

Méthode Quantitative Avancées en Environnement Marin

David Nerini

COM

second semestre

Motivation de ce cours

- Une des grandes tâches en statistiques consiste en l'analyse de grands tableaux de données construits à partir de l'observation de *variables* échantillonnées sur un grand nombre d'*individus*. On cherchera par exemple, à construire des indices de qualité susceptibles de résumer au mieux l'information contenue dans ces tableaux, tant du point de vue de leur structure en ligne qu'en colonne.

Motivation de ce cours

- Une des grandes tâches en statistiques consiste en l'analyse de grands tableaux de données construits à partir de l'observation de *variables* échantillonnées sur un grand nombre d'*individus*. On cherchera par exemple, à construire des indices de qualité susceptibles de résumer au mieux l'information contenue dans ces tableaux, tant du point de vue de leur structure en ligne qu'en colonne.
- Une autre tâche importante des statistiques consiste à la recherche de représentations graphiques en faible dimension qui permettent également de donner une approximation optimale de la structure d'un nuage de points. A partir de l'analyse de ces graphiques, on détectera des observations qui se ressemblent, des variables corrélées.

Motivation de ce cours

- Une des grandes tâches en statistiques consiste en l'analyse de grands tableaux de données construits à partir de l'observation de *variables* échantillonnées sur un grand nombre d'*individus*. On cherchera par exemple, à construire des indices de qualité susceptibles de résumer au mieux l'information contenue dans ces tableaux, tant du point de vue de leur structure en ligne qu'en colonne.
- Une autre tâche importante des statistiques consiste à la recherche de représentations graphiques en faible dimension qui permettent également de donner une approximation optimale de la structure d'un nuage de points. A partir de l'analyse de ces graphiques, on détectera des observations qui se ressemblent, des variables corrélées.
- En océanographie, la plupart des résultats de campagnes d'échantillonnage peuvent être consignées sous la forme de tableaux et les analyses qui en découlent font appels aux notions précédentes.

Exemple : structure d'une masse d'eau

- On dispose d'un échantillon $E = \{\mathbf{x}_i, i = 1, \dots, n\}$ construit à partir de mesures de température, salinité et oxygène dissous pendant n jours sur une même station. On cherche à résumer l'information contenue dans cette matrice de données d'abord par la recherche d'un petit nombre de variables non corrélées, combinaison linéaire des variables initiales.

Exemple : structure d'une masse d'eau

- On dispose d'un échantillon $E = \{\mathbf{x}_i, i = 1, \dots, n\}$ construit à partir de mesures de température, salinité et oxygène dissous pendant n jours sur une même station. On cherche à résumer l'information contenue dans cette matrice de données d'abord par la recherche d'un petit nombre de variables non corrélées, combinaison linéaire des variables initiales.
- On va ensuite chercher à regrouper les jours qui se ressemblent le plus du point de vue de leur structure hydrologique.

Exemple : structure d'une masse d'eau

Exemple

D. Nerini et al. / C. R. Acad. Sci. Paris, Sciences de la Terre et des planètes / Earth and Planetary Sciences 332 (2001) 457–464

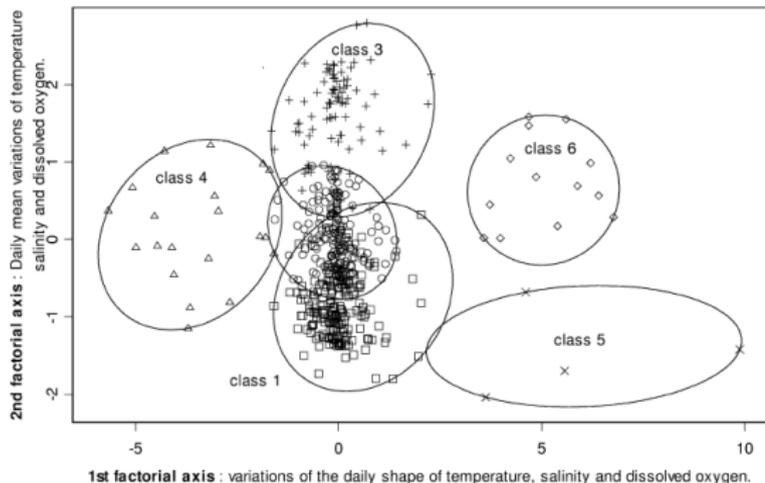
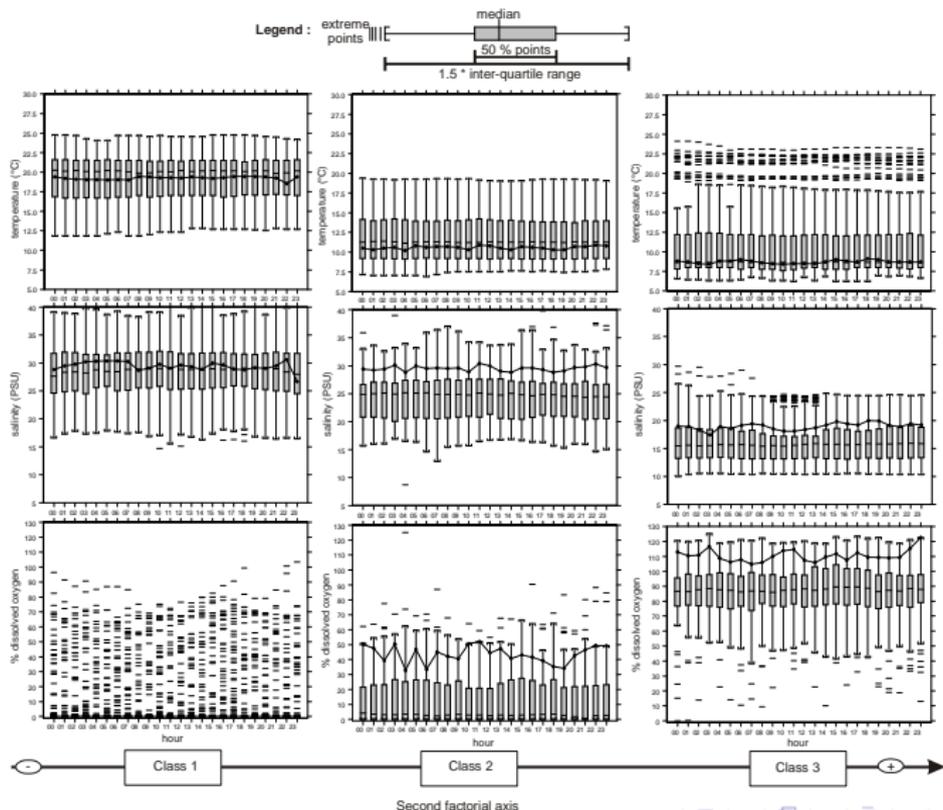


