

## Parametric Functional Analysis of Variance for Fish Biodiversity Assessment

T. Di Battista<sup>1,\*</sup>, F. Fortuna<sup>1</sup>, and F. Maturò<sup>2</sup>

<sup>1</sup>*Department of Philosophical, Pedagogical and Economic-Quantitative Sciences, University of Chieti-Pescara, 65127, Italy*

<sup>2</sup>*Department of Management and Business Administration, University of Chieti-Pescara, 65127, Italy*

Received 5 April 2015; revised 15 July 2015; accepted 30 August 2015; published online 4 December 2016

**ABSTRACT.** Due to the increasing impact of human activities, water conservation has become a primary aim of environmental management policies. In this context, fish biodiversity represents a good measure of water quality because changes in ecological factors involve qualitative modifications in species composition. For this reason, the analysis of the interaction between biodiversity and environmental characteristics becomes crucial. This paper aims to analyse the effects of habitat and seasonality on fish biodiversity in freshwater environments. In particular, we applied functional data analysis to the beta diversity profile. The proposed approach allows us to overcome the limitations of the classical biodiversity indices, highlighting its multidimensional aspect. More in detail, our research focuses on the functional data analysis of variance in order to quantify the effects exerted on a functional observation by some factors. This model is applied to a real data set concerning ichthyic biodiversity of 104 streams in the province of Arezzo (Central Italy). We consider first the fish zonation and then the seasonality as factors; the results show that species diversity fluctuates seasonally whereas the zonation has no significant effect in influencing the biodiversity. Our proposal is a powerful tool for the analysis of the relationship between qualitative variables and a functional response. Since the diversity profile is a function of the relative abundance vector in a fixed domain, this method is particularly suited to the beta profile and it could be very helpful to monitor or to identify areas of high environmental risk.

*Keywords:* beta profile, fANOVA models, functional data analysis approach, fish biodiversity, permutation F-Test, water quality

### 1. Introduction

The conservation of natural resources and biodiversity, especially in river systems, is a great challenge for the coming century (Ormerod, 1999). In order to identify priority areas for restoration or conservation, scientists are asked to propose simple, synthetic and cheap tools for the evaluation of the ecological status of rivers (Darwall and Vie, 2005). The rivers contribute significantly in fulfilling the basic human needs such as water for drinking and industrial use, irrigation, flood control, hydro power generation, inland navigation, fishing and recreation (Basavaraja et al., 2014). Indeed, many rivers, streams, lakes, and reservoirs have been damaged as a consequence of the increasing impact of human activities (McAllister et al., 1997). This situation is particularly noticeable in the urban environment, where watercourses show highly degraded water quality, receiving not only a great amount of domestic and industrial wastewater, but also sediments and trash (Pompeu et al., 2005). In this context, the analysis of the relationship among physical, chemical, and biological character-

istics of the environment is becoming crucial (Huang and Chang, 2003). Many methods for environmental monitoring have been proposed in recent studies, demonstrating that the debate among scholars on this issue is very lively (Li et al., 2015; Yang et al., 2015; Rege et al., 2015). Since changes in environmental factors involve qualitative modifications in species composition, the most direct and effective measure of water body condition is the status of its living systems (Pompeu et al., 2005). Indeed, it is well known that biodiversity is essential for the stabilization of ecosystems and for the protection of overall environmental quality (Shukla and Singh, 2013; Di Battista et al., 2016). In particular, fishes have widely been used as ecological indicator to assess the level of degradation and health of water bodies (Vijaylaxmi et al., 2010). The impact of the anthropogenic activities, habitat degradation, exotic species introduction, water diversions, pollution and global climate change are the main causative agents for the aquatic species rapid decline (Pawara et al., 2014). Since freshwater biodiversity has declined faster than either terrestrial or marine biodiversity over the past 30 years (Jenkins, 2003), protection and management of water environments has become a primary objective. For this reason many scholars have recently focused on pollution of rivers (Forsythe et al., 2015; Xiao et al., 2015). However, there is an urgent need for proper quantification of fish diversity in order to determine changes that can be used to predict population declines and

---

\* Corresponding author. Tel.: +39-0854537551; fax: +39-0854537551.

E-mail address: dibattis@unich.it (T. Di Battista).

loss of environmental resources (Burger et al., 2013) and to develop policies of natural breed conservation.

In this paper we aim to assess water quality through biodiversity measures. Many indices have been proposed in the literature to evaluate the biological health of a community (Gove et al., 1994). A diversity index is, in general terms, a mathematical expression that combines species richness and species evenness. However, these two components are confounded when a single index is considered. The problem is that a community with few species and high evenness could have the same diversity measure as another community with many species and low evenness (Pielou, 1977). Therefore different indices could lead to different community ranking (Patil and Taillie, 1982). The choice of an index must be considered with care. The most frequently used diversity indices are the species richness index, the Shannon index (Shannon, 1948) and the Simpson index (Simpson, 1949). The richness index provides the number of species present in a given area; its main limitation is that the evenness is not considered. The Simpson diversity index is a good measure of dominance, but not a good predictor of species richness, whereas the Shannon index is affected by both the number of species and their evenness (Evangelista et al., 2012). However, the Shannon index is particularly sensitive to the presence of rare species in a community, while the Simpson index is particularly sensitive to changes in the relative abundances of the most dominant species. This highlights how the use of a single indicator greatly reduces the complexity of the ecological systems and hides the multidimensional aspect of biodiversity (Patil and Taillie, 1979; Gattone and Di Battista, 2009; Gove et al., 1994). For these reasons, a universally accepted measure has not been established (Ricotta, 2005; Di Battista et al., 2016). A possible solution is to consider parametric families of diversity indices (Hill, 1973; Patil and Taillie, 1979), which are usually referred to as diversity profiles. We highlight that we focus on  $\alpha$ -diversity, that is the within habitat diversity (Magurran and McGill, 2011). In particular, we propose the combined use of parametric biodiversity indices and the functional data analysis approach. This method allows us to consider the multi-dimensional aspect of biodiversity and to use statistical techniques, such as functional linear models, for studying the relationships between biodiversity and environmental characteristics. The proposed method provides a better knowledge of the fish biodiversity profile and represents a useful tool for conservation planning of aquatic environments.

