

Nouveautés de la plateforme PAMPA WP2

Annexe du « Guide d'utilisation du programme de calculs des métriques biodiversité et ressources »

Auteur : Yves REECHT (Yves.Reecht@ifremer.fr)

Résumé : Ce document présente les principales modifications récentes apportées à la plateforme de calcul d'indicateurs et métriques pour le WP2 du projet PAMPA. Il est complémentaire du « Guide d'utilisation du programme de calculs des métriques biodiversité et ressources » avec lequel il sera fusionné ultérieurement.

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.

*Des modifications importantes étant généralement apportées entre 2 versions successive, **vous devriez au minimum lire ces encadrés avant chaque nouvelle installation.***

Les nouveaux systèmes de création des graphiques et d'analyses des données étant voués à remplacer les anciens (plus maintenus), vous êtes invités à les utiliser préférentiellement. Tous les retours concernant l'ergonomie, d'éventuels bugs rencontrés ou encore des fonctionnalités manquantes sont les bienvenus.

Table des matières

1. Installation.....	3
2. Chargement.....	4
a) Packages.....	4
b) Catégories benthiques supplémentaires.....	5
c) Dossiers et fichiers de données.....	5
3. Calculs divers.....	7
a) Sélection et re-calcul.....	7
b) Zéros.....	7
c) tailles d'après les classes de tailles.....	7
4. Graphiques.....	8
a) Cas général.....	8
b) Benthos.....	11
5. Analyses statistiques.....	12
a) Lancer les analyses.....	12
b) Résultats.....	15
i) Informations sur le modèle.....	16
ii) détails sur les facteurs significatifs et leurs coefficients	16
iii) Valeurs prédites.....	17
iv) Comparaisons multiples (2 facteurs).....	17
v) Comparaisons multiples (1 facteur).....	18
c) Graphiques diagnostiques et valeurs aberrantes.....	18
6. Autres.....	19
a) Comportement de l'interface.....	19

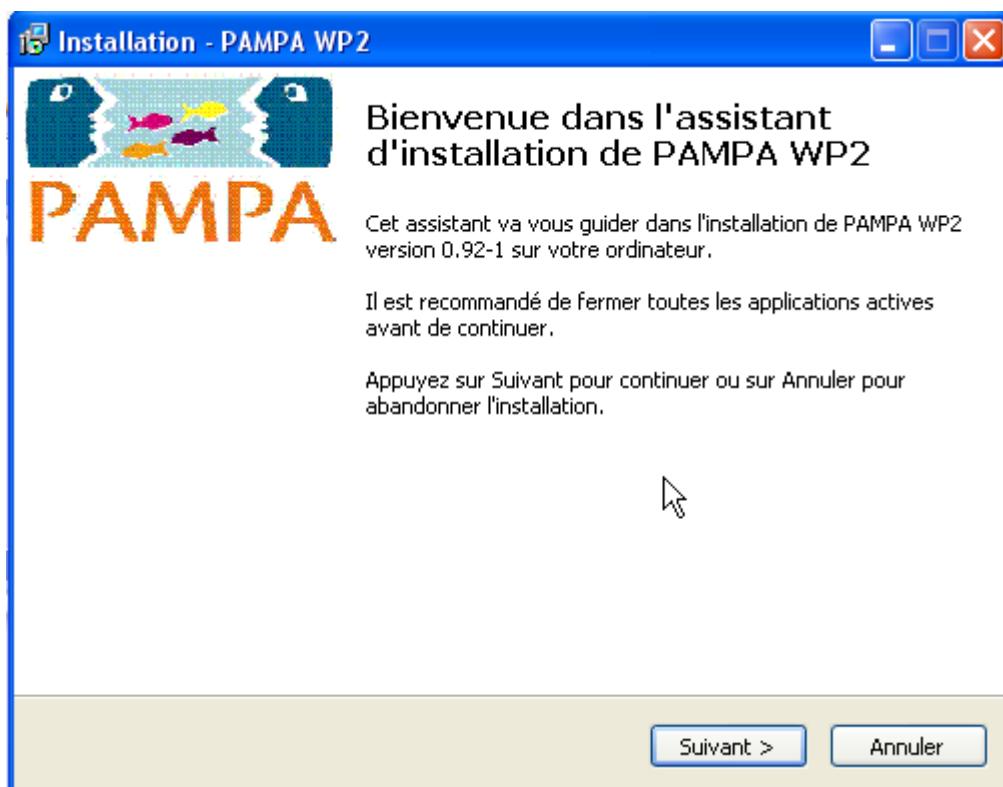
1. Installation

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.

C'est en particulier le cas du fichier « config.r » qui doit contenir les noms de vos fichiers de données. 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_PAMPA_WP2.exe » vous serez dirigé vers une procédure classique (sous Windows) d'installation :

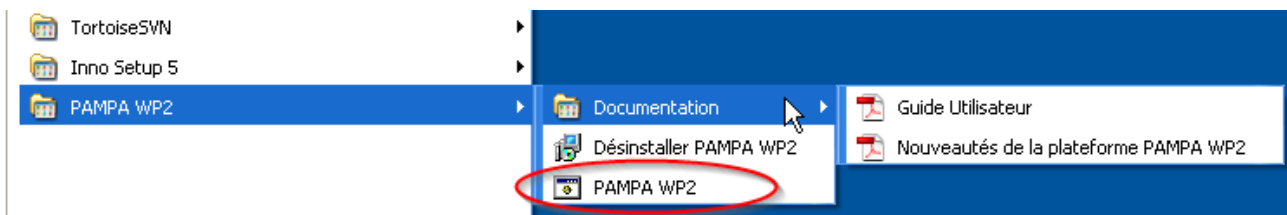


Je vous fais grâce de copies d'écran de chaque étape ; les différentes options sont suffisamment explicites et les valeurs par défaut devraient suffire pour la grande majorité des usages.

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.

