Politecnico di Torino ***Università degli Studi di Torino***

Corso di Laurea Magistrale in Ingegneria per l'Ambiente e il Territorio

TESI DI LAUREA MAGISTRALE

Eco-Idraulica Fluviale: legami tra la struttura del reticolo idrografico e la distribuzione spaziale dei macroinvertebrati bentonici in Piemonte.

Relatori:
Prof. Luca Ridolfi
Ing. Riccardo Vesipa
Prof. Stefano Fenoglio (Università degli Studi di Torino)
Dr. Alberto Doretto (Università degli Studi di Torino)

Candidato:
Fabio Tarena

Marzo 2020

INDICE

1. Introduzione 6
1.1 Posizione del problema 6
1.2 Fasi di lavoro 7
1.3 Limiti 9

2. Teorie Ecologiche 10
2.1 Problema Matematico: riconoscimento di pattern spaziali 10
2.2 Cenni di Ecologia Comunitaria 12
2.2.1 Determinanti Locali e Regionali 13
2.2.1.1 Struttura Dendritica: Dispersione ed Interconnessione 15
2.2.1.2 Fattori Ambientali 18
2.2.1.3 Interazione Biotica 21
2.3 Teoria delle Metacomunità 22
2.3.1 Concetto di metacomunità 22
2.3.2 Definire una metacomunità nei casi reali 22
2.3.3 Ecosistema Fluviale come metacomunità 24
2.3.4 I quattro paradigmi della teoria 26
2.3.4.1 Patch Dynamics 26
2.3.4.2 Species Sorting 26
2.3.4.3 Mass Effects (Source-Sink Dynamics) 27
2.3.4.4 Neutral Theory 28
2.3.5 Conclusioni della teoria 28

3. Database Ecologico 31
3.1 Dati Disponibili 31
3.1.1 Struttura preliminare del Database 31
3.1.1.1 Uso dei macroinvertebrati nel biomonitoraggio dei corsi d'acqua 31
3.1.1.2 Metodo di campionamento AQEM 32
3.1.1.3 Matrice dei dati grezzi 33
3.1.2 Dati Progetto RENERFOR 34
3.2 Georeferenziazione del Database 35
3.2.1 Importazione dei dati in QGIS 35
3.2.2 Congruenza tra Database e Reticolo 36
3.3 Completamento del Database 37
3.3.1 Metriche Biologiche 37

3.3.2 Script GRASS 40
3.3.2.1 Python-GRASS – Estrazione bacini idrografici 40
3.3.2.2 Python-GRASS – Estrazione descrittori dei bacini 47
3.3.3 Aggiunta ultimi parametri 51

4. Variabili 52
4.1 Premessa 52
4.2 Variabili Morfologiche 53
4.2.1 Area del bacino 53
4.2.1.1 Impatto ecologico 53
4.2.1.2 Metodo di calcolo 54
4.2.2 Densità di abitanti al kilometro quadrato 55
4.2.2.1 Impatto ecologico 55
4.2.2.2 Metodo di calcolo 55
4.2.3 Esposizione dei versanti 57
4.2.3.1 Impatto ecologico 57
4.2.3.2 Metodo di calcolo 57
4.2.4 Luce Solare Media 58
4.2.4.1 Impatto ecologico 58
4.2.4.2 Metodo di calcolo 59
4.2.5 Pendenza 61
4.2.5.1 Impatto ecologico 61
4.2.5.2 Metodo di calcolo 61
4.2.6 Quota 62
4.2.6.1 Impatto ecologico 62
4.2.6.2 Metodo di calcolo 63
4.2.7 Uso del suolo 64
4.2.7.1 Impatto ecologico 64
4.2.7.2 Metodo di calcolo 64
4.3 Variabili Idrologiche 67
4.3.1 Precipitazione Media Annua 67
4.3.1.1 Impatto ecologico 67
4.3.1.2 Metodo di calcolo 68
4.3.2 Portata Media Annua 68
4.3.2.1 Impatto ecologico 68
4.3.2.2 Metodo di calcolo 69
4.4 Variabili Idrauliche 69
4.4.1 Ordine di Strahler 70
4.4.1.1 Impatto ecologico 70
4.4.1.2 Metodo di calcolo 70
4.4.2 Morfologia dell'alveo 71

4.4.2.1 Impatto ecologico 71
4.4.2.2 Metodo di calcolo 72
4.4.3 Naturalità del fiume 72
4.4.3.1 Impatto ecologico 72
4.4.3.2 Metodo di calcolo 73

5. Analisi Lineare e Correlazioni 74
5.1 Introduzione 74
5.2 Studio qualitativo dei parametri ecologici 75
5.2.1 Correlogrammi 76
5.3 Regressione Lineare Semplice 77
5.3.1 Cenni teorici – Modello di regressione lineare semplice 78
5.3.1.1 Minimi Quadrati – Stima dei coefficienti di regressione 78
5.3.1.2 Adeguatezza del modello – Coefficiente R^2 80
5.3.1.3 Test di significatività – t-Student 81
5.3.2 Applicazione ai dati ecologici 82
5.3.2.1 Regressione lineare semplice in Matlab 82
5.4 Regressione Lineare Multipla 88
5.4.1 Cenni teorici – Modello di regressione lineare multipla 89
5.4.1.1 Multicollinearità e scelta delle variabili 89
5.4.1.2 Minimi Quadrati – Stima dei coefficienti di regressione parziali 90
5.4.2 Applicazione ai dati ecologici 93
5.4.2.1 Scelta delle variabili – Correlazioni tra le variabili idromorfologiche 93
5.4.2.2 Analisi Multivariata – Modello Lineare 96
5.4.2.2.1 Metodo 96
5.4.2.2.2 Descrizione dello script Matlab 98
5.4.2.2.3 Risultati 100
5.4.2.3 Analisi Multivariata – Modello Esponenziale 104
5.4.2.3.1 Risultati 105
5.4.2.4 Zonazioni 107

6. Analisi Spaziale e Semivariogrammi 111
6.1 Introduzione 111
6.2 Concetto di *Continuum Fluviale* 112
6.3 Semivariogrammi 113
6.4 Applicazione ai dati ecologici 115
6.4.1 Distanze Euclidee 116
6.4.2 Distanze lungo il reticolo 117
6.4.2.1 Preparazione dei dati preliminari e assegnazione delle stazioni al reticolo 118
6.4.2.2 Algoritmo per il calcolo delle distanze (bidirezionali) 122

6.4.2.3 Calcolo delle distanze monodirezionali 124
6.4.3 Risultati (semivariogrammi) 125
6.4.3.1 Script Matlab - Classi di distanza e generazione dei semivariogrammi 125
6.4.3.2 Sottobacino del Tanaro 128
6.4.3.3 Sottobacino del Po 133

7. Conclusioni 137

Appendice 141
Allegato 1 – Script python-GRASS per l'estrazione dei bacini idrografici 142
Allegato 2 – Script python-GRASS per l'estrazione dei parametri descrittori dei bacini 143
Allegato 3 – Script Matlab per le regressioni monoparametriche 145
Allegato 4 – Script Matlab per valutare la collinearità tra variabili idromorfologiche 147
Allegato 5 – Script Matlab per la regressione multivariata 149
Allegato 6 – Script Matlab per l'assegnazione delle stazioni al reticolo idrografico 161
Allegato 7 – Script Matlab per il calcolo delle distanze lungo il reticolo fluviale 162
Allegato 8 – Funzione Matlab per il calcolo della distanza tra punti proiettati su una superficie di riferimento 165
Allegato 9 – Script Matlab per il calcolo dei semivariogrammi 166

Bibliografia 167

Sitografia 168

Capitolo 1

Introduzione

1.1 Posizione del problema

Tra tutte le tipologie di ambienti che caratterizzano e compongono le differenti regioni climatiche del pianeta, quello fluviale rappresenta uno dei più sfaccettati e complessi, sia dal punto di vista dello studio dei processi fisici legati all'acqua che si muove all'interno dell'alveo (campo dell'*idraulica fluviale*), che da quello delle interazioni bioecologiche che si realizzano all'interno di esso (campo dell'*ecologia comunitaria*).

Da un lato, è presente, dunque, la complessità dell'idraulica fluviale. Essa è data dalla stretta relazione di interdipendenza che c'è tra moto dell'acqua e ed evoluzione della morfologia del fondale (uno influenza l'altro e viceversa). L'analisi dal punto di vista matematico dell'evoluzione di tale sistema fisico porta ad un sistema accoppiato di equazioni: la corrente dei fiumi si muove tipicamente in moto turbolento, che dipende dalla forma del fondale, e la sua dinamica è matematicamente regolata dalle equazioni di Navier-Stokes (la cui risoluzione in forma chiusa rappresenta uno dei più grandi ed importanti problemi aperti della fisica moderna); l'evoluzione dell'alveo è regolato dalle equazioni della morfodinamica, che a sua volta dipendono dal moto dell'acqua (e dunque dalla presenza di turbolenza).

Dall'altro, c'è il problema relativo allo studio dell'ecosistema fluviale, ossia il comprendere come il sistema fisico e le caratteristiche ambientali si interfacciano coi processi biologici di vita, sviluppo e sopravvivenza delle comunità che vivono nei pressi degli ambienti fluviali. In ecologia, quello fluviale è il classico esempio di ecosistema aperto, ovvero esso non risulta isolato, ma in continuo scambio di materia ed energia con tutti gli ambienti che lo circondano. Questo fatto fa sì che le caratteristiche ambientali (sia fisiche che chimiche) di un sistema fluviale non siano mai costanti, ma sempre estremamente variabili nel tempo (da stagione a stagione) e nello spazio (da monte a valle). A sua volta, le peculiarità dell'habitat influenzano la vita e la distribuzione delle specie lungo il territorio. La varietà di specie che partecipano alle dinamiche ecosistemiche è, inoltre, estremamente ampia e diversificata.

La trattazione che si intende eseguire in questa tesi si pone all'interno di questo secondo tipo di problema, riguardante lo studio degli ecosistemi fluviali. L'idea è quella di cercare di trovare un qualche tipo di legame matematico tra componenti ambientali (morfologiche e idrauliche, nello specifico) e distribuzione delle specie biologiche all'interno di una regione di interesse. Si intende cercare di capire come le scelte organizzative da parte delle specie siano o meno influenzate dalle caratteristiche, cosiddette, abiotiche dell'ecosistema (sarebbero le componenti non viventi di un ecosistema, quali, appunto, quelle ambientali e fisiche). In particolare, lo studio che verrà proposto in questa tesi riguarda l'analisi, mediante varie tecniche statistiche, di una serie di dati biologici rilevati in campo a scala regionale. In questo caso, la zona di interesse rappresenta la regione

Piemonte (o alcuni suoi bacini idrografici) e i dati da analizzare rappresentano quelli di misura di abbondanza delle famiglie di macroinvertebrati bentonici lungo le aste del reticolo fluviale della regione.

Anche se la tipologia di problema non risulta particolarmente astratta (infatti, il concetto di adattamento di specie diverse ad habitat diversi rappresenta, ormai, quasi universalmente una conoscenza di base di chiunque), si vedrà, nello sviluppo dei successivi capitoli della tesi, come, in realtà, esistono dei risvolti ecologici molto più profondi, di cui non ci si accorge attraverso un semplice sguardo superficiale al problema.

In letteratura scientifica internazionale, esistono molti esempi di articoli accreditati in questo campo di analisi, che può essere denominato *eco-idraulica fluviale*. Si tratta di un campo di indagine non molto antico, che ha subito numerosi progressi nel corso degli ultimi anni, sia in termini di esperimenti fatti che di teorie sviluppate. In particolare, il più grande passo avanti è rappresentato dallo sviluppo della *teoria delle metacomunità*, che permette di studiare le varie comunità biologiche viventi lungo un fiume non come isolate tra di loro, ma come mutuamente interagenti. Quello che la teoria ha permesso, rispetto ad altri studi svolti in passato, è il passaggio da una piccola e limitata scala spaziale di indagine (fenomeni locali di competizione, predazione e sopravvivenza) ad una molto più ampia (fenomeni di influenza regionale). Ad ogni modo, tale discorso (assieme alle definizioni dettagliate dei vari concetti espressi) verrà trattato, per esteso, nel corso nel *Capitolo 2* della tesi.

Dunque, partendo dalla comprensione di questi concetti teorici e delle metodologie di analisi sviluppate al fine di poter analizzare in maniera completa e significativa la strutturazione spaziale delle comunità ecologiche fluviali, si è cercato, nel corso della tesi, di intraprendere il cammino percorso nelle pubblicazioni scientifiche di riferimento, seguendo la strada che è stata, in parte, già tracciata dai ricercatori. In particolare, il lavoro, che verrà presentato nel corso dei successivi capitoli, è caratterizzato da una doppia fase di procedimento: una prima fase di organizzazione, ordinamento e completamento dei dati ed una seconda fase di interpretazione vera e propria.

Si precisa che il lavoro svolto è il risultato di una collaborazione tra *Politecnico di Torino* e *Università degli Studi di Torino*. Infatti, i dati biologici di partenza di tutte le analisi, che sono state successivamente svolte, sono stati ottenuti grazie all'opera di ricerca e misura in campo del gruppo di ricerca del prof. Stefano Fenoglio del dipartimento di *Scienze della Vita e Biologia dei Sistemi* dell'Università degli Studi di Torino. Inoltre, il problema che si vuole studiare si pone, come già sottolineato, a cavallo tra due ambiti scientifici diversi ma complementari: l'idraulica fluviale e l'ecologia comunitaria.

1.2 Fasi di lavoro

Il privilegio di cui si è potuto disporre nell'iniziare il lavoro contenuto in questa tesi è stato l'accesso a dei dati preliminari già raccolti e accorpati a formare un database georeferenziato (secondo sistema di coordinate geografiche). Tali dati sono inerenti ad un progetto di mappatura della fauna bentonica dei maggiori corsi d'acqua piemontesi. Si tratta di dati raccolti in più stazioni di campionamento e con una metodologia standard e ripetibile nel tempo (che verrà meglio descritta nel *Capitolo 3*).

La prima fase del lavoro è consistita nel cercare di entrare nella mentalità di ragionamento di un'analisi in campo eco-idraulico. È stata eseguita, dunque, una disamina approfondita, su motori di ricerca appositi (quali *Web of Science* o *Scopus*), di articoli scientifici accreditati in questo ambito di ricerca. La lettura e la comprensione delle principali tecniche di analisi utilizzate dai ricercatori hanno permesso di entrare nel merito della scienza che sta dietro al problema esposto.

Una volta compresi i concetti ecologici principali, la tipologia di dati normalmente utilizzati nelle analisi e le tecniche statistiche tipicamente adoperate si è passati all'operazione di aggiornamento del database con valori di parametri che sarebbero poi potuti risultare utili nelle successive analisi. La scelta della natura dei parametri da utilizzare si è basata su quelli che tipicamente venivano riscontrati in letteratura. In particolare, molti parametri citati negli articoli (principalmente chimici) richiedevano delle misurazioni precise in campo di cui non si poteva disporre. Per cui, l'aggiunta della maggior parte di informazioni al database è stata possibile solo grazie a delle analisi spaziali in remoto (mediante software di analisi topografica e morfologica). Tali parametri sono poi stati l'oggetto fondamentale per le successive analisi spaziali. Poiché calcolati in remoto, essi sono risultati di natura idraulica, idrologica e morfologica (e, quindi, non direttamente chimica, ma a cui possono essere legati indirettamente, come sarà possibile spiegare al *Capitolo 4*).

Altro passo che è stato fatto, prima delle vere e proprie analisi, è stato cercare di accorpare le centinaia di dati ecologici (che racchiudevano informazioni su 150 famiglie biologiche diverse) in pochi parametri rappresentativi. Ciò è stato possibile attraverso l'ottenimento di quattro metriche ecologiche molto usate in campo internazionale.

A questo punto, si è trattato di scegliere delle tecniche di analisi che permettessero di legare i dati ecologici a quelli ambientali (idromorfologici). Per confronto con la letteratura scientifica e anche la semplicità di applicazione, si è scelto di eseguire, come primo passo verso l'interpretazione dei dati, delle analisi di regressione lineare: prima monoparametriche, poi multiple. Il senso del passaggio da una all'altra è da ricercare nella volontà di creare dei modelli che avessero sempre più significatività statistica (in termini di coefficiente R^2). Sempre in quest'ottica, sono stati creati più modelli di regressione multipla (lineare ed esponenziale) e, infine, sono state eseguite due analisi, non più sul database completo, ma solo su una parte di esso. Lo scopo è stato verificare se l'uniformità territoriale (dal punto di vista delle caratteristiche ambientali) potesse avere una qualche influenza sui risultati. Una prima zonazione ha riguardato la zona del cuneese, la seconda, quella del Po (in particolare, si è trattato di due sottobacini, uno del Tanaro e uno del Po). La scelta di queste due zone non è stata casuale ma ha avuto a che vedere con la densità di informazioni ivi presenti, piuttosto che in altre parti più scarne del database.

Terminate le analisi mediante regressione, da un confronto con un altro articolo scientifico, si è scelto di affinare ulteriormente la metodologia di analisi. Per la complessità di applicazione del metodo, esso è stato limitato alle sole due zonazioni già considerate alla fine delle analisi lineari (invece di considerare l'intero database). Il metodo è consistito nell'analisi spaziale delle aste di un reticolo fluviale mediante la costruzione di particolari grafici statistici, detti semivariogrammi. Al fine di ottenere dei risultati completi è stato necessario sviluppare dei nuovi algoritmi per il calcolo delle distanze tra punti lungo il reticolo fluviale (in *Matlab*) e applicarli ai dati reali forniti dal database. Il calcolo delle distanze, infatti, è parte dei passaggi da compiere per la costruzione di un semivariogramma. In particolare, sono state costruite più tipologie di semivariogramma (che riflettono il metodo con cui sono state calcolate le distanze tra punti di campionamento, euclidee o

lungo il reticolo), in modo tale da ottenere dei risultati con significato differente e complementare, in termini di analisi spaziale. Il risultato di tale analisi è stata una comprensione più approfondita dei fenomeni spaziali che influenzano la distribuzione delle comunità di macroinvertebrati lungo un bacino idrografico.

1.3 Limiti

Nonostante la molteplicità e la varietà dei risultati ottenuti nel corso del lavoro, gli ostacoli al raggiungimento di conclusioni complete non sono stati pochi.

Una prima ragione è da attribuire alla natura dei dati di partenza. La rilevazione e i campionamenti in campo hanno riguardato i soli dati biologici di abbondanza delle specie (senza associazione con quelli ambientali). Mancano tutta una serie di parametri, soprattutto di natura chimica, la cui assenza impedisce la creazione di un legame completo (mediante le regressioni) tra territorio in esame e distribuzione delle comunità biologiche. È, infatti, innegabile che la chimica delle acque abbia una notevole influenza sulla vita di un organismo. Escludere parametri come pH, contenuto di ossigeno disciolto o temperatura implica certamente l'incompletezza di un modello di regressione, in campo eco-idraulico. Pertanto, le analisi sono state condotte usando solo parametri calcolabili in remoto (mediante software di interpolazione spaziale), fisici (idrologici e idraulici) e/o morfologici (struttura del territorio). È vero che tali parametri di natura idromorfologica possono essere, in qualche modo, correlati alla natura chimica del territorio in esame, ma ciò rappresenta comunque un legame indiretto con l'ecologia della zona.

Un'altra ragione risiede nell'estrema complessità del fenomeno studiato. Quando si trattano i fenomeni naturali, è praticamente impossibile determinare un modello che abbia una struttura tale da contenere in sé tutte le possibili variabili che influenzano ciò che avviene nel sistema reale. La realtà è un qualcosa di continuo (nello spazio e nel tempo). Ogni volta che si crea un modello si va a discretizzare la realtà, rompendo quella componente di continuità che però sarebbe imprescindibile per una comprensione completa e totale dei fatti. Tuttavia, nonostante l'impossibilità di creare un modello totalmente rappresentativo, sono state effettuate delle scelte che hanno portato all'assunzione di certi modelli matematici, invece che altri, sulla base dell'applicabilità ai dati che si avevano a disposizione. Infatti, è la natura dei dati in ingresso che spesso determina delle restrizioni sulle tipologie di analisi che possono essere eseguite e di conclusioni che possono essere tratte.

Si parla, in questo senso, di generalità delle conclusioni, in quanto esse riguardano i quattro parametri ecologici in cui sono stati accorpati i dati campionati in campo. Le analisi, come è stato già detto, sono state effettuate su questi parametri, invece che sulle singole misurazioni specie per specie, in modo tale da poter avere una panoramica globale sull'intera zona coperta dal database. Naturalmente, l'aumento in generalità ha portato inevitabilmente alla perdita di un certo grado di dettaglio, che, invece, era presente nelle informazioni di partenza.

Capitolo 2

Teorie Ecologiche

2.1 Problema Matematico: riconoscimento di pattern spaziali

L'aspetto principale, su cui si intende basare tutta la discussione che verrà sviluppata in questa tesi, riflette un problema matematico molto più generale e applicabile in vari campi, non solo a quello ecologico. Il concetto generale è l'estrazione da un numero (generalmente elevato) di dati grezzi, inerenti al particolare ambito scientifico che si sta studiando, di particolari associazioni statistiche che permettano una classificazione di essi secondo caratteristiche comuni. I dati sono definiti grezzi, in quanto, inizialmente, sono riconducibili alle sole misurazioni effettuate direttamente in campo (o estrapolate, in remoto, per mezzo di software), senza ancora nessuna successiva elaborazione di tipo teorico-matematico. Anzi, sono proprio queste elaborazioni ad essere l'oggetto principale dell'analisi che si intende eseguire.

In generale, in ambito statistico, tale problema è definito attraverso il termine di *pattern recognition* (o, in italiano, *riconoscimento di pattern*). Esso consiste nell'analisi e identificazione di pattern all'interno di un'ampia area di dati sperimentali al fine di determinarne una classificazione di carattere significativo.

Il concetto di *pattern* è estremamente importante all'interno della scienza dei dati, ossia quell'insieme di principi metodologici volti a interpretare e caratterizzare il significato di una serie di informazioni empiriche (*Data Science*). La traduzione italiana letterale sarebbe "disposizione", ma il concetto matematico-scientifico è molto più profondo. A seconda del contesto, il termine può essere inteso in svariati modi (come "modello", "struttura", "schema"), però, generalmente, si riferisce alla ripetizione di una determinata sequenza di informazioni all'interno di una serie di dati grezzi, con una regolarità osservabile nello spazio e/o nel tempo.

Si evince come le applicabilità di questa definizione siano estremamente varie e non ristrette alla sola ricerca scientifica. Riconoscere dei pattern risulta una procedura universale nel momento in cui ci si voglia riferire ad uno schema di comportamenti, azioni, fenomeni o situazioni ricorrenti secondo una certa regolarità. In questo senso, si può spaziare dall'ambito medico a quello politico, dall'ambito sociale a quello ecologico, dall'ambito fisico a quello informatico.

L'obiettivo fondamentale è, dunque, quello di creare un qualche ordinamento di questi dati, secondo determinati criteri basati su informazioni inferibili dai dati stessi (mediante analisi e processi di natura statistica). Si tratta di passare dal disordine, dalla complessità e dall'ambiguità dei dati singoli, all'ordine dato dall'identificazione di un certo numero di regolarità e dalla conseguente classificazione.

Questo discorso si interfaccia bene con le tematiche ecologiche che verranno trattate e analizzate nel corso di questa tesi. In questo caso, i concetti appena espressi si traducono nell'analisi statistica di misure di abbondanza di determinate famiglie di macroinvertebrati fluviali (che sarebbero i dati grezzi), al fine di scoprire quali siano i fenomeni (parametri) che regolano la disposizione di tali comunità lungo il reticolo fluviale. Il concetto matematico, che sta alla base del problema espresso, è il medesimo, così come l'obiettivo finale. Si vogliono trovare delle regolarità sulle modalità e tipologie di strutturarsi di tali comunità viventi nei pressi dell'ambiente fluviale.

Comprendere l'organizzazione non casuale delle comunità biologiche nello spazio è un tema centrale in ecologia. I pattern composizionali che si osservano in ogni punto dello spazio (o del tempo) derivano da fattori operanti in un vasto intervallo di scale spaziali e temporali, includendo *determinanti locali* (condizioni dell'habitat, rapporti di predazione e competizione), *fattori regionali* (dispersione, geologia e clima) e *vincoli riguardanti la storia e l'evoluzione della zona* (glaciazioni ed evoluzioni di uso e copertura del suolo). Chiarire quali siano i determinanti chiave dei pattern riconoscibili non rappresenta solamente un interesse di natura teorica, ma risulta anche molto importante nel campo dell'ecologia applicata, in quanto può fornire robuste informazioni in campo di cambiamenti climatici o di azioni risanatrici per l'ecosistema.

In ogni caso, si tratta di concetti molto complessi e, talvolta, apparentemente astratti, che verranno spiegati e descritti in maniera più approfondita nel corso di questo capitolo.

Lo studio della relazione che esiste tra tali determinanti e l'organizzazione spaziale delle comunità ecologiche viene eseguito da un particolare (nonché abbastanza recente) sotto ambito dell'ecologia comunitaria, la cosiddetta *teoria delle metacomunità*. Il suo scopo generale è il tentativo di predizione dell'assemblamento delle comunità secondo differenti prospettive e modelli teorici. Essa studia come l'interfacciarsi di fattori regionali e locali regola i cambiamenti a livello di ecosistema e in quale misura.

Nei successivi paragrafi verrà presentata una panoramica generale riguardante i vari tipi di determinanti, le differenti teorie ecologiche e i vari modelli e assunzioni teorici elaborati nel cercare di dare una metodologia standard per lo studio del problema di *pattern recognition* in ambito di ecologia comunitaria.

2.2 Cenni di Ecologia Comunitaria

In ecologia, con *comunità* si intende un gruppo o associazione di popolazioni di due o più specie differenti, occupanti la stessa area geografica in un particolare tempo. Con *metacomunità*, si intende, invece, un insieme di comunità (di specie diverse) spazialmente interconnesse da meccanismi di dispersione tra habitat (concetto che verrà approfondito nel seguito). L'*ecologia comunitaria* o *sinecologia* è lo studio delle interazioni tra specie viventi in comunità, su differenti scale spaziali e temporali, includendo distribuzione, struttura, abbondanza e catena alimentare.

In particolare, l'ecosistema viene considerato come un tutto unico, del quale vengono studiate le dinamiche e i rapporti tra le comunità. L'*ecosistema*, in generale, viene suddiviso in due componenti principali:

1) *geocenosi* o *geotopo*, che è la sfera costituita dalle componenti non viventi (quelle ambientali di acqua, suolo, aria);
2) *biocenosi*, che è la comunità vera e propria, costituita dagli organismi presenti all'interno dell'ecosistema.

La definizione di *ecosistema* è, dunque, un insieme biologico costituito da organismi viventi interagenti fra loro e con l'ambiente che li circonda. Gli organismi facenti parte di un ecosistema sono organizzati secondo la cosiddetta *piramide trofica* (***Figura 2.1***).

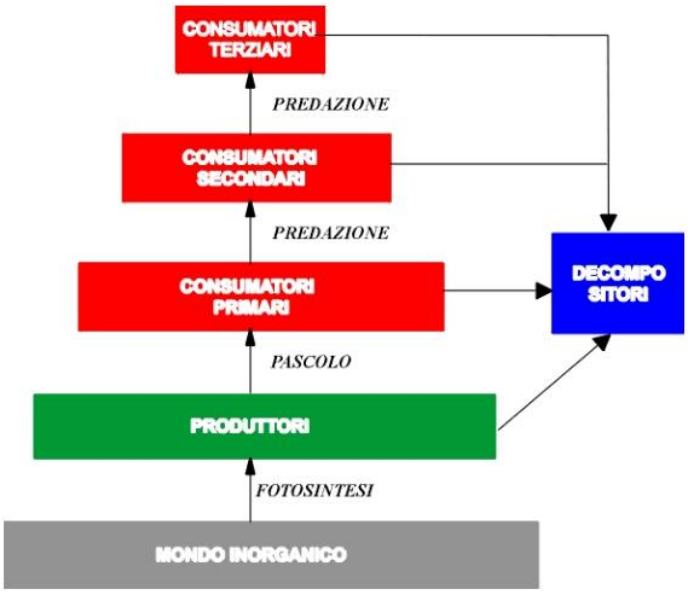

Figura 2.1 *– Struttura schematica della piramide trofica. [Immagine tratta da https://it.wikipedia.org/wiki/Sinecologia]*

Gli organismi autotrofi sono detti *produttori* e trasformano la materia inorganica in materia organica, attraverso il processo di fotosintesi. Essi introducono la materia organica all'interno della biocenosi. I *consumatori* sono gli organismi eterotrofi, ossia quelli che utilizzano come nutrimento altri organismi. Quelli che si nutrono di produttori, mediante il *pascolo*, sono detti *consumatori primari*. I consumatori che si nutrono di consumatori primari, mediante la *predazione*, si

dicono *consumatori secondari*. Analogamente si definiscono i *consumatori terziari* e così via. I resti e le escrezioni di tutti gli organismi vengono utilizzati dai *decompositori* che li remineralizzano.

Naturalmente, la complessità delle tipologie di interazioni possibili tra le comunità componenti un ecosistema non è riconducibile alla sola catena alimentare. Ci sono molti meccanismi e dinamiche diversi in gioco che regolano questi rapporti (che non dipendono solamente dalla mera interazione biotica e di competizione per le risorse). Per questa ragione, è necessario introdurre i concetti che stanno alla base della *teoria delle metacomunità*, focalizzandosi, ovviamente, sulla struttura e sulle meccaniche relative all'ecosistema fluviale.

Dapprima, verranno presentati i meccanismi principali che regolano questi equilibri (*fattori locali* e *regionali*). Poi, dopo aver dato una prima definizione di *metacomunità*, verrà fornita una panoramica più approfondita sui concetti che stanno dietro alla teoria stessa e sui principali modelli di spiegazione sviluppati.

2.2.1 Determinanti Locali e Regionali

Gli ecosistemi fluviali rappresentano un esempio lampante della complessità spaziale dell'organizzazione delle comunità all'interno dell'ecosistema. La singolare forma del reticolo fluviale è stata dimostrata, in studi passati, essere uno degli attori principali che influenzano le dinamiche delle comunità e i pattern di ricchezza locale in specie. Tuttavia, ciò non è sufficiente ad ottenere una spiegazione completa ed esaustiva. Studi più recenti sottolineano che la relativa importanza del reticolo fluviale regionale e le caratteristiche locali dell'habitat sia, in qualche modo, dipendente dal contesto, in funzione dei tratti caratteristici delle specie e dalle condizioni particolari di uso del suolo e interazione biotica.

Nonostante questi studi enfatizzino l'importanza di considerare sia *fattori locali*, che *regionali*, nello spiegare le variazioni osservate nelle comunità fluviali, il potere esplicativo generale fornito dalle teorie rimane tendenzialmente basso. Qui si arriva al nocciolo della questione, in cui bisogna distinguere tra *approccio classico* e *approccio moderno*.

L'*approccio classico* ha da sempre considerato i due meccanismi come separati e distinti, ognuno agente per conto suo in modo separato dall'altro e senza nessun legame con esso. Molte volte, inizialmente, i fattori regionali (come dispersione e fattori climatici) venivano del tutto trascurati, in ragione dei soli fattori locali (ambientali e di interazione biotica). Recentemente, non è stato più così, ma il metodo si è evoluto sotto forma di dicotomia tra i due meccanismi di influenza. In sostanza, viene considerata la relativa importanza di entrambi ma solo al fine di vedere quali siano i fattori principali che influenzano il processo, ignorando le relazioni di interdipendenza che nella realtà si vengono a creare. Questo modo di procedere considera, sì, tutti gli effetti, ma tralascia degli aspetti, altrettanto importanti, che si celano dietro alla catena di influenza diretta e continua tra determinanti locali e regionali. L'*approccio moderno* parte proprio da queste considerazioni. L'obiettivo è quello di estrarre informazioni anche su questi aspetti nascosti.

Ad esempio, un parametro come la quota non genera un effetto diretto sulla struttura di una comunità, ma influenzerà dei fattori locali (quali temperatura dell'acqua, o disponibilità di ossigeno)

che a loro volta influenzeranno le scelte delle comunità. Il suo effetto complessivo sulla struttura spaziale sarà, dunque, di natura *indiretta*.

È proprio la definizione di questa catena virtuale di effetti a cascata che permette di distinguere tra determinanti *locali* o *regionali*. Un determinato fattore sarà definito come *locale* se, nella catena degli effetti, esso provocherà un'influenza diretta sui processi biologici; un fattore sarà invece definito come *regionale* se influenza i processi biologici in maniera non diretta, ossia attuando il proprio effetto su fattori inferiori, nel senso della gerarchia degli effetti. Ovviamente, la scala di azione di un determinato fattore è importantissima nel definirne la natura. Un fattore agente a scala regionale sarà sicuramente di tipologia omonima. Ne consegue che anche variabili, che, a primo impatto, sembrano poco importanti e per nulla legate al processo che si sta studiando, possono invece agire da trasmettitore dell'effetto di altri fattori.

Lo schema è, dunque, il seguente: i fattori regionali influenzano i fattori locali che, a loro volta, agiscono sulle comunità (il loro strutturarsi nello spazio). Tutta la rete di influenze indirette si propaga all'interno della comunità per via dei rapporti di interazione biotica (competizione per le risorse, predazione e colonizzazione di nuovi territori).

Resta da capire quali siano i fattori principali che ricadono nelle due categorie e quali meccanismi caratteristici permettono di definirli e riconoscerli. I fattori locali, tipicamente, si identificano attraverso i due meccanismi di *influenza ambientale* (con la quale si intende tutta la rete di meccanismi di adattamento alle condizioni esterne) e di *interazione biotica* (ossia le relazioni di competizione, colonizzazione e predazione). Quelli regionali, invece, si riassumono nei meccanismi di *dispersione* (che dipendono sia da caratteristiche intrinseche che estrinseche alle specie) e di *interconnessione* del reticolo dendritico (in cui si inserisce anche il fattore della posizione delle comunità lungo la rete).

È già stato detto come, tradizionalmente, le condizioni ambientali locali siano state considerate il motore principale dei cambiamenti di comunità dei sistemi di acqua dolce. Adesso, è sempre più riconosciuto, nonché tenuto in conto nelle analisi spaziali, che i determinanti regionali (come interconnessione della rete fluviale e dispersione), in aggiunta ai filtri locali, contribuiscano a definire le comunità fluviali attraverso un vasto intervallo di estensioni spaziali. Il riconoscimento di dispersione e grado di connessione come fattori chiave è in perfetto accordo, come si vedrà, coi concetti espressi nell'ecologia metacomunitaria, che sottolinea fortemente come l'interazione tra dispersione e fattori locali contribuisca allo strutturarsi delle organizzazioni spaziali delle comunità. Ad esempio, un prerequisito fondamentale perché possa realizzarsi un efficiente filtro locale è che ci sia un grado di dispersione sufficiente affinché le comunità possano risentire delle differenze ambientali a più scale spaziali ed organizzarsi secondo le più adeguate.

È chiaro come il cercare di comprendere l'organizzazione spaziale delle comunità lungo una rete fluviale sia un'operazione tutt'altro che semplice. Essa dipende dall'interfacciarsi di numerosissimi meccanismi di natura diversa e di scala spaziale diversa. Prima di giungere al punto finale del capitolo, ossia alle vere e proprie teorie metacomunitarie, è conveniente trattare un po' più in dettaglio questi meccanismi di influenza locale e regionale.

2.2.1.1 Struttura dendritica: dispersione e interconnessione

Con l'evoluzione delle teorie ecologiche, gli ecologi si sono resi conto dell'impossibilità di non considerare i fiumi come strutture dendritiche per spiegare la biodiversità e le dinamiche generatrici di essa. Si tratta della particolare struttura ramificata dei sistemi fluviali (di forma arborescente, con un canale principale che si suddivide in rami via via meno importanti, procedendo verso monte), che può esercitare un forte controllo sulle dinamiche metacomunitarie, in particolare, regolando l'estensione e la quantità di dispersione all'interno del reticolo.

Figura 2.2 *– La particolare conformazione dendritica tipica dei sistemi fluviali.*

Con *dispersione* si intende, in generale, il movimento degli individui da un sito ad un altro e può essere considerata in due modi diversi: la dispersione *potenziale* e quella *effettiva*. La prima è quella legata alle caratteristiche fisiologiche delle specie, ossia a quei fattori genetici e morfologici che possono predisporre una determinata specie allo spostarsi all'interno dell'habitat in cui vive (ad esempio, avere delle ali di grosse dimensioni). Essa determina la suddivisione tra specie itineranti e residenziali. La dispersione effettiva, invece, dipende da una combinazione di caratteristiche intrinseche ed estrinseche, come il successo riproduttivo o la capacità di reclutamento (nel senso di aggiunta di ulteriori individui alla comunità) di una specie, piuttosto che le caratteristiche ambientali (che possono generare delle barriere allo spostamento anche per specie con un elevato potenziale di dispersione).

Al concetto di dispersione è direttamente collegato quello di *isolamento* all'interno della rete fluviale. Esso può verificarsi alle scale spaziali più piccole, quando il grado di interconnessione della rete fluviale non è molto elevato. È il caso tipico delle cosiddette acque di testa (*headwaters*), in cui la scarsa connessione con la restante parte della rete determina un'apertura minore all'arrivo di nuovi individui e pertanto un maggior grado di isolamento rispetto alle zone più a valle (più centrali, come posizione nella rete). Una diretta conseguenza è una maggiore componente di turnover della biodiversità in specie (la cosiddetta β-*diversità*) nelle acque di testa che non in quelle di valle, meno

ricche da questo punto di vista. In sostanza, c'è un maggior cambio di specie, tra siti limitrofi, nelle zone alte e più isolate del reticolo.

Risulta necessario fare una precisazione. In ecologia, la biodiversità viene suddivisa in tre componenti:

- α è la *diversità locale*, quella che tiene in conto del numero di specie in una ristretta area geografica;
- β è la *diversità di turnover*, quella che tiene in conto della variazione di ricchezza in specie da un habitat ad un altro;
- γ è la *diversità regionale*, che tiene conto del numero di specie in una regione (intesa come area che non include barriere alla dispersione delle specie).

Il valore di queste tre componenti rappresenta il prodotto finale dell'azione dei meccanismi descritti in precedenza. È chiaro, infatti, come tantissimi aspetti della struttura dendritica possano pesare sull'organizzazione della biodiversità, che a sua volta può estendersi a processi ecosistemici e trasmissione di malattie.

Bisogna, ora, cercare di capire il modo in cui la struttura dendritica si confronti coi meccanismi di dispersione. In generale, in letteratura, si identificano quattro modelli principali per descrivere l'interconnessione ecologica all'interno dei sistemi fluviali. Essi sono stati formulati, in principio, per giustificare la dinamica delle popolazioni, ma possono essere estesi alle metacomunità se accoppiati ad una comprensione della relativa dominanza dei vari tratti correlati ai fenomeni di dispersione delle specie componenti le comunità.

Il primo è il *modello gerarchico*. Esso afferma che la connessione tra località del fiume riflette la natura dendritica del reticolo, in modo tale che bacini di drenaggio maggiormente connessi dall'idrologia superficiale sperimentano scambi più frequenti di organismi. Secondo questa teoria, si assume che ci sia una minima dispersione fuori dal reticolo e, quindi, essa viene applicata in presenza di organismi caratterizzati da una pura (o preponderante) dispersione acquatica.

Il secondo è il, cosiddetto, *modello Death Valley*. Esso è stato sviluppato per quelle specie che sono acquatiche in tutte le fasi del ciclo di vita, ma che non sono tipiche percorrere grandi distanze in termini di dispersione, a causa dell'estremo isolamento di piccoli habitat acquatici (come nelle tipiche pozze sorgive della Death Valley), oppure per estrema specificità di habitat.

Il terzo è il *modello delle acque di testa*. Esso consiste, sostanzialmente, nell'opposto del modello gerarchico. Considera una maggiore connessione tra i rami più piccoli, senza curarsi della connessione idrologica dei bacini di drenaggio. In generale, questo modello viene applicato per quelle specie che si specializzano in certi tipi di habitat delle acque di testa che hanno anche qualche abilità nella dispersione non acquatica in delle fasi della loro vita.

L'ultimo è il *modello del flusso genico diffuso*. Esso predice una generale interconnessione tra tutti i rami del reticolo di una determinata zona (a scala, quindi, regionale). In queste condizioni, la dispersione avviene sempre, indipendentemente dalla connessione e dalla struttura dei rami della rete. Tale modello viene applicato per quelle specie con elevata capacità di dispersione aerea e con poca specificità di habitat.

Dalla descrizione dei quattro modelli, si capisce bene come il grado di connessione del reticolo abbia un peso diverso sulla risposta comunitaria, a parità di struttura, a seconda della tipologia e dei tratti caratteristici delle specie, legati alle modalità di dispersione. Dato il ventaglio enorme di specie differenti che compongono una metacomunità, i meccanismi e le direzioni di dispersione risultano essere altamente variabili. Ciò può avvenire anche all'interno della medesima specie, in fasi diverse della vita. Per esempio, esistono specie di insetti fluviali che, allo stadio larvale, sono limitati al puro trasporto passivo dato dalla corrente, lungo il reticolo, mentre possono disperdersi sia lungo la rete che fuori da essa allo stadio adulto (avendo subito la metamorfosi e sviluppato le ali).

Dunque, comprendere le modalità di dispersione delle specie può condurre ad una maggiore definizione di come il reticolo fluviale mantenga la funzionalità in termini di interconnessione delle comunità. La capacità e possibilità di tenere in conto dei cosiddetti *tratti caratteristici* (fisiologici e funzionali) di esse sarebbe un primo passo verso questo obiettivo. Si è visto, infatti, che caratteristiche morfologiche (come la forma del corpo, la dimensione delle ali) possano agire in maniera molto pesante sulla capacità di dispersione, che rimane però da correlare alla possibilità di dispersione (data anche da fattori estrinsechi alle specie, come la presenza di barriere topografiche).

Inoltre, la modalità preferenziale di dispersione (acquatica, terrestre o aerea) può rappresentare essa stessa un indice dell'importanza del reticolo e della sua connessione. Infatti, una specie puramente acquatica sarà molto influenzata dalla struttura del reticolo e dal grado di interconnessione (in quanto le sue modalità di spostamento sono limitate al solo alveo fluviale). Cosa non vera per una specie a dispersione aerea preponderante, per la quale la morfologia del reticolo risulta assolutamente ininfluente. Dunque, la struttura metacomunitaria di organismi acquatici tenderà a riflettere maggiormente la struttura reticolare del fiume (secondo un possibile modello gerarchico), mentre, per gli altri tipi di dispersori, sarà più realistico uno degli altri tre modelli.

Capire quale dei modelli si adatti maggiormente alla situazione obbliga a legare il grado di interconnessione ad un gradiente climatico. Infatti, la connessione idraulica tra i rami della rete non dipende esclusivamente dalla struttura morfologica del sistema fluviale, ma anche dalla disponibilità e quantità di acqua che scorre nella rete. In questo senso, un gradiente climatico che vada da arido a umido assume un forte controllo della connessione idraulica. In aree umide, i fiumi tenderanno a possedere una maggiore connessione longitudinale all'interno della rete (senza considerare, ovviamente, la presenza di barriere naturali o antropiche). Al contrario, le zone aride saranno caratterizzate da pozze isolate di acqua permanente. Da non trascurare sono le connessioni stagionali dovute ad eventi eccezionali di pioggia, che forniscono ai sistemi aridi un grado di connessione aggiuntivo (longitudinale, ma anche trasversale e fuori dalla rete, lungo la piana alluvionale circostante, nel caso di zone pianeggianti). Risulta, dunque, verosimile che l'ecoregione di appartenenza di una zona abbia un ruolo nel regolare l'influenza della struttura dendritica sulla dispersione, in termini di connessione idraulica tra i rami della rete.

Nello specifico, all'aumentare dell'aridità, la struttura del reticolo spiega sempre meno l'organizzazione spaziale delle comunità dovuta alla dispersione lungo il reticolo fluviale. Al contrario, i modelli che includono modalità di dispersione aerea ne sono meno influenzati (che si traduce nel fatto che le specie volanti si adattano più facilmente a condizioni climatiche più aride e con meno presenza d'acqua in alveo).

Un riassunto generale del discorso può essere osservato nel grafico di ***Figura 2.3***.

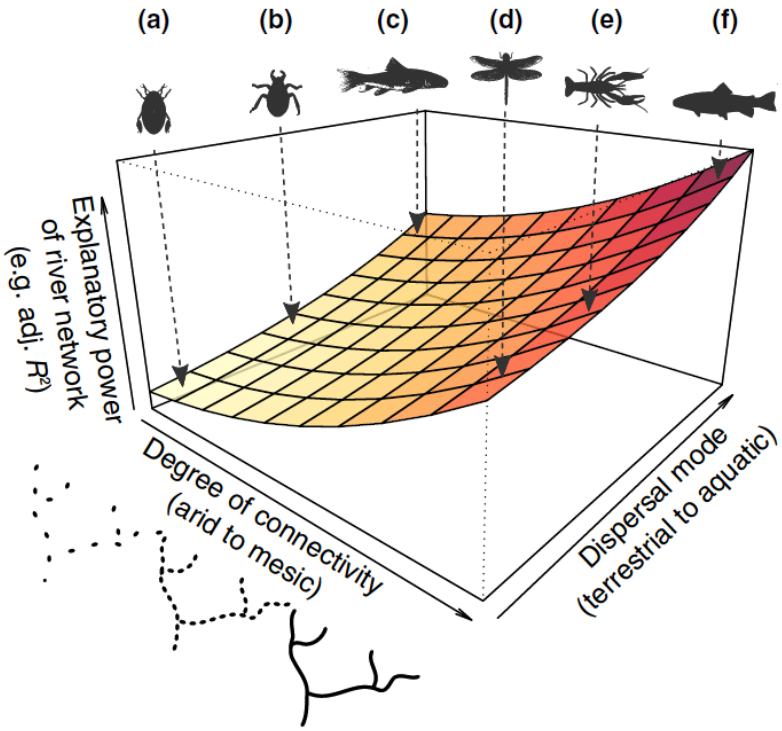

Figura 2.3 *– Relazione concettuale tra il grado di interconnessione idraulica nella rete (da sistemi aridi e umidi), la modalità di dispersione (che passa da terrestre ad acquatica lungo l'asse) e il potere esplicativo del reticolo fluviale nei confronti dei pattern di organizzazione della biodiversità (misurato mediante il coefficiente di correlazione R^2). In alto, sono indicate schematicamente le principali specie fluviali associate alle differenti parti del grafico (passando da invertebrati terrestri o volanti, fino ai pesci). [Immagine tratta da Tonkin et al., The role of dispersal in river network metacommunities: Patterns, processes, and pathways, 2017]*

Si nota come l'importanza del reticolo cresca all'aumentare del grado di umidità (quindi di interconnessione) e sia massima per i dispersori acquatici. In ambienti aridi, dove aumenta la frammentazione del reticolo (in termini di connessione), la rete perde di importanza e le modalità di dispersione tendono a quelle aeree. Ovviamente, si tratta di un grafico riassuntivo che semplifica molto le reali relazioni tra i tre parametri. Esso assume che ci sia una proporzionalità inversa tra connettività e modalità di dispersione (mentre può accadere, nella realtà, che specie con modalità di dispersione aerea seguano il fiume piuttosto che altri percorsi trasversali).

2.2.1.2 Fattori Ambientali

Con fattori ambientali si intende tutta quella rete complessa di caratteristiche del territorio in cui le metacomunità vivono e interagiscono. È noto, infatti, come diverse specie rispondano in modo diverso alle medesime inferenze esterne derivanti dall'ambiente circostante. Pertanto, la strutturazione di una comunità (o metacomunità) biologica è inevitabilmente definita dalle condizioni generali dell'habitat, siano esse di tipo fisico, chimico, climatico o geomorfologico.

Si tratta, in pratica, di entrare nel campo dei cosiddetti *determinanti locali*, ossia quelli che agiscono a scala spaziale più ristretta (rispetto all'intero reticolo fluviale regionale). Nella descrizione di natura ed effetti della dispersione, si era già accennato al discorso riguardante l'interrelazione tra quest'ultima e gli effetti ambientali. È stato detto come i due meccanismi non siano in rapporto dicotomico fra di loro, ossia non agiscono uno in modo indipendente dall'altro, ma sono inseribili in una catena di continua e reciproca influenza. Anzi, se non fosse presente questo interfacciarsi tra dinamiche a scale spaziali diverse, l'influenza complessiva dei fattori ambientali sarebbe molto meno determinante rispetto a quello che è (come si vedrà nel paragrafo relativo alla descrizione della *teoria delle metacomunità*). Affinché la loro azione risulti significativa nella determinazione dell'organizzazione spaziale di una comunità ecologica, I fattori locali necessitano di quelli regionali e viceversa (proprio per il discorso che le specie devono avere una potenzialità di dispersione tale da poter sperimentare condizioni ambientali diverse).

In questo senso, l'impostazione narrativa di questo paragrafo sarà indirizzata nella descrizione dello stretto rapporto di influenza che c'è tra ogni fattore di natura ambientale e struttura dendritica e dispersione.

La tipologia dei meccanismi di influenza ambientale può essere di varia natura. Esistono dei meccanismi fisici, che, nel caso di ecosistemi fluviali, si riferiscono alla branca dell'idraulica (la fisica dell'acqua e del suo movimento) e dell'idrologia (relazione tra l'apporto d'acqua atmosferico e le portate in alveo a seconda del clima e della stagionalità). Ci sono poi meccanismi chimici, relativi quindi all'apporto ionico in soluzione acquosa (derivante dall'interazione del fiume con la geologia della zona e l'evoluzione dell'uso del suolo). Infine, c'è un discorso molto recente relativo ai cambiamenti climatici e all'adattamento che le popolazioni devono attuare in risposta ad essi (cambi improvvisi di temperatura delle acque o dei regimi idrici).

Dunque, il primo fattore ambientale col quale le comunità fluviali devono fare i conti è sicuramente l'*acqua* e il suo *movimento* da monte verso valle. Il ciclo di vita di una comunità fluviale è imprescindibilmente legato alla quantità di acqua presente in alveo e, quindi, ai valori di portata, livello, velocità e alla tipologia di flusso. Ci saranno, infatti, specie acquatiche privilegianti i tratti a basse velocità del flusso, specie adattate a flussi di tipo turbolento o specie in cui la presenza o meno di acqua non influenza la loro vita o il movimento lungo la rete (come alcune specie di anfibi o di insetti volanti).

Un altro fattore fisico col quale le comunità si trovano ad interagire è la *morfologia* dell'habitat. In particolare, il movimento lungo il sistema fluviale (in termini di dispersione) è estremamente influenzato dalle caratteristiche topografiche della zona circostante. L'aumento della complessità topografica può infatti rappresentare un forte ostacolo al movimento sia lungo la rete che oltre essa (quindi sia per dispersori acquatici che aerei). Un caso tipico di questa situazione sono i torrenti himalayani, in cui l'elevata limitazione al movimento e alla dispersione delle specie è prevalentemente causato dalla presenza di vere e proprie barriere topografiche, rappresentate dalle montagne (e ciò accade nonostante la maggioranza delle specie presenti una preponderanza di tratti di dispersione aerea). Considerare la topografia sottolinea anche l'importanza dello scegliere un'adeguata scala spaziale di riferimento in un'analisi metacomunitaria. In caso di morfologie particolarmente complesse di un habitat, considerare un'estensione spaziale troppo limitata può condurre a conclusioni errate riguardanti i meccanismi di dispersione inter-habitat.

Data l'importanza della chimica delle acque nel discorso relativo all'adattamento delle specie, entra in gioco il fattore dell'*uso del suolo*. Gli effetti recenti dell'antropizzazione hanno, infatti, portato a dei cambiamenti sostanziali in composizione chimica dei terreni circostanti gli ecosistemi fluviali. Considerare, quindi, l'evoluzione dell'uso del suolo (da naturale ad agricolo o da naturale ad artificiale) può aiutare a capire i cambiamenti in proprietà chimiche delle acque, che derivano direttamente dalla natura del substrato su cui esse scorrono (temperatura, pH, contenuto ionico, ossigeno disciolto). Inoltre, il cambiamento di suolo determina anche un cambiamento in quantità e qualità di sostanza organica contenuta in esso, dalla quale molte comunità traggono nutrimento e sostentamento. Oltre a determinare un cambio nelle strategie di adattamento all'habitat, l'evoluzione di uso del suolo risulta avere un peso anche nella definizione della possibilità di dispersione (tipicamente quella aerea). Un esempio sono le zone di foresta. In territori altamente forestali, la dispersione aerea può essere nettamente limitata, rispetto ad una zona agricola aperta e pianeggiante. C'è da considerare anche il risvolto opposto, ossia quello in cui le specie necessitano della foresta e della sua copertura ombrosa. In questo caso, il passaggio da territori forestali ad agricoli avrebbe un effetto negativo sulla dispersione.

Anche se non è propriamente un fattore legato all'uso del suolo, l'effetto della *frammentazione della rete fluviale* risulta comunque avere un ruolo importante nel discorso dei fattori locali (e dell'antropizzazione degli habitat fluviali). Essa può avere infatti un profondo impatto sul modo in cui le specie possono disperdere lungo la rete. La frammentazione (sia essa naturale, come per le dighe dei castori, o artificiale, come negli interventi di sistemazione dell'alveo) può avere un ruolo analogo alla topografia, creando anch'essa delle barriere al movimento delle specie itineranti. In questo caso, però, ne risentiranno di più le specie a dispersione acquatica, in quanto quelle a dispersione aerea possono riuscire ad aggirare le barriere, attraverso altri percorsi di dispersione oltre la rete dendritica.

Come ultima tipologia di determinante locale, c'è il discorso relativo agli effetti locali della fascia climatica della zona e al *cambiamento climatico*. Da studi recenti, emerge fortemente l'interrelazione tra dispersione e cambiamenti climatici. Infatti, le specie caratterizzate da una minore capacità o tendenza alla dispersione risultano essere quelle più minacciate dagli effetti relativi ai cambiamenti climatici. In particolare, le specie costrette alla dispersione acquatica lungo la rete sono le più minacciate in assoluto. La capacità e velocità di reazione ai cambiamenti ambientali è, infatti, direttamente proporzionale alla possibilità di muoversi lungo i vari habitat presenti (e, quindi, alla dispersione). Per esempio, specie adattate a temperature fredde delle acque, in caso di aumento repentino di temperatura dell'acqua, potrebbero tendere all'estinzione nel caso in cui non riuscissero a cambiare habitat a causa di limitazione alla dispersione. Inoltre, l'effetto della crescente aridità dei sistemi fluviali può condurre ad una minore connessione idraulica tra i rami della rete e, dunque, a comunità sempre più frammentate nello spazio (più isolate e meno interagenti).

2.2.1.3 Interazione Biotica

Tra i fattori locali di determinazione dell'organizzazione spaziale delle comunità assume un ruolo estremamente importante l'interazione biotica, ossia quella rete di meccanismi e dinamiche che riguardano la vita delle specie in comunità. Essa, generalmente, si identifica nei tre rapporti principali di competizione per le risorse, predazione e colonizzazione di nuovi territori.

La *competizione*, in biologia, può essere definita come quell'interazione tra specie per cui l'attitudine di una è ridotta a causa della presenza dell'altra. Questa meccanica di influenza reciproca nasce dal fatto che esiste una limitatezza delle risorse disponibili in habitat. Essa può avvenire tra individui della stessa specie (e allora si parla di competizione intraspecifica) oppure tra specie differenti (competizione interspecifica). L'importanza di tale fattore è dovuta al fatto che molte volte può soppiantare il ruolo dei determinanti ambientali nella definizione della struttura spaziale di una comunità. Infatti, l'interazione biotica può determinare l'esclusione di alcune specie, in favore di altre, a prescindere da discorsi riguardanti l'adattamento alle condizioni ambientali dell'habitat. Dunque, il solo rapporto di competizione trofica può determinare la presenza o meno di una specie in un determinato territorio (proprio perché esiste una seconda specie che le occlude un adeguato accesso alle risorse).

Estremamente connesso a quest'ultimo discorso è il concetto biologico di *predazione*. Si tratta di un meccanismo di interazione biotica per cui un organismo di una specie usa come fonte di cibo un organismo di specie differente (sia in campo animale che vegetale). In ecologia comunitaria, essa assume un particolare ruolo all'interno della catena alimentare, in termini di regolazione della popolazione. Ciò che fa è tenere sotto controllo l'eccessivo sviluppo di prede e favorire la spinta evolutiva, portando allo sviluppo di comportamenti antipredatori. Si capisce, quindi, come il rischio di predazione abbia una grande influenza nel determinare il comportamento decisionale degli organismi (prede e predatori), inclusa anche la scelta di dispersione all'interno dell'habitat. Ad esempio, ricerche recenti hanno evidenziato la possibilità che l'elevata densità di pesci carnivori tra fiumi tributari abbia condotto le specie predate (di minore taglia) a diminuire il loro range di spostamento o, al contrario, abbia incoraggiato il movimento verso rami di ordine inferiore (più isolati). In sostanza, risulta molto importante capire la maniera in cui le prede alterano i loro equilibri di emigrazione-immigrazione all'interno della rete fluviale, in risposta ai predatori, al fine di progredire nella comprensione dell'organizzazione delle metacomunità fluviali.

Anche l'attitudine di una specie alla *colonizzazione* di nuovi habitat è in connessione con il concetto di dispersione. In generale, il successo della colonizzazione di un nuovo territorio permette di distinguere tra i due modi di intendere la dispersione visti in precedenza (quella potenziale e quella effettiva). Quindi, sarà l'attitudine di una specie alla colonizzazione ad essere indice di quella che è la dispersione effettiva di una specie. In particolare, le specie maggiormente colonizzatrici saranno quelle che richiedono una minore specificità di habitat, in termini di caratteristiche ambientali particolari. Specie che, quindi, si adattano facilmente e velocemente alle nuove condizioni esterne che trovano nel loro itinerario di spostamento.

2.3 Teoria delle Metacomunità

2.3.1 Concetto di metacomunità

È stato visto come l'ecologia comunitaria rappresenti un campo estremamente complesso per quanto riguarda la ricerca scientifica. Il suo obiettivo principale è la spiegazione dei pattern di distribuzione, abbondanza ed interazione delle specie in una determinata regione di analisi. Tali pattern possono ricorrere a differenti scale spaziali e variare a seconda della scala di osservazione, suggerendo che, al fine di comprendere in modo esaustivo il problema, possano essere applicati differenti metodi a differenti scale spaziali.

Tuttavia, gran parte della comunità scientifica, che si è occupata del problema fino a pochi decenni fa, ha focalizzato l'attenzione delle analisi ad una scala spaziale singola ed univoca, assumendo che le comunità locali fossero chiuse ed isolate. All'interno di tali comunità, le popolazioni erano immaginate interagire fra loro influenzando i tassi di nascite e decessi, come spiegano i classici modelli teorici di *Lotka-Volterra*. Si tratta di un sistema di equazioni differenziali non lineari del primo ordine che forniscono un modello matematico in grado di descrivere il rapporto di predazione tra due specie (una specie predata ed una predatrice).

Come è stato visto, però, nei precedenti paragrafi di questo capitolo, con l'evolversi della ricerca in tale ambito, ci si è resi conto che ci fossero molti altri processi ecologici che coinvolgevano specie e comunità differenti, a varie scale spaziali. Ad esempio, ci sono interazioni tra specie, che possono realizzarsi all'interno di una rete di comunità locali, dove influenzano i pattern di probabilità di colonizzazione ed estinzione a scale spaziali ben più grandi di quelle considerate tradizionalmente.

Le interazioni interspecifiche e la demografia sono, infatti, il prodotto dell'effetto di una vastissima rete di fattori interdipendenti oggetto di dinamiche a scala spaziale diversa, come, per esempio, si è visto con il grado di interconnessione della rete fluviale e la possibilità di dispersione al suo interno.

Tali dinamiche che coinvolgono interazioni tra comunità locali a scale spaziali multiple sono definite *metacomunitarie*. Si definisce, quindi, *metacomunità* un insieme di comunità locali interconnesse grazie a meccanismi di dispersione di multiple specie potenzialmente interagenti.

2.3.2 Definire una metacomunità nei casi reali

Il concetto di metacomunità risulta, dalla sua definizione, abbastanza teorico e non ha ricevuto moltissima attenzione da parte delle ricerche scientifiche (in campo), eccetto che negli ultimi anni.

È facile definire una comunità locale quando le specie interagiscono tra loro influenzando i relativi tassi demografici. Allo stesso modo, una metacomunità viene intesa come un insieme di tali comunità locali che si scambiano coloni di specie differenti.

Tale definizione impone che ci siano almeno due livelli discreti di integrazione comunitaria. A livello locale, è presente, in letteratura, una vasta gamma di ricerche riguardanti le interazioni interspecifiche (es: Lotka-Volterra) e conseguenti elaborazioni attinenti al concetto di struttura

trofica e organizzazione della catena alimentare. A scala regionale, invece, si verifica il fenomeno della dispersione in modalità e quantità difficilmente stimabili. Quando il tasso di dispersione è basso, gli effetti principali riguarderanno degli eventi di colonizzazione che regolano la storia dell'assemblamento delle comunità locali (e c'è un'ampia area della letteratura che si è occupata di questo). Quando tale tasso risulta alto, entrano in gioco anche altri fenomeni (come gli effetti di massa e i fenomeni di reclutamento, che verranno descritti in seguito). Essi modificano sia l'abbondanza delle specie che il loro modo di interagire, influenzando, di conseguenza, l'organizzazione spaziale della metacomunità.

Applicare queste definizioni di metacomunità a situazioni empiriche non è, però, così facile e immediato. Ci sono due evidenti problemi che saltano subito all'occhio. Il primo è che le comunità reali non sempre hanno dei confini discreti e netti. Il secondo è che specie differenti possono rispondere ai processi a differente scala spaziale.

In generale, si tende a suddividere i casi reali in tre differenti tipologie di situazioni:

i. assemblamento di patch discrete e permanenti;
ii. patch temporanee, distinte da un habitat matrice di sfondo, varianti in posizione e frequenza nel tempo;
iii. habitat permanenti con confini indistinti.

Il concetto di *patch*, che verrà ripreso nel seguito, definisce una qualche località con caratteristiche tali da permettere la vita di una comunità al suo interno.

Un esempio del caso *(i)* è quello di un insieme di isole oceaniche, con gli oceani che fungono da barriere alla dispersione (a differenti livelli a seconda delle specie considerate). La vasta letteratura sulla biogeografia insulare rappresenta un punto di partenza per trovare tanti potenziali esempi di metacomunità dove l'isolamento a certe scale spaziali ed in certi habitat è stato dimostrato essere fattore principale dell'organizzazione spaziale della biodiversità.

La categoria *(ii)* è rappresentata da assemblamenti che occupano delle habitat patch temporanee, che possono cambiare in posizione nel corso del tempo e che sono distinte da una matrice di sfondo. Le specie presenti in questi ambienti possono essere fortemente regolate dai tratti correlati alle dinamiche spaziali di dispersione. È il caso tipico delle metacomunità fluviali, in cui la variazione spaziale rappresenta un processo continuo e differenziato a seconda del grado di interconnessione lungo la rete (che varia da monte verso valle). Esistono, però, anche casi in cui le comunità possiedono continuità nello spazio, ma soffrono di frequenti disturbi che possono causare l'estinzione di un certo numero di comunità dominanti e alterare, così, gli equilibri metacomunitari precedenti.

La categoria *(iii)* rappresenta la più problematica di tutte, in quanto consiste in sistemi dove gli habitat tendono a rimanere costanti nel tempo e dove i limiti metacomunitari sono poco riconoscibili. Un esempio sono le barriere coralline, ossia habitat che contengono specie operanti a scale spaziali estremamente diverse a causa della mancanza di confini netti. In questo caso, anche la dispersione ha un ruolo poco chiaro, sempre per il fatto che opera su scale spaziali molto varie (da molto piccole a molto grandi). In tali sistemi, infatti, il livello a cui le dinamiche spaziali risultano importanti è variabile a seconda del grado di specializzazione in habitat, che influenza la percezione che gli organismi hanno dell'ambiente esterno e dell'isolamento spaziale.

