

Résultats et pistes

ESSAI DE SIMPLIFICATION DU PROTOCOLE FLORE DE LA  
BOITE A OUTIL RHO<sub>ME</sub>O POUR LES INDICATEURS I2  
(ENGORGEMENT) ET I6 (FERTILITE)

Janvier 2017

Gilles PACHE





# ESSAI DE SIMPLIFICATION DU PROTOCOLE FLORE DE LA BOITE A OUTIL RHOME0 POUR LES INDICATEURS I2 (ENGORGEMENT) ET I6 (FERTILITE)

1	Rappels et objectifs .....	5
2	Un préalable : avoir des données homogènes et filtrées.....	6
2.1	Problèmes au niveau des sites .....	7
2.2	Problèmes au niveau des relevés.....	7
2.3	Problèmes au niveau des espèces.....	7
3	Définition d'un cadre typologique pour les sites .....	9
3.1	Objectifs.....	9
3.2	Méthodes employées.....	9
3.3	Résultats .....	11
3.4	Proposition d'une typologie synthétique.....	18
4	Réduction des listes d'espèces.....	21
4.1	Méthodes de filtrages possibles.....	21
4.2	Filtrer pour ne retenir que les espèces ZH au sens de l'arrêté de 2008 .....	21
4.3	Filtrer les espèces à large amplitude .....	23
4.4	Filtres sur la fréquence des espèces.....	27
4.4.1	Impacts sur les informations disponibles .....	27
4.4.2	Comparaisons des valeurs diagnostiques selon le niveau de filtre et le type de zone humide à l'échelle des sites.....	28
4.4.3	Comparaisons des valeurs diagnostiques selon le niveau de filtre à l'échelle des relevés .....	35
4.4.4	Discussion et conclusions .....	40
4.5	Filtres sur la valeur indicatrice (IndVal) ou le coefficient Phi .....	41
4.5.1	Découpage des valeurs diagnostiques en classes .....	42

4.5.2	Résultats pour l'engorgement.....	44
4.5.3	Validation des résultats pour l'engorgement.....	47
4.5.4	Résultats pour la fertilité.....	51
4.5.5	Validation des résultats pour la fertilité.....	52
4.6	Analyse des listes d'espèces.....	55
4.7	Conséquences possibles sur l'interprétation des évolutions.....	61
5	Conclusion générale.....	65

## Bibliographie

ANNEXE 1 : données relatives aux nombres d'observations, nombre d'espèces et nombre d'espèces en commun selon les principales méthodes (filtre sur la richesse à 50, 100 ou 150 espèces, filtre avec IndVal, filtre avec Phi), par type de zone humide et globalement.

ANNEXE 2 : liste des espèces données indicatrices par l'une des sept méthodes de filtre, par type de zone humide.

# 1 Rappels et objectifs

---

Le protocole flore actuel de la BAO RhoMÉO nécessite de noter toutes les espèces présentes sur les placettes étudiées. La question posée est : est-il possible de reproduire les valeurs calculées avec l'ensemble des espèces en n'utilisant qu'un lot réduit d'espèces, sans perdre trop de précision dans l'estimation des valeurs de niveau d'engorgement et de fertilité, et si oui, comment ? Etant donné la grande diversité des types de zones humide et des contextes climatiques, une seule liste, valable sur l'ensemble du bassin versant et pour tous les types de zones humides, et comportant peu d'espèces, paraît impossible à atteindre. L'option la plus simple consiste à établir plusieurs listes en fonction d'une typologie des zones humides. Une typologie « RhoMÉO » a été proposée (basée en grande partie sur la typologie SDAGE). Dans un premier temps, sa pertinence pour notre problématique doit être évaluée. Dans un deuxième temps, l'établissement de listes restreintes par « type » de zone humide sera mené avec différentes méthodes, et la pertinence des résultats obtenus sera testée par référence aux évaluations obtenues avec les données complètes.

Les objectifs poursuivis peuvent être résumés ainsi :

- fournir des listes d'espèces fermées (100 espèces au maximum) qui permettent de reproduire aux mieux les évaluations obtenues avec les listes complètes pour les indicateurs I2 (niveau d'engorgement) et I6 (fertilité des sols) ;
- fournir un cadre typologique pour les sites qui permette d'avoir des listes les plus adaptées possibles ;
- évaluer la pertinence des évaluations obtenues ;
- définir les types de zones humides (ZH) pour lesquelles la pertinence est insuffisante ;
- fournir, le cas échéant, les clés d'interprétation pour l'analyse temporelle des indicateurs.

Tous les calculs et graphiques ont été réalisés avec le logiciel libre R version 3.1.3 (R Core Team, 2015). Outre les fonctions de base, les bibliothèques de fonctions (package dans la terminologie de R) suivantes ont été utilisées :

- 'ade4' (DRAY & DUFOUR, 2007), 'vegan' (OKSANEN *et al.*, 2015), 'gclus' (HURLEY, 2012) et 'optart' (ROBERTS, 2016) pour les analyses multivariées ;
- 'plyr' (WICKHAM, 2011) et 'reshape2' (WICKHAM, 2007) pour la mise en forme des données ;
- 'lattice' (SARKAR, 2008) pour certains graphiques ;
- 'labdsv' (ROBERTS, 2015) et 'indicspecies' (DE CACERES & LEGENDRE, 2009) pour les calculs de IndVal et Phi et certaines manipulations des tables.

## 2 Un préalable : avoir des données homogènes et filtrées

La base de données mise à disposition par le CEN Rhône-Alpes comporte 190 sites renseignés au 21/09/2016 (tableau 1).

**Tableau 1** : bilan global des données par type de zone humide avant tout filtrage.

Code ZH	Type ZH	Nb. Sites total	Nb. relevés total	Nb. relevés	Nb. obs total	Nb. obs.	Nb. taxons total	Nb. taxons
-	Non renseigné	7	82	82	817	817	145	145
3.2	Péri-lagunaire	4	162	157	834	754	161	150
3.3	Péri-lagunaire avec apport d'eau	4	269	260	1.376	1.237	208	190
5	Bordures de cours d'eau	21	334	324	5.131	4.746	708	668
6	Plaines alluviales [inondable]	27	756	752	11.277	10.933	744	723
7.1	Zones humides de bas-fonds en tête de BV : altitude	8	117	117	1.693	1.600	264	246
7.2	Zones humides de bas-fonds en tête de BV : tourbières acides	25	579	518	8.542	7.994	531	503
7.3	Zones humides de bas-fonds en tête de BV : tourbières alcalines	24	420	396	6.405	6.347	467	461
7.4	Zones humides de bas-fonds en tête de BV : pentes et sources	8	90	88	1.389	1.370	328	322
7.5	Zones humides de bas-fonds en tête de BV : combes et bordure de ruisseau	4	53	53	1.264	1.208	219	203
7.3_7.5		1	9	9	154	154	76	76
8	Régions d'étangs	3	34	34	317	315	119	118
9.2	Bordures de plan d'eau (lac) : ZH alcaline	6	154	154	3.557	3.546	365	362
10.1	Marais et landes humides de plaine : tourbières de plaine	15	287	286	5.586	5.479	447	436
10.2	Marais et landes humides de plaine : prairies humides	14	226	217	4.181	4.092	544	527
11.11	Zones humides ponctuelles : mares temporaires saumâtres	1	11	11	70	68	29	29
11.12	Zones humides ponctuelles : mares temporaires alcalines	6	125	120	696	593	129	122
11.13	Zones humides ponctuelles : mares temporaires acides	6	147	146	1.532	1.392	241	
11.2	Zones humides ponctuelles : mare permanente	4	17	15	181	134	111	49
13	Zones humides artificielles	2	24	24	266	266	123	123

Le bilan des données par type de ZH fait apparaître :

- des disparités énormes des nombres d'observations par type de ZH (70 à 10.977 – moyenne de 2.638 – médiane de 1.233) ;

- des écarts importants de la richesse spécifique (29 à 744 – moyenne de 287 – médiane de 217) ;
- certains types sont vraiment trop peu présents (1 ou 2 sites) dans l'échantillon pour pouvoir être analysés (7.3\_7.5 ; 11.11 et 13) ;
- pour d'autres, peu présents, le regroupement avec d'autres types mérite d'être analysé.

Par ailleurs, trois sources de problèmes existent (au niveau des sites, au niveau des relevés et enfin au niveau des espèces), impliquant un nettoyage préalable des données ; nous les examinerons successivement.

## 2.1 Problèmes au niveau des sites

---

Treize sites ne sont pas renseignés pour le type de zone humide ou pour l'abondance des espèces (soit les deux) : pour 6 d'entre eux que nous connaissons, nous avons attribué le type de zone humide, mais nous avons écartés les 7 autres des analyses.

Pour dix-neuf sites, le nombre de données est très faible à faible ; bien qu'il n'y ait pas de rupture évidente dans la courbe du cumul des distributions du nombre de taxons, nous avons estimé que les sites avec moins de 50 données devaient être écartés.

## 2.2 Problèmes au niveau des relevés

---

Sur les 3.711 relevés, 166 ne comportent qu'une espèce, soit parce que le relevé est très incomplet, soit dans quelques cas car il n'y a effectivement qu'une seule espèce détectée ; 130 autres contiennent deux espèces, également pour les mêmes raisons. Nous avons procédé à un tri manuel de ces environs 300 relevés et écarté ceux qui étaient manifestement des relevés très incomplets (un cas fréquent est le signalement d'une espèce patrimoniale et/ou protégée) et conservé ceux qui correspondaient de toute évidence à des milieux très pauvres en espèces (certaines roselières ou cariçaies).

Dernier problème rencontré, l'existence de relevés semblant assez complets mais dépourvus de coefficients d'abondance-dominance ; ceci est réhibitoire pour certaines analyses mais pas pour d'autres, et nous avons choisi de les inclure ou de les exclure selon les cas.

## 2.3 Problèmes au niveau des espèces

---

Au total, 55.268 observations sont disponibles, mais 1.362 n'ont pas de code CD\_RHOME0 qui permet de faire le lien avec les valeurs indicatrices. Sur ces données non renseignées, 36 concernent d'autres groupes (characées, algues diverses), environ 900 concernent des plantes déterminées dans l'immense majorité des cas au genre (*Agrostis*, *Alchemilla*, ...), environ 300 concernent des données de « mousses ». Pour la centaine de données restantes, il s'agit d'un problème nomenclatural ou taxonomique et nous avons effectué les correspondances adaptées.

Si des valeurs indicatrices ont été renseignées pour la Région Rhône-Alpes pour l'ensemble des espèces, ce n'est pas le cas dans les autres régions ; ainsi, près de 310 taxons (hors mousses et données au genre) ne sont pas renseignés ; pour un certain nombre d'entre eux, il s'agit d'autonymes non gérés (par exemple utilisation de *Pimpinella major subsp. major* au lieu de *Pimpinella major*) mais pour beaucoup, il s'agit d'espèces n'existant pas en Rhône-Alpes ou qui n'ont pas été contactées sur les sites tests de cette région ; pour l'essentiel, ce sont les régions Languedoc-Roussillon et Provence-Alpes-Côte d'Azur qui sont concernées, et plus particulièrement les milieux littoraux, lagunaires et spécifiques aux climat méditerranéen (tableau 1). Pour environ 250 taxons, nous avons amendé la table de correspondance utilisée en Rhône-Alpes pour permettre aux noms cités d'être correctement rattachés ; pour quelques autres, nous avons complété les valeurs indicatrices d'après FLORA INDICATIVA (LANDOLT *et al.*, 2012) et surtout JULVE (2016). Cela porte à 1.618 taxons présents au moins une fois et renseignés. **Pour les autres, faute de connaissances suffisantes de ces espèces et des sites, nous n'avons pas jugé possible de combler les manques. Cela laisse environ 53 taxons dépourvus de valeurs indicatrices, mais ces 53 taxons ne représentent que moins d'une centaine d'observations.**

Au total, une fois finis le nettoyage et les adaptations, ce sont 52.137 observations qui sont disponibles avec des valeurs indicatrices renseignées pour les indicateurs I2 et I6 pour 165 sites. Ces données seront utilisées pour construire le cadre typologique que l'on a choisi de baser sur le nombre d'occurrences des espèces par site. Si on ne conserve que les données avec un coefficient d'abondance-dominance, le jeu de données comporte donc 51.465 observations disponibles pour le travail de simplification.

