



Application of Ordinal Correspondence Analysis for Submerged Aquatic Vegetation Monitoring

C. Manté, David Nerini

► To cite this version:

C. Manté, David Nerini. Application of Ordinal Correspondence Analysis for Submerged Aquatic Vegetation Monitoring. Journal of Applied Statistics, 2013, 40 (8), pp.1619-1638. 10.1080/02664763.2013.789494 . hal-00912773

HAL Id: hal-00912773

<https://hal.science/hal-00912773>

Submitted on 3 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Application of Ordinal Correspondence Analysis for Submerged Aquatic Vegetation Monitoring

Claude Manté¹, Guillaume Bernard², Patrick Bonhomme³ and D. Nerini¹

¹Aix-Marseille Université, Université du Sud Toulon-Var, CNRS/INSU, IRD, MIO, UM 110, 13288, Marseille, Cedex 09, France ;²GIPREB, Cours Mirabeau - 13130 Berre-l'étang, France ; ³GIS Posidonie, Campus de Luminy - 13009 Marseille, France

Corresponding author. Fax : (+33) 491 829 119 ; email : claudemante@univ-amu.fr

Application of Ordinal Correspondence Analysis for Submerged Aquatic Vegetation Monitoring

Abstract

*The European Water Framework states that macrophyte communities (seaweeds and seagrass) are key indicators of the ecological health of lagoons. Furthermore the restoration of these communities, especially the *Zostera* meadows, is one of the main objectives of the Berre lagoon restoration plan. Consequently a monitoring program of the main *macrophyte* species still present in the lagoon was initiated in 1996. This monitoring resulted in a sequence of eleven spatially structured annual tables consisting of the observed density of these species. These tables are processed in this study.*

First, we specify the principles of Beh's Ordinal Correspondence Analysis (OCA), designed for ordered row/column categories, and compare this method to classical Correspondence Analysis (CA). Then, we show that OCA is straightforwardly adaptable for processing a sequence of ordered contingency tables like ours. Both OCA and CA are afterwards used to reveal and test the main patterns of spatiotemporal changes of two *macrophyte* species in the Berre lagoon: *Ulva* and *Zostera*. The results we obtained are compared and discussed.

Keywords: ordinal correspondence analysis, tests, Chi-squared partitioning, ecological monitoring, *Macrophytes*, Berre lagoon

1. Introduction

Since the early to mid-20th century coastal lagoons and estuaries have become among the most disturbed coastal ecosystems in the world. The initial causes are relatively well identified: eutrophication and organic pollution as a result of increasing agriculture, urbanisation in river catchments and port facilities, aquaculture, turbidity and over-sedimentation [16, 11]. Seagrass beds, which are common dwellers of these habitats, are declining throughout the world [33] at the expense of macroalgae. Indeed,

eutrophication created suitable conditions for macroalgae blooms in several coastal lagoons, bays and estuaries over the last decades [37, 7, 18]. In general, the process of eutrophication leads to a shift in the macrophytobenthic community from slow-growing seagrasses to phytoplankton and fast-growing macroalgae such as *Ulva*. The high surface-area to volume ratio of these macroalgae favours rapid nutrient uptake, high production and rapid growth rates, which enables them to outcompete the original vegetation [29, 25].

Monitoring of submerged aquatic vegetation distribution and abundance ranges from coarse assessments of presence/absence or area distribution in large areas – based on remotely sensed data and presented as macroscale maps - to fine-scale diver assessments of depth limits and cover (in % of the bottom surface area), biomass or shoot density along depth gradients [22, 6]. The cover is indeed the most frequently used measure for vegetation surveys, since it is not destructive and requires relatively little effort compared to other measurements of vegetation health.

The Berre lagoon (Provence, France) is one of the largest Mediterranean brackish lagoons (155 km²) and was occupied at the turn of the 20th century by extensive seagrass meadows (perhaps over 6000 ha). Subsequently the lagoon was disturbed by urban and industrial pollution. The monitoring of *macrophytes* in this lagoon is based on visual censuses in scuba-diving along permanent transects, each compounded of 20 spatial units. A coding in six classes (intervals of cover) was adopted for studying the temporal variations (1996 to 2008) of the cover by eight species (or groups of species) of *macrophytes*. In this paper, we will focus on only two species of cardinal importance for evaluating the health of this ecosystem: *Zostera* and *Ulva*. Our goal is double:

1. from the spatio-temporal side, we want to investigate whether or not the distance to the shoreline (indirectly associated with coastal topography, depth, currents, etc.) is related to the abundance of each species and, when the answer is positive, we will try to identify homogeneous regions where its dynamics is important (regions of interest)
2. from the purely temporal side, it is capital to investigate the ways the new policy for freshwater discharge established since 2004 (the 8th observation) impacted the abundance of each species in its regions of interest.

We chose to investigate these points through the Ordinal Correspondence Analysis (**OCA**) of Beh [1, 2, 3] and classical Correspondence Analysis (**CA**). Because of the above decision-making concerns, we will focus on the different decompositions of the Pearson chi-squared statistics associated with CA and OCA. For that purpose, we first compare decomposition and reconstitution formulas associated with both of these methods and propose original representations of the chi-squared statistics apportionment. We also propose original representations of the interaction terms in OCA, and a variant of this method suited for analyzing a sequence of doubly-ordered contingency tables.

2. Material and methods

2.1 Study site and sampling procedures

In 1966 the diversion of the Durance River towards a hydroelectric power plant and then into the Berre lagoon resulted in a heavy input of freshwater. Surface water salinity declined from 24-36 to 1-22 resulting in haline stratification, eutrophication and unstable ecological conditions [26, 27, 36]. The seagrass beds underwent a dramatic decline. In fact they are functionally extinct since 1998; in 2004 they occupied a total surface area of 1.5 ha [5]. During that time, large macroalgae blooms mainly constituted

of *Ulva* were observed in the lagoon. Consequently enormous amounts of drift algae were washed up on the shore [30, 35] (G. Bernard and P. Bonhomme pers. obs.). A new policy for freshwater discharges aiming at restoring the lagoon started up in 2004. The policy induced strong changes in the lagoon: freshwater and silt inputs from the St-Chamas hydroelectrical plant were heavily reduced. As a result, the global salinity of the lagoon increased while its stratification and the frequency of anoxia phenomena in its deepest part decreased.

[Figure 1 near here]

Thirty one permanent transects [35] (G. Bernard & P. Bonhomme, pers. obs.), 100 m long by 1 m width, and regularly distributed along the shore (see Fig. 1), were annually explored by scuba-diving, in July, from 1996 to 1998 and from 2001 to 2008 (11 years of sampling). Each transect is divided in 20 successive individual units (5m² each). The cover of *Zostera* and *Ulva* (in % of the bottom surface area) has been evaluated by visual census for each individual unit and has been coded with an ordinal scale in 6 classes, ranging from 0 to 5 (see table 1). We used a “à la Braun-Blanquet” [31] cover score adapted for marine ecology, because the field of vision of divers is reduced.

[Table 1 near here]

Thus, for each species, the data consists in 11 annual contingency tables, of common size 6×20 ; the i^{th} row of each table lists the overall frequency (≤ 31) of the abundance code $i - 1$ at each one of the 20 successive units composing transects. Thus, both the variables associated with rows and column of these tables are ordinal.

2.2 Statistical methods

Notice first that the data set associated with each *macrophyte* consist in a “cube” of size $I \times J \times K$, with $I=6$, $J=20$, $K=11$. There are in the literature a lot of statistical methods suited for such three-ways data, which can be essentially divided into two large families. The first one is a whole family of genuine three-way methods: CANDECOMP/PARAFAC, TUCKER, *etc.* [23], which are extensions of PCA to three-ways quantitative data. These methods were not designed for contingency tables, but Carlier and Kroonenberg [12] proposed an extension of CA to three-way contingency tables. Nevertheless, to our knowledge, none of these methods has been adapted to ordinal data.

The second family was designed for processing a sequence of partial tables. Roughly speaking, these methods consist in analyzing a “compromise” table built from this sequence. For instance, in the setting of Multiple Factor Analysis and related methods, the compromise is obtained by weighting the partial tables according to their dominant eigenvalues [38], while in the setting of STATIS [24] the compromise is obtained from a preliminary PCA of operators associated with the partial tables. In the same line, Foucart [15] proposed a method consisting in performing CA of the average table, the rows and columns of the partial tables being represented as supplementary elements. This method is frequently used by ecologists [28], and we will show that it is straightforwardly adaptable for ordinal data. The relationships between both these families were investigated by Kiers [20, 21].

Let us now introduce the notation used. Depending on the context, $I \times J$ will denote either the dimensions of a table of I rows and J columns, or the Cartesian product of the sets $\{1, \dots, I\}$ and $\{1, \dots, J\}$. We shall denote $\vec{1}_K$ the constant vector $(1, 1, \dots, 1)^t$ of dimension K , $U \otimes V := UV' = \left(U_i V_j \right)_{(i,j) \in I \times J}$ the tensor product of two vectors of respective length I and J , and \circ the element-wise (Hadamard) product of

matrices: $(A \circ B)_{i,j} := A_{i,j} B_{i,j}$, where $A_{i,j}$ denotes the entry of the matrix A at the i^{th} row and j^{th} column (for further details, see for instance [19]). At last, the expression “ $V \equiv \ell$ ” will mean that the random variable V obeys the distribution ℓ , and δ_m^p will denote the Kronecker symbol ($\delta_m^p = 0$ if $m \neq p$ and $\delta_p^p = 1$).

2.2.1 Correspondence Analysis in a nutshell (Benzécri, 1967; Beh, 2004; Greenacre, 1984)

It is time now to remind the reader of essential principles of CA. Let T be some $I \times J$ contingency table of grand total n , $P := T / n$ be the associated probabilities table, and $P_i := \sum_{j=1}^J P_{i,j}$ (resp. $P_j := \sum_{i=1}^I P_{i,j}$) be the marginal probabilities. The aim of CA is to highlight the ways P possibly differs from the matrix $P_I \otimes P_J$ of general entry $P_i P_j$ or, in other words, to investigate the ways the model of complete independence between rows and columns of T diverges from reality.