## 2. Materials and Methods

### 2.1. Beta Diversity Profile

Biodiversity is a multidimensional concept accounting for both species richness (the number of different species present in an ecological community) and species evenness (a measure of the relative abundance of each species in an area). It represents a good indicator of ecosystem quality (Burger et al., 2013) because, generally, a high level of biodiversity is associated with natural environmental conditions, and vice ver-

sa. Thus, biodiversity management becomes essential for assessing and predicting biological impacts of ecological damages. In this context, the resulting imperative is to provide a suitable measure for biodiversity in order to develop management solutions for its protection.

Since the classical indices neglect the multidimensional aspect of biodiversity, a common solution in the literature is the use of diversity profiles. A diversity profile is a curve depicting in a single graph several values of diversity indices, including the most commonly used indices: Shannon index, Simpson index and species richness. Therefore, the diversity profile is a family of measures that is a family of diversity indices dependent upon a single continuous parameter that is sensitive to both rare and common species. The plot of diversity profiles plays a fundamental role in comparing different communities. Indeed, if the diversity profiles do not intersect, the higher curve corresponds to the community with greater diversity.

In particular, the  $\beta$  diversity profile model proposed by Patil and Taillie (1979) has been used:

$$f(\beta) = \sum_{j=1}^s \frac{(1-p_j)^\beta}{\beta} p_j \quad \text{for } \beta \geq -1 \quad (1)$$

where,  $p_j$  is the relative abundance of the  $j$ -th specie ( $j = 1, 2, \dots, s$ ), with  $p_j = P_j / \sum P_j$  such that  $0 \leq p_j \leq 1$  and  $\sum p_j = 1$ ;  $P_j$  is the absolute abundance of the  $j$ -th species (the number of individuals belonging to the species  $j$ ) and the values of  $\beta$  denote the different levels of the influence of dominant species in computing the biodiversity of a community. Indeed, some of the most frequently used indices of biodiversity are special cases of Equation (1):  $\beta = -1$  generates the richness index;  $\lim_{\beta \rightarrow 0}$  leads to the Shannon diversity index;  $\beta = 1$  provides the Simpson index. The restriction,  $\beta \geq -1$ , assures certain desirable properties for the  $\beta$  profile (Patil and Taillie, 1979). The plot of Equation (1) versus  $\beta$  provides the diversity profile which is a decreasing and convex curve. It provides a faithful graphical representation of the shape of a community, showing how the perceived diversity changes as the emphasis shifts from rare to common species. Moreover, communities can usefully be compared by comparing their diversity profiles (Leinster and Cobbold, 2012).

### 2.2. Functional Data Analysis Approach

Functional data analysis (FDA) addresses problems in which the observations are described by functions rather than finite dimensional vectors (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006). This kind of data is very common in different fields for example environmental science, biology, medicine, meteorology, among many others. In particular, the FDA approach assumes the existence of certain smooth functions,  $f(\beta)$ , which generate the observations. However, in real applications, functional data are often observed as a sequence of data points. In this context, the FDA approach is able to

convert discrete observations to functional form by means of suitable techniques. A popular method to represent smooth functions is through linear combinations of  $K$  known basis functions,  $\phi_k$ , which are linearly independent of each other (Ramsay and Silverman, 2005):

$$f(\beta) = \sum_{k=1}^K c_k \phi_k(\beta) \tag{2}$$

where  $c_k$  is the vector of coefficients defining the linear combination,  $\phi_k(\beta)$  is the vector of the basis functions and  $K$  represents the dimension of the expansion. Several basis functions can be used, including polynomials, regression splines, Fourier series and wavelets, according to the characteristics of the data. In particular, B-spline basis functions are the most used to represent non-periodic functions because of their flexibility and easy implementation (Wegman and Wright, 1983).

The FDA approach treats the whole curve as a single entity instead of a sequence of observations; moreover, it is possible to use functional tools to obtain more information about the data, such as the analysis of the slopes of the functions, reflected in their first derivative.

However, we focus on a particular aspect of functional data analysis, i.e. when the functional datum is expressed by a specific function known in advance (De Sanctis and Di Battista, 2012; Di Battista and Fortuna, 2013). In this case, the observations belong to a parametric family of functions, called  $S$ , with  $s$  real parameters that is:

$$S = \{f(\boldsymbol{\theta}, \beta)\} \tag{3}$$

where  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_s)^T$  represents a set of unknown parameters taking values in a convex parameter space  $\Theta$ , while  $\beta$  is the functional domain. Thus, the functional space  $S$  is constituted by a set of functions belonging to the same family. In this framework, functional data constitute a subset  $S$  of some  $L^p$  space, with  $0 < p < \infty$  and with the usual  $L^p$ -norm,  $\|f\|_p$  (Rudin, 2006). In particular, we consider the Banach space that is every  $L^p$  space with  $p > 0$ . A Banach space is a normed linear space over the field of real functions. It is a complete metric space with respect to the metric derived from its norm that is a Cauchy sequence of vectors always converges to a well defined limit in the space. Since diversity profile expresses diversity as a function of the relative abundance vector in a fixed domain, it can be analyzed in a functional context (Gattone and Di Battista, 2009). Thus, in an ecological setting, the functional space  $S$  could be the family of diversity profiles, such as the  $\beta$  profile in Equation (1), and for each  $i$ -th sites,  $i = 1, 2, \dots, N$ , each relative abundance vector can be assumed as a single parameter,  $\mathbf{p}_i = (p_{i1}, \dots, p_{is}) = \boldsymbol{\theta}_i$ , so that,  $\mathbf{p} = \boldsymbol{\theta}$ , whereas  $\beta$  represents the functional domain (Di Battista and Fortuna, 2013). In this framework, the approximation by means of basic functions is not suitable because the underlying data process is known in advance and it is important to preserve its parametric form. In particular, we assume that

there is a bi-univocal correspondence between the family  $S$  and the convex parameter space  $\Theta$ , so that each function  $f(\boldsymbol{\theta}; \beta)$  of  $S$  is unequivocally defined by the parameter vector  $\boldsymbol{\theta}$ . Therefore, the functions vary in  $S$  only by means of the parameters  $\boldsymbol{\theta}_i$ .