## 3 Définition d'un cadre typologique pour les sites

---

### 3.1 Objectifs

---

Face à la diversité des types de zones humides présentes sur l'ensemble du bassin versant du Rhône, il est peu probable de trouver une liste d'espèces qui soit à la fois d'ampleur restreinte et valable sur l'ensemble de ce territoire. Il faut donc essayer de grouper d'une manière ou d'une autre les sites. L'approche la plus simple est de s'en tenir aux types de zones humides tels que définis dans la boîte à outils RhoméO. Il n'est toutefois pas certain *a priori* que ce soit la manière la plus pertinente d'opérer une stratification des sites car cette typologie est basée sur des logiques ou des critères pas forcément pertinents par rapport à la flore : ainsi, le type « zone humide de tête de bassin versant » recouvre des réalités très variées (altitude, substrat, conditions climatiques, ...) alors que le type « lagunes et marais côtiers » semble plus homogène ; *a contrario*, il est possible que des similitudes entre types de zones humides permettent au contraire de les regrouper. D'autre part, dans une logique de recherche d'espèces végétales indicatrices, il semble juste qu'une typologie de sites soit basée en partie sur des critères floristiques. Dans le même temps, il ne s'agit pas de remettre en cause la typologie RhoméO, **mais bien de rechercher les regroupements possibles qui permettent d'assurer la meilleure homogénéité possible des sites**. Il ne s'agira donc pas de remettre en cause l'assignement d'un site à un type de zone humide non plus. Enfin, cette procédure permettra de franchir l'écueil des types de zones humides très peu représentés dans l'échantillon (1 ou quelques sites) et celui des types ayant une composition floristique très originale.

### 3.2 Méthodes employées

---

Trois **méthodes d'analyses multivariées d'ordination** ont été employées successivement.

- L'analyse factorielle des correspondances (AFC), basée sur une mesure de distance (entre sites ou entre espèces) dite du Khi-deux ; elle présente l'avantage d'être bien connue, de se prêter à des interprétations graphiques simples (la proximité entre sites sur les plans des analyses implique une ressemblance entre les cortèges d'espèces présents sur ces sites) ; en revanche, cette technique tend à donner beaucoup de poids aux espèces rares ; pour y remédier en partie, seules les espèces présentes au moins 5 fois dans les jeux de données ont été retenues. Enfin, les analyses ont porté sur un tableau croisant les sites et les espèces, le nombre de fois où une espèce a été contactée sur l'ensemble des placettes échantillonnées sur ce site figurant à l'intersection site-espèce.
- L'analyse en coordonnées principales (PCOA), basée sur une mesure de distance entre les sites ; plusieurs distances peuvent être utilisées et nous avons retenu celle de Chord (ou de la corde) ; cette analyse, couplée à la distance de Chord permet une meilleure dispersion des points que l'AFC et est beaucoup moins sensible aux espèces rares que l'AFC (par exemple, GILLET, 2000) ; l'interprétation se fait comme pour l'AFC essentiellement sur la base des graphiques des plans factoriels (plutôt appelés dimensions dans le cas de la PCOA). La PCOA a été employée suite à deux premières AFC qui ont permis d'isoler nettement plusieurs lots de sites assez homogènes entre eux mais avec des flores très originales, alors que la troisième AFC portant sur les sites restants ne permettait plus de séparations évidentes (phénomène d'agglutination, autre défaut fréquent de l'AFC).

- L'AFC inter-classes (DOLEDEC & CHESSEL, 1989 par exemple), permet de découper le tableau croisant les sites et les espèces selon un autre critère (ici le type de zone humide) et de masquer cet effet ou de le mettre en exergue de manière explicite.

Dans tous les cas, les espèces présentes dans moins de 5 sites ont été exclues pour ne pas trop peser sur les AFC et pour avoir des résultats de PCOA issues du même jeu de données.

Pour aider à l'interprétation des analyses et rechercher si des groupes de sites ressortent, deux **méthodes de classification** ont été employées :

- des classifications hiérarchiques (CAH) ; suivant les préconisations de AHO *et al.* (2008), et de BORCARD *et al.*(2011), nous avons retenu comme mesure de dissimilarité entre les sites celles de Chord et Bray-Curtis et comme méthodes d'agrégation celles de Ward (valable en théorie uniquement pour les dissimilarités euclidiennes comme celle de Chord) et celle de l'association moyenne (UPGMA) pour Bray-Curtis ;
- une classification non hiérarchique, la partition autour des médoïdes (PAM – REYNOLDS *et al.*, 1992) ; cette méthode est proche de la méthode dite des « k-means » (MacQUEEN, 1967), plus ancienne et connue, mais est plus robuste (GILLET, 2000).

Le choix de la distance, de la méthode de classification et du nombre optimal de classes a été fait en combinant plusieurs approches :

- le calcul de la corrélation cophénétique, valeur indiquant dans quelle mesure la matrice de distance originale est préservée par la combinaison « dissimilarité – méthode d'agrégation » ; plus cette corrélation est élevée, meilleure est la combinaison ; cette mesure est en général maximale pour la méthode de l'association moyenne ;
- le calcul du coefficient d'agglomération, évaluant le niveau de structure hiérarchique d'un arbre ; plus ce coefficient est élevé, plus l'arbre présente des structures compactes et distinctes entre elles ; cette mesure est en général maximale pour la méthode de Ward ;
- pour les CAH, l'examen des graphiques des niveaux de fusion, qui montrent les valeurs de dissimilarités auxquelles les fusions entre deux branches successives se réalisent ; le nombre optimal de classes correspond au segment de longueur maximal ;
- la méthode des silhouettes (ROUSSEUW, 1987) ; pour cela, un nombre croissant de classes (entre 1 et 16, 16 étant le nombre de catégories de zones humides de la typologie RhoméO) est choisi, et on évalue pour chacune des partitions obtenues la qualité d'assignement de chaque site à une classe, sous la forme d'un indice variant de +1 à -1 (les valeurs positives indiquant que le site est bien classé dans ce groupe, les valeurs négatives indiquant que le site est probablement mal classé dans le groupe, les valeurs proches de zéro indiquant un classement incertain) ; l'évaluation globale de la partition est obtenue en calculant la moyenne des valeurs sur l'ensemble des sites ;
- la méthode 'partana' de ROBERTS (*in* ROBERTS, 2016), basée notamment sur les rapports des similarités moyennes entre objets d'une même classe, la similarité moyenne entre classes et le rapport des similarités à l'intérieur des classes sur les similarités entre classes ; plus la statistique est élevée, meilleure est la qualité globale de la partition.

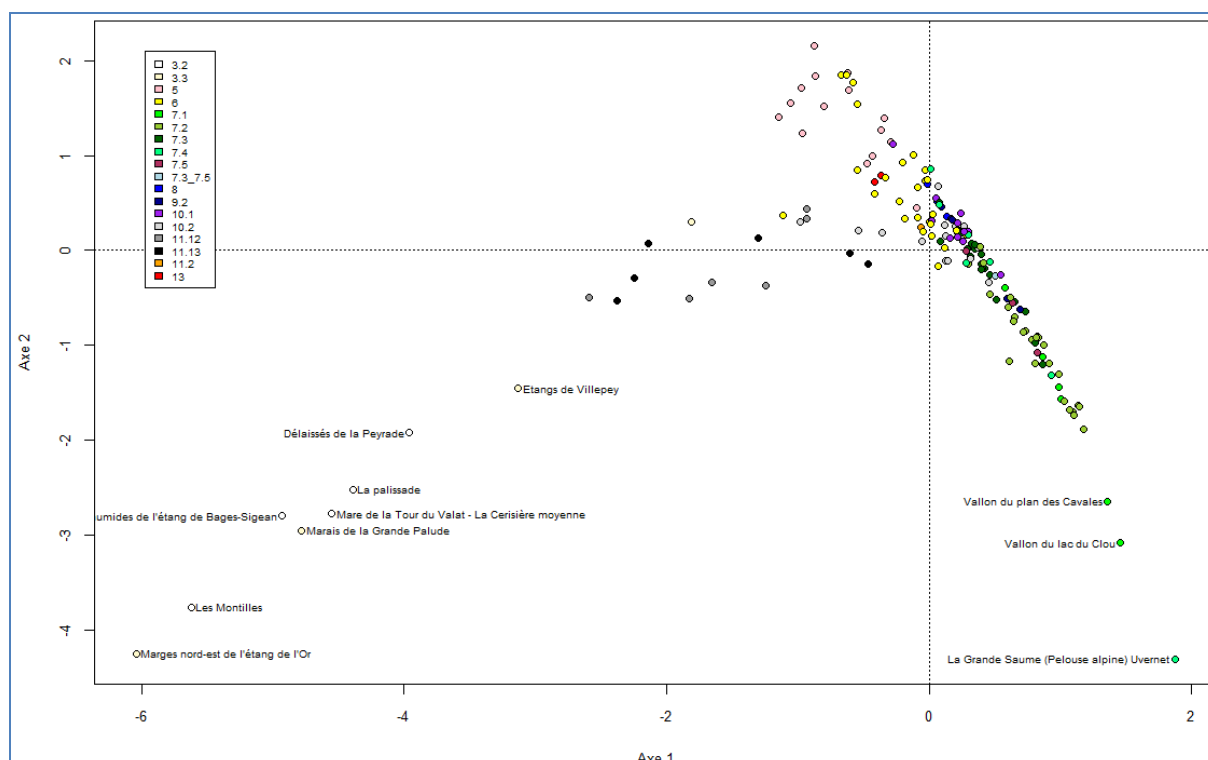
Pour évaluer la cohérence des différentes partitions obtenues avec les différentes méthodes et indices de dissimilarité, des matrices de confusion ont été calculées pour permettre de rechercher des groupes de sites qui restent souvent ou toujours classés ensemble. Enfin, les coefficients Phi d'association entre les groupes de sites ou les types de zones humides et les espèces (CHITRY & TYCHY, 2002) ont été calculés afin d'évaluer le nombre d'espèces liées aux unes ou aux autres.

### 3.3 Résultats

Une première AFC globale porte sur 165 sites et 1.596 taxons avant filtrage, ce dernier permettant de réduire les taxons à 725. Le plan des axes 1 et 2 (valeurs propres de 0,70 et 0,60 pour des taux d'inertie respectifs de 5,8% et 5,0%) montre que deux lots de sites ressortent très différents de par leur flore de l'ensemble des autres (figure 1).

- un petit lot de 3 sites de très haute altitude (caractérisé par un cortège d'espèces arctico-alpines), La Grande Saume, le vallon du lac du Clou et celui du plan des Cavalles, classés dans les types tourbières et marais, tourbière et tourbière respectivement ; pour ces sites, la pertinence du classement est discutable, et surtout, l'originalité de la flore de ce type de zones humides d'altitude doit conduire à en faire une catégorie à part dans l'idée de rechercher des espèces indicatrices vraiment spécifiques et en petit nombre ;
- un lot plus important de 8 sites (Délaissés de la Peyrade, Etangs de Villepey, La palissade, Les Montilles, Marais de la Grande Palude, Mare de la Tour du Valat - La Cerisière moyenne, Marges nord-est de l'étang de l'Or, Zones humides de l'étang de Bages-Sigean), appartenant tous au type de zone humide « péri-lagunaire», avec ou sans apport d'eau, sauf le site de la Mare de la Tour du Valat (classé en ZH ponctuelles) ; réciproquement, la quasi-totalité de sites classés en péri-lagunaires en font partie.