Practically, it consists [3, 17] in performing the Generalized Singular Value Decomposition of the matrix Θ of general entry $\theta_{i,j} := \frac{P_{i,j}}{P_i P_j}$ (Pearson ratios), giving rise to a system of singular values and singular vectors (also named principal axes) $(\lambda_m; A_m, B_m): 0 \leq m \leq M$, with $M := \min[I-1, J-1]$ (remember that $\lambda_0 = 1$, $A_0 = \overrightarrow{1_I}$ and $B_0 = \overrightarrow{1_J}$). We can now write:

$$\theta_{i,j} = \sum_{m=0}^M A_{m,i} \lambda_m B_{m,j}$$

or, using tensor products [4, 19] (see also the appendix):

$$\Theta = \sum_{m=0}^M \lambda_m A_m \otimes B_m \quad (1)$$

and also the reconstitution formula:

$$P = (P_I \otimes P_J) \circ \Theta = P_I \otimes P_J + \sum_{m=1}^M \lambda_m (P_I \otimes P_J) \circ (A_m \otimes B_m) \quad (2).$$

The singular vectors are centred and normed, *i.e.* fulfil:

$$\forall (m, p) \in M \times M, \sum_{i=1}^I A_{m,i} P_i A_{p,i} = \sum_{j=1}^J B_{m,j} P_j B_{p,j} = \delta_m^p \quad (3).$$

They also verify:

$$\lambda_m \delta_m^p = \sum_{(i,j) \in I \times J} A_{m,i} P_{i,j} B_{p,j} \quad (4).$$

Consider now the Pearson chi-squared statistics

$$X^2 := n \sum_{(i,j) \in I \times J} P_i \times P_j (\theta_{i,j} - 1)^2.$$

Because of formula (1), we obtain an alternative decomposition of this statistics:

$$\frac{X^2}{n} = \sum_{m=1}^M \lambda_m^2 \quad (5).$$

2.2.2 Correspondence Analysis with ordered categories

Since the principal axes issued from CA do not take into account the order of the categories associated with the rows and the column of P , Beh [1] proposed an alternative decomposition of X^2 , where the role of principal axes is played by orthonormal polynomials. To compare his Ordinal CA (OCA) with CA, we will use the same notations as in the previous section, denoting $\{A_m : 1 \leq m \leq I-1\}$ the system of row polynomials, and $\{B_p : 1 \leq p \leq J-1\}$, the one of column polynomials. Each one of these systems only depends on a **score** [2], and on the associated marginal probability, either P_I or P_J . The chosen score reflects the ordered structure of the categories. The polynomials, resulting from Emerson's recurrence formula [14, 2], are centred and verify:

$$\forall (m, p) \in (I-1) \times (J-1), \sum_{i=1}^I A_{m,i} P_i A_{p,i} = \sum_{j=1}^J B_{m,j} P_j B_{p,j} = \delta_m^p \quad (6)$$

There are substantial differences between principal axes of CA and these polynomials: as singular vectors, axes are **paired**, as shown in formula (1); moreover they are naturally **sorted** according to their common variance. Emerson's polynomials don't possess such properties, and the decomposition of the variance associated with OCA includes all the **bivariate moments** between row and column polynomials, giving rise to a whole $(I-1) \times (J-1)$ matrix $\Lambda_{Rows, Columns}$, of general entry:

$$\lambda_{m,p} := \sum_{(i,j) \in I \times J} A_{m,i} P_{i,j} B_{p,j} \quad (7).$$

We have also a reconstitution formula [1]:

$$P = P_I \otimes P_J + \sum_{(m,p) \in (I-1) \times (J-1)} \lambda_{m,p} (P_I \otimes P_J) \circ (A_m \otimes B_p) \quad .$$

While the interest of this saturated formula is only theoretical, it can be coupled with tests in order to get some “de-noised table”. More precisely, suppose a set $S \subset I \times J$ of relevant interactions has been determined. The associated de-noised table is given by the unsaturated reconstitution formula [1]:

$$\tilde{P}_S = P_I \otimes P_J + \sum_{(m,p) \in S} \lambda_{m,p} (P_I \otimes P_J) \circ (A_m \otimes B_p) \quad (8).$$

Notice the similarity between formulas (7) and (4), (3) and (6), and (8) and (2).

Representation of a sequence of ordered tables

For this purpose, we will straightforwardly adapt Foucart's Correspondence Analysis [15, 28]. Remember that the processed table is the sum of 11 annual partial tables, with identical rows and columns (“slices”) of the data cube $T := \sum_{k=1}^K T^k$, and let us denote $\pi^k := \sum_{(i,j) \in I \times J} T_{i,j}^k$ the sum of the k^{th} contingency table, and $P^k := T^k / \pi^k$ the associated k^{th} partial probability. The table analyzed [15, 28] is $P = \sum_{k=1}^K \alpha^k P^k$, with

$\alpha^k = \pi^k / \sum_{k=1}^K \pi^k$. While Foucart [15] analyzed T through CA, we will use OCA instead.

Let $B := (B_0, \dots, B_{J-1})$ denotes the matrix of column polynomials, and $A := (A_0, \dots, A_{I-1})$ the corresponding matrix for rows. The column profile coordinates are given by: $G := D_{P_j}^{-1} P A$ [1]; consequently, because $G = \sum_{k=1}^K \alpha^k D_{P_j}^{-1} P^k A$, the column profiles obtained from the OCA of P are situated at the barycentre of the partial probabilities projected as supplementary elements.

2.2.3 Tests

The unique rigorous independence test in CA is based on X^2 which, under the hypothesis (**H**) of independence of the columns and rows of T , asymptotically obeys a chi-square distribution: $X^2 \cong \chi^2((I-1)(J-1))$. Benzécri [4] derived from this statistics and (5) a heuristic for determining the number of significant principal axes in CA of contingency tables. It consists in seeking for the smaller m such that, for some fixed $\tau \in]0, 0.5[$:

$$n \sum_{k=1+m}^M \lambda_k^2 < Q(\chi^2((I-1)(J-1)), \tau) \quad (9)$$

where the threshold $Q(\ell, \tau)$ denotes the quantile of order τ of the probability law ℓ (Benzécri [4] merely proposed to use the mean $(I-1)(J-1)$ as a threshold). This heuristic is useful for data representation and reconstitution [4, 1, 3] from the set of significant axes.

Contrary to CA, OCA gives rise to a three-level battery of tests. Firstly, at the cell-level, Rayner & Best [32] proved that under (H), we have asymptotically:

$$\forall (m, p) \in (I-1) \times (J-1), \sqrt{n} \lambda_{m,p} \cong N(0,1) \quad (10)$$

and that all such bivariate moments are asymptotically independents. Thus, since every polynomial $\{A_m : 1 \leq m \leq I-1\}$ can be correlated with every $\{B_p : 1 \leq p \leq J-1\}$, we will use (10) to test whether or not each interaction can be considered as significant.

At the intermediary level we can also test whether or not each row polynomial A_m (respectively column polynomial B_p) significantly contributes to the Pearson chi-squared statistics, since we have under (H):

$$\eta_m := \sum_{p=1}^{J-1} \lambda_{m,p}^2 \cong \frac{\chi_{(J-1)}^2}{n} \quad (11)$$

$$\mu_p := \sum_{m=1}^{I-1} \lambda_{m,p}^2 \cong \frac{\chi_{(I-1)}^2}{n} \quad (12).$$

Finally, at the global level, we find again the classical test, associated with a third decomposition:

$$\frac{X^2}{n} = \sum_{m=1}^{I-1} \eta_m = \sum_{p=1}^{J-1} \mu_p.$$

Thus, η_m (resp. μ_p) is the part of the total information bore by A_m (resp. B_p), and these quantities are relevant criteria for selecting best projection planes for the columns (resp. the rows) profiles (because of the duality between rows and columns).

3. Data analysis

3.1 Scores and weights

As we said in the previous section, the system $\{A_m : 1 \leq m \leq I-1\}$ of row polynomials only depends on P_I and on a score s_I reflecting the gaps between successive categories of cover. Similarly, the system $\{B_p : 1 \leq p \leq J-1\}$ only depends on the pair (P_J, s_J) . Because all the individual units of the transect have the same length, we chose the “natural score” [2] $s_J(k) := k$ to build the column polynomials. On the contrary, since the cover modalities are not equally spaced, we defined $s_I(k)$ as the centre of the $(k-1)^{\text{th}}$ interval of cover if $k \geq 2$, while $s_I(1) := 0$ (see Table 1). Notice

that, due to the sampling conditions, $P_j = \frac{1}{J} \overline{1_j}$ does not depend on the data, while P_i does.

3.2 Spatial and temporal dynamics of *Ulva*

We will now analyze the 6×20 contingency table $T := \sum_{k=1}^{11} T^k$ obtained by summing the 11 annual “slices” of the cube of type $6 \times 20 \times 11$ associated with the data. This will be done first through CA, and afterwards through OCA; issues from the second analysis will be explained in greater detail.

3.2.1 Results of CA of the *Ulva* data

The overall Pearson statistics was 346.05: this is highly significant since, under (H), $X^2 \cong \chi^2(95)$. To determine the number of principal axes to keep, we fixed in formula (9) either $\tau = 0.05$, or $\tau = 0.5$. In both cases, formula (9) indicated that only two axes are significant, associated respectively with 67.2% and 13.5% of the total inertia. The corresponding components are plotted on Figure 2; the individual units (in grey) are labelled by their distance to the shoreline along the sampling transects (from 0-5m to 95-100m), while the cover is labelled by the original code (from 0 to 5).

[Figure 2 near here]

We can infer from Figure 2 that the first 20 meters of the transects are equally associated with the codes {2,3,4,5} and concentrates most of the *Ulva* abundances trough the time series, while the last 50 meters are associated with the absence of *Ulva* (code 0).

3.2.2 Results of OCA

We will first use formulas (11) and (12) to determine the relevant polynomials. On Figure 3 are plotted the contributions $\{n \hat{\mu}_j : 1 \leq j \leq J-1\}$ on the left, and $\{n \hat{\eta}_i : 1 \leq i \leq I-1\}$ on the right; on these plots, grey and black lines respectively mark

the quantiles of order 0.5 and 0.95 of the associated chi-squared distributions. It is noteworthy that while $\hat{\mu}_1$ corresponds to about 60.5% of total variance (close to the first principal axis: 67.2%), $\hat{\eta}_1$ corresponds to only 29.4% of the same quantity. This stresses the complexity of column profiles of this table.

