

Guide d'utilisation de la plateforme PAMPA WP2

Aide au calculs, à la représentation et l'analyse des métriques biodiversité et ressources

Auteurs : Yves REECHT (Yves.Reecht@ifremer.fr), Romain DAVID, Jérémie HABASQUE et Bastien PREUSS

Résumé : *Ce document vise à orienter les utilisateurs (gestionnaires et scientifiques) dans l'utilisation de la plateforme WP2 afin de calculer des métriques biodiversité et ressources à partir de leur jeu de données. Il les guidera également dans l'utilisation de l'interface pour générer des graphiques et conduire des analyses statistiques.*

Informations importantes : *ce document contient des avertissements importants (ayant la même apparence que ce cadre) – notamment en ce qui concerne l'installation de la plateforme – que vous êtes invités à lire attentivement avant installation et/ou utilisation.*

Ce document doit être cité comme suit :

Reecht Y., David R., Habasque J. et Preuss B. Guide d'utilisation de la plateforme PAMPA WP2 – Aide au calculs, à la représentation et l'analyse des métriques biodiversité et ressources. Document interne PAMPA/WP2/Meth/4. Version du 19/04/2011. 38 p.

Table des matières

1. Présentation.....	3
2. Données analysées.....	4
A. Type de données d'entrée.....	4
B. Format des fichiers de données.....	4
3. Installation de l'environnement de travail et configuration.....	5
A. Prérequis : installation de R.....	5
B. Installation de la plateforme.....	5
C. Création du dossier de travail.....	6
D. Configuration.....	7
i. Dossier de travail et fichiers de données.....	7
ii. Catégories benthiques supplémentaires.....	8
4. Chargement de l'interface.....	8
A. Démarrage de l'application.....	8
B. Packages.....	9
5. Importation des données.....	10
A. Dossiers et fichiers de données.....	10
B. Importation des fichiers par défaut.....	11
C. Informations de chargement.....	11
D. Opérations « spéciales » lors du chargement et calculs divers.....	12
i. tailles d'après les classes de tailles.....	12
ii. Zéros.....	12
iii. Année de campagne.....	13
E. Sélection et re-calcul.....	13
6. Graphiques.....	15
A. Métriques agrégées par unité d'observation par espèce (et éventuellement par classe de taille).....	15
i. Boîtes à moustaches ou Boxplots.....	16
ii. Diagrammes en barres ou Barplots.....	17
iii. Remarques.....	18
B. Métriques agrégées par unité d'observation (et éventuellement par classe de taille).....	18
i. Boîtes à moustaches ou Boxplots.....	18
ii. Diagrammes en barres ou barplots.....	19
iii. Remarques.....	20
C. Remarques générales sur les graphiques.....	22
i. Rang d'utilisation du facteur « statut de protection » :.....	22
ii. Options graphiques.....	22
7. Analyses statistiques.....	24
A. Métriques agrégées par unité d'observation par espèce (et éventuellement par	

classe de taille).....	25
B. Métriques agrégées par unité d'observation (et éventuellement par classe de taille)	27
C. Résultats.....	28
i. Informations sur le modèle.....	29
ii. détails sur les facteurs significatifs et leurs coefficients	29
iii. Valeurs prédites.....	30
iv. Comparaisons multiples (2 facteurs).....	30
v. Comparaisons multiples (1 facteur).....	32
D. Graphiques diagnostiques et valeurs aberrantes.....	32
i. GLMs et graphiques diagnostiques.....	33
ii. ANOVAs et graphiques diagnostiques.....	34

1. Présentation

Un programme de calcul d'indicateurs de ressources et biodiversité a été développé sous le logiciel R dans le cadre du projet PAMPA. Ce document a pour but de guider l'utilisateur (gestionnaire et scientifique) dans l'utilisation de ce programme. Le programme fonctionne sous Windows, Mac et Linux.

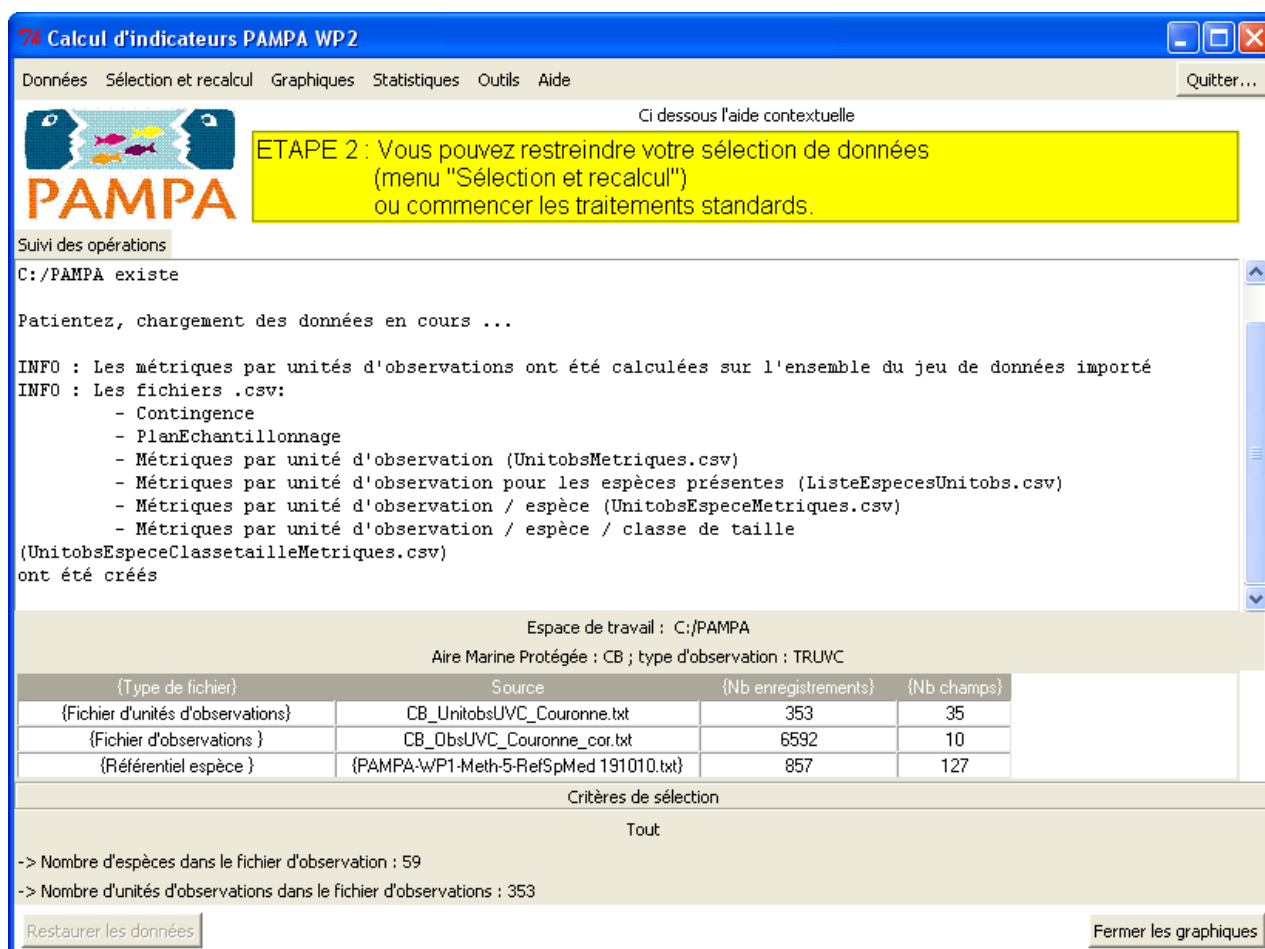
Le détail des métriques calculées figurent dans le document : « Métriques biodiversité et ressources PAMPA/WP2/Meth/1 ».

L'équipe coordinatrice du projet fournit aux utilisateurs :

- le programme R de calcul des métriques,
- l'aide aux utilisateurs sur le formatage des données (à la fin de ce document)
- le référentiel espèces.

L'interface est constituée d'un menu « déchirable » (comprendre « qui s'ouvre »), contenant plusieurs rubriques. Les rubriques correspondent à différentes étapes d'analyse de vos jeux de données. Chacune d'entre elle contient des sous menus correspondant à des types d'analyses et de graphiques.

Si l'interface n'a, au premier abord, que peu changé au cours des derniers mois, les menus ont été – dans sa nouvelle mouture – complètement réorganisées, le code a été nettoyé et de nombreuses fonctionnalités ajoutées. Il n'y a notamment plus maintenant de différence, dans les menus actifs après chargement, pour des données benthos ou autres. Seules les métriques disponibles dépendent des données.



Calcul d'indicateurs PAMPA WP2

Données Sélection et recalcul Graphiques Statistiques Outils Aide

Quitter...

Ci dessous l'aide contextuelle

ETAPE 2 : Vous pouvez restreindre votre sélection de données (menu "Sélection et recalcul") ou commencer les traitements standards.

Suivi des opérations

C:/PAMPA existe

Patientez, chargement des données en cours ...

INFO : Les métriques par unités d'observations ont été calculées sur l'ensemble du jeu de données importé

INFO : Les fichiers .csv:

- Contingence
- PlanEchantillonnage
- Métriques par unité d'observation (UnitobsMetriques.csv)
- Métriques par unité d'observation pour les espèces présentes (ListeEspecesUnitobs.csv)
- Métriques par unité d'observation / espèce (UnitobsEspeceMetriques.csv)
- Métriques par unité d'observation / espèce / classe de taille (UnitobsEspeceClassetailleMetriques.csv)

ont été créés

Espace de travail : C:/PAMPA

Aire Marine Protégée : CB ; type d'observation : TRUVC

{Type de fichier}	Source	{Nb enregistrements}	{Nb champs}
{Fichier d'unités d'observations}	CB_UnitobsUVC_Couronne.txt	353	35
{Fichier d'observations}	CB_ObsUVC_Couronne_cor.txt	6592	10
{Référentiel espèce}	{PAMPA-WP1-Meth-5-RefSpMed 191010.txt}	857	127

Critères de sélection

Tout

-> Nombre d'espèces dans le fichier d'observation : 59

-> Nombre d'unités d'observations dans le fichier d'observations : 353

Restaurer les données

Fermer les graphiques

Le reste de l'interface (en dessous des menus) est constituée de deux champs de texte, dont une aide contextuelle en jaune et un champs d'information en blanc intitulé « suivi des opérations » qui vous permet de suivre le déroulement de vos principales analyses.

Trois zones d'informations supplémentaires vous permettent en un coup d'œil de visualiser l'état du jeux de données que vous traitez :

- une zone texte indiquant l'emplacement de votre zone de travail, l'AMP considérée et le type d'observation en cours de traitement.
- un tableau présentant les fichiers observation, unités d'observation et référentiel que vous avez saisis avec leurs noms, leurs nombre de champs et d'enregistrements. Plus, le cas échéant, une quatrième colonne indiquant le nombre d'éléments (observations) de la sélection.
- une zone « critère de sélection » qui affiche le critère que vous avez sélectionné le cas échéant et le nombre d'espèces et unités d'observation restantes dans votre fichier d'observations.

2. Données analysées

A. Type de données d'entrée

Plusieurs types de données – qui nécessitent des traitements particuliers – sont supportés par la plateforme :

- **UVC** (*underwater visual censuses* – comptages visuels sous-marins) : utilisé pour la faune sous-marine... essentiellement les poissons.
- **LIT** (*Line Intercept Transect*) : type d'UVC utilisé pour le benthos et l'habitat.
- **SVR** : vidéos rotatives, avec séparation des rotations en secteurs. Nécessite un traitement particulier, et notamment des interpolations des secteurs non-valides (pas analysables).
- Données de pêche, qui se subdivisent en plusieurs sous-catégories :
 - **DEB** : débarquements.
 - **EMB** : échantillonnage par observateur(s) embarqué(s).
 - **PSCI** : pêches scientifiques.
 - **PecRec** : enquêtes de pêche récréative.

Un seul type de donnée peut-être analysé à la fois, en raison, notamment, des différences de traitement et, dans certains cas, d'unités des métriques calculées.

B. Format des fichiers de données

Quelque soit le type de données traitées, l'information doit être répartie en 3 fichiers texte (.txt) distincts : Table contenant les unités d'observation. Table contenant les observations (comptages, captures, etc.) Le référentiel espèces (Méditerranée ou Outre-Mer) fourni par l'équipe coordinatrice.

Le format d'entrée des données pour exécuter les programmes de calcul est standardisé

(Cf. Formats de données – PAMPA/WP2/Meth/3). Des problèmes peuvent se poser lorsque le format des données n'est pas respecté. Ceux ci sont détaillés en fin de document.

Quelques rappels importants :

- Le séparateur décimal des fichiers texte est le point.
- Le séparateur de champ des fichiers texte est la tabulation.
- Les fichiers doivent être préalablement inspectés pour ne contenir ni espaces (à remplacer par des *underscores*, « _ »), ni « ; ».

3. Installation de l'environnement de travail et configuration

A. Prérequis : installation de R

Le logiciel R peut être récupéré en vous rendant sur la page du projet <http://cran.cict.fr/> (vous pouvez choisir un [autre miroir plus proche de chez vous](#)). Vous y trouverez le programme et/ou les instructions d'installation pour votre système.

Il est pour l'instant conseillé d'installer la [version 2.11.1](#) qui se trouve être la plus récente permettant le lancement de l'interface à l'aide des raccourcis (voir section suivante).

Il n'est désormais plus nécessaire d'installer manuellement les *packages* requis (cf. 4.B.)... à moins que vous ne disposiez pas des droits d'administration sur votre ordinateur, auquel cas ils devront être installés par votre administrateur.

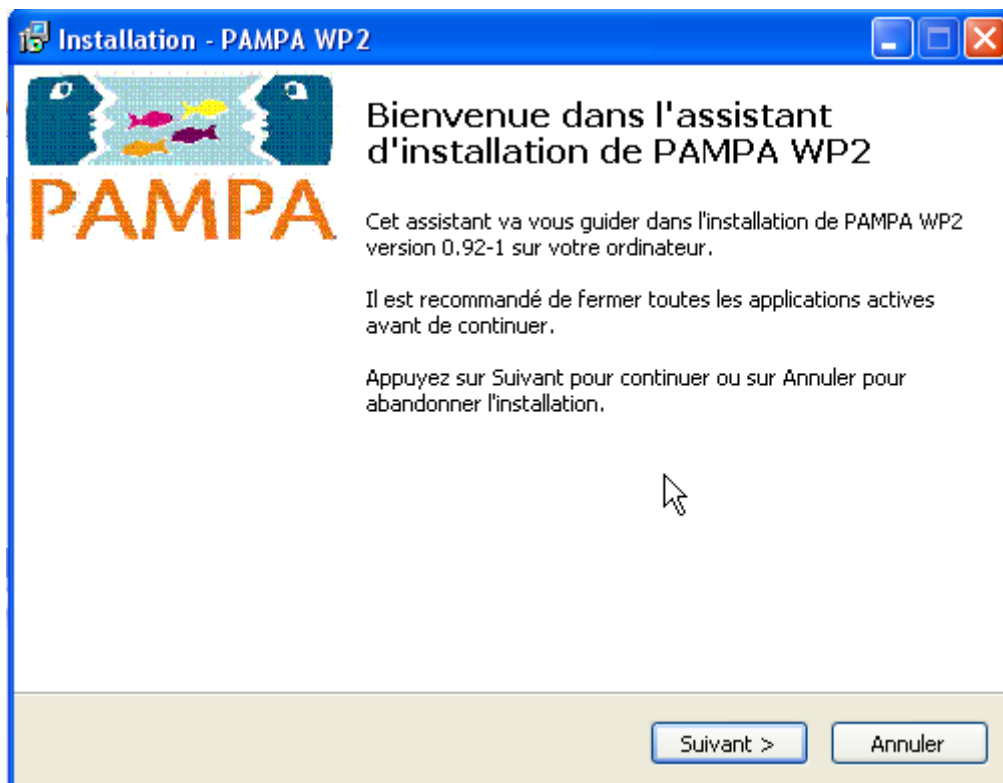
B. Installation de la plateforme

La plateforme est maintenant dotée d'un installeur pour Windows, qui place les scripts dans les dossiers adéquats et crée des raccourcis pour lancer R et l'interface graphique.