Figure 1 : AFC globale sur l'ensemble des 165 sites ; les couleurs correspondent à des types de zones humides.

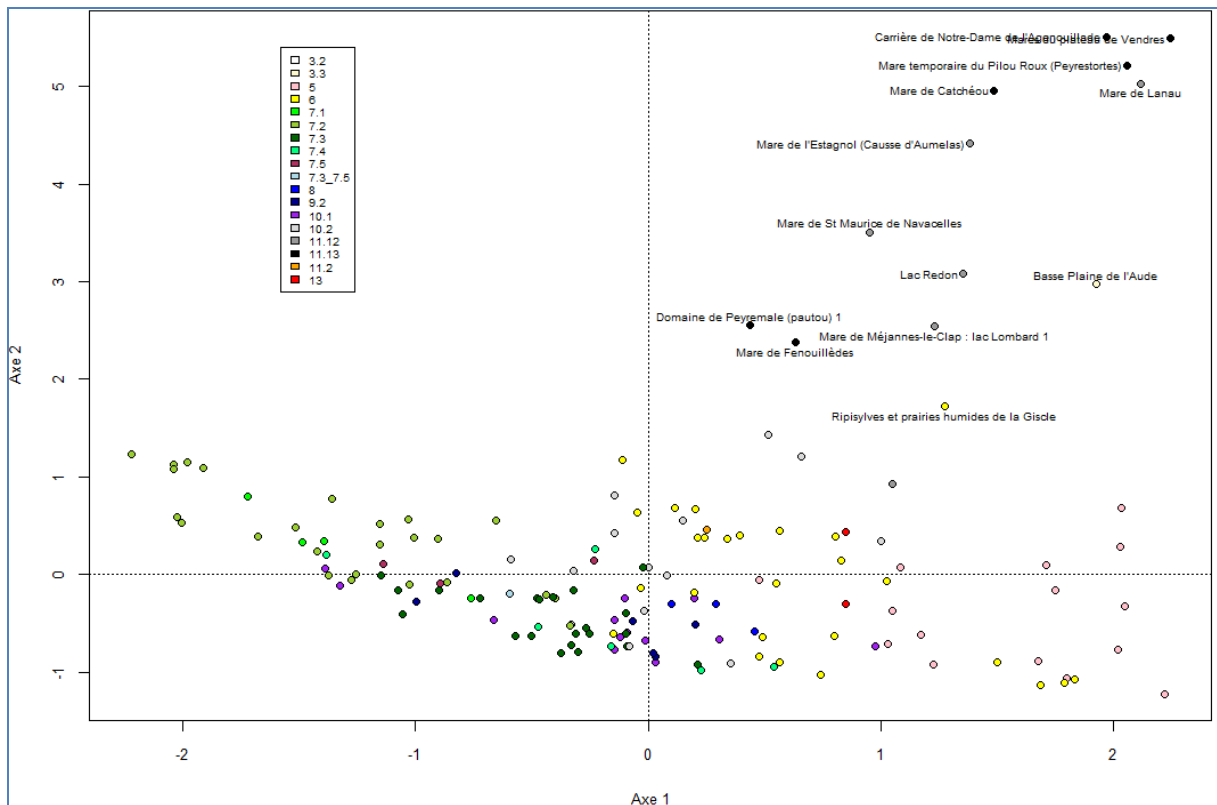


Une deuxième AFC a été menée sur les 154 sites restants après exclusion des 11 sites séparés par la première AFC, (figure 2), portant sur 1.506 taxons avant filtrage et 696 après. Les deux premiers axes (valeurs propres de 0.636 et 0.483 pour des taux d'inertie respectifs de 6,1 % et 4,6 %) sont seuls présentés. Un ensemble de 12 sites se dégage sur les axes 1-2, mais cette homogénéité est rompue sur les autres axes, le site de la basse plaine de l'Aude se détachant des 11 autres. On a donc un ensemble homogène de 11 sites, constitué de 8

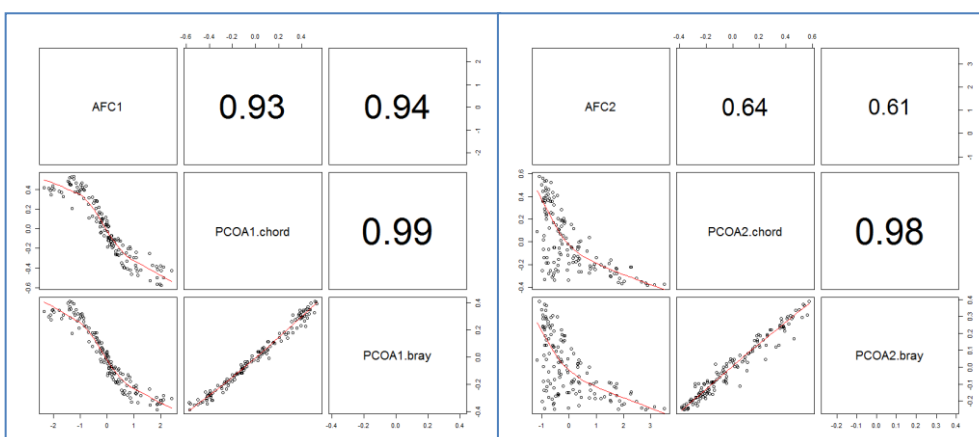
mares temporaires et de 3 sites classés comme zones humides ponctuelles. La quasi-totalité des mares temporaires (11.11 à 11.13) sont regroupées ici, sauf :

- l'étang salé de Courthézon, qui semble très différent des autres mares temporaires ;
- la mare de la Tour du Valat, écartée au vu de sa position lors de l'analyse précédente, qui est proche d'après sa flore des zones humides péri-lagunaires.

**Figure 2** : plan 1-2 de la deuxième AFC



Une troisième analyse a été donc conduite, portant sur les 143 sites restants pour 1.440 espèces avant filtrage et 666 après. L'AFC correspondante (non montrée) permet de mettre en évidence certains groupes homogènes mais les autres sites restent agglutinés vers le centre de l'axe 1. Des PCOA sur les distances de Bray-Curtis et de Chord ont permis un meilleur éclatement, sans remettre en cause la cohérence globale entre les analyses. La figure 3 donne les diagrammes de dispersion et les coefficients de corrélation entre les coordonnées des trois analyses sur les axes 1 et 2, respectivement.

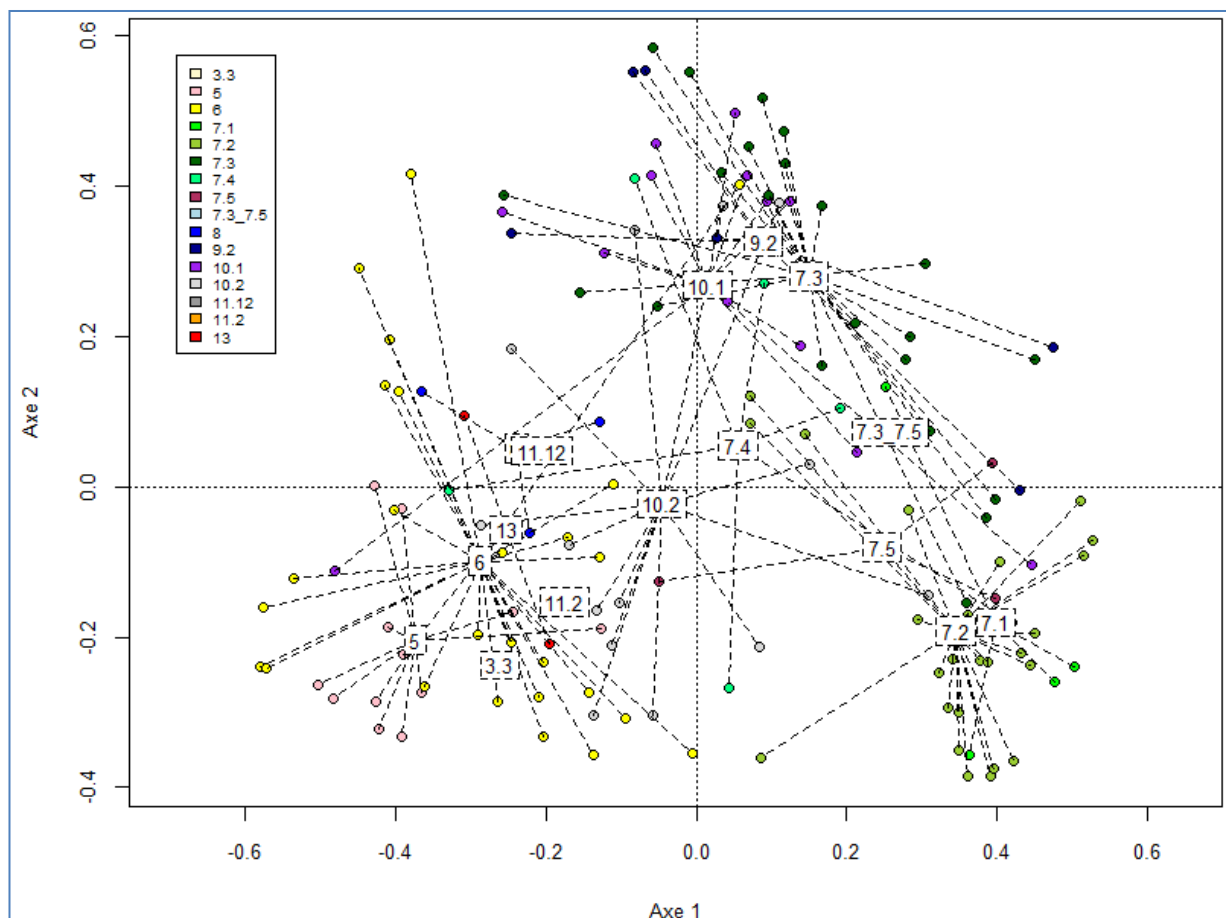


**Figure 3** : relations entre les coordonnées factorielles de l'AFC, la PCOA sur distance de Chord et la PCOA sur distance de Bray-Curtis et coefficient de corrélation, sur les axes 1 (à gauche) et 2 (à droite).

On constate que les deux PCOA donnent des résultats quasiment identiques sur les deux axes (corrélations de 0.99 et 0.98 respectivement), et que les deux PCOA sont également fortement corrélées, sur l'axe 1, avec l'AFC (0.93 et 0.94) ; en revanche, la corrélation est moins bonne sur l'axe 2 (0.64 et 0.61), ce qui s'explique par la tendance qu'a l'AFC à produire, notamment dans le cas de structures floristiques fortes, un deuxième axe qui est corrélé quadratiquement avec le premier, défaut dont est exempt la PCOA. Nous ne discuterons plus que les résultats issus de la PCOA sur distance de Chord (figure 4).

Sur le plan des axes 1-2 de la PCOA sur distance de Chord, une couleur différente est attribuée aux sites en fonction du type de zone humide auxquels ils ont été attribués ; le centroïde des sites de chaque type est indiqué par un texte encadré et les sites appartenant à ce type sont reliés par des traits en pointillés au centroïde de leur type.