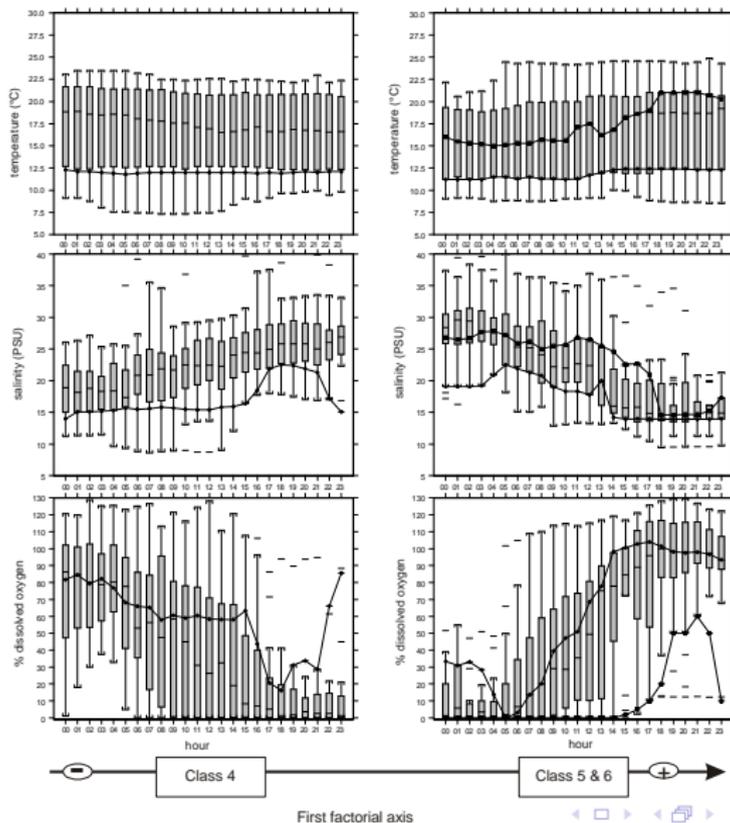
Figure 2. Plan factoriel 1/2 de l'AFM et ellipses de volume minimum [12] associées à chacune des classes issues de la CAH sur les huit premiers facteurs de l'AFM.

Figure 2. Bivariate plot (principal components 1/2) and minimum volume ellipsoids [12] associated to the hierarchical clustering of the data using eight principal components of the MFA.

