la macrofauna che per la meiofauna e delle variabili ambientali principali nei siti studio, vengono riportate a seguire due tabelle riassuntive dei valori medi delle variabili ambientali e dei valori medi delle variabili target per ciascuna spiaggia (Tab. 3 e 4).

| Spiaggia | | TOM | Capulerio (g) | Mz (mm) | Classazione (phi) | 1/slope (1/m) | BDI |
|---------------|-------|-------|------------------|------------|----------------------|------------------|---------|
| Cesentico | Media | 0,576 | 38,969 | 0,161 | 0,583 | 37,713 | 166,228 |
| | ES | 0,015 | 4,320 | 0,004 | 0,032 | 2,599 | 7,755 |
| Cervia | Media | 0,847 | 96,540 | 0,160 | 0,532 | 43,278 | 244,027 |
| | ES | 0,055 | 12,131 | 0,006 | 0,053 | 1,938 | 7,621 |
| Lido di Dante | Media | 0,590 | 215,480 | 0,312 | 0,513 | 15,529 | 45,710 |
| | ES | 0,016 | 22,281 | 0,005 | 0,028 | 0,553 | 2,569 |
| Lido di Spina | Media | 0,814 | 17,450 | 1,937 | 0,733 | 28,250 | 15,085 |
| | ES | 0,063 | 1,595 | 0,027 | 0,019 | 1,648 | 0,928 |
| Bellocchio | Media | 4,203 | 30,782 | 4,328 | 1,002 | 50,513 | 13,530 |
| | ES | 0,581 | 3,673 | 0,214 | 0,106 | 4,126 | 1,370 |

Tabella 3: Valori medi ed errore standard delle principali variabili abiotiche campionate nelle 5 spiagge (TOM=contenuto percentuale di materia organica dei sedimenti, Capulerio=frazione conchigliare residua, Mz=mediana granulometrica, classazione del sedimento, 1/slope= inverso della pendenza delle spiagge, BDI=indice morfodinamico delle spiagge (citazione))

| Spiaggia | Macrofauna | | Meiofauna | | |
|---------------|------------|------------|-----------|------------|---------|
| | Specie | Abbondanza | Specie | Abbondanza | |
| Cesentico | Media | 3,069 | 10,389 | 4,889 | 300,574 |
| | ES | 0,227 | 1,244 | 0,166 | 45,986 |
| Cervia | Media | 3,542 | 13,194 | 5,056 | 457,185 |
| | ES | 0,181 | 1,850 | 0,119 | 53,009 |
| Lido di Dante | Media | 3,625 | 21,736 | 5,796 | 272,204 |
| | ES | 0,231 | 3,630 | 0,207 | 48,687 |
| Lido di Spina | Media | 3,953 | 11,172 | 7,500 | 416,625 |
| | ES | 0,298 | 1,527 | 0,202 | 39,546 |
| Bellocchio | Media | 6,292 | 453,813 | 6,396 | 733,167 |
| | ES | 0,406 | 104,874 | 0,349 | 220,088 |

Tabella 4: Valori medi per campione ed errore standard delle variabili target della macrofauna e della meiofauna, come n°di specie ed abbondanza, nelle 5 spiagge campionate.

Per fornire un'ulteriore informazione circa le relazioni che intercorrono fra le suddette variabili ambientali ed i target indagati sono state calcolate delle regressioni secondo polinomiali di primo e secondo ordine fra tutti i target e tutte le variabili ambientali utilizzando i valori medi per transetto. Dalle regressioni (Figg. 5 – 8) il dato che emerge è che non esiste una variabile ambientale in grado da sola di spiegare bene il grado di

dipendenza delle componenti biotiche, di contro tutte le regressioni mostrano come esistano solo porzioni piccole d'informazione legate all'effetto delle singole variabili ambientali.

Si può quindi pensare che l'effetto complessivo di causalità del sistema fisico/chimico su quello biologico sia in parte esprimibile da interazioni anche non lineari fra le variabili abiotiche.

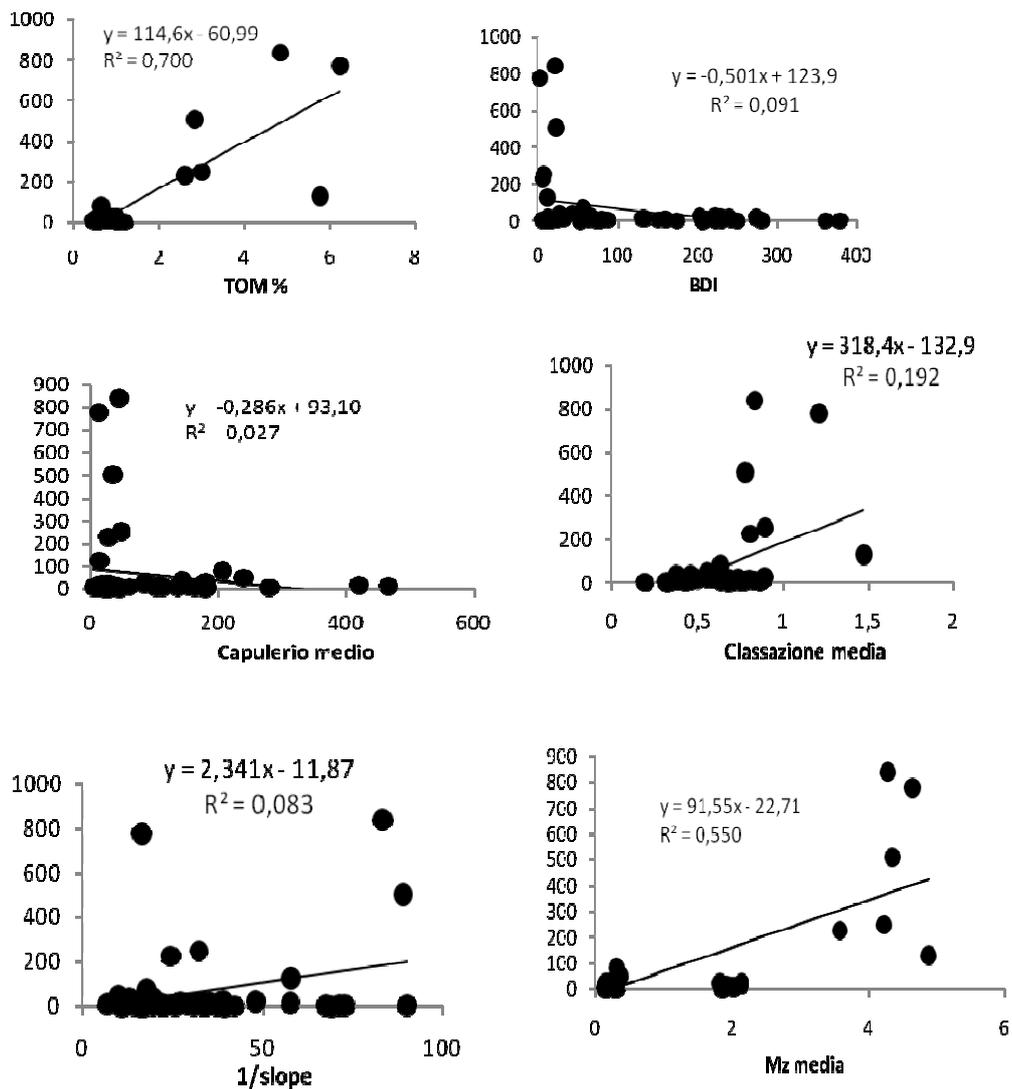


Figura 5: Regressioni fra abbondanza media della macrofauna per transetti e valori medi per transetto delle principali componenti abiotiche.

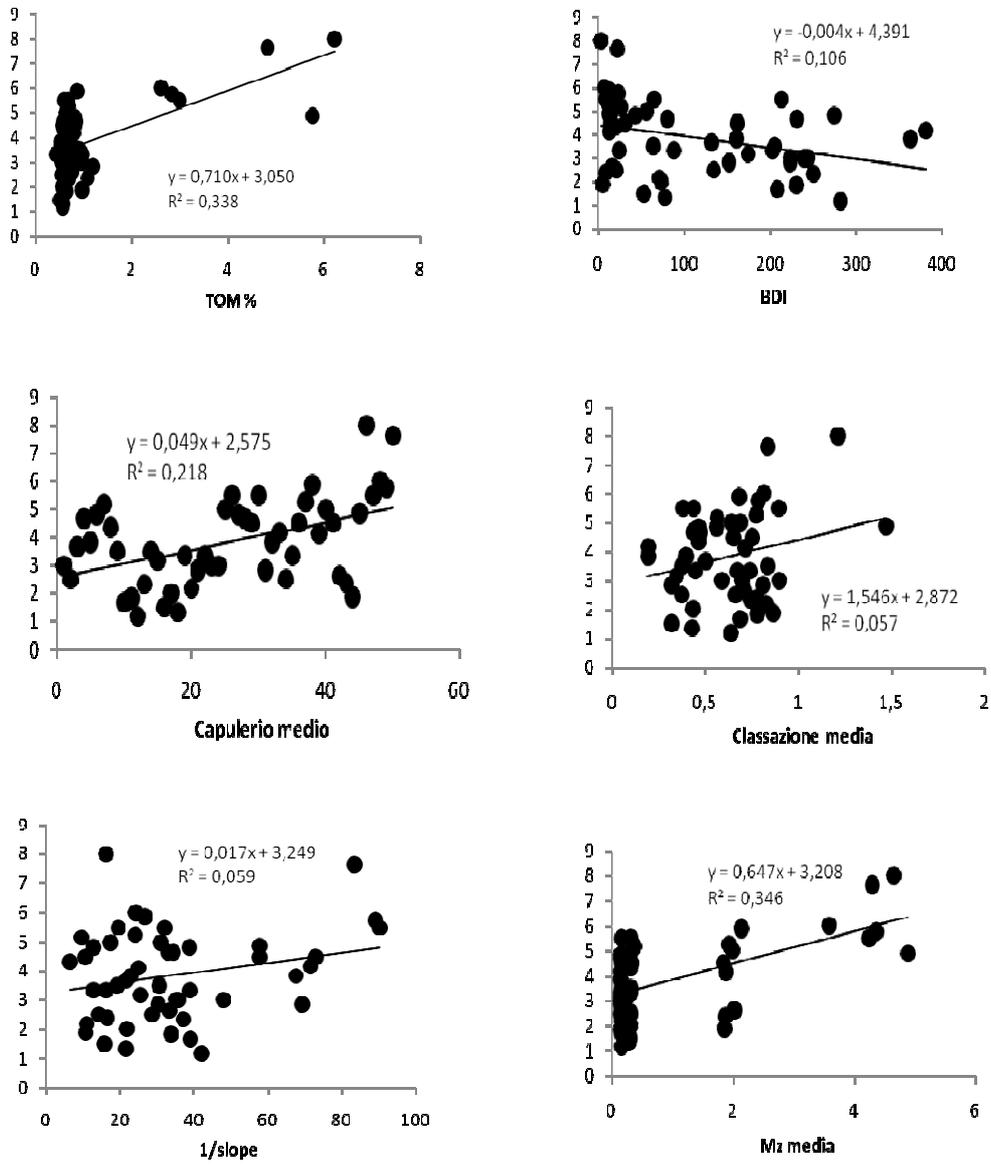


Figura 6: Regressioni fra n° di taxa medi della macrofauna per transetto e valori medi per transetto delle principali componenti abiotiche.

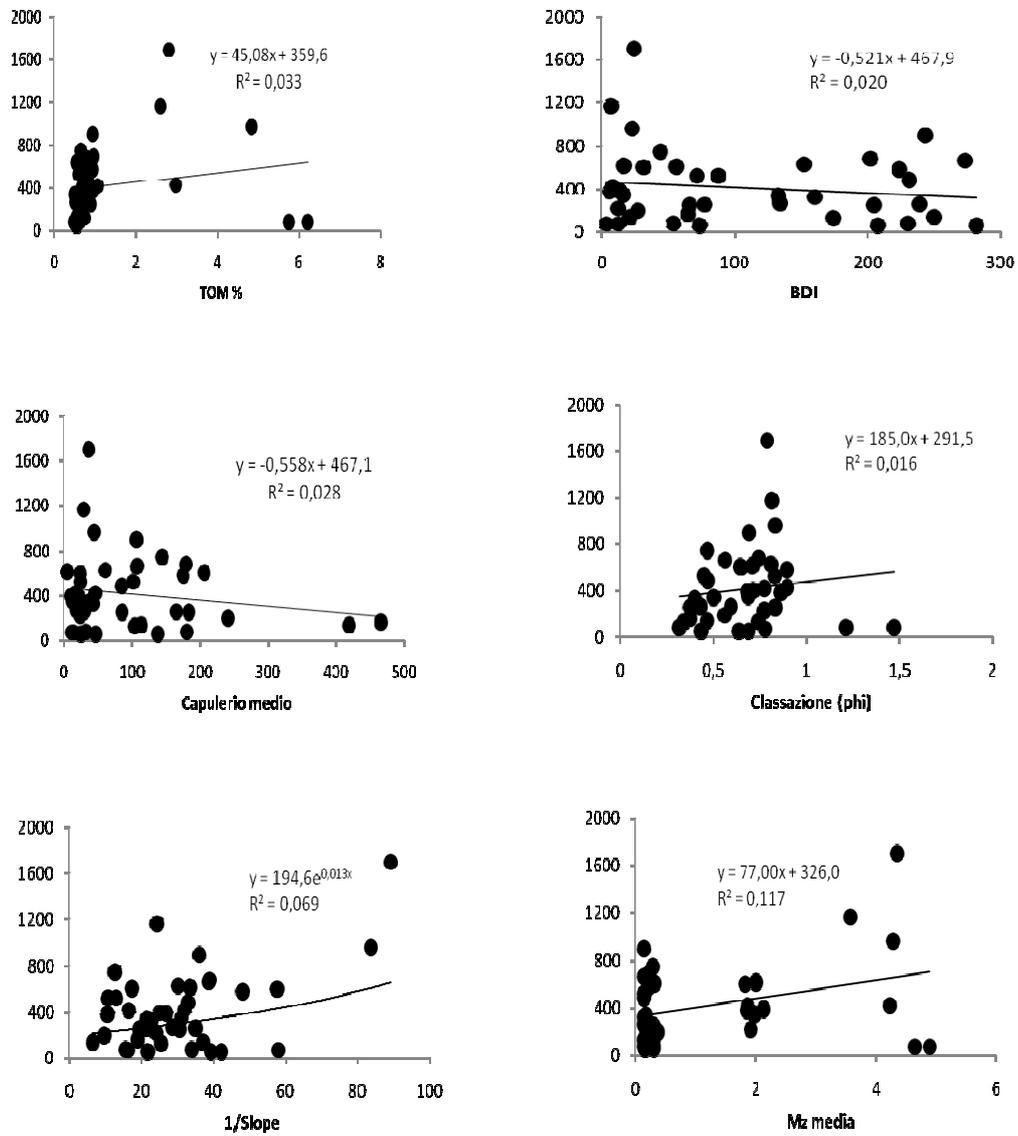


Figura 7: Regressioni fra abbondanza media di meiofauna per transetto e valori medi per transetto delle principali componenti abiotiche.

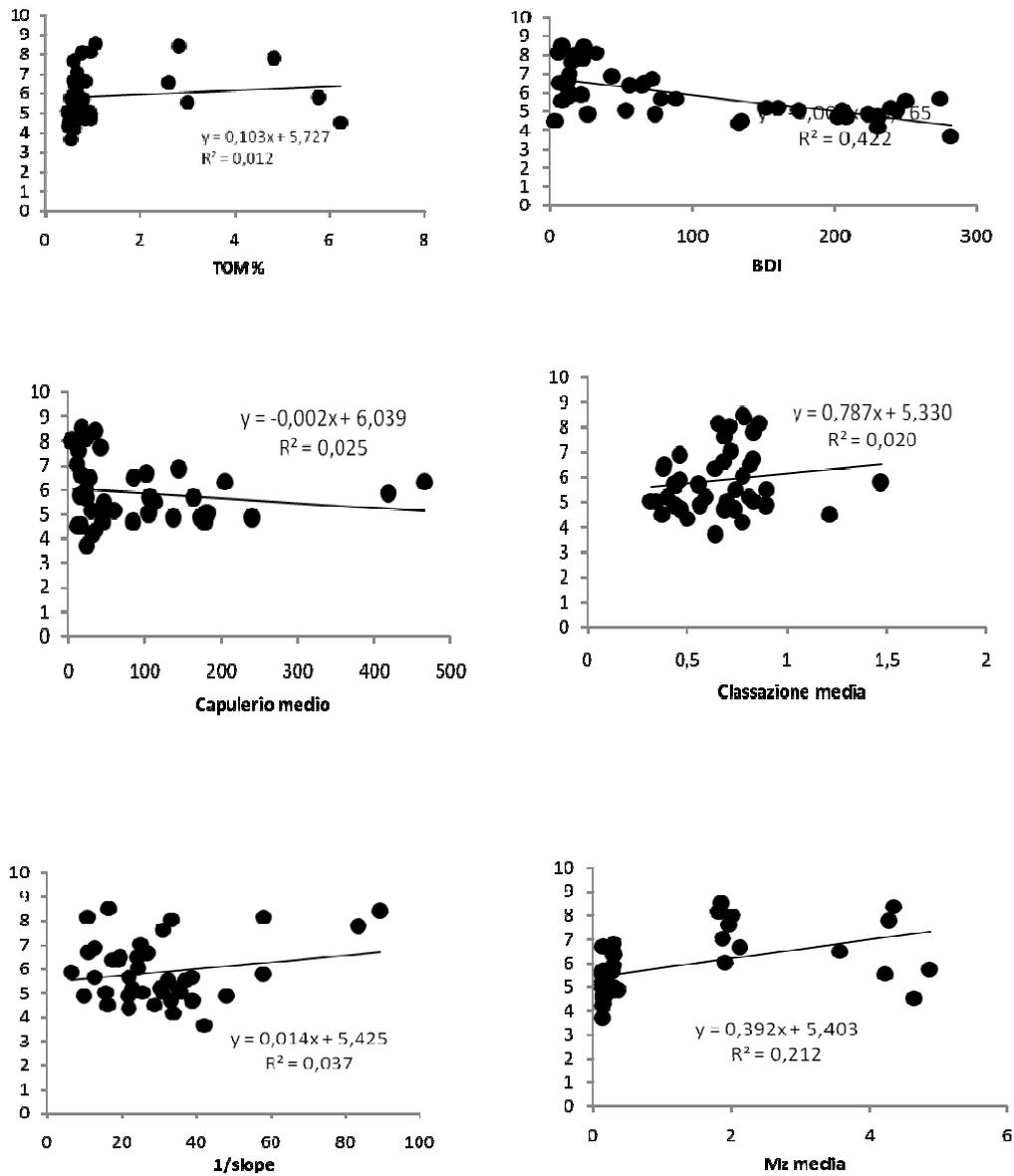


Figura 8: Regressioni fra n° di taxa medi di meiofauna per transetto e valori medi per transetto delle principali componenti abiotiche.