2.3.3 Ecosistema fluviale come metacomunità

La *teoria delle metacomunità* descrive, dunque, dei processi che avvengono a scala di metacomunità e suggerisce dei nuovi modi di pensare alle interazioni tra specie. Essa studia come il fatto che le comunità locali siano inserite all'interno di un biota regionale più ampio influenzi e determini fenomeni locali e pattern di variazione di organizzazione spaziale.

Il pilastro fondamentale di questa teoria è il fatto che la struttura spaziale di una comunità non può essere compreso se si considerano solo i processi a scala locale. È necessario considerare processi a più scale spaziali e, per farlo, come già visto, si parte dal presupposto che esistano due forze principali che influenzano questi meccanismi: i fattori locali e i fattori regionali. L'interconnessione degli effetti generati da questi insiemi di forze genera i pattern spaziali che si realizzano all'interno dell'ecosistema. Non c'è nessun bilancio tra i due processi, bensì essi agiscono uno in contemporanea all'altro, influenzandosi reciprocamente (secondo lo schema di effetti a cascata descritto precedentemente).

Questi concetti si adattano, a livello teorico, molto bene ai *sistemi fluviali*. La struttura fisica dei bacini idrografici e della rete dendritica ivi presente crea dei collegamenti fra le scale spaziali attraverso il trasporto sia di organismi che di materiale. Infatti, le dinamiche di dispersione che avvengono all'interno della rete percorrono scale spaziali diverse tra loro a seconda del grado di isolamento del ramo fluviale e delle caratteristiche ontogeniche delle specie.

La seguente ***Figura 2.4*** evidenzia bene cosa significa applicare la teoria della metacomunità ad un sistema fluviale.

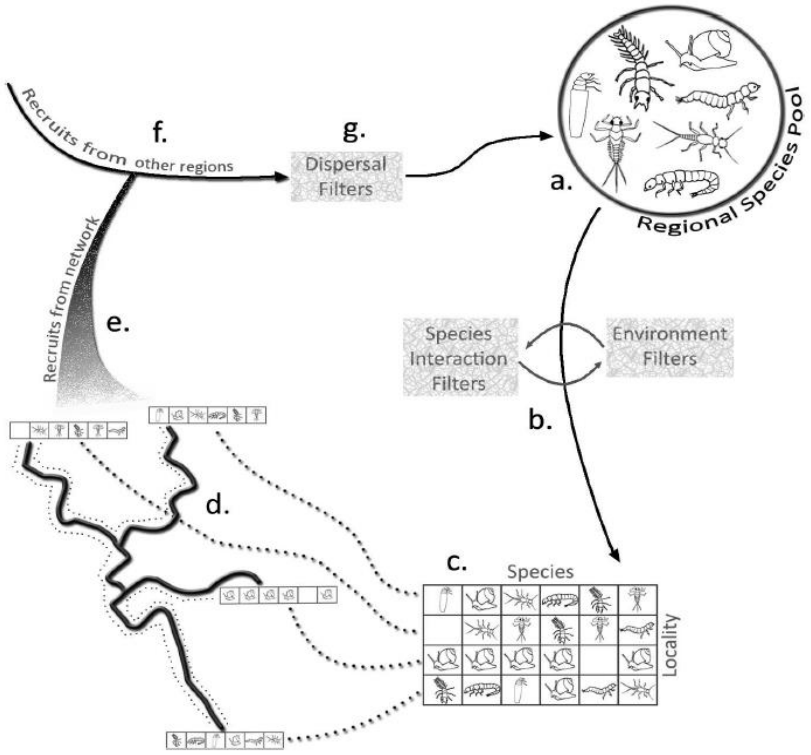

Figura 2.4 *– Interazione tra fattori locali (condizioni ambientali ed interazione biotica) e regionali (dispersione) nella determinazione della composizione locale delle comunità fluviali. [Immagine tratta da Brown et al, Metacommunity theory as a multispecies, multiscale framework for studying the influence of river network structure on riverine communities and ecosystems, 2011]*

In figura è stato riportato l'esempio dei macroinvertebrati bentonici, ma il medesimo concetto si può applicare in generale a qualsiasi metacomunità fluviale. Il significato della ***Figura 2.4*** può essere descritto per punti:

a. il punto di partenza è la *regional species pool*, ossia l'insieme regionale di organismi potenzialmente capaci di far parte di una comunità locale;
b. tale insieme di organismi viene filtrato dalle condizioni ambientali locali (in base alle tolleranze che caratterizzano ogni specie) e dalle interazioni biotiche interspecifiche;
c. si definisce una comunità locale, come un insieme di specie in un determinato habitat (in questo caso, ogni cella della tabella rappresenta un particolare microhabitat abitato da una particolare specie o lasciato inoccupato - celle bianche);
d. entra in gioco il ruolo del reticolo fluviale (interazione delle comunità per via della dispersione), dove le linee tratteggiate grosse indicano la localizzazione di ogni comunità all'interno del reticolo, quelle tratteggiate piccole indicano la dispersione;
e. reclutamento all'interno del reticolo, tramite la dispersione, di ulteriori individui (aggiunti alla comunità e, quindi, alla regional species pool);
f. reclutamento da altre regioni (esterne al reticolo), tramite la funzione dei dispersori aerei;
g. filtri alla dispersione, che possono essere rappresentati da barriere o limiti alla dispersione (possono esserlo per tutte le specie o solo per alcune), che fanno sì che non tutte le specie, quindi, determineranno un apporto alla regional species pool e alla biodiversità della zona.

La descrizione di questi processi come ciclica e ordinata è un'approssimazione. Ovviamente, le dinamiche reali all'interno di una metacomunità avvengono simultaneamente ed in maniera continua (non in fasi discrete e successive).

L'applicazione di questa teoria a comunità fluviali reali si è dimostrata essere molto efficace nel fornire nuove prospettive di interpretazione dell'organizzazione spaziale di tali comunità. Talvolta (come nel caso di uno studio fatto su una comunità di pesci del fiume Mississippi-Missouri) è stato evidenziato come i meccanismi regionali predominino su quelli locali (dispersione); altre volte sono le condizioni ambientali locali a strutturare la disposizione delle comunità (studio su macroinvertebrati bentonici lungo il bacino dell'Adige).

In generale, il risultato finale di applicazione della teoria dipende molto dalla scala spaziale di risoluzione dell'analisi (sarebbe a dire, il modo in cui vengono presi i dati e all'interno di quale estensione territoriale) e dalle caratteristiche delle specie presenti (i cosiddetti tratti funzionali).

Tuttavia, è possibile fare delle generalizzazioni teoriche. Studiando i comportamenti tipici delle metacomunità fluviali, i ricercatori sono giunti alla conclusione che esistono quattro meccanismi principali di strutturazione di una metacomunità. Essi sono noti come i *quattro paradigmi* della teoria delle metacomunità e sono i seguenti:

1. *patch dynamics*;
2. *species sorting*;
3. *mass effects* (*source-sink dynamics*);
4. *neutral theory*.

Essi verranno descritti nel prossimo paragrafo di questo capitolo.

2.3.4 I quattro paradigmi della teoria

2.3.4.1 Patch Dynamics

Con il termine *patch* (intraducibile, in senso biologico, in italiano) si intende, in generale, una località che possiede le caratteristiche adatte ad ospitare la vita di una popolazione o comunità. L'approccio, che si intende descrivere in questo paragrafo, assume che siano presenti multiple e identiche *patch*, in cui si verificano estinzioni sia stocastiche che deterministiche che possono essere influenzate dalle interazioni biotiche e contrastate dalla dispersione. Per descrivere queste dinamiche sono stati tradizionalmente usati due approcci.

Sovente, i modelli basati sulla *patch dynamics* si concentrano sul fatto che esse siano vacanti o occupate da popolazioni al loro equilibrio, in modo che ci sia una netta distinzione, in scala temporale, tra dinamiche locali e dinamiche regionali di colonizzazione-estinzione. Una limitazione molto spinta di questo approccio è che le patch sono assunte essere identiche. Una versione semplificata di tale modello considera solo la coesistenza locale in sistemi dove le specie competono per le risorse, senza prendere in considerazione alcun altro tipo di meccanismo di influenza locale.

Il secondo modello considera che, per metacomunità competitive in un ambiente omogeneo, la coesistenza risulta possibile solo grazie ad un adeguato bilancio tra capacità di competizione e capacità dispersiva. Esso considera anche come il rapporto di predazione influisce sulla persistenza regionale di una comunità nell'occupazione delle patch presenti. Un'elevata presenza di predatori in determinate patch può causare l'estinzione delle specie preda, generando una diminuzione del grado di dispersione grazie al quale era possibile la persistenza regionale. In questa situazione, le prede dovranno essere capaci di colonizzare le patch più velocemente della loro possibilità di estinguersi, mantenendo un certo grado di capacità dispersiva.

Ricapitolando, la prospettiva della *patch dynamics* considera una serie di habitat tutti uguali ed equivalenti tra loro, occupati da comunità locali, la cui composizione in specie è il risultato di una dinamica di bilancio tra competizione interspecifica e colonizzazione di nuovi territori.

2.3.4.2 Species Sorting

Questo approccio si fonda sulle teorie ecologiche che considerano i cambiamenti in composizione delle comunità causati dai gradienti ambientali esterni. Esso evidenzia l'influenza dei fattori abiotici sui cicli vitali delle specie e sulle interazioni interspecifiche. In questa prospettiva, la teoria della *species sorting* considera gli habitat presenti come eterogenei in qualche fattore (a differenza del paradigma precedente). Dunque, il risultato finale delle interazioni tra specie dipenderà dalle caratteristiche dell'ambiente abiotico.

Se specie diverse possono abitare solo habitat esclusivi tipici, allora non ha senso parlare di metacomunità unica, ma dovrà essere divisa in due indipendenti. Se invece, una stessa comunità può occupare differenti tipologie di habitat ci potrà essere una grande varietà di risultati dipendenti da interazioni a più larga scala (dispersione).

Una maniera per modellare queste dinamiche è considerare le teorie di assemblamento comunitario applicate a sistemi con multipli tipi di patch. Analogamente al modello di patch

dynamics, questo approccio assume che ci sia una distinzione in scala temporale tra dinamiche locali e dinamiche di colonizzazione-estinzione. Le popolazioni sono assunte essere al loro punto di equilibrio tra gli eventi di colonizzazione e quelli di perturbazione ambientali estrema che possono causare una repentina estinzione. Tale prospettiva si concentra sui bilanci tra specie che consentono loro di adattarsi alle condizioni ambientali presenti nelle varie patch.

Il paradigma metacomunitario della *species sorting* risulta avere molto in comune con le classiche teorie di separazione delle nicchie ecologiche e di coesistenza delle specie in un ecosistema. La differenza principale (e fondamentale) è dovuta all'importanza che viene data ai fattori regionali. Infatti, è grazie alla potenzialità di dispersione che le specie possono sperimentare cambiamenti ambientali a differenti scale spaziali lungo la rete fluviale (per esempio dovuti a perturbazioni locali o a cambiamenti graduali lungo le aste, da monte verso valle). In questo modo, si verifica la corrispondenza tra condizioni locali e composizione comunitaria.

La conclusione finale che si trae dall'applicazione di tali concetti è che l'organizzazione delle specie risulta essere estremamente collegata alle condizioni ambientali locali e largamente indipendente dagli effetti puramente spaziali (anche se, come si è visto, è sempre necessaria la presenza di un certo grado di dispersione perché le comunità riescano a risentire delle variazioni ambientali).

2.3.4.3 Mass Effects (Source-Sink Dynamics)

Se la patch dynamics e la species sorting partivano dal presupposto che ci fosse una netta separazione di scala temporale di azione tra dinamiche locali e processi di colonizzazione-estinzione, il terzo paradigma descrive come possa emergere una nuova importante dinamica regionale quando ciò non è più vero e, quindi, le dinamiche locali risultano pesantemente influenzate dalla dispersione. Tale teoria prevede l'instaurarsi di cosiddetti *effetti di massa*, dovuti alla dispersione, richiedenti che le differenti patch possiedano differenti condizioni ambientali (in un determinato tempo) e che possiedano un certo grado di interconnessione tra loro (fornito dalla capacità di dispersione delle specie). Il risultato è una cosiddetta relazione *source-sink* (sorgente-scarico) tra le varie patch, in cui le specie riescono a sopravvivere in ambienti *sink* (ossia con condizione esterne anche poco adatte alla loro vita) grazie alla costante e massiva emigrazione di individui (con alta capacità dispersiva) da ambienti *source* (quelli con condizioni ambientali privilegiate).

Il ruolo della dispersione risulta essere duplice (ed è dovuto al fatto che essa sia così elevata). Da un lato, l'immigrazione di individui può supplire al tasso di nascita locale e aumentare la densità di popolazione in comunità che potevano sembrare chiuse e isolate, mentre, dall'altro, l'emigrazione può aumentare il tasso di perdita di popolazione in comunità considerate più aperte all'arrivo di individui.

Al contrario del modello della species sorting, in cui venivano enfatizzate le differenze locali tra le specie (dovute alle relazioni di nicchia), in questo paradigma si vuole sottolineare come le differenze locali (pur sempre presenti) siano soppiantate da un maggior somiglianza a scala regionale (dovuta alla forte capacità dispersiva delle varie specie). Dunque, la coesistenza di tali specie è dovuta ad una compensazione regionale delle abilità competitive locali. Diretta conseguenza è che le specie risultano localmente diverse, ma regionalmente simili.

Tuttavia, esiste un limite al concetto di *mass-effects* ed è proprio dato dalla dispersione. La coesistenza delle comunità richiede un certo grado di abilità competitiva, ma non può esistere se la dispersione tra le patch è troppo elevata. Per gradi eccessivi di capacità dispersiva, infatti, il risultato finale sarebbe una omogeneizzazione della metacomunità. Gli effetti di massa, in questo caso, possono ridurre la coesistenza di specie diverse su scala regionale, con conseguente e parallela distruzione della diversità locale.

2.3.4.4 Neutral Theory

Tutti i precedenti approcci partivano dal fatto che esistesse una qualche differenza tra specie, in termini di relazioni di nicchia con i fattori locali o di abilità di dispersione. La dinamica che ne risultava era il prodotto dell'azione di entrambi gli effetti locali e quelli regionali. In assenza di tutte le differenze precedenti, la prospettiva metacomunitaria può cambiare.

Il paradigma della *neutral theory* vuole risolvere questo potenziale caso, che rimane escluso dai modelli trattati in precedenza. Esso intende spiegare le dinamiche che si instaurano all'interno di certe popolazioni composte da specie così simili tra loro da poter essere considerate equivalenti. Infatti, nonostante la mancanza di differenze in speciazione, si registra comunque una certa organizzazione spaziale di tali comunità.

Questa è dovuta al fatto che non è proprio vero che i processi di speciazione ed immigrazione siano totalmente assenti. Essi sono presenti ma con grado di manifestazione molto basso e lento nel corso del tempo. Sono presenti dei pattern casuali di cambiamento composizionale nel corso del tempo, prevalentemente ascrivibili a processi di stocasticità demografica. In questo senso, il lento manifestarsi di speciazione ed immigrazione sembra contrastare l'estinzione in tali comunità ecologiche composte da specie estremamente simili tra loro.

2.3.5 Conclusioni della teoria

Le situazioni reali, naturalmente, sono molto più complesse dei singoli meccanismi teorici appena descritti. Ciascuna di esse possiederà qualche caratteristica derivante da ognuna delle dinamiche descritte nei quattro paradigmi (qualcosa da una, qualcosa dall'altra). Tuttavia, è dimostrato dai risultati di numerose ricerche scientifiche (svolte partendo dai dati raccolti in campo), che cercare di assegnare un singolo paradigma come spiegazione della struttura spaziale di una particolare metacomunità non possa portare che contributi aggiuntivi, in termini di nuove prospettive di comprensione dell'ecologia comunitaria. In particolare, tali ricerche sottolineano come l'applicazione di questi concetti alle comunità fluviali sia particolarmente prosperoso in termini di risultati ottenibili.

La validità di tali affermazioni, ovviamente, dipende dalle conoscenze, che si hanno, riguardanti le caratteristiche di capacità dispersiva degli organismi, dalla scala spaziale di analisi e dal tipo di sistema studiato.

Il primo punto di attenzione riguarda la conoscenza della dispersione. Le abilità di dispersione degli organismi sono, generalmente, poco conosciute a livello di intere comunità. Questo si riflette,

inevitabilmente, sulla capacità della ricerca scientifica di studiare l'influenza dei fattori regionali nel regolare la vita delle comunità (presupposto fondamentale per l'analisi metacomunitaria). È questa la ragione per cui i ricercatori devono fare affidamento su parametri o espedienti metodologici che fungano da approssimazione del concetto di dispersione nello studio dell'organizzazione metacomunitaria (come il calcolo di vari tipi di distanza o di suddivisione delle specie sulla base di tratti dispersivi omogenei). In questo modo, essi possono quindi proseguire nelle analisi, anche se con una minore capacità risolutiva, avendo l'accortezza di considerare il concetto di *limitatezza alla dispersione*. Esso determinerà un peso nei metodi di approssimazione, che sarà tanto maggiore quanto una specie è poco propensa a disperdere.

La propensione alla *dispersione* è anche funzione della scala spaziale di analisi. In un contesto di organizzazione metacomunitaria, ci si aspetta che ci possano essere differenti aspetti relativi alla dispersione a differenti scale spaziali considerate (che, ovviamente, dipendono anche dalle capacità dispersive delle singole specie). Dapprima, si fa notare che, all'aumentare della distanza geografica tra siti (e dell'estensione spaziale di analisi), aumenta anche la *limitatezza alla dispersione*. Bisogna, però, considerare anche che l'estensione spaziale a cui ci si aspetta che si verifichi un fenomeno di questo tipo dipende dalle caratteristiche fisiologiche degli organismi coinvolti nell'analisi. Al diminuire della scala spaziale, invece, è ragionevole pensare che aumentino gli *effetti di massa* (terzo paradigma della teoria). Tale estensione sarà tanto maggiore, quanto lo è la propensione delle specie a disperdere.

Parte importante della teoria metacomunitaria è anche l'*eterogeneità ambientale*. Sia le prospettive dei *mass-effects* che della *species sorting* incorporano le differenze ambientali come presupposto della teoria, ma possiedono delle differenze sostanziali di ipotesi in termini di dispersione. Da un lato, secondo gli effetti di massa, un'eccessiva quantità di dispersione sovrasta gli effetti dell'eterogeneità ambientale (omogeneizzazione). Dall'altro, è necessaria la presenza di una quantità sufficiente di dispersione affinché le specie possano rilevare e sperimentare le differenze in cambiamenti ambientali. Un obiettivo fondamentale della teoria è dunque riuscire a separare le influenze della limitatezza alla dispersione e della sua intensità da quelle della species sorting, sull'organizzazione delle metacomunità.

Il discorso è riassunto sotto forma di grafico in ***Figura 2.5***.

In essa viene usato il parametro della scala spaziale come surrogato della dispersione (in pratica, il grado di dispersione è alto quando i siti sono vicini tra loro, mentre è basso quando sono lontani). A basse estensioni spaziali, un elevato grado di dispersione genera un elevato segnale spaziale (curva blu), a significare che siti vicini tra loro tendono all'omogeneizzazione data dalla dispersione (*mass-effects*). A scale spaziali intermedie, gradi ridotti di dispersione tendono a non omogenizzare più le comunità, ma permettono alle specie di sperimentare i cambiamenti in fattori ambientali tra i siti (*species sorting*). Ad estensioni spaziali massime, i valori molto bassi di dispersione conducono al fatto che le specie non riescono più a sperimentare tali cambiamenti in eterogeneità ambientale (*dispersal limitation*). Si entra, quindi, nel campo dei puri effetti spaziali (e assenza di condizionamento ambientale).

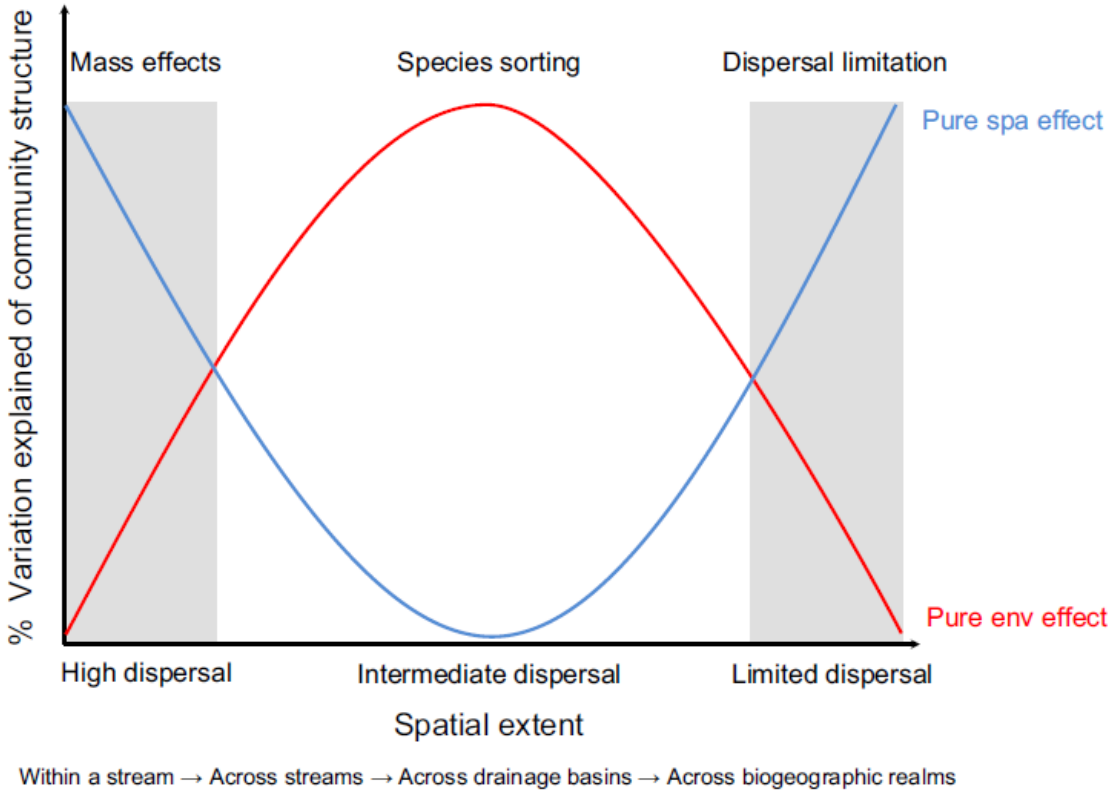

Figura 2.5 *– Una visione schematica della scala spaziale dei processi che influenzano l'organizzazione delle metacomunità. La curva rossa rappresenta l'andamento dei fattori ambientali (influenza locale), quella blu l'andamento di quelli spaziali (influenza regionale). [Immagine tratta da Heino et al., Metacommunity organisation, spatial extent and dispersal in aquatic systems: patterns, processes and prospects, 2015]*

Capitolo 3

Database Ecologico

3.1 Dati Disponibili

3.1.1 Struttura preliminare del Database

Il punto di partenza del lavoro affrontato in questa tesi è stato la fornitura, da parte del gruppo di ricerca del prof. Stefano Fenoglio del dipartimento di *Scienze della Vita e Biologia dei Sistemi* dell'*Università degli Studi di Torino*, di un database, in continua costruzione e aggiornamento, riguardante la mappatura della fauna bentonica nei maggiori corsi d'acqua piemontesi.

In particolare, l'intero lavoro di raccolta dei dati si riferisce all'attività per il monitoraggio degli effetti di centraline idroelettriche nei confronti dell'ecosistema fluviale. Non esiste un vero e proprio progetto a scala nazionale o regionale, che abbia dato vita a questa attività di sistematizzazione. Nonostante, dunque, essa sia stata svolta per committenti diversi e con tempistiche variabili, i dati sono stati ottenuti seguendo una metodologia di campionamento standard e ripetibile.

Il complesso delle misure effettuate (in termini di abbondanza di specie presenti) comprende informazioni che si estendono alla quasi totalità del Piemonte e ad alcune aree di Liguria e Lombardia. Tuttavia, in questa tesi, sono state analizzate solo le informazioni relative ai corsi d'acqua piemontesi, per maggiore uniformità territoriale.

3.1.1.1 Uso dei macro-invertebrati nel biomonitoraggio dei corsi d'acqua

Nel corso degli ultimi decenni, il monitoraggio biologico ha assunto un ruolo cruciale, nonché complementare alla tradizionale analisi dei parametri chimico-fisici, nella definizione dell'impatto antropico sui corsi d'acqua.

Nonostante ci siano tantissime possibilità in termini di metodi e organismi da studiare e campionare, una prassi che sta diventando sempre più comune è quella di riferirsi alla *fauna bentonica* come indicatore di salute di un fiume. In ecologia fluviale, con il termine *benthos* si intende il complesso degli organismi acquatici che trascorrono gran parte della loro vita sul fondale o a contatto con un substrato solido. Si tratta di un gruppo eterogeneo e ubiquitario di organismi relativamente facile da campionare e riconoscere.

Il vantaggio di usare questa categoria di macro-invertebrati nel valutare il grado di funzionalità e salute del corso d'acqua risiede nella loro differente sensibilità ai cambiamenti dei parametri chimici

della colonna d'acqua e di quelli fisici relativi all'habitat. Inoltre, la relativa lunghezza del ciclo di vita di molte specie e la loro costante presenza in località con caratteristiche simili rende l'analisi della loro struttura comunitaria uno strumento effettivo per valutare l'occorrenza di pressioni antropiche in anche lunghi periodi di tempo.

3.1.1.2 Metodo di campionamento AQEM

La metodologia standard a cui ci si è riferiti in precedenza prende forma a partire dalla direttiva europea *WFD 2000/60/EC* (*Water Framework Directive*), relativa alla valutazione dello stato di qualità delle acque di territorio europeo.

In questa prospettiva, viene così definito il metodo AQEM, una procedura di campionamento quantitativa volta a valutare lo stato ecologico di un corso d'acqua e basata su un approccio, cosiddetto, *multi-habitat*. Un definito numero di campioni viene estratto da differenti microhabitat a seconda della loro percentuale di occorrenza lungo la sezione fluviale precedentemente esaminata a livello visivo. L'aspetto quantitativo è un'esplicita richiesta della direttiva WFD, mentre l'approccio proporzionale multi-habitat è frutto della volontà di valutare il sito di studio in maniera affidabile e generalizzabile.

In sostanza, la procedura è divisa in due fasi distinte:

(i) valutazione a priori delle caratteristiche del fiume, del tipo di monitoraggio richiesto e della idroecoregione HER (concetto relativo all'area di appartenenza del fiume, in termini di morfologia e parametri chimico-fisici delle acque). A seconda di queste informazioni, viene definito il numero di campioni, l'area totale di campionamento e gli habitat differenti da campionare. Per l'importanza dell'aspetto quantitativo dell'analisi, viene richiesto un certo numero di campioni ripetuti (che varia in base al tipo di monitoraggio);

(ii) attività di campo, in cui viene dapprima definita (a livello visivo) la percentuale di occorrenza di ogni singolo microhabitat lungo l'area da campionare. Alla fine, vengono considerati solo quegli habitat con una percentuale di occorrenza maggiore del 10% (il consorzio AQEM è provvisto di una classificazione dei microhabitat tipici). Definite queste percentuali, il numero di campioni (ripetuti e non) vengono suddivisi sulla base dell'estensione spaziale dell'habitat considerato.

I campionamenti effettivi vengono effettuati mediante una *rete Surber*, un retino fatto ad imbuto montato su un telaio in alluminio che va posizionato sul fondo e con l'imboccatura controcorrente (così facendo i macroinvertebrati si accumulano all'interno del retino trasportati dalla corrente).

3.1.1.3 Matrice dei dati grezzi

La fase di raccolta dati in campo permette di identificare i macroinvertebrati presenti a livello di *famiglia*, oppure *genere* o *sottogenere* (ma solo per alcuni particolari individui). Si tratta di dati relativi all'abbondanza di ogni famiglia in corrispondenza del sito campionato.

Come detto in precedenza, il sito di campionamento è quasi sempre situato in corrispondenza di una centralina idroelettrica e per ognuna di esse sono stati effettuati più campioni (generalmente a monte/valle della centralina e talvolta in corrispondenza di essa, il cosiddetto campione sotteso).

Per ogni campione è stata registrata anche la posizione tramite GPS, l'idroecoregione di appartenenza del corso d'acqua, mese e anno di campionamento e comune e località in cui è stato effettuato.

Tutti questi dati sono stati raccolti nella cosiddetta *matrice dei dati grezzi*, in questo caso un file excel, riportante, lungo le colonne, tutte le informazioni ricavate durante il campionamento e, lungo le righe, ciascun sito campionato. In totale risultano 322 campionamenti (nei quali sono comprese misure ripetute nello stesso sito in diverse stagioni o diversi anni) e 160 colonne (di cui 10 di informazioni di natura geografica e 150 di abbondanza di famiglie, una per ogni famiglia).

Questa matrice dei dati grezzi non è nient'altro che il database preliminare, da cui e in funzione del quale sono state fatte tutte le analisi successive.

Fiume	Comune	Località	Stazione	Mese	Anno	HER	X	Y	COD4	ACROLOXIDAE
Agogna	Novara	Brida dei Cavalletti	S1 - monte	Marzo	2017	6	4530249	8343287	G	0
Agogna	Novara	Brida dei Cavalletti	S2 - valle	Marzo	2017	6	45291499	8344426	G	0
Agogna	Novara	Agognate	S1	Dicembre	2015	6	45291575	8344301	G	0
Agogna	Novara	Agognate	S2	Dicembre	2015	6	4528212	8352311	G	0
Agogna	Novara	Agognate	S3	Dicembre	2015	6	45281783	8351754	G	0
Agogna	Novara	Agognate	S1	Giugno	2016	6	45291575	8344301	G	0
Agogna	Novara	Agognate	S2	Giugno	2016	6	4528212	8352311	G	0
Agogna	Novara	Agognate	S3	Giugno	2016	6	45281783	8351754	G	0
Agogna	Novara	Agognate	S1	Marzo	2016	6	45291575	8344301	G	0
Agogna	Novara	Agognate	S2	Marzo	2016	6	4528212	8352311	G	0
Agogna	Novara	Agognate	S3	Marzo	2016	6	45281783	8351754	G	0
Agogna	Novara	Agognate	ST1	Settembre	2015	6	45291575	8344301	G	0
Agogna	Novara	Agognate	ST2	Settembre	2015	6	4528212	8352311	G	0
Agogna	Novara	Brida dei Cavalletti	S1 - monte	Ottobre	2016	6	4530249	8343287	G	0
Agogna	Novara	Brida dei Cavalletti	S2 - valle	Ottobre	2016	6	45291499	8344426	G	0
Bedale	Celle di Macra		MONTE	Agosto	2014	4	44282798	792259	G	0
Bedale	Celle di Macra		SOTTESO	Agosto	2014	4	44284171	793488	G	0
Bedale	Celle di Macra		VALLE	Agosto	2014	4	44285996	710112	G	0
Borbera	Cabella Ligure	Rosano	MONTE	Settembre	2018	10	4439170	906329	G	0
Borbera	Cabella Ligure	Rosano	VALLE	Settembre	2018	10	4440025	906047	G	0
Borbera	Cantalupo Ligure		S1	Dicembre	2016	10	4442225	903184	G	0

Figura 3.1 – *Struttura preliminare del database (le prime dieci colonne contengono informazioni relative alla geolocalizzazione della stazione di misura – comprese le coordinate x e y di latitudine e longitudine – e poi, dalla undicesima, iniziano ad essere riportate le misure di abbondanza delle famiglie biologiche).*

3.1.2 Dati progetto RENERFOR

A completamento delle informazioni ecologiche derivanti dal database appena descritto, sono stati utilizzati i dati creati in precedenza da RENERFOR, progetto di cui uno degli scopi è stata la valutazione dell'impatto dei prelievi idrici in alveo (per lo sviluppo di energia idroelettrica) sull'ecosistema fluviale.

In particolare, sono stati forniti dei dati idro-morfologici di partenza necessari per le successive analisi territoriali:

(i) file DTM del Piemonte (con una risoluzione di 100m);

(ii) reticolo idrografico del settore nord-occidentale dell'Italia ;

(iii) script per estrazione dei bacini idrografici e dei descrittori idro-morfologici dei bacini (che verranno trattati nel dettaglio in un paragrafo successivo di questo capitolo).

Il supporto di questi dati ha permesso di risolvere un problema di base del database in forma grezza, ossia che non fosse interfacciabile con le caratteristiche idro-morfologiche che differenziano i vari siti di campionamento (in termini di posizionamento lungo le aste fluviali sulle quali i dati reali sono stati presi).

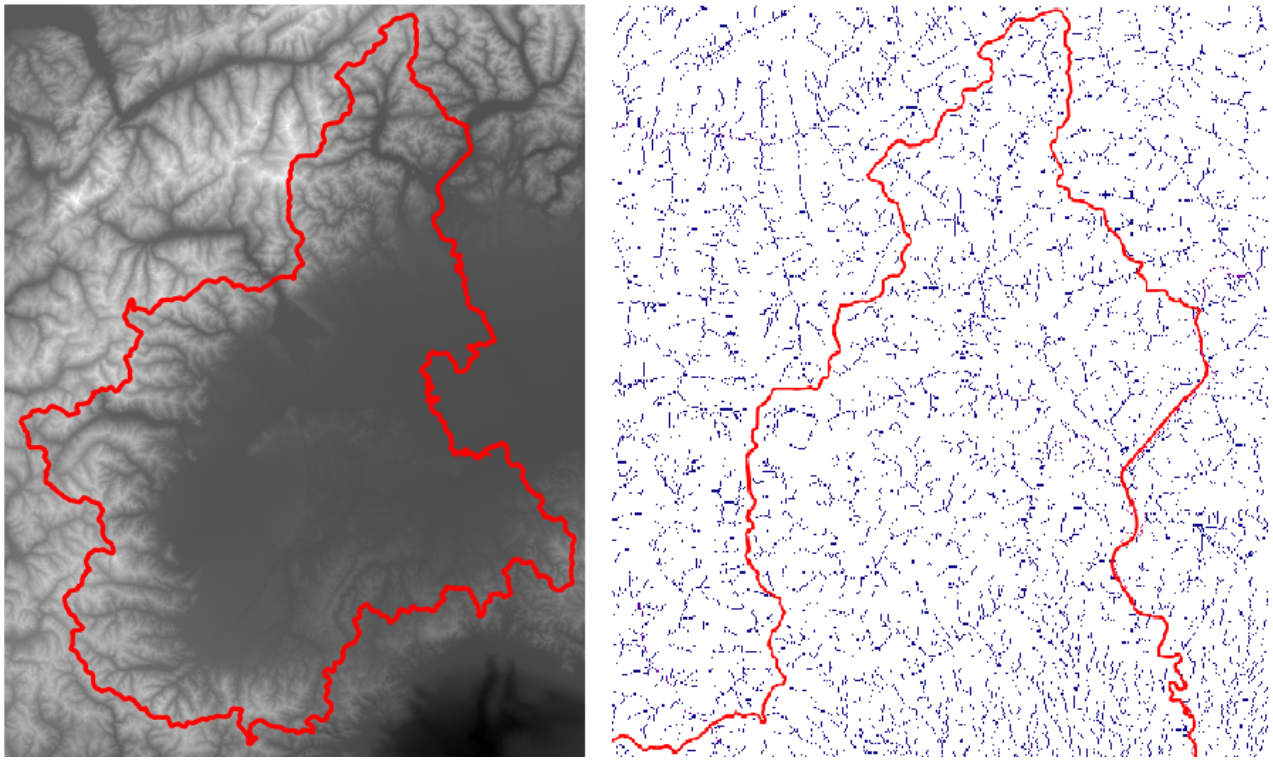

Figura 3.2 – *A sinistra: Modello Digitale del Terreno (DTM) comprendente la regione Piemonte. A destra: reticolo RENERFOR (zoom nella zona del VCO).*

3.2 Georeferenziazione del Database

3.2.1 Importazione dei dati in QGIS

Come è stato detto nel precedente paragrafo, una delle informazioni raccolte e inserite all'interno del database è la posizione geografica dei punti di campionamento. Poter localizzare correttamente nello spazio la posizione di un punto di misura è un'informazione fondamentale in campo ambientale. Infatti, è da questo presupposto che bisogna partire per poter poi eseguire correttamente e agevolmente le successive analisi spaziali e statistiche.

Le coordinate sono state registrate in campo mediante sistema satellitare e, successivamente, trascritte all'interno del database in forma di coordinate geografiche (longitudine, latitudine). Il formato di registrazione è stato quello di gradi-primi-secondi.

Per poter essere importate nel sistema informativo territoriale *QGIS* (software che permette analisi territoriali e spaziali) è stato necessario convertirle in un formato leggibile dal software (che non è quello di partenza del database grezzo). Dunque, le coordinate sono state convertite in formato decimale mediante una semplice proporzione matematica e poi importate in QGIS con sistema di coordinate geografico *EPSG:4326* (*Datum: WGS84*).

In ***Figura 3.3*** può essere apprezzata la disposizione spaziale dei siti di campionamento presenti nel database all'interno della regione Piemonte (come già detto, sono stati esclusi i pochi campioni presi in Liguria e Lombardia, in quanto in numero eccessivamente basso per poter eseguire delle analisi spaziali valide).

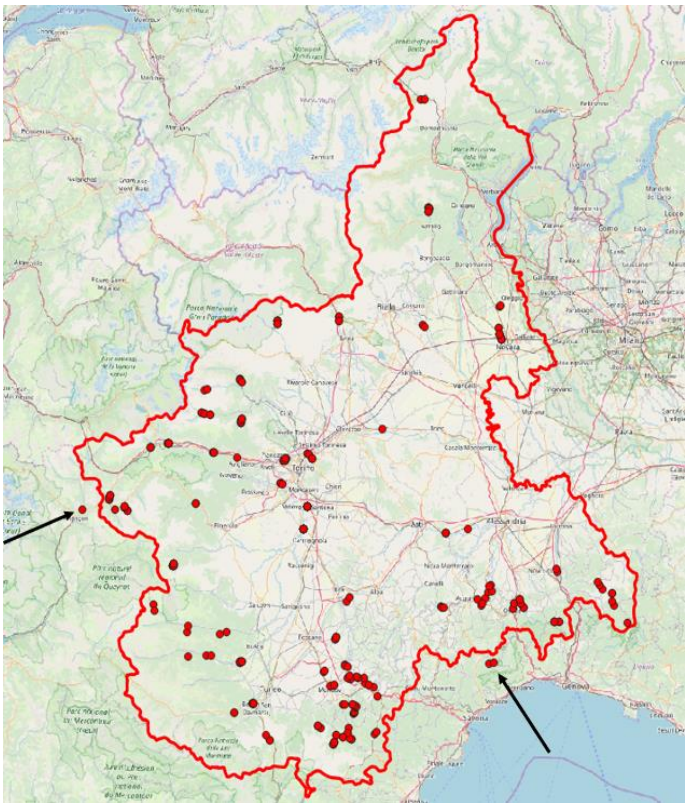

Figura 3.3 *– Siti di campionamento (rossi). Si notano i punti posizionati non correttamente (indicati dalle frecce nere).*

Si può notare come alcuni punti (per esempio quelli indicati dalle frecce), pur essendo stati registrati in località appartenenti al Piemonte, vengano posizionati sulla mappa esternamente ad esso oppure in posizioni non corrette (in questo caso, internamente al Piemonte ma in posti scorretti). Questo è un classico errore di trascrizione delle coordinate. I suddetti punti sono stati individuati e riposizionati manualmente nei pressi delle località indicate sul database.

3.2.2 Congruenza tra Database e Reticolo

Il passaggio successivo è stato verificare la congruenza nella posizione spaziale dei punti del database appena georeferenziato e le aste del reticolo fluviale fornito da RENERFOR.

Infatti, per una serie di problemi come la riproduzione non esatta dell'andamento del fiume da parte del reticolo digitale (per esempio nelle zone maggiormente meandrificate, come nel caso della Dora Riparia) oppure la trascrizione delle coordinate GPS lungo strade vicine (e quindi non esattamente dentro l'alveo), può risultare che la posizione del punto di campionamento non ricada esattamente dentro l'asta fluviale corrispondente.

Poiché per effettuare le successive analisi di estrazione dei bacini e dei parametri idro-morfologici è necessario che ogni punto ricada all'interno del reticolo (lungo le aste), è stato fondamentale riposizionare manualmente i punti collocati in maniera errata (cambiando il valore delle coordinate del file excel del database).

A seguito di questa operazione, si ottiene in risultato riportato in ***Figura 3.4***, in cui viene mostrato, sullo sfondo, il reticolo RENEFOR e, in primo piano, la posizione dei punti scorretti e di quelli aggiustati manualmente.

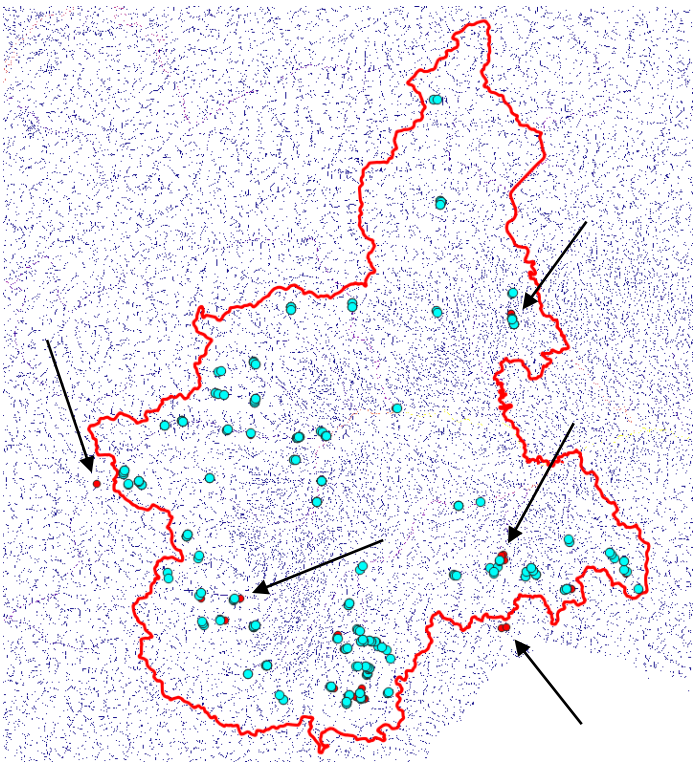

Figura 3.4 *– Database corretto. Sono riportati in azzurro i punti corrispondenti alle coordinate corrette e in rosso quelli corrispondenti alle coordinate non corrette. Si nota una certa discrepanza (frecce nere).*

3.3 Completamento del Database

Il database risultante dalle ultime operazioni può essere considerato completato dal punto di vista della georeferenziazione. Ogni punto di campionamento è correttamente inserito nella posizione giusta, sia in termini di località geografica che in termini di asta fluviale del reticolo complessivo.

Quello che manca adesso è l'aggiunta di informazioni quantitative di carattere fisico. Infatti, preso così come si mostra per il momento, oltre ad informazioni di abbondanza di famiglie di macroinvertebrati questo database non contiene nulla, o meglio, nulla di utilizzabile all'interno del discorso delle correlazioni e delle analisi spaziali a cui si è accennato in precedenza.

È stato dunque necessario ragionare su come aggiungere delle informazioni che andassero a completare il valore del database in funzione delle successive analisi statistiche.

3.3.1 Metriche Biologiche

Il database dispone di centinaia di dati di abbondanza delle famiglie di macroinvertebrati, che, presi ciascuno per sé, sarebbero molto difficili da trattare matematicamente. Un procedimento analitico molto conveniente è quello di accorpare tutti i dati in pochi parametri rappresentativi.

In particolare, in letteratura sono molto usate le quattro metriche ecologiche elencate di seguito:

(i) abbondanza totale di invertebrati (N);
(ii) ricchezza in "specie" (S);
(iii) abbondanza individui EPT (EPTN);
(iv) ricchezza in "specie" EPT (EPTS).

Tali parametri sono definiti come segue:

- N rappresenta la somma di tutti gli individui presenti in ciascun sito;
- S rappresenta il numero totale di specie per ciascun sito;
- EPT rappresenta la somma di tutti gli individui EPT per ogni sito;
- EPTS rappresenta il numero totale di specie EPT per ogni sito.

Per *EPT* si intende gli insetti acquatici appartenenti ai gruppi degli *Efemerotteri*, *Plecotteri* e *Tricotteri*, che, oltre ad essere gli invertebrati più sensibili alle modificazioni in habitat, rappresentano anche la componente faunistica più rappresentativa dei corsi d'acqua (specialmente quelli montani).

Questi quattro parametri riassumono in un valore unico le caratteristiche in termini di composizione faunistica di ogni sito del database. Saranno poi messi in correlazione, nelle successive analisi statistiche, con i parametri che descrivono il bacino in cui ciascun campione è stato preso.

Il valore di questi quattro parametri è stato calcolato su excel e per ognuno di essi è stata aggiunta una nuova colonna al database. I parametri N ed EPTN (corrispondenti ad un valore totale di individui) sono stati valutati mediante la funzione *somma()* di excel, mentre i due parametri di ricchezza in specie (S ed EPTS) sono stati valutati mediante la funzione *conta.se()*. Tutti e quattro i parametri sono stati calcolati per ogni riga della matrice (ossia per ogni sito campionato).

La disposizione spaziale in mappa del valore di questi indici è riportata nelle successive quattro figure, in cui il valore numerico dell'indice è direttamente proporzionale alla dimensione del punto.

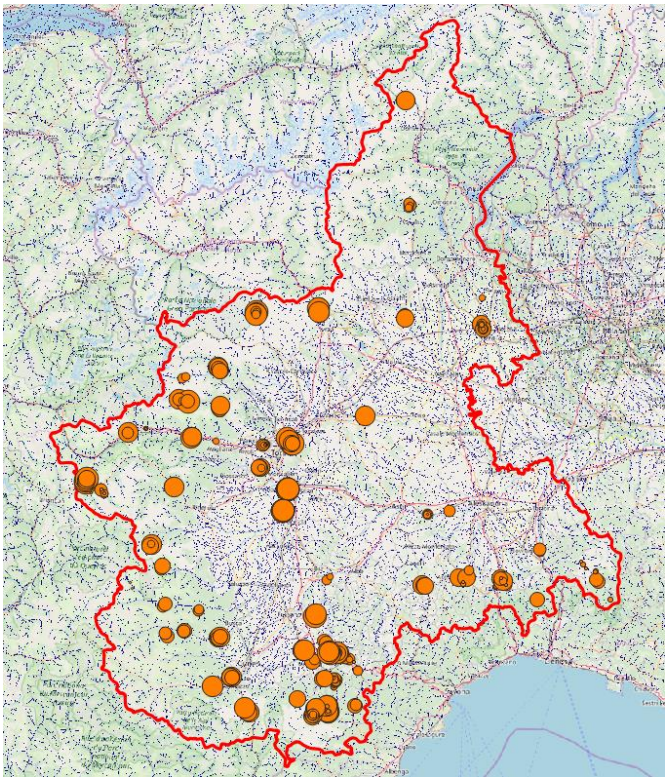

Figura 3.5 *– Abbondanza Totale Individui (N)*

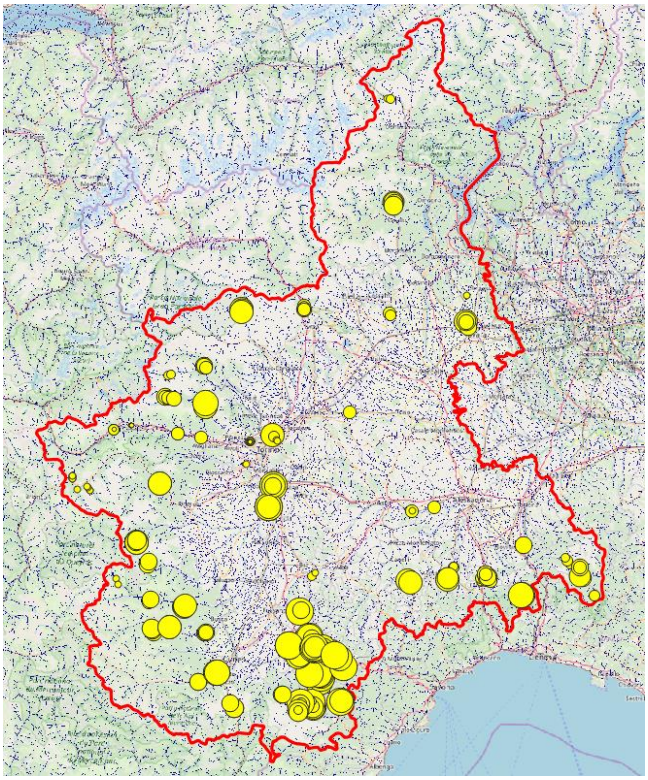

Figura 3.6 *– Ricchezza in Specie (S)*

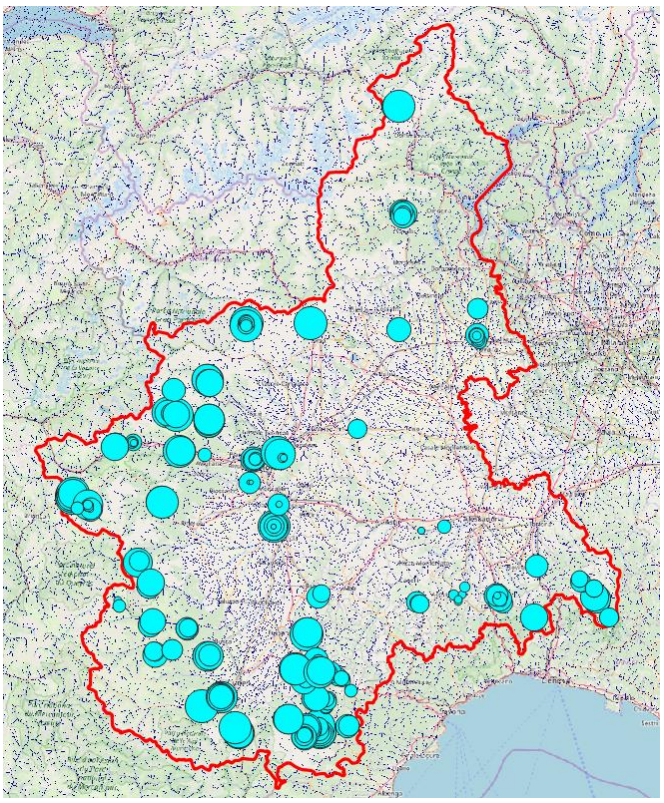

Figura 3.7 *– Abbondanza Totale Individui EPT (EPTN)*

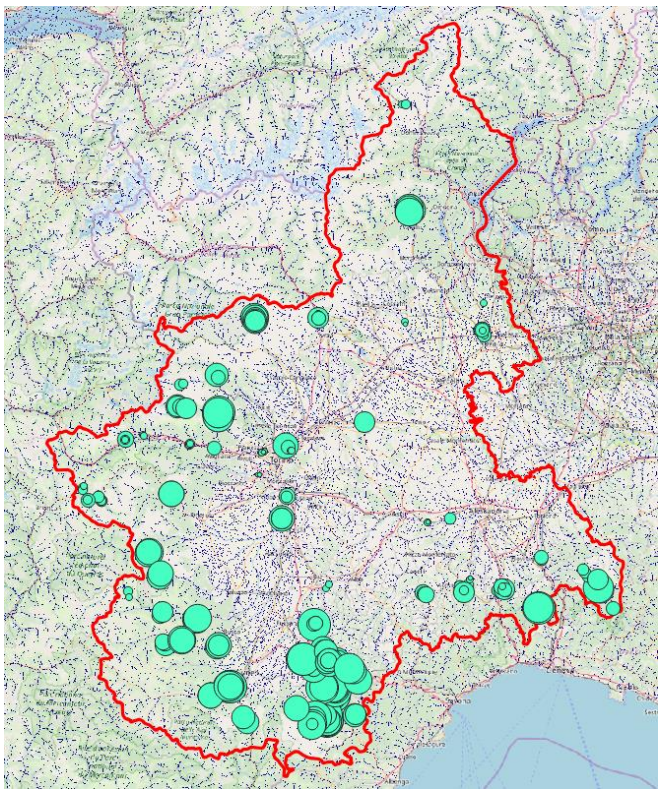

Figura 3.8 *– Ricchezza in Specie EPT (EPTS)*

3.3.2 Script GRASS

Il materiale di supporto ottenuto grazie al lavoro del progetto RENERFOR comprendeva, come già accennato, degli script utili per l'estrazione di ulteriori informazioni da inserire a completamento del database. Si tratta di due codici scritti in linguaggio GRASS il cui utilizzo ha previsto due fasi consecutive:

(i) utilizzo del primo script per l'estrazione dei bacini idrografici corrispondenti ad ogni sito di misura del database. Le coordinate delle stazioni devono essere inserite nello script, assumendo il significato di sezione di chiusura del bacino;

(ii) utilizzo del secondo script, che parte dai dati del primo (cioè dai bacini estratti), per l'estrazione di informazioni idro-morfologiche di ogni bacino.

Dunque, sulla carta, il problema di estrazione di informazioni risulterebbe risolto. È sufficiente fornire al terminale GRASS, creato dal progetto RENERFOR, le coordinate della sezione di chiusura per prima estrarre il bacino e, poi, fornire il nome dello shapefile del bacino generato al secondo script per generare un file di testo contenente le informazioni idro-morfologiche.

Tuttavia, ci sono un paio di problemi. Il primo è che il terminale è inserito all'interno di una macchina virtuale, che rende già di per sé più scomodo il lavoro di estrazione. Il secondo è che il database dispone di 322 campionamenti (che vuol dire altrettante sezioni di chiusura e shapefile di bacini da inserire manualmente, caso per caso, all'interno del terminale).

Per risolvere questi problemi, è stato necessario esportare i due script GRASS fuori dalla macchina virtuale, per poter lavorare più comodamente dal sistema ospite, e, soprattutto, automatizzare il processo di importazione di dati e conseguente esportazione dei risultati.

A tal fine, è stato fondamentale sfruttare il "gemellaggio" che il linguaggio di GRASS possiede con un linguaggio di più alto livello: *Python*. Più in generale, si è trattato di convertire il linguaggio dei due script nel cosiddetto linguaggio *Python-GRASS*.

3.3.2.1 Python-GRASS – Estrazione Bacini Idrografici

Come operazione base, ma fondamentale se si vuole che lo script funzioni correttamente, è stato necessario convertire tutte le coordinate delle stazioni di misura da WGS84 a UTM32N. Infatti, il primo codice legge queste ultime e da esse esegue tutta una serie di operazioni sequenziali che permettono di arrivare alla definizione del bacino.

Per effettuare questa conversione è sufficiente entrare su QGIS e utilizzare il comando di *Aggiunta Colonne Geometriche* del menu *Analisi Vettore*. Una volta aperto il comando si specifica cosa si vuole convertire, il sistema di riferimento di destinazione e il risultato è una nuova colonna della tabella degli attributi contenente le coordinate nel SR scelto.

Fatta questa operazione base, si è potuto passare alla conversione del codice. In questa parte di tesi si procederà in maniera estremamente didascalica, in modo che, se qualcuno volesse replicare la procedura eseguita, possa trovare in questo paragrafo tutte le informazioni utili per poterlo fare.

La prima cosa da fare è aprire GRASS (in questo caso si tratta della versione *GRASS GIS 7.6.1*).

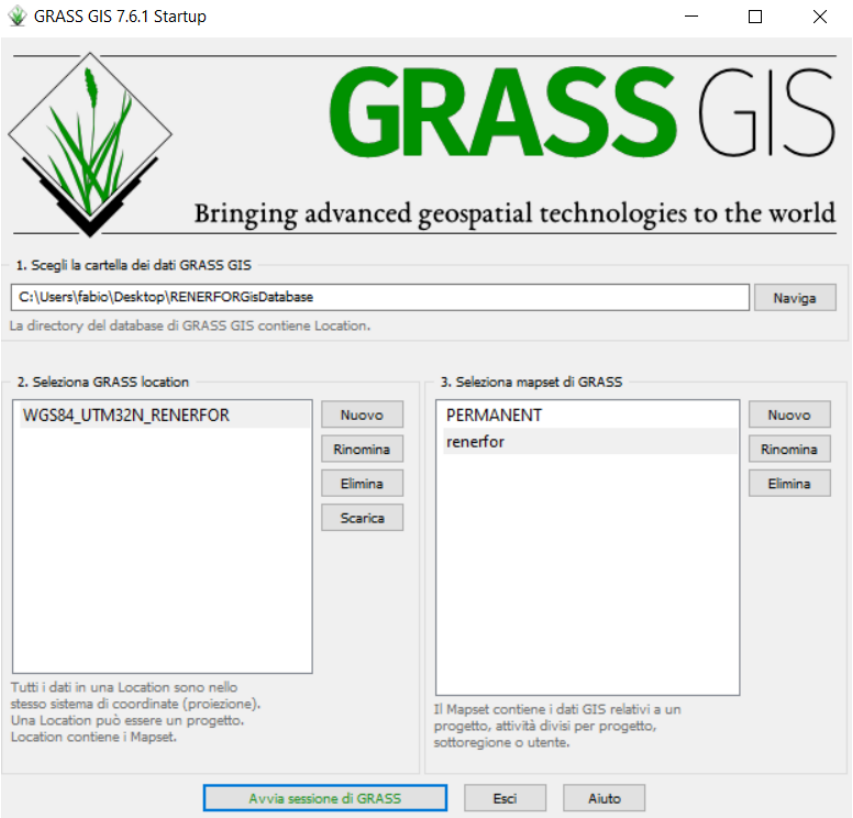

Figura 3.9 *– Configurazione iniziale di GRASS*

GRASS necessita di una sua unica cartella di lavoro, all'interno della quale salvare tutti i dati su cui si svolgeranno le successive elaborazioni. Questo spazio di lavoro di è composto a tre livelli:

(i) cartella dei dati GRASS;
(ii) location;
(iii) mapset.

La prima è la cartella globale, quella all'interno della quale devono essere inseriti tutti i dati su cui si intende lavorare. In questo caso, si tratta di una cartella già generata e strutturata dal progetto RENERFOR (che contiene tutti i dati raster e vettoriali di interesse).

All'interno di essa è possibile creare delle *Location*, ovvero dei sottoinsiemi dei dati, distinti ciascuno secondo il proprio sistema di riferimento. Se non ne fosse presente alcuna, sarebbe necessario crearla mediante il tasto *Nuova Location* e successivamente definire il sistema di proiezione. Solitamente, risulta conveniente importare questo SR da un file già esistente nella cartella principale (uno vale l'altro se, come in questo caso, sono tutti proiettati nello stesso SR).

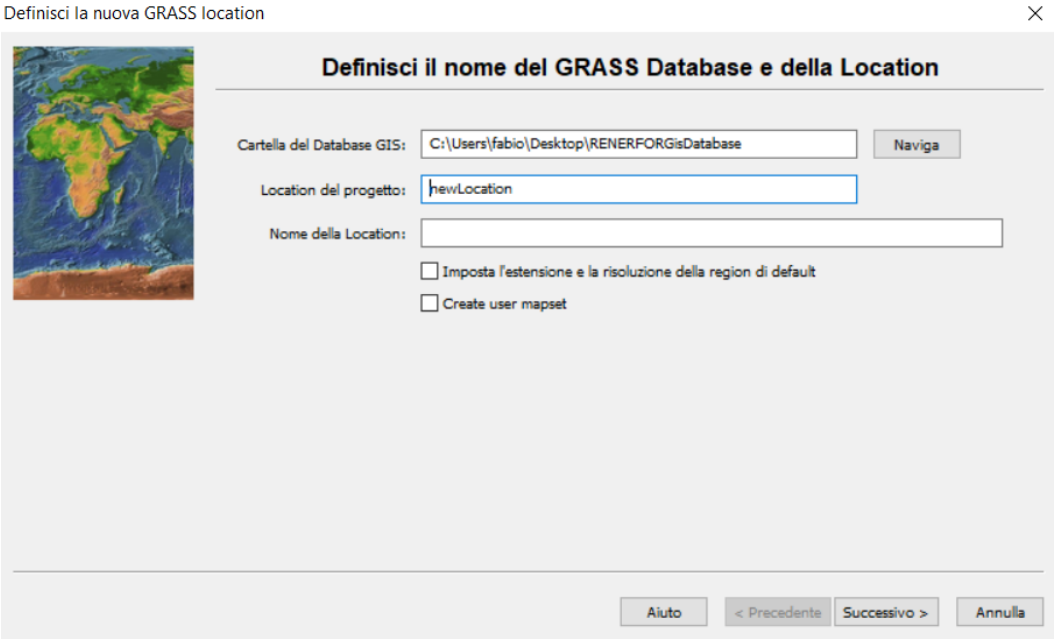

Figura 3.10 *– Creazione della Location*

Quindi, dopo aver dato un nome alla *Location*, si preme su *successivo* e si sceglie l'opzione *leggi proiezioni e datum da file georeferito*.

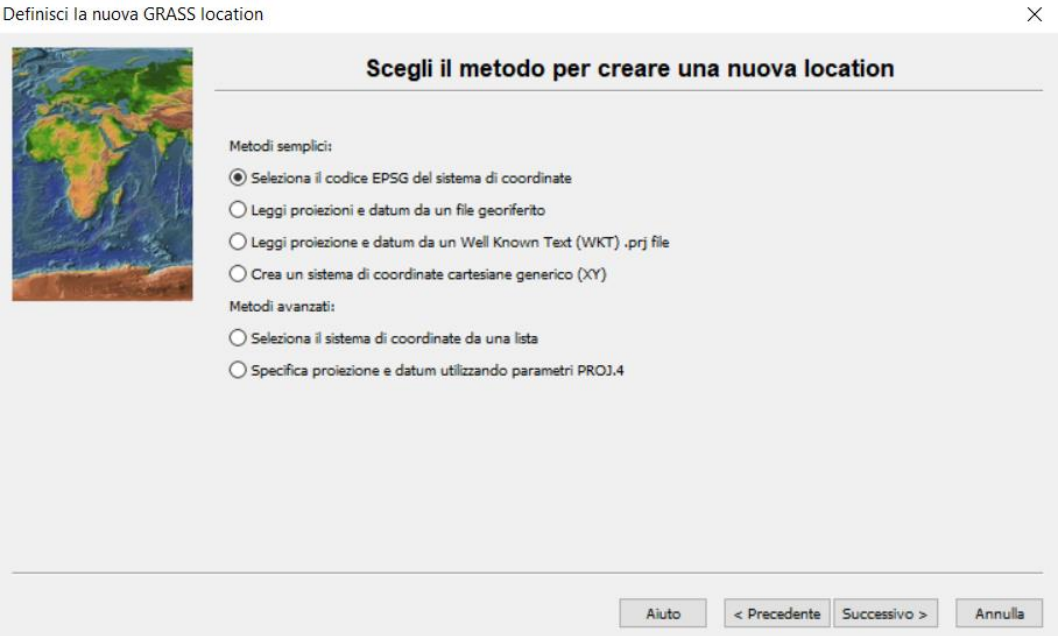

Figura 3.11 *– Creazione della Location*

Premendo ancora su successivo, la *Location* sarà visualizzabile sull'interfaccia iniziale di GRASS.

Nel nostro caso, la *Location* è già presente nella cartella iniziale di RENERFOR, e il sistema di coordinate creato in fase di definizione è visualizzato.

Si nota, dopo aver selezionato la *Location*, che sono presenti due *Mapset*, ossia gli spazi riservati alle elaborazioni dei dati. Il primo (*PERMANENT*) è quello che viene creato di default dal software, il secondo (*renerfor*) è quello del progetto RENERFOR ed è pertanto da selezionare.

Creata la struttura dei dati (e selezionati *Location* e *Mapset* corretti), si può avviare la sessione di GRASS mediante il tasto omonimo. Una volta avviata la sessione, si nota che l'interfaccia del software è composta da tre finestre iniziali: una *Console* di comando, una finestra di *Layer Manager* e una di *Map Display*. Ai nostri fini, di trascrizione di codici, è sufficiente usare la sola *Console*. È infatti da quest'ultima che sarà possibile lanciare i comandi e ottenere le elaborazioni volute.

