



POLITECNICO DI TORINO

MASTER THESIS

---

**A novel approach to model the  
macroinvertebrate distribution in  
braided rivers**

---

*Author:*

Beatrice PINNA

*Supervisor:*

Prof. Paolo VEZZA

Eng. Giovanni NEGRO

Dr. Alex LAINI

Prof. Pierluigi VIAROLI

Dr. Gemma BURGAZZI

Department of Environment, Land and Infrastructure Engineering

A.A. 2019/2020



*"Not just a chemical compound, but a fundamental part of Nature, with aspects that are serene, enchanting, enlivening, profound, spiritual and even terrible. In the voice of the babbling streams, is a music of Humanity."*

*The water book - Alok Jha*

## *Abstract*

Rivers are often subject to hydromorphological pressures due to engineering activities such as derivations, dams or arrangement or defensive works. In order to ensure a proper water resources management and the conservation of both biodiversity and the ecological services provided by rivers, the definition of an appropriate e-flow downstream to the intake structures, plays an essential role. In particular, for the e-flow definition, in Italy it's currently used a well-established method for the river habitat modelling, named MesoHABSIM (Mesohabitat SIMulation Model) (Parasiewicz, 2001), designed to evaluate the habitat space-time availability for fish fauna. The aim of the methodology is the assessment of the river habitat suitability to host target species, as a function of the hydromorphological characteristics of the water course. The selected scale is the morphological unit (or meso-habitat), which is useful to describe the relationships between physical and biological elements in river environment, according to what suggested by the MesoHABSIM methodology (Vezza, Zanin, and Parasiewicz, 2017).

This work aims to extend the application of the MesoHABSIM methodology to the macroinvertebrate community, representative of the state of ecological health of the river. The work is part of the Benthab Project, which sees the collaboration between the Politecnico of Turin and the University of Parma. Three rivers in Emilia Romagna region (Italy), characterized by pluricursal morphology and braided channels: Trebbia River (PC), Taro River (PR) and Enza River (RE), had been used as case study within the BentHab project. The choice of braided rivers had been dictated by the need and the will to represent a complex system, particularly sensitive to flow rate changes during the spring-summer irrigation season. The field survey activities had been carried out between June and August 2019, in order to describe the most critical situations before and during the local watering activity. The hydromorphological parameters used for the physical description of the habitat are current velocity, water depth, substrate composition and hydraulic connectivity. The biological sampling has been carried out jointly to the morphological and topographic ones, with the purpose of obtaining a complete description in the same flow rate conditions.

The 2D hydrodynamic modelling of the considered river reach, performed through the Hec-Ras software, had allowed to obtain the spatialization of the target variables (depth and velocity). Thanks to the field measurement of current velocity and water depth, it had been possible to validate the hydrodynamic model and, consequently, to use the modelling results to develop



the biological models.

The random forest method, a well-developed machine learning technique, was used to build the statistical models predicting the habitat availability. Thanks to this one, it has been possible to link hydromorphological characteristics of the river and distribution of the macroinvertebrate community, expressed by the use of two synthetic indices: LIFE and LEN-LOT indices.

The statistical analysis has lead to obtain ecologically consistent results, proving that some classes of velocity, depth and substrates are significantly important to predict the taxa distribution. In particular, it was found how values of current velocity close to zero, fine sediments (clay and silt) and poor hydraulic connectivity to the main channel, have strong negative influence on taxa diversity. The model performance of the LEN-LOT index ( $R^2=0.70$ ) highlights a robust relationship between the macroinvertebrate community distribution and the meso-scale habitat morphology. For this reason, it is recommended the use of this index for future activities concerning the description and modelling of macroinvertebrate community habitat.

## Sommario

I corsi d'acqua sono sottoposti a diverse pressioni idromorfologiche dovute alla presenza di opere idrauliche quali derivazioni, sbarramenti, opere di sistemazione e difesa. Al fine di una corretta gestione delle risorse idriche e della conservazione della biodiversità e dei servizi ecosistemici forniti dai fiumi, la definizione di un appropriato deflusso ecologico nei tratti a valle delle opere di presa riveste un ruolo di fondamentale importanza. In particolare, per la definizione dei deflussi ecologici, in Italia è comunemente usato un metodo di modellazione dell'habitat fluviale denominato MesoHABSIM (Mesohabitat Simulation Model) (Parasiewicz, 2001), finalizzato alla valutazione della disponibilità spazio-temporale di habitat per fauna ittica. Obiettivo della metodologia è quello di valutare l'idoneità dell'habitat fluviale per le determinate specie, in funzione delle caratteristiche idromorfologiche del corso d'acqua. La scala scelta è quella dell'unità morfologica (o meso-habitat), utile a rappresentare le relazioni che intercorrono tra elementi fisici e biologici in ambiente fluviale (Vezza, Zanin e Parasiewicz, 2017).

L'obiettivo di questo lavoro è quello estendere l'applicazione del metodo MesoHABSIM alla comunità macrobentonica, rappresentativa dello stato di qualità ecologica del corso d'acqua. Il lavoro si inserisce all'interno del Progetto BentHab, che vede la collaborazione tra Politecnico di Torino e Università di Parma.

Tre fiumi emiliani, caratterizzati da morfologia pluricursale, a canali intrecciati, sono stati utilizzati come casi studio nell'ambito del progetto BentHab: Fiumi Trebbia, Taro ed Enza, in tratti rispettivamente situati in provincia di Piacenza, Parma e Reggio Emilia. La scelta della tipologia di fiumi a canali intrecciati è stata dettata dall'esigenza e dalla volontà di rappresentare un sistema complesso, particolarmente sensibile alle variazioni di portata in alveo durante la stagione irrigua primaverile-estiva. Si tratta, inoltre, dei tre principali fiumi emiliani in destra orografica del fiume Po. Le attività di rilievo in campo sul fiume Trebbia sono state svolte tra i mesi di giugno e di agosto 2019, al fine di rappresentare le situazioni di criticità maggiori prima e durante la stagione irrigua.

I parametri idromorfologici utilizzati per la descrizione fisica dell'habitat sono velocità di deflusso, profondità dell'acqua, e composizione del substrato. I campionamenti biologici sono stati effettuati contestualmente ai rilievi morfologici e topografici, così da ottenere una descrizione completa relativa ad un'unica condizione di portata del corso d'acqua.

La modellazione idrodinamica 2D del tratto fluviale analizzato, realizzata

mediante l'utilizzo del software Hec-Ras, ha permesso di ottenere la spazializzazione delle variabili descrittive (profondità e velocità). Attraverso la misurazione in campo di profondità e velocità della corrente, è stato possibile validare il modello idrodinamico e, di conseguenza, utilizzare i risultati della modellazione per la costruzione del modello biologico.

Per lo sviluppo di modelli statistici di previsione della disponibilità di habitat per la comunità macrobentonica, è stata scelta la tecnica di apprendimento automatico Random Forest. Tale tecnica ha permesso di mettere in relazione le caratteristiche idromorfologiche dei corsi d'acqua con la distribuzione della comunità macrobentonica, espressa tramite l'utilizzo di due indici sintetici, LIFE e LEN-LOT.

L'analisi statistica ha portato ad ottenere risultati ecologicamente consistenti, dimostrando come alcune classi di velocità, profondità e substrato, siano significativamente importanti nella previsione della distribuzione dei taxa. In particolare si è riscontrato come valori di velocità prossime a zero, sedimento fine (limo o argilla) e scarsa connessione idraulica con l'alveo di magra influiscano negativamente sulla diversità dei taxa presenti in alveo. Inoltre, l'aumento della frequenza di tali caratteristiche fisiche dell'habitat orientano la comunità macrobentonica verso una composizione prossima a quella di ambienti lentic, quali stagni e aree umide, scostandosi nettamente rispetto ad una composizione di comunità caratteristica dei corsi d'acqua a canali intrecciati. La performance del modello statistico dell'indice LEN-LOT ( $R^2=0.70$ ) evidenzia un'affidabile relazione tra la distribuzione della comunità macrobentonica e le caratteristiche degli habitat alla meso-scala. Per questa ragione, si consiglia l'utilizzo di questo indice per future attività di valutazione e modellazione dell'habitat della comunità macrobentonica.

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Sommario</b>	<b>iv</b>
<b>List of Figures</b>	<b>viii</b>
<b>List of Tables</b>	<b>xiv</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Material and Methods</b>	<b>5</b>
2.1 Study area . . . . .	5
2.2 Meso-scale Habitat Modelling . . . . .	8
MesoHABSIM . . . . .	9
2.3 Field surveys . . . . .	9
2.3.1 Morphological surveys . . . . .	10
Bathymetry and flow field measurement . . . . .	10
Sediment description . . . . .	12
2.3.2 Topographic surveys . . . . .	15
2.3.3 Biological surveys . . . . .	16
2.4 Hydromorphological data analysis . . . . .	18
2.4.1 Granulometric size distribution . . . . .	18
2.4.2 GIS analysis . . . . .	20
2.4.3 Hydrodynamic simulation . . . . .	23
2.5 Biological modelling . . . . .	27
2.5.1 Biological indices . . . . .	27
LIFE and LENLOT indices . . . . .	28
2.5.2 Introduction to machine learning techniques . . . . .	32
2.5.3 Random Forest . . . . .	32
2.5.4 Model calibration and cross-validation . . . . .	34
<b>3 Results</b>	<b>35</b>
3.1 Hydromorphological description . . . . .	35

3.1.1	Granulometric size distribution . . . . .	35
3.1.2	GIS outputs . . . . .	35
3.1.3	Hydrodynamic model . . . . .	41
3.2	Biological models . . . . .	43
3.2.1	Model calibration . . . . .	44
3.2.2	Parsimonious model selection . . . . .	52
<b>4</b>	<b>Discussion and Conclusions</b>	<b>62</b>
4.1	Data processing . . . . .	62
4.1.1	GIS . . . . .	62
4.1.2	Hydrodynamic model . . . . .	64
4.2	Statistical Model . . . . .	68
4.3	Conclusions . . . . .	72
	<b>Bibliography</b>	<b>74</b>
	<b>Ringraziamenti</b>	<b>77</b>

# List of Figures

2.1	Geographical introduction of Trebbia River . . . . .	5
2.2	Detailed geographical introduction of the study area on Trebbia river . . . . .	6
2.3	Orthophoto representing the studied reach of Trebbia river (June 2019). . . . .	7
2.4	ADCP on the floating platform acquiring depth and velocity data. . . . .	11
2.5	Outputs provided by the software RiverSurveyor Live for a section of the Trebbia river. On top there's the representation the stream bed, thus the depth values measured by both the vertical beam (VB) and the bottom track (BT); in the middle we can observe the track on a N-E plan; in the last square it's visualized the grid of calculation of mean current velocity. The calculated flow rate is $12.97 \text{ m}^3/\text{s}$ . . . . .	12
2.6	Example of zigzag pattern on an isle of Trebbia river, according to Wolman pebble count method. . . . .	13
2.7	Axes of the pebble ( <i>West Virginia department of environmental protection</i> ) . . . . .	14
2.8	<i>DJI Mavic 2 Pro</i> drone used in the surveys in June 2019 on Trebbia river (on the left) and its usage on field (on the right) . . .	15
2.9	<i>TruPulse laser rangefinder 360B</i> telemeter used during the biological surveys (on the left) and measuring of velocity and depth in a sample point with <i>Marsh-McBirney Flow-Mate</i> current meter (on the right) . . . . .	16
2.10	Position of the 180 macroinvertebrates sample points on the Trebbia river reach . . . . .	17
2.11	Moments of the biological sampling. On the left are the emptying of the surber net, sieving of the material and collecting in the bottles. On the right is visible the bottle containing the sampled material under denatured alcohol. . . . .	18

2.12	Different formulas used for the calculation of the Manning coefficient . . . . .	19
2.13	Functioning of the refraction of light between two different media. . . . .	21
2.14	Bathymetric model calibration for Trebbia river. The red line represents the linear regression, that provides the equation, applied for the correction of the stream bed elevation. The real depth values are those collected during the bathymetric surveys, and the drone depth are those extrapolated from the DTM. . . . .	22
2.15	Correction of the bathymetric model: the black profile represent the stream bed elevation provided by the topographic survey. The red one is the stream bed elevation measured during the morphological survey: this is considered the real shape of the bed stream. . . . .	23
2.16	Mesh built for the 2-D hydrodynamic simulation of the Trebbia river reach. The black feature represent the perimeter of the calculation domain and the grey one is the squared grid. . . .	26
2.17	Benthic freshwater macroinvertebrate flow groups, ecological associations and defined current velocities (Extence, Balbi, and Chadd, 1999) . . . . .	29
2.18	Standard EA macroinvertebrate abundance categories (Extence, Balbi, and Chadd, 1999) . . . . .	29
2.19	Scores ( fs) for different abundance categories of taxa associated with flow groups I–VI (Extence, Balbi, and Chadd, 1999)	30
2.20	Correlation between the studied indices and the mean velocity ( $V_m$ ), bed velocity ( $V_f$ ) and water depth (P). The represented indices are: number of families (nfam); Shannon (Sha), Simpson (Sim) and Margalef (Mar); ASPT, 1-GOLD, EPT, $\log_{10}(\text{SelEPTD}+1)$ and LIFE; richness, dispersion and equiripartizione funzionale; LENLOT index. . . . .	31
2.21	Structure and functioning of a decision tree (Choudary, 2020)	33
3.1	Granulometric curves resulting from the substrate composition analysis of the emerged zones of Trebbia river. . . . .	35
3.2	Water depth distribution, acquired by ADCP and current meter, imported and visualized on GIS software. The acquisition frequency is 4 seconds. . . . .	36

3.3	Current velocity distribution, acquired by ADCP and current meter, imported and visualized on GIS software. The acquisition frequency is 4 seconds. . . . .	37
3.4	Water depth distribution of the dataset characterized by 1 second of acquisition frequency, resulting from the interpolation of the raw dataset . . . . .	38
3.5	Current velocity distribution of the dataset characterized by 1 second of acquisition frequency, resulting from the interpolation of the raw dataset . . . . .	39
3.6	Representation of the designed HMUs on the Trebbia river . .	40
3.7	Correct DTM, obtained thanks to the linear regression equation, that relates the fictitious riverbed elevation (and corresponding water depth) to the real ones. The wet stream bed generally turns to have a lower elevation, so that the corresponding zones of the river are characterized by higher water depth. . . . .	41
3.8	Water depth distribution raster, coming from the hydrodynamic simulation. . . . .	42
3.9	Current velocity distribution raster, coming from the hydrodynamic simulation. . . . .	43
3.10	Output of the regression model predicting the target variable LIFE index. <i>Var explained</i> represents the $R^2$ correlation coefficient equal to 46.55 % . . . . .	44
3.11	Ranking of statistical importance of independent variables, using LIFE index as target variable. The importance of the variables is related to the percentage of increasing in MSE ( <i>Mean Squared Error</i> ). This importance is a measure of how much including a variable increases accuracy of the model. . . . .	45
3.12	Error committed in the prediction of target variable LIFE index, as a function of the number of trees creating the random forest. Increasing the number of trees, the error percentage tends to become constant. . . . .	46
3.13	Partial Dependence Plots of the 9 classes of depth, as independent variables in the LIFE index prediction model. . . . .	47
3.14	Partial Dependence Plots of the 9 classes of velocity, as independent variables in the LIFE index prediction model. . . . .	47



3.15	Partial Dependence Plots of the classes of substrate composition, as independent variables in the LIFE index prediction model. . . . .	48
3.16	Partial Dependence Plots of the discrete variable <i>Connectivity</i> [YES] or [NO], in the LIFE index prediction model. . . . .	48
3.17	Output of the regression model predicting the target variable LENLOT index. $R^2 = 69.16\%$ . . . . .	49
3.18	Ranking of statistical importance of independent variables, using LENLOT index. . . . .	49
3.19	Error committed in the prediction of target variable LENLOT index, as a function of the number of trees creating the random forest. . . . .	50
3.20	Partial Dependence Plots of the 9 classes of depth, as independent variables in the LENLOT index prediction model. . . . .	50
3.21	Partial Dependence Plots of the 9 classes of velocity, as independent variables in the LENLOT index prediction model. . . . .	51
3.22	Partial Dependence Plots of the classes of substrate composition, as independent variables in the LENLOT index prediction model. . . . .	51
3.23	Partial Dependence Plots of the discrete variable <i>Connectivity</i> [YES] or [NO], in the LENLOT index prediction model. . . . .	52
3.24	Correlation matrix of the entire variables dataset, used to build the generic biological model. It graphically represents the correlation between variables, calculated with the Spearman method (0.8 - 1) . . . . .	53
3.25	Output of the parsimonious regression model predicting the target variable LIFE index. $R^2 = 49.84\%$ . . . . .	54
3.26	Ranking of statistical importance of independent variables, using LIFE index as target variable, in the parsimonious regression model. The variables represent the selected dataset, as a results of considerations about the generic model. . . . .	55
3.27	Error committed in the prediction of target variable LIFE index, as a function of the number of trees creating the random forest in the parsimonious model. . . . .	56
3.28	Partial Dependence Plots of independent variables used to build the parsimonious LIFE index model, selected from the generic model dataset, ordered by importance. . . . .	57

3.29	Output of the parsimonious regression model predicting the target variable LENLOT index. $R^2 = 71.17\%$ . . . . .	57
3.30	Ranking of statistical importance of independent variables, using LENLOT index, in the parsimonious regression model . .	58
3.31	Error committed in the prediction of target variable LENLOT index, as a function of the number of trees creating the random forest in the parsimonious model. . . . .	59
3.32	Partial Dependence Plots of independent variables used to build the parsimonious LENLOT index model, selected from the generic model dataset, ordered by importance. . . . .	59
4.1	Comparison between two frequency histograms of depth values: the first refers to the values collected every 4 seconds during the acquisition timeframe, the second refers to the values collected every second. The hydro-morphological unit taken into account is a glide (GLIDE 1). It's quite easy to note the similarity between the distributions. . . . .	62
4.2	Comparison between two frequency histograms (4 seconds and 1 second of acquisition interval) of a riffle (RIFFLE 8). . . . .	63
4.3	Comparison between two frequency histograms of velocity values: the first refers to the values collected every 4 seconds during the acquisition timeframe, the second refers to the values collected every second. The hydro-morphological unit taken into account is a glide (GLIDE 1). It's quite easy to note the similarity between the distributions. . . . .	63
4.4	Comparison between two frequency histograms (4 seconds and 1 second of acquisition interval) of a riffle (RIFFLE 8). . . . .	64
4.5	The bi-dimensional model validation was carried out by analyzing this scatter plot. The figure shows the correlation degree between the simulated depth values and the real ones (from the morphological survey). The correlation coefficient is sufficiently close to one, that makes the model valid. The red line represents the ideal situation in which the simulated values perfectly correspond to the real ones. . . . .	65
4.6	Comparison between depth values distribution and the simulated distribution of the same variable, within the study area. It's a relevant result the similarity of distributions: in particular the deeper and more shallow zones are located in the same positions. . . . .	66

4.7	Comparison between velocity values distribution and the simulated distribution of the same variable, within the study area. It's a relevant result the similarity of distributions: in particular the faster and slower zones are located in the same positions.	66
4.8	Comparison between two frequency histograms (measured depth data at 1 second of acquisition interval and simulated depth) of a glide (GLIDE 1).	67
4.9	Comparison between two frequency histograms (measured depth data at 1 second of acquisition interval and simulated depth) of a riffle (RIFFLE 8).	67
4.10	Comparison between two frequency histograms (measured velocity data at 1 second of acquisition interval and simulated velocity) of a glide (GLIDE 1).	68
4.11	Comparison between two frequency histograms (measured velocity data at 1 second of acquisition interval and simulated velocity) of a riffle (RIFFLE 8).	68
4.12	Comparison between the PDPs of the LIFE and LENLOT indices as a function of the independent variable <i>connectivity</i> .	69
4.13	Scatter-plot of the LIFE regression model representing the correlation degree between the calculated and predicted index values.	70
4.14	Scatter-plot of the LENLOT regression model representing the correlation degree between the calculated and predicted index values.	71

# List of Tables

2.1	Sediment classification according to Hauer et al. . . . .	14
2.2	Number of ASTM sieves and mesh dimensions used for the granulometric analysis . . . . .	19
3.1	Comparison between calculated ( <i>LIFE</i> and <i>LENLOT</i> ) and pre- dicted ( <i>LIFE-pred</i> and <i>LENLOT-pred</i> ) values of both <i>LENLOT</i> and <i>LIFE</i> indices, for each hydro-morphological units of the sampling area. . . . .	61

# Chapter 1

## Introduction

Within the environment and nature management and protection, a modelling-based approach plays an essential role. It allows to analyze existing situations and to simulate hypothetical ones, in order to be able to better deal with them.

Particularly, the river habitat modelling aims to relate the presence of a species (or community) to the main hydro-morphological variables (current velocity, water depth and substrates distribution) of a studied river reach. In water resources management, the HSMs (Habitat Suitability Models) represent an effective tool to assess the impact of a given hydro-morphological alteration on the river ecosystem. These habitat models were developed since the eighties, and to date there are different types oriented to the micro or meso-scale. The river habitat modelling aims to find a linkage between the presence of a species or community and the main hydro-morphological features that characterize the river systems: current velocity, water depth and substrates composition. The presence or absence of the species is predicted using the Species Distribution Models.

Among the meso-scale Habitat Suitability Models, particularly relevant is the MesoHABSIM methodology (Meso-habitat Simulation Model) (Parasiewicz, 2001), developed by the Rushing Rivers Institute (MA, USA) and adapted to the Italian context from the Politecnico of Turin (Vezza et al., 2014). According to this method, the main ecological target is the fish fauna. However, a similar process can be implemented for other communities, such as macroinvertebrates, birds etc. Among them, the macroinvertebrate community is very diversified and complex, and high efforts are required to identify these individuals at species level. In fact, in literature many studies about macroinvertebrates habitat availability are based on the classification at family or order level.

Examples of macroinvertebrates distribution models at the species level are

restricted to big size protected species only (i.e freshwater crayfish or bivalves). In particular, in the USA Suitability Habitat Models had been developed for some species of freshwater molluscs and crustaceans, such as *Margaritifera margaritifera* (BiodrawiversityLLC, 2010) and *Alasmidonta heteron* (Parasiewicz et al., 2017). In Italy, two studies in Lombardia and Abruzzo regions have lead to build a species distribution models at the meso-scale for freshwater cryfish or European cryfish (*Austropotamobius pallipes*) (Vezza, Ghia, and Fea, 2015). Still on the subject of the meso-scale habitat modelling, a first simplified study about that was conducted on the Lamprey river for the New Hampshire Department of Environmental Services (Parasiewicz et al., 2008). This analysis focused on four orders of macroinvertebrates: Odonata, Ephemeroptera, Plecoptera and Tricoptera.

Some researches, instead, were carried out about the micro-scale habitat modelling. For example, in France on the Rhone river, M rigoux et al., 2009 have built distribution models for some macroinvertebrates species and orders. These models relate the density of individuals only to one variable: the bed shear stress. In New Zealand, Shearer et al., 2015 have developed habitat suitability curves (HSC) using Generalised Additive Models (GAMs) for 9 macroinvertebrates genera, based on three independent variables: current velocity, water depth and substrate composition. A similar approach, aimed to define e-flow, has been submitted from Li et al., 2008. They have worked on the Xiangxi river (China) and used the *Baetis* genus as ecological target following the IFIM (Instream Flow Incremental Methodology) guidelines. The IFIM was developed in the late 1970's (Bovee and Milhous, 1978) and is based on habitat quality, related to hydraulic characteristics, and the relationship between water flow variations and habitat availability (Navarro, McCauley, and Blystra, 1994). Within the IFIM methodology, the most implemented micro-scale simulation model is the PHABSIM, implemented by Bovee in 1982 (Bovee, 1982). It was developed in United States with applications in 38 states, and aims to provide a method for the definition of appropriate e-flow values, based on the relationship between physical geometry and fish community distribution. Another more recent simulation method at the micro-scale of the river habitat is the CASiMiR (Jorde et al., 2001). As an improvement and complement of the PHABSIM, between the 2000 and 2001 the MesoHABSIM was developed and issued (Parasiewicz, 2001). This methodology will be described in detail in a following section (*Meso-scale habitat modelling*).

The MesoHABSIM methodology (*Mesohabitat Simulation Model*) (Parasiewicz, 2001), has been developed to study the fish fauna habitat availability, but a similar approach may be applied also for other community equally ecologically relevant.

The macroinvertebrates community is of fundamental importance for the ecological health of a river, and it's considered a water quality indicator. This community is a very diversified and complex system, difficult to be classified at species level. It includes organisms larger than 1 mm, such as dragonflies, crayfishes, ephemerals. They're an important link of the food chain and primary processors of organic material, and represent important source of food for many fish. They are indicators of the stream degradation, function of the water oxygenation (Wallace and Webster, 1996). Their presence or abundance, their distribution or absence may be taken into account in a complete evaluation of a river.

The main difference between the existing MesoHABSIM methodology and the described BentHab project, is the typology of statistical models on which they're based on. The MesoHABSIM exploits the *Habitat Suitability Models*, that relate the presence or absence of a species (or community) to the hydro-morphological characteristics of the studied river; the models that predict that presence or absence are the *Species distribution models*.

In order to acquire all the necessary data for the species distribution modelling, morphological, topographic and biological surveys had been carried out on Trebbia River. This natural water course is a braided river, characterised by a strong relationship with the trend of rainfall and derivations, so that it represents a complex system. This choice eventually allows to build prediction models which can later describe a wide range of rivers with different hydro-morphological characteristics and flow fields.

The field works took action first in June 2019, aiming to represent the situation before the critical situation due to the seasonal irrigation. Then, the same measurements had been conducted again in August 2019, which usually hosts the irrigation period. This second field acquisition campaign had provided data for the validation analysis, that will be carried out in autumn 2020. All the field sampling operations were carried out simultaneously in order to completely describe each aspect at the same flow rate conditions.

The main descriptors of a river system to be modeled according to the BentHab project and the MesoHABSIM methodology, are the current velocity, the water depth and the substrates distribution. These parameters had been

collected during the field acquisition activities, together with the macroinvertebrates sampling, carried out by a research group of Parma University.

The first steps of the data analysis were to subdivide the considered river reach in Hydro-Morphological Units (HMU) as suggested by the MesoHAB-SIM methodology, and to determine the classes extremes where calculate the relative frequencies of the measured variables (velocity, depth and substrates).

The collected data had been elaborated to obtain some necessary results: the digital terrain model (DTM) of the study area, the hydrodynamic model that describes the flow field, and the relative frequency of presence of each class of current velocity, water depth and substrates distribution.