3.5 Scenari di flooding

Le simulazioni da modello fuzzy naive-Bayes si basano sugli scenari di flooding stimati dagli ingegneri costieri del DICAM (Dipart. di Ingegneria Civile Ambientale) di Bologna in relazione a quanto emerso dalle simulazioni climatiche portate a termine dal gruppo di modellistica oceanografica e climatica del CNR di Trieste e dell'istituto CORILA di Venezia.

Le simulazioni sono state organizzate in 4 scenari corrispondenti ai livelli di innalzamento del livello medio marino degli anni 2010(0,00 m), 2020(0,07 m), 2050(0,13 m) e 2080(0,23 m);

per ogni scenario sono descritte le variabili di flooding relative a 8 eventi, estratti dalla curva cumulativa degli eventi meteorologici previsti per il periodo considerato (2010-2100), e associati a frequenze di accadimento di eventi (storm surge, altezza dell'onda) corrispondenti ai tempi di ritorno di 2, 5, 10, 20, 25, 30, 50 e 100 anni.

Per ogni evento di ogni scenario sono state effettuate simulazioni fuzzy-bayesiane della risposta di macrofauna e meiofauna bentoniche di cui sono state considerate le variabili di abbondanza totale, di numero di taxa e di diversità stimata con l'indice di Shannon.

Per i medesimi eventi sono stati costruiti separatamente i 4 scenari relativi a Cesenatico e quelli relativi a Bellocchio, tenendo in considerazione le diverse, specifiche morfologie dei due siti per quanto riguarda intensità e durata dei flooding ed altezza dell'onda frangente.

3.6 Cesenatico: applicazione FBEM

L' applicazione modellistica per la zona intertidale del sito di Cesenatico è stata basata su un campionamento effettuato su 3 siti del nord Adriatico : Lido di Dante, Cervia e Cesenatico.

I siti sono stati scelti in modo da rappresentare un gradiente nel continuum di dissipatività che caratterizza le spiagge del nord Adriatico. In particolare le tre spiagge seguono un gradiente di dissipatività crescente da Lido di Dante a Cervia passando per Cesenatico. Inoltre in ottica gestionale i tre siti rappresentano situazioni diverse in quanto Lido di Dante può essere considerato sito naturale e non gestito, mentre

Cesenatico e Cervia sono sottoposte a gestioni diverse, in quanto a Cesenatico è presente un sistema di barriere soffolte mentre a Cervia si è optato per un sistema di barriere rigide emerse.

La scelta dei tre siti è stata effettuata per consentire la stima dei parametri su situazioni ambientali contigue in cui ogni spiaggia potrebbe rappresentare un proxy dell'evoluzione delle altre a fronte di variazioni nelle componenti legate all'idrodinamismo costiero ed al sea level rise.

Sono state prese in considerazione come variabili abiotiche : la mediana granulometrica, la cassazione della granulometria, il contenuto in percentuale di materia organica, il peso della frazione conchigliare residua, la slope della zona intertidale, l'altezza dell'onda frangente, l'altezza della colonna d'acqua al momento del campionamento, la durata del ricoprimento in funzione dei flussi di marea, l'esposizione delle spiagge. Per quanto riguarda le variabili target di volta in volta considerate, sono stati costruiti in totale 6 modelli separati, rispettivamente per: l'abbondanza totale, il numero di taxa riconosciuti ed il valore dell'indice di diversità di Shannon rispettivamente per macrofauna bentonica e meiofauna bentonica.

Domini delle variabili considerate

Per ciascuna variabile è stato stabilito aprioristicamente un dominio di valori costruito a partire dalle osservazioni sui tre siti e, in ragione della necessità di effettuare simulazioni generative su condizioni mai osservate, di conoscenze provenienti da studi bibliografici. La determinazione dei domini parte dall'assunzione secondo la quale eventi estremi climatici, quali sono gli eventi di flooding, potrebbero favorire all'interno del sistema complesso intertidale una cascata di eventi estremi e determinare valori nel target anch'essi estremi e quindi lontani dal range determinabile a partire dalle osservazioni effettuate in situ.

I domini così ottenuti sono stati standardizzati nell'intervallo di valori fra 0 e 120 in modo da far emergere le reali variabilità delle componenti del modello e renderle reciprocamente utilizzabili nella stima dei parametri evitando problemi legati alle diverse unità di misura ed alle differenze dovute alle caratteristiche di variazione proprie di ogni variabile.

Partizione fuzzy dei domini

Nel rendere i valori osservati secondo il formalismo fuzzy è determinante la scelta delle funzioni d'appartenenza. Le funzioni d'appartenenza descrivono l'appartenenza dei valori registrati ai fuzzy set in cui è diviso il dominio d'appartenenza delle diverse variabili. Per ragioni di coerenza è stato deciso di utilizzare per tutte le variabili considerate la stessa partizione fuzzy (5 fuzzy set) nello spazio degli eventi e di standardizzare tutti i domini nel medesimo intervallo [0 – 120]. La partizione fuzzy può essere determinata anche a partire da un algoritmo di fuzzy-clustering che rende la miglior partizione fuzzy in ragione dei dati osservati, nel nostro caso si è deciso di non procedere con l'utilizzo di algoritmi di fuzzy clustering in modo da aumentare le capacità generative dei modelli fuzzy- naive Bayes rendendoli così in grado di simulare anche possibili eventi estremi delle variabili target.

Per quanto riguarda la funzione d'appartenenza è stato deciso di applicarne una delle più semplici descritta da una forma a triangolo isoscele (Fig. 1) con il vertice corrispondente al centro dell'intervallo fuzzy. La forma a triangolo isoscele, oltre ad essere più facile dal punto di vista computazionale, permette di determinare in maniera più oggettiva possibile l'appartenenza del dato ai fuzzy set senza indulgere in interpretazioni più soggettive proprie di altre funzioni d'appartenenza desumibili da sistemi di cui si abbia una conoscenza più completa delle dinamiche che intercorrono fra le variabili.

La funzione a triangolo isoscele descrive l'appartenenza come definita dalla vicinanza del dato al centro dell'intervallo (il vertice del triangolo) in cui l'appartenenza è massima ($\mu = 1$) fino agli estremi dell'intervallo in cui è minima ($\mu = 0$). La partizione fuzzy con le funzioni a triangolo isoscele è costruita in modo che gli intervalli si compenetrino nel punto medio per cui, sul dominio, il punto centrale di ogni intervallo fuzzy, corrispondente al vertice del triangolo isoscele, si va a sovrapporre all'estremo minore dell'intervallo successivo ed all'estremo maggiore del precedente.

Il valore d'appartenenza tramite funzione “a triangolo isoscele” è dato dalle seguenti funzioni:

$$\mu = \frac{x - x_{\min}}{x_c - x_{\min}}$$

$$\mu = \frac{x_{\max} - x}{x_{\max} - x_c}$$

In cui μ è il coefficiente d'appartenenza, x_{\min} e x_{\max} sono i limiti minore e maggiore dell'intervallo considerato e x_c è il centro dell'intervallo corrispondente al vertice del triangolo. I valori d'appartenenza al di fuori dell'intervallo $[0,1]$ sono da considerare pari a 0 in quanto sono generati da valori esterni all'intervallo per i quali il coefficiente d'appartenenza assume quindi valore 0.

Impostazione del modello generativo

In considerazione delle variabili di flooding e del loro diverso “timing” di effetto su organismi recettori sensibili, come gli organismi bentonici, è stato necessario effettuare la modellizzazione in due fasi successive (Fig. 9).

Una prima fase in cui si è tenuto conto della forzante “sea level rise” e del suo effetto continuo nel tempo ed indipendente dal verificarsi o meno degli eventi di flooding che compongono gli scenari.

Tenendo quindi conto dell'effetto continuo dell'innalzamento medio marino sui popolamenti bentonici è stata effettuata una prima simulazione in cui, a partire dai valori corrispondenti ad ogni scenario, il sea level rise è stato “spalmato” sull'intero periodo 2010 – 2080 e secondo un intervallo temporale annuale sono stati simulati i popolamenti in funzione solamente dell'aumento del livello medio marino. La fase di apprendimento delle simulazioni in funzione del sea level rise è stata integrata ad ogni “run” del modello coi risultati ottenuti dalla simulazione precedente in modo da considerare in fase di apprendimento la successione temporale del sea level rise.

In questo modo sono stati stimati i popolamenti bentonici su cui andranno poi ad incidere gli eventi di flooding contenuti negli scenari.

Gli eventi di flooding, a differenza del “sea level rise”, sono eventi a grande intensità fisica ma a limitata durata temporale per cui nel continuum temporale possono essere considerati eventi puntiformi ad effetto immediato, che possono anche avere un effetto

cumulativo, come si vedrà nei prossimi capitoli, solo in presenza di una successione abbastanza ravvicinata di eventi.

Le variabili proprie degli eventi previsti, quali altezza dell'onda frangente, intensità e durata del flooding sono state utilizzate per simulare le risposte della fauna bentonica intertidale ai singoli eventi di flooding in cui le fasi di apprendimento sono state effettuate considerando come training set i popolamenti stimati nella simulazione secondo "sea level rise" corrispondenti agli anni degli scenari (2010, 2020, 2050, 2080). Fra le variabili presenti all'interno delle istanze alcune sono state modificate secondo le forzanti di flooding e sea level rise mentre altre sono state mantenute costanti per caratterizzare i siti in modo da stimare parametri descrittivi della sito-specificità dell'habitat intertidale. In particolare le variabili che risentono numericamente delle forzanti sono : altezza dell'onda frangente, altezza della colonna d'acqua e durata del ricoprimento, mentre le altre variabili abiotiche sono state mantenute costanti. E' opportuno a questo punto spiegare come il mantenimento costante dei valori di alcune variabili come TOM, mediana granulometrica, classazione granulometrica, capulerio, introduca inevitabilmente un errore nei modelli. Data la loro posizione intermedia nella struttura gerarchica di dipendenze nel sistema, si vedrà nell'ultimo capitolo di questa tesi come esse verranno trattate, attraverso l'implementazione dell'approccio, al fine di ottenere modelli annidati uno dentro l'altro secondo la struttura di dipendenze dei sistemi ambientali presi in esame. Altre variabili come l'esposizione, stimata a partire da considerazioni di tipo geografico, possono invece essere mantenute costanti senza il rischio di inserire errori nelle simulazioni.



Figura 9: Schema concettuale di funzionamento del modello ibrido FBEM costruito per Cesenatico.

La variabile target

Come già spiegato all'interno dell'habitat recettore il descrittore è rappresentato dalla fauna bentonica considerata nelle due componenti della macrofauna e della meiofauna. Per entrambe le componenti sono stati considerate come variabili target i principali parametri di struttura delle comunità: abbondanze totali, numero di taxa e diversità calcolata secondo l'indice di Shannon.

Mentre per le prime due, i valori delle variabili target nelle simulazioni sono stati trattati come dipendenti direttamente dalla struttura delle variabili abiotiche del sistema. La modellizzazione della terza variabile target ha richiesto un aumento della dimensionalità della struttura del sistema. In particolare, essendo la diversità dipendente dai valori delle altre due variabili target e dalle capacità di autoorganizzazione proprie delle componenti biologiche, è stato necessario addentrarsi ulteriormente all'interno della struttura complessa inserendo come valori del terzo target anche abbondanze totali e numero di taxa, aumentando così il numero di variabili e di conseguenza il numero di parametri stimati.

3.7 Validazione dei modelli generativi (Figg. 10 – 15)

L'accuratezza e la validità dei modelli generativi costruiti sono state testate attraverso il calcolo del "G-value", un test di fitting che mette a confronto valore osservato e valore predetto in relazione alla variabilità osservata nei dati di apprendimento. I data set totali sono stati suddivisi casualmente in 2 parti equivalenti una delle quali è stata utilizzata per stimare i parametri dei modelli e l'altra per il test tramite calcolo del G-value.

Il valore del G-value è dato dalla seguente relazione :

$$G = 1 - \frac{\sum_{i=1}^n |Z_{(xi)} - \bar{Z}|^2}{\sum_{i=1}^n |Z_{(xi)} - Z_{(xi)}|^2}$$

dove $Z_{(x_i)}$ è il valore osservato nell' i -esimo campione, $Z_{(X_i)}$ è il valore stimato e \bar{Z} è la media dei valori osservati. Il valore del G-value può variare da 0 per una simulazione completamente sbagliata a 1 per una simulazione perfettamente accurata (Schloeder et al., 2001).

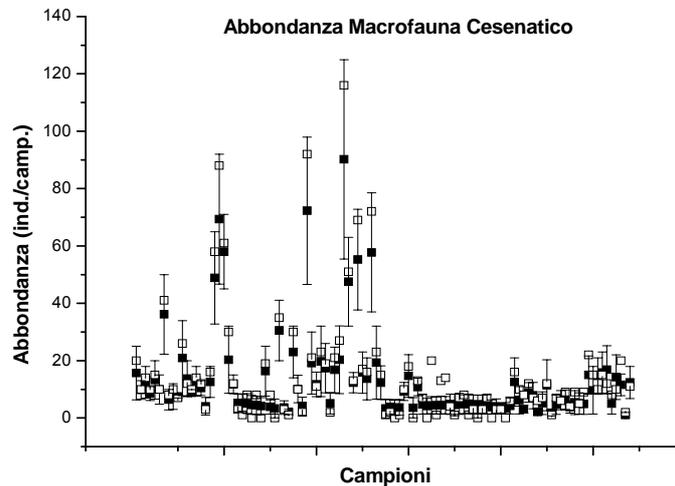


Figura 10 Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target abbondanza di macrofauna per Cesenatico. Le barre verticali indicano il range d'incertezza stimato dal modello

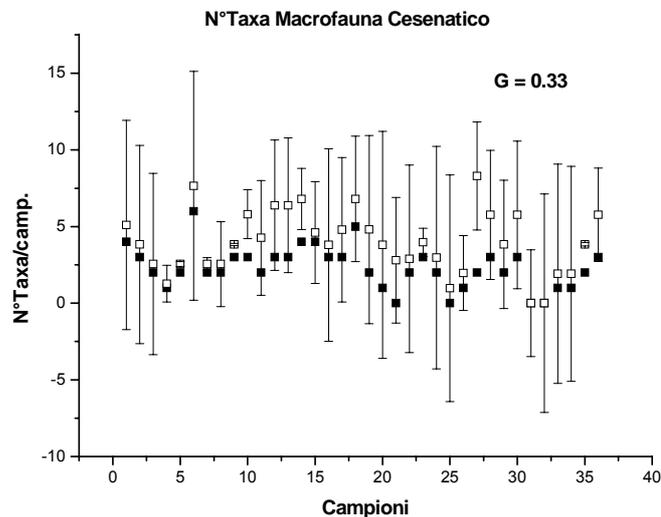


Figura 11 Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target n° di taxa di macrofauna per Cesenatico. Le barre verticali indicano il range d'incertezza stimato dal modello

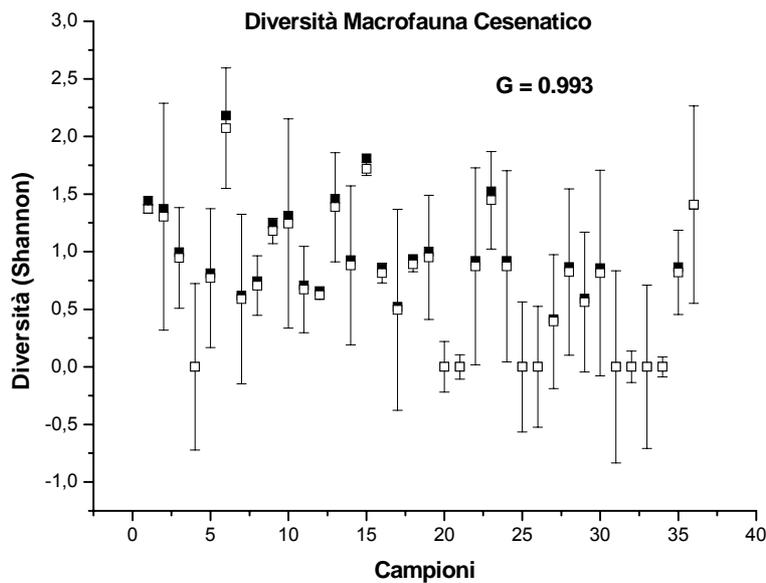


Figura 12 Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target diversità di macrofauna per Cesenatico. Le barre verticali indicano il range d'incertezza stimato dal modello

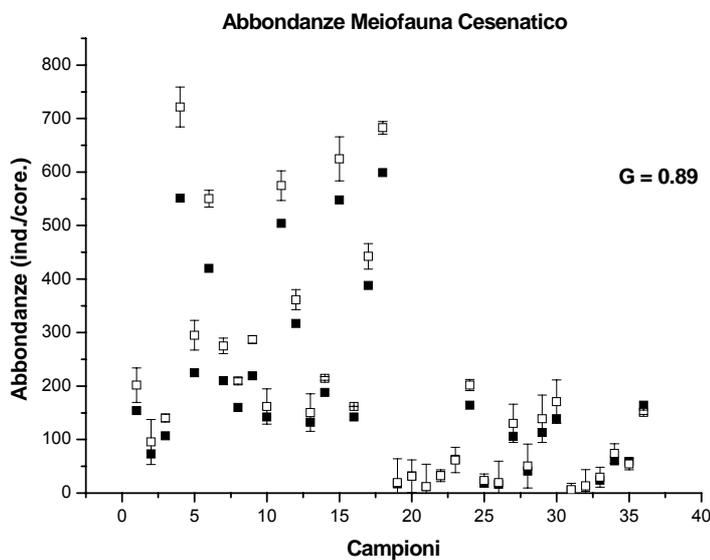


Figura 13 Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target abbondanza di meiofauna per Cesenatico. Le barre verticali indicano il range d'incertezza stimato dal modello