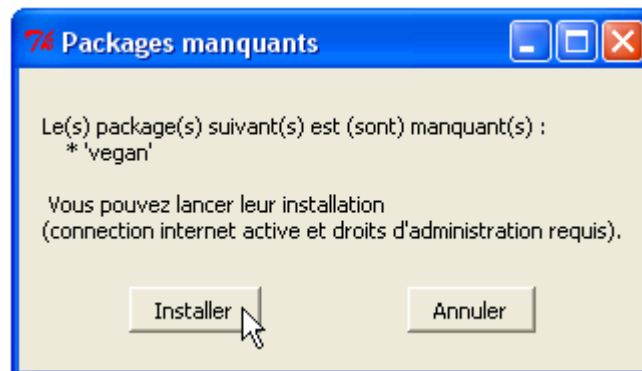
Une fois l'installation terminée, vous disposez de raccourcis pour lancer R et l'interface graphique :



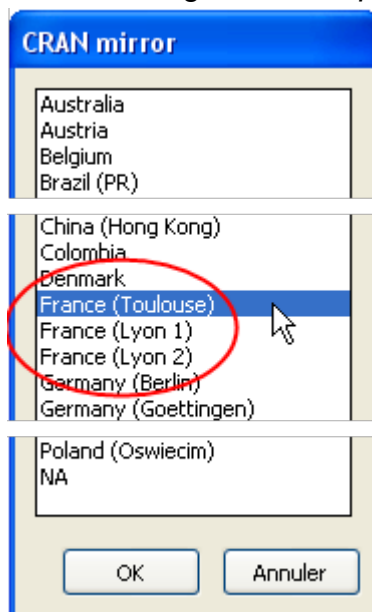
2. Chargement

a) 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 relancer la plateforme dans des conditions qui permettent l'installation.

b) Catégories benthiques supplémentaires

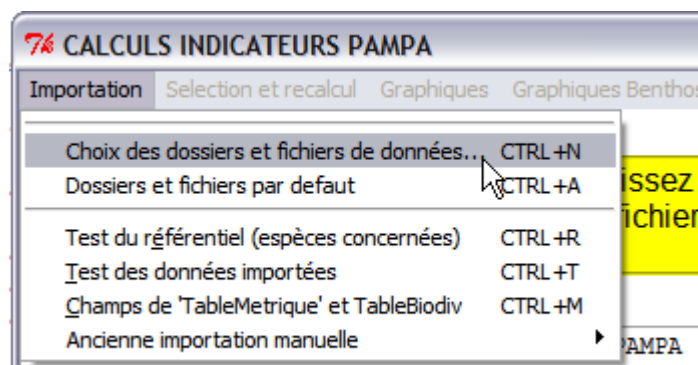
Il est maintenant possible d'ajouter facilement des catégories benthiques supplémentaires, correspondant à des regroupements des catégories existantes. Celles-ci sont définies dans un fichier « corresp-cat-benth.csv » (dans [C:/PAMPA/Exec](#)) qui peut être édité avec un tableur de type Excel. Ce tableau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

	A	B	C
1	Cath_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
14	CM	CV	NACR

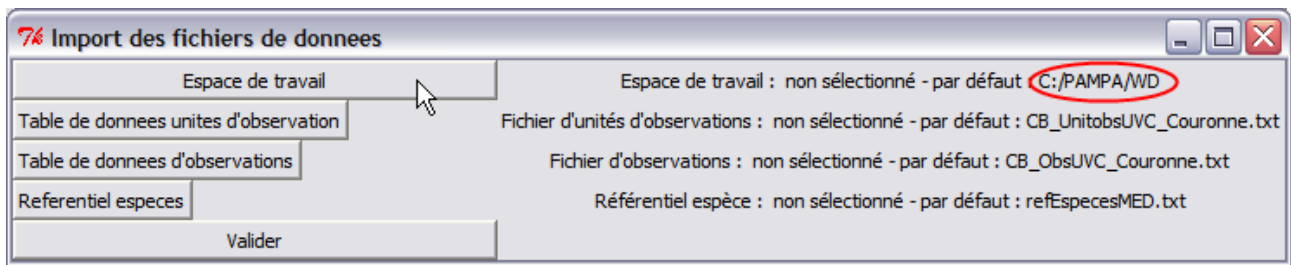
La première colonne doit contenir obligatoirement les valeurs du champ « Cath_benthique » du référentiel espèce. Les suivantes – non limitées en nombre – représentent les catégories agrégées. Des nouveaux facteurs seront alors disponibles sous les noms affichés en première ligne (éviter les espaces et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

c) Dossiers et fichiers de données

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

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

comme cela est expliqué dans le guide utilisateur à la section 3.c (pages 7-8).

Ce fichier faisant également 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).

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

```
#### CB
SiteEtudie <- "CB"
fileName1 <- "CB_UnitobsUVC_Couronne.txt"
fileName2 <- "CB_ObsUVC_Couronne.txt"
fileName3 <- "refEspecesMED.txt"
nameWorkspace <- "C:/PAMPA/WD"
```

N'oubliez pas de commenter (faire précéder de « # ») les autres lignes de configuration.

Ces fichiers peuvent alors être (i) directement chargés par depuis le menu « Importation » 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 ».

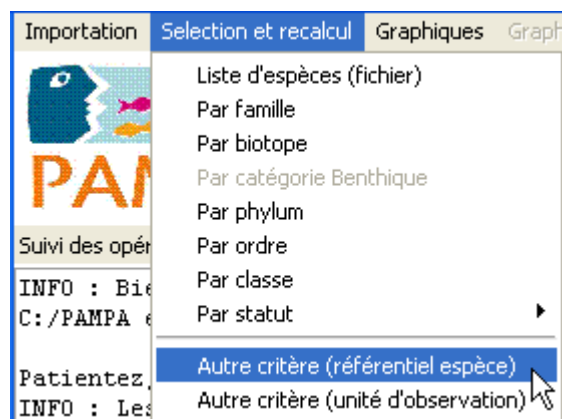
3. Calculs divers

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

a) Sélection et re-calcul

Un bug important a été détecté dans le mécanisme de sélection des observations selon « Autre critère (...) » (que ce soit un critère du référentiel espèce ou des unités d'observations) : Il se traduisait par une sélection partielle des observations selon le critère choisi (souvent moins d'un quart des observations normalement sélectionnées).

Ce bug est maintenant corrigé dans les nouvelles versions.



Ceci implique que tous les graphiques et analyses précédemment faits à partir de ce type de sélection sont potentiellement faux !

b) Zéros...