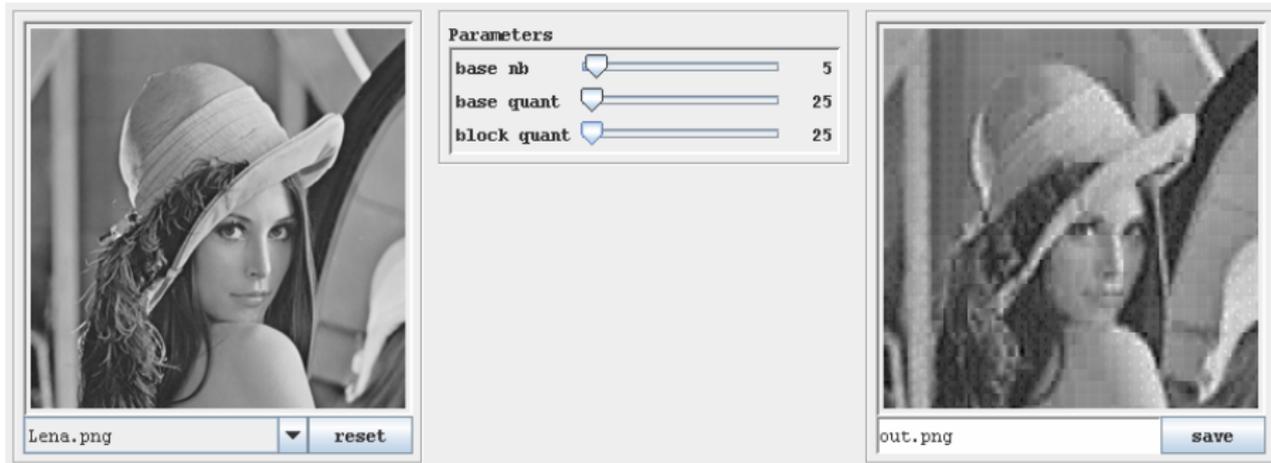
Exemple : structure d'une masse d'eau



Exemple : structure d'une masse d'eau



Exemple : compression d'images



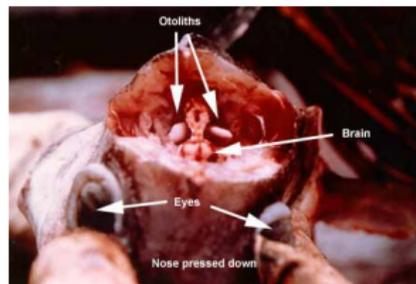
The screenshot shows a web-based image compression tool. On the left is the original image 'Lena.png'. In the center is a 'Parameters' control panel with three sliders: 'base nb' (set to 5), 'base quant' (set to 25), and 'block quant' (set to 25). On the right is the compressed image 'out.png', which is visibly more compressed and has a 'save' button below it.

Parameters		
base nb	<input type="range"/>	5
base quant	<input type="range"/>	25
block quant	<input type="range"/>	25

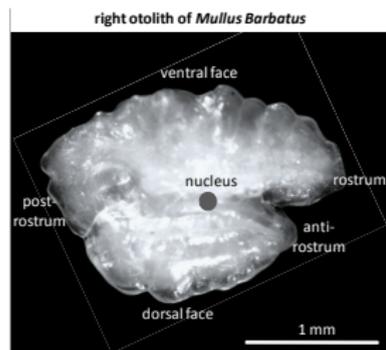
- On cherche à conserver au mieux la structure de l'image en diminuant sa taille au maximum

Exemple : analyse de forme

Exemple (position et forme)

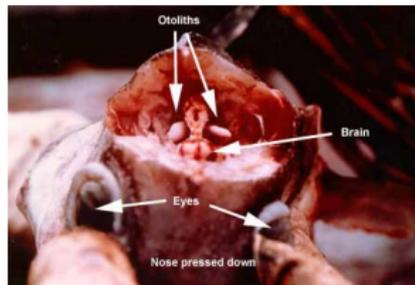


- Un otolithe est une petite pièce de carbonate de calcium situé dans l'oreille interne des Téléostéens

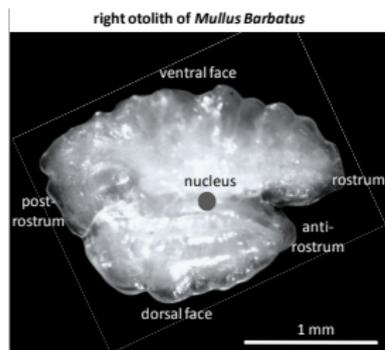


Exemple : analyse de forme

Exemple (position et forme)

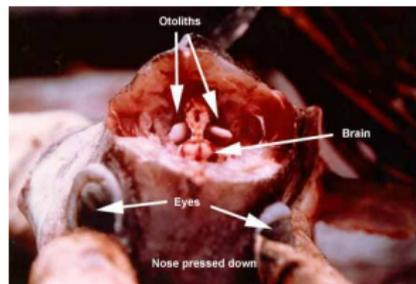


- Un otolithe est une petite pièce de carbonate de calcium situé dans l'oreille interne des Téléostéens
 - perception des sensations d'accélération et sens de l'équilibre

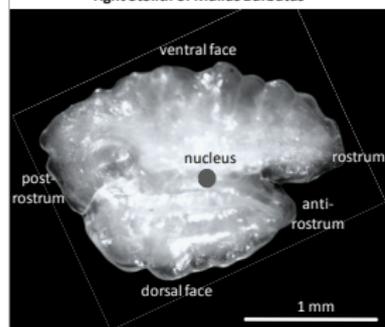


Exemple : analyse de forme

Exemple (position et forme)



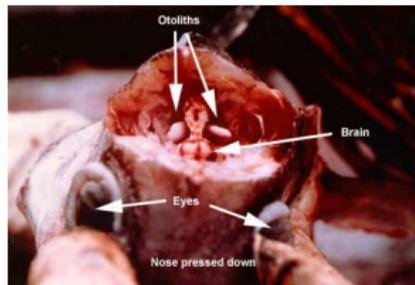
right otolith of *Mullus Barbatus*



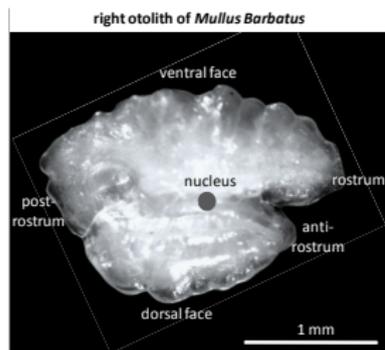
- Un otolithe est une petite pièce de carbonate de calcium situé dans l'oreille interne des Téléostéens
 - perception des sensations d'accélération et sens de l'équilibre
 - aide à l'audition et perception tridimensionnelle

Exemple : analyse de forme

Exemple (position et forme)

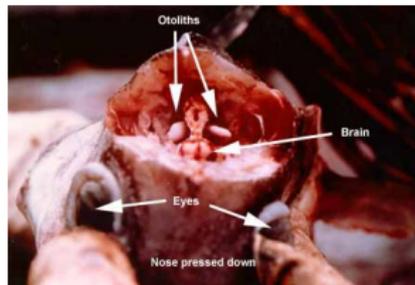


- Un otolithe est une petite pièce de carbonate de calcium situé dans l'oreille interne des Téléostéens
 - perception des sensations d'accélération et sens de l'équilibre
 - aide à l'audition et perception tridimensionnelle
- 3 paires