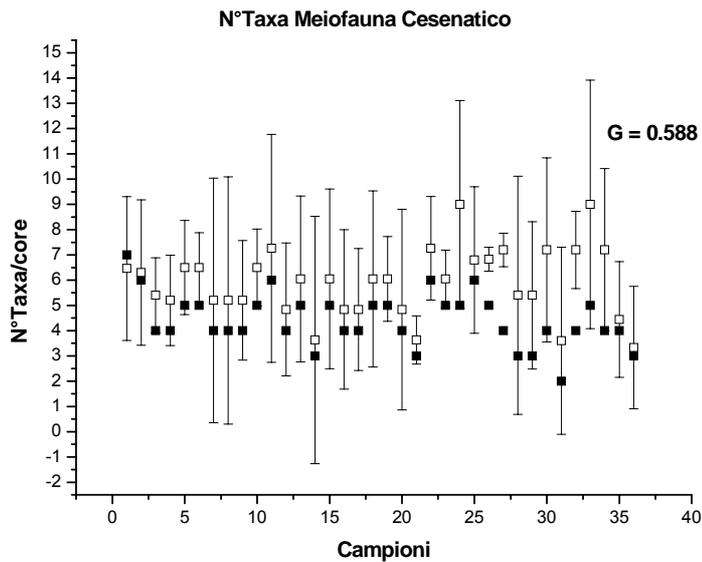


Figura 14 Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target n° di taxa di meiofauna per Cesenatico. Le barre verticali indicano il range d'incertezza stimato dal modello

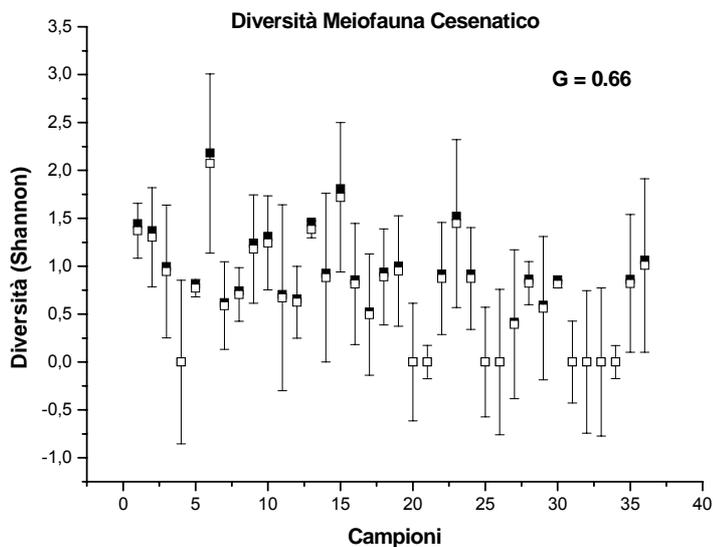


Figura 15 Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target diversità della meiofauna per Cesenatico. Le barre verticali indicano il range d'incertezza stimato dal modello

Dai risultati delle validazioni si nota come l'accuratezza del modello sia migliore in situazioni in cui il pattern da simulare presenti un'alta variabilità. E' importante inoltre notare anche come, anche in casi di non buon fitting, il valore osservato tranne in rari casi rientri sempre all'interno del range d'incertezza stimato dal modello.

3.8 Risultati: Cesenatico

3.8.1 Macrofauna (Figg. 16 a-c)

I risultati delle simulazioni di flooding riferite alla macrofauna mostrano andamenti diversi per i 3 parametri di struttura di comunità.

L'abbondanza segue andamenti analoghi nei primi 3 scenari, in particolare emerge un trend discendente nei valori medi di abbondanza stimata. Il quarto scenario (innalzamento LMM: 0,22 m) segue l'andamento dei primi tre fino al tempo di ritorno 50 a. per poi invertire decisamente la tendenza al tempo di ritorno 100 a. in corrispondenza del quale il modello individua un probabile valore soglia corrispondente ad un deciso aumento dei valori medi di abbondanza che si assestano attorno a 37,9 ind./camp. in luogo di un valore iniziale, per lo scenario 4 a., pari a 16,07 ind./camp.

Il numero di taxa rimane pressoché costante variando nei primi 3 scenari da un valore medio di 2,46 n°/camp. ad un valore uguale a 3,37 n°/camp. mostrando anch'esso una soglia in corrispondenza del tempo di ritorno 100 a. del quarto scenario, in cui il numero di taxa "crolla" a 1 n°/camp. con un range d'incertezza che va da 0 a 3,45. n°/camp.. L'andamento del numero di taxa sembra condizionare anche la diversità di Shannon il cui valore costante per i primi 3 scenari e per i primi 7 tempi di ritorno del quarto scenario diminuisce nettamente in corrispondenza della simulazione relativa al tempo di ritorno 100 a. dell'ultimo scenario considerato.

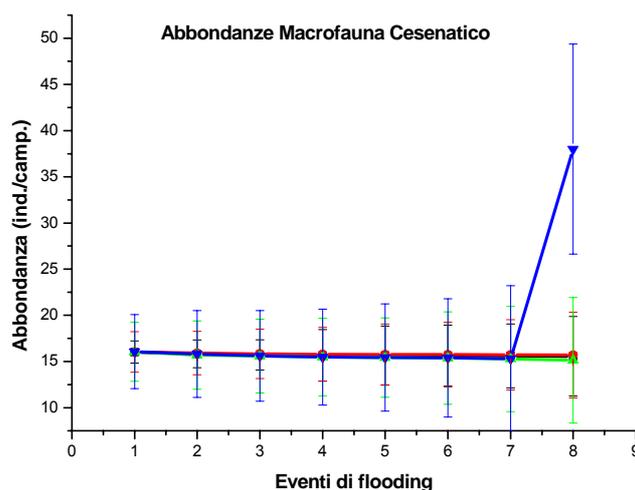


Figura 16a: Risultati delle simulazioni per il target dell'abbondanza della macrofauna a Cesenatico (NERO=scenario 1, ROSSO=scenario 2, VERDE=scenario 3, BLU = scenario 4)

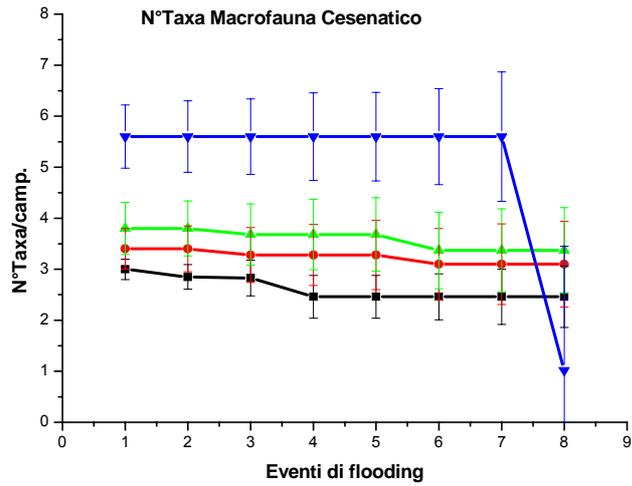


Figura 16b: Figura 15a: Risultati delle simulazioni del target n° di taxa di macrofauna a Cesenatico (NERO=scenari01, ROSSO=scenari02, VERDE=scenari03, BLU=scenari04)

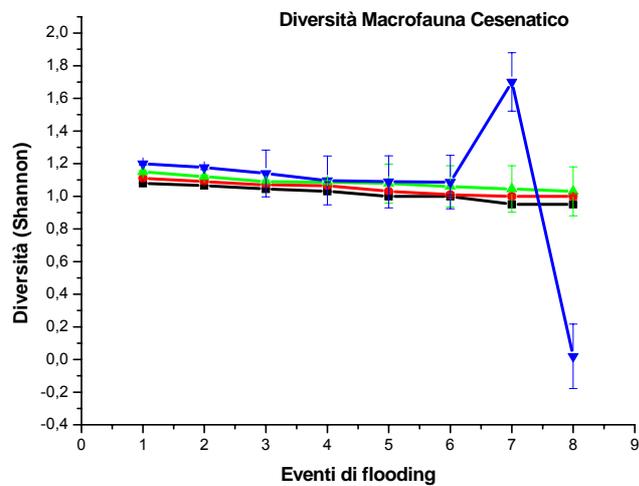


Figura 16c: Risultati delle simulazioni per il target della diversità della macrofauna a Cesenatico (NERO=scenari01, ROSSO=scenari02, VERDE=scenari03, BLU=scenari04)

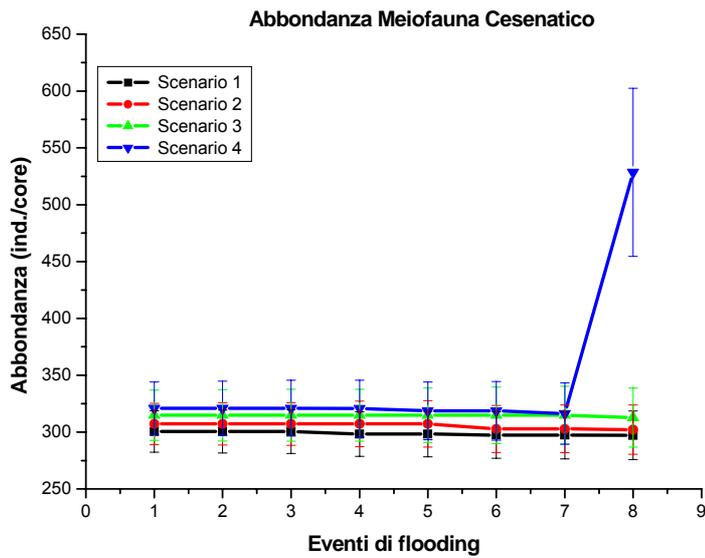


Figura 17a: Risultati delle simulazioni per il target abbondanza di meiofauna a Cesenatico (NERO=scenario1, ROSSO=scenario2, VERDE=scenario3, BLU=scenario4)

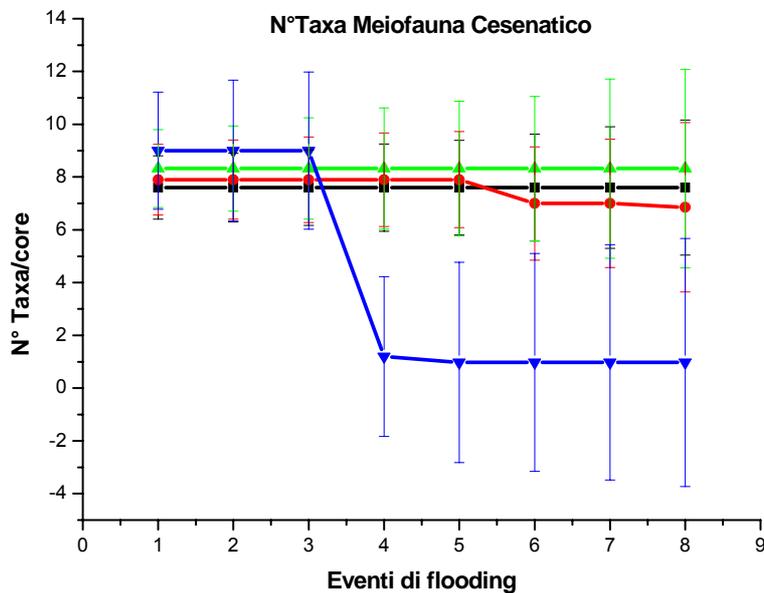


Figura 17b: Risultati delle simulazioni per il target n° di taxa di meiofauna a Cesenatico (NERO=scenario1, ROSSO=scenario2, VERDE=scenario3, BLU=scenario4)

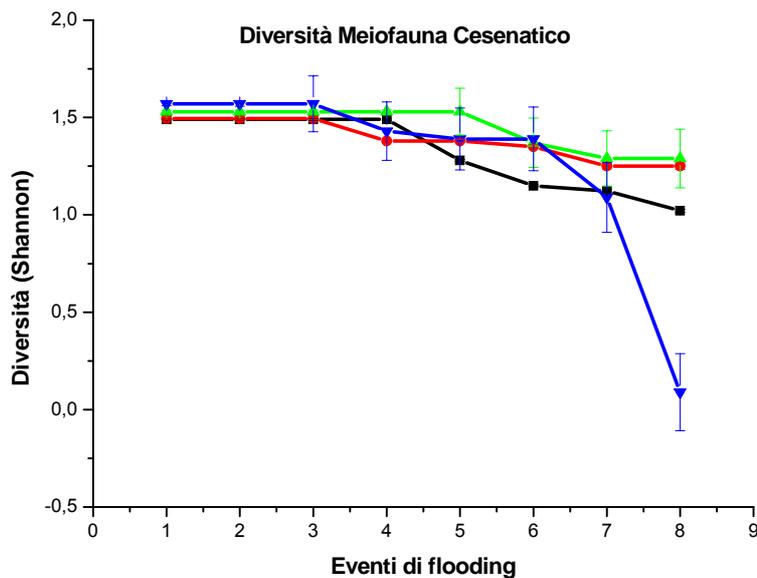


Figura 17b: Risultati delle simulazioni per il target diversità della meiofauna a Cesenatico (NERO=scenario1, ROSSO=scenario2, VERDE=scenario3, BLU=scenario4)

3.8.2 Meiofauna (Figg. 17 a-c)

Dalle simulazioni relative alla meiofauna risultano andamenti, in generale, assimilabili a quelli stimati per la macrofauna con alcune differenze nella distribuzione delle variazioni. In particolare, sia per quanto riguarda l'abbondanza media sia per quanto riguarda il numero di taxa, le differenze, nei valori medi dei range d'incertezza in output, emergono non in corrispondenza di ogni evento di flooding ma in maniera discontinua (e senza apparente linearità) dopo un certo numero di eventi, rispettando comunque il trend discendente già osservato per le abbondanze della macrofauna. Il numero di taxa risponde agli scenari di flooding in maniera speculare al numero di taxa della macrofauna così come la diversità, che dopo essersi sostanzialmente mantenuta intorno ai valori iniziali per ogni scenario, incontra una soglia, al tempo di ritorno 100 a., passando da un valore di 1,08 ad un valore di 0,34.

3.9 BELLOCCHIO: applicazione FBEM

Per il sito di Bellocchio, comprendente il tratto di spiaggia antistante l'abitato di Lido di Spina fino, spostandosi verso sud, al tratto antistante la laguna costiera propriamente detta laguna di Bellocchio, i modelli sono stati costruiti a partire da un tipo di campionamento diverso da quello utilizzato per stimare i parametri del modello di Cesenatico. Infatti è stato campionato solo il sito di interesse nelle due macrozone caratterizzate da situazioni, gestioni ed evoluzione del profilo di spiaggia profondamente differenti. Mentre il tratto di fronte all'abitato presenta una gestione fortemente impattante, in quanto sottoposto a continui ripascimenti negli anni, il sito a sud, di fronte alla laguna costiera di Bellocchio presenta una situazione "in divenire" su cui è opportuno spendere qualche riga di spiegazione. A sud di Lido di Spina ci si trova di fronte ad una situazione ibrida ancora lontana da una precisa definizione. L'osservazione dell'evoluzione temporale del sito mostra un tratto di costa sottoposto ad una forte spinta erosiva che ha fatto arretrare la linea di costa di un centinaio di metri negli ultimi 50 anni. La situazione si presenta "ibrida" e non ben definita perché in passato il mare è già penetrato nella laguna costiera dando luogo ad una zona intertidale in cui il limite dell'antica laguna è diventato l'attuale linea di costa. L'elevata eterogeneità spaziale, indotta dal passaggio da un habitat tipicamente lagunare ad un habitat più marino, ha favorito la formazione di una spiaggia, antistante la laguna attuale, che in alcuni punti sembra addirittura aver invertito il trend erosivo. Per quanto riguarda i popolamenti bentonici, si è osservata una maggiore complessità strutturale delle comunità probabilmente correlata con l'accentuata eterogeneità spaziale osservata, di cui si è dovuto tener conto al momento della modellizzazione.

Variabili ambientali

Per la ragione suddetta, oltre alle variabili considerate per il sito di Cesenatico sono state registrate anche informazioni circa la presenza di residui, ormai morenti, di piante di origine lagunare e circa le formazioni, favorite dalla presenza dei vegetali, di letti di mitili. E' stato necessario aumentare il numero delle variabili in quanto le due variabili

aggiuntive, aumentando l'eterogeneità spaziale dell'habitat aggiungevano gradi di libertà nella determinazione delle caratteristiche ambientali che influenzano i popolamenti descrittivi di stato. Nel contesto del modello, i rilevamenti aggiuntivi hanno dato origine a due nuove variabili rispettivamente “peso delle piante” e “presenza/assenza di mitili” (a seconda dei transetti di campionamento). Purtroppo non è stato possibile ottenere misure migliori dei due nuovi attributi in quanto la loro presenza è stata rilevata in corso d'opera per cui le due nuove variabili sono da considerare dei proxy e non delle misure dirette dell'eterogeneità dell'habitat.

Partizione fuzzy dei domini

Come per il sito di Cesenatico, è stata applicata la funzione d'appartenenza “a triangolo isoscele” secondo però una partizione fuzzy a 8 intervalli fuzzy in modo da riflettere meglio l'elevata variabilità delle variabili abiotiche e biotiche riscontrata durante l'attività di campionamento. I domini sono stati anche in questo caso standardizzati in un intervallo [0, 120].

