Functional Data Analysis of Curve Asymmetry with Application to the Color Pattern of *Hydropsyche contubernalis* Head Capsule

Champely Stéphane,[1] Guinand Bruno,[2] Thioulouse Jean,[3] and Clermidy Annabelle[2]

[1]URA CNRS 399, "Méthodes Mathématiques pour l'Analyse des Systèmes—Equipe Statistique,"
Université de Metz, Ile du Saulcy, 57045 Metz Cedex 1, France
[2]URA CNRS 1974, "Écologie des Grands Fleuves et des Eaux Douces" and
[3]UMR CNRS 5558, "Biométrie, Génétique et Biologie des Populations,"
43 bvd du 11 Novembre 1918, 69622 Villeurbanne Cedex, France

SUMMARY

The head capsule of the aquatic larvae of *Hydropsyche contubernalis* (Trichoptera: Hydropsychidae) shows color pattern asymmetries. Several sites have been sampled on four French rivers (the Rhine, Doubs, Loire, and Rhône rivers) in order to answer the following question: Do the average asymmetries in a head-capsule color pattern among sites differ in their expression of selected stress responses by the organisms? For each sampled individual, the observed color pattern is measured by a gray level curve. As the curves are not measured at identical abscissae, functional data analysis (Ramsay and Dalzell, 1991, *Journal of the Royal Statistical Society, Series B* **53,** 539–572) is an interesting framework. In the first step, we estimate the continuous gray level curve for each insect and extract its asymmetrical component. In the second step, we perform a functional analysis of site mean curves. In the first step, the curves are fitted by a linear smoother. We consider a partitioning of the resulting function into two parts, one containing the symmetrical components of the gray level curves and the other containing the asymmetrical components. In the second step, we directly consider the integral distances or integral inner products between continuous asymmetry curves, computed by evaluation over a fine regular grid, to analyze the set of curves. This very simple approach of numerical integration leads to easy and classical matrix computations. A permutation test demonstrates significant differences in asymmetry patterns according to sampling sites. Finally, an analysis similar to functional discriminant analysis (Kiiveri, 1992, *Technometrics* **34,** 321–331) details these differences.

1. Introduction

The study of asymmetries is an interesting problem in ecology, as it represents one of the fundamental tools for exploring the development stability of a given organism. It can be used either at the level of one individual or, extended as an indicator, at the population or species scale (Leary and Allendorff, 1989; Zakharov, 1989, 1992, for review). The presence of one or several quantitative asymmetries is generally considered as an indicator of the stress that an individual or a population sustains; this stress can be genetic or/and environmental (see, e.g., Palmer and Strobeck, 1986; Zakharov, 1989; Parsons, 1990, 1992; Clarke, 1993). One of the most frequent measurements used to express such a stress is the degree of fluctuating asymmetry, i.e., small random departures from symmetry for a bilateral character (see references above), but other measurements are also possible (listed in Graham, Freeman, and Emlen, 1993). The characteristics or the traits measured are, for the most part (but see Essafi, Mathieu, and Legay, 1991), quantitative characteristics considered at a definite position on the left- and the right-hand side of the organism.

Key words: Asymmetry; Between-groups PCA; Color pattern; Curve PCA; Functional data analysis; *Hydropsyche contubernalis*; Linear smoother; Permutation test.

Champely, S., B. Guinand, J. Thioulouse, and A. Clermidy. 1997. Functional data analysis of curve asymmetry with application to the color pattern of *Hydropsyche contubernalis* head capsule. Biometrics **53**:294-305.

However, some traits cannot really be considered in such a scheme. Here we consider the asymmetry of the color pattern of an aquatic insect larvae (*Hydropsyche contubernalis*; Trichoptera: Hydropsychidae). For each sampled individual, the observed color pattern is measured by a (sampled) gray level curve using an image analysis technique. From an ecological and environmental point of view, the main problem is twofold: Can a mean asymmetrical profile from a complex color pattern (i.e., a mean asymmetry curve) be expressed for a given population, and can the mean asymmetrical profiles of the different populations be compared (i.e., curve comparison) and these differences vizualized (i.e., graphical representation of a curve collection)?

We propose an exploratory method along the lines of the functional data analysis (FDA in Ramsay and Dalzell, 1991). The central idea is to replace the usual vectorial units (sampled gray level curves) of data analysis by functional units (continuous gray level curves). This allows the study of observations that are functions. We think directly in functional terms and "the fact that function values are obtained from measurements at discrete points is a practical limitation to the problem but not part of the problem formulation" (Buell, 1971). Ensuring that results are independent of the sampling scheme is one of the main objectives (Castro, Lawton, and Sylvestre, 1986), but taking into account the successive nature of the measurements is another important goal.

The application of FDA to real life studies raises new difficulties that are, up to now, seldom mentioned in the statistical literature:

(i) Data are collected over the same domain (time or space, for example), but the sampling scheme may vary from one curve to the next. In the simplest case, some values are missing. In the present case, the abscissae of observations differ. Theoretical papers have focused on situations when all subjects are measured at the same points (Besse and Ramsay, 1986). Our goal using FDA is to find a common space for the entire set of units in which an (integral) inner product can be defined in order to allow a curve comparison.

(ii) We are only interested in the asymmetrical components of the curves. This is a specific feature of our investigation.

We intend to solve both statistical problems by using FDA. This approach benefits from the contribution of both non parametric modelling techniques and multivariate data analysis methods.

In Section 2, we describe the main goals of the study and the data set. Our practical approach to FDA is presented in Section 3. Section 4 concentrates on the method of curve partitioning into symmetrical and asymmetrical components. In Section 5, we discuss the functional analysis of variance and the associated permutation test to compare the mean asymmetry curves. We propose in Section 6 a technique called between-groups principal components analysis to detail these differences. In Section 7, we present the biological results.

2. Biological Problem

The data acquisition method and treatment using the "Image" version 1.47 software is described by Sagnes (1995). Figure 1a shows the chosen color pattern. It is a rectangular area located between stable and well-defined reference marks on the head capsule of *Hydropsyche contubernalis*. For each insect, the image analysis software produces an array of gray level values belonging to the interval $[0, 255]$. These values are measured at the nodes of a fixed rectangular grid superimposed on the image of the head capsule (Figure 1b). The longitudinal means of these values make an average cross-section gray level curve (Figure 1c). The data considered in this paper are the resulting 321 sampled curves of this type.

The final aim is the comparison of the curve asymmetry between populations/stations; a brief description of the stations is given in the Appendix. We must consider two difficulties. The first one is the necessity of using a flexible and powerful statistical procedure for curve comparisons, although these curves are not identically sampled. Small larvae gray level curves are described by fewer pixels than large larvae (the range of the number of pixels per curve is [197; 300]). Therefore, we decided to scale the curves in the interval $[-1/2, 1/2]$ for pattern comparison. This transformation leads to equispaced but different sampling schemes. The second difficulty is in really considering asymmetry; for this purpose, it can be postulated that the mean gray level curve for a given individual is the sum of a symmetrical component, resulting from individual developmental stability, and an asymmetrical component, due to stress and/or developmental instability. The functional data analysis should thus reveal in which populations a stress has occurred (they will show a greater asymmetry) and whether this stress is expressed by the same changes in the chosen color pattern.