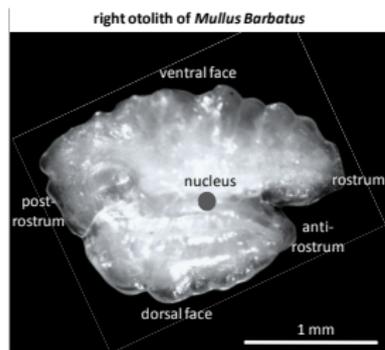
Exemple : analyse de forme

Exemple (position et forme)



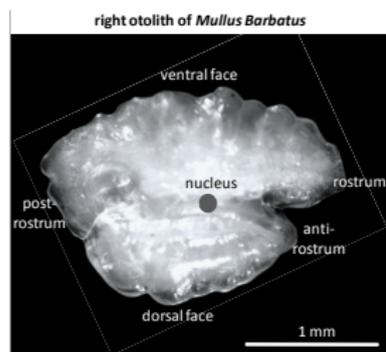
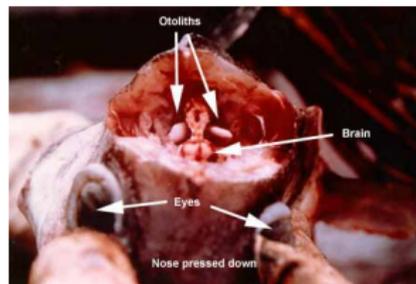
- Un otolithe est une petite pièce de carbonate de calcium situé dans l'oreille interne des Téléostéens
 - perception des sensations d'accélération et sens de l'équilibre
 - aide à l'audition et perception tridimensionnelle

- 3 paires
- paire de *sagittae*



Exemple : analyse de forme

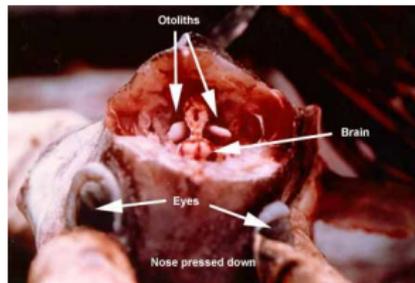
Exemple (position et forme)



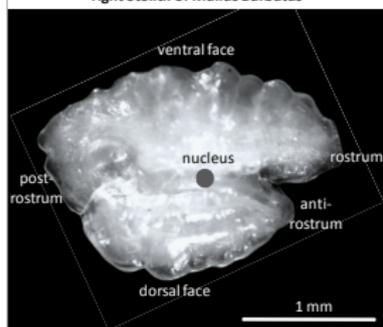
- Un otolithe est une petite pièce de carbonate de calcium situé dans l'oreille interne des Téléostéens
 - perception des sensations d'accélération et sens de l'équilibre
 - aide à l'audition et perception tridimensionnelle
- 3 paires
- paire de *sagittae*
 - la plus grosse

Exemple : analyse de forme

Exemple (position et forme)



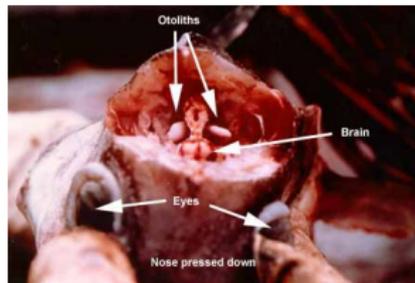
right otolith of *Mullus Barbatus*



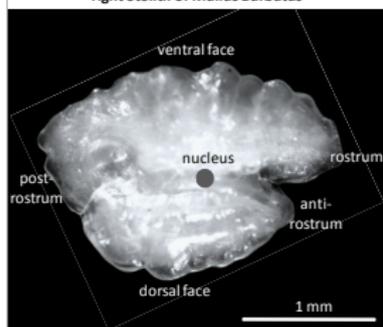
- Un otolithe est une petite pièce de carbonate de calcium situé dans l'oreille interne des Téléostéens
 - perception des sensations d'accélération et sens de l'équilibre
 - aide à l'audition et perception tridimensionnelle
- 3 paires
- paire de *sagittae*
 - la plus grosse
 - grossit depuis le stade larvaire jusqu'à la mort de l'animal

Exemple : analyse de forme

Exemple (position et forme)



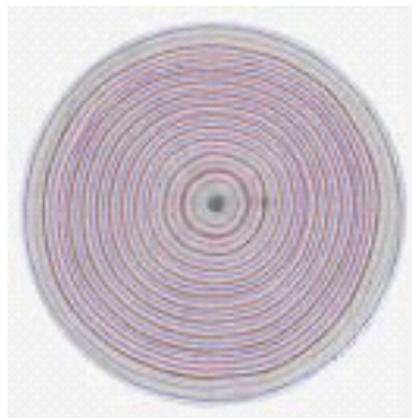
right otolith of *Mullus Barbatius*



- Un otolithe est une petite pièce de carbonate de calcium situé dans l'oreille interne des Téléostéens
 - perception des sensations d'accélération et sens de l'équilibre
 - aide à l'audition et perception tridimensionnelle
- 3 paires
- paire de *sagittae*
 - la plus grosse
 - grossit depuis le stade larvaire jusqu'à la mort de l'animal
 - croissance à partir du nucleus

Pourquoi étudier la forme des otolithes ?