Variabili ambientali

Come già accennato, oltre alle variabili considerate per il sito di Cesenatico sono state registrate anche informazioni circa la presenza di residui vegetali di origine lagunare e quella di letti di mitili. Nel contesto del modello, i rilevamenti aggiuntivi sono rappresentati da due nuove variabili, rispettivamente: “peso delle piante” e “presenza di mitili”.

Impostazione del FBEM

Le variabili proxy della complessità spaziale avrebbero dovuto creare all'interno del modello un canale per l'informazione che avrebbe dovuto permettere al modello di discriminare fra situazioni a differente livello di complessità.

Purtroppo a causa della scarsa replicazione di tali situazioni il risultato è stato quello di aumentare il rumore delle distribuzioni “confondendo” il modello ed uniformando il calcolo dei parametri su valori equiprobabili per ogni fuzzy set di ogni variabile.

Si è deciso quindi di intervenire direttamente sul dataset di apprendimento oltre che sull'architettura del modello (Fig. 18). Il data set derivante dal campionamento è stato inizialmente spezzato in due sub-data set uno contenente le osservazioni provenienti dal tratto di costa antistante Lido di Spina e l'altro proveniente dal tratto di costa antistante la laguna di Bellocchio. Il data set ottenuto dall'intertidale adiacente alla laguna è stato ulteriormente scomposto utilizzando come criterio di divisione la mediana granulometrica nel valore di 4.5ϕ . In questo modo sono state ottenute 3 matrici di parametri delle stesse dimensioni ognuna specifica per una delle tre differenti combinazioni di variabili ambientali rilevate durante il campionamento. Per permettere al modello di effettuare una scelta rispetto a quale matrice di parametri utilizzare durante la modellizzazione è stato necessario effettuare un altro ciclo di apprendimento in cui la variabile target è la matrice da utilizzare per modellizzare la nuova istanza. In questo primo passaggio il modello fuzzy naive Bayes è stato utilizzato come un vero e proprio classificatore su 3 classi discrete in cui le classi indicano la matrice di parametri su cui costruire il modello generativo finale.

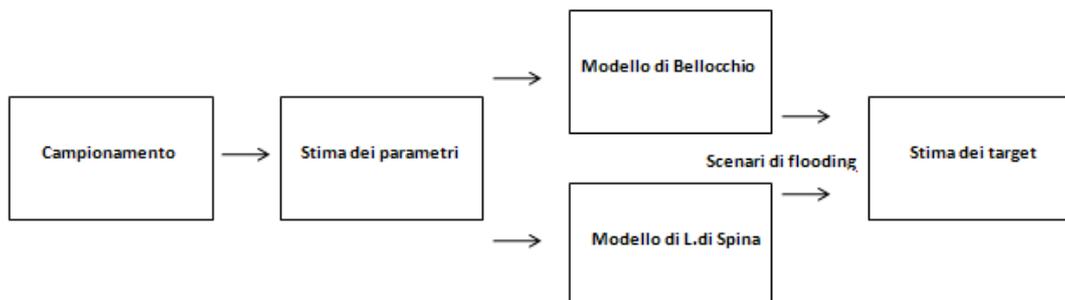


Figura 18: Schema concettuale di funzionamento per il modello di Bellocchio

3.10 Validazione dei modelli generativi

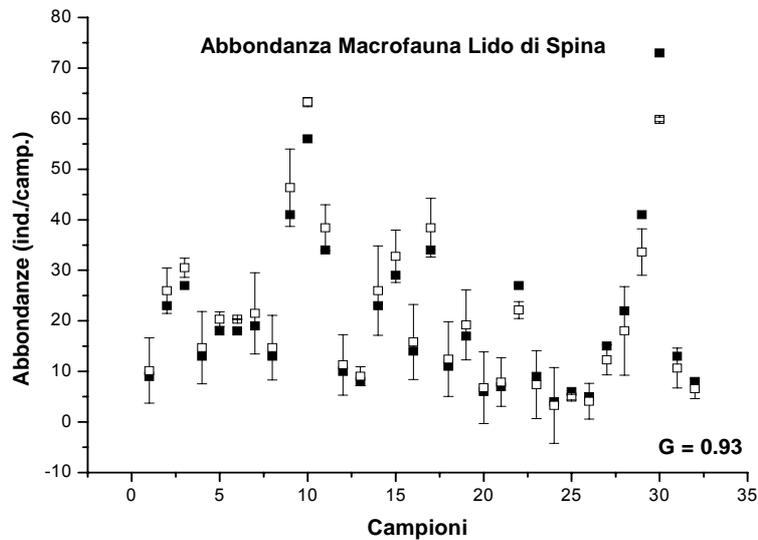


Figura 19 : Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target abbondanza di macrofauna per Lido di Spina. Le barre verticali indicano il range d'incertezza stimato dal modello

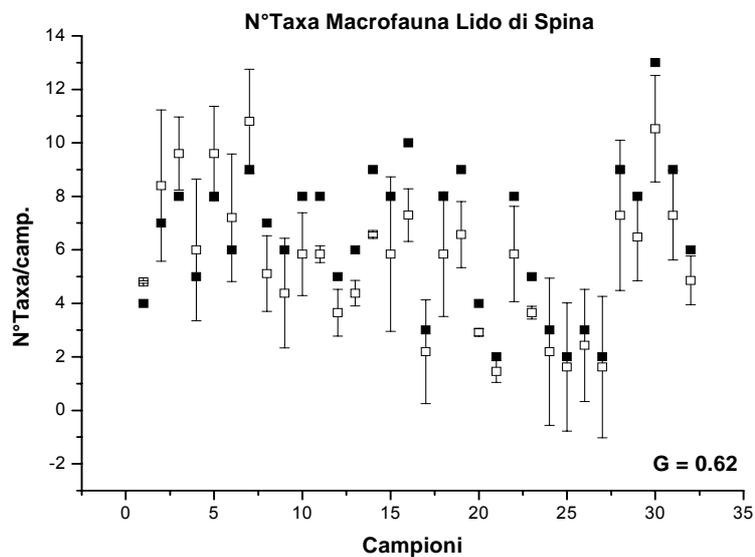


Figura 20 : Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target n° di taxa di macrofauna per Lido di Spina. Le barre verticali indicano il range d'incertezza stimato dal modello

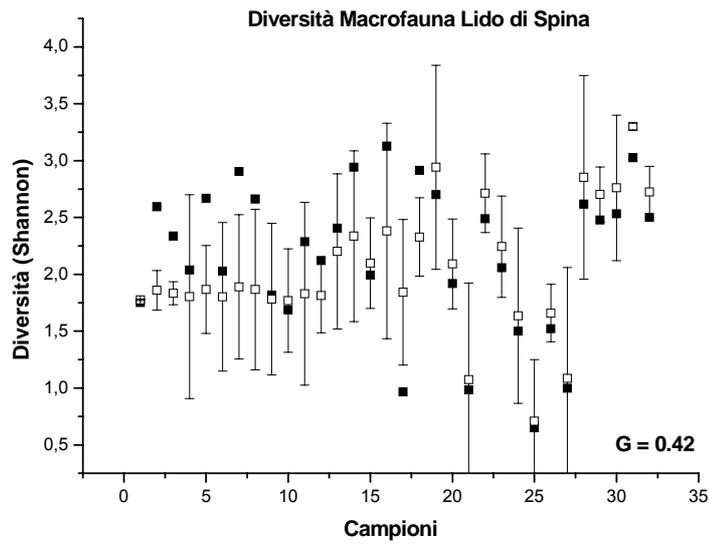


Figura 21 Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target diversità della macrofauna per Lido di Spina. Le barre verticali indicano il range d'incertezza stimato dal modello

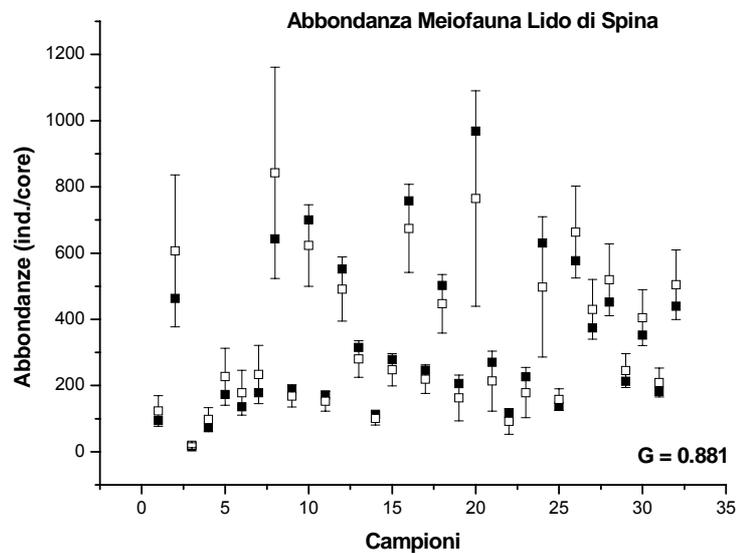


Figura 22 Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target abbondanza di meiofauna per Lido di Spina. Le barre verticali indicano il range d'incertezza stimato dal modello

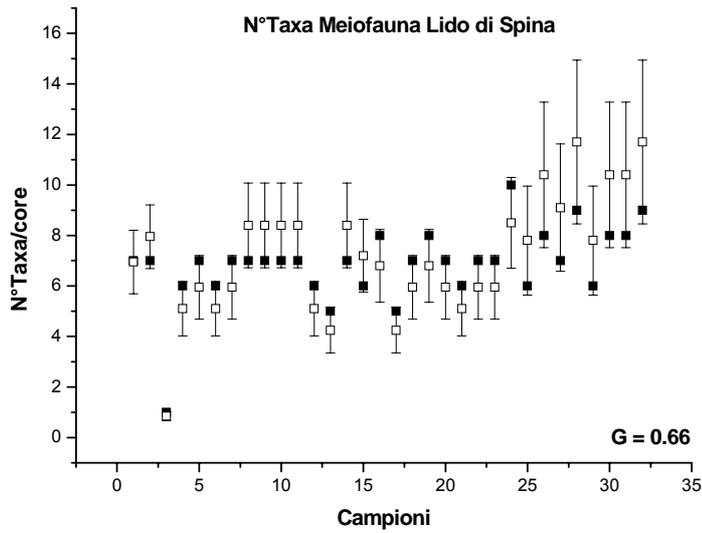


Figura 23 : Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target n° taxa di meiofauna per Lido di Spina. Le barre verticali indicano il range d'incertezza stimato dal modello

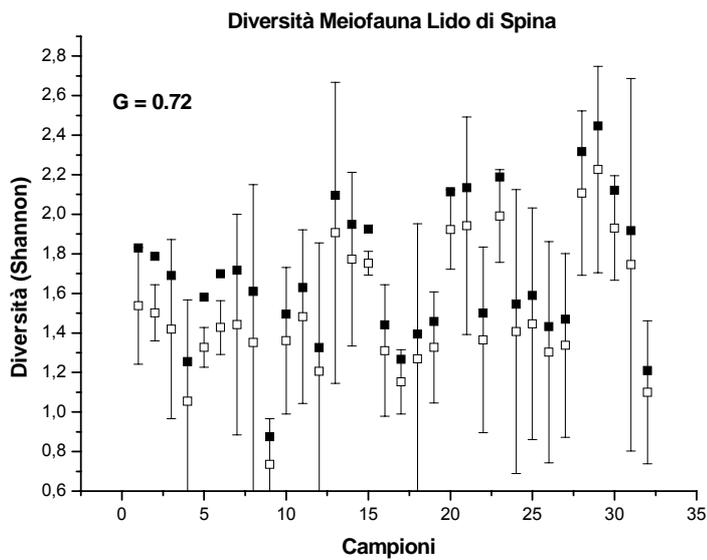


Figura 24 : Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target diversità della meiofauna per Lido di Spina. Le barre verticali indicano il range d'incertezza stimato dal modello

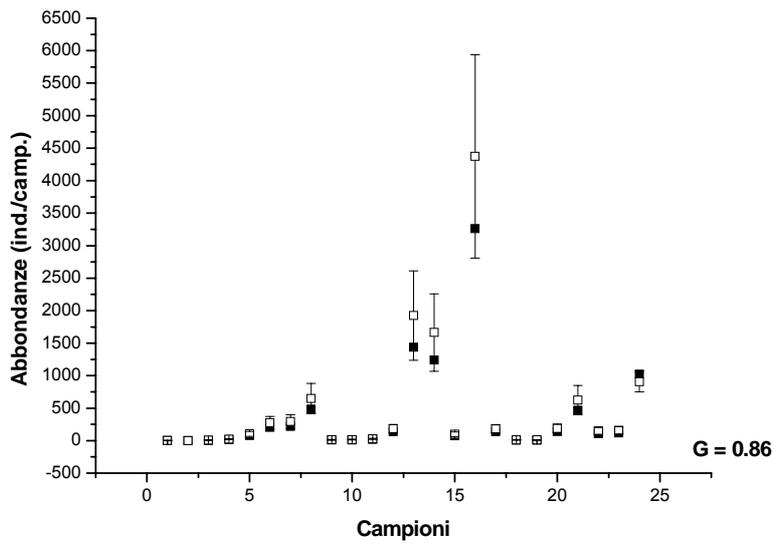


Figura 25 : Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target abbondanza di macrofauna a Bellocchio. Le barre verticali indicano il range d'incertezza stimato dal modello

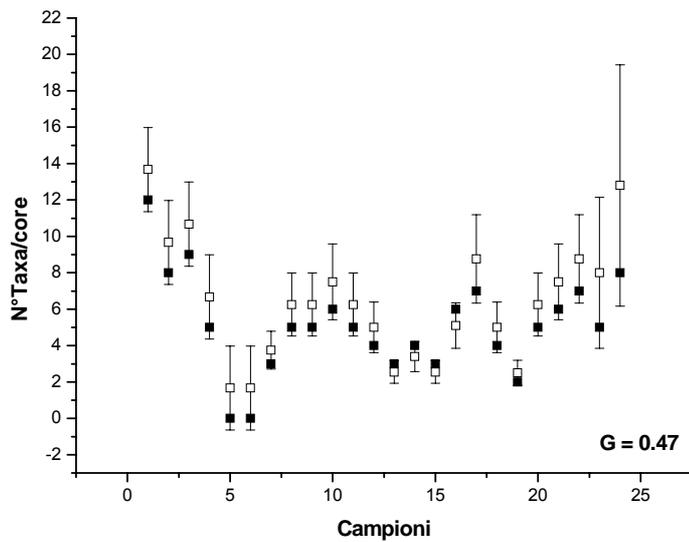


Figura 26 : Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target n° di taxa di macrofauna a Bellocchio. Le barre verticali indicano il range d'incertezza stimato dal modello

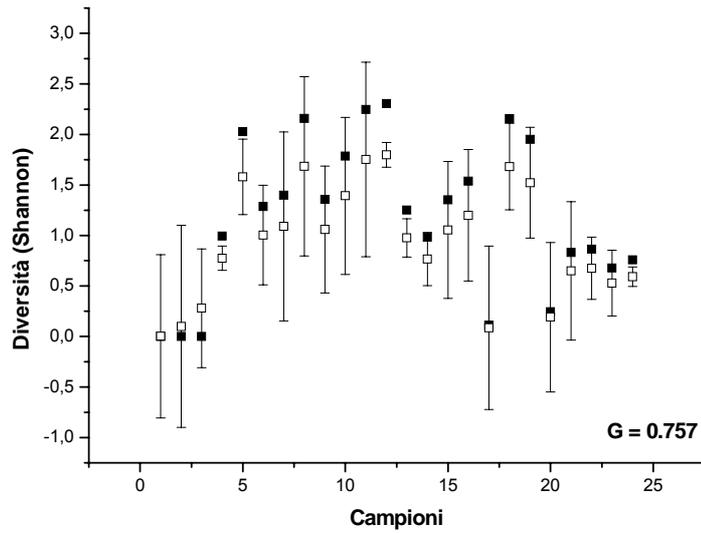


Figura 27 : Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target diversità della macrofauna a Bellocchio. Le barre verticali indicano il range d'incertezza stimato dal modello

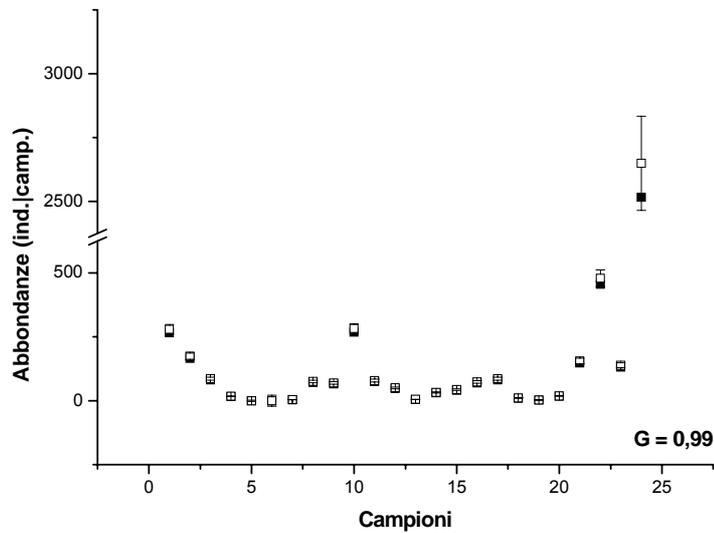


Figura 28 : Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target abbondanza di meiofauna a Bellocchio. Le barre verticali indicano il range d'incertezza stimato dal modello