The knowledge of the parametric form underlying the functions allows us to overcome some typical issues of the classic FDA approach. Indeed, in the classical FDA framework, the primary question of interest is the choice of the suitable method for converting the original data in a functional form. Generally, it is possible to approximate any function by taking a linear combination of a sufficiently large number  $K$  of basis functions (Ramsay and Silverman, 2005). The number of basis functions is itself a parameter that we choose according to the characteristics of the data. Obviously, a high  $K$  corresponds to a high fit of the model to the data, but we aim to achieve a satisfactory approximation using a comparatively small number of basis functions. Moreover, a great deal depends on the choice of the basis functions and on how they are able to reproduce the behaviour of the original functions. Indeed, there is no best choice of the basis function. Furthermore, in the specific case of the beta profile, the smooth function  $f(\beta)$  in Equation (2) must satisfy the non-egativity and the convexity of diversity profile. These restrictions can be obtained arbitrarily in several ways (see for example Barlow et al, 1972; Ramsay, 1988; Gattone and Di Battista, 2009). The way of pre-processing the original data is an important issue due to the impact that this process might have on the subsequent statistical analysis (Pérez and Vilar, 2013). For example, in functional clustering methods, the results depend on how the curves are smoothed to the data (Tarpey, 2007). Our method allows us to overcome this limitation. Indeed, the pre-processing step is not necessary because the functional form of the observations is known in advance in its explicit parametric form.

### 3. Parametric fANOVA Model

In order to quantify the effects exerted on a functional observation by some factors, each at multiple levels, a parametric functional analysis of variance (fANOVA) can be used (Ramsay and Silverman, 2005). This procedure involves testing for possible differences among population mean curves under  $G$  different conditions (treatment levels) over the whole functional domain. In this framework, we are interested in generalizations of common statistical tests and of common statistical models. In particular, the aim of this paper is to understand whether two or more groups of functions are statistically distinguishable. For this purpose, we want to test the null hypothesis that the functional groups have the same functional mean, against the alternative hypothesis that there is some difference among them:

$$H_0 : \mu_1(\beta) = \mu_2(\beta) = \dots = \mu_G(\beta)$$

$$H_1 : \mu_g(\beta) \neq \mu_{g^*}(\beta) \text{ for at least one } \beta \text{ and } g \neq g^*$$

We assume that there is a single factor with  $G$  different levels or groups ( $g = 1, 2, \dots, G$ ) and  $N = \sum n_g$  total observations, with  $n_g$  observations within each group; then, the model for the  $i$ -th observation ( $i = 1, 2, \dots, N$ ) in the  $g$ -th group can be expressed as follows:

$$f_{ig}(\beta) = \mu(\beta) + \alpha_g(\beta) + \varepsilon_{ig}(\beta) \quad (4)$$

where  $f_{ig}(\beta)$  is the functional response in the  $g$ -th group,  $\mu(\beta)$  is the grand mean function (i.e. the average function across all observations),  $\alpha_g(\beta)$  represent the functional effect of being in a specific treatment and the residual functions  $\varepsilon_{ig}(\beta)$  are the unexplained variations for the  $i$ -th observation within the  $g$ -th group. The model in Equation (4) can be written in matrix notation:

$$\mathbf{f}(\beta) = \mathbf{Z}\boldsymbol{\gamma}(\beta) + \boldsymbol{\varepsilon}(\beta) \quad (5)$$

where  $\mathbf{f}(\beta)$  is the  $N$  dimensional vector of functional observations, and  $\boldsymbol{\gamma}(\beta) = (\mu(\beta), \alpha_1(\beta), \dots, \alpha_G(\beta))^T$  is the  $(G + 1)$  dimensional vector of the functional parameters;  $\boldsymbol{\varepsilon}(\beta)$  is a vector of  $N$  residual functions and  $\mathbf{Z}$  is a design matrix of dimension  $(N, G + 1)$ , coding the group membership. In particular, each row of the matrix  $\mathbf{Z}$  corresponds to a single observation; the first column consists entirely of “1” to represent the overall mean and the  $G$  columns correspond to different groups with value “1” if the observation belongs to the  $g$ -th group, “0” otherwise.

In order to ensure the identifiability of the functional effects  $\alpha_g(\beta)$ , the sum to zero constrained is imposed:

$$\sum_{g=2}^{G+1} \boldsymbol{\gamma}_g(\beta) = 0, \quad \forall \beta \quad (6)$$

The model is equivalent to the standard ANOVA, with the difference that the vector of parameters  $\boldsymbol{\gamma}(\beta)$ , and hence the predicted observations  $\hat{\mathbf{f}}(\beta) = \mathbf{Z}\boldsymbol{\gamma}(\beta)$ , are vectors of functions rather than vectors of numbers.