[Figure 3 near here]

On Figure 3, the only significant column polynomials are B_1 , B_2 and, at the very outside, B_3 , while all the $\{A_m : 1 \leq m \leq 5\}$ seem statistically significant. Since in both cases both the first polynomials correspond to higher variances, it is reasonable to simultaneously display on Figure 4 the row profiles projected on the plane (B_1, B_2) , and the column profiles on the plane (A_1, A_2) , i.e. the location and dispersion components in the terminology of Beh [1].

[Figure 4 near here]

Note that, even if Figure 4 is quite different from Figure 2, its interpretation is similar: the same clusters of rows and columns are evidenced in both plots.

Issues of the tests associated with formula (10) are represented on Figure 5. On this figure is represented a table of type $\Lambda_{Density, Space}$ of bivariate moments, where each $\lambda_{m,p}$ is associated with some cell whose grey level depends on its estimated value.

Precisely, each statistics $Gl(m, p) := \frac{1}{\tau} \text{Min} \left[\tau, P \left(\chi_1^2 > n \left(\hat{\lambda}_{m,p} \right)^2 \right) \right] \in [0,1]$ has been computed and coded in gray levels for producing Figure 5 (and Figure 12 too).

[Figure 5 near here]

On this figure, non-significant bivariate moments (bilateral test at the threshold 0.05) correspond to white cells ($Gl(m, p) \approx 1$), while highly significant ones

($Gl(m, p) \approx 0$) correspond to black ones, and moderate interactions correspond to grey cells. It is noteworthy that issues of these tests have implications for simultaneous representation of the row and column. For instance, a glance on Figure 5 indicates that the simultaneous representation of rows and columns on Figure 4 would undoubtedly make sense if the bivariate moments $\hat{\lambda}_{1,1}$ and $\hat{\lambda}_{2,2}$ were high, while $\hat{\lambda}_{1,2}$ and $\hat{\lambda}_{2,1}$ would be low; since it is not the case, we must be cautious! Furthermore, one should also be careful with polynomials of degree greater than 4, because the estimation variance of these moments can be excessive [32]. We can indeed see on Figure 5 that no column polynomial of degree greater than 3 is very important for *Ulva*, excepted B_8 , since $\hat{\lambda}_{4,8}$ is highly significant. The very dominant term in this decomposition is $\hat{\lambda}_{1,1} \approx -8.4$, corresponding to about 20.5% of the total inertia. Thus, the salient feature of T is that the more we move away from the shoreline, the more the cover by *Ulva* decreases (see the left panel of Figure 6).

[Figure 6 near here]

The second dominant interaction is the quadratic by linear one $\hat{\lambda}_{2,1} \approx 7.35$. In this case, when we move away from the shoreline (see the right panel of Figure 6), the cover, which is initially medium, gradually concentrates on extreme values: the dispersion increases. More insights about interactions are provided in the appendix.

The set S of significant interactions, displayed on Figure 5, enabled us to reconstitute the table by using formula (8); the result is displayed on Figure 7.

[Figure 7 near here]

Finally, we display on Figure 8 the spatio-temporal variations of cover by *Ulva*, projected onto the plane (A_1, A_2) . Each panel of this figure represent the variations of

cover of an individual unit of the transect. Each year is represented by its index, and the coordinates of the average point are the same as in Figure 4. The reader can see on Figure 8 that most of the dynamics of *Ulva* took place in the 20 first meters of the transect, which is consistent with the maximal abundances observed (see Figures 2&4). To save place, we discarded the units from 50 to 100m, where the dynamics is less important.

[Figure 8 near here]

Note that until now, the time was **passive** (projection of supplementary elements). But, since the units 0 to 20m consist in an homogeneous region, we can make the time play an **active** part by considering the “transposed cube” of type $I \times K \times J$ ($I=6$, $J=20$, $K=11$), and keeping only its first four slices $\{U^1, \dots, U^4\}$. We will then analyse the 6×11 contingency table $U := \sum_{k=1}^4 U^k$ in the same way as T ; notice that while the row polynomials remain the same as in the previous analysis, the column polynomials $\{B_p : 1 \leq p \leq J-1\}$ are now associated with **time**, which is thus **active**. To save place, we will only mention the main issues of OCA and CA of U .

The overall chi-squared statistics associated with U is highly significant (under (H), $X^2=527.8$ should obey $\chi^2(50)$). All the row polynomials are significant, and the highest contributions of polynomials, $\hat{\mu}_1$ and $\hat{\eta}_1$ correspond to only 44% of total variance, while $\hat{\mu}_2$ and $\hat{\eta}_2$ correspond to respectively 15% and 9.8% of total variance. Thus, CA seems more efficient than OCA for displaying the data: the first eigenvalues of CA corresponded to 83.6% and 10.12% of the total variance while, according to the heuristic (9), the next ones were not significant. We can see on Figure 9 that through the whole time series, the *Ulva* abundance increased trough time, from 1996 to the years

2001 and followings (unfortunately, 1999 and 2000 are missing). The year 2005 is singular, while 2002 is associated with extreme abundances of *Ulva*.

[Figure 9 near here]

From the inference point of view, complementary insights can be derived from OCA, thanks to the table of bivariate moments (of type $\Lambda_{Density, Time}$). The most important moments are: $\hat{\lambda}_{5,1} \approx 10.2$, $\hat{\lambda}_{1,1} \approx 8.5$ and $\hat{\lambda}_{1,10} \approx 8.8$. Thus, the most important and reliable interactions are associated with B_1 , i.e. the flow of time. Since $\hat{\lambda}_{1,1} > 0$ indicates an increasing linear by linear interaction, the significance of this moment confirms that the abundance of *Ulva* increased with time in the lagoon. The moment $\hat{\lambda}_{5,1}$ is also very important, and corresponds to a progressive regularization of the cover in the region of interest: extreme values (cover around 20% or 80%) got rarer, while medium values (about 50%) became more frequent.

3.3 Spatial and temporal dynamics of *Zostera*

We proceeded the same way for *Zostera* as for *Ulva*, processing first the table $T := \sum_{k=1}^{11} T^k$ successively through CA and OCA.

3.3.1 Results of CA of the *Zostera* data

Notice first that the overall Pearson statistics is now 228.4: this is significant for some $\chi^2(95)$, but remember it was 346.05 for *Ulva*. Correlatively, we found from (9) that a single principal axis is significant, which is associated with 75.3% of the total inertia. The first principal plane is displayed on Figure 10.

[Figure 10 near here]

This plane essentially highlights that *Zostera* was mostly absent in the lagoon (the code 0 represent 97.3% of the observations) and that the highest modalities of cover

were extremely rare for this species; it stresses that most *Zostera* abundances were observed in the first part of the sampling transects from 5 to 25 m, from the shore, in very shallow waters.

3.3.2 Results of OCA of the *Zostera* data

One can see on Figure 11 that four column polynomials $\{B_1, B_2, B_5, B_6\}$ and two row polynomials $\{A_1, A_5\}$ are statistically significant. Consequently, the best representation of the column profiles is obtained in the plane (A_1, A_5) . The row profiles can also be displayed on the associated plane (B_1, B_5) , but notice that the planes (B_1, B_2) and (B_1, B_6) would convey as much information.

[Figure 11 near here]

We separately plotted on Figure 12 the row and column profiles because, as one can see on Figure 13, simultaneous representation could be dubious: $\hat{\lambda}_{5,5}$ is hardly significant, while $\hat{\lambda}_{1,5}$ and $\hat{\lambda}_{5,1}$ are really big.

[Figure 12 near here]

The reader can also see on Figure 13 that only a few of the polynomials of degree greater than 4 are important for *Zostera*. Nevertheless, the very dominant term in the decomposition is $\hat{\lambda}_{1,1} \approx -7.4$, corresponding to about 23.8% of the total inertia. Thus, the salient feature of T is that the more we move away from the shoreline the more the cover by *Zostera* decreases. It is noteworthy that the same phenomenon appears for *Ulva* and *Zostera* with similar proportions of variance.

[Figure 13 near here]

The second important interaction in this $\Lambda_{Density, Space}$ table is the quintic by linear one $\hat{\lambda}_{5,1} \approx -4.6$. Plotting the associated function, we saw that this interaction is dominated by very low values of the cover due to the rarity of *Zostera*. Indeed, the probability of not observing it decreases when we move away from the shoreline. But we also have $\hat{\lambda}_{1,5} \approx 4.2$ and $\hat{\lambda}_{1,6} \approx -4.5$. We noted that in these two cases, interactions are dominated by low values of the cover. In these cases, the probability of not observing *Zostera* is highly oscillating along the transect, without any trend. This erratic spatial structure is probably linked with the rarity of *Zostera*, and its patchiness. On the other hand, notice that $\hat{\mu}_1$ corresponds to about 38.7% of total variance (much less than the first principal axis of CA: 75.3%), and $\hat{\eta}_1$ corresponds to 53.6% of the same quantity.

Thanks to formulas (8) and (10), the table could be reconstituted from the set S of significant interactions, displayed on Figure 13. Since the “independence structure” $P_I \otimes P_J$ is very strong, we preferred to plot the discrepancy of P from independence, $P - P_I \otimes P_J$, and its approximation

$$\tilde{P}_S - P_I \otimes P_J = \sum_{m \in S} \lambda_m (P_I \otimes P_J) \circ (A_m \otimes B_m).$$

One can see on Figure 14 how the original discrepancy is filtered this way, and how much the position of null densities on the transect is important.

[Figure 14 near here]

Next, we display on Figure 15 the temporal variations of cover by *Zostera*, projected onto the optimal plane (A_1, A_5) . We discarded the segment 50-100m, where there is practically no *Zostera*.

[Figure 15 near here]

We can see on Figure 15 that for *Zostera* too, through the whole time series, most part of the dynamics took place in the 20 first meters of the transect, very close to the shoreline, and that these four units were those where this species was most abundant (see Figures 10&12). Thus we will proceed the same way as for *Ulva*, keeping only the first four slices of the transposed cube for investigating further the temporal variations of *Zostera*. To save place, we will only mention the main results of OCA and CA of *U* .