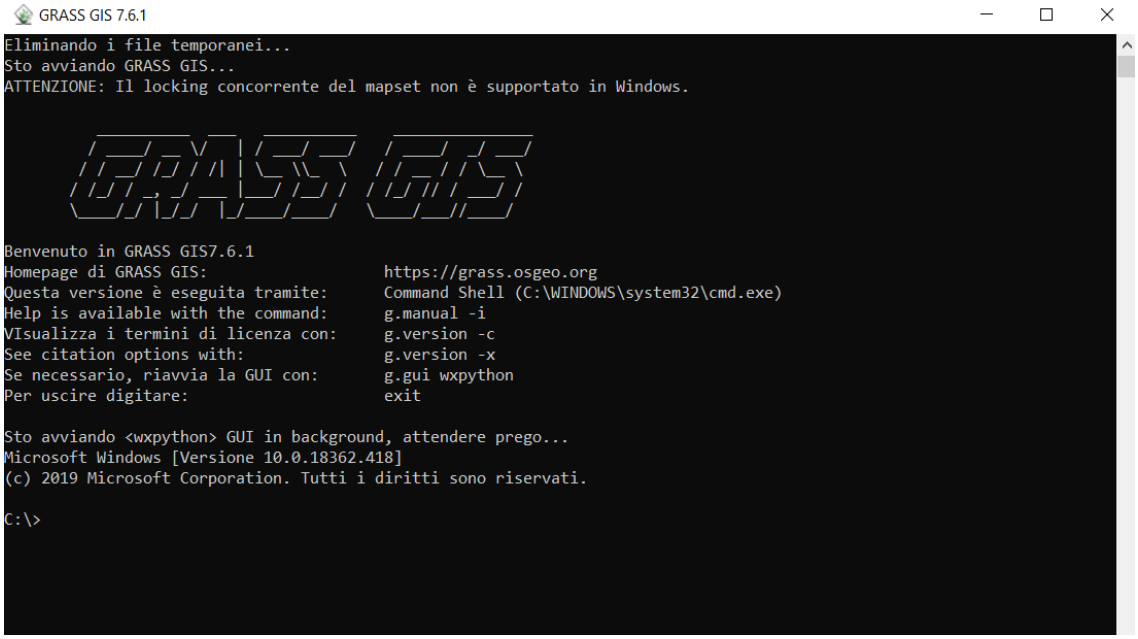

Figura 3.12 *– Console di comando di GRASS GIS*

Grazie a questa è possibile inserire tutti i comandi dello script, una volta che li si è trascritti in *python*.

La prima cosa da fare adesso è cambiare la directory di lavoro. In particolare, bisogna dire a GRASS che tutti i file di lavoro saranno inseriti nella stessa cartella, che è quella del Mapset appena creata (*renerfor*). Quindi si procede al cambio di directory attraverso il seguente comando

cd Percorso/Cartella/Mapset

Questo passaggio è fondamentale per il corretto funzionamento dello script finale.

I prossimi passi, che consistono nel commento dello script python, saranno scritti sotto forma di elenco per maggiore leggibilità (script che viene messo per intero come *Allegato 1* alla fine di questa tesi).

Primo passo: dire al software il tipo di linguaggio in cui si intende programmare. Siccome, si vuole scrivere in python, è sufficiente digitare ***python*** in console e premere invio.

Secondo passo: importare le librerie python per programmare attraverso i comandi GRASS. Ciò si fa digitando i due comandi seguenti uno dietro l'altro e premendo invio alla fine

```
from grass.pygrass import raster
import grass.script as gscript
```

Si tratta di due librerie che consentono di usare i comandi GRASS attraverso il linguaggio python (il primo per importare dei file raster in python, da cui lo script parte per fare le sue elaborazioni, il secondo per tradurre in python qualsiasi comando GRASS).

In particolare, un comando fondamentale per eseguire l'operazione di "traduzione" in python-GRASS è il ***g.script.parse_command***, che si vedrà essere tema estremamente ricorrente nel codice. In sostanza, esso permette di passare a python un comando esterno al suo linguaggio (che è, per l'appunto, un comando appartenente a GRASS) e di fare comunque eseguire tutte le operazioni che farebbe in GRASS.

Terzo passo: convertire i dati iniziali (excel) in formato txt. Questo permette un'importazione più agevole all'interno dell'ambiente py-GRASS. Non è da convertire tutto il database. È sufficiente creare un file txt con le coordinate UTM32N (come prima colonna si copia la coordinata est, come seconda la coordinata nord) e uno con i nomi dei fiumi, che poi servirà per dare il nome allo shapefile del bacino estratto (importante che il nome dei fiumi sia senza spazi). In questo caso, i due file di testo creati sono stati chiamati "Data" e "Fiumi_nome" (e verranno direttamente richiamati nello script mediante comandi appositi, come si vedrà nel prossimo step).

Quarto passo: ciclo per importare in python i dati da file esterno (testo). Si definiscono due vettori vuoti **coord_E** e **coord_N**, che saranno riempiti riga per riga ad ogni giro del ciclo. Ciclo che legge le coordinate da file txt e le inserisce all'interno dei due vettori (la prima colonna, con le coordinate est, nel primo vettore, la seconda, con le coordinate nord, nel secondo). Stesso procedimento per l'import dei nomi dei fiumi, solo che in questo caso c'è un'unica colonna.

```
coord_E = []
coord_N = []

with open('Data.txt', 'r') as f:
n_bacini = f.readlines()
for x in n_bacini:
raw = x.split()
coord_E.append(float(raw[0]))
coord_N.append(float(raw[1]))


fiume = []

with open('Fiumi_nome.txt', 'r') as f2:
N = f2.readlines()
for y in N:
righe = y.split()
fiume.append(righe[0])
```

Il comando **f.readlines** legge tutte le righe del file txt in ingresso al ciclo. L'inserimento delle coordinate lette all'interno dei vettori avviene attraverso la funzione **append** da posporre al nome del vettore in aggiornamento. Il comando **split** divide la stringa in ingresso in una lista lunga quanto il numero totale di elementi nella stringa (quindi quanto il numero di coordinate estratte), e il comando **float** converte la stringa in un numero decimale, permettendo di comporre il vettore numerico in uscita dal ciclo.

Quinto passo: definire il ciclo di estrazione dei bacini (che vada ad ogni giro a prendere le coordinate di un sito ed estragga il bacino corrispondente). È conveniente usare un *ciclo for* lungo quanto la dimensione dei vettori appena creati

```
for bacino in range(0,len(n_bacini)):
bacino
```

Il termine **bacino** è l'indice che va da 0 al numero totale di bacini (322).

Sesto passo: inserire dentro al ciclo la sequenza di comandi da eseguire per l'estrazione completa di un bacino ad ogni giro. Questo si fa traducendo i comandi GRASS attraverso il già visto comando ***g.script.parse_command***.

```
for bacino in range(0,len(n_bacini)):
bacino
gscript.parse_command('g.region', raster='piemonte_drain_r100', res=100)
gscript.parse_command('r.water.outlet', input='piemonte_drain_r100', output='Bacino_' +str(bacino),
coordinates=str(coord_E[bacino])+','+str(coord_N[bacino]), overwrite=True)
gscript.parse_command('g.region', zoom='Bacino_' +str(bacino), res=100, overwrite=True)
gscript.parse_command('r.mask', raster='Bacino_' +str(bacino), maskcats='1', overwrite=True)
gscript.parse_command('r.to.vect', input='Bacino_' +str(bacino), output=fiume[bacino]+'_'+str(bacino),
type='area', overwrite=True)
gscript.parse_command('v.out.ogr', flags='c', input=fiume[bacino]+'_'+str(bacino), output='shp', type='area',
format='ESRI_Shapefile', overwrite=True)
gscript.parse_command('r.mask', flags='r')
```

La sintassi del comando è la seguente: ***gscript.parse_command('nome comando GRASS da tradurre', input1=nome1, input2=nome2 …)***. In sostanza, si passa prima il nome del comando GRASS da tradurre (inserito tra apici) e poi tutti i parametri richiesti in sequenza (come nome del file raster da elaborare, valore di risoluzione, nome del file di output uscente dal comando etc.). In particolare, è importante inserire tutti gli input al comando tra due apici, in quanto la sintassi richiede input sotto forma di stringa (eccetto che per i valori richiesti in formato numerico, che possono essere inseriti senza modifiche). Si può notare come, nei comandi per l'esportazione, il numero progressivo da inserire in coda al nome del file in uscita sia convertito in stringa (**str()**).

Le operazioni che vengono fatte ad ogni ciclo sono le seguenti:

- zoom sulla regione di interesse (***g.region***), a cui si passa il file col reticolo di drenaggio;
- definizione del reticolo di drenaggio (***r.water.outlet***), che esporta il file col reticolo di drenaggio zoomato sul singolo bacino (corrispondente alla sezione di chiusura data dal ciclo). Le coordinate della sezione di chiusura sono passate al comando *coordinates*;
- maschera sul singolo bacino (***r.mask***), per diminuire la quantità di pixel da convertire col comando successivo;
- conversione da raster a vettore (***r.to.vect***);
- esportazione del vettore del bacino appena creato come shapefile (***v.out.ogr***), a cui si dà il nome del fiume (dato dal giro del ciclo a cui si è) seguito da un numero progressivo. Importante definire la cartella di destinazione (in *output*) e il formato di esportazione (*format*).

In ogni caso, una spiegazione completa e dettagliata di ogni comando e dei suoi vari input può essere trovata nel manuale online di GRASS.

Il risultato finale dello script così scritto è l'estrazione, per ogni sezione di chiusura passata al ciclo, del bacino idrografico corrispondente in formato shapefile (che può essere poi inserito su QGIS per ulteriori analisi).

Lo script intero è riportato come *Allegato 1* alla fine della tesi. Le varie cartelle utilizzate, contenenti i dati di input e successivamente elaborati, sono inserite, invece, in un *CD* allegato all'elaborato.

3.3.2.2 Python-GRASS – Estrazione Descrittori dei bacini

Il procedimento di conversione in linguaggio python appena effettuato per lo script GRASS di estrazione dei bacini è stato poi ripetuto per il secondo script fornito da RENERFOR, quello di estrazione dei parametri descrittori di ogni bacino.

Si tratta dei classici parametri idro-morfologici che vengono calcolati nelle analisi regionali dei bacini per la definizione delle curve di durata delle portate (CDP). In particolare, il lancio dello script permette di calcolare valori di quota, valori di precipitazione media annua (MAP), valori dei coefficienti delle curve di possibilità pluviometrica (IDF) e valori statistici relativi al regime pluviometrico della zona (coefficiente di Fourier e coefficiente di variazione del regime pluviometrico).

Si discuterà nel dettaglio il significato e l'uso di queste variabili nel *Capitolo 4* di questa tesi.

Come in precedenza, si procederà al commento di questo secondo script per punti, per una maggiore leggibilità del discorso.

Primo passo: configurazione preliminare standard. Collegamento alla directory di lavoro, definizione del linguaggio in cui si vuole scrivere all'interno della console GRASS e importazione delle librerie di interesse. Importazione di coordinate e nomi dei fiumi da file txt (codice uguale al precedente).

Secondo passo: definizione dei vettori vuoti dei descrittori, da riempire ad ogni giro del successivo ciclo di estrazione. Definizione fatta in maniera analoga a quella dei 3 vettori di coordinate e nomi dei fiumi (eseguita nel primo script).

```
quota_media_v = []
quota_max_v = []
MAP_v = []
IDFa_v = []
IDFa_devstd_v = []
fourier_B1_v = []
cv_rp_v = []
```

Terzo passo: definizione del ciclo di estrazione e configurazione della regione di analisi. Si tratta anche in questo caso di un ciclo for, definito come in precedenza (che vada cioè dal primo all'ultimo bacino).

```
for bacino in range(0,len(n_bacini)):
bacino

# Caricamento shapefile
gscript.parse_command('v.in.ogr', flags='o', input='shp', layer=fiume[bacino]+'_'+str(bacino), output='Bacino_n_'
+str(bacino), min_area=0, overwrite=True)

# Configura regione
gscript.parse_command('g.region', flags='d', overwrite=True)
gscript.parse_command('g.region', vector='Bacino_n_' +str(bacino), res=100, overwrite=True)
gscript.parse_command('v.to.rast', input='Bacino_n_' +str(bacino), output='Bacino_n_' +str(bacino),   use='cat',
type='area', overwrite=True)
gscript.parse_command('r.mask', raster='Bacino_n_' +str(bacino), overwrite=True)
```

Si importa ogni shapefile creato col precedente script (uno per giro del ciclo) mediante il comando ***v.in.ogr*** dichiarando il nome dello shapefile da importare (*layer*) e in quale cartella della directory principale cercarlo (*input*). Lo si salva poi all'interno di GRASS (*output*), dandogli un nuovo nome.

Poi, si configura la regione su cui fare le operazioni successive, analogamente a come è stato fatto in precedenza (***g.region***). In particolare, il file vettoriale importato viene convertito in raster (***v.to.rast***) per permettere al software l'utilizzo del successivo comando di analisi statistica ***r.univar*** (che elabora dati raster).

Quarto passo: estrazione dei parametri e popolamento dei vettori ad ogni ciclo. L'estrazione di tutti e sette i parametri viene fatta mediante il comando di calcolo statistico ***r.univar*** , a cui si passa come input, di volta in volta, il nome del raster richiesto (già presente all'interno del dataset RENERFOR). Nel seguito è riportato l'esempio dell'estrazione di *quota media* e *quota massima* (lo script è analogo per tutti i parametri, cambia solo, per l'appunto, il file di ingresso al comando ***r.univar***).

```
# QUOTA MEDIA e MAX
  gscript.parse_command('g.region', raster='Bacino_n_' +str(bacino), res=100, overwrite=True)
  stats_1 = gscript.parse_command('r.univar', flags='g', map='piemonte_dem_r100', overwrite=True)
  quota_media = float(stats_1.mean)
  quota_max = float(stats_1.max)
```

Si esegue lo zoom con ***g.region*** sul bacino in analisi importato, si calcolano le statistiche del raster col comando ***r.univar*** (in questo caso è stato dichiarato in *map* il file DTM del Piemonte come raster da elaborare) e si salva il valore in esame in una variabile di trasporto, richiamando la statistica richiesta mediante l'apposizione del suo nome (***.mean*** oppure ***.max***) al vettore in cui sono state salvate le statistiche, ***stats_1*** (dopo aver convertito il valore risultante dall'analisi statistica in decimale, ***float***).
Il codice per il calcolo dei rimanenti cinque descrittori è analogo a quanto appena descritto. Ciò che cambia è solo il nome del raster da passare in *map* al comando ***r.univar***.
Alla fine del ciclo, si aggiorna il vettore totale con il nuovo valore salvato nella variabile di trasporto calcolato nel ciclo corrente (mediante il comando ***append*** già descritto in precedenza).

```
# Popolare vettori descrittori (aggiunto un valore ogni ciclo)
  quota_media_v.append(quota_media)
  quota_max_v.append(quota_max)
  MAP_v.append(MAP)
  IDFa_v.append(IDFa)
  IDFa_devstd_v.append(IDFa_devstd)
  fourier_B1_v.append(fourier_B1)
  cv_rp_v.append(cv_rp)
```

Quinto passo: aggiornamento automatico del database con i valori dei descrittori appena calcolati. Quello che si fa è, dapprima, importare il database non aggiornato come shapefile di punti (dopo aver salvato lo shapefile all'interno della cartella di lavoro collegata alla console GRASS), mediante il comando ***v.in.ogr,*** specificando in *input* la cartella dove è allocato lo shapefile, in *layer* il nome dello shapefile nella cartella e in *output* il nome in uscita dal comando. Poi si aggiunge una nuova colonna alla tabella degli attributi per ogni nuovo parametro calcolato dentro al ciclo (comando ***v.db.addcolumn***). Ad ogni colonna si dà il nome del parametro corrispondente.

```
# Import database e aggiunta nuove colonne descrittori
gscript.parse_command('v.in.ogr', flags='o', input='shp', layer='database', output='Database_Aggiornato', min_area=0,
overwrite=True)
gscript.parse_command('v.db.addcolumn', map='Database_Aggiornato', columns='Q_Med double precision, Q_Max double
precision, MAP double precision, IDFa_m double precision, IDFa_std double precision, B1 double precision, cv_rp double
precision, Area double precision', overwrite=True)
```

È importante definire all'interno di ***v.db.addcolumn*** il formato dei valori inseribili nella nuova colonna, oltre che al nome (in questo caso si tratta di valori numerici reali, per cui si è scelto per tutti il formato *double precision*).

Infine, è stata creata una ***query SQL*** (ossia una operazione di selezione scritta in linguaggio SQL – un particolare linguaggio GRASS) che permette di aggiungere ogni valore dei vettori dei parametri nella colonna e riga corrispondente della tabella degli attributi in aggiornamento. Questa query è stata inserita all'interno di un nuovo ciclo for che completa il database una riga per volta (sito per sito), aggiornando tutte le colonne col nuovo valore del parametro.

Anche in questo caso, si riporta a titolo di esempio il solo caso della quota, in quanto tutte le altre query inserite nel ciclo sono analoghe alla prima.

```
for i in range(len(quota_media_v)):
   # QUOTA MEDIA
   query0="UPDATE Database_Aggiornato SET Q_Med=" + str(quota_media_v[i]) + " WHERE cat = " + str(i+1)
   gscript.parse_command('db.execute', sql=query0)
```

In sostanza, il codice SQL dice a GRASS di aggiornare il file *Database_Aggiornato* (è il nome del database dopo essere stato importato all'interno del software coi comandi precedenti), mettendo il valore *quota_media_v[i]* (cioè quello corrispondente al giro i-esimo del ciclo) in corrispondenza della colonna *Q_Med* (creata in precedenza col comando ***v.db.addcolumn***) e della riga in cui il valore della variabile *cat* è uguale all'indice i+1 (si aggiunge uno perché in python la numerazione degli indici parte da zero).

La variabile *cat* è semplicemente un indice progressivo (che parte da 1 e arriva fino al numero totale di bacini estratti, cioè 322), che viene aggiunto in automatico (come nuova colonna della tabella degli attributi) dal comando ***v.db.addcolumn***. In questo caso risulta utile sfruttarla per dire al software in quale riga posizionare il valore i-esimo di quota (che altrimenti verrebbe riprodotto uguale per tutte le 322 righe della tabella).

Il codice SQL così scritto viene poi eseguito dal comando ***db.execute***.

Sesto passo: esportazione dello shapefile aggiornato, come layer di punti (comando ***v.out.ogr***).

Quindi, per riassumere, il risultato finale dello script è l'aggiornamento automatico del database QGIS con i nuovi parametri idro-morfologici calcolati in *py-GRASS* (inseriti come nuove colonne del database stesso).

Anche questo script è riportato nella sua forma completa in *Appendice*, come *Allegato 2*. In questo paragrafo ci si è voluto soffermare sul solo funzionamento dello script, che, essendo molto lungo, avrebbe appesantito parecchio la lettura.

3.3.3 Aggiunta ultimi parametri

A conclusione, per completare il database in maniera definitiva, sono stati calcolati e aggiunti gli ultimi parametri ritenuti utili per le successive fasi del lavoro:

- *area del bacino*;
- *percentuale delle classi di uso del suolo*;
- *portata media annua defluente nel bacino*;
- *esposizione dei versanti*;
- *quota puntuale dei siti di campionamento*;
- *numero di abitanti per kilometro quadrato (per ogni bacino)*;
- *ordine di Strahler*;
- *illuminazione naturale media giornaliera*;
- *pendenza*;
- *indice di naturalità del fiume*;
- *indice di morfologia dell'alveo*.

La descrizione completa del metodo di ottenimento di questi parametri è fornita, nel dettaglio, all'interno del *Capitolo 4*, insieme ad un commento sul significato ecologico che assumono in un'analisi spaziale come quella che si intende fare in questa tesi. In particolare, si vedrà, nel proseguo del lavoro, come tali parametri diventeranno i veri e propri pilastri su cui si baseranno le analisi di regressione lineare che verranno effettuate.

Si vuole, inoltre, sottolineare come gli unici due parametri, calcolati con lo script di estrazione dei descrittori, che assumeranno il significato di variabile, saranno i valori di *quota* e di *precipitazione*. Anch'essi saranno descritti e commentati all'interno del *Capitolo 4*. Gli altri parametri statistici ottenuti dallo script verranno usati come dati di partenza per l'ottenimento di altre variabili poi assunte nei modelli lineari (come ad esempio la *portata media* defluente all'interno di un bacino).

Ad ogni modo, le colonne del database sono state aggiornate col valore di questi parametri. Il database completo e ultimato viene inserito all'interno di un *CD*, assieme alle varie cartelle usate nell'applicazione degli script GRASS, come materiale allegato al presente elaborato di tesi.

Capitolo 4

Variabili

4.1 Premessa

L'operazione di completamento del database ecologico attraverso le informazioni idrauliche, idrologiche e geo-morfologiche, attinenti al territorio in cui sono state eseguite le misurazioni dei dati in campo, ha portato all'aggiunta di numerosi parametri che non sono stati ancora del tutto spiegati o descritti nel dettaglio.

Si tratta di quell'insieme di informazioni fisico-ambientali che sono state ritenute necessarie per poter raggiungere quello che è l'obiettivo principale di questa dissertazione, ossia, come spiegato nel dettaglio al *Capitolo 2*, il riconoscimento e la definizione di pattern ed organizzazioni spaziali di comunità biologiche viventi a ristretto contatto con i sistemi fluviali. Infatti, è noto come, in ecologia, l'organizzazione delle comunità all'interno degli ecosistemi sia notevolmente influenzata dalle condizioni fisiche, chimiche, topografiche, morfologiche e climatiche che caratterizzano l'habitat in cui esse vivono e che regolano le funzioni che esse svolgono all'interno dell'ecosistema.

Naturalmente, come si è già ampiamente discusso, i fattori ambientali non sono gli unici ad avere un ruolo in questo scenario. Per avere un quadro completo e dettagliato della situazione, bisognerebbe poter disporre di ulteriori informazioni, che permettano di considerare anche altri *fattori locali*, complementari a quelli ambientali (come l'interazione biotica tra le comunità e, quindi, tutti i meccanismi relativi alla colonizzazione di nuovi territori e alla competizione intraspecifica ed interspecifica), oppure quelli, cosiddetti, *regionali*, come la dispersione delle specie all'interno dell'ecosistema (che dipende da caratteristiche sia intrinseche che estrinseche alle specie). In realtà, tutti questi meccanismi non sono stati del tutto scartati. Essi sono presenti, seppur in maniera indiretta, e contenuti nel significato di alcune delle variabili utilizzate.

Si tratta di un delicato gioco di equilibri tra tutte queste meccaniche. È il bilancio finale tra tutte le varie dinamiche ecosistemiche che permette di spiegare in maniera maggiormente esaustiva il problema descritto all'inizio.

In questa tesi, non ci si è proposti di trovare una spiegazione definitiva (sarebbe impossibile, in quanto mancano dei dati fondamentali per poterlo fare). Piuttosto, si intende fornire un inquadramento generale su quali siano le principali problematiche relative a questo ambito di ricerca e cercare di dare una spiegazione generale su quali siano i principali fattori che possono influire sulla strutturazione di questi pattern spaziali. Per fare ciò, sono sufficienti le informazioni relative ai parametri ambientali.

In particolare, si inizierà a chiamare questi parametri attraverso il termine, più corretto, di *variabili*. Infatti, essi saranno inseriti, come si vedrà in dettaglio al *Capitolo 5*, all'interno dei modelli lineari di

regressione, al fine di studiare come ciascuno di loro influisca sulla conformazione spaziale dei dati ecologici. Pertanto, essi diventeranno i cosiddetti regressori dei modelli che verranno definiti (e saranno le vere e proprie variabili dei modelli).

Tali variabili non sono state scelte in maniera casuale. È stata eseguita, dapprima, un'attenta revisione della letteratura scientifica più eminente nel campo dell'eco-idraulica fluviale, in modo da poter individuare le variabili normalmente considerate come più rappresentative, in ottica delle analisi statistiche successive. Poi, di tutte quelle raccolte, si è ragionato su quelle che sarebbe stato possibile ottenere come dato iniziale (mediante il lavoro in campo del gruppo di ricerca dei biologi) e quelle possibili da calcolare attraverso software e/o espedienti matematici.

Questa attività di ricerca ha portato alla definizione delle variabili che verranno descritte in questo capitolo. In particolare, esse saranno presentate secondo due punti di vista: il metodo attraverso il quale sono state calcolate e/o misurate; l'influenza che possono avere all'interno dell'ecosistema fluviale e del ciclo di vita delle comunità.

4.2 Variabili Morfologiche

Con *morfologiche*, si vuole intendere quell'insieme di variabili che caratterizzano il bacino idrografico dal punto di vista della conformazione ed estensione territoriale o che dipendono da esse.

Si tratta dei parametri relativi alla topografia (nel senso della distribuzione spaziale dei valori di quota e di pendenza), all'esposizione dei versanti verso nord, ma anche all'estensione spaziale del bacino (quindi area, uso del suolo e densità di abitanti) e alla presenza di ombra (legata alla quantità di illuminazione naturale giornaliera, dipendente a sua volta dalla topografia e dalla presenza di vegetazione).

In particolare, tutte le variabili ricadenti in questa categoria sono state calcolate in remoto (e non misurate in campo), attraverso i software *QGIS* e *GRASS GIS*.

4.2.1 Area del bacino

4.2.1.1 Impatto ecologico

Considerare l'area del bacino è la maniera più immediata ed intuitiva per tenere in considerazione la scala spaziale di riferimento su cui si sta eseguendo lo studio. Infatti, la distinzione, che essa permette di fare (tra bacini grandi e bacini piccoli), consente a chi studia il problema di tenere in conto (seppure in maniera non diretta e abbastanza sommaria) di quello che si è chiamato fattore di *interazione biotica* tra le comunità. Il punto del problema è il seguente.

In aree di grosse dimensioni, l'estensione spaziale del bacino permette, in generale, un contatto minore tra le comunità e quindi un'interazione minore, in termini di competizione per le risorse. In tali condizioni, sarà dunque favorita la coesistenza di numerose nicchie ecologiche ognuna distinta dalle altre e complementare, come funzione, alle altre.

Al contrario, in bacini di piccole dimensioni, le comunità saranno molto più interagenti tra di loro. Le nicchie risulteranno più specializzate (con funzioni maggiormente differenziate e distinte). I ruoli ecologici risulteranno meno simili e più intrecciati e, tali, risulteranno anche i rapporti di competizione (che saranno, pertanto, più intensi).

Si tratta del più generale principio della teoria *CWCD* (*Community Wide Character Displacement*), secondo cui specie coesistenti localmente sono più diverse da quelle coesistenti in aree più ampie, a causa proprio dei rapporti ecologici di nicchia e di competizione interspecifica che si instaurano in aree più limitate geograficamente.

Ovviamente, come si è detto, l'area del bacino è un indicatore generale e qualitativo per quanto riguarda questi rapporti di equilibrio ecologico. Essa fornisce un certo grado di approssimazione di questa particolare dinamica.

Si può dire, in linea con quanto spiegato nel *Capitolo 2*, che l'area del bacino è un cosiddetto *propagatore principale*, all'interno del sistema di influenza intrecciata tra fattori regionali e fattori locali. Infatti, essa è indice dei rapporti di interazione biotica, in quanto fattore regionale che, a cascata, influenza un fattore locale. In generale, questo fattore locale (l'interazione tra comunità) dipenderà in una certa misura dall'area del bacino, ma una sua caratterizzazione completa richiederebbe sicuramente conoscenze più dettagliate riguardo alle specie presenti nel territorio.

4.2.1.2 Metodo di calcolo

Nel *Capitolo 3*, relativo alla costruzione ed al completamento del database ecologico, si è descritto come, per ogni stazione di misura, sia stato estratto il corrispettivo bacino idrografico di appartenenza (che, per molte stazioni, è risultato essere lo stesso). Tale estrazione ha permesso la creazione di tutta una serie di shapefile, contenenti ognuno la forma di un bacino.

La presenza di questi shapefile ha permesso il calcolo, attraverso il software *QGIS*, del valore dell'area di ogni bacino.

La prima cosa fatta è stata unire tutti gli shapefile in un unico (che racchiudesse tutti i bacini, mantenendoli però differenziati l'uno dall'altro). Questa non è stata un'operazione particolarmente complessa. È stato sufficiente importare tutti gli shapefile assieme (col comando di importazione di file vettoriale) e *QGIS* ne ha automaticamente creato uno unico, mantenendo tutte le features (ossia le singole geometrie) come distinte.

Avendo, adesso, a disposizione un file vettoriale, in cui, ad ogni feature, corrisponde un singolo bacino, è stato possibile calcolare l'area attraverso il *calcolatore di campi* di QGIS. In particolare, è stata aggiunta una nuova colonna alla tabella degli attributi ed è stata calcolata l'area mediante la funzione ***$area*** (che calcola l'area di ogni feature, quindi di ogni bacino, per tutto il layer vettoriale).

In questo modo, è stata aggiunta al database la nuova colonna coi valori dell'area del bacino idrografico di ogni stazione (intesa, come già detto, come sezione di chiusura del bacino individuato).

4.2.2 Densità di abitanti al kilometro quadrato

4.2.2.1 Impatto ecologico

L'idea che sta alla base della definizione di questa variabile è molto simile a quella dell'uso dell'area per quanto riguarda la considerazione dell'interazione biotica. Si tratta, anche in questo caso, di fare uso di un propagatore principale, per quanto riguarda, questa volta, lo studio dell'influenza dell'impatto antropico sull'ecosistema fluviale.

In generale, l'influenza dell'antropizzazione del territorio nei confronti di un ecosistema è un problema estremamente complesso e sfaccettato. Esso dipende da moltissimi fattori (relativi per esempio al cambio di chimica delle acque o del regime di portata conseguito a seguito di interventi di sistemazione in alveo). Tenere conto di tutto risulterebbe molto dispendioso in termini di tempo e anche non particolarmente utile.

È stato ritenuto conveniente utilizzare un indicatore che riassumesse adeguatamente questo discorso attraverso un valore semplice da calcolare. Quindi, è stato scelto, come indice, la densità di abitanti al kilometro quadrato (in cui ogni valore di numero di abitanti è riferito all'area del bacino della stazione di interesse).

Questa variabile è da interpretare in maniera abbastanza semplice e lineare. Maggiore è il suo valore, maggiore sarà la presenza antropica all'interno del bacino e maggiore sarà l'influenza dell'antropizzazione sull'ecosistema.

Naturalmente, si tratta di un'approssimazione (che verrà migliorata, in seguito, considerando altri fattori attinenti al problema dell'antropizzazione del territorio, come l'evoluzione dell'uso del suolo), ma che permette di avere un'idea dell'ordine di grandezza della presenza del problema per ogni bacino di analisi.

4.2.2.2 Metodo di calcolo

Sono stati scaricati dal sito dell'ISTAT i valori di popolazione residente (al primo Gennaio 2019), comune per comune, per tutta la regione Piemonte. Di questi dati è stato creato un file excel, in cui, per ogni riga, ci fosse nome del comune, numerosità di residenti e coordinate geografiche (latitudine nord e longitudine est). La presenza delle coordinate geografiche è stata necessaria per completare tutta la seguente procedura effettuata in *QGIS*.

La tabella excel, così creata, è stata importata in QGIS (con sistema di coordinate geografico *WGS84*), come layer di punti (in cui, per ogni punto, risulta definito il campo coi valori del numero di abitanti, nella tabella degli attributi).

È stata poi eseguita l'*intersezione* del layer degli abitanti con quello dei bacini idrografici, così da ottenere un layer di punti contenente i comuni che effettivamente risultassero all'interno dei bacini di interesse.

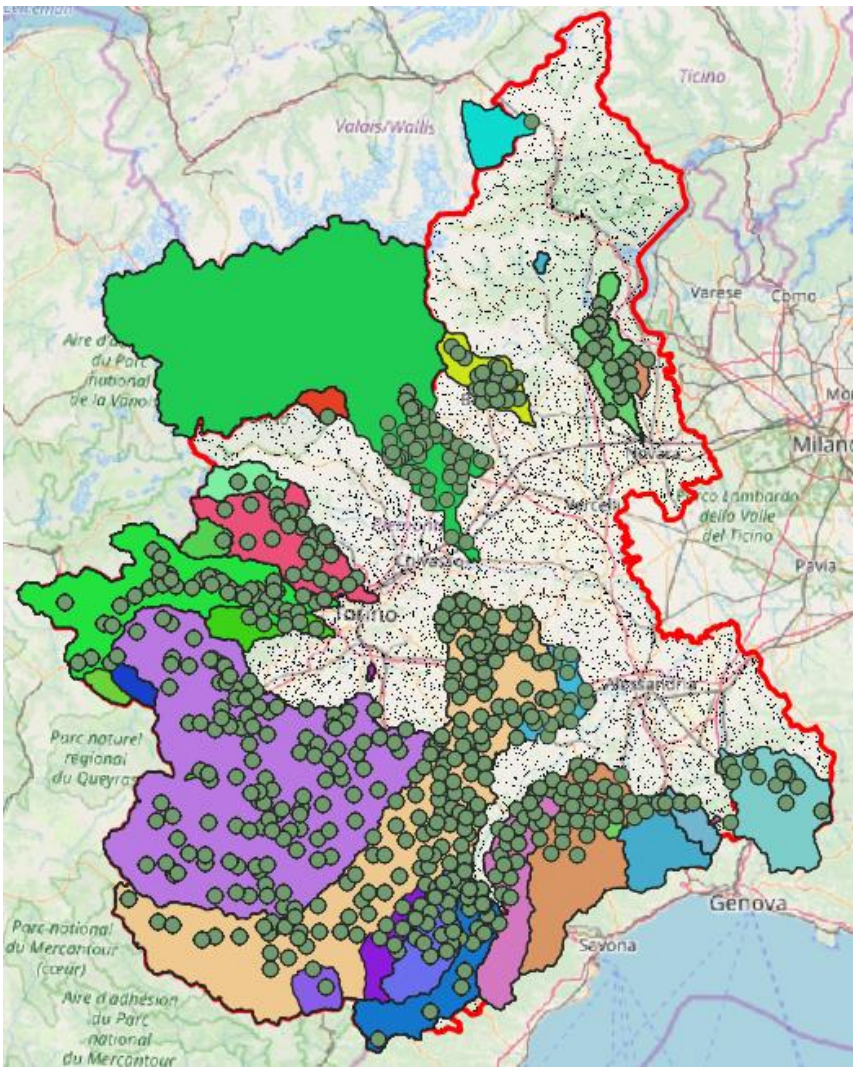

Figura 4.1 – *Risultato dell'operazione di intersezione in QGIS del layer del numero di abitanti per comune con quello dei bacini idrografici. I punti verdi rappresentano il layer degli abitanti per comune ricadenti all'interno dei bacini.*

Eseguita l'intersezione, si è proceduto al calcolo della densità di popolazione mediante il *calcolatore di campi* della tabella degli attributi del nuovo layer appena creato. È stato, quindi, creato un nuovo campo e popolato coi valori derivanti dal rapporto tra il valore di popolazione al primo Gennaio e l'area del bacino corrispondente (in km^2).

Il risultato finale è stato una nuova colonna aggiunta al database excel, contenente, per ogni stazione, la densità media di abitanti per km^2 di area del bacino a cui la stazione apparteneva.

4.2.3 Esposizione dei versanti

4.2.3.1 Impatto ecologico

L'esposizione di un versante rappresenta la sua orientazione rispetto ad una direzione di riferimento (che tipicamente è assunta essere quella relativa al nord geografico). Si tratta di un parametro estremamente importante in biologia ed ecologia, in quanto l'orientazione di un versante rispetto al nord determina in quanta e quale quantità i raggi solari incidono sulla superficie del suolo durante il dì.

Tale parametro si inserisce, nuovamente, nell'ottica di influenza a cascata tra fattori regionali e fattori locali, in quanto la quantità di raggi solari che incidono su una superficie risulta a sua volta proporzionale a due variabili ambientali (di carattere locale, secondo i concetti espressi al *Capitolo 2*) estremamente importanti dal punto di vista del ciclo di vita degli insetti fluviali: umidità e temperatura. Infatti, risultano entrambe due parametri strettamente collegati alla disponibilità di risorse che si possono trovare lungo un corso d'acqua e alle strategie di adattamento nei confronti delle avversità ambientali, che vengono attuate dalle comunità. Inoltre, la presenza di umidità consente lo sviluppo di muschi e licheni, favorendo anche processi di degradazione ed evoluzione naturale del suolo (e quindi sviluppo di risorse nutritive).

In questo senso, il parametro dell'esposizione risulta estremamente importante ed attuale in campo di analisi ecosistemica (naturalmente, sempre considerato con le dovute attenuazioni del caso, in quanto non determina una caratterizzazione completa dal punto di vista delle dinamiche ecologiche, che sarebbero molto più complesse).

4.2.3.2 Metodo di calcolo

Se il significato dell'esposizione dei versanti dal punto di vista ecologico rappresenta una tematica abbastanza complessa e profonda, la procedura effettiva di calcolo del suo valore è stata estremamente rapida.

Tutto è stato fatto attraverso QGIS, tramite il comando omonimo del software (***Esposizione***). Si tratta di un algoritmo che parte dal DTM della regione di interesse, in formato raster (a disposizione grazie ai dati forniti dal progetto RENERFOR), e calcola, per ogni pixel del raster di partenza, il valore dell'orientazione del versante rispetto al nord geografico.

Dunque, il valore del parametro finale sarà espresso in gradi.

La mappa finale, prodotta su QGIS, ha l'aspetto della seguente ***Figura 4.2***, in cui è stata modificata la simbologia standard, definendone manualmente una nuova, in cui i colori più caldi rappresentassero i versanti maggiormente esposti verso il sud geografico, mentre, quelli più freddi, verso il nord geografico.

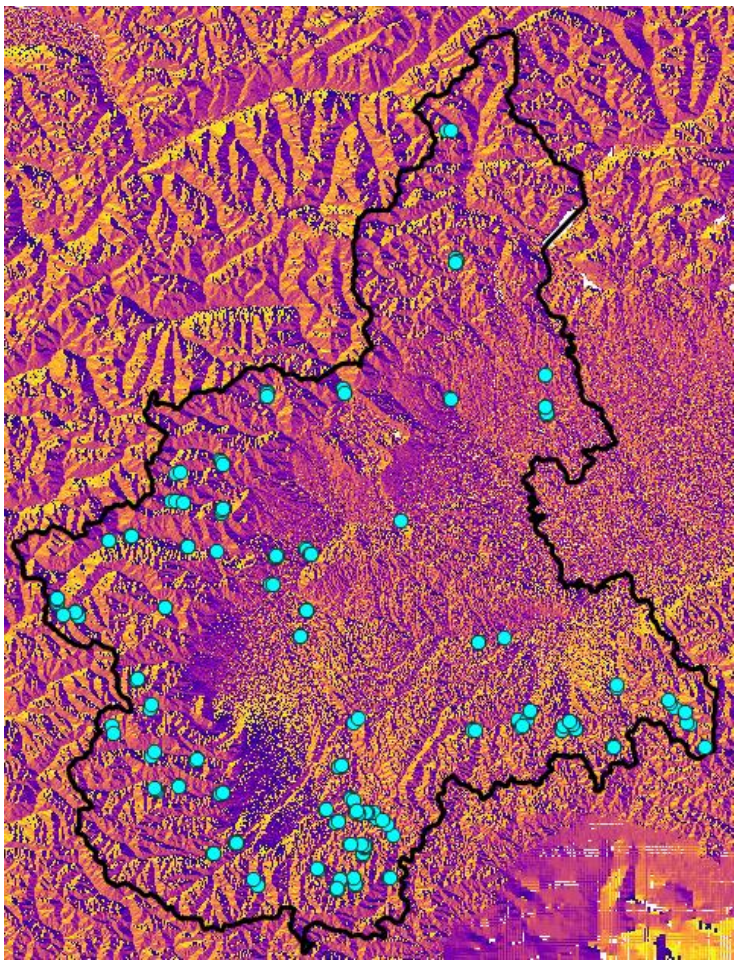

Figura 4.2 – *Risultato del calcolo dell'esposizione dei versanti su QGIS, partendo dal modello digitale del terreno. I colori più caldi (tendenti al rosso) rappresentano i versanti orientati verso il sud, mentre quelli più freddi (blu, viola) verso il nord.*

Il calcolo effettivo del valore di esposizione (in gradi) per ogni stazione di misura è stato effettuato attraverso l'algoritmo di QGIS ***campiona valori raster*** , che ha permesso di ottenere il valore puntuale del raster per ogni punto del layer del database ecologico (creando una nuova colonna nella tabella degli attributi del layer vettoriale).

4.2.4 Luce Solare Media

4.2.4.1 Impatto ecologico

Questa variabile può essere considerata come uno degli effetti diretti di quella appena descritta, ossia l'esposizione dei versanti. Essa rappresenta una stima (media) di quella che è la quantità effettiva di illuminazione naturale che perviene al suolo, lungo il territorio di analisi.

Il suo significato ecologico è, pertanto, direttamente legato a quello descritto per quanto riguarda l'impatto dell'esposizione. La luce solare regola, infatti, il ciclo di vita degli insetti e, a sua volta, è indice indiretto della prossimità della vegetazione terrestre ai lati dell'alveo (infatti, maggior presenza di vegetazione implica maggior copertura ombrosa e minore quantità di luce a terra). La

presenza di vegetazione risulta essere estremamente importante per quanto riguarda la crescita di alcuni tipi di larve, che hanno bisogno di risorse che crescono solo in determinate zone adeguatamente ombreggiate. Essa riflette anche, indirettamente, la posizione all'interno del reticolo fluviale. Infatti, la presenza di vegetazione arborea (e quindi di ampia copertura d'ombra) è la situazione tipica dei rami di basso ordine (di strahler) del reticolo fluviale, che sono altamente influenzati dal materiale terrestre e dalla vegetazione riparia.

4.2.4.2 Metodo di calcolo

L'illuminazione naturale rappresenta il prodotto di una numerosa serie di fattori: la topografia (quindi il discorso legato a quote, esposizione e pendenze), la presenza di vegetazione (difficile da quantificare, a meno di tecniche satellitari) e vari fenomeni di natura atmosferica (copertura nuvolosa, diffusione della luce etc..).

Tenere in conto di tutti questi fattori, nel calcolo della quantità di ore di luce giornaliere in una determinata zona, risulterebbe sconveniente e complesso.

Pertanto, si è scelto di usare un indicatore medio, rappresentato dalla *luce solare media annuale*.

In sostanza, è stato scelto un giorno di riferimento, che, visto l'obiettivo di creare un valore indicativo medio, è stato assunto essere il giorno dell'equinozio di primavera (quindi il 21 Marzo), e si è valutato il quantitativo di ore di luce in quel determinato giorno, sulla base dei valori di esposizione, elevazione e pendenza del territorio in esame. Inseguito, si è assunto il valore calcolato in questo modo come valore medio riferito all'anno.

Tutto ciò è stato fatto attraverso un piccolo script GRASS, scritto in python-GRASS, che andasse a estrarre, per ogni bacino idrografico definito in precedenza, un raster rappresentante la distribuzione delle ore di luce del giorno 81 dell'anno (ossia il 21 Marzo), attraverso il comando ***r.sun***. Il codice scritto è il seguente.

```
a = range(0,len(n_bacini))

for bacino in a:
bacino
gscript.parse_command('v.in.ogr', flags='o', input='shp', layer=fiume[bacino]+'_'+str(bacino), output='Bacino_n_'
+str(bacino), min_area=0, overwrite=True)
gscript.parse_command('g.region', flags='d', overwrite=True)
gscript.parse_command('g.region', vector='Bacino_n_' +str(bacino), res=100, overwrite=True)
gscript.parse_command('v.to.rast', input='Bacino_n_' +str(bacino), output='Bacino_n_' +str(bacino), use='cat', type='area',
overwrite=True)
gscript.parse_command('r.mask', raster='Bacino_n_' +str(bacino), overwrite=True)
gscript.parse_command('r.slope.aspect', elevation='piemonte_dem_r100', slope='Slope_'+fiume[bacino]+'_'+str(bacino),
aspect='Aspect_'+fiume[bacino]+'_'+str(bacino), format='degrees', precision='CELL', min_slope=5, overwrite=True)
gscript.parse_command('r.sun', elevation='piemonte_dem_r100', aspect='Aspect_'+fiume[bacino]+'_'+str(bacino),
slope='Slope_'+fiume[bacino]+'_'+str(bacino), insol_time='Insolation_Time_'+fiume[bacino]+'_'+str(bacino), day=81,
overwrite=True)
gscript.parse_command('r.out.gdal', input='Insolation_Time_'+fiume[bacino]+'_'+str(bacino),
output='Insolation_Time_'+fiume[bacino]+'_'+str(bacino), format='GTiff', overwrite=True)
```

La funzione dei vari comandi usati nello script è già stata commentata nel *Capitolo 3* (nella fase di spiegazione dei due script per estrazione dei bacini idrografici e dei descrittori). In breve, essi servono ad importare i vettori contenenti i bacini idrografici, convertire gli shapefile in raster, fare una maschera sulla regione di interesse su cui eseguire il successivo calcolo, definire esposizione e pendenza (sulla base del DTM), calcolare le ore di sole durante il giorno specificato (con il comando ***r.sun***) e, infine, esportare tutti i raster di illuminazione appena calcolati (tutti i comandi sono stati inseriti all'interno di un ciclo che importa uno shapefile alla volta).

Se si uniscono tutti i raster di illuminazione media diurna in uno unico, si ottiene il risultato riportato nella seguente ***Figura 4.3***.

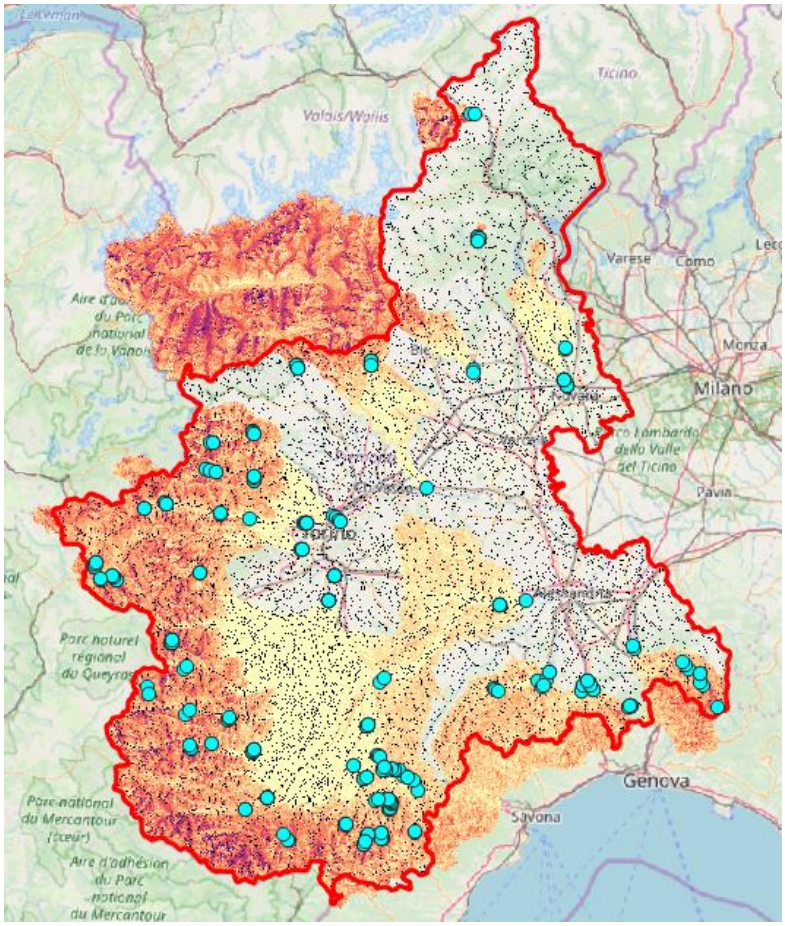

Figura 4.3 *– Risultato del calcolo dell'illuminazione media diurna su QGIS. I valori più chiari (tendenti al giallo) rappresentano le zone in cui le ore di sole sono massime (12h). Man mano che la sfumatura tende verso colori più scuri (rosso), le ore di luce diminuiscono.*

La simbologia del raster è relativa al valore di ore di luce/giorno calcolate dall'algoritmo. I valori massimi, poiché ci si è posti, per il calcolo, nel giorno dell'equinozio di primavera, sono di 12 ore (e sono rappresentati dal colore giallo). I valori più bassi sono quelli delle zone che, in mappa, tendono al colore rossiccio (in cui le ore di luce diminuiscono).

Si precisa che, nell'uso dell'algoritmo, si è tenuto conto solamente della topografia del territorio (attraverso il calcolo intermedio di esposizione e pendenza dei versanti) e non della presenza

effettiva di vegetazione ai lati del reticolo fluviale. Questo ha sicuramente provocato una sovrastima del quantitativo effettivo di ore di luce.

Il calcolo del valore della variabile per ogni stazione è stato eseguito, analogamente a prima, attraverso il comando ***campiona valori raster***.

4.2.5 Pendenza

4.2.5.1 Impatto ecologico

La distribuzione delle pendenze lungo il reticolo fluviale risulta essere di cruciale importanza per quanto riguarda le strategie di adattamento delle comunità all'ambiente circostante.

Un valore maggiore o minore di essa può rappresentare un ostacolo oppure un'agevolazione al raggiungimento di un determinato sito privilegiato, determinando, così, una delle quattro dinamiche metacomunitarie descritte al *Capitolo 2*.

L'influenza della topografia (in particolare delle pendenze), infatti, può causare un'impossibilità allo spostamento di comunità anche caratterizzata da elevate potenzialità alla dispersione inter-habitat e da capacità di colonizzazione. Elevate pendenze possono rappresentare una vera e propria barriera al raggiungimento di siti strategici.

Un esempio di questo problema è profondamente attuale e riguarda la risposta delle comunità ai cambiamenti climatici. Delle comunità adattate ad acque più fredde si trovano a dover rispondere ad aumenti inaspettati di temperatura dell'acqua e a dover risalire la corrente in cerca di zone più adeguate al loro ciclo di vita. In questo senso, zone con elevato grado di pendenza rappresentano un enorme ostacolo all'adattamento di tali comunità ai cambiamenti ambientali. Ovviamente, la gravità del problema dipenderà dalle caratteristiche peculiari delle specie. Specie volanti ne saranno meno influenzate, rispetto a specie che sono relegate al movimento lungo la corrente.

In generale, però, si capisce come la pendenza assuma un ruolo importante nella costruzione e sopravvivenza (in situazioni estreme) delle comunità fluviali.

4.2.5.2 Metodo di calcolo

Il calcolo della carta di distribuzione delle pendenze è stato eseguito su *QGIS* attraverso il comando di analisi geomorfologica ***Pendenza***.

Esso permette di calcolare la carta delle pendenze (in gradi o percentuale), partendo dal modello digitale del terreno della zona di interesse. La pendenza viene calcolata considerando la variazione di quota tra una cella (pixel) e le otto adiacenti. Alla fine, viene assegnato, come valore definitivo, quello massimo.

Il risultato di questa operazione viene riportato in ***Figura 4.4***.

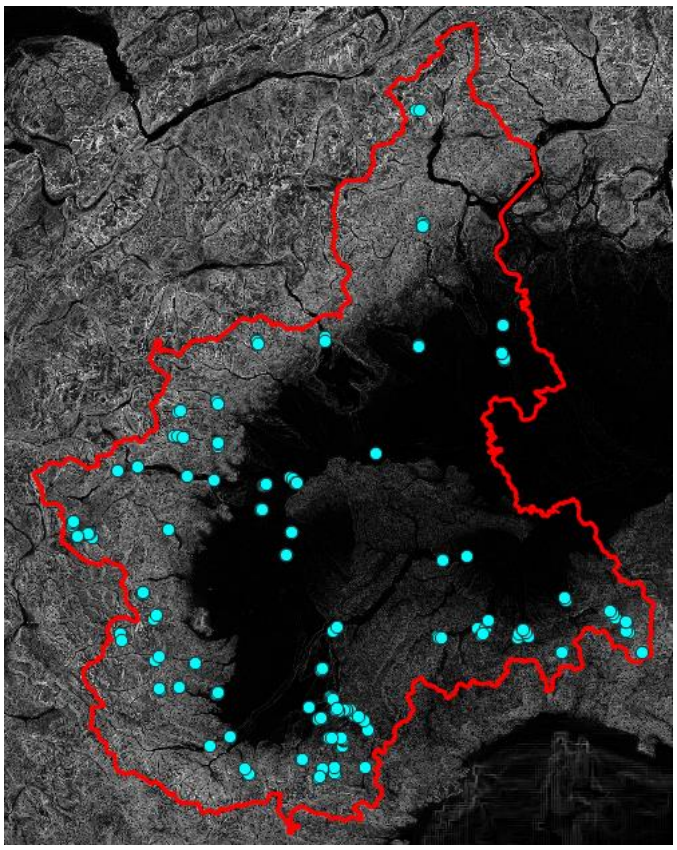

Figura 4.4 – *Carta delle pendenze (calcolata attraverso QGIS). I valori scuri (tendenti al nero) rappresentano le zone più pianeggianti, mentre quelli chiari (bianco) rappresentano le pendenze più elevate.*

Dal calcolo risulta che il valore massimo di pendenza sia dell'ordine dell'80% (in zone di torrente, prevalentemente).

Infine, è stato eseguito il campionamento, stazione per stazione, dei valori calcolati, attraverso il solito comando del menu di analisi raster ed è stata aggiunta una nuova colonna al database.

4.2.6 Quota

4.2.6.1 Impatto ecologico

Dalla consultazione degli articoli scientifici più recenti ed accreditati in campo di eco-idraulica, la quota risulta essere una delle variabili maggiormente influenzanti lo strutturarsi delle comunità ecologiche fluviali.

Le specie di macroinvertebrati che si trovano nelle zone montane o pedemontane (come tutte quelle della regione di analisi di questa tesi) risultano essere profondamente adattate alla quota in cui svolgono le proprie funzioni vitali ed ecosistemiche. Al variare di essa, infatti, si possono constatare notevoli cambiamenti in termini di composizione delle comunità (di specie, quindi), ma anche, e soprattutto, di caratteristiche trofiche e funzionali (ossia di quelle caratteristiche interspecifiche che delineano il ruolo che ogni individuo svolge all'interno della comunità di appartenenza, in termini anche di catena alimentare).

In particolare, la struttura di una comunità è fortemente dipendente dal contesto e la quota può essere interpretata come indicatore di questo cambiamento di scenario ambientale.

Per esempio, utilizzando il ragionamento degli *effetti a cascata* (tra fattori regionali e fattori locali), un valore maggiore di quota può essere inteso come una maggiore distanza dal punto di sbocco a valle del corso d'acqua e quindi un valore inferiore dell'influenza antropica sull'ambiente circostante. Una minore antropizzazione, a sua volta, determina un effetto diretto sulla struttura dei gruppi funzionali delle comunità, determinata da strategie diverse di adattamento al contesto (più o meno antropizzato).

Considerando la quota, si può dunque ricostruire, in maniera qualitativa, la catena di dominanza di una comunità lungo la rete fluviale. Nelle zone più in alto in quota ,ad esempio, ci saranno quelle comunità caratterizzate da gruppi funzionali adattati ad ambienti con pendenze più elevate, vegetazione più abbondante e granulometria più grossolana del fondale. Nelle zone di valle, invece, ci saranno comunità più adattate al confronto con le modificazioni artificiali in alveo (e conseguenti pendenze più dolci, granulometrie più fini, risorse meno grossolane etc.).

In questo senso, considerare il valore di quota può essere inteso come testimonianza indiretta del cambio di bioma, che avviene, da monte verso valle, lungo un corso d'acqua.

4.2.6.2 Metodo di calcolo

La quota è uno di quei parametri idromorfologici che sono stati ottenuti mediante l'implementazione del programma, scritto in *py-GRASS*, fornito assieme ai dati del progetto *RENERFOR*.

In particolare, sono stati calcolati due valori di quota, quello puntuale (relativo ad ogni stazione di misura) e quello medio (in cui il calcolo è stato effettuato bacino per bacino, e non su tutta la regione di analisi).

Lo script intero (in cui oltre alla quota sono stati calcolati altri parametri idrologici) è riportato in *Appendice*. Il commento allo script e al suo funzionamento si trova al *Capitolo 3* (e non è altro, per quanto riguarda il calcolo della quota, che la traduzione in python di due comandi GRASS per l'estrazione del valore di quota da un DTM, inseriti all'interno di un ciclo for al fine di eseguire il calcolo bacino per bacino).

4.2.7 Uso del suolo

4.2.7.1 Impatto ecologico

Questa variabile determina il completamento del discorso riguardante gli effetti dell'antropizzazione nei confronti degli ecosistemi fluviali.

L'evoluzione dell'uso del suolo (da naturale ad urbanizzato o agricolo) determina un cambiamento profondo nella composizione chimica degli strati che lo compongono e anche delle sostanze nutritive che gli organismi possono trovare ed avere a disposizione.

La chimica dei suoli, poi, si riflette in maniera pressoché diretta in chimica delle acque, che risulta inevitabilmente influenzata dalla tipologia di substrato lungo cui la corrente scorre.

I cambiamenti in temperatura, pH, quantità di ossigeno disciolto, presenza di inquinanti (nitrati, solfati, fosfati) sono un forte indizio di antropizzazione e un forte fattore di cambiamento a cui le specie (e quindi le comunità ecologiche spazialmente organizzate) devono saper rispondere.

In generale, esiste un limite a questa influenza. Come è naturale immaginare, infatti, solo le zone più vicine all'alveo saranno quelle che interessano e influenzano i cambi in strutturazione spaziale delle comunità. In ambito eco-idraulico, questa distanza è assunta pari a 5 km.

Dunque, la percentuale delle classi di uso del suolo diventa un indice indiretto (ma molto significativo) di questi cambiamenti di degradazione ed impoverimento del suolo e conseguente riflesso sulla qualità delle acque.

4.2.7.2 Metodo di calcolo

La definizione della percentuale delle diverse classi di uso del suolo è partita dallo scaricamento dal sito del progetto europeo *Copernicus* della classificazione totale della *Corine Land Cover 2018*, in formato shapefile. Si tratta di una classificazione del suolo standardizzata sulla base di 44 classi.

Come viene descritto nel documento dei metadati, allegato nella stessa cartella di download dello shapefile del progetto Copernicus, queste 44 classi sono, in generale, raggruppabili in cinque principali:

1) territori boscati e ambienti semi-naturali;
2) superfici artificiali;
3) superfici agricole;
4) zone umide;
5) corpi idrici.

Ognuna delle classi possiede un codice numerico, detto *codice CLC*. Il valore di questo codice permette di associare ognuna della 44 classi ad una delle 5 principali appena elencate. Ad esempio,

se il codice è compreso tra 100 e 200, allora la classe apparterrà alla tipologia 1, tra 200 e 300 alla tipologia 2 e via così.

L'obiettivo finale, per la definizione delle percentuali di uso del suolo, è ottenere queste cinque classi e derivarne la percentuale areale di copertura del suolo, bacino per bacino.

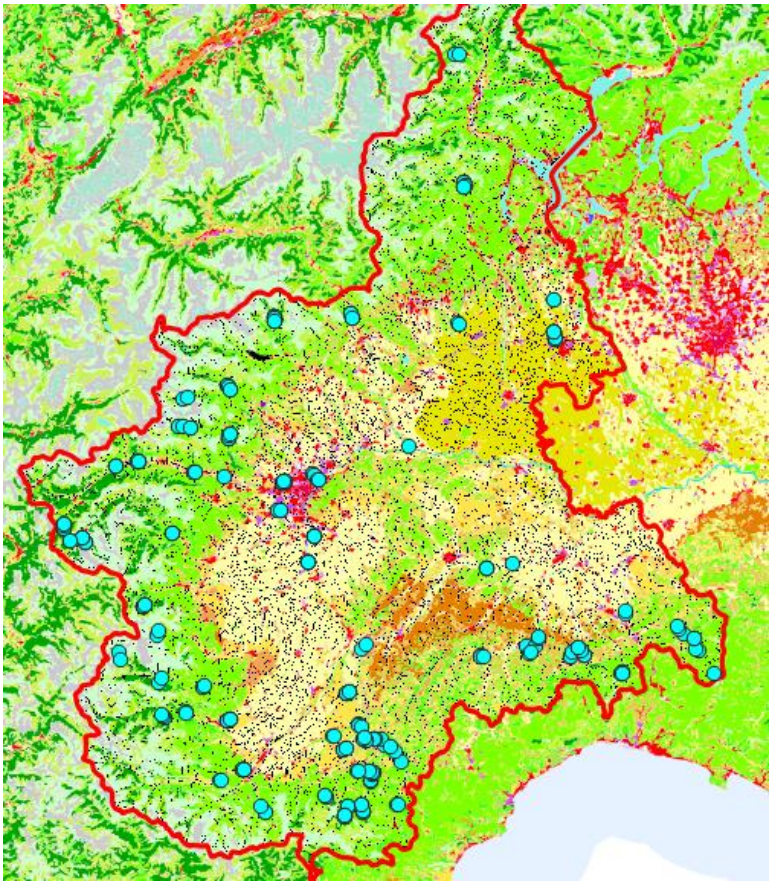

Figura 4.5 – *Le 44 classi di Corine Land Cover 2018 nella zona del Piemonte.*

Il procedimento è stato eseguito interamente in *QGIS*.

La prima operazione fatta è stata ritagliare la mappa totale (comprendente l'intera Europa) sulla sola zona di interesse (il Piemonte). Ciò è stato fatto attraverso il comando ***Ritaglia da estensione***.

La successiva operazione è stata eseguita sul ritaglio appena fatto ed è stata l'operazione di *dissoluzione* delle 44 classi, nelle 5 principali. Dissoluzione significa raggruppamento di ogni pezzo di geometria, che compone il vettore delle classi CLC, avente uno stesso valore di un particolare campo della tabella degli attributi.

Considerando il fatto che ogni classe possiede un codice numerico che la associa ad una particolare tipologia di copertura, è stata dapprima creata, mediante il *calcolatore di campi*, una nuova colonna della tabella degli attributi che andasse ad associare il valore 1 a tutte le geometrie con codice CLC compreso tra 100 e 200, valore 2 a quelle con codice compreso tra 200 e 300 e via di conseguenza fino alla classe 5, in cui la condizione era che il codice fosse maggiore di 500 (e a cui è stato associato valore 5).

Una volta creata la nuova colonna, è stato possibile associare le geometrie con la stessa tipologia di copertura (da 1 a 5) tramite il comando ***Dissoluzione***, che, appunto, ha consentito la creazione di un nuovo layer di classi CLC contenente le sole cinque tipologie principali (così come mostrato e descritto nel commento di ***Figura 4.6***).

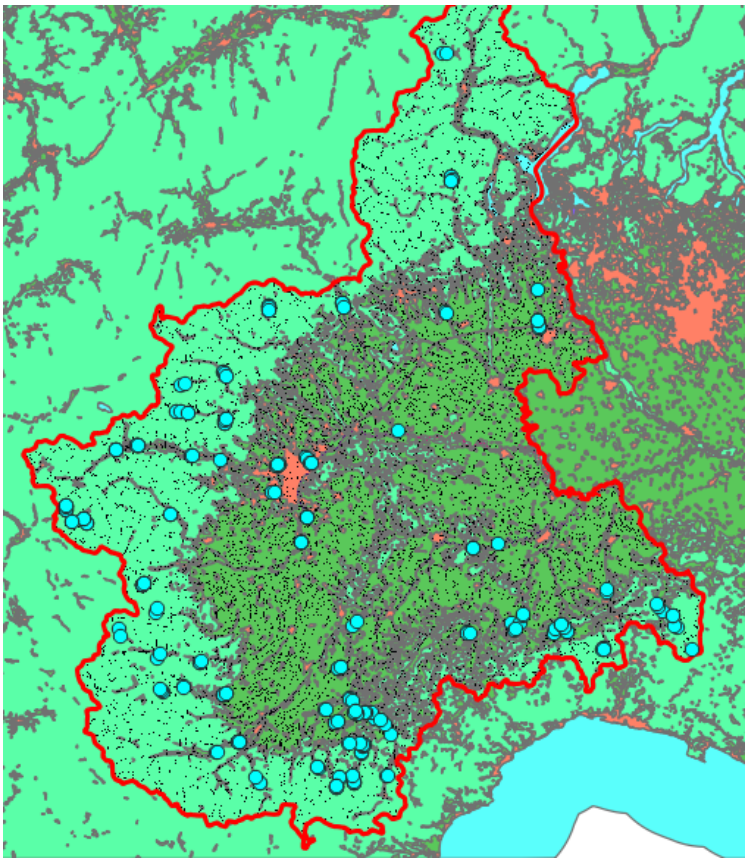

Figura 4.6 – *Classi di uso del suolo CLC dopo l'operazione di dissoluzione. Si notano le superfici artificiali (rosso), le agricole (verde scuro), le naturali (verde chiaro), le zone umide (blu scuro) e i corpi idrici (azzurro).*

In seguito, è stata eseguita l'intersezione tra il layer delle classi CLC e quello dei bacini, in modo da ottenere le classi interne ad ogni bacino. È stato usato il comando ***Intersezione***.