Champely, S., B. Guinand, J. Thioulouse, and A. Clermidy. 1997. Functional data analysis of curve asymmetry with application to the color pattern of *Hydropsyche contubernalis* head capsule. Biometrics **53**:294-305.

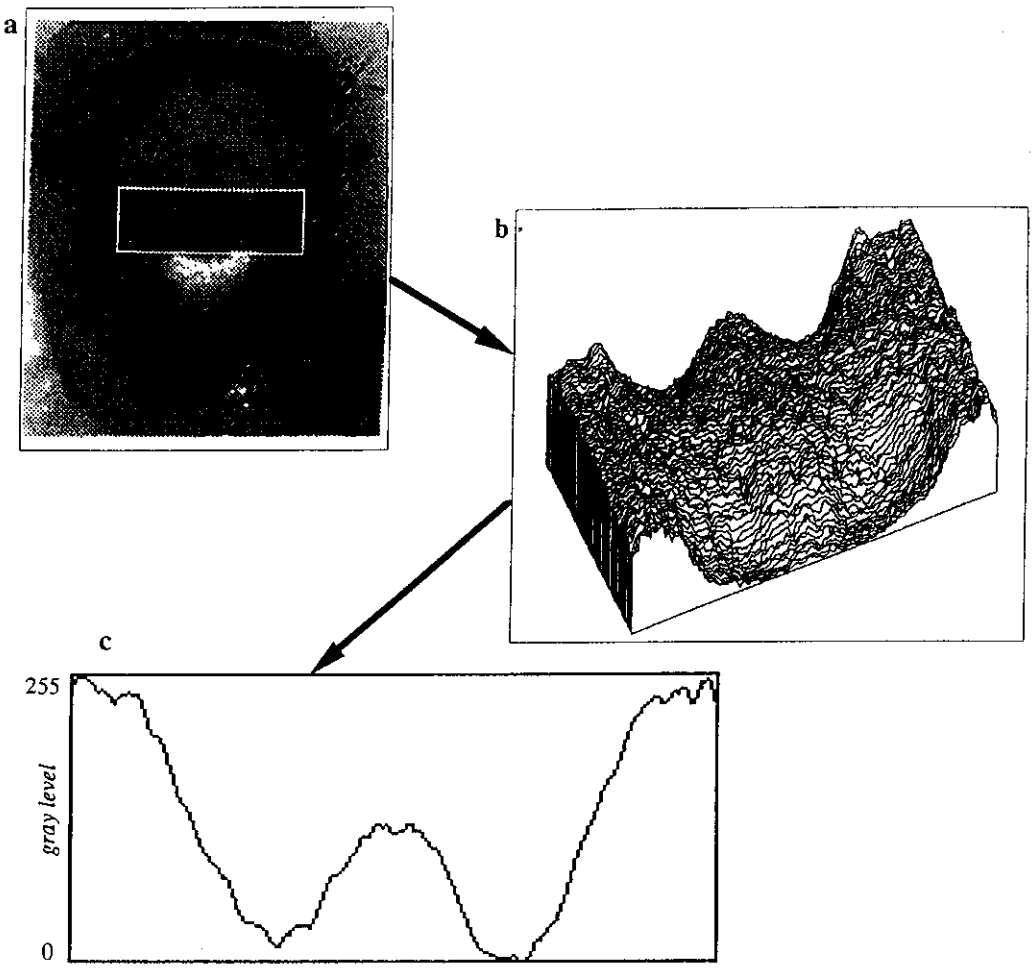

Figure 1. **a.** Color pattern area (rectangle) on the head capsule of *H. contubernalis*. **b.** Gray level family of pixel curves obtained on the area of the color pattern for one individual. **c.** Mean curve of gray levels for one individual.

3. Practical Functional Data Analysis

3.1 *Motivations*

Functional data analysis is an extension of multivariate data analysis suitable when the data are curves. This framework offers two important advantages over traditional multivariate analysis methods: (i) the results are independent of the sampling scheme of the set of curves and (ii) one can use functional (and time series) concepts like derivatives and deseasonalization to augment the tool kit for data exploration.

In the present study, we need these two features. First, the sampling points are different on each curve and the functional framework not only leads to stable results but allows results to be obtained when a direct multivariate approach fails because the same variables must be studied on each individual. Second, only the asymmetrical part of the curves is of interest. This original modelling is natural in functional terms; the partitioning principle (data = structure + residual) is essential in the functional potentialities and is the main object of Ramsay and Dalzell's (1991) paper.

The usual process of a functional data analysis consists in three successive steps: (1) generating continuous functions from finite observations, (2) formulating a theoretical analysis of the set of curves implying the computation of integral distances or integral inner products, for instance a principal components analysis (PCA) leads to an integral eigen equation, and (3) solving by a nonclassical multivariate method.

Champely, S., B. Guinand, J. Thioulouse, and A. Clermidy. 1997. Functional data analysis of curve asymmetry with application to the color pattern of *Hydropsyche contubernalis* head capsule. Biometrics **53**:294-305.

3.2 *A Practical Approach to Functional Data Analysis*

The various works on functional data analysis differ in the practical solving of step 3. Considering the context of principal components analysis, as in the paper of Besse and Ramsay (1986), the use of a spline smoother allows an exact computation of the implied integral eigen equation from a PCA of the original sampling values using a particular metric derived from spline reproducing kernels. Caussinus and Ferré (1992) study a set of functions resulting from nonlinear modelling of the original data; they propose a PCA of the parameters with a metric inherited from the dispersion matrix of the nonlinear model. Castro et al. (1986) solve the integral eigen equation using simple quadrature formulas.

We propose a simple approach (Karl, Koscielny, and Diaz, 1982, in a PCA context) for computing integral distances and integral inner products between continuous functions by evaluation over a fine regular grid $G = \{\tau_1 = -1/2, \tau_2, \ldots, \tau_N = 1/2\}$ using the following crude numerical integration approximation for any continuous function $f(t)$:

$$\int_{-1/2}^{1/2} f(t)\,dt \approx \frac{1}{N}\sum_{k=1}^{N} f(\tau_k). \tag{1}$$

The effective advantages of this approach are fourfold: (i) It is a simple one and can be explained to nonstatisticians (smoothing to define new identical variables and then performing classical computations with an interpretation in an intuitive metric). (ii) It can be reproduced because no specific software is needed. The potential user can concentrate on statistical problems rather than on technical ones. (iii) The best smoother for the particular application can be considered, for example the STL smoother (Cleveland et al., 1990) is a natural competitor to study seasonal data. (iv) It is easy to consider a complex initial sampling scheme of the set of curves: in a lot of observational studies in biology, curves can't be measured synchronously. The interpolating or smoothing step and the subsequent identical regular discretization automatically solve this practical problem.