The overall chi-squared statistics associated with this table is hardly significant ($X^2=70.16$ while, under (H), it would obey $\chi^2(50)$); remember that in the case of *Ulva* it was 527.8! The decomposition of *U* through OCA is also very different from that obtained for *Ulva*: only the two first row and column polynomials are significant. They correspond to moderate percentages of the total variance: 25.2 and 29.1% for the column polynomials, 26.5 and 27.6% for the row polynomials. In the case of CA, only the first component (62.7%) is significant; we will use OCA for displaying the data on Figure 16, where the observations posterior to 2004 are clearly opposed to the previous ones.

[Figure 16 near here]

The very dominant bivariate moment in the $\Lambda_{Density, Time}$ table is $\hat{\lambda}_{1,1} \approx -3.4$, corresponding to 16% of the total variance (Pvalue= 8.10^{-4}). The significance of this moment shows that the abundance of *Zostera* decreased with time in the lagoon. The moments $\hat{\lambda}_{1,2}$ and $\hat{\lambda}_{2,2}$ are significant too, but with much bigger Pvalues (0.037 and 0.045); each of them is associated with about 6% of total variance.

Thus, through the whole time series, there is a significant decrease of *Zostera* abundance, even if it remains in low levels of abundances, close to 0. On Figure 16, the

years 2004 to 2008 are associated with the score 0, indicating a slight but indubitable rarefaction of *Zostera*.

3.4 Ecological significant results for submerged aquatic vegetation monitoring

Among submerged aquatic vegetation, *Ulva* and *Zostera* are identified as biological indicators of ecological status in the Water Framework Directive for transition waters. As a consequence, their monitoring concerns most of the lagoon's managers worldwide. We have shown that the abundances of *Ulva* and *Zostera* in the Berre lagoon are statistically dependent of the distance to the shore, and that this spatial structuring is stronger but more complicated for *Ulva* than for *Zostera*. Most of the cover of these two species is located in the first 20 meters of the sampling transects, very close to the shore in very shallow water. The main part of the variability is also concentrated along the first 20 meters of the sampling transects, and remain high up to 50m from the shore. This is probably due to the *Ulva* way of growing, which spends most of its life-cycle as a free drift, transported by the currents and accumulated on the shore depending on the hydrodynamic conditions. As a consequence, for each year, bathymetry and coast-line topology could be related to *Ulva* density in a point. Concerning *Zostera*, the present day extent of the seagrass meadows in the lagoon, restricted to very shallow waters, suggest light as the limiting factor [37, 5]. Light reduction could be due either to the silt input being still too high or to sediment resuspension. It could also be related to eutrophication of the lagoon which results in micro- and macro-algae blooms together with high levels of colonization of *Zostera* leaves by epiphytes and light reduction for the seagrass meadows.

Trough time, *Ulva* cover shows a clear increase from 1998 to 2008 despite a high interannual variability: years with high *Ulva* abundances (2002) can be distinguished from years with weak abundances (1996-1998). From the first years of the

time-series, the *Zostera* cover remained very weak and the abundances decrease with time. There is no clear sign of re-colonization since 2005, when the environmental pressure (inputs of freshwater, silts and nutrients) have been strongly reduced. The reasons could be, in addition to continuing nutrient inputs, the resuspension of silt, no longer trapped under the seagrass canopy, during wind episodes, which are frequent in the area, and/or the release of nutrients from the bare silt habitat, which would constitute an indication of a possible hysteresis of the system. Competition between the two species could also be one of the constraining factors for *Zostera* growth, the suffocation by an enteromorph-like *Ulva* of a *Zostera* meadow has been described in the literature [34, 8, 9, 13, 10].

4. Conclusion

We studied an 11 year long sequence of the abundance of *Ulva* and *Zostera* in a heavily disturbed Mediterranean lagoon, the Berre lagoon, through the Ordinal Correspondence Analysis of Beh. The results are compared with those issued from classical Correspondence Analysis. Both of these methods consist in distinct decompositions of the Pearson chi-squared statistics and of the corresponding contingency table, according to different systems.

In the case of CA, since the systems consist in singular vectors, there is no interaction between factors of different orders, which is optimal in a purely descriptive perspective. The price to pay lies in the complexity of the factors.

The systems associated with OCA are very simple (orthogonal polynomials), but it takes into account numerous interactions between row and column systems, giving rise to representation difficulties (simultaneous projections, for instance). Consequently, we propose a number of graphical guidelines (χ^2 plots, interactions plots, partial table

plots) to help the user to select the relevant components and interpret the associated interactions.

Indeed the main advantages of OCA over CA (in addition to simplicity of the factors) is that it takes into account the ordinal nature of row and/or column variables, and that it is possible to test the significance of all the pairs of factors. We could thus obtain denoised tables given by the unsaturated reconstitution formula, and determine significant patterns of spatial (firstly) and temporal (secondly) changes of *Ulva* and *Zostera* abundance.

The spatial location of *Ulva* and *Zostera* seems to be strongly dependant on the distance to the shore, and indirectly to the depth. Main abundances and variability are observed very close to the shore since this is where free-floating *Ulva* drift can accumulate and limit light for the deeper growing *Zostera*. Through time *Ulva* shows a clear increase of its abundances while *Zostera* remained in very weak abundances; this suggests a high level of eutrophication despite recent reduction of nutrient inputs in the lagoon.

Acknowledgements

The authors wish to thank the referees for their stimulating comments and suggestions, and Starrlight Augustine for help with the English text.

References

- [1] E. J. Beh, *Simple Correspondence Analysis of ordinal cross-classifications using orthogonal polynomials*, Biom. J. 39, 5 (1997), pp. 589-613.
- [2] E. J. Beh, *A comparative study of Scores for Correspondence Analysis with ordered categories*, Biom. J. 40, 4, (1998), pp.413-429.
- [3] E. J. Beh, *Simple Correspondence Analysis : a bibliographic review*, Internat. Stat. Rev. 72, 2, (2004), pp. 257-284.
- [4] J. P. Benzecri, *L'Analyse des Données*, Vol. 2, 2nd edn. Dunod, Paris, 1976.
- [5] G. Bernard, CF. Boudouresque, and P. Picon. *Long term changes in Zostera meadows in the Berre lagoon (Provence, Mediterranean Sea)*, Estuar., Coast. Shelf Sci. 73 (2007), pp. 617-629.
- [6] CF. Boudouresque, G. Bernard, P. Bonhomme, E. Charbonnel, G. Diviacco, A. Meinesz, G. Pergent, C. Pergent-Martini, S. Ruitton, and L. Tunesi, *Préservation et conservation des herbiers à Posidonia oceanica*, RAMOGE publ., Monaco, 2006.
- [7] J.L. Bowen and I. Valiela, *The ecological effects of urbanization of coastal watersheds: Historical increases in nitrogen loads and eutrophication of Waquoit Bay estuaries*, Canad. J. Fish. Aquat. Sci. 58 (2001), pp. 1489–1500.
- [8] F.G. Brun, I. Hernández, and J.J. Vergara Pérez-Lloréns, *Resource allocation and proteolytic activity in the seagrass Zostera noltii Hornem. subjected to light reduction by Ulva canopies*, Funct. Plant Biol. 30,5 (2003b), pp. 551-560.
- [9] F.G. Brun, G. Navarro, J.J. Vergara J.J., I. Hernández, and J.L. Pérez-Lloréns, *Effect of shading by Ulva canopies on growth and carbon balance of the seagrass Zostera noltii*, Mar. Ecol. Progr. Ser. 265 (2003c), pp. 85-96.
- [10] M.L. Cambridge, J.R. How, P.S. Lavery, and M.A. Vanderklift, *Retrospective analysis of epiphyte assemblage in relation to seagrass loss in a eutrophic coastal embayment*, Mar. Ecol. Progr. Ser. 346 (2007), pp. 97-107.
- [11] P.G. Cardoso, M.A. Pardal, A.I. Lillebø, S.M. Ferreira, D. Raffaelli, and J.C. Marques, *Dynamic exchanges in seagrass assemblages under eutrophication and implications for recovery*, J. Exp. Mar. Biol. Ecol. 302 (2004), pp. 233-248.
- [12] A. Carlier and P. M. Kroonenberg, *Decompositions and biplots in three-way correspondence analysis*, Psychometrika, 61, 2 (1996), pp. 355-373.
- [13] M.L. De Casabianca, M. Rabotin, and R. Rigault, *Preliminary results on eelgrass regression and red seaweed dominance under increasing eutrophication (Thau Lagoon, France)*, Acta Adriat. 44 (2003), pp. 33- 40.
- [14] P.L. Emerson, *Numerical construction of orthogonal polynomials from a general recurrence formula*, Biometrics 24 (1968), pp. 695-701.
- [15] T. Foucart, *Sur les suites de tableaux de contingence indexés par le temps*, Statistique et analyse des données, [3](#), [2](#) (1978), pp. 67-85
- [16] W.B.J.T Giesen, M.M. Katwijk van, and C. Hartog den., *Eelgrass condition and turbidity in the Dutch Wadden Sea*, Aquat. Bot. 37 (1990), pp. 71-85.
- [17] M. J. Greenacre, *Theory and Applications of Correspondence Analysis*, Academic Press, Orlando, 1984.
- [18] J. Hauxwell, J. Cebrian, and I. Valiela, *Light dependence of Zostera marina annual growth dynamics in estuaries subject to different degrees of eutrophication*, Aquat. Bot. 84 (2006), pp. 17-25.
- [19] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