The final aim is to obtain a predictive biological model of the macro-benthic community distribution, as a function of the hydro-morphological characteristics of the river reach. The target variables, calculated from the collected data, and predicted by the statistical models, are two synthetic indicators: LIFE and LEN-LOT indices. They describe the affinity to the current velocity of the macroinvertebrate community in each meso-habitat and represent the linkage between the physical description and the biological characteristics of the river.



## Chapter 2

# Material and Methods

## 2.1 Study area

The study was carried out on the Trebbia river, near the village of Rivergaro (PC) in Emilia Romagna region (ITALY). The biological model was built using the data collected on this site. Two other watercourses, Taro river and Enza river, were studied in order to provide data for the model validation. The Trebbia river spring is on the Antola massif, in the Ligurian Apennines, on the south slopes of Monte Prelà. It is one the most important northwest right-bank tributaries of the Po river. Its basin is 1150 Km<sup>2</sup> wide distributed among Liguria, Emilia Romagna and Lombardia regions. It flows into the Po river on the west side of Piacenza.

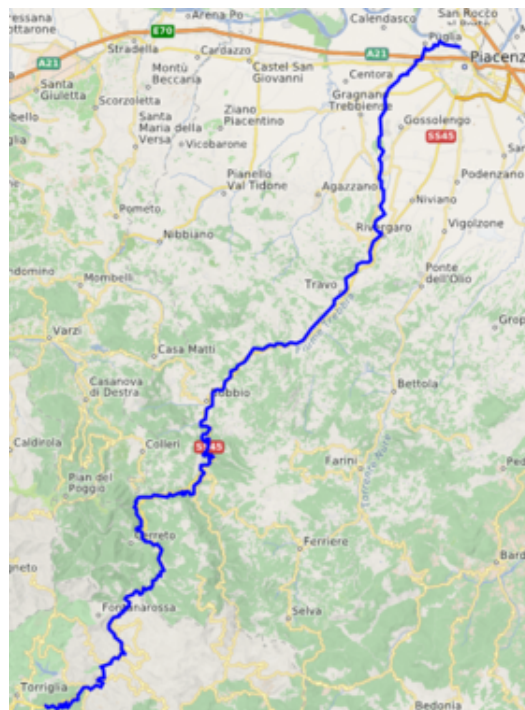


FIGURE 2.1: Geographical introduction of Trebbia River

In particular, the river reach under study is located near the village of Canneto Sotto, just downstream of the Tuna bridge and an intake structure for irrigation purpose and just north of Rivergaro.

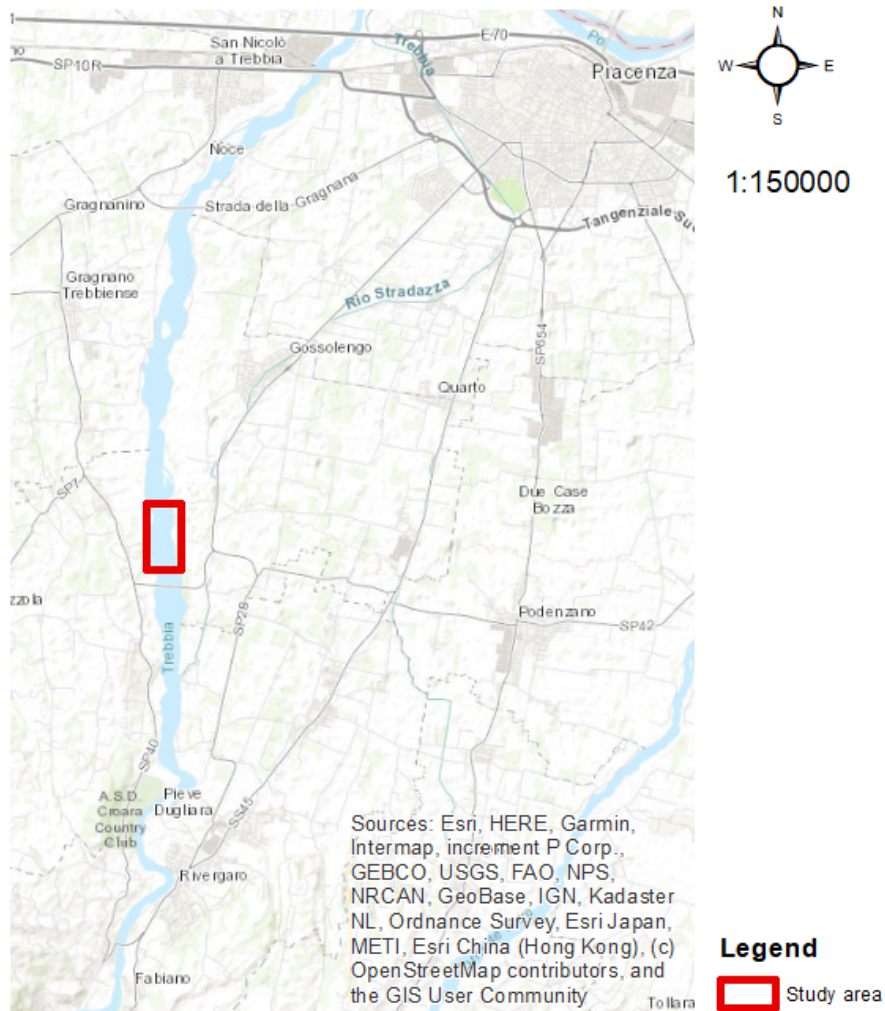


FIGURE 2.2: Detailed geographical introduction of the study area on Trebbia river

The stream section under investigation is approximately 1,200 meters long and 500 meters wide. It is quite a large extension, involving significant sampling effort, but it permits to obtain a variety of type of habitats, class of velocity, depth and substrate composition sufficient and necessary to build a robust biological model.

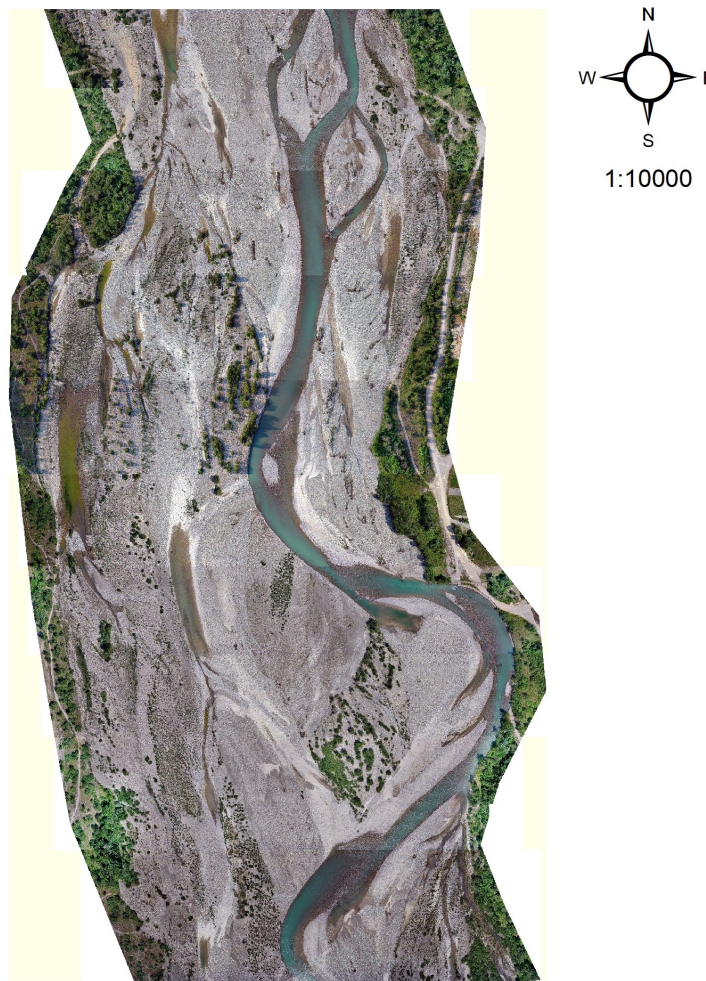


FIGURE 2.3: Orthophoto representing the studied reach of Trebbia river (June 2019).

The northern Apennines are a belt of northeast-southwest mountains on the coast of northwest Italy, composed primarily of sedimentary rock. For much of their lengths, the braided rivers in this area run through montane terrain, which sediments are eroded from the mountains, deposited in alluvial fans and terraces in the Po valley. Significant lengths of these rivers flow above ground because of the intense deposition within the stream bed.

The Trebbia River is an environment strongly conditioned by the seasonal trend of rainfall. The torrential course influences the surrounding vegetation so that, in the sections less exposed to flood streams, it is possible to observe luxuriant flora growth, such as, for example, abundant blooms of wild orchids. However, the rest of the landscape is composed of large river terraces where grasslands and shrubs alternate with cultivated fields.

The abundance of farmlands in this area involves the need of water for irrigation purposes, like most of the Pianura Padana plain. In order to provide sufficient amount of water, this river is subjected to numerous intake structures, that have to release a certain flow rate, named DMV (*Deflusso Minimo Vitale*). This one represents "*the instantaneous flow rate to be determined in each homogeneous section of the watercourse, which must ensure the preservation of the physical characteristics of the water body, chemical - physical parameters and the maintenance of biocoenoses typical of local natural conditions*" (Art.7, Comma 1). The definition is established by the Italian reference legislation Law 152/2006 (in Italian *Testo Unico Ambientale*). This project, just like the MesoHABSIM, aims to find such a way to contribute in the definition of the DMV (also named e-flow), through the modelling of macroinvertebrate community.

## 2.2 Meso-scale Habitat Modelling

The meso-scale habitat modelling had been developed since the last years of the past century, as a response to some problems and lacks of the traditional micro-scale methodologies. The main idea of these new approach is to assess the habitat availability for species at a larger scale, instead of associating the criteria of suitability to small units of microhabitats.

The mesohabitat is defined as a *functional habitat*, which is that portion of river surface where an animal can be observed for a significant part of its diurnal routine (Kemp, Harper, and Crosa, 1999). For what concerns the superficial rivers, the mesohabitat dimensionally coincides with the Hydro-Morphological Unit (HMU), such as *rapid, riffle or pool* (Rinaldi et al., 2016). These structures have a spatial extension in the range between  $10^{-1}$  and  $10^3$  m, much larger compared to the microscale structures in the order of 10cm. According to what indicated Parasiewicz, 2007a, the meso-habitat is that portion of a river in which, due to the homogeneity of the morphological characteristics and the hydrodynamic configurations (HMU), and the presence of physical attributes (environmental descriptors), favourable conditions for the survival and growth of a particular aquatic species or its vital stage are created.

## MesoHABSIM

Within the meso-scale habitat modelling methodology, an essential role is played by the MesoHABSIM, developed from the year 2000 as a complement of the PHABSIM, and described and issued one year later (Parasiewicz, 2001). This methodology directly comes from the traditional physic micro-habitat models, with different data acquisition process and analytical resolution approach. The model structure was obtained by operating for successive reductions in the collection of field data compared to the classical models, going to select only those physical and environmental descriptors that more explicitly condition the preferences of habitat for the fluvial community (Parasiewicz et al., 2009). The spatial and temporal resolution became more adapt to the quantification of available habitat for aquatic species, and the perception of details is less.

Thus, the MesoHABSIM is a physical modelling system and involves strategy of data collection and analytical techniques that allows the user to assess the suitability of river habitat to host target species, as a function of environmental existing conditions (Parasiewicz, 2007a). Through the application of the method it's possible to evaluate the ecological spatial-temporal integrity of a reach of river, obtaining a quantification of the Habitat integrity Index (IH) (Organization, 2019).

Within the Italian context, the MesoHABSIM had been improved and adapted to the Italian rivers by the Politecnico of Turin (Department of land, environment and infrastructure, DIATI) from 2010. To date, it represents the river habitat suitability model for the whole national context, and the procedural guide for its application is the manual of ISPRA (Istituto Superiore per la Protezione e la Ricerca Ambientale) written in 2017 (Vezza, Zanin, and Parasiewicz, 2017).

## 2.3 Field surveys

The field surveys were carried out between the 3rd and the 12th of June 2019 on the chosen reach of Trebbia river, with the aim to obtain a morphological, biological and topographic description. The flow rate conditions of the river were pretty representative of the period of the year, close to  $13 \text{ m}^3/\text{s}$ . An important prerogative to be able to operate the surveys was the behaviour of the hydrograph, which has had to be constant in order to ensure a steady state flow rate during the surveys themselves. The sampling efforts, for all the

surveys types, were significant, also due to the spatial extension of the study area: more than 1 km long and 500m wide. The number of mapped hydro-morphological units is 85, including the isolated ponds not connected to the main channel. Also the number of biological sampling points was huge, they were collected samples of macroinvertebrates in 180 different positions along the entire river reach. At last, the effort of sampling the velocity and depth values had made possible to collect more than 10.000 measure points, after the elaboration.

### 2.3.1 Morphological surveys

The morphological surveys have provided a complete and high detailed description of the flow field and the bathymetry, of the granulometric composition and the substrates classification.

#### **Bathymetry and flow field measurement**

The technique used for the water depth and current velocity data acquisition is based on the RiverSurveyor M9 (Sontek, San Diego, CA, USA) that is made up of the device HydroSurveyor Acoustic Doppler Profiler (ADP) and of a multi-beam precision depth sounder. This system allows to acquire punctual depth data, in particular two types of depth are measured: Bottom Track and Vertical Beam depth. The first method is used to record the vessel speed over the ground (especially when the GPS is lost), and provides also rough depth data. This hydroboard, however, has the RTK GPS option, an ultra precise positioning, to be used as an alternative to bottom tracking, and the vertical beam to obtain more precise water depth values. At the same time it's calculated a velocity profile along the water column, respective to each sample point: all these water columns along which the mean velocity is measured, create the grid of mean velocity of the entire section. The functioning is based on the Doppler effect at the interaction between the emitted waves and the water waves. The temporal acquisition frequency is 1 second and it's possible to keep track of the relative displacements with respect to the initial measure point. In order to obtain a georeferenced morphological description, the acquisition system was completed by the use of a double frequency antenna GNSS (Global Navigation Satellite System). The temporal frequency of acquisition is 4 seconds, so that every 3 points of depth/velocity measured by the boat, there's a georeferenced one. In order to georeference all the measured points, it was operated an interpolation of the coordinates,



exploiting the recorded track of the boat. The set of devices was assembled on a floating platform, as shown in the following figure.



FIGURE 2.4: ADCP on the floating platform acquiring depth and velocity data.

For both post-processing data elaboration and real-time data monitoring, it's available the software RiverSurveyor Live, that allows to extract the acquired data and to visualize the track followed by the vessel. Several outputs are provided by this software: it's possible to visualize the georeferenced track of the devices, the shape of the stream bed and the depth of water, or the grid of calculation of the current velocity in each section. Furthermore, the system allows the operators to obtain the flow rate value by operating on an appropriate section of the river. It follows an example of these results.

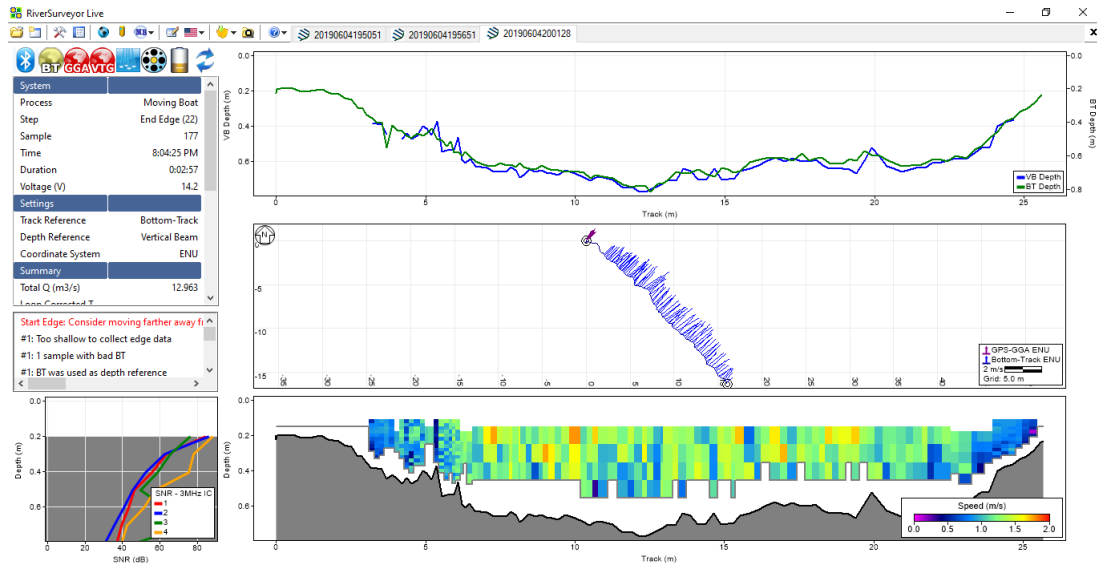


FIGURE 2.5: Outputs provided by the software RiverSurveyor Live for a section of the Trebbia river. On top there's the representation the stream bed, thus the depth values measured by both the vertical beam (VB) and the bottom track (BT); in the middle we can observe the track on a N-E plan; in the last square it's visualized the grid of calculation of mean current velocity. The calculated flow rate is 12.97  $\text{m}^3/\text{s}$

Among the limits of this acquisition method there's the absence of recorded data in correspondence of the zones characterized by depth less than 20 cm circa. In that zones, the lack of data was integrated with electromagnetic current meter acquisition (*MF PRO, OTT Loveland, CO, USA*). The spatial frequency of sampling is much lower compared to the boat method, but it allows to obtain a more complete dataset.

### Sediment description

The sediment and substrate classification and analysis was conducted following two different method on the emerged zones and the submerged zones: the sampling of the first ones had lead to obtain the granulometric description of the considered reach of Trebbia river, assuming as homogeneous the distribution of the substrates on the entire riverbed; the submerged zones were sampled in order to obtain a precise description of the substrates in each hydro-morphological unit.

In order to operate the sampling and the granulometric classification of the emerged zones, it was followed the method of Wolman named *Pebble count*. It's based upon an analysis of the relative area covered by given sizes (Wolman, 1954), in order to obtain representative samples of an entire reach of river



characterized by a wide range of sediments. The application of the method involves some steps:

- Establishment of a grid, or subdivision in hydro-morphological units (HMUs) to be described.
- Picking up the pebbles from the emerged river bed, following a pattern that allows to homogeneously sample the area (zigzag pattern). It's important to ensure randomness while picking up, for instance by not looking at the ground in that moment.

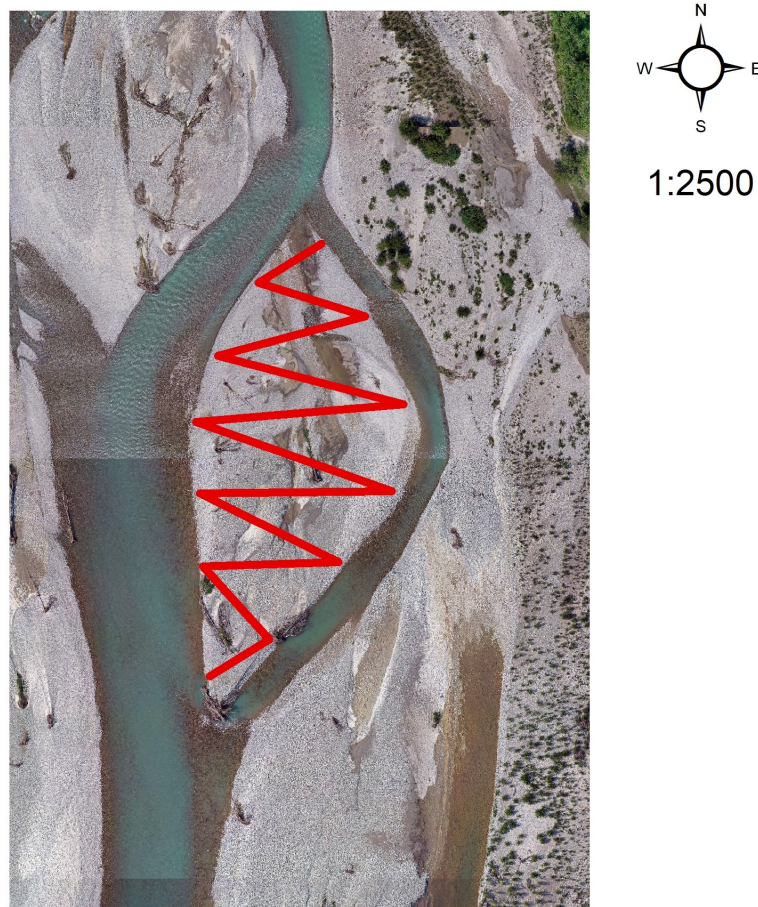


FIGURE 2.6: Example of zigzag pattern on an isle of Trebbia river, according to Wolman pebble count method.

- Measuring the intermediate axis of the pebble.

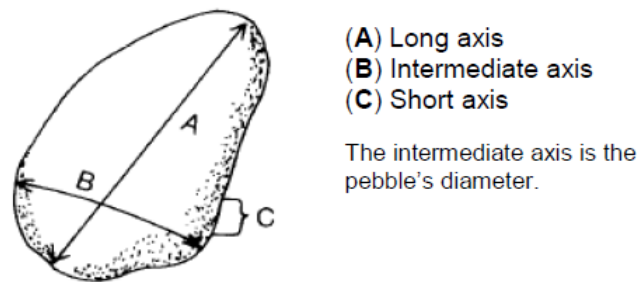


FIGURE 2.7: Axes of the pebble (West Virginia department of environmental protection)

- Classification according to the scale in the following table.

<i>Substrate</i>	<i>Size range</i>
<i>Gigalithal</i>	(bedrock)
<i>Megalithal</i>	>40 cm
<i>Macrolithal</i>	20 - 40 cm
<i>Mesolithal</i>	6 - 20 cm
<i>Microlithal</i>	2 - 6 cm
<i>Akal</i>	0.2 - 2 cm (gravel)
<i>Psammal</i>	0.06 - 0.2 cm (sand)
<i>Pelal</i>	<0.06 cm (clay)
<i>Detritus</i>	(organic matter)
<i>Xylal</i>	(woody debris)
<i>Sapropel</i>	(anoxic mud)
<i>Phytal</i>	(submerged plants)

TABLE 2.1: Sediment classification according to Hauer et al.

The finer samples (sand and fine gravel) were collected in order to conduct in laboratories a detailed granulometric analysis, obtaining the granulometric curves, later showed and described.

On the submerged zones of the water stream was applied another classification method based on what suggests the MesoHABSIM methodology. The definition of the sampled substrates from the riverbed was carried out according to a stratified random technique, at the HMU level. This particular stratified random sampling method involves the division of the entire population being studied in subgroups, named *strata*. This process allows to obtain a representative sample of a large area, by using the subdivision in

strata that, in this case, are the hydro-morphological units. The strata have to be sufficiently homogeneous mainly in current velocity, in order to ensure the reliability of the random sampling technique.

### 2.3.2 Topographic surveys

The photogrammetric surveys, carried out simultaneously to the morphological and biological ones, have played an essential role in order to obtain the Digital Terrain Model of the river reach. The used instruments are basically a double frequency antenna, the same used on the vessel with the hydroboard, and a remote piloted aircraft *DJI Mavic 2 Pro*.



FIGURE 2.8: *DJI Mavic 2 Pro* drone used in the surveys in June 2019 on Trebbia river (on the left) and its usage on field (on the right)

This field work has taken place in two main steps: the first has aimed to place a consistent number of Ground Control Points on the entire area of interest, in order to precisely georeference the orthophoto provided by the drone footage; the second step has consisted in the drone flight, with a previous flight plan.

The post-processing of the recorded images has provided the RGB orthomosaic of the Trebbia river reach (spatial ground resolution 1.3 cm) and the corresponding DSM (*Digital Surface Model*) and DTM (*Digital Terrain Model*), of fundamental importance in order to perform the hydrodynamic modelling. The processing of the photograms was carried out with the use of *PIX4D* software (*PIX4D SA*, Lausanne, Switzerland) and applying *structure from motion* techniques, that utilizes 2D images in series in order to reconstruct a 3D scene.

### 2.3.3 Biological surveys

The macroinvertebrates sampling was carried out from the 5th to the 13th of June 2019 on the Trebbia river reach, in correspondence of a part of HMUs individuated according to the MesoHABSIM methodology. The samples were collected using a surber net (frame 0.05 m<sup>2</sup>, mesh 500 µm). Contextually to the biological samples, it has been recorded substrate type, position, depth, medium and bed current velocity, in each sample point. The substrate was classified according to Buffagni and Erba, 2007; depth and velocity values were measured using a current meter with graduated rod (Marsh-McBirney Flow-Mate, USA); the georeferencing has been operate with a laser rangefinder (TruPulse laser rangefinder 200, Laser Technology, USA) that provides horizontal distance and azimuth degree from the measure point.



FIGURE 2.9: *TruPulse laser rangefinder 360B* telemeter used during the biological surveys (on the left) and measuring of velocity and depth in a sample point with *Marsh-McBirney Flow-Mate* current meter (on the right)



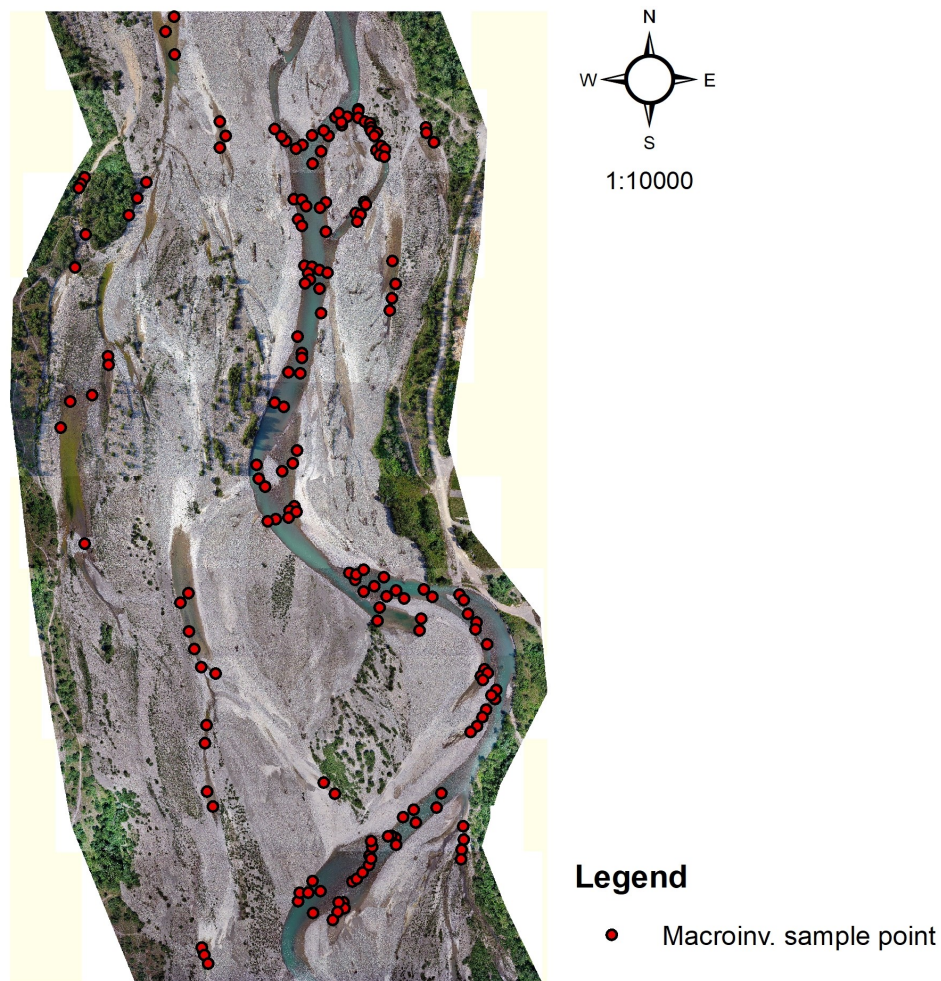


FIGURE 2.10: Position of the 180 macroinvertebrates sample points on the Trebbia river reach

The collected material during the sample was filtered with a sieve and impurities or amount of algae in excess were removed. The sampled macroinvertebrates were collected in plastic bottles (visible in the following figures) and preserved in denatured alcohol 90°, in order to be later analyzed in the laboratories of the University of Parma. The number of collected samples was 180, distributed on the entire area of interest, that's a discrete sampling effort and makes this survey's results a rich and robust dataset.