- Les stocks exploités, les populations et les espèces peuvent être étudiées à partir de la composition et de la forme des otolithes
- La composition chimique et la forme renseignent sur l'environnement du poisson (temperature, voies migratoires) et sur son régime alimentaire
- Une section latérale fait apparaître des cernes journalières de croissance qui enregistrent les traits de vie de l'animal



Analyse de forme

- On cherche à comparer la forme d'un ensemble d'otholites

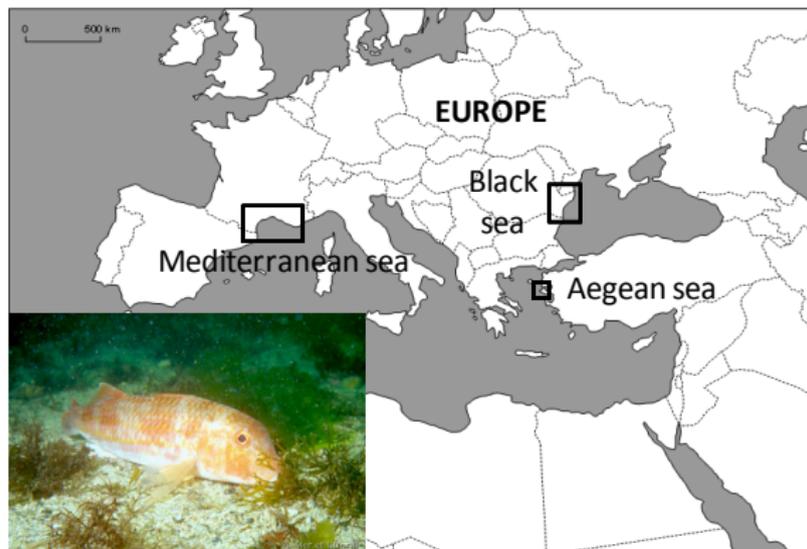
Analyse de forme

- On cherche à comparer la forme d'un ensemble d'otholites
 - **objectif** : étudier la variabilité des formes à partir d'individus collectés à différents endroits

Analyse de forme

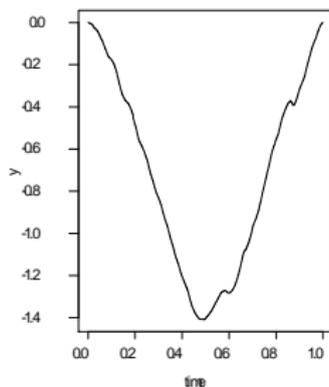
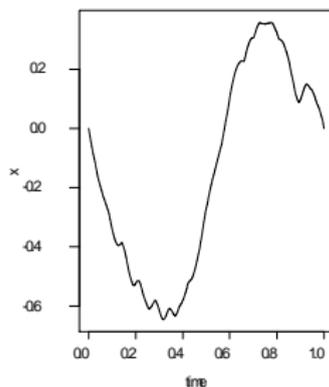
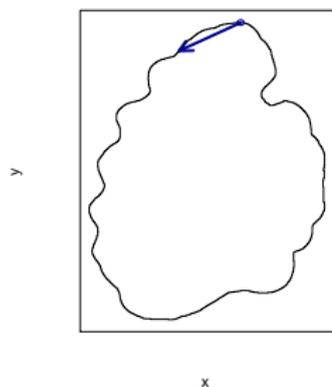
- On cherche à comparer la forme d'un ensemble d'otholites
 - **objectif** : étudier la variabilité des formes à partir d'individus collectés à différents endroits

Exemple (Localisation des échantillons & *Mullus Barbatus*)



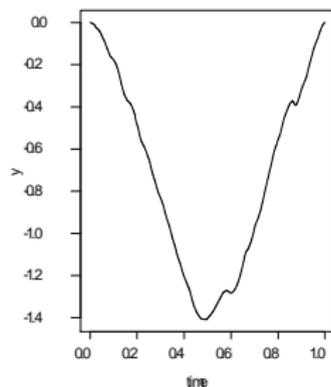
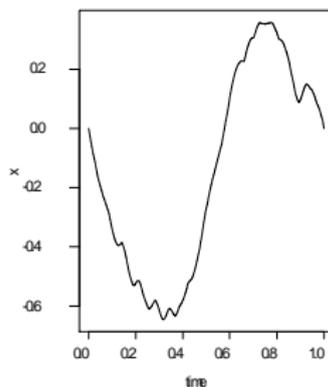
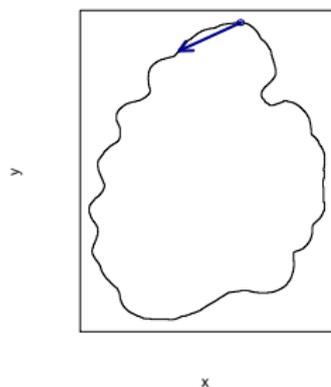
Analyse de forme

- On dispose d'un échantillon de $n = 305$ images haute résolution d'otolithes droit de *Mullus barbatus*



Analyse de forme

- On dispose d'un échantillon de $n = 305$ images haute résolution d'otolithes droit de *Mullus barbatus*
- Les contours sont extraits par traitement d'image en utilisant des techniques de seuillage



Approximation fonctionnelle

- Les fonctions $x(t)$ et $y(t)$ peuvent être décomposées comme une combinaison linéaire de K fonctions ϕ_k :

$$\begin{cases} x(t) = \sum_{k=1}^K \alpha_k \phi_k(t) \\ y(t) = \sum_{k=1}^K \beta_k \phi_k(t) \end{cases}$$

