STATISTICA, anno LXXIII, n. 1, 2013

ASSESSING BIODIVERSITY PROFILE THROUGH FDA

Tonio Di Battista
Department of Quantitative-Economic and Philosophical-Educational Sciences, University G. D'Annunzio, Pescara, Italy
Francesca Fortuna
Department of Quantitative-Economic and Philosophical-Educational Sciences, University G. D'Annunzio, Pescara, Italy.

1. INTRODUCTION

The concept of diversity arises in several disciplines (Pielou, 1975), however, in this paper it is discussed with reference to the ecological framework.
Since rates of habitat and species destruction continue to rise, the need of conserving biodiversity has become increasingly imperative during the last decade (Nagendra, 2001). In light of this new requirement, the first problem that arises is to define and quantify a complex concept such as biodiversity. Although many diversity indices have been proposed, nowadays there is not yet a universally accepted measure.
In this paper we emphasize the multidimensional aspect of diversity and the consequent inadequacy of the classical diversity indices which capture only one aspect of it. The classical indices, in fact, represent diversity as a single statistic in which the number of species and the evenness are confounded, providing a reductive and contradictory vision of it. For this reason, Patil and Taillie (1979, 1982) have proposed to quantify diversity by means of diversity profiles.
A diversity profile is a non-negative and convex curve which expresses diversity as a function of the relative abundance vector.
As a result, Gattone and Di Battista (2009) proposed an alternative way of understanding biological diversity through the classical functional data analysis (FDA). In this way, diversity profile is not simply a sequence of observations, but a function in a fixed domain and it is possible to analyze the intrinsic structure of the data rather than their explicit form.
In successive works, the authors have addressed the problem of estimate functional statistics of the same functional form of the data highlighting that, with the classical functional approach, it is not always possible to achieve this result (Di Battista *et al.*, 2010). For this reason they propose to adopt a parametric FDA approach which, under suitable assumptions, allows us to use the parameter space in order to transport the statistics of the parameters to the functional space.
The paper is organized as follows: in Section 2 we present a review of the fundamental measures of diversity with an emphasis on the limits of classical indices. The use of

FDA to deal with diversity profiles is described in Section 3; while in Section 4 we focus on the parametric FDA setting and on the functional statistics. Moreover, in Section 5 we present a case study carried out in the province of Florence, using the biodiversity of epiphytic lichens. In this application the diversity of biological populations is studied through some methods discussed in the paper with particular attention to the parametric FDA approach. Especially we propose a classification of diversity profiles in a parametric functional setting.

2. Ecological diversity measures

Diversity is considered a wealth of mankind. For this reason, it must be analyzed and quantified to make possible its protection. Diversity is related to the apportionment of some quantity into a number of categories (Patil and Taillie, 1979). In particular, in an ecological framework, the diversity concept relies on the variety of living organisms in a delineated study area (Pielou, 1975).
Since in recent years the attention in environmental protection is increasing, the preserving of ecological communities diversity becomes one of the fundamental objectives of environmental policy. In this context, the arising problem is to define a broad and complex concept such as biodiversity and to measure it by proper indices. As a consequence, a huge number of diversity indices have been developed to quantify and to assess the biological health of a given community (Gove *et al.*, 1994). However, there is no universally accepted biodiversity measure. Traditional indices, such as species richness, the Shannon index (Shannon, 1948), and the Simpson index (Simpson, 1949), characterize diversity through the number of different species (species richness) and the distribution of the number of organism per species (species evenness).
Several studies (Gattone and Di Battista, 2009; Gove *et al.*, 1994; Patil and Taillie, 1979) have highlighted the inadequacy of the classical diversity indices. In fact, the latter incorporates a particular degree of sensitivity to rare and common species but, especially, they represent diversity as a single statistic omitting the multidimensionality of the ecological system. As a result, measures of diversity may lead to different ordering of communities according to the selected index of diversity. Indeed, ecological diversity is a multidimensional concept accounting for both species richness and species evenness, and the use of a single index greatly reduces the complexity of the ecological systems (Gattone and Di Battista, 2009).
In order to overcome the inappropriateness of a single index, Patil and Taillie (1979, 1982) have proposed to quantify diversity by means of diversity profiles, allowing all diversity measures to be encompassed into a single diversity spectrum. A diversity profile is a non-negative and convex curve which expresses diversity as a function of the relative abundance vector. In particular the authors proposed a general class of diversity indices by defining species diversity as the average species rarity within the community. Let us suppose that the ecological population is made up of N units and it is partitioned into s species. Let N_i be the abundance of the i-th species ($i = 1, 2, .., s$); hence $\boldsymbol{N} = (N_1, ..., N_s)'$ denotes the abundance vector, while $\boldsymbol{p} = (p_1, ..., p_s)'$ represents the relative abundance vector with $p_i = N_i / \sum_{i=1}^{s} N_i$, such that $0 \leq p_i \leq 1$ and $\sum_{i=1}^{N} p_i = 1$. Given a community $C = \{s; p_1, .., p_s\}$ and defining $R(i, \mathbf{p})$ as a measure of the rarity for the i-th species,

the average rarity of species in the community is given by:

$$\triangle(\mathbf{p}) = \sum_{i=1}^{s} p_i R(i;\mathbf{p}) \tag{1}$$

Patil and Taillie (1979, 1982) discuss two types of rarity measures: dichotomous-type and rank-type. In the first case the rarity of the *i*-th species depends only on its own relative abundance p_i; thus rarity is denoted by $R(p_i)$.
A general formulation of $R(p_i)$ is $R(p_i) = 1 - p_i^{\beta}/\beta$ so that we get the β diversity profile for community C as:

$$\triangle_{\beta} = \sum_{i=1}^{s} \frac{(1-p_i^{\beta})}{\beta} p_i = \frac{1-\sum_{i=1}^{s} p_i^{\beta+1}}{\beta} \quad \beta \geq -1 \tag{2}$$

where the value of β denotes the relative importance of richness and evenness. The restriction that $\beta \geq -1$ assures that $\triangle_{\beta}$ has certain desirable properties (Patil and Taillie, 1979, 1982). The plot of $\triangle_{\beta}$ versus β provides the diversity profile which is a decreasing and convex curve.
Some of the most frequently used indices of diversity are special cases of equation (2); in fact for $\beta = -1$ we get the richness index, for $\lim_{\beta \to 0}$ the Shannon diversity index and for $\beta = 1$ the Simpson index.
In the case of rank-type measures, however, species rarity depends only on its rank and community diversity is given as $\triangle = \sum_{i=1}^{s} R(i) p_i^{\#}$ where $\boldsymbol{p}^{\#}$ is the vector of relative abundances ranked in descending order $(p_1^{\#} \geq .... \geq p_s^{\#})$ and the rarity of the *i*-th species is given by $R(i) = \{r(i)\}$ where $\{r(i)\}$ represents the rank of p_i in the sequence $(p_1^{\#} \geq .... \geq p_s^{\#})$.
A diversity approach has been developed by Patil and Taillie (1979, 1982) using a rank-type rarity measure of type $R(i) = 1$ if $i > l$ and $R(i) = 0$ if $i \leq l$, for $1 \leq l \leq s$. In this case the right-hand tail sum family of diversity indices is obtained:

$$T_l = \sum_{i=l+1}^{s} p(i)^{\#} \qquad l = 0, 1, ..., s \tag{3}$$

where $T_0 = 1$ and $T_s = 0$.
The plot of the (l, T_l) pairs for each community gives rise to the intrinsic diversity profiles which make possible to compare diversity among communities.
According to Patil and Taillie (1982) a community C is intrinsically more diverse than C', if C' leads to C through a finite sequence of operations:

- transferring abundance from more to less abundant species without reversing the rank-order of the species;
- transferring abundances to a new species;
- relabeling species (i.e., permuting the components of the abundance vector).

We can generalize the formulation of a parametric family of diversity indices as:

$$P_x : x \in \Re^m \tag{4}$$

where P_x are various diversity measures obtained by varying x in the domain $\Re^m$. The curve which joins the (x, P_x) pairs for $x \in \Re^m$ gives us the diversity profile. It depicts in a single picture simultaneous values of diversity measures with varying sensitivities to the rare and abundant species as a function of the parameter x (Gattone and Di Battista, 2009).

3. FUNCTIONAL DIVERSITY PROFILES

In order to consider the multidimensional aspect of biodiversity, Gattone and Di Battista (2009) suggested to explain the diversity profiles through functional data analysis (FDA). As pointed out by Ramsay and Silverman (2005), FDA refers to the analysis of information on functions or curves in a fixed domain. In this setting, the functional datum is regarded as a single entity, instead of sequences of observations; hence this approach focuses on the characteristics of the functions rather than on the simple data.
The classical FDA approach assumes the existence of certain smooth functions $f(\cdot)$ which generate the observations. However, in the real case of study, functional data are often observed as a sequence of point data, then the function denoted by $y = f(x)$ reduces to a record of discrete observations that we can label by the T pairs (x_t, y_t) where $x \in \Re$ and y_t is the values of the function computed at the point x_t, $t = 1, ..., T$. Thus, the first task in a functional data analysis is to convert the discrete measures values, $y_{j1}, ..., y_{jT}$, for each unit $(j = 1, ..., J)$, to a functional form, $f_j(x)$, computable at any desired point $x \in \Re$. For this purpose, the basis function expansion technique allows us to represent a function $f(x)$ in terms of K known basis functions ϕ_k which are linearly independent of each other (Ramsay and Silverman, 2005):

$$f(x) = \sum_{k=1}^{K} c_k \phi_k(x) \tag{5}$$

where c_k is the coefficient vector defining the linear combination, $\phi_k(x)$ is the vector of basis functions and K represents the dimension of the expansion.
In particular, B-spline basis functions are the most used to represent functions. They are piecewise polynomials constructed by dividing the interval of observation into subintervals, with boundaries at points called break points. B-spline are generally applied to non-periodic functional data because of their flexibility and easy implementation (Wegman and Wright, 1983).
Gattone and Di Battista (2009) have studied the diversity profile in the traditional nonparametric FDA setting by using basis functions which have been suitably adapted in order to respect the constraints of the functional diversity profiles (such as non-negativity, decreasing monotonicity, and convexity).
Let us consider a biological population in a delineated study area which is partitioned into J sub-areas $(j = 1, ..., J)$. For each j-th environmental site we observe a relative abundance vector, $\boldsymbol{p} = (p_1, .., p_s)'$. To quantify the diversity of the population the diversity profile in (4) have been used.

Classic FDA represents the diversity profile as a functional on the set $\Re^m$ leading to the non-parametric regression model:

$$P_{jx} = f_j(x) + \epsilon_{jx} \quad j = 1,...,J; \tag{6}$$

where $f_j(x)$ is an arbitrary smooth function and ϵ denotes an unknown independent zero-mean error term with variance σ^2 (Gattone and Di Battista, 2009).