Attention : cet installeur étant pour l'instant très basique, il écrase les fichiers existants sans demander confirmation. Vous devez donc soit copier à un autre endroit, soit renommer les fichiers d'une ancienne installation que vous avez modifiés par vous même, si vous souhaitez ne pas perdre les informations qu'il contiennent.

Si le nom de l'installeur contient « **update** » le fichier `config.r` (qui doit contenir les noms de vos fichiers de données) sera cependant conservé. Si ce n'est pas le cas, il est recommandé de le sauvegarder. Vous pourrez ensuite copier votre configuration dans le fichier nouvellement installé (*ne pas le remplacer, il vous manquerait des lignes de code nécessaires au bon fonctionnement de la plateforme*).

En exécutant le fichier « <Setup|Update>_PAMPA_WP2_<version>.exe » vous serez dirigé vers une procédure d'installation classique (sous Windows) :



Je vous fais grâce de copies d'écran de chaque étape ; les différentes options sont suffisamment explicites et les valeurs par défaut devraient suffire pour la grande majorité des usages.

Attention : ne modifiez pas le chemin d'installation par défaut (« `C : / PAMPA / Exec` »), la plateforme ne serait pas fonctionnelle... à moins que vous ne soyez prêt à faire vous même les modifications qui s'imposeraient dans le code.

Il est prévu d'assouplir, dans un avenir proche, la gestion des dossiers, mais cette fonctionnalité n'est, à l'heure actuelle, pas implémentée.

Une fois l'installation terminée, vous disposez de raccourcis pour lancer R et l'interface graphique, ainsi que d'accès directs aux documentations et un modèle de rapport de *bug* :



C. Création du dossier de travail

La plateforme utilise un dossier de travail dans lequel elle va chercher les répertoires de données et de résultats. Par défaut, il s'agit de `C : / PAMPA /`, mais vous pouvez en choisir un autre (et même en avoir plusieurs pour l'analyse de différents jeux de données). Vous devez créer, dans ce répertoire, un dossier « *Data* » qui contiendra vos jeux de données.

Il est ensuite nécessaire de renseigner ce dossier de travail dans le fichier de configuration « `C : / PAMPA / Exec / config.r` » (voir la section suivante, « 3.D. Configuration »).

D. Configuration

i. Dossier de travail et fichiers de données

Le dossier d'installation contient un fichier dédié à la configuration des fichiers de données et du répertoire de travail.

[C:/PAMPA/Exec/config.r](#)

En paramétrant correctement ce fichier, vous pourrez recharger vos données en un clic à chaque lancement de l'interface !!

Il est conseillé afin de recharger les données en un seul clic, d'indiquer au programme le nom de vos fichiers d'unité d'observation, d'observation et de référentiel.

Dans ce fichier de configuration, changer les valeurs correspondante comme indiqué ci-dessous, en y mettant vos noms de fichier (en les respectant les minuscules et majuscules !).

Le dossier de travail peut également être renseigné à ce niveau là. Les lignes éditables ressemblent donc à ce qui suit :

```
#### CB
SiteEtudie <- "CB" # Identification du site.
fileName1 <- "CB_UnitobsUVC_Couronne.txt" # Unités d'observation.
fileName2 <- "CB_ObsUVC_Couronne.txt" # Fichier d'observations.
fileName3 <- "refEspecesMED.txt" # Référentiel espèce.
nameWorkspace <- "C:/PAMPA/WD" # Répertoire de travail.
```

En utilisant le système de commentaire de R (voir les descriptions de fichier dans le précédent cadre), vous pouvez conserver la trace de plusieurs configurations et y revenir lorsque vous le souhaitez. Toute ligne qui commence par # sera considérée comme du commentaire, et donc ignorée :

```
## SiteEtudie <- "CB"
## fileName1 <- "unitobs_CB_peche_pro.txt"
## fileName2 <- "obs_biodiv_CB_peche_pro.txt"
## fileName3 <- "refEspecesMED.txt"
## nameWorkspace <- "C:/PAMPA/Pêche_pro"
```

Après chaque modification, n'oubliez pas de commenter (faire précéder de « # ») les lignes de configuration devenues inutiles.

Ces fichiers peuvent alors être (i) directement chargés par depuis le menu « Données » en cliquant sur l'entrée « Dossiers et fichiers par défaut » (anciennement « C:/PAMPA/ par défaut », (ii) modifiés avec l'entrée « Choix des dossiers et fichiers de données ».

Ce fichier faisant parfois l'objet de modifications profondes, vous ne devriez pas l'écraser avec une ancienne version après une nouvelle installation. Il est toutefois recommandé de faire une sauvegarde de votre ancien « config.r » pour pouvoir par la suite copier votre configuration dans le nouveau fichier (cf. encadré ci-dessous).

ii. Catégories benthiques supplémentaires

Il est maintenant possible d'ajouter facilement des catégories benthiques supplémentaires, correspondant à des regroupement des catégories existantes. Celles-ci sont définies dans un fichier « `corresp-cat-benth.csv` » (dans `C:/PAMPA/Exec`) qui peut être édité avec un tableur de type Excel. Ce tableau donne les correspondances entre le champs « `Cat_benthique` » et les nouvelles catégories et se présente comme suit :

	A	B	C	
1	Cat_benthique	CategB_general	CategB_groupe	
2	AA	PLANT	AA	
3	ACB	CV	ACR	
4	ACD	CV	ACR	
5	ACE	CV	ACR	
6	ACS	CV	ACR	
7	ACT	CV	ACR	
8	CA	PLANT	NACR	
9	CB	CV	NACR	
10	CBL	CV	CBL	
11	CE	CV	NACR	
12	CF	CV	NACR	
13	CHL	CV	NACR	

La première colonne doit contenir obligatoirement les valeurs du champs « `Cat_benthique` » du référentiel espèce. Les suivantes – non limitées en nombre – représentent les catégories agrégées. Des nouveaux facteurs seront alors disponibles sous les noms affichés en première ligne (éviter les espaces et caractères spéciaux) pour produire des graphiques ou faire des analyses statistiques.

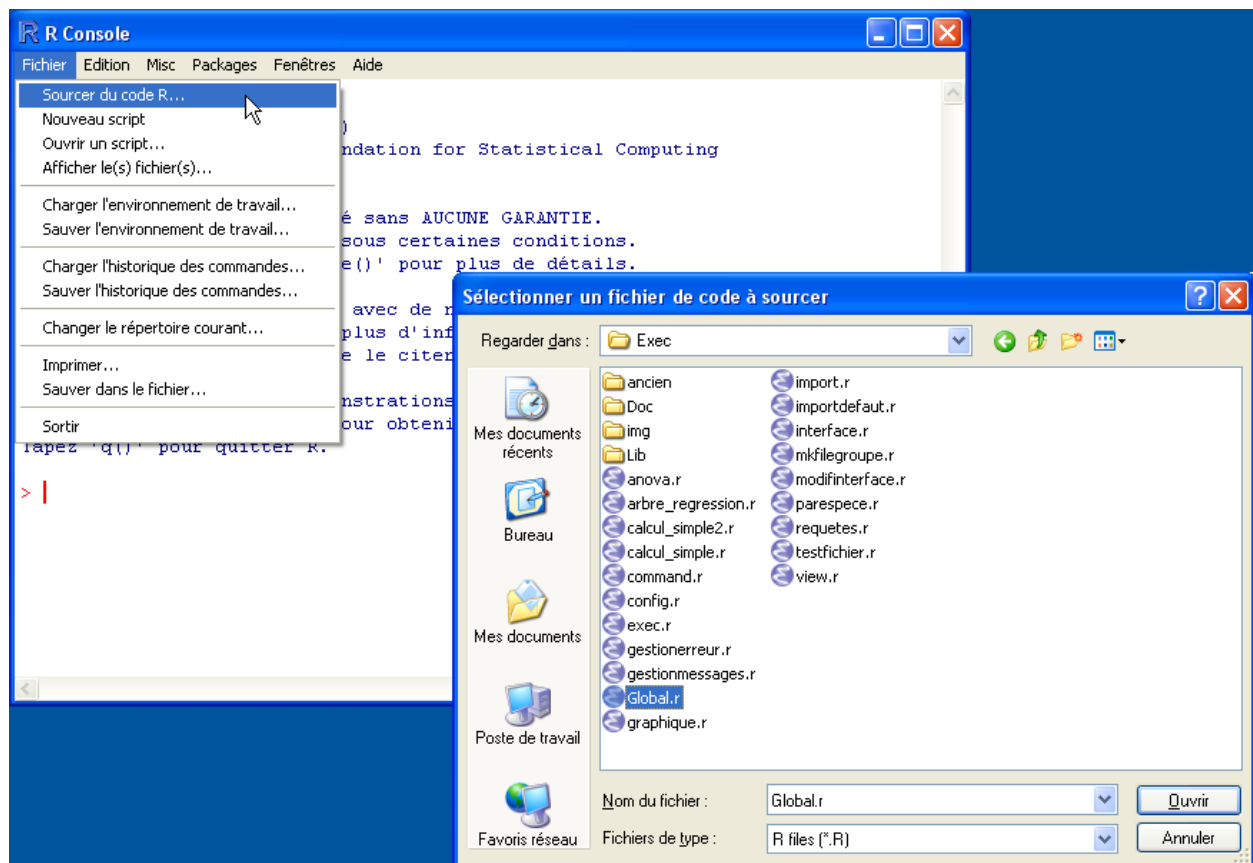
4. Chargement de l'interface

A. Démarrage de l'application

Comme c'est expliqué plus haut, vous disposez sous Windows de raccourcis pour le chargement de l'interface.

Si toutefois ils n'étaient pas fonctionnels sur votre machine, vous pouvez la charger manuellement après avoir lancé R :

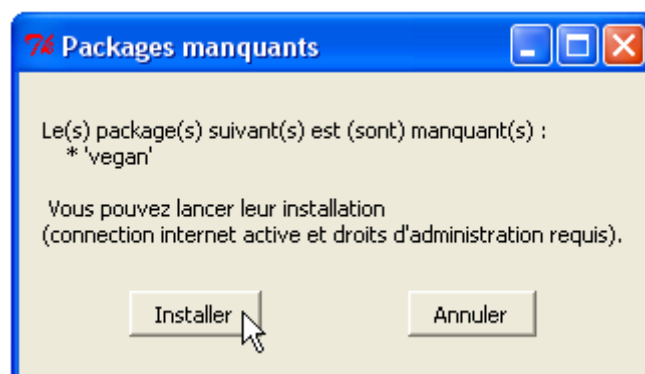
- Soit à l'aide du menu de la console R :
 1. Menu « Fichier » cliquez sur « Sourcer du code R... ».
 2. ...en sélectionnant le fichier « `C:/PAMPA/Exec/Global.r` ».



- soit en collant la ligne suivante dans la console :

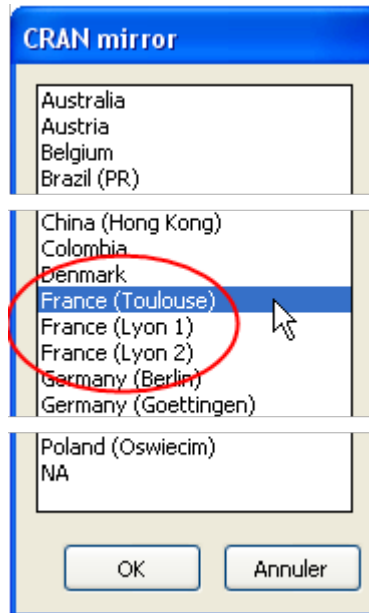
```
source("C:/PAMPA/Exec/Global.r")
```

B. Packages



Le chargement des *packages* requis présente une petite nouveauté : si des *packages* manquent à l'appel, la plateforme propose leur installation :

Si vous choisissez l'installation, R vas probablement vous demander de choisir un dépôt pour le téléchargement des *packages* :



Choisissez un dépôt le plus près de chez vous possible.

Si vous annulez ou si une erreur se produit durant l'installation, la liste des *packages* à installer manuellement est affichée sur la console R :

```
> source("C:\\\\PAMPA\\\\Exec\\\\Global.r")
Erreur dans switch(res, ok = invisible(sapply(requiredPack, library, character.only = TRUE)), :
  Vous devez installer manuellement le(s) package(s) :

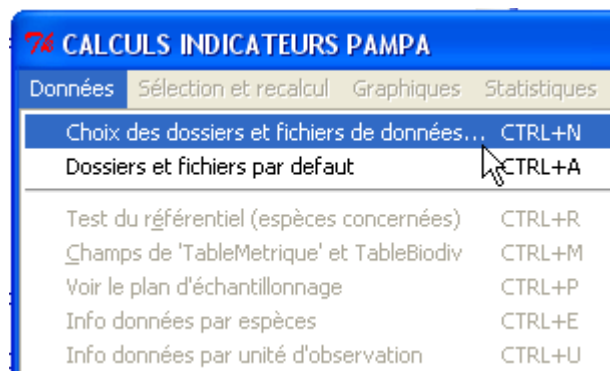
  * 'vegan'
> |
```

Notez que dans ce cas, plutôt que de procéder à une installation manuelle, vous pouvez également relancez la plateforme dans des condition qui permettent l'installation (*i.e.* avec les droits d'administration et une connexion internet).

5. Importation des données

A. Dossiers et fichiers de données

L'entrée « Choix des dossiers et fichiers de données » (anciennement « Nouveau fichier source », très peu explicite) du menu « Données » a été débuggée.



Elle tient compte du répertoire de travail qui peut maintenant être configuré dans le fichier « config.r » :

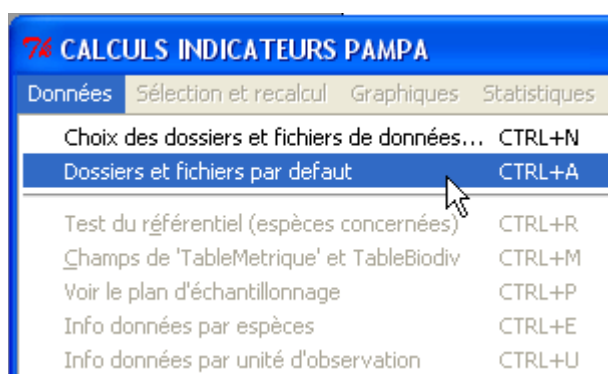


Les modifications faites à l'aide de cette boîte de dialogue ne seront cependant conservées que dans la session courante.

Il est donc conseillé de renseigner les noms de fichiers de données dans le fichier de configuration, comme cela est expliquée dans la section 3.D. Configuration.