- [20] H. A. L. Kiers, *Comparison of “Anglo-saxon” and “French” three-mode methods*, Statist. Anal. Donn. 13, 3 (1988), pp. 14-32.
- [21] H. A. L. Kiers, *Hierarchical relations among three-way methods*, Psychometrika 56, 3 (1991), pp. 449-470.
- [22] D. Krause-Jensen, A.L. Quaresma, A. Cunha, and T.M. Greve, *How are seagrass distribution and abundance monitored?*, in *European seagrasses-an introduction to monitoring and management*, J. Borum, C.M. Duarte, D. Krause-Jensen and T. Greve, eds, Publication by EU Monitoring and Managing of European Seagrasses EVK3-CT-2000-00044, 2004, pp 45-54.
- [23] P. M. Kroonenberg, *Three-mode Principal Component Analysis*, DSWO Press, Leiden, 1989.
- [24] C. Lavit, *Analyse Conjointe de Tableaux Quantitatif*, Masson, Paris, 1988.
- [25] E.J. Malta, J.M. Verschuure, and P.H. Nienhuis, *Regulation of spatial and seasonal variation of macroalgal biomass in a brackish, eutrophic lake*, Helgol. Mar. Res. 56 (2002), pp. 211–220.
- [26] D. Nérini, *Analyse statistique de processus physiques et chimiques en océanologie côtière à l'aide d'une méthode de régression et de classification par arbre décisionnel. Application à l'étude d'un milieu fortement perturbé: l'étang de Berre*, PhD diss., Université de la Méditerranée, 2000.
- [27] D. Nérini, J.P. Durbec, and C. Manté, *Analysis of oxygen rate time series in a strongly polluted lagoon using a regression tree method*, Ecol. Model. 133 (2000), pp. 95–105.
- [28] S. Pavoine, J. Blondel, M. Baguette, and D. Chessel, *A new technique for ordering asymmetrical three-dimensional data sets in Ecology*, Ecology 88, 2 (2007), pp. 512-523.
- [29] M. F. Pedersen and J. Borum, *Nitrogen dynamics of eelgrass *Zostera marina* during a latesummer period of high growth and low nutrient availability*, Mar. Ecol. Progr. Ser. 80 (1992), pp. 65-73.
- [30] C. Pergent-Martini, R. Semroud, V. Rico-Raimondino, and G. Pergent, *Localisation et évolution des peuplements de phanérogames aquatiques de l'étang de Berre (Bouches du Rhône-France)*, in *Proceedings of the 39th national meeting of the Association Française de Limnologie*, Université de Corse & Agence de l'Eau Rhône-Méditerranée-Corse publ., Fr., 1995, pp. 169-179.
- [31] J. Podany, *Braun-Blanquet's legacy and data analysis in vegetation science*, Journal of Vegetation Science, 17 (2006), pp. 113-117.
- [32] J.C.W. Rayner and D.J. Best, *Smooth extensions of Pearson's product moment correlation and Spearman's Rho*, Statist. Probab. Lett. 30 (1996), pp. 171-177.
- [33] F.T. Short and S. Wyllie-Echeverria, *Natural and human-induced disturbance of seagrasses*, Environ. Conserv. 23, 1 (1996), pp. 17-27.
- [34] K. Silberstein, A.W. Chiffings, and A.J. Mc Comb, *The loss of seagrasses in cockburn sound, Western Australia. III. The effect of epiphytes on productivity of *Posidonia australis* Hook f.*, Aquat. Bot. 24 (1986), pp. 355–371.
- [35] D. Soltan, and P. Francour, *Monitoring system of benthic macrophytes communities in Berre and Vaine ponds: development of a new strategy*, J. Rech. Océanogr. 25 (2000), pp. 53-57.
- [36] G. Stora, A. Arnoux, and M. Galas, *Time and spatial dynamics of Mediterranean lagoon macrobenthos during an exceptionally prolonged interruption of freshwater inputs*, Hydrobiol. 300-301 (1995), pp. 123-132.

- [37] I. Valiela, G. Collins, J. Kremer, M. Lajtha, M. Geist, B. Seely, J. Brawley, and C.H. Sham, *Nitrogen loading from coastal watersheds to receiving estuaries: Review of methods and calculation of loading to Waquoit Bay*, Ecol. Appl. 7 (1997), pp. 358–380.
- [38] A. Zarraga, and B. Goitisolo, *Simultaneous analysis and multiple factor analysis for contingency tables: two methods for the joint study of contingency tables*, Comput. Statist. Data Anal. 53 (2009), pp. 3171-3182.

Appendix

1. *Why tensor products?*

Since $U \otimes V := UV'$, one should think that only regular matrix-vector algebra is necessary for our purpose. Nevertheless, we preferred to use the tensor product notation for three reasons:

1. as Benzécri noticed [4, p. 58], this is a natural notation for bilinear applications, and it is straightforwardly extensible to three (or more) ways tables (see for instance [12] or most papers about PARAFAC)
2. suppose E is a vector space equipped with the basis $\{e_1, \dots, e_m\}$ and F is another vector space, equipped with the basis $\{f_1, \dots, f_p\}$; then $E \otimes F$ is a third vector space, equipped with the basis $\{e_i \otimes f_j : 1 \leq i \leq m, 1 \leq j \leq p\}$: this perfectly fit the bivariate moments tables structure (see figures 5 and 12)
3. the tensor product of two vectors is straightforwardly extensible to functional spaces: if $f(y) \in H_1$ and $g(x) \in H_2$, the tensor product $f \otimes g \in H_1 \otimes H_2$ of these functions is $f \otimes g(y, x) := f(y)g(x)$; we will use such products in the next paragraph.

2. *About bivariate moments*

Remember the aim of OCA is to highlight the ways P possibly differs from the matrix $P_I \otimes P_J$. As we saw in Section 2.2.2, this difference can be decomposed (see formula (8)) in a sum of contributions of general form: $\lambda_{m,p} (P_I \otimes P_J) \circ (A_m \otimes B_p)$.

Notice that in that expression, the vectors A_m and B_p actually stem from the sampling of orthonormal polynomials. The column polynomial B_p is obtained from a recurrence formula [2], where the score $s_I(i)$ plays the part of the abscissa in Emerson's

recurrence formula [14], while the score $s_j(j)$ plays the part of the ordinate for obtaining the $A_m(y)$. Consequently, it is natural to define the polynomial $A_m \otimes B_p(y, x)$ on the rectangle $[s_j(1), s_j(J)] \times [s_I(1), s_I(I)]$. Proceeding the same way with the sampled probability, we can finally describe each interaction with the bi-dimensional function $\lambda_{m,p} (P_I \otimes P_J)(y, x) (A_m \otimes B_p)(y, x)$ defined on this domain. As an illustration, we have shown on Figure 6 the dominant interactions obtained in the analysis of the spatio-temporal distribution of *Ulva* (see section 3.2.2).

Table 1. Classes and scores corresponding to the cover of the bottom surface area by *Ulva* and *Zostera* for each 5m² successive units of the 31 permanent transects along the Berre lagoon shore. See 3.1 for the definition of the scores.

Cover (%)	0	< 5	5 to 20	20 to 50	50 to 90	> 90
Class	0	1	2	3	4	5
Score	0	2.5	12.5	35	70	95

Figure 1. location of the permanent transects along shore for monitoring the submerged aquatic vegetation in the Berre lagoon.

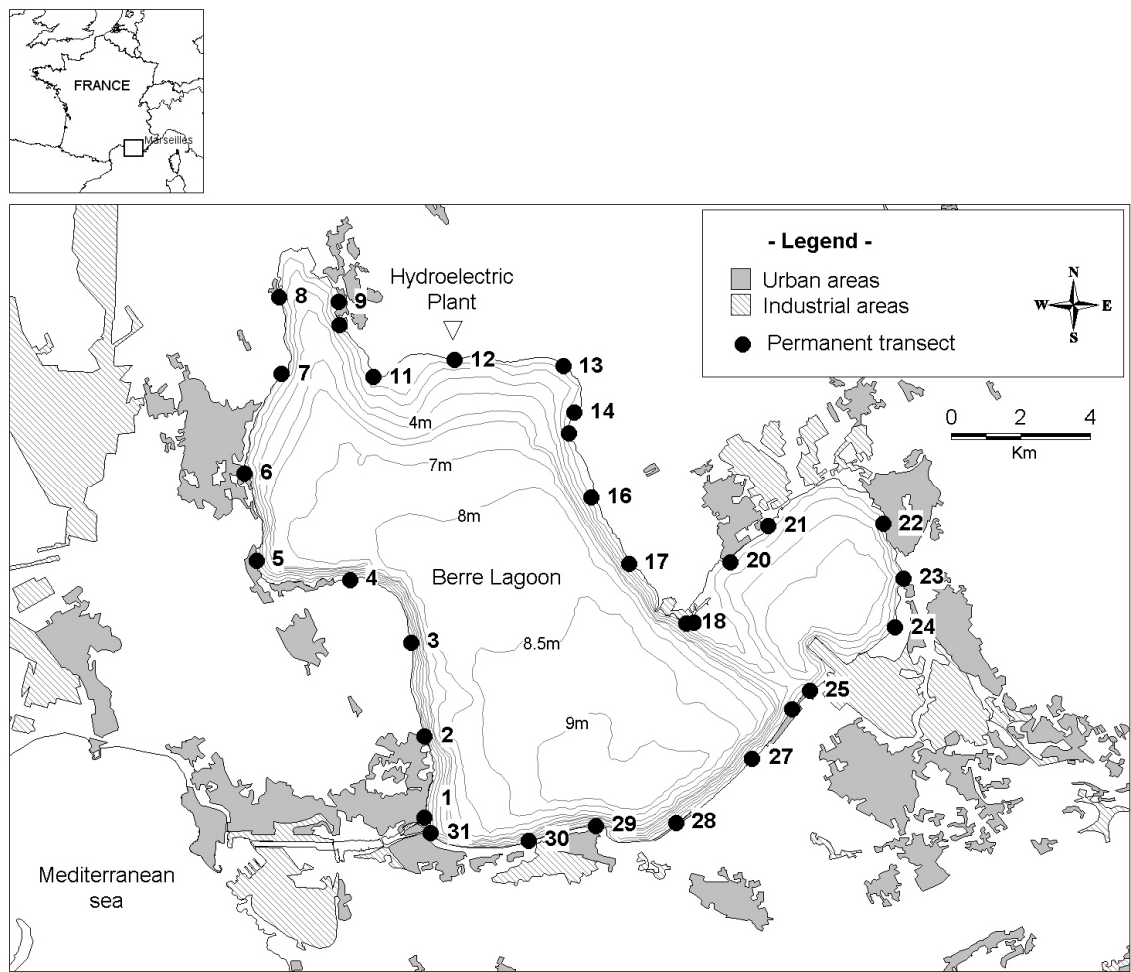


Figure 2 *Ulva*: the first principal plane of CA. The individual sampling units clearly constitute three groups: 0 to 20m, 20 to 50m and 50 to 100m.