**Figure 4** : axes 1 et 2 de la PCOA sur les 143 sites.



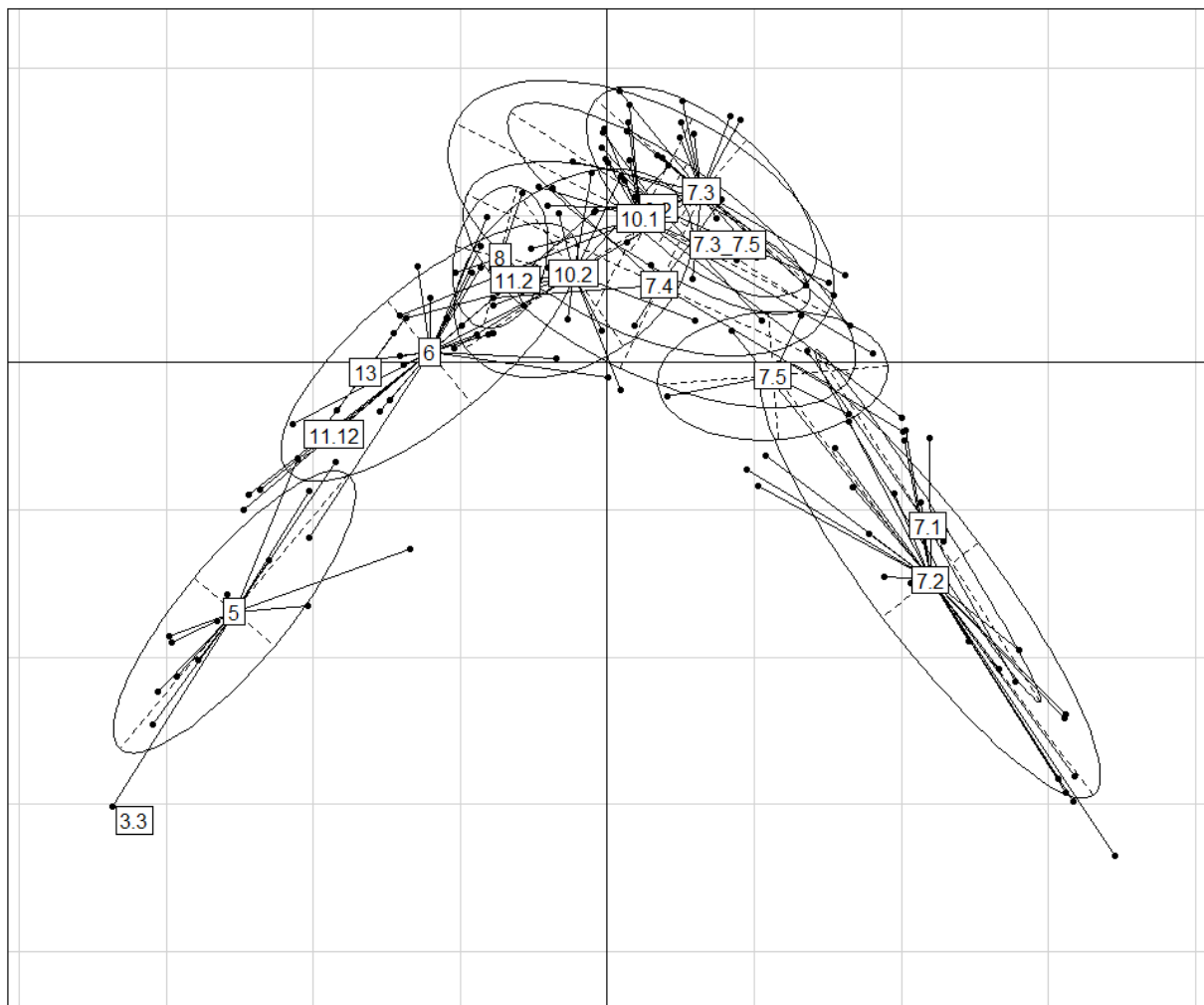
Notons au préalable que les types 3.3, 7.3\_7.5, 11.12, 11.2 et 13 ne sont représentés que par un ou deux sites ; nous n'interpréterons donc pas trop leur position. Les sites se répartissent le long de l'axe 1 depuis les sites alluviaux et de bords de mer (côté négatif) jusqu'aux tourbières. On a donc ici un gradient principal écologique d'engorgement et de fertilité (sites plutôt secs et/ou riches côté négatif de l'axe 1 versus sites très humides et oligotrophes côté positif). L'axe 2 oppose surtout les sites situés sur substrats calciques ou calcaires (côté positif de l'axe 2) aux substrats neutres à acides ou salés.

Quelques ensembles plus ou moins cohérents se dégagent. Côté négatif de l'axe 1, les sites des types 5 et 6 sont concentrés, même si la cohésion est plus faible pour les sites du type 6 (sur l'axe 1 et plus encore sur l'axe 2).

Côté positif de l'axe 1, les sites des types 7.1 et 7.2 forment un ensemble assez net (seuls 4 sites s'en écartent). Les sites du type 7.3 forment un nuage étendu sur l'axe 1, plutôt en position centrale, mais globalement tous du côté positif de l'axe 2. Les sites du type 7.4 (au nombre de 5) et 7.5 (au nombre de 3) sont situés globalement, comme les autres sites du type 7 au sens large, côté positif de l'axe 1, mais semblent assez hétérogènes. Les sites de type 8 sont proches des sites du type 6 (le centroïde est masqué par celui du type 11.12 sur la figure). Les sites de type 9.2 sont mêlés, côté positif de l'axe 2, aux sites de type 7.3 et 10.1. Les sites du type 10.1 et 10.2 forment deux nuages allongés le long de l'axe 2 mais en position centrale sur l'axe 1.

Les résultats de l'AFC inter-types de zones humides sont présentés figure 5.

**Figure 5** : résultats de l'AFC inter-types de zones humides ; projections des points-sites avec barycentres par type de zone humide et ellipses de confiance contenant environ les 2/3 des sites.



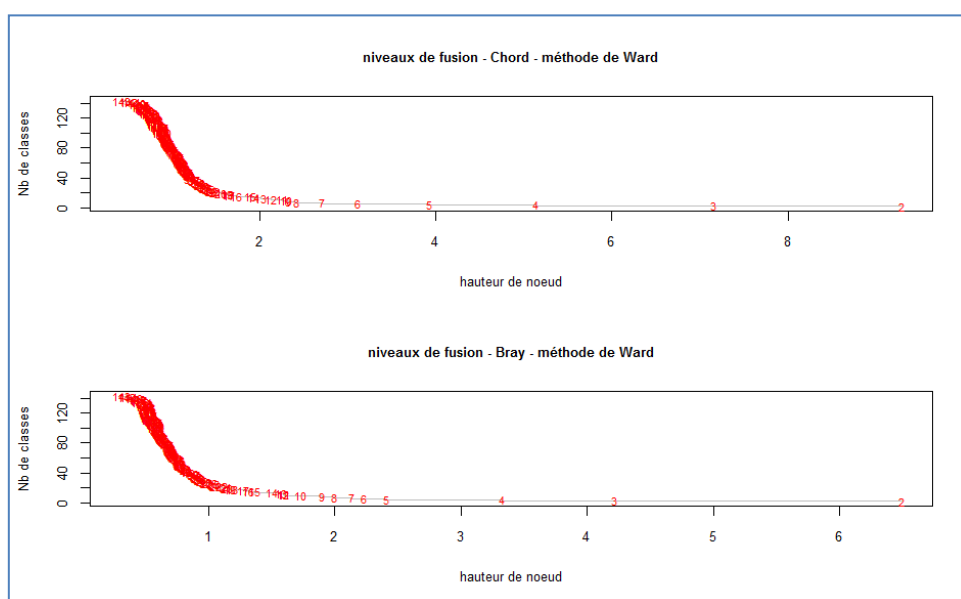
Les corrélations entre les axes 1 et 2 de l'analyse inter et ceux de l'AFC initiales sont très bonnes (résultats non montrés). On retrouve globalement la même structuration qu'avec la PCOA, à savoir la proximité des sites des types 7.1 et 7.2, celle, également quoique moins nette, entre les types 5 et 6, le regroupement possible des types 7.3, 7.4 et 7.5 auxquels on peut joindre le seul site de type intermédiaire 7.3\_7.5, et enfin la situation intermédiaire et peu claire des sites de type 10.1 et 10.2.

La structuration qui semble se dégager peut être affinée avec les résultats des classifications. Les résultats des critères évaluatifs des différentes classifications sont résumés tableau 2.

**Tableau 2** : évaluation des méthodes de classification et nombre de classes optimal.

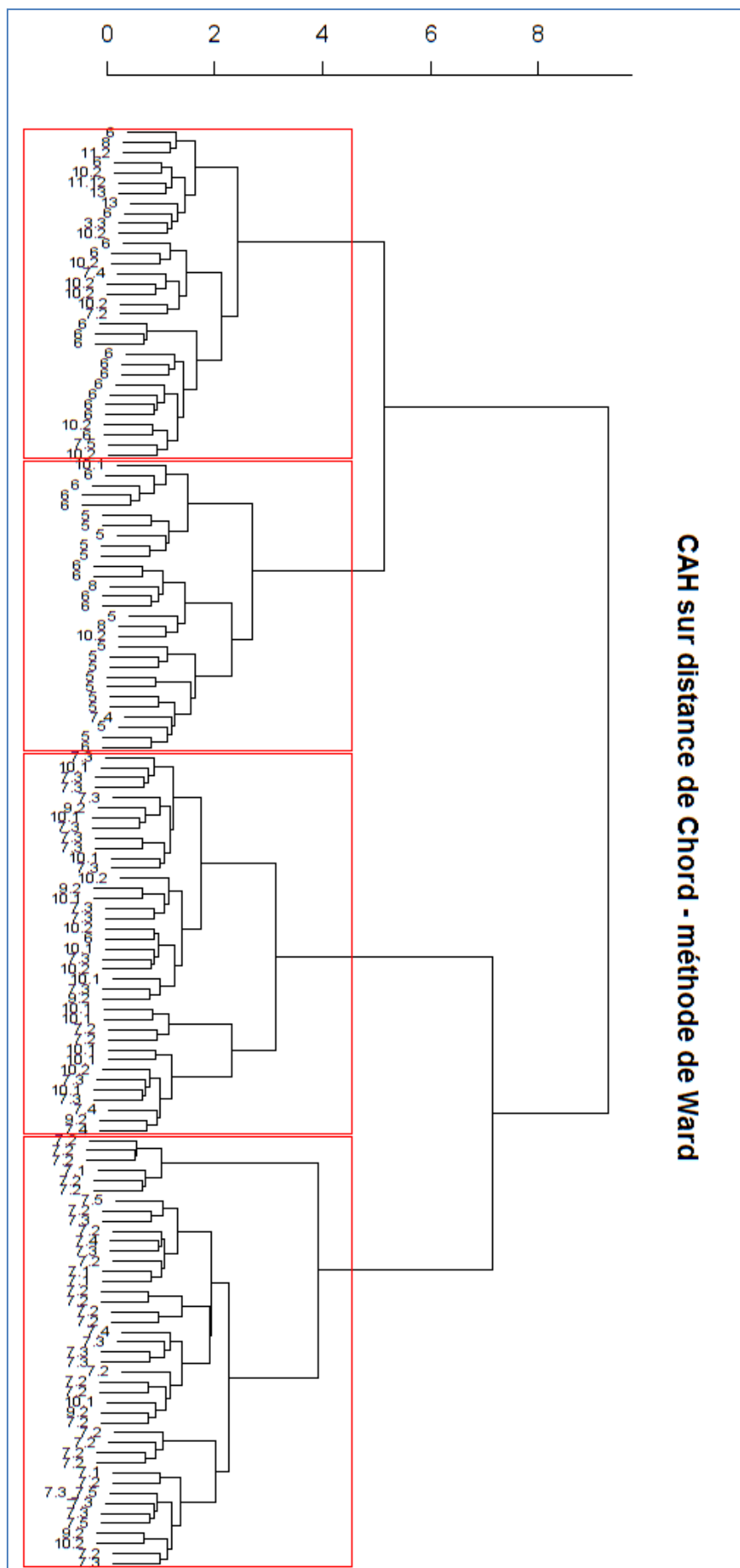
méthode	corrélation cophénétique	coefficient d'agglomération	nb de classes optimal - graphe de niveaux de fusion	nb de classes optimal – (silhouette moyenne)	nb de classes optimal – méthode PARTANA
CAH - distance de Chord - méthode de Ward	0,575	0,904	3 ; 4	5 (0,091)	8 ; 5 ; 4 (2,67)
CAH - distance de Bray - méthode de Ward	0,497	0,909	3 ; 5	2 (0,087)	8 ; 5 (2,20)
CAH - distance de Chord - association moyenne	0,76	0,340	5	3 (0,089)	3 à 7 (2,70)
CAH - distance de Bray – association moyenne	0,691	0,371	5	3 (0,085)	8 ; 7 ; 6 (2,19)
PAM - distance de Chord	-	-	-	4 (0,097)	4 ; 5 ; 8 (2,73)
PAM - distance de Bray	-	-	-	4 (0,090)	8 ; 6 ; 4 (2,19)
PARTANA – distance de Chord	-	-	-	4	5 (2,63)