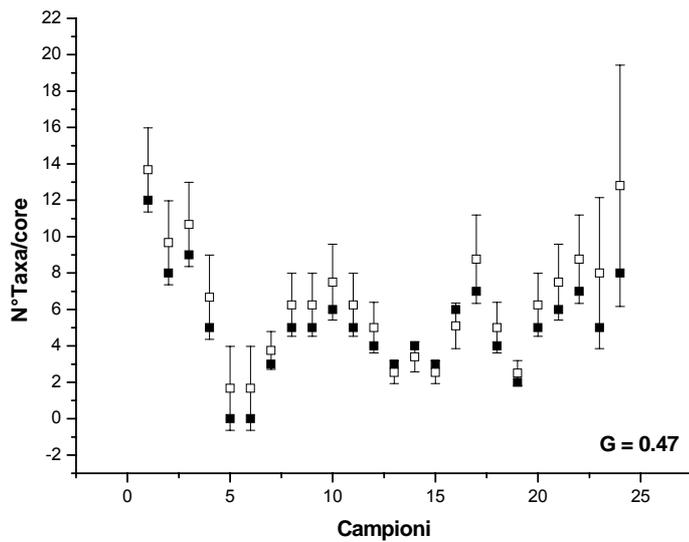


Figura 29 : Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target n° di taxa di meiofauna a Bellocchio. Le barre verticali indicano il range d'incertezza stimato dal modello

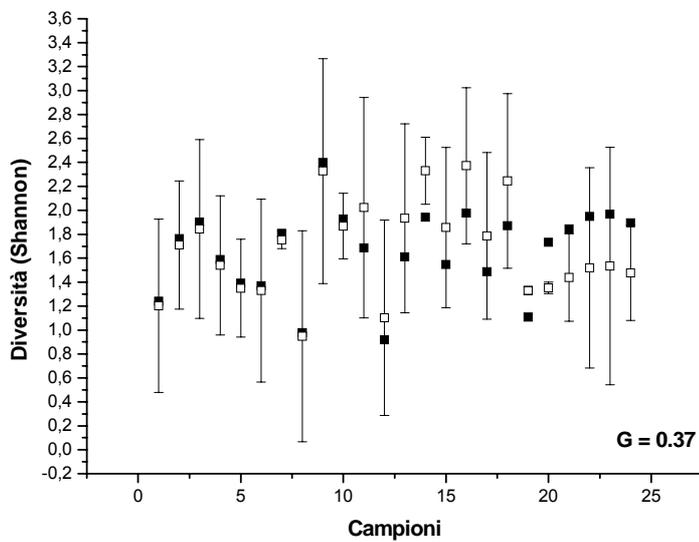


Figura 30 : Valori osservati (quadrati neri) e valori stimati (quadrati bianchi) del target diversità della meiofauna a Bellocchio. Le barre verticali indicano il range d'incertezza stimato dal modello

3.11 Risultati: Lido di Spina

3.11.1 Macrofauna (Figg. 31 a-c)

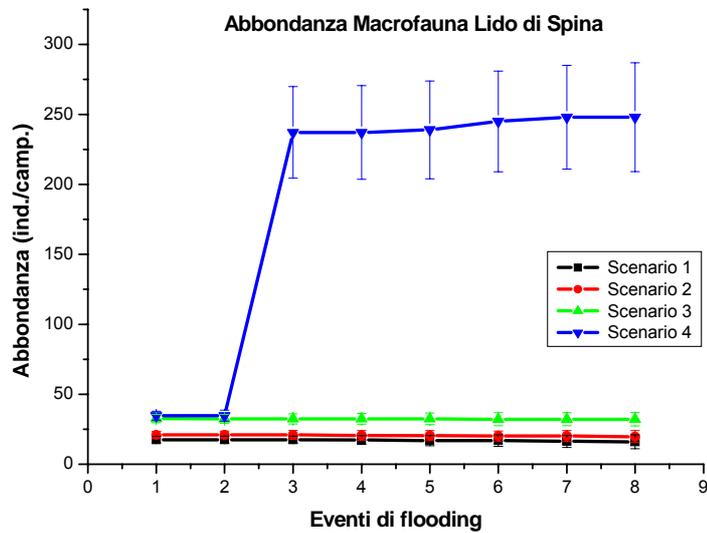


Figura 31a : Risultati delle simulazioni per il target abbondanza di macrofauna a Lido di Spina (NERO=scenario1, ROSSO=scenario2, VERDE=scenario3, BLU=scenario4)

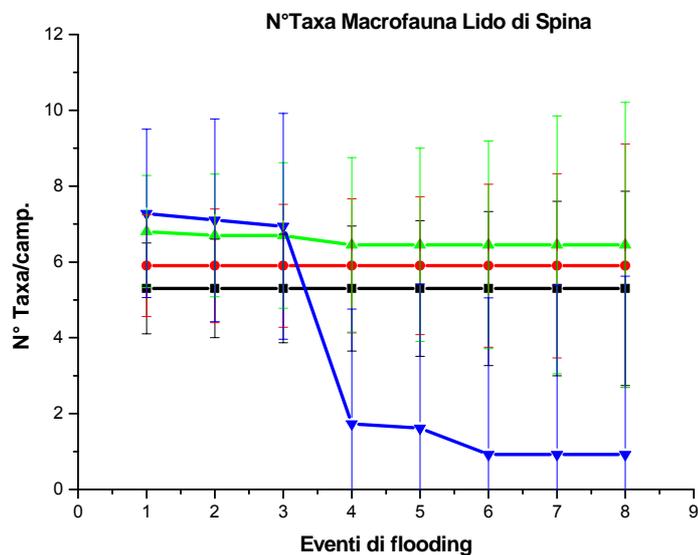


Figura 31b : Risultati delle simulazioni per il target n°di taxa di macrofauna a Lido di Spina (NERO=scenario1, ROSSO=scenario2, VERDE=scenario3, BLU=scenario4)

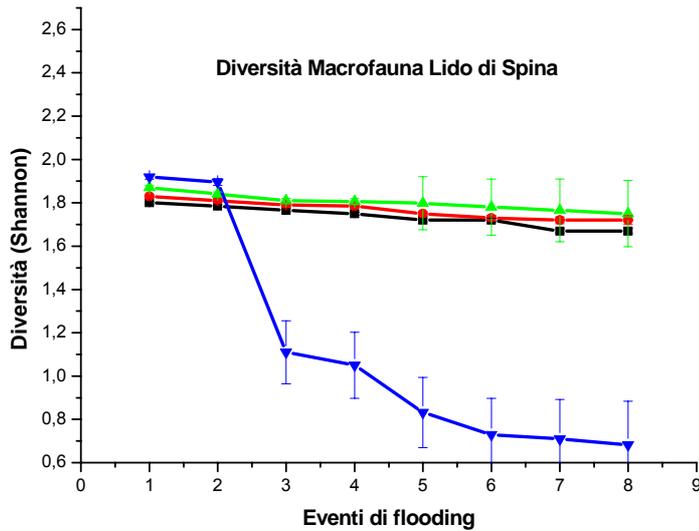


Figura 31c : Risultati delle simulazioni per il target diversità della macrofauna a Lido di Spina (NERO=scenario1, ROSSO=scenario2, VERDE=scenario3, BLU=scenario4)

Il modello generativo parametrizzato sui dati provenienti da Lido di Spina mostra, secondo l'intensità degli eventi sopracitati, per l'abbondanza della macrofauna un trend leggermente crescente nei primi 3 scenari in cui si passa rispettivamente da valori medi pari a 16,1 ind./camp. fino a 17,6 ind.camp. nel primo scenario, da 19,78 ind./camp. a 21 ind./camp. nel secondo e da valori pressochè 32,2 a 32,64 nel terzo scenario. La simulazione relativa al quarto ed ultimo scenario mostra una soglia di non linearità in corrispondenza del tempo di ritorno 10 in cui i valori medi di abbondanza macrofaunale passano da 34,74 a 237,12 ind.camp. per poi variare leggermente fino a 250 in corrispondenza del tempo di ritorno 100.

Riguardo al numero di taxa, dalle simulazioni si osserva come questo viene mantenuto all'interno del range [5,3 – 6,75 n°/camp.] nei primi 3 scenari per poi incontrare un valore soglia, al tempo di ritorno 20 a., del quarto scenario in cui da valori pari a 6,94n°/camp. si passa a valori decisamente più bassi pari a 2,73 n°/camp.fino a 0,9n°/core al tempo di ritorno 50 a..

La diversità di Shannon non sembra risentire particolarmente delle forzanti di flooding fino al quarto scenario in cui, in corrispondenza del tempo di ritorno 20 a., i valori di diversità mostrano un decremento fino ad un valore pari a 1,003.

3.11.2 Meiofauna (Figg. 32 a-c)

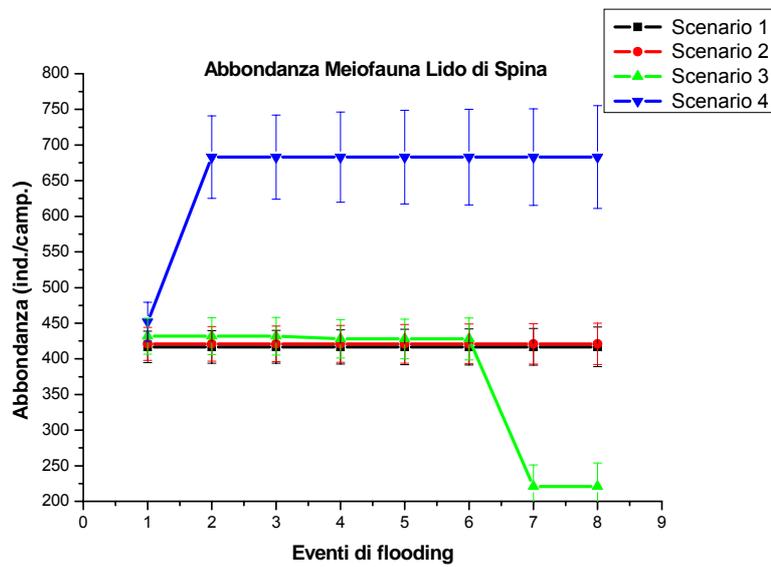


Figura 32a : Risultati delle simulazioni per il abbondanza di meiofauna a Lido di Spina (NERO=scenario1, ROSSO=scenario2, VERDE=scenario3, BLU=scenario4)

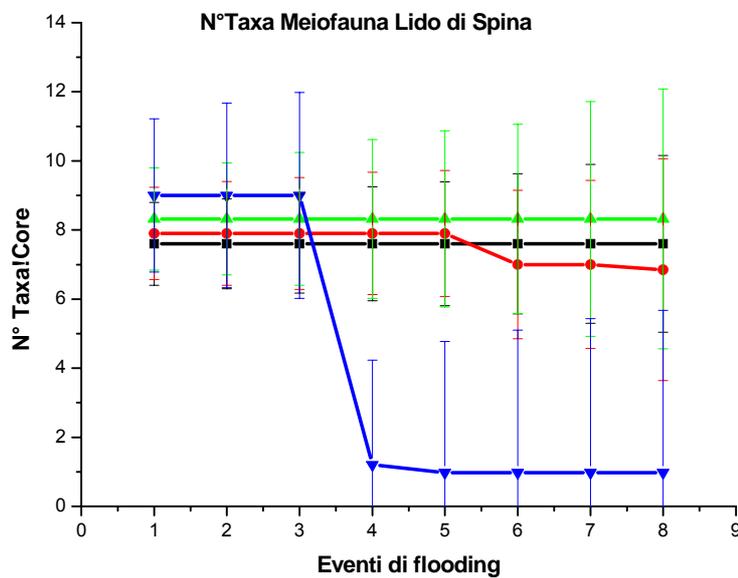


Figura 32b : Risultati delle simulazioni per il target n° di taxa di meiofauna a Lido di Spina (NERO=scenario1, ROSSO=scenario2, VERDE=scenario3, BLU=scenario4)

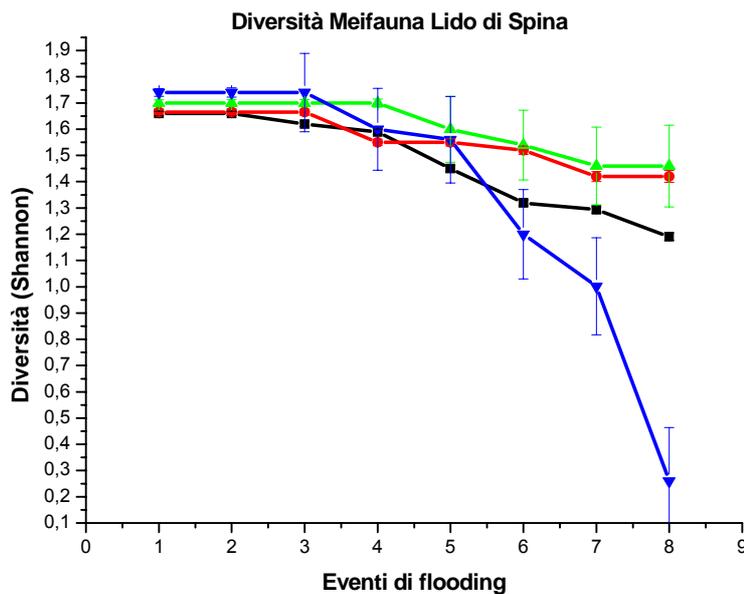


Figura 32c : Risultati delle simulazioni per il target diversità della meiofauna a Lido di Spina (NERO=scenario1, ROSSO=scenario2, VERDE=scenario3, BLU=scenario4)

Le abbondanze di meiofauna per Lido di Spina incontrano il primo valore soglia, al tempo di ritorno 50 a. del terzo scenario, in cui le abbondanze passano da valori medi di 398 ind./core a 190 ind./core. I trend si inverte nel quarto scenario in cui, a partire dal tempo di ritorno 5 a., i valori di abbondanza si assestano attorno a 683 ind./core.

Il numero di taxa segue il trend già osservato per la macrofauna in cui, dopo lievi variazioni all'interno del range [7,6 – 8,32 n°/core] nei primi 3 scenari simulati, viene rilevata una soglia nel quarto scenario, in corrispondenza del tempo di ritorno 5 a., in cui il numero di taxa si assesta all'interno del range [0,973 – 1,2 n°/core] che viene mantenuto fino all'ultimo evento dello scenario.

La diversità anche in questo caso varia leggermente con un trend crescente fino alla soglia al livello della quale il numero di taxa cala rapidamente. In corrispondenza della stessa soglia anche la diversità scende fino a valori simulati pari a 0,99.

3.12 Risultati: Bellocchio

3.12.1 Macrofauna (Figg. 33 a-c)

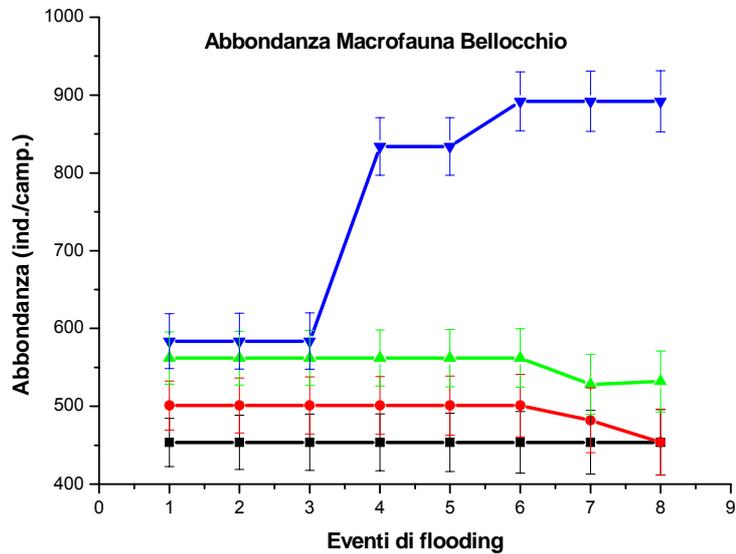


Figura 33a : Risultati delle simulazioni per il target abbondanza di macrofauna a Bellocchio (NERO=scenario1, ROSSO=scenario2, VERDE=scenario3, BLU=scenario4)

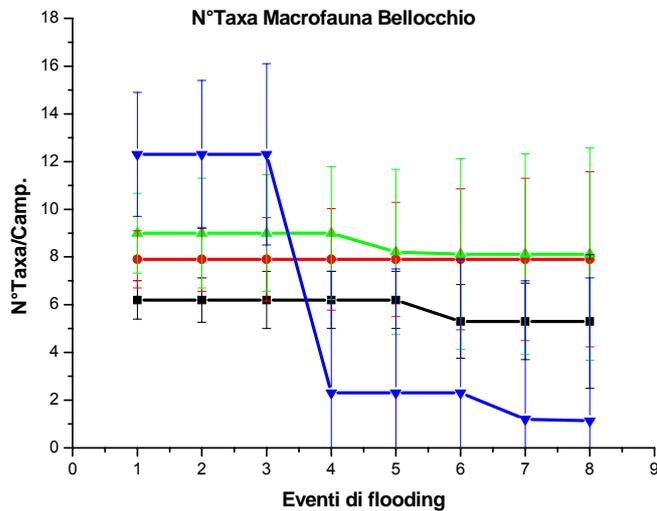


Figura 33b : Risultati delle simulazioni per il target n°di taxa di macrofauna a Bellocchio (NERO=scenario1, ROSSO=scenario2, VERDE=scenario3, BLU=scenario4)

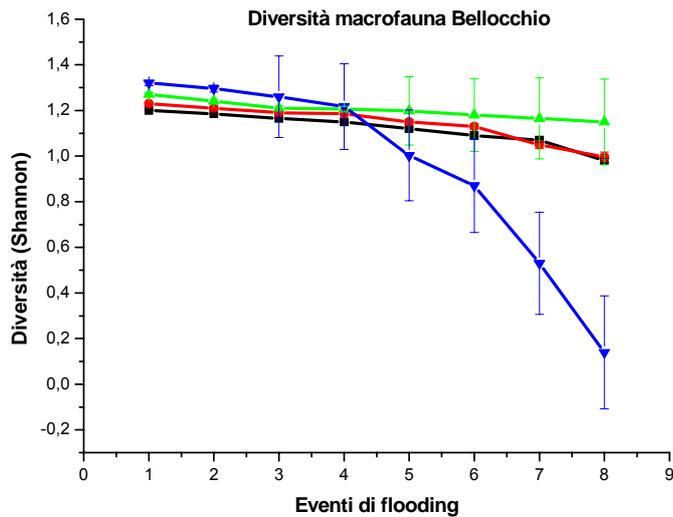


Figura 33 Risultati delle simulazioni per il target diversità della macrofauna a Bellocchio (NERO=scenario1, ROSSO=scenario2, VERDE=scenario3, BLU=scenario4)

Pur facendo rilevare valori di molto superiori rispetto a quanto osservato a Cesenatico e Lido di Spina anche l'abbondanza della macrofauna di Bellocchio segue, nelle simulazioni, lo stesso andamento. Dopo un trend discendente secondo il quale, nei primi 3 scenari, i valori di abbondanza di macrofauna rimangono all'interno del range [453 - 523 ind./core], emerge un valore soglia nel quarto scenario, corrispondente al tempo di ritorno 25 a. in corrispondenza del quale vengono raggiunti valori di abbondanza pari a 834 ind./core.