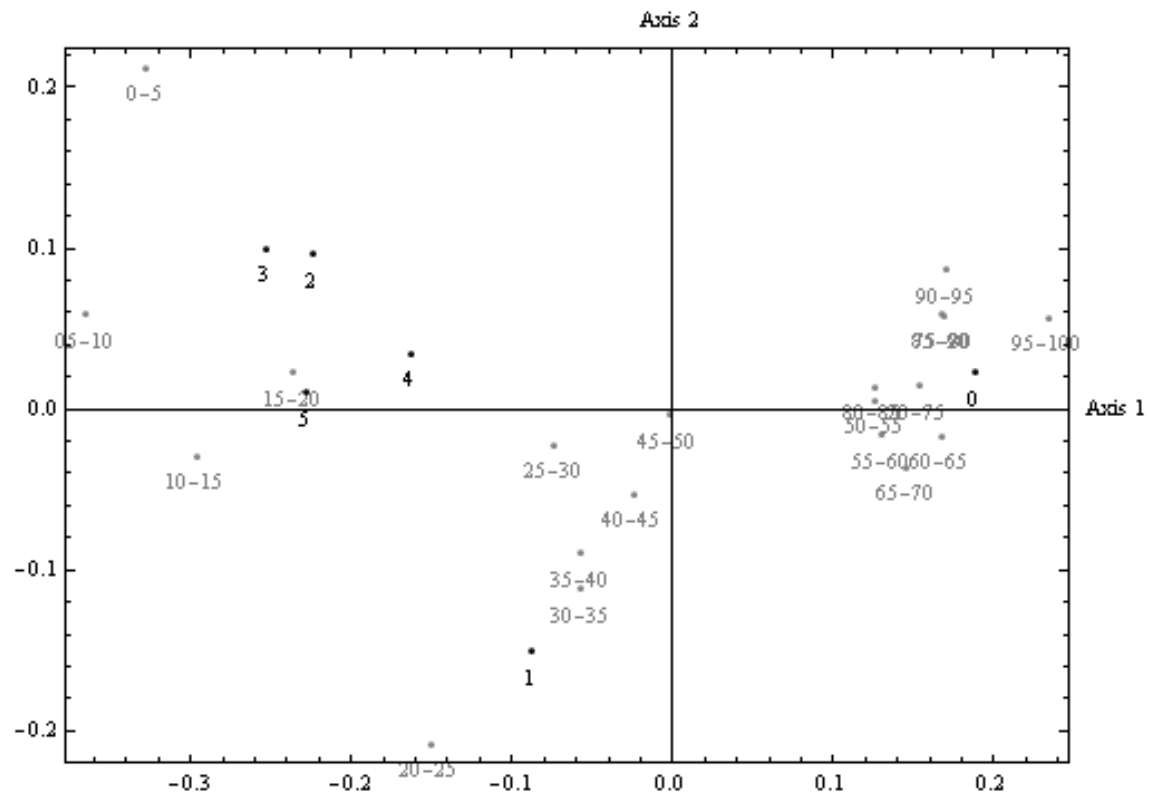


Figure 3: Chi-square tests on row and column polynomials for *Ulva*.

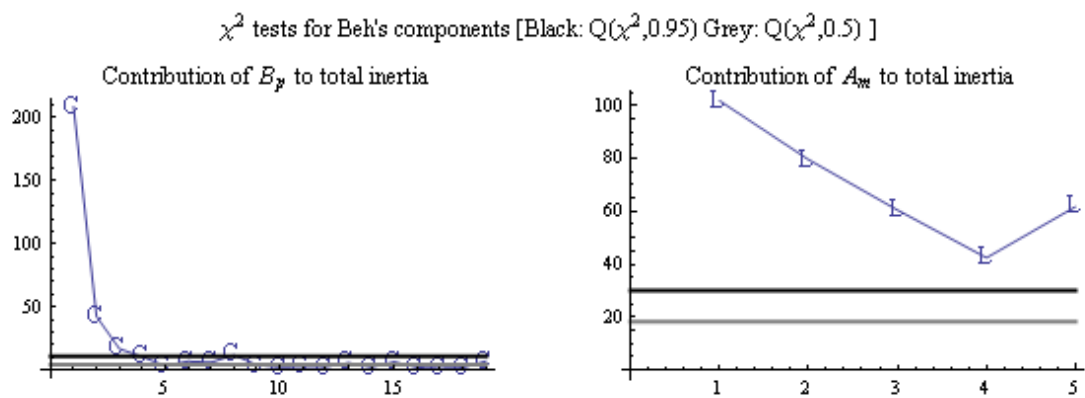


Figure 4: location and dispersion components for *Ulva*.

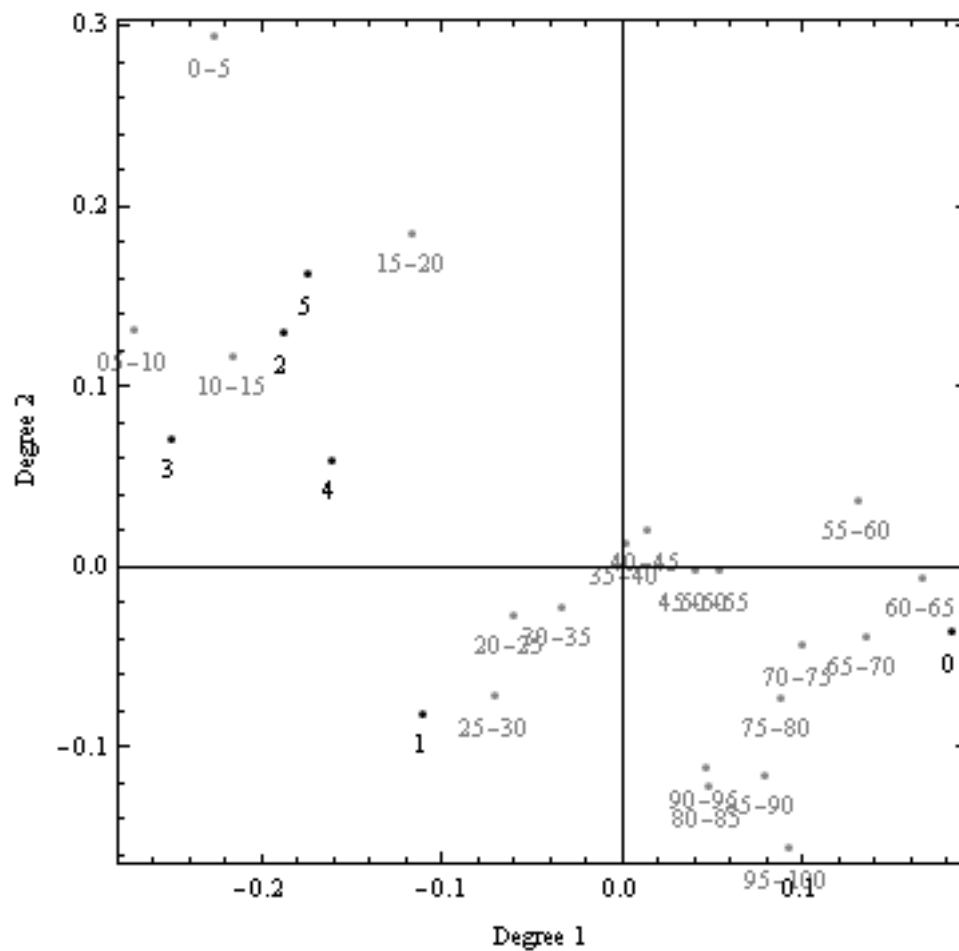


Figure 5: significance of bivariate moments at the threshold 0.05 for *Ulva*.

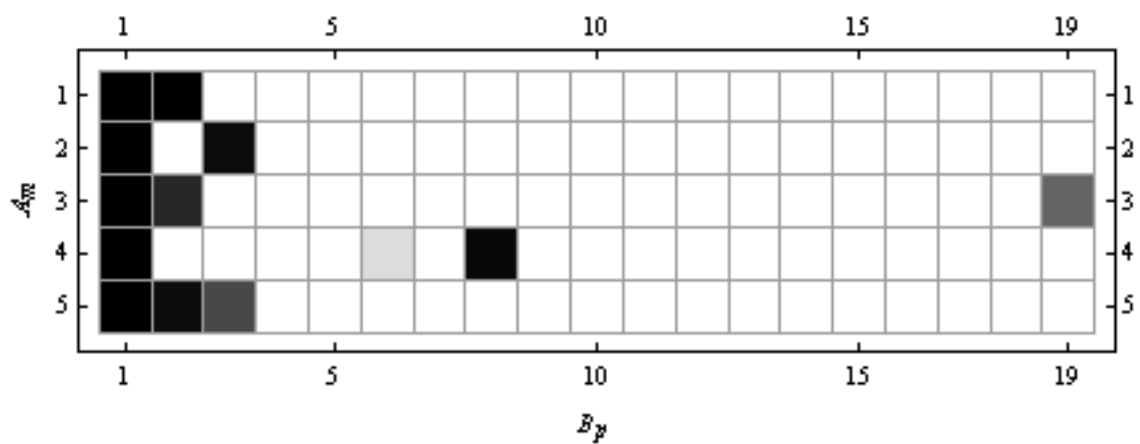


Figure 6 *Ulva*: the two main bivariate moments of the spatio-temporal OCA.

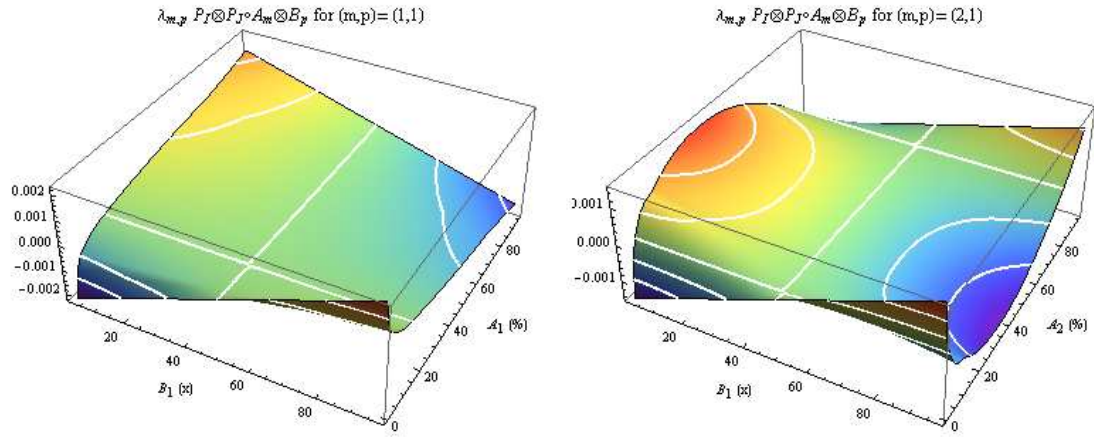


Figure 7: approximation of P by \tilde{P}_S , from formula (8), for *Ulva*.