Infine, per ottenere le percentuali di copertura del suolo riferite all'area di ogni bacino, è stato usato nuovamente il *calcolatore di campi*. Dapprima, è stato calcolato il valore dell'area di ogni geometria (quindi, di ogni tipologia di classe CLC), mediante la funzione ***$area***. Poi, è stata definita una nuova colonna della tabella degli attributi e popolata coi valori derivanti dal rapporto tra area della classe CLC (interna al bacino, per via dell'operazione di intersezione) e area totale del bacino.

Il risultato finale rappresenta la percentuale di uso del suolo (della classe considerata), all'interno del singolo bacino. Quindi, per ogni stazione del database, è stato ottenuto il valore della percentuale di classi di uso del suolo (per ognuna delle cinque tipologie definite prima).

In particolare, come si vedrà poi nelle successive analisi statistiche, verranno considerate solo le due classi più influenti dal punto di vista dello scopo di questa tesi, ossia la *percentuale di zone agricole* e *la percentuale di zone urbanizzate*.

4.3 Variabili Idrologiche

Con *idrologiche*, si intendono quelle variabili che si riferiscono alla caratterizzazione e alla definizione delle precipitazioni atmosferiche e delle portate in alveo. In particolare, l'idrologia studia i rapporti che intercorrono tra le due fasi, le cosiddette trasformazioni da afflussi (precipitazioni) in deflussi (portata disponibile in alveo), considerando tutte le varie dispersioni che avvengono in mezzo (evaporazione, infiltrazione, ruscellamento etc.).

L'approccio usato nella definizione di queste variabili è, in genere, di tipo statistico, in quanto si tratta di fenomeni fortemente stocastici e molto diversi da zona a zona a seconda delle caratteristiche morfologiche e atmosferiche.

In questo paragrafo, verranno presentate due variabili relative a quanto appena detto e verranno descritti brevemente i metodi con cui sono state ottenute (e l'uso che se ne può fare), nonché l'impatto che esse hanno, in generale, a livello ecosistemico.

4.3.1 Precipitazione Media Annua

4.3.1.1 Impatto ecologico

Le piogge sono una componente fondamentale della regolazione dei regimi idrici dei corsi d'acqua superficiali. Una maggiore o minore quantità di acqua, apportata dalle precipitazioni, può determinare condizioni di magra o di piena per un fiume. Condizioni che, a loro volta, inevitabilmente influenzano la fauna, che, in quei corsi d'acqua, vive e regola le proprie funzioni.

In particolare, essa assume un ruolo cruciale in condizioni di estrema aridità, in quelle zone in cui la magra del fiume permane per la maggior parte delle stagioni. In queste condizioni, anche apporti poco frequenti di acqua dalle precipitazioni possono garantire le condizioni di umidità necessarie alla sopravvivenza delle comunità adattate a quel tipo di clima.

All'altro estremo, il susseguirsi di eventi straordinari di precipitazione, in zone particolarmente umide (con corsi d'acqua sempre a regime), comporta l'avvenimento di frequenti eventi di piena, che causano l'alluvionamento stagionale anche di parti della piana alluvionale circostante il fiume (con impatti ecologici notevoli, in termini di adattamento delle specie presenti alle condizioni estreme).

L'apporto meteorico risulta, quindi, una variabile estremamente dipendente dal clima della zona e, quindi, dalla latitudine. Nel caso dei corsi d'acqua montani e pedemontani, tipici delle zone che si stanno considerando in questa analisi, la precipitazione assume un ruolo cruciale nella definizione delle condizioni ambientali.

Si tratta di un parametro molto difficile da misurare e quantificare, ma molto importante dal punto di vista ecologico, per le latitudini che si stanno considerando.

4.3.1.2 Metodo di calcolo

Anche nel caso delle piogge, risulta conveniente riferirsi ad una variabile che rappresenti le condizioni medie che si verificano all'interno di ogni bacino.

Il progetto RENERFOR, tra i suoi dati, fornisce anche delle mappe di distribuzione della precipitazione estesa a tutta la regione di interesse. I valori ottenuti sono il risultato dell'applicazione di metodi di calcolo statistici, con l'assunzione di una certa distribuzione di frequenza delle piogge.

Grazie a questa mappa, è stato possibile valutare il valore *precipitazione media annua*, la cosiddetta MAP, per ogni bacino idrografico (e, quindi, stazione di misura) di interesse. Il calcolo è stato effettuato, come già detto nel *Capitolo 3* (in cui il codice viene presentato e commentato), mediante l'applicazione dello script GRASS per il calcolo dei descrittori principali dei bacini. In particolare, è stata estratta la media dei valori di precipitazione (in *mm*) presenti in ogni bacino.

Lo script intero viene messo in *Appendice*, come *Allegato 2*.

4.3.2 Portata Media Annua

4.3.2.1 Impatto ecologico

Concetto estremamente legato a quello di precipitazione è quello della portata di un corso d'acqua. Essa dipende, in maniera inderogabile, dalle condizioni climatiche della zona circostante l'alveo e, in particolar modo, dall'apporto atmosferico.

La portata è un valore indice della quantità di acqua che scorre mediamente in un fiume. È chiaro che, per le comunità ecologiche che vivono nei pressi di un corso d'acqua, l'azione del flusso è una componente fondamentale per quanto riguarda le strategie di adattamento. Infatti, può succedere che una comunità si sia adattata alla vita nelle zone umide ma con poca acqua che scorre e, nel caso di verificarsi di un evento estremo, non avrebbe le capacità adatte alla sopravvivenza. Oppure, ci possono essere altre situazioni in cui la presenza di una notevole quantità di acqua che scorre sia necessaria al ciclo di vita delle comunità e alla regolazione dell'ecosistema.

Un caso particolare è rappresentato dalle piane alluvionali. Esse sono caratterizzate, tipicamente, dai cosiddetti alluvionamenti stagionali, che permettono la connessione idraulica di habitat normalmente sconnessi, favorendo la dispersione di comunità verso zone nuove (e regolando, dunque, le dinamiche metacomunitarie descritte al *Capitolo 2*).

Dunque, il verificarsi sempre più frequente di eventi estremi di piena, in determinate fasce climatiche, può, da un lato, favorire e garantire un certo grado di connessione longitudinale e trasversale tra i rami del reticolo idrografico (agendo sul fattore della dispersione e aumentando la connessione nei sistemi aridi), ma, dall'altro, andare a determinare delle condizioni inadatte alla vita di comunità abituate ad un modesto scorrimento di acqua.

In generale, però, l'influenza della portata non dipende solo dalle caratteristiche idro-morfologiche e climatiche, ma anche (e, forse, soprattutto) da quelle fisiologiche e anatomiche delle varie specie che compongono le comunità. Le specie capaci di volare, che, quindi, non relegano il proprio

movimento e le proprie funzioni al solo reticolo fluviale, saranno molto meno influenzate dalle variazioni stagionali di portata, rispetto alle specie puramente acquatiche.

4.3.2.2 Metodo di calcolo

Nuovamente, come indicatore della variabile, si è assunto il valore medio annuo, a scala di bacino. La valutazione deriva dai metodi di *regionalizzazione delle portate*, procedure di analisi statistica che, partendo dalla definizione e calcolo dei cosiddetti parametri descrittori di un bacino idrografico (quali quota, uso del suolo e pioggia), permettono di determinare, alla fine, una stima del valore di portata media defluente all'interno del bacino.

In questo caso, si è partiti dai dati del progetto RENERFOR e sono stati valutati tutti i descrittori necessari all'applicazione delle seguenti due formule di regressione finali (per la stima di deflusso medio *Y*, espresso in mm, e portata media $\bar{Q}$, espressa in m^3/s)

$$Y = -7.3605 \cdot 10^2 + 1.2527 \cdot MAP + 3.2569 \cdot 10^{-1} \cdot h_m + 5.2674 \cdot fourier_{B_1} - 6.7185 \cdot clc_2$$

$$\bar{Q} = \frac{Y \cdot A}{31536}$$

dove *MAP* è la precipitazione media annua, h_m è la quota media all'interno del bacino, $fourier_{B_1}$ è il coefficiente dello sviluppo in serie di Fourier (derivante dall'applicazione dei metodi di regionalizzazione), clc_2 è la percentuale di suolo naturale e *A* è l'area del bacino (espressa in km^2).

Tutti i coefficienti per il calcolo del deflusso medio *Y* (eccetto che la percentuale di suolo naturale, ottenuta in QGIS) sono stati valutati mediante lo script GRASS per il calcolo dei descrittori (partendo dalle mappe di distribuzione dei parametri fornite dal progetto RENERFOR e calcolandone la media su ogni bacino). La portata media è stata valutata in excel dopo aver esportato i parametri calcolati con lo script GRASS come colonne nuove del database.

4.4 Variabili Idrauliche

Le variabili idrauliche si riferiscono alla descrizione del fiume in maniera più specifica, dal punto di vista di morfologia ed evoluzione dell'alveo, presenza o assenza di vegetazione arborea e/o riparia, granulometria del fondale (ed eventuale forma di esso), tipologia di flusso, ordine dei rami del reticolo e livelli idrici.

Purtroppo, molte di queste informazioni possono essere ottenute solo mediante esperienza diretta in campo e una loro valutazione quantitativa precisa è stata talvolta preclusa. Si è cercato di sopperire a queste mancanze mediante delle strategie qualitative, quali l'uso di software di visualizzazione satellitare, come *Google Earth*, o il supporto quantitativo degli algoritmi di *QGIS*.

In questo paragrafo, verranno presentate tre variabili, che si riferiscono agli aspetti descritti sopra, sempre secondo il procedimento assunto per le variabili morfologiche ed idrologiche.

4.4.1 Ordine di Strahler

4.4.1.1 Impatto ecologico

Il sistema di ordinamento di Strahler rappresenta una tipologia di classificazione dei rami di un reticolo idrografico secondo un metodo gerarchico. Esso si basa su alcune semplici regole e risulta pertanto di comoda applicabilità pratica:

a) si definiscono rami di *primo ordine* le aste che hanno origine dalle sorgenti;
b) dall'unione di due rami di ordine *n* ne deriva uno di ordine *n+1*;
c) dall'unione di due rami di ordine differente, il confluente con ordine maggiore darà l'ordine al tratto di canale immediatamente a valle;
d) la successione di due o più rami, caratterizzati dallo stesso ordine *n*, costituisce canali del loro stesso ordine;
e) il canale caratterizzato dall'ordine più elevato *N* determina l'ordine complessivo dell'intero bacino idrografico.

Tale definizione dei rami del reticolo assume, in ecologia, un significato molto preciso, in termini di isolamento delle popolazioni e di possibilità di dispersione lungo il reticolo.

Infatti, i rami di primo ordine rappresentano, ecologicamente, quelli maggiormente isolati, in cui il grado di interconnessione col resto della rete risulta più basso. Le comunità che abitano queste parti della rete fluviale (le cosiddette acque di testa, *headwaters*) saranno, perciò, caratterizzate da un maggior grado di isolamento spaziale. Questo però non implica che la loro dispersione lungo la rete sia nulla (in quanto essa dipende soprattutto dalle caratteristiche intrinseche e fisiologiche delle specie).

La struttura della rete fluviale (la cosiddetta *struttura dendritica*) rappresenta, semplicemente, un indice sul controllo dell'estensione e della quantità della dispersione delle comunità all'interno dell'ecosistema. Alle scale spaziali più piccole, si tende, in media, ad un grado di isolamento maggiore rispetto che a quelle più grandi (rappresentate dai *main stems*, i rami centrali del reticolo).

In questo senso, l'ordine di Strahler (che è un parametro che descrive e riassume questa struttura dendritica della rete fluviale) caratterizza molto la disposizione spaziale della biodiversità negli ecosistemi fluviali, sottolineando la centralità degli aspetti riguardanti la connessione tra i rami e la dispersione nella spiegazione dei pattern spaziali di organizzazione delle comunità fluviali.

4.4.1.2 Metodo di calcolo

La valutazione dell'ordine di Strahler è stata particolarmente problematica. Il reticolo idrografico raster, fornito da RENERFOR, risulta molto dettagliato, in termini di ramificazione dei canali, e difficile da gestire dall'algoritmo ***Strahler Order*** del gruppo *SAGA* del software *QGIS*.

È stata provata la sua applicazione, ma il risultato non è stato soddisfacente. I rami non venivano definiti secondo le cinque regole previste dal metodo originale. Il problema era dovuto al fatto che, nella costruzione dell'algoritmo, è stata prevista l'applicazione ad un DTM e non ad un reticolo idrografico già estratto (e senza alcuna indicazione di quota). L'algoritmo, infatti, prende il DTM e

restituisce i rami principali del corso d'acqua, assegnando ad ognuno un numero indicativo dell'ordine di Strahler (che però, come detto, non sempre rispetta le definizioni del metodo teorico).

Pertanto, con estrema pazienza, l'ordine di Strahler è stato valutato visivamente per tutte le stazioni (con la fortuna che, molte volte, le misurazioni si ripetevano nello stesso punto o in punti vicini, in cui, di fatto, non cambiava la struttura del reticolo e, quindi, l'ordine di Strahler).

Il risultato finale è stato il completamento della rispettiva colonna del database. L'ordine massimo registrato è stato 8, in prossimità di Quassolo, lungo il bacino idrografico della Dora Baltea (che comprende l'intera zona della Valle d'Aosta).

4.4.2 Morfologia dell'alveo

4.4.2.1 Impatto ecologico

In molti articoli scientifici riguardanti la teoria delle metacomunità e l'eco-idraulica, viene tenuto in conto della struttura del canale, in cui vengono eseguite le misurazioni biologiche, e della sua morfologia. La forma dei canali può essere intesa, in una descrizione più generale possibile, come prevalentemente *rettilinea*, a *deboli meandri*, a *forti meandri* o a *canali intrecciati*.

Essa rappresenta un altro degli ormai celebri propagatori principali, in quanto sulla base della sola morfologia dell'alveo fluviale possono essere dedotte molte proprietà che caratterizzano l'evoluzione nel tempo di un corso d'acqua.

La prima di tutte è la capacità erosiva della corrente. Più la struttura di un fiume presenta dei meandri accentuati, maggiore sarà l'erosione della corrente sulle sponde (e, a sua volta, maggiore sarà il trasporto solido, in quanto l'erosione dipende da quello).

Poi, un fiume ad elevato trasporto solido potrà essere caratterizzato da frequenti fenomeni di sovralluvionamento, dovuti all'aumento del battente idrico, causato dall'improvviso deposito di materiale solido trasportato, in zone di ristagno della corrente. Questo aumento di livello non riuscirà più ad essere contenuto nell'alveo principale e si riverserà nelle aree golenali. Si è descritto in precedenza il risvolto ecologico dei fenomeni di allagamento. Analogamente, si può capire come il fenomeno di sovralluvionamento, nelle zone di pianura, comporti l'aumento della connessione idraulica e della possibilità di dispersione per le specie acquatiche.

Al contrario, nei fiumi a struttura rettilinea, tutto ciò non avviene (o può avvenire in misura minore e con minore frequenza). La struttura rettilinea è proprio dovuta al fatto che la corrente non possiede abbastanza energia per trasportare materiale solido e provocare erosione spondale e modellamento dell'alveo.

In tutto questo, rientra il discorso dell'adattamento delle comunità alle caratteristiche ambientali (di flusso, portata, trasporto di materiale, modellamento delle sponde e del fondale).

4.4.2.2 Metodo di calcolo

La valutazione della morfologia dell'alveo è stata eseguita, in maniera qualitativa, attraverso l'osservazione da *Google Earth*. Possedendo le coordinate geografiche di ogni stazione di misura, si è potuto risalire all'effettiva posizione del punto di campionamento e visualizzare (in un'estensione assunta pari a 100 metri a monte e valle del punto) la struttura e la morfologia dell'alveo di ogni fiume.

Poiché per ogni fiume sono presenti più misurazioni, è stata assunta la stessa morfologia per ogni campionamento relativo allo stesso fiume.

Poiché non sono state trovate informazioni visive su fiumi a canali intrecciati, sono state definite solo tre classi di morfologia dell'alveo:

1. alveo rettilineo;
2. alveo mediamente meandrificato;
3. alveo altamente meandrificato.

Si tratta di cosiddette *variabili categoriche*, ossia che non sono caratterizzate da un contenuto quantitativo (numerico), ma da uno puramente descrittivo (e, non per questo inferiore, anzi, come si vedrà nel successivo *Capitolo 5*, esse assumono un ruolo importantissimo nel determinare i modelli di regressione). Quello che si fa, in genere, è un'operazione di indicizzazione delle variabili. In sostanza, si considera ogni variabile, si valuta il peso che può avere nel determinare ed influenzare il fenomeno che si sta studiando e gli si assegna un valore numerico.

In questo caso, per motivi legati al funzionamento degli script Matlab per il calcolo delle regressioni lineari multivariate, è stato assegnato alla morfologia dell'alveo un valore che aumentasse al diminuire del grado di meandrificazione. In particolare, è stato assegnato valore 5 alla prima classe, valore 4 alla seconda e valore 3 alla terza.

4.4.3 Naturalità del fiume

4.4.3.1 Impatto ecologico

Quest'ultima variabile non è altro che un approfondimento, a scala locale (invece che a scala di bacino), di informazioni e concetti già espressi nei paragrafi precedenti. Essa è, infatti, legata alla presenza ed estensione delle varie tipologie di uso del suolo (in questo caso, vicinanza dei centri abitati, presenza di strade vicine, presenza di aree agricole e presenza di vegetazione circostante l'alveo).

Si è già visto come tutte queste informazioni assumano un significato molto profondo dal punto di vista ecologico. Un significato legato ai cambiamenti in caratteristiche chimiche (dei suoli e delle acque) ed agli adattamenti delle comunità all'evoluzione ambientale.

Si è scelto di effettuare questa analisi di dettaglio (seppure in modo qualitativo), poiché, in genere, negli articoli scientifici, è riportato come l'influenza delle varie classi di uso del suolo sia maggiore se si considerano scale spaziali piccole (quindi dei cosiddetti *buffer*, attorno alla zona di misurazione dei dati biologici).

4.4.3.2 Metodo di calcolo

La procedura di valutazione è stata analoga a quella appena descritta, per quanto riguarda la forma dell'alveo. Si è proceduto su *Google Earth*, cercando di considerare, di volta in volta, stazione per stazione, un'estensione di circa 1 km^2 attorno al fiume (si è scelto questo valore sulla base di studi precedenti eseguiti in campo ecologico).

Quindi, tenendo in considerazione informazioni relative a distanza dei centri abitati, presenza di strade vicine al punto di misura, estensione delle aree coltivate e presenza abbondante o meno di vegetazione riparia o arborea, si è composta una classificazione dei fiumi sulla base di un indice (qualitativo) di *naturalità*.

Seguendo il ragionamento precedentemente esposto (riguardante le variabili categoriche), è stato assegnato un valore della variabile che aumentasse al diminuire della naturalità. Nello specifico, è stato assegnato valore 3 alle zone con naturalità considerata alta (totale assenza di influenza antropica, ambienti lontani dai centri abitati e con preponderante presenza di vegetazione), valore 4 alle situazioni intermedie e valore 5 agli ambienti risultanti maggiormente vicini ai centri abitati e con elevata estensione areale di terreni coltivati.

Capitolo 5

Analisi Lineare e Correlazioni

5.1 Introduzione

A seguito della definizione delle variabili esplicative, ossia di quei parametri idromorfologici (qualitativi e quantitativi) ritenuti significativi al fine della comprensione e della spiegazione dei pattern spaziali di organizzazione delle metacomunità ecologiche, si è passati alla vera e propria fase di analisi ed interpretazione dei dati. Sono stati considerati, come dati da interpretare, le quattro metriche biologiche definite nel *Capitolo 3* per accorpare le centinaia di misurazioni di abbondanza di specie.

Una fase preliminare di interpretazione ha previsto la creazione di quattro scatterplot globali (cioè che tenessero in considerazione i dati relativi a tutti i bacini estratti nelle precedenti fasi del lavoro) dei soli parametri ecologici, al fine di studiare il legame tra uno e l'altro a scala regionale.

In seguito, per semplicità di applicazione e, allo stesso tempo, confrontabilità con studi già eseguiti sull'argomento, è stato scelto, come primo approccio all'interpretazione quantitativa dei dati raccolti , un'analisi di carattere lineare.

In particolare, sono stati creati, in sequenzialità, dei modelli lineari a complessità crescente.

Dapprima, è stata eseguita una regressione lineare semplice, ossia considerando una sola variabile esplicativa alla volta. In questo modo, sono stati ottenuti diversi modelli che legassero il parametro ecologico alla variabile esplicativa (in sostanza, si tratta della classica linea di regressione della distribuzione di punti in uno scatterogramma bidimensionale).

Poi, è stata eseguita una regressione multi-parametrica, in modo da creare un modello che tenesse in considerazione più variabili contemporaneamente. Nello specifico, è stato applicato un approccio di analisi multivariata e sono stati creati due tipi di modello, uno lineare ed uno esponenziale.

La significatività delle analisi è stata tenuta in considerazione mediante opportuni coefficienti statistici (in particolare, di assoluta importanza è stato il parametro R^2).

A conclusione delle analisi svolte, sono state eseguite due zonazioni (una nella zona del bacino del Tanaro, l'altra nella zona del bacino del Po) per verificare la crescita in significatività della regressione qualora si considerassero solo dati appartenenti a bacini con caratteristiche simili (quindi, sarebbe il caso di un'analisi a scala locale, invece che regionale).

5.2 Studio qualitativo dei parametri ecologici

La fase preliminare del lavoro di interpretazione ha riguardato lo studio dei dati ecologici stessi. Come già detto, si tratta delle quattro metriche ecologiche precedentemente calcolate, che hanno permesso di raggruppare in pochi parametri rappresentativi e significativi le numerose rilevazioni biologiche eseguite in campo. In questo modo, è stato possibile eseguire delle analisi su scala regionale, senza dover considerare specie per specie (o meglio, famiglia per famiglia).

Quello che si è fatto è stato lo studio (seppure visivo e indicativo) della correlazione tra le quattro variabili, attraverso la costruzione di uno *scatterogramma*. Esso è un tipo di grafico di dispersione bidimensionale che permette di visualizzare in maniera efficace come varia un parametro in funzione di un altro. In base alla forma della cosiddetta nuvola di punti che si genera all'interno del diagramma si può riconoscere un maggiore o minore legame tra le due variabili. Se la nuvola risulta ampia e dispersa (di forma tendente al circolare), allora le due variabili sono poco correlate tra di loro (ossia, una influenza poco l'altra). Se la nuvola risulta di forma allungata e direzionata lungo un particolare verso, allora le due variabili possiedono un certo grado di correlazione. Generalmente, questa direzione è data dalla retta di regressione dei dati.

Lo scopo di questa analisi non è stato di tipo quantitativo (infatti, non è stato valutato alcun parametro al fine di giustificare la forma della dispersione), bensì essa è servita a dare un'informazione visuale, quindi immediata, su quanto fossero linearmente indipendenti le quattro variabili risposta sotto analisi.

Il problema della *multicollinearità* è, infatti, estremamente importante quando si svolge un'analisi di regressione (e verrà trattato nel dettaglio più avanti in questo capitolo, quando si giungerà a descrivere i modelli di regressione multipla). Per ora, è sufficiente tenere presente che maggiore è la correlazione tra due variabili, minore è il potere esplicativo derivante da una regressione che le coinvolge. Ciò è dovuto al fatto che una parte di spiegazione di una variabile è già contenuta nella regressione dell'altra variabile (a cui la prima è legata). Dunque, il quantitativo di informazioni potenzialmente estraibile da modelli contenenti variabili correlate sarebbe inferiore.

5.2.1 Correlogrammi

Sono stati creati in Matlab gli scatterplot bidimensionali riferiti a tutte le combinazioni possibili tra i quattro parametri (N, S, EPTN ed EPTS), in modo da verificare la rilevanza del problema della collinearità tra le variabili.

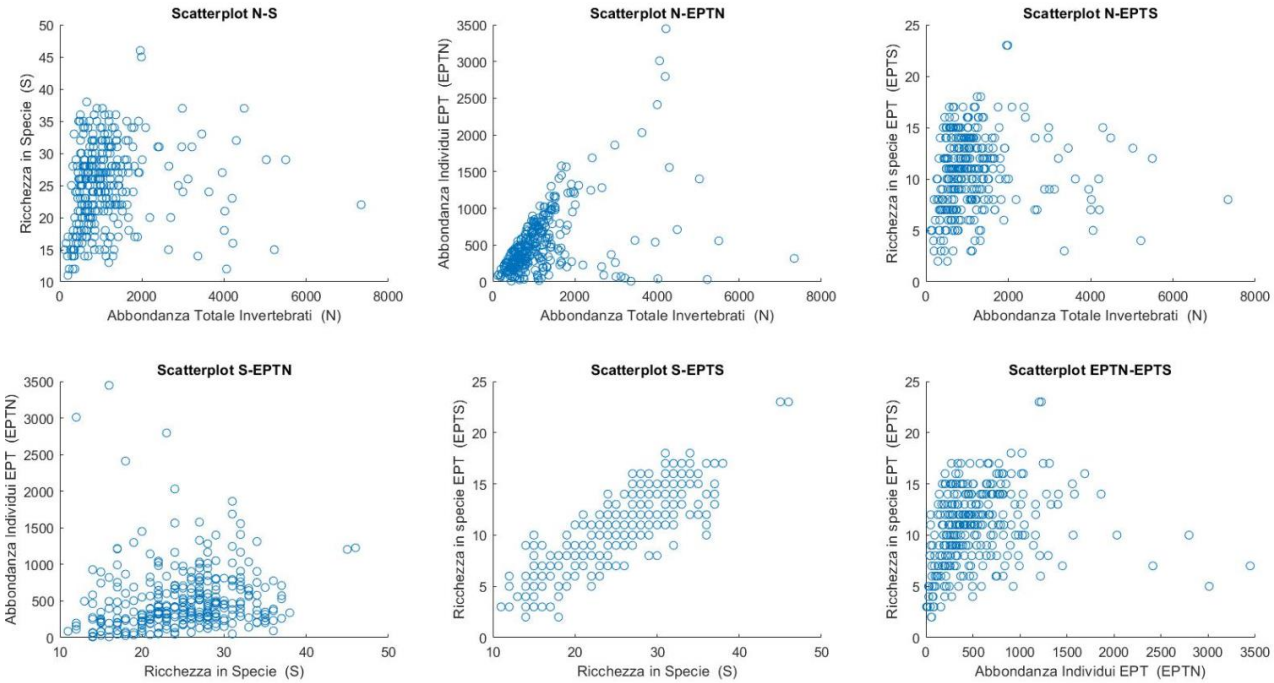

Figura 5.1 *– Scatterplot dei parametri ecologici (tutte le combinazioni)*

Da un'ispezione visiva immediata di ***Figura 5.1*** si nota come solo una delle nuvole di punti create sia riconducibile alla casistica descritta prima (ossia con orientamento verso una particolare direzione).

Si tratta dello scatterogramma tra le variabili S (ricchezza totale in specie) ed EPTS (ricchezza in specie EPT), nel quale si riconosce il tipico orientamento della nuvola lungo una retta (in questo caso, la bisettrice del diagramma). Questa conformazione evidenzia il fatto che, al crescere del valore della variabile S, cresce anche quello della variabile EPTS (rendendo, di fatto, lineare il legame tra le due).

Tuttavia, il riconoscimento di questo legame non è sufficiente a permettere di scartare una variabile in favore dell'altra. Infatti, tutti gli altri scatterogrammi mostrano delle nuvole con forma molto meno definita e, quindi, senza alcun particolare legame di linearità tra variabili.

Proprio per questa ragione, si è scelto di svolgere le successive analisi di regressione utilizzando tutte e quattro le variabili ecologiche. Ciò significa che è stato creato, ogni volta, un modello per ognuna delle quattro variabili.

5.3 Regressione Lineare Semplice

La fase di interpretazione preliminare si è conclusa con la conferma dell'importanza di tutte le variabili ecologiche nella spiegazione dei pattern spaziali di distribuzione delle comunità biologiche oggetto di studio. Quello che occorre fare adesso è trovare un qualche legame tra le misurazioni di abbondanza eseguite in campo e le caratteristiche della zona in cui le misurazioni sono state eseguite.

Si tratta cioè di studiare come si comportano le variabili risposta al variare dei parametri idro-morfologici che caratterizzano (a livello di bacino idrografico) il territorio di presa dei dati. Questi parametri sono stati assunti come variabili esplicative dei modelli che sono stati creati e che verranno descritti nel seguito di questo capitolo.

Il comportamento di una comunità biologica è, infatti, inevitabilmente determinato da fattori esterni (ambientali, climatici, topografici). Studiare la correlazione tra un parametro che caratterizza la distribuzione spaziale delle comunità (vale a dire quello ecologico) e uno che caratterizza l'habitat in cui esse vivono (quello idro-morfologico) equivale ad analizzare esattamente questa influenza. Ovviamente c'è molto di più in gioco. Una stima completa e totale prevede la considerazione effettiva anche di fattori legati all'interazione biotica tra le comunità (i vari meccanismi di competizione per le risorse e di colonizzazione di nuovi habitat) e alla dispersione, così come è stato già spiegato nel *Capitolo 2*. Tuttavia, come è stato mostrato nel *Capitolo 4*, le informazioni idro-morfologiche, contenute dalle variabili descritte, riescono a tenere conto (seppur con un certo grado di generalità) degli altri due tipi di meccanismi ecologici, rendendo, di fatto, robusta, dal punto di vista ecologico, l'analisi eseguita.

In questo senso, a seguito della consultazione degli articoli di eco-idraulica più in vista in campo letterario internazionale, è stato scelto di effettuare, come primo passo all'interpretazione quantitativa dei dati, un'analisi di *regressione lineare semplice*.

Tale analisi è stata fatta prendendo in considerazione tutte le variabili definite nel *Capitolo 4* e creando un modello lineare per ogni variabile. Sono stati, dunque, creati tanti modelli monoparametrici (di fatto, sarebbero le linee di regressione di ogni dispersione bidimensionale creata, considerando di volta in volta una differente variabile risposta – parametro ecologico – ed esplicativa – parametro idromorfologico).

Si tratta della metodologia più semplice per interpretare dei dati, ma ciò non nega che talvolta anche ciò che non è troppo elaborato possa portare a risultati comunque soddisfacenti (in termini di quello che si vedrà essere l'indice di "potere" della regressione, ossia il coefficiente R^2).

Nei prossimi paragrafi viene mostrata la teoria matematica che sta dietro al metodo e, poi, l'applicazione vera e propria ai dati ecologici raccolti in campo.

Si vuole precisare che le formule matematiche ed informazioni teoriche presentate nel seguito (*paragrafo 5.3.1* e relativi sottoparagrafi) sono state tratte dal libro *Introduction to Linear Regression Analysis - Fifth Edition* (Montgomery, Peck, Vining, 2012).

5.3.1 Cenni teorici – Modello di regressione lineare semplice

Il modello più semplice che può risultare da un'analisi lineare è quello che lega un singolo regressore *x* (in questo caso, il parametro idromorfologico) ad una singola variabile risposta *y* (variabile ecologica) mediante una linea retta, chiamata, per l'appunto, *retta di regressione*. Un modello così creato risulta avere la seguente forma matematica

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{5.1}$$

dove β_0 e β_1 sono costanti da determinare ed ε è una componente di errore casuale.

Questa componente di errore è sempre presente quando la legge ottenuta non si adatta perfettamente a tutti i dati misurati (come accade in tutti i casi reali). Generalmente, sia assume che questi errori abbiano media zero, un certo valore di varianza σ^2 e anche che siano non correlati tra di loro.

Tuttavia, nella pratica ingegneristica, è conveniente considerare il regressore come controllato e registrato dall'analista con una precisione tale da poter assumere la componente di errore pari a zero. In ogni caso, se si considera la media della distribuzione (5.1), si ottiene una legge priva di componente di errore

$$E(y|x) = \beta_0 + \beta_1 x \tag{5.2}$$

dove $E(y|x)$ sta a significare la media del modello ottenuto facendo la regressione della variabile *y*, usando come regressore la variabile *x*. La varianza del modello (5.1) è invece pari a

$$Var(y|x) = Var(\beta_0 + \beta_1 x + \varepsilon) = Var(\varepsilon) = \sigma^2 \tag{5.3}$$

I parametri β_0 e β_1 sono chiamati *coefficienti di regressione* e hanno una interessante interpretazione matematica: β_1, ossia la pendenza della retta di interpolazione, rappresenta la variazione della media della distribuzione y prodotta da una variazione unitaria di x; β_0, ossia l'intercetta della retta, rappresenta la media della distribuzione *y* quando x è pari a zero (ma solo se tra i dati in ingresso di x è compreso il valore zero, altrimenti non possiede nessun particolare significato).

5.3.1.1 Minimi Quadrati – Stima dei coefficienti di regressione

Il problema fondamentale che si deve affrontare in fase di creazione del modello di regressione è quello della stima dei coefficienti β_0 e β_1. Essi devono essere determinati partendo dai dati che si hanno a disposizione.

Il metodo più efficiente per determinare una stima robusta del loro valore è quello, cosiddetto, dei *minimi quadrati*. Esso consiste nel calcolare i valori di pendenza (β_1) e quota (β_0) della retta di interpolazione dei dati, che minimizzano la somma dei quadrati degli scarti tra retta e punti.

Immaginando di generare tante di queste rette, per poi trovare quella con i requisiti appena descritti, si può scrivere

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{5.4}$$

Imponendo il criterio dei minimi quadrati alla (5.4), si ottiene

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i \right)^2 \tag{5.5}$$

che sarebbe la somma dei quadrati degli scarti tra modello e dati. Il minimo di questa funzione si ottiene facendone la derivata rispetto ai due parametri ignoti β_0 e β_1

$$\left. \frac{\partial S}{\partial \beta_0} \right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) = 0 \tag{5.6}$$

$$\left. \frac{\partial S}{\partial \beta_1} \right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) x_i = 0 \tag{5.7}$$

Elaborando e semplificando le (5.6) e (5.7) si ottengono le successive *equazioni dei minimi quadrati*

$$\begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} y_i \\ \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} y_i x_i \end{aligned} \tag{5.8}$$

la cui soluzione permette di trovare il valore dei due coefficienti di regressione

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} \tag{5.9}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i \right) \left(\sum_{i=1}^{n} x_i \right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i \right)^2}{n}} \tag{5.10}$$

dove $\overline{y}$ e $\overline{x}$ rappresentano le medie campionarie delle due variabili. Il modello finale che si ottiene sarà dunque il seguente

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{5.11}$$

Inoltre, poiché a denominatore della (5.10) si può riconoscere la somma corretta dei quadrati degli scarti di x e a numeratore quella del prodotto incrociato $y \cdot x$, si può riscrivere la definizione del coefficiente β_1 in una maniera più compatta

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \tag{5.12}$$

dove

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$S_{xy} = \sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n} = \sum_{i=1}^{n} y_i (x_i - \bar{x})$$

5.3.1.2 Adeguatezza del modello – Coefficiente R^2

Una volta creato il modello, è indispensabile poter disporre di uno strumento che permetta di capire in maniera rapida quanto la regressione ottenuta spieghi della distribuzione originaria. A questo proposito si definisce il *coefficiente di determinazione* della regressione, R^2.

Esso rappresenta un indice che descrive il quantitativo di varianza della distribuzione originaria che viene catturata e spiegata mediante la creazione del modello lineare. Pertanto, si definisce il coefficiente come il rapporto tra la varianza spiegata dal modello e quella totale

$$R^2 = \frac{\sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} = \frac{(varianza\ spiegata\ dal\ modello)}{(varianza\ totale\ originaria)} \tag{5.13}$$

dove y_i rappresentano i dati raccolti, $\hat{y}_i$ i dati stimati dal modello e $\bar{y}$ la media dei dati.

Per come è stato definito, il coefficiente R^2 può assumere valori compresi tra 0 ed 1: valori del coefficiente prossimi allo zero significano che il modello creato non si adatta bene ai dati forniti, mentre valori prossimi ad uno significano che è stato creato un modello che spiega bene i dati iniziali.

Generalmente, però, in statistica si preferisce attuare una piccola modifica alla definizione di R^2. Esso, infatti, ha il problema che al crescere del numero di regressori (ossia di variabili aggiunte dentro il modello di regressione) il suo valore cresce indipendentemente dal fatto che la variabile aggiunta abbia una qualche significatività nella spiegazione o meno. In altri termini, la definizione di R^2 non dipende dal numero di variabili incluse nel modello e, pertanto, non è considerata statisticamente robusta. Questo problema, ovviamente, non si pone nel caso di regressione monoparametrica (come quella appena descritta), ma assume un ruolo estremamente delicato nel caso di regressione multipla.

Dunque, per rendere assoluto e sempre significativo il valore di R^2, si definisce un nuovo coefficiente, detto R^2_{adj} (*R^2 aggiustato*)

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \cdot \frac{n-1}{n-k-1} \tag{5.14}$$

dove *n* è il numero delle osservazioni e *k* il numero di variabili assunte nel modello.

I valori di R^2 che sono stati calcolati nell'applicazione ai dati ecologici sono sempre riferiti alla definizione (5.14).

5.3.1.3 Test di significatività – t-Student

Si è visto come con il parametro R^2 si riesce ad avere un'idea sull'adeguatezza del modello alla distribuzione dei dati originari, indipendentemente dal legame reale che c'è tra le due variabili in gioco (variabile risposta e variabile di regressione).

Lo studio di questo legame lineare tra le variabili incluse nel modello è invece permesso dall'applicazione di alcuni particolari metodi statistici, chiamati *test di significatività*. In particolare, in questo paragrafo si vuole accennare ad uno dei più semplici (nel senso di comodità di applicazione), nonché completi, tra i test utilizzati nell'analisi dei dati: il *test t-Student*.

Come tutti i test, esso parte dalla definizione di una, cosiddetta, ipotesi zero (da rigettare, se il test risultasse passato) e una ipotesi uno (da confermare)

$$H_0: \beta_1 = 0, \quad H_1: \beta_1 \neq 0$$

In questo caso, volendo verificare l'esistenza di un legame lineare significativo tra le due variabili in gioco, si definisce, come ipotesi zero (ossia quella che si vuole rigettare), il fatto che la retta di regressione abbia una pendenza $\beta_1 = 0$. Ciò significa dire che non esiste legame lineare tra le due variabili, in quanto la retta di regressione corrispondente all'ipotesi nulla sarebbe orizzontale. Rigettare l'ipotesi nulla significa dire, invece, che questo legame lineare esiste (e, pertanto, il modello risulta significativo dal punto di vista statistico).

Il punto fermo del test è l'assunzione che gli errori ε_i del modello di regressione siano normalmente distribuiti (secondo una campana gaussiana, quindi), con media nulla e un certo valore di varianza σ^2. In matematica, questo concetto si traduce con la seguente dicitura : NID(0,σ^2).

Di conseguenza, se vale questa ipotesi, pure le osservazioni y_i sono NID(β_0+$\beta_1 x_i$, σ^2). Poiché $\hat{\beta}_1$ è definito come una combinazione lineare delle osservazioni, anch'esso risulta normalmente distribuito con media β_1 e varianza σ^2/S_{xx} . Definendo la seguente statistica

$$Z = \frac{\hat{\beta}_1}{\sigma^2/S_{xx}}$$

risulta che anch'essa sia normalmente distribuita, se l'ipotesi nulla è vera. Il problema è che il valore di σ^2 è ignoto per la maggior parte delle volte. Per questa ragione, si preferisce definire una statistica che dipenda non direttamente dalla varianza, ma da una sua stima, definita come segue

$$\hat{\sigma}^2 = \frac{SS_{\text{Res}}}{n-2} = MS_{\text{Res}}$$

dove SS_{Res} rappresenta la somma dei quadrati dei residui (tra valore misurato e calcolato dal modello) ed n il numero di osservazioni totale. Se si definisce una nuova statistica

$$t_0 = \frac{\hat{\beta}_1}{MS_{Res}/S_{xx}} \tag{5.15}$$

e si può dimostrare che essa segue una distribuzione *t* a n-2 gradi di libertà.

La (5.15) può essere riscritta come segue

$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} \tag{5.16}$$

dove $se(\hat{\beta}_1) = MS_{Res}/S_{xx}$ rappresenta la deviazione standard di $\hat{\beta}_1$.

La procedura prevista dal test è quella di comparare il valore di t_0 ottenuto dalla (5.16) con il valore della distribuzione t_{n-2} corrispondente e con grado di significatività α (generalmente pari al 5%).

L'ipotesi nulla (ossia di insignificanza della regressione) viene scartata se

$$|t_0| > t_{\alpha,n-2} \tag{5.17}$$

5.3.2 Applicazione ai dati ecologici

Ora che è stata spiegata un po' di matematica che sta dietro alla costruzione di un modello di regressione lineare semplice, si tratta di applicare il procedimento ad una situazione reale. In questo caso, si tratta di provare a studiare se e in che modo le caratteristiche idromorfologiche dei bacini idrografici possano portare un contributo significativo alla comprensione della disposizione spaziale dei dati ecologici raccolti.

In questa prima fase, si prenderà in considerazione una sola variabile idromorfologica alla volta, in modo da seguire la teoria appena descritta e costruire dei modelli lineari monoparametrici.

5.3.2.1 Regressione lineare semplice in Matlab

Il punto di partenza di questa analisi è stata la creazione di uno script in Matlab per la generazione di tutti i correlogrammi di tutte le combinazioni possibili tra variabile ecologica (variabile risposta) e variabile ambientale (regressore). Per ogni correlogramma è stata eseguita la regressione lineare, trovati i vari coefficienti di interesse (R^2 e coefficiente angolare *m* della retta) e fatto il test di Student per verificare la significatività della regressione.

Il codice, scritto per intero e commentato, viene riportato in *Appendice*, come *Allegato 3*. In questo paragrafo si intende procedere in maniera descrittiva e sintetica, giusto per far capire i passaggi che sono stati seguiti per la generazione degli scatterplot.

La prima cosa fatta è stato importare i dati dal database excel a Matlab. In questo modo, sono state definite due matrici (una, contenente le variabili ecologiche, l'altra quelle idromorfologiche). Poi, dopo aver definito le variabili di interesse da calcolare (R^2,m,t_0), è stato creato un doppio ciclo for (in modo che quello esterno rimanesse fisso sul parametro ecologico, mentre quello interno andasse di volta in volta su ogni variabile ambientale) così da poter creare, per ogni coppia parametro ecologico-parametro ambientale, il corrispondente correlogramma, la corrispondente linea di regressione (della quale è stato fatto il test t-Student) e calcolare i parametri detti sopra.

Il risultato finale è riportato nelle seguenti figure, in cui la linea di interpolazione è colorata in rosso, qualora il test fosse passato, mentre risulta nera in caso contrario.

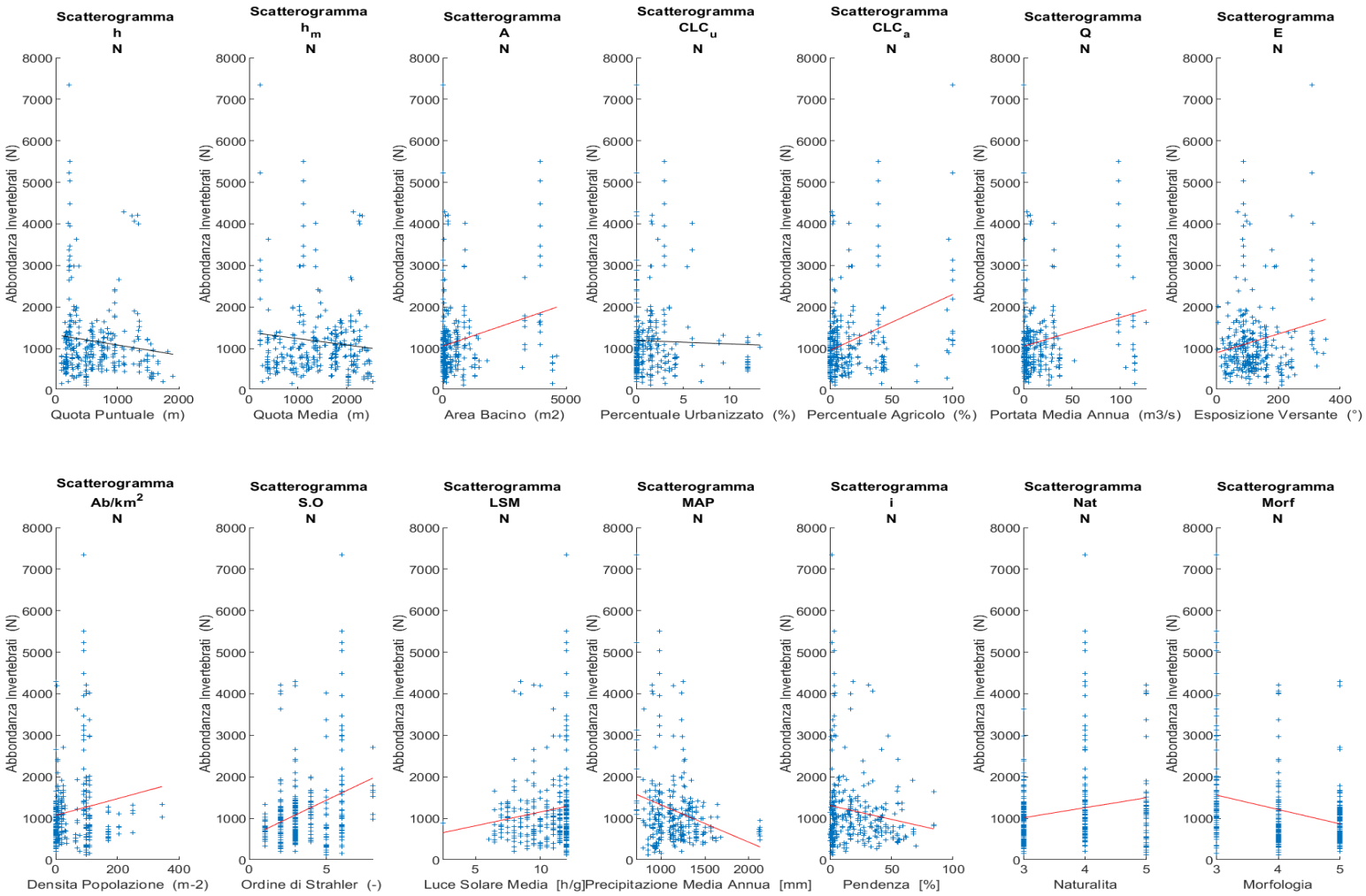

Figura 5.2 *– Regressione lineare semplice. Correlogrammi e rette di regressione per il parametro N (sono colorate in rosso le rette che passano il t-test)*

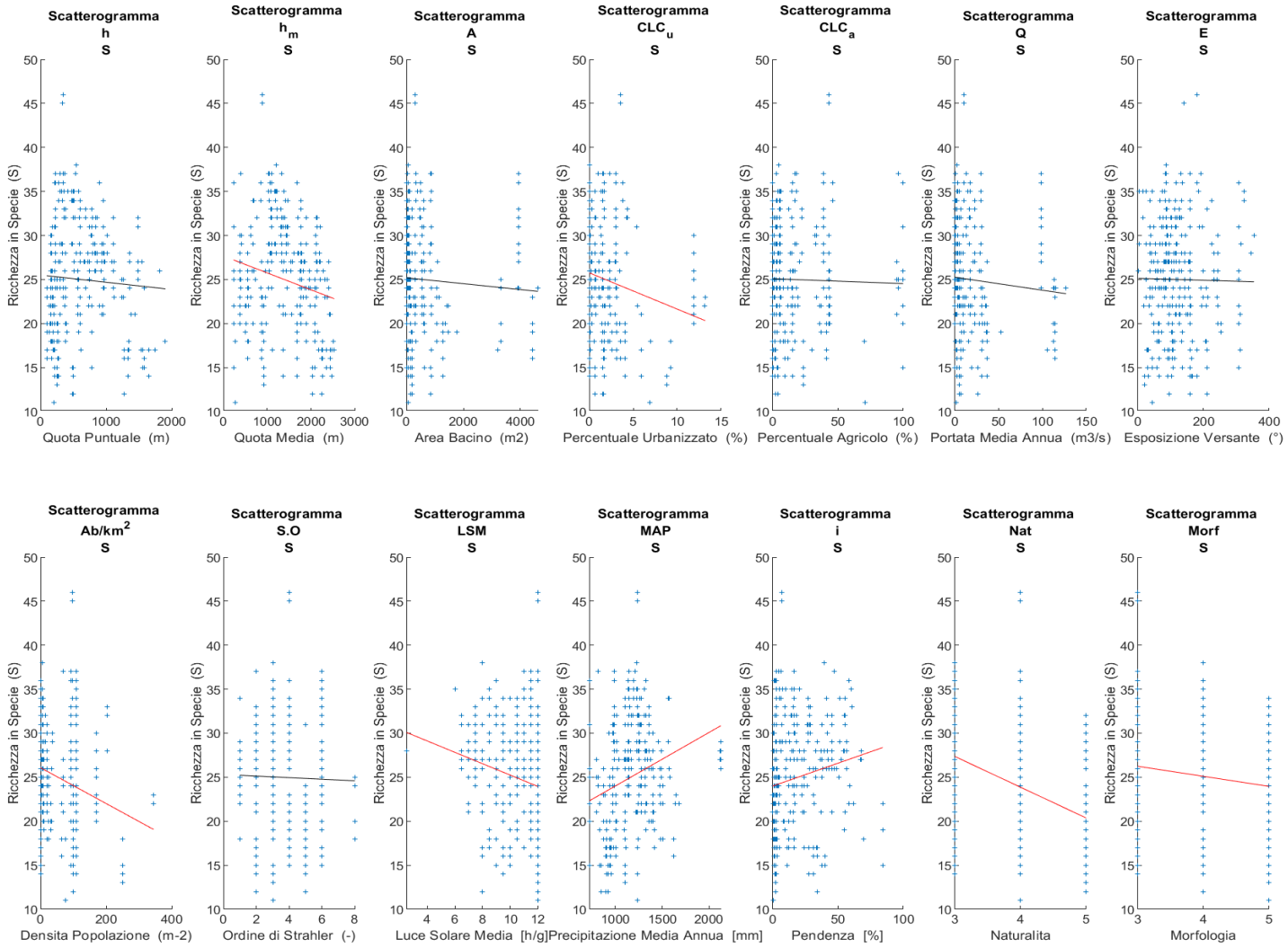

Figura 5.3 *– Regressione lineare semplice. Correlogrammi e rette di regressione per il parametro S (sono colorate in rosso le rette che passano il t-test).*

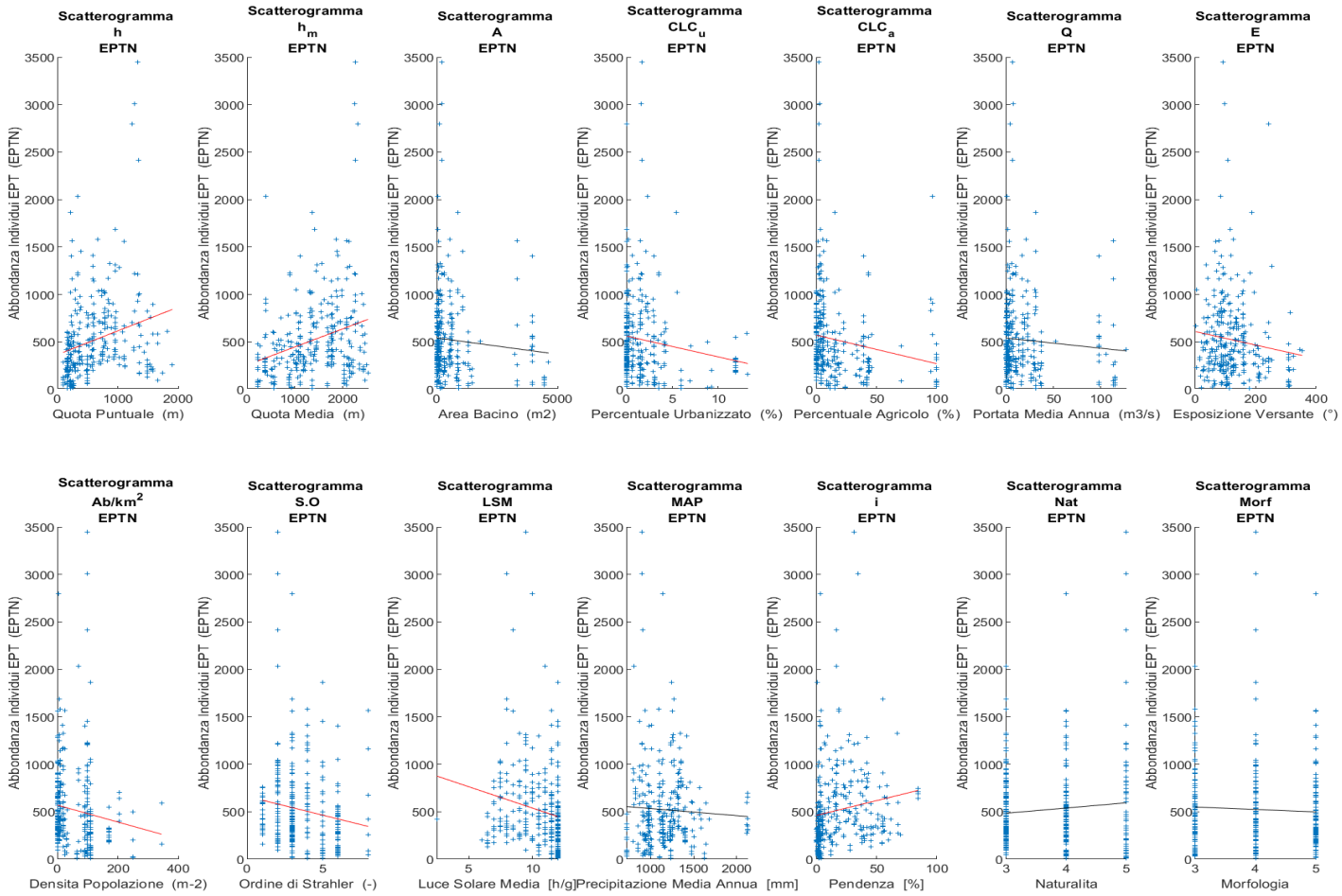

Figura 5.4 *– Regressione lineare semplice. Correlogrammi e rette di regressione per il parametro EPTN (sono colorate in rosso le rette che passano il t-test).*

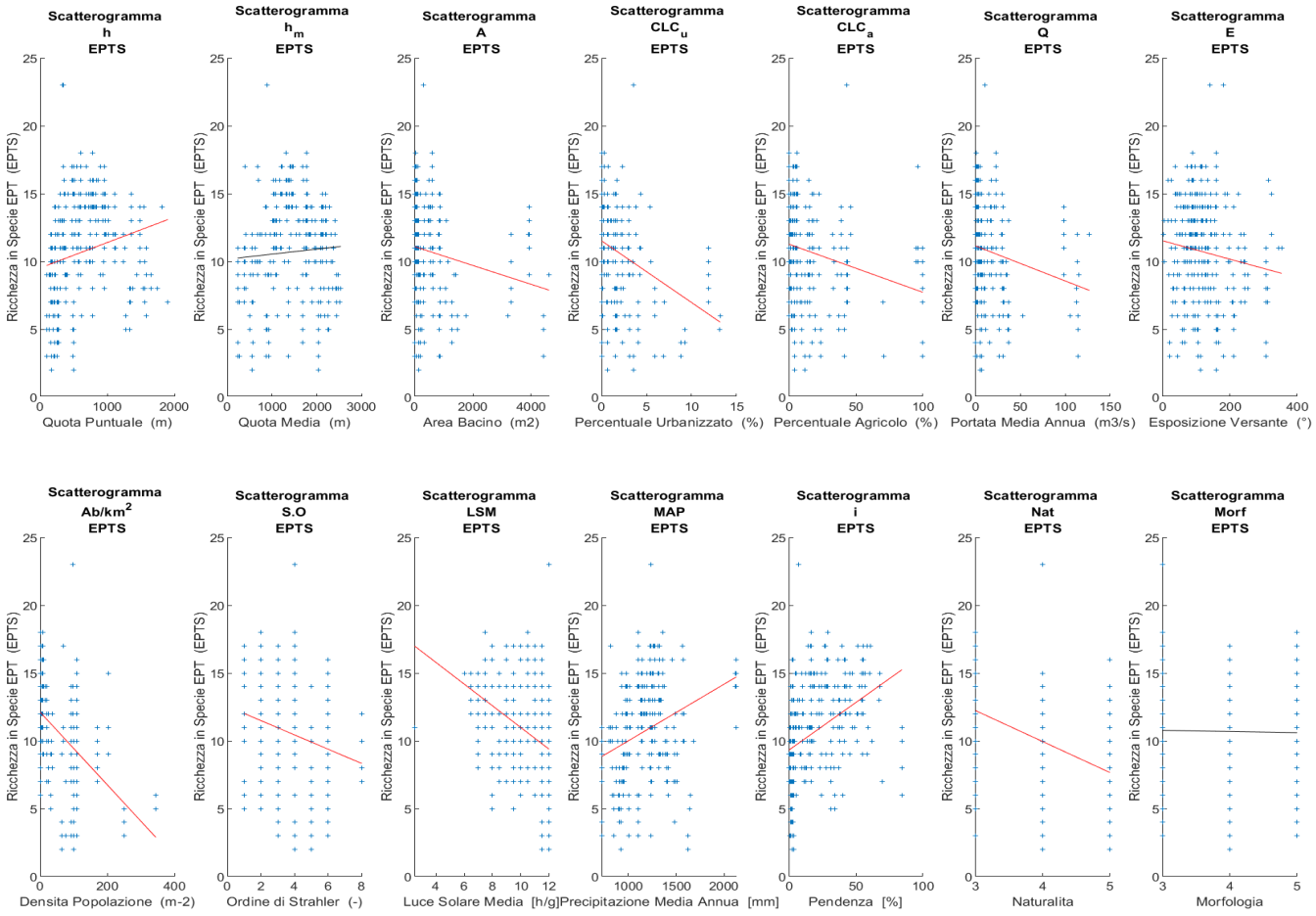

Figura 5.5 *– Regressione lineare semplice. Correlogrammi e rette di regressione per il parametro EPTS (sono colorate in rosso le rette che passano il t-test).*

Il risultato ottenuto per via grafica nelle precedenti figure può essere sintetizzato in un'unica tabella come di seguito.

Tabella 5.1 *– Risultati della regressione lineare semplice. Valori dei coefficienti R^2 ed m per ogni regressione monoparametrica (i valori di m in grassetto indicano le regressioni che hanno superato il test t-Student).*

	PARAMETRI ECOLOGICI							
	N		**S**		**EPTN**		**EPTS**	
PARAMETRI IDRO-MORFOLOGICI	R^2_{adj}	***m***	R^2_{adj}	***m***	R^2_{adj}	***m***	R^2_{adj}	***m***
h	0.008	-0.246	0.000	-0.001	0.057	***0.250***	0.046	***0.002***
h_{media}	0.007	-0.154	0.035	***-0.002***	0.073	***0.190***	0.001	0.000
A_{bacino}	0.043	***0.203***	0.000	0.000	0.003	-0.034	0.036	***-0.001***
Q_{annua}	-0.003	-8.409	0.033	***-0.410***	0.018	***-21.866***	0.128	***-0.453***
$CLC_{urbanizzato}$	0.108	***13.453***	-0.003	-0.006	0.025	***-3.004***	0.052	***-0.035***
$CLC_{agricolo}$	0.037	***6.931***	0.001	-0.015	0.002	-1.066	0.037	***-0.026***
Esposizione Versante	0.024	***2.273***	-0.003	-0.001	0.010	***-0.712***	0.014	***-0.007***
Abitanti/kmq	0.013	***2.010***	0.037	***-0.020***	0.013	***-0.888***	0.203	***-0.027***
Ordine di Strahler	0.081	***176.937***	-0.003	-0.090	0.018	***-39.481***	0.051	***-0.527***
Luce Solare Media	0.011	***66.494***	0.029	***-0.646***	0.028	***-44.939***	0.142	***-0.800***
MAP	0.055	***-0.899***	0.061	***0.006***	-0.001	-0.075	0.087	***0.004***
Pendenza	0.015	***-6.659***	0.024	***0.052***	0.017	***3.114***	0.142	***0.070***
Naturalità	0.028	***240.356***	0.160	***-3.493***	0.005	54.814	0.203	***-2.283***
Morfologia	0.078	***-345.696***	0.019	***-1.141***	-0.001	-25.845	-0.003	-0.085

I valori di *m* in grassetto sono quelli corrispondenti alle regressioni che hanno passato il t-test. In particolare, è stato eseguito un test a due code, considerando un α = 5% (*n* pari a 320). Il valore corrispondente è stato ottenuto mediante le tabelle della distribuzione di Student ed è risultato pari a t_{400} = 1.97.

Per quanto riguarda il valore degli altri parametri (compreso il valore di t_0), sono state implementate in Matlab le formula descritte nella teoria (inserite all'interno del doppio ciclo for). I valori calcolati sono stati poi esportati su excel ed organizzati in modo da formare la ***Tabella 5.1***.

Da un'osservazione della tabella, si nota come i valori di R^2 ottenuti attraverso un modello lineare monoparametrico non siano in generale soddisfacenti. Nella stragrande maggioranza dei casi, si arriva, a malapena, al 10%. Questo era immaginabile. Infatti, nei fenomeni naturali, il numero di variabili in gioco è esteso ed estremamente vario e, pertanto, non sarebbe realistico condensare l'azione di tutto questo ventaglio di variabili in una unica.

Di conseguenza, si è scelto di passare ad un livello successivo, attraverso un approccio di regressione lineare multiparametrica.

5.4 Regressione Lineare Multipla

Dopo aver constatato l'inefficienza dei modelli lineari ad una variabile nel rappresentare l'andamento di fenomeni altamente stocastici, quali quelli eco-idraulici, si è scelto di aumentare il livello di complessità delle analisi.

I modelli scelti sono sempre di carattere lineare, per le ragioni legate al confronto con articoli scientifici certificati e riconosciuti spiegate in precedenza, ma è stato aumentato il numero di parametri da tenere in considerazione durante la fase di regressione. Per questa ragione, è stata effettuata un'analisi di *regressione lineare multipla*.

Si tratta sempre di modelli di regressione, per cui il punto di partenza sono sempre i dati raccolti in campo (quindi, in questo caso, le quattro metriche ecologiche). Quello che cambia è il numero di variabili inserite nel modello lineare, che sarà superiore ad uno. Si vedrà, nel proseguire di questo capitolo, secondo quali criteri vengano scelte queste variabili e le relazioni di indipendenza che devono soddisfare tra di loro (problema della *collinearità*).

La necessità di passare da un'analisi monoparametrica ad una multiparametrica nasce dalle ultime osservazioni fatte nel precedente paragrafo (quando sono stati commentati i risultati della regressione semplice). I valori dei vari coefficienti R^2 risultano estremamente bassi e rasentano a malapena l'ordine del 10% (che non è per niente significativo statisticamente).

Il nuovo obiettivo è, dunque, provare a verificare se, con l'aggiunta di più variabili dentro ad un singolo modello, aumenti anche il livello di significatività della regressione.

In particolare, si vedranno, dapprima, due tipologie di modello di regressione multipla: un modello lineare (in cui i termini della regressione sono sommati tra di loro) ed uno, cosiddetto, esponenziale (in cui i termini sono inseriti all'interno di un prodotto). Poi, si mostrerà un altro risultato della stessa regressione, effettuato però su un campione ristretto dei dati iniziali, limitato su due zonazioni di natura geografica (fatte sul bacino del Po e sul bacino del Tanaro).

Tuttavia, prima dell'applicazione ai dati reali, verrà fatta, come nel precedente paragrafo, una breve introduzione contenente dei cenni teorici alla matematica che sta dietro al metodo e ai problemi di selezione delle variabili su cui basare la costruzione del modello finale.