3.3 *Generating Functions from Finite Observations*

A cubic spline smoother was used to fit the 321 gray level curves. For each curve, the smoothing parameter was selected by generalized cross validation (Hastie and Tibshirani, 1990). The resulting fits are fairly interpolating because of the continuous nature of the measurements (see Figure 2).

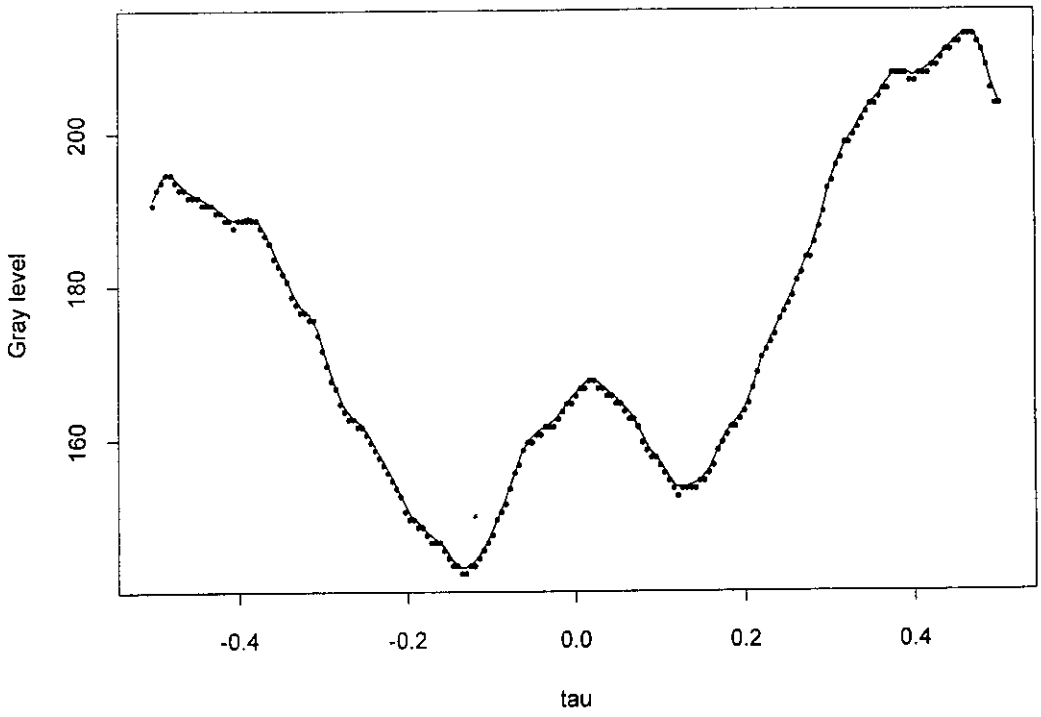

Figure 2. Generalized cross validation fit of a smoothing spline to a sampled gray level curve.

Champely, S., B. Guinand, J. Thioulouse, and A. Clermidy. 1997. Functional data analysis of curve asymmetry with application to the color pattern of *Hydropsyche contubernalis* head capsule. Biometrics **53**:294-305.

However, the complete gray level curve is not the main point of our study; only the asymmetrical component is of importance, and we tackle now the partitioning technique.

4. Partitioning Principle

The quantification of individual asymmetry will not be summarized here by a "skewness coefficient" as usual. We propose to study some more complete information by considering instead a functional pattern of asymmetry extracted from the original (spline) curve. The resulting asymmetry curve appears as a continuous sequence of deviations from the mean individual symmetry curve obtained by averaging the corresponding twin values of the original curve with reference to zero. Complex forms of asymmetry can be exhibited in this way.

4.1 *A Partitioning of Continuous Functions*

Any continuous function $f(t)$ defined on $[-1/2, 1/2]$ may be decomposed into two continuous functions

$$\forall t \in [-1/2, 1/2], f(t) = f_s(t) + f_a(t),$$

where

$$f_s(t) = (f(t) + f(-t))/2 \quad \text{and} \quad f_a(t) = (f(t) - f(-t))/2.$$

The continuous function $f_s(t)$ is even,

$$\forall t \in [-1/2, 1/2], f_s(t) = f_s(-t),$$

and the continuous function $f_a(t)$ is odd,

$$\forall t \in [-1/2, 1/2], f_a(t) = -f_a(t).$$

The resulting odd function $f_a(t)$ will be the asymmetrical component of interest in the following study.

4.2 *Evaluation Over a Fine Regular Grid*

The regular grid $G = \{\tau_1 = -1/2, \tau_2, \cdots, \tau_N = 1/2\}$ is symmetrical with reference to zero, i.e., if $\tau_k \in G$, then $-\tau_k \in G$. Therefore, it is very simple to compute the values of the two components over the same grid G:

$$\forall \tau_k \in G, f_a(\tau_k) = (f(\tau_k) - f(\tau_{N-k+1}))/2 \quad \text{and} \quad f_s(\tau_k) = (f(\tau_k) + f(\tau_{N-k+1}))/2. \qquad (2)$$

The asymmetry curves could therefore be analyzed using the methods proposed in Section 3. We will use $N = 100$.

4.3 *Biological Application*

Using formula (2), we can now define two curves (see Figure 3) corresponding to the asymmetrical and symmetrical components from each original curve.

Note the different scaling of the asymmetrical component, which appears as a small distortion of the basic symmetrical pattern. The main problem for the ecologist is to detect these informative errors in an evolution usually respecting a "natural symmetry law," e.g., the Bauplan concept.

5. Permutation Test

The goal of the test is to determine if mean asymmetry curves differ by sites. A statistic Z (El Faouzi and Escoufier, 1991), the functional equivalent of a coefficient of determination, allows us to compare the sum of between-groups squared distances and the total sum of squared distances. One obtains a reference distribution for hypothesis testing by a permutation procedure: new values of Z are evaluated after randomly assigning individuals to groups.

5.1 *Notations*

In the present study, the $n = 321$ asymmetry curves $f_{aij}(t)$, $i = 1 \cdots n_j$, $j = 1 \cdots q$ for the insects (statistical units) are assigned to $q = 13$ sites (groups). The group mean curves are $\bar{f}_{aj}(t) = (1/n_j) \sum_{i=1}^{n_j} f_{aij}(t)$, $j = 1 \cdots q$ and the grand mean curve is $\bar{f}_a(t) = (1/n) \sum_{ij} f_{aij}(t)$.

The test is based on the following decomposition of the sum of square distances:

$$\sum_{ij} \int (f_{aij}(t) - \bar{f}_a(t))^2 \, dt = \sum_j n_j \int (\bar{f}_{aj}(t) - \bar{f}_a(t))^2 \, dt + \sum_{ij} \int (f_{aij}(t) - \bar{f}_{aj}(t))^2 \, dt = B + W.$$

Champely, S., B. Guinand, J. Thioulouse, and A. Clermidy. 1997. Functional data analysis of curve asymmetry with application to the color pattern of *Hydropsyche contubernalis* head capsule. Biometrics **53**:294-305.