Approximation fonctionnelle

- Les fonctions $x(t)$ et $y(t)$ peuvent être décomposées comme une combinaison linéaire de K fonctions ϕ_k :

$$\begin{cases} x(t) = \sum_{k=1}^K \alpha_k \phi_k(t) \\ y(t) = \sum_{k=1}^K \beta_k \phi_k(t) \end{cases}$$

- Les vecteurs de paramètres α et β sont estimés par régression polynomiale

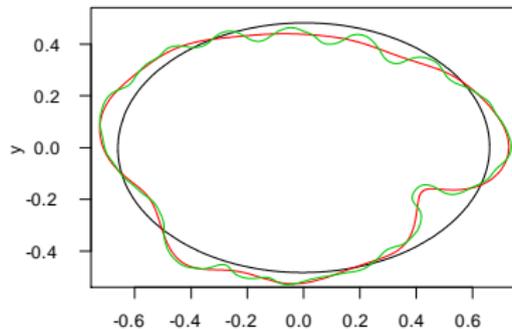
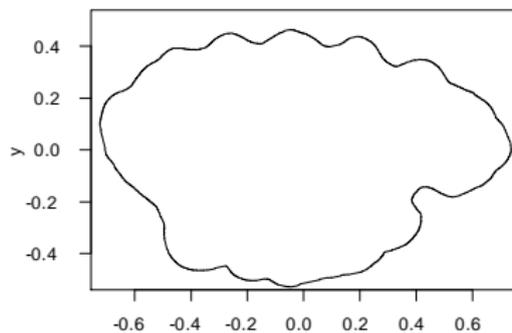
Approximation fonctionnelle

- Les fonctions $x(t)$ et $y(t)$ peuvent être décomposées comme une combinaison linéaire de K fonctions ϕ_k :

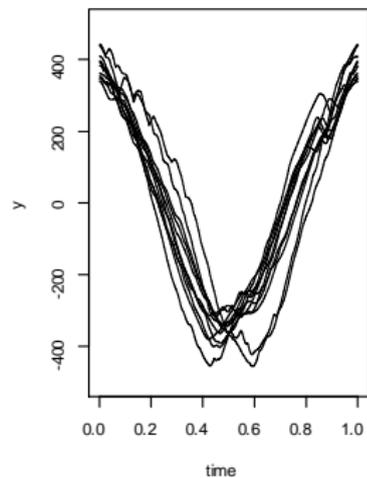
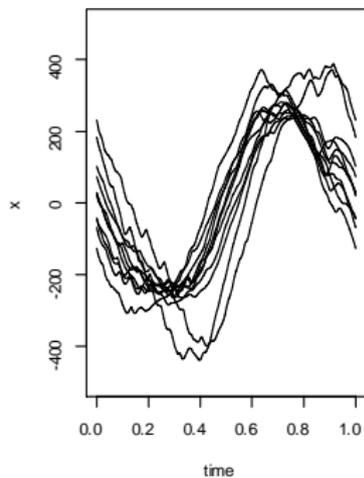
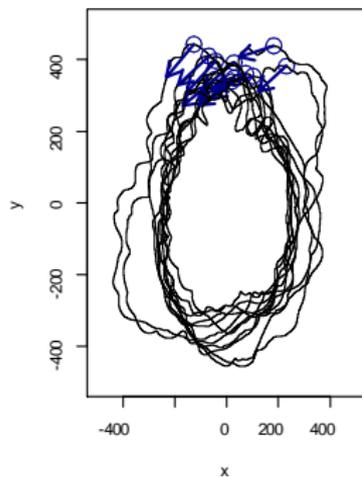
$$\begin{cases} x(t) = \sum_{k=1}^K \alpha_k \phi_k(t) \\ y(t) = \sum_{k=1}^K \beta_k \phi_k(t) \end{cases}$$

- Les vecteurs de paramètres α et β sont estimés par régression polynomiale
- Le graphique (x, y) donne le contour ajusté de l'otolithe

Approximation fonctionnelle

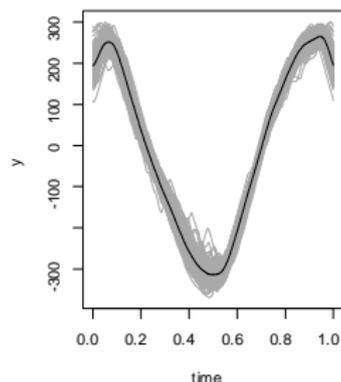
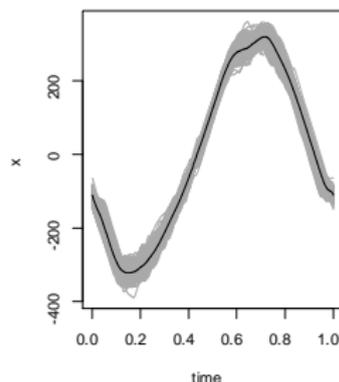
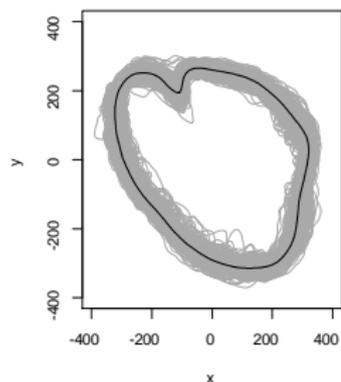