Under the assumption of independent errors and constant variance, the vector of parameters  $\boldsymbol{\gamma}(\beta)$ , can be estimated using the standard least squares criterion. Thus, for each  $\beta$ , the minimizing problem becomes:

$$\text{LMSSE}(\boldsymbol{\gamma}) = [\mathbf{f}(\beta) - \mathbf{Z}\boldsymbol{\gamma}(\beta)]^T [\mathbf{f}(\beta) - \mathbf{Z}\boldsymbol{\gamma}(\beta)] \quad (7)$$

Minimizing Equation (7) subject to the constraint in Equation (6), the least squares estimates of the functional parameters at each point of the functional domain are given by:

$$\hat{\boldsymbol{\gamma}}(\beta) = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{f}(\beta) \quad (8)$$

In order to assess whether there are significant differences between the groups, a pointwise F statistic can be used

(Ramsay and Silverman, 2005):

$$F(\beta) = \frac{[SSY(\beta) - SSE(\beta)] / (G - 1)}{SSE(\beta) / (N - G)} \quad (9)$$

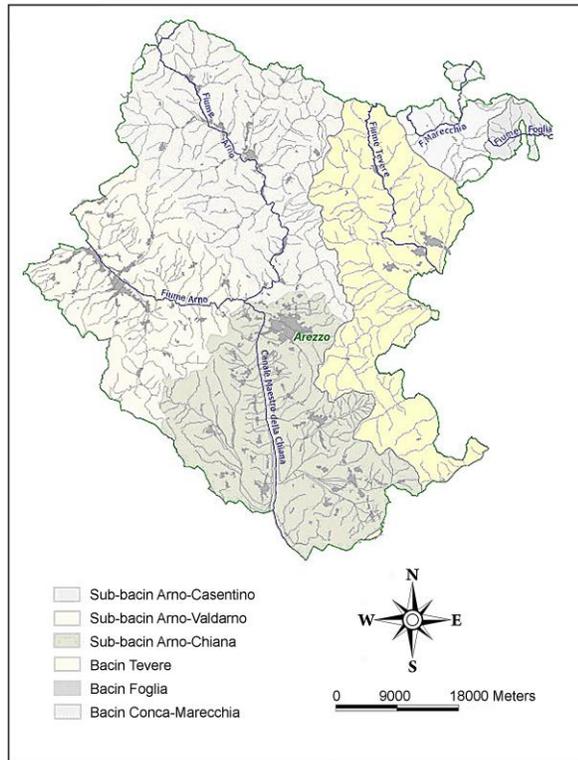
$SSY(\beta)$  represents the variance between groups in the functional context:

$$SSY(\beta) = \sum_{i=1}^N \sum_{g=1}^G (f_{ig}(\beta) - \hat{\mu}(\beta))^2 \quad (10)$$

where the  $\hat{\mu}(\beta)$  is an estimate of the overall mean function.  $SSE(\beta)$  is the functional within variance:

$$SSE(\beta) = \sum_{i=1}^N \sum_{g=1}^G (f_{ig}(\beta) - \hat{f}_{ig}(\beta))^2 \quad (11)$$

Equation (9) provides the observed F-statistic function; it is built by calculating the Fisher test statistic for each point of the domain. As in the classical ANOVA, a high value of  $F$  indicates that the variance explained by the model is greater than the non-explained variance. The main difference between this procedure and the standard univariate or multivariate ANOVA is that the value of  $F$  is not fixed, but it varies over the whole domain. The classical significance level was designed to be used for a single hypothesis test, rather than in a continuum. Therefore, we need to protect ourselves against falsely declaring significance somewhere in the interval. At this purpose a possible solution is the use of the permutation test (Ramsay and Silverman, 2005) that is the functional equivalent of the univariate F-test statistic. Thus, the permutation F-statistic allows us to assess if there are any significant differences between the groups. The basic idea is to compute the Fisher test statistic (Equation 9) as a function built from the series of point estimates for each part of the domain. However, in order to formally test the null hypothesis that there is no relationship between the functional variables, a single test statistic is required. Using the maximum of the observed F-statistic function, a distribution of the test statistic under the null hypothesis can be obtained by calculating the test-statistic several times, each time using random permutations of curves. In detail the procedure is the following: first we calculate the observed F-statistic function using Equation (9) obtaining the “observed F” whose maximum is necessary to compute the  $p$ -value of the test. In the second step, we randomly re-label the curves with different curve numbers, without changing the grouping structure; then for the set of re-labelled curves we calculate the F-statistics for each part of the domain and the maximum of these functions. This re-labeling procedure is repeated several times, and for each one we calculate the pointwise F-statistic function and its maximum. In the third step we find the pointwise 0.05 critical value of the null distribution at each point of the domain and calculate the 95th percentile of the F-statistic values corre-



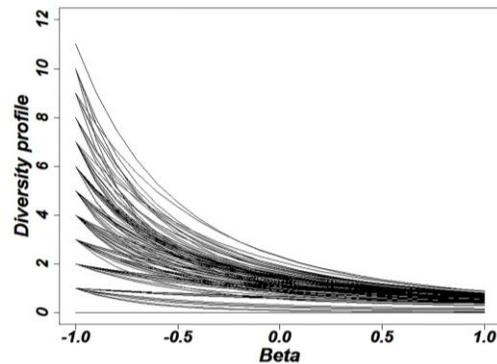
**Figure 1.** The basins of the four rivers in the Province of Arezzo.

sponding to that point. The last step allows us to find the maximum 0.05 critical value of the null distribution, calculating the 95th percentile of the distribution obtained by the permutations of the second step. Taking the 95th percentile of the distribution we find the value of the critical threshold without using the statistical tables. The  $p$ -value corresponding to this test is the proportion of occasions where the maximum value of the permutation F-statistic function is greater than the maximum of the observed F-statistic function. Of course, the  $p$ -value provides the degree of consistency with the null hypothesis like the classical ANOVA.