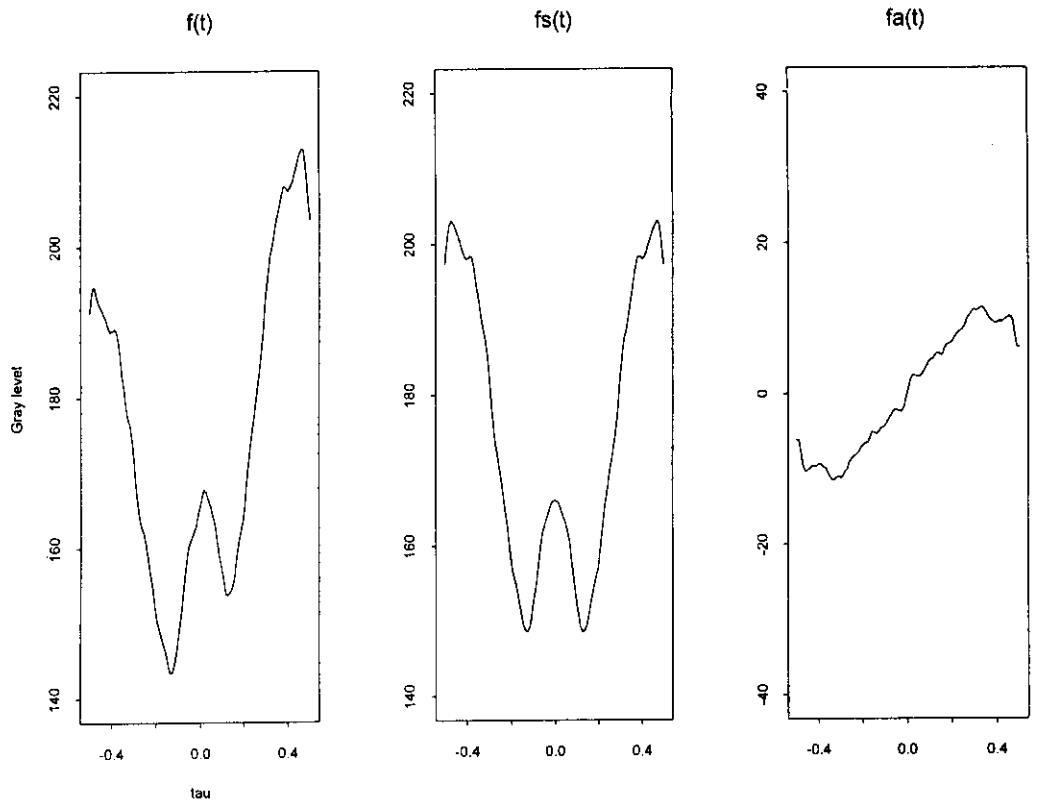

Figure 3. Partitioning of a gray level curve. The left panel shows the complete model, the center one is the symmetrical component, and the right one the asymmetrical component.

5.2 *Functional Analysis of Variance*

Consider the null and alternative hypotheses: H_0—no difference between the mean (asymmetry) curves; H_1—some difference.

The test statistic

$$Z = \sum_j n_j \int (\bar{f}_{aj}(t) - \bar{f}_a(t))^2 \, dt / \sum_{ij} \int (f_{aij}(t) - \bar{f}_a(t))^2 \, dt = B/(B+W)$$

reveals the importance of the contribution of the groups to data variation. Being a monotonic function of the more classical $F = ((n-q)/(q-1))(B/W)$, this test is equivalent to the one used in the functional analysis of variance of Zerbe and Walker (1977).

Using the regular grid G and formula (1) for computing all integrals, Z is approximated by

$$\tilde{Z} = \mathrm{tr}(\mathbf{P}_U \mathbf{F}_a \mathbf{F}_a^t \mathbf{P}_U^t)/\mathrm{tr}(\mathbf{F}_a \mathbf{F}_a^t), \tag{3}$$

where $\mathbf{F}_a = \left[f_{aij}(\tau_k) - \bar{f}_a(\tau_k)\right]_{ij=1\cdots n, k=1\cdots N}$ is a centered $(n \times N)$ evaluation matrix, and $\mathbf{P}_U$ the orthogonal projector onto the linear subspace of R^n spanned by the columns of $\mathbf{U}$ the $(n \times q)$ design matrix.

A p-value can be determined by evaluating Z for random assignments of statistical units to groups and by comparing the observed value Z_0 with the induced sampled distribution. It is equivalent to permute the rows of matrix $\mathbf{U}$ in formula (3) and to compute the corresponding values of $\tilde{Z}$.

The null hypothesis is clearly rejected here in a sample of 1000 randomized assignments; no value exceeds the observed value $\tilde{Z}_0 = 0.170$. Under the null hypothesis, the first two exact moments of $\tilde{Z}$ can in fact be computed using formulas given by Kazi-Aoual et al. (1995). They are similar to those estimated from the Monte-Carlo procedure proposed here ($\mathrm{E}(Z) = 0.0375$, $\bar{\tilde{Z}} = 0.0378$, $\sigma_Z = 1.195 \times 10^{-2}$ and $s_Z = 1.186 \times 10^{-2}$) and lead to the same conclusion: the existence of actual differences in the mean asymmetry curves. Nevertheless, the favored alternative hypothesis remains vague and an exploratory multivariate analysis may help to better understand the structure of the data.

Champely, S., B. Guinand, J. Thioulouse, and A. Clermidy. 1997. Functional data analysis of curve asymmetry with application to the color pattern of *Hydropsyche contubernalis* head capsule. Biometrics **53**:294-305.

6. Between-Groups PCA

6.1 *The Theoretical Functional PCA*

We first consider the functional analog of principal components analysis because our proposals are direct consequences of this method.

Given the centered collection of continuous curves $\{f_i(t)\}_{i=1,\dots,n}$ defined on $[-1/2, 1/2]$, we want to find a basis $\{v_\alpha(t)\}_\alpha$ of orthonormal functions defined on $[-1/2, 1/2]$, called empirical orthogonal functions, and showing the following property:

Each curve $f_i(t)$ is projected onto $v_\alpha(t)$, resulting in an n-dimensional vector $\mathbf{R}_\alpha$ ($r_{\alpha i} = \int_{-1/2}^{1/2} f_i(t)v_\alpha(t)\,dt$, $i = 1 \dots n$) called row scores. The successive row scores are of maximum variance; that is to say, they maximize $\mathrm{var}(\mathbf{R}_\alpha) = (1/n)\,\Sigma_i\,(r_{\alpha i} - \bar{r}_\alpha)^2$, *where* $\bar{r}_\alpha = (1/n)\,\Sigma_i\, r_{\alpha i}$.