FIGURE 2.11: Moments of the biological sampling. On the left are the emptying of the surber net, sieving of the material and collecting in the bottles. On the right is visible the bottle containing the sampled material under denatured alcohol.

## 2.4 Hydromorphological data analysis

All the collected data about flow field, bathymetry, substrates classification and granulometry are successively elaborated. For what concerns the granulometric analysis, that's the one concrete laboratory analysis conducted at the Politecnico of Turin. The other elaborations were carried out using software and digital instruments. The main objective of this type of analysis was to obtain a complete description of each physical habitat in terms of hydromorphologic characteristics. In order to reach in the best way this aim, the MesoHABSIM methodology was followed: 9 classes of current velocity, 9 classes of water depth and 13 classes of substrate were created, and for all the HMUs were calculated the relative frequencies of each class.

These HMUs were designed first of all according to the definition of hydromorphological units provided by the manual of ISPRA (Rinaldi et al., 2016) that foresees 8 different type of units: *pothole*, *cascade*, *rapid*, *riffle*, *step*, *pool*, *glide*, *dune system*, characterized by habitat homogeneity.

### 2.4.1 Granulometric size distribution

The procedure followed during this analysis is the classical method that involves the use of ASTM sieves with different mesh, and a balance used to weigh the material passing by each sieve. The series of mesh that marks the ranges of passing diameters is in the following table.

ASTM sieve n	Mesh [mm]
3.5	5.66
6	3.36
7	2.83
8	2.58
10	2
20	0.84
40	0.42
50	0.297
60	0.25

TABLE 2.2: Number of ASTM sieves and mesh dimensions used for the granulometric analysis

This analysis was conducted considering the samples of substrates collected on the emerged zones of the river bed. This choice is dictated by a feasibility reason: the difficulty to collect fine sediment in certain wet deeper zones have made almost impossible operating this analysis on the entire reach. Nevertheless, it's possible to assume that the medium granulometric conditions are closely the same in every zones of the area of interest. The aim of this analysis is, as well as the mere classification, the calculation of the Manning's coefficient, of essential importance for the purpose of the hydrodynamic simulation. This coefficient is a measure of the roughness of the river bed. There are several applied formulas that allow to calculate the coefficient depending on the granulometric characteristics.

Strickler	Meyer-Peter & Muller	Julien (d50)	Julien (d90)	Crow (d50)	Crow (d90)
$\frac{1}{21,1} \cdot \frac{d_{50}^{\frac{1}{6}}}{d_{50}^{\frac{1}{6}}}$	$\frac{1}{26,0} \cdot \frac{d_{90}^{\frac{1}{6}}}{d_{90}^{\frac{1}{6}}}$	$0,062 \cdot d_{50}^{\frac{1}{6}}$	$0,038 \cdot d_{90}^{\frac{1}{6}}$	$0,034 \cdot d_{50}^{\frac{1}{6}}$	$0,036 \cdot d_{90}^{\frac{1}{6}}$

FIGURE 2.12: Different formulas used for the calculation of the Manning coefficient

The first results were compared to the value of the same coefficient, calculated using the Chezy formula:

$$Q = \chi S \sqrt{Rj}$$

Where:

$$\chi = K_s R^{1/6}$$

and the Manning coefficient is:

$$n = \frac{1}{K_s}$$

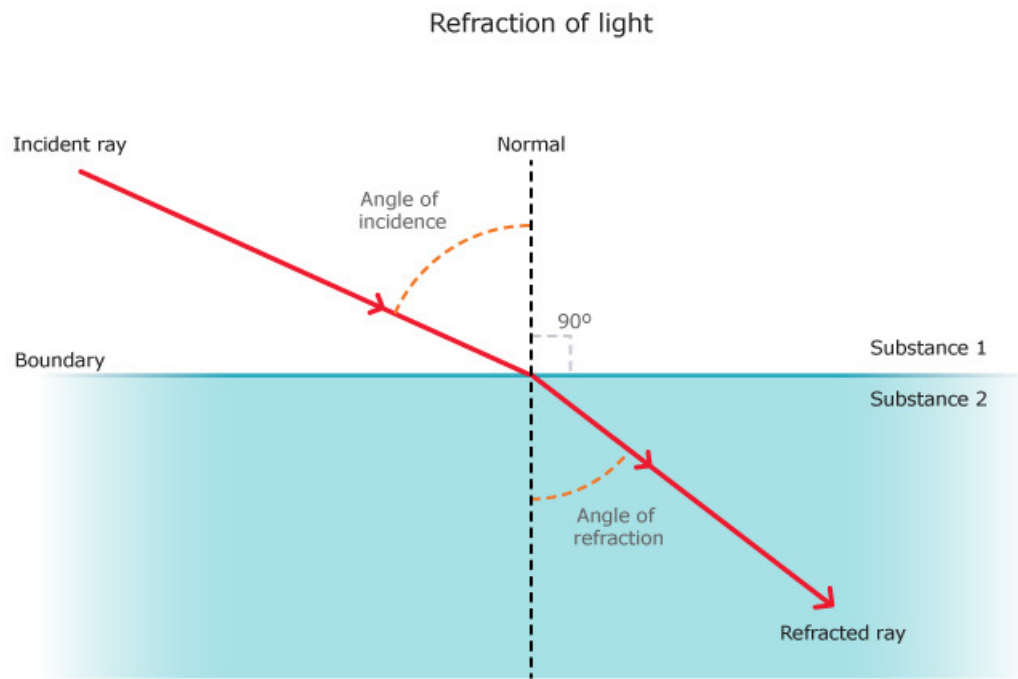
After all the consideration about that, it was established that the optimal Manning coefficient is  $0.036 \text{ m}^{-1/3} \cdot \text{s}$ .

### 2.4.2 GIS analysis

The analysis was conducted using the software ArcGIS (ESRI, ArcGIS Desktop: Release 10.7.1 Redlands, CA: Environmental Systems Research Institute). An important step to be carried out is the definition of the HMUs, previously described in within the MesoHABSIM section. The criteria on which the design of the units was operated are: the homogeneity in current velocity within the area, the description of Rinaldi et al., 2016 (riffle, glide, pool, etc) and the physical conformation of the riverbed. The HMUs were manually designed first of all during the field surveys, in order to organize the surveys themselves and the biological sampling, and successively adjusted using the ArcGIS editor.

Within the hydromorphological data analysis, a calibration of the bathymetric model was performed in order to reconstruct the topographic surface of the riverbed. The aim of the analysis was to obtain the correct dtm (Digital Terrain Model) of the study area, to be used as geometrical basis for the hydrodynamic modelling. It was necessary to merge the georeferenced information coming from the photogrammetric (or topographic) survey and the bathymetric acquisition. In fact, the analysis was designed to obtain the linear regression model which best describes the relationship between the measured depth, considered as real (*ground truth*) and the fictitious ones, estimated through the photogrammetric analysis and finally provides the correction equation. This equation allowed to adjust the extrapolated DTM, obtaining a new DTM that correctly describes the submerged zones elevation. The need of the DTM correction comes from the error committed during the acquisition of the riverbed elevation related to the refraction effect.





© Copyright. 2012. University of Waikato. All Rights Reserved.

FIGURE 2.13: Functioning of the refraction of light between two different media.

According to the refraction effect, that modifies the beam direction in the passage from a less dense (air) to a more dense mean (water), it was recorded an higher elevation of the submerged riverbed instead of the real one. Thus, the water depth in those zones, obtained as a raster difference (on GIS) between the reconstructed free surface raster and the DTM, is lower than the real one. The reconstruction of the free water surface was built on a GIS software through a spatial interpolation, previously defining the contours of the wet zones. Then, a raster difference provided the water depth distribution, to be corrected in order to eliminate the error.

The higher the water depth, the greater is the error on the elevation due to refraction and the behaviour is described by the following formula:

$$D_f = D_r \frac{n_{\text{air}}}{n_{\text{water}}}$$

Where  $n$  is the refraction coefficient of the means,  $D_f$  is the fictitious depth and  $D_r$  is the real one, obtained with the correction. In the practical applications of this correction, it is not used the refraction coefficient because of the presence of other environmental variables, such as air humidity, water

turbidity, resolution of the acquisition instruments ecc. Consequently, in order to define the relation between the fictitious and the real depth, it may be built a statistical model that allows to reconstruct the real elevation of the riverbed. This model is named *bathymetric model* and usually involves a linear regression function.

Precisely, the equation provided by the statistical analysis, in which the fictitious depth are those acquired by the drone, and the real ones are those collected during the hydromorphological surveys, is shown below.

$$D_r = 1.163D_f - 0.024$$

In particular, it was found that for depth up to 20 cm the correction was not necessary, so that the linear regression equation was applied only to the zones characterized by water depth greater than 20 cm, since they may be considered correct. The bathymetric model function follows:

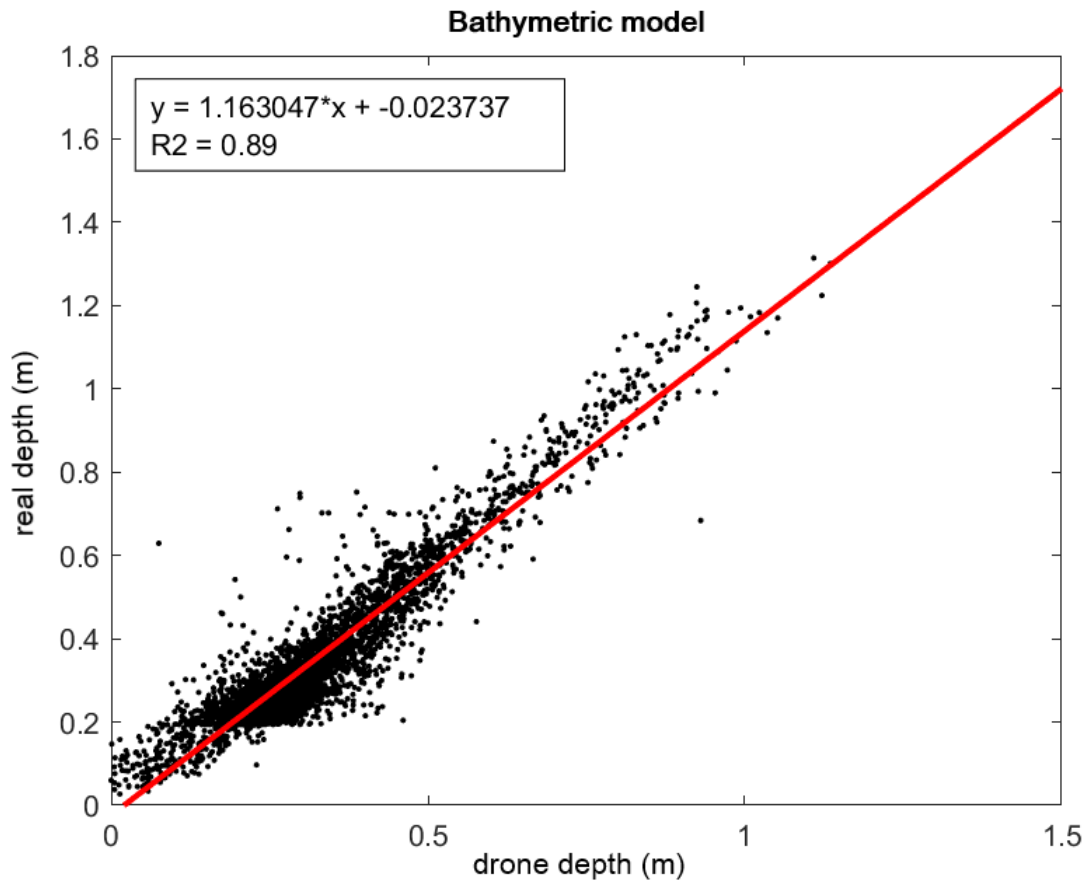


FIGURE 2.14: Bathymetric model calibration for Trebbia river. The red line represents the linear regression, that provides the equation, applied for the correction of the stream bed elevation. The real depth values are those collected during the bathymetric surveys, and the drone depth are those extrapolated from the DTM.

Anyway, it's important to point out the high value of the determination coefficient  $R^2$  (0.89), that indicates a significant correlation between the data sample distribution and the implemented statistical model.

As an example in the following figure it is shown a section of the wet riverbed of Trebbia, in which is possible to observe the correction of the bed elevation operated through the bathymetric model. The black dots represent the fictitious elevation and the red ones are those collected (real elevation).

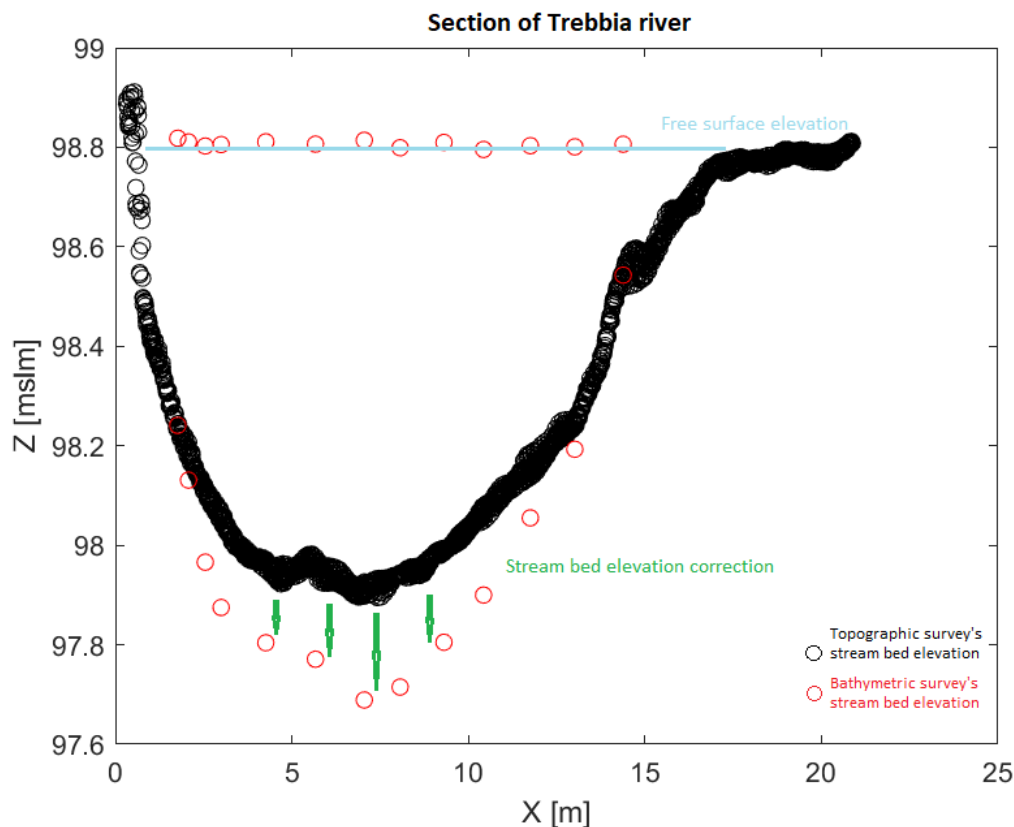


FIGURE 2.15: Correction of the bathymetric model: the black profile represent the stream bed elevation provided by the topographic survey. The red one is the stream bed elevation measured during the morphological survey: this is considered the real shape of the bed stream.

### 2.4.3 Hydrodynamic simulation

The 2-D hydrodynamic simulation on the Trebbia river reach was performed using the software HEC-RAS (US Army Corps of Engineers, Hydrologic Engineering Center, Davis, California) version 5.0.7. The simulation is based on the DTM of both the emerged and wet zones and provides a flow field reconstruction. This software is widely used all over the world for the simulation

of fluid flow in both natural and artificial channels network. It offers the opportunity to simulate in mono or bi-dimensional systems, in either steady or unsteady flow conditions, and has a lot of different interpretation ways of the results.

Basically, the software resolves the main equations of the fluid dynamic: the mass and momentum conservation equations (Brunner, 2016).

- case 1-D, steady flow conditions: the simulation is based on the 1-D energy equation, with a iterative procedure called *standard step method*.
- case 1-D, unsteady flow conditions: the solved equation are both the mass (continuity) and momentum conservation.
- case 2-D, unsteady flow conditions: the continuity and momentum conservation equations are respectively:

$$\frac{\delta H}{\delta t} + \frac{\delta(hu)}{\delta x} + \frac{\delta(hv)}{\delta y} + q = 0$$

Where  $H$  is the water surface elevation,  $u$  and  $v$  are the horizontal components of the velocities, and  $q$  is a source/sink flux term. The momentum equations for both the horizontal directions are:

$$\frac{\delta u}{\delta t} + u \frac{\delta u}{\delta x} + v \frac{\delta u}{\delta y} = -g \frac{\delta H}{\delta x} + v_t \left( \frac{\delta^2 u}{\delta x^2} + \frac{\delta^2 u}{\delta y^2} \right) - c_f u + f v$$

$$\frac{\delta v}{\delta t} + u \frac{\delta v}{\delta x} + v \frac{\delta v}{\delta y} = -g \frac{\delta H}{\delta y} + v_t \left( \frac{\delta^2 v}{\delta x^2} + \frac{\delta^2 v}{\delta y^2} \right) - c_f v + f u$$

Where  $u$  and  $v$  are the velocities in the Cartesian directions,  $g$  is the gravitational acceleration,  $v_t$  is the horizontal eddy viscosity coefficient,  $c_f$  is the bottom friction coefficient and  $f$  is the Coriolis parameter. On the left hand side of the equations are the acceleration terms, on the right hand side we can find the internal or external forces acting on the fluid.

A simplified form of the momentum equation is the single differential vector form:

$$\frac{\delta V}{\delta t} + V \nabla V = -g \nabla H + v_t \nabla^2 V - c_f V + f k \times V$$

These equations, combined with the initial and boundary conditions, complete the hydrodynamic system.

The initial condition is defined by a completely without water area (the study

area), that will be gradually filled by the fluid as a result of the flow upstream release, coming from the initial section of the considered grid.

The boundary conditions, instead, are of fundamental importance for the resolution of the differential equations, which requests the definition of the hydrodynamic conditions of the upstream and downstream sections. For what concerns the upstream one, the condition is represented by the input hydrogram, which defines the input water volume at each step of the simulation. Precisely, the flow rate of the hydrogram set as input of the simulation is  $13 \text{ m}^3/\text{s}$ , as provided by the River Surveyor software. The final section condition is found in the *normal depth* mode, known the slope of the bed, downstream the same section.

Fixed these conditions, the used hydrodynamic simulation model is bidimensional, fix-bottomed, with unsteady state conditions. The choice of a modelling of this type is strongly related to the particular hydro-morphology of the considered river. The Trebbia river is, in fact, characterized by complex behaviour of fluid current, due to the pluricursal planimetric asset, and this peculiar situation is surely better described by a bi-dimensional model approach. By applying this type of simulation the flow field is simulated in both the horizontal directions, resulting in a more completed and detailed reconstruction of the water surface and its moving.

The HEC-RAS software allows the user to use the elevation information provided by the Digital Terrain Model, for the definition of the geometry of the grid on which realize the hydrodynamic simulation. In particular, through the RAS Mapper, the GIS section of the software, is possible to import a DTM to be used as tridimensional representation of the bed stream. Thus, the spatial resolution of the DTM represents a potential limiting factor for the simulation performance. In this case, the used corrected DTM has a spatial resolution of 10 cm.

In order to define the boundary conditions of the simulated geometry, it was necessary to determine the mesh on which the resolution of the equations is based. The mesh has to contain the entire wet area of the river, in particular the portion of the active riverbed, and represents the reference domain in which the hydrodynamic equations are solved. In this case, the mesh size is 1 meter.

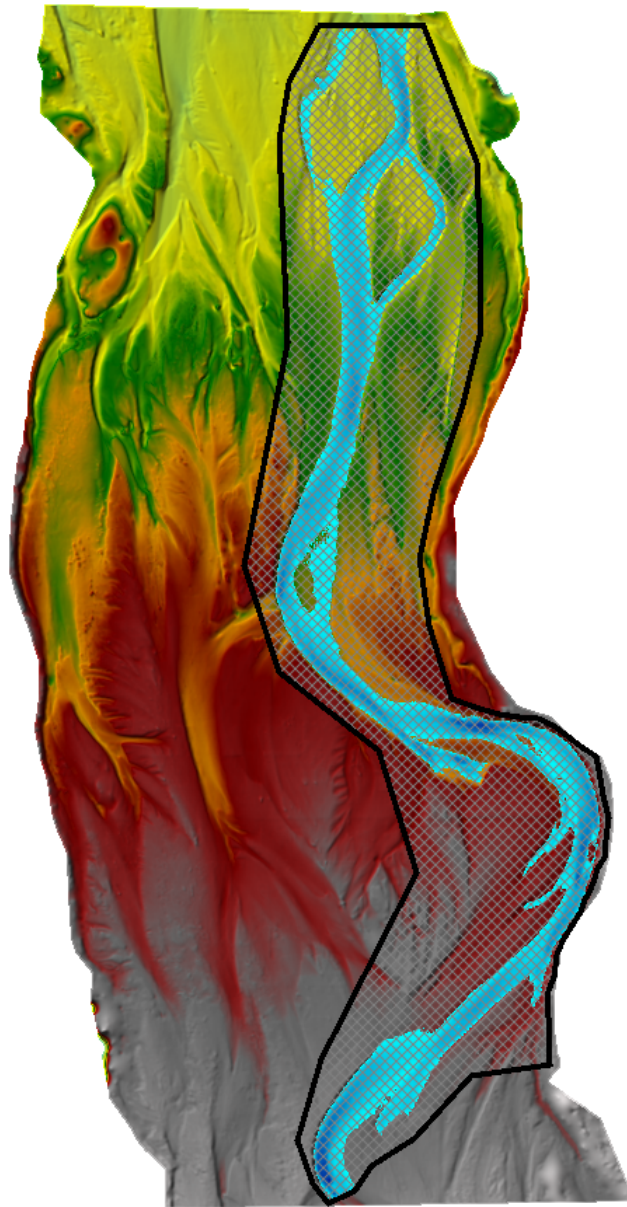


FIGURE 2.16: Mesh built for the 2-D hydrodynamic simulation of the Trebbia river reach. The black feature represent the perimeter of the calculation domain and the grey one is the squared grid.

Another fundamental parameter to be defined and insert among the input settings, is the Manning coefficient, determined during the granulometric analysis.

This hydrodynamic simulation provided a complete flow field and water depth distribution dataset, that correctly represents the real analysed situation, being compared with the measured velocity and depth data. For this reason, the simulated data was used as input for the biological model calibration. This choice allowed to base the statistical modelling on a more complete and representative dataset.

## 2.5 Biological modelling

### 2.5.1 Biological indices

Downstream the biological sampling, it was necessary to analyze and classify the collected macroinvertebrates, in order to obtain some needed parameters that allowed to chose and define the target variable to be modeled. This target variable is that parameter which is predicted by the model and describes the species distribution. So, the first step in order to extent the MesoHABSIM methodology to the macroinvertebrates, was to find that parameter which ensures:

- A response of the target variable itself to physical independent variables (current velocity, water depth and substrates distribution);
- The transferability of the results to more than one different rivers.

By following these two main goals, it was decided to use as target variables summary indices, able to summarize one or more characteristics of the community in a value. The second prerogative of this research (the transferability) is totally allowed by these summary indices, because the the provided information is not related to the taxonomic composition of the starting data. Thus, it was chosen to not model the distribution of all or a subset of the taxa of the community, due to the need of very detailed information about the habitat preferences of the numerous different types of taxa that compose the community. Only for the most representative taxa of the community can be possible to collect that needed detailed information, but the rare of at low frequency taxa provide much less information. Furthermore, in order to use taxa as indicators, it would be necessary to chose taxa at high taxonomic level (i.e. family) widespread in the Italian rivers, in order to ensure the transferability of the predictive models.

For what concerns the macrobenthos community, there's a huge amount of summary indices among which it was possible to identify 3 categories:

- Richness, diversity and distribution indices;
- Bio-monitoring indices in the strict sense, which represent the indices directly developed for bio-monitoring needs;
- Functional indices or explicitly based on biological and ecological characteristics of organisms (defined as "traits" in the international literature).

It was also decided to give priority to the indices of the STAR-ICMi index (Buffagni and Erba, 2007), which represents the bio-monitoring methodology currently used in Italy for the assessment of the ecological status of watercourses. For what concerns the indices of the first group, they had been selected the family richness and Shannon, Simpson and Margalef indices (Magurran, 2013). Among the biomonitoring indices, the ASPT, EPT, 1-GOLD,  $\log_{10}(\text{EPTD}+1)$  and the LIFE (Extence, Balbi, and Chadd, 1999) indices were selected. As functional indices they were used the richness, dispersion and evenness indices (Villegier, Mason, and Mouillot, 2008) and an additional index named LENLOT, specifically developed within the Benthab project.

### **LIFE and LENLOT indices**

Among all the considered and calculated indices, that could describe in the right way the the affinity of macroinvertebrate community to flow velocity, two of these were selected as target variables: LIFE and LENLOT. The first is a widely known and easily understandable index, the second index was implemented specially during the data elaboration, so tailored on the collected information.