Comme attendu, les corrélations cophénétiques sont les plus élevées pour l'association moyenne et surtout pour la distance de Chord. Les coefficients d'agglomérations sont les plus forts pour la méthode de Ward, sans différence notable entre les deux distances. Sur le critère de silhouette, les distances de Chord présentent des valeurs également plus élevées que celles obtenues avec l'association moyenne (0,091 et 0,089 contre 0,087 et 0,085). La valeur de silhouette est également plus élevée pour la distance de Chord que celle de Bray avec la classification PAM. La valeur du critère silhouette est la plus élevée sur l'ensemble des méthodes testées pour la PAM sur distance de Chord, puis en second vient la CAH sur distance de Chord avec la méthode de Ward. L'évaluation avec PARTANA donne la préférence à la distance de Chord et la méthode PAM (rapport de similarité intra-classe sur inter-classe de 2,73), suivie par la CAH sur distance de Chord avec association moyenne (rapport de 2,70) puis avec la méthode de Ward (rapport de 2,67). Le nombre de classes optimal oscille entre trois et cinq, sauf avec PARTANA où ce nombre est souvent de 8 mais avec des valeurs pour 4 ou 5 classes assez proches. A titre d'illustration, les graphes de niveaux de fusion sont présentés pour la méthode de Ward figure 6.



**Figure 6** : graphes des niveaux de fusion pour la méthode de Ward avec la distance de Chord ou la dissimilarité de Bray-Curtis.

Au vu des différents éléments, une structuration globale en 4 grandes classes, basée sur la distance de Chord et la méthode de Ward (figure 7) ou la partition PAM sur distance de Chord en 4 classes est retenue.



**Figure 7** : arbre de classification hiérarchique ascendante sur distance de Chord et méthode de Ward ; les types de zone humide des sites sont indiqués à l'extrémité des branches ; les 4 rectangles rouges matérialisent les blocs issus de la coupure de l'arbre en 4.

Le tableau 3 donne la ventilation croisée par classes pour les deux meilleures solutions. On constate que la très grande majorité des sites classés dans un groupe d'une des deux solutions l'est dans un autre groupe de l'autre solution (l'inversion des groupes 3 et 4 n'a aucune importance car les classes issues de la PAM peuvent être numérotées comme on le souhaite car elles ne reflètent pas de structure hiérarchique) ; **en d'autres termes, la cohésion des deux classifications est bonne.**

**Tableau 3** : table croisée du nombre de sites classés dans les classes par une méthode (Chord et Ward en lignes) et dans les classes par l'autre méthode (Chord et PAM en colonnes).

		Chord - PAM			
		1	2	3	4
Chord - Ward	1	42	1		
	2	5	33		
	3	1	4	4	24
	4		6	20	3

Nous avons également calculé les ventilations des sites par type de zone humide dans les groupes issus des 2 meilleures classifications (tableau 4).

**Tableau 4** : répartition croisée des sites dans les groupes issues de la CAH sur distance de Chord avec la méthode de Ward et avec la PAM sur distance de Chord en fonction des types de zones humides (en lignes). Les cases encadrées en rouges correspondent à celles où les sites d'un type de zone humide sont presque entièrement concentrés.

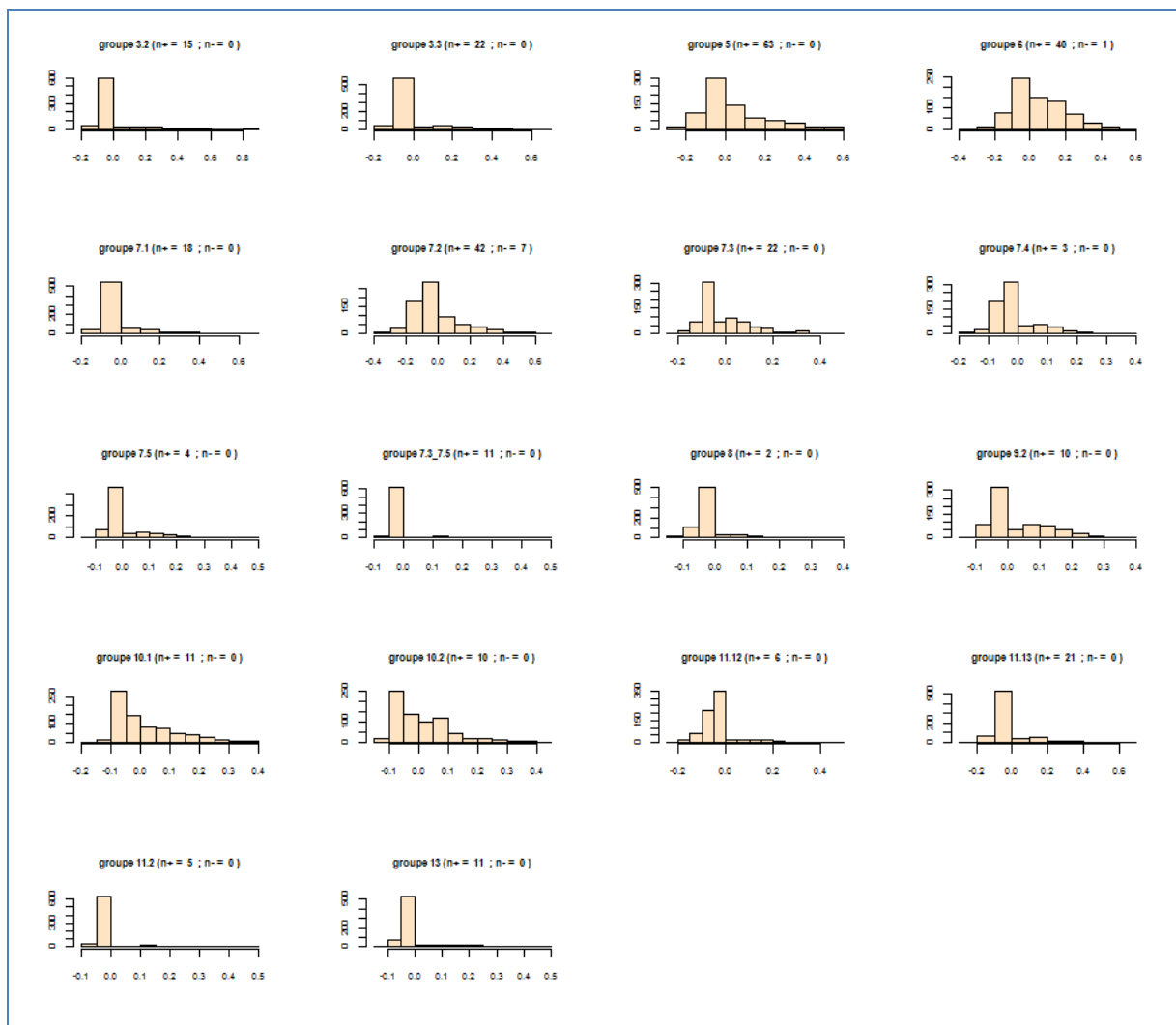
		Groupes Chord – CAH Ward				Groupes Chord - PAM			
		1	2	3	4	1	2	3	4
Type de ZH	3.3			1				1	
	5				15			13	2
	6		1	16	9		6	7	13
	7.1	4				4			
	7.2	22	2	1		23	1		1
	7.3	8	14			8	14		
	7.4	2	2	1	1	2	2	1	1
	7.5	2		1		2			1
	7.3_7.5	1				1			
	8			1	2		2		1
	9.2	2	4			2	4		
	10.1	1	11		1	3	9	1	
	10.2	1	4	8	1	3	4	1	6
	11.12			1			1		
	11.2			1					1
13			2			1		1	

On constate que la cohésion globale entre groupes et types de zones humides est légèrement meilleure avec la CAH qu'avec la PAM ; nous ne détaillerons plus que les résultats issus de la première méthode de classification.

### 3.4 Proposition d'une typologie synthétique

Au vu des différentes analyses, et en prenant en compte également les données disponibles et la connaissance des sites et de la flore associée, nous discuterons chaque type de zone humide et proposerons soit de les garder tels quels, soit de les fusionner ensemble, soit de ne pas les prendre en compte dans les analyses suivantes. Une dernière aide au diagnostic est donnée figure 8. Il s'agit des histogrammes des valeurs des coefficients Phi entre les espèces présentes dans un type de zone humide et ce type, ainsi que du nombre d'espèces positivement liées au type (arbitrairement fixé à  $\Phi \geq 0.3$ ) ou négativement liées au type (arbitrairement fixé à  $\Phi \leq 0.3$ ).

**Figure 8** : histogramme des valeurs du coefficient Phi d'association entre les espèces et les types de zones humides et nombres d'espèces positivement liées (n+) ou négativement liées (n-).



Les sites de type 3.2 (*Péri-lagunaire*) et 3.3 (*Péri-lagunaire avec apport d'eau*) sortent assez bien groupés lors de la deuxième AFC, à l'exception d'un site. Les deux types ne sont pas séparés dans l'analyse. De plus, peu de données étant disponibles, il ne parait pas fiable d'analyser les types initiaux séparément. **Nous regrouperons donc les types 3.2 et 3.3.**

Les sites du type 5 (*Bordures de cours d'eau*) sont concentrés dans le groupe 4 de la CAH, même si un tiers des sites du type 6 (*Plaines alluviales [inondable]*) est classé dans ce groupe. La parenté entre ces deux types est

évidente sur le plan 1-2 de la PCOA (figure 4), un peu moins sur le plan de l'AFC inter-classes (figure 5). Dans le détail, les compositions floristiques de ces deux types montrent :

- que le type 5 est caractérisé par des fréquences plus élevées d'un ensemble d'espèces méridionales ou à répartition actuelle méridionale (*Laurus nobilis*, *Osyris alba*, *Celtis australis*, *Scirpoides holoschoenus*, *Smilax aspera*, *Amorpha fruticosa*) ;
- que le type 6 est lui caractérisé par des espèces plus septentrionales, qui sans être nécessairement absentes du sud du bassin versant, y sont soit plus rares soit manquent dans ces types de zones humides (*Carex acuta*, *Carex acutiformis*, *Thalictrum flavum*, ...) ; ces différences de composition sont corroborées par la localisation des sites, ceux du type 5 étant surtout dans la moitié sud du bassin (13 des 15 sites analysés), ceux du type 6 étant surtout dans la moitié nord du bassin (19 sites sur 26 analysés) ;
- mais avant tout, que pour un lot assez important d'espèces, les fréquences sont peu différentes dans les deux types.