Il numero di taxa fa rilevare una sola soglia, nel quarto scenario, al tempo di ritorno 25 a. a partire dalla quale viene raggiunto il valore di 1,13 n°/core.

La diversità di Shannon incontra, nei vari scenari, 2 valori soglia: il primo nel terzo scenario, al tempo di ritorno 30 a. in cui il trend decresce e la diversità passa da 1,2 a 1,02 ed il secondo nel quarto scenario, al tempo di ritorno 20 a. in cui al calare del numero di taxa, come già osservato per gli altri siti, la diversità si assesta anch'essa su valori meno elevati all'interno del range [0,8 - 1].

3.12.2 Meiofauna (Figg. 34 a-c)

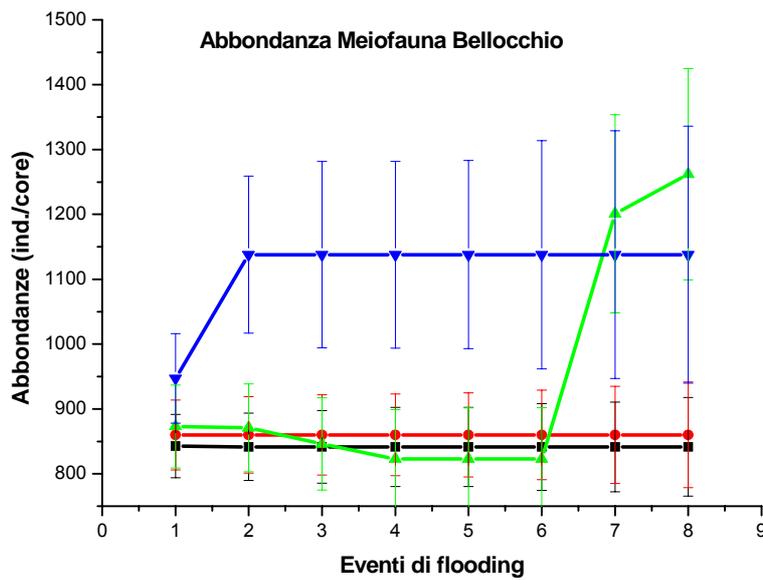


Figura 34a : Risultati delle simulazioni per il target abbondanza di meiofauna a Bellocchio (NERO=scenario1, ROSSO=scenario2, VERDE=scenario3, BLU=scenario4)

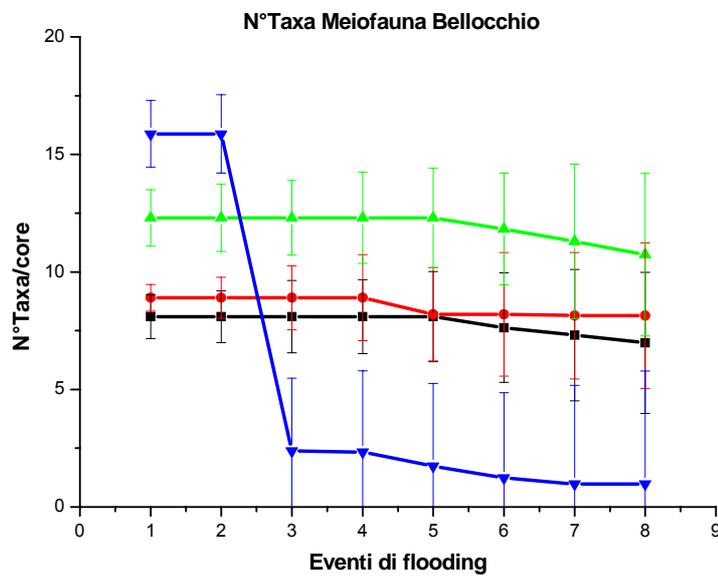


Figura 34b : Risultati delle simulazioni per il target n° di taxa di meiofauna a Bellocchio (NERO=scenario1, ROSSO=scenario2, VERDE=scenario3, BLU=scenario4)

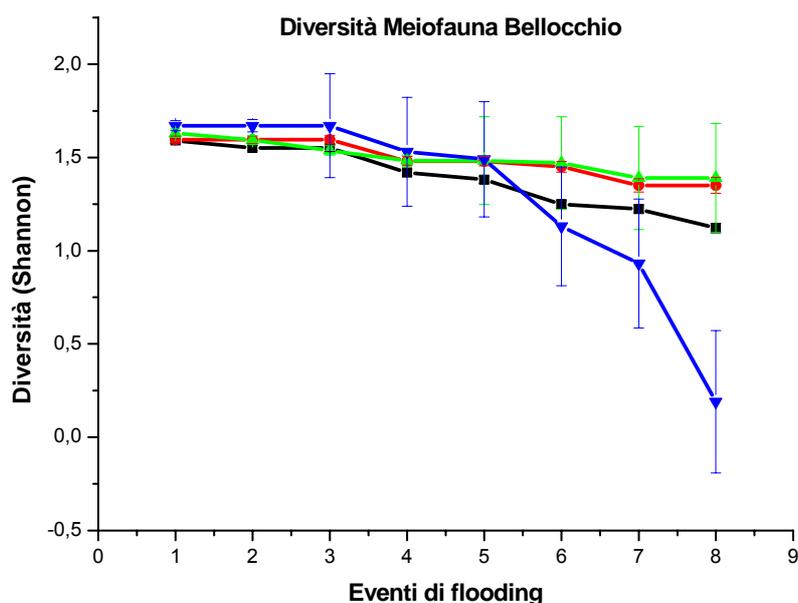


Figura 34c : Risultati delle simulazioni per il target diversità della meiofauna a Bellocchio (NERO=scenario1,ROSSO=scenario2, VERDE=scenario3, BLU=scenario4)

La risposta ai flooding della meiofauna di Bellocchio contiene 2 punti di discontinuità: uno all'interno del terzo scenario al tempo di ritorno 25 a. ed uno nel quarto scenario al tempo di ritorno 5 a.; entrambi i punti di discontinuità invertono il trend fino ad allora osservato per cui si passa nel terzo scenario da 823 ind./core a 1201 ind./core e nel quarto da 947 ind./core a 1138 ind./core.

Il numero di taxa segue lo stesso andamento simulato nei due siti precedentemente descritti, per cui, alla soglia di discontinuità localizzata dal modello nel quarto scenario, al tempo di ritorno 10 a., al “crollo” del numero di taxa corrisponde il “crollo” della diversità. Infatti, le due componenti strutturali delle comunità raggiungono rispettivamente i valori di 0,98 n° di taxa per core e di 1,03 per il valore di diversità di Shannon.

3.13 Funzioni di danno

I risultati delle simulazioni da modello dopo la validazione sito specifica sono stati utilizzati per costruire delle curve di danno per implementare il DSS .

Le curve di danno sono funzioni in cui il danno, ecologico in questo caso, è la variabile dipendente funzione di una delle variabili driver dei fenomeni che eventualmente possono creare un danno. Nel caso specifico del DSS del progetto “Theseus” le curve di danno sono riferite ad una delle variabili di flooding. Nel caso specifico della fauna bentonica della zona intertidale, la variabile di flooding prescelta è stata l’intensità di flooding, cioè l’altezza della colonna d’acqua che frange sulla spiaggia.

Le curve di danno rientrano nell’ambito della valutazione del rischio, una disciplina che porta con sé importanti implicazioni riguardo l’incertezza dei risultati, quali :

- Forti influenze della stocasticità del mondo reale.
- Errori nella valutazione di “cosa è il danno”.
- Un grado d’incertezza dovuto alla conoscenza incompleta della variabile in esame.

Nel caso delle curve di danno del progetto “Theseus” è stato deciso di utilizzare come standard del rischio ambientale un indice denominato Environmental Vulnerability Index (EVI). L’EVI può assumere 4 valori corrispondenti a 4 categorie di danno rispettivamente : 0, 1, 2, 3.

Le categorie dell’EVI sono definite come segue :

EVI = 0 → nessun danno all’habitat

EVI = 1 → danno lieve inferiore alle capacità di resilienza dell’habitat

EVI = 2 → danno rilevante recuperabile attraverso intervento dell’uomo

EVI = 3 → perdita dell’habitat

Per cui le curve di danno sono esprimibili come :

$$EVI = f(Hw)$$

dove “Hw” sta per “altezza raggiunta dalla colonna d’acqua durante l’evento o “magnitudo” del flooding.

La scelta dell’EVI è stata determinata dal fatto che tale indice non ha una definizione formale ma solo la definizione concettuale sopra riportata e per tale motivo è applicabile all’ampia varietà di habitat e di siti da cui provengono i risultati del progetto “Theseus”, così come anche dall’altrettanto varia gamma di approcci su cui tali studi si sono basati. Ogni partner ha quindi avuto la possibilità di definire la forma dell’indice nella maniera più adatta ai risultati ottenuti ed all’approccio di studio utilizzato.

Per l’intertidale sabbioso si è deciso di impostare la categorizzazione dell’ EVI in base alla differenza fra i valori di abbondanza totale della macrofauna bentonica dei popolamenti campionati e quelli di risposta alle forzanti di flooding, le categorie EVI sono quindi state formalizzate come segue :

$$|P - P_s| \leq 25\%(P) \rightarrow EVI = 0$$

$$25\%(P) < |P - P_s| \leq 50\%(P) \rightarrow EVI = 1$$

$$50\%(P) < |P - P_s| \leq 75\%(P) \rightarrow EVI = 2$$

$$|P - P_s| > 75\%(P) \rightarrow EVI = 3$$

A partire dalla categorizzazione e dalle simulazioni effettuate sono state ricavate 4 curve di danno ecologico per ogni sito, ognuna corrispondente ad uno scenario specifico.

La definizione del danno secondo il concetto EVI presenta problemi concettuali e tecnici nella sua implementazione all’interno del DSS quali:

1. La definizione del danno in categorie discrete riduce la complessità del risultato aumentando l’incertezza del responso.
2. L’ EVI non tiene conto della “direzione” del cambiamento inserendo una rilevante criticità soprattutto a livello del valore EVI 3 , non è dunque esplicativo circa il tipo di cambiamento dell’habitat.
3. L’EVI non tiene conto della sito specificità degli habitat, non considera l’influenza sull’evoluzione degli habitat delle variabili ambientali al contorno i cui valori influenzano in maniera importante la risposta dell’ecosistema.

4. La categorizzazione, in conseguenza alla scarsa dinamicità del metodo, non è validabile su scala continentale come richiesto nelle specifiche del progetto.

3.14 Implementazione dei modelli generativi FBEM nel DSS

Le criticità introdotte nel sistema di supporto decisionale dall'utilizzo dell'EVI per valutare il danno ecologico sono state risolte implementando nel DSS la struttura dell'approccio fuzzy naive Bayes . L'approccio FBEM si presenta più dinamico e preciso rispetto ad una categorizzazione del danno basata su criteri soggettivi quale è l'EVI. Inoltre l'utilizzo dei modelli FBEM permetterebbe di utilizzare lo stesso metodo per tutti gli habitat riconosciuti nei siti studio del progetto.

All'interno del DSS è stata implementata, per il momento solo per l'habitat intertidale di spiagge sabbiose, la matrice dei parametri del modello costruita in base ai campionamenti effettuati a Lido di Dante, Cervia e Cesenatico e le tre matrici di parametri ottenute dai campionamenti effettuati a Lido di Spina.

L'assegnamento delle matrici è operato in funzione della tipologia di habitat riportato sulle mappe GIS del DSS.

In ultima istanza, nel DSS è stata inclusa l'equazione vera e propria dei modelli generativi i cui valori parametrici sono assegnati a partire dai valori riportati nelle matrici e secondo le variazioni decise dall'utilizzatore cui viene richiesto di determinare i valori sito-specifici per ogni variabile d'istanza; valori che vanno a modificare i parametri riportati nelle matrici:

$$NT_1 = T_1 \times \prod_{j=1}^m \left(\sum_{i=1}^n \frac{C_1 E_{ji}}{E_{ji}} \mu_{E_{ji}} \right)$$

Il procedimento può essere definito come un apprendimento ridotto, in cui una conoscenza aprioristica viene ricalibrata in funzione di poche nuove informazioni provenienti dai siti d'interesse.

L'utilizzo dei modelli generativi direttamente nel DSS permette di ottenere responsi dinamici e che tengano in considerazione la probabile influenza delle condizioni locali

sulla risposta della macrofauna bentonica. Inoltre la particolare natura dell'output fuzzy bayesiano consente una valutazione dell'incertezza portata dalle forzanti di flooding e dell'incertezza legata alle dinamiche interne dell'habitat rendendo, in questo modo, un'ulteriore ed importante informazione all'utilizzatore del DSS.

L'approccio FBEM, funzionando in continuo e non secondo classi discrete, descrive più che un danno un'evoluzione dell'habitat rendendo così un'informazione relativa alla direzione del cambiamento.

4. SCENARIO ECOLOGICO SUL LUNGO PERIODO

Una prima applicazione completa del DSS “Theseus” verrà portata a termine per 4 siti del progetto Cesenatico, Bordeaux, Amburgo e Varna. L’applicazione è limitata a soli 4 siti in quanto rappresenta una dimostrazione delle peculiarità e dei possibili utilizzi del DSS. Solo per il sito di Cesenatico, unico fra i siti test del DSS, accanto agli scenari fisici, climatici, sociali verrà presentato uno scenario ecologico della zona intertidale di spiaggia sabbiosa, basato su una previsione a lungo termine effettuata tramite approccio fuzzy naive Bayes. La simulazione a lungo termine è basata sulle stime a lungo termine dei meteorologi e dei climatologi delle variabili di flooding dalla cui curva cumulativa sono stati estratti i tempi di ritorno degli eventi di flooding di differente intensità utilizzati per costruire gli scenari di riferimento del progetto.

Il modello applicato è, per quanto riguarda i dati di training, la determinazione dei domini, la fuzzyficazione dei dati e delle variabili ambientali al contorno, lo stesso utilizzato per la valutazione del danno ecologico a Cesenatico.

Importanti variazioni riguardano invece il numero di variabili considerate in input e lo schema di funzionamento del modello.

Alle variabili di input è stata aggiunta una nuova variabile che riguarda lo stato della variabile target al tempo immediatamente precedente il tempo della previsione. La nuova variabile serve a valutare la parte del flusso d’informazione “inficiata” (affected) dall’autocorrelazione dei valori della variabile biologica con se stessa. Come già detto, di norma, le componenti biologiche degli ecosistemi si ritengono influenzate fortemente dai valori delle componenti abiotiche; al momento però di effettuare una simulazione “in continuum” è opportuno tenere conto del fatto che i valori della componente biologica nelle condizioni precedenti al momento dell’osservazione non possono prescindere da quelli che la componente stessa aveva mostrato nell’osservazione precedente. I campionamenti di Cervia, Cesenatico e Lido di Dante hanno occupato uno spazio temporale di 2 anni che hanno permesso di definire 2 stati annuali dell’habitat. A questo punto è stata definita la scala temporale delle simulazioni decidendo, coerentemente con la tempistica di campionamento, di utilizzare una scansione annuale. La fase di apprendimento è stata condotta utilizzando le istanze del secondo anno di

campionamento, ma inserendo fra le variabili di input i valori di abbondanza della macrofauna registrati nel primo anno di campionamento.

A livello dello schema di funzionamento è stato inserito un procedimento di apprendimento iterativo secondo il quale ad ogni “run” del modello i risultati vengono immessi nel “ training data set” e utilizzati per una nuova stima dei parametri che verranno utilizzati per la simulazione dell’anno successivo. L’inserimento della nuova variabile di input e della continua ricalibrazione dei parametri unitamente alla diminuzione dell’intervallo temporale fra le simulazioni, ha consentito di evitare la simulazione usando solo il dato di “sea level rise”, il che avrebbe appesantito in maniera eccessiva l’aspetto computazionale del modello.

4.1 Risultati

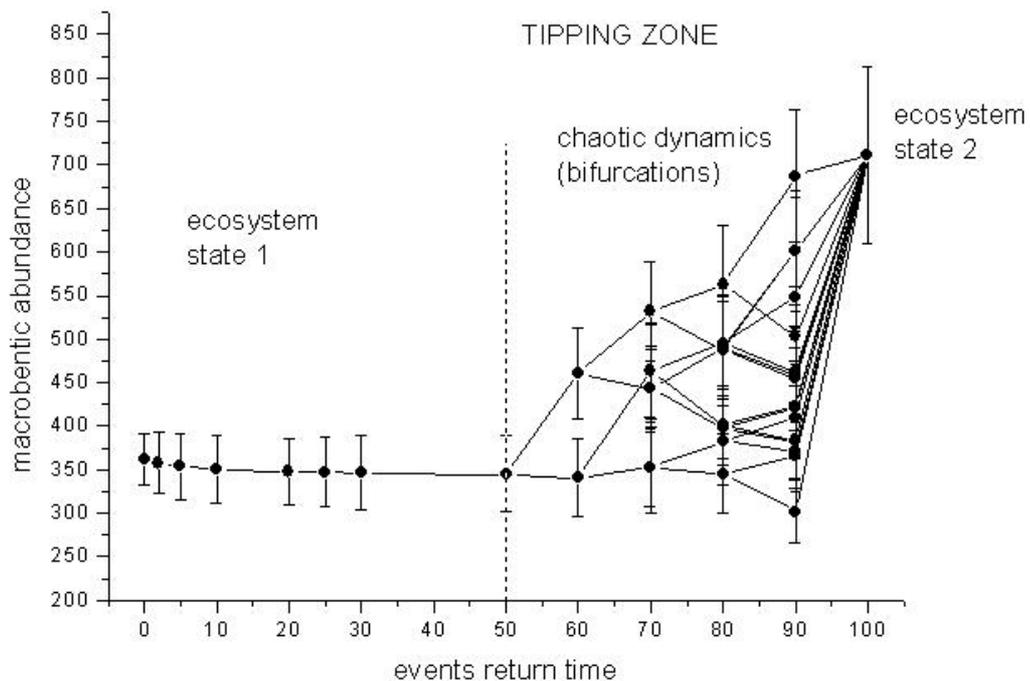


Figura 35 : Risultati della simulazione sul lungo periodo effettuata per il sito di Cesenatico sul target abbondanza della macrofauna. Il grafico descrive il passaggio da uno stato stabile ad un altro attraverso una zona ad alta incertezza (tipping zone) caratterizzata da una cascata di biforcazioni.