Si vedrà, anche, che la tipologia di regressione effettuata rappresenta un'affinazione del metodo generale di regressione multipla, in cui ogni variabile inserita nel modello rappresenta quella che porta un contributo maggiore in termini di R^2. Si parla di *analisi lineare multivariata*.

In analogia a quanto specificato per la teoria sulla regressione monoparametrica, si precisa, nuovamente, che la teoria matematica esposta nel seguito (*paragrafo 5.4.1* e relativi sottoparagrafi) è stata tratta dal libro *Introduction to Linear Regression Analysis - Fifth Edition* (Montgomery, Peck, Vining, 2012).

5.4.1 Cenni Teorici – Modello di regressione lineare multipla

Un modello lineare che coinvolge una variabile riposta e più di un regressore (in questo caso, più di una variabile idromorfologica) è detto di *regressione multipla*. La sua forma generale può essere la seguente

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \tag{5.18}$$

in cui i vari x_i rappresentano le variabili della regressione, i β_i rappresentano i coefficienti della regressione (da stimare) e ε rappresenta, come nel caso di regressione semplice, la componente di errore casuale del modello.

Se, nel caso di regressione monoparametrica, era possibile visualizzare geometricamente il modello di regressione (si trattava infatti di una retta su un grafico bidimensionale), non è sempre così nel caso di regressione multipla. La visualizzazione è possibile solo nel caso di due regressori, in cui la superficie di interpolazione (non è più una linea) rappresenta un piano nello spazio tridimensionale. Qualora, come nella maggior parte dei casi, si superasse questo numero (da 3 variabili in poi) la superficie assumerebbe la natura di un iperpiano in uno spazio k-dimensionale generato dalle variabili x_i (e quindi, non visualizzabile ad occhio).

Nuovamente, è interessante notare il significato matematico dei coefficienti β_i. Un singolo coefficiente rappresenta il cambiamento atteso nella variabile risposta *y* per variazione unitaria di regressore x_i, quando tutte le rimanenti variabili vengono mantenute costanti. Per questa ragione, i parametri β_i vengono spesso anche chiamati *coefficienti di regressione parziali*.

5.4.1.1 Multicollinearità e scelta delle variabili

Una delle fasi più critiche riguardante i modelli di regressione multipla è quella della scelta delle variabili da includere all'interno del modello. Infatti, la scelta di variabili poco rappresentative o linearmente dipendenti tra di loro può portare alla creazione di modelli che hanno un debole legame coi dati iniziali.

Come si è già accennato all'inizio del *paragrafo 5.3*, il problema della multicollinearità è estremamente critico in fase di costruzione di un modello di regressione. Con *multicollinearità* si intende la dipendenza lineare tra due o più tra le variabili scelte da inserire nel modello.

Come si vedrà nel successivo paragrafo, in cui viene mostrato il metodo di stima dei coefficienti di regressione del modello, la (5.18) può essere riscritta in forma matriciale nel seguente modo

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

in cui ***X*** rappresenta la matrice dei regressori e **y**, **β**, **ε** i vettori di variabili risposta, coefficienti di regressione e componenti di errore.

Nel corso della stima dei parametri, la matrice ***X*** dovrà essere resa quadrata e, poi, invertita, passando per il calcolo del prodotto per la matrice trasposta ***X'X***, che risulterebbe una matrice singolare (in presenza di variabili correlate), cioè in cui alcune righe/colonne sono linearmente dipendenti le une dalle altre.

Questo problema matematico impatta in maniera estremamente negativa sulla capacità esplicativa del modello. Infatti, quando una variabile risulta legata linearmente ad un'altra, essa non aggiunge un contributo totalmente nuovo al modello, in quanto una parte del suo apporto nello spiegare la varianza iniziale dei dati risulta già fornito dalla variabile alla quale è dipendente.

Questo problema è inevitabilmente legato alla precisione con cui vengono, poi, stimati i coefficienti di regressione β_i.

Per questo motivo, in fase di scelta delle variabili, è sempre preferibile utilizzare quelle che risultano più *ortogonali* tra loro. Si usa il termine ortogonali per il significato che assume in geometria analitica. Infatti, due vettori ortogonali sono anche linearmente indipendenti (ossia, in parole povere, significa che uno dei due vettori non può essere scritto in forma di combinazione lineare dell'altro).

In un'analisi statistica, il concetto di *ortogonalità* può essere tradotto in termini di bassa correlazione tra variabili. Dunque, se si eseguono tutte le cross-correlazioni tra le variabili ambientali, ossia i regressori (come verrà mostrato nel seguito di questo capitolo), e si calcolano i valori di R^2, allora le variabili che si preferirà usare saranno quelle la cui correlazione genera un valore di R^2 più basso.

5.4.1.2 Minimi Quadrati – Stima dei coefficienti di regressione parziali

La stima dei parametri della regressione multipla viene eseguita in maniera analoga al caso di regressione lineare semplice. Si deve sempre partire dai dati disponibili e il metodo più efficace è sempre quello dei *minimi quadrati*.

Assumendo che siano stati scelti i regressori migliori (secondo il criterio spiegato nel precedente paragrafo), si può scrivere la seguente forma del modello di regressione

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \\ &= \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i, \quad i = 1, 2, \ldots, n \end{aligned} \tag{5.19}$$

che è come riscrivere la (5.18).

Si definisce, come nel caso di modello monoparametrico, la funzione dei minimi quadrati *S*, ossia la sommatoria dei quadrati degli scarti tra modello e dati iniziali

$$S(\beta_0, \beta_1, \ldots, \beta_k) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2 \tag{5.20}$$

e si calcolano le derivate rispetto ai parametri β_i , in modo da ottenere il minimo della funzione *S*. I parametri β_i devono dunque rispettare le seguenti equazioni

$$\left. \frac{\partial S}{\partial \beta_0} \right|_{\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_j x_{ij} \right) = 0 \tag{5.21a}$$

$$\left.\frac{\partial S}{\partial \beta_j}\right|_{\hat{\beta}_0,\hat{\beta}_1,\ldots,\hat{\beta}_k} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k}\hat{\beta}_j x_{ij}\right)x_{ij} = 0, \quad j = 1, 2, \ldots, k \tag{5.21b}$$

Semplificando ed esplicitando i vari termini delle precedenti equazioni (5.21), si ottengono le *equazioni normali dei minimi quadrati*

$$\begin{aligned}
&n\hat{\beta}_0 + \hat{\beta}_1\sum_{i=1}^{n}x_{i1} + \hat{\beta}_2\sum_{i=1}^{n}x_{i2} + \cdots + \hat{\beta}_k\sum_{i=1}^{n}x_{ik} = \sum_{i=1}^{n}y_i \\
&\hat{\beta}_0\sum_{i=1}^{n}x_{i1} + \hat{\beta}_1\sum_{i=1}^{n}x_{i1}^2 + \hat{\beta}_2\sum_{i=1}^{n}x_{i1}x_{i2} + \cdots + \hat{\beta}_k\sum_{i=1}^{n}x_{i1}x_{ik} = \sum_{i=1}^{n}x_{i1}y_i \\
&\quad\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
&\hat{\beta}_0\sum_{i=1}^{n}x_{ik} + \hat{\beta}_1\sum_{i=1}^{n}x_{ik}x_{i1} + \hat{\beta}_2\sum_{i=1}^{n}x_{ik}x_{i2} + \cdots + \hat{\beta}_k\sum_{i=1}^{n}x_{ik}^2 = \sum_{i=1}^{n}x_{ik}y_i
\end{aligned} \tag{5.22}$$

La soluzione di queste equazioni rappresenta il valore dei parametri β_i.

Quando si tratta di regressione multipla, è, però, conveniente usare la notazione matriciale per rappresentare i passaggi del metodo dei minimi quadrati. Dunque, la (5.19) può essere scritta, in maniera compatta, nel modo seguente

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

in cui

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

In generale, ***y*** rappresenta il vettore dei dati inziali, ***X*** la matrice dei regressori, **β** il vettore dei coefficienti e **ε** il vettore degli errori.

Si vogliono trovare i valori di **β** che minimizzano la funzione ***S*** dei minimi quadrati

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n}\varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

che può essere scritta anche come

$$\begin{aligned}
S(\boldsymbol{\beta}) &= \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\
&= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}
\end{aligned}$$

Le equazioni dei minimi quadrati in forma matriciale diventano

$$\left.\frac{\partial S}{\partial \boldsymbol{\beta}}\right|_{\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0}$$

e si semplificano in

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \tag{5.23}$$

La (5.23) rappresenta la forma matriciale delle *equazioni normali dei minimi quadrati* ed è una maniera analoga di scrivere la (5.22). Per risolvere queste equazioni, è necessario poter invertire la matrice ***X'X*** per poter moltiplicare entrambi i membri per $\boldsymbol{(X'X)^{-1}}$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{5.24}$$

ottenendo la stima corretta dei coefficienti di regressione.

Il discorso ha matematicamente senso se la matrice $\boldsymbol{(X'X)^{-1}}$ esiste. Perché la matrice ***X'X*** risulti invertibile, bisogna che i regressori scelti all'inizio siano linearmente indipendenti. In questo modo, nessuna colonna della matrice sarà una combinazione lineare delle altre e, dunque, ***X'X*** risulta non singolare (e invertibile).

Il modello finale, che si ottiene, assume, dunque, la seguente forma

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y} \tag{5.25}$$

in cui è stata definita la matrice ***H***, che effettua la trasformazione da variabile osservata a variabile calcolata attraverso il modello.

5.4.2 Applicazione ai dati ecologici

Una volta compresa la teoria alla base del metodo di regressione, si è trattato di applicare i concetti appresi ai dati che si avevano a disposizione.

È stato seguito un procedimento che rispecchiasse quello che è stato spiegato, per ora, solo a livello teorico: dapprima, sono state scelte le variabili secondo il criterio dell'ortogonalità, poi è stata eseguita la regressione multipla. In particolare, è stato applicato un particolare tipo di regressione, che verrà descritto nel seguito, che andasse ad affinare i risultati ottenibili con la normale regressione multipla: la regressione multivariata. Da questo procedimento si otterranno le due tipologie di modelli già accennate (lineare ed esponenziale).

Infine, verranno descritte le due zonazioni.

Lo scopo principale di tutte queste elaborazioni non è stato altro che studiare la variazione del coefficiente di correlazione R^2 all'aumentare del livello di dettaglio preso in considerazione (aggiunta di variabili al modello, creazione di più tipologie di modello e omogeneizzazione dei dati).

5.4.2.1 Scelta delle variabili – Correlazioni tra variabili idromorfologiche

La definizione delle variabili di regressione è stata fatta seguendo il criterio definito in precedenza, quello dell'*ortogonalità*, che, in statistica, si traduce attraverso il calcolo delle correlazioni tra variabili.

È stato quindi creato uno script matlab che andasse a calcolare tutte le correlazioni tra le variabili idromorfologiche presentate nel *Capitolo 4*. In realtà, questo script non è stato creato da nulla, è stato semplicemente modificato quello già creato in fase di analisi lineare semplice. Infatti, si tratta, anche in questo caso, di calcolare la correlazione tra due singoli parametri.

In sostanza, è stato solo messo, al posto della matrice coi valori dei parametri ecologici, quella coi parametri idromorfologici. In questo modo, invece di correlare, ad ogni giro del ciclo, un parametro idromorfologico ed uno ecologico, si effettua la correlazione tra soli parametri idromorfologici.

Il codice completo e commentato è riportato come *Allegato 4* alla fine della dissertazione.

Il risultato finale dello script è sempre il valore del coefficiente R^2 di ogni singola correlazione, in questo caso non serve il valore m di inclinazione della retta di regressione (si sta studiando la collinearità, è sufficiente il solo parametro R^2). Riportando ed ordinando i valori di R^2 si può ottenere un risultato come quello mostrato in ***Tabella 5.2***.

Tabella 5.2 *– Multicollinearità. Valori di R^2 risultanti dalle correlazioni tra variabili idromorfologiche. I colori delle celle corrispondono alle classi di correlazione (R^2) definite.*

	h	h_{media}	A_{bacino}	Q_{annua}	$CLC_{urbanizzato}$	$CLC_{agricolo}$	Esposizione Versante	Abitanti/kmq	Ordine di Strahler	Luce Solare Media	MAP	Pendenza	Naturalità	Morfologia
PARAMETRI IDRO-MORFOLOGICI	R^2_{adj}	R^2_{adj}	R^2_{adj}	R^2_{adj}	R^2_{adj}	R^2_{adj}	R^2_{adj}	R^2_{adj}	R^2_{adj}	R^2_{adj}	R^2_{adj}	R^2_{adj}	R^2_{adj}	R^2_{adj}
h	1.00	0.59	0.17	0.23	0.29	0.18	0.07	0.29	0.32	0.49	0.00	0.369802	0.0464733	0.1174022
h_{media}	0.59	1.00	0.00	0.18	0.51	0.00	0.09	0.17	0.03	0.20	0.03	0.138214	-0.000939	0.0283371
A_{bacino}	0.17	0.00	1.00	0.03	0.03	0.97	0.01	0.06	0.43	0.15	0.08	0.117562	0.0119798	0.0716749
Q_{annua}	0.23	0.18	0.03	1.00	0.11	0.03	0.01	0.64	0.01	0.20	0.01	0.171536	0.0930885	0.000873
$CLC_{urbanizzato}$	0.29	0.51	0.03	0.11	1.00	0.02	0.13	0.20	0.12	0.22	0.11	0.191391	0.0119201	0.1043072
$CLC_{agricolo}$	0.18	0.00	0.97	0.03	0.02	1.00	0.01	0.05	0.49	0.15	0.07	0.123557	0.0128198	0.0633541
Esposizione Versante	0.07	0.09	0.01	0.01	0.13	0.01	1.00	0.03	0.04	0.04	0.01	0.067567	0.0341317	0.0273017
Abitanti/kmq	0.29	0.17	0.06	0.64	0.20	0.05	0.03	1.00	0.13	0.30	0.02	0.217994	0.1579767	0.0497014
Ordine di Strahler	0.32	0.03	0.43	0.01	0.12	0.49	0.04	0.13	1.00	0.29	0.21	0.253473	0.04337	0.2571939
Luce Solare Media	0.49	0.20	0.15	0.20	0.22	0.15	0.04	0.30	0.29	1.00	0.07	0.521001	0.1009936	0.0735568
MAP	0.00	0.03	0.08	0.01	0.11	0.07	0.01	0.02	0.21	0.07	1.00	0.083987	0.0990877	0.1020382
Pendenza	0.369802	0.1382141	0.117562	0.1715	0.1913909	0.123557147	0.067566686	0.217994328	0.253472538	0.521001078	0.083987	1	0.1071082	0.0474437
Naturalità	0.046473	-0.000939	0.01198	0.0931	0.0119201	0.012819776	0.034131707	0.157976699	0.043369962	0.10099362	0.099088	0.107108	1	0.018723
Morfologia	0.117402	0.0283371	0.071675	0.0009	0.1043072	0.063354073	0.027301698	0.049701393	0.257193902	0.073556767	0.102038	0.047444	0.018723	1

Si nota come, lungo la diagonale principale della tabella, siano presenti tutti valori unitari (in quanto la correlazione di una variabile con sé stessa è ovviamente massima).

Sono stati poi definiti quattro intervalli di valore del coefficiente R^2, in modo da ottenere quattro classi di correlazione tra variabili.

Tabella 5.3 *– Multicollinearità. Definizione delle quattro classi di* R^2

Colore	Intervallo	Collinearità
	$0 < R^2_{adj} < 0.2$	Assente
	$0.2 < R^2_{adj} < 0.4$	Bassa
	$0.4 < R^2_{adj} < 0.6$	Media
	$R^2_{adj} > 0.6$	Alta

Valori di R^2 inferiori a 0.2 sono stati considerati come ininfluenti dal punto di vista della correlazione. Poi è stato possibile raccogliere le differenti coppie di variabili ricadenti nelle varie classi appena definite. Il risultato è quello presentato in ***Tabella 5.4***.

Tabella 5.4 *– Multicollinearità tra variabili (coppie di variabili ricadenti nelle classi di correlazione definite)*

Altamente correlate	Mediamente correlate	Poco correlate
A bacino - CLC_a	h_{media} - h	h - Q
Q_{annua} - Ab/km2	Luce solare media - h	h - CLC_u
	h_{media} - CLC_u	h - ab/km^2
	A_{bacino} - SO	h - SO
	CLC_a - SO	CLC_u - LSM
	LSM - Pendenza	Ab/km2 - LSM
		SO - LSM
		MAP - SO
		Ab/km^2 - Pendenza
		h - Pendenza
		SO - Pendenza
		Morf - SO

Si nota come la maggior parte delle variabili ricada nel campo della medio-bassa correlazione (che è un risultato positivo). Ovviamente, non sono stati riportati tutti i termini della diagonale principale (che sarebbero ricaduti nella classe di variabili altamente correlate).

Mediante la costruzione di questa ***Tabella 5.4***, la successiva definizione delle variabili di interesse per la regressione è stata particolarmente agevole.

Come prima cosa, sono state scartate tutte le coppie di variabili ricadenti nella classe di alta correlazione (gialla). Si è passati direttamente a ragionare dalla media ($R^2 < 0.4$).

Quello che si è fatto è considerare una coppia di variabili alla volta e sceglierne una delle due da tenere (sulla base dei ragionamenti sulla possibile influenza di una variabile, piuttosto che un'altra, sull'alterazione della vita delle comunità studiate – come spiegato nel *Capitolo 4*, in cui queste variabili sono state descritte). Si è passati così di coppia in coppia, mantenendo, di volta in volta, le variabili già scelte prima e scartando quelle con cui queste (già scelte) risultassero in correlazione.

Il procedimento è stato eseguito, dapprima, sulla classe media e poi sulla classe bassa.

Per rendere tutto più chiaro è conveniente fare un esempio. Partendo dalla classe di media correlazione, ci si trova di fronte alla prima coppia di variabili, *quota media* e *quota puntuale*. Si sceglie, ad esempio, di mantenere la *quota media* (in quanto rappresenta un indicatore generico definito a scala di bacino) e di scartare la *quota puntuale*. Passando alla coppia successiva (*illuminazione solare media* e *quota puntuale*), si dovrà per forza scartare la *quota puntuale*, in quanto, altrimenti, si andrebbe in contrasto con l'assunzione fatta al passaggio precedente. Alla terza coppia (*quota media* e *percentuale di urbanizzato*) si sceglie di mantenere la quota media, per lo stesso ragionamento. Si procede così, finché non si esauriscono le coppie da valutare.

L'applicazione completa di questa procedura, porta alla scelta delle seguenti otto variabili :

- quota media (h_m);
- luce solare media giornaliera (LSM);
- ordine di strahler (SO);
- precipitazione media annua (MAP);
- esposizione del versante (E);
- portata media annua (Q);
- indice di naturalità del fiume (Nat);
- indice di morfologia del fiume (Morf).

Sono quelle su cui verrà eseguita la successiva regressione multivariata.

5.4.2.2 Analisi Multivariata – Modello Lineare

5.4.2.2.1 Metodo

Si definisce regressione lineare multivariata un particolare tipo di regressione multipla, eseguita in modo tale da inserire nel modello una variabile alla volta tra quelle risultanti dal processo di selezione (ossia le otto appena elencate, nel precedente paragrafo). Si tratta di un processo sequenziale, in cui, ad ogni passaggio, viene aggiunta una variabile.

Il metodo procede per fasi successive, creando un modello che si arricchisce di informazioni (ossia di variabili inserite) ad ogni successivo passaggio della regressione.

Il primo passo ripercorre quello che è già stato fatto in precedenza, ossia è la regressione lineare semplice (ad un parametro). Si eseguono le correlazioni monoparametriche, secondo tutte le

combinazioni possibili tra variabile risposta (ecologica) e variabile esplicativa (ambientale), e alla fine si decide di definire, come prima variabile da inserire nel modello, quella dalla cui regressione è stato determinato il coefficiente R^2 con valore maggiore.

In questo modo si arriva a disporre di un modello preliminare del tipo

$$y = \beta_0 + \beta_1 x_1$$

in cui β_0 è l'intercetta del modello, mentre β_1 e x_1 rappresentano il coefficiente di regressione e la variabile di regressione, che corrispondono alla massima correlazione (in termini di R^2).

A questo punto, non si parte più da zero, ma si cerca di arricchire il modello di altre informazioni, tenendo in considerazione quello che si è appena trovato. Dunque, si cerca, adesso, un modello a due variabili di questo tipo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

in cui x_1 è la variabile di prima (perché appunto è già stata assunta come dato, nel modello), mentre β_0, β_1, β_2 e x_2 rappresentano intercetta, coefficiente di regressione della prima variabile, coefficiente di regressione della seconda variabile e seconda variabile di regressione, corrispondenti alla regressione (multipla, questa volta) che determina il valore maggiore di R^2.

In sostanza, tenendo fissa la variabile assunta al passaggio precedente (la x_1), sono state eseguite tutte le regressioni multiple a due parametri, variando ogni volta la nuova variabile x_2 inserita nel modello, e alla fine è stata assunta, come nuova variabile da aggiungere al modello, quella che ha prodotto una regressione con coefficiente di correlazione massimo.

Di fatto, il procedimento è sempre questo. Ad ogni passaggio, cioè ogni volta che si vuole aggiungere una nuova variabile al modello, si ripete quello che è appena stato descritto.

Per esempio, ad adesso, si è arrivati ad un modello a due variabili. Per aggiungerne una terza, il metodo è analogo. Si tengono fisse le due variabili già assunte, si varia la terza (eseguendo quindi delle regressioni multiple a tre parametri) e alla fine si inserisce nel modello la variabile che ha prodotto la regressione più significativa in termini di R^2. E via così per la quarta, quinta, sesta, settima e ottava variabile (come nel caso reale a cui questo metodo sarà applicato).

Al termine del procedimento di *analisi multivariata*, è possibile costruire un grafico riepilogativo dell'andamento degli R^2, in funzione del numero di variabili inserite nel modello. A seconda della forma della curva che si ottiene, della sua concavità, è possibile avanzare delle ipotesi su quali siano le variabili maggiormente influenti nella regressione (e, quindi, maggiormente influenti nella spiegazione della distribuzione dei parametri ecologici).

Se, per esempio, la concavità della curva risulta verso il basso, allora significa che il grosso delle informazioni lo catturo già solo con le prime variabili inserite nel modello, senza la necessità di aggiungerne altre (che non porterebbero nessun contributo significativo alla regressione, appesantirebbero solo il modello).

Se, invece, risultasse una concavità verso l'alto, allora significherebbe che il coefficiente R^2 sta continuando a crescere man mano che si aggiungono variabili. Quindi, avrebbe senso supporre che

l'aggiunta al modello di ulteriori variabili non ancora considerate potrebbe portare alla deduzione di nuove informazioni (in quanto la significatività del modello di regressione sta crescendo).

A prescindere dalla forma del grafico, potrebbe anche succedere che, alla fine della regressione, il valore dell'ultimo R^2 calcolato sia comunque basso. In questo caso, sarebbe obbligatorio ritornare sui propri passi e cercare di trovare nuove informazioni da aggiungere al modello.

5.4.2.2.2 Descrizione dello Script Matlab

Quello che si è fatto è stato tradurre in un codice Matlab la sequenza di passaggi appena descritti. In matlab, purtroppo, non esiste una funzione che esegua la regressione multivariata così come è stata definita nel paragrafo precedente. Esiste solo la funzione *regress* del toolbox *Statistics and Machine Learning*, che però esegue la regressione multipla prendendo tutti le variabili insieme e restituendo i valori dei coefficienti di regressione.

In questo paragrafo ci si propone di dare un'idea generale del funzionamento dello script (che alla fine non è altro che un ripercorrere i passaggi del metodo di analisi multivariata). Il codice Matlab completo e commentato viene riportato in *Appendice*, come *Allegato 5*.

Il primo passo è sempre l'importazione dei dati dal database excel e la creazione delle matrici di lavoro (quella coi parametri ecologici e quella coi parametri idromorfologici). Poi è importante definire anche una variabile contenente i nomi delle otto variabili derivanti dal processo di selezione. Essa servirà a riconoscere quali variabili sono già state assunte nel modello e quali non ancora, nei successivi cicli a cascata delle regressioni.

Il primo step è quello di aggiungere al modello la prima variabile. Quindi, è stato creato un doppio *ciclo for*, in modo tale da tenere fisso nel ciclo esterno il parametro ecologico e fare variare, in quello interno, ad ogni giro, il parametro di regressione (idromorfologico).

Qui però è sorto il primo problema. Per poter eseguire la regressione multipla correttamente, il comando *regress* ha bisogno che gli si fornisca il vettore delle variabili risposta (e questo non è un problema, perché è stato importato da excel nella fase di estrazione dei dati dal database) e una matrice contenente una prima colonna fatta di soli valori 1 e le successive colonne con i valori delle variabili da usare nella regressione multipla (la colonna di 1 è dovuta al fatto di voler generare un modello con intercetta diversa da zero). Quindi, prima del doppio *ciclo for*, è stato creato un *ciclo while* che andasse a generare una nuova matrice costituita nella maniera appena descritta (prima la colonna di 1 e poi le variabili).

Creata questa matrice, è stato possibile eseguire correttamente la prima regressione, che in questo primo passaggio è stata ad un solo parametro, e di calcolare i valori di R^2 ad ogni ciclo. Alla fine del ciclo interno (ossia alla fine del calcolo di tutte le regressioni per il primo parametro ecologico – si ricorda che ce ne sono quattro totali, una per ogni ciclo esterno) è stato valutato il valore massimo tra gli R^2 e salvato in un'apposita variabile di lavoro. Poi, è stata salvata pure la sua posizione, in termini di indice, all'interno della stringa coi nomi delle variabili definita in precedenza. Questo servirà, nei passaggi successivi, a riconoscere quali variabili sono già state assunte e quali lo devono ancora essere.

Ovviamente, al termine del ciclo for esterno, ci saranno quattro valori di R^2 massimo e quattro indici di posizione della variabile estratta, uno per ogni parametro ecologico (si sta usando lo stesso procedimento, in contemporanea per tutti e quattro i parametri ecologici).

Definita la prima variabile assunta nel modello, è stato necessario procedere col trovare un modo di dire al software di tenere fissa, nelle successive regressioni multiple (doppio ciclo for), questa variabile e far variare le altre (non ancora assunte), in modo da creare quella successione di modelli a complessità crescente che sono stati descritti (in cui, appunto, viene aggiunta una varabile alla volta).

Vengono quindi definite, in base a quale fase dell'analisi si è giunti (regressione a doppia, tripla etc..), un numero di matrici pari al numero di variabili già assunte nel modello nei passaggi precedenti. Ad esempio, se si è in fase di regressione multipla a due parametri, viene definita una sola matrice (in quanto, nei passaggi precedenti, è stata estratta solo una variabile). In fase di regressione a tre parametri vengono definite due matrici (in quanto dai passaggi precedenti sono state estratte due variabili): una è quella già definita all'inizio della parte di script dedicata alla regressione multipla a due variabili (per cui è già pronta), l'altra è una nuova, che si ottiene grazie al vettore con gli indici di posizione salvati alla fine del doppio ciclo for della regressione a due variabili (sarebbe quella uscente dal ciclo precedente, che in questo caso è quello della regressione doppia).

Dunque, in generale, la prima parte di script di ogni step della regressione è composto in questa maniera, in cui si definisce una nuova matrice il cui compito è far capire all'interno del doppio ciclo for quale variabile è già assunta nel modello (una sola ad ogni step, perché le altre sono già definite dallo step precedente).

Si può adesso parlare del doppio ciclo for, che è il cardine di ogni step dello script. È infatti grazie ad esso che, per ogni giro del ciclo interno, è possibile tenere fisso il parametro contenente le variabili già assunte nel modello ai cicli precedenti e far variare solo quelle ancora da considerare.

A questo punto, si ripropone il problema della matrice dei regressori da passare al comando *regress*. Questa, come detto, deve contenere una colonna di uno e poi tutte le variabili (in questo caso, ossia dalla regressione a due parametri in poi) dovrà contenere ad ogni step (di regressione) una variabile in meno rispetto a quello precedente, poiché ad ogni step una viene assunta nel modello. Questa matrice viene composta ad ogni giro del ciclo interno nelle prime righe di script. Infatti, risulta diversa ad ogni giro.

Per capire questo concetto è necessario tenere bene a mente come funziona il comando *regress*. Si consideri il caso di regressione a due variabili. La matrice che verrà composta all'inizio del ciclo interno dovrà essere così: la prima colonna sarà di soli valori 1; la seconda colonna deve tenere fisso (nel ciclo interno) i valori della variabile assunta nello step precedente; la terza colonna (che è l'unica a variare nel ciclo interno) cambia ad ogni giro del ciclo, andando a prendere di volta in volta i valori delle variabili non ancora assunte nel modello (e tra le quali si deve trovare quella che produce un valore di R^2 maggiore al termine di tutte le regressioni). La seconda colonna, invece, varierà solo quando cambia il giro del ciclo esterno (poiché, in quanto rappresenta i valori della variabile già assunta nel modello, essa deve variare solo quando cambia il parametro ecologico che si sta considerando).

In questo modo vengono eseguite tutte le regressioni nella maniera corretta. Nei casi di regressioni multiple a più di due variabili, cambia solo la dimensione della matrice da definire prima del comando *regress*. Nel caso a tre parametri avrà quattro colonne (la prima con gli 1, la seconda e la terza con le due variabili già assunte dalle regressioni ad 1 parametro e a 2, la quarta che varia ad ogni giro sulle sole variabili ancora escluse).

Al termine del ciclo for, vengono sempre salvati i quattro valori di R^2 massimo (per la regressione in corso) e l'indice di posizione (della nuova variabile assunta) riferito alla stringa iniziale delle variabili.

Lo script si ripete in modo uguale per tutte le regressioni multiple (da due a otto variabili), cambiano solo le dimensioni delle varie matrici da definire per l'esecuzione corretta del doppio ciclo for. Lo script finale risulta abbastanza lungo proprio per questo motivo, ossia non è stato possibile automatizzare il tutto poiché i comandi per la creazione automatica delle matrici devono essere aggiornati a mano ad ogni step successivo della regressione.

L'unico cambiamento nella struttura dello script è stato effettuato quando si giunge al doppio ciclo for della regressione ad otto variabili (l'ultimo step). Infatti, in questo caso, è stato necessario salvare anche i valori dei coefficienti di regressione β_i che permetteranno di scrivere il modello ottenuto sotto forma di equazione (come verrà mostrato nel paragrafo dedicato ai risultati della regressione multivariata).

La parte finale dello script è composta dalla creazione dei quattro grafici (uno per ogni variabile ecologica), raffiguranti l'andamento del valore di R^2 in funzione del numero di variabili assunte nel modello, e dai comandi per la trascrizione diretta in excel dei parametri utili: valore di R^2, posizione delle variabili nella stringa iniziale e valori dei coefficienti di regressione β_i.

5.4.2.2.3 Risultati

Il risultato finale di un'analisi lineare multivariata è formato da tre prodotti principali:

- i grafici dell'andamento del coefficiente R^2 in funzione del numero di variabili assunte nel modello;
- le equazioni lineari del modello di regressione multipla (una per ogni variabile risposta);
- i grafici di confronto tra valore misurato e valore previsto dal modello.

I grafici dell'andamento di R^2 servono a tirare le somme della regressione, ossia a capire dove ha portato (in termini di informazioni nuove aggiunte) e con che livello di significatività. Come già spiegato in fase di descrizione teorica del metodo, in base alla forma del grafico si può capire la potenza o meno del modello e anche il suo livello di completezza. In particolare, si capisce quali siano le variabili più influenti all'interno del modello.

Invece, la formulazione analitica delle equazioni di regressione permette di creare un grafico di dispersione che mette a confronto valore misurato e valore calcolato (attraverso il modello) delle quattro variabili risposta.

I grafici del coefficiente R^2 sono riportati nella seguente ***Figura 5.6***.

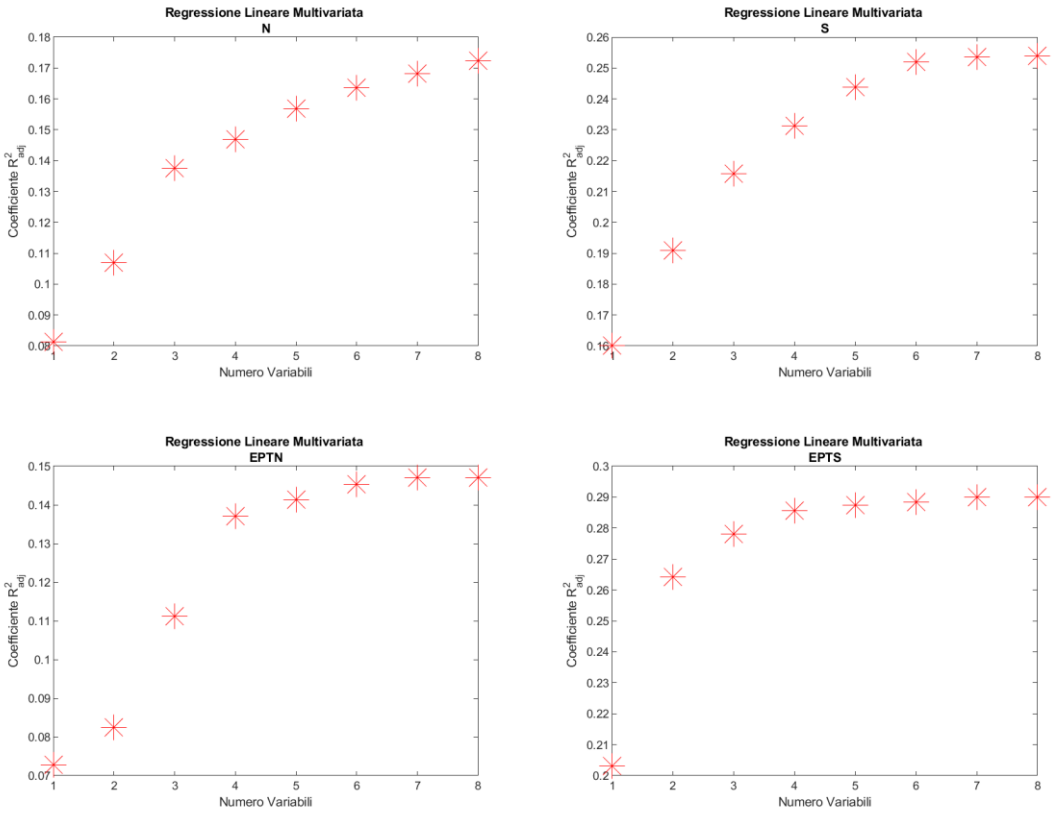

Figura 5.6 *– Grafici con l'andamento del valore del coefficiente R^2 in funzione del numero di variabili aggiunte al modello (ogni grafico rappresenta il risultato della regressione multivariata di una delle quattro variabili ecologiche).*

Da un'analisi di ***Figura 5.6***, si può notare come tutti e quattro i grafici abbiano la tipica forma con la concavità verso il basso. Ciò significa che gran parte dell'informazione viene catturata già nelle prime fasi della regressione (cioè con 4, 5 variabili, a seconda del caso). Infatti, ci si accorge di come, nelle fasi iniziali della regressione (fino a 4,5 variabili), il valore di R^2 cresca velocemente, per poi iniziare a stabilizzarsi, nelle ultime fasi, verso un valore di stazionarietà.

Le stesse informazioni (contenute a livello visivo nei quattro grafici) si possono trovare in ***Tabella 5.5***, dove è stato aggiunto, a fianco di ogni valore di R^2, anche il nome della variabile assunta nel modello a quel passaggio dell'analisi multivariata. Questo modo di visualizzare i risultati permette di osservare quali siano le variabili ambientali che maggiormente influenzano la distribuzione della variabile ecologica studiata.

Tabella 5.5 *– Analisi Multivariata – Valori di R^2 del modello lineare e progressiva aggiunta di variabili al modello*

N		S		EPTN		EPTS	
Variabili Modello	R_{ADJ}^2	**Variabili Modello**	R_{ADJ}^2	**Variabili Modello**	R_{ADJ}^2	**Variabili Modello**	R_{ADJ}^2
S.O	0.08	Nat	0.16	h_m	0.07	Nat	0.20
S.O , Morf	0.11	Nat , h_m	0.19	h_m , S.O	0.08	Nat, LSM	0.26
S.O , Morf, Nat	0.14	Nat , h_m , LSM	0.22	h_m , S.O , Morf	0.11	Nat, LSM, MAP	0.28
S.O , Morf, Nat, LSM	0.15	Nat , h_m , LSM, SO	0.23	h_m , S.O , Morf, Nat	0.14	Nat, LSM, MAP, Morf	0.286
S.O , Morf, Nat, LSM, h_m	0.16	Nat , h_m , LSM, SO, MAP	0.24	h_m , S.O , Morf, Nat, E	0.14	Nat, LSM, MAP, Morf, Q	0.287
S.O , Morf, Nat, LSM, h_m, MAP	0.16	Nat , h_m , LSM, SO, MAP, Morf	0.252	h_m , S.O , Morf, Nat, E, LSM	0.145	Nat, LSM, MAP, Morf, Q, E	0.288
S.O , Morf, Nat, LSM, h_m, MAP, Q	0.168	Nat , h_m , LSM, SO, MAP, Morf, Q	0.2535	h_m , S.O , Morf, Nat, E, LSM, Q	0.1470	Nat, LSM, MAP, Morf, Q, E, SO	0.290
S.O , Morf, Nat, LSM, h_m, MAP, Q, E	0.172	Nat , h_m , LSM, SO, MAP, Morf, Q, E	0.2539	h_m , S.O , Morf, Nat, E, LSM, Q, MAP	0.1471	Nat, LSM, MAP, Morf, Q, E, SO, h_m	0.2901

Si nota come le due variabili qualitative (naturalità e morfologia del fiume) abbiano un ruolo pesante nella definizione del modello. Esse compaiono, infatti, quasi sempre nelle prime fasi della regressione multivariata. Ciò sottolinea il fatto che anche un parametro ottenuto in maniera semplice (come descritto nel *Capitolo 4*) possa portare un contributo significativo alle analisi, se scelto e definito con criterio. In questo caso, viene evidenziata l'importanza della struttura dell'alveo e dell'influenza antropica nella regolazione e nell'adattamento all'habitat del ciclo di vita dei macroinvertebrati.

La stessa importanza vale per ordine di strahler ed illuminazione media diurna, variabili assunte anch'esse alle prime fasi della regressione. L'ordine di strahler, come già detto, rappresenta un indice dell'isolamento di una comunità (e, dunque, influisce sulla possibilità di una comunità di interagire con altre in termini di competizione per le risorse e colonizzazione), mentre l'illuminazione naturale è imprescindibilmente legata al ciclo di vita degli insetti fluviali e alla disponibilità di risorse.

Discorso diverso è quello relativo alla qualità della regressione. Infatti, soffermandosi sul valore numerico del coefficiente R^2 finale ottenuto, si ottengono risultati abbastanza diversi in base alla variabile ecologica considerata. I valori più alti sono sempre relativi alle due variabili relative alla ricchezza in specie (S ed EPTS) , piuttosto che all'abbondanza (N ed EPTN). Anche il valore massimo che si ottiene è abbastanza basso (è dell'ordine del 30%). Tuttavia, se si considerano i valori ottenuti da esperienze di ricerca di caratura internazionale, si può capire come anche un valore di questo ordine di grandezza sia significativo in questo ambito di ricerca (raramente, in campo eco-idraulico, si superano valori di R^2 pari a 0.4 o 0.5).

Si riporta, infine, la forma analitica dei quattro modelli di regressione ottenuti, che hanno permesso la creazione dei quattro grafici di confronto tra previsione teorica e valore misurato in campo delle quattro variabili ecologiche:

1) $N = 2988.55 + 31.98 \cdot S.O - 289.68 \cdot Morf + 282.95 \cdot Nat - 116.4 \cdot LSM - 0.22 \cdot h_m - 0.4 \cdot MAP + 4.56 * Q + 1.02 \cdot E$;

2) $S = 42.63 - 2.37 \cdot Nat - 0.002 \cdot h_m - 0.9 \cdot LSM + 0.8 \cdot SO + 0.003 \cdot MAP - 0.83 \cdot Morf - 0.01 \cdot Q - 0.002 \cdot E$;

3) $EPTN = 972.29 + 0.14 \cdot h_m - 75.09 \cdot S.O - 152.04 \cdot Morf + 131.96 \cdot Nat - 0.36 \cdot E - 23{,}42 \cdot LSM + 0.97 \cdot Q + 0.01 \cdot MAP$;

4) $EPTS = 20.59 - 1.51 \cdot Nat - 0.52 \cdot LSM + 0.002 \cdot MAP - 0.42 \cdot Morf - 0.01 \cdot Q - 0.0022 \cdot E + 0.17 \cdot S.O + (5.09 \cdot 10^{-5}) \cdot h_m$.

Si riportano, in ***Figura 5.7***, i quattro grafici suddetti.

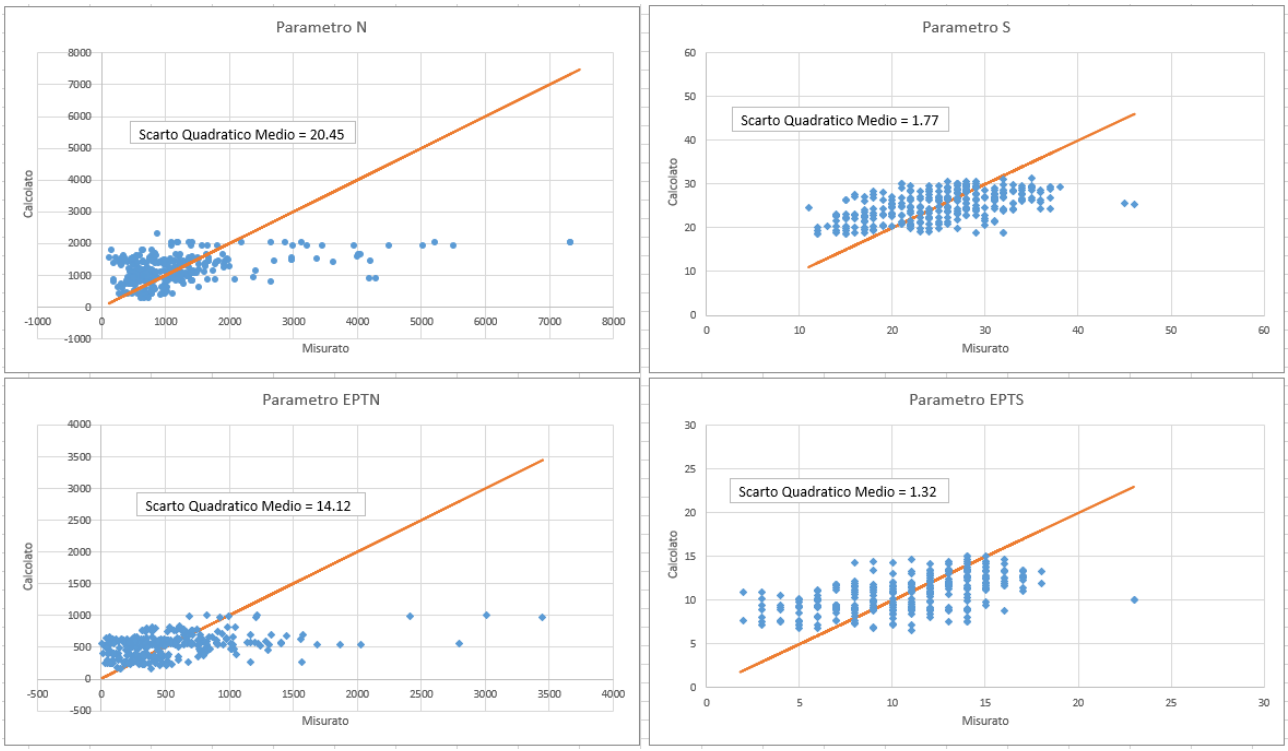

Figura 5.7 *– Grafici di confronto tra valore misurato e previsto dal modello dei quattro parametri ecologici oggetto della regressione multivariata. In blu sono rappresentati i punti della dispersione, in arancione la bisettrice del grafico (ovvero la situazione ideale, in cui modello e osservazione coincidono). È indicato anche il valore di sqm.*

La situazione ideale sarebbe quella rappresentata dalla linea arancione (ossia la bisettrice del grafico). Essa rappresenta il caso in cui il valore misurato è sempre uguale a quello previsto dal modello. Osservando gli scarti della dispersione dalla bisettrice si può avere un'idea sull'efficacia del modello nell'adattamento ai dati misurati. Per avere un riscontro quantitativo, è stato calcolato il valore dello scarto quadratico medio (come somma dei quadrati degli scarti dei punti dalla retta del caso ideale) e riportato in figura.

Come si era già concluso prima, i due casi di miglior regressione sono rappresentati dalle variabili di ricchezza in specie (S ed EPTS), dove si osservano dei punti molto più ravvicinati al caso ideale (e con un valore di scarto quadratico medio inferiore).

In ogni caso, risulta evidente l'aumento di significatività della regressione rispetto alle semplici correlazioni monoparametriche derivanti dalla regressione semplice.

Infatti, come già discusso, aumentare la complessità del modello (includendo un numero sempre maggiore di variabili) significa andare ad indagare aspetti dei fenomeni naturali non ancora presi in considerazione in precedenza. Siccome la natura non è discreta, come nei modelli matematici che si fanno per tentare di spiegarla, ma continua, aggiungere informazioni sempre più dettagliate porta le analisi statistiche a considerare il fenomeno, che si sta studiando, in maniera che tende sempre più ad una rappresentazione realistica.

5.4.2.3 Analisi Multivariata – Modello Esponenziale

Un'altra maniera di studiare il problema consiste nella creazione di una differente tipologia di modello di regressione. Si tratta di un metodo complementare ed analogo al precedente, che potrebbe, però, portare a risultati diversi (e anche migliori, talvolta).

Si vuole costruire un cosiddetto *modello esponenziale*, in cui, al posto della somma tra le variabili di regressione, ciascuna moltiplicata per il proprio coefficiente, si mette il prodotto di esse, ciascuna elevata ad un proprio esponente.

Un modello, così definito, può essere rappresentato nella seguente forma analitica

$$y = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2} \cdot x_3^{a_3} \cdot \ldots \cdot x_k^{a_k}$$

dove le varie x_i rappresentano i regressori, y la variabile risposta e gli a_k sono i coefficienti della regressione da determinare.

La comodità nel generare un modello del genere è dovuta al fatto che esso può essere interpretato come lineare in campo logaritmico. Infatti, applicando il logaritmo naturale ad entrambi i membri dell'equazione, si ottiene

$$\ln(y) = \ln\left(a_0 \cdot x_1^{a_1} \cdot x_2^{a_2} \cdot x_3^{a_3} \cdot \ldots \cdot x_k^{a_k}\right)$$

e, ricordando le proprietà di prodotto ed elevamento a potenza degli argomenti dei logaritmi,

$$\ln(y) = \ln(a_0) + a_1 \ln(x_1) + a_2 \ln(x_2) + a_3 \ln(x_3) + \cdots + a_k \ln(x_k)$$

che è un modello con una struttura identica a quella lineare di prima. La sola differenza rispetto a prima è che, al posto della variabile libera, c'è il suo logaritmo naturale.

Una volta capita e definita la struttura del modello che si vuole ottenere, l'applicazione ai dati reali è immediata. Infatti, il procedimento di analisi multivariata è esattamente lo stesso, gli stessi sono i passaggi e lo stesso è il tipo di risultato.

Anche l'applicazione dello script Matlab è immediata. Poiché i passaggi dell'analisi sono identici, anche lo script resta lo stesso. Ciò che deve cambiare è solo la natura dei dati in ingresso all'analisi, che devono essere in valore logaritmico.

Adesso, è ancora più chiara la comodità della definizione di un modello siffatto. Basta cambiare una linea del codice Matlab appena scritto, inserendo il comando per il calcolo del logaritmo delle due matrici contenenti i valori di variabili ecologiche e idromorfologiche, e si ottiene l'analisi secondo il nuovo modello di interpretazione.

Così è stato fatto e, nel successivo paragrafo, sono riportati i risultati e confrontati con quelli ottenuti dal precedente modello.

5.4.2.3.1 Risultati

Dall'analisi multivariata, eseguita in Matlab, risultano i seguenti andamenti del coefficiente R^2 in funzione del numero di variabili incluse nel modello esponenziale.

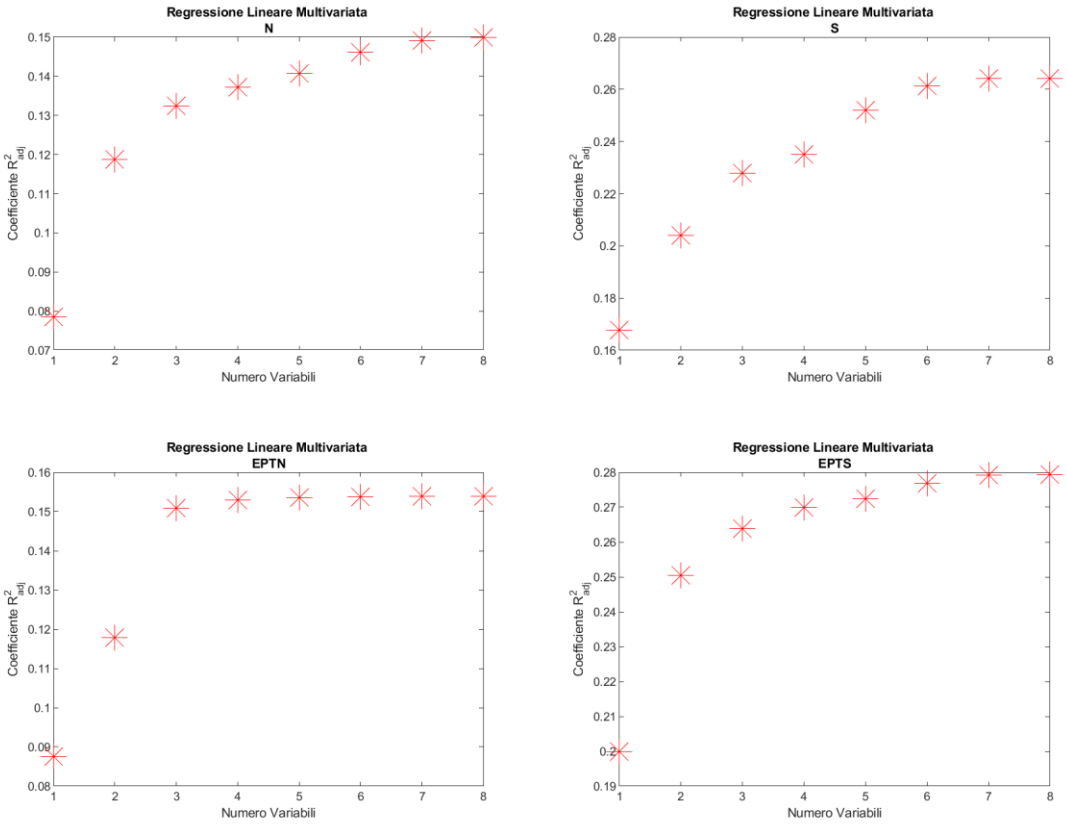

Figura 5.8 *– Grafici con l'andamento del valore del coefficiente R^2 in funzione del numero di variabili aggiunte al modello (ogni grafico rappresenta il risultato della regressione multivariata esponenziale di una delle quattro variabili ecologiche).*

Si nota la forma a concavità verso il basso degli ultimi tre grafici e quella più aperta (e ancora in crescita, apparentemente) del primo. Ciò significa che per spiegare e comprendere correttamente la disposizione spaziale del primo parametro (N) si avrebbe bisogno di informazioni aggiuntive riguardo alla idromorfologia dei vari bacini in studio.

I valori finali (massimi di R^2) sono confrontabili, come ordine di grandezza, con quelli ottenuti attraverso l'applicazione del modello precedente. Per la precisione, sono leggermente più alti per quanto riguarda la regressione dei parametri S ed EPTN, come si può notare confrontando i valori riportati in ***Tabella 5.5*** con quelli della seguente ***Tabella 5.6*** (che riassume i valori precisi di R^2 ottenuti dalla regressione esponenziale e il nome delle variabili aggiunte ad ogni passo della regressione multivariata).

Tabella 5.6 *– Analisi Multivariata – Valori di R^2 del modello esponenziale e progressiva aggiunta di variabili al modello*

N		S		EPTN		EPTS	
Variabili Modello	R_{ADJ}^2	**Variabili Modello**	R_{ADJ}^2	**Variabili Modello**	R_{ADJ}^2	**Variabili Modello**	R_{ADJ}^2
Morf	0.08	Nat	0.17	h_m	0.09	Nat	0.20
Morf, Nat	0.12	Nat , MAP	0.20	h_m , S.O	0.12	Nat, LSM	0.25
Morf, Nat, Q	0.13	Nat , MAP, Morf	0.23	h_m , S.O , Morf	0.151	Nat, LSM, MAP	0.26
Morf, Nat, Q, E	0.137	Nat , MAP, Morf, S.O	0.24	h_m , S.O , Morf, Nat	0.153	Nat, LSM, MAP, h_m	0.270
Morf, Nat, Q, E, LSM	0.141	Nat , MAP, Morf, S.O, Q	0.25	h_m , S.O , Morf, Nat, LSM	0.1536	Nat, LSM, MAP, h_m, Morf	0.272
Morf, Nat, Q, E, LSM, SO	0.146	Nat , MAP, Morf, S.O, Q, LSM	0.261	h_m , S.O , Morf, Nat, LSM, MAP	0.1537	Nat, LSM, MAP, h_m, Morf, Q	0.277
Morf, Nat, Q, E, LSM, SO, MAP	0.149	Nat , MAP, Morf, S.O, Q, LSM, h_m	0.26422	h_m , S.O , Morf, Nat, LSM, MAP, E	0.15383	Nat, LSM, MAP, h_m, Morf, Q, S.O	0.2793
Morf, Nat, Q, E, LSM, SO, MAP, h_m	0.150	Nat , MAP, Morf, S.O, Q, LSM, h_m, E	0.26424	h_m , S.O , Morf, Nat, LSM, MAP, E, Q	0.15383	Nat, LSM, MAP, h_m, Morf, Q, S.O, E	0.2794

Nuovamente, si osserva l'influenza preponderante delle variabili qualitative di morfologia dell'alveo e naturalità, ma anche della luce solare media, dell'ordine di strahler e della quota media. Anche in questo caso i valori massimi di R^2 non superano il 30% e sono maggiori per le due variabili di ricchezza in specie (*S* ed *EPTS*).

Si riportano, di seguito, i quattro nuovi modelli ottenuti:

1. $N = 1503.8 \cdot Morf^{-1.1} \cdot Nat^{0.97} \cdot Q^{-0.05} \cdot E^{-0.088} \cdot LSM^{-0.36} \cdot S.O^{0.17} \cdot MAP^{0.17} \cdot {h_m}^{-0.04}$
2. $S = 5.28 \cdot Nat^{-0.39} \cdot MAP^{0.4} \cdot Morf^{-0.15} \cdot S.O^{0.12} \cdot Q^{-0.01} \cdot LSM^{-0.21} \cdot {h_m}^{-0.03} \cdot E^{0.001}$
3. $EPTN = 131.39 \cdot {h_m}^{0.38} \cdot S.O^{-0.53} \cdot Morf^{-1.04} \cdot Nat^{0.29} \cdot LSM^{-0.16} \cdot MAP^{0.05} \cdot E^{-0.01} \cdot Q^{0.0012}$
4. $EPTS = 1.53 \cdot Nat^{-0.66} \cdot LSM^{-0.29} \cdot MAP^{0.39} \cdot {h_m}^{0.1} \cdot Morf^{-0.11} \cdot Q^{-0.02} \cdot S.O^{0.06} \cdot E^{0.006}$

Lo script Matlab ha restituito i valori dei vari coefficienti secondo il modello scritto sotto forma logaritmica (quindi simil-lineare).

Per ottenere la forma appena riportata è stato necessario convertire il coefficiente iniziale (quello senza variabile) sotto forma esponenziale (semplicemente si è elevato il numero *e* per il valore del coefficiente). Gli altri coefficienti assumono lo stesso valore sia in una forma che nell'altra, quindi sono stati semplicemente messi ad esponente delle variabili moltiplicate fra di loro, senza bisogno di alcuna ulteriore conversione.

Infine, si sono usate queste equazioni per la composizione dei grafici di confronto tra previsione del modello e misurazione effettiva in campo.

I risultati ottenuti sono visualizzabili nella successiva ***Figura 5.9***.

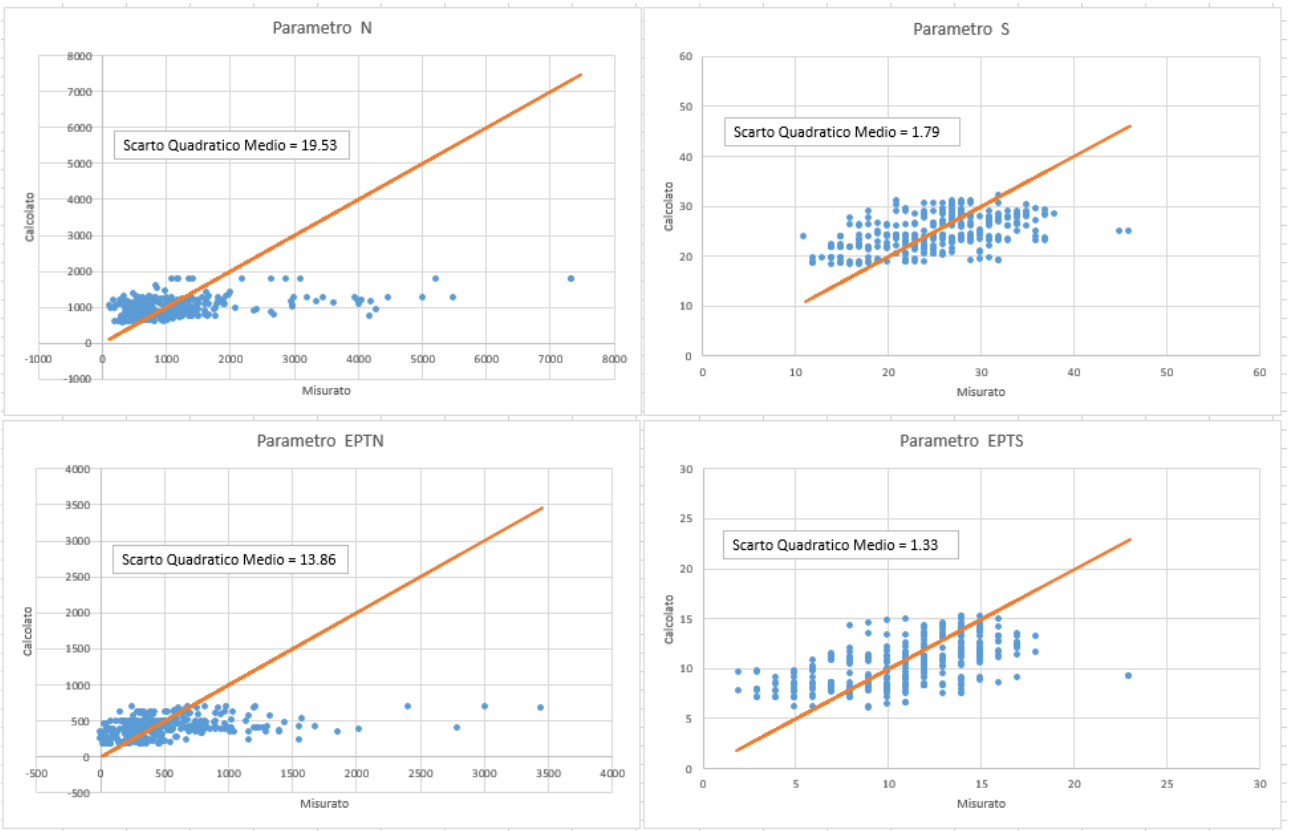

Figura 5.9 *– Grafici di confronto tra valore misurato e previsto dal modello esponenziale dei quattro parametri ecologici oggetto della regressione multivariata. In blu sono rappresentati i punti della dispersione, in arancione la bisettrice del grafico (ovvero la situazione ideale, in cui modello e osservazione coincidono). Sono riportati anche i valori di sqm.*

Da un'analisi dei valori di scarto quadratico medio riportati, si nota, nuovamente, come la dispersione sia sempre maggiore per i parametri di abbondanza (N, EPTN) che per quelli di ricchezza in specie (S, EPTS), in analogia, sempre, con l'andamento dei valori di R^2 di ***Tabella 5.6***.

5.4.2.4 Zonazioni

Come ultimo test per verificare la possibilità di aumentare la significatività di queste analisi statistiche, sono state eseguite due regressioni su una estensione territoriale più ristretta, rispetto a quella del database intero (che comprende all'incirca tutto il Piemonte).

Sono state quindi definite delle zonazioni sulla base di un *criterio geografico* (ossia di vicinanza delle stazioni di misura tra di loro e di appartenenza allo stesso bacino idrografico). La vicinanza spaziale, però, si traduce anche come somiglianza dal punto di vista delle caratteristiche ambientali e fisiche del territorio all'interno del quale le misurazioni sono state fatte. Per cui, questo criterio di scelta assume un doppio significato, sia in termini di vicinanza geografica, che in quelli di *omogeneità idromorfologica*.

L'ultima regola, di cui si è tenuto conto per la scelta definitiva delle zonazioni, è che ciascuna comprendesse non meno di 60-70 stazioni (per questioni di rappresentatività dei risultati ottenibili).

Le uniche zone che rispettassero questi criteri sono risultate due: la zona del cuneese (rappresentata da un sottobacino del Tanaro) e la zona del bacino del Po.

Nella seguente ***Figura 5.10***, possono essere visualizzate le stazioni appartenenti alle due zonazioni eseguite.

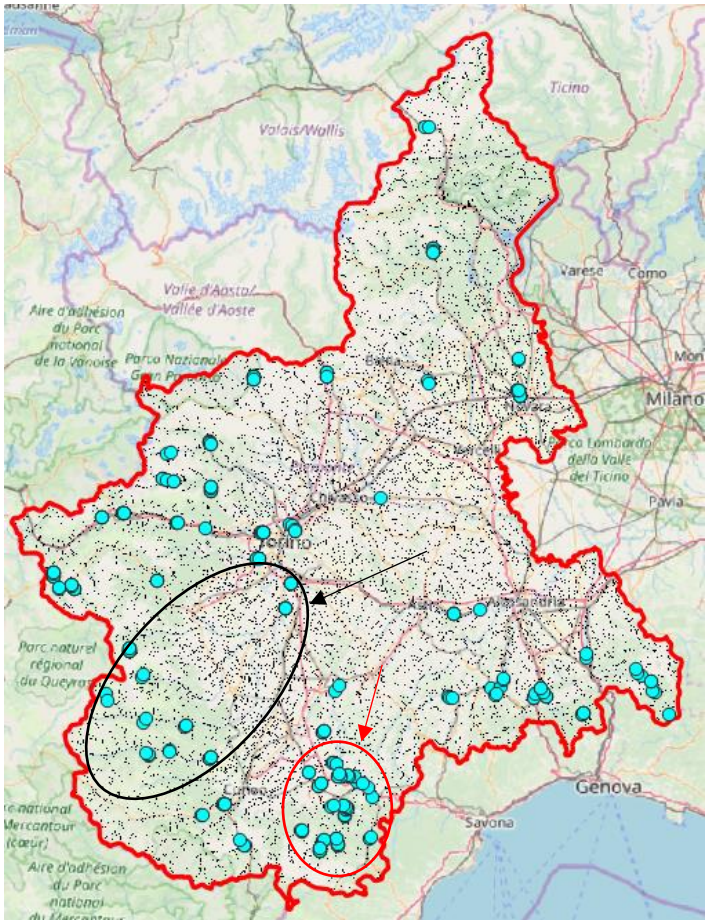

Figura 5.10 *– Posizione delle due zonazioni all'interno della zona di analisi (In rosso è evidenziata la zona del cuneese, in nero la zona del bacino del Po).*

Per entrambi i raggruppamenti, è stata eseguita la regressione multivariata (secondo modello lineare). Ciò è stato fatto per verificare se, andando a restringere il campo di indagine verso una zona a risposta più omogenea dal punto di vista delle caratteristiche fisico-ambientali, si potesse ottenere una regressione con un livello di significatività più alto (e quindi valori di R^2 maggiori rispetto al modello globale).