Notons également que ces deux types de zones humides sont peu homogènes en leur sein (surtout le type 6), comme en attestent la grande dispersion des points sur la PCOA (figure 4), l'ellipse contenant 65% des sites sur l'AFC inter-classes (figure 5) ou encore la scission des sites entre deux ou trois groupes dans les deux classifications retenues (tableau 4) ; enfin, le rattachement à un type ou à l'autre est parfois ardu et assez arbitraire. **Pour ces raisons (différences peu nettes et hétérogénéité du type 6, classement problématique), nous fusionnerons les types 5 et 6.**

Le type 7.2 (*zones humides de bas-fonds en tête de BV : tourbières acides*) est quasi exclusif du groupe 1, mais là aussi environ un tiers des sites du type 7.3 (*Zones humides de bas-fonds en tête de BV : tourbières alcalines*) sont classés avec ce groupe. Toutefois, les sites du type 7.3 sont nettement distincts sur les analyses et les conditions édaphiques assez différentes (acides versus alcalines) les deux types possèdent de nombreuses espèces fidèles (figure 8) ; **nous garderons donc ces types, 7.2 et 7.3, distincts.**

Les sites du groupe 7.1 (*Zones humides de bas-fonds en tête de BV : altitude*) sont aussi classés avec le type 7.2 (*Zones humides de bas-fonds en tête de BV : tourbières acides*) dans cette analyse, mais les autres sites du type 7.1 se sont avérés nettement différents et ont été exclus après la première AFC (vallon du lac du Clou et du plan des Cavales). Deux solutions sont donc possibles : soit ne pas traiter ce groupe, soit le fusionner avec le type 7.2, à l'exclusion des deux sites mentionnés ci-dessus. Rappelons qu'un indicateur spécifique avait été proposé pour les bas-marais artico-alpins mais qu'il n'avait finalement pas été retenu. **Nous garderons donc le type 7.1 à part.**

La situation des sites de type 7.4 (*Zones humides de bas-fonds en tête de BV : pentes et sources*), 7.5 (*Zones humides de bas-fonds en tête de BV : combes et bordure de ruisseau*) et 7.3\_7.5 (*intermédiaire entre 7.3 et 7.5*) est assez confuse, mais ils sont globalement situés dans les analyses entre les types 7.2 et 7.3. Leur composition « interne » ne semble pas très homogène (par exemple, il n'y a quasiment pas d'espèces fidèles (figure 8)), mais ils forment un ensemble cohérent dans les analyses. Comme très peu de sites sont concernés, il ne paraît pas possible de les maintenir séparés, sauf à ne plus les analyser. **Nous fusionnerons ces types ensemble.**

Les sites de type 8 (*Régions d'étangs*), sont très peu représentés dans l'échantillon (3 sites). Ils sont proches dans la PCOA du type 6 (*Plaines alluviales [inondable]*) et leurs sites sont répartis dans les groupes de manière peu nette ; en outre, il n'y a quasiment aucunes caractéristiques de ce type (figure 8). **Nous écarterons ce type des analyses ultérieures.**

Les sites de type 9.2 (*Bordures de plan d'eau (lac) : ZH alcaline*) sont plutôt regroupés dans les classifications dans le groupe 2, avec les sites de types 7.3 (*Zones humides de bas-fonds en tête de BV : tourbières alcalines*) et

10.1 (*Marais et landes humides de plaine : tourbières de plaine*), les autres sites étant dans le groupe 1, avec d'autres sites du type 7.3. Par ailleurs, les sites de ce type sont en général essentiellement des bas-marais tourbeux alcalins dans lesquels les communautés des berges qui auraient pu donner une originalité à ce type n'existent pas ou peu. **Nous fusionnerons le type 9.2 avec le type 7.3.**

La situation des sites de type 10.1 et 10.2 (*marais et landes humides de plaine, tourbières et prairies* respectivement) est confuse sur les analyses car les sites concernés semblent avoir des compositions floristiques variables. Toutefois, un positionnement central sur l'axe 1 de la PCOA nous incite à proposer de les **conserver mais en les regroupant.**

Les sites des types 11.11, 11.12 et 11.13 ont tous été distingués lors de la deuxième (1 site) ou la troisième AFC (11 sites), seul le site de l'étang salé de Courthézon semblant affiner d'autres types. Il n'y a pas de discrimination totalement nette entre ces types sur l'AFC de la figure 3, les sites étant assez éclatés au sein d'un même type (la différence inter-sites est élevée). De plus, il y a relativement peu de sites par type. **Pour ces raisons, nous proposons de regrouper ces trois types lors des analyses ultérieures.**

Les sites de type 11.2 (*zones humides ponctuelles : mare permanente*) est trop peu représenté dans l'échantillon (4 sites dont 1 seul avec suffisamment de données). **Nous proposons de l'écartier des analyses ultérieures.**

Les sites de type 13 (*Zones humides artificielles*) sont très peu nombreux (2). Ils se situent dans la région des sites de type 6 sur les analyses factorielles mais leur positionnement n'est pas clair dans les classifications. **Nous ne les traiterons donc pas dans les analyses ultérieures.**

Le tableau 5 donne la nouvelle ventilation des sites par catégories synthétiques et différentes statistiques générales après filtrages sur le nombre de données par site (supérieur à 50). Seuls 5 types sont véritablement bien échantillonnés (5\_6, 7.2, 7.3\_9.2 et 10) mais nous avons tenu à conserver les autres, fussent-ils représentés que par un seul site afin d'avoir une vision assez large des espèces que l'on pourrait retenir.

**Tableau 5 :** regroupement des types de sites et statistiques descriptives de ces nouveaux types.

Code ZH	Type(s) ZH RhoméO	Nb. sites	Nb. relevés	Nb. obs.	Nb. taxons
3	3.2 (Péri-lagunaire) et 3.3 (Péri-lagunaire avec apport d'eau)	1	62	449	108
5_6	5 (Bordures de cours d'eau) et 6 (Plaines alluviales [inondable])	41	1.047	15.476	972
7.1	Zones humides de bas-fonds en tête de BV : altitude	4	57	970	200
7.2	Zones humides de bas-fonds en tête de BV : tourbières acides	25	561	7.954	501
7.3_9.2	7.3 (Zones humides de bas-fonds en tête de BV : tourbières alcalines) et 9.2 (Bordures de plan d'eau (lac) : ZH alcaline)	28	546	9.723	539
7.autres	7.4 (Zones humides de bas-fonds en tête de BV : pentes et sources), 7.5 (Zones humides de bas-fonds en tête de BV : combes et bordure de ruisseau) et 7.3_7.5	10	137	2.685	407
10	10.1 (Marais et landes humides de plaine : tourbières de plaine) et 10.2 (Marais et landes humides de plaine : prairies humides)	27	495	9.516	655
11.autres	11.11 (Zones humides ponctuelles : mares temporaires saumâtres), 11.12 (Zones humides ponctuelles : mares temporaires alcalines) et 11.13 (Zones humides ponctuelles : mares temporaires acides)	1	32	153	60

## 4 Réduction des listes d'espèces

---

Rappelons que l'objectif idéal est d'essayer de reproduire au mieux avec des listes fermées d'au plus 100 espèces par type de zone humide (au sens de notre typologie simplifiée présentée ci-dessus) permettant de reproduire au mieux les résultats obtenus avec le jeu des données complètes. Par jeu des données complètes, nous entendons retenir les relevés pour lesquels les données sont renseignées pour l'abondance et pour les valeurs indicatrices. La question de « reproduire au mieux » n'est également pas évidente mais un des outils pour l'ajuster est le nombre de taxons réellement retenus. Si l'objectif est d'une centaine d'espèces par type de zone humide, il semble nécessaire de tester plusieurs autres valeurs pour évaluer l'impact de ces valeurs sur la perte de précision des estimations ; nous essayerons donc, en fonction des pistes testées, de sélectionner des listes de 50, 100 ou 150 taxons.

Les analyses ultérieures portent sur les types de zones humide retenues à la suite de l'analyse précédente, en ne retenant que les sites considérés comme suffisamment bien connus (au moins 50 observations par site), et en ne retenant que les observations avec recouvrement, soit 137 sites pour 46.926 observations et 1.420 taxons.

### 4.1 Méthodes de filtrages possibles

---

Plusieurs méthodes ont été envisagées pour la réduction du nombre d'espèces ; elles sont listées ci-dessous. Nous en mettrons en œuvre quelques-unes dans les paragraphes suivants. Les méthodes que nous avons identifiées consistent à :

- filtrer en ne retenant que les espèces figurant à l'arrêté ZH de 2008 ;
- filtrer sur le caractère plus ou moins spécialiste des espèces (sténioïcité) en supprimant les espèces à large tolérance ;
- filtrer sur la fréquence des espèces en supprimant les espèces les plus rares (avec trois seuils) ;
- filtrer sur la difficulté de détermination des espèces en supprimant celles jugées difficiles ;
- rechercher par des méthodes statistiques (corrélation, fidélité, co-inertie, ...) les espèces permettant de reconstituer au mieux les données ;
- de manière plus ou moins transversale aux points précédents, tester l'impact du recouvrement des espèces versus la seule présence.

### 4.2 Filtrer pour ne retenir que les espèces ZH au sens de l'arrêté de 2008

---

Sur les données totales, pour tous les sites suffisamment connus, on dispose de 24.208 observations de taxons inscrits sur l'arrêté de définition des zones humides de 2008 sur les 52.2254 données correspondantes, pour 389 taxons. La ventilation du nombre d'observations et de taxons par type de zone humide pour notre jeu de données test est fournie tableau 6. Un point intéressant à noter est que le nombre de taxons « ZH » est globalement peu élevé par type de zone humide, dépassant de peu au maximum 200, ce qui implique que peu d'espèces seraient à filtrer encore. La réduction du nombre d'espèces par rapport au pool total par type de zone humide est typiquement compris entre un facteur 2 à un facteur 3, hormis pour les types les plus pauvres (car souvent aussi les moins échantillonnés) ou cette réduction est beaucoup plus faible.

**Tableau 6** : nombres d'observations et de taxons par type de zone humide globalement et valeurs correspondantes pour les observations et taxons inscrits sur l'arrêté zone humide.

Code ZH	Type(s) ZH RhoméO	Nb. obs	Nb. obs taxons ZH	Nb. taxons	Nb. Taxons ZH
3	3.2 (Péri-lagunaire) et 3.3 (Péri-lagunaire avec apport d'eau)	449	81	109	26
5_6	5 (Bordures de cours d'eau) et 6 (Plaines alluviales [inondable])	15.476	5.185	972	232
7.1	Zones humides de bas-fonds en tête de BV : altitude	970	575	245	92
7.2	Zones humides de bas-fonds en tête de BV : tourbières acides	7.954	4.186	501	181
7.3_9.2	7.3 (Zones humides de bas-fonds en tête de BV : tourbières alcalines) et 9.2 (Bordures de plan d'eau (lac) : ZH alcaline)	9.723	6.333	539	214
7.autres	7.4 (Zones humides de bas-fonds en tête de BV : pentes et sources), 7.5 (Zones humides de bas-fonds en tête de BV : combes et bordure de ruisseau) et 7.3_7.5	2.685	1.508	412	135
10	10.1 (Marais et landes humides de plaine : tourbières de plaine) et 10.2 (Marais et landes humides de plaine : prairies humides)	9.516	4.773	655	231
11.autres	11.11 (Zones humides ponctuelles : mares temporaires saumâtres), 11.12 (Zones humides ponctuelles : mares temporaires alcalines) et 11.13 (Zones humides ponctuelles : mares temporaires acides)	153	79	299	18

La difficulté principale avec cette approche est qu'elle écarte par principe les espèces mésophiles qui témoignent d'une altération du niveau de nappe. Le tableau 7 ci-dessous illustre, pour un exemple fictif, les résultats des calculs faits en ne prenant en compte que les espèces ZH ou toutes les espèces dans le cas d'une réduction importante du niveau d'engorgement entre les deux dates des suivis, dans le cas où seuls des changements des recouvrements des espèces interviennent.