#### 4. Application: Fish Biodiversity in the Province of Arezzo

The proposed functional approach is applied to a real data set concerning ichthyic biodiversity in the province of Arezzo, Central Italy (see Carini et al., 2006). In 2006 the fish abundance data were collected for a total of 32 species and 104 streams which belong to the basins of four important rivers of Central Italy: Arno, Tevere, Marecchia and Foglia (Figure 1). Ichthyic biodiversity in the province of Arezzo is evaluated through the  $\beta$ -profiles in Equation (1). Figure 2 displays the functional beta profiles for the 104 streams of the study area while due to the presence of a large number of communities, the figure is not easily interpretable; it is possible to highlight only the extremes cases. In particular, the

highest curve represents the community of the stream “Esse” (basin of Chiana); it is composed of 12 different species, indeed for  $\beta = -1$ , the diversity profile is equal to 11 (the number of species minus 1). Although it is not clear from the plot, due to the overlap of the curves, we can identify five communities with only one species (for  $\beta = -1$  the ordinates are equal to 0); three of them are situated on the basin of Casentino (the streams “Gorgone”, “Oia” and “Staggia”), whereas the other two are located in the basins of Chiana (the stream “Esse V.”) and of Tevere (the stream “Minimella”). Figure 3 shows the beta profiles of the four basins; the great part of the streams belongs to the Arno and the Tevere basins whereas in Marecchia and Foglia basins there are only two and one streams, respectively.

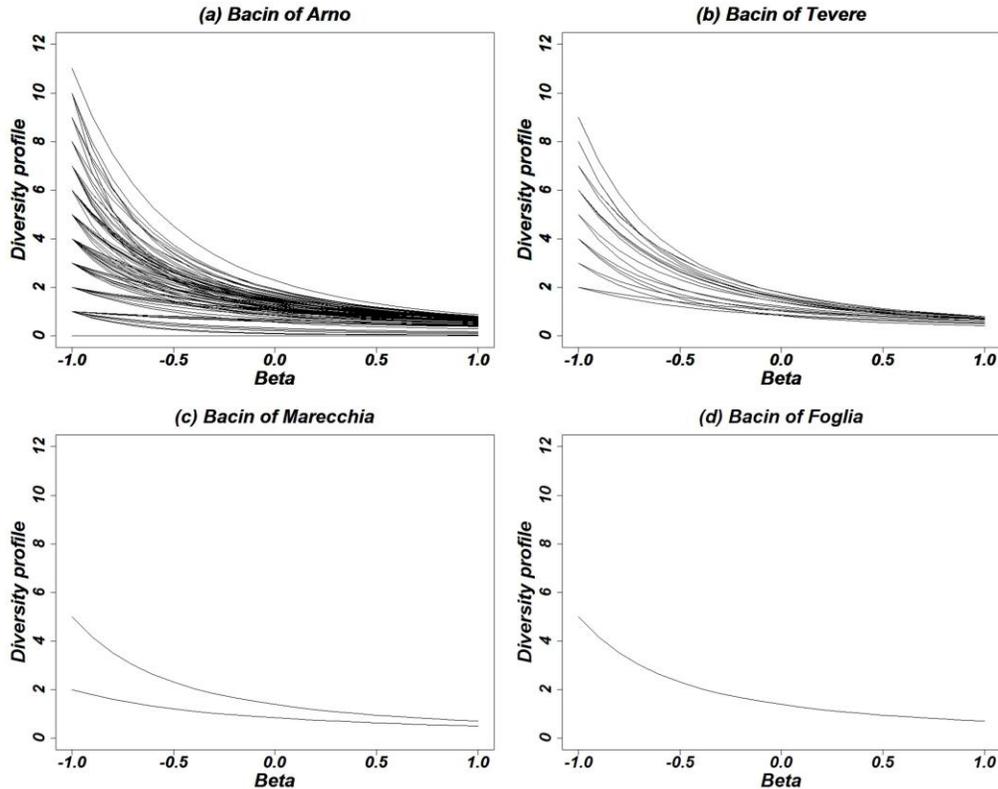


**Figure 2.** The functional beta profiles of the 104 streams in the province of Arezzo.

Several variables could affect biodiversity and the analysis of their interaction is complex. In the literature it is well known that biodiversity is influenced by the environmental characteristics of habitat, geomorphology, climate and species interactions (Pompeu et al., 2005; Shukla and Singh, 2013; Basavaraja et al., 2014). For this reason, the fANOVA model, introduced in Section 3, is applied to explore possible differences in freshwater fish biodiversity among sites in the province of Arezzo. In this paper we analyse the effects of two different categorical variables on the beta diversity profile. At this purpose, we present two different fANOVA models: in the first one the factor is the fish zonation with two treatments (Cyprinids and Salmonids), whereas in the second one, the factor is the seasonality with its four modalities (spring, summer, autumn and winter).

##### 4.1. The Functional Effect of Fish Zonation on Biodiversity

In order to evaluate the effect of habitat characteristics on the fish biodiversity, we consider the zonation system of the streams. According to this approach, the rivers can be divided into separate zones, each with typical fish species, physical and chemical parameters such as the slope, the width, the depth, the current velocity and the water temperature (Huet, 1949). Indeed, the zonation system of rivers is based on the assumption that the physical, chemical and biological characteristics of every natural river change gradually along its



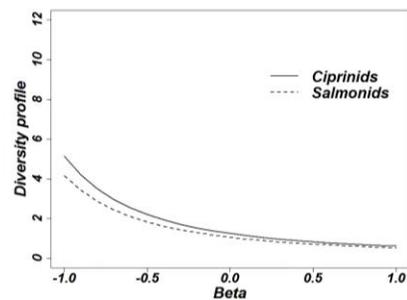
**Figure 3.** The functional beta profiles of the 4 basins of the province of Arezzo: (a) Arno, (b) Tevere, (c) Marecchia and (d) Foglia.