4. FUNCTIONAL STATISTICS IN PARAMETRIC FDA SETTING

In the functional setting, one of the main objectives is to define measures of synthesis and variability of the phenomenon under study and to use the functional analysis to obtain more information on the data such as derivative and so on. However, Di Battista *et al.* (2010) emphasize that, with a standard FDA approach (Ramsay and Silverman, 2005), it is not possible to obtain summary statistics of the same functional form of the observed data and this, obviously, leads to erroneous interpretations of the final functional statistic. To overcome this problem, the authors suggest to adopt a parametric FDA approach.
In fact, there are cases in which the observed functions are known in their explicit form and fixed for each unit. Therefore a functional summary statistic should take into account this issue and should be a function of the same functional form of the observed data. In the parametric FDA setting, the functional data observed for each unit belong to a parametric family of functions, say S, with s real parameters, that is:

$$S = \{f(\theta; x)\} \tag{7}$$

where $\theta = (\theta_1, \theta_2, ..., \theta_s)'$ represents a set of unknown parameters taking values in a parameter space Θ while $x \in \Re^m$ is the domain of the functions.
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):

$$||f||_p = \left\{ \int_X |f|^p d\mu \right\}^{\frac{1}{p}} < \infty \tag{8}$$

where X is an arbitrary measure space with a positive measure μ.
In particular, in this work every L^p space with $p > 0$ (Banach-space) has been considered. A subset of functions, S, is a subspace if it is itself a vector space, that is, given two functions f and g:

1. whenever $f \in S$ and $g \in S$, then $f + g \in S$;
2. whenever $f \in S$ and α is a scalar, then $\alpha f \in S$.

In addition it is possible to define the L^p distance of two functions f and g as:

$$d(f, g) = ||f - g||_{L^p} = \int_X |f(x) - g(x)|^p dx \tag{9}$$

In an ecological setting, S could be the family of Δ_x diversity profiles (defined in (2) and (4) and it is assumed that the biological population under study is fixed and has got a list frame of s species. Thus, data are considered fixed (Thompson and Seber, 1996) and each relative abundance vector can be assumed as a single parameter $p_j = (p_{j1}, p_{j2}, ...p_{js}) = \theta_j$, so that $p = \theta$.
Starting from J functional data, $f(\theta_1, x), ..., f(\theta_J, x)$, belonging to S, the aim is to find a functional statistic, $\widehat{f(\theta, x)} = H(f(\theta_1, x), ..., f(\theta_J, x))$, which is an element of S.
When S is a linear vectorial subspace in L^p, it is possible to express the functional statistics as a straightforward statistics of the functions. For example, the functional mean has been obtained as:

$$\overline{f(\theta, x)} = \frac{\sum_{j=1}^{J} f(\theta_j, x)}{J} \tag{10}$$

In this case $\overline{f(\theta, x)} \in S$ because S is closed with respect to linear combinations.
However, in general the functional data constitute a subspace which is not a linear vectorial subspace of L^p, and the functional statistic in (10) does not necessarily lead to an element belonging to S.
Inspired by mathematical tools, and assuming a monotonic dependence from the parameters, it is possible to use the parameter space in order to transport the statistics of the parameters to the functional space. In particular it is assumed that:

1. the parameter space Θ is a convex subset of $\Re^s$ that is $(\theta_1, \theta_2, ..., \theta_J)'$, where $\theta_j = (\theta_{j1}, \theta_{j2}, ..., \theta_{js})$; α_{ji} is a scalar with $0 < \alpha_{ji} < 1$ and $\sum_{j=1}^{J} \alpha_{ji} = 1$ for each $i = 1, 2, ..., s$; then $\sum_{j=1}^{J} \alpha_{ji} \theta_{ji} = \overline{\theta}_i \in \Theta$ for each $i = 1, 2, ...s$;

2. there is a bi-univocal correspondence between the family S and a convex parameter space Θ, so that each functional datum $f(\theta, x)$ of S is unequivocally defined by the parameter θ.

Under these assumptions, a functional statistic for the functional data is given by a suitable statistic of the parameters $\theta_1, ..., \theta_J$:

$$\widehat{\theta} = K(\theta_1, ..., \theta_J) \tag{11}$$

and the functional statistic will be an element of S with the statistic $\widehat{\theta}$ as parameter.
We can define the function $K(\cdot)$ through the analogy criterion, thus, for the functional mean, the function $K(\cdot)$ would be the mean of the parameters $\widehat{\theta} = (\overline{\theta}_1, ..., \overline{\theta}_s)$ where $\overline{\theta}_i = \sum_{j=1}^{J} \frac{p_{ij}}{J}$, for $i = 1, ..., s$.
The advantage of this approach is that it is possible to require for the functional mean the same properties of the mean of the parameters.
The monotonic dependence assumption ensures the internality property of the functional mean $\left(f(\theta_1, x) \leq f(\overline{\theta}, x) \leq f(\theta_J, x)\right)$. However, in many cases, this assumption is too strong. To overcome this drawback we can introduce a weak monotonic dependence; in fact if it is possible to define a finite partition of $X : P = \{X_i\}$, where:

1. $\bigcup_i X_i = X$;
2. $X_i \bigcap X_k = \emptyset$ if $i \neq k$;
3. for each X_i the monotonic dependence of the parameters and, accordingly, the internality property of the mean are verified;

then the internality property of the mean is ensured in all the partition X. Therefore, if S is a parametric functional space in which the functions satisfy the weak monotonic dependence assumption, then $f(\overline{\theta}, x)$ satisfies the weak internality property.
In order to consider a measure of functional variability, Di Battista *et al.* (2010) express the r-th order algebraic deviation between the observed functional data $f(\theta_j, x)$ and the functional statistic $f(\overline{\theta}, x)$ as:

$$v_j^r(x) = |f(\theta_j, x) - f(\overline{\theta}, x)|^r \tag{12}$$

Then, the functional variability can be defined pointwise by the r-th functional moment by placing r=2:

$$V^r(x) = \frac{1}{J}\sum_{j=1}^{J} v_j^r(x) \tag{13}$$

Defining the L^p distance of two functions as in (9), the function $V^r(x)$ has the following property: if $f(\theta_j, x) = f(\overline{\theta}, x)$ *a.e.*, $\forall j = 1.2...,J$ then $V^r(x) = 0$ *a.e.*.