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.

Ici aussi, si votre jeu de données d'observation contient de telles lignes, des résultats précédents sont potentiellement faux.

c) tailles d'après les classes de tailles

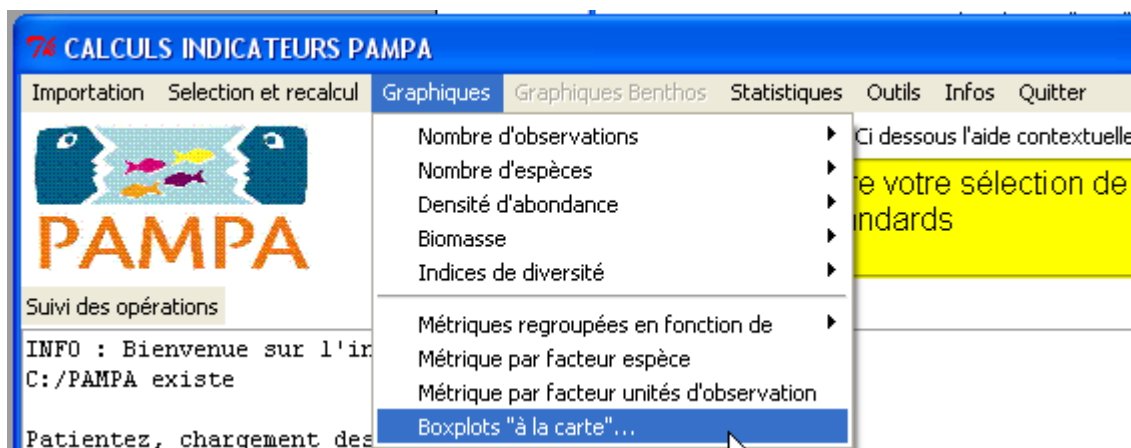
Les tailles ne peuvent pas être systématiquement estimées avec précision lors d'observations *in situ*. En l'absence d'enregistrement des tailles, celles-ci sont maintenant estimées d'après les classes de taille (si celles-ci sont renseignées et s'y prêtent, bien entendu). Les formats de classes pris 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.

4. Graphiques

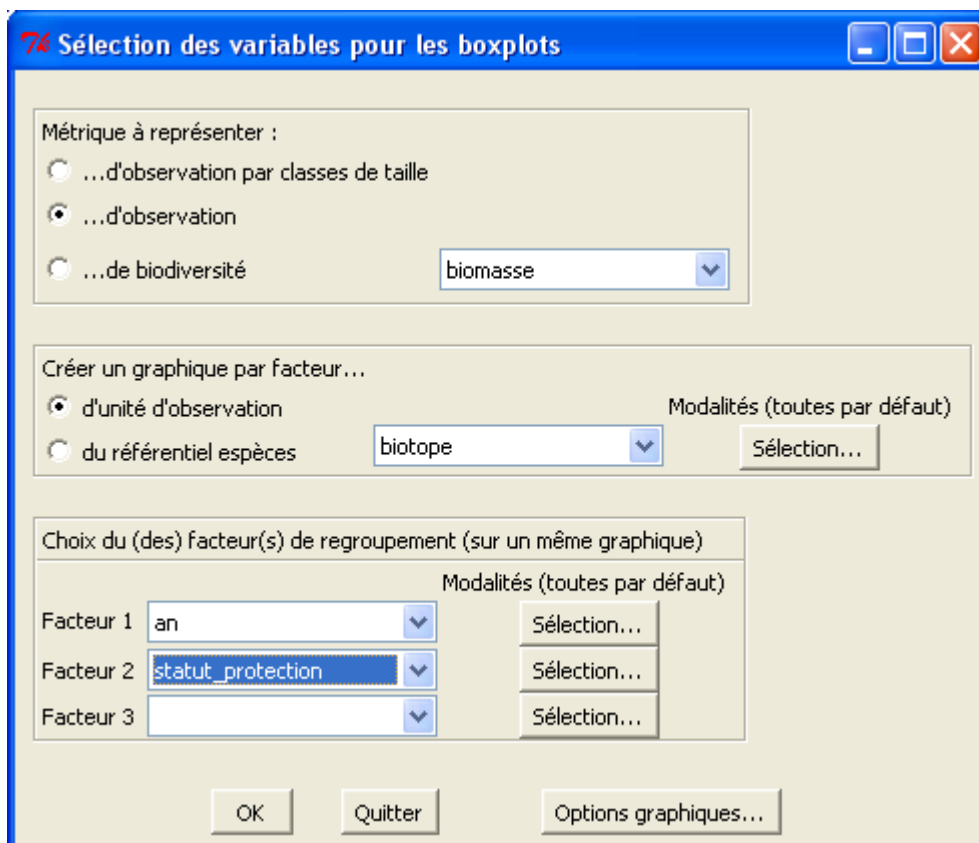
a) Cas général

Un nouveau système de production des « BoxPlots » ou « boîtes à moustaches » a été mis en place, qui permet de faire quasiment tout ce que faisaient les précédents menus, et plus encore.

Il est accessible dans le menu « Graphiques » (sauf pour le benthos) de la plateforme sous le nom « Boxplots "à la carte"... » :



Cette entrée dirige l'utilisateur vers une fenêtre de choix des variables :



Les choix se font à trois niveaux :

1. La métrique à représenter : les choix dans le menu déroulant dépendent du type de métrique/indicateur sélectionné.
2. Un éventuel facteur de séparation des graphiques : un graphique sera créé pour chaque modalité (parmi les modalités sélectionnées) de ce facteur.
3. Un ou plusieurs facteur(s) de regroupement pour les boîtes à moustaches.

Il n'y a pas de limite du nombre de facteurs de regroupement... mais gardez à l'esprit qu'au delà de deux facteurs, les graphiques deviennent vite illisibles.