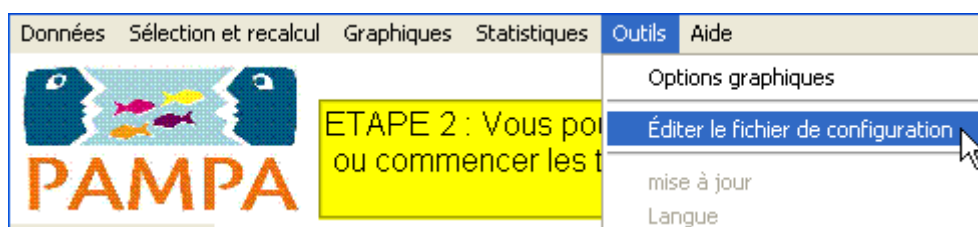
B. Importation des fichiers par défaut

Ainsi, lorsque les fichiers par défaut sont correctement configurés, ils peuvent être directement chargés sans passer par l'étape de choix des fichiers un à un. Il suffit d'utiliser la seconde entrée du menu « Données » ou bien le raccourci « Ctrl+a » :



Lors d'une modification de votre fichier de configuration, vous n'êtes désormais plus obligé de relancer la plateforme pour une prise en compte de la nouvelle configuration. À chaque fois que vous utiliserez ce mode de chargement, le fichier de configuration sera lu à nouveau.

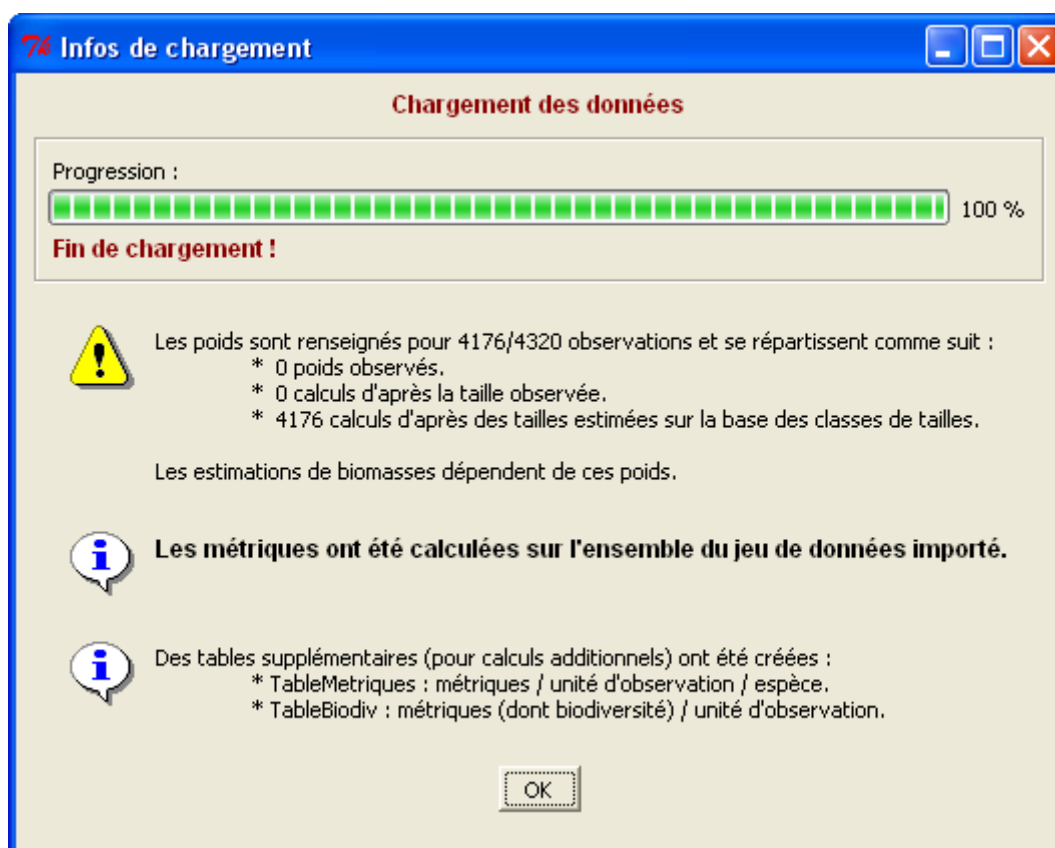
Pour modifier les dossier et fichiers par défaut, vous pouvez directement ouvrir le fichier « C:/PAMPA/Exec/config.r » depuis le menu « outils » de l'interface principale :



Une fois la nouvelle configuration enregistrée, il vous suffira de recharger les « Dossier et fichiers par défaut ».

C. Informations de chargement

Au cours du chargement des données, il n'y a désormais plus qu'une seule fenêtre d'information, qui résume et attire votre attention, au fur et à mesure, sur les éléments importants :



Lorsque le bouton « OK » apparaît, c'est que le chargement est terminé.

La barre de progression ne constitue qu'une information grossière et ne donne pas une estimation précise par rapport au temps total de chargement.

D. Opérations « spéciales » lors du chargement et calculs divers

De nombreux bugs ont été corrigés ainsi que quelques fonctionnalités ajoutées. Parmi lesquels...

i. tailles d'après les classes de tailles

Les tailles ne peuvent pas être systématiquement estimées avec précision lors d'observations *in situ*. En l'absence d'enregistrement des tailles, celles-ci sont maintenant estimées d'après les classes de taille (si celles-ci sont renseignées et s'y prêtent, bien entendu). Les formats de classes pris en compte sont pour l'instant de la forme "5-10", "5_10", "40-", "40_", "_5", etc. Les classes ouvertes vers le bas sont considérées comme "0-<valeur>" et pour celles ouvertes vers le haut, la taille est supposée être la borne inférieure (e.g. "40-" -> 40 ; je ne vois pas de moyen de faire mieux). Pour tout le reste, la taille est estimée comme la moyenne des deux bornes.

ii. Zéros...

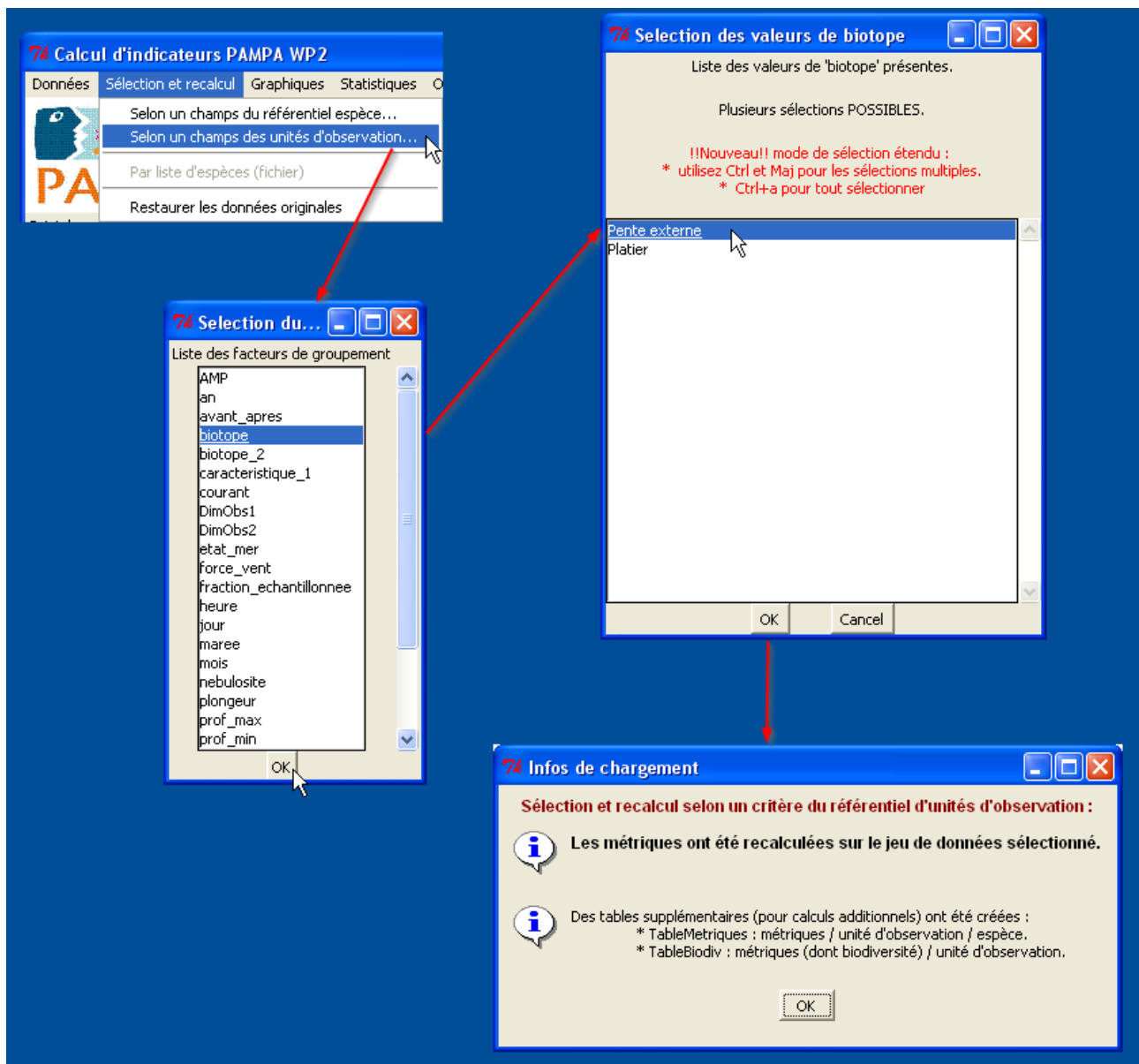
Dans d'anciennes versions, les observations correspondant à zéro individus d'une espèce (*i.e.* cherchée mais non trouvée) étaient supprimées des données, conduisant à des résultats potentiellement faux concernant les métriques / unité d'observation / espèce. Ce comportement inadéquat de la plateforme a été corrigé.

iii. Année de campagne

Certaines campagnes d'acquisition de données peuvent être à cheval sur deux années civiles. Il peut donc dans ce cas être préférable d'utiliser une « année de campagne » à l'année civile comme facteur explicatif de la variabilité des métriques. Il a donc été convenu de placer dans ce cas l'année de campagne dans la colonne « caractéristique_2 » du référentiel des unités d'observations, sous la forme « C<année> » ou bien « c<année> » (*e.g.* "C2004", "c1999",...). Le fait de placer ce type d'informations précisément dans cette colonne assure une bonne prise en compte de leur caractère temporel lors de analyses statistiques (*cf.* 7.C.iv. Comparaisons multiples (2 facteurs)).

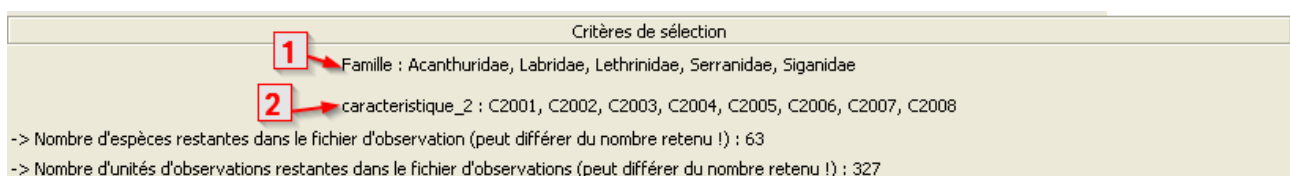
E. Sélection et re-calcul

Les « Sélections et recalcul », accessible par le menu du même nom, vous permettent de réduire les données selon un critère du référentiel espèces et/ou du référentiel d'unités d'observation :

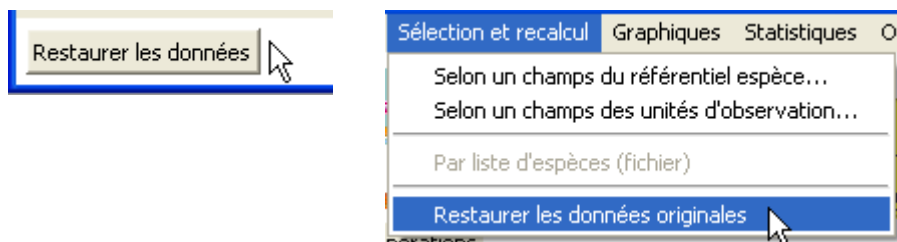


Elles ont été corrigées et fortement optimisées. Seules les calculs réellement nécessaire sont maintenant refaits, tandis que la plupart des tables de métriques sont simplement réduites en fonction des critères (temps de calcul réduit de plusieurs dizaines de fois dans certains cas).

Ces sélections peuvent être imbriquées, c'est à dire que plusieurs critères peuvent être appliqués à la suite. Dans ce cas, la zone d'information sur les critères de sélection (dans l'interface principale) garde maintenant trace de toutes les sélections actives :



Pour restaurer vos données originales sans recharger vos jeux de données, et ainsi gagner du temps, vous pouvez utiliser soit le bouton sur l'interface principale, soit l'entrée dans le menu « Sélection et recalcul » :



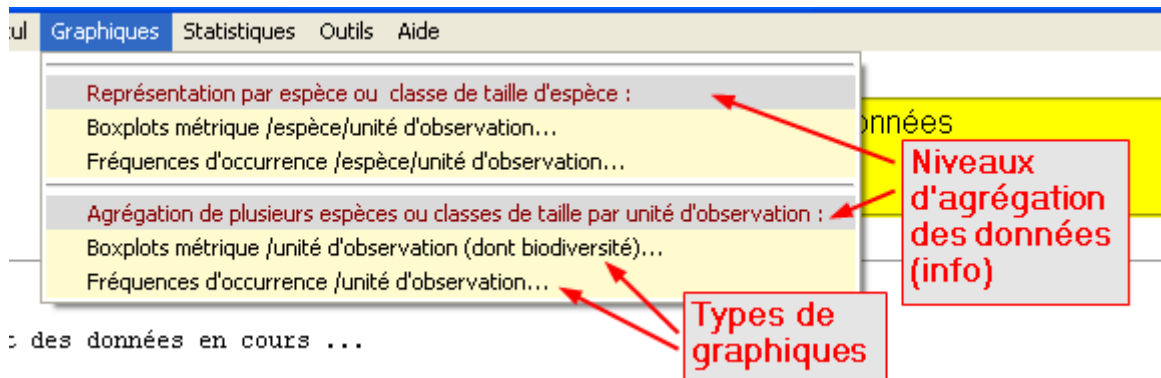
Il n'est pas possible ni prévu de pouvoir annuler juste un niveau de sélection.

Il est cependant envisagé de remplacer, à l'avenir, ce système par un système intégré aux sous-interfaces de production de graphiques et d'analyses. Les sélections seront donc faites juste avant ces opérations, et donc modifiables en permanence.

6. Graphiques

Des interfaces de création de graphiques ont été créées pour faciliter la production en masse de graphiques suivant un modèle prédéfini.