5. Epiphytic lichen biodiversity profile: a case study in the province of Florence

Diversity profiles and FDA approach have been applied to evaluate the diversity of epiphytic lichens, i.e. lichens which live on trees bark.
Lichens are composite organisms consisting of a fungus (the mycobiont) and a photosynthetic partner (the photobiont) growing together in a symbiotic relationship. Lichens are long-lived and slow-growing organisms, that present a good constancy of morphology over time (Conti and Cecchetti, 2001). The vast majority of them shows a wide range of tolerance to environmental extremes and, for this reason, they are able to colonise habitats where few other macroscopic organisms can grow (Asta *et al.*, 2002). Nevertheless lichens are particularly sensitive to environmental stresses, especially with regard to pollution, eutrophication and climate change (ANPA, 2001), because their metabolism is directly dependent on gas exchange. Since lichens respond to phytotoxic gases at cellular, individual and community level (Van Dobben *et al.*, 2001), in the last decades, biodiversity of epiphytic lichens has been considered a good indicator of air pollution (Nimis *et al.*, 1989; Giordani *et al.*, 2002; Cristofolini *et al.*, 2008). In particular, bioindication techniques are based on the observation of decreased richness of lichen communities, in relation to increasing concentration of atmospheric pollutants (Cristofolini *et al.*, 2008).
Generally lichen biodiversity studies are based on the analysis of the Lichen Diversity Value (LDV) which is calculated as the sum of lichen frequencies in a sampling grid on

a tree and, then, it is converted into naturality/alteration classes in order to define different zones of environmental quality (ANPA, 2001; Asta *et al.*, 2002). However, as the classic biodiversity indices, it is a scalar which can provide a narrow view of biodiversity.

In this paper we consider epiphytic lichen biodiversity of Tuscany region, in central Italy. Data on lichen abundance were collected by ARPAT (Regional Agency for Environmental Protection of Tuscany) in partnership with SIRA (Regional Environment Information System of Tuscany) following the standards suggested by Asta *et al.* (2002). The survey lasted from 2003 to 2009 and involved 65 stations planned by the national protocol. Further details on the data may be found in http://sira.arpat.toscana.it/sira/biomonitoraggio/ucp.htm.

For practical purposes, here we consider an extract of the more general Tuscan lichen data set; in particular we focus on the province of Florence for a total of 11 stations involving 10 municipalities (because two stations fall in the municipality of Firenzuola). Figure (1) shows the location of the sample units in the province and reports the legend of the station codes which will be used again in following graphs.

Within each environmental site, the closest trees to the center of the station and con-

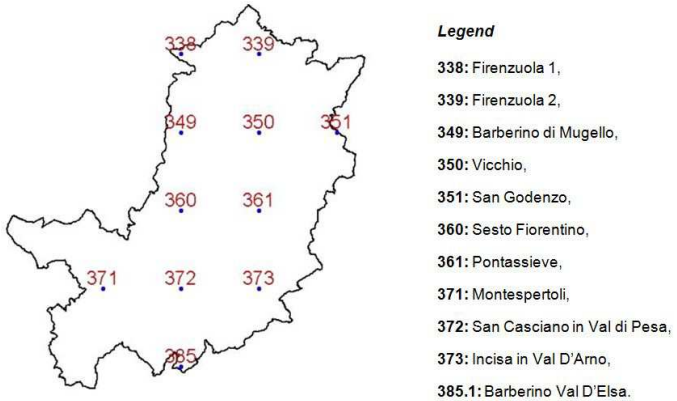

Figure 1 – Location of the 11 stations in the province of Florence.

forming to minimum standards (circumference > 60 cm; inclination of the bole $< 10°$; absence of damage and decorticated areas on the trunk) have been considered. In the survey area, 34 trees have been collected, three for each site with the exception of San Casciano in Val di Pesa with four trees. All the trees are *Quercus* except for a single tree which is a *Tilia sp.* on the Barberino Val D'Elsa station.

For each tree, the abundance of every lichen species has been recorded using a sampling grid consisting of a 10×50 cm ladder divided into five 10×10 cm quadrants. This ladder grid has been placed systematically on the North, East, South and West side of the bole of each tree (four per tree), with the top edge 1.5 m above ground, as suggested by Asta *et al.* (2002). Species frequency has been calculated for each tree as the number of quadrants in which the species is present. Thus, the abundance vector of the i-th species ranges from 0 to 20 for each tree.