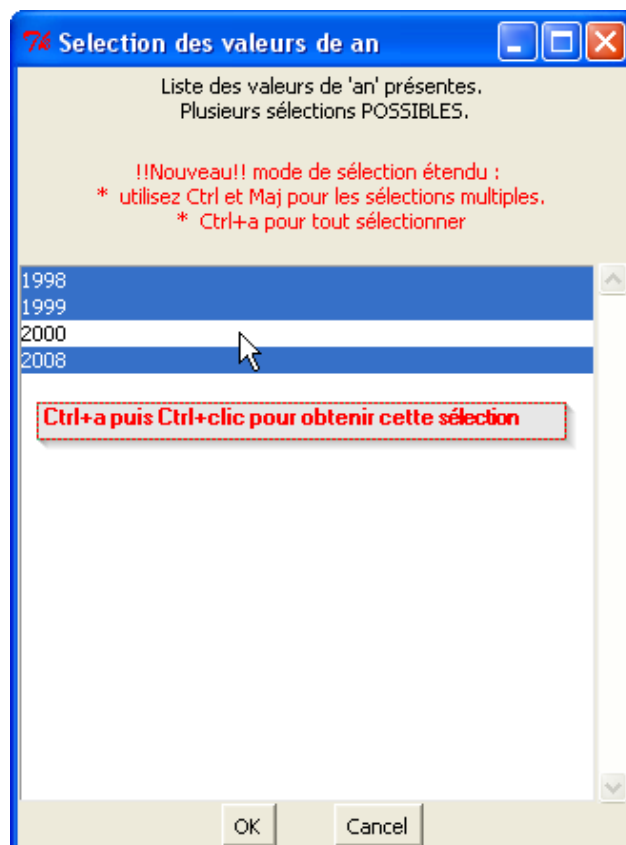
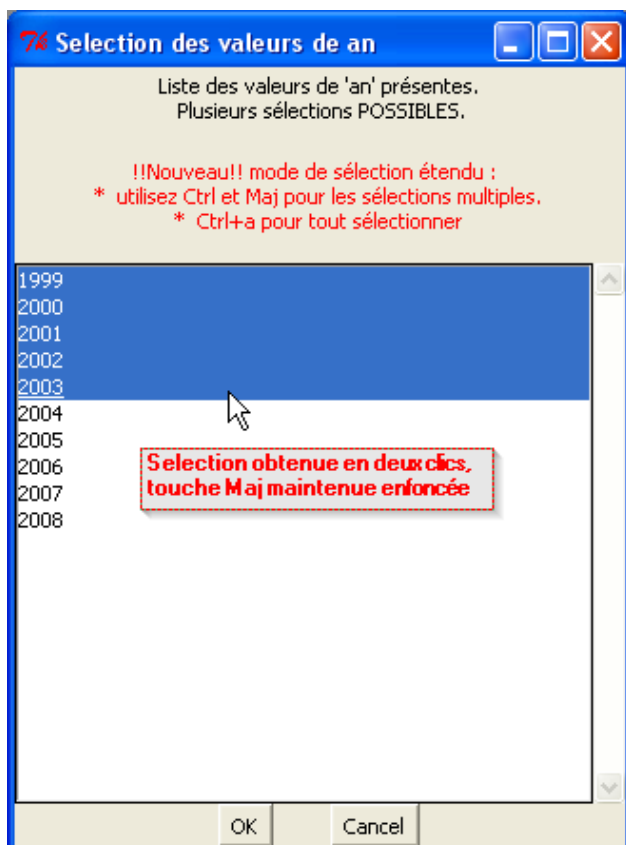
Il n'y a pas d'ordre imposé pour le choix des facteurs/métriques. Les choix de facteurs s'adaptent dynamiquement à la métrique choisie. Si un facteur déjà sélectionné n'est pas cohérent avec une métrique nouvellement choisie (e.g. un facteur du référentiel espèce n'a pas de sens pour un indice de biodiversité), il sera automatiquement supprimé.

Un facteur peut être supprimé en sélectionnant une ligne vide :

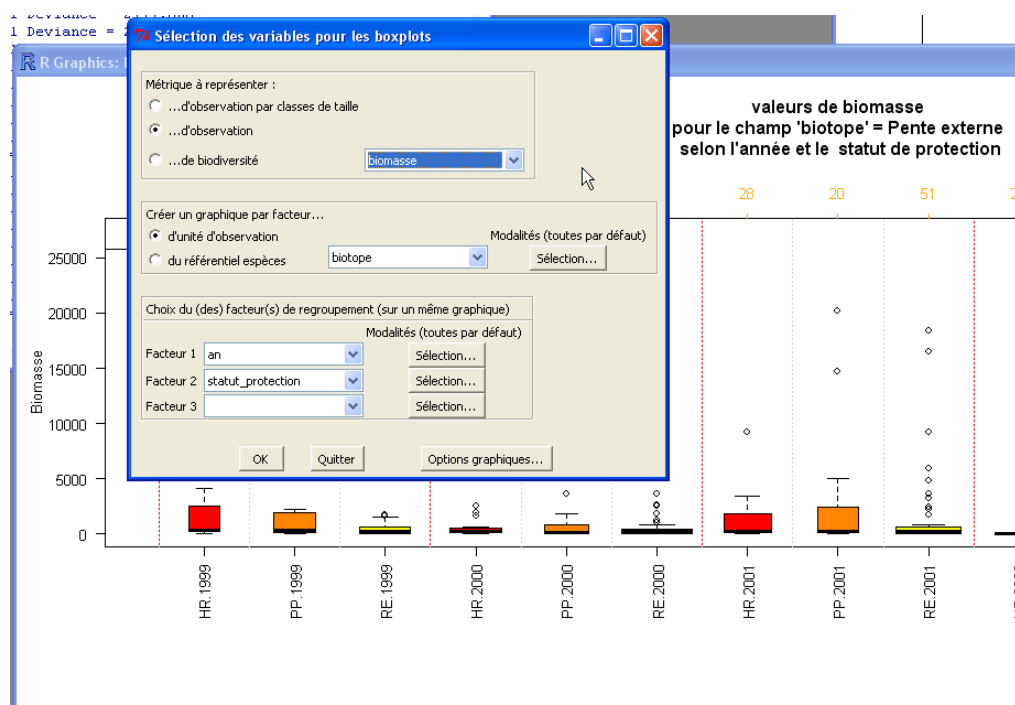
Choix du (des) facteur(s) de regroupement (sur un même graphique)	
	Modalités (toutes par défaut)
Facteur 1 an	Sélection...
Facteur 2 [blue bar]	Sélection...
Facteur 3 statut_protection	Sélection...
Facteur 4	Sélection...