The LENLOT index is calculated starting from the current velocity affinity values proposed by Tachet et al., 2010 and free downloadable on the site [www.freshwaterecology.info/abbreviations.php](http://www.freshwaterecology.info/abbreviations.php). The Tachet's database contains information about biological, ecological and functional characteristics of, among the others, 500 macroinvertebrates taxa. The method suggested by Tachet et al. is based on the aim to associate a numeric value to the biological and ecological features of sampled taxa. In order to do this, each *trait* is composed by a number of categories. A fuzzy logic assigns to each category a score that describes the affinity of the taxon for the category of the trait. This coding method, allows to take into account at the same time the biological, ecological and functional characteristics of the considered taxonomic level.

In the case of LENLOT calculation, the considered trait is the current velocity, composed by 4 affinity categories: null, slow, medium and fast. The scores of the taxa are ranged from "0" indicating *no affinity* to 3 indicating *high affinity*. Within the calculation of LENLOT index, the scores were normalized, by dividing the values by the sum of the corresponding raw, in order to normalize the final value, between 0 and 1, and facilitate the interpretation of the



result. Then, an average of the normalized scores of each category (*community weighted mean*) was operated, and the sum of the averages for fast and medium categories of velocity was divided by the sum of the averages of all the velocity categories, obtaining the LENLOT index value. Each row of the scores table refers to a species, and the calculated index refers to the community, composed by the species involved in the calculation.

The LIFE (*Lotic-invertebrate Index for Flow Evaluation*) "method is primarily based on recognized flow associations of different macroinvertebrate species and families" (Extence, Balbi, and Chadd, 1999). Its calculation is based first of all on the classification of the collected samples, at the species or family level, and then, on the *ecological flow association* to each recognised taxa: the main goal is to individuate which degree of preference characterize each taxa.

Group	Ecological flow association	Mean current velocity
I	Taxa primarily associated with rapid flows	Typically $> 100 \text{ cm s}^{-1}$
II	Taxa primarily associated with moderate to fast flows	Typically $20\text{--}100 \text{ cm s}^{-1}$
III	Taxa primarily associated with slow or sluggish flows	Typically $< 20 \text{ cm s}^{-1}$
IV	Taxa primarily associated with flowing (usually slow) and standing waters	—
V	Taxa primarily associated with standing waters	—
VI	Taxa frequently associated with drying or drought impacted sites	—

FIGURE 2.17: Benthic freshwater macroinvertebrate flow groups, ecological associations and defined current velocities (Extence, Balbi, and Chadd, 1999)

Subsequently, the need is to estimating the abundance of each species or family, and, consequently, obtaining the *flow scores*.

Category	Estimated abundance
A	1–9
B	10–99
C	100–999
D	1000–9999
E	10 000 +

FIGURE 2.18: Standard EA macroinvertebrate abundance categories (Extence, Balbi, and Chadd, 1999)

Flow groups		Abundance categories			
		A	B	C	D/E
I	Rapid	9	10	11	12
II	Moderate/fast	8	9	10	11
III	Slow/sluggish	7	7	7	7
IV	Flowing/standing	6	5	4	3
V	Standing	5	4	3	2
VI	Drought resistant	4	3	2	1

FIGURE 2.19: Scores ( $f_s$ ) for different abundance categories of taxa associated with flow groups I–VI (Extence, Balbi, and Chadd, 1999)

Thus, the index of the sample is calculated using the following formula:

$$LIFE = \frac{\sum f_s}{n}$$

Where  $\sum f_s$  is the sum of all of the flow scores of the whole sample.

The choice of the LIFE and LENLOT indices as target variables for the prediction model building, can be found in the Spearman correlation between the physical parameters (current velocity, water depth) measured in the sampling points and the described indices. The visualization of that correlation follows.

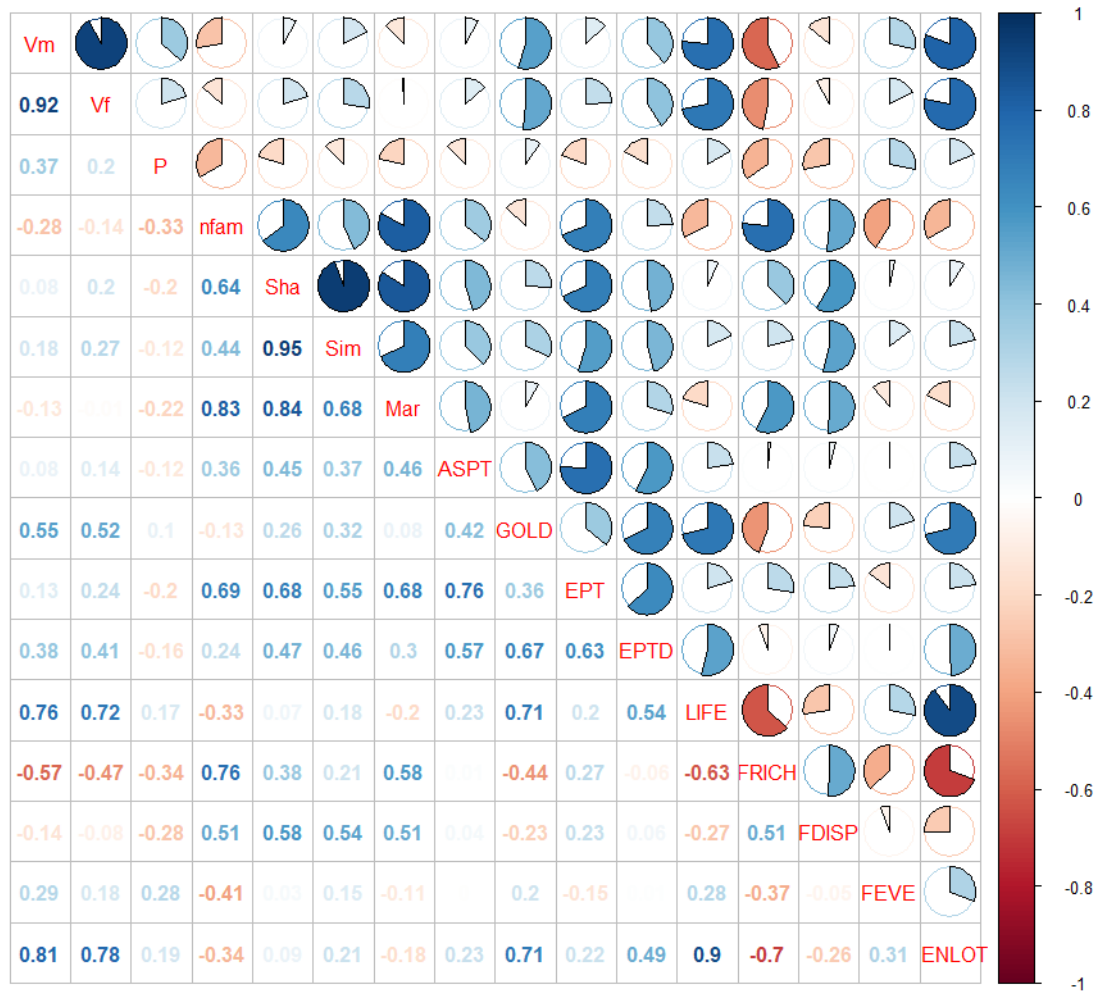


FIGURE 2.20: Correlation between the studied indices and the mean velocity ( $V_m$ ), bed velocity ( $V_f$ ) and water depth ( $P$ ). The represented indices are: number of families (nfam); Shannon (Sha), Simpson (Sim) and Margalef (Mar); ASPT, 1-GOLD, EPT,  $\log_{10}(\text{SeLEPTD}+1)$  and LIFE; richness, dispersion and equiripartizione funzionale; LENLOT index.

In the previous figure it's possible to notice that the most important relationships are those between the velocity ( $V_m$  mean velocity and  $V_f$  bed velocity) and LENLOT ( $\rho = 0.81$  and  $\rho = 0.72$ ) and LIFE ( $\rho = 0.76$  and  $\rho = 0.78$ ). For what concerns the water depth and its correlation with the indices, there are values of  $\rho$  coefficient significantly lower: never more than 0.35. The best performance of both LIFE and LENLOT indices are due to their correlation with the current velocity.

One of the issue to be taken care of is the *upscaling* from the micro-scale, at which the sampling is carried out, to the meso-scale, at which the model has to be developed. This upscaling has to be operated in order to associate to the HMUs (meso-scale) a averaged of the variable values punctually measured

within the HMU itself (micro-scale). In the case of this elaboration, the up-scaling was realized operating an average of the calculated punctual indices, for each HMU. This operation was made possible thanks to the georeferencing of each biological sampling point, performed during the field surveys.

### 2.5.2 Introduction to machine learning techniques

Machine learning techniques enable systems to learn from experience and improve their ability to observe data and informations learning from them, rather than being programmed with that knowledge.(Shapiro, 1992). Machine learning is widely used in many disciplines: commercial and statistical systems, programming and model building, and, as the aim of this thesis, computer science applied to ecology. These methods intelligently record and analyse collected data, in order to acquire new knowledge about the behaviour and the organization of that data. They're able to understand the ranking of importance of the input variables, so that the user is able to decide which ones worth to be used or to be excluded from the dataset. The importance of this approach is the possibility to continuously improve efficiency and effectiveness.

### 2.5.3 Random Forest

Random forest is a widely used supervised machine learning algorithm for classification, regression and other tasks that operates by constructing a multitude of decision trees (Yiu, 2019). In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making (Choudary, 2020. An example follows: at each node there's a splitting that divides the node itself in two or more sub-nodes based on if-else conditions. The process continues until there's a *terminal node* where it cannot be split.

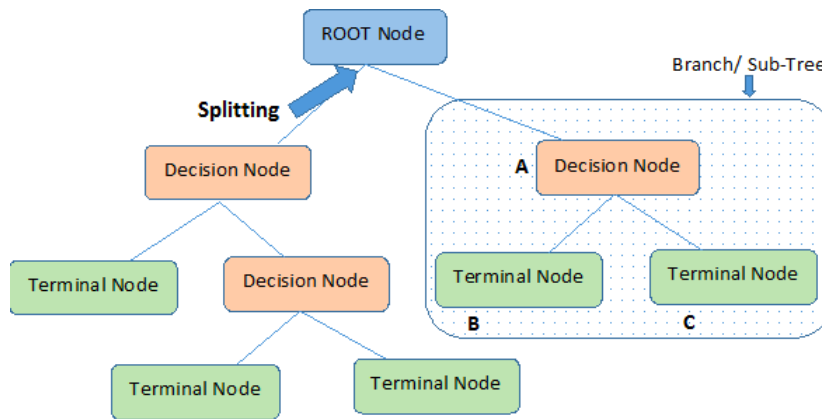


FIGURE 2.21: Structure and functioning of a decision tree (Choudary, 2020)

The "forest" it builds, is an ensemble of these trees, usually trained with the *bagging* method. This is a technique that combines the predictions from multiple machine learning algorithms together to make more accurate predictions than any individual model. Bagging (or Bootstrap Aggregation) is a general procedure that can be used to reduce the variance of those algorithms, such as Classification and Regression Trees (CART), used in this case for the biological models building. One big advantage of random forest is that it can be used for both classification and regression problems, which form the majority of current machine learning systems. The classification method aims to find the belonging class of the test parameter, considering the majoring vote of all the decision trees. Instead, the regression process involves the calculation (prediction) of the test parameter by operating an average of all the results coming from the decision trees of the forest. This type of data prediction is based on an input dataset, called *training data* and made of the measured values of the target variable.

In the case of the study, the test parameter is represented by the biological indices, and the needed method is the regression one, operated through R coding based on the Random Forest library package. The open source compiler used for the program coding is R studio. So, the main difference between the existing Habitat Suitability Models, exploit by the MesoHABSIM approach, and the "BentHab" prediction models is the RF method used within the building: in the first case, it is the *classification* method, instead of the *regression* one, like in the case of this study.

#### 2.5.4 Model calibration and cross-validation

The selection of the variables dataset used to build the parsimonious biological model, for both LIFE index and LEN-LOT indicator, has been lead following two main criteria. The first is based on the correlation coefficient matrix: this instrument allows to understand the degree of correlation between the variables, in order to avoid including two or more correlated variables among the prediction data. It was fixed a threshold of the correlation coefficient equal to 0.8: all the variables characterized by correlation less than that value are considered not much correlated. Another criteria on which the selection of variables is based is the ranking of variable importance. This variable importance depends on the MSE and means how much a variable influences the prediction of target variables.

One of the most important advantages of the random forest prediction method is the *cross-validation* functioning, that automatically allows the user to obtain a validation of the created model. This is, precisely, a validation of the model based on a subset randomly selected by the system itself: 40% of the prediction data provided by the user are not used as training data but "kept out of the bag" in order to operate a first cross-validation of the model results. This subset is in fact named Out of Bag (OOB) and it's a particularity of this machine learning method. That's not a real validation of the model because the environmental conditions are the same for both the training and the OOB subset, so that a validation based on a different dataset is needed. The field surveys on Taro and Enza rivers for the purpose to collect these data were already carried out and the elaboration will be lead in Autumn 2020.

## Chapter 3

# Results

### 3.1 Hydromorphological description

#### 3.1.1 Granulometric size distribution

The granulometric analysis provided as a main result the granulometric curves of the emerged zones of the Trebbia river reach under study.

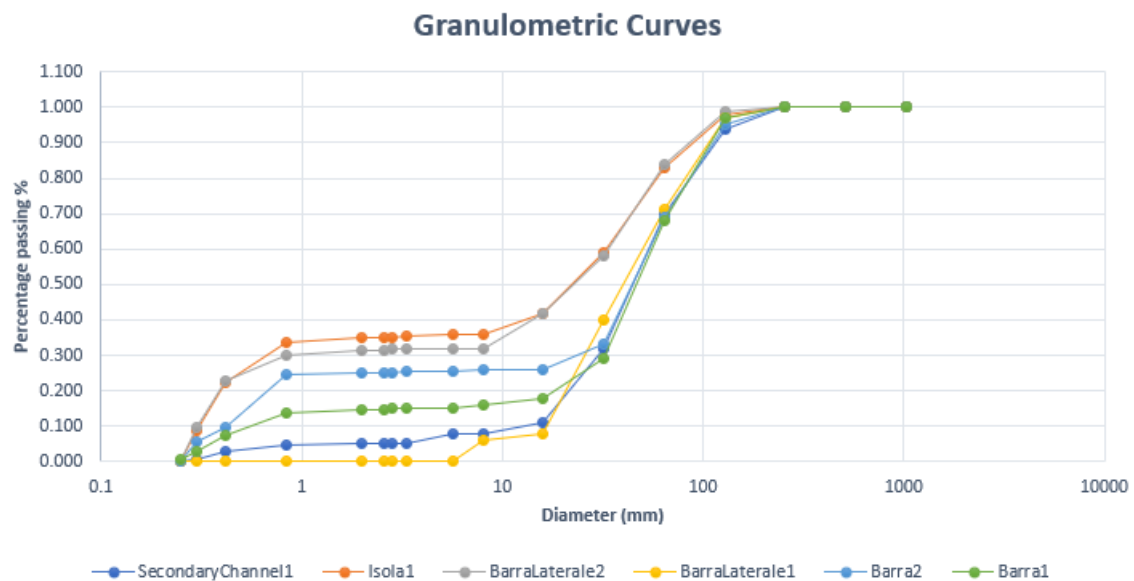


FIGURE 3.1: Granulometric curves resulting from the substrate composition analysis of the emerged zones of Trebbia river.

Another important result of this type of elaboration is the value of the Manning coefficient, of fundamental importance for the hydrodynamic simulation building. This optimal value was fixed at  $0.036 \text{ m}^{-1/3}\text{s}$ .

#### 3.1.2 GIS outputs

The first GIS analysis results obtained from the elaborations are the depth and velocity values distributions, collected by the ADCP and recorded by

the River Surveyor software. The raw dataset involves the data collected with a frequency of acquisition of 4 seconds, which corresponds to the acquisition frequency of the antenna. The measured distributions are shown in the following figures.

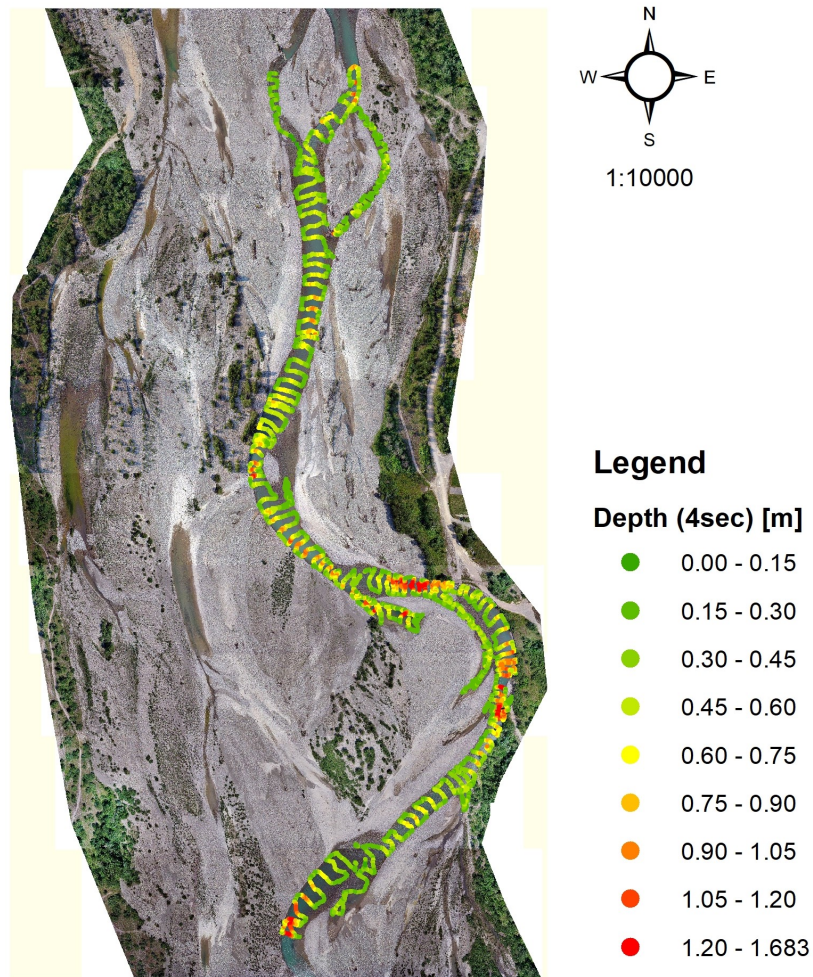


FIGURE 3.2: Water depth distribution, acquired by ADCP and current meter, imported and visualized on GIS software. The acquisition frequency is 4 seconds.



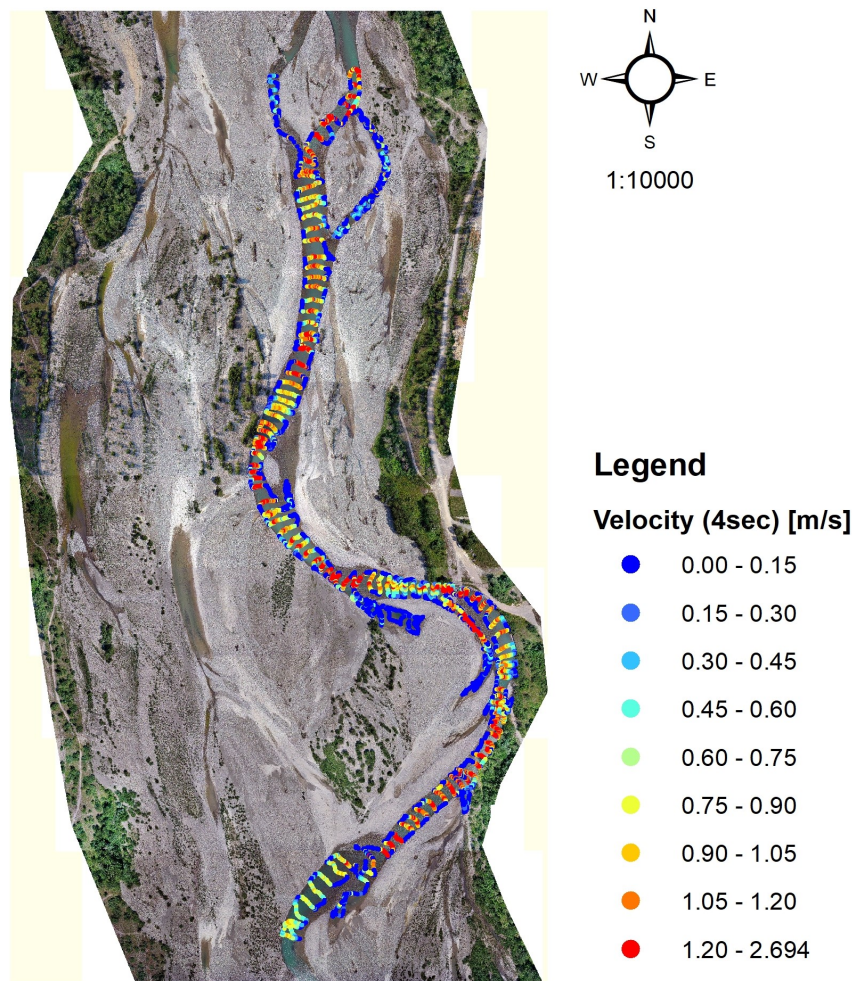


FIGURE 3.3: Current velocity distribution, acquired by ADCP and current meter, imported and visualized on GIS software. The acquisition frequency is 4 seconds.

This raw dataset was manually clean and selected, in order to eliminate some eventual outliers or acquisition errors. This corrected distributions were implemented in order to obtain a more complete dataset characterized by 1 second of acquisition frequency. This operation was made possible thanks to the frequency of recording of the ADCP (1 second) and the possibility of the same instrument to collect the tracking of the boat relative to the initial measure point. This interpolated dataset was used as term of comparison with the simulated dataset obtained from the hydrodynamic simulation. The GIS visualization of the 1 second distribution follow. In the following figures is possible to notice the integration of the data coming from the bathymetric survey with the data collected with the current meter.

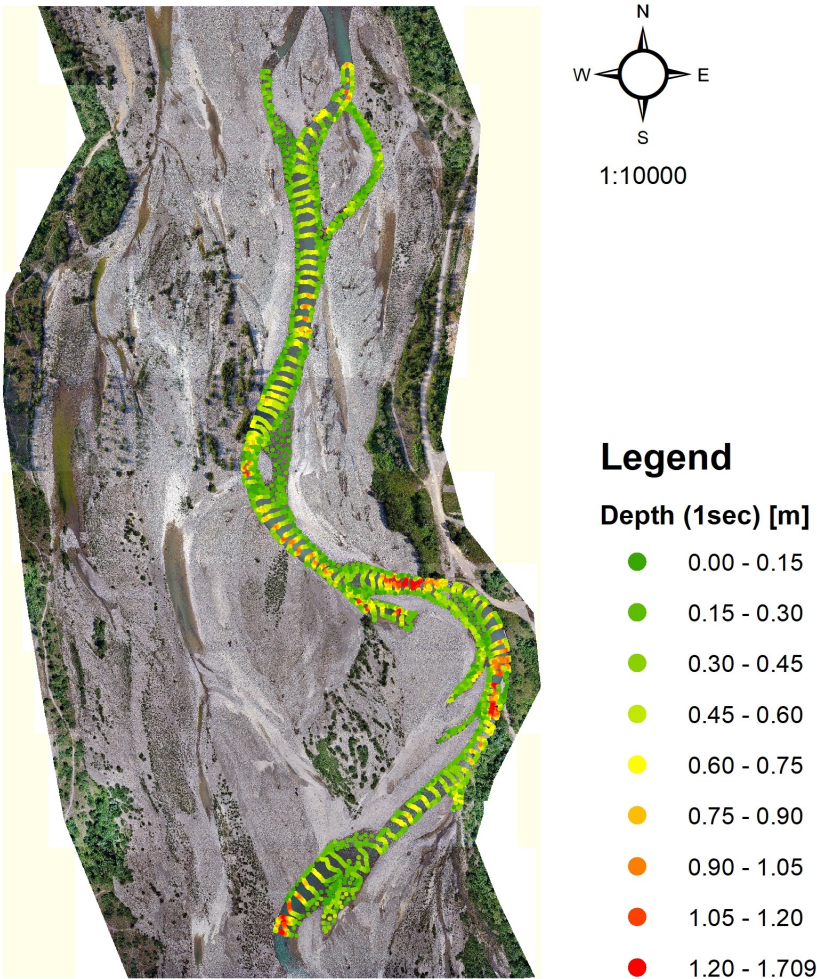


FIGURE 3.4: Water depth distribution of the dataset characterized by 1 second of acquisition frequency, resulting from the interpolation of the raw dataset

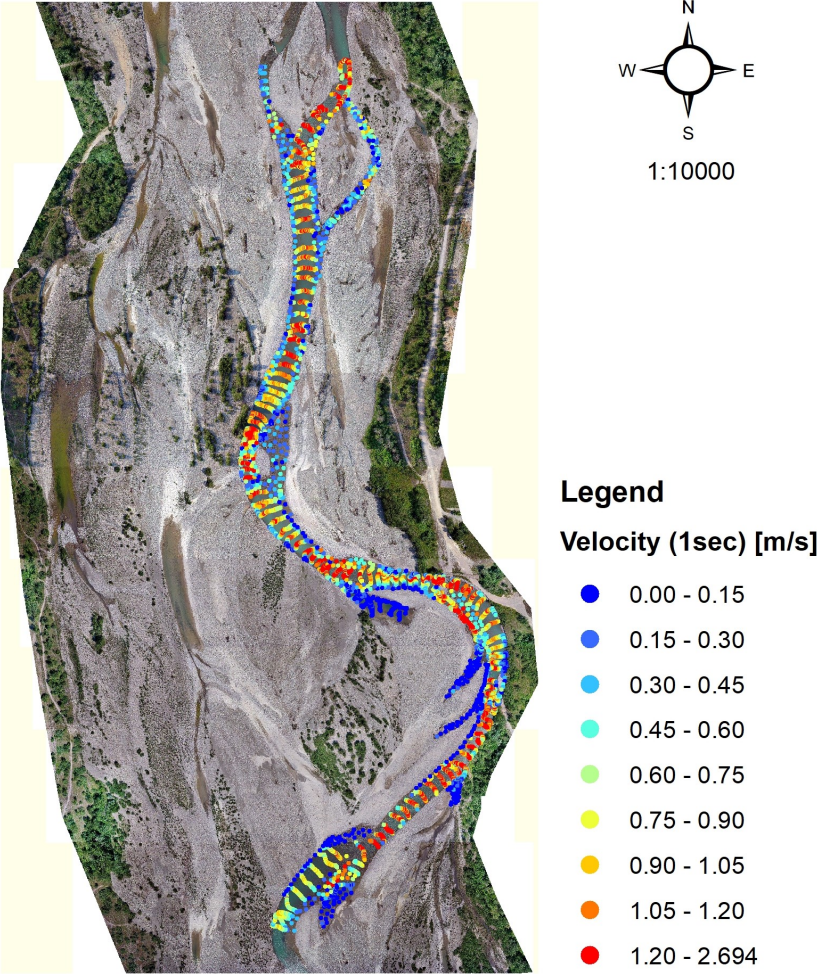


FIGURE 3.5: Current velocity distribution of the dataset characterized by 1 second of acquisition frequency, resulting from the interpolation of the raw dataset



Another important result coming from the GIS elaborations is the distribution of the Hydromorphological Units (HMU), shown in the following figure.