I risultati della simulazione (Fig. 35) mostrano come eventi di flooding successivi causino variazioni nell’abbondanza di macrofauna solo una volta superato un valore soglia di intensità dell’evento.

Per l'evento corrispondente al tempo di ritorno di 50 a., il modello inizia a fornire un risultato non chiaro dando così luogo ad una serie di biforcazioni. Dal punto di vista del modello una biforcazione sta ad indicare una situazione non ben definita in relazione alle dinamiche parametrizzate durante l'apprendimento, per cui il classificatore individua due range equiprobabili di soluzioni dell'equazione d'inferenza. A partire dall'intensità di flooding corrispondente al tempo di ritorno di 50 a., si innesca una cascata di biforcazioni successive che descrive una zona di soluzione dell'equazione d'inferenza del modello denominata "tipping zone" in quanto qualunque punto all'interno della zona suddetta può essere definito un singolo "tipping point" del sistema ecologico. Il risultato che comunque desta il maggior interesse è la convergenza delle simulazioni corrispondente alla risposta dell'ecosistema all'ultimo evento di flooding. La convergenza indica il passaggio da una zona ad alta incertezza ad un nuovo stato in cui le abbondanze dovrebbero aumentare considerevolmente.

4.2 Discussione

La particolarità di questo approccio fuzzy naïve Bayes è che la classificazione viene effettuata non su classi discrete ma su un dominio continuo descritto da una partizione fuzzy. Il risultato è dunque un range di valori nel dominio della variabile target che rappresenta il range d'incertezza della simulazione. Una nuova istanza sottoposta al metodo determina il responso della simulazione fuzzy naïve Bayes in maniera correlata con la somiglianza statistica fra la nuova istanza e le istanze di training. Al crescere dell'estremismo delle nuove istanze si osserva inizialmente un andamento lineare dei valori in output e dei range d'incertezza associati, i quali a loro volta crescono in modo proporzionale alla dissimilarità delle istanze. L'approccio appare in grado di individuare le soglie di non linearità del sistema rispetto alle quali la risposta dell'indicatore biologico segue dinamiche diverse legate ai cambiamenti di stato dell'ecosistema.

La non linearità individuata dal modello fuzzy naïve Bayes esiste in corrispondenza di alti livelli d'incertezza MAP e di scarsa verosimiglianza con le istanze di training, evidenziando come il risultato dei modelli fuzzy naïve Bayes siano legati alla bontà dei dati raccolti ed al disegno sperimentale adottato. L'attendibilità delle soglie di non

linearità individuate dal compilatore per situazioni specifiche è funzione di quanto le variabili considerate sono tenute sotto controllo in fase sperimentale.

La modellizzazione di eventi estremi deve affrontare come primo problema quello della mancanza di osservazioni e comunque dell'alta dissimilarità fra le istanze osservate e quelle previste dal “cambiamento globale”.

Il primo problema è risolto analiticamente con l'inserimento in fase di calcolo dei parametri di una correzione di Laplace, una semplice procedura di smoothing, attraverso la quale l'informazione totale viene distribuita sull'intero dominio delle variabili trasformando gli zeri in valori molto piccoli non significativi a livello statistico ma che mantengono il modello attivo sull'intero dominio di riferimento.

L'evento estremo viene invece trattato dal modello in relazione all'incertezza che porta con sé.

Inizialmente, al crescere dell'estremismo dell'evento, si osserverà un andamento lineare dei valori simulati insieme ad un andamento crescente dei range d'incertezza associati. L'autocalibrazione dei parametri permette di tenere sotto controllo l'aumento dell'incertezza fino al raggiungimento di possibili valori soglia in corrispondenza dei quali le istanze estreme possono originare degli “shift” dell'incertezza al di fuori di quello che ci si aspetterebbe dal criterio MAP. Per forti “shift” dell'incertezza associata al valore simulato, il modello effettua simulazioni per piccoli incrementi, mettendo in evidenza una cascata di biforcazioni che descrivono nel tempo campi d'incertezza simili ad “attrattori strani”, cioè aree del dominio della variabile all'interno delle quali non è possibile individuare valori puntiformi ma aree di possibilità che secondo il criterio MAP dovrebbero contenere il risultato della simulazione. L'evento molto estremo e/o l'evento estremo permanente possono causare un cambiamento di stato del sistema ambientale che corrisponde per le variabili ecologiche al superamento della “capacità portante” dell'ecosistema, tale superamento innesca nella variabile target un comportamento all'apparenza di tipo caotico che rappresenta la risposta delle variabili biotiche durante un cambiamento di stato permanente o comunque di grande intensità. Il carattere ricorsivo dell'algoritmo permette in via teorica di stabilire quanto tempo ed in che modo, secondo le osservazioni di training e l'iterazione del processo di apprendimento, il sistema ecologico si riasserterà su di un nuovo stato stabile.

Il comportamento caotico descritto dall'approccio fuzzy naïve bayes dipende da variabili che possiamo chiamare variabili di controllo per cui per un certo grado di

variazione di tali variabili cambia il comportamento qualitativo del sistema aprendo in tal modo una strada verso il caos. In quest'ottica nei sistemi ambientali sono da considerare variabili di controllo quelle variabili di sistema ad alto grado di indipendenza e associate o a bassa variabilità o a range di variabilità costanti nel tempo. La stabilità (o i bassi valori di variabilità) di alcune variabili indipendenti crea (creano) le condizioni perché altre variabili maggiormente annidate nella struttura del sistema, e per questo maggiormente sottoposte a vincoli di dipendenza, possano variare reciprocamente secondo rapporti costanti. Considerando le diverse variabilità dal punto di vista temporale si possono distinguere variabili a bassa o ad alta frequenza di oscillazione che chiameremo rispettivamente variabili "lente" (es.: salinità, temperatura) e "veloci" (es.: popolazioni biologiche). Essendo parte di un sistema, queste ultime dovranno limitare le loro oscillazioni in un intervallo temporale determinato dalle oscillazioni a bassa frequenza delle variabili lente. Da questo ne deriva che le dinamiche a lungo termine si potrebbero semplificare in considerazione della subordinazione delle variabili veloci rispetto a quelle lente. In questo modo, rimarrebbero da tener sotto controllo solo le variabili lente descrivibili da un sistema con poche dimensioni (es: interrelazione tra modelli oceanografici e fitoplancton). Bisogna però considerare che le variabili biotiche aggiungono al sistema un'ulteriore dimensione di complessità. Esse, infatti, rappresentano a loro volta dei sistemi complessi, sia a livello individuale che collettivo, poiché posseggono capacità auto-organizzative non rilevabili a livello delle variabili "lente" ma fondamentale nella comprensione delle dinamiche di quelle "veloci". Se fossero oggetto di studio della fisica, ed in parte lo sono essendo stati studiati da un punto di vista termodinamico, gli ecosistemi sarebbero classificabili come "sistemi complessi dinamici e dissipativi" (Bendoricchio e Jorgensen, 2007). La definizione stessa esclude qualsiasi approccio riduzionistico, secondo cui è sufficiente studiare le dinamiche "lente" per comprendere quelle "veloci", a meno di non conoscere perfettamente tutte le dinamiche che intercorrono fra le variabili all'interno del ecosistema. Complessità ed autoorganizzazione degli ecosistemi impongono all'osservatore di identificare dei sottosistemi coi propri domini che saranno più piccoli di tutto il campo a disposizione cioè più piccoli dell'intero sistema. La dinamica complessiva risulterà quindi una collezione di domini, ognuno con la propria dinamica, che potranno essere "debolmente" correlati fino a completamente indipendenti: nel primo caso si parlerà di complessità, nel secondo si entra nel campo dei sistemi

tendenzialmente chiusi propri del paradigma termodinamico. La stessa descrizione sistemica valida per i laser, per un fluido scaldato o per una reazione chimica risulta pertanto insufficiente per descrivere un sistema complesso le cui componenti elementari risultano esse stesse sistemi complessi. L'emergenza dell'ordine per effetto delle interazioni fra sottosistemi è stata chiamata auto-organizzazione (Prigogine, 1987); l'auto-organizzazione ecosistemica emerge (Whitfield, 2005) come si è visto ai livelli più elevati di complessità. Il caos originato dalle interazioni dei sottosistemi non può quindi essere deterministico. Le dinamiche caotiche osservate (Fig. n) non possono nemmeno descrivere una situazione di caos strettamente stocastico in quanto non solamente determinato dall'incertezza. La dinamica caotica mostrata dalla simulazione può essere inclusa nella definizione di "ipercaos" (Rossler, 1979). In una situazione di ipercaos, l'aumento del rumore statistico, e conseguentemente dell'incertezza, descrive un salto fra bacini di attrazione o fra stati stabili (dove per stabilità si intende una diminuzione del rumore statistico) dando origine a regioni ad alta incertezza nello spazio delle fasi del sistema. Il salto da un bacino d'attrazione all'altro può avvenire attraverso diverse traiettorie possibili caratterizzate da equiprobabilità.

Con l'assestamento delle variabili di controllo su nuovi valori centrali le simulazioni possono evolvere verso situazioni via via meno complesse fino ad un punto fisso oltre il quale le dinamiche di sistema assumeranno andamenti lineari o lineari a tratti secondo i nuovi valori delle variabili di controllo.

Le dinamiche caotiche riprodotte dall'approccio fuzzy naïve bayes sono dinamiche caotiche stocastiche generate secondo il criterio probabilistico in base all'incertezza generata dall'estremismo delle previsioni relative alle forzanti ambientali.

I modelli fuzzy naïve Bayes mostrano una stabilità computazionale simile a quella associata alle reti di Hopfield (J. Hopfield, 1982). Si nota come il risultato finale non sia dovuto tanto alla struttura fuzzy naïve Bayes imposta quanto alle osservazioni su cui questa viene istruita. In accordo con le teorie di Hopfield le simulazioni mostrano come gli stati stabili degli ecosistemi funzionino da attrattori in situazioni di alta incertezza che per il modello istruito rappresentano situazioni di indecisione seguenti al "tipping point".

La formazione di un attrattore caotico è legata alla bontà dei dati di training set e può essere interpretata in due modi. Nel caso in cui il training data set sia rappresentativo della disomogeneità delle varianze di sistema, l'attrattore caotico ed il campo

d'incertezza associato descrivono una situazione d'indeterminismo che precede il raggiungimento di un nuovo stato stabile. Se, al contrario, la stima dei parametri è stata effettuata su un data set non rappresentativo del sistema, l'emergere dell'attrattore potrebbe essere semplicemente dovuto ad una questione di natura matematica associata alla simulazione di dinamiche complesse nel tempo. Il carattere naïve dei modelli FNB non rende una spiegazione deterministica di quanto realmente accade in natura. Il modello, però, può sicuramente fornire indicazioni importanti, può essere utilizzato per testare ipotesi e soprattutto, in virtù della sua elasticità può essere utilizzato come "box" di osservazioni e conoscenze che vengono così contestualizzate in ottica complessa e rese utilizzabili al di là del loro aspetto sperimentale.

Un problema affrontato durante la formulazione della metodologia riguarda l'assunzione naïve bayes di indipendenza delle variabili, assunzione che in natura è praticamente impossibile rispettare. Da letteratura, comunque, si rileva come il compilatore naïve bayes sembra funzionare sorprendentemente bene anche nei casi in cui l'assunzione di indipendenza non venga rispettata. In particolare, nel caso del compilatore fuzzy naïve Bayes si può assumere che l'indipendenza sia rispettata in quanto il calcolo delle probabilità condizionate (i parametri del modello) viene effettuato non fra domini completi delle variabili ma fra eventi fuzzy di variabili diverse. Le variabili non sono sempre indipendenti mentre è vero che eventi fuzzy di variabili diverse sono indipendenti in quanto un evento fuzzy dipende dal dominio della stessa variabile, dagli altri eventi fuzzy di quella stessa variabile e dall'intero dominio di altre variabili di sistema.

5. ULTERIORI APPLICAZIONI DELL'APPROCCIO FUZZY NAIVE BAYES

Questo capitolo ha lo scopo di illustrare applicazioni dell'approccio a dati di altro genere e provenienza rispetto a quelli già utilizzati nel progetto "Theseus", così come applicazioni congiunte con altre metodologie statistiche.

5.1 Non linear power law

Come mostrato nei capitoli precedenti gli ecosistemi sono mantenuti, grazie alle loro proprietà di auto-organizzazione, in una situazione di equilibrio dinamico instabile. Essi, infatti, vengono definiti sistemi ai margini del caos, in cui l'aumento di rumore statistico dovuto ad apporti esterni può causare un salto fra bacini di attrazione di cui è possibile conoscere il punto iniziale, il punto finale (cioè i punti stabili di attrazione) ma non il percorso che viene seguito durante il passaggio di stato ecosistemico.

Rispetto a sistemi il cui descrittore è rappresentato da variabili "velocissime" come la fauna bentonica, altri ecosistemi si prestano ad un'applicazione ibrida di approccio fuzzy naive Bayes e "power laws" potenzialmente in grado di definire la traiettoria percorsa nel salto fra bacini.

Tra questi rientrano gli ecosistemi estuarini definibili come "corpi d'acqua semi chiusi, in cui si ritrovano acque a salinità intermedia a causa dell'incontro tra acqua salata ed acqua dolce" (Cameron e Pritchard, 1963; Cornacchia, 2012) . In particolare gli ecosistemi di estuario a dominio tidale possono essere descritti come sistemi in stato

stabile pulsante (Odum et al., 1995) in cui le pulsazioni interne alla variabile target dipendono in gran parte dalle intermittenze esterne alle dinamiche proprie del target e strettamente connesse con l'idroperiodo (Cornacchia, 2012).

Partendo da questi presupposti, l'idea di incrociare l'approccio fuzzy bayesiano con l'approccio "power laws" è nata dai risultati della tesi di laurea specialistica in biologia marina della dottoressa Loreta Cornacchia (2012) in cui l'approccio "power law" è stato utilizzato per descrivere i pattern di distribuzione di specie vegetali in sistemi intertidali di zona estuarina.

La "power law", o legge di potenza, è una relazione lineare che lega l'intensità di particolari eventi alla loro probabilità di accadimento secondo una legge di invarianza di scale, vale a dire secondo il principio per cui, in relazione ad un dato fenomeno, ad una variazione di scala d'intensità degli eventi la realizzazione della successione degli eventi seguirà in maniera proporzionale la realizzazione degli eventi alla scala precedente.

La legge di potenza è una relazione del tipo:

$$P(x) \propto ax^k + o(x^k)$$

In cui a e k sono costanti e $o(x^k)$ è un intorno piccolo di x^k con k chiamato esponente di scala.

Considerando trascurabile l'intorno piccolo e , a costante pari a 1, la legge di potenza è riducibile a

$$P(x) \propto x^k \quad \text{o equivalentemente} \quad P(x) \propto x^{-k}$$

per valori di k compresi nell'intervallo [2;3].

Le leggi di potenza indagano dinamiche essenzialmente auto correlative di eventi anche complessi. Come già spiegato, le variabili ecologiche devono buona parte delle loro proprietà dinamiche all'elevata dipendenza dalle condizioni ambientali al contorno, determinate in prevalenza da variabili fisiche e chimiche; nel momento in cui l'effetto di tali variabili assume importanza rilevante, l'approccio "power law" può risultare insufficiente alla descrizione delle dinamiche ecosistemiche.

Dai risultati dello studio di tesi sopra citato, emerge, infatti, come in determinate occasioni, la distribuzione reale non segua la "power law" per tutto il dominio di afferenza dell'evento considerato per cui nelle code della distribuzione i dati, a volte, "escono" dalla legge di potenza mostrando una non linearità o, in altre parole, una

“varianza nell’invarianza di scala”.

Viste le potenzialità mostrate dall’approccio fuzzy Bayesiano, è possibile utilizzarlo in maniera congiunta con l’approccio “power law” per individuare le possibili soglie di non linearità e quantificare la variazione di scala nella distribuzione indagando, in questo modo, le proprietà specifiche delle code della distribuzione osservata.

L’approccio congiunto dovrebbe essere in grado di aggiungere alle “power laws” una componente di non linearità specifica degli eventi considerati nel livello di complessità raggiunto dalle variabili ambientali “esterne. Tale componente può essere utile per la comprensione delle dinamiche di coda della distribuzione, a loro volta utilizzabili in chiave previsionale in relazione agli effetti di eventi estremi. Le capacità generative dei modelli fuzzy bayesiani irrobustirebbero le “power laws” soprattutto per rilevare gli effetti degli eventi estremi, previsti ma non osservati perchè non c’è riscontro empirico, in corrispondenza dei quali acquistano rilievo le dinamiche di distribuzione dominanti nelle code e considerate alla stregua di anomalie statistiche secondo i tradizionali approcci e tecniche di modellizzazione.

L’approccio congiunto con “power laws” permetterebbe di formalizzare i modelli fuzzy bayesiani annidati in cui l’approccio viene utilizzato per penetrare nella struttura complessa del sistema aumentando, run dopo run, il grado di complessità della variabile target.

Le “non linear power laws” possono essere applicate a dati ricavati da immagini satellitari in riferimento ai pattern di distribuzione di diversa ampiezza di specie vegetali all’interno di un ambiente estuarino; in accoppiamento con i dati da satellite dovrebbero essere rilevati anche valori di piovosità, regime fluviale, temperatura, direzione ed intensità del vento ed altezza dell’onda frangente al fine di stabilire le condizioni ambientali al contorno.