Le fait qu'un « facteur vide » s'intercale entre deux autres n'aura pas d'influence : seul l'ordre des champs renseignés compte (ceci permet de supprimer un facteur tout en conservant les sélections pour les autres, voir plus bas, la remarque sur les sélections persistantes).

Par défaut, toutes les modalités d'un facteur sont retenues. Les boutons « Sélection... » permettent de n'en sélectionner que certaines. Les boîtes de dialogue ont été améliorés pour faciliter les sélections multiples :



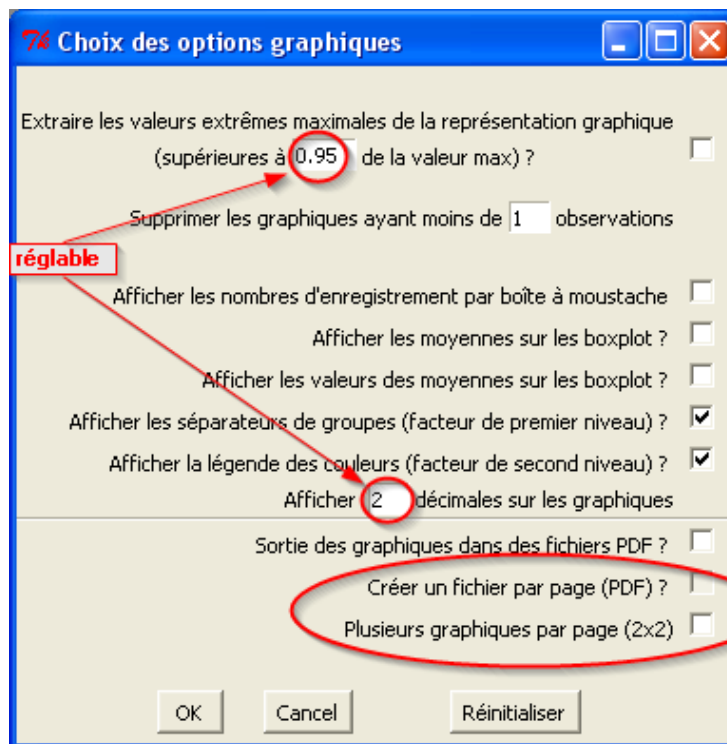
Ces sélections sont persistantes tant que le facteur n'est pas modifié (changé de champs, supprimé) et que la fenêtre de choix de variables n'est pas fermée. Ce comportement permet de réutiliser une sélection pour plusieurs séries de graphiques puisque le fait de lancer la création des graphiques conserve la fenêtre de choix des variables ouverte :



Différentes séries de graphiques présentant de légères variations peuvent donc être produites très rapidement avec ce système.

Les options graphiques – accessibles par le bouton en bas à droite – sont maintenant également persistante. Et ce tout au long de la session R.

Des améliorations ont également été apportées de ce côté-là, avec davantage de seuils modifiables et d'options pour les types de sorties (capture d'écran ci-dessous). En particulier, lors de la création de plusieurs graphiques en pdf, ceux-ci peuvent maintenant être stockés dans des fichiers séparés – plus faciles à intégrer dans d'autres documents – avec un numéro en suffixe. Plusieurs graphique par fenêtre/page de périphérique graphique sont maintenant possibles (limité pour l'instant à 2x2).



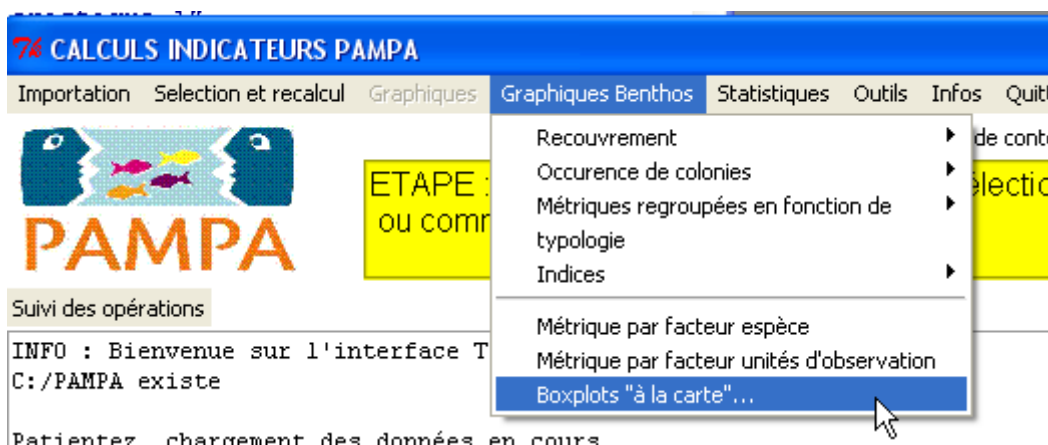
Dans les cas de sorties pdf, les fichiers sont créés dans le répertoire C:/PAMPA/FichiersSortie/ avec un nom de la forme

<préfixe>_<métrique>_[<facteur de séparation>_]<facteur 1>[-<facteur 2>...]-<numéro>].pdf

(les parties entre [] sont optionnelles).