Un problème de recalage

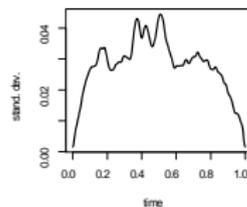
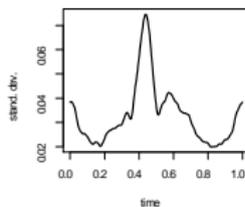
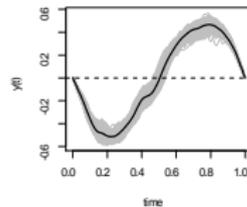
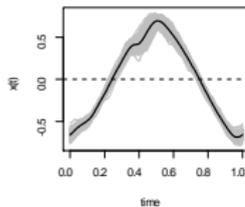
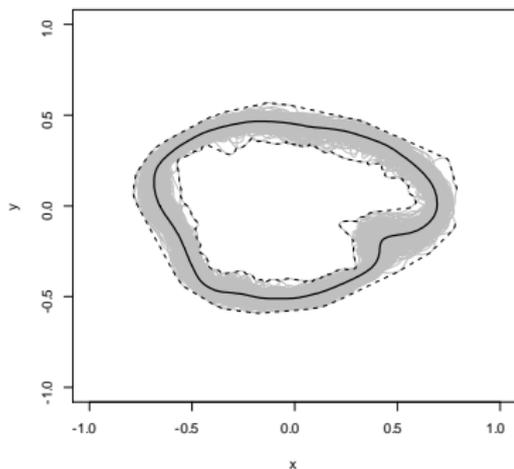


Analyse procuste généralisée

- On s'affranchit des problèmes de positionnement des objets entre eux et des différences de taille

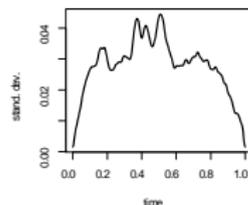
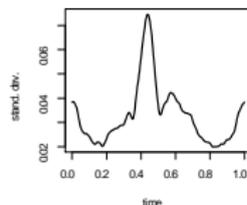
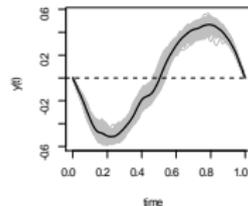
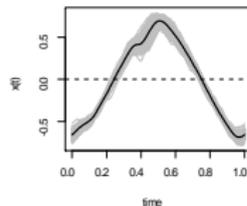
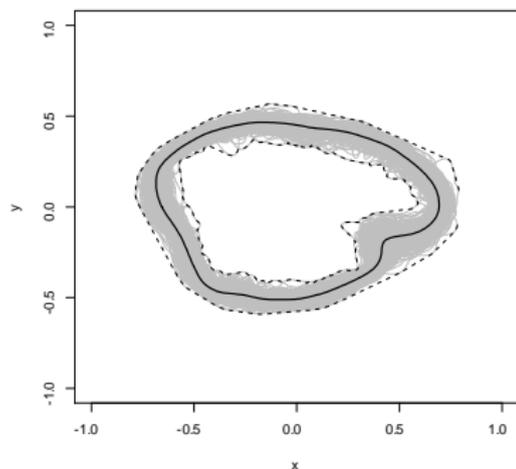


Basic statistics



- The average otolith is defined as $\{(\bar{x}(t), \bar{y}(t)), t \in [0; 1]\}$

Basic statistics



- The average otolith is defined as $\{(\bar{x}(t), \bar{y}(t)), t \in [0; 1]\}$
- The variance can be computed and allows to identify variability zones

PCA in functional spaces

- With each function $y_i(t)$ expressed in terms of K known basis functions, searching for an $\zeta(s)$ leads to a simplified problem :

$$\frac{1}{N} C' C W b = \lambda b$$

where C denotes the $N \times K$ matrix of centred coefficients of the basis decomposition,

$W = \int \phi' \phi$ is the matrix of inner products of basis functions

b is the coefficient vector of $\zeta(s)$

PCA in functional spaces

- With each function $y_i(t)$ expressed in terms of K known basis functions, searching for an $\zeta(s)$ leads to a simplified problem :

$$\frac{1}{N} C' C W b = \lambda b$$

where C denotes the $N \times K$ matrix of centred coefficients of the basis decomposition,

$W = \int \phi' \phi$ is the matrix of inner products of basis functions

b is the coefficient vector of $\zeta(s)$

- The eigenfunction $\zeta(s)$ is constructed with b as $\zeta(t) = \sum_{k=1}^K b_k \phi_k(t)$

PCA in functional spaces

- With each function $y_i(t)$ expressed in terms of K known basis functions, searching for an $\zeta(s)$ leads to a simplified problem :

$$\frac{1}{N} C' C W b = \lambda b$$

where C denotes the $N \times K$ matrix of centred coefficients of the basis decomposition,

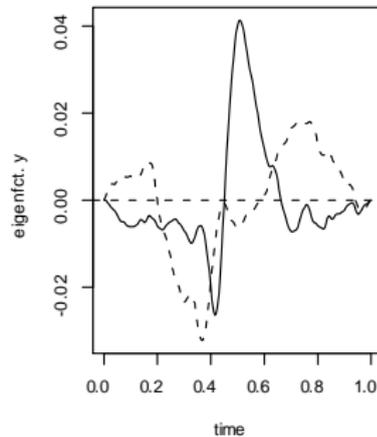
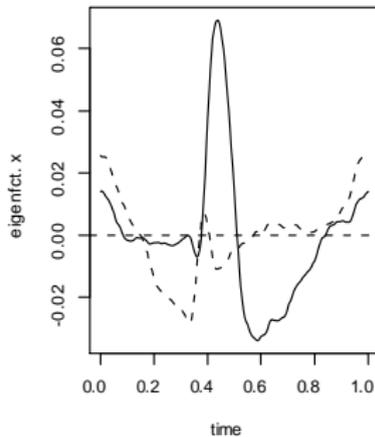
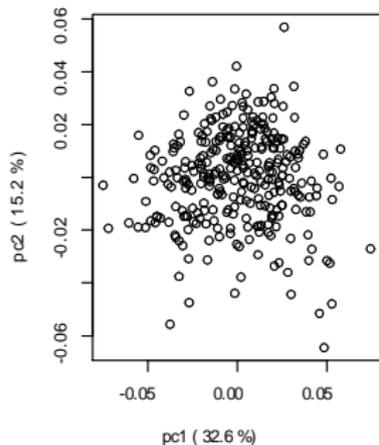
$W = \int \phi' \phi$ is the matrix of inner products of basis functions