Plusieurs sous menus du menu « Graphiques » donnent accès à différents types de graphiques et niveaux d'agrégation des données :



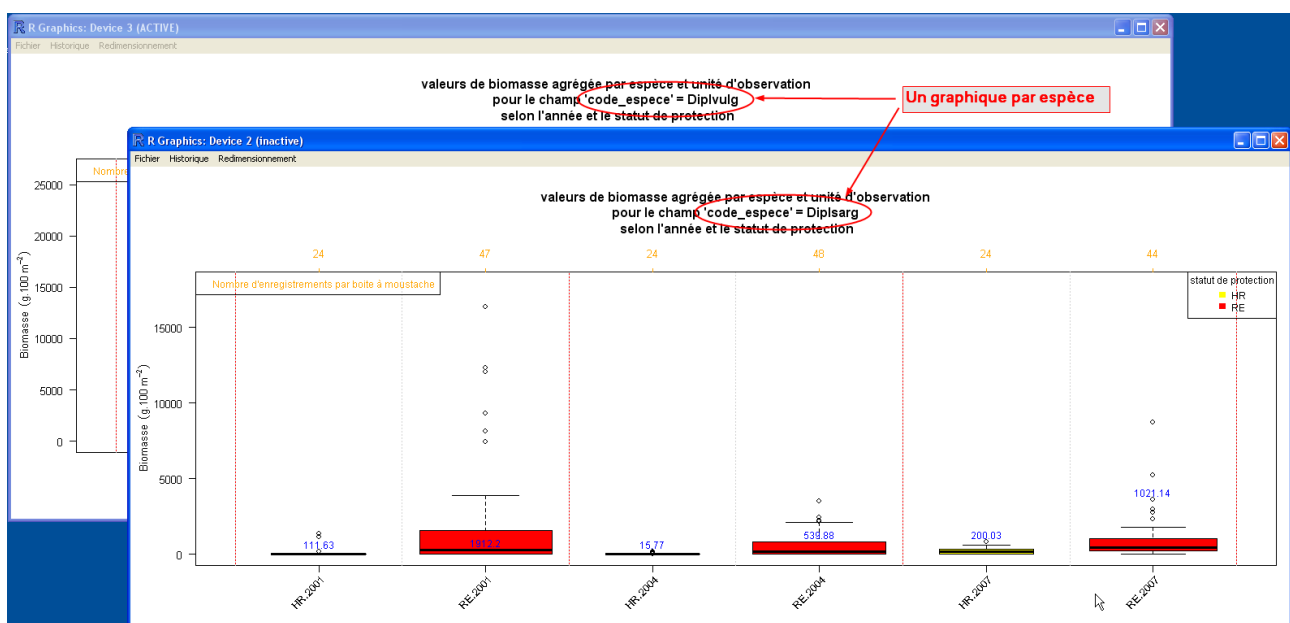
A. Métriques agrégées par unité d'observation par espèce (et éventuellement par classe de taille)

Il n'y a pour l'instant pas d'avertissement lorsque vous utilisez ce niveau d'agrégation à mauvais escient. Ne l'utilisez que pour les graphiques et analyses que vous ne pouvez faire avec le niveau d'agrégation supérieur (par unité d'observation).

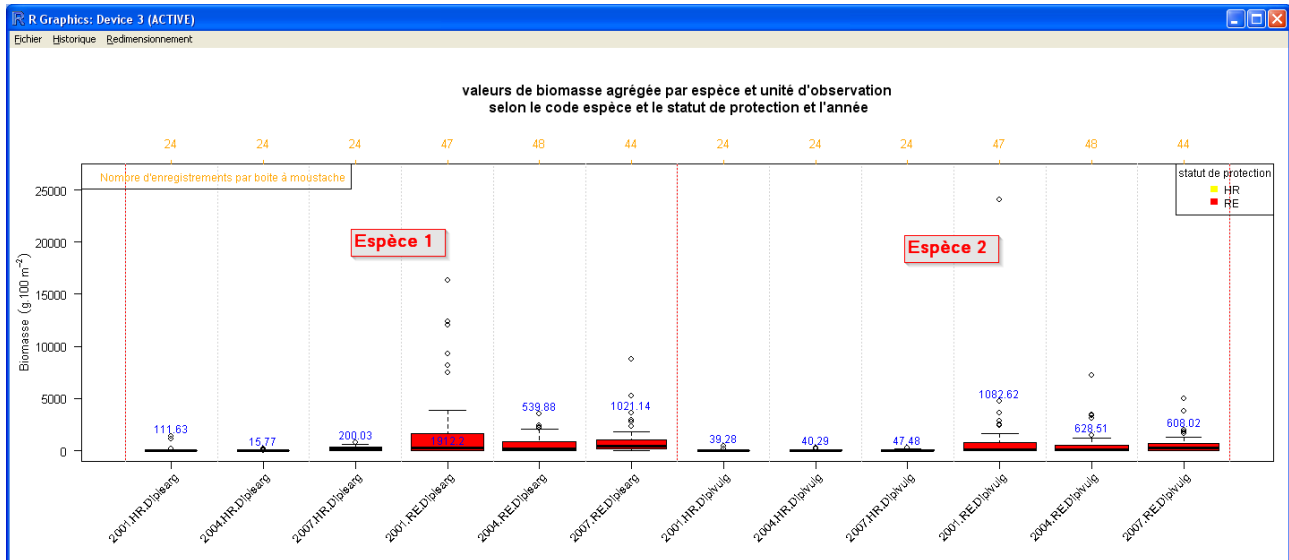
Si vous n'utilisez pas ici les facteurs « code_espece » (ou « espece ») et, le cas échéant, « classe_taille », les résultats seront certainement faux.

Celles-ci sont à utiliser lorsque l'on souhaite :

- représenter une espèce (ou classe de taille d'espèce) par graphique :



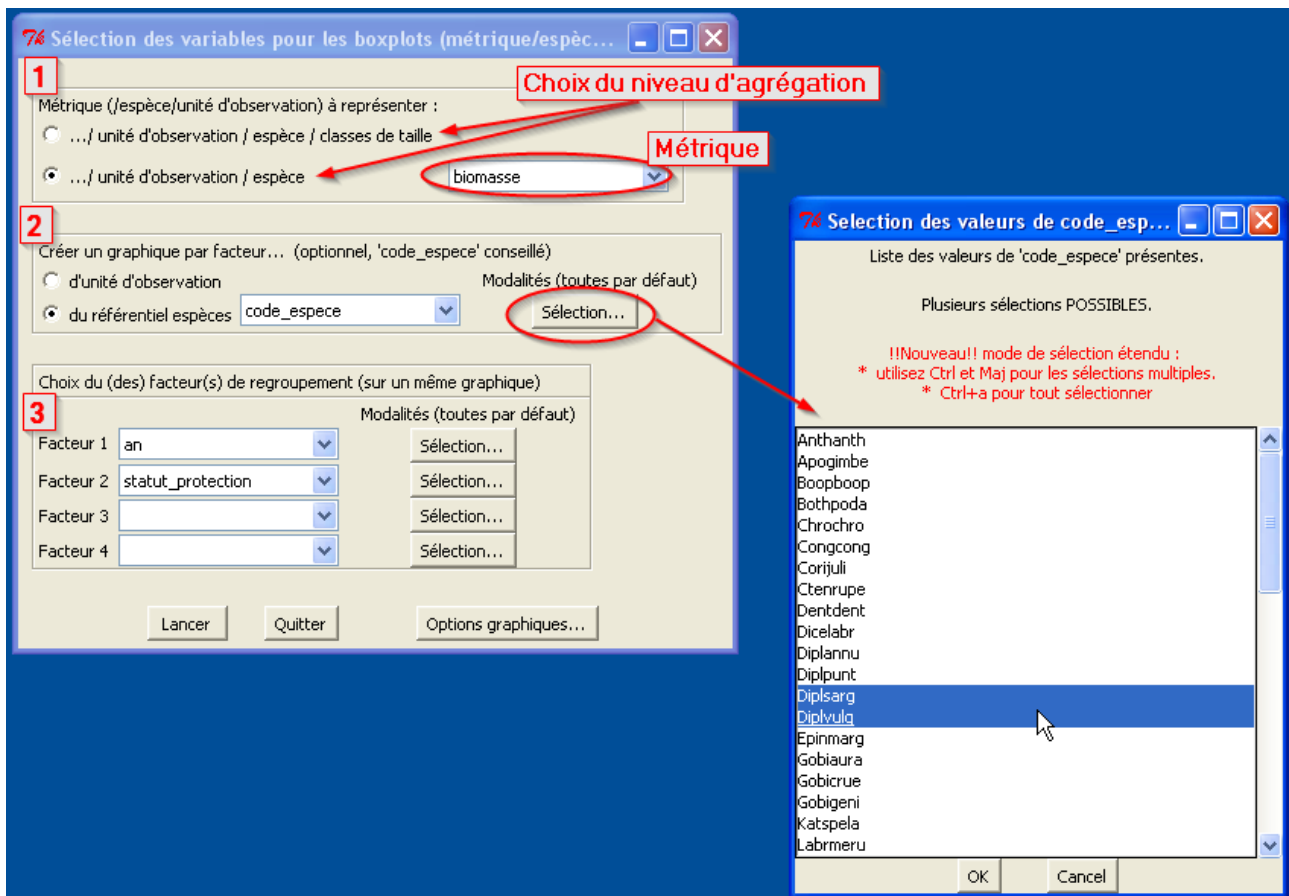
- utiliser l'espèce (plus éventuellement la classe de taille) comme « facteur de regroupement » des boxplots (ou barplots) :



i. Boîtes à moustaches ou Boxplots

Adaptés pour la représentation de la plupart des métriques (comme la biomasse dans les exemples ci-dessus).

L'interface se présente comme suit :



1. cadre de sélection de la métrique et du niveau d'agrégation (espèce ou classe de taille d'espèce). Deux niveaux d'agrégation possibles.
2. cadre de sélection du facteur (**optionnel**) de séparation des graphiques. Un

nouveau graphique est créé pour chaque modalité (suivant éventuellement une sélection) du facteur choisi ici.

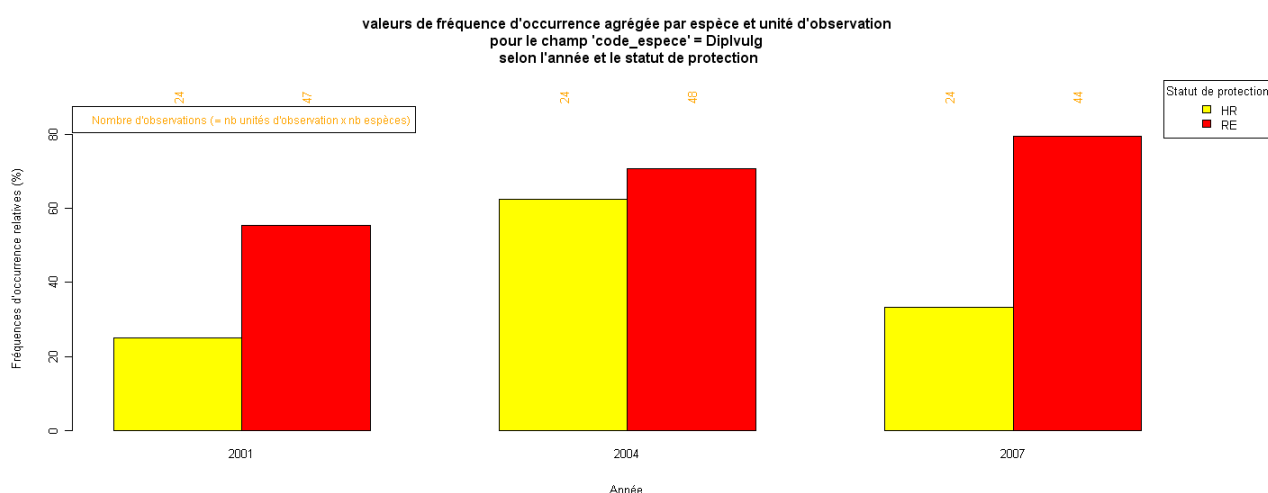
- cadre de sélection du(des) facteur(s) de regroupement et éventuellement sélection des modalités retenues (création d'une boîte à moustache pour chaque combinaison des modalités des facteurs). Un facteur – au minimum – doit être sélectionné ici.

ii. Diagrammes en barres ou *Barplots*

Uniquement utilisés pour représenter les fréquences d'occurrence car les *boxplots* ne sont pas adaptés pour celles-ci.

L'interface est donc similaire à cela près que la métrique ne peut être choisie :

Ce qui donne (ici pour une des espèces sélectionnées) :



iii. Remarques

Classes de tailles :

B. Métriques agrégées par unité d'observation (et éventuellement par classe de taille)

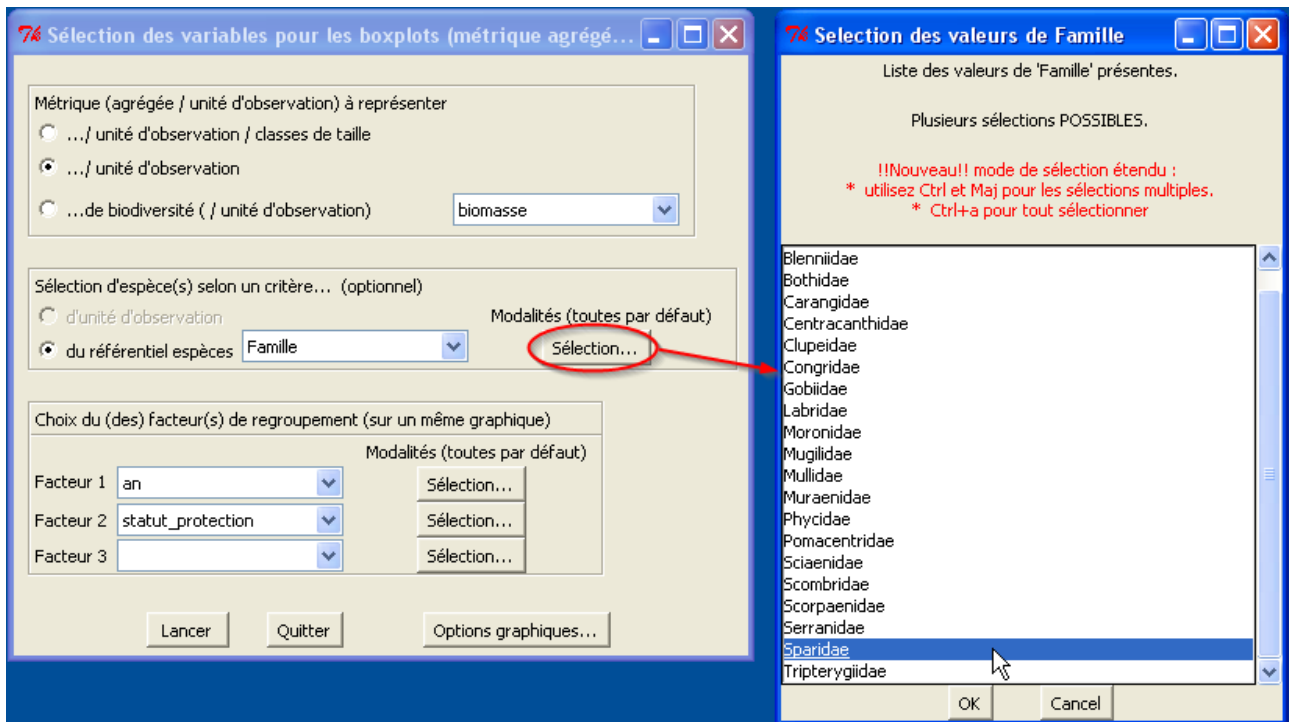
Dans ce cas-ci, la métrique est re-calculée pour chaque unité d'observation, après une éventuelle sélection d'individus selon un critère du référentiel espèces ou bien de taille (classes de tailles P, M et G pour « petits », « moyens » et « grands »).