In this application, for every j-th environmental site ($j = 1,..,11$) and for each i-th

species ($i = 1, .., 54$), we consider an abundance vector, $\boldsymbol{N}_j = (N_{j1}, .., N_{js})'$, calculated as the sum of the lichen frequencies found on every tree belonging to the plot.
A total of 54 epiphytic lichen species has been found on the 11 sampled plots. As shown in Table (1), the most common species in the province are *Physcia adscendens, Lepraria incana* and *Lecidella elaochroma* with a frequency of 374, 307 and 232 respectively. Furthermore *Physcia adscendens* and *Lecidella elaochroma* are the only species present in all sites. The rarest species, instead, belong to *Caloplaca sp.* with an abundance equal to 1.
Regard to the distribution of species within the sites, 31.5% of the species is present

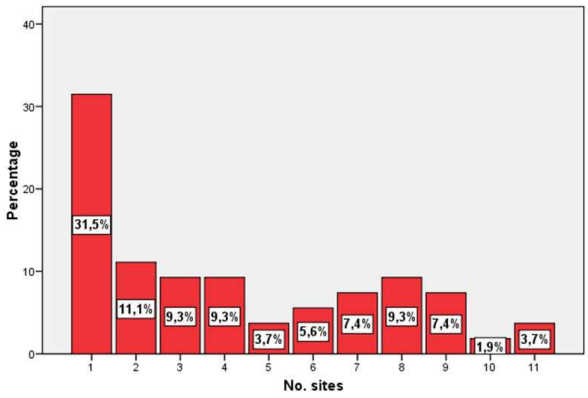

Figure 2 – Distribution of the species within the sites: percentage of species present at each site.

only in one of them while only 3.7% is found in all sites (figure 2).
According to a preliminary descriptive analysis is evident the prevalence of rare species both in terms of abundance both in terms of distribution between sites.
Initially lichen biodiversity has been evaluated through traditional diversity indices with regard to the relative abundance vector $p_j = (p_{j1}, ..., p_{js})'$. As shown in Table (2), Sesto Fiorentino presents the lowest diversity for all indices. On the other hand, Barberino Val D'Elsa reflects the greatest diversity for both the Shannon and the Simpson index. According to the richness index and lichen biodiversity index, San Casciano Val di Pesa is the site with the higher biodiversity. The analysis of the classical diversity indices emphasizes their inadequacy, because they return a different ordering of sites according to the particular aspect of diversity under consideration. It is also evident that LDV captures only one type of diversity, omitting the species composition and showing results similar to the richness index.
To overcome these limits, the β-profile expressed in equation (2) has been used to assess lichen biodiversity. Figure (3) shows the curves obtained for each environmental site. The analysis of the β-profile does not return a clear ordering between sites. However it seems that the discrimination between them is mainly explained by species richness. The study of certain β values confirms the results obtained by classical indices. In fact, San Casciano in Val di Pesa shows a greater diversity in species richness (for $\beta = -1$); while Sesto Fiorentino presents the lowest richness in species. Furthermore, for β from -0.3 to 0.4, Barberino Val D'Elsa reflects the most different profile.
In order to compare diversity between communities, the intrinsic diversity profile in (3)

TABLE 1
Provincial abundance and distribution between sites for species.

Species	Provincial total	Min	Max	Number of sites
Physcia adscendens	374	10	50	11
Lepraria incana	307	0	65	9
Lecidella elaeochroma	233	4	40	11
Hyperphyscia adglutinata	167	0	30	9
Candelariella xanthostigma	161	0	43	10
Parmelia sulcata Taylor	150	0	51	8
Normandina pulchella	143	0	38	7
Candelaria concolor	140	0	52	9
Flavoparmelia caperata	136	0	43	7
Lecanora carpinea	128	0	45	8
Physconia grisea	95	0	26	9
Lecanora chlarotera Nyl.	92	0	22	8
Flavoparmelia soredians	87	0	45	3
Phaeophyscia orbicularis	87	0	18	7
Lecanora expallens Ach.	77	0	22	7
Physcia tenella	70	0	37	4
Punctelia subrudecta	63	0	16	8
Xanthoria parietina	62	0	33	6
Evernia prunastri	52	0	14	8
Pertusaria albescens	52	0	27	5
Physcia aipolia	51	0	20	4
Lecanora horiza	42	0	22	6
Leprocaulon microscopicum	39	0	11	6
Melanelia subaurifera	39	0	37	3
Parmotrema chinense	38	0	15	5
Parmelina pastillifera	33	0	15	4
Lepraria sp	29	0	15	2
Parmelia saxatilis	28	0	11	4
Catillaria nigroclavata	24	0	24	1
Parmelina tiliacea	22	0	13	3
Punctelia borreri	20	0	20	1
Ramalina fastigiata	19	0	14	3
Gyalecta	17	0	7	4
Pleurosticta acetabulum	13	0	8	3
Phaeophyscia chloantha	12	0	7	2
Physcia biziana	10	0	10	1
Pertusaria pertusa	10	0	9	2
Phaeophyscia hirsuta	8	0	8	1
Physconia distorta	7	0	4	2
Lecanora meridionalis H.Magn.	4	0	4	1
Parmelina quercina	4	0	4	1
Pertusaria amara	4	0	4	1
Collema sp	3	0	3	1
Crustose	3	0	3	1
Melanelia sp	3	0	2	2
Physconia venusta	3	0	3	1
Buellia schaereri De Not.	2	0	1	2
Opegrapha atra Pers.	2	0	2	1
Physcia stellaris	2	0	2	1
Phaeophyscia cernohorskyi	2	0	2	1
Physconia perisidiosa	2	0	2	1
Rinodina sp	2	0	2	1
Caloplaca cerina	1	0	1	1
Caloplaca sp	1	0	1	1

TABLE 2
Classical diversity indices and LDV for the 11 stations.