It is shown in Ramsay (1982) that the solutions are given by the spectral analysis of the (estimated) covariance function,

$$\int_{-1/2}^{1/2} c(t,t')v_\alpha(t') = \lambda_\alpha v_\alpha(t) \quad \text{with} \int_{-1/2}^{1/2} v_\alpha(t)v_\beta(t) = \delta_{\alpha\beta}, \tag{4}$$

where $c(t,t') = (1/n)\,\Sigma_{i=1}^n f_i(t)f_i(t')$ is the (estimated) covariance function and $\delta_{\alpha\beta}$ the Kronecker symbol. The original formulation in terms of stochastic processes is in Loève (1945).

6.2 *The Approximate Functional PCA*

Let $\mathbf{F} = [f_i(\tau_k)]_{i=1\cdots n, k=1\cdots N}$ be the centered $(n \times N)$ evaluation matrix. The integral eigen equation (4) is approximated using formula (1) in R^N by

$$\frac{1}{nN}\mathbf{F}^t\mathbf{F}\tilde{\mathbf{v}}_\alpha = \tilde{\lambda}_\alpha\tilde{\mathbf{v}}_\alpha \quad \text{with } \frac{1}{N}\tilde{\mathbf{v}}_\alpha^t\tilde{\mathbf{v}}_\beta = \delta_{\alpha\beta}. \tag{5}$$

Estimated empirical orthogonal functions can be formed by appropriate interpolation of the $(\tau_k, \tilde{\mathbf{v}}_{\alpha k})_{k=1\cdots N}$.

Approximate row scores $\tilde{\mathbf{R}}_\alpha$ are now

$$\tilde{r}_{\alpha i} = \frac{1}{N}\sum_{k=1}^{N} f_i(\tau_k)\tilde{v}_{\alpha k}. \tag{6}$$

Formulas (5) and (6) are equivalent to the PCA of the centered evaluation matrix $\mathbf{F}$ using uniform systems of weights $1/n$ and $1/N$.

6.3 *Between-Groups PCA*

When the data are curves, discriminant analysis is well known (for two groups, see Rao (1958), and for the general case, Kiiveri (1992)). However, discriminant analysis can produce unstable results due to the dimension of the curve space. Here we propose a stable method, the between-groups PCA (Dolédec and Chessel, 1989) in a functional framework. This technique belongs to the family of PCA with respect to instrumental variables (Rao, 1964; Sabatier, Lebreton, and Chessel, 1989).

We consider the set of curves $f_{ij}(t)$, $i = 1 \cdots n_j$, $j = 1 \cdots q$, resulting from an experiment described by the design matrix $\mathbf{U}$ $(n \times q)$ and the corresponding orthogonal projector $\mathbf{P}_U$. We want to find a basis $\{v_\alpha(t)\}_\alpha$ of empirical orthogonal functions showing the following property:

The successive row scores are of maximum variance between groups; that is to say, they maximize $\mathrm{var}_B(\mathbf{R}_\alpha) = (1/n)\,\Sigma_j\, n_j(\bar{r}_{\alpha j} - \bar{r}_\alpha)^2$, *where* $\bar{r}_{\alpha j} = (1/n_j)\,\Sigma_{i=1}^{n_j} r_{\alpha i(j)}$ *are the group mean scores.*

This is equivalent to a PCA of mean curves using a sequence of weights equal to the number of units in each group. Indeed, this is an exploration of the relationships between group mean curves and of the numerator of the test statistics Z (Section 5).

Using results from 6.2 and notations from 5.2, the eigen equation in R^N is

$$\frac{1}{nN}\mathbf{F}_a^t\mathbf{P}_U\mathbf{F}_a\tilde{\mathbf{v}}_\alpha = \tilde{\lambda}\tilde{\mathbf{v}}_\alpha \quad \text{subject to } \frac{1}{N}\tilde{\mathbf{v}}_\alpha^t\tilde{\mathbf{v}}_\beta = \delta_{\alpha\beta}.$$

The approximate row scores $\tilde{\mathbf{R}}_\alpha$ in R^n,

$$\tilde{\mathbf{R}}_\alpha = \frac{1}{N}\mathbf{P}_U\mathbf{F}_\alpha\tilde{\mathbf{v}}_\alpha,$$

are of maximum variance between groups.

Champely, S., B. Guinand, J. Thioulouse, and A. Clermidy. 1997. Functional data analysis of curve asymmetry with application to the color pattern of *Hydropsyche contubernalis* head capsule. Biometrics **53**:294-305.

The scores $\tilde{\mathbf{R}}_{\alpha U}$ in R^q are

$$\tilde{\mathbf{R}}_{\alpha U} = \frac{1}{N}(\mathbf{U}^t\mathbf{U})^{-}\mathbf{U}^t\mathbf{F}_\alpha\tilde{\mathbf{v}}_\alpha,$$

and allow a graphical representation of mean curves in a low-dimensional space.

7. Results

We apply the between-groups PCA to the set of asymmetry curves. Figure 4 shows the first two empirical orthogonal functions (98% of total variation). It provides an explanation for the asymmetrical patterns relative to these axes. The F1 axis expresses the differences lower (left part, inferior asymmetry) or higher (right part, superior asymmetry) than the mean curve. This axis presents two types of asymmetry according to the position of the populations on the factor plane. The F2 axis expresses another, more complex description of the variation with respect to the mean curve.

The interpretation of the first factor map of the between-groups PCA is facilitated by plotting some interesting mean asymmetry curves (Figure 5). The sites situated at the ends of the axes have the most asymmetrical gray level curves (with an exception for the negative side of the F2 axis): "Rhin 90" for the F1 axis, "Rhin 89" for the F2 axis, and "Crissey" for these two axes simultaneously. It is interesting to note that, relative to the F1 axis, the asymmetries are clearly marked and are expressed by different patterns in the Rhine River and the Doubs River. This axis reveals less pronounced asymmetries for the Loire River sites except for Gennes.

From an ecological point of view, it is particularly remarkable to see that several complex types of asymmetry exist (Figure 4). The populations belong to one of these complex types, as shown by their position on the between-groups PCA factor plane.

It is not a biological surprise to observe the higher levels of asymmetry in the Rhine and Doubs rivers. These rivers are, in the sampled sections, very polluted, whereas the Loire river is still relatively unpolluted, due to the absence of both great industrial and human installations along its course. On the Doubs river, the highly polluted reaches of the Dole agglomeration separate the upstream station of Rochefort from the downstream station of Crissey. This can explain their clear