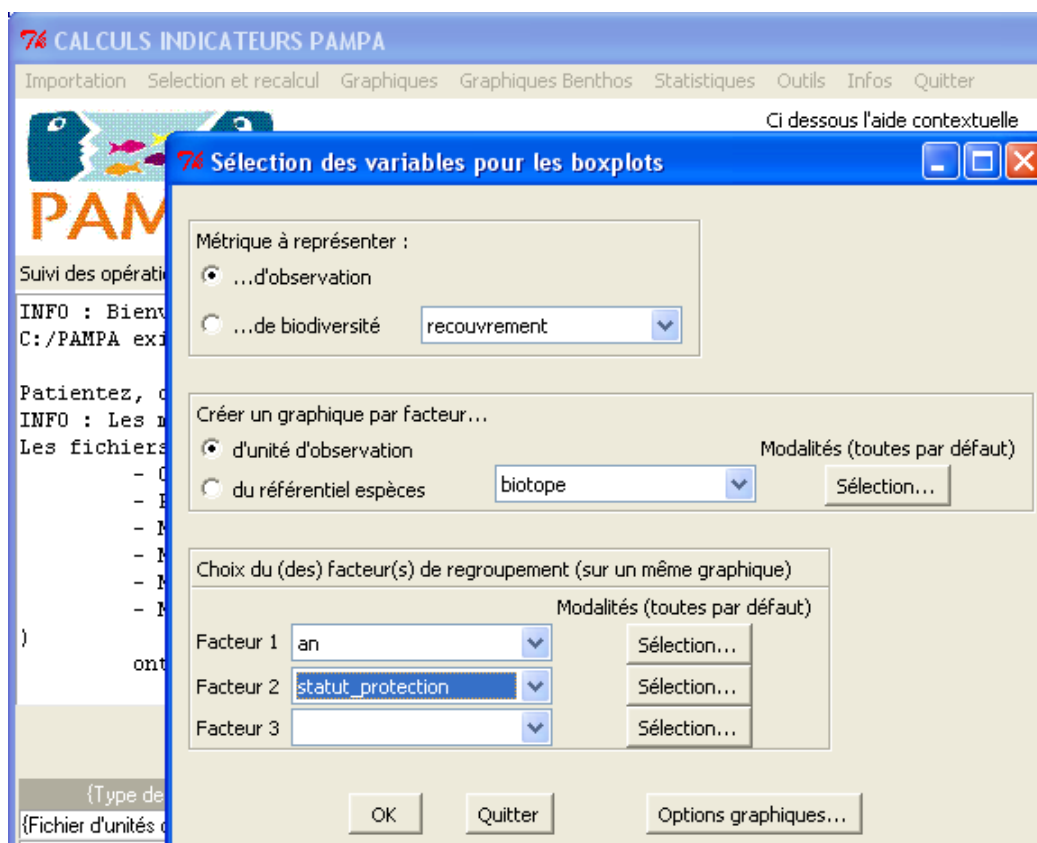
b) Benthos

Le système développé pour le cas général a été adapté pour les données benthos et est accessible dans le menu « Graphiques Benthos », ici aussi sous le nom « Boxplots "à la carte"... » :



La boîte de dialogue de choix des variables est très similaire à la précédente et fonctionne

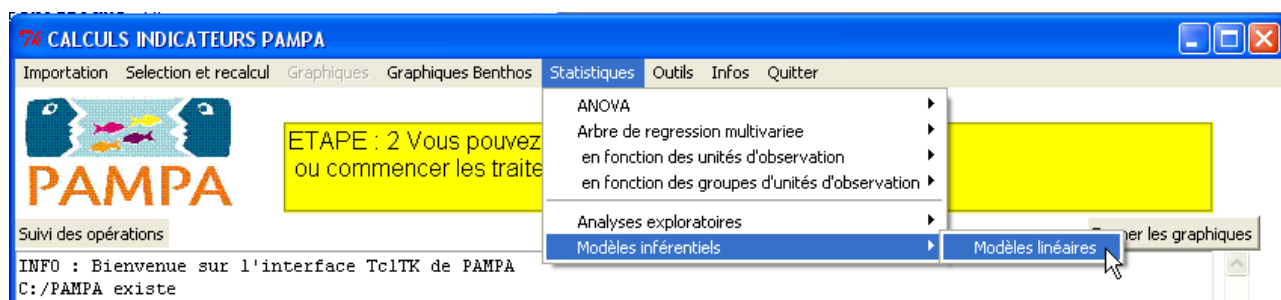
de la même façon. Seules les métriques par classes de tailles ne sont pas disponibles (car non-calculées) :



5. Analyses statistiques

a) Lancer les analyses

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



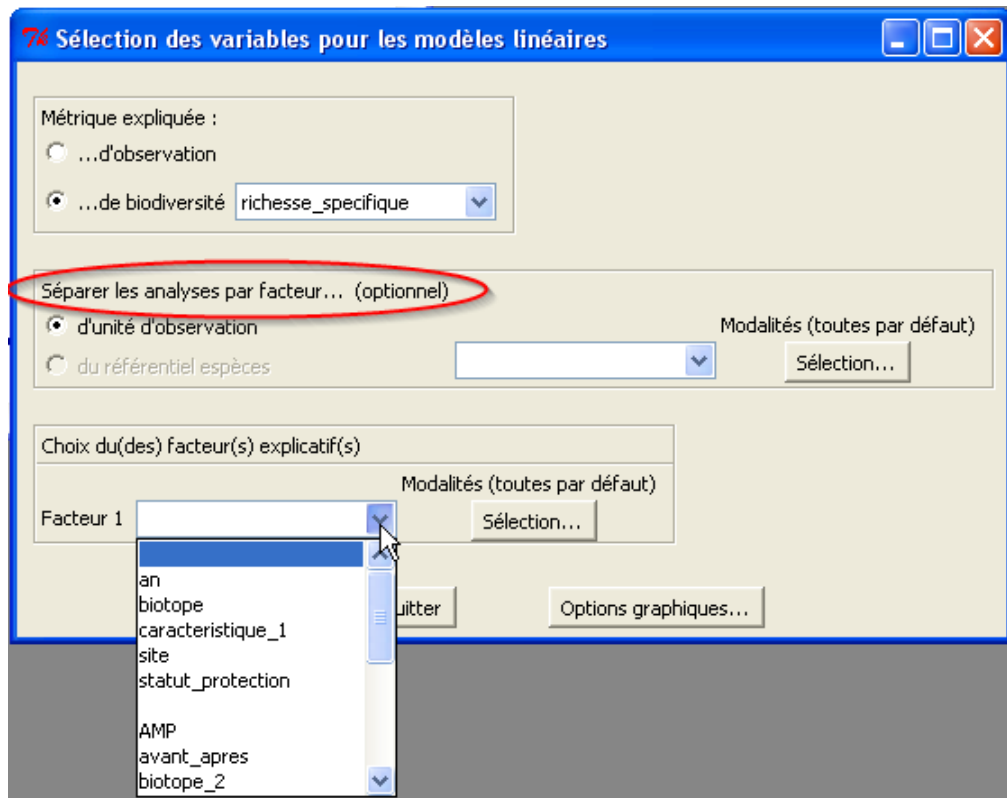
Les principales améliorations apportées sont :

- davantage de souplesse dans le choix des facteurs de l'analyse (nombre, table d'origine, etc.)
- une interface plus intuitive.
- le choix du type d'analyse en fonction de la distribution des données de la métrique choisie. Des modèles linéaires généralisés (GLMs) peuvent maintenant être utilisés

à la place de modèles linéaires simples lorsqu'ils sont plus adaptés.