b is the coefficient vector of $\zeta(s)$

- The eigenfunction $\zeta(s)$ is constructed with b as $\zeta(t) = \sum_{k=1}^K b_k \phi_k(t)$
- The generalization in dimension 2 is easily tractable since

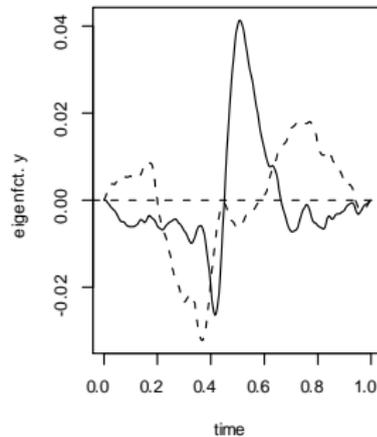
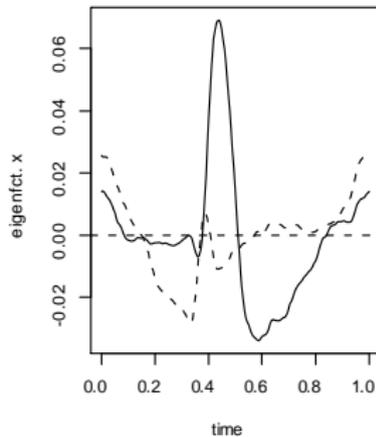
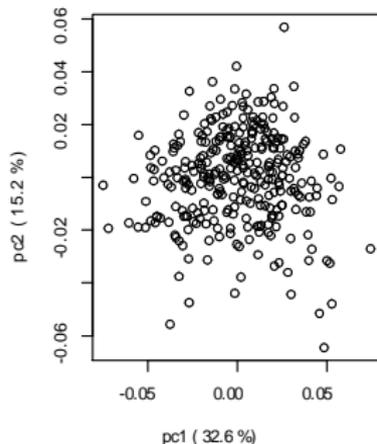
$$\zeta = \begin{pmatrix} \zeta^x \\ \zeta^y \end{pmatrix}'$$

PCA in functional spaces



- Rostrum shape variations ... and after ?

PCA in functional spaces



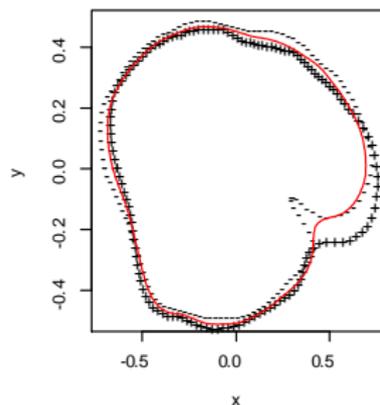
- Rostrum shape variations ... and after ?
- Each otolith is a linear combination of these independent sources of variability

PCA in functional spaces

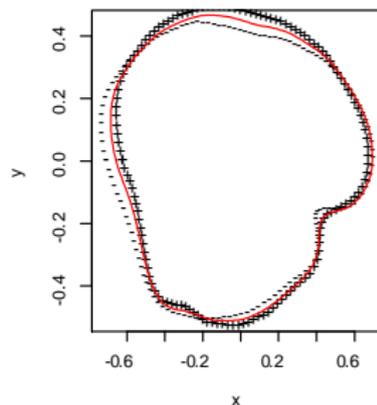
The display of an eigenfunction m is achieved by perturbation of the mean functions such that :

$$\begin{cases} \bar{x}(t) \pm \sqrt{\lambda_m} \zeta_m^x(t) \\ \bar{y}(t) \pm \sqrt{\lambda_m} \zeta_m^y(t) \end{cases}$$

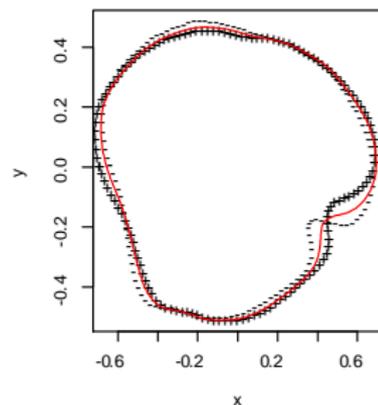
1st comp. 32.6 %



2nd comp. 15.2 %



3rd comp. 9.6 %



Plan du cours

- 1 Notation de base
- 2 Rappel d'algèbre
 - 1 Distance
 - 2 produit scalaire
 - 3 forme quadratique
- 3 Décomposition d'une matrice
 - 1 rang d'une matrice
 - 2 décomposition d'une matrice de rang 2
 - 3 généralisation
 - 4 interprétation géométrique
- 4 Analyse en composante principale
 - 1 centrage et inertie
 - 2 approximation globale du tableau
 - 3 décomposition ligne et colonne
 - 4 mesure de la qualité
 - 5 illustration