Municipalities	Code	$\triangle_{Rich}$	$\triangle_{Sh}$	$\triangle_{Si}$	LDV
Firenzuola 1	338	19	2.65	0.91	92.33
Firenzuola 2	339	19	2.57	0.90	84.67
Barberino di Mugello	349	17	2.52	0.90	102.00
Vicchio	350	21	2.65	0.91	92.67
San Godenzo	351	19	2.42	0.88	97.67
Sesto Fiorentino	360	14	2.32	0.88	63.33
Pontassieve	361	17	2.68	0.92	82.33
Montespertoli	371	26	2.81	0.92	91.00
San Casciano in Val di Pesa	372	29	2.90	0.93	118.25
Incisa in Val D'Arno	373	22	2.68	0.91	93.00
Barberino Val D'Elsa	385_1	26	2.93	0.94	112.00

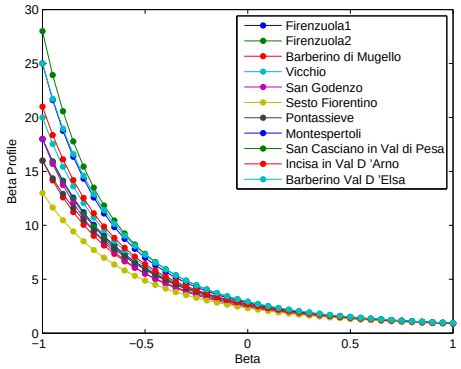

Figure 3 – β-profiles ($\triangle_\beta$) for the 11 sites in the province of Florence.

has been applied. Figure (4) displays the intrinsic profiles for each environmental site, with species abundances on the abscissa. Since the profiles cross each other, it is not possible to distinguish a site with greater diversity. Where possible, it could be carry on with a pairwise comparison. However, to get an overview of the study area, we proceeded to identify groups among sites on the basis of the abundances. For this purpose parametric FDA approach has been applied on account of the limitations highlighted by Di Battista et al. (2010), with reference to the classical FDA. In this setting, it is possible to define functional groups applying clustering algorithms on the parameters instead of on the functions themselves. In particular, agglomerative Ward's hierarchical clustering method (Ward, 1963) has been implemented using Euclidean distance as dissimilarity measure among observations. As shown by the dendrogram in Figure (5), it is possible to recognize two groups for a distance of 0.35.
The first group is composed of seven sites: Firenzuola2, Pontassieve, Vicchio, Barberino di Mugello, Montespertoli, Barberino Val D'Elsa and Sesto Fiorentino. The second group, instead, includes only four sites: San Casciano in Val di Pesa, Incisa in Val D'Arno,

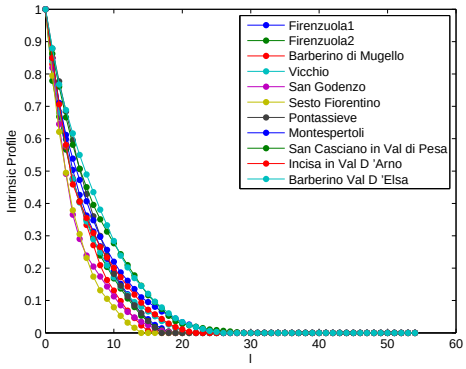

Figure 4 – Intrinsic profiles (T_l) for the 11 sites in the province of Florence.

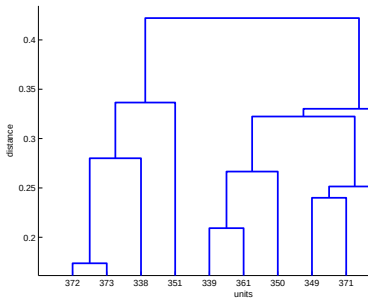

Figure 5 – Dendrogram for Ward's method applied to the 11 sites.

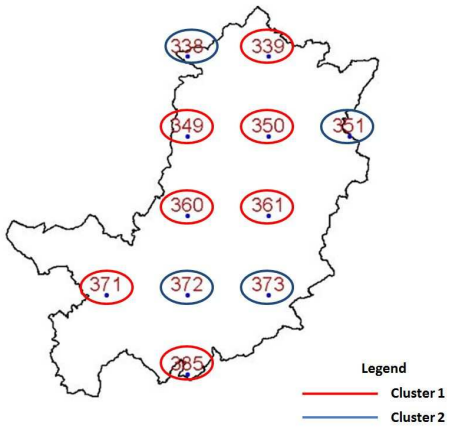

Figure 6 – Clusters spatial distribution in the province of Florence.

Firenzuola and San Godenzo. The first cluster involves 45 species with a nonzero abundance vector, while the second 38. The spatial distribution of the groups in the province area does not show a clear spatial pattern (Figure 6).
Figure (7) displays the intrinsic profiles distinguishing among sites of the first (in red) and second group (in blue).
In order to highlight the specificity of a single group respect to the entire area, the functional mean profile has been estimated. Following the parametric functional approach, it is possible to calculate the mean of the relative abundance vector for each group and for the entire study area on the parameters rather than on the functions. Since the num-

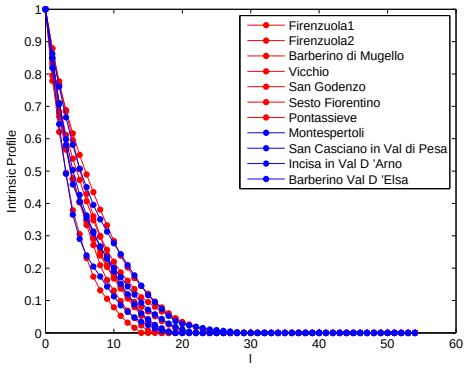

Figure 7 – T_l for sites belonging to cluster 1 (in red) and cluster 2 (in blue).