Per raggiungere il livello di complessità desiderato verrebbe portata a termine una prima modellizzazione fuzzy Bayes utilizzando come variabile target il “quantile”, rispetto al pattern di distribuzione, in corrispondenza del quale si stimerebbe, come più probabile, l’emergenza di una non linearità rispetto alla legge di potenza ricavata. I risultati ottenuti diventerebbero, insieme alle variabili già utilizzate nel primo modello, degli input d’istanza nel modello fuzzy bayes più annidato la cui variabile target verrebbe ad essere il coefficiente di scala della “power law” dominante nelle code della distribuzione.

L'approccio congiunto power laws – fuzzy naive Bayes verrebbe così ad avere come scopo la predizione dei valori delle variabili ambientali che potrebbero portare il sistema verso una soglia di non linearità. L'identificazione, tramite simulazione, delle soglie potrebbe allora essere utilizzata in fase di monitoraggio per tenere sotto controllo i sistemi in osservazione, nell'eventualità che si avvicinassero a tali valori soglia.

In un'ottica di "Climate Change", la nuova tecnica potrà servire per predire eventuali variazioni nei sistemi estuarini indotte da variazioni nei normali regimi di precipitazione, temperatura e salinità.

5.2 .Classificazione di banchi di pesce mediante descrittori strutturali e parametri ambientali in relazione a dati acustici

Un'ulteriore applicazione dell'approccio fuzzy – naive Bayes sarà effettuata per costruire sistemi esperti in grado, potenzialmente, di classificare la composizione in specie di banchi di pesce secondo dati acustici provenienti da sonar e misurazioni di variabili ambientali in situ corrispondenti alla misurazione acustica.

L'applicazione andrà ad integrare quanto già fatto dalla dottoressa D'Elia M. nell'ambito di un progetto di ricerca volto a

Applicando diversi algoritmi di classificazione sono stati riscontrati problemi nella classificazione di segnali corrispondenti a mix di specie mentre sono stati ottenuti buoni risultati per segnali monospecifici.

L'intenzione è quella di costruire un sistema esperto funzionante a due step, il primo focalizzato sulla natura mono o plurispecifica del segnale ed il secondo in cui due sistemi esperti indipendenti,

uno per segnali monospecifici e l'altro per segnali plurispecifici dovranno individuare la composizione in specie (nel caso di segnali classificati come plurispecifici) o la tipologia (nel caso di segnali considerati nel primo step monospecifici).

I data set di apprendimento verranno basati sui risultati ottenuti dalle campagne in corso nel Canale di Sicilia in cui al rilevamento del segnale acustico, e della variabili ambientali, corrispondono i conteggi quantitativi e qualitativi dei banchi di pesce che dopo essere stati misurati tramite sonar sono stati raccolti, contati e riconosciuti a livello di specie.

La suddivisione in due step è necessaria in quanto la composizione dei banchi risulta dipendente dalla natura del banco stesso ,da interazioni particolari e sito specifiche fra le variabili ambientali considerate e dal momento della rilevazione.

E' stato infatti osservato che forma e misure dei banchi dipendono fortemente dall'attività dei pesci al momento della rilevazione in particolare il rumore statistico sembra aumentare al momento della predazione in cui i banchi variano le misure caratteristiche rilevate in momenti di "calma" e si mischiano. L'aumento del rumore implica una variazione nei parametri di classificazione corrispondente, dal punto di vista del sistema "banco di pesce", ad un aumento della complessità.

La scomposizione effettuata al primo step crea di fatto due canali in cui l'informazione totale viene suddivisa riducendo un problema molto complesso a due problemi a complessità ridotta. Rispetto all'implementazione dell'approccio mostrata in questo lavoro di tesi la funzione d'appartenenza non sarà basata sulla forma a "triangolo isoscele" ma sulla forma a "trapezio isoscele" avvicinando così la classificazione fuzzy ad una classificazione per classi discrete propria dell'algoritmo naive-Bayes.

6. CONSIDERAZIONI CONCLUSIVE

L'approccio proposto in questo lavoro di dottorato è pensato come un approccio generale potenzialmente applicabile ad un'ampia varietà di ecosistemi ed in genere a problematiche ambientali complesse sia in ottica di ricerca che gestionale.

Fra le diverse categorie di modelli descritte nel capitolo introduttivo i modelli generativi costruiti secondo l'approccio fuzzy naive Bayes si possono collocare fra i modelli ibridi di cui, non a caso, non sono state riportate né caratteristiche vantaggiose né svantaggiose in quanto definite dalla specifica ibridazione effettuata.

Rispetto ad altri approcci una prima caratteristica vantaggiosa dell'approccio è la robustezza rispetto al rumore statistico presente nei dati. La robustezza, quantificata dal calcolo del G-value per ogni modello costruito, dipende in primis da due fattori: la disgregazione dell'informazione e di conseguenza l'elevato numero di parametri stimati. L'utilizzo di una "standardizzazione" fuzzy dei dati raccolti permette infatti di considerare separatamente eppure in un dominio continuo porzioni diverse dei domini delle variabili in gioco e, di stimare parametri descrittivi di relazioni condizionali fra porzioni dell'informazione totale. In questo modo i modelli riescono a discriminare fra dinamiche relazionali differenti, se non opposte, agenti fra le componenti degli ecosistemi. La maggioranza degli altri approcci statistici (metodi correlativi, PCA, indici aggregati, etc.) e degli approcci riduzionistici rappresentano invece metodologie aggregative dell'informazione, magari molto precise nel caso specifico o nelle condizioni in cui sono state formalizzate ma difficilmente trasportabili al di fuori della problematica, della posizione geografica o delle caratteristiche specifiche per cui sono stati formalizzati.

Un'altra importante caratteristica dell'approccio è quella di riuscire a replicare diverse tipologie di dinamiche simulando la non linearità di risposta ecosistemica. La capacità di replicare dinamiche diverse tra loro è dovuta alla natura naive Bayes del processo inferente che è generalmente considerato uno dei procedimenti col più ampio spettro di funzionalità.

Al processo naive Bayes è dovuta anche la capacità di individuare eventuali punti soglia nel sistema della variabile target, infatti il naive Bayes è nato per individuare le soglie che determinano la classificazione di pattern multivariati in una categoria piuttosto che in un'altra.

L'approccio si presta inoltre a diverse strutture applicative come si nota dalle differenti strutture concettuali dei modelli utilizzati nei tre siti d'interesse.

Di contro l'approccio fuzzy naive Bayes non è un approccio ad applicazione "automatica" cioè non è mai uguale a se stesso in quanto richiede, perché si abbia una accurata stima dei target scelti, una buona conoscenza del sistema che si intende modellizzare ed un'altrettanto buona conoscenza del procedimento formale. La conoscenza della problematica ambientale e dell'approccio risultano per cui basilari nella determinazione del campionamento dai cui risultati vengono stimati i parametri dei modelli generativi.

La fase di apprendimento non richiede necessariamente, perché sia efficace, di un data set di training molto ampio mentre è importante che sia rappresentata la variabilità del sistema. I risultati dei modelli generativi non rendono una spiegazione precisa circa i rapporti di causa effetto che sono comunque in parte deducibili dalla matrice di parametri.

L'approccio, nei limiti dell'applicazione mostrata per la zona intertidale di spiagge sabbiose ha risposto abbastanza bene a quanto richiesto dalla problematica esaminata e sembra potenzialmente in grado di operare come base modellistica e come box dove integrare informazioni di diversa provenienza per quanto riguarda la loro origine e gli ambiti di studio.

I risultati ottenuti riguardo la zona intertidale come recettore di eventi di flooding nell'interfaccia mare – terra possono risultare banali nella loro interpretazione ecologica e fisica (in relazione al "sea level rise") ma acquisiscono interesse in relazione alla natura "naive" della metodologia che in maniera "ingenua" è riuscita a simulare un'evoluzione logica secondo le conoscenze fin qui acquisite irrobustendola dal punto di vista formale ed analitico. Un ulteriore risultato di interesse è l'individuazione sito specifica dei punti soglia all'interno di dinamiche analoghe del tutto coerente con la sito specificità delle dinamiche ecosistemiche.

L'individuazione di punti soglia specifici risulta fondamentale in chiave gestionale permettendo, in via teorica, di ridurre l'incertezza di responsi basati unicamente su

principi di massima, sul giudizio di esperti e su risultati di esperimenti lontani dall'effettivo livello di complessità degli ecosistemi.

Riguardo i descrittori scelti l'habitat intertidale, nelle simulazioni, è risultato resistente agli eventi di flooding predetti evidenziando risposte di notevole intensità per valori elevati di altezza dell'onda e "sea level rise". Sarà interessante in un'eventuale continuazione del lavoro di valutazione e gestione fin qui compiuto considerare ,oltre che le intensità dei singoli eventi, anche la loro successione da cui si potranno ottenere risultati simili a quello mostrato per la previsione a lungo termine di Cesenatico ottenendo presumibilmente percorsi possibili diversi in luogo di punti di attrazione analoghi specifici dei siti. I risultati delle previsioni a lungo termine o comunque basate su una successione di eventi andranno integrati da attività di monitoraggio utili per effettuare una validazione delle simulazioni, per discriminare fra le traiettorie possibili simulate dai modelli e per ricalibrare la stima dei parametri attraverso l'assimilazione dei dati provenienti dai monitoraggi tenendo così sotto controllo l'espansione dell'incertezza delle simulazioni nel tempo.

BIBLIOGRAFIA

- Altenberg, L. 1991. Chaos from linear frequency-dependent selection. *American Naturalist* 138: 51-68.
- Bendoricchio G., Jorgensen S.E. 1997. Energy as goal of ecosystem dynamics. *Ecological Modelling* 102: 5–15.
- Bezdek, J.C. 1981. *Pattern Recognition with fuzzy objective function algorithms*. Plenum, New York.
- Boyd, I. L. 2012. The Art of Ecological Modeling. *Science* 337:306-307
- Bustamante, C., L. Garrido, and R. Soto. 2007. Fuzzy naive Bayesian classification in RoboSoccer 3D: a hybrid approach to decision making. Pages 507-515 in G. Lakemeyer, E. Sklar, D. G. Sorrenti, and T. Takahashi editors. *RoboCup 2006*, LNAI 4434, Springer-Verlag Berlin Heidelberg.
- Camara, B. I. 2011. Waves analysis and spatiotemporal pattern formation of an ecosystem model. *Nonlinear Analysis-Real World Applications* 12:2511-2528
- Clark, J.S., S.R. Carpenter, M. Barber, S. Collins, A. Dobson, J. A. Foley, D. M. Lodge, M. Pascual, R. Pielke Jr., W. Pizer, C. Pringle, W. V. Reid, K.A. Rose, O. Sala, W. H. Schlesinger, D. H. Wall, and D. Wear. 2001. Ecological forecasts: an emerging imperative. *Science* 293: 657-660.
- Clauset, A., C. Moore, and M. E. J. Newman. 2008. Hierarchical structure and the prediction of missing links in networks. *Nature* 453: 98-101.

- Cressie, N., C.A. Calder, J.S. Clark, J.M. Ver Hoef, and C.K. Wikle. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modelling. *Ecological Applications* 19: 553-570.
- Defeo O., A. McLachlan, D. S.Schoeman, T. A.Schlacher, J.Dugan , A. Jones, M.Lastra, F.Scapini. 2009. Threats to sandy beach ecosystem: A review.*Estuarine Coastal and Shelf Science* 81:1-12.
- Dennis, B., R. A. Desharnais, J.M. Cushing, S.M. Henson, and R.F. Costantino. 2003. Can noise induce chaos? *Oikos* 102: 329-339.
- Gomez, J., and O. Defeo. 2012. Predictive distribution modeling of the sandy-beach supralittoral amphipod *Atlantorchestoidea brasiliensis* along a macroscale estuarine gradient. *Estuarine Coastal and Shelf Science* 98: 84-93.
- Gotway, J.C., R.B. Ferguson, G.W. Hergert, and T.A. Peterson. 1996. Comparison of kriging and inverse-distance methods for mapping soil parameters. *Soil Science Society of America* 60: 1237-1247.
- Grimm, V., E. Revilla, U. Berger, F. Jeltsch, W. M. Mooij, S. F. Railsback, H. H. Thulke, J. Weiner, T. Wiegand, and D. L. DeAngelis. 2005. Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science* 310: 987-991.
- Hopfield, J.J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of Natural Academy of Science USA* 79: 2554-2558.
- Hughes, E. J. 2001. Evolutionary multi-objective ranking with uncertainty and noise. *Evolutionary Multi-Criterion Optimization, Proceedings* 1993: 329-343.
- Jensen, H. J. 1998. *Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems*. Cambridge University Press, Cambridge, UK.
- John, G.H. and P. Langley. 1995. Estimating continuous distributions in Bayesian classifiers. Pages 338-345 *in* P. Besnard and S. Hanks editors. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann Publishers, San Mateo.

- Kohavi, R. 1996. Scaling up the accuracy of naive-Bayes classifier: A decision-tree hybrid. Pages 202-207 in AAAI Press. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining
- Levin, S. A. 1998. Ecosystems and the biosphere as complex adaptive systems. *Ecosystems* 1: 431-436.
- Levin, S. A. 2005. Self-organization and the emergence of complexity in ecological systems. *Bioscience* 55: 1075-1079.
- Lewis, D.D. 1998. Naïve (Bayes) at forty: the independence assumption in information retrieval. *Lecture Notes in Computer Science* 1398: 4-15.
- Liu, B. 2006. A survey of credibility theory. *Fuzzy Optimization and Decision Making* 5: 387-408.
- McLachlan A. 1983 Sandy beach ecology : A review. In:McLachlan, A., Erasmus, T.(Eds.), *Sandy beaches as ecosystems*. W. Junk, The Hague, pp. 321-380
- McLachlan A., A. Brown 2006. *The Ecology of Sandy Shores*. Elsevier, Amsterdam
- May, R. M. 1995. Necessity and chance: deterministic chaos in ecology and evolution. *Bulletin of the American Mathematical Society* 32: 291-308
- May, R. M. 1999. Unanswered questions in ecology. *Philosophical Transactions of Royal Society* 354: 1951-1959.
- McLachlan, A., and A. Dorvlo. 2005. Global patterns in sandy beach macrobenthic communities. *Journal of Coastal Research* 21: 674-687.
- Migliavacca, M., O. Sonnentag, T. F. Keenan, A. Cescatti, J. O'Keefe, and A. D. Richardson. 2012. On the uncertainty of phenological responses to climate change, and implications for a terrestrial biosphere model. *Biogeosciences* 9: 2063-2083.
- Pazzani. M. 1995. Searching for attribute dependencies in bayesian classifiers. *In Fifth International Workshop on Artificial Intelligence and Statistics* : 424-429

- Planque, B., E. Bellier, and C. Loots. 2011. Uncertainties in projecting spatial distributions of marine populations. *Ices Journal of Marine Science* 68:1045-1050.
- Prigogine, I. 1987. Exploring complexity. *European Journal of Operational Research* 30: 97-103
- Randon, N.J., J. Lawry, K. Horsburgh, and I.D. Cluckie. 2008. Fuzzy Bayesian modelling of sea-level along the east coast of Britain. *IEEE Transactions on Fuzzy Systems* 16: 725-737.
- Rossler, O.E. 1979. An equation for hyperchaos. *Physics Letters A* 71: 155-157.
- Ruelle, D. 2006. What is...a Strange Attractor? *Notices of the American Mathematical Society* 53: 764–765.
- Sarewitz, D. and Pielke, R.A. Jr. 2001. Extreme events: a research and policy framework for disasters in context. *International Geology Review* 43: 406-418.
- Schlacher, T. A., D. S. Schoeman, J. Dugan, M. Lastra, A. Jones, F. Scapini, and A. McLachlan. 2008. Sandy beach ecosystems: key features, sampling issues, management challenges and climate change impacts. *Marine Ecology-An Evolutionary Perspective* 29: 70-90.
- Schloeder, C.A., N.E. Zimmermann, and M.J. Jacobs. 2001. Comparison of methods for interpolating soil properties using limited data. *Soil Science Society of America* 65: 470-479.
- Sibani, P., J. C. Schon, P. Salamon, and J. O. Andersson. 1993. Emergent hierarchical structures in complex-system dynamics. *Europhysics Letters* 22: 479-485.
- Störr, H.P. 2002. A compact fuzzy extension of the Naive Bayesian classification algorithm. Pages 172-177 *in* Proceedings InTech/VJFuzzy.
- Svirezhev, Y. M. 2008. Nonlinearities in mathematical ecology: Phenomena and models - Would we live in Volterra's world? *Ecological Modelling* 216: 89-101.

- Thom, R. 1975. *Structural Stability and Morphogenesis*. Benjamin–Addison Wesley, New York
- Viertl R. and D. Hareter. 2004. Fuzzy information and imprecise probability. *ZAMM. Zeitschrift für Angewandte Mathematik und Mechanik* 84: 731-739.
- Whitfield, J. 2005. Order out of chaos. *Nature* 436: 905-907.
- Widyantoro, D. H., and Yen, J. 2000. A fuzzy similarity approach in text classification task. *In Proceedings of the IEEE International Conference on Fuzzy Systems, FUZZ-IEEE*.
- Wolfram, S. 1986. *Theory and applications of cellular automata*. Advanced Series of Complex Systems: World Scientific Publication, Singapore.
- Woodard, R., D. E. Newman, R. Sanchez, and B. A. Carreras. 2007. Persistent dynamic correlations in self-organized critical systems away from their critical point. *Physica A: Statistical Mechanics and Its Applications* 373: 215-230.
- Wu, J. G., and J. L. David. 2002. A spatially explicit hierarchical approach to modeling complex ecological systems: theory and applications. *Ecological Modelling* 153: 7-26.
- Zadeh, L. A. 1965. Fuzzy sets. *Information and Control* 8: 338-353.
- Zadeh, L. A. 1968. Probability measures of fuzzy events. *Journal of mathematical analysis and applications* 23: 421-427.