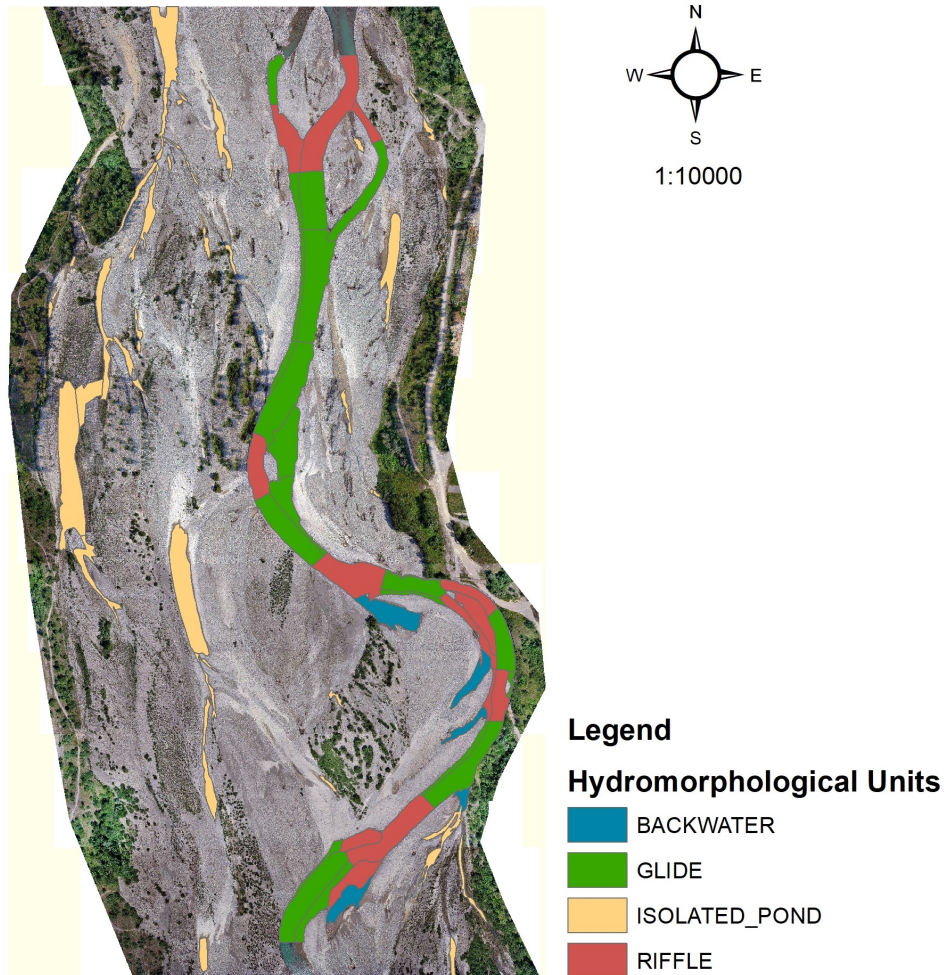


FIGURE 3.6: Representation of the designed HMUs on the Trebbia river

The application of the correction equation of the riverbed elevation, aiming to obtain the correct water depth values distribution within the submerged zones, provided the new Digital Terrain Model (DTM), which is shown below.

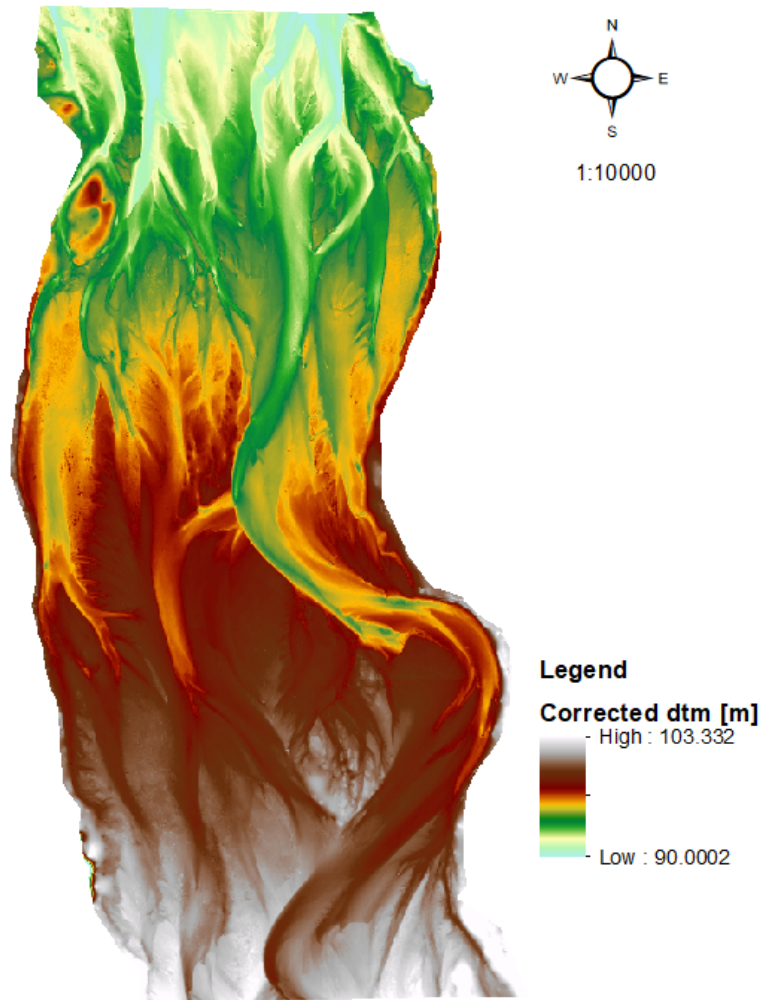


FIGURE 3.7: Correct DTM, obtained thanks to the linear regression equation, that relates the fictitious riverbed elevation (and corresponding water depth) to the real ones. The wet stream bed generally turns to have a lower elevation, so that the corresponding zones of the river are characterized by higher water depth.

### 3.1.3 Hydrodynamic model

The hydrodynamic simulation provided results highly detailed and complete, for both the wet area calculation and the flow field description. A comparison between the simulated wet area and the high resolution orthophoto of the Trebbia river follows.

For what concerns the depth and velocity distribution simulated by HEC-RAS modelling, the resulting rasters are shown in the following pages. It's important to point out the similarity in the distribution between the simulated dataset and the collected one: this was an opportunity to use the first one as input data for the statistical model.

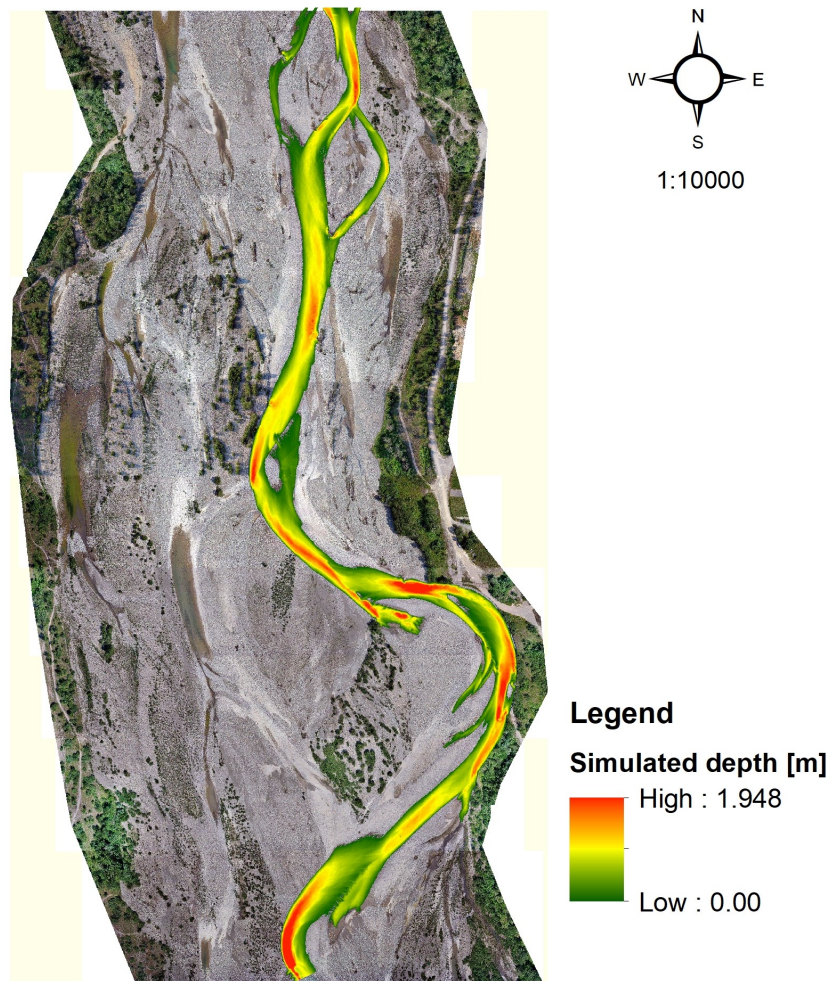


FIGURE 3.8: Water depth distribution raster, coming from the hydrodynamic simulation.



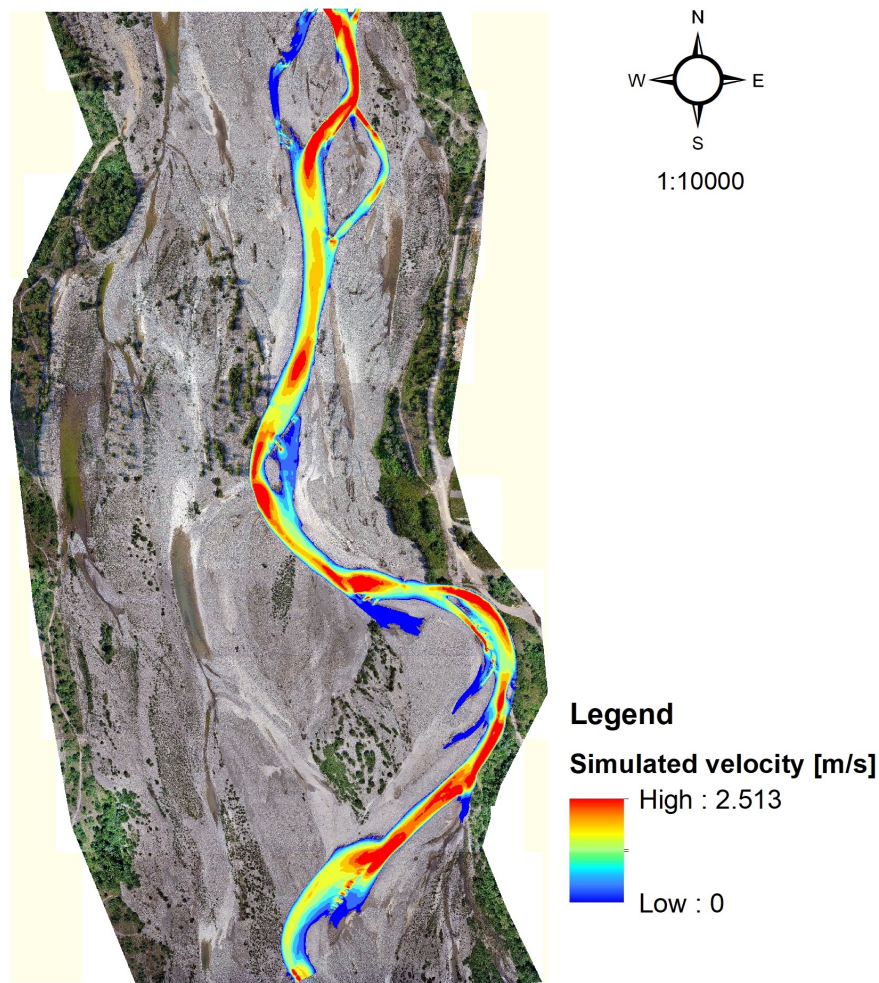


FIGURE 3.9: Current velocity distribution raster, coming from the hydrodynamic simulation.

An instrument to evaluate the similarity between the simulated and measured distributions was the comparison between the frequency histograms, considering 9 frequency classes of depth and velocity. So, histograms of both the distributions follow.

## 3.2 Biological models

The biological models building has lead to obtain knowledge about the structure of the input data and eventual pattern among them. The main outputs useful to understand how the collected data may predict the LIFE and LENLOT indices are:

- Correlation coefficient  $R^2$ , also known as *explained variance* for each prediction model. This coefficient describes how much the model fit the calculated indices.

- Ranking visualization of the variable importance, as a function of the increase of MSE (Mean Squared Error) introduced by keeping every single variable among the input data.
- Error in percentage that the regression model makes as a function of the number of trees created in the random forest: the more populated is the forest, the lower and more stabilized is the error. It allows to set the number of trees, which is one of the required parameters to be entered during random forest programming, in order to obtain a more robust output.
- Partial Dependence Plots, suggested by Breiman, 2001, that represent the partial relationship between the index predicted values and the habitat descriptors. In this case, the habitat descriptors are relative frequencies of each independent variable (current velocity, water depth and substrates distribution).

### 3.2.1 Model calibration

This model was developed using the independent variables (current velocity, water depth, substrates composition and connectivity to the main channel) to predict the LIFE and LENLOT indices values, for each meso-habitat. The following figures show the outputs previously described in the case of both the indices. In particular, here below is shown the summary of the main parameters resulting from the LIFE index prediction model, among which the most important is the *Var explained* ( $R^2$ ).

```
Call:
randomForest(x = pred.data, y = LIFE, ntree = 5000, mtry = m,
  nodesize = 5, importance = T, nPerm = 2, proximity = TRUE, do.
  trace = 50, keep.forest = TRUE, data = pred.data, outscale = TRUE)
  Type of random forest: regression
    Number of trees: 5000
No. of variables tried at each split: 10

  Mean of squared residuals: 0.2884163
    % var explained: 46.55
```

FIGURE 3.10: Output of the regression model predicting the target variable LIFE index. *Var explained* represents the  $R^2$  correlation coefficient equal to 46.55 %

The following figure shows the ranking of the variables importance, successively used to select the most representative parameters to be involved in the parsimonious biological model for the LIFE index.



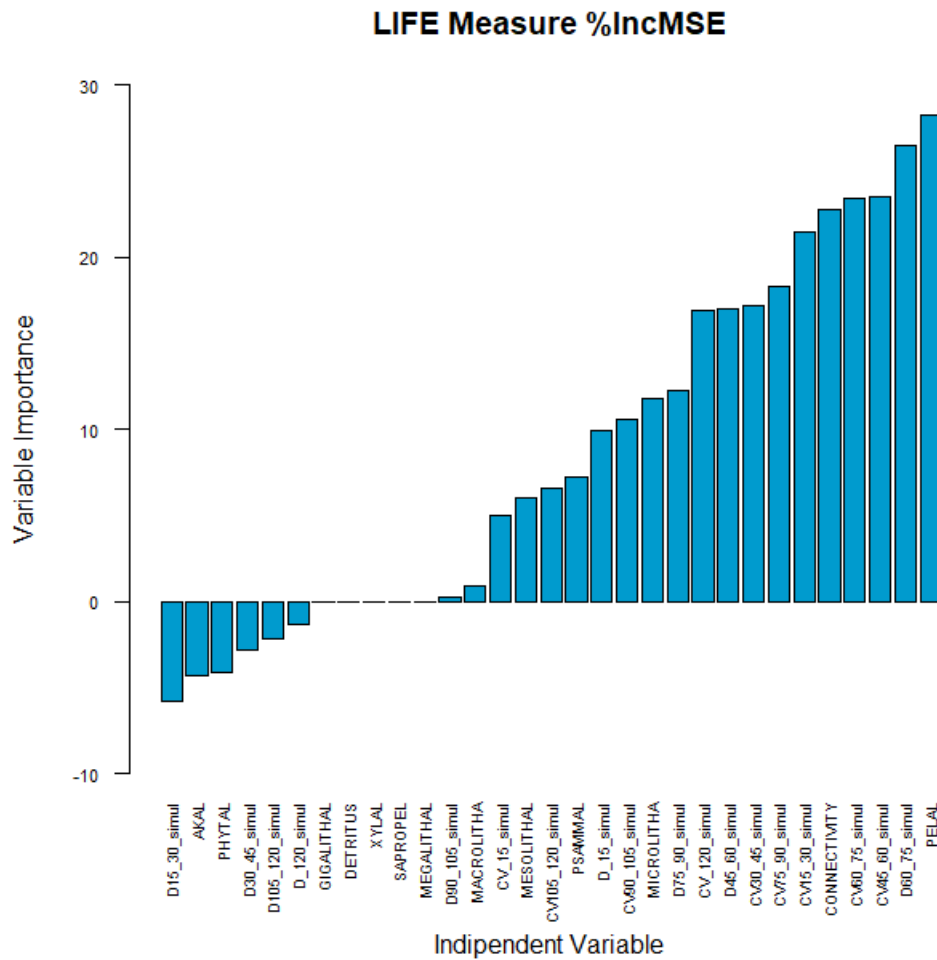


FIGURE 3.11: Ranking of statistical importance of independent variables, using LIFE index as target variable. The importance of the variables is related to the percentage of increasing in MSE (*Mean Squared Error*). This importance is a measure of how much including a variable increases accuracy of the model.

It's also reported the behaviour of the error committed by the model in the prediction, as a function of the number of regression trees. It is easy to note the decreasing function: the selected number (5000) is widely sufficient to ensure the minimum error in the random forest regression.

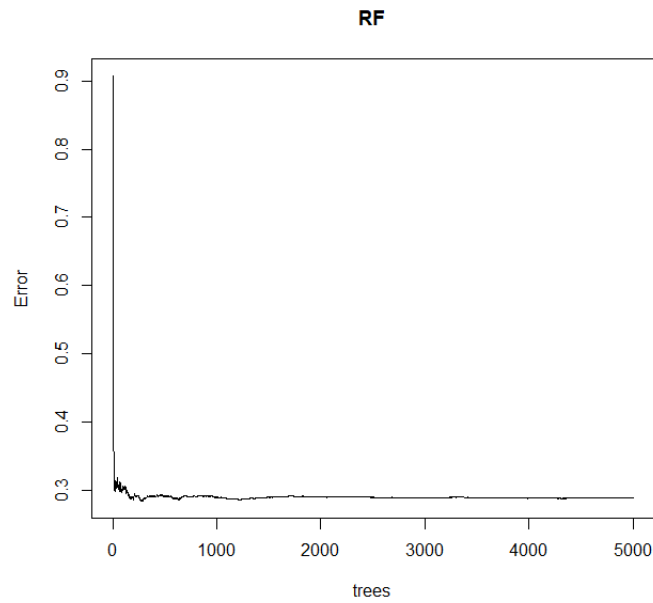


FIGURE 3.12: Error committed in the prediction of target variable LIFE index, as a function of the number of trees creating the random forest. Increasing the number of trees, the error percentage tends to become constant.

The most important results of the RF modelling are the Partial Dependence Plots, previously described. The following figures represent these particular type of plots referred to the LIFE index, as a function of the water depth, of the current velocity, the substrates distribution and the connectivity to the main channel.

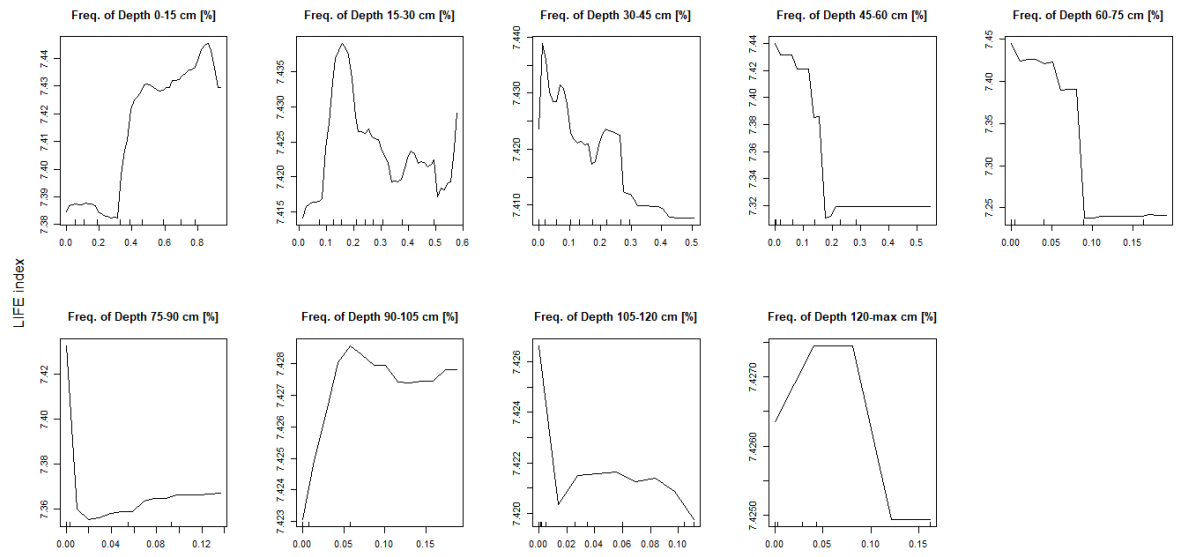


FIGURE 3.13: Partial Dependence Plots of the 9 classes of depth, as independent variables in the LIFE index prediction model.

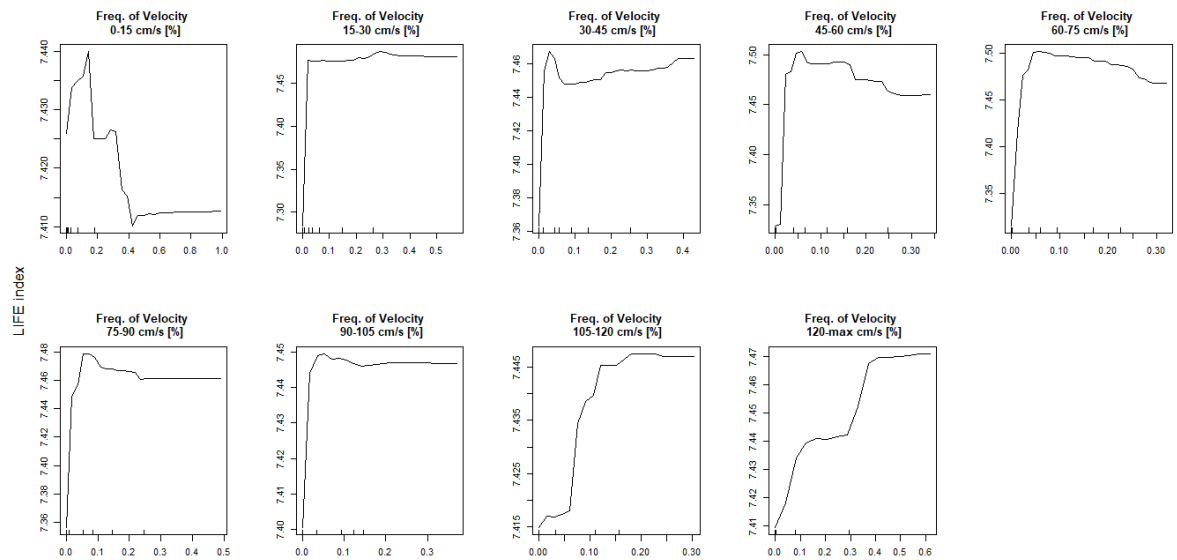


FIGURE 3.14: Partial Dependence Plots of the 9 classes of velocity, as independent variables in the LIFE index prediction model.

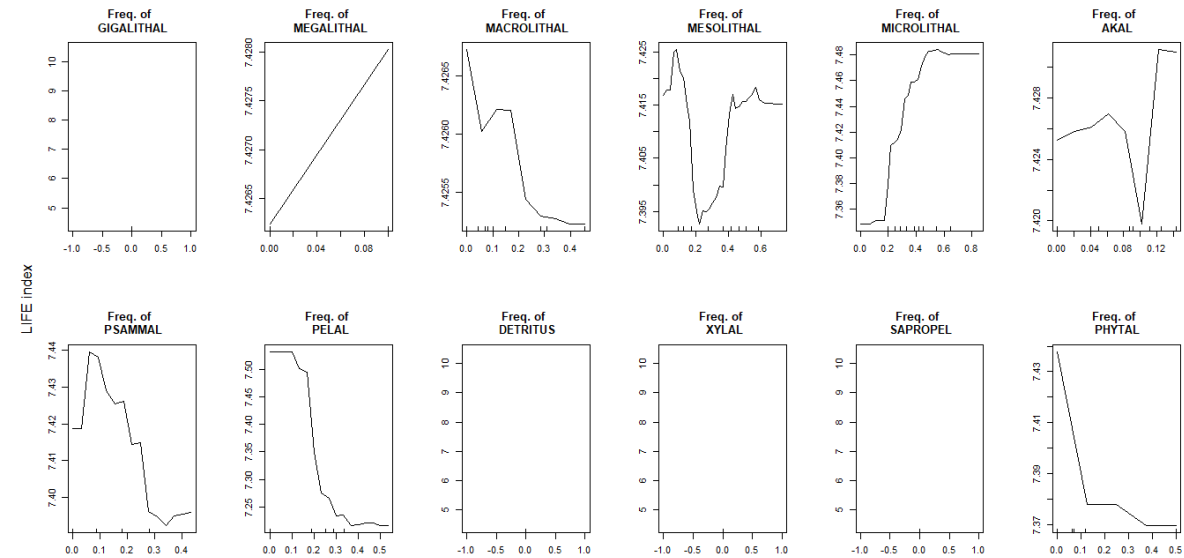


FIGURE 3.15: Partial Dependence Plots of the classes of substrate composition, as independent variables in the LIFE index prediction model.

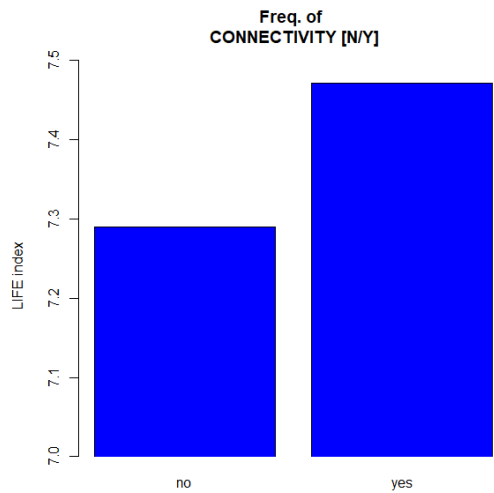


FIGURE 3.16: Partial Dependence Plots of the discrete variable *Connectivity* [YES] or [NO], in the LIFE index prediction model.