Dès lors que plusieurs espèces sont amenées à être représentées dans une même boîte à moustache d'un *boxplot* ou la même barre d'un *barplot*, vous devriez avoir recours à ce niveau d'agrégation des métriques.

i. Boîtes à moustaches ou *Boxplots*

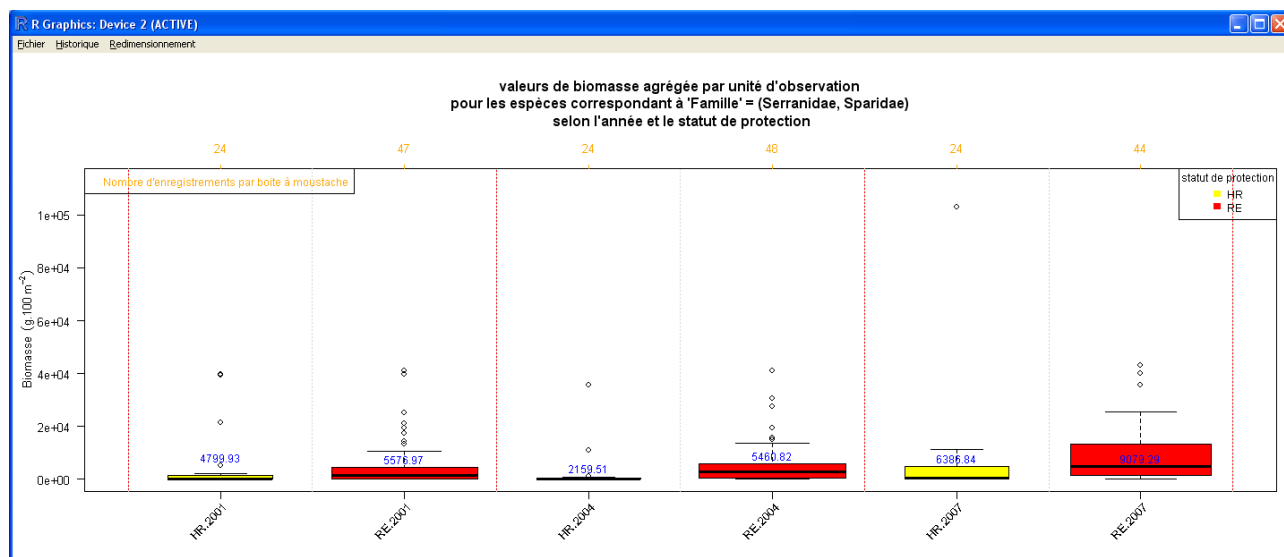
L'interface est proche de celle pour les métriques agrégées par unité d'observation par espèce mais le second cadre n'a pas ici la même fonction. Il sert dans ce cas à faire une sélection des individus à conserver, généralement selon un critère du référentiel espèce (sinon sur un critère de classe de taille). Les données seront ensuite agrégées par unité d'observation pour tous les individus correspondant aux critères. Si ce cadre est laissé vide, les données de l'ensemble des espèces et classes de tailles seront agrégées par unité d'observation et représentées en fonction des critères de regroupement.

Par exemple pour avoir la biomasse d'une famille en fonction de l'année et du statut de protection :



Ceci fonctionne également pour avoir la biomasse de plusieurs familles à la fois (e.g. biomasse agrégée de *Sparidae* et *Serranidae*).

Avec cette sous-interface, un seul graphique est produit à la fois :



Dans le premier cadre, des indices de biodiversités (calculés au niveau de l'unité d'observation) peuvent également être sélectionnés (troisième « bouton radio ») puis choix de la métrique).

Le facteur du second cadre est optionnel : si aucun facteur n'y figure ou aucune modalité n'est sélectionnée, toutes les espèces (ou classes de taille) sont conservées.

ii. Diagrammes en barres ou *barplots*

Le principe des différences avec l'interface pour les métriques agrégées par espèce par unité d'observation (et éventuellement par classe de taille) est le même que dans le cas des *boxplots*. Le second cadre est ici aussi destiné à procéder à une sélection d'individus selon un critère du référentiel espèce ou de classe de taille :

7% Sélection des variables pour les fréquences d'occurrences

Métrique calculée : "fréquence d'occurrence"

Sélection d'espèce(s) selon un critère... (optionnel)

☐ d'unité d'observation

☒ du référentiel espèces

Famille

Modalités (toutes par défaut)

Sélection...

Choix du(des) facteur(s) explicatif(s)/de regroupement

Facteur 1 : an

Facteur 2 : statut_protection

Facteur 3 :

Modalités (toutes par défaut)

Sélection...

Sélection...

Lancer Quitter Options graphiques...

7% Sélection des valeurs de Famille

Liste des valeurs de 'Famille' présentes.

Plusieurs sélections POSSIBLES.

!!Nouveau!! mode de sélection étendu :

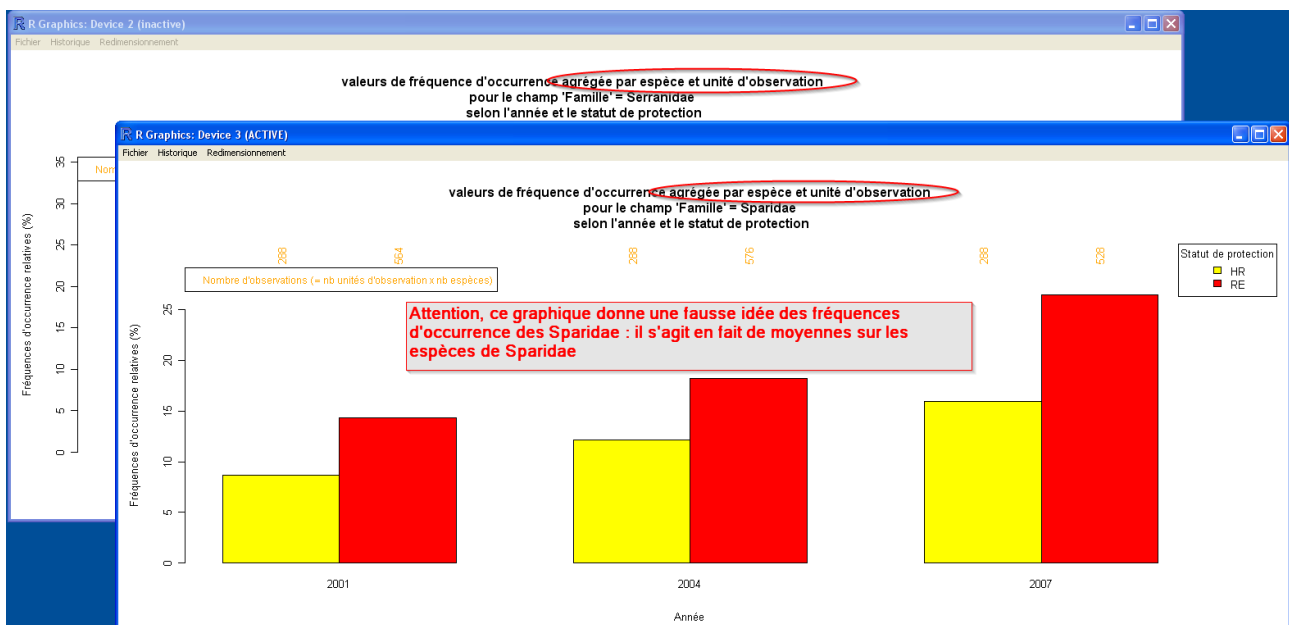
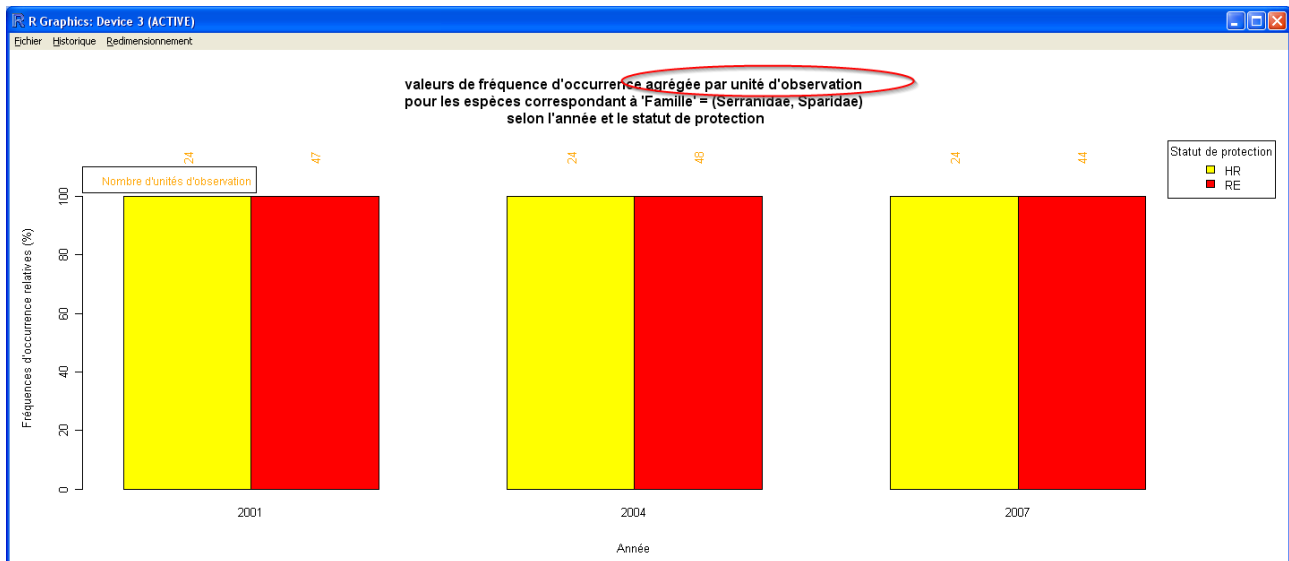
- * utilisez Ctrl et Maj pour les sélections multiples.
- * Ctrl+a pour tout sélectionner

Cichlidae
Clupeidae
Congridae
Gobiidae
Labridae
Moronidae
Mugilidae
Mullidae
Muraenidae
Octopodidae
Palinuridae
Phycidae
Pomacentridae
Sciaenidae
Scombridae
Scorpaenidae
Sepiidae
Serranidae
Sparidae
Tripterygiidae

OK Cancel

Ce qui nous donne un (premier) graphique sans intérêt puisque l'une ou l'autre des deux familles au moins est représentée dans chaque unité d'observation (fréquence

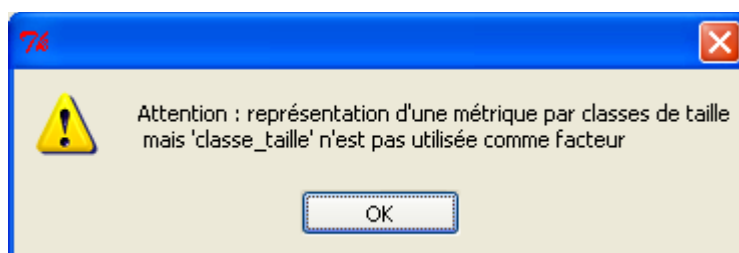
d'occurrence de 100 %), mais nous montre bien que c'est la fréquence d'occurrence de l'ensemble des familles, et non la « moyenne » des espèces qui les composent (seconds graphiques obtenus avec l'interface pour les métriques agrégées par espèce et unité d'observation) qui est représentée :



iii. Remarques

Classes de tailles :

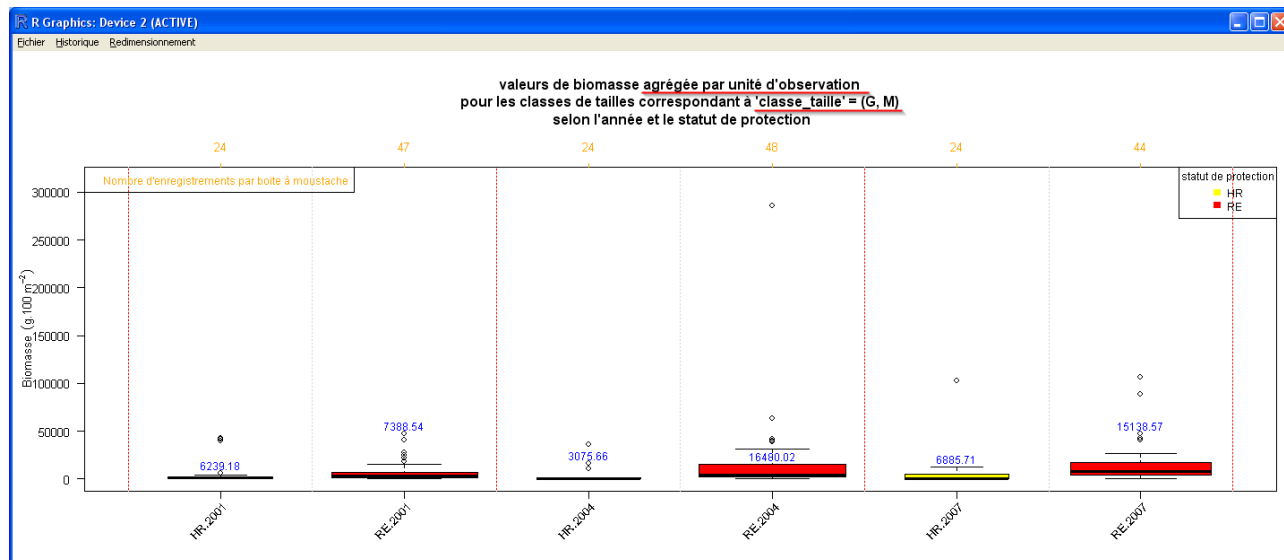
Si vous choisissez dans le premier cadre (*boxplots* uniquement pour l'instant) les métriques « .../unité d'observation/classe de taille » et que vous n'utilisez nulle part le facteur « classe_taille », vous aurez l'avertissement suivant :



Et les résultats (moyenne sur les classes de tailles) ne correspondront certainement pas à ce que vous souhaitez.

Dans le cas contraire, deux cas peuvent être distingués :

1. le facteur « classe_taille » est utilisé dans le second cadre (sélection des individus) et la métrique sera alors agrégée par unité d'observation, après sélection des classes de tailles souhaitées :

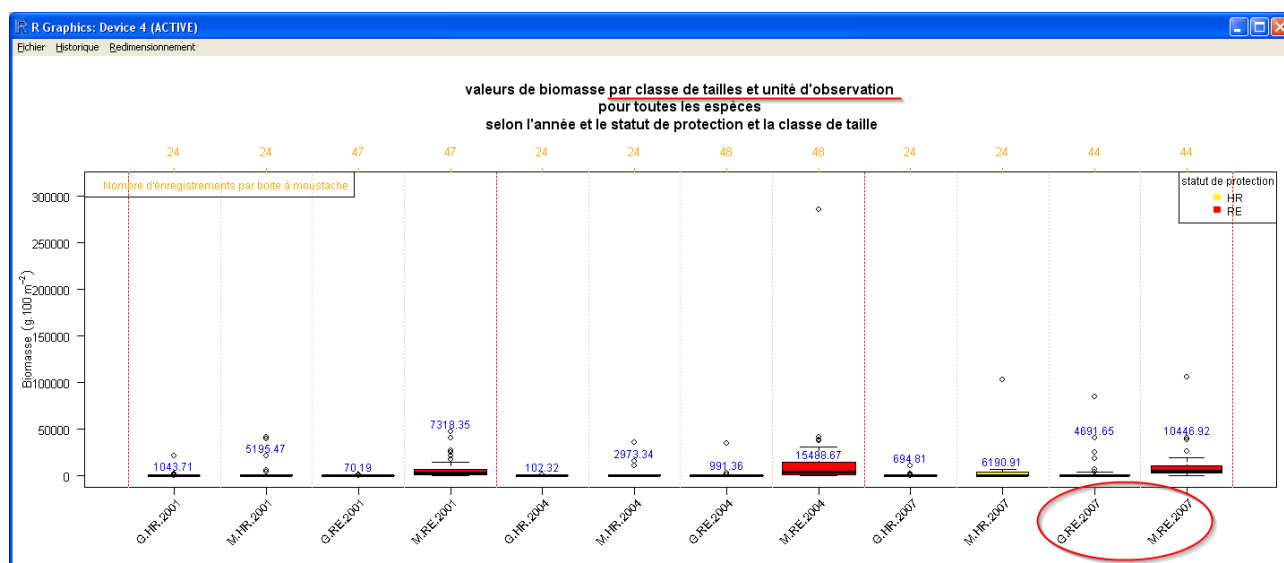


2. le facteur « classe_taille » est utilisé dans le troisième cadre (facteurs de regroupement), la métrique est donc agrégée par unité d'observation par classe de taille et boîtes à moustaches séparées par classe de taille :

Choix du (des) facteur(s) de regroupement (sur un même graphique)

Modalités (toutes par défaut)

Facteur 1	an	Sélection...
Facteur 2	statut_protection	Sélection...
Facteur 3	classe_taille	Sélection...
Facteur 4		Sélection...