ber of lichen species is fixed ($s = 54$), the same number of lichens has been considered for each site by assuming an abundance equal to zero for species that were not present in the specific site. This assumption allows us to remove the possibility that the species observed in each site may be different. For this purpose it points out that the biodiversity decreases when the dominance of some species increases, while it increases when the equitability of the species distribution and the species richness increase. Since all species for each site have been considered, the effect of the diversity has been amplified and this allows us to highlight differences between clusters. Starting from these considerations, three mean vectors have been estimated. In particular, $\overline{\theta} = (\overline{\theta}_1, ..., \overline{\theta}_s)$ represents the mean vector of the entire study area, where $\overline{\theta}_i = \frac{1}{J}\sum_{j=1}^{J} p_{ij}$ for $i = 1, ..., s$. Whereas, defining by G the number of groups ($g = 1, ..., G$), with $G = 2$, two mean vectors, $\overline{\theta}_1$ and $\overline{\theta}_2$ have been obtained, where $\overline{\theta}_{ij} = \frac{1}{J_g}\sum_{j=1}^{J_g}$ for $g = 1, ..., G$ and $i = 1, ..., s$. Thereafter, using these estimates, the mean intrinsic profile for cluster one, cluster two and for the entire area has been built (Figure 8) applying the equation (3) to the mean vectors, after having sorted their elements in descending order.

Since the intrinsic profile requires a descending sort of the relative abundance vector, a different parameter ordering from site to site has been obtained. This condition compromises the possibility to trace a general ordering of the mean abundance vector of the entire area. Accordingly the assumption of monotonic dependence from the parameters would seem fail, together with the internality properties of the functional mean.
Indeed, the internality property of the functional mean profile is ensured within the subspace S of L^p constituted by the functional data. In fact, the functional mean profile lies between the upper and lower bound. The upper bound represents the maximum diversity which occurs when all of the observed species present the same relative abundance value ($p_i = 1/s$ for $i = 1, ..., s$). On the contrary the lower limit denotes a situation of maximum dominance, that is, when there is only one species within a community.
Figure (8) shows that the mean profile of the entire area is more diverse from that of the single groups, and this is clearly due to the increasing equitability, given that the richness is fixed. Focusing on the analysis of the two groups, also in this case, it is not possible to

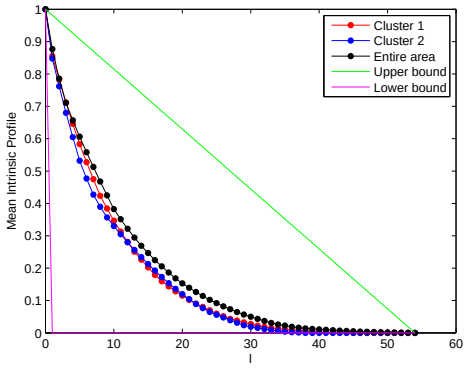

Figure 8 – Mean T_l for cluster 1 (in red), cluster 2 (in blue) and for the entire area (in black).

distinguish an ordering among the profiles since they continue to cross each other. This is probably due to the uniformity of the study area, which leads to the impossibility of identifying homogeneous site groups. To verify this, we proceeded with the analysis of the functional variability measures, decomposing the functional total variance into the within and between components:

$$Var_T(x) = Var_B(x) + Var_B(x) \tag{14}$$

In particular, the functional within variability, $Var_W(x)$, has been calculated as:

$$Var_W(x) = \sum_{g=1}^{G} V_g^2(x) \frac{J_g}{J} \tag{15}$$

where:

$$V_g^2(x) = \frac{1}{J_g} \sum_{j=1}^{J_g} \Big(f(\theta_{ig}, x) - f(\overline{\theta}_g, x) \Big)^2 \tag{16}$$

represents the functional variance of the g-th group ($g = 1, ..., G$).
Finally, the functional between variability, $Var_B(x)$, has been computed as:

$$Var_B(x) = \frac{J_g}{J} \sum_{g=1}^{G} \Big(f(\overline{\theta}_g, x) - f(\overline{\theta}, x) \Big)^2 \tag{17}$$

Figure (9) displays the single functional variances of the two groups ($V_1^2(x)$ in red and $V_2^2(x)$ in blue) together with the functional within variability ($Var_W(x)$ in black) and shows less variability in the second group than in the first. Since the horizontal axis represents the species permutations according to a descending ordering of the relative abundance vector, it is not possible to establish which species is more variable than the others. Therefore the functional variances should be interpreted according to the magnitude of the area under the variability curve. Figure (10) shows that the functional

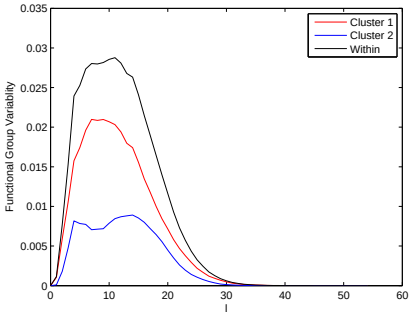

Figure 9 – Functional variability for each group and functional within variability.

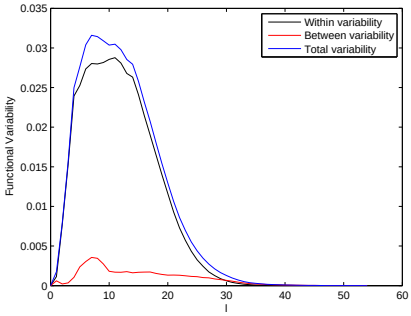

Figure 10 – Measures of variability: within, between and total functional variability.

between variability (in red) is very low while the functional within variability (in black) is almost equal to the total functional variability (in blue). This confirms the impossibility of classifying the sites of the study area considering only the abundance vector. Since the profiles are not comparable and it is not possible to distinguish exhaustive groups, the best way to study the diversity in the province of Florence is through the mean profile and the functional variability of the entire area. This application has identified homogeneity among the sites in terms of diversity. This could be due to the fact that a single province has been analyzed. In fact, more considerations could be obtained increasing the spectrum of the area by considering the entire Tuscany region.