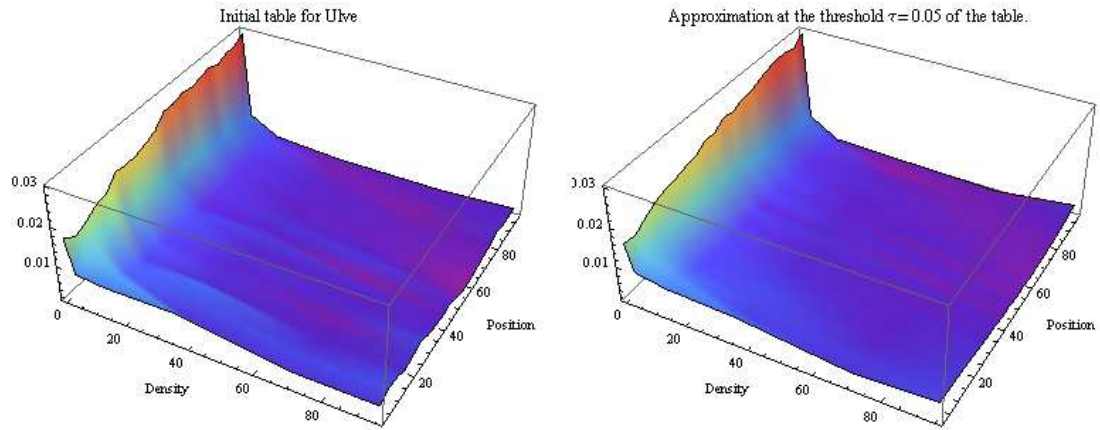


Figure 8: projection of the partial tables onto the plane (A_1, A_2) , for *Ulva*. Years of sampling are represented by their index (1 to 11). To save place, only the first 50m of the transects are represented.

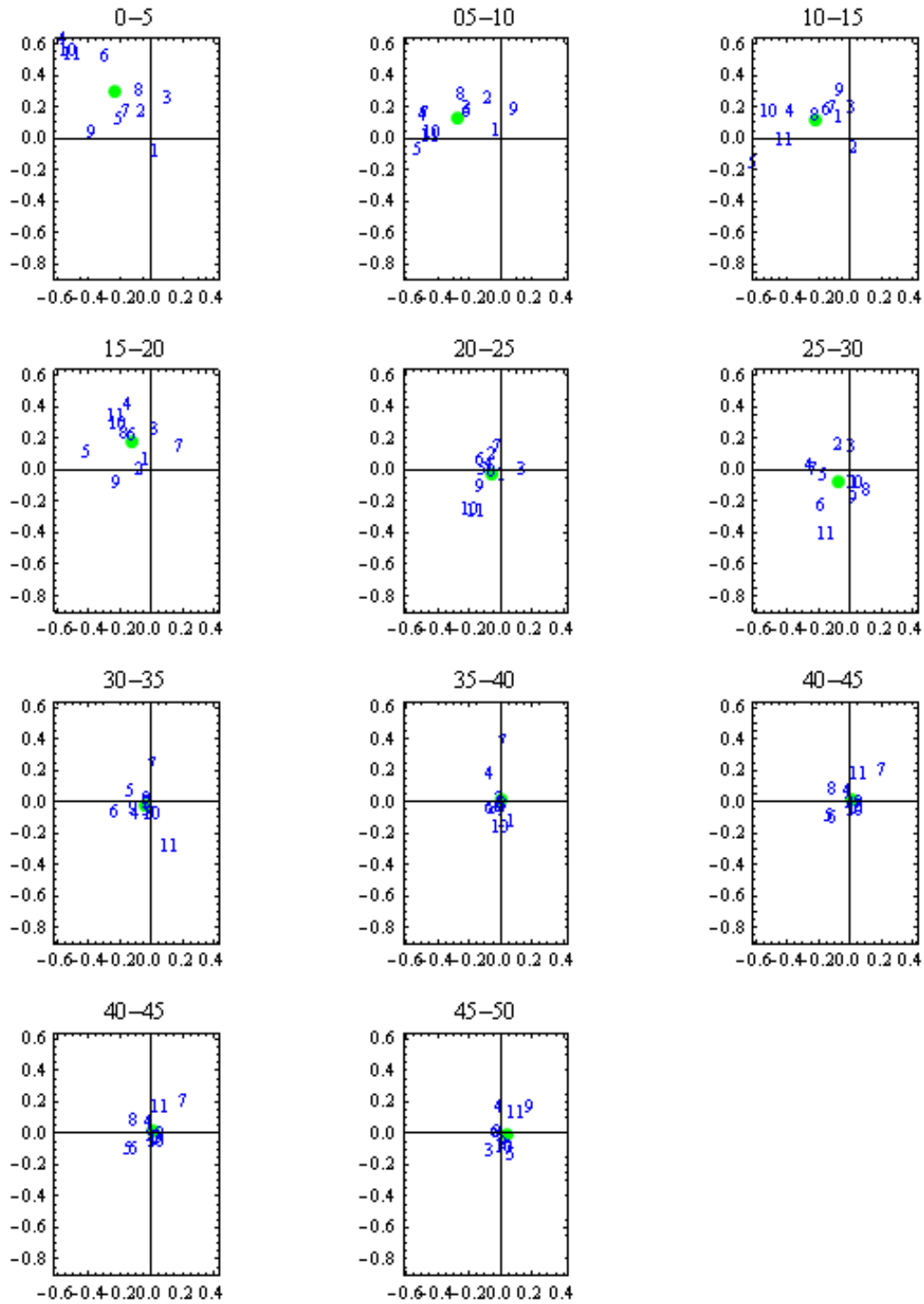


Figure 9: first plane of temporal CA for *Ulva* (83,6%, 10.1 %).

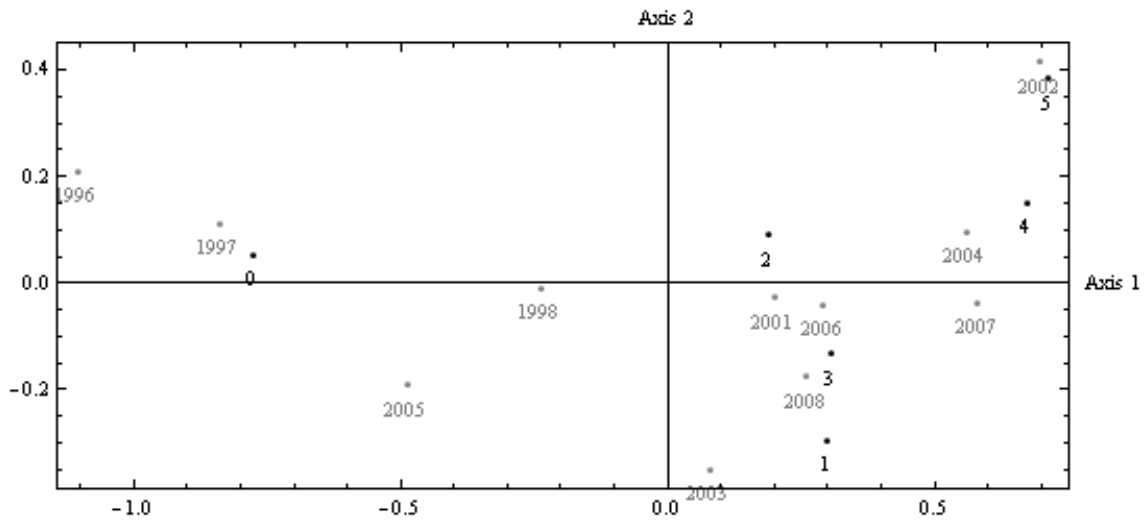


Figure 10: *Zostera*: the first principal plane of CA.

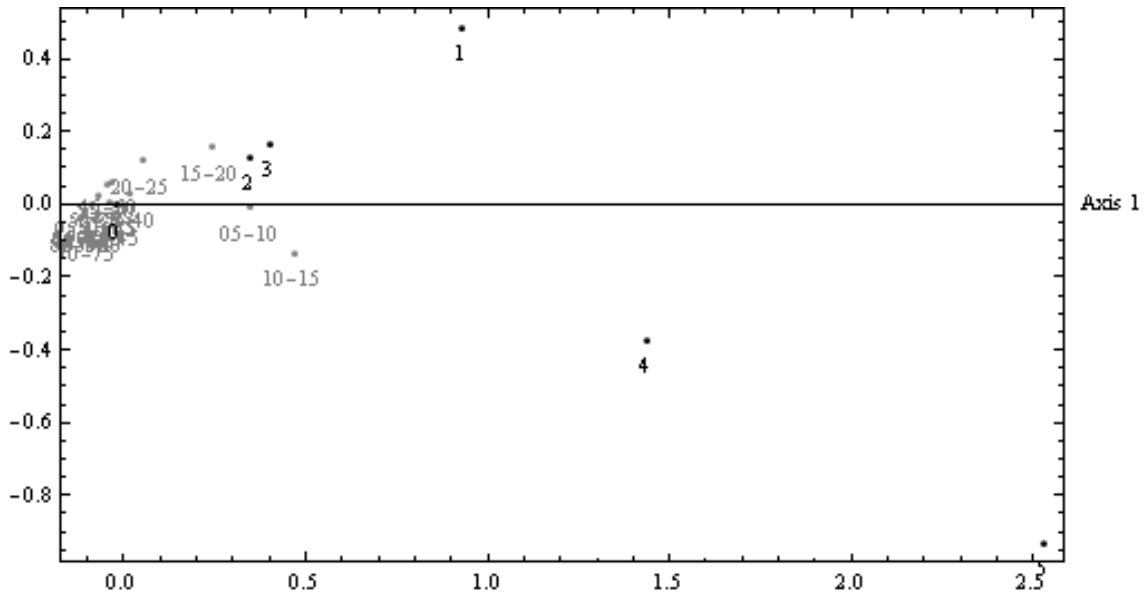


Figure 11: Chi-square tests on row and column polynomials for *Zostera*.