course from headwaters to mouth. One important result of these shifting environmental properties is that most fish species find suitable living conditions in only a selected stretch of the entire river (Bram and Piet, 2003). In the province of Arezzo it is possible to identify only two different zones: Salmonids and Cyprinids (Carini et al., 2006). The first one is usually characterized by slope and cold rivers (max 15 °C), high and well oxygenated water, fast water stream, uneven substrate (with rock, stones, pebbles and gravel) and absence of aquatic vegetation. In the second one, the rivers present slight slope, warm waters in summer, very slow water stream and substrate prevalently muddy. In order to quantify how much of the pattern of biodiversity variation is explainable by the two levels of fish zonation, the fANOVA model in Equation (5) is applied under the constraint in Equation (6). Figure 4 shows the predicted  $\beta$ -profiles for the two zones. The lower diversity is present in the group of Salmonids. Since the two profiles do not intersect, we can conclude that in the Cyprinids zone there is higher biodiversity. The two functional effects of being in a specific zone are displayed in Figure 5. The Cyprinids zone exerts a positive effect on fish biodiversity, whereas the effect of Salmonids zone is negative. The gap between the two functional effects is present over the whole domain, and it is particularly high for  $\beta < 0$ .

Figure 6 shows the functional statistics for testing the null hypothesis that there are no significant differences among the mean group functions. The solid line represents the ob-

served F statistic calculated as in Equation (10); the dashed curve indicates the 0.05 point-wise critical value computed with the permutation test and the dotted line is the 0.05 maximum critical value (see Section 3).

This test is based on the null distribution which has been constructed using 1000 random permutations of the curve labels. It is evident from Figure 6 that the observed F statistic is lower than the maximum critical value for each point of the domain. Although the point-wise F statistic intersects the observed F statistic, we accept the null hypothesis and conclude that there are not statistical differences between the zonation groups in terms of their mean functions. In a functional context, the  $p$ -value is the proportion of occasions where the maximum value of the permutation F-statistic function is



**Figure 4.** The estimated  $\beta$  profiles of the two groups according to the fish zonation.

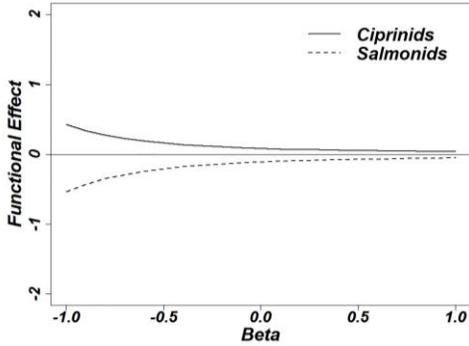


Figure 5. The functional effects of the fish zonation.

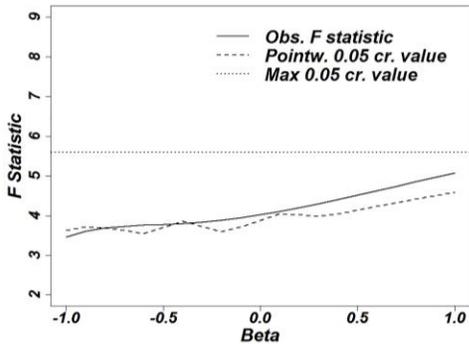


Figure 6. The functional F-test of the fANOVA model using fish zonation as factor.

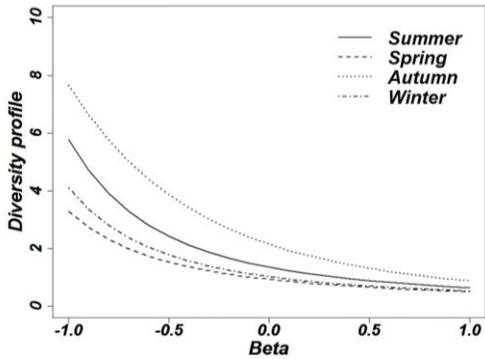


Figure 7. The estimated  $\beta$  profiles of the four groups according to seasonality.

greater than the maximum of the observed F-statistic function. Therefore, in this case, the  $p$ -value tends to infinity and is highly consistent with the null hypothesis.

#### 4.2 The Functional Effect of Seasonality on Biodiversity

The previous fANOVA highlights that the fish zonation does not influence significantly the functional beta profiles in the study area. However, biodiversity can be affected by other factors such as the seasonality because the physico-chemical characteristics of the water vary during different seasons. For this reason, the fANOVA model is applied to test biodiversity

differences in different seasons.

The predicted  $\beta$ -profiles for the four seasons are plotted in Figure 7. The lower biodiversity is present in spring whereas the higher one is given by autumn. Since the four profiles do not intersect, we can easily rank them according to the different seasons. Figure 8 shows the four functional effects of the specific seasons. Autumn and summer present a positive effect on fish biodiversity for the whole domain; whereas winter and spring effects are negative on the entire interval.

Figure 9 shows the observed F statistic (solid line), the point-wise critical value (dashed curve) and the 0.05 maximum critical value (dotted line). Also in this case the test is based on 1000 random permutations of the curve labels. The solid line lies above the point-wise critical level and the maximum critical value over the whole domain, except for  $\beta$  close to 1. Therefore, there is sufficient evidence to reject the null hypothesis and to conclude that there are statistically significant differences among the groups in terms of the seasonality levels. Obviously, in this case, the  $p$ -value tends to zero and it is not consistent with the null hypothesis.

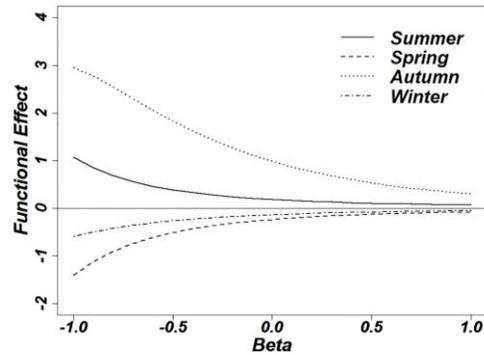


Figure 8. The functional effects of the four seasons.