La regressione è stata fatta utilizzando lo stesso script matlab descritto nei paragrafi precedenti, con la sola modifica dei dati in ingresso (che non sono stati più il database totale, ma solo quelli relativi alle stazioni ricadenti all'interno della zonazione).

Si riportano, di seguito, i risultati relativi al cuneese e, poi, quelli relativi al bacino del Po.

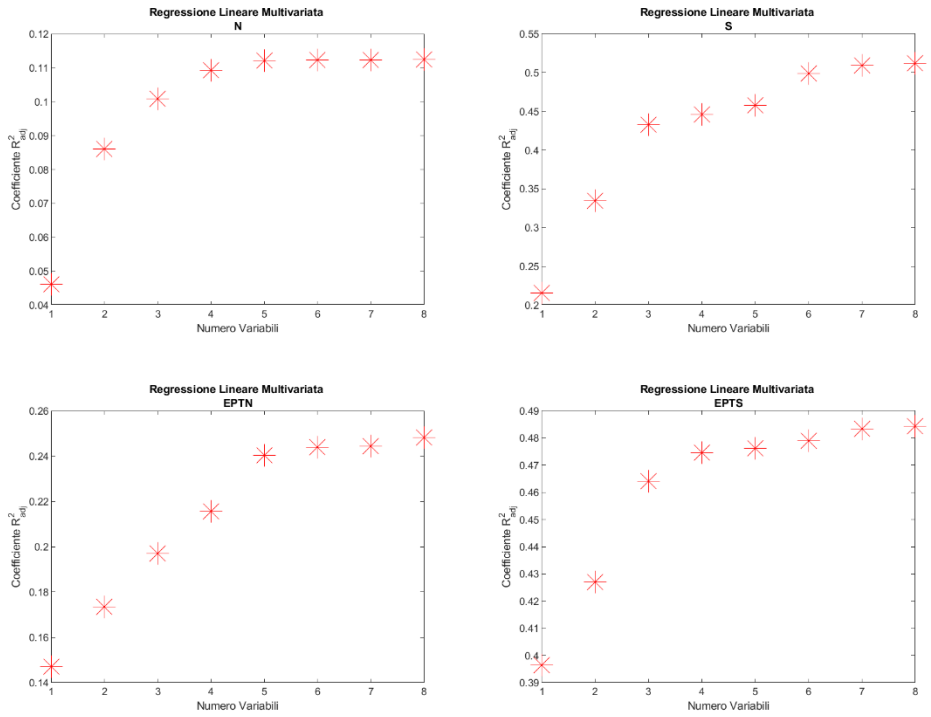

Figura 5.11 *– Zonazione cuneese. Grafici con l'andamento del valore del coefficiente R^2 in funzione del numero di variabili aggiunte al modello (ogni grafico rappresenta il risultato della regressione multivariata lineare di una delle quattro variabili ecologiche).*

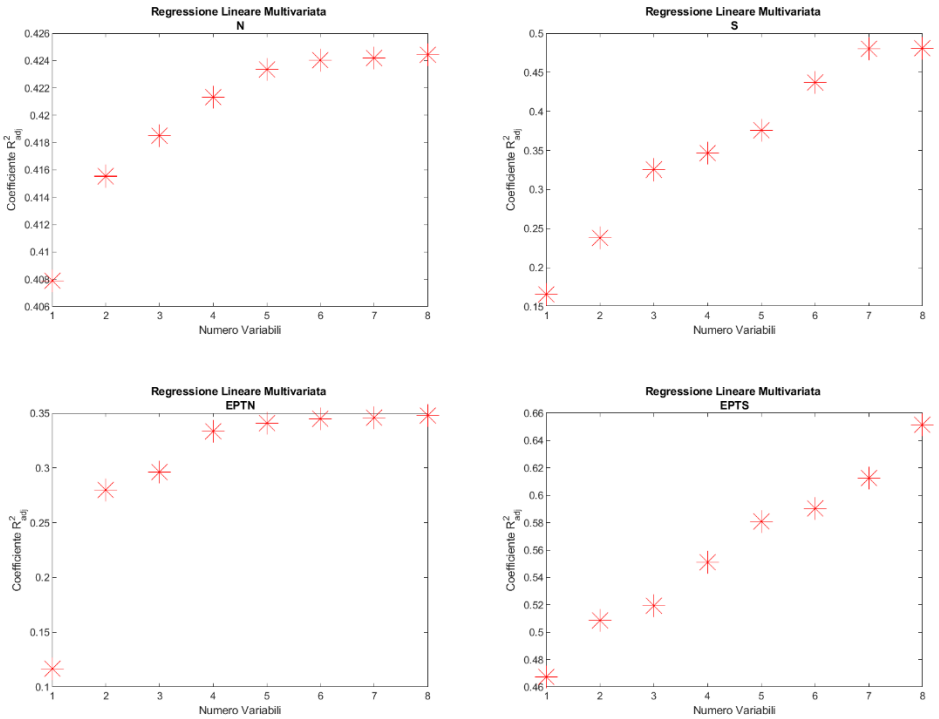

Figura 5.12 *– Zonazione bacino del Po. Grafici con l'andamento del valore del coefficiente R^2 in funzione del numero di variabili aggiunte al modello (ogni grafico rappresenta il risultato della regressione multivariata lineare di una delle quattro variabili ecologiche).*

Si nota immediatamente come il valore del coefficiente R^2 sia aumentato (e anche di molto) per quasi tutte le variabili ed in entrambe le zonazioni (si raggiunge un valore massimo del 66% per la variabile EPTS nella zona del Po). L'unica eccezione risulta essere la variabile N nella zona del Tanaro, che, invece, subisce un brusco abbassamento a seguito della zonazione (se nella regressione globale aveva raggiunto un valore del 18%, ora risulta del 12%). Inoltre, sembra che il modello della variabile EPTS possa aumentare ancora la propria significatività se fosse possibile tenere in conto di ulteriori fattori (dato l'andamento crescente del grafico).

Questa piccola restrizione del campo di indagine (si è, di fatto, passati da un'analisi di tipo regionale ad una di tipo locale) testimonia come l'uniformità di informazioni e la somiglianza in caratteristiche ambientali siano due aspetti estremamente importanti nella definizione dei pattern spaziali delle comunità di macroinvertebrati fluviali.

Capitolo 6

Analisi Spaziale e Semivariogrammi

6.1 Introduzione

L'analisi di tipo lineare ha rivelato come esista un certo rapporto statistico tra le variabili descrittrici dei corsi d'acqua e le variabili ecologiche. I valori del coefficiente di correlazione R^2 evidenziano un principio di correlazione significativa (si è visto che in certi casi si poteva arrivare anche al 30%, se non 60%, in caso di zonazione dei dati in ingresso ai modelli). I modelli creati possiedono, dunque, un fondamento di significatività statistica.

Nel *Capitolo 4*, si è visto come le variabili esplicative selezionate per comporre questi modelli lineari fossero intrinsecamente dotate di una duplice natura. Infatti, esse non solo forniscono informazioni locali, relative alle caratteristiche dell'habitat in cui le comunità biologiche vivono, ma, indirettamente, riflettono le modalità di azione di quelli che sono stati definiti determinanti regionali (ossia quei processi ecologici che influenzano la strutturazione delle comunità con una scala di azione più grande rispetto a quella del singolo habitat).

In questo capitolo, ci si propone di andare più nel dettaglio riguardo a questo discorso, ossia si vuole cercare di delineare con maggior nettezza l'influenza, che le variazioni spaziali a scale diverse comportano nell'organizzazione delle comunità locali. Si è visto, infatti, al *Capitolo 2*, come la relativa azione di determinanti locali e regionali dipendesse molto dalla scala spaziale da cui (e attraverso cui) si analizza il problema. Non si tratta di un bilancio dei due fattori, di uno che fa da complemento all'altro, ma, si tratta di una continua e costante azione di entrambi, la cui modalità ed entità dipende dalla posizione lungo il ramo fluviale, dalle caratteristiche delle specie, dalla fascia climatica della zona, dal grado di interconnessione della rete dendritica e molto altro (come è stato già detto). Considerare l'influenza della scala spaziale significa tenere in considerazione l'azione e il risultato di tutti questi fattori.

Un modo per farlo è attraverso l'uso dello strumento statistico dei *semivariogrammi*. Si tratta di particolari diagrammi che permettono di riconoscere, sulla base della forma della curva che ne risulta, la scala spaziale di influenza dei determinanti sulla disposizione spaziale delle comunità. Essi vengono costruiti sulla base del calcolo di distanze all'interno del sistema fluviale e delle differenze in valore delle variabili a tali distanze considerate. In particolare, nel corso del capitolo, verranno considerati due tipi di distanze: quelle *euclidee* e quelle *lungo il reticolo*. La scelta di queste due tipologie di distanza non è casuale e il suo significato verrà definito meglio in seguito.

In generale, si può dire che l'uso di questi semivariogrammi è ottimale a testare una teoria di analisi spaziale delle comunità molto assodata in ambito di eco-idraulica fluviale: il *continuum fluviale* (*river continuum concept*, in letteratura). Essa è molto utile per capire le modalità con cui le diverse scale spaziali interagiscono al livello delle comunità.

Per rendere tutto più chiaro, si partirà dalla descrizione a livello teorico del concetto di continuum fluviale (e del suo legame con la struttura dei semivariogrammi) per poi arrivare alla vera e propria applicazione ai dati ecologici rilevati in campo.

In particolare, si vuole precisare che il metodo di analisi utilizzato in questo capitolo è tratto dallo studio descritto, per il bacino dell'Adige, nell'articolo di Larsen et al., *Testing the River Continuum Concept with geostatistical stream-network models* (2019).

6.2 Concetto di *Continuum Fluviale*

La distribuzione e la diversità delle specie acquatiche dei sistemi fluviali è principalmente influenzata dal movimento unidirezionale dell'acqua e dai cambiamenti fisico-ambientali che si verificano spostandosi lungo le aste del fiume. È questo il concetto che sta alla base dei modelli teorici che vogliono idealizzare la struttura di un fiume e la funzione delle comunità che abitano lungo i rami della rete.

Una delle teorie più accreditate in questo ambito è il cosiddetto *continuum fluviale*. Tale teoria immagina il reticolo fluviale come un ecosistema aperto caratterizzato da dei cambiamenti continui (da monte verso valle) in fattori fisici e associate risposte ecologiche. In sostanza il modello associa il realizzarsi di tali risposte, in termini di disponibilità e tipologia di materia organica (legata al concetto di *catena trofica* già espresso), struttura spaziale delle comunità biologiche e distribuzione delle risorse, ad un cambiamento longitudinale continuo e graduale (in cui ciò che sta a monte influenza ciò che sta a valle). Per questo motivo, si parla di esistenza di un *gradiente ambientale longitudinale*.

I corsi d'acqua, però, sono sistemi caratterizzati da una struttura fisica molto variabile nel tempo e, quindi, da una elevata diversità ambientale. Questi cambiamenti repentini causano, in generale, nelle situazioni reali, degli spostamenti dallo schema ideale appena tracciato. La transizione delle caratteristiche fisiche dalla sorgente alla foce, infatti, raramente risulta totalmente uniforme, ma può mostrare accelerazioni o rallentamenti. Dunque, risulta indispensabile tenere in conto delle particolarità locali all'interno del modello RCC (*river continuum concept*).

Nonostante il continuum fluviale sia una semplificazione dei processi reali, che sottostima l'importanza del ruolo della struttura eterogenea e cangiante dei sistemi fluviali (associata alla geologia locale, agli effetti dei tributari e agli input dalle piane alluvionali adiacenti), tale teoria rimane una delle più influenti e accreditate in ambito di ecologia fluviale.

Nel corso degli anni, sono stati effettuati numerosi esperimenti atti a testare le ipotesi e le previsioni della teoria. Alcuni di essi forniscono dei risultati positivi, altri sottolineano l'importanza del discorso accennato prima riguardo alle eterogeneità locali dei sistemi fluviali. In particolare, in tali esperimenti, viene evidenziato il fatto che certi fattori locali possano superare, in termini di influenza verso le comunità, il suddetto gradiente longitudinale, in modo tale che comunità anche molto distanti tra loro possano risultare più simili di alcune più vicine tra loro lungo la rete.

6.3 Semivariogrammi

Il risultato teorico fornito da tali esperimenti apre la strada al problema della quantificazione della relativa contribuzione dei gradienti globali e delle eterogeneità locali. In questo senso, i modelli lineari classici forniscono risultati interessanti fino ad un certo punto. Tenere in considerazione tutti questi concetti in un'analisi spaziale è tutt'altro che semplice. La difficoltà dei modelli statistici classici nell'adattarsi alla teoria RCC risiede in due ragioni principali:

- il fatto che la struttura tipica dei corsi d'acqua reali è altamente complessa e ramificata;
- il movimento unidirezionale dell'acqua da monte verso valle.

La forte autocorrelazione spaziale generata da questi due fattori risulta particolarmente elevata per i sistemi fluviali reali e la sua intensità varia a seconda dell'interconnessione tra i rami e della direzionalità del flusso. Modelli che considerano una metrica di distanze slegata da quella del reticolo dendritico non portano alcuna informazione in questo senso.

Queste considerazioni risultano di particolare interesse quando si parla di continuum fluviale. Infatti, alla base della teoria sta il fatto che i processi ecologici che avvengono a valle siano imprescindibilmente legati a quelli che avvengono a monte. Inoltre, si è visto come le eterogeneità ambientali influiscano a scale spaziali multiple e molto variabili a seconda della situazione.

Recenti sviluppi nel campo della statistica geospaziale (adattata ai reticoli dendritici) hanno fornito le basi teoriche necessarie per quantificare le principali scale di variazione spaziale nei sistemi fluviali e, quindi, per testare modelli teorici (quali il *RCC*).

Uno degli strumenti più usati, a tal fine, in campo scientifico internazionale, è il cosiddetto *semivariogramma*. In particolare, ciò, che è stato fatto, è adattare il diagramma al caso particolare del reticolo fluviale (in tal caso, in letteratura, lo sia chiama anche *torgegramma*).

In generale, un *variogramma* serve a riconoscere delle particolari strutture spaziali lungo un'area di campionamento di dati e, quindi, può essere usato per rilevare le scale spaziali dominanti in un processo ambientale.

Nello specifico, il semivariogramma definisce l'andamento dell'autocorrelazione di una data variabile, calcolata usando la semivarianza tra coppie di osservazioni contenute in un intervallo di distanza *h*. Si definisce, così, la seguente funzione di autocorrelazione

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N} [z(s_i) - z(s_i + h)]^2$$

dove $N(h)$ è il numero di osservazioni contenute nell'intervallo di distanza *h* scelto, $z(s_i)$ è il valore che assume la variabile nel punto di osservazione *i* e $z(s_i + h)$ è il valore assunto dalla variabile in un punto a distanza *h* da s_i.

Si capisce come il concetto di semivariogramma, così definito, sia intrinsecamente legato al calcolo di distanze all'interno del reticolo fluviale. Infatti, si possono ottenere semivariogrammi diversi - per le stesse stazioni - semplicemente variando la definizione degli intervalli *h* di distanze considerati. A

sua volta, si possono ottenere forme diverse di semivariogramma, a seconda del metodo di calcolo di tali distanze. Nei reticoli fluviali, sono generalmente usate tre tipologie di distanze:

1. *connessa al flusso* (ossia quella calcolata lungo il reticolo che considera i soli punti connessi dallo scorrimento unidirezionale dell'acqua);
2. *non connessa al flusso* (quella lungo il reticolo, a prescindere dal flusso d'acqua);
3. *euclidea* (ottenuta collegando i due punti con una linea retta, a prescindere dalla struttura della rete).

La prima tipologia è utile a descrivere gli effetti del trasporto di materiale e del collegamento monte-valle, evidenziando i casi in cui la presenza di un gradiente ambientale longitudinale è particolarmente importante e rappresenta la principale scala spaziale di variabilità dei processi ecologici (materia organica e catena trofica). Le altre due tipologie, invece, portano principalmente informazioni riguardo all'influenza dei sistemi tributari (affluenti) e alle caratteristiche ambientali circostanti (e indipendenti dalla struttura del reticolo o dalla connessione idraulica tra le varie parti del sistema).

Quindi, il riconoscimento di questi pattern spaziali risulta essere di fondamentale utilità nel formulare ipotesi riguardanti i principali meccanismi ecologici che sono descritti dalle variabili che si stanno analizzando. Infatti, dalla forma del diagramma si possono trarre molte informazioni riguardanti l'organizzazione spaziale delle comunità lungo un reticolo. Il confronto dei risultati ottenuti in campo con le previsioni teoriche ha portato, in particolare, alla definizione di quattro situazioni generali di distribuzione dell'autocorrelazione spaziale di una variabile. Esse sono riassunte in ***Figura 6.1*** (dove, come esempio, viene mostrata la distribuzione spaziale della variabile che indica l'abbondanza di specie trituratrici lungo il reticolo).

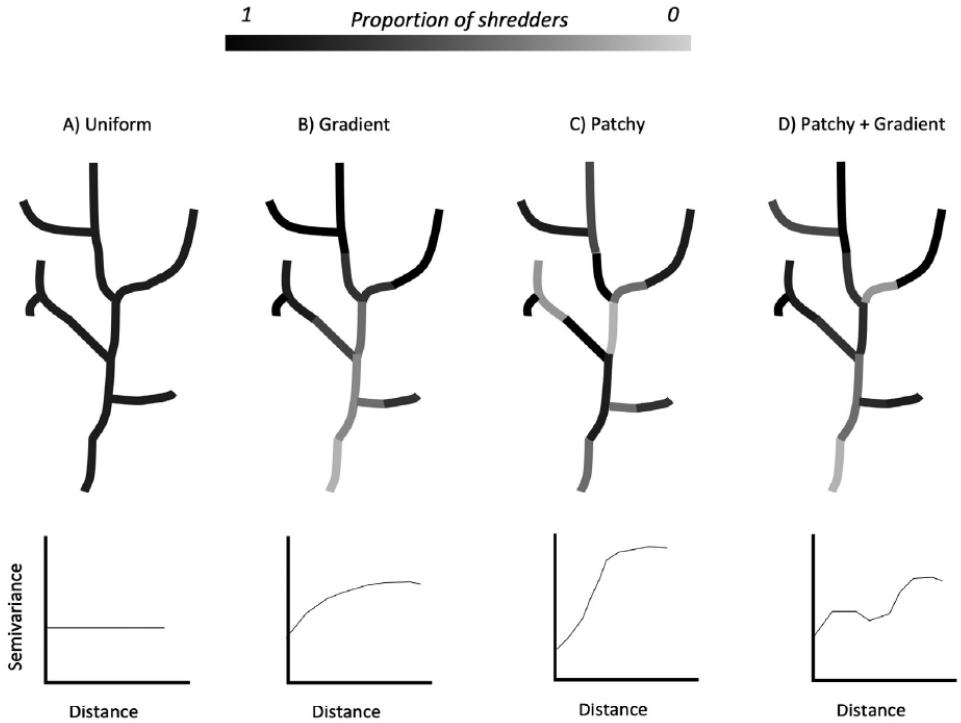

Figura 6.1 *– Le quattro tipologie di distribuzione dell'autocorrelazione spaziale lungo un reticolo fluviale. [Immagine tratta da Larsen et al., Testing the River Continuum Concept with geostatistical stream-network models, 2019]*

La *Figura A* rappresenta una situazione di completa uniformità e assenza di autocorrelazione spaziale (indicata anche dal fatto che la funzione γ sia costante a tutti i range di distanza considerati). Il valore dell'intercetta della retta rappresenta la varianza dovuta ad errori di campionamento o a variazioni a distanza più piccole di quella minima considerata.

La *Figura B* evidenzia come la dipendenza spaziale a larga scala rifletta la presenza di un singolo gradiente di variazione longitudinale dominante da monte verso valle. Il punto, in cui la funzione γ raggiunge un valore stazionario, indica il valore di distanza oltre la quale i valori della variabile a valle possono essere considerati indipendenti da quelli a monte (ossia, il punto oltre il quale si "spegne" l'autocorrelazione).

La *Figura C* rappresenta la situazione opposta alla *B*, ossia quella in cui le eterogeneità a piccola scala riflettono la struttura a mosaico e discontinua, in cui i fattori ambientali influenzano le variabili ad una scala spaziale piccola. Si nota come la forma del grafico sia simile a quella della *B*, ma, in questo caso, si raggiunge la stazionarietà (ossia la fine dell'autocorrelazione con ciò che sta a monte) ad una distanza molto minore.

La *Figura D*, infine, rappresenta una situazione mista tra le due appena descritte. Si è in presenza di una combinazione di eterogeneità a piccola scala inserite in una zona caratterizzata da una variazione a gradiente longitudinale a più larga scala. Lo si può riconoscere dalla tipica forma del grafico a multipli punti di inflessione. In questo caso, che, in generale, risulta essere il più adatto a descrivere le situazioni reali, i pattern sono influenzati da fattori operanti a scale spaziali multiple.

6.4 Applicazione ai dati ecologici

Compresa la base teorica che sta dietro al concetto di continuum fluviale ed alla sua analisi attraverso i variogrammi, si tratta di applicare il tutto a dei dati reali. I dati di partenza, che sono stati considerati, sono sempre gli stessi, ossia le quattro metriche biologiche che riassumono la raccolta di dati di abbondanza di famiglie di macroinvertebrati, fatta su tutto il territorio di studio.

L'obiettivo delle successive analisi è, dunque, la comprensione dei fenomeni spaziali che si possono dedurre dalla distribuzione dei dati, in termini di fattori di influenza regionale o locale.

Prima di tutto, verrà spiegato il procedimento che è stato seguito per la definizione delle distanze tra le stazioni di campionamento (euclidee e lungo il reticolo). In particolare, il calcolo delle distanze lungo la rete ha richiesto l'implementazione di un algoritmo di nuova costruzione (a differenza delle euclidee, per le quali era già implementato in un software). A partire dalle differenti tipologie di distanze definite, verranno, poi, costruiti i semivariogrammi della distribuzione delle quattro variabili ecologiche (*N*, *S*, *EPTN* ed *EPTS*). Infine, sulla base dell'interpretazione della forma di questi quattro grafici e dell'andamento generale della curva, si cercheranno di comprendere le scale spaziali principali che influenzano l'ecologia del territorio in esame.

In particolare, analogamente a quanto fatto nell'articolo di *Larsen et al.*, non verrà considerato tutto il territorio descritto dal database, ma ci si concentrerà su due singoli bacini idrografici (che sono gli stessi delle zonazioni per l'analisi lineare, un sottobacino del Tanaro e uno del Po), in modo da

riuscire ad apprezzare meglio le variazioni in eterogeneità e gradienti regionali lungo le aste fluviali delle due zone. Le ragioni della scelta di tali zone sono già state discusse nel *Capitolo 5*.

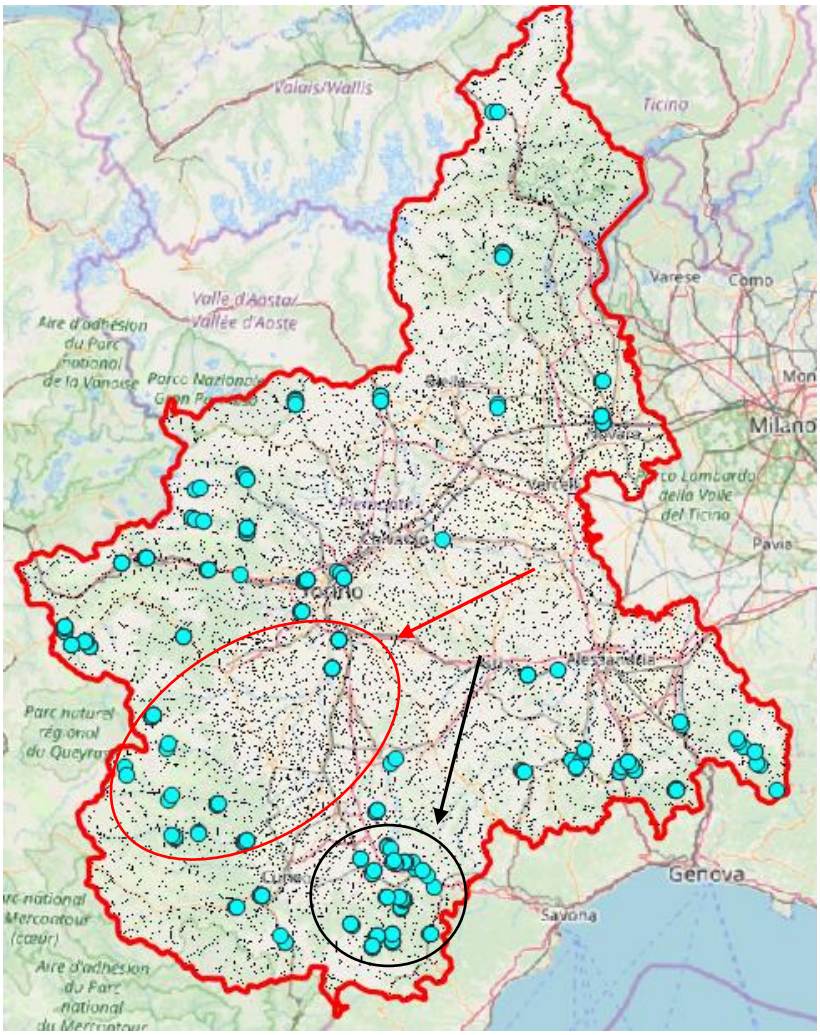

Figura 6.2 *– Stazioni del sottobacino del Tanaro (indicate in nero) e del sottobacino del Po (indicate in rosso).*

6.4.1 Distanze Euclidee

È stato detto come la costruzione di un variogramma sia intrinsecamente legato al calcolo delle distanze che intercorrono tra i punti di misurazione dei dati in ingresso. Il metodo più semplice per il calcolo di queste distanze è tracciare la linea retta che li congiunge e definirne la lunghezza: si tratta della cosiddetta *distanza euclidea*.

Il significato di un variogramma ottenuto mediante il calcolo di questo tipo di distanza è stato già discusso nel precedente paragrafo. In sostanza, esso fornisce informazioni sull'ecologia della zona di analisi in relazione all'ambiente che circonda le aste del reticolo fluviale (quindi gli affluenti e la piana alluvionale circostante ad esempio). Viene, di fatto, tralasciata la funzione della struttura dendritica del reticolo fluviale in ragione di ciò che gli sta attorno.

Per il calcolo effettivo di queste distanze, ci si è affidati ad un comando del software QGIS, chiamato **Matrice di distanza**. Si tratta di un algoritmo che prende come input due layer di punti (anche

uguali), calcola tutte le distanze euclidee tra ogni punto del reticolo in ingresso ed ogni punto di quello in uscita e fornisce come risultato una tabella a tre colonne, in cui, nella prima e seconda colonna, c'è il nome dei due punti tra i quali è stata calcolata la distanza e, nella terza, il valore della distanza (che, se i punti sono dati in coordinate geografiche, risulta misurata in chilometri).

In questo caso, è stato passato, come dato in ingresso all'algoritmo, il layer di punti che geolocalizza le stazioni del database sulla mappa (in sostanza, sono stati effettuati due ritagli del database completo in corrispondenza delle sole due zone di interesse, che sarebbero rappresentate dalle stazioni di ***Figura 6.2***, e poi passati come input all'algoritmo di calcolo delle distanze). L'algoritmo ha calcolato le distanze tra tutte le stazioni, per ogni zona, e il risultato è stato esportato come documento excel. Questo è stato il punto di partenza per la successiva costruzione dei semivariogrammi.

È stato successivamente creato uno script matlab, che, partendo dalla definizione del numero di classi di distanza *h,* ha permesso di costruire i quattro semivariogrammi per il sottobacino del Tanaro e i quattro per il sottobacino del Po (quattro, perché ce n'è uno per ogni variabile ecologica in analisi). Il commento al funzionamento dello script e il risultato, in termini di semivariogrammi, verranno presentati nel paragrafo conclusivo di questo capitolo.

6.4.2 Distanze lungo il reticolo

La seconda tipologia di analisi spaziale, che è stata effettuata, si è basata sulla definizione ed il calcolo delle distanze tra le stazioni seguendo l'andamento del reticolo idrografico.

Se la distanza euclidea rappresenta il percorso più breve che congiunge due punti (si tratta, come detto, di tracciare una linea retta che li unisca) e permette di studiare gli effetti trasversali alla rete dendritica sui processi ecologici, la distanza lungo il reticolo consente di prendere in piena considerazione come la struttura e la morfologia del complesso di aste fluviali agisca nell'influenzare le dinamiche di strutturazione delle comunità biologiche. In generale, siccome questo tipo di distanza segue l'andamento del fiume in ogni sua parte (compresi curve e meandri), essa risulterà quasi sempre maggiore di quella euclidea. Dunque, il semivariogramma, che ne deriverà, consentirà un'analisi a scala spaziale maggiore.

In particolare, possono essere calcolate due sotto tipologie di distanza lungo il reticolo:

a) *monodirezionale*, ossia quella che considera l'effettivo movimento dell'acqua da monte verso valle e non il percorso inverso;
b) *bidirezionale* (quella che, dati due punti *A* e *B*, tiene in considerazione il percorso di andata e anche di ritorno, ossia entrambi i percorsi *A-B* e *B-A*).

Entrambe le distanze sono particolarmente adatte a studiare gli effetti della rete fluviale sui processi di strutturazione spaziale e a rilevare la presenza di gradienti ambientali longitudinali. Inoltre, la seconda tipologia può portare anche informazioni riguardanti gli effetti dei sistemi tributari a quello principale (in quanto considera più direzioni possibili di movimento).

Il calcolo di queste distanze è risultato molto più complesso rispetto al caso precedente (in cui l'algoritmo di calcolo delle distanze euclidee era già implementato all'interno di un comando del

software *QGIS*). In questo caso, invece, è stato necessario creare, in *Matlab*, un programma ad hoc che andasse a calcolare queste distanze così come sono state appena definite.

Nei prossimi paragrafi viene spiegato nel dettaglio il procedimento che è stato seguito per la costruzione dell'algoritmo. In particolare, si vuole sottolineare che verranno descritti i passaggi per la definizione delle distanze bidirezionali, in quanto il caso monodirezionale risulta essere analogo (cambia solo il metodo di definizione dell'interconnessione tra le varie stazioni). Ad ogni modo, questo discorso verrà approfondito in seguito.

6.4.2.1 Preparazione dei dati preliminari e assegnazione delle stazioni al reticolo

Il punto di partenza per il calcolo delle distanze sono i dati georeferiti forniti dalle ricerche in campo del gruppo di biologi e dal progetto RENERFOR, ossia le stazioni di misura e il reticolo in formato raster della regione Piemonte. Tali file sono riportati nel *CD* allegato all'elaborato finale.

Poiché si vuole concentrare l'analisi su una parte ristretta e uniforme di territorio (che corrisponde alle due zonazioni descritte in precedenza), la prima cosa da fare è stata effettuare un ritaglio lungo la zona di interesse dei due layer sopracitati. Ciò è stato fatto in maniera agevole tramite il comando ***Ritaglia*** di QGIS.

I due dati, però, sono di natura differente: uno è un layer vettoriale di punti, l'altro è un raster. Per rendere più agevole il successivo calcolo delle distanze in Matlab, è conveniente disporre di entrambi i dati in formato layer di punti. Dunque, si è scelto di convertire il reticolo RENERFOR in layer di punti mediante il comando QGIS ***Da pixel raster a punti***. In sostanza, tale algoritmo converte ogni cella del raster (pixel) in un punto corrispondente al centroide del pixel di partenza. Il risultato finale è il reticolo fluviale dei due bacini, rappresentato come susseguirsi di punti georeferiti (di numero uguale a quello dei pixel che componevano il raster di partenza).

Per quanto riguarda la zona del Tanaro, l'operazione di ritaglio e conversione ha portato alla definizione di 29 stazioni di misura e 846 punti di rappresentazione del reticolo idrografico. I punti di stazione, in realtà, erano di più (circa 60), ma molti di essi si riferivano a misurazioni effettuate nella stessa stazione ma in stagioni diverse. Tali punti "doppione" sono stati esclusi, in quanto avrebbero creato problemi al funzionamento dell'algoritmo che verrà descritto in seguito. Per il Po, invece, le stazioni risultano essere 23 e i punti costituenti il reticolo 3100 (il bacino considerato è molto più esteso).

Il risultato finale, per entrambe le zone, viene mostrato nelle successive ***Figura 6.3*** e ***Figura 6.4***.

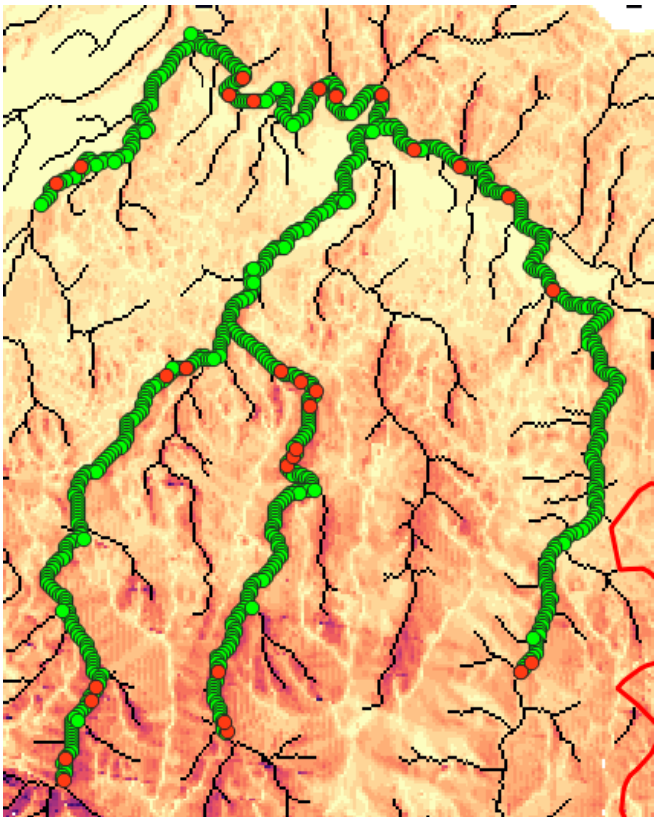

Figura 6.3 *– Visualizzazione in QGIS del primo bacino di riferimento (Tanaro) per l'analisi con il calcolo delle distanze lungo il reticolo. In verde, è rappresentato il layer di punti che costituiscono il reticolo, in rosso, le stazioni di misura.*

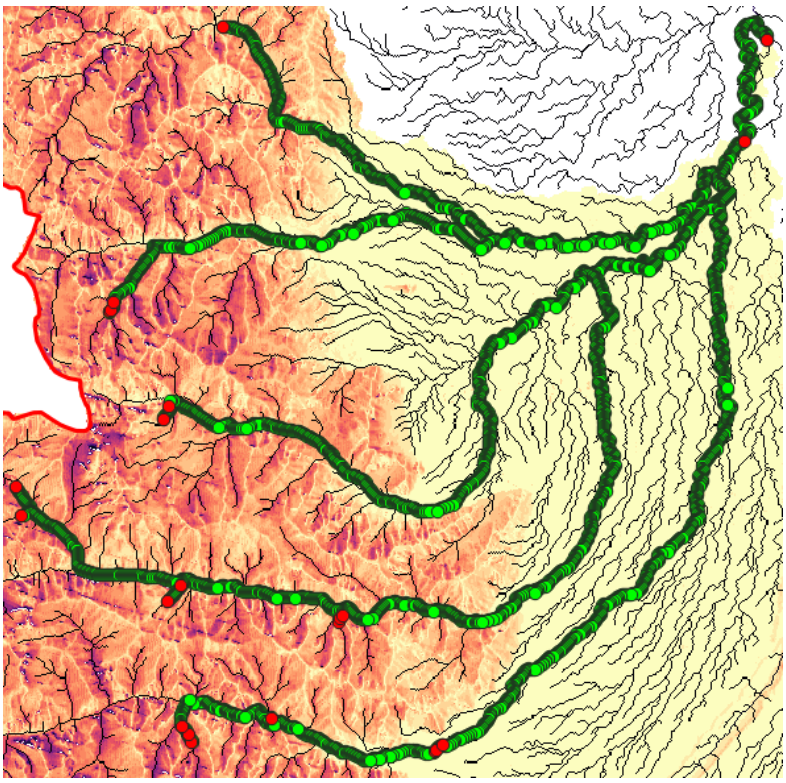

Figura 6.4 *– Visualizzazione in QGIS del secondo bacino di riferimento (Po) per l'analisi con il calcolo delle distanze lungo il reticolo. In verde, è rappresentato il layer di punti che costituiscono il reticolo, in rosso, le stazioni di misura.*

Passo preliminare è stato disporre di entrambi i dati di punti-stazione e punti-reticolo in formato tabellare (in excel). Per cui, i due layer appena descritti sono stati esportati in excel.

Si sottolinea che i passaggi seguiti sono stati gli stessi per entrambi i bacini. Dunque, a titolo di esempio, essi verranno descritti per il caso del Tanaro (in quanto analoghi per l'altro caso).

Poiché l'idea generale del successivo script Matlab è di calcolare le distanze tra tutte le coppie di stazioni sfruttando le distanze euclidee tra punti successivi del reticolo idrografico (rappresentato in verde in ***Figura 6.3***), è necessario che le tutte le stazioni siano posizionate esattamente sul reticolo. Dal successivo zoom di ***Figura 6.5*** si vede che non è così da subito.

Pertanto, prima del calcolo vero e proprio delle distanze, si è dovuto effettuare un passaggio intermedio attraverso la creazione di un algoritmo che assegnasse i punti-stazione ai punti-reticolo. In particolare, l'idea dello script è quella di assegnare ogni stazione al punto più vicino del reticolo, creando così un nuovo layer con le coordinate delle stazioni "riposizionate" sul reticolo.

In sostanza, il codice Matlab, dopo aver importato i dati di coordinate delle stazioni e del reticolo, si compone di un doppio *ciclo for* che permette di calcolare, per ogni stazione, tutte le distanze coi punti-reticolo, estrarre la minima ed assegnare al punto di partenza le coordinate del reticolo che corrispondono al punto che produce la distanza minima. Il risultato sono, appunto, le stazioni riposizionate sul reticolo (***Figura 6.6***).

Il codice completo e commentato viene messo in *Appendice*, come *Allegato 6*.

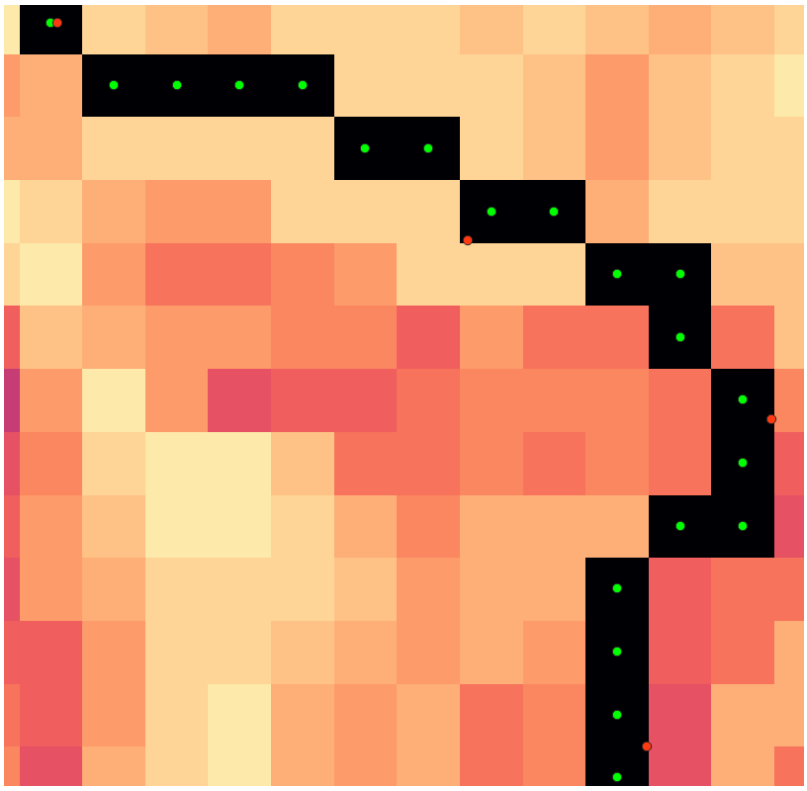

Figura 6.5 *– Zoom su un'asta del reticolo del Tanaro. Si vede come le stazioni (in rosso) non coincidano esattamente coi punti del reticolo (in verde).*

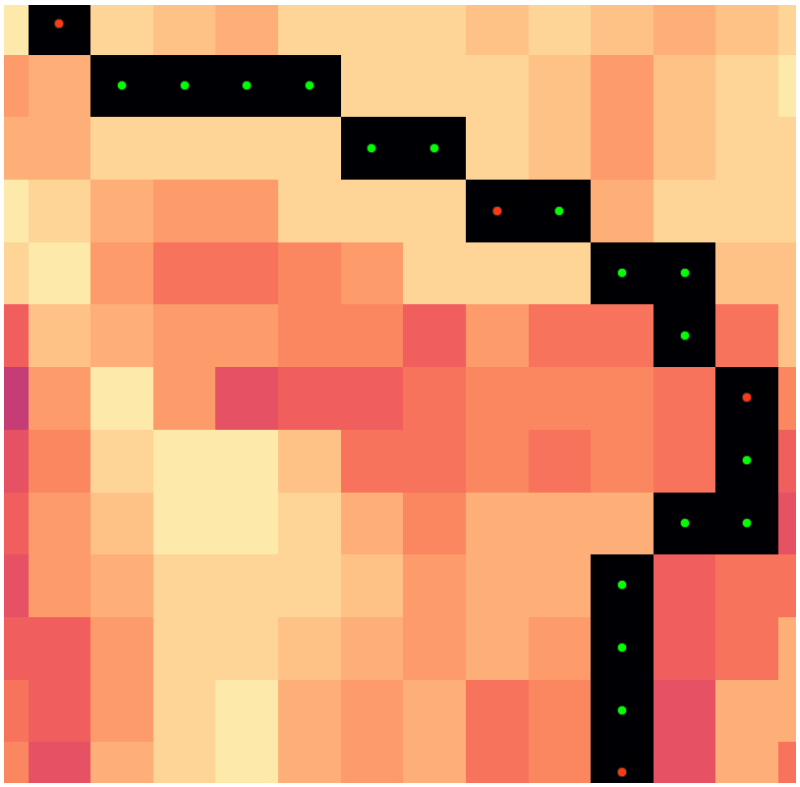

Figura 6.6 *– Zoom su un'asta del reticolo del Tanaro. Stazioni (rosso) riposizionate sul reticolo idrografico. Si nota la differenza con la precedente* ***Figura 6.5****.*

Il vettore ottenuto in Matlab con le nuove coordinate delle stazioni è stato poi esportato in excel sotto forma di tabella.

Ulteriore operazione, necessaria per il successivo funzionamento dello script di calcolo delle distanze, è stata l'attribuzione ad ogni punto del reticolo di un numero identificativo (ID) progressivo (da 1 a 846, nel caso del Tanaro). In questo modo, anche tutte le 29 stazioni riposizionate avranno un ID, corrispondente a quello del punto del reticolo con cui coincidono.

In conclusione, la struttura delle due matrici contenenti i dati preliminari (punti-stazione e punti-reticolo), uscenti da questa fase preparatoria, è a tre colonne:

- la prima con i numeri ID;
- la seconda con le coordinate di longitudine;
- la terza con le coordinate di latitudine.

Inoltre, si fa notare che i punti del reticolo (in tabella excel) non sono ordinati in nessun modo (hanno un numero identificativo, che non rispetta però la struttura precisa e sequenziale con cui vengono visualizzati in mappa). Per Matlab, essi sono punti "a caso" posizionati nello spazio. Anche se, poi, proiettandoli in mappa su QGIS, l'occhio riconosce la struttura del fiume, per il software non è così. Dunque, siccome si dispone di un reticolo disordinato di punti, tutti i ragionamenti successivi sono stati basati sul calcolo delle distanze tra i punti-stazione e tutti i punti-reticolo, di volta in volta, e sull'imposizione di limiti a queste distanze in modo da riconoscere quali punti del reticolo fossero significativi per una determinata stazione (e consecutivi rispetto al punto precedente).

Nel prossimo paragrafo vengono presentati i passaggi che compongono lo script che permette il calcolo delle distanze lungo il fiume. Il codice Matlab completo e commentato viene messo in *Appendice*.

6.4.2.2 Algoritmo per il calcolo delle distanze (bidirezionali)

Il commento degli step consecutivi (comandi Matlab) che compongono l'algoritmo è fornito, nel dettaglio, in *Allegato 7*. Poiché si tratta di uno script molto lungo e complesso, si è pensato di fornire, in questo paragrafo, una descrizione qualitativa di ciò che è il suo funzionamento generale. Infatti, commentare passaggio per passaggio senza avere sotto occhio il codice risulterebbe eccessivamente astratto e anche poco utile. Si rimanda, pertanto, all'*Allegato 7* per la descrizione dei comandi Matlab step per step. Adesso, si cercherà di delineare, in modo pur sempre completo e dettagliato, l'idea generale che sta alla base del funzionamento di questo script Matlab per il calcolo di distanze lungo un determinato percorso. Si prenderà nuovamente, a titolo di esempio, il caso del sottobacino del Tanaro.

Si hanno a disposizione gli 860 punti che compongono il reticolo fluviale e le 29 stazioni presenti lungo esso. I punti del reticolo corrispondono ai centroidi dei pixel che componevano il raster originario del reticolo (RENERFOR). Dalla conversione in layer di punti, si è ottenuta una tabella con l'ID di ogni punto (numero progressivo da 1 a 846) e le coordinate geografiche. Tali punti non sono ordinati in nessun modo, dunque per il software sono punti "a caso" posizionati nello spazio. Dunque, l'idea è di creare un doppio *ciclo for* (con numerosi cicli interni concatenati) e di procedere all'individuazione, di volta in volta, delle coordinate del punto più vicino (in particolare si ragionerà in termini di ID del punto, per avere un valore solo, invece che due, di coordinate) e salvare gli ID individuati in una struttura. Siccome però sono possibili più percorsi (infatti, non si conosce a priori la direzione giusta, i punti è come se fossero disposti a caso per il software), bisogna individuare tutti quelli possibili e poi alla fine scegliere quello giusto (che sarà quello che conterrà l'ID della stazione di arrivo come ultimo punto). Alla fine di ogni giro del ciclo, si ottiene un vettore con tutti gli ID dei punti (consecutivi) che compongono il percorso che collega la stazione di partenza con quella di arrivo. Per ottenere la distanza tra le due stazioni non si deve far altro che la somma di ogni tratto presente tra due punti consecutivi del vettore.

In altre parole, si parte da un punto (che, al primo giro del ciclo, è la stazione di partenza). Da esso, si deve cercare il punto più vicino. Bisogna, ora, capire quali caratteristiche debba soddisfare un qualsiasi punto appartenente al reticolo per poter essere considerato "vicino" a quello di partenza (ossia consecutivo in termini di visualizzazione a mappa del reticolo di punti complessivo). Considerando che si parte da un layer, i cui punti rappresentano il centro dei pixel del raster originario, il punto più vicino sarà sicuramente ad una distanza minore o uguale al valore della diagonale del pixel (in generale, un punto consecutivo può essere distante, al minimo, un lato del pixel, che è circa 1 m, o, al massimo, una diagonale del pixel, circa 1.41 m). Individuato questo punto, al giro dopo si riparte da esso e si trova il nuovo punto più vicino (avendo cura di escludere quello da cui si è partiti prima, in modo da non tornare indietro). Si procede così finché non si arriva al punto di arrivo (che è l'altra stazione). Quando si giunge alla stazione di arrivo, bisogna uscire dal ciclo e calcolare la distanza complessiva tra le due stazioni (di partenza e arrivo) lungo la rete, sommando tutti i pezzi di distanza euclidea tra punti consecutivi del reticolo. Si salva il valore di

distanza (e gli ID delle due stazioni che l'hanno generata) in una matrice finale e poi si riinizia il ciclo, cambiando la stazione di partenza e/o arrivo.

È importante ribadire un concetto. È stato detto come, in generale, dal punto di partenza si troveranno da uno a due punti vicini (che rispettano la condizione di distanza minore o uguale alla diagonale del pixel). Siccome a priori non si sa quale sia la direzione giusta (se bisogna andare da una parte o dall'altra per arrivare alla stazione di arrivo, seguendo il percorso del fiume), si mantengono entrambi i percorsi, salvando di volta in volta gli ID dei punti che si trovano in entrambe le direzioni, ma mantenendoli distinti (si creerà una struttura che permette di contenere i due percorsi diversi). Solo alla fine, quando effettivamente si giunge alla stazione di arrivo, si potrà vedere quale dei due percorsi permette di raggiungerla. Ciò consentirà di scegliere un percorso invece dell'altro e di calcolare la distanza finale (nel modo che si è detto).

Un ultimo problema, di cui si è dovuto tenere conto nella creazione dello script, è stato come gestire le *confluenze*. Esse sono state riconosciute come quei punti che invece di uno o due punti vicini, ne presentavano almeno tre (***Figura 6.7***). Nel momento in cui si giunge ad una confluenza, i percorsi disponibili diventano tre, invece di uno o due. Dunque, lo script ha dovuto tenere conto anche di questi casi particolari e aumentare il numero di percorsi da seguire qualora si trovasse una confluenza. Nel codice ciò si traduce nel fatto che la struttura contenente gli ID dei punti consecutivi, che compongono i vari percorsi, avrà dimensione maggiore (conterrà tre vettori di ID, invece che uno o due).

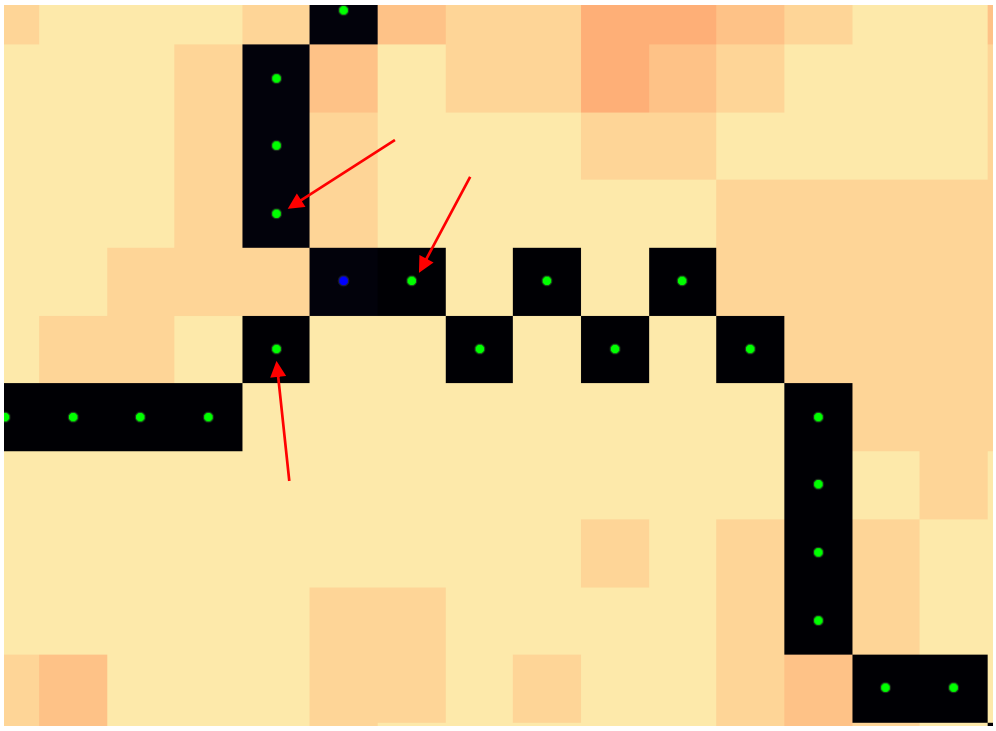

Figura 6.7 *– Zoom su un'asta del reticolo del Tanaro. Esempio di confluenza (blu). Si nota come il numero di punti vicini sia pari a tre (indicati dalle frecce). I punti in verde rappresentano il reticolo.*

Il risultato finale dello script è, quindi, il riconoscimento degli ID dei punti consecutivi del reticolo idrografico, che collegano la stazione di partenza con quella di arrivo. Avendo a disposizione questo vettore con i punti già ordinati, il calcolo della distanza lungo il reticolo è immediato. Basta sommare, alla fine di ogni giro del ciclo, tutti i tratti di distanza euclidea tra punto precedente e successivo del vettore.

Tuttavia, il calcolo della distanza non può essere effettuato con la classica formula geometrica derivante dal teorema di Pitagora. Bisogna tenere conto del fatto che si tratta di punti georeferenziati e proiettati su una superficie curva di riferimento (geoide).

Per tenere in considerazione la curvatura, la distanza tra due punti A e B, dati in coordinate geografiche, viene calcolata attraverso la seguente formula

$$d(A,B) = R \cdot arccos(\ (\sin(latA) \cdot \sin(latB) + \cos(latA) \cdot \cos(latB) + \cos(latA) \cdot \cos(latB) \cdot \cos(lonA - lonB)\ )$$

dove *R* è il raggio quadratico medio della Terra (approssimato a 6372.8 km). Le coordinate devono essere preliminarmente trasformate da gradi a radianti, in modo da poter utilizzare correttamente la formula. Il valore finale di distanza risulta espresso in chilometri.

Tale formula è stata implementata in una funzione a parte (chiamata *distanza*), in modo da alleggerire un po' lo script. Prendendo in input le coordinate di un punto qualsiasi del reticolo e la matrice dei punti di tutto il reticolo, tale funzione consente di calcolare tutte le distanze tra il punto di partenza e gli altri che compongono il reticolo.

Anche tale funzione viene riportata in *Appendice*, come *Allegato 8*.

L'oggetto finale, risultante dall'applicazione dello script Matlab a tutte le coppie di stazioni, è una matrice a tre colonne, così composta:

- prima colonna con ID della stazione di partenza;
- seconda colonna con ID della stazione di arrivo;
- terza colonna con distanza tra le due stazioni (in km).

Tale matrice è stata poi esportata in excel ed ha permesso la costruzione dei semivariogrammi per le distanze bidirezionali lungo fiume.

6.4.2.3 Calcolo delle distanze monodirezionali

I comandi che compongono l'algoritmo per il calcolo delle distanze monodirezionali sono gli stessi del caso bidirezionale. Infatti, il metodo di calcolo è sempre quello delle distanze lungo fiume e, pertanto, l'algoritmo resta uguale nei suoi passaggi fondamentali. Quello che cambia è che si devono considerare solo i percorsi che seguono l'andamento della corrente.

È stato, dunque, necessario comprendere, prima di far partire lo script, quale fosse il percorso dell'acqua nel reticolo considerato. Comprendere la direzione dell'acqua ha permesso poi di definire i nuovi collegamenti tra le stazioni. Infatti, stazioni che precedentemente risultavano collegate lungo le aste del reticolo (in quanto si teneva in conto puramente la struttura del reticolo, senza

considerare la direzionalità data dal moto dell'acqua da monte verso valle), adesso, non lo sono più. Tantissime distanze, quindi, vengono "perse" nel passaggio a questo tipo di analisi, poiché il collegamento tra due stazioni è dato dal mero passaggio di acqua da una verso l'altra (non è possibile risalire gli affluenti ad esempio).

Questo ragionamento è stato effettuato visivamente. Le stazioni interessate sono state trascritte e passate, di volta in volta (ramo per ramo), allo script.

Anche in questo caso, la matrice finale delle distanze è stata, poi, esportata in excel in modo tale da avere la tabella coi dati pronti per l'inserimento all'interno dell'algoritmo per la generazione del semivariogramma.

6.4.3 Risultati (Semivariogrammi)

La parte finale di quest'analisi spaziale è costituita dalla generazione dei semivariogrammi per entrambi i bacini. In particolare, per ogni variabile ecologica, è stato possibile costruire tre differenti tipologie di grafico: *euclideo*, *bidirezionale* e *monodirezionale*.

Nei successivi paragrafi verrà presentato l'algoritmo per la definizione, in *Matlab*, dei semivariogrammi e verranno commentati i risultati ottenuti.

Si precisa che, nel seguito, si useranno anche i termini *flow-connected* o *flow-unconnected* per riferirsi alle due tipologie di distanze lungo il reticolo, a sottolineare la considerazione o meno dello scorrimento dell'acqua nel generare la connessione tra stazioni.

6.4.3.1 Script Matlab - Classi di distanza e generazione dei semivariogrammi

La parte iniziale dello script riguarda l'importazione dei dati excel all'interno del software. Vengono, così, definite due matrici: la prima, contenente i nomi delle stazioni (ID) e il valore delle quattro variabili ecologiche ad ogni stazione (matrice a cinque colonne); la seconda, con i valori di distanza tra le stazioni e gli ID della coppia che genera ogni distanza (matrice a tre colonne). Ovviamente, la dimensione della seconda matrice varia a seconda del metodo di calcolo della distanza (le distanze euclidee sono molte di più di quelle monodirezionali, ad esempio).

Importati i dati di partenza, sono state poi definite le *classi di distanza h* per la costruzione dei semivariogrammi. Esse risultano diverse, in termini di limiti tra una classe e l'altra, in base alla tipologia di distanza considerata (euclidea, bidirezionale o monodirezionale). Tuttavia, si può fare un discorso generale, in quanto il metodo per la definizione di queste classi è stato il medesimo in tutti i casi. Quello che è stato fatto è trovare la distanza massima e dividerla per un valore, pari al numero di classi di distanza che si voleva ottenere. Infatti, si è scelto di definire molteplici suddivisioni in classi di distanza, a parità di tipologia di distanza calcolata. In particolare, sono stati definiti, per entrambi i bacini, i semivariogrammi a dieci classi di distanza (in analogia con quanto fatto nell'articolo di *Larsen et al.*) e ad otto classi di distanza. Si è scelto di eseguire questa operazione di sdoppiamento dei risultati in modo da studiare l'influenza del *data binning* iniziale sul resto del procedimento e, quindi, verificare la dipendenza o meno del risultato ottenuto dalla scelta

aleatoria del numero di classi. Tale numero è stato lo stesso per entrambi i bacini (otto e dieci), così da rendere confrontabili i vari grafici. Ad esempio, per quanto riguarda le distanze euclidee per la zona del Tanaro, volendo costruire un semivariogramma a dieci classi e avendo notato che la distanza massima tra quelle calcolate con l'algoritmo *QGIS* fosse pari a 24 km (circa), sono state definite 10 classi di distanza, da 0 a 25km, intervallate l'una con l'altra di 2,5 km.

Ad ogni modo, tutte le classi di distanza definite sono riassunte in ***Tabella 6.1***, ***Tabella 6.2***, ***Tabella 6.3*** e ***Tabella 6.4*** (in cui è anche riportato il numero effettivo di distanze contate dallo script, per ogni classe). Si nota il differente ordine di grandezza tra le classi, a seconda del tipo di distanza considerato. Le distanze lungo fiume (più elevate) permettono un'analisi a scala spaziale molto maggiore rispetto a quelle euclidee.

Tabella 6.1 *– Zona del Tanaro. Dieci classi di distanza e distanze trovate dall'algoritmo all'interno di ogni classe.*

EUCLIDEE	
Classi di Distanza	**Distanze Trovate**
0 - 2.5 km	310
2.5 - 5 km	240
5 - 7.5 km	150
7.5 - 10 km	256
10 - 12.5 km	396
12.5 - 15 km	190
15 - 17.5 km	48
17.5 - 20 km	92
20 - 22.5 km	196
22.5 - 25 km	192

FLOW UNCONNECTED	
Classi di Distanza	**Distanze Trovate**
0 - 6 km	104
6 - 12 km	72
12 - 18 km	122
18 - 24 km	158
24 - 30 km	84
30 - 36 km	106
36 - 42 km	106
42 - 48 km	14
48 - 54 km	20
54 - 60 km	26

FLOW CONNECTED	
Classi di Distanza	**Distanze Trovate**
0 - 4.2 km	41
4.2 - 8.4 km	16
8.4 - 12.6 km	21
12.6 - 16.8 km	31
16.8 - 21 km	23
21 - 25.2 km	25
25.2 - 29.4 km	16
29.4 - 33.6 km	18
33.6 - 37.8 km	17
37.8 - 42 km	14

Tabella 6.2 *– Zona del Po. Dieci classi di distanza e distanze trovate dall'algoritmo all'interno di ogni classe.*

EUCLIDEE	
Classi di Distanza	**Distanze Trovate**
0 - 7 km	36
7 - 14 km	90
14 - 21 km	70
21 - 28 km	70
28 - 35 km	58
35 - 42 km	40
42 - 49 km	22
49 - 56 km	38
56 - 63 km	28
63 - 70 km	10

FLOW UNCONNECTED	
Classi di Distanza	**Distanze Trovate**
0 - 17.4 km	62
17.4 - 34.8 km	22
34.8 - 52.2 km	0
52.2 - 69.6 km	20
69.6 - 87 km	26
87 - 104.4 km	30
104.4 - 121.8 km	34
121.8 - 139.2 km	114
139.2 - 156.6 km	80
156.6 - 174 km	40

FLOW CONNECTED	
Classi di Distanza	**Distanze Trovate**
0 -12 km	18
12- 24 km	4
24 - 36 km	0
36 - 48 km	0
48 - 60 km	0
60 - 72 km	7
72 - 84 km	13
84 - 96 km	10
96 - 108 km	5
108 - 120 km	5

Tabella 6.3 *– Zona del Tanaro. Otto classi di distanza e distanze trovate dall'algoritmo all'interno di ogni classe.*

EUCLIDEE	
Classi di Distanza	**Distanze Trovate**
0 - 3.125 km	394
3.125 - 6.25 km	296
6.25 - 9.375 km	130
9.375 - 12.5 km	532
12.5 - 15.625 km	238
15.625 - 18.75 km	6
18.75 - 21.875 km	278
21.875 - 25 km	196

FLOW UNCONNECTED	
Classi di Distanza	**Distanze Trovate**
0 - 7.5 km	124
7.5 - 15 km	108
15 - 22.5 km	164
22.5 - 30 km	144
30 - 37.5 km	132
37.5 - 45 km	94
45 - 52.5 km	18
52.5 - 60 km	28

FLOW CONNECTED	
Classi di Distanza	**Distanze Trovate**
0 - 5.25 km	46
5.25 - 10.5 km	18
10.5 - 15.75 km	36
15.75 - 21 km	32
21 - 26.25 km	30
26.25 - 31.5 km	19
31.5 - 36.75 km	21
36.75 - 42 km	20

Tabella 6.4 *– Zona del Po. Otto classi di distanza e distanze trovate dall'algoritmo all'interno di ogni classe.*

EUCLIDEE	
Classi di Distanza	**Distanze Trovate**
0 - 8.75 km	42
8.75 - 17.5 km	120
17.5 - 26.25 km	104
26.25 - 35 km	58
35 - 43.75 km	42
43.75 - 52.5 km	34
52.5 - 61.25 km	48
61.25 - 70 km	14

FLOW UNCONNECTED	
Classi di Distanza	**Distanze Trovate**
0 - 21.75 km	72
21.75 - 43.5 km	12
43.5 - 65.25 km	2
65.25 - 87 km	44
87 - 108.75 km	44
108.75 - 130.5 km	78
130.5 - 152.25 km	118
152.25 - 174 km	58

FLOW CONNECTED	
Classi di Distanza	**Distanze Trovate**
0 - 15 km	19
15 - 30 km	3
30 - 45 km	0
45 - 60 km	0
60 - 75 km	10
75 - 90 km	15
90 - 105 km	10
105 - 120 km	5

A questo punto, è stato creato un doppio *ciclo for* che permette il calcolo della funzione di autocorrelazione e la generazione automatica dei semivariogrammi.