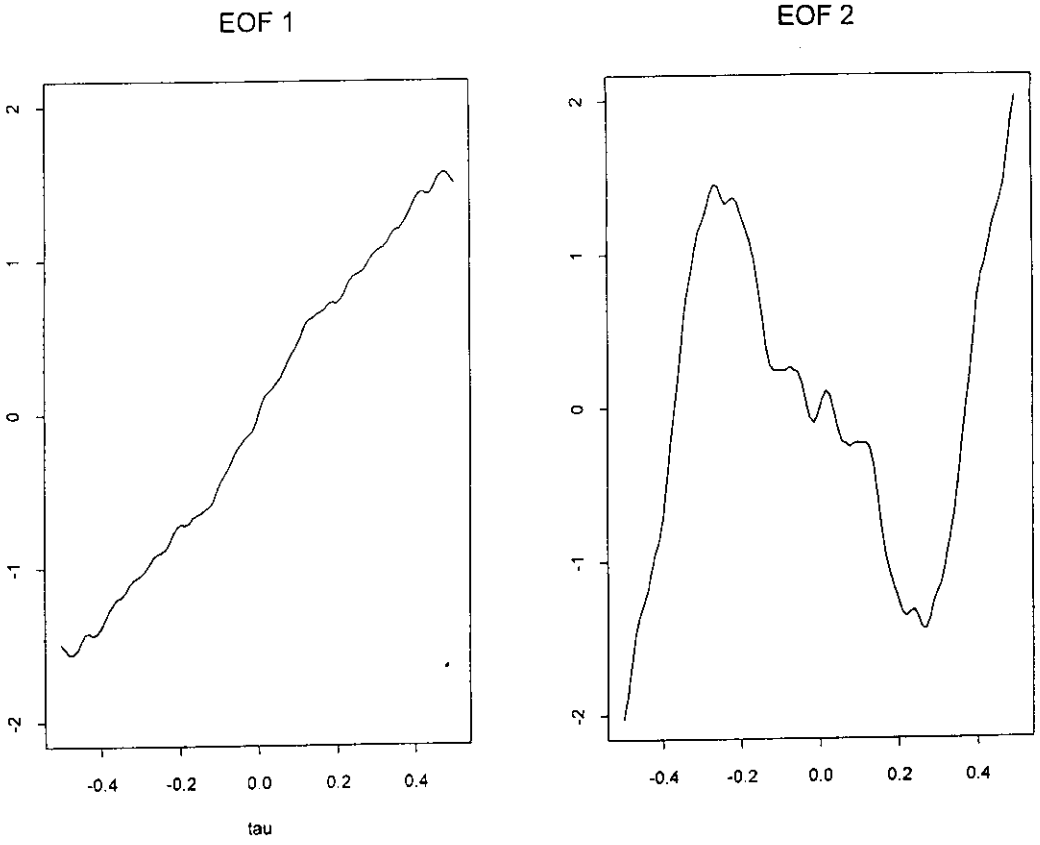

Figure 4. First two orthogonal empirical functions resulting from the between-groups PCA.

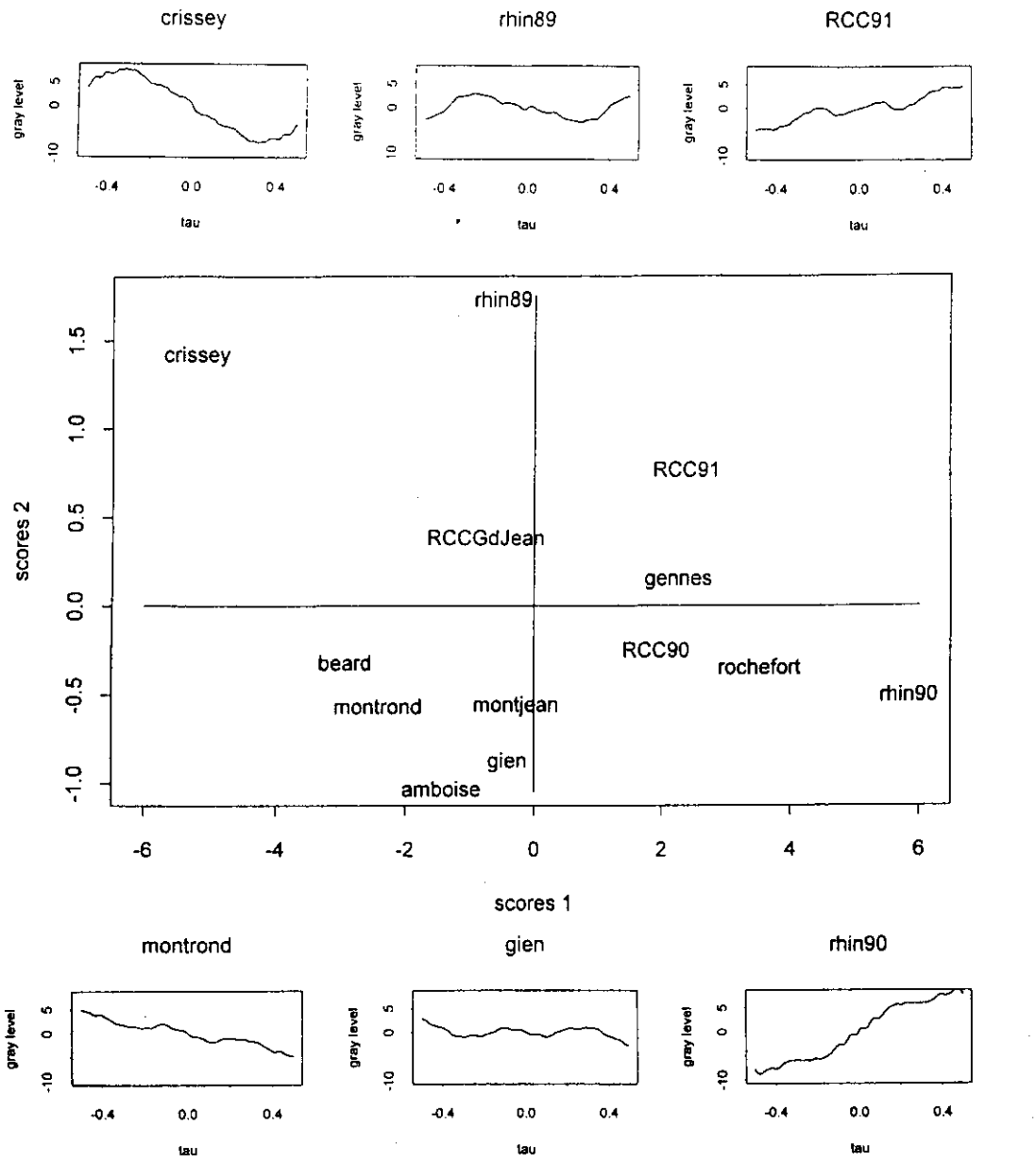

Figure 5. F1/F2 factor map of the between-groups PCA.

opposition on the factor map despite the geographical proximity of the sites (see the Appendix). Other studies (Décamps, Besch, and Vobis, 1983; Petersen and Petersen, 1983; Vuori, 1993, 1994) have demonstrated the influence of pollution on behavioral and morphological characteristics of *Hydropsyche*.

8. Conclusions

Although the fundamental concepts of FDA have been developed in the last 30 years, the present application is among the first ones to deal with real life data. Here we have exposed some practical arguments for the use of this kind of analysis. With the simple approach proposed here, functional data analysis becomes a simple and efficient tool: (i) the distances between curves can be easily compared and (ii) symmetrical and asymmetrical components can be extracted (other decompositions are also possible, e.g., monotonicity).