REFERENCES

ANPA (2001). *I.B.L. Indice di Biodiversità Lichenica*. Manuali e Linee Guida. no. 2.

J. ASTA, W. ERHARDT, M. FERRETTI, F. FORNASIER, U. KIRSCHBAUM, P. L. NIMIS, O. W. PURVIS, S. PIRINTOS, C. SCHEIDEGGER, C. V. HALUWYN, V. WIRTH (2002). *Mapping lichen diversity as an indicator of environmental quality*. In P. L. NIMIS, C. SCHEIDEGGER, P. A. WOLSELEY (eds.), *Monitoring with Lichens-Monitoring Lichens*, Nato Science Program-IV, Kluwer Academic Publisher, The Netherlands, pp. 273–279.

M. E. CONTI, G. CECCHETTI (2001). *Biological monitoring: lichens as bioindicators of air pollution assessment - a review*. Environmental Pollution, 114, pp. 471–492.

F. CRISTOFOLINI, P. GIORDANI, E. GOTTARDINI, P. MODENESI (2008). *The response of epiphytic lichens to air pollution and subsets of ecological predictors: a case study from the italian prealps*. Environmental Pollution, 151, pp. 308–317.

T. DI BATTISTA, S. A. GATTONE, A. DE SANCTIS (2010). *Dealing with fda estimation methods*. In S. INGRASSIA, R. ROCCI, M. VICHI (eds.), *New Perspectives in Statistical Modeling in Data Analysis*, Springer, New York, pp. 357–365.

S. A. GATTONE, T. DI BATTISTA (2009). *A functional approach to diversity profiles*. Journal of the Royal Statistical Society, 58, pp. 267–284.

P. GIORDANI, G. BRUNIALTI, D. ALLETEO (2002). *Effects of atmospheric pollution on lichen biodiversity (lb) in a mediterranean region (liguria, nw italy)*. Environmental Pollution, 118, pp. 53–64.

J. H. GOVE, G. P. PATIL, D. F. SWINDEL, C. TAILLIE (1994). *Ecological diversity and forest management*. In G. P. PATIL, C. R. RAO (eds.), *Handbook of Statistics, vol.12, Environmental Statistics*, Elsevier, Amsterdam, pp. 409–462.

H. NAGENDRA (2001). *Using remote sensing to assess biodiversity*. Int. J. Rem. Sens., 22, pp. 2377–2400.

L. NIMIS, A. CICCARELLI, G. LAZZARIN, R. BARGAGLI, A. BENEDET, M. CASTELLO, D. GASPARO, D. LAUSI, S. OLIVIERI, M. TRETTACH (1989). *I licheni come bioindicatori di inquinamento atmosferico nell'area di schio-thiene-breganze (vi)*. Bollettino del Museo Civico di Storia Naturale di Verona.

G. P. PATIL, C. TAILLIE (1979). *An overview of diversity*. In J. F. GRASSLE, G. P. PATIL, W. SMITH, C. TAILLIE (eds.), *Ecological Diversity in Theory and Practice*, International Co-operative Publishing House, Fairland, MD, pp. 23–48.

G. P. PATIL, C. TAILLIE (1982). *Diversity as a concept and its measurement*. Journal of the American Statistical Association, 77, pp. 548–567.

E. C. PIELOU (1975). *Ecological Diversity*. John Wiley & Sons, New York.

J. O. RAMSAY, B. W. SILVERMAN (2005). *Functional Data Analysis, 2nd edn*. Springer, New York.

W. RUDIN (2006). *Real and complex analysis*. McGraw-Hill, New York.

C. SHANNON (1948). *A mathematical theory of communication*. Bell Syst. tech. J., 27, pp. 379–423.

E. H. SIMPSON (1949). *Measurement of diversity*. Nature, 163, p. 688.

S. K. THOMPSON, G. A. F. SEBER (1996). *Adaptive sampling*. John Wiley & Sons, New York.

H. F. VAN DOBBEN, H. T. WOLTERBEEK, G. W. WAMELINK, C. J. T. BRAAK (2001). *Relationship between epiphytic lichens, trace elements and gaseous atmospheric pollutants*. Environmental Pollution, 112, no. 2, pp. 163–169.

J. H. WARD (1963). *Hierarchical grouping to optimize an objective function*. Journal of the American Statistical Association, 58, pp. 236–244.

J. E. WEGMAN, W. I. WRIGHT (1983). *Splines in statistics*. Journal of the American Statistical Association, 78, pp. 351–365.

SUMMARY

Assessing biodiversity profile through FDA

The past few years have highlighted the need to protect diversity in its broader concept and, in particular, with reference to the ecological context for environmental protection. In this context, the problem of the definition and the measurement of diversity becomes fundamental. This paper provides a general picture of the main biodiversity indices proposed in literature and shows their limits in favor of diversity profiles. Since the diversity profile is a curve which expresses the diversity as a function of the relative abundance vector, it may be studied through the functional approach. In particular, we point out the advantages of the parametric FDA which, under suitable assumptions, allow us to obtain summary statistics of the same functional form of the observed data. Diversity profiles and parametric FDA approach have been applied to evaluate the diversity of epiphytic lichens in the province of Florence, providing an alternative way of understanding the biological diversity.