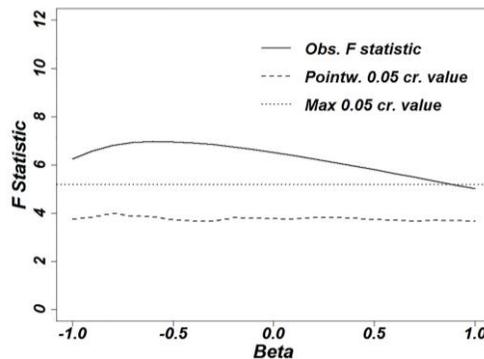


Figure 9. The functional F-test of the fANOVA model using seasonality as factor.

## 5. Conclusions

Ecosystems and marine resources, as common resources, need collective management. Thus, it is necessary to have adequate information on the amount of resources available

and on the dynamics of fish populations and their interaction with the marine environment. Qualitative and quantitative ecological models can help in assessing resources and in regularization of human activities that can have significant impacts on marine ecosystems. Therefore, biodiversity assessment is a key component of environmental sustainability. Our research introduces an innovative method to assess the relationships between functional data and categorical variables of particular interest. Specifically, we suggest the joint use of beta diversity profiles and functional analysis of variance. This method allows us to consider biodiversity in its multidimensional aspect and to evaluate its relationship with environmental factors. In particular, we studied the ichthyic biodiversity of 104 Italian streams in the province of Arezzo. Two environmental factors were considered: fish zonation and seasonality. The first one has no statistically significant effects on biodiversity, contrary to our beliefs. Otherwise, the seasonality highly influences our functional response. In particular, autumn seems to improve fish biodiversity, whereas spring is a limiting treatment. The hypothesis testing was conducted using the permutation test that is the functional equivalent of the univariate Fisher statistical test. The proposed method is particularly suited for monitoring water quality, because the lower beta profiles indicates the lower biodiversity in the graphical comparison and gives an immediate indication of areas of high environmental risk. Moreover, another important feature of our proposal is its ability in comparing the functional effects of different factors in different communities.

## References

- Barlow, R., Bartholomew, D., Brenner, J., and Brunk, H. (1972). Statistical inference under Order Restrictions: the Theory and Application of Isotonic Regression, Wiley, New York.
- Basavaraja, D., Narayana, J., Kiran, B.R., and Puttaiah, E.T. (2014). Fish diversity and abundance in relation to water quality of Anjanapura reservoir, Karnataka, India. *Int. J. Curr. Microbiol. App. Sci.* 3(3), 747-757.
- Bram, G.W.A. and Piet, H.N. (2003). Fish zonation and guilds as the basis for assessment of ecological integrity of large rivers. *Hydrobiologia*, 500(1-3), 157-178. <http://dx.doi.org/10.1023/A:1024638726162>
- Burger, J., Gochfeld, M., Powers, C., et al. (2013). Determining environmental impacts for sensitive species: Using iconic species as bioindicators for management and policy. *J. Environ. Prot.*, 4, 87-95. <http://dx.doi.org/10.4236/jep.2013.48A2011>
- Carini, F., Guffanti, M., Porcellotti, S., et al. (2006). Carta delle vocazioni ittiche della provincia di Arezzo, Provincia di Arezzo.
- Darwall, W.R.T. and Vie, J.C. (2005). Identifying important sites for conservation of freshwater biodiversity: extending the species-based approach. *Fish. Manage. Ecol.*, 12(5), 287-293. <http://dx.doi.org/10.1111/j.1365-2400.2005.00449.x>
- De Sanctis, A. and Di Battista, T. (2012). Functional analysis for parametric families of functional data. *Int. J. Bifurcat. Chaos*, 22(9), 348-355. <http://dx.doi.org/10.1142/S0218127412502264>
- Di Battista, T. and Fortuna, F. (2013). Assessing biodiversity profile through FDA. *Stat.*, 1(1), 69-85. <http://dx.doi.org/10.6092/issn.1973-2201/3986>
- Di Battista, T., Fortuna, F., and Maturo, F. (2016). Environmental monitoring through functional biodiversity tools. *Ecol. Indicators*, 60, 237-247. <http://dx.doi.org/10.1016/j.ecolind.2015.05.056>
- Evangelista, H., Thomaz, S., and Evangelista, L. (2012). Comparison of diversity indices applied to macrophyte incidence-based data. *Braz. Arch. Biol. Technol.*, 55(2), 277-282. <http://dx.doi.org/10.1590/S1516-89132012000200014>
- Ferraty, F. and Vieu, P. (2006). Nonparametric functional data analysis, Springer, New York.
- Forsythe, K.W., Irvine, K.N., Atkinson, D.M., et al. (2015). Assessing lead contamination in Buffalo River sediments. *J. Environ. Inf.*, 26(2), 106-111. <http://dx.doi.org/10.3808/jei.201500300>
- Gattone, S. and Di Battista, T. (2009). A functional approach to diversity profiles. *J. Roy. Stat. Soc. Ser.*, 58(2), 267-284. <http://dx.doi.org/10.1111/j.1467-9876.2009.00646.x>
- Gove, J.H., Patil, G.P., Swindel, B.F., and Taillie, C. (1994). Ecological diversity and forest management. In: Patil, G.P. and Rao, C. (Eds.), *Handb. Stat.*, 12, 409-462. [http://dx.doi.org/10.1016/s0169-7161\(05\)80014-8](http://dx.doi.org/10.1016/s0169-7161(05)80014-8)
- Hill, M. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecol.*, 54(2), 427-432. <http://dx.doi.org/10.2307/1934352>
- Huang, G.H. and Chang, N.B. (2003). Perspectives of environmental informatics and systems analysis. *J. Environ. Inf.*, 1(1), 1-6. <http://dx.doi.org/10.3808/jei.200300001>
- Huet, M. (1949). Aperçu des relations entre la pente et les populations piscicoles des eaux courantes. *Schweiz. Zeitschr. Hydrol.*, 11(3), 333-351. <http://dx.doi.org/10.1007/bf02503356>
- Jenkins, M. (2003). Prospect of Biodiversity. *Science*, 302, 1175-1177. <http://dx.doi.org/10.1126/science.1088666>
- Leinster, T. and Cobbold, C.A. (2012). Measuring diversity: the importance of species similarity. *Ecol.*, 93(3), 477-489. <http://dx.doi.org/10.1890/10-2402.1>
- Li, W., Zhang, H.T., Zhu, Y., et al. (2015). Spatiotemporal classification analysis of long-term environmental monitoring data in the northern part of Lake Taihu, China by using a self-organizing map. *J. Environ. Inf.*, 26(1), 71-79. <http://dx.doi.org/10.3808/jei.201500297>
- Magurran, A.E. and McGill, B.J. (2011). *Biological Diversity*, Oxford University Press.
- McAllister, D.E., Hamilton, A.L., and Harvey, B. (1997). Global freshwater biodiversity: striving for the integrity of freshwater ecosystems. *Sea Wind*, 11(3), 1-142. <http://hdl.handle.net/10625/14024>
- Ormerod, S.J. (1999). Three challenges for river conservation. *Aquat. Conserv.: Mar. Freshw. Ecosyst.*, 9(6), 551-558. [http://dx.doi.org/10.1002/\(SICI\)1099-0755\(199911/12\)9:6<551::AID-AQC387>3.0.CO;2-E](http://dx.doi.org/10.1002/(SICI)1099-0755(199911/12)9:6<551::AID-AQC387>3.0.CO;2-E)
- Patil, G.P. and Taillie, C. (1979). An overview of diversity. In: Grassle, J., Patil, G.P., Smith, W., Taillie, C. (Eds.), *Ecological Diversity in Theory and Practice*, International Co-operative Publishing House, Fairland, MD, 23-48.
- Patil, G.P. and Taillie, C. (1982). Diversity as a concept and its measurement. *J. Am. Stat. Assoc.*, 77(379), 548-567. <http://dx.doi.org/10.1080/01621459.1982.10477845>
- Pawara, R.H., Patel, N.G., and Patel Y.E. (2014). Review on freshwater fish diversity of Maharashtra (India). *J. Entomol. Zool. Stud.*, 2(5), 358-364.
- Pérez, G.E. and Vilar, J.A. (2013). Functional ANOVA starting from discrete data: an application to air quality data. *Environ. Ecol. Stat.*, 20(3), 495-517. <http://dx.doi.org/10.1007/s10651-012-0231-2>
- Pielou, E. (1977). *Mathematical Ecology*, Wiley, New York.
- Pompeu, P.S., Alves, C.B.M., and Callisto, M. (2005). The effects of urbanization on biodiversity and water quality in the Rio das Velhas Basin, Brazil. *Am. Fish. Soc. Symp.*, 47, 11-22.
- Ramsay, J.O. (1988). Monotone regression splines in action. *Stat. Sci.*, 3(4), 425-461. <http://dx.doi.org/10.1214/ss/1177012761>

- Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis*, 2nd edn, Springer, New York.
- Rege, S., Arenz, M., Marvuglia, A., Vázquez-Rowe, I., Benetto, E., Igos, E., and Koster, D. (2015). Quantification of agricultural land use changes in consequential life cycle assessment using mathematical programming models following a partial equilibrium approach. *J. Environ. Inf.*, 26(2), 121-139.
- Ricotta, C. (2005). Through the jungle of biological diversity. *Acta Biotheor.*, 53(53), 29-38. <http://dx.doi.org/10.1007/s10441-005-7001-6>
- Rudin, W. (2006). *Real and complex analysis*, McGraw-Hill, New York.
- Shannon, C. (1948). A mathematical theory of communication. *Bell Syst. tech. J.*, 27(3), 379-423. <http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Shukla, P. and Singh, A. (2013). Distribution and Diversity of Freshwater Fishes in Aami River, Gorakhpur, India. *Adv. Biol. Res.*, 7(2), 26-31. <http://dx.doi.org/10.5829/idosi.abr.2013.7.2.7186>
- Simpson, E. (1949). Measurement of diversity. *Nature*, 163, 688. <http://dx.doi.org/10.1038/163688a0>
- Tarpey, T. (2007). Linear transformations and the k-means clustering algorithm: applications to clustering curves. *Am. Stat.*, 61(1), 34-40. <http://dx.doi.org/10.1198/000313007X171016>
- Vijaylaxmi, C., Rajshekhar, M., and Vijaykumar, K. (2010). Freshwater fishes distribution and diversity status of Mullameri River, a minor tributary of Bheema River of Gulbarga District Karnataka. *Int. J. Syst. Biol.*, 2(2), 1-9.
- Wegman, J.E. and Wright, W.I. (1983). Splines in statistics. *J. Am. Stat. Assoc.*, 78, 351-365. <http://dx.doi.org/10.1080/01621459.1983.10477977>
- Xiao, W.H., Wang, J.H., Huang, Y.H., Sun, S.C., and Zhou, Y.Y. (2015). An approach for estimating the nitrobenzene (NB) emission effect in Frozen Rivers: A case study of nitrobenzene pollution in the Songhua River, China. *J. Environ. Inf.*, 26(2), 140-147. <http://dx.doi.org/10.3808/jei.201500323>
- Yang, Y., Zhang, C.T., Zhang, R.X., and Christakos, G. (2015). Improving environmental prediction by assimilating auxiliary information. *J. Environ. Inf.*, 26(2), 91-105. <http://dx.doi.org/10.3808/jei.201500316>