Il *ciclo for* esterno è quello che serve a tenere fissa la variabile ecologica per la quale si vuole costruire il variogramma (in pratica, tiene fissa la colonna della matrice con i dati delle variabili ecologiche). Il *ciclo for* interno è quello che permette di definire tutte le coppie di punti che ricadono dentro ad ognuna delle classi di distanza. Infatti, all'interno di quest'ultimo è presente un altro *ciclo for* che permette di considerare, una ad una, tutte le righe della matrice delle distanze, valutare se la distanza sia compresa tra i due limiti (inferiore e superiore) della classe di distanza *h,* che il ciclo sta considerando in quel giro e, nel caso lo fosse, andare a ricercare, poi, nella matrice dei dati ecologici, il valore che assume la variabile ecologica (considerata e bloccata in quel momento dal ciclo esterno) in corrispondenza dei due punti che generano tale distanza. I due valori trovati rappresentano i parametri $z(s_i)$ e $z(s_i + h)$ della funzione gamma. Quest'ultimo ciclo (il terzo, quello più interno) permette, in sostanza, di calcolare la sommatoria della definizione della funzione $\gamma(h)$ e di valutare il valore di $N(h)$, ossia di contare il numero di distanze che ricadono nella classe che si sta considerando. Al termine di esso, si esplicita la formula della funzione $\gamma(h)$ e si calcola il valore di essa per quanto riguarda la classe di distanza h considerata da quel giro del ciclo. Al termine del secondo ciclo interno, si plottano tutti i valori della funzione gamma calcolati per ogni classe, mediante la serie di cicli precedenti, e si costruisce il semivariogramma per la variabile ecologica che è appena stata considerata.

In altre parole, si procede per classi di distanza successive (dalla prima all'ultima), una variabile ecologica alla volta. L'obiettivo è calcolare la funzione di autocorrelazione *γ(h)* per ciascuna classe e salvare poi tutti i valori in uno stesso vettore (viene aggiunto un valore ad ogni ciclo). Si parte dalla prima classe (che andrà da 0 ad un certo valore di km). Si percorre la matrice di distanze e si trovano tutte le distanze comprese nel range (una alla volta, ossia una per ciclo). Ogni volta che se ne trova una si guarda quali due stazioni (ID di esse) hanno generato quella distanza. Quindi, si prendono i valori della variabile ecologica di interesse (N, S, EPTN, EPTS) in corrispondenza di tali stazioni e si aggiunge un pezzo alla sommatoria della funzione *γ(h)*. Si procede così finché si sono trovate tutte le distanze ricadenti nella classe uno. Poi si passa alla successiva classe (la seconda) e si procede fino all'ultima classe. Alla fine, si ottengono i quattro semivariogrammi per la tipologia di distanze considerata (uno per ogni variabile ecologica).

Lo script completo e commentato è riportato, come quelli precedenti, in *Appendice*, come *Allegato 9*.

6.4.3.2 Sottobacino del Tanaro

L'applicazione dell'algoritmo appena descritto ai dati ecologici reali, relativi alla zona del sottobacino del Tanaro, ha portato alla definizione di tre semivariogrammi diversi per ogni variabile considerata. Ogni semivariogramma corrisponde ad un metodo diverso di calcolo della distanza ed ha, pertanto, un significato ecologico diverso.

Partendo dal calcolo effettuato per la suddivisione in dieci classi, si riporta il risultato in ***Figura 6.8***, in cui sono riassunti tutti i semivariogrammi ottenuti per ogni variabile. Si possono analizzare i tre casi separatamente.

In primis, se si considera il caso delle *distanze euclidee* (in nero), si nota come la forma dei grafici in figura possieda un significato estremamente chiaro. Innanzitutto, è importante considerare che, in termini di analisi spaziale, un grafico di questo tipo è da interpretare nel senso indicato in precedenza, quando è stata trattata la teoria dei semivariogrammi (ossia, la forma finale dei grafici non sta ad indicare una proprietà del reticolo fluviale, ma, piuttosto, di ciò che gli sta attorno, delle connessioni bidimensionali che caratterizzano l'intero bacino).

Per tutte e quattro le variabili, quello che si registra è un andamento molto vario della funzione di autocorrelazione (funzione γ). Essa non risulta strettamente crescente o decrescente (come nel caso teorico *B*, ad indicare la presenza di un'unica scala spaziale di variazione lungo tutto il bacino), ma presenta numerosi valori di picco e cambi di andamento (a volte cresce, a volte decresce, a seconda della classe di distanza). Tale risultato non assomiglia neanche al caso *C*, in cui, per lo meno, ci dovrebbe essere una piccola crescita inziale (a testimoniare la presenza di autocorrelazione spaziale della variabile per piccole distanze).

In definitiva, si tratta di semivariogrammi di tipologia *D*, ossia si è in presenza di una struttura spaziale mista, in cui eterogeneità locali e gradiente ambientale a vasta scala competono nella regolazione degli equilibri ecologici delle comunità.

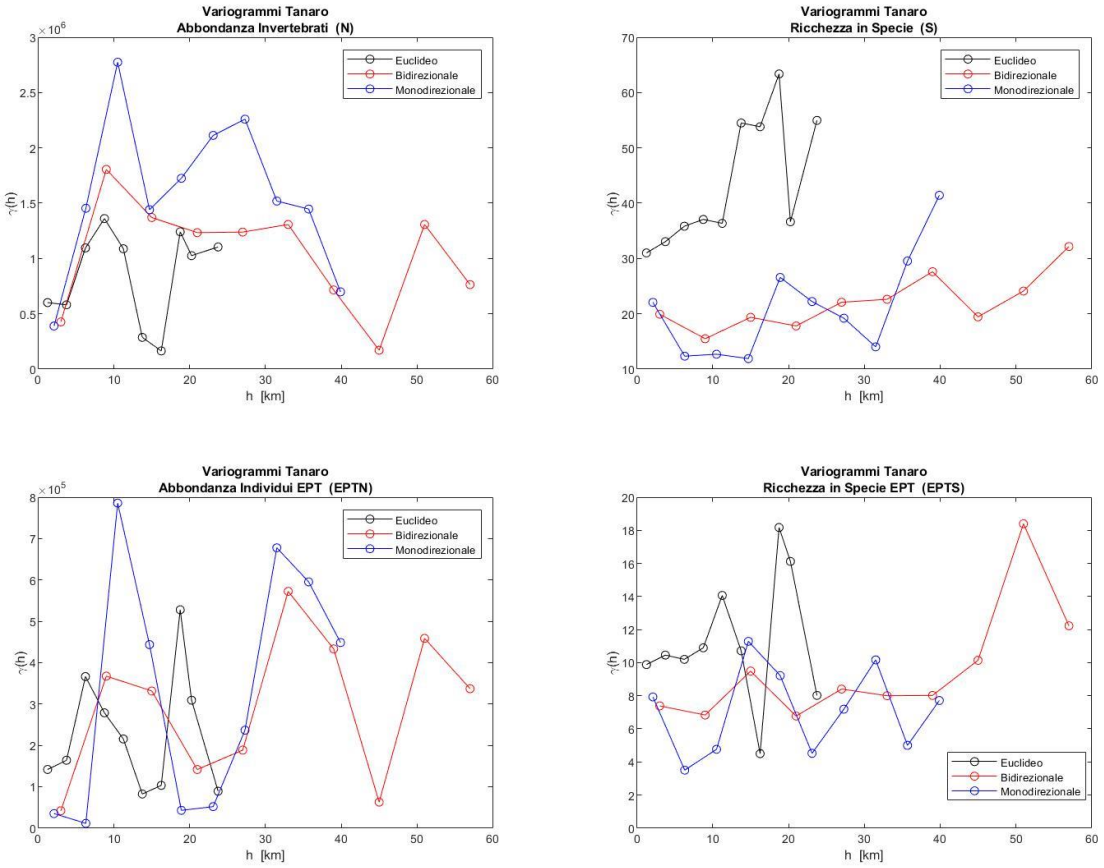

Figura 6.8 *– Semivariogrammi a dieci classi per la zona del Tanaro. Sono riportate le curve ottenute considerando le distanze euclidee (nero), bidirezionali (rosso) e monodirezionali (blu). Ogni grafico si riferisce ad una delle quattro variabili ecologiche (N, S, EPTN ed EPTS).*

Dunque, il risultato finale dell'analisi spaziale con le distanze euclidee testimonia, in generale, la presenza di variazioni a multiple scale spaziali (grandi e piccole). È ciò che tipicamente si ottiene nei casi reali, in cui tutte le ipotesi e le discretizzazioni dei meccanismi agenti nei modelli teorici sono un po' meno definiti. Infatti, studiando un ecosistema reale, raramente si osserva la presenza di un'unica scala spaziale di azione dei meccanismi di regolazione dell'assemblamento comunità, in quanto la presenza di entrambi i determinanti locali e regionali è troppo invasiva ed importante (e, come si è già detto, i due fattori non si bilanciano l'uno con l'altro, ma agiscono uno insieme all'altro).

Per quanto riguarda, invece, il caso di distanze *flow-unconnected* (bidirezionali, in rosso), già ad un primo sguardo della ***Figura 6.8***, si notano molte differenze rispetto ai semivariogrammi calcolati mediante le distanze euclidee. Infatti, se per le distanze euclidee era riconoscibile un'unica tipologia di semivariogramma (delle quattro teoriche possibili), qui si nota la compresenza di più forme diverse delle curve.

Si ricorda che, utilizzando le distanze lungo fiume, si sta esplicitamente tenendo in conto di tutti quei meccanismi metacomunitari che sono influenzati dalla struttura dendritica del reticolo fluviale

(a differenza delle distanze euclidee, in cui la struttura del reticolo veniva totalmente trascurata nell'analisi spaziale). Dunque, l'andamento della funzione di autocorrelazione spaziale *γ(h)* risulta un prodotto diretto della posizione delle stazioni di misura lungo le aste del reticolo idrografico (e anche del valore della variabile ecologica in quelle stazioni, naturalmente).

Si riconosce, in particolare, nei grafici delle variabili di ricchezza in specie (*S* ed *EPTS*), una preponderanza della struttura tipica del semivariogramma di tipologia *B* (eccetto che per alcune classi di distanza in cui l'andamento monotono subisce delle variazioni repentine, a testimoniare, comunque, l'influenza delle eterogeneità locali nei casi reali). In altri termini, per quanto riguarda l'autocorrelazione spaziale delle variabili di ricchezza in specie, si nota una dinamica dovuta prevalentemente alla presenza di un gradiente ambientale longitudinale, che influenza l'ecologia della zona, procedendo da monte verso valle. Le proprietà ambientali di ciò che sta a monte cambiano gradualmente procedendo verso valle ed influenzano gli equilibri comunitari e la distribuzione spaziale delle comunità (in termini di specie diverse, in questo caso).

Per quanto riguarda le variabili di abbondanza (*N* ed *EPTN*) si notano due andamenti diversi.

La variabile *N* presenta, perlomeno per distanza piccole, un andamento di tipologia *C* (la cosiddetta struttura a mosaico, *patchy*). Infatti, si riconosce una piccola crescita del semivariogramma fino al raggiungimento di un piccolo plateau, a testimoniare lo spegnimento dell'autocorrelazione spaziale di tipo gradiente per distanza molto piccole (circa 10 km) e l'inizio di una preponderante influenza delle eterogeneità locali. Procedendo verso distanze sempre maggiori, si nota la transizione verso la tipologia *D*, in cui si riconosce un'influenza combinata tra gradienti ambientali e situazioni locali.

La variabile *EPTN*, invece, presenta, analogamente alle distanze euclidee, in tutto lo sviluppo del semivariogramma, una struttura di tipo *D*, che è poi quella tipica dei casi reali. Infatti, si nota l'andamento ad alternanza di continui picchi (di massimo o minimo), che testimoniano la compresenza di dinamiche ecologiche a scale spaziali multiple.

Infine, nel considerare le *distanze monodirezionali* si compie un passo in avanti verso il raggiungimento di un'analisi spaziale che sia il più completa possibile. Per ora, sì, è stato studiato il bacino idrografico nel suo complesso, in termini sia di struttura e morfologia del reticolo (distanze bidirezionali) che di zone circostanti ad esso (distanze euclidee), ma non si è tenuto conto di un fattore essenziale nell'analisi ecologica di un sistema fluviale: il movimento dell'acqua.

È abbastanza scontato da dire quando si sta parlando di fiumi, ma l'organizzazione spaziale di una comunità legata all'ecosistema fluviale è inevitabilmente, nonché primariamente, influenzata dal flusso dentro l'alveo. In termini di distanze lungo il reticolo, risulta evidente come il moto dell'acqua determini una connessione univoca tra le stazioni (in maniera, appunto, monodirezionale). L'acqua scorre in un verso solo, da monte verso valle. Se nelle distanze bidirezionali era possibile trovare collegamenti tra certe stazioni a valle e certe stazioni a monte, in questo caso non è più possibile (ad esempio non si possono risalire gli affluenti dei sistemi principali). È questa la ragione per cui, in letteratura, ci si riferisce a questo tipo di distanza come *flow-connected* (ossia connessa dal flusso).

I semivariogrammi costruiti secondo questo ragionamento, rappresentati in blu ***Figura 6.8***, denotano tutti la stessa struttura, di tipologia *D*. A differenza di quelli bidirezionali, in cui era stato possibile riconoscere dei meccanismi spaziali ad ampio raggio (gradiente ambientale) per due delle quattro variabili ecologiche, in questo caso di connessione idraulica si riconosce la totale influenza

delle eterogeneità locali a scale spaziali piccole. Le forme dei semivariogrammi risultano molto irregolari (un continuo sali-scendi della funzione di autocorrelazione all'aumentare della classe di distanza), a testimoniare come basti spostarsi di poco lungo le aste del fiume (e nel senso della corrente) per trovare caratteristiche biologiche molto diverse (in termini di abbondanza o di tipologia di specie presenti).

A conclusione, è possibile spendere due parole sul modo in cui è stata definita la funzione di autocorrelazione *γ(h)*, dal calcolo della quale sono stati costruiti tutti i grafici in ***Figura 6.8***. Dalla definizione posta al *paragrafo 6.3*, tale funzione è stata costruita sulla falsa riga di una media statistica, di tipo quadratico. Infatti, essa risulta data dalla sommatoria di una differenza, elevata al quadrato, di due valori di una stessa variabile, il tutto diviso per il numero di coppie di valori complessivo. Ne consegue che i semivariogrammi ottenuti dall'applicazione dell'algoritmo Matlab rappresentano un andamento medio dell'autocorrelazione lungo il bacino. Dunque, tutte le conclusioni a cui si è giunti descrivono ciò che in media si può trovare all'interno del reticolo.

Al fine di completare queste informazioni, è stato definito il valore di *scarto quadratico medio* della funzione di autocorrelazione. Esso è stato calcolato in Matlab come un classico valore di deviazione standard statistico, ossia come radice quadrata della sommatoria dei quadrati degli scarti dei valori singoli (calcolati di volta in volta, per ogni singola distanza) rispetto al valore medio (che si conosce a priori, essendo già stato calcolato per costruire i semivariogrammi), divisa per il valore totale (cioè comprensivo di tutte le dieci classi) di coppie di distanze trovate. In sostanza, è stato calcolato un valore unico per l'intero diagramma (considerando, però, in base alla classe di distanza, il valore medio di *γ(h)* riferito a quella classe). Il risultato finale è riportato nella seguente ***Tabella 6.5***.

Tabella 6.5 *– Zona del Tanaro, semivariogramma a dieci classi. Valori di deviazione standard della funzione di autocorrelazione γ(h), per ogni variabile ecologica e per ogni tipologia di semivariogramma.*

	Scarto Quadratico Medio Semivariogrammi (Tanaro)		
	Euclideo	**Flow Unconnected**	**Flow Connected**
N	1.28E+06	1.48E+06	1.64E+06
S	49.68	26.12	25.16
EPTN	3.84E+05	4.77E+05	4.89E+05
EPTS	14.70	9.14	8.39

Analogamente ai risultati delle analisi di regressione lineare multivariata (*Capitolo 5*), in cui si erano ottenuti, dai grafici variabile prevista - variabile misurata, degli scostamenti inferiori rispetto alla media da parte delle due variabili di ricchezza in specie (S ed EPTS), anche in questo caso si notano dei valori di scarto quadratico medio minori per le medesime variabili. Tali valori riflettono direttamente la maggiore significatività statistica di tali variabili rispetto alle altre due (per le quali, si osservano dei valori di deviazione standard estremamente elevati).

Ora che il commento del grafico dei semivariogrammi a dieci classi è completo, si possono analizzare i risultati ottenuti definendo otto classi di distanza (***Figura 6.9***).

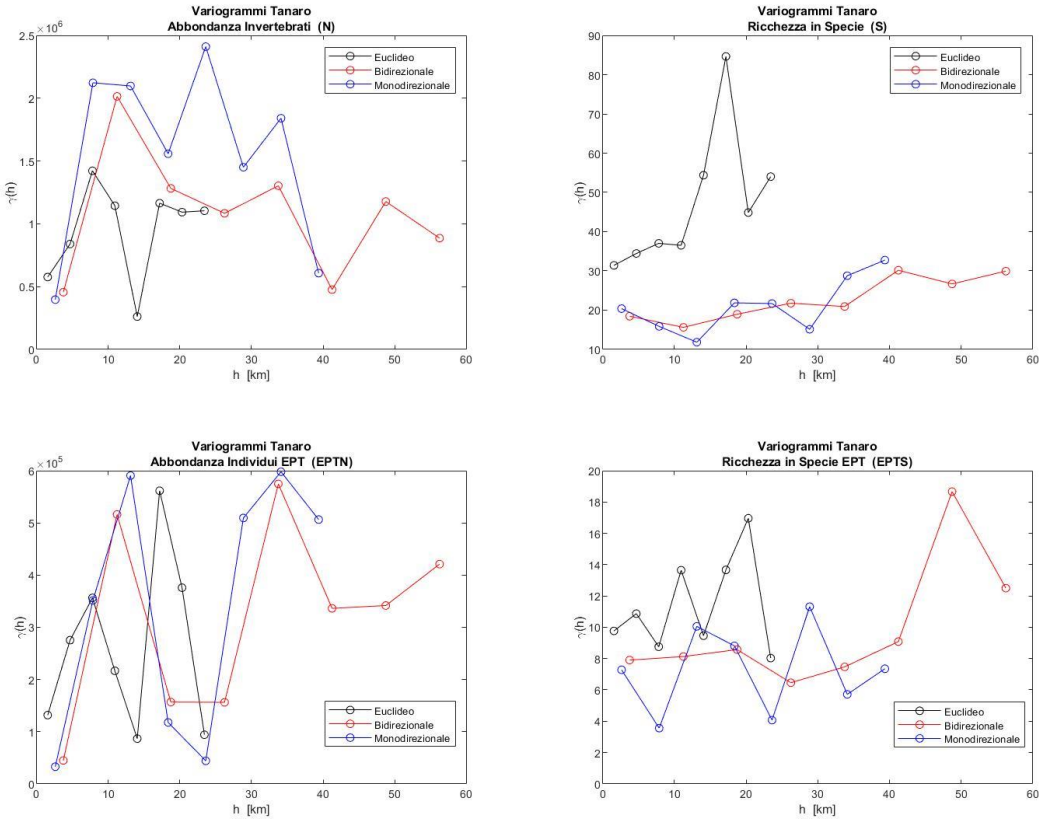

Figura 6.9 *– Semivariogrammi a otto classi per la zona del Tanaro. Sono riportate le curve ottenute considerando le distanze euclidee (nero), bidirezionali (rosso) e monodirezionali (blu). Ogni grafico si riferisce ad una delle quattro variabili ecologiche (N, S, EPTN ed EPTS).*

Da un confronto con i semivariogrammi a dieci classi, ci si rende subito conto come la situazione complessiva non cambi. Nonostante il valore dei vari picchi sia diverso (infatti, essendoci due classi in meno, i limiti tra classi sono diversi e le distanze sono ridistribuite in classi diverse), l'andamento generale delle curve risulta lo stesso. Pertanto, l'interpretazione ecologica dei grafici è la medesima e risulta indipendente dalla scelta delle classi fatta all'inizio delle analisi.

Per completezza, si riportano anche i valori degli scarti quadratici medi per i semivariogrammi ad otto classi (***Tabella 6.6***). Le conclusioni a cui si giunge sono, nuovamente, le stesse già descritte.

Tabella 6.6 *– Zona del Tanaro, semivariogramma a otto classi. Valori di deviazione standard della funzione di autocorrelazione γ(h), per ogni variabile ecologica e per ogni tipologia di semivariogramma.*

	Scarto Quadratico Medio Semivariogrammi (Tanaro)		
	Euclideo	**Flow Unconnected**	**Flow Connected**
N	1.28E+06	1.46E+06	1.63E+06
S	49.87	26.06	25.67
EPTN	3.86E+05	4.69E+05	5.08E+05
EPTS	14.72	9.16	8.42

6.4.3.3 Sottobacino del Po

Lo stesso algoritmo di generazione dei semivariogrammi è stato applicato ai dati del sottobacino del Po. Si tratta di un bacino molto più esteso di quello del Tanaro appena considerato (come si è già potuto notare dal confronto di ***Figura 6.3*** e ***Figura 6.4***), ma che possiede meno stazioni di campionamento (23 invece di 29).

Analogamente alla descrizione dei risultati per il bacino del Tanaro, si parte dal commento dei semivariogrammi a dieci classi. I risultati sono riassunti nella successiva ***Figura 6.10***. Anche in questo caso, sono riportati, per ogni grafico, le tre tipologie diverse di semivariogramma (euclideo, bidirezionale e monodirezionale).

Per quanto riguarda le *distanze euclidee*, si nota un andamento simile a quello registrato per la zona del Tanaro, ossia un semivariogramma di tipologia *D*, per tutte le variabili, eccetto che la *N* (abbondanza totale di invertebrati).

Infatti, è evidente l'andamento fortemente eterogeneo delle 3 variabili *S*, *EPTN* ed *EPTS* (con picchi alti e bassi alternati della funzione di autocorrelazione, a testimoniare l'influenza di processi a scale spaziali molto varie), che si contrappone a quello più regolare della variabile *N*, in cui, dopo una prima fase di quasi totale assenza di correlazione (per piccole distanze, fino a circa 30km), si delinea una crescita continua della funzione $\gamma(h)$. Tale particolare forma del grafico permette di riconoscere un semivariogramma di tipologia *B*, la cui analisi teorica prevede, come struttura spaziale dominante, la presenza di un gradiente ambientale longitudinale (per distanze elevate). Ciò significa che, nella zona del Po, per quanto concerne l'abbondanza di invertebrati, è più ragionevole trovare caratteristiche ecologiche simili quando si considerano stazioni più distanti di 30km.

Per quanto riguarda le *distanze bidirezionali* lungo il reticolo, si registra molta meno eterogeneità nella struttura dei semivariogrammi, rispetto al medesimo caso, precedentemente commentato, del sottobacino del Tanaro. Anzi, in questo caso del sottobacino del Po, le conclusioni deducibili sono estremamente analoghe a quelle generate dai semivariogrammi delle distanze euclidee, appena commentati. Tutti i semivariogrammi si presentano della tipologia *D* (la cosiddetta struttura spaziale mista, *patchy + gradient*), eccetto che per, nuovamente, il caso della variabile N. Per tale variabile, infatti, si registra un andamento crescente della funzione (tipologia *B*), dopo un'iniziale assenza di correlazione, fino al raggiungimento della stazionarietà (per distanze comprese tra circa 70 e 100 km) e poi una decrescita finale, sempre circa lineare, che porta all'annullamento progressivo dell'autocorrelazione. Complessivamente, si può concludere che le scale spaziali dominanti nei processi ecologici lungo le aste del Po siano estremamente influenzate dalle caratteristiche locali, che determinano accensione e spegnimento dell'autocorrelazione spaziale per distanze molto variabili.

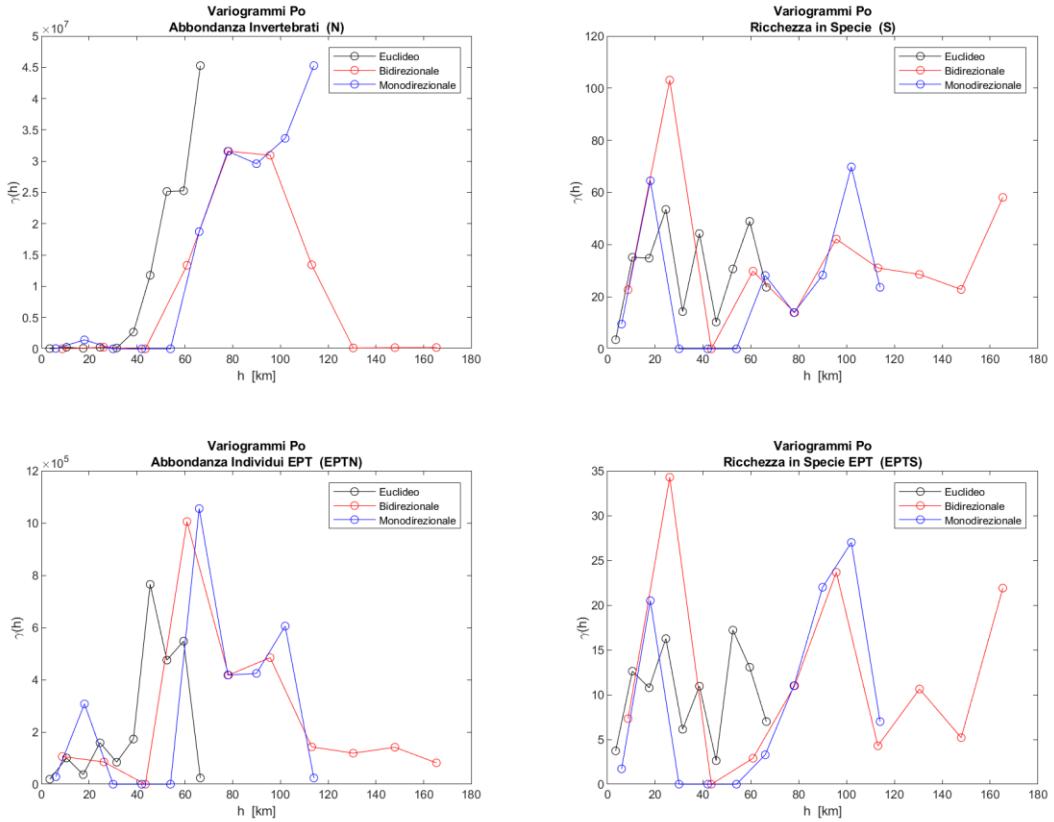

Figura 6.10 *– Semivariogrammi a dieci classi per la zona del Po. Sono riportate le curve ottenute considerando le distanze euclidee (nero), bidirezionali (rosso) e monodirezionali (blu). Ogni grafico si riferisce ad una delle quattro variabili ecologiche (N, S, EPTN ed EPTS).*

Anche nel caso di distanze lungo il reticolo *flow-connected* (rappresentato in blu in ***Figura 6.10***), si registrano dei semivariogrammi con caratteristiche simili a quelli ottenuti in precedenza con le distanze euclidee (forma simile, ma, naturalmente significato diverso nei due casi, in quanto uno, quello euclideo, si concentra sui legami trasversali e non correlati alla struttura del fiume o al movimento dell'acqua, l'altro, quello lungo fiume, su quelli longitudinali e nel senso della corrente). Nuovamente, l'unica variabile che presenta un semivariogramma di tipologia diversa dalla *D* è quello della variabile *N*, per la quale si registra un andamento di tipologia *B* dopo un certo valore di distanza (circa 50 km). Fino a tale valore sembra che ci sia proprio assenza totale di correlazione spaziale tra le stazioni di campionamento. Dopo di esso, la correlazione si accende e denota un meccanismo di influenza ad ampia scala spaziale (forma crescente del grafico).

A conclusione, analogamente a quanto fatto per il sottobacino del Tanaro, si procede, anche ora, ad un'analisi in termini di scarto quadratico medio dei semivariogrammi (in quanto l'esposizione dei risultati appena effettuata si riferisce sempre a delle caratteristiche medie del bacino).

Tabella 6.7 *– Zona del Po, semivariogrammi a dieci classi. Valori di deviazione standard della funzione di autocorrelazione γ(h), per ogni variabile ecologica e per ogni tipologia di semivariogramma.*

	Scarto Quadratico Medio Semivariogrammi (Po)		
	Euclideo	**Flow Unconnected**	**Flow Connected**
N	7.14E+06	7.65E+06	8.58E+06
S	39.20	37.68	24.79
EPTN	2.49E+05	2.59E+05	3.66E+05
EPTS	13.41	12.04	11.04

Da un'osservazione della ***Tabella 6.7***, si nota come si giunga alle medesime conclusioni riportate in precedenza per il sottobacino del Tanaro. Gli scostamenti inferiori riguardano sempre le due variabili S ed EPTS (in maniera perfettamente concorde, anche in questo caso, ai grafici valore previsto - valore misurato presentati alla fine del *Capitolo 5*).

Si riportano, ora, i risultati ottenuti per i semivariogrammi ad otto classi. Nuovamente, non si registra nulla di nuovo, gli andamenti dei grafici e i valori di scarto quadratico medio ricalcano esattamente quanto già descritto per le dieci classi di distanza.

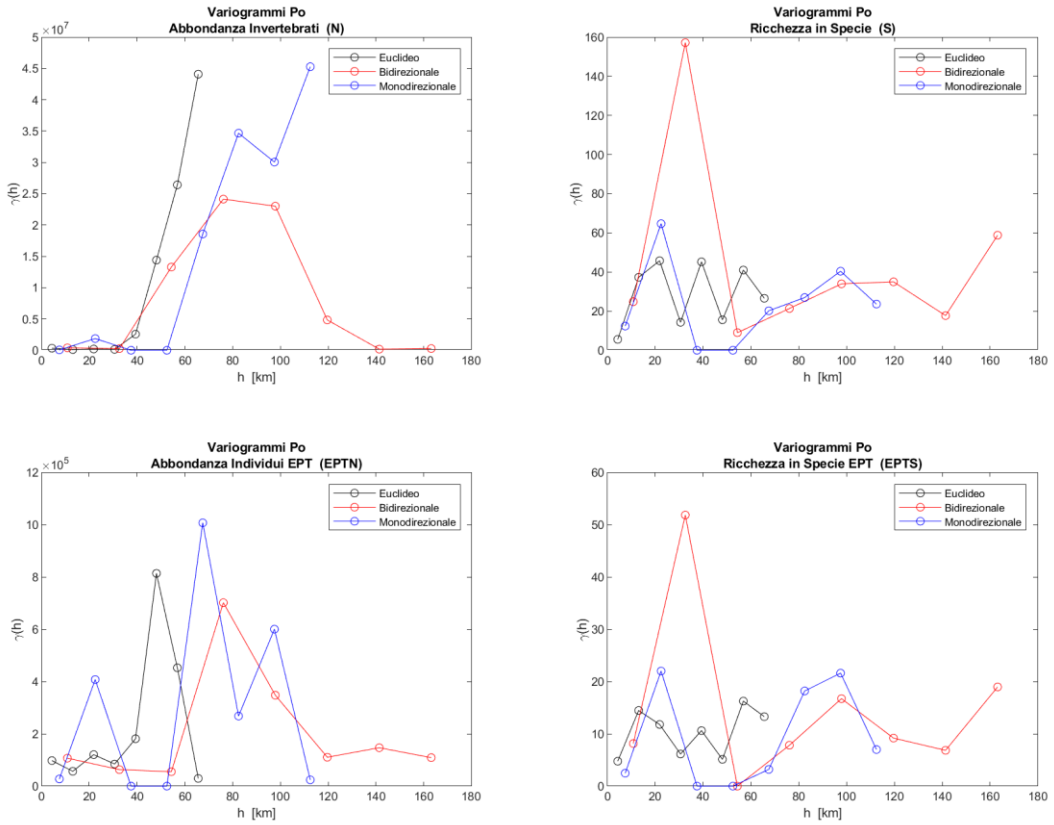

Figura 6.11 *– Semivariogrammi a otto classi per la zona del Po. Sono riportate le curve ottenute considerando le distanze euclidee (nero), bidirezionali (rosso) e monodirezionali (blu). Ogni grafico si riferisce ad una delle quattro variabili ecologiche (N, S, EPTN ed EPTS).*

Tabella 6.8 *– Zona del Po, semivariogrammi a otto classi. Valori di deviazione standard della funzione di autocorrelazione γ(h), per ogni variabile ecologica e per ogni tipologia di semivariogramma.*

	Scarto Quadratico Medio Semivariogrammi (Po)		
	Euclideo	**Flow Unconnected**	**Flow Connected**
N	6.79E+06	9.46E+06	7.95E+06
S	39.61	34.42	28.42
EPTN	2.39E+05	2.80E+05	3.38E+05
EPTS	13.54	11.96	11.44

Capitolo 7

Conclusioni

Lo scopo della trattazione eseguita e poi descritta in questo lavoro di tesi è stato, da subito, quello di cercare di gettare un po' di luce su un argomento di scottante attualità e di mutevoli e molteplici interessi. L'organizzazione spaziale delle comunità appartenenti agli ecosistemi fluviali è, infatti, una tematica ricchissima di risvolti sia in ambito scientifico, per quel che riguarda lo studio dal punto di vista biologico ed ecologico dei meccanismi naturali che regolano tale strutturazione spaziale, che in ambito ambientale-sociale, per quel che riguarda, invece, il rispetto della naturalità di un fiume (e, conseguentemente, della vita all'interno dell'ecosistema) e un suo eventuale sovrasfruttamento da parte delle attività antropiche.

La comprensione delle teorie ecologiche già assodate in letteratura ha permesso di entrare adeguatamente nel merito del problema. Studiando la *teoria delle metacomunità*, si può capire come ogni singolo aspetto di un ecosistema sia legato, tramite una catena di effetti a cascata (tipo tessere di un domino), a tutti gli altri. Le comunità non sono più viste come tanti gruppi di individui che occupano regioni distinte e divise dello spazio, ma come un tutt'uno, un qualcosa in continua interazione ed in costante mutazione (adattamento ai cambiamenti). Gli attori di questi fenomeni di adattamento sono, come detto, di due tipi: i, cosiddetti, determinanti locali, ossia quella serie di fenomeni che agisce sulla piccola scala (come le condizioni ambientali o le interazioni biotiche tra le comunità), e quelli regionali (ossia quei meccanismi a più ampia scala spaziale che hanno a che fare con la struttura dendritica dei sistemi fluviali e la possibilità e capacità delle specie presenti al movimento, alla dispersione).

Si capisce, dunque, come il tema, tanto attuale, del mantenimento della biodiversità in una determinata zona di territorio sia molto più complesso di quello di cui si sente parlare molte volte ed in maniera superficiale, attraverso i media. Come quasi sempre accade quando si conosce a fondo un argomento scientifico, c'è molto di più da considerare. La biodiversità non è il prodotto del solo benessere dell'habitat, in termini di qualità delle caratteristiche fisiche o chimiche. Una degradazione della biodiversità non è solo dovuta alla denaturalizzazione di un ambiente (da parte di attività umane o altro), ma è il risultato anche (e, talvolta, soprattutto) di una serie di meccanismi e dinamiche che con gli aspetti locali ed ambientali centrano ben poco. In parole povere, molte volte, per risanare un ecosistema non è sufficiente risanare l'habitat.

Il database, da cui è partita l'analisi spaziale successivamente eseguita, è stato creato proprio a questo scopo: studiare il mantenimento della biodiversità spaziale lungo i maggiori fiumi del reticolo idrografico piemontese. I dati sono quasi sempre stati campionati nei pressi di centraline idroelettriche, in zone moderatamente influenzate da interventi antropici. L'elevata estensione areale delle informazioni disponibili ha permesso, inoltre, di effettuare uno studio su scala

territoriale sufficientemente ampia da poter registrare gli effetti di meccanismi a diverse dinamiche spaziali (e tentare anche di studiarli in maniera generale).

A questo scopo, sono state eseguite le prime analisi statistiche, mediante la tecnica di regressione lineare. Dopo aver convertito e riassunto le centinaia di dati ecologici in quattro parametri sintetici (ma comunque rappresentativi dei dati iniziali), si è cercato di creare un modello matematico che si adattasse ai dati in maniera adeguata e permettesse di trovare un legame tra la distribuzione delle comunità (dati ecologici di abbondanza e ricchezza in specie) e fattori idromorfologici (che, come detto, hanno permesso di tenere in considerazioni sia le dinamiche locali, in maniera diretta, sia regionali, in maniera a volte diretta, a volte indiretta). La creazione di queste leggi ha inevitabilmente condotto ad una semplificazione del fenomeno reale, ossia, all'ipotesi iniziale che i meccanismi di strutturazione ecologica delle comunità variassero nello spazio in maniera lineare. Per misurare e definire l'adeguatezza del modello alla distribuzione di dati di partenza, è stato usato il coefficiente di determinazione R^2.

In particolare, si è constatato come questo coefficiente variasse, in maniera più o meno elevata, a seconda della variabile ecologica considerata (N, S, EPTN, EPTS) o del modello assunto (monoparametrico o multiparametrico, lineare od esponenziale), ma, soprattutto, a seconda della omogeneità dei dati idromorfologici in ingresso. L'aumento di quest'ultima componente, mediante le due zonazioni effettuate alla fine delle analisi di regressione, ha mostrato infatti un netto aumento di valore nella correlazione tra variabili (e, quindi, nella significatività dei modelli).

Un'ulteriore affinazione dei risultati è stata eseguita a seguito della costruzione dei modelli lineari, mediante la definizione dei semivariogrammi. Essi sono particolari tipologie di diagrammi statistici serviti a studiare i concetti relativi alla teoria del *continuum fluviale*, una teoria riguardante le modalità e tipologie di cambiamenti in caratteristiche spaziali lungo le aste di un reticolo, e costruiti mediante il calcolo dell'autocorrelazione spaziale delle variabili ecologiche (che contengono già in sé informazioni relative alla distribuzione spaziale delle comunità). La definizione della funzione di autocorrelazione (funzione γ) viene eseguita tramite il calcolo di vari tipi di distanze tra le stazioni di campionamento (euclidee o lungo il reticolo). Dall'analisi della forma finale dei semivariogrammi ottenuti, è stato possibile effettuare un'analisi spaziale a livello di due singoli bacini idrografici (gli stessi delle zonazioni precedentemente effettuate durante le regressioni). Si è notato, in particolare, come, in alcuni casi, fosse riscontrabile la presenza di fenomeni spaziali ad ampia scala (gradienti ambientali longitudinali, che implicano cambiamenti graduali lungo il fiume, da monte verso valle) e, in altri (la maggioranza), fosse presente una forte influenza delle situazioni locali. Infatti, nei sistemi naturali reali, la presenza di eterogeneità locali (in termini di cambiamenti improvvisi in condizioni ambientali o di situazioni particolari di interazione tra le specie) è una costante quasi mai trascurabile.

Si ricorda che tutte le tipologie di analisi spaziale, messe in pratica in questo lavoro, sono riconducibili ad esperienze già assodate nel corso di studi internazionali in campo di eco-idraulica fluviale. In particolare, esse rappresentano il passo preliminare e fondamentale da compiere prima di poter eseguire studi più complessi, qualora si disponesse di dati maggiormente dettagliati e della possibilità di fare verifiche e campionamenti sul campo di ulteriori parametri utili.

Il risultato complessivo è stato, dunque, uno studio (a scala sia regionale, su tutto il database, che locale, nel caso dei singoli bacini idrografici) sull'influenza che fattori idraulici, idrologici e

morfologici possono avere sulla vita e la distribuzione delle comunità ecologiche di macroinvertebrati bentonici lungo i maggiori corsi d'acqua del Piemonte.

La particolarità dei risultati ottenuti, in termini di differenti approcci di analisi a scala spaziale diversa, non esclude però la possibilità di maggiori approfondimenti sull'argomento (su dati, appunto, più dettagliati e con tecniche differenti). Questo lavoro può essere visto come un punto di partenza generale, per quanto riguarda un argomento estremamente interessante ed attuale (dai vari punti di vista che sono stati detti), nonché complesso, così come lo sono tutti i fenomeni naturali riguardanti i sistemi fluviali.

Appendice

ALLEGATO 1

Script python-GRASS per l'estrazione dei bacini idrografici

```
# Si ricorda che il commento completo dello script è inserito (passo per passo) all'interno del Capitolo 3.

#### LETTURA coordinate EST e NORD ####
python

coord_E = []
coord_N = []

with open('Data.txt', 'r') as f:
    n_bacini = f.readlines()
    for x in n_bacini:
        raw = x.split()
        coord_E.append(float(raw[0]))
        coord_N.append(float(raw[1]))

#### LETTURA nomi Fiumi ####
fiume = []

with open('Fiumi_nome.txt', 'r') as f2:
    N = f2.readlines()
    for y in N:
        righe = y.split()
        fiume.append(righe[0])

#### ESTRAZIONE BACINO ####
from grass.pygrass import raster
import grass.script as gscript

for bacino in range(0,len(n_bacini)):
    bacino
    gscript.parse_command('g.region', raster='piemonte_drain_r100', res=100)
    gscript.parse_command('r.water.outlet', input='piemonte_drain_r100', output='Bacino_' +str(bacino), coordinates=str(coord_E[bacino])+','+str(coord_N[bacino]),
overwrite=True)
    gscript.parse_command('g.region', zoom='Bacino_' +str(bacino), res=100, overwrite=True)
    gscript.parse_command('r.mask', raster='Bacino_' +str(bacino), maskcats='1', overwrite=True)
    gscript.parse_command('r.to.vect', input='Bacino_' +str(bacino), output=fiume[bacino]+'_'+str(bacino), type='area', overwrite=True)
    gscript.parse_command('v.out.ogr', flags='c', input=fiume[bacino]+'_'+str(bacino), output='shp', type='area', format='ESRI_Shapefile', overwrite=True)
    gscript.parse_command('r.mask', flags='r')
```

ALLEGATO 2

Script python-GRASS per l'estrazione dei parametri descrittori dei bacini

```
# Si ricorda che il commento completo dello script è inserito (passo per passo) all'interno del Capitolo 3.


#### LETTURA coordinate EST e NORD ####

# Collegare il terminale alla directory di lavoro (comando "cd percorso_cartella")
cd C:\Users\fabio\Desktop\RENERFORGisDatabase\WGS84_UTM32N_RENERFOR\renerfor

python
from grass.pygrass import raster
import grass.script as gscript

coord_E = []
coord_N = []

with open('Data.txt', 'r') as f:
   n_bacini = f.readlines()
   for x in n_bacini:
      raw = x.split()
      coord_E.append(float(raw[0]))
      coord_N.append(float(raw[1]))

#### LETTURA nomi Fiumi ####

fiume = []

with open('Fiumi_nome.txt', 'r') as f2:
   N = f2.readlines()
   for y in N:
      righe = y.split()
      fiume.append(righe[0])

##################################################

# Definisco vettori dei descrittori
quota_media_v = []
quota_max_v = []
MAP_v = []
IDFa_v = []
IDFa_devstd_v = []
fourier_B1_v = []
cv_rp_v = []
for bacino in range(0,len(n_bacini)):
   bacino
   gscript.parse_command('v.in.ogr', flags='o', input='shp', layer=fiume[bacino]+'_'+str(bacino), output='Bacino_n_' +str(bacino), min_area=0, overwrite=True)
   gscript.parse_command('g.region', flags='d', overwrite=True)
   gscript.parse_command('g.region', vector='Bacino_n_' +str(bacino), res=100, overwrite=True)
   gscript.parse_command('v.to.rast', input='Bacino_n_' +str(bacino), output='Bacino_n_' +str(bacino), use='cat', type='area', overwrite=True)
   gscript.parse_command('r.mask', raster='Bacino_n_' +str(bacino), overwrite=True)
   # QUOTA MEDIA e MAX
   gscript.parse_command('g.region', raster='Bacino_n_' +str(bacino), res=100, overwrite=True)
   stats_1 = gscript.parse_command('r.univar', flags='g', map='piemonte_dem_r100', overwrite=True)
   quota_media = float(stats_1.mean)
   quota_max = float(stats_1.max)
   # MEDIA AFFLUSSO ANNUO
   gscript.parse_command('g.region', raster='Bacino_n_' +str(bacino), res=250, overwrite=True)
   stats_2 = gscript.parse_command('r.univar', flags='g', map='piemonte_MAP_r250', overwrite=True)
   MAP = float(stats_2.mean)
   # CPP (Media e Deviazione Standard)
   gscript.parse_command('g.region', raster='Bacino_n_' +str(bacino), res=250, overwrite=True)
   stats_3 = gscript.parse_command('r.univar', flags='g', map='piemonte_IDFa_r250', overwrite=True)
   IDFa = float(stats_3.mean)
   IDFa_devstd = float(stats_3.stddev)
   # MEDIA COEFFICIENTE REGIME PLUVIOMETRICO B1
   gscript.parse_command('g.region', raster='Bacino_n_' +str(bacino), res=50, overwrite=True)
   stats_4 = gscript.parse_command('r.univar', flags='g', map='piemonte_fourierB1_r50', overwrite=True)
   fourier_B1 = float(stats_4.mean)
   # MEDIA COEFFICIENTE VARIAZIONE REGIME PLUVIOMETRICO
```

```
    gscript.parse_command('g.region', raster='Bacino_n_' +str(bacino), res=50, overwrite=True)
    stats_5 = gscript.parse_command('r.univar', flags='g', map='piemonte_rp_cv_r50', overwrite=True)
    cv_rp = float(stats_5.mean)
    # Popolare vettori descrittori (aggiunto un valore ogni ciclo)
    quota_media_v.append(quota_media)
    quota_max_v.append(quota_max)
    MAP_v.append(MAP)
    IDFa_v.append(IDFa)
    IDFa_devstd_v.append(IDFa_devstd)
    fourier_B1_v.append(fourier_B1)
    cv_rp_v.append(cv_rp)

# Import database e aggiunta nuove colonne descrittori
gscript.parse_command('v.in.ogr', flags='o', input='shp', layer='database', output='Database_Aggiornato', min_area=0, overwrite=True)
gscript.parse_command('v.db.addcolumn', map='Database_Aggiornato', columns='Q_Med double precision, Q_Max double precision, MAP double precision, IDFa_m
double precision, IDFa_std double precision, B1 double precision, cv_rp double precision, Area double precision', overwrite=True)



# Inserimento valori nelle colonne dello shapefile (uno per descrittore)
for i in range(len(quota_media_v)):
    # QUOTA MEDIA
    query0="UPDATE Database_Aggiornato SET Q_Med=" + str(quota_media_v[i]) + " WHERE cat = " + str(i+1)
    gscript.parse_command('db.execute', sql=query0)
    # QUOTA MAX
    query1="UPDATE Database_Aggiornato SET Q_Max=" + str(quota_max_v[i]) + " WHERE cat = " + str(i+1)
    gscript.parse_command('db.execute', sql=query1)
    # MAP
    query2="UPDATE Database_Aggiornato SET MAP=" + str(MAP_v[i]) + " WHERE cat = " + str(i+1 )
    gscript.parse_command('db.execute', sql=query2)
    # IDFa
    query3="UPDATE Database_Aggiornato SET IDFa_m=" + str(IDFa_v[i]) + " WHERE cat = " + str(i+1)
    gscript.parse_command('db.execute', sql=query3)
    # IDFa_std
    query4="UPDATE Database_Aggiornato SET IDFa_std=" + str(IDFa_devstd_v[i]) + " WHERE cat = " + str(i+1)
    gscript.parse_command('db.execute', sql=query4)
    # B1
    query5="UPDATE Database_Aggiornato SET B1=" + str(fourier_B1_v[i]) + " WHERE cat = " + str(i+1)
    gscript.parse_command('db.execute', sql=query5)
    #cv_rp
    query6="UPDATE Database_Aggiornato SET cv_rp=" + str(cv_rp_v[i]) + " WHERE cat = " + str(i+1)
    gscript.parse_command('db.execute', sql=query6)

# Esportazione shapefile database aggiornato
gscript.parse_command('v.out.ogr', flags='c', input='Database_Aggiornato', output='shp', type='point', format='ESRI_Shapefile', overwrite=True)
```

ALLEGATO 3

Script Matlab per le Regressioni Monoparametriche

```
% IMPORT DATI - Modificarlo se aggiunti nuovi parametri

% STRUTTURA DATI DA IMPORTARE: Creare un foglio excel con inseriti
% all'interno tutti i valori dei dati da importare. Creare una tabella in
% cui le righe rappresentano le differenti stazioni di campionamento e ogni
% colonna è dedicata ad una variabile (ecologica o idromorfologica). Ad
% esempio, in questo caso, ci sono 320 stazioni di campionamento e 14
% variabili idromorfologiche. La tabella sarà composta da 320 righe (una
% per stazione) e 14 colonne (una per variabile). Ogni cella rappresenta il
% valore di una determinata variabile in una determinata stazione. Lo
% stesso vale per le 4 variabili ecologiche.


% Creazione delle due matrici contenenti i dati. M_idro sarà una matrice
% contenente su ogni colonna i valori di una delle 14 variabili idromorfologiche scelte
% per l'analisi. M_eco sarà una matrice contenente su ogni colonna i
% valori di una delle 4 variabili ecologiche --> L'import viene fatto
% direttamente dal file excel specificato, nel foglio specificato e nel
% range di celle specificato.

M_idro = readmatrix('Database_Matlab_Correlazioni.xlsx','Sheet',1,'Range','G1:T320'); % MOD - legge i parametri
idromorfologici
M_eco = readmatrix('Database_Matlab_Correlazioni.xlsx','Sheet',1,'Range','C1:F320');  % legge i parametri
ecologici

% Nomi per TITOLI ASSI
Nomi_idro = readvars('Database_Matlab_Correlazioni.xlsx','Sheet',2,'Range','A1:A14'); % importa i nomi delle
variabili idromorf. con cui mettere il titolo agli assi dei grafici
Nomi_eco = readvars('Database_Matlab_Correlazioni.xlsx','Sheet',2,'Range','D1:D4');

% Nomi per TITOLO GRAFICO
Titoli_idro = readvars('Database_Matlab_Correlazioni.xlsx','Sheet',2,'Range','B1:B14'); % importa i nomi delle
variabili idromorf. per titolare i grafici
Titoli_eco = readvars('Database_Matlab_Correlazioni.xlsx','Sheet',2,'Range','D9:D12');


%%%% GENERAZIONE SCATTERPLOT E CALCOLO CORRELAZIONI %%%%

k=1; % indice riempimento vettori (serve per salvare i valori calcolati nel ciclo nella posizione giusta)
m = zeros(size(M_eco,2)*size(M_idro,2),1); % vettore che verrà riempito dai coefficienti angolari delle rette di
regressione
R_adj = zeros(size(M_eco,2)*size(M_idro,2),1); % vettore che verrà riempito dagli R_adj
R = zeros(size(M_eco,2)*size(M_idro,2),1); % vettore R
t_value = zeros(size(M_eco,2)*size(M_idro,2),1); % vettore per il test t-Student

% Ciclo che genera tutti gli scatterplot --> Mette in correlazione ogni
% variabile ecologica (4 totali) con ogni variabile idromorfologica (14
% totali) --> Genera 4 figure totali (una per var. ecologica), in ognuna
% delle quali ci sono 14 scatterplot.
for i=1:size(M_eco,2)% ciclo che tiene fisso il parametro ecologico
    for j=1:size(M_idro,2) % ciclo che va su ogni parametro idro-morf
    figure(i);

    % Coefficiente m della linea di regressione
    p = polyfit(M_idro(:,j),M_eco(:,i),1);
    y_fit = polyval(p,M_idro(:,j));
    m(k)=p(1); % salvataggio del valore calcolato nel vettore degli m

    % Calcolo R e R_adj --> Applicazione formule dalla teoria
    y_resid = M_eco(:,i) - y_fit;
    SS_resid = sum(y_resid.^2);
    SS_total = (length(M_eco(:,i))-1) * var(M_eco(:,i));
    R(k) = 1 - SS_resid/SS_total;
    R_adj(k) = 1 - ((SS_resid/SS_total)*((length(M_eco(:,i))-1)/(length(M_eco(:,i))-2)));

    % Calcolo test t-Student --> Formule dalla teoria
    MS_res = SS_resid/(length(M_eco(:,i))-2);
    Sxx = sum((M_idro(:,j)-mean(M_idro(:,j))).^2);
    t_value(k) = abs(m(k))/sqrt(MS_res/Sxx);

    % Generazione del grafico finale --> E' stato messo l'if per
    % distinguere i due casi in cui la regressione passa il t-test oppure
    % no. Se lo passa, il comando colora la retta di regressione in rosso,
    % altrimenti la colora di nero.
    if t_value(k)> 1.97
        subplot(2,7,j) % MODIFICARE se aggiungo parametri idro-morf (poichè ci saranno più grafici da inserire in
figura)
        scatter(M_idro(:,j),M_eco(:,i),8,'+');
```

```
        hold on
        plot(M_idro(:,j),y_fit,'r'); % linea regressione
        xlabel(Nomi_idro(j));
        ylabel(Nomi_eco(i));
%         text_1=sprintf('R_{adj}^2 = %f', R_adj(k)); % comando per inserire il valore di R_adj nel grafico
%         text(0.52*max(M_idro(:,j)),0.96*max(M_eco(:,i)),text_1); % comando per inserire il valore di R_adj nel
grafico
        title(['Scatterogramma ',Titoli_idro(j),Titoli_eco(i)]);
    else
        subplot(2,7,j) % MODIFICARE se aggiungo parametri idro-morf
        scatter(M_idro(:,j),M_eco(:,i),8,'+');
        hold on
        plot(M_idro(:,j),y_fit,'k'); % linea regressione
        xlabel(Nomi_idro(j));
        ylabel(Nomi_eco(i));
%         text_1=sprintf('R_{adj}^2 = %f', R_adj(k));
%         text(0.52*max(M_idro(:,j)),0.96*max(M_eco(:,i)),text_1);
        title(['Scatterogramma ',Titoli_idro(j),Titoli_eco(i)]);
    end

    k=k+1; % Aumento indice riempimento vettori

    end

    % Salvare figure in automatico .png (jpeg)
    set(gcf, 'PaperPositionMode', 'manual');
    set(gcf, 'PaperUnits', 'centimeters');
    set(gcf, 'PaperOrientation', 'portrait');
    set(gcf, 'PaperPosition', [0 0 35 25]);%formato del foglio 35x25 cm
    print(gcf,['Scatterogramma_Monoparametrico_' num2str(i) '.png'],'-dpng');

end

% Scrivere i risultati di interesse come colonna excel
writematrix(R_adj,'Tabella_Correlazioni.xlsx','Sheet',1,'Range','L2:L500');
writematrix(m,'Tabella_Correlazioni.xlsx','Sheet',1,'Range','M2:M500');
writematrix(t_value,'Tabella_Correlazioni.xlsx','Sheet',1,'Range','N2:N500');
```

ALLEGATO 4

Script Matlab per valutare la collinearità tra variabili idromorfologiche

```
% IMPORT DATI - Modificarlo se aggiunti nuovi parametri

% STRUTTURA DATI DA IMPORTARE: Creare un foglio excel con inseriti
% all'interno tutti i valori dei dati da importare. Creare una tabella in
% cui le righe rappresentano le differenti stazioni di campionamento e ogni
% colonna è dedicata ad una variabile (idromorfologica). Ad
% esempio, in questo caso, ci sono 320 stazioni di campionamento e 14
% variabili idromorfologiche. La tabella sarà composta da 320 righe (una
% per stazione) e 14 colonne (una per variabile). Ogni cella rappresenta il
% valore di una determinata variabile in una determinata stazione.

% Creazione della matrice contenente i dati. M_idro sarà una matrice
% contenente su ogni colonna i valori di una delle 14 variabili
% idromorfologiche scelte per l'analisi --> L'import viene fatto direttamente
% dal file excel specificato, nel foglio specificato e nel range di celle
% specificato.

% Definizione matrice M_idro
M_idro = readmatrix('Database_Matlab_Correlazioni_2.xlsx','Sheet',1,'Range','G1:T320'); % MOD - legge i parametri
idromorfologici

% Nomi per TITOLI ASSI
Nomi_idro = readvars('Database_Matlab_Correlazioni_2.xlsx','Sheet',2,'Range','A1:A14'); % MOD

% Nomi per TITOLO GRAFICO
Titoli_idro = readvars('Database_Matlab_Correlazioni_2.xlsx','Sheet',2,'Range','B1:B14'); % MOD


%%%% GENERAZIONE SCATTERPLOT %%%%

k=1; % indice riempimento vettori
m = zeros(size(M_idro,2)*size(M_idro,2),1); % vettore coi coefficienti angolari
R_adj = zeros(size(M_idro,2)*size(M_idro,2),1); % vettore coi R_adj
R = zeros(size(M_idro,2)*size(M_idro,2),1); % vettore coi R
t_value = zeros(size(M_idro,2)*size(M_idro,2),1); % vettore per il test t-Student

% Ciclo che genera tutti gli scatterplot
for i=1:size(M_idro,2)% ciclo che tiene fisso il parametro idro-morf
    for j=1:size(M_idro,2) % ciclo che va su ogni parametro idro-morf
    figure(i);

    % Coefficiente m della linea di regressione
    p = polyfit(M_idro(:,j),M_idro(:,i),1);
    y_fit = polyval(p,M_idro(:,j));
    m(k)=p(1);

    % Calcolo R e R_adj
    y_resid = M_idro(:,i) - y_fit;
    SS_resid = sum(y_resid.^2);
    SS_total = (length(M_idro(:,i))-1) * var(M_idro(:,i));
    R(k) = 1 - SS_resid/SS_total;
    R_adj(k) = 1 - ((SS_resid/SS_total)*((length(M_idro(:,i))-1)/(length(M_idro(:,i))-2)));

    % Calcolo test t-Student
    MS_res = SS_resid/(length(M_idro(:,i))-2);
    Sxx = sum((M_idro(:,j)-mean(M_idro(:,j))).^2);
    t_value(k) = abs(m(k))/sqrt(MS_res/Sxx);

    % grafico finale
    if t_value(k)> 1.97
        subplot(2,7,j) % MODIFICARE se aggiungo parametri idro-morf
        scatter(M_idro(:,j),M_idro(:,i),8,'+');
        hold on
        plot(M_idro(:,j),y_fit,'r'); % linea regressione
        xlabel(Nomi_idro(j));
        ylabel(Nomi_idro(i));
        text_1=sprintf('R_{adj}^2 = %f', R_adj(k));
        text(0.52*max(M_idro(:,j)),0.96*max(M_idro(:,i)),text_1);
        title(['Scatterogramma ',Titoli_idro(j),Titoli_idro(i)]);
    else
        subplot(2,7,j) % MODIFICARE se aggiungo parametri idro-morf
        scatter(M_idro(:,j),M_idro(:,i),8,'+');
        hold on
        plot(M_idro(:,j),y_fit,'k'); % linea regressione
        xlabel(Nomi_idro(j));
        ylabel(Nomi_idro(i));
        text_1=sprintf('R_{adj}^2 = %f', R_adj(k));
```

```
        text(0.52*max(M_idro(:,j)),0.96*max(M_idro(:,i)),text_1);
        title(['Scatterogramma ',Titoli_idro(j),Titoli_idro(i)]);
    end

    k=k+1; % Aumento indice riempimento vettori

    end
end


% Scrivere i risultati di interesse come colonna excel
writematrix(R_adj,'Tabella_Collinearita_2.xlsx','Sheet',1,'Range','S2:S500');   % MOD - Specificare le colonne
giuste
writematrix(m,'Tabella_Collinearita_2.xlsx','Sheet',1,'Range','T2:T500');       % MOD - Specificare le colonne
giuste
writematrix(t_value,'Tabella_Collinearita_2.xlsx','Sheet',1,'Range','U2:U500'); % MOD - Specificare le colonne
giuste
```

ALLEGATO 5

Script Matlab per la Regressione Multivariata

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% IMPORT DATI - Modificarlo se aggiunti nuovi parametri

% STRUTTURA DATI DA IMPORTARE: Creare un foglio excel con inseriti
% all'interno tutti i valori dei dati da importare. Creare una tabella in
% cui le righe rappresentano le differenti stazioni di campionamento e ogni
% colonna è dedicata ad una variabile (ecologica o idromorfologica). Ad
% esempio, in questo caso, ci sono 320 stazioni di campionamento e 14
% variabili idromorfologiche. La tabella sarà composta da 320 righe (una
% per stazione) e 14 colonne (una per variabile). Ogni cella rappresenta il
% valore di una determinata variabile in una determinata stazione. Lo
% stesso vale per le 4 variabili ecologiche.

% Creazione delle due matrici contenenti i dati. M_idro sarà una matrice
% contenente su ogni colonna i valori di una delle 14 variabili idromorfologiche scelte
% per l'analisi. M_eco sarà una matrice contenente su ogni colonna i
% valori di una delle 4 variabili ecologiche --> L'import viene fatto
% direttamente dal file excel specificato, nel foglio specificato e nel
% range di celle specificato.


% Creazione delle due matrici coi valori delle variabili (import da excel).
M_idro = readmatrix('Database_Matlab_Multivariata.xlsx','Sheet',1,'Range','G1:T320'); % MOD - legge i parametri
idromorfologici
M_eco = readmatrix('Database_Matlab_Multivariata.xlsx','Sheet',1,'Range','C1:F320');  % legge i parametri
ecologici

% Nomi per TITOLI ASSI
Nomi_idro = readvars('Database_Matlab_Multivariata.xlsx','Sheet',2,'Range','A1:A14'); % MOD
Nomi_eco = readvars('Database_Matlab_Multivariata.xlsx','Sheet',2,'Range','D1:D4');

% Nomi per TITOLO GRAFICO
Titoli_idro = readvars('Database_Matlab_Multivariata.xlsx','Sheet',2,'Range','B1:B14'); % MOD
Titoli_eco = readvars('Database_Matlab_Multivariata.xlsx','Sheet',2,'Range','D9:D12');

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% REGRESSIONE AD 1 PARAMETRO %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Creazione matrice con le variabili di interesse (non correlate)
variabili = ["h_m", "Q", "E", "S.O", "LSM", "MAP", "Nat", "Morf"]; % stringa coi nomi delle variabili scelte per
la regressione
j=2; % indice, parte da 2 perchè la prima colonna è riservata agli 1
k=1; % indice per confronto (nell'if dentro al ciclo)
Idro = zeros(length(M_idro),length(variabili)+1); % def. nuova matrice da riempire nel ciclo
Idro(:,1)=ones(length(M_idro),1); % prima colonna di 1
for i=1:length(Titoli_idro)
    if variabili(k)==Titoli_idro(i)
        Idro(:,j)=M_idro(:,i);
        j=j+1;
        k=k+1;
        if j==length(variabili)+2
            break
        end
    end
end % Risultato --> Matrice Idro --> E' quella che ci serve per usare correttamente il comando regress.
    % Essa conterrà la prima colonna con gli 1 e poi tutte le 8 variabili scelte per la regressione.

% Regressione --> Calcolo R_adj

k=1; % indice riempimento vettori
R_adj = zeros(size(M_eco,2)*(size(Idro,2)-1),1); % vettore coi R_adj
R = zeros(size(M_eco,2)*(size(Idro,2)-1),1); % vettore coi R
R_max = zeros(size(M_eco,2),1); % Vettore che conterrà i 4 valori di R massimi (uno per variabile ecologica).
                                % Ogni R_max corrisponderà alla regressione
                                % di una determinata variabile, la cui
                                % posizione nella stringa iniziale
                                % "variabili" sarà salvata in pos.
pos = zeros(size(M_eco,2),1); % posizione nella stringa "variabili" della variabile migliore
                              % ossia quella assunta nel modello perché ha
                              % prodotto una regressione con R^2 massimo.
                              % Ci saranno quattro valori di posizione (uno
                              % per ogni variabile ecologica)--> Ricordiamo
                              % che si stanno facendo quattro regressioni
                              % diverse.
Idro_2 = zeros(length(Idro),2); % è la matrice che viene definita (riempita)
```

```
                                % ad ogni ciclo. Essa, per via del funzionamento del    comando regress, conterrà
sempre la
                                % prima colonna di 1, e poi (siccome siamo
                                % nel caso di regressione ad 1 variabile)
                                % conterrà ad ogni giro del ciclo una
                                % variabile idromorfologica diversa (in
                                % modo da poter svolgere tutte le
                                % regressioni e trovare, alla fine, quella
                                % che ha prodotto R^2 massimo). --> Il
                                % valore di R^2 e la posizione della
                                % variabile (che l'ha prodotto) sono
                                % salvati in R e pos (definiti prima).