The application of this FDA technique to an ecological problem is centered on the study of asymmetry. We describe a color pattern on a complex data set and the existence of an important asymmetrical component for some populations of *H. contubernalis*. This asymmetrical component

Champely, S., B. Guinand, J. Thioulouse, and A. Clermidy. 1997. Functional data analysis of curve asymmetry with application to the color pattern of *Hydropsyche contubernalis* head capsule. Biometrics **53**:294-305.

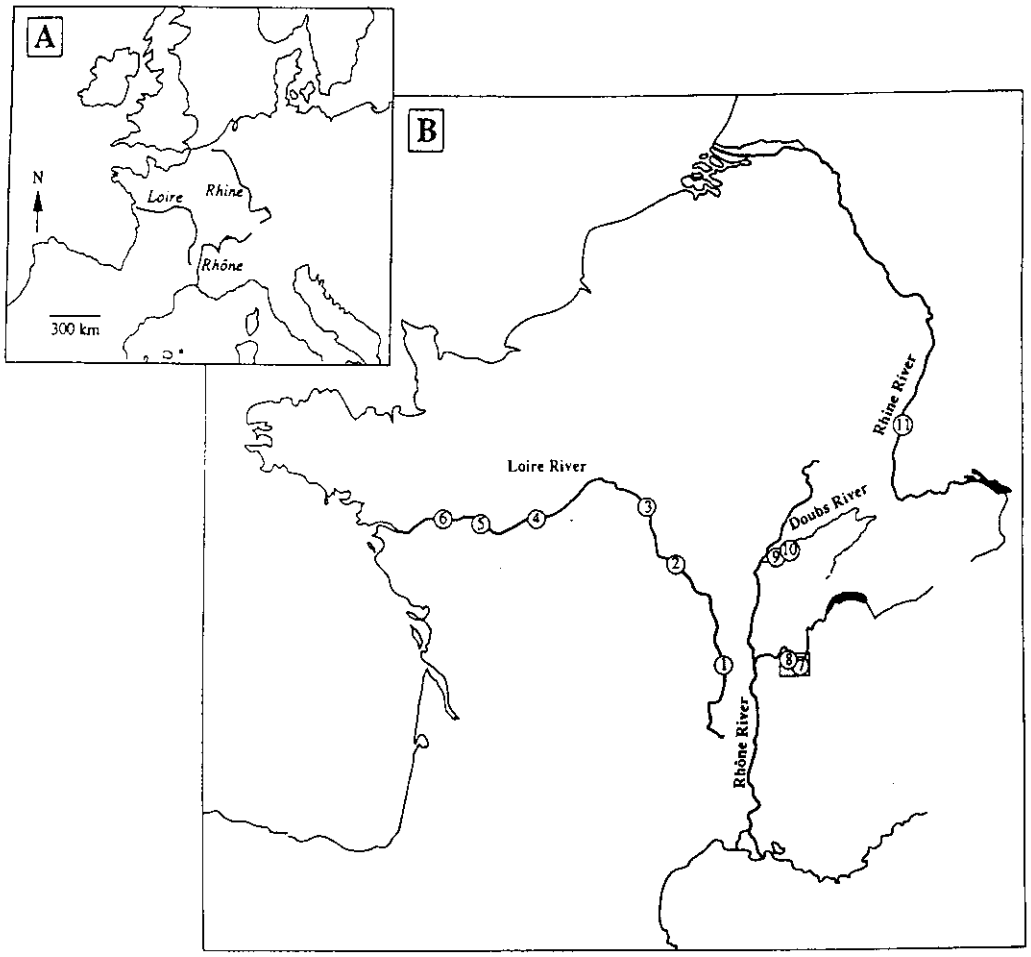

Figure 6. Locations of the sampled sites on four French rivers (the Rhine, Doubs, Loire, and Rhône rivers). Detailed explanations are given in the Appendix.

also allows differentiating the sites. This sustains the theory that a stress can be a factor responsible for the developmental instability of the organism (e.g., Zakharov, 1989). Since the most polluted sites present the highest level of asymmetry ("Crissey", "Rhin 90", ...), we suggest that the stress may be due to pollution, although *H. contubernalis* is quite resistant to oxygen depletion (Becker, 1987). The possibility of using an organism's asymmetry level as a biological indicator of stress seems confirmed here.

The pattern recognition problem considered here demonstrates the practical interest of combining linear smoothing and FDA in the study of a collection of curves. Such families are increasingly met in biometry. They are collected to study the dynamics of a biological system, e.g., growth curves or cycles in population biology. We think that these flexible statistical tools are worth considering because the conditions for a classical experimental design (particularly identical sampling times) are seldom satisfied in ecology.

ACKNOWLEDGEMENTS

The authors wish to thank R. Tomassone and H. Tachet for their helpful comments. We are indebted to L. Hill for correcting the English. We also thank the referees for their very constructive suggestions.

RÉSUMÉ

La capsule céphalique des larves d'*Hydropsyche contubernalis* (Trichoptera; Hydropsychidae) montre des asymétries dans ses patrons de coloration. Plusieurs sites ont été échantillonnés sur quatre rivières françaises (le Rhin, le Doubs, la Loire et le Rhône) afin de répondre à la question suivante: est-ce que les asymétries dans les patrons de coloration de la capsule céphalique diffèrent

Champely, S., B. Guinand, J. Thioulouse, and A. Clermidy. 1997. Functional data analysis of curve asymmetry with application to the color pattern of *Hydropsyche contubernalis* head capsule. Biometrics **53**:294-305.

en moyenne de site à site, pour exprimer des réponses des organismes à un certain stress? Pour chaque individu, le patron de coloration mesuré est une courbe de niveaux de gris. Comme les courbes ne sont pas mesurées aux mêmes abscisses, l'Analyse de Données Fonctionnelles (Ramsay et Dallzell, 1991, *Journal of the Royal Statistical Society, Series B* **53,** 539–572) est un cadre intéressant. Dans une première étape, nous estimons une courbe continue de niveau de gris pour chaque insecte et extrayons sa composante asymétrique. Dans la seconde étape, nous réalisons une analyse de données fonctionnelles des courbes moyennes d'asymétrie (par site). Lors de la première étape, les courbes sont ajustées avec un lisseur linéaire. Nous considérons un partitionnement des fonctions en deux composantes, l'une symétrique et l'autre asymétrique. Dans la seconde étape nous utilisons une analyse fonctionnelle de la variance pour prendre en compte directement les distances entre les courbes d'asymétrie moyennes, calculées par intégration. Le calcul est fait par évaluation sur une grille fine et régulière. Cette approche simple d'intégration numérique conduit à des calculs matriciels faciles et classiques. Un test de permutation montre des différences significatives dans les patrons d'asymétrie suivant le site échantillonné. Enfin, une analyse proche de l'analyse discriminante fonctionnelle (Kiiveri, 1992, *Technometrics* **34,** 321–331) détaille ces différences.

References

Becker, G. (1987). Net-building behaviour, tolerance and development of two caddisfly species from the River Rhine (*H. contubernalis* and *H. pellucidula*) in relation to the oxygen content. *Œcologie* **73,** 242–250.

Besse, P. and Ramsay, J. O. (1986). Principal components analysis of sampled functions. *Psychometrika* **51,** 285–311.