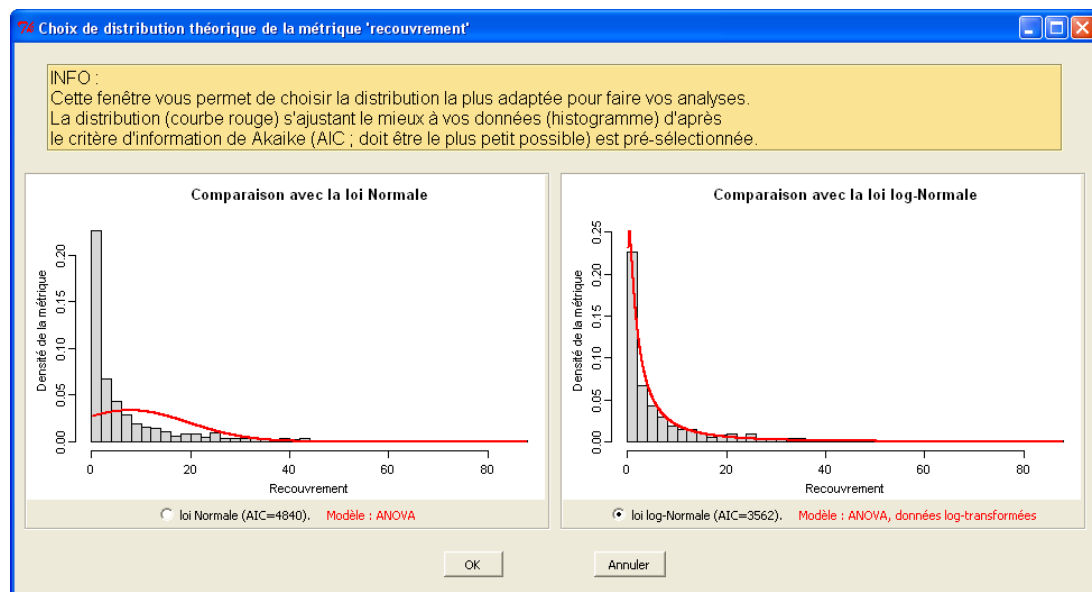
- Comparaisons multiples lorsque les facteurs « statut_protection » et « an » sont conjointement sélectionnés.

La boîte de dialogue de choix des métriques/indicateurs et facteurs est très similaire celle développée pour les graphiques. C'est d'ailleurs la même fonction qui opère dans les deux cas (c'est pourquoi un bouton « Options graphiques... » – inutile ici – s'affiche, qui devrait à terme être remplacé par un bouton pour le choix des sorties graphiques des analyses) :

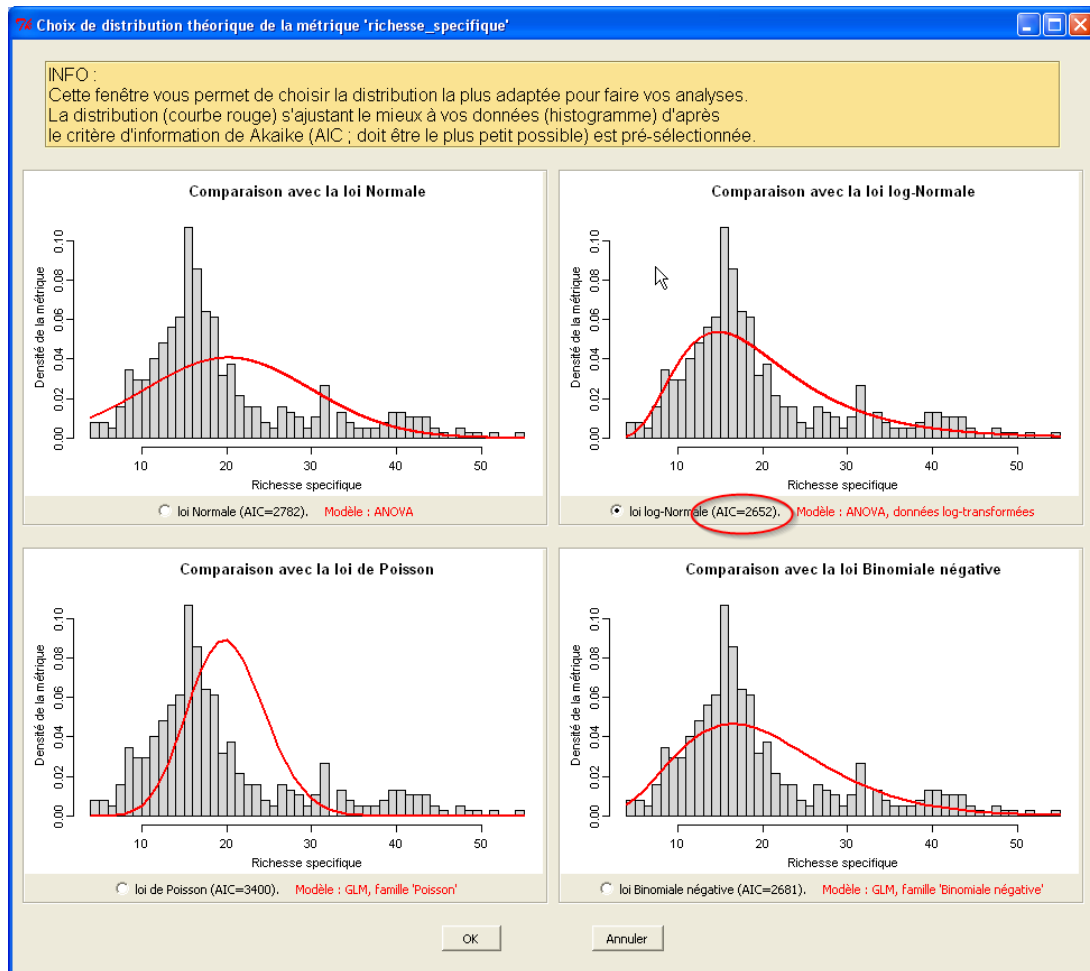


Ici également, le facteur de séparation est optionnel. Dans la plupart des cas, vous ne devriez pas en avoir besoin.