**Tableau 7** : exemple fictif de calcul.

	ZH	Valeur indicatrice d'engorgement	Recouvrement 2010	Recouvrement 2015
Espèce 1	oui	8	25	1
Espèce 2	oui	6	15	1
Espèce 3	oui	7	1	1
Espèce 4	non	3	10	40
Espèce 5	non	3	20	30
Recouvrement total			71	73
Recouvrement espèces ZH			41	3
I2 sur totalité			5,45	3,16
I2 sur espèces ZH			7,20	7,00
I2 sur ZH pondéré			4,15	0,28

Malgré une réduction drastique des trois espèces ZH, la valeur d'engorgement calculée uniquement avec ces espèces ne diminue quasiment pas (0,20), alors qu'avec le cortège total, cette diminution est forte (2,21). Pour pallier à ceci, on pourrait envisager de pondérer la note par le rapport des recouvrements entre espèces ZH sur le recouvrement total ; les résultats (ligne « I2 sur ZH pondéré ») montrent une chute très importante de la

valeur de l'indicateur (passage de 4,15 à 0,28 soit une variation de 3,87) ; cette procédure pourrait donc être appliquée mais la significativité des évolutions reste à étudier dans le détail. Toutefois, si on ne retient plus qu'une partie des espèces ZH (pour diminuer le nombre d'espèces à noter sur le terrain), cette procédure devient fautive car le recouvrement des espèces ZH est minoré.

Une autre difficulté, majeure et sans solution, elle, est que cette approche n'est pas applicable à l'indicateur I6 (fertilité du sol). **Pour ces raisons, cette approche nous semble difficilement applicable.**

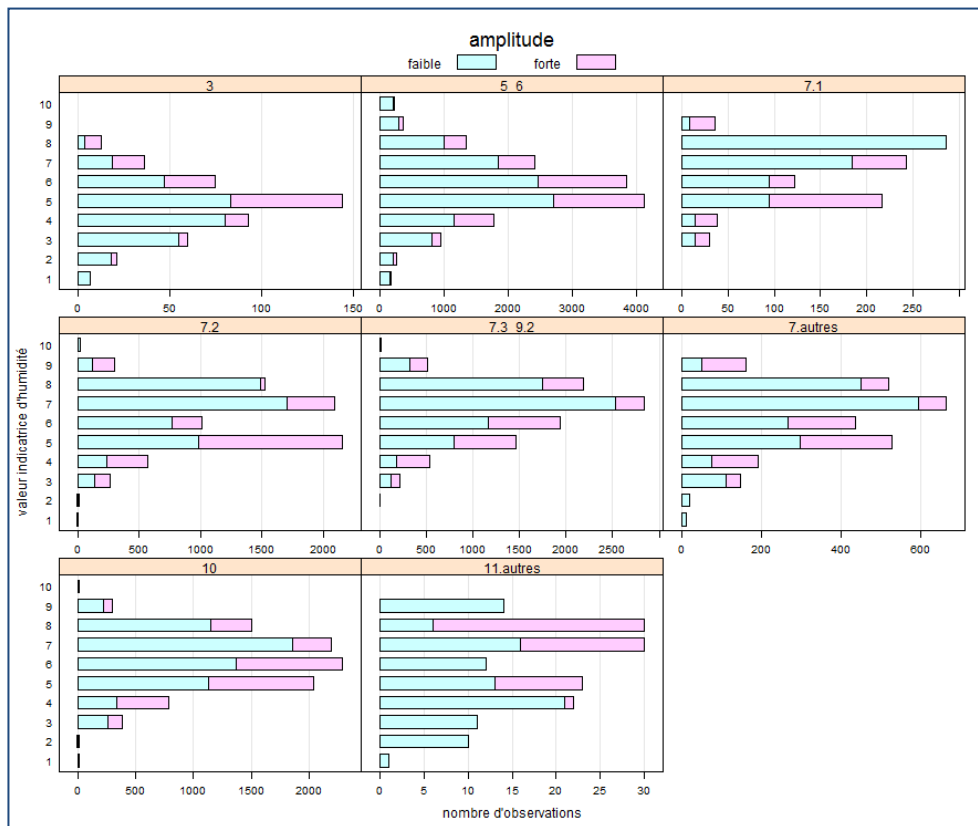
### 4.3 Filtrer les espèces à large amplitude

Les valeurs indicatrices de LANDOLT (établies pour la flore de la Suisse) pour le niveau d'engorgement et de fertilité du sol sont accompagnées, en partie, d'une évaluation de leur amplitude écologique par rapport à ces deux facteurs. On qualifie de sténoèce une espèce qui présente par rapport à un facteur écologique une amplitude faible. On pourrait donc envisager de filtrer les espèces en supprimant celles à large amplitude. Le tableau 8 donne, par type de ZH des statistiques générales. La réduction des nombres d'observations est comprise entre 33% et 16,5% en moyenne pour le niveau d'engorgement et le niveau de fertilité, respectivement. On constate que la réduction du nombre d'espèces est également beaucoup plus faible, tant pour le niveau d'engorgement que pour le niveau de fertilité ; elle est typiquement de 25% pour le premier, et seulement de 15% pour le second. Dans les deux cas (observations et taxons), la réduction est beaucoup plus nette pour les zones humides de type 3 car beaucoup d'espèces de ces milieux ont été renseignées avec les valeurs indicatrices de JULVE (car absentes de Suisse) qui ne donne pas d'évaluation du caractère plus ou moins sténoèce des espèces ; en conséquence, beaucoup ne sont pas renseignées.

**Tableau 8** : nombres d'observations, de taxons, d'observations et de taxons sténoèces par type de zone humide, pour l'engorgement (H) et pour la fertilité (N).

Code ZH	Type(s) ZH RhoméO	Nb. obs	Nb.obs sténoèces H	Nb.obs. sténoèces N	Nb. taxons	Nb. taxons sténocès H	Nb. taxons sténocès N
3	3.2 (Péri-lagunaire) et 3.3 (Péri-lagunaire avec apport d'eau)	449	313	347	108	80	85
5_6	5 (Bordures de cours d'eau) et 6 (Plaines alluviales [inondable])	15.476	10.873	12.105	972	805	863
7.1	Zones humides de bas-fonds en tête de BV : altitude	970	698	870	200	146	174
7.2	Zones humides de bas-fonds en tête de BV : tourbières acides	7.954	5.489	7.095	501	380	443
7.3_9.2	7.3 (Zones humides de bas-fonds en tête de BV : tourbières alcalines) et 9.2 (Bordures de plan d'eau (lac) : ZH alcaline)	9.723	6.887	8.092	539	418	474
7.autres	7.4 (Zones humides de bas-fonds en tête de BV : pentes et sources), 7.5 (Zones humides de bas-fonds en tête de BV : combes et bordure de ruisseau) et 7.3_7.5	2.685	1.876	2.269	407	320	365
10	10.1 (Marais et landes humides de plaine : tourbières de plaine) et 10.2 (Marais et landes humides de plaine : prairies humides)	9.516	6.360	7.850	655	506	683
11.autres	11.11 (Zones humides ponctuelles : mares temporaires saumâtres), 11.12 (Zones humides ponctuelles : mares temporaires alcalines) et 11.13 (Zones humides ponctuelles : mares temporaires acides)	153	104	92	60	52	50

Une autre difficulté potentielle de cette approche est le biais de sélection qui pourrait en résulter si les espèces sténoèces n'ont globalement pas la même distribution d'abondance par valeur indicatrice : par exemple, si les occurrences d'espèces sténoèces ayant une valeur indicatrice de 8 (espèces assez nettement hygrophiles) sont en proportion plus faibles que les occurrences de toutes les espèces ayant une valeur indicatrice de 8, l'application du filtre sur les sténoèces pourrait entraîner une surreprésentation de celles-ci dans les données filtrées et conduire à une estimation par excès des valeurs d'humidité édaphique. Ceci est illustré figure 9.



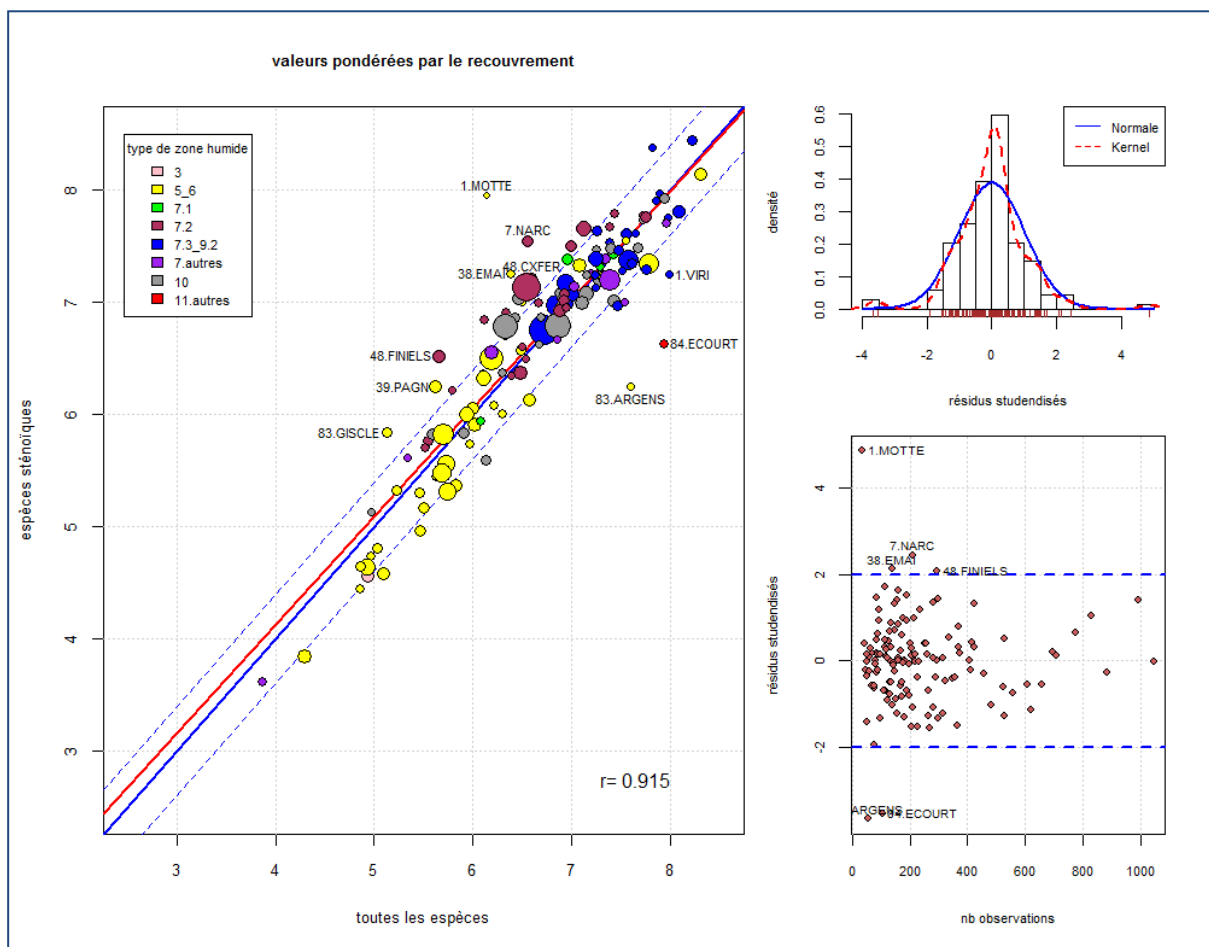
**Figure 9 :** nombre d'occurrences de taxons (en abscisses) par valeur indicatrice d'engorgement (en ordonnées, de 1 à 10), pour chaque type de zone humide, pour les espèces sténoèces (en bleu) et les espèces à large amplitude (en rose); le filtre sur l'amplitude fait disparaître les observations en rose.