(Notez qu'en faisant la somme des moyennes – en bleu – pour les groupes cerclés de rouge, on retrouve bien les 15138,57 g.100m⁻² du précédent graphique pour les zones en

réserve et l'année 2007.)

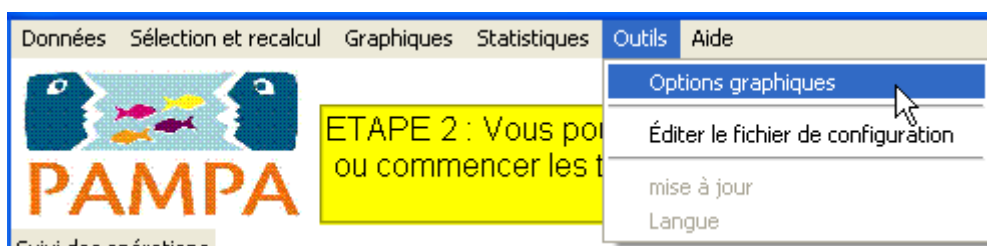
C. Remarques générales sur les graphiques

i. Rang d'utilisation du facteur « statut de protection » :

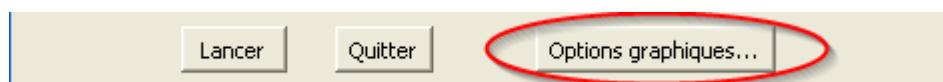
Que ce soit pour les *boxplots* ou les *barplots*, il est recommandé de toujours mettre le statut de protection comme deuxième facteur de regroupement (troisième cadre) afin que les codes couleurs correspondent à des niveaux de protection. Ceci n'est bien évidemment pas obligatoire, mais permet de mettre en lumière l'effet de la protection.

ii. Options graphiques

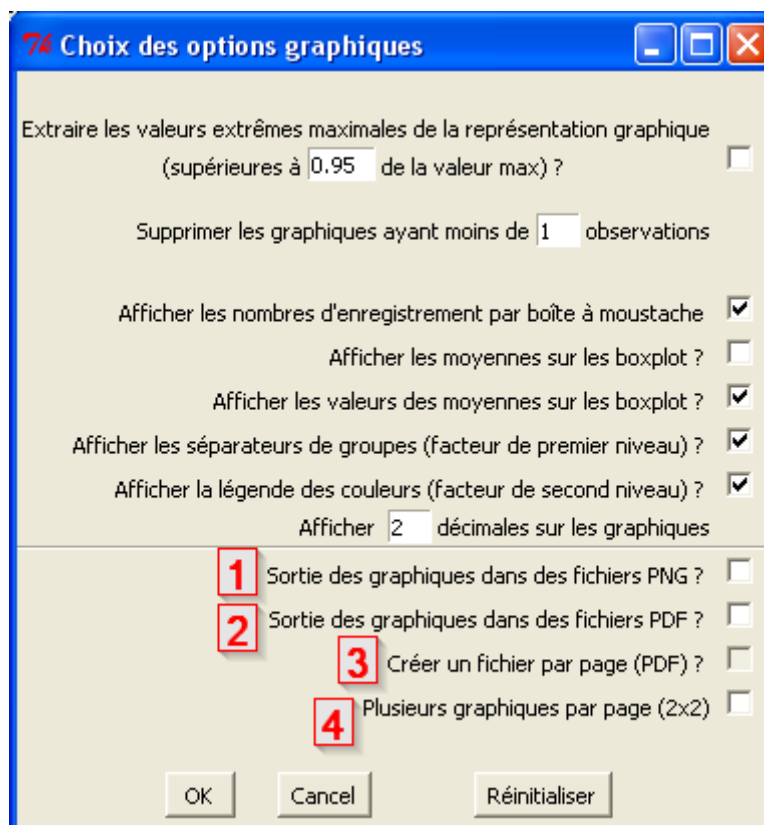
Vous pouvez modifier les options graphiques à partir, soit de l'interface principale, soit des sous-interfaces de création de graphiques ; respectivement :



et



qui ouvrent une sous-interface de gestion des options graphiques (ici avec les options par défaut) :



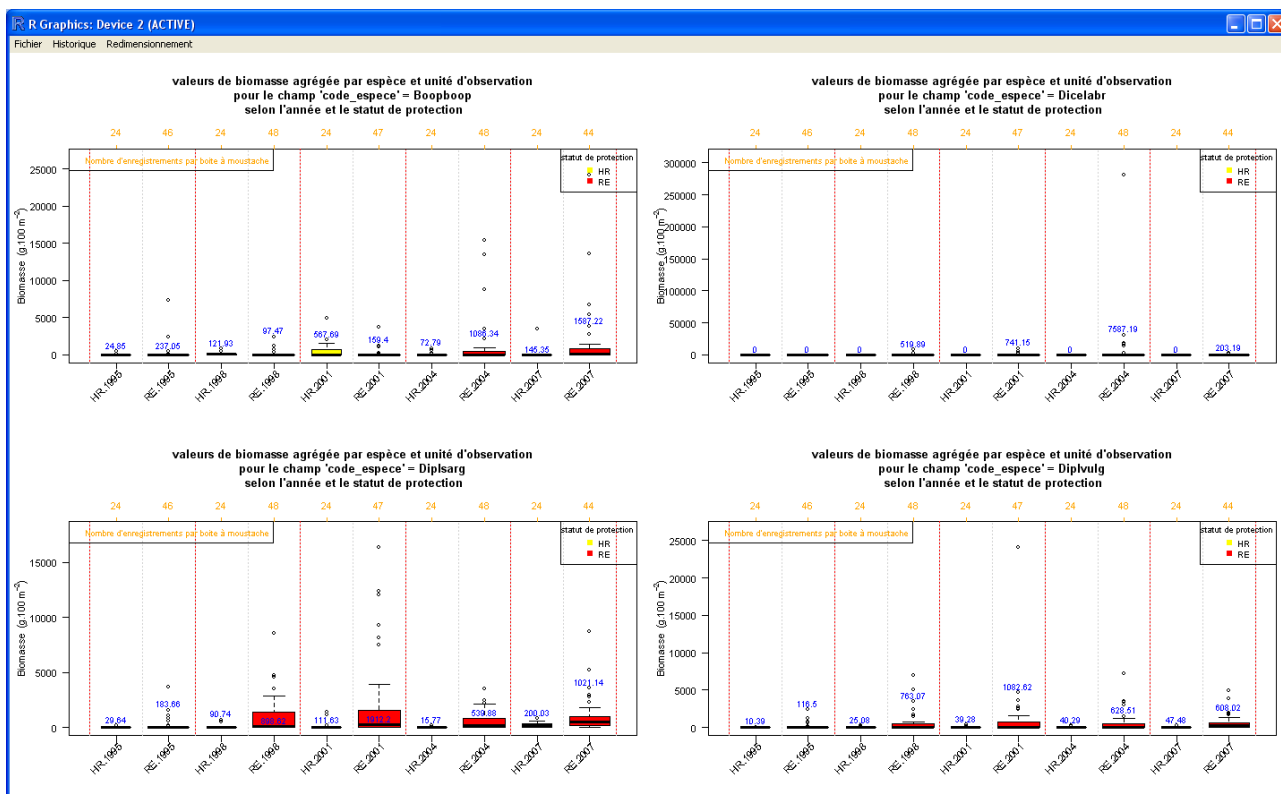
La plupart sont assez parlantes par elles-mêmes. Les option numérotées en rouge – sous le séparateur – concernent les périphériques graphiques :

1. Le graphique n'est pas affiché mais un fichier au format **PNG** (insérable dans tout type de document, ou presque) est créé dans le dossier « <dossier de travail>/FichiersSortie » avec un nom aussi explicite que possible, contenant :
 - le type de graphique.
 - la métrique.
 - le niveau d'agrégation.
 - d'éventuels facteurs de séparation des graphiques/sélections d'individus + les modalités sélectionnées.
 - le(s) facteur(s) de regroupement.

Par exemple :

```
boxplot_biomasse_Agr-  
CL+unitobs_Famille(Lethrinidae)_statut_protection-classe_taille.png
```

2. Le graphique n'est pas affiché mais un fichier au format **PDF** est créé dans le dossier « <dossier de travail>/FichiersSortie ». Option mutuellement exclusive avec l'option 1.
3. Option active uniquement si 2. est active, et utile en cas de création de graphiques par lots (métriques agrégées par espèce...). Force les graphiques PDF d'un même lot à être créés dans des fichiers séparés. Si inactive, ils sont créés dans des pages séparées d'un même fichier.
4. Pour tout type de périphérique graphique. Permet de placer jusqu'à quatre graphiques sur la même page/périphérique graphique. De même que l'option 3., celle-ci n'est utilisée que pour le traitement de graphiques par lots (*ie.* pour les métriques agrégées par espèce, séparés par « code_espece »).



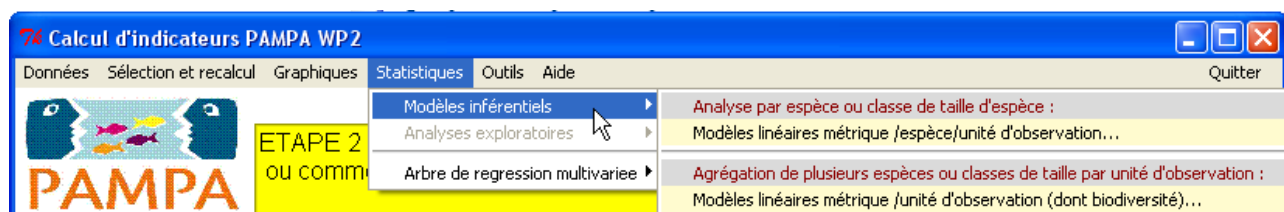
Les choix d'option ne seront effectifs qu'après avoir cliqué sur le bouton « OK ». Le bouton « Réinitialiser » permet de retrouver les options par défaut.

Les options graphiques personnalisées sont persistantes pour la sessions en cours, c'est-à-dire qu'elles resteront les mêmes après fermeture puis ouverture d'une quelconque sous-interface, « sélection et recalcul », rechargement de données, etc.

Elles sont en revanche réinitialisées aux valeurs par défaut à chaque chargement de l'interface principale.

7. Analyses statistiques

Sous le menu « Statistiques > Modèles inférentiels » se trouvent de nouvelles fonctionnalités qui ont notamment vocation à remplacer l'ancien système d'ANOVAs :



Les boîtes de dialogue de choix des métriques/indicateurs et facteurs sont très similaires à celles développées pour les graphiques. Et ce sont d'ailleurs les mêmes critères qui doivent mener au choix de métriques agrégées par espèce par unité d'observation ou par unité d'observation uniquement (plus éventuellement par classe de taille). Reportez vous aux explications concernant les choix de niveaux d'agrégation pour les graphiques.

A. Métriques agrégées par unité d'observation par espèce (et éventuellement par classe de taille)

Le principe de fonctionnement de l'interface est le même que pour les graphiques avec le même niveau d'agrégation des métriques à cela près que les options graphiques n'y sont pas disponibles :

Sélection des variables pour les modèles linéaires (métriq...)

Métrique expliquée :

☐ .../ unité d'observation / espèce / classes de taille

☒ .../ unité d'observation / espèce biomasse

Séparer les analyses par facteur... (optionnel)

☐ d'unité d'observation

☒ du référentiel espèces code_espece Sélection...

Modalités (toutes par défaut)

Choix du(des) facteur(s) explicatif(s)

Facteur 1 an Sélection...

Facteur 2 statut_protection Sélection...

Facteur 3 Sélection...

Modalités (toutes par défaut)

Lancer Quitter Options graphiques...

En cliquant sur « Lancer », vous lancez un module de choix de la distribution théoriques des données de la métrique choisie, qui va vous orienter vers le type d'analyse adéquat. Pour des données continues, les choix sont entre Anova (modèle linéaire), Anova sur données log-transformées (modèle linéaire sur données log-transformées) et Modèle Linéaire Généralisé (GLM) de la famille Gamma. Les sorties de ces trois types de modèles se présentent sous des formes similaires.

Remarque : dans le cas des présences/absences

Métrique expliquée :

☐ .../ unité d'observation / espèce / classes de taille

☒ .../ unité d'observation / espèce

pres_abs

biomasse

densite

nombre

poids

poids_moyen

pres_abs

taille_moy

– qui permettent de conduire des analyses correspondant aux fréquences d'occurrence pour la partie graphique – vous n'aurez pas de choix de modèle à faire car la distribution binomiale est automatiquement retenue.

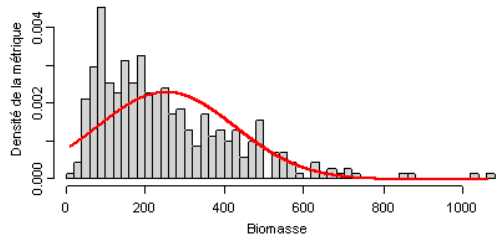
Choix de distribution théorique de la métrique 'biomasse'

INFO :

Cette fenêtre vous permet de choisir la distribution la plus adaptée pour faire vos analyses.

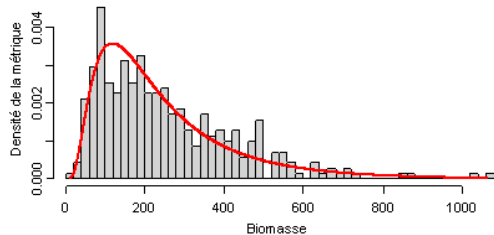
La distribution (courbe rouge) s'ajustant le mieux à vos données (histogramme) d'après le critère d'information de Akaike (AIC ; doit être le plus petit possible) est pré-sélectionnée.

Comparaison avec la loi Normale



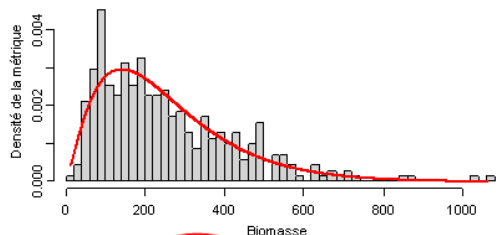
☐ loi Normale (AIC=4642). **Modèle : ANOVA**

Comparaison avec la loi log-Normale



☐ loi log-Normale (AIC=4515). **Modèle : ANOVA, données log-transformées**

Comparaison avec la loi Gamma



☒ loi Gamma (AIC=4508). **Modèle : GLM, famille 'Gamma'**

OK

Annuler

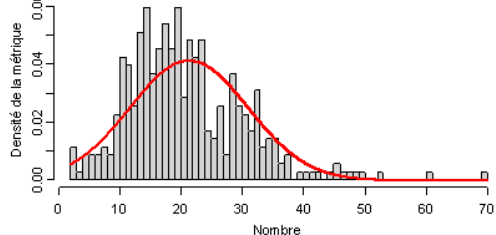
Choix de distribution théorique de la métrique 'nombre'

INFO :

Cette fenêtre vous permet de choisir la distribution la plus adaptée pour faire vos analyses.

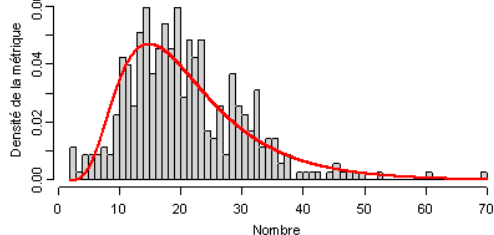
La distribution (courbe rouge) s'ajustant le mieux à vos données (histogramme) d'après le critère d'information de Akaike (AIC ; doit être le plus petit possible) est pré-sélectionnée.