En cliquant sur « OK », vous lancez un module de choix de la distribution théorique des données de la métrique choisie, qui va vous orienter vers le type d'analyse adéquat :



Selon la nature de la métrique sélectionnée, le nombre de choix peut-être variable. Pour les données de comptage (entières), les lois de Poisson et binomiale négative sont en plus disponibles :



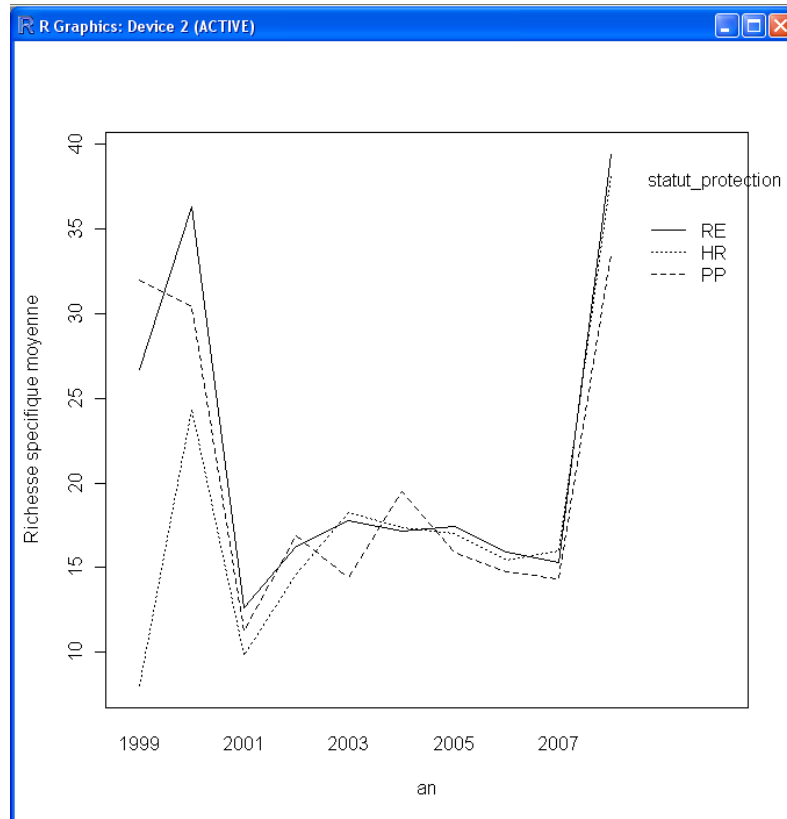
La distribution qui s'ajuste le mieux à vos données selon le critère d'information de Akaike (AIC) – qui doit être le plus petit possible – est déjà présélectionnée. Pour en changer, il vous suffit de cliquer sur le graphique ou le bouton correspondant à votre choix.

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.

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

b) Résultats

Si deux facteurs sont sélectionnés, un *interaction plot* est produit (plus uniquement pour les facteurs « an » et « statut_protection » dorénavant), qui peut servir de support visuel à l'interprétation des comparaisons multiples :



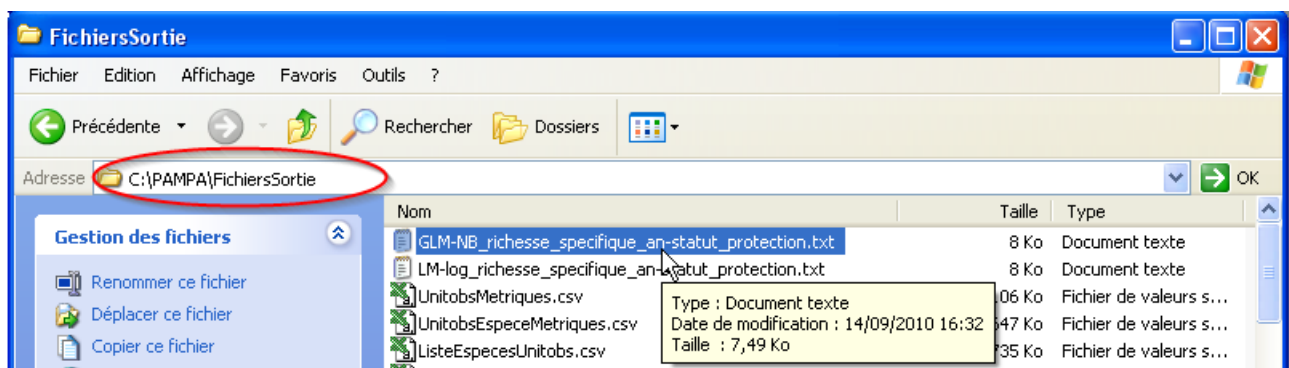
Les résultats de l'analyse sont stockés dans le dossier C:/PAMPA/FichiersSortie/ dans un fichier texte de la forme

<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.



Ce fichier contient plusieurs parties :

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.

iv) Comparaisons multiples (2 facteurs)

Uniquement dans le cas où le statut de protection et l'année sont conjointement sélectionnés comme facteurs du modèle. Lorsque deux facteurs sont sélectionnés :

Comparaisons multiples :

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 « caracteristique_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.

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 :

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')

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

Si des données potentiellement aberrantes (mais pas forcément) existent, il vous sera offert la possibilité de relancer les mêmes analyses en supprimant les données provoquant les plus forts résidus. Le fichier de résultats aura le même nom, plus un suffixe « _(red) » (pour « réduit »).

6. Autres

a) Comportement de l'interface

Le placement des fenêtres ainsi que leur mise au premier plan (pour éviter d'avoir des fenêtre qui sortent en arrière plan à n'importe endroit de l'écran) ont été améliorés (essentiellement sur la fenêtre principale et les nouvelles boîtes de dialogue).