This last PDP, related to the connectivity variable, is different from the others because the variable is categorical.

The same results of the biological model are reported referring to the LENLOT index, characterize by an higher value of correlation coefficient  $R^2$ .

```
call:
randomForest(x = pred.data, y = LEN_LOT, ntree = 5000, mtry = m,          nodesize = 5,
importance = T, nPerm = 2, proximity = TRUE,          do.trace = 50, keep.forest = TRUE,
data = pred.data, outscale = TRUE)
Type of random forest: regression
Number of trees: 5000
No. of variables tried at each split: 10

Mean of squared residuals: 0.005714807
% var explained: 69.16
```

FIGURE 3.17: Output of the regression model predicting the target variable LENLOT index.  $R^2 = 69.16\%$

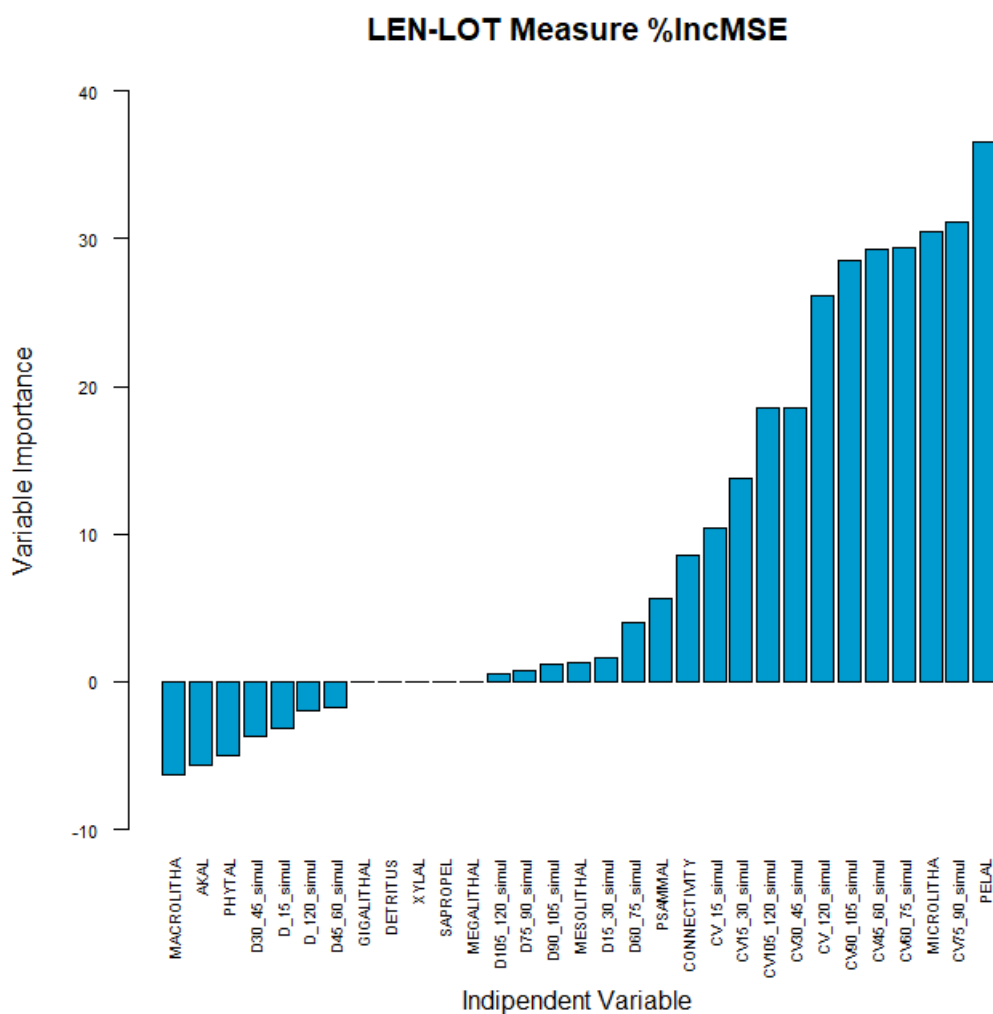


FIGURE 3.18: Ranking of statistical importance of independent variables, using LENLOT index.

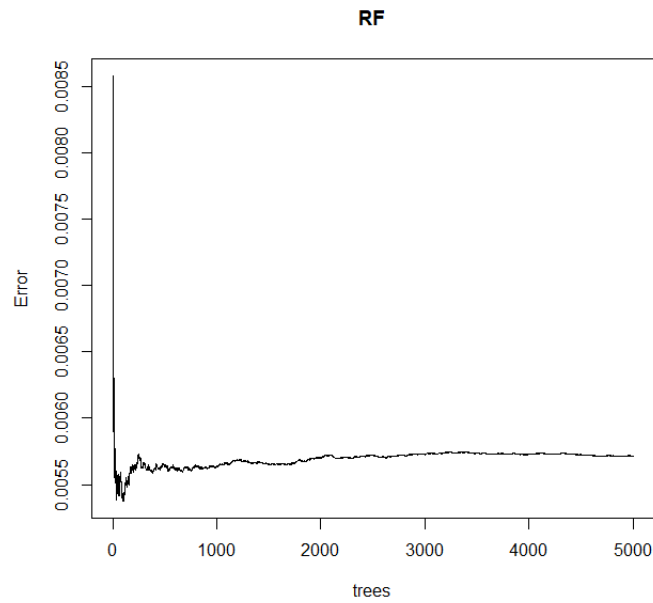


FIGURE 3.19: Error committed in the prediction of target variable LENLOT index, as a function of the number of trees creating the random forest.

Below are shown the PDP of the LENLOT index.

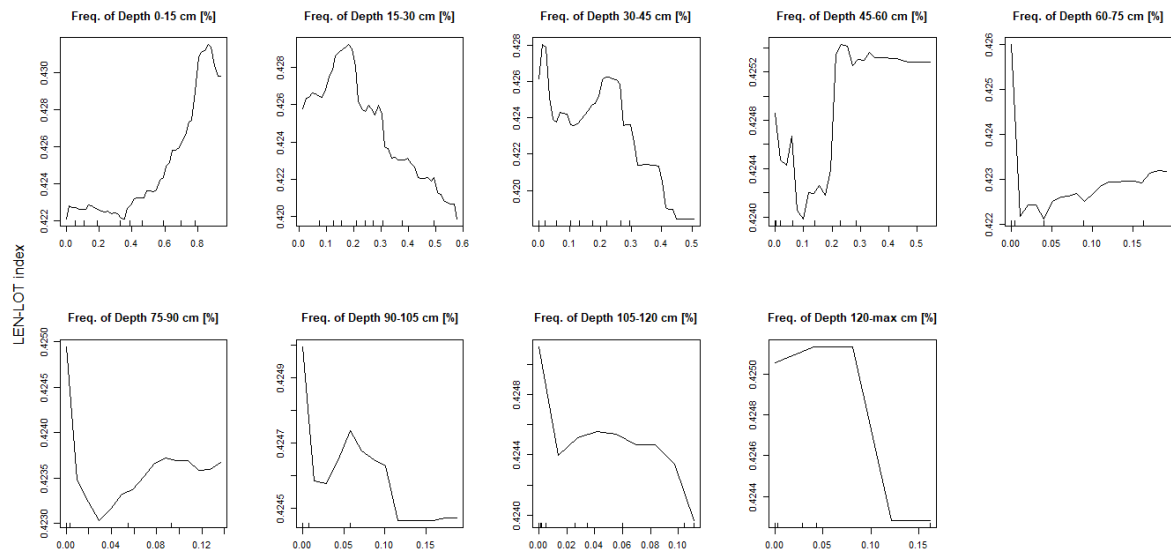


FIGURE 3.20: Partial Dependence Plots of the 9 classes of depth, as independent variables in the LENLOT index prediction model.

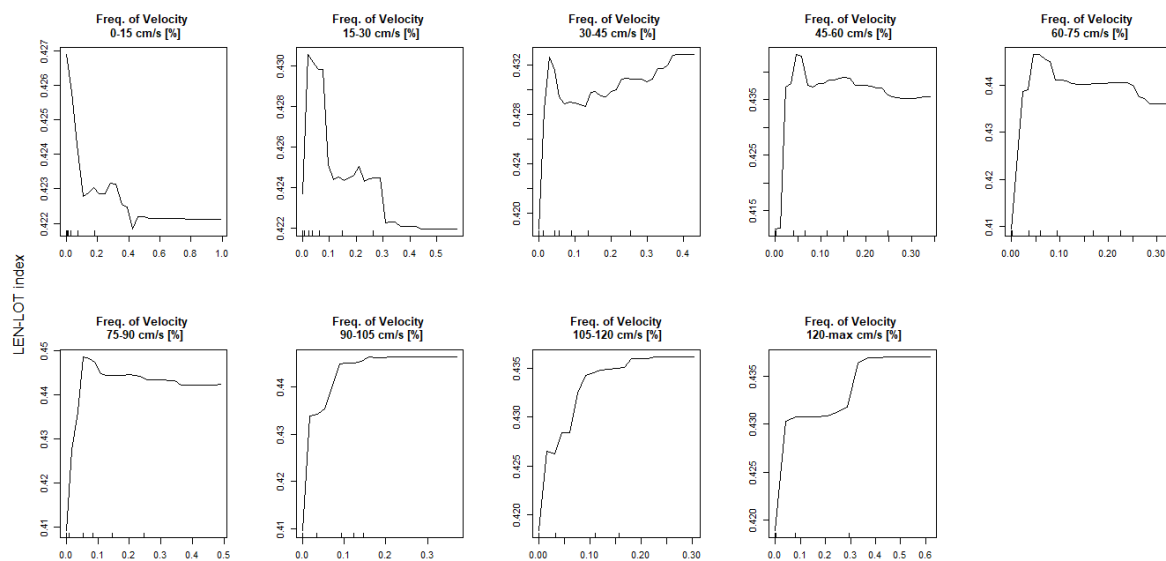


FIGURE 3.21: Partial Dependence Plots of the 9 classes of velocity, as independent variables in the LENLOT index prediction model.

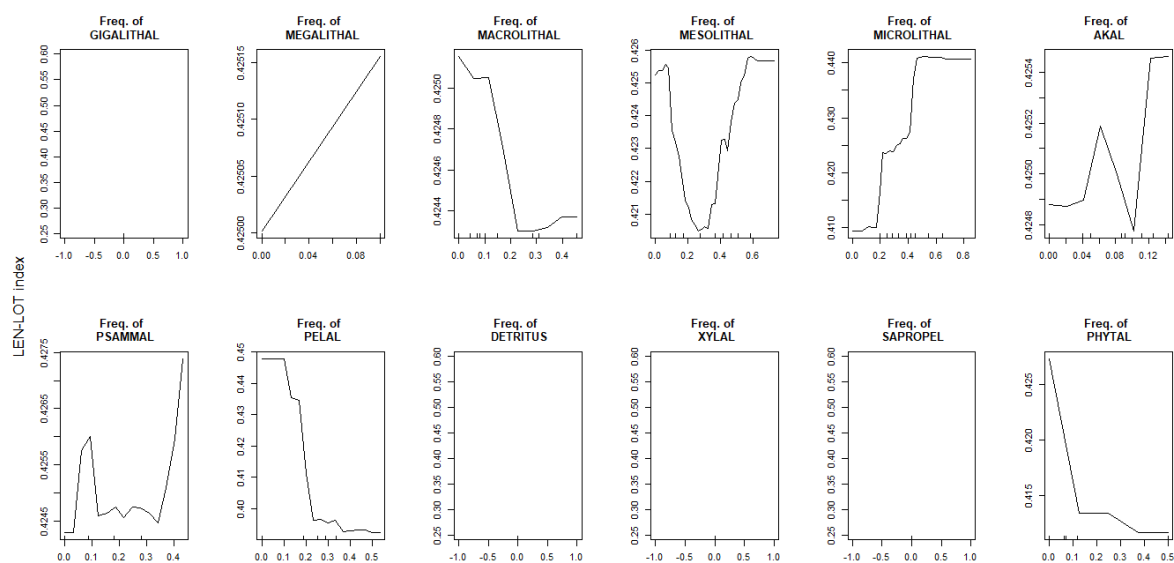


FIGURE 3.22: Partial Dependence Plots of the classes of substrate composition, as independent variables in the LENLOT index prediction model.

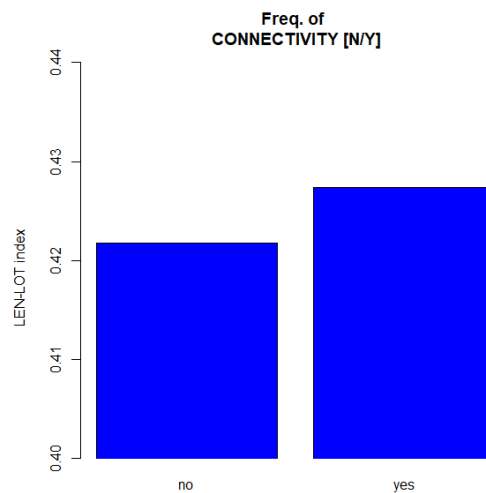


FIGURE 3.23: Partial Dependence Plots of the discrete variable *Connectivity* [YES] or [NO], in the LENLOT index prediction model.

### 3.2.2 Parsimonious model selection

The RF method generates all the possible models involving all the different combinations of independent variables. The best of these models was selected taking into account the best predictive power, in terms of determination coefficient  $R^2$ , calculated in cross-validation, and the thrift of independent variables (Vezza et al., 2014, Vezza, Ghia, and Fea, 2015). In this way, it is possible to isolate and ecologically interpret those variables that mainly influence the spatial distribution of the macroinvertebrate community. Among the criteria of choice of the "parsimonious" independent variables involved in the final selected model, there's the correlation degree between the variables themselves. In the following figure is represented the correlation matrix, built using the Spearman method for the calculation of the correlation coefficient. A threshold of tolerance of the degree of correlation was fixed to 0.8, so that the variables less correlated are considered not sufficiently linked to be excluded from the dataset. The coefficient was calculated as absolute value, in order to take into account also the negative correlations between the variables.



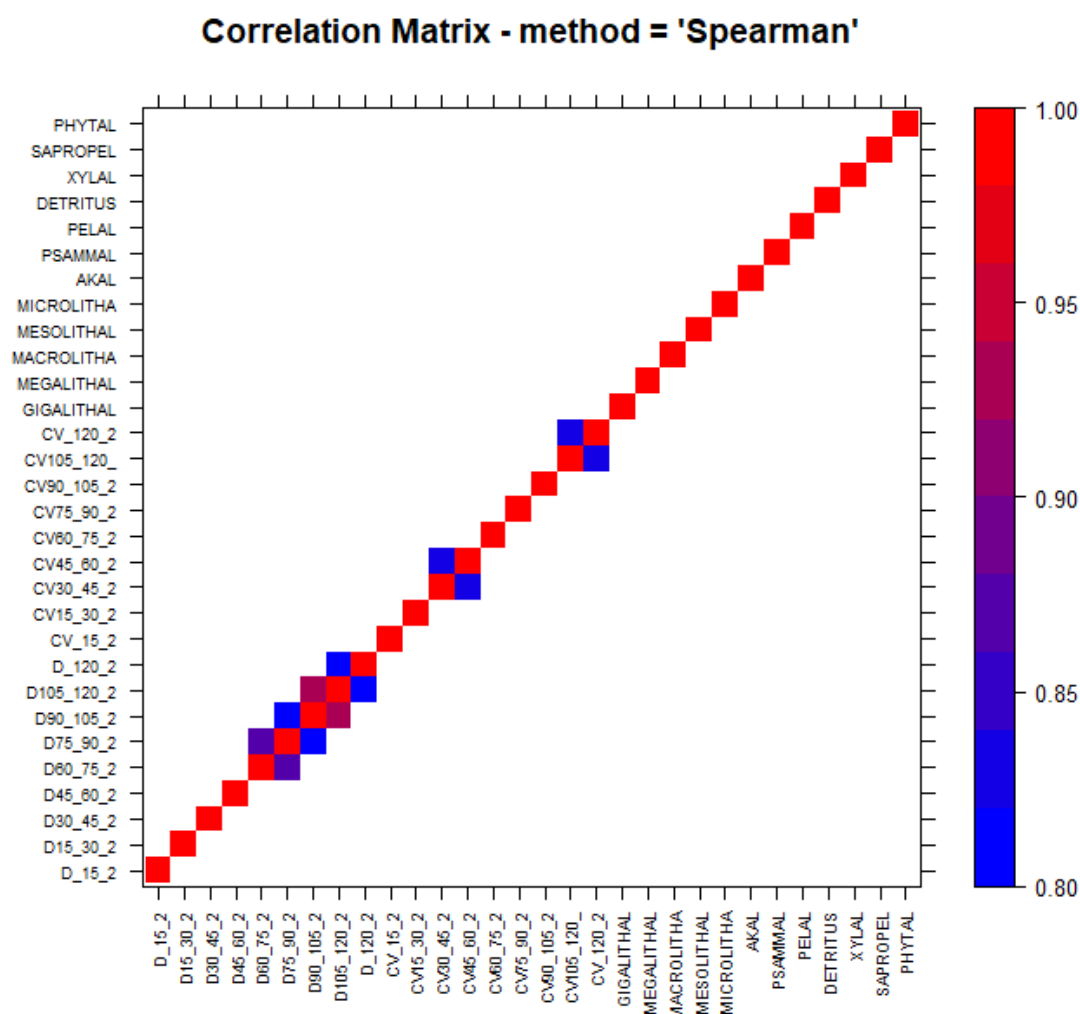


FIGURE 3.24: Correlation matrix of the entire variables dataset, used to build the generic biological model. It graphically represents the correlation between variables, calculated with the Spearman method (0.8 - 1)

As an example the class of current velocity CV45\_60 was not considered in the group of the parsimonious variables for the LIFE model, because of its correlation degree greater than 0.8 with the class CV30\_45. The same rule was applied in the case of the LENLOT regression model between the correlated variables CV105\_120 and CV\_120, the last included in the parsimonious model.

In the following figures are shown the main outputs of the LIFE and LENLOT indices parsimonious models.

```
Call:
randomForest(x = pred.data, y = LIFE, ntree = 5000, mtry = m,
             nodesize = 5, importance = T, nPerm = 2, proximity = TRUE,
             do.trace = 50, keep.forest = TRUE, data = pred.data, outscale =
             TRUE)
             Type of random forest: regression
             Number of trees: 5000
No. of variables tried at each split: 4

             Mean of squared residuals: 0.2706491
             % var explained: 49.84
```

FIGURE 3.25: Output of the parsimonious regression model predicting the target variable LIFE index.  $R^2 = 49.84\%$

It is immediately possible to note the increase, even if minimal, in  $R^2$  that became 49.84 % in respect of the 46.55 % in the generic model. The reason of the increment is found in the selection of the variables, as described previously, eliminating those ones that are not ecologically significant, or reduce the accuracy of the model. In fact of that, the following figure shows the importance of the included variables.

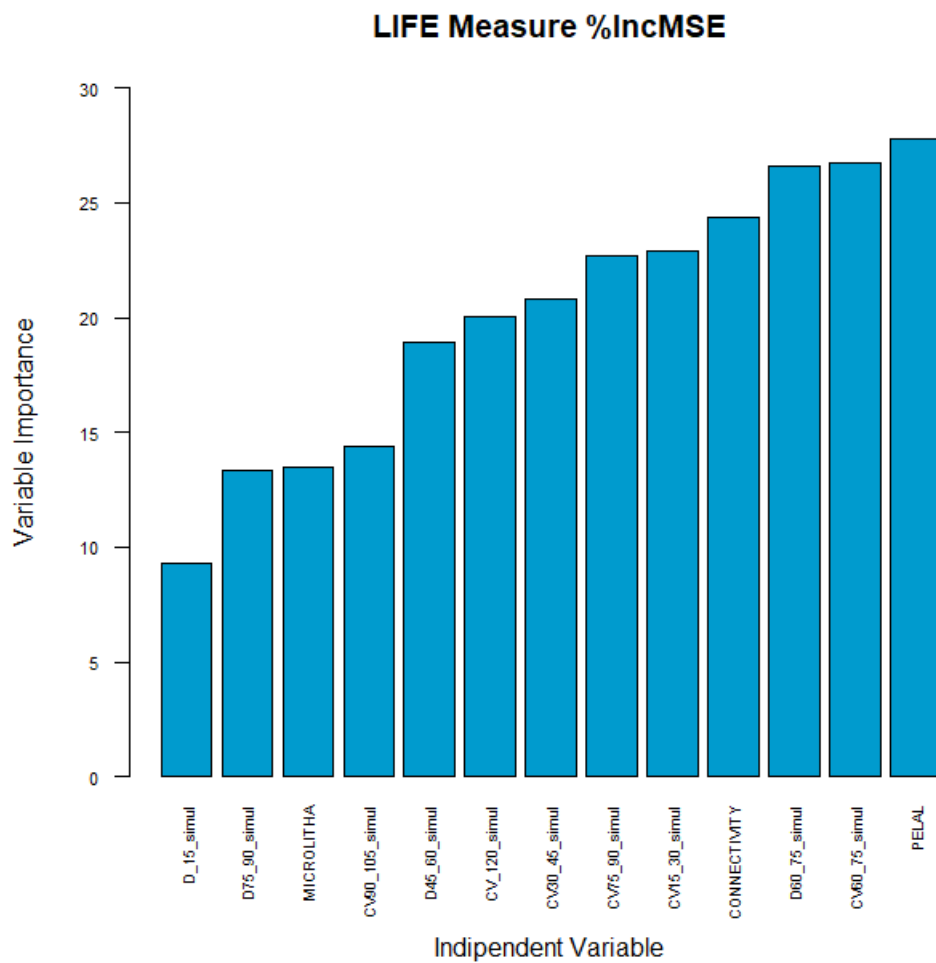


FIGURE 3.26: Ranking of statistical importance of independent variables, using LIFE index as target variable, in the parsimonious regression model. The variables represent the selected dataset, as a results of considerations about the generic model.

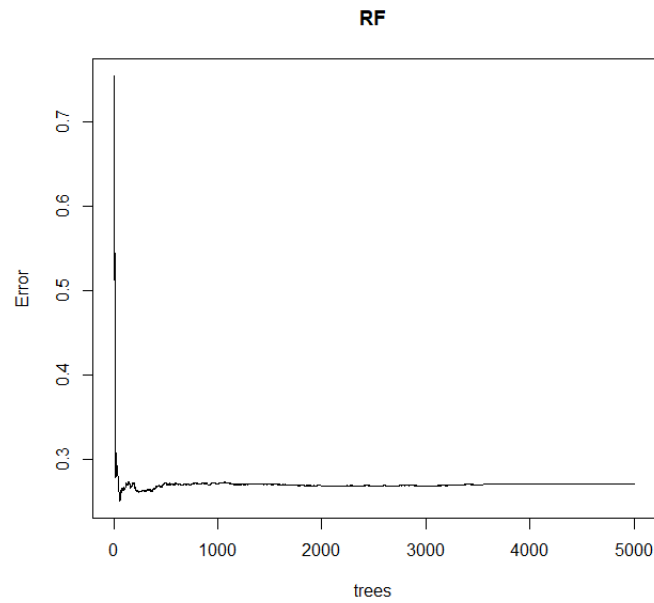


FIGURE 3.27: Error committed in the prediction of target variable LIFE index, as a function of the number of trees creating the random forest in the parsimonious model.

The curves reported in the following PDP (Breiman, 2001) were built for each variable considered as representative, while keeping the others constant. The PDP are not explanatory of the complete behaviour of the model in the multi-space. They can, however, been ecologically interpreted in order to study and understand the relationship between the single habitat variable and the distribution of the entire macrobenthos community.

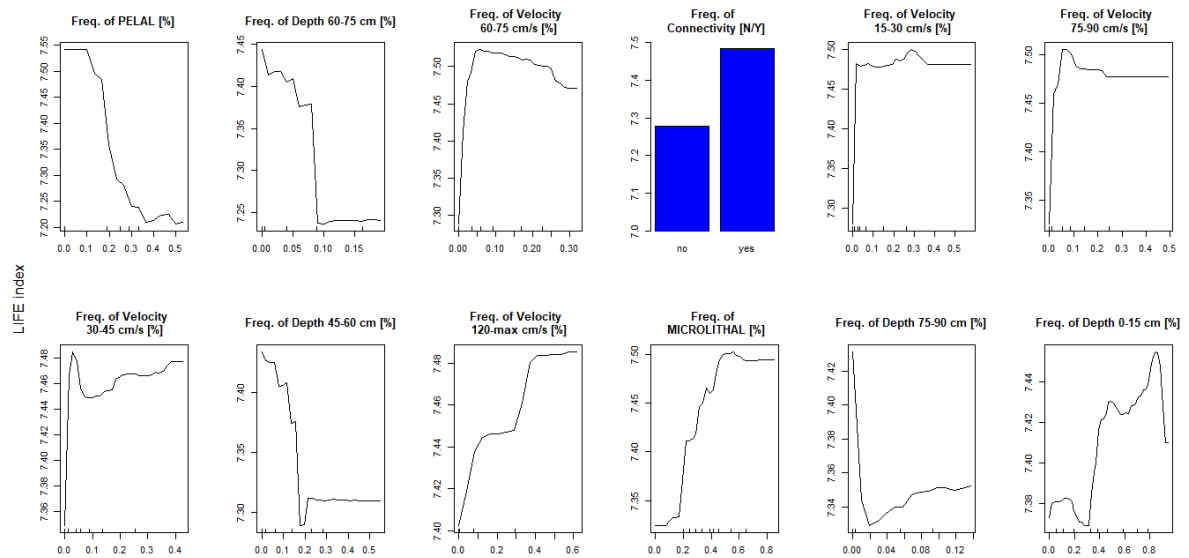


FIGURE 3.28: Partial Dependence Plots of independent variables used to build the parsimonious LIFE index model, selected from the generic model dataset, ordered by importance.