Comparaison avec la loi Normale



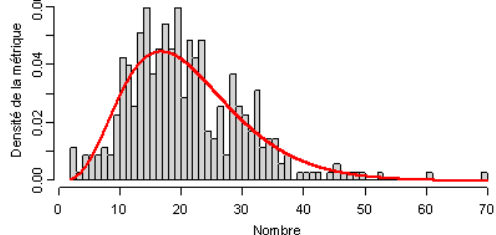
☐ loi Normale (AIC=2605). **Modèle : ANOVA**

Comparaison avec la loi log-Normale



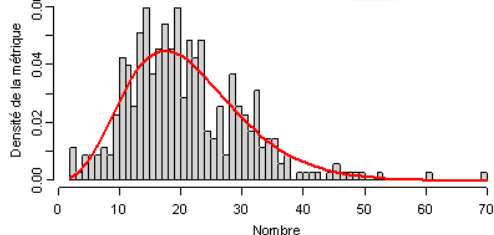
☐ loi log-Normale (AIC=2602). **Modèle : ANOVA, données log-transformées**

Comparaison avec la loi Gamma



☐ loi Gamma (AIC=2566). **Modèle : GLM, famille 'Gamma'**

Comparaison avec la loi Binomiale négative



☒ loi Binomiale négative (AIC=2561). **Modèle : GLM, famille 'Binomiale négative'**

OK

Annuler

La distribution qui s'ajuste le mieux à vos données selon le critère d'information de Akaike (AIC) – qui doit être le plus petit possible – est déjà présélectionnée dans l'interface ci-après. Pour en changer, il vous suffit de cliquer sur le graphique ou le bouton correspondant à votre choix.

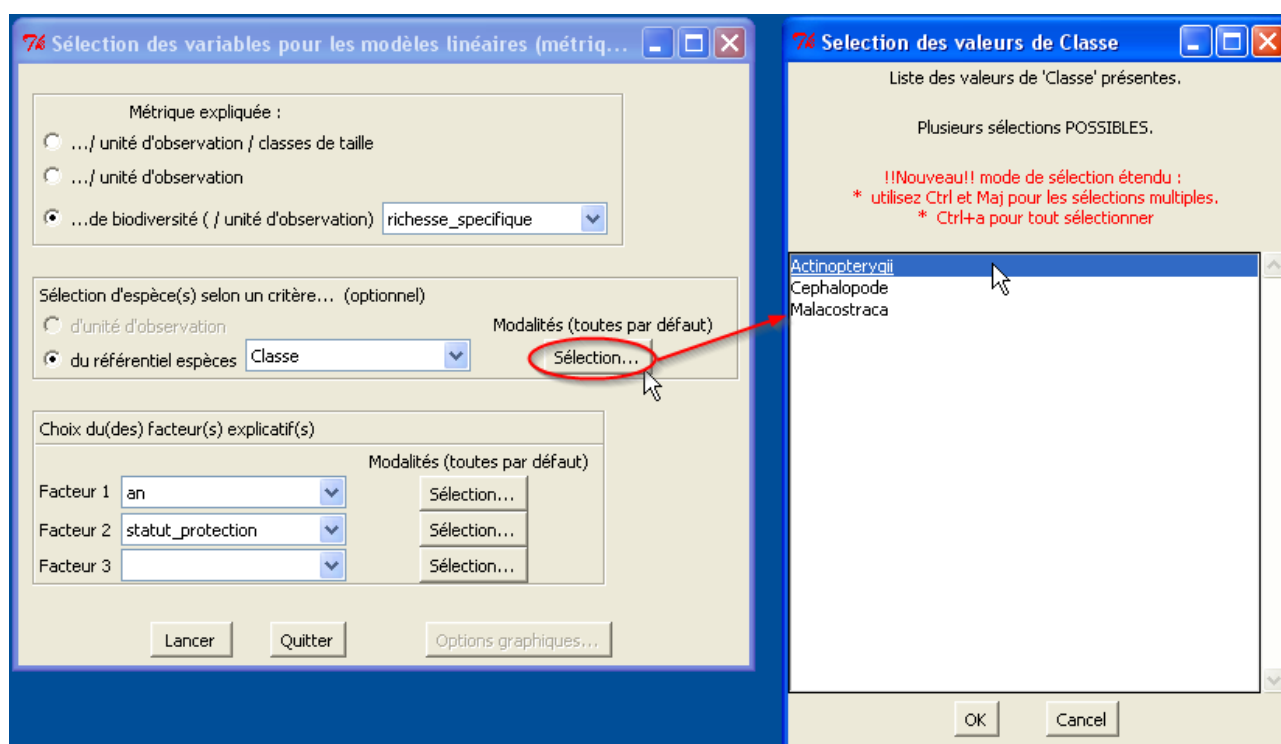
En cliquant sur « OK » vous lancez l'analyse correspondant à la distribution choisie.

Selon la nature de la métrique sélectionnée, le nombre de choix peut-être variable. Pour les données de comptage (entières, eg. « nombre »), la loi binomiale négative est en plus disponible.

Dans les cas où les lois log-normale et binomiale négative sont toutes deux disponibles et donnent des ajustement de qualités relativement similaire, il est préférable de sélectionner la seconde.

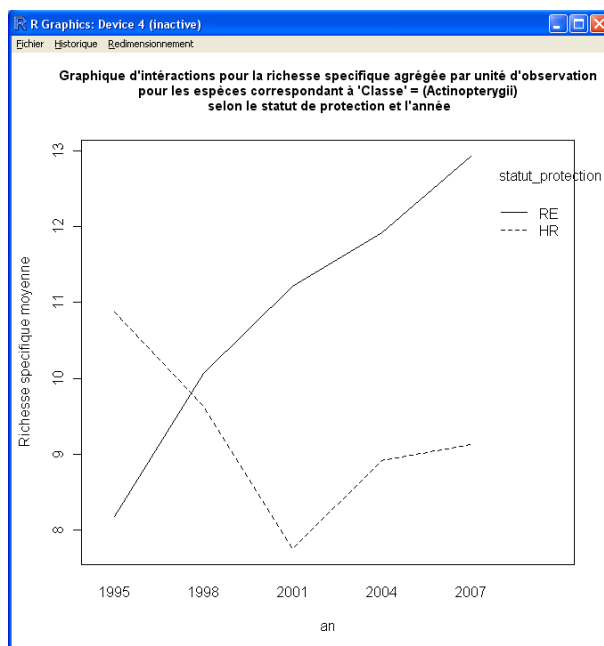
B. Métriques agrégées par unité d'observation (et éventuellement par classe de taille)

Ici aussi, l'interface ressemble à celle des *boxplots* pour les mêmes niveaux d'agrégation et le principe reste le même. Par exemple pour analyser les effets année et statut de protection sur la richesse spécifique de **poissons** :



C. Résultats

Si deux facteurs sont sélectionnés, un *interaction plot* est produit, qui peut servir de support visuel à l'interprétation des comparaisons multiples :



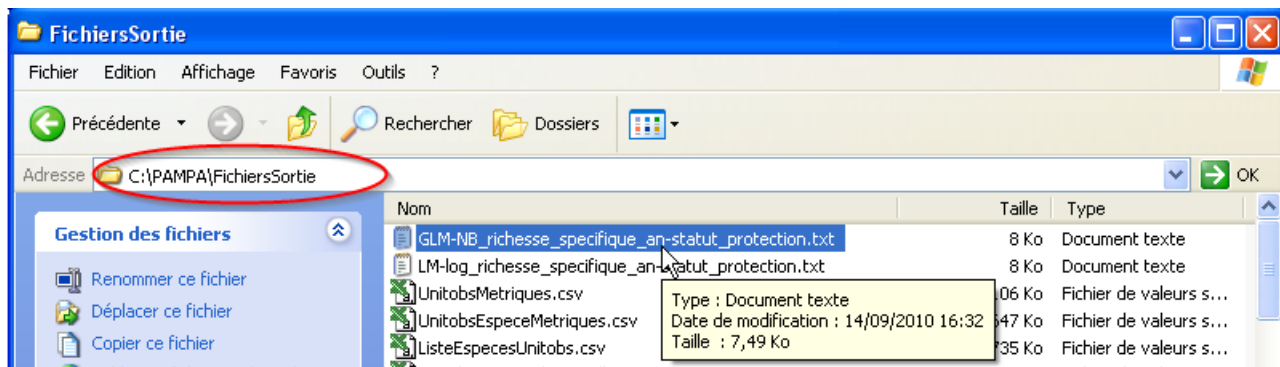
Les résultats de l'analyse sont stockés dans le dossier C:/PAMPA/FichiersSortie/ dans un fichier texte de la forme

<préfixe>_<métrique>_[<facteur de séparation>(<modalité>)]<facteur 1>[-<facteur 2>...].txt

(les parties entre [] sont optionnelles).

Avec comme préfixe :

- LM pour le modèle linéaire simple (ou ANOVA),
- LM-log pour le modèle linéaire sur données log-transformées,
- GLM-P pour le glm avec la distribution de Poisson,
- GLM-NB pour le glm avec la distribution binomiale négative.
- GLM-Ga pour le glm avec la distribution Gamma.



Ce fichier contient plusieurs parties détaillées dans les sous-sections suivantes.

i. Informations sur le modèle

Modèle ajusté :

```
glm.nb(formula = richesse_specifique ~ an * statut_protection,  
data = tmpDataMod, init.theta = 280.5474544, link = log)
```

(ici un glm avec la distributions binomiale négative sur la richesse spécifique, avec l'année et le statut de protection comme facteurs).

Dans le cas d'un modèle linéaire, les statistiques globales du modèle sont également affichées dans cette partie.

ii. détails sur les facteurs significatifs et leurs coefficients

Table d'analyse de la déviance :

Modèle : Binomiale négative(280.5475), lien : log

Réponse : richesse_specifique

Termes ajoutés séquentiellement (premier au dernier)

	Df	Deviance	Resid. Df	Resid. Dev	P(> Chi)	
NULL			374	1499.16		
an	9	1044.86	365	454.30	< 2.2e-16	***
statut_protection	2	17.98	363	436.33	0.0001249	***
an:statut_protection	18	88.36	345	347.97	2.839e-11	***

Significativités des paramètres

(seuls ceux correspondant à des facteurs/intéractions significatifs sont représentés) :

	z value	Pr(> z)	
(Intercept)	10.0449	< 2.2e-16	***
an2000	5.1418	2.721e-07	***
an2001	0.9085	0.3635972	
an2002	2.7556	0.0058577	**
an2003	3.7057	0.0002108	***
an2004	3.4696	0.0005213	***
an2005	3.4386	0.0005848	***
an2006	3.0279	0.0024624	**
an2007	3.1885	0.0014303	**
an2008	7.3832	1.545e-13	***
statut_protectionPP	6.2847	3.286e-10	***
statut_protectionRE	5.6507	1.598e-08	***
an2000:statut_protectionPP	-4.8547	1.205e-06	***

```
an2001:statut_protectionPP -4.8689 1.122e-06 ***
...
```

iii. Valeurs prédites

Valeurs prédites par le modèle :

```

1999:HR  1999:PP  1999:RE  2000:HR  2000:PP  2000:RE  2001:HR  ...
8.00000  32.00000  26.73333  24.36364  30.50000  36.35294  9.80000  ...
2003:PP  2003:RE  2004:HR  2004:PP  2004:RE  2005:HR  2005:PP  ...
14.41667  17.76190  17.33333  19.50000  17.16667  17.00000  15.88889  ...
2007:RE  2008:HR  2008:PP  2008:RE
15.27778  38.20000  33.55556  39.44444

```

Valeur prédite par le modèle pour chaque combinaisons des modalités des facteurs de l'analyse. À part pour les ANOVAs sur données log-transformées (Log-LM) – pour lesquelles elles sont données dans l'échelle logarithmique – ces valeurs prédites le sont dans l'échelle d'observation des données (*i.e.* non-transformées).

iv. Comparaisons multiples (2 facteurs)

Lorsque deux facteurs sont sélectionnés :

Comparaisons multiples :

Attention : les estimations de différences sont sur les logarithmes :
 $(\log(A) - \log(B))$

Comparaisons pour les différences spatiales (statut de protection) par année :

Hypothèses linéaires :

	Estimate	Std. Error	z value	Pr(> z)
1999 : PP - HR == 0	1.386294	0.220584	6.285	<1e-04 ***
2000 : PP - HR == 0	0.224635	0.092732	2.422	0.336
2001 : PP - HR == 0	0.135514	0.131672	1.029	1.000
2002 : PP - HR == 0	0.145634	0.108530	1.342	0.990
2003 : PP - HR == 0	-0.234257	0.112120	-2.089	0.613
...				

Comparaisons pour les différences temporelles par statut de protection :

Hypothèses linéaires :

	Estimate	Std. Error	z value	Pr(> z)
--	----------	------------	---------	----------


```

HR : 2008 - 1999 == 0  1.56339  0.21175  7.383  < 0.001 ***
HR : 2008 - 2007 == 0  0.87025  0.07992  10.889  < 0.001 ***
HR : 2007 - 2006 == 0  0.03390  0.09462  0.358  1.00000
HR : 2006 - 2005 == 0 -0.09453  0.09873  -0.957  0.99994
HR : 2005 - 2004 == 0 -0.01942  0.10957  -0.177  1.00000
...

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(P-valeurs ajustées -- méthode 'single-step')

```

Ces résultats indiquent par exemple une richesse spécifique supérieure en protection partielle (PP) par rapport à l'extérieur de la réserve (HR) en 1999. Et une augmentation significative hors réserve sur la période 1999-2008.

Comme vous pouvez le constater, les facteurs temporels sont traités d'une façon différente de tous les autres. Les comparaisons temporelles ne sont faites qu'entre le début et la fin de la série (e.g. 2008 - 1999) et entre deux « pas de temps » successifs (e.g. 2007 - 2006, 2006 - 2005 mais pas 2007 - 2005).

Sont considérés comme facteurs temporels :

- le champs « an »
- le champs « caractéristique_2 » uniquement s'il a le format suivant :
 "C<année sur quatre chiffre>" (e.g. "C1998")
 Ceci permet de renseigner la campagne – lorsqu'elle est à cheval sur deux années civiles – correspondant à l'unité d'observation.

Lorsque le modèle est un GLM ou une ANOVA sur données log-transformées, un avertissement est affiché au début des comparaisons multiples, qui précise dans quel espace de transformation sont estimées les différences. Dans le cas des GLMs de la famille Gamma, la fonction de lien est l'inverse :

```

Attention : les différences sont estimées dans la fonction de lien
(inverse) :
(1/A) - (1/B)  =>  *inversion du signe des différences*

```

Dans ce cas (et uniquement celui-là), le signe des différences est changé par rapport aux différences dans l'espace des observations. Par exemple :

```

Hypothèses linéaires :
      Estimate Std. Error z value Pr(>|z|)
RE - HR == 0  -0.2864      0.1186  -2.415   0.0157 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(P-valeurs ajustées -- méthode 'single-step')

```

signifie que la métrique est significativement plus élevée en réserve (RE) que hors réserve (HR).

v. Comparaisons multiples (1 facteur)

Lorsqu'un seul facteur est sélectionné, des comparaisons sont faites

- entre modalités successives et entre état final et initial si le facteur est temporel (voire encadré ci-dessus).
- entre toutes les paires de modalités sinon.