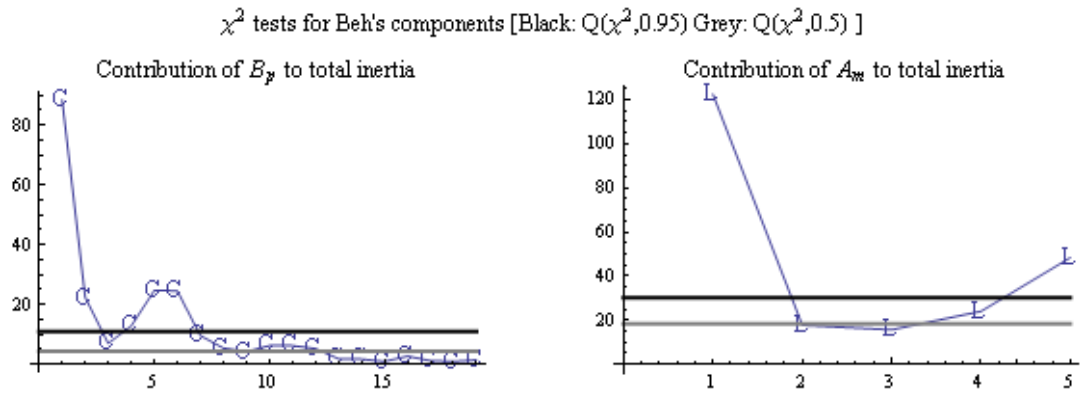


Figure 12: row and column profiles projected onto the plane (1,5) of *Zostera*.

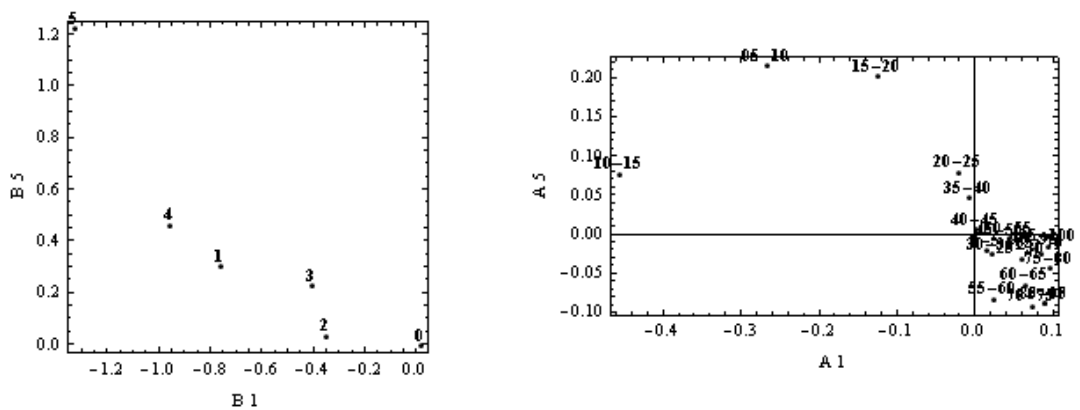


Figure 13: significance of bivariate moments at the threshold 0.05, for *Zostera*.

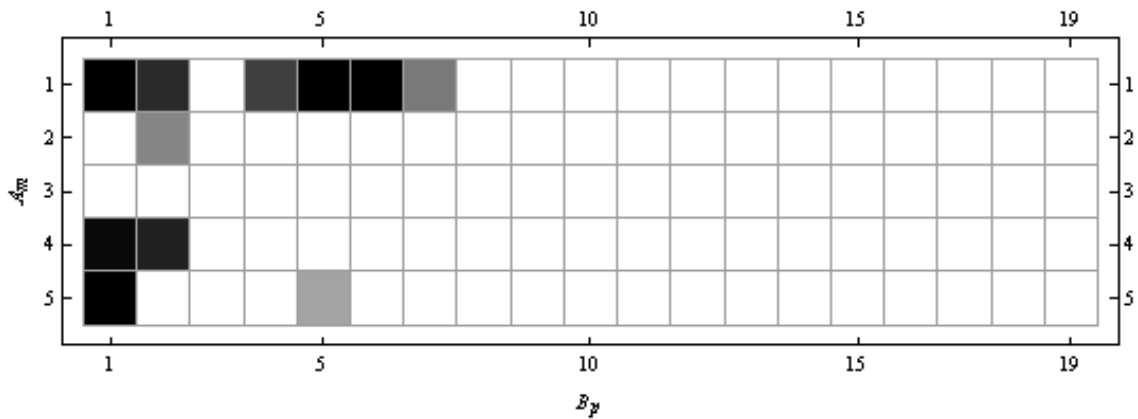


Figure 14: the discrepancy, and its approximation for *Zostera*.

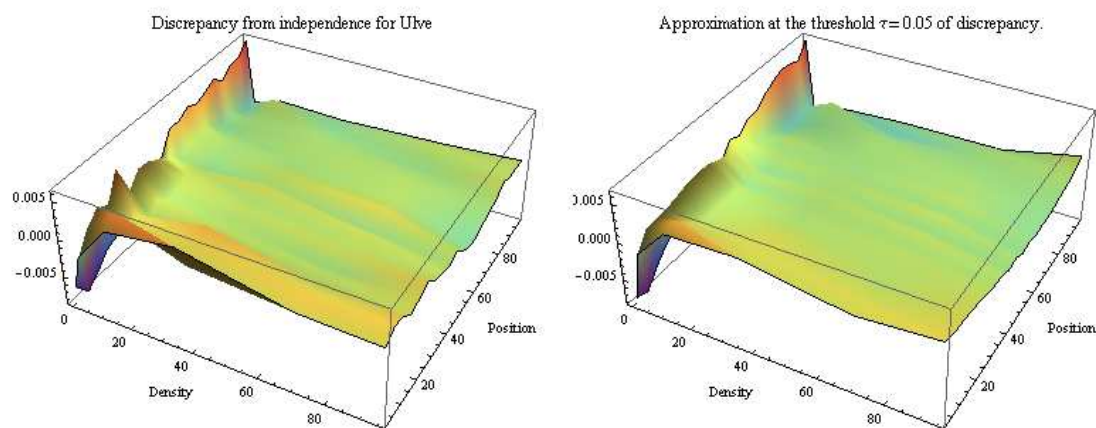


Figure 15: projection of the partial tables onto the plane (A_1, A_5) for *Zostera*. Years of sampling are represented by their index (1 to 11). To save place, only the first 50 m of the transects are represented.

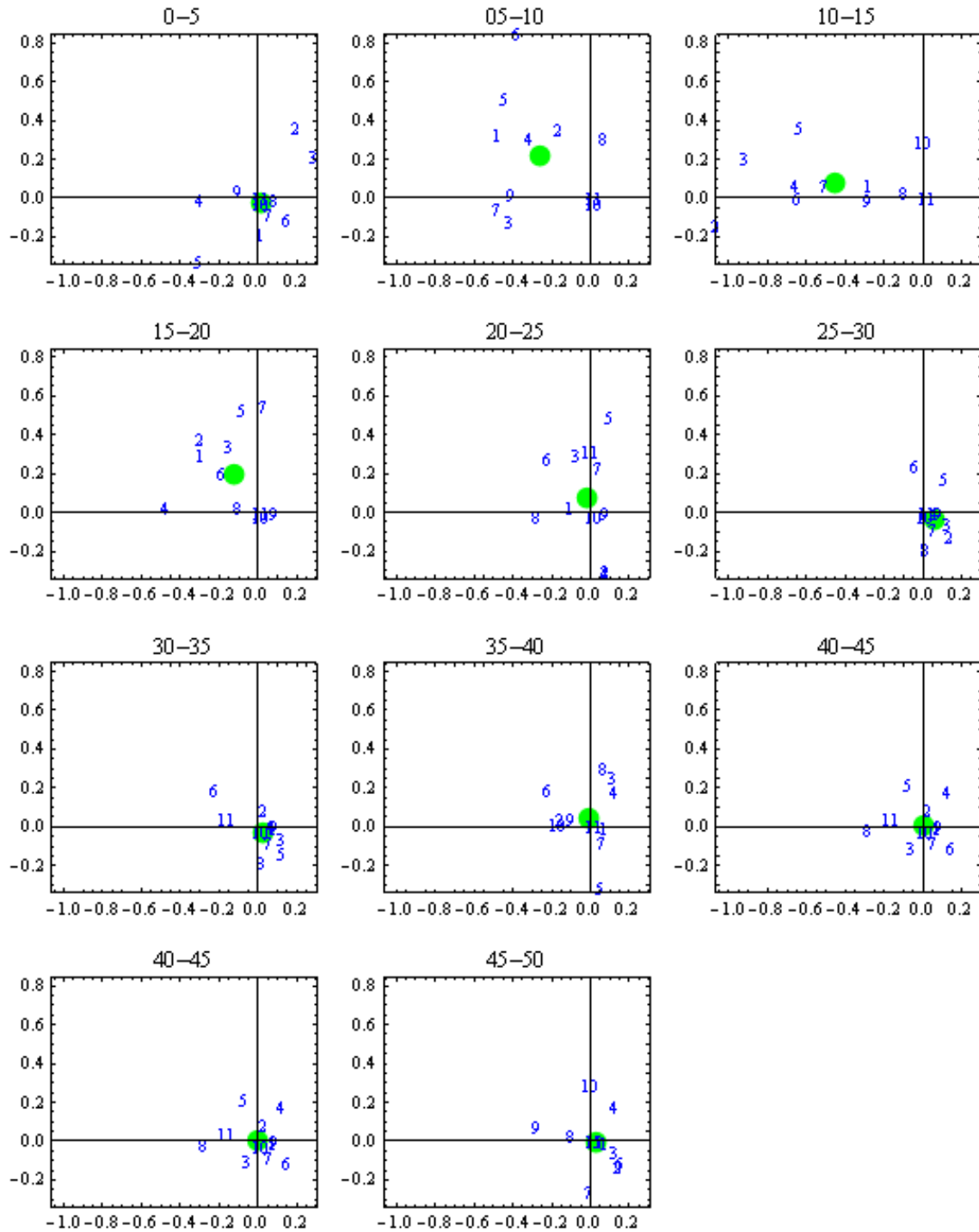


Figure 16: first plane of temporal OCA for *Zostera*.