Below are shown the analogous results obtained for the LENLOT regression model.

```
Call:
randomForest(x = pred.data, y = LEN_LOT, ntree = 5000, mtry = m, nodesize = 5,
importance = T, nPerm = 2, proximity = TRUE, do.trace = 50, keep.forest = TRUE,
data = pred.data, outscale = TRUE)
Type of random forest: regression
Number of trees: 5000
No. of variables tried at each split: 4

Mean of squared residuals: 0.00534242
% var explained: 71.17
```

FIGURE 3.29: Output of the parsimonious regression model predicting the target variable LENLOT index.  $R^2 = 71.17\%$

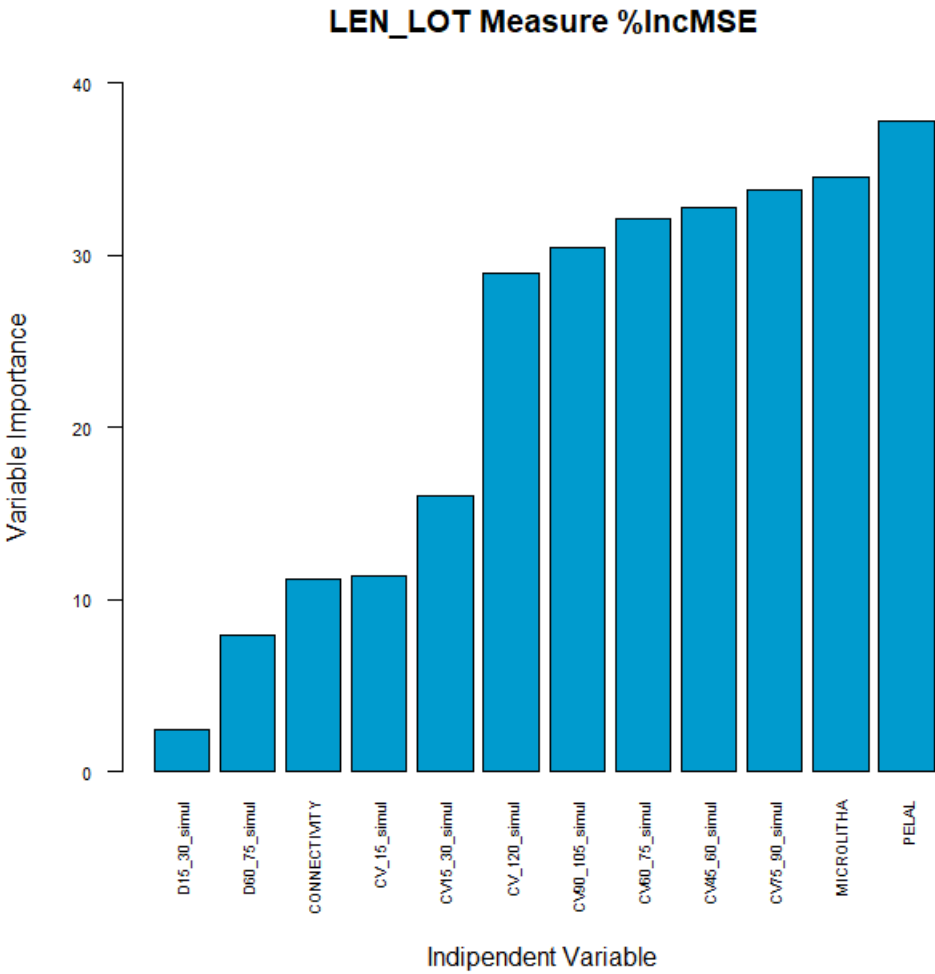


FIGURE 3.30: Ranking of statistical importance of independent variables, using LENLOT index, in the parsimonious regression model

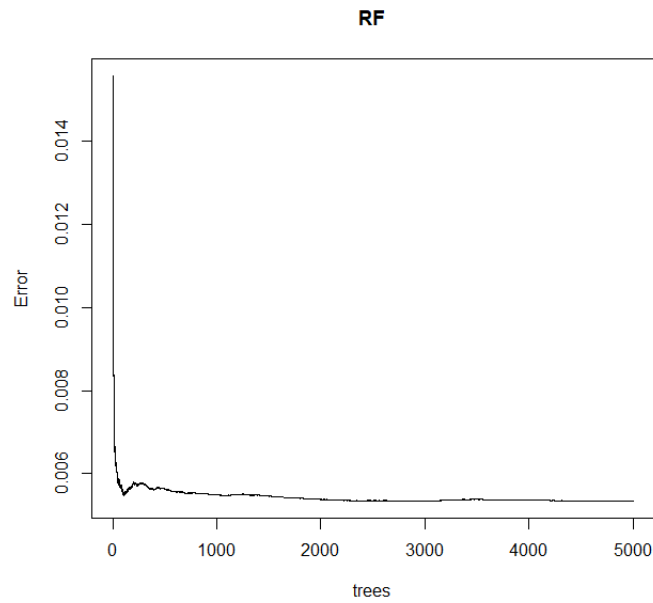


FIGURE 3.31: Error committed in the prediction of target variable LENLOT index, as a function of the number of trees creating the random forest in the parsimonious model.

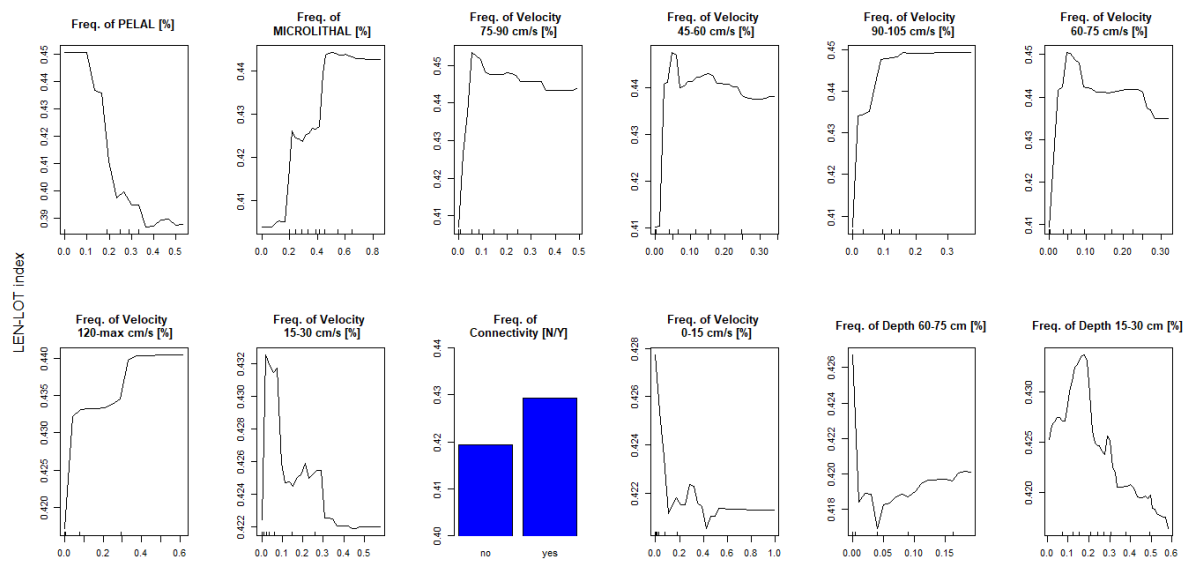


FIGURE 3.32: Partial Dependence Plots of independent variables used to build the parsimonious LENLOT index model, selected from the generic model dataset, ordered by importance.

Finally, they were compared the calculated and the predicted values of both the LIFE and LENLOT indices. In the following table are shown the number, type and name of the sampled HMUs and the calculated and predicted values of both LIFE and LENLOT indices.

HMU_NUM	HMU_TYPE	HMU_NAME	LIFE	LIFE_pred	LEN_LOT	LEN_LOT_pred
1	GLIDE	GLIDE1DX	8.2000	7.8135	0.4161	0.5306
2	GLIDE	GLIDE1	7.9429	7.8722	0.5459	0.4835
3	GLIDE	GLIDE1SX	7.6000	7.7772	0.3992	0.4578
4	RIFFLE	RIFFLE1	7.8833	7.8356	0.5498	0.5422
5	BACKWATER	BACKWATER1	7.1458	7.5494	0.3389	0.3974
6	RIFFLE	RIFFLE3	7.7500	7.8198	0.5674	0.4856
7	RIFFLE	RIFFLE2	8.0278	8.1618	0.6175	0.6103
9	GLIDE	GLIDE2	8.0833	7.6664	0.6410	0.5491
11	BACKWATER	BACKWATER3	7.3323	7.2810	0.3309	0.3269
12	BACKWATER	BACKWATER4	7.3444	7.5509	0.3547	0.4403
13	RIFFLE	RIFFLE4SX	8.0218	7.6383	0.5279	0.4229
14	RIFFLE	RIFFLE4	8.0833	7.9629	0.5436	0.5490
17	RIFFLE	RIFFLE5	7.7333	7.8436	0.5603	0.4980
18	RIFFLE	RIFFLE7	8.0333	7.8525	0.5541	0.5435
19	RIFFLE	RIFFLE6	8.3333	8.0962	0.6379	0.5828
21	GLIDE	GLIDE5	7.5500	7.9320	0.4517	0.5394
22	GLIDE	GLIDE5SX	7.5833	7.5826	0.4633	0.4209
23	RIFFLE	RIFFLE8	7.9306	7.9616	0.5507	0.5631
24	BACKWATER	BACKWATER5	7.4205	6.1828	0.3403	0.2885
25	GLIDE	GLIDE6	8.1714	7.9205	0.5895	0.5339
26	GLIDE	GLIDE6DX	7.8185	7.7433	0.4866	0.4401
27	GLIDE	GLIDE7	7.6556	7.6261	0.4199	0.4632
28	RIFFLE	RIFFLE9	8.4000	8.0375	0.6389	0.5842
29	GLIDE	GLIDE8	7.9014	8.0074	0.5400	0.5457
31	GLIDE	GLIDE9	7.8354	7.9258	0.5366	0.5176
32	GLIDE	GLIDE9SX	7.7407	7.6645	0.4193	0.4394
33	GLIDE	GLIDE10SX	7.5381	7.7681	0.4036	0.4679
34	GLIDE	GLIDE10	7.7767	7.8827	0.4418	0.5560
35	RIFFLE	RIFFLE10	8.2447	8.0876	0.6077	0.6076
36	RIFFLE	RIFFLE11	8.1033	7.8386	0.5991	0.4429
38	GLIDE	GLIDE12	7.6263	7.8753	0.4246	0.5340
39	RIFFLE	RIFFLE12	7.8305	7.9058	0.5569	0.5293



HMU_NUM	HMU_TYPE	HMU_NAME	LIFE	LIFE_pred	LEN_LOT	LEN_LOT_pred
40	ISOLATED_POND	ISOLATED_POND1	4.4444	6.7407	0.2662	0.2903
41	ISOLATED_POND	ISOLATED_POND2	7.3667	6.8329	0.4071	0.2883
42	ISOLATED_POND	ISOLATED_POND3	6.9129	6.1282	0.3107	0.2748
43	ISOLATED_POND	ISOLATED_POND4	6.0000	6.7219	0.1537	0.2538
44	ISOLATED_POND	ISOLATED_POND5	6.1667	6.3072	0.1930	0.2851
46	ISOLATED_POND	ISOLATED_POND7	7.5114	6.8654	0.3412	0.2851
47	ISOLATED_POND	ISOLATED_POND8	7.3472	6.6083	0.3629	0.2742
48	ISOLATED_POND	ISOLATED_POND9	7.3083	6.6439	0.3613	0.2793
50	ISOLATED_POND	ISOLATED_POND11	6.5556	6.6764	0.2230	0.2525
51	ISOLATED_POND	ISOLATED_POND12	6.6000	7.1068	0.2453	0.3222
55	ISOLATED_POND	ISOLATED_POND16	6.7359	6.9719	0.2670	0.3189
58	ISOLATED_POND	ISOLATED_POND19	6.9264	6.6628	0.2942	0.2381
59	ISOLATED_POND	ISOLATED_POND20	7.2238	6.8873	0.3985	0.2897
70	ISOLATED_POND	ISOLATED_POND31	6.5056	7.0059	0.2267	0.3209
71	ISOLATED_POND	ISOLATED_POND32	6.2000	6.7730	0.1722	0.2650
72	ISOLATED_POND	ISOLATED_POND33	6.3333	7.2238	0.2302	0.3447
73	ISOLATED_POND	ISOLATED_POND34	7.0000	7.0999	0.2971	0.3248
74	ISOLATED_POND	ISOLATED_POND35	7.7500	7.0185	0.4672	0.3335

TABLE 3.1: Comparison between calculated (*LIFE* and *LENLOT*) and predicted (*LIFE-pred* and *LENLOT-pred*) values of both *LENLOT* and *LIFE* indices, for each hydro-morphological units of the sampling area.

## Chapter 4

# Discussion and Conclusions

### 4.1 Data processing

#### 4.1.1 GIS

The morphological analysis of the considered section of the stream, carried out by GIS software use, has provided the visualization and elaboration of the water depth and current velocity values distribution. The raw data coming from the boat survey together with the current meter measurements, were elaborated in order to obtain an higher temporal frequency of acquisition. This last distribution, characterized by 1 second of acquisition frequency, is a pretty faithful representation of what we may call the reality of the Trebbia situation. It's possible to note the similarity between the data provided by two acquisition interval by visualizing the frequency histograms of depth of velocity for each HMUs. An example follows:

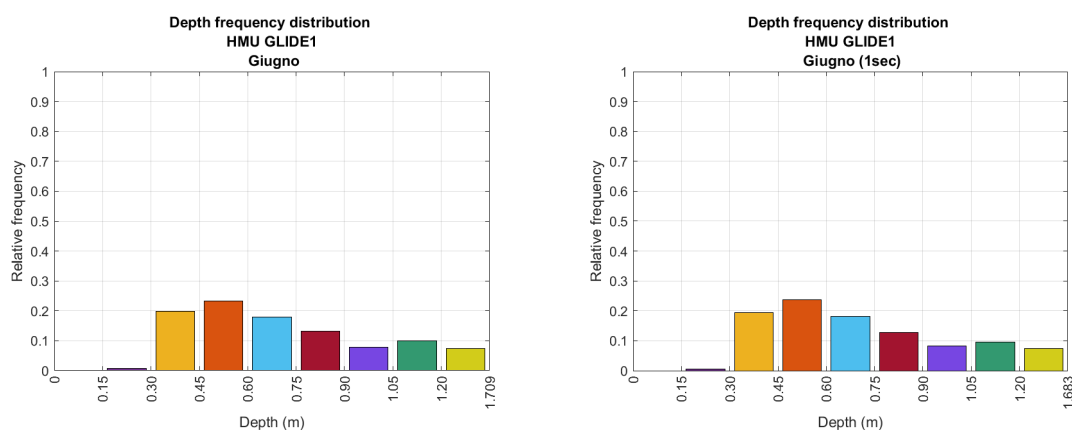


FIGURE 4.1: Comparison between two frequency histograms of depth values: the first refers to the values collected every 4 seconds during the acquisition timeframe, the second refers to the values collected every second. The hydro-morphological unit taken into account is a glide (GLIDE 1). It's quite easy to note the similarity between the distributions.

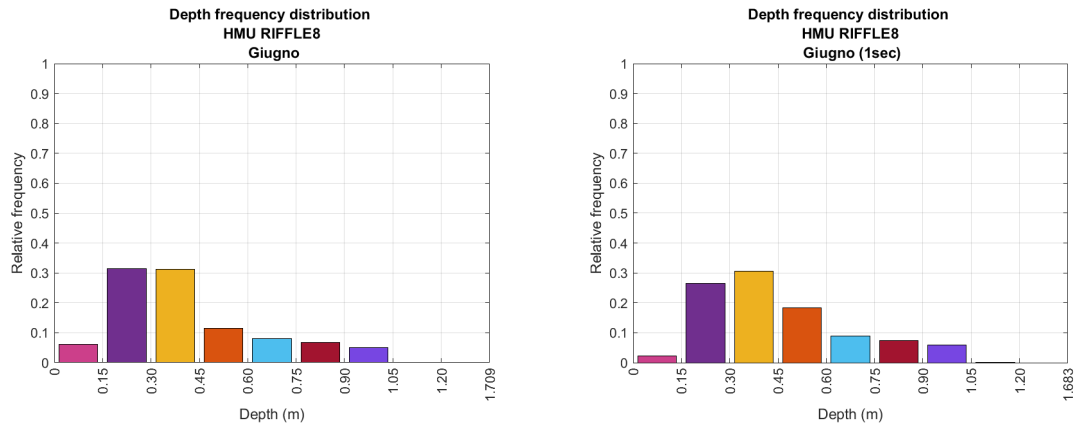


FIGURE 4.2: Comparison between two frequency histograms (4 seconds and 1 second of acquisition interval) of a riffle (RIFFLE 8).

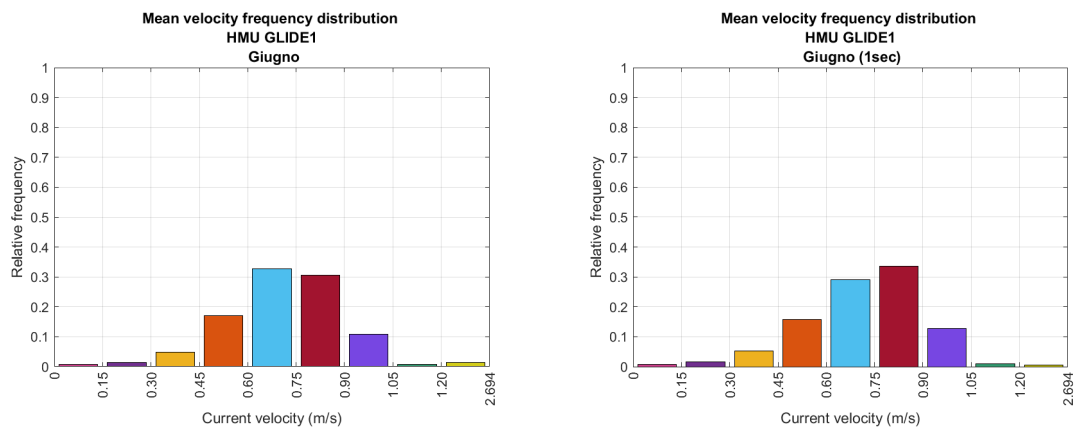


FIGURE 4.3: Comparison between two frequency histograms of velocity values: the first refers to the values collected every 4 seconds during the acquisition timeframe, the second refers to the values collected every second. The hydro-morphological unit taken into account is a glide (GLIDE 1). It's quite easy to note the similarity between the distributions.

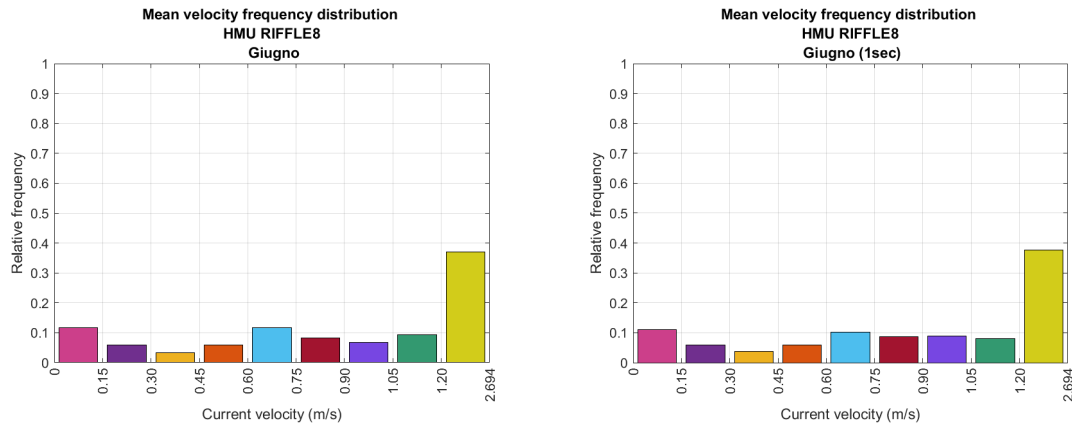


FIGURE 4.4: Comparison between two frequency histograms (4 seconds and 1 second of acquisition interval) of a riffle (RIFLE 8).

Another issue which was taken care of is the bathymetric model correction, by finding the linear regression equation (fig. 2.14) that provides the correction coefficients. Once applied this calculation to the elevations extrapolated from the DTM, the real elevation of the bed stream was achieved. It's important to note that this correction is of fundamental importance in order to obtain the real DTM and so the correct results from the bi-dimensional hydrodynamic simulation.

#### 4.1.2 Hydrodynamic model

The flow field and the depth of water were simulated, obtaining therefore the hydrodynamic bi-dimensional model, that has had to be validated. This validation had ensured the reliability ( $R^2=0.91$  and  $RMSE=0.08$ ) of the simulation outputs, so that this one could be used as input dataset for the statistical model building.

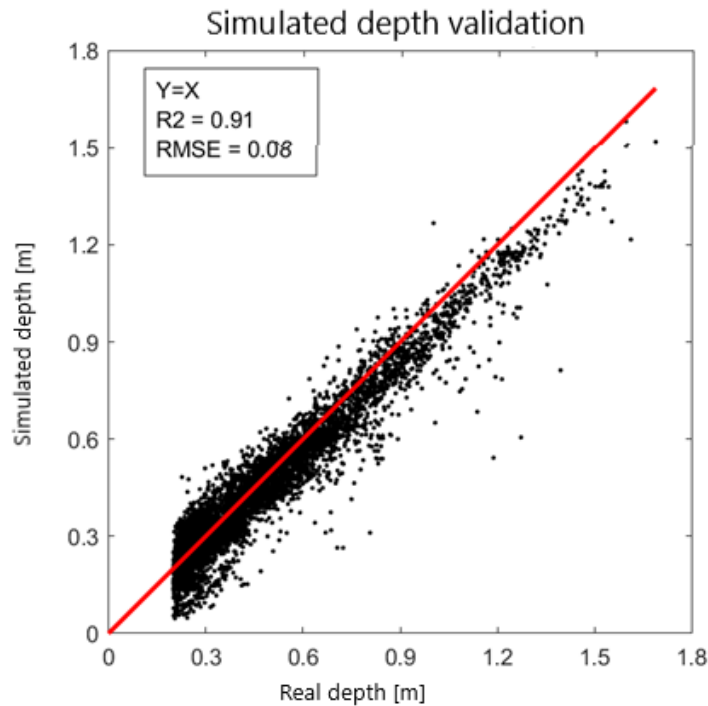


FIGURE 4.5: The bi-dimensional model validation was carried out by analyzing this scatter plot. The figure shows the correlation degree between the simulated depth values and the real ones (from the morphological survey). The correlation coefficient is sufficiently close to one, that makes the model valid. The red line represents the ideal situation in which the simulated values perfectly correspond to the real ones.

In order to better understand the choice of the simulated depth and velocity data as input data for the statistical model building, a comparison between the measured and the simulated dataset follows. In the figures is possible to point out that the modeled distribution of velocity and depth of the stream water is analogous to the measured ones, that may be considered the real situation. This is an important result because it allows to use the simulated distribution for the biological modelling, thus a much larger dataset.

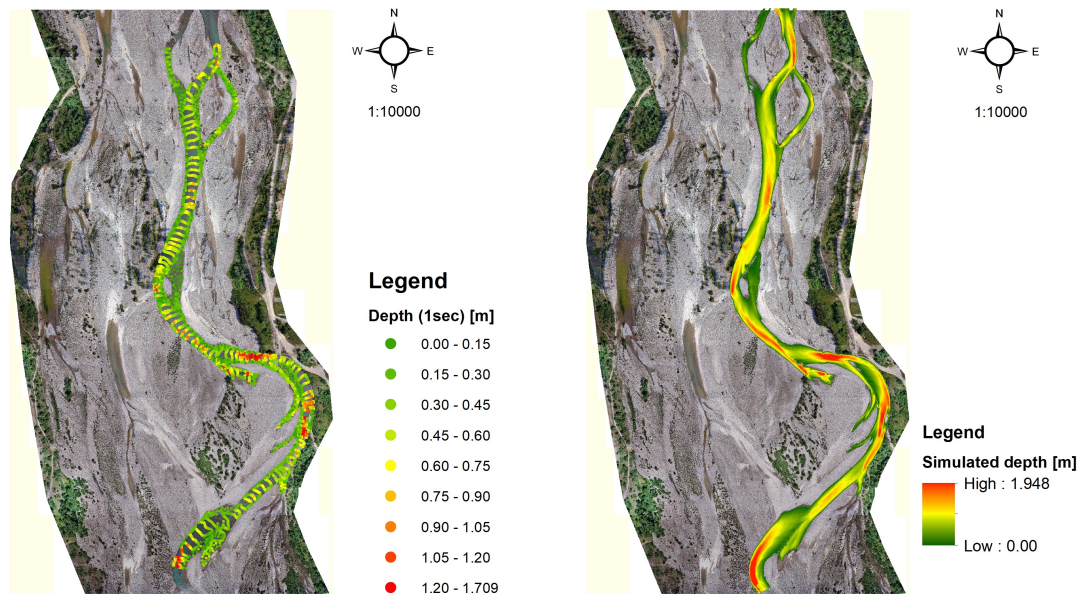


FIGURE 4.6: Comparison between depth values distribution and the simulated distribution of the same variable, within the study area. It's a relevant result the similarity of distributions: in particular the deeper and more shallow zones are located in the same positions.

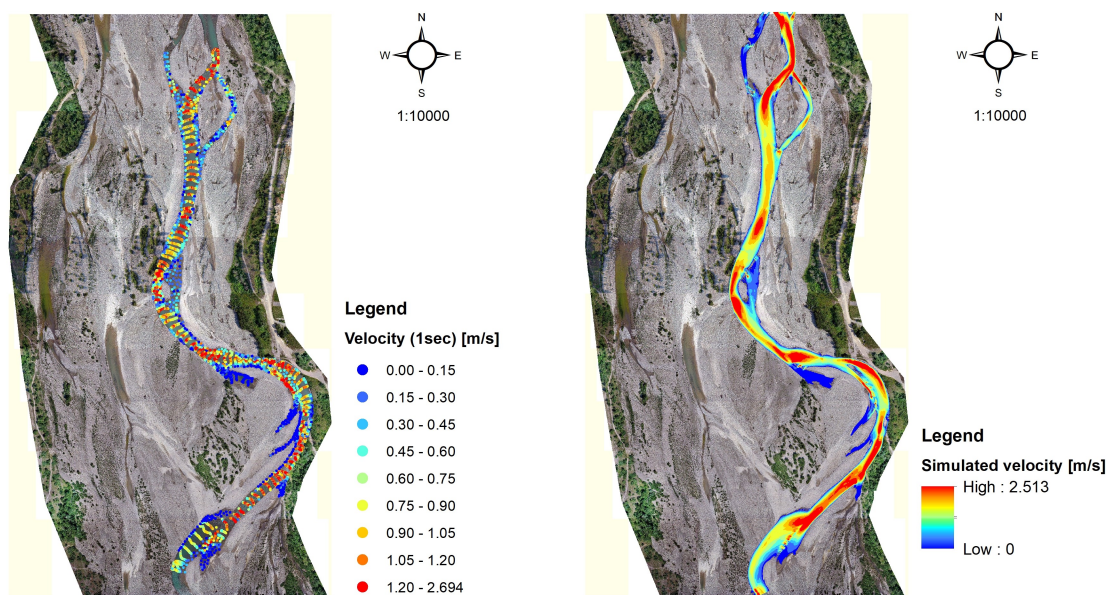


FIGURE 4.7: Comparison between velocity values distribution and the simulated distribution of the same variable, within the study area. It's a relevant result the similarity of distributions: in particular the faster and slower zones are located in the same positions.