Buell, C. G. (1971). Integral equation representation for factor analysis. *Journal of Atmospheric Science* **28,** 1502–1505.

Castro, P. E., Lawton, W. H., and Sylvestre, E. A. (1986). Principal modes of variations for processes with continuous sample curves. *Technometrics* **28,** 329–337.

Caussinus, H. and Ferré, L. (1992). Comparing the parameters of a model for several units by means of principal components analysis. *Computational Statistics and Data Analysis* **13,** 269–280.

Clarke, G. M. (1993). The genetic basis of developmental stability. I. Relationships between stability, heterozygosity and genomic coadaptation. *Genetica* **89,** 15–23.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). STL: A seasonal trend decomposition procedure based on Loess. *Journal of Official Statistics* **6,** 3–73.

Décamps, H., Besch, W. K., and Vobis, H. (1983). Influence de produits toxiques sur la construction du filet des larves d'*Hydropsyche* (Trichoptera, Hydropsychidae). *Comptes Rendus de l'Académie des Sciences, Paris* **B276,** 375–378.

Dolédec, S. and Chessel, D. (1989). Rythmes saisonniers et composantes stationnelles en milieu aquatique II—Prise en compte et élimination d'effets dans un tableau faunistique. *Acta Œcologica, Œcologia Generalis* **10,** 207–232.

El Faouzi, N. and Escoufier, Y. (1991). Modélisation I-spline et comparaison de courbes de croissance. *Revue de Statistique Appliquée* **39,** 51–64.

Essafi, K., Mathieu, J., and Legay, J.-M. (1991). Asymmetries and locomotor behaviour for *Niphargus* (stygobiont amphipod). *Stygologia* **6,** 91–95.

Graham, J. H., Freeman, D. C., and Emlen, J. M. (1993). Developmental stability: A sensitive indicator of populations under stress. In *Environmental Toxicology and Risk Assessment,* W. G. Landis, J. S. Hughes, and M. A. Lewis (eds), 136–158. Philadelphia: American Society for Testing and Materials.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models.* London: Chapman and Hall.

Karl, T. R., Koscielny, A. J., and Diaz, H. F. (1982). Potential errors in the application of principal components analysis to geophysical data. *Journal of Applied Meteorology* **21,** 1183–1186.

Kazi-Aoual, F., Hitier, S., Sabatier, R., and Lebreton, J. D. (1995). Refined approximations to permutation tests for multivariate inference. *Computational Statistics and Data Analysis* **20,** 643–656.

Kiiveri, H. T. (1992). Canonical variate analysis of high-dimensional spectral data. *Technometrics* **34,** 321–331.

Leary, R. F. and Allendorf, F. W. (1989). Fluctuating asymmetry as an indicator of stress: Implications for conservation biology. *Trends in Ecology and Evolution* **4,** 214–217.

Loève, M. (1945). Fonctions aléatoires de second ordre. *Comptes Rendus de l'Académie des Sciences Série I—Mathématique* **220,** 469.

Champely, S., B. Guinand, J. Thioulouse, and A. Clermidy. 1997. Functional data analysis of curve asymmetry with application to the color pattern of *Hydropsyche contubernalis* head capsule. Biometrics **53**:294-305.

Palmer, A. R. and Strobeck, C. (1986). Fluctuating asymmetry: Measurement, analysis, patterns. *Annual Review of Ecology and Systematics* **17,** 391–421.

Parsons, P. A. (1990). Fluctuating asymmetry: An epigenetic measure of stress. *Biological Review* **65,** 131–145.

Parsons, P. A. (1992). Fluctuating asymmetry: A biological monitor of environmental and genomic stress. *Heredity* **68,** 361–364.

Petersen, L. B. M. and Petersen, R. C. (1983). Anomalies in hydropsychid capture nets from polluted streams. *Freshwater Biology* **13,** 185–192.

Ramsay, J. O. (1982). When the data are functions. *Psychometrika* **47,** 379–396.

Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society, Series B* **53,** 539–572.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics* **14,** 1–17.

Rao, C. R. (1964). The use and interpretation of principal components analysis in applied research. *Sankhya A* **26,** 329–358.

Sabatier, R., Lebreton, J. D., and Chessel, D. (1989). Principal components analysis with instrumental variables as a tool for modelling composition data. In *Multiway Data Analysis,* R. Coppi and S. Bolasco (eds), 341–351. Amsterdam: North Holland.

Sagnes, P. (1995). Un outil de prise de données sur une image numérisée: exemple de la morphométrie des poissons. *Bulletin Français de la Pêche et de la Pisciculture* **337**.

Vuori, K.-M. (1993). Influence of water quality and feeding habits on the whole-body metal concentrations in lotic trichopteran larvae. *Limnologica* **23,** 301–308.

Vuori, K.-M. (1994). Rapid behavioural and morphological responses of Hydropsychid larvae (Trichoptera, Hydropsychidae) to sublethal cadmium exposure. *Environmental Pollution* **84,** 291–299.

Zakharov, V. M. (1989). Future prospects for population phenogenetics. *Soviet Scientific Reviews, Section F, Physiology and General Biology Reviews* **4,** 1–79.

Zakharov, V. M. (1992). Population phenogenetics: Analysis of developmental stability in natural populations. *Acta Zoologica Fennica* **191,** 7–30.

Zerbe, G. O. and Walker, S. H. (1977). A randomization test for comparison of groups of growth curves with different polynomial design. *Biometrics* **33,** 653–657.

Received February 1995; *revised July* 1996; *accepted September* 1996.

Appendix

Study Sites

These sites are distributed on four French rivers (see Figure 6).

—Rhine River: two sites (number 11 in Figure 6) in two parts of a hydroelectric power dam (Rhinau-Kappel); one was sampled in 1989 (denoted as "Rhin 89") and the other in 1990 (denoted as "Rhin 90"). The distance between the two sites is approximately 100 meters.

—Doubs River: two sites, "Crissey" (9) and "Rochefort" (10), separated by 10–12 kilometers (sampled in September 1992).

—Rhône River: two sites in the Upper Rhône between lake Geneva and the town of Lyon, separated by 5 kilometers and denoted, respectively, as "RCC Grand-Jean" (8) (sampled in November 1993) and "RCC" (sampled at the same date during two successive years, 1990 and 1991, thus denoted as "RCC 90" and "RCC 91") (7).

—Loire River: six sites located from upstream to downstream along approximately 700 kilometers, named as follows: "Montrond" (1), "Béard" (2), "Gien" (3), "Amboise" (4), "Gennes" (5), and "Montjean" (6). These sites were all sampled in March 1993, except for the first and last sites, which were sampled in June/July 1991. This represents, considering the temporal repetition for one Rhône River site, 13 distinct sites. The 321 color patterns have been analyzed.

Champely, S., B. Guinand, J. Thioulouse, and A. Clermidy. 1997. Functional data analysis of curve asymmetry with application to the color pattern of *Hydropsyche contubernalis* head capsule. Biometrics **53**:294-305.