% Ciclo che calcola le 4 variabili migliori --> In totale si fanno 4 giri
% del ciclo esterno (che corrispondono alle 4 regressioni, una per N, una per S, una
% per EPTN ed una per EPTS). Per ogni giro del ciclo esterno si fa un
% numero di cicli interni par al numero di variabili non ancora assunte dal
% modello (in questo caso, in cui non ne è stata assunta ancora nessuna) se
% ne fanno 8 (uno per ogni var. idromorf.). Alla fine del ciclo si salvano
% i valori dell'R^2 massimo e della posizione della variabile migliore.

for i=1:size(M_eco,2)% ciclo che tiene fisso il parametro ecologico
    for j=2:size(Idro,2) % ciclo che va su ogni parametro idro-morf

    % Definizione della matrice per la regressione
    Idro_2(:,1)=Idro(:,1); %creazione della prima colonna di 1
    Idro_2(:,2)=Idro(:,j); % seconda colonna (diversa ad ogni giro), con la variabile idromorfologica
    % Regressione
    b = regress(M_eco(:,i),Idro_2(:,1:2)); % vettore coi coeff di regressione --> b = [b(1),b(2)]
    y_fit = b(1) + b(2)*Idro(:,j);
    % Calcolo R e R_adj --> Formule dalla teoria
    y_resid = M_eco(:,i) - y_fit;
    SS_resid = sum(y_resid.^2);
    SS_total = (length(M_eco(:,i))-1) * var(M_eco(:,i));
    R(k) = 1 - SS_resid/SS_total;
    R_adj(k) = 1 - ((SS_resid/SS_total)*((length(M_eco(:,i))-1)/(length(M_eco(:,i))-2)));

    k=k+1; %aumento indice riempimento vettori

    end

    if i==1
        R_max(i) = max(R_adj(1:length(variabili))); % valore di R^2 massimo
        pos(i) = find(R_adj(1:length(variabili))==R_max(i)); % posizione della var. che l'ha determinato.
    else
        R_max(i) = max(R_adj((((i-1)*length(variabili))+1):(i*length(variabili)))); % valore di R^2 massimo
        pos(i) = find(R_adj((((i-1)*length(variabili))+1):(i*length(variabili)))==R_max(i)); % posizione della
var. che l'ha determinato.
    end
end % Del ciclo ci interessa mantenere pos (con le posizioni riferite alla stringa
    % "variabili" definita all'inizio dello script ed il vettore di gli Rmax (uno ciascuno, in ordine
    % per N, S, EPT, EPTS). --> Nel caso di questo primo step della
    % regressione (in questo caso monoparametrica) non è necessario il
    % passaggio ulteriore (che sarà necessario nei prossimi step) di
    % riferire la variabile pos alla posizione effettiva della stringa
    % "variabili" inziale. Lo è già di suo.


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% REGRESSIONE A 2 PARAMETRI %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Definizione di due matrici --> Una (Idro_i) che contiene le variabili già assunte nel
% modello (che varia solo quando varia il parametro ecologico--> Ciclo
% esterno) --> L'altra (Idro_j) è l'unione di 4 matrici contenenti una di seguito
% all'altra i parametri non ancora inclusi nel modello (sarà una matrice
% che conterrà una colonna di 1, poi tutte le variabili non ancora incluse
% nel modello della prima variabile ecologica, poi di nuovo una colonna di
% 1 seguita da tutte le variabili non ancora incluse nel modello della
% seconda variabile ecologica.. e uguale per la 3 e 4 var. ecologica) -->
% In questo caso (in cui si ha una sola variabile già assunta) Idro_j sarà
% una matrice di 8*4=32 colonne (4 di soli 1, 7*4=28 delle variabili non
% ancora assunte).

Idro_i = zeros(length(M_idro),size(M_eco,2)); % def. matrice che contiene il primo parametro già assunto nel
modello
Idro_j = zeros(length(M_idro),size(M_eco,2)*(length(variabili)-1)); % matrice spiegata sopra (necessaria per
utilizzare correttamente
                                                                    % il
                                                                    % comando
                                                                    % regress).

% Creazione Matrice che varia con il par ecologico (ciclo esterno) -->
% Matrice contenente le quattro variabili (una per regressione) assunte nei
% modelli allo step precedente della regressione.
for i=1:size(M_eco,2)
    Idro_i(:,i)=Idro(:,pos(i)+1);
end
```

```
% Definizione matrice che va su ogni parametro non ancora assunto nel
% modello --> Idro_j
n=1;
for i=1:size(M_eco,2) % per ogni parametro ecologico
    j=1;
    m=1;
    while m~=length(variabili)+1
        if variabili(m)==Titoli_idro(j) && variabili(m)~=variabili(pos(i))
            Idro_j(:,n)=M_idro(:,j);
            m=m+1;
            n=n+1;
        else if variabili(m)==variabili(pos(i))
                m=m+1;
            end
        end

        j=j+1;
    end
end % Risultato --> Nuova matrice Idro_j con 7 parametri (1 tolto)

%%%%%%%%%%%%%%%%%%
k=1; % indice riempimento vettori R,R_adj
R_adj = zeros(size(Idro_j,2),1); % vettore coi R_adj
R = zeros(size(Idro_j,2),1); % vettore coi R
R_max_2 = zeros(size(M_eco,2),1); % Vettore che conterrà i nuovi quattro valori di R^2 massimo trovati.
                                  % Sta volta saranno relativi ad un
                                  % modello a 2 parametri però.

Idro_2 = zeros(length(M_idro),3); % Matrice definita come prima, ma con una colonna in più.
                                  % Infatti, questa volta ci sarà una
                                  % colonna di 1, una colonna con la
                                  % variabile già assunta (che non deve
                                  % variare nel ciclo interno, ma solo in
                                  % quello esterno, quando varia il
                                  % parametro ecologico e quindi il modello
                                  % che si sta considerando) e la terza
                                  % colonna che varia ad ogni ciclo interno
                                  % e va su tutte le variabili non ancora
                                  % assunte nel modello.

pos_2 = zeros(size(M_eco,2),1); % Vettore delle posizioni delle nuove 4 variabili migliori.

j=1;
cambio_j =[7 14 21 28]; % vettore di indici che serve a dire al ciclo while
                        % il momento in cui si è arrivati (all'interno
                        % della matrice Idro_j, che, come è stato spiegato
                        % prima, è come se fosse l'unione di 4
                        % sottomatrici, una per var. ecologica) al cambio
                        % di sottomatrice. Infatti, la prima matrice sarà
                        % considerata da un indice j che va da 1 a 7, la
                        % seconda da un j che va da 7 a 14, la terza da 14
                        % a 21 e la quarta da 21 a 28. Questo è il motivo
                        % della condizione del ciclo while.

% Ciclo che calcola le nuova variabile migliore (una per N,S,EPTN,EPTS)
for i=1:size(M_eco,2)% ciclo che tiene fisso il parametro ecologico

    while j <= cambio_j(i)% ciclo che va su ogni parametro idro-morf
        % Definizione della matrice per la regressione
         Idro_2(:,1) = ones(length(M_idro),1); %prima colonna di 1
         Idro_2(:,2) = Idro_i(:,i); %variabile già assunta per il parametro i (fissa nel ciclo interno)
         Idro_2(:,3) = Idro_j(:,j); %variabile non ancora assunta (varia ogni volta)
         % Regressione
         b = regress(M_eco(:,i),Idro_2(:,1:3)); % vettore coi coeff di regressione
         y_fit = b(1) + b(2)*Idro_i(:,i) + b(3)*Idro_j(:,j);
         % Calcolo R e R_adj
         y_resid = M_eco(:,i) - y_fit;
         SS_resid = sum(y_resid.^2);
         SS_total = (length(M_eco(:,i))-1) * var(M_eco(:,i));
         R(k) = 1 - SS_resid/SS_total;
         R_adj(k) = 1 - ((SS_resid/SS_total)*((length(M_eco(:,i))-1)/(length(M_eco(:,i))-2)));

         k=k+1;
         j=j+1;
    end

    if i==1
        R_max_2(i) = max(R_adj(1:(length(variabili)-1)));
        pos_2(i) = find(R_adj(1:(length(variabili)-1))==R_max_2(i));
    else
        R_max_2(i) = max(R_adj((((i-1)*(length(variabili)-1))+1):(i*(length(variabili)-1))));
        pos_2(i) = find(R_adj((((i-1)*(length(variabili)-1))+1):(i*(length(variabili)-1)))==R_max_2(i));
    end
end % All'uscita ci serve --> pos_2, con la posizione della nuova var migliore
    %                     --> R_max_2 con i valori massimi

% Riferire le posizioni del nuovo parametro pos_2 alla stringa "variabili"
% iniziale
for i=1:size(M_eco,2)
```

```
        if i~=1
            t=find(Idro_j(1,(((i-1)*(length(variabili)-1))+ pos_2(i)))==Idro(1,:));
            pos_2(i)=t-1;
        else
            t=find(Idro_j(1,pos_2(i))==Idro(1,:));
            pos_2(i)=t-1;
        end
end

% IMPORTANTE : Si è arrivati alla fine del secondo step della regressione
% (ossia quello che ha portato a creare un modello a due variabili). Adesso, gli step che
% seguiranno saranno, di fatto, un ripetersi della parte di script appena
% descritta (quello della regressione a 2 parametri). Quello che cambia
% sono solo alcuni numeri, che indicano la dimensione delle matrici di cui
% si è discusso e dei vettori per salvare i valori di interesse. In
% particolare, ad ogni step successivo viene aggiunta una nuova matrice
% contenente le variabili assunte nel modello al ciclo prima.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% REGRESSIONE A 3 PARAMETRI %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Definizione delle matrici --> Una che contiene le variabili assunte nel
% modello (che variano solo quando varia il parametro ecologico--> Ciclo
% esterno) --> L'altra è l'unione di 4 matrici contenenti una di seguito
% all'altra i parametri non ancora inclusi nel modello.

Idro_i;
Idro_k = zeros(length(M_idro),size(M_eco,2)); % seconda var aggiunta
Idro_j = zeros(length(M_idro),size(M_eco,2)*(length(variabili)-2)); % (il -2 è dovuto al fatto che sono state
assunte già due variabili)
% Quindi, adesso, ci sara una colonna di 1, poi 6 variabili (del primo parametro ecologico), altra colonna
% di 1 ed altre 6 variabili e via così ...

%Definizione Matrice 2a variabile aggiunta
for i=1:size(M_eco,2)
    Idro_k(:,i)=Idro(:,pos_2(i)+1);
end

% Definizione matrice che va su ogni parametro non ancora assunto nel
% modello
n=1;
for i=1:size(M_eco,2) % per ogni parametro ecologico
    j=1;
    m=1;
    while m~=length(variabili)+1
        if variabili(m)==Titoli_idro(j) && variabili(m)~=variabili(pos(i)) && variabili(m)~=variabili(pos_2(i))
            Idro_j(:,n)=M_idro(:,j);
            m=m+1;
            n=n+1;
        else if or(variabili(m)==variabili(pos(i)),variabili(m)==variabili(pos_2(i)))
                m=m+1;
            end
        end

        j=j+1;
    end
end % Risultato --> Nuova matrice Idro_j con 6 parametri (2 tolti)

%%%%%%%%%%%%%%%%%%

k=1; % indice riempimento vettori R,R_adj
R_adj = zeros(size(Idro_j,2),1); % vettore coi R_adj
R = zeros(size(Idro_j,2),1); % vettore coi R
R_max_3 = zeros(size(M_eco,2),1);
Idro_2 = zeros(length(M_idro),4); % 4 colonne sta volta (c'è una variabile in più assunta nel modello)
pos_3 = zeros(size(M_eco,2),1);

j=1;
cambio_j =[6 12 18 24]; % stesso concetto di prima --> Definisce il cambio di sottomatrice (in Idro_j)

% Ciclo che calcola le nuova variabile migliore (una per N,S,EPTN,EPTS)
for i=1:size(M_eco,2)% ciclo che tiene fisso il parametro ecologico

    while j <= cambio_j(i)% ciclo che va su ogni parametro idro-morf
        % Definizione della matrice per la regressione
         Idro_2(:,1) = ones(length(M_idro),1); %prima colonna di 1
         Idro_2(:,2) = Idro_i(:,i); %variabile assunta 1 (fissa)
         Idro_2(:,3) = Idro_k(:,i); %variabile assunta 2 (fissa)
         Idro_2(:,4) = Idro_j(:,j); % variabile da cercare (varia)
         % Regressione
         b = regress(M_eco(:,i),Idro_2(:,1:4)); % vettore coi coeff di regressione
         y_fit = b(1) + b(2)*Idro_i(:,i) + b(3)*Idro_k(:,i) + b(4)*Idro_j(:,j);
         % Calcolo R e R_adj
         y_resid = M_eco(:,i) - y_fit;
         SS_resid = sum(y_resid.^2);
         SS_total = (length(M_eco(:,i))-1) * var(M_eco(:,i));
```

```
        R(k) = 1 - SS_resid/SS_total;
        R_adj(k) = 1 - ((SS_resid/SS_total)*((length(M_eco(:,i))-1)/(length(M_eco(:,i))-2)));

        k=k+1;
        j=j+1;
    end

    if i==1
        R_max_3(i) = max(R_adj(1:(length(variabili)-2)));
        pos_3(i) = find(R_adj(1:(length(variabili)-2))==R_max_3(i));
    else
        R_max_3(i) = max(R_adj((((i-1)*(length(variabili)-2))+1):(i*(length(variabili)-2))));
        pos_3(i) = find(R_adj((((i-1)*(length(variabili)-2))+1):(i*(length(variabili)-2)))==R_max_3(i));
    end
end % All'uscita ci serve --> pos_3, con la posizione della nuova var migliore
    %                       --> R_max_3 con il valori massimi

% Riferire le posizioni del nuovo parametro alla stringa "variabili"
% iniziale

for i=1:size(M_eco,2)
        if i~=1
            t=find(Idro_j(1,(((i-1)*(length(variabili)-2))+ pos_3(i)))==Idro(1,:));
            pos_3(i)=t-1;
        else
            t=find(Idro_j(1,pos_3(i))==Idro(1,:));
            pos_3(i)=t-1;
        end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% REGRESSIONE A 4 PARAMETRI %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Definizione di quattro matrici --> Una che contiene le variabili assunte nel
% modello (che variano solo quando varia il parametro ecologico--> Ciclo
% esterno) --> L'altra è l'unione di 4 matrici contenenti una di seguito
% all'altra i parametri non ancora inclusi nel modello.

Idro_i; % prima var aggiunta
Idro_k; % seconda var aggiunta
Idro_l = zeros(length(M_idro),size(M_eco,2)); % terza var
Idro_j = zeros(length(M_idro),size(M_eco,2)*(length(variabili)-3));

%Definizione Matrice 3a variabile aggiunta
for i=1:size(M_eco,2)
    Idro_l(:,i)=Idro(:,pos_3(i)+1);
end

% Definizione matrice che va su ogni parametro non ancora assunto nel
% modello
n=1;
for i=1:size(M_eco,2) % per ogni parametro ecologico
    j=1;
    m=1;
    while m~=length(variabili)+1
        if variabili(m)==Titoli_idro(j) && variabili(m)~=variabili(pos(i)) && variabili(m)~=variabili(pos_2(i)) &&
variabili(m)~=variabili(pos_3(i))
            Idro_j(:,n)=M_idro(:,j);
            m=m+1;
            n=n+1;
        else if variabili(m)==variabili(pos(i)) | variabili(m)==variabili(pos_2(i)) |
variabili(m)==variabili(pos_3(i))
            m=m+1;
            end
        end

        j=j+1;
    end
end % Risultato --> Nuova matrice Idro_j con 5 parametri (3 tolti)

%%%%%%%%%%%%%%%%%%

k=1; % indice riempimento vettori R,R_adj
R_adj = zeros(size(Idro_j,2),1); % vettore coi R_adj
R = zeros(size(Idro_j,2),1); % vettore coi R
R_max_4 = zeros(size(M_eco,2),1);
Idro_2 = zeros(length(M_idro),5);
pos_4 = zeros(size(M_eco,2),1);

j=1;
cambio_j =[5 10 15 20];

% Ciclo che calcola le nuova variabile migliore (una per N,S,EPTN,EPTS)
for i=1:size(M_eco,2)% ciclo che tiene fisso il parametro ecologico

    while j <= cambio_j(i)% ciclo che va su ogni parametro idro-morf
```

```
        % Definizione della matrice per la regressione
         Idro_2(:,1) = ones(length(M_idro),1); %prima colonna di 1
         Idro_2(:,2) = Idro_i(:,i); %variabile assunta 1 (fissa)
         Idro_2(:,3) = Idro_k(:,i); %variabile assunta 2 (fissa)
         Idro_2(:,4) = Idro_l(:,i); %variabile assunta 3 (fissa)
         Idro_2(:,5) = Idro_j(:,j); % variabile da cercare (varia)
         % Regressione
         b = regress(M_eco(:,i),Idro_2(:,1:5)); % vettore coi coeff di regressione
         y_fit = b(1) + b(2)*Idro_i(:,i) + b(3)*Idro_k(:,i) + b(4)*Idro_l(:,i) + b(5)*Idro_j(:,j);
         % Calcolo R e R_adj
         y_resid = M_eco(:,i) - y_fit;
         SS_resid = sum(y_resid.^2);
         SS_total = (length(M_eco(:,i))-1) * var(M_eco(:,i));
         R(k) = 1 - SS_resid/SS_total;
         R_adj(k) = 1 - ((SS_resid/SS_total)*((length(M_eco(:,i))-1)/(length(M_eco(:,i))-2)));

         k=k+1;
         j=j+1;
    end

    if i==1
        R_max_4(i) = max(R_adj(1:(length(variabili)-3)));
        pos_4(i) = find(R_adj(1:(length(variabili)-3))==R_max_4(i));
    else
        R_max_4(i) = max(R_adj((((i-1)*(length(variabili)-3))+1):(i*(length(variabili)-3))));
        pos_4(i) = find(R_adj((((i-1)*(length(variabili)-3))+1):(i*(length(variabili)-3)))==R_max_4(i));
    end
end % All'uscita ci serve --> pos_4, con la posizione della nuova var migliore
    %                      --> R_max_4 con il valori massimi

% Riferire le posizioni del nuovo parametro alla stringa "variabili"
% iniziale

for i=1:size(M_eco,2)
        if i~=1
            t=find(Idro_j(1,(((i-1)*(length(variabili)-3))+ pos_4(i)))==Idro(1,:));
            pos_4(i)=t-1;
        else
            t=find(Idro_j(1,pos_4(i))==Idro(1,:));
            pos_4(i)=t-1;
        end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% REGRESSIONE A 5 PARAMETRI %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Definizione di 5 matrici --> Una che contiene le variabili assunte nel
% modello (che variano solo quando varia il parametro ecologico--> Ciclo
% esterno) --> L'altra è l'unione di 4 matrici contenenti una di seguito
% all'altra i parametri non ancora inclusi nel modello.

Idro_i; % prima var aggiunta
Idro_k; % seconda var aggiunta
Idro_l; % terza var aggiunta
Idro_m = zeros(length(M_idro),size(M_eco,2)); % quarta var aggiunta
Idro_j = zeros(length(M_idro),size(M_eco,2)*(length(variabili)-4));

%Definizione Matrice 4a variabile aggiunta
for i=1:size(M_eco,2)
    Idro_m(:,i)=Idro(:,pos_4(i)+1);
end

% Definizione matrice che va su ogni parametro non ancora assunto nel
% modello
n=1;
for i=1:size(M_eco,2) % per ogni parametro ecologico
    j=1;
    m=1;
    while m~=length(variabili)+1
        if variabili(m)==Titoli_idro(j) && variabili(m)~=variabili(pos(i)) && variabili(m)~=variabili(pos_2(i)) &&
variabili(m)~=variabili(pos_3(i)) && variabili(m)~=variabili(pos_4(i))
            Idro_j(:,n)=M_idro(:,j);
            m=m+1;
            n=n+1;
        else if variabili(m)==variabili(pos(i)) | variabili(m)==variabili(pos_2(i)) |
variabili(m)==variabili(pos_3(i)) | variabili(m)==variabili(pos_4(i))
            m=m+1;
            end
        end

        j=j+1;
    end
end % Risultato --> Nuova matrice Idro_j con 4 parametri (4 tolti)

%%%%%%%%%%%%%%%%%%
```

```
k=1; % indice riempimento vettori R,R_adj
R_adj = zeros(size(Idro_j,2),1); % vettore coi R_adj
R = zeros(size(Idro_j,2),1); % vettore coi R
R_max_5 = zeros(size(M_eco,2),1);
Idro_2 = zeros(length(M_idro),6);
pos_5 = zeros(size(M_eco,2),1);

j=1;
cambio_j =[4 8 12 16];

% Ciclo che calcola le nuova variabile migliore (una per N,S,EPTN,EPTS)
for i=1:size(M_eco,2)% ciclo che tiene fisso il parametro ecologico

    while j <= cambio_j(i)% ciclo che va su ogni parametro idro-morf
        % Definizione della matrice per la regressione
         Idro_2(:,1) = ones(length(M_idro),1); %prima colonna di 1
         Idro_2(:,2) = Idro_i(:,i); %variabile assunta 1 (fissa)
         Idro_2(:,3) = Idro_k(:,i); %variabile assunta 2 (fissa)
         Idro_2(:,4) = Idro_l(:,i); %variabile assunta 3 (fissa)
         Idro_2(:,5) = Idro_m(:,i); %variabile assunta 4 (fissa)
         Idro_2(:,6) = Idro_j(:,j); % variabile da cercare (varia)
         % Regressione
         b = regress(M_eco(:,i),Idro_2(:,1:6)); % vettore coi coeff di regressione
         y_fit = b(1) + b(2)*Idro_i(:,i) + b(3)*Idro_k(:,i) + b(4)*Idro_l(:,i) + b(5)*Idro_m(:,i) +
b(6)*Idro_j(:,j);
         % Calcolo R e R_adj
         y_resid = M_eco(:,i) - y_fit;
         SS_resid = sum(y_resid.^2);
         SS_total = (length(M_eco(:,i))-1) * var(M_eco(:,i));
         R(k) = 1 - SS_resid/SS_total;
         R_adj(k) = 1 - ((SS_resid/SS_total)*((length(M_eco(:,i))-1)/(length(M_eco(:,i))-2)));

         k=k+1;
         j=j+1;
    end

    if i==1
        R_max_5(i) = max(R_adj(1:(length(variabili)-4)));
        pos_5(i) = find(R_adj(1:(length(variabili)-4))==R_max_5(i));
    else
        R_max_5(i) = max(R_adj((((i-1)*(length(variabili)-4))+1):(i*(length(variabili)-4))));
        pos_5(i) = find(R_adj((((i-1)*(length(variabili)-4))+1):(i*(length(variabili)-4)))==R_max_5(i));
    end
end % All'uscita ci serve --> pos_5, con la posizione della nuova var migliore
    %                     --> R_max_5 con il valori massimi

% Riferire le posizioni del nuovo parametro alla stringa "variabili"
% iniziale

for i=1:size(M_eco,2)
        if i~=1
            t=find(Idro_j(1,(((i-1)*(length(variabili)-4))+ pos_5(i)))==Idro(1,:));
            pos_5(i)=t-1;
        else
            t=find(Idro_j(1,pos_5(i))==Idro(1,:));
            pos_5(i)=t-1;
        end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% REGRESSIONE A 6 PARAMETRI %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Definizione di 6 matrici --> Una che contiene le variabili assunte nel
% modello (che variano solo quando varia il parametro ecologico--> Ciclo
% esterno) --> L'altra è l'unione di 4 matrici contenenti una di seguito
% all'altra i parametri non ancora inclusi nel modello.

Idro_i; % prima var aggiunta
Idro_k; % seconda var aggiunta
Idro_l; % terza var aggiunta
Idro_m; % quarta var aggiunta
Idro_n = zeros(length(M_idro),size(M_eco,2)); % quinta var aggiunta
Idro_j = zeros(length(M_idro),size(M_eco,2)*(length(variabili)-5));

%Definizione Matrice 5a variabile aggiunta
for i=1:size(M_eco,2)
    Idro_n(:,i)=Idro(:,pos_5(i)+1);
end

% Definizione matrice che va su ogni parametro non ancora assunto nel
% modello
n=1;
for i=1:size(M_eco,2) % per ogni parametro ecologico
    j=1;
    m=1;
    while m~=length(variabili)+1
```

```
        if variabili(m)==Titoli_idro(j) && variabili(m)~=variabili(pos(i)) && variabili(m)~=variabili(pos_2(i)) &&
variabili(m)~=variabili(pos_3(i)) && variabili(m)~=variabili(pos_4(i)) && variabili(m)~=variabili(pos_5(i))
            Idro_j(:,n)=M_idro(:,j);
            m=m+1;
            n=n+1;
        else if variabili(m)==variabili(pos(i)) | variabili(m)==variabili(pos_2(i)) |
variabili(m)==variabili(pos_3(i)) | variabili(m)==variabili(pos_4(i)) | variabili(m)==variabili(pos_5(i))
            m=m+1;
            end
        end

        j=j+1;
    end
end % Risultato --> Nuova matrice Idro_j con 3 parametri (5 tolti)

%%%%%%%%%%%%%%%%%%

k=1; % indice riempimento vettori R,R_adj
R_adj = zeros(size(Idro_j,2),1); % vettore coi R_adj
R = zeros(size(Idro_j,2),1); % vettore coi R
R_max_6 = zeros(size(M_eco,2),1);
Idro_2 = zeros(length(M_idro),7);
pos_6 = zeros(size(M_eco,2),1);

j=1;
cambio_j =[3 6 9 12];

% Ciclo che calcola le nuova variabile migliore (una per N,S,EPTN,EPTS)
for i=1:size(M_eco,2)% ciclo che tiene fisso il parametro ecologico

    while j <= cambio_j(i)% ciclo che va su ogni parametro idro-morf
        % Definizione della matrice per la regressione
         Idro_2(:,1) = ones(length(M_idro),1); %prima colonna di 1
         Idro_2(:,2) = Idro_i(:,i); %variabile assunta 1 (fissa)
         Idro_2(:,3) = Idro_k(:,i); %variabile assunta 2 (fissa)
         Idro_2(:,4) = Idro_l(:,i); %variabile assunta 3 (fissa)
         Idro_2(:,5) = Idro_m(:,i); %variabile assunta 4 (fissa)
         Idro_2(:,6) = Idro_n(:,i); %variabile assunta 5 (fissa)
         Idro_2(:,7) = Idro_j(:,j); % variabile da cercare (varia)
         % Regressione
         b = regress(M_eco(:,i),Idro_2(:,1:7)); % vettore coi coeff di regressione
         y_fit = b(1) + b(2)*Idro_i(:,i) + b(3)*Idro_k(:,i) + b(4)*Idro_l(:,i) + b(5)*Idro_m(:,i) +
b(6)*Idro_n(:,i) + b(7)*Idro_j(:,j);
         % Calcolo R e R_adj
         y_resid = M_eco(:,i) - y_fit;
         SS_resid = sum(y_resid.^2);
         SS_total = (length(M_eco(:,i))-1) * var(M_eco(:,i));
         R(k) = 1 - SS_resid/SS_total;
         R_adj(k) = 1 - ((SS_resid/SS_total)*((length(M_eco(:,i))-1)/(length(M_eco(:,i))-2)));

         k=k+1;
         j=j+1;
    end

    if i==1
        R_max_6(i) = max(R_adj(1:(length(variabili)-5)));
        pos_6(i) = find(R_adj(1:(length(variabili)-5))==R_max_6(i));
    else
        R_max_6(i) = max(R_adj((((i-1)*(length(variabili)-5))+1):(i*(length(variabili)-5))));
        pos_6(i) = find(R_adj((((i-1)*(length(variabili)-5))+1):(i*(length(variabili)-5)))==R_max_6(i));
    end

end % All'uscita ci serve --> pos_6, con la posizione della nuova var migliore
    %                     --> R_max_6 con il valori massimi

% Riferire le posizioni del nuovo parametro alla stringa "variabili"
% iniziale

for i=1:size(M_eco,2)
        if i~=1
            t=find(Idro_j(1,(((i-1)*(length(variabili)-5))+ pos_6(i)))==Idro(1,:));
            pos_6(i)=t-1;
        else
            t=find(Idro_j(1,pos_6(i))==Idro(1,:));
            pos_6(i)=t-1;
        end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% REGRESSIONE A 7 PARAMETRI %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Definizione di 7 matrici --> Una che contiene le variabili assunte nel
% modello (che variano solo quando varia il parametro ecologico--> Ciclo
% esterno) --> L'altra è l'unione di 4 matrici contenenti una di seguito
% all'altra i parametri non ancora inclusi nel modello.
```

```
Idro_i; % prima var aggiunta
Idro_k; % seconda var aggiunta
Idro_l; % terza var aggiunta
Idro_m; % quarta var aggiunta
Idro_n; % quinta var aggiunta
Idro_o = zeros(length(M_idro),size(M_eco,2)); % sesta var aggiunta
Idro_j = zeros(length(M_idro),size(M_eco,2)*(length(variabili)-6));

%Definizione Matrice 6a variabile aggiunta
for i=1:size(M_eco,2)
    Idro_o(:,i)=Idro(:,pos_6(i)+1);
end

% Definizione matrice che va su ogni parametro non ancora assunto nel
% modello
n=1;
for i=1:size(M_eco,2) % per ogni parametro ecologico
    j=1;
    m=1;
    while m~=length(variabili)+1
        if variabili(m)==Titoli_idro(j) && variabili(m)~=variabili(pos(i)) && variabili(m)~=variabili(pos_2(i)) &&
variabili(m)~=variabili(pos_3(i)) && variabili(m)~=variabili(pos_4(i)) && variabili(m)~=variabili(pos_5(i)) &&
variabili(m)~=variabili(pos_6(i))
            Idro_j(:,n)=M_idro(:,j);
            m=m+1;
            n=n+1;
        else if variabili(m)==variabili(pos(i)) | variabili(m)==variabili(pos_2(i)) |
variabili(m)==variabili(pos_3(i)) | variabili(m)==variabili(pos_4(i)) | variabili(m)==variabili(pos_5(i)) |
variabili(m)==variabili(pos_6(i))
            m=m+1;
            end
        end

        j=j+1;
    end
end % Risultato --> Nuova matrice Idro_j con 2 parametri (6 tolti)

%%%%%%%%%%%%%%%%%%

k=1; % indice riempimento vettori R,R_adj
R_adj = zeros(size(Idro_j,2),1); % vettore coi R_adj
R = zeros(size(Idro_j,2),1); % vettore coi R
R_max_7 = zeros(size(M_eco,2),1);
Idro_2 = zeros(length(M_idro),8);
pos_7 = zeros(size(M_eco,2),1);

j=1;
cambio_j =[2 4 6 8];

% Ciclo che calcola le nuova variabile migliore (una per N,S,EPTN,EPTS)
for i=1:size(M_eco,2)% ciclo che tiene fisso il parametro ecologico

    while j <= cambio_j(i)% ciclo che va su ogni parametro idro-morf
        % Definizione della matrice per la regressione
         Idro_2(:,1) = ones(length(M_idro),1); %prima colonna di 1
         Idro_2(:,2) = Idro_i(:,i); %variabile assunta 1 (fissa)
         Idro_2(:,3) = Idro_k(:,i); %variabile assunta 2 (fissa)
         Idro_2(:,4) = Idro_l(:,i); %variabile assunta 3 (fissa)
         Idro_2(:,5) = Idro_m(:,i); %variabile assunta 4 (fissa)
         Idro_2(:,6) = Idro_n(:,i); %variabile assunta 5 (fissa)
         Idro_2(:,7) = Idro_o(:,i); %variabile assunta 6 (fissa)
         Idro_2(:,8) = Idro_j(:,j); % variabile da cercare (varia)
         % Regressione
         b = regress(M_eco(:,i),Idro_2(:,1:8)); % vettore coi coeff di regressione
         y_fit = b(1) + b(2)*Idro_i(:,i) + b(3)*Idro_k(:,i) + b(4)*Idro_l(:,i) + b(5)*Idro_m(:,i) +
b(6)*Idro_n(:,i) + b(7)*Idro_o(:,i) + b(8)*Idro_j(:,j);
         % Calcolo R e R_adj
         y_resid = M_eco(:,i) - y_fit;
         SS_resid = sum(y_resid.^2);
         SS_total = (length(M_eco(:,i))-1) * var(M_eco(:,i));
         R(k) = 1 - SS_resid/SS_total;
         R_adj(k) = 1 - ((SS_resid/SS_total)*((length(M_eco(:,i))-1)/(length(M_eco(:,i))-2)));

         k=k+1;
         j=j+1;
    end

    if i==1
        R_max_7(i) = max(R_adj(1:(length(variabili)-6)));
        pos_7(i) = find(R_adj(1:(length(variabili)-6))==R_max_7(i));
    else
        R_max_7(i) = max(R_adj((((i-1)*(length(variabili)-6))+1):(i*(length(variabili)-6))));
        pos_7(i) = find(R_adj((((i-1)*(length(variabili)-6))+1):(i*(length(variabili)-6)))==R_max_7(i));
    end

end % All'uscita ci serve --> pos_7, con la posizione della nuova var migliore
    %                     --> R_max_7 con i valori massimi

% Riferire le posizioni del nuovo parametro alla stringa "variabili"
```

```
% iniziale

for i=1:size(M_eco,2)
        if i~=1
            t=find(Idro_j(1,(((i-1)*(length(variabili)-6))+ pos_7(i)))==Idro(1,:));
            pos_7(i)=t-1;
        else
            t=find(Idro_j(1,pos_7(i))==Idro(1,:));
            pos_7(i)=t-1;
        end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% REGRESSIONE A 8 PARAMETRI %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Definizione di 8 matrici --> Una che contiene le variabili assunte nel
% modello (che variano solo quando varia il parametro ecologico--> Ciclo
% esterno) --> L'altra è l'unione di 4 matrici contenenti una di seguito
% all'altra i parametri non ancora inclusi nel modello.

Idro_i; % prima var aggiunta
Idro_k; % seconda var aggiunta
Idro_l; % terza var aggiunta
Idro_m; % quarta var aggiunta
Idro_n; % quinta var aggiunta
Idro_o; % sesta var aggiunta
Idro_p = zeros(length(M_idro),size(M_eco,2)); % settima var aggiunta
Idro_j = zeros(length(M_idro),size(M_eco,2)*(length(variabili)-7));

%Definizione Matrice 7a variabile aggiunta
for i=1:size(M_eco,2)
    Idro_p(:,i)=Idro(:,pos_7(i)+1);
end

% Definizione matrice che va su ogni parametro non ancora assunto nel
% modello
n=1;
for i=1:size(M_eco,2) % per ogni parametro ecologico
    j=1;
    m=1;
    while m~=length(variabili)+1
        if variabili(m)==Titoli_idro(j) && variabili(m)~=variabili(pos(i)) && variabili(m)~=variabili(pos_2(i)) &&
variabili(m)~=variabili(pos_3(i)) && variabili(m)~=variabili(pos_4(i)) && variabili(m)~=variabili(pos_5(i)) &&
variabili(m)~=variabili(pos_6(i)) && variabili(m)~=variabili(pos_7(i))
            Idro_j(:,n)=M_idro(:,j);
            m=m+1;
            n=n+1;
        else if variabili(m)==variabili(pos(i)) | variabili(m)==variabili(pos_2(i)) |
variabili(m)==variabili(pos_3(i)) | variabili(m)==variabili(pos_4(i)) | variabili(m)==variabili(pos_5(i)) |
variabili(m)==variabili(pos_6(i)) | variabili(m)==variabili(pos_7(i))
            m=m+1;
            end
        end

        j=j+1;
    end
end % Risultato --> Nuova matrice Idro_j con 1 parametri (7 tolti)

%%%%%%%%%%%%%%%%%%

k=1; % indice riempimento vettori R,R_adj
R_adj = zeros(size(Idro_j,2),1); % vettore coi R_adj
R = zeros(size(Idro_j,2),1); % vettore coi R
R_max_8 = zeros(size(M_eco,2),1);
Idro_2 = zeros(length(M_idro),9);
pos_8 = zeros(size(M_eco,2),1);

j=1;
cambio_j =[1 2 3 4];

b_finale = zeros(size(Idro,2)*size(M_eco,2),1); % Definizione vettore finale coi coefficienti della regressione
completa
u=1; % indice per riempimento in sequenza di b_finale

% Ciclo che calcola le nuova variabile migliore (una per N,S,EPTN,EPTS)
for i=1:size(M_eco,2)% ciclo che tiene fisso il parametro ecologico

    while j <= cambio_j(i)% ciclo che va su ogni parametro idro-morf
        % Definizione della matrice per la regressione
         Idro_2(:,1) = ones(length(M_idro),1); %prima colonna di 1
         Idro_2(:,2) = Idro_i(:,i); %variabile assunta 1 (fissa)
         Idro_2(:,3) = Idro_k(:,i); %variabile assunta 2 (fissa)
         Idro_2(:,4) = Idro_l(:,i); %variabile assunta 3 (fissa)
         Idro_2(:,5) = Idro_m(:,i); %variabile assunta 4 (fissa)
         Idro_2(:,6) = Idro_n(:,i); %variabile assunta 5 (fissa)
         Idro_2(:,7) = Idro_o(:,i); %variabile assunta 6 (fissa)
```

```
        Idro_2(:,8) = Idro_p(:,i); %variabile assunta 7 (fissa)
        Idro_2(:,9) = Idro_j(:,j); % variabile da cercare (varia)
        % Regressione
        b = regress(M_eco(:,i),Idro_2(:,1:9)); % vettore coi coeff di regressione
        y_fit = b(1) + b(2)*Idro_i(:,i) + b(3)*Idro_k(:,i) + b(4)*Idro_l(:,i) + b(5)*Idro_m(:,i) +
b(6)*Idro_n(:,i) + b(7)*Idro_o(:,i) + b(8)*Idro_p(:,i) + b(9)*Idro_j(:,j);
        % Calcolo R e R_adj
        y_resid = M_eco(:,i) - y_fit;
        SS_resid = sum(y_resid.^2);
        SS_total = (length(M_eco(:,i))-1) * var(M_eco(:,i));
        R(k) = 1 - SS_resid/SS_total;
        R_adj(k) = 1 - ((SS_resid/SS_total)*((length(M_eco(:,i))-1)/(length(M_eco(:,i))-2)));

        k=k+1;
        j=j+1;
    end

    if i==1
        R_max_8(i) = max(R_adj(1:(length(variabili)-7)));
        pos_8(i) = find(R_adj(1:(length(variabili)-7))==R_max_8(i));
    else
        R_max_8(i) = max(R_adj((((i-1)*(length(variabili)-7))+1):(i*(length(variabili)-7))));
        pos_8(i) = find(R_adj((((i-1)*(length(variabili)-7))+1):(i*(length(variabili)-7)))==R_max_8(i));
    end

    % Riempimento vettore finale coi coefficienti di regressione
    for x=1:size(Idro,2)% ciclo che salva un coeff alla volta (b(1),b(2),..)
        b_finale(u)=b(x);
        u=u+1;
    end

end % All'uscita ci serve --> pos_8, con la posizione della nuova var migliore
    %                     --> R_max_8 con i valori massimi
    %                     --> Vettore coi valori finali dei coefficiente di
    %                     regressione (vettore b).

% Riferire le posizioni del nuovo parametro alla stringa "variabili"
% iniziale

for i=1:size(M_eco,2)
        if i~=1
            t=find(Idro_j(1,(((i-1)*(length(variabili)-7))+ pos_8(i)))==Idro(1,:));
            pos_8(i)=t-1;
        else
            t=find(Idro_j(1,pos_8(i))==Idro(1,:));
            pos_8(i)=t-1;
        end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%% GRAFICO FINALE R_adj %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

R_MAX = cat(1,R_max,R_max_2,R_max_3,R_max_4,R_max_5,R_max_6,R_max_7,R_max_8);

h=[1:length(variabili)]; % vettore col numero di variabili
R_f=zeros(length(variabili),1);

% Creazione vettore R_f con, in sequenza, gli 8 valori di R_adj per il
% parametro N, poi quelli del parametro S e via così..
for i=1:size(M_eco,2)
    x=0;
    j=1;

    while x < length(R_MAX)
        R_f(j) = R_MAX(i+x);
        x=x+4;
        j=j+1;
    end

    % Plot figure (4 totali, una per ciclo)
    figure(1)
    subplot(2,2,i)
    plot(h,R_f,'r*','MarkerSize',20);
    xlabel('Numero Variabili');
    ylabel('Coefficiente R_{adj}^2');
    title(['Regressione Lineare Multivariata ' Titoli_eco(i)]);


end

 % Salvare figura in automatico
    set(gcf, 'PaperPositionMode', 'manual');
    set(gcf, 'PaperUnits', 'centimeters');
    set(gcf, 'PaperOrientation', 'portrait');
    set(gcf, 'PaperPosition', [0 0 35 25]);%formato del foglio 35x25 cm
    print(gcf,['Regressione_Multivariata_8_par'],'-dpng');

pos_finale = cat(1,pos,pos_2,pos_3,pos_4,pos_5,pos_6,pos_7,pos_8);
```

```
% Esportazione valori coefficienti b di regressione e associazione alla
% variabile giusta.
writematrix(variabili','Database_Matlab_Multivariata.xlsx','Sheet',3,'Range','B2:B500');
writematrix(b_finale,'Database_Matlab_Multivariata.xlsx','Sheet',3,'Range','L2:L500');
writematrix(pos_finale,'Database_Matlab_Multivariata.xlsx','Sheet',3,'Range','G2:G500');
writematrix(R_MAX,'Database_Matlab_Multivariata.xlsx','Sheet',3,'Range','H2:H500');
```

ALLEGATO 6

Script Matlab per l'assegnazione delle stazioni al reticolo idrografico

```
% IMPORT DATI

% STRUTTURA DATI DA IMPORTARE: Due tabelle in excel. La prima (stazioni)
% è a 2 colonne (la prima con la LON, la seconda con la LAT --> coordinate delle stazioni di campionamento). La
seconda (reticolo) ha la stessa struttura ma contiene tutti i punti del reticolo.

stazioni = readmatrix('Stazioni_Tanaro.xlsx','Sheet',1,'Range','G1:H31'); % Vettore (a due colonne) con LON,LAT
delle stazioni
reticolo = readmatrix('Reticolo_Punti_Tanaro.xlsx','Sheet',1,'Range','B1:C847'); % Vettore con coordinate LON,LAT
del reticolo
R = 6372.8; % Raggio quadratico medio della Terra

% Trasformazioni in radianti per il calcolo delle distanze
stazioni = stazioni*(pi/180);
reticolo = reticolo*(pi/180);

for i=1:size(stazioni,1) % ciclo che va su tutte le stazioni

    % Inizializzazione della distanza d1 al primo valore di distanza
    d1 = R*acos( sin(stazioni(i,2))*sin(reticolo(1,2)) + cos(stazioni(i,2))*cos(reticolo(1,2))*cos(stazioni(i,1)-
reticolo(1,1)) );

    for j=1:size(reticolo,1) % ciclo che calcola tutte le distanze tra stazione corrente e reticolo e salva la
posizione della minima
      d2 = R*acos( sin(stazioni(i,2))*sin(reticolo(j,2)) +
cos(stazioni(i,2))*cos(reticolo(j,2))*cos(stazioni(i,1)-reticolo(j,1)) );
      if d2 <= d1
          pos=j; % Salva la posizione della nuova distanza minore trovata.
          d1=d2; % Mette ,come nuova d1, la d2 (che è quella minore, appena trovata).
                 % In questo modo, si arriva alla fine del ciclo con la posizione
                 % del punto del reticolo che è più vicino alla stazione.
      end
    end

    % Creazione matrice con le nuove coordinate delle stazioni (sul
    % reticolo)
    stazioni_riposizionate(i,1) = reticolo(pos,1)*(180/pi); % riconversione in gradi
    stazioni_riposizionate(i,2) = reticolo(pos,2)*(180/pi); % riconversione in gradi
end

% Esportazione in excel delle nuove posizioni delle stazioni
writematrix(stazioni_riposizionate,'Stazioni_Riposizionate.xlsx','Sheet',1,'Range','A2:B500');
```

ALLEGATO 7

Script Matlab per il calcolo delle distanze lungo il reticolo fluviale

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%% DISTANZE LUNGO IL RETICOLO FLUVIALE %%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%% DISTANZE FLOW UNCONNECTED %%%%%%%%%%%%%%%%%%%%%%%%%%%

% IMPORT DATI

% STRUTTURA DATI DA IMPORTARE: Due tabelle in excel. La prima (stazioni)
% è a 3 colonne (la prima con gli ID delle stazioni, seconda e terza con
% longitudine e latitudine) --> Si riferisce alle stazioni riposizionate
% sul reticolo. La seconda (coordinate_M) ha la stessa struttura ma
% contiene tutti i punti del reticolo.

stazioni = readmatrix('Stazioni_Riposizionate.xlsx','Sheet',1,'Range','A1:C30'); % Vettore (a 3 colonne) con
ID,LON,LAT delle stazioni
coordinate_M = readmatrix('Reticolo_Punti_Tanaro.xlsx','Sheet',1,'Range','A1:C847'); % Vettore con  ID,LON,LAT del
reticolo
R = 6372.8; % Raggio quadratico medio della Terra
lato_pixel=0.0998; % Distanza minima tra due punti consecutivi del reticolo (pari al lato del pixel del raster
originale)
x=1; % indice per il riempimento progressivo della matrice finale con le distanze complete tra le coppie di
stazioni

% Trasformazioni in radianti per il calcolo delle distanze (LON e LAT
% traformate da gradi a radianti)
stazioni(:,2) = stazioni(:,2)*(pi/180);
stazioni(:,3) = stazioni(:,3)*(pi/180);
coordinate_M(:,2) = coordinate_M(:,2)*(pi/180);
coordinate_M(:,3) = coordinate_M(:,3)*(pi/180);

% IDEA DI FUNZIONAMENTO DEL SUCCESSIVO CICLO PER IL CALCOLO DELLE DISTANZE:
% Si hanno a disposizione gli 860 punti che compongono il reticolo fluviale
% e le 30 stazioni presenti lungo esso. I punti del reticolo corrispondono
% ai centroidi dei pixel che componevano il raster originario del reticolo
% (RENERFOR). Dalla conversione in layer di punti, si è ottenuta una
% tabella con ID di ogni punto (numero progressivo da 1 a 846) e coordinate
% LON,LAT. Tali punti non sono ordinati in nessun modo, dunque per il
% software sono punti "a caso" posizionati nello spazio.
% L'idea di del successivo doppio ciclo for (con numerosi cicli interni
% concatenati) è di procedere all'individuazione, di volta in volta, delle
% coordinate del punto più vicino (in particolare si ragionerà in termini
% di ID del punto, per avere un valore solo) e salvare gli ID individuati
% in una struttura. Siccome però sono possibili più percorsi (infatti, non
% si conosce a priori la direzione giusta --> i punti è come se fossero
% disposti a caso per il software), bisogna individuare tutti quelli
% possibili e poi alla fine scegliere quello giusto (che sarà quello che
% conterrà l'ID della stazione di arrivo come ultimo punto). Alla fine di
% ogni giro del ciclo, si ottiene un vettore con tutti gli ID dei punti
% (consecutivi) che compongono il percorso che collega la stazione di
% partenza con quella di arrivo. Per ottenere la distanza tra le due
% stazioni non si deve far altro che la somma di ogni tratto presente tra
% due punti consecutivi del vettore.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%% CICLO PER IL CALCOLO DELLA DISTANZE %%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

for i=1:size(stazioni,1) % Doppio ciclo for (indici i e j) che va su tutte le stazioni importate
    for j=1:size(stazioni,1)

        if stazioni(i,1)~=stazioni(j,1) % procedo solo se staz. partenza è diversa da quella di arrivo

        indice_partenza=stazioni(i,1); % indice stazione di partenza
        index_arrivo=stazioni(j,1);    % indice stazione di arrivo

        x_A=coordinate_M(find(coordinate_M(:,1)==indice_partenza),2); % Trova la riga di coordinate_M in cui c'è
l'ind_partenza --> LON
        y_A=coordinate_M(find(coordinate_M(:,1)==indice_partenza),3); % Trova la riga di coordinate_M in cui c'è
l'ind_partenza --> LAT

        distanza_M = distanza(x_A,y_A,coordinate_M); % funzione che calcola le distanze tra punto di partenza e
tutto il reticolo

        index_a=find((distanza_M(:,2)<(2^0.5)*lato_pixel*1.1).*(distanza_M(:,2)>0.1*lato_pixel)); % prodotto
elemento per elemento
                                                                                    % --> Trova le 0 < distanze
<= diagonale_pixel
```

```
        ID_vicini_a{1}=distanza_M(index_a(1),1); % struttura con gli ID dei punti vicini a quello di partenza.
                                                 % In generale, trova da 2 a tre punti (nel caso di confluenza).

        try % Lo fa solo se c'è un secondo valore minore della diagonale del pixel
        ID_vicini_a{2}=distanza_M(index_a(2),1);
        catch

        end

        indice_precedente_a=[indice_partenza indice_partenza]; % inizializzo l'indice precedente a quello di
partenza
        check_arrivo=0; % valore binario che dice al ciclo se si è giunti alla stazione di arrivo (e quindi
bisogna fermarsi)
        numero_percorsi = max(length(index_a)); % il numero di percorsi possibili è uguale al numero di distanze
trovate < diagonale_pixel
        check_convergenze=0; % valore binario che dice al ciclo se si è giunti ad una confluenza

        while check_convergenze==0 % ci dice che il giro deve essere fatto per tutti i nuovi percorsi creati dalla
convergenza (se c'è)
            check_convergenze=1; % se non trova convergenze, si esce dal ciclo while
            while check_arrivo==0 % entra solo se non sono arrivato alla stazione di arrivo al giro prima

                for cont_ID_vicini=1:numero_percorsi % ciclo che va sulle direzioni trovate (dal punto di
partenza) --> 1 o 2
                    indice_partenza=ID_vicini_a{cont_ID_vicini}(end); % diventa quello del nuovo punto vicino
trovato
                    x_A=coordinate_M(find(coordinate_M(:,1)==indice_partenza),2); % coordinata x del nuovo punto
partenza
                    y_A=coordinate_M(find(coordinate_M(:,1)==indice_partenza),3); % coordinata y del nuovo punto
partenza
                    distanza_M=distanza(x_A,y_A,coordinate_M); % distanza con tutto il reticolo
                    index_a=find((distanza_M(:,2)<(2^0.5)*lato_pixel*1.1).*(distanza_M(:,2)>0)); % posizione 0 <
distanze <= diagonale_pixel
                    index_a=distanza_M(index_a,1); % vettore con due valori di indice di distanza minima
                    index_a=index_a(find(index_a~=indice_precedente_a(cont_ID_vicini))); % filtro: tiene solo
l'indice nuovo (diverso precedente)
                    indici_trovati=[]; % vettore vuoto inizializzato

                    for k=1:numero_percorsi % popolare vettore con i valori indice trovati
                       indici_trovati=[indici_trovati ID_vicini_a{k}]; % al primo giro mette solo quelli vicino al
punto di part iniziale
                    end

                    if isempty(index_a)==1 % se non ha trovato nuovi punti vicini --> Non fa nulla

                    elseif length(index_a)>1 % se ne ha trovati più di uno = è una CONFLUENZA
                        index_a;
                        indici_trovati;
                        index_a=index_a(find(min((index_a~=indici_trovati)')>0)); % trova quando i nuovi indici
sono != da quelli già trovati al giro prima
                    end

                    if isempty(index_a)==1 % se non ci sono punti vicini, non fa nulla

                    elseif length(index_a)>1 % se è una CONFLUENZA
                        check_convergenze=0; % ciclo while esterno deve continuare --> Altro giro
                        numero_percorsi=numero_percorsi+1; % percorso in più dato dallo sdoppiamento del ramo
                        salva=ID_vicini_a{cont_ID_vicini}; % salvo gli ID vicini trovati nel giro precedente
                        ID_vicini_a{cont_ID_vicini}=[salva index_a(1)]; % aggiorno con i valori trovati al giro
corrente (tenendo i precedenti)
                        indice_precedente_a(cont_ID_vicini)=indice_partenza; % nel nuovo ciclo, l'indice
precedente sarà quello di partenza del vecchio ciclo
                        ID_vicini_a{numero_percorsi}=[salva index_a(2)]; % inserisco nella struttura anche gli
altri due percorsi
                        indice_precedente_a(numero_percorsi)=indice_partenza; % aggiorno il nuovo indice di
partenza del nuovo percorso
                        break
                    else % se c'è un solo percorso
                        indice_precedente_a(cont_ID_vicini)=indice_partenza;
                        ID_vicini_a{cont_ID_vicini}=[ID_vicini_a{cont_ID_vicini} index_a];
                    end

                    if index_a==index_arrivo % controllo se sono arrivato alla stazione di arrivo
                        check_arrivo=1;
                    end

                    ID_vicini_a{cont_ID_vicini};
                end
            end
        end

        % Usciti dal ciclo per la ricostruzione di tutti i percorsi
        % possibili, bisogna individuare quale sia quello corretto (che
        % collega la stazione(i)=partenza e la stazione(j)=arrivo. Una
        % volta individuato, si fanno le somme di tutti gli elementi del
        % percorso trovato (che sono punti consecutivi sul reticolo) e si
        % ottiene la distanza finale lungo il reticolo fluviale.

        dist_finale=0; % valore inizializzato a zero
```

```
        for cont_n=1:numero_percorsi % ho ottenuto (generalmente) più di un percorso --> col successivo if capisco
qual è quello giusto
            if find(ID_vicini_a{cont_n}==index_arrivo) % il percorso giusto è quello in cui è presente l'indice
della stazione di arrivo
                ID_vicini_a{cont_n};
                percorso_corretto = ID_vicini_a{cont_n}; % salvo il vettore con gli ID dei punti consecutivi del
percorso giusto
                for m=1:(length(percorso_corretto)-1)
                pos_1 = find(percorso_corretto(m)==coordinate_M(:,1));   % posizione nella matrice del reticolo
del punto precedente
                pos_2 = find(percorso_corretto(m+1)==coordinate_M(:,1)); % posizione nella matrice del reticolo
del punto successivo
                LON_A = coordinate_M(pos_1,2); % longitudine del punto precedente
                LAT_A = coordinate_M(pos_1,3); % latitudine del punto precedente
                LON_B = coordinate_M(pos_2,2); % longitudine del punto successivo
                LAT_B = coordinate_M(pos_2,3); % latitudine del punto successivo

                % Calcolo della distanza finale come valore del giro
                % precedente (al primo giro pari a zero) più il nuovo
                % tratto trovato al giro corrente. Lo si fa per tutti i
                % punti presenti nel vettore percorso_corretto

                dist_finale = dist_finale + R*acos( sin(LAT_A)*sin(LAT_B) + cos(LAT_A)*cos(LAT_B)*cos(LON_A-LON_B)
);
                end
                % Aggiungere il tratto di distanza tra stazione di partenza e primo
                % punto del reticolo (che rimaneva esclusa dal ciclo)
                pos_1 = find(stazioni(i,1)==coordinate_M(:,1));
                pos_2 = find(percorso_corretto(1)==coordinate_M(:,1));
                LON_A = coordinate_M(pos_1,2);
                LAT_A = coordinate_M(pos_1,3);
                LON_B = coordinate_M(pos_2,2);
                LAT_B = coordinate_M(pos_2,3);
                dist_finale = dist_finale + R*acos( sin(LAT_A)*sin(LAT_B) + cos(LAT_A)*cos(LAT_B)*cos(LON_A-LON_B)
);
            end
        end

        % Creazione di una matrice a tre colonne:
        % (1) Prima colonna con ID della stazione di partenza
        % (2) Seconda colonna con ID della stazione di arrivo
        % (3) Terza colonna con valore della distanza appena calcolata (km)

        distanze_reticolo(x,1) = stazioni(i,1);
        distanze_reticolo(x,2) = stazioni(j,1);
        distanze_reticolo(x,3) = dist_finale;
        x=x+1;

        end
    end
end

% Esportazione della matrice finale delle distanze in un foglio excel
writematrix(distanze_reticolo,'Tabella_Distanze_Reticolo.xlsx','Sheet',5,'Range','A2:C2000');
```

ALLEGATO 8

Funzione Matlab per il calcolo della distanza tra punti proiettati su una superficie di riferimento

```
% FORMULA PER IL CALCOLO DELLA DISTANZA TRA DUE PUNTI (date LON e LAT):
% distanza (A,B) = R * arccos( sin(latA)*sin(latB) + cos(latA)*cos(latB)*cos(lonA-lonB) )

% Si pasano alla funzione le coordinate del punto di partenza (x_A = LON e y_A = LAT) e la matrice con le
coordinate di tutto il reticolo.

function [distanza_M] = distanza(x_A,y_A,coordinate_M)
distanza_M=zeros(size(coordinate_M,1),2); % righe = quelle della matrice del reticolo
                                          % colonne = 2 --> ID e valore di distanza
R = 6372.8; % Raggio quadratico medio della Terra

    % Ciclo che calcola tutte le distanze tra stazione di partenza e tutti
    % i punti del reticolo fluviale. Si crea la suddetta matrice a due
    % colonne --> distanza_M
    for i=1:size(coordinate_M,1)
        distanza_M(i,1)=coordinate_M(i,1);
        distanza_M(i,2)= R*acos( sin(coordinate_M(i,3))*sin(y_A) +
cos(coordinate_M(i,3))*cos(y_A)*cos(coordinate_M(i,2)-x_A) );
    end
end
```

ALLEGATO 9

Script Matlab per il calcolo dei Semivariogrammi

```
% IMPORT DATI

% STRUTTURA DATI DA IMPORTARE: Creare due tabelle in excel. La prima
% (M_eco) sarà a 5 colonne (la prima con gli ID delle stazioni, le altre 4
% con le variabili ecologiche). La seconda (distanze) sarà a tre colonne
% (prima e seconda colonna con gli ID delle due stazioni, la terza col
% valore di distanza in km).


M_eco = readmatrix('Variogramma_Tanaro.xlsx','Sheet',1,'Range','B1:F79'); % legge i parametri ecologici
% Nomi per TITOLI ASSI
Nomi_eco = readvars('Variogramma_Tanaro.xlsx','Sheet',2,'Range','D1:D4'); % nomi dei parametri ecologici
% Matrice coi nomi della coppia e valore della distanza
distanze = readmatrix('Distanze_Punti_Tanaro.xlsx','Sheet',1,'Range','A1:C2071'); % distanza e nome di ogni coppia
di punti


%%%%%%%%% VARIOGRAMMA %%%%%%%%%%%%

% Definizione delle classi di distanza (aggiornare a seconda delle distanze
% importate) --> Modificare n_classi.

n_classi = 10; % numero di classi
h_max = max(distanze(:,3)); % distanza massima calcolata

F = h_max/n_classi; % intervallo tra le classi

lim_inf(1)=0; % primo valore del vettore coi limiti inferiori delle classi
lim_sup(1)=F; % primo valore del vettore coi limiti superiori delle classi
h(1)=F/2;     % primo valore del vettore per plottare gamma (punti medi di ogni classe)

% Ciclo per creare in automatico il resto delle classi
for i=2:n_classi
    lim_inf(i)=lim_inf(i-1)+F;
    lim_sup(i)=lim_sup(i-1)+F;
    h(i)=h(i-1)+F;
end

gamma_h = zeros(n_classi,1); % inizializzazione del vettore coi valori della funzione gamma(h)



for i=2:size(M_eco,2) % su ogni variabile --> serve per definire la variabile in quel momento
    for j=1:n_classi % su ogni classe (per ogni variabile) --> serve per definire i limiti inf e sup
        sommatoria=0; % inizializzare a zero il valore di sommatoria per il calcolo di gamma(h)
        N_h=0; % inizializzare a zero il numero di distanze da trovare nella classe h di distanza
        for k=1:size(distanze,1) % ciclo per cercare tutte le distanze che ricadono entro una det. classe
            if distanze(k,3)>=lim_inf(j) && distanze(k,3)<=lim_sup(j) % limiti sup e inf della classe dati
dall'indice j
                s_1 = find(distanze(k,1)==M_eco(:,1)); % trova la posizione s_i della prima variabile
                s_2 = find(distanze(k,2)==M_eco(:,1)); % trova la posizione s_i+1 della seconda variabile
                sommatoria = sommatoria + (M_eco(s_1,i) - M_eco(s_2,i))^2;
                N_h = N_h+1;
            end
        end
        gamma_h(j)= sommatoria/(N_h);

    end

    % Metto tutti e 4 i grafici nella stessa figura (aggiunto uno ogni
    % ciclo i)
    subplot(2,2,i-1)
    plot(h,gamma_h,'r-o');
    xlabel('h  [km]');
    ylabel('\gamma(h)');
    title(['Variogramma ',Nomi_eco(i-1)]);
end

 % Salvare figure in automatico
    set(gcf, 'PaperPositionMode', 'manual');
    set(gcf, 'PaperUnits', 'centimeters');
    set(gcf, 'PaperOrientation', 'portrait');
    set(gcf, 'PaperPosition', [0 0 35 25]);%formato del foglio 35x25 cm
    print(gcf,['Variogramma_Tanaro'],'-dpng');
```

Bibliografia

1. de Mendoza G, Kaivosoja R, Groenroos M, et al., 2018. *Highly variable species distribution models in a subarctic stream metacommunity: Patterns, mechanisms and implications*. Freshwater Biol. 2018;63:33–47.

2. Harvey E, Altermatt F, 2019. *Regulation of the functional structure of acquatic communities across spatial scales in a major river network*. Ecology 100(4):e02633. 10.1002/ecy.2633.

3. Goethe E, Baattrup-Pedersen A, Wiberg-Larsen P, Graeber D, Kristensen E, Friberg N, 2017. *Environmental and spatial controls of taxonomic versus trait composition of stream biota*. Freshwater Biology (2017) 62, 397–413.

4. Tonkin JD, Altermatt F, Finn D, et al., 2018. *The role of dispersal in river network metacommunities: Patterns, processes, and pathways*. Freshwater Biol. 2018;63:141–163.

5. Larsen S, Bruno M, Vaughan I, Zolezzi G, 2019. *Testing the River Continuum Concept with geostatistical stream-network*. Ecological Complexity 39 (2019) 100773.

6. Tonkin JD, Heino J, Altermatt F. *Metacommunities in river networks: The importance of network structure and connectivity on patterns and processes*. Freshwater Biol. 2018;63:1–5.

7. Logue J, Mouquet N, Peter H, Hillebrand H, et al., 2011. *Empirical approaches to metacommunities: a review and comparison with theory*. Trends in Ecology and Evolution, September 2011, Vol. 26, No. 9

8. Leibold M A, et al, 2004. *The metacommunity concept: a framework for multi-scale community ecology*. Ecology Letters, (2004) 7: 601–613.

9. Brown B, Swan C, Auerbach D, Campbell Grant E, Hitt N, Maloney K, Patrick C, 2011. *Metacommunity theory as a multispecies, multiscale framework for studying the influence of river network structure on riverine communities and ecosystems*. Journal of the North American Benthological Society, 30(1):310-327.

10. Heino J, Melo A, Siqueira D, Soininen J, Valanko S, Bini L, 2015. *Metacommunity organisation, spatial extent and dispersal in aquatic systems: patterns, processes and prospects*. Freshwater Biology (2015) 60, 845–869.

11. Montgomery D, Peck E, Vining G, 2012. *Introduction to Linear Regression Analysis*. John Wiley & Sons, Inc., Hoboken, New Jersey.

12. Bo T, Doretto A, Laini A, Bona F, Fenoglio S, 2017. *Biomonitoring with macroinvertebrate communities in Italy: what happened to our past and what is the future?* J. Limnol., 2017; 76(s1): 21-28.

13. Ganora D, Gallo E, Laio F, Masoero A, Claps P, 2014. *Analisi Idrologiche e valutazione del potenziale idroelettrico dei bacini piemontesi - Manuale Operativo*. Dipartimento di Ingegneria dell'Ambiente, il Territorio e le Infrastrutture. Politecnico di Torino.

Sitografia

1. https://docs.qgis.org/3.4/en/docs/training_manual/index.html
2. https://grass.osgeo.org/grass76/manuals/libpython/
3. https://grass.osgeo.org/
4. https://it.mathworks.com/
5. https://it.wikipedia.org/wiki/Bacino_idrografico#Sistema_di_ordinamento_Horton-Strahler
6. https://it.wikipedia.org/wiki/Coefficiente_di_determinazione
7. https://en.wikipedia.org/wiki/Pattern_recognition
8. https://en.wikipedia.org/wiki/Community_(ecology)
9. https://it.wikipedia.org/wiki/Predazione
10. https://it.wikipedia.org/wiki/Competizione
11. http://www.arpa.vda.it/images/stories/ARPA/acquesuperficiali/IFF/manualeiff/iffcap02.pdf