This 1 second distribution was used, indeed, as term of comparison with the simulated dataset resulting from the hydrodynamic simulation model.

Below, a comparison between the simulated and the measured distribution of depth and current velocity is shown. The referring HMUs are the same as mentioned above: GLIDE 1 and RIFFLE 8.

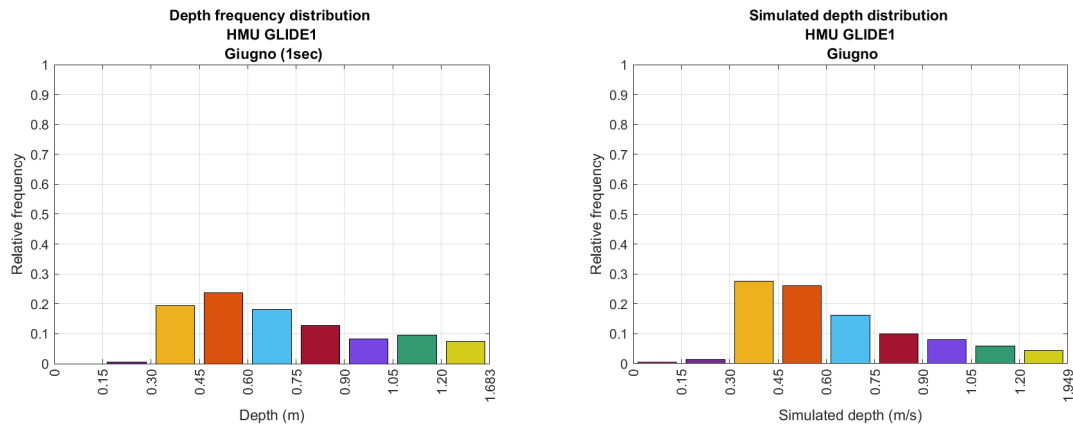


FIGURE 4.8: Comparison between two frequency histograms (measured depth data at 1 second of acquisition interval and simulated depth) of a glide (GLIDE 1).

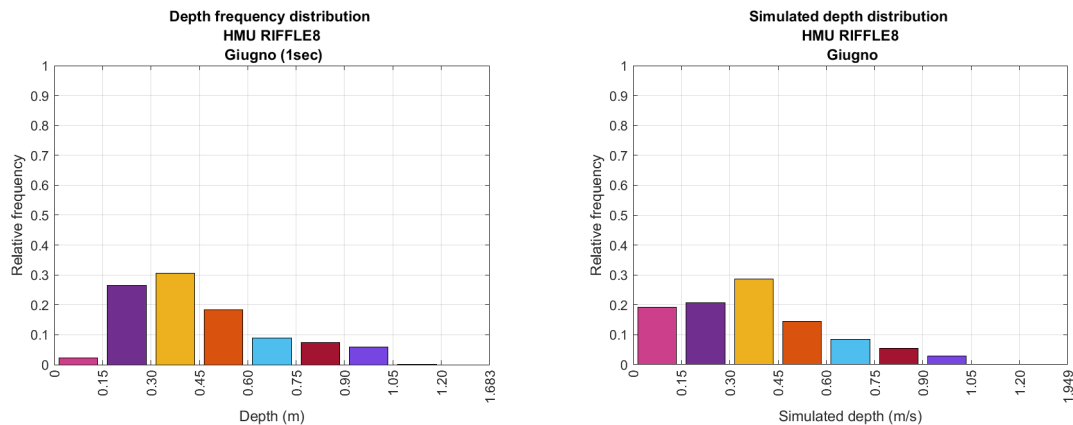


FIGURE 4.9: Comparison between two frequency histograms (measured depth data at 1 second of acquisition interval and simulated depth) of a riffle (RIFFLE 8).

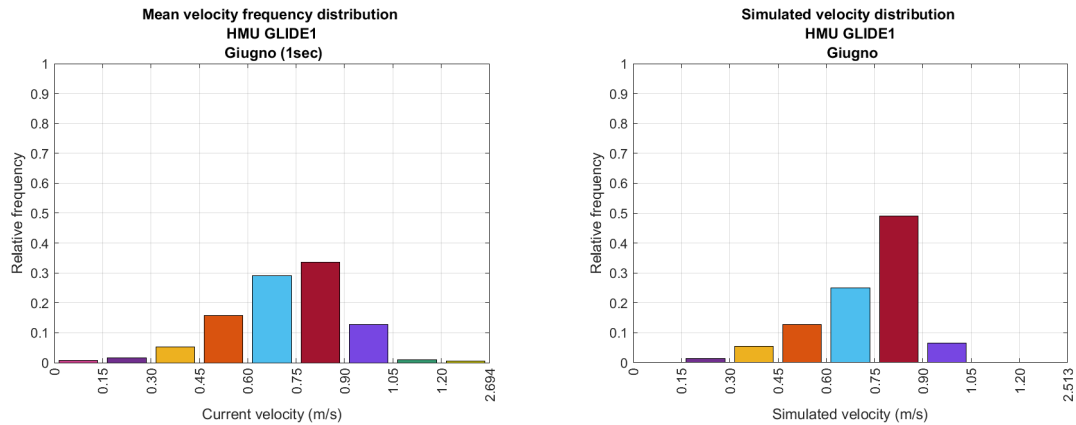


FIGURE 4.10: Comparison between two frequency histograms (measured velocity data at 1 second of acquisition interval and simulated velocity) of a glide (GLIDE 1).

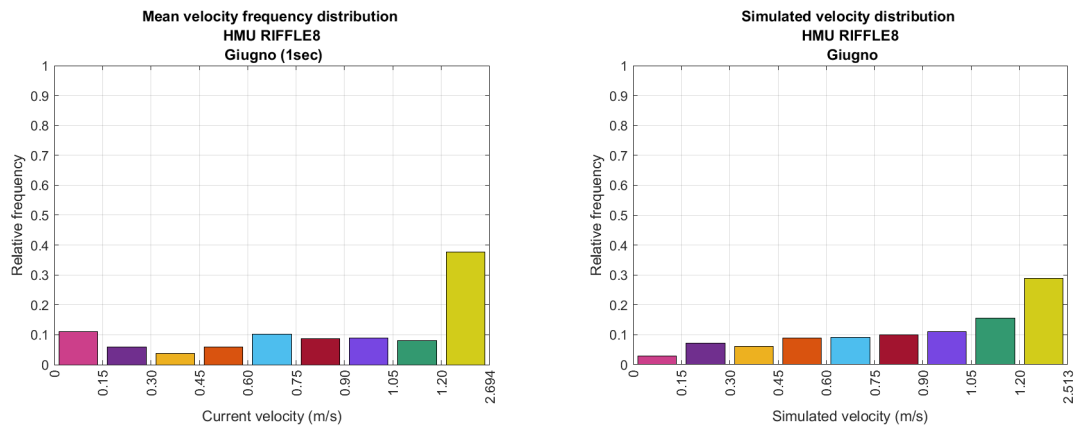


FIGURE 4.11: Comparison between two frequency histograms (measured velocity data at 1 second of acquisition interval and simulated velocity) of a riffle (RIFFLE 8).

## 4.2 Statistical Model

The Random Forests method provides easily understandable outputs that allow the user to find out about the trend of the target variable, so, in this case, of the LIFE and LEN-LOT indices. As an example, looking at the following figure, both the indices are much higher if the considered hydro-morphological unit is connected with the principal channel (discrete variable *CONNECTIVITY*). In fact of that, in the table 3.1, it follows that to the *Isolated ponds* are associated lower values of this index.



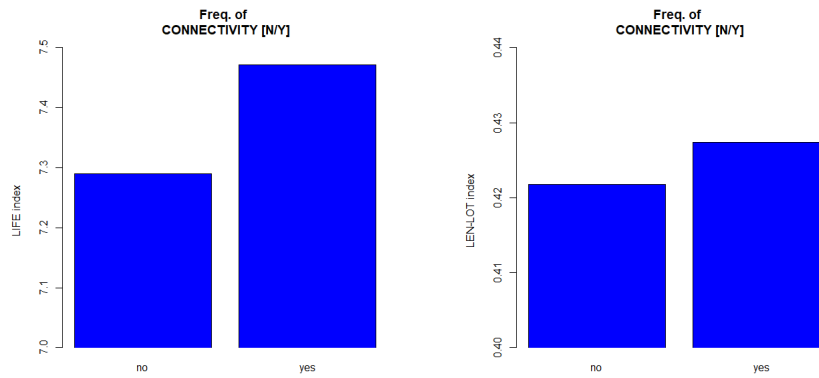


FIGURE 4.12: Comparison between the PDPs of the LIFE and LENLOT indices as a function of the independent variable *connectivity*

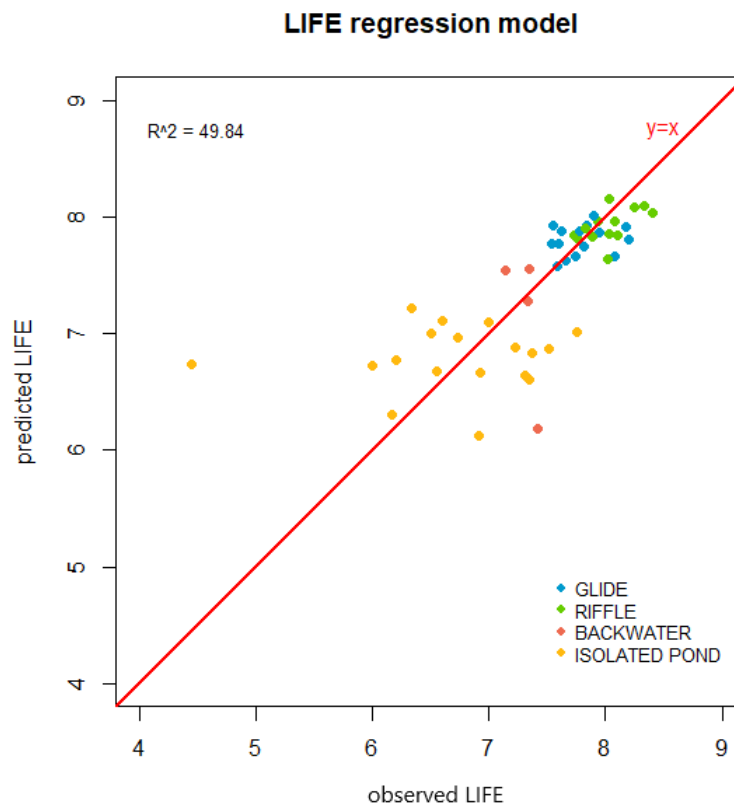
Both the target variables behave similarly, in presence of *PELAL* substrate. In particular, they negatively respond to the increase in relative frequency of this fine sediment and the behaviour is decreasing monotone. Furthermore, *PELAL* substrate plays an essential role considering the variable importance ranking. Different trend (increasing monotone) was obtained for the *MICROLITHAL* class of substrates.

For what concerns the classes of velocity, many of these are considered in the parsimonious model, and some of these provide a similar behaviour of both the indices. The relationship between the target variable and the relative frequencies of *CV45\_60*, *CV60\_75*, *CV75\_90* is similar: there's a maximum at 10% of relative frequency of presence of the class, then a slight decrease. The indices have an increasing behaviour as a function of higher velocity classes (*CV\_120* or *CV90\_105*). The positive behaviour of both LIFE and LEN-LOT with respect to this classes of velocity is a proof of the affinity to the current velocity, based on which both of the indices are calculated. Differently, instead, the response to the presence of lower velocity classes depends on which index is taken into account. An example is the *CV15\_30* class: LEN-LOT index has an increasing then stabilized trend, but LIFE index has a bell-shaped distribution, so that the highest values occur with quite low frequency of presence. This particularity is due to the dependence of LEN-LOT index on higher velocity classes, rather than LIFE index.

The depth values distribution plays a different role in the description of the meso-habitat but some classes are present and of relevant importance. The common class between the models (LIFE and LEN-LOT) is the *D60\_75*, with a negative trend. The other classes in the medium ranges of depth have a similar behaviour. Instead, the lower depth classes are belt-shaped (*D\_15*

with the maximum around 80% of presence and *D15\_30* around 20% of presence).

In order to better understand and visualize the performance of prediction of the RF models, two scatter-plots, in which is illustrated the degree of correlation between the calculated and the predicted indices, are shown. The  $y=x$  line equation is red, and the coloured dots have coordinates (calculated index, predicted index). In an ideal situation, the points are expected to coincide with the red line: the dispersion around it provides the determination coefficient written on top left.



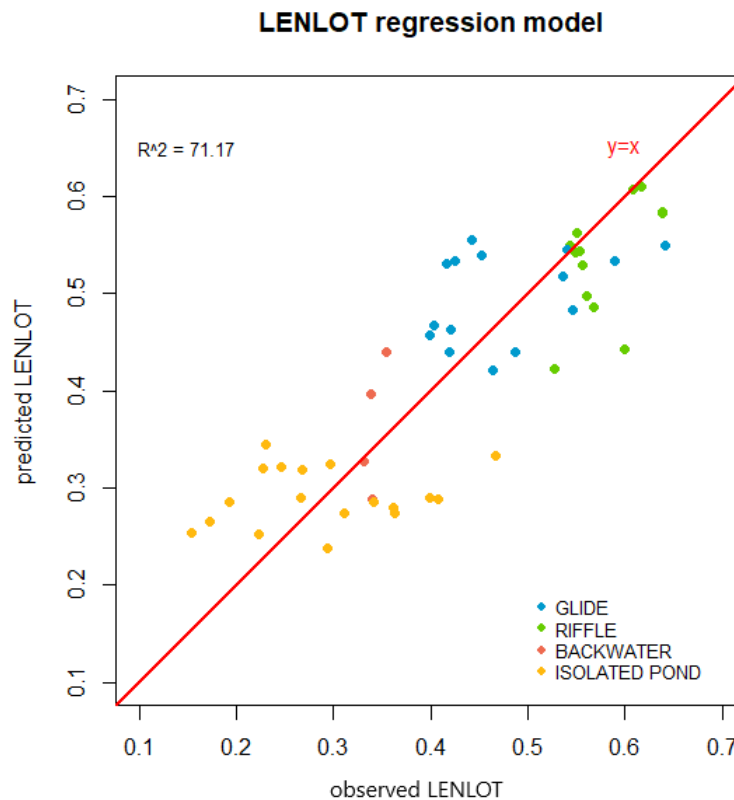


FIGURE 4.14: Scatter-plot of the LENLOT regression model representing the correlation degree between the calculated and predicted index values.

In the previous graphs it's possible to find out the distribution of macroinvertebrate community related to the meso-habitat and, more generally, strongly influenced by the current velocity. As expected, the isolated ponds and the backwaters occupy the positions characterized by the lowest values of both the indices. Conversely, glides and riffles provide higher values of the indices, and, in the case of the LIFE one, less distributed and so more correct in the prediction. These results represent an excellent starting point for the implementation of the MesoHABSIM method with the macroinvertebrates community.

For what concerns the LIFE regression model scatter plot, the most evident outlier corresponds to the LIFE index value calculated and predicted in an isolated pond. In particular, the calculated value is 4.444, while the predicted is nearly 6.74. This point causes a drop in the  $R^2$  and the prediction power of the model, which is, instead, generally more precise.

The most important thing that can be assumed by analysing these plots is the ability of the RF prediction models to find a reliable relationship between the

biological characteristics (represented by the indices) and the morphological ones (meso-habitats).

### 4.3 Conclusions

The main goal of the present study and, more generally, of the BentHab project, is to obtain a new method for modelling the river habitat, based on the macroinvertebrate community. The innovation of this approach lies in the will to determine the composition of the community at the morphological unit scale (meso-habitat), by extending the well-established MesoHABSIM methodology. The MesoHABSIM method allows to model the habitat availability for fish fauna, as a function of the physical characteristics of the river. The main objective of the study is to develop a new procedure of the MesoHABSIM methodology to model habitat availability for the macroinvertebrate community, which represents an important indicator of the ecological health of river systems. Thus, the objective is to define the environmental needs of the target community, as a function of the hydromorphological characteristics of the river (current velocity, water depth and substrate distribution).

For the purpose, data from hydromorphological, topographic and biological surveys were collected and analyzed. The wide dataset provided by the hydrodynamic simulation performed in HEC-RAS, had been used as input in the statistical biological model building.

The functioning of this model is based on a machine learning technique, named *Random Forest*, applied as a predictive regression version. In order to develop the predictive model, some target variables had to be selected: the choice ended up on LIFE and LENLOT indices.

The output models of the community distribution prediction at the morphological unit scale, revealed to have a moderate (for the LIFE case) and good (for the LENLOT case) predictive power and represent, therefore, an effective modelling instrument. The results show how, indeed, the community is distributed in relation with the meso-habitats and, even more strongly, with the current velocity. They are the first try to extend the MesoHABSIM methodology to the macroinvertebrate community, although it is still necessary the validation process of the models.

The greatest potentials of the BentHab project are to be searched in the development of the new approach for the macroinvertebrate meso-habitat modelling; furthermore, the new model-based approach can be eventually applied for studies on intermittent rivers, widely diffused on the mediterranean basin, arid and semi-arid areas. The existing river habitat models are basically inapplicable on intermittent rivers, thus the new approach of macroinvertebrate community modelling would fill this gap and be instrument for environmental flow assessment on IR. These rivers represent a resource often overlooked and delicate ecosystems to be monitored and safeguarded, in particular in view of climate change phenomena.

Future developments of the project will be directed to validate the developed methodology using data from the Enza and the Taro rivers. Moreover, the methodology will be applied to describe the macroinvertebrate spatial distribution during other seasons and bioperiods within the year.

# Bibliography

- [1] BiodrawiversityLLC. "Distribution, Demographics, and Habitat of the Eastern Pearlshell (*Margaritifera margaritifera*) in Upper Wekepeke Brook (Sterling, Massachusetts)". In: (2010).
- [2] K.D. Bovee. "A guide to stream habitat analysis using the Instream Flow Incremental Methodology". In: *U.S. Fish and Wildlife Service* 12 (1982).
- [3] K.D. Bovee and R. Milhous. "Hydraulic simulation in instream flow studies: Theory and techniques. Instream Flow Information". In: *Cooperative Instream Flow Service Group, Fort Collins* (1978).
- [4] L. Breiman. "Random Forests". In: *Machine Learning* 45 (2001), pp. 5–32. URL: <https://doi.org/10.1023/A:1010933404324>.
- [5] G.W. Brunner. "HEC-RAS, River Analysis System Hydraulic Reference Manual". In: (2016).
- [6] A. Buffagni and S. Erba. "Notiziario dei metodi analitici–Macroinvertebrati acquatici e Direttiva 2000/60/EC (WFD)". In: *CNR-IRSA* 1 (2007).
- [7] D.H. Choudary. "Decision Trees Explained With a Practical Example". In: *Towards AI—Multidisciplinary Science Journal* (2020). URL: <https://towardsai.net/>.
- [8] C.A. Extence, D.M. Balbi, and R.P. Chadd. "River flow indexing using British benthic macroinvertebrates: a framework for setting hydroecological objectives." In: *Regulated Rivers: Research and Management* 15 (1999), pp. 543–574.
- [9] K. Jorde et al. "Fuzzy based Models for the Evaluation of Fish Habitat Quality and Instream Flow Assessment". In: (2001).
- [10] J.L. Kemp, D.M. Harper, and G.A. Crosa. "Use of 'functional habitats' to link ecology with morphology and hydrology in river rehabilitation". In: *Aquatic conservation: Marine and Freshwater Ecosystem* (1999). URL: [https://doi.org/10.1002/\(SICI\)1099-0755\(199901/02\)9:1<159::AID-AQC319>3.0.CO](https://doi.org/10.1002/(SICI)1099-0755(199901/02)9:1<159::AID-AQC319>3.0.CO).

- [11] F. Li et al. "Construction of habitat suitability models (HSMs) for benthic macroinvertebrate and their applications to instream environmental flows: A case study in Xiangxi River of Three Gorges Reservoir region, China". In: *Progress in Natural Science* 19.3 (2008), pp. 359–367. URL: <https://doi.org/10.1016/j.pnsc.2008.07.011>.
- [12] A.E. Magurran. "Measuring biological diversity". In: *John Wiley sons* (2013).
- [13] S. M  rigoux et al. "Invertebrate hydraulic preferences and predicted impacts of changes in discharge in a large river". In: *Freshwater Biology* (2009), pp. 1343 –1356.
- [14] J.E. Navarro, D.J. McCauley, and A.R. Blystra. "Instream Flow Incremental Methodology (IFIM) for Modelling Fish Habitat". In: *Journal of water management modeling* 2 (1994). URL: 10.14796/JWMM.R176-01.
- [15] WMO (World Meteorological Organization. "Guidance on Environmental Flows: Integrating E-flow Science with Fluvial Geomorphology to Maintain Ecosystem Services". In: 1235 (2019).
- [16] P. Parasiewicz. "MesoHABSIM: A concept for application of instream flow models in river restoration planning". In: *Fisheries* 26 (2001), pp. 6–13. URL: [https://doi.org/10.1577/1548-8446\(2001\)026<0006:M>2.0.CO;2](https://doi.org/10.1577/1548-8446(2001)026<0006:M>2.0.CO;2).
- [17] P. Parasiewicz. "The MesoHABSIM model revisited." In: *River Research and Applications* 23.8 (2007a), pp. 893–903.
- [18] P. Parasiewicz et al. "Implementation of the Natural Flow Paradigm to Protect Dwarf Wedgemussel (*Alasmidonta heterodon*) in the Upper Delaware River". In: *River Research and Application* 33 (2017), pp. 277–291. URL: <https://doi.org/10.1002/rra.3112>.
- [19] P. Parasiewicz et al. "Lamprey River Protected Instream Flow Report". In: *Report NHDES-R-WD-08-26 for New Hampshire Department of Environmental Services, Concord* (2008).
- [20] P. Parasiewicz et al. "MesoHABSIM: una herramienta eficaz para la gesti  n de r  os y cuencas fluviales". In: (2009).
- [21] M. Rinaldi et al. "Sistema di rilevamento e classificazione delle Unit   Morfologiche dei corsi d'acqua (SUM) - Versione aggiornata 2016". In: *Istituto Superiore per la Ricerca e la Protezione Ambientale (ISPRA) - Manuali e Linee guida 132/2016* (2016).

- [22] K.A. Shearer et al. "Habitat suitability curves for benthic macroinvertebrates from a small New Zealand river". In: *New Zealand Journal of Marine and Freshwater Research* 49.2 (2015), pp. 178–191. URL: <https://doi.org/10.1080/00288330.2014.988632>.
- [23] H. Tachet et al. "Invertébrés d'eau douce - systématique, biologie, écologie". In: (2010).
- [24] P. Vezza, D. Ghia, and G. Fea. "Quantitative Habitat Models for the Conservation of the Endangered European Crayfish *Austropotamobius pallipes* Complex (Astacoidea: Astacidae)". In: *A Global Overview of the Conservation of Freshwater Decapod Crustaceans* (2015), pp. 339–358.
- [25] P. Vezza, A. Zanin, and P. Parasiewicz. "Manuale tecnico-operativo per la modellazione e la valutazione dell'integrità dell'habitat fluviale." In: *ISPRA – Manuali e Linee Guida* 154/2017 (2017). URL: [https://www.isprambiente.gov.it/files2017/pubblicazioni/manuali-linee-guida/MLG\\_154\\_17\\_MesoHABSIM.pdf](https://www.isprambiente.gov.it/files2017/pubblicazioni/manuali-linee-guida/MLG_154_17_MesoHABSIM.pdf).
- [26] P. Vezza et al. "Habitat modeling in high-gradient streams: the mesoscale approach and application". In: *Ecological Applications* 24.4 (2014), pp. 844–861. URL: <https://www.jstor.org/stable/24432195>.
- [27] S. Villeger, N.W. Mason, and D. Mouillot. "New multidimensional functional diversity indices for a multifaceted framework in functional ecology". In: *Ecology* 89.8 (2008), pp. 2290–2301.
- [28] J.B. Wallace and J.R. Webster. "The Role of Macroinvertebrates in Stream Ecosystem Function". In: *Annual Review of Entomology* (1996). URL: [10.1146/annurev.en.41.010196.000555](https://doi.org/10.1146/annurev.en.41.010196.000555).
- [29] M.G. Wolman. "A method of sampling coarse river-bed material". In: *Transactions, American Geophysical Union* 35.6 (1954). URL: <https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.1201%2F9781439833865.ch219>.
- [30] T. Yiu. "Understanding Random Forest". In: *Towards AI—Multidisciplinary Science Journal* (2019). URL: <https://towardsai.net/>.



## *Ringraziamenti*

I miei primi ringraziamenti vanno senz'altro a Paolo ed a tutto il gruppo di ricerca, di Torino e di Parma, per avermi dato preziose opportunità di crescita professionale ma anche personale e culturale. Devo anche a loro la fiducia nel mio futuro, per niente scontata.

Ringrazio Giovanni per la costante presenza, per il paziente aiuto, ma anche per l'allegria disponibilità.

Grazie a Luca e Riccardo, Gemma ed Alex, per il supporto e la sempre gradita compagnia.

Grazie, poi, a tutti gli amici, quelli lontani, quelli vicini, quelli di sempre e quelli recenti: credo davvero che siano sostegno e sicurezza in ogni momento della vita.

Ringrazio di cuore Matti, presenza discreta ma sicura, supporto vero e porto sereno.

Ultima, ma davvero non meno importante, la mia Famiglia, su cui posso sempre fare affidamento.

Mamma e papà, che mai mi hanno fatta sentire sola, soprattutto nei momenti di sconforto. Vi ringrazio per tutti i sacrifici, per le gioie, per i pianti e per i sorrisi. Lo sapete, se non fosse per voi non sarei dove sono ora.

Le mie amate sorelle, la piccola Leti che con la sua semplicità è sempre qui, per me esempio di determinazione e fiducia; e la Mati che lascia il cuore dove passa e io so sempre dove trovarlo. Grazie sorè.

Grazie, infine, a chi con silenzio e bontà è riuscito ad insegnarmi cos'è l'umiltà, cosa significa vivere non tanto per se stessi, quanto per lasciare un segno positivo. A chi è presente ad ogni condizione, in ogni occasione, sempre in quel rispettoso silenzio, che ora ci racconta la vita.

Allo zio Gianni.