Comparaisons des modalités :

Attention : les estimations de différences sont sur les logarithmes :

$(\log(A) - \log(B))$

Facteur 'caracteristique 2' (temporel) :

Hypothèses linéaires :

	Estimate	Std. Error	z value	Pr(> z)	
C2008 - C1999 == 0	0.38385	0.05323	7.212	<0.001	***
C2008 - C2007 == 0	0.90064	0.05105	17.641	<0.001	***
C2007 - C2006 == 0	-0.01081	0.05926	-0.182	1.000	
C2006 - C2005 == 0	-0.07685	0.05815	-1.322	0.810	
C2005 - C2004 == 0	-0.08583	0.06832	-1.256	0.847	
C2004 - C2003 == 0	0.14008	0.06733	2.081	0.280	
C2003 - C2002 == 0	0.04923	0.05376	0.916	0.971	
C2002 - C2001 == 0	0.07912	0.06025	1.313	0.815	
C2001 - C2000 == 0	-0.99037	0.06098	-16.241	<0.001	***
C2000 - C1999 == 0	0.37864	0.05891	6.427	<0.001	***

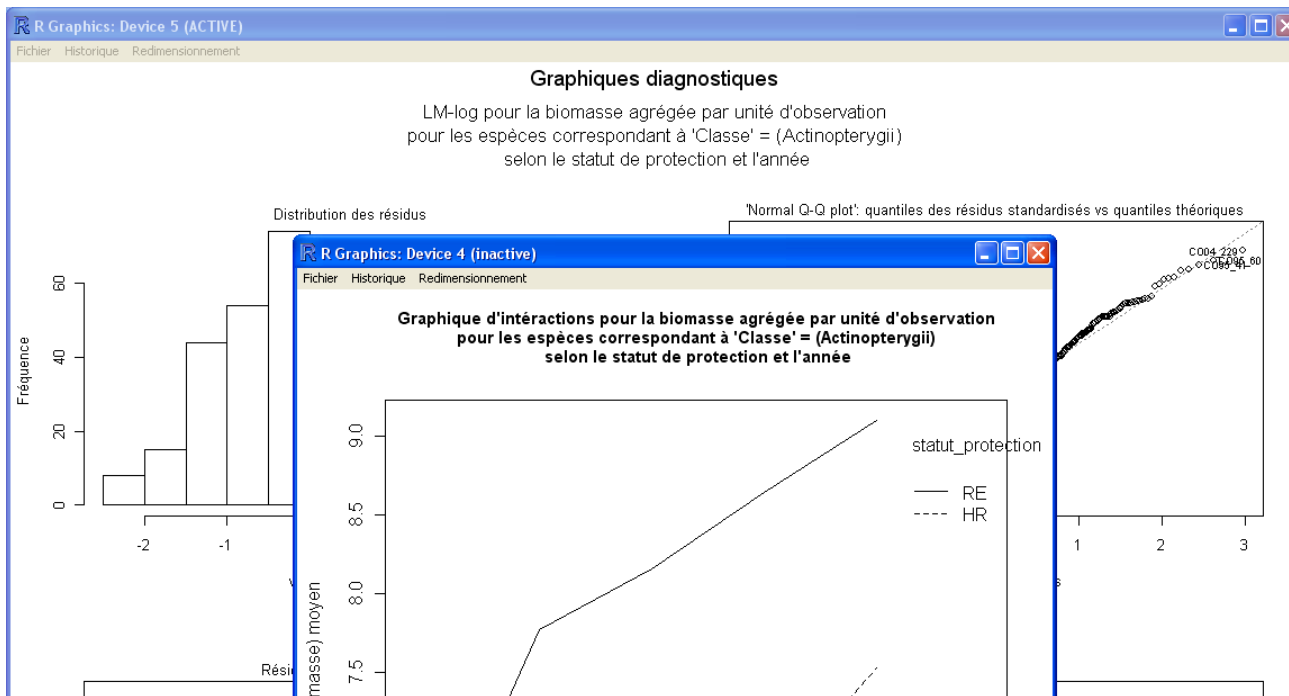
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(P-valeurs ajustées -- méthode 'single-step')

D. Graphiques diagnostiques et valeurs aberrantes

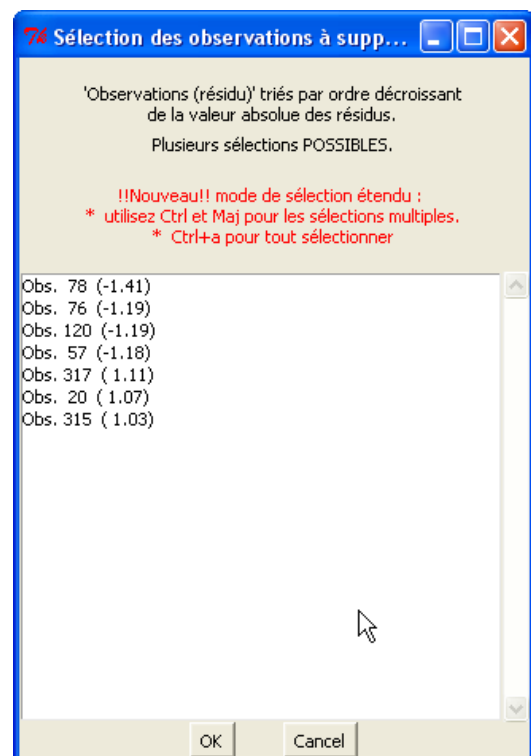
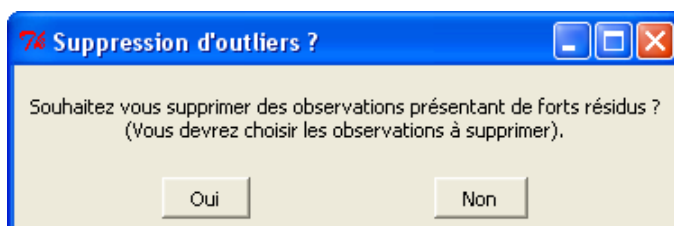
Afin d'offrir les mêmes possibilités que l'ancien système, les graphiques diagnostiques sont créés à la fin des analyses. Ils permettent de tester la qualité de l'ajustement et la pertinence du modèle utilisé.

Ces graphiques sont utiles avec les modèles linéaires simples (avec données log-transformées ou non).

Ces graphiques présentent maintenant des titres explicites qui permettent de ne pas les confondre lorsque plusieurs analyses sont lancées simultanément :



Si des données potentiellement aberrantes (mais pas forcément) existent, il vous sera offert la possibilité de relancer les mêmes analyses en supprimant les observations provoquant les plus forts résidus :



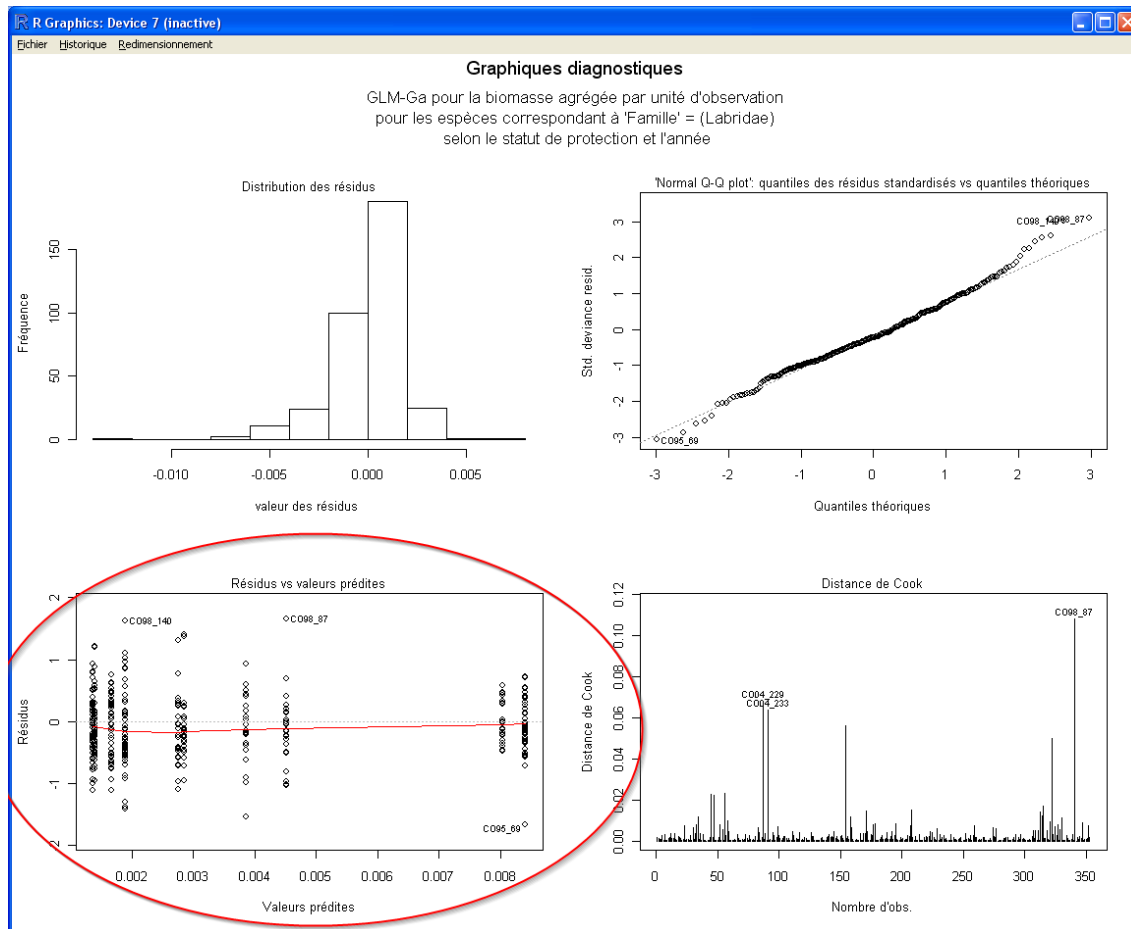
Le fichier de résultats obtenu après suppression des observations aura le même nom, plus un suffixe « `_(red)` » (pour « réduit »).

i. GLMs et graphiques diagnostiques

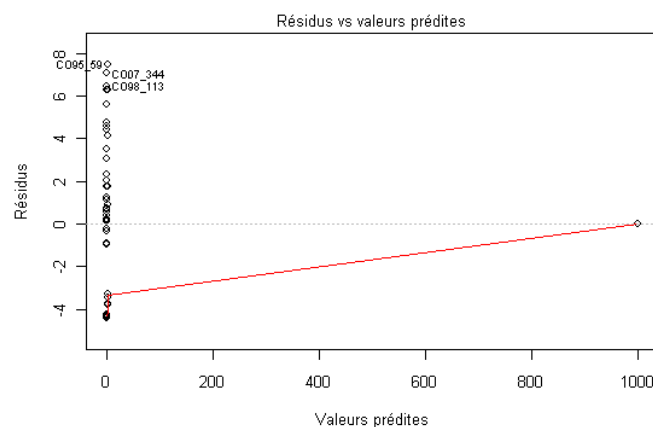
Pour les GLMs, la normalité des résidus – évaluée à l'aide du « Normal Q-Q plot » – en particulier, n'est pas nécessairement attendue.

Dans ce cas, le graphique qui présente le plus d'intérêt pour évaluer la qualité d'ajustement du modèle est celui des résidus en fonction des valeurs prédites. Si l'ajustement est bon, il ne doit pas y avoir de tendance.

Voici par exemple un ajustement correct :



Tandis que le suivant est mauvais :

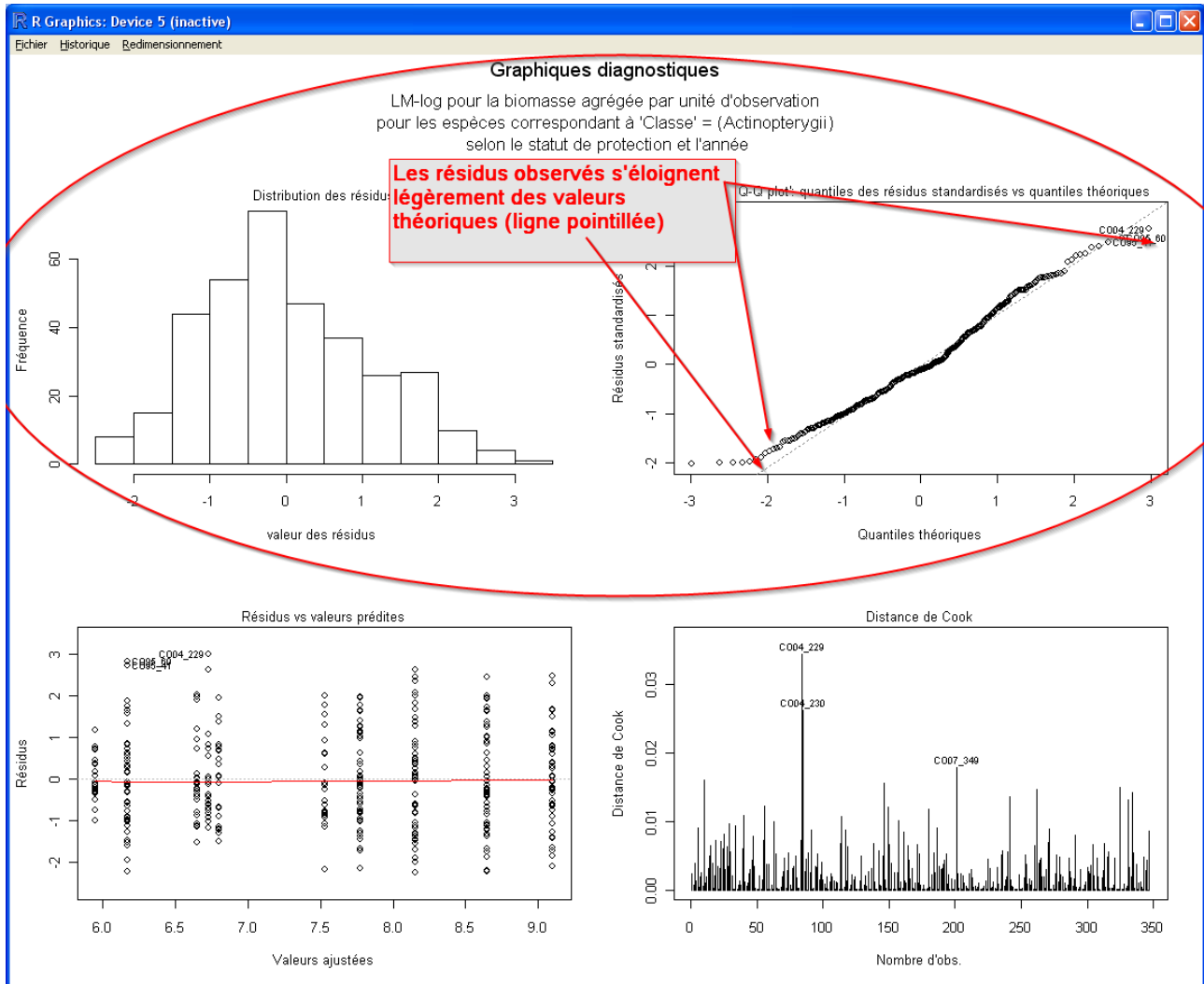


ii. ANOVAs et graphiques diagnostiques

Ici aussi, l'absence de tendance dans la représentation des résidus en fonction des valeurs prédites constitue un critère de bon ajustement. Mais à cela vient s'ajouter un critère de normalité des résidus du modèle, sans laquelle les p-valeurs risquent d'être mal estimées.

Dans l'exemple ci-dessous – un ajustement somme-toute correct – la distribution des

résidus est légèrement dissymétrique, comme le montrent les deux graphiques de la première ligne :



Mais rien d'aussi mauvais que dans l'exemple suivant (distribution bimodale) :