Pour les tourbières acides (7.2) et les marais (10), ce sont surtout les valeurs indicatrices moyennes et faibles qui sont filtrées ; on aura donc une surestimation du niveau d'engorgement. Cette tendance existe également pour les tourbières alcalines (7.3\_9.2) les sources et suintements (7.autres), mais de manière moins nette. A l'opposé, ce sont surtout les valeurs indicatrices élevées qui sont filtrées pour les types 3 (marais salés) et 11.autres (zones humides temporaires) mais ceci ne concerne que très peu de données. Pour les milieux alluviaux, aucune tendance nette ne se dégage.

Pour aller plus loin, il faut descendre à l'échelle des sites. Nous avons étudié la relation entre les valeurs calculées avec l'ensemble des espèces et avec les espèces sténoèces avec un modèle linéaire simple. Les coefficients de corrélation linéaire de Pearson sont indiqués, mais comme les données filtrées et non filtrées ne sont pas indépendantes, les valeurs indiquées sont à prendre comme des ordres de grandeur descriptifs ; nous n'avons notamment pas testé leur significativité. Nous avons également fait figurer sur les graphiques des lignes matérialisant l'égalité des valeurs filtrées - non filtrées (ligne pleine bleue), et deux autres lignes à  $\pm 0.4$  en dessus ou dessous de cette ligne d'égalité, représentant un « intervalle de confiance empirique », encapsulant les erreurs liées à l'effet opérateur, à la mauvaise relocalisation des placettes., et que pour l'interprétation des tendances, les écarts entre dates constatés doivent (souvent) être supérieurs à  $\pm 0.4$ . Nous avons rajouté quelques éléments de diagnostics des régressions, comme l'histogramme des résidus (en haut à droite) qui doit être plus ou moins normal, et la relation entre les résidus studentisés et le nombre d'observations disponibles par site après filtrage. Les résidus studentisés, au sens de la fonction 'rstudent' de R

sont les écarts entre les valeurs prédites par le modèle et les valeurs vraies, divisés par les produits des écarts-types des résidus par les effets levier (c'est l'effet qu'a une observation sur la pente de la droite de régression). Les résidus studentisés suivent alors une loi de Student et un modèle linéaire peut être jugé correct si moins de 5% des sites sont situés à  $\pm 2$  ; ces lignes de  $\pm 2$  sont matérialisées sur les graphique du bas à droite par les droites en tirets bleus. Les sites situés à plus de  $\pm 2$  sont donc des sites qui sont soit mal prédits, soit exercent une forte influence sur la droite de régression, soit les deux. Nous avons produit les graphiques de ces résidus studentisés en fonction du nombre d'observations car il semblait probable *a priori* que ce nombre influence la qualité des prédictions.

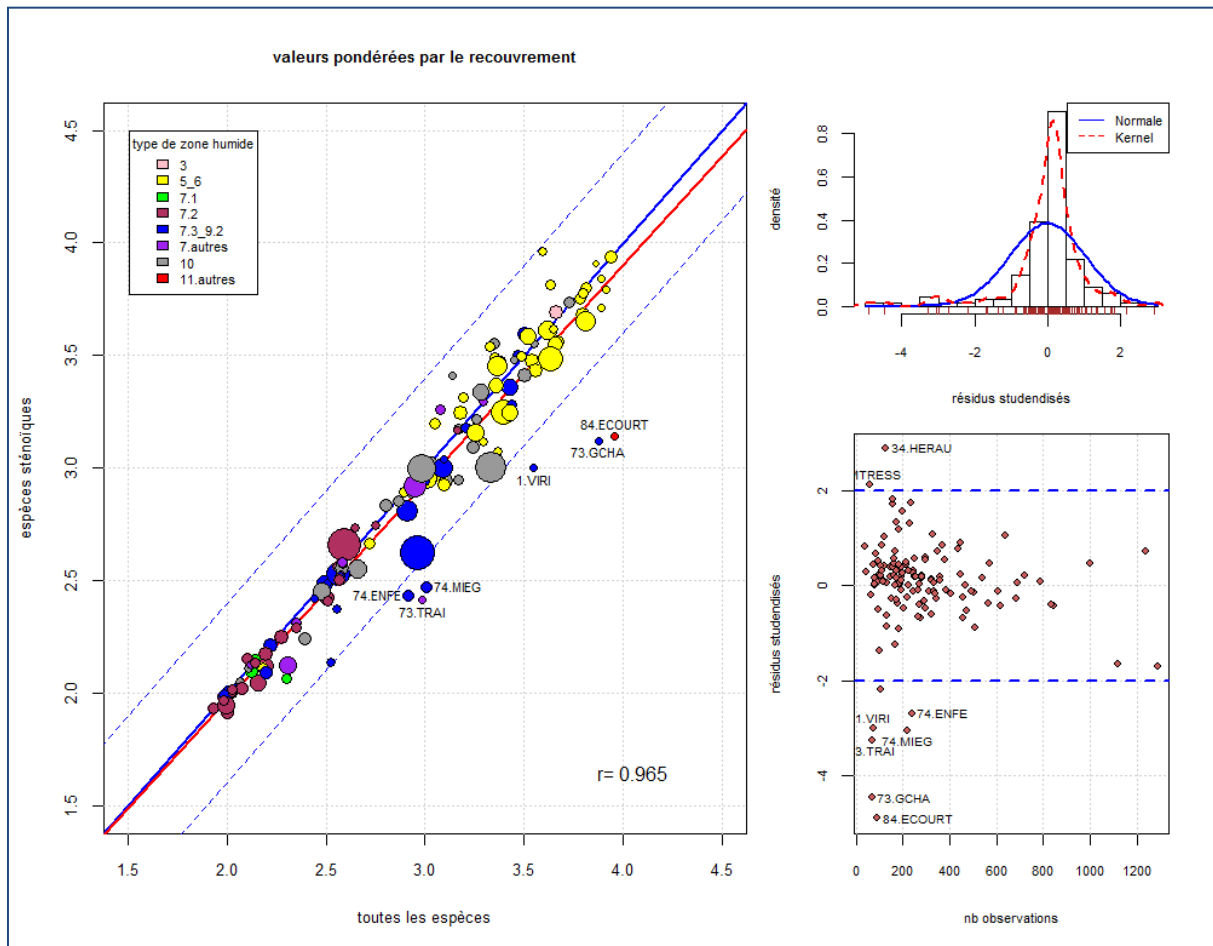
**Figure 10** : relation entre les valeurs observées pour l'engorgement avec l'ensemble des données sur chaque site (en abscisses) et les valeurs prédites avec les seules données des espèces sténoïques (en ordonnées) ; la taille des points est proportionnelle au nombre d'observations d'espèces sténoèces disponibles, la droite rouge est la droite de régression linéaire correspondante, la droite bleue pleine matérialise le cas où les valeurs prédites sont égales aux valeurs observées, les deux droites bleues en tirets représentant l'enveloppe de l'erreur empirique ( $\pm 0.4$ ).



Pour l'engorgement, la figure 10 montre que la droite de régression est quasiment alignée sur la bissectrice (absence de biais dans les prédictions) et une corrélation bonne ( $r = 0,915$ ). Toutefois, les écarts sont assez souvent supérieurs à  $\pm 0.4$  (environ 25 sites), particulièrement pour quelques sites (MOTTE, NARC, ARGENS, ECOURT, FINIELS) ; il s'agit pour la quasi-totalité de sites avec peu de données. L'histogramme des résidus montre une légère dissymétrie. Les résidus studentisés ont tendance à être plus élevés lorsqu'il y a peu d'observations, et quelques sites (l'île de la Motte, Argens,...) ont une forte influence sur la pente de la droite, sans toutefois que le nombre de sites ayant des résidus studentisés supérieurs à  $\pm 2$  soit critique.

La figure 11 montre les résultats pour la fertilité.

**Figure 11** : relation entre les valeurs observées pour la fertilité avec l'ensemble des données sur chaque site (en abscisses) et les valeurs prédites avec les seules données des espèces sténoïques (en ordonnées) ; la taille des points est proportionnelle au nombre d'observations d'espèces sténoïques disponibles, la droite rouge est la droite de régression linéaire correspondante, la droite bleue pleine matérialise le cas où les valeurs prédites sont égales aux valeurs observées, les deux droites bleues en tirets représentant l'enveloppe de l'erreur empirique ( $\pm 0.4$ ).



La corrélation est très bonne ( $r = 0,965$ ), et la droite de régression est quasiment alignée sur la droite d'égalité des valeurs. Toutefois, on remarque la présence de quelques résidus négatifs très forts et un nombre assez important de sites ayant des résidus studentisés importants (9 sites sur 137).

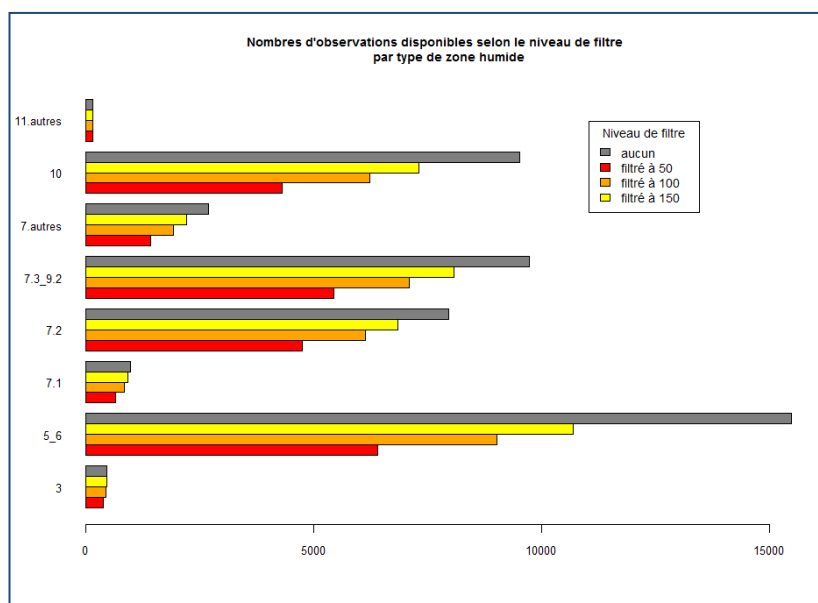
Malgré les très bonnes valeurs prédites avec les seules espèces sténoïques, cette méthode de filtre ne suffit pas à elle seule car le nombre de taxons impliqués reste beaucoup trop élevé par type de zone humide (tableau 8). En revanche, on pourrait utiliser ces valeurs de sténoïcité pour affiner des listes d'espèces obtenues par d'autres moyens.

## 4.4 Filtres sur la fréquence des espèces

Pour chaque type de zone humide, nous avons calculé les 50, 100 ou 150 espèces les plus fréquentes. Lorsque, pour un type de zone humide, le nombre total de taxon est compris entre deux de ces seuils (70 par exemple, soit entre 50 et 100), les 50 premiers sont retenus pour la liste filtrée à 50 et 70 le sont pour la liste filtrée à 100 ; **ces valeurs seuils sont donc des maxima**. Cette procédure permet de limiter le nombre de relevés qui sont exclus car plus aucune espèce ne s'y retrouve. Nous avons ensuite calculé, par relevé, les valeurs diagnostiques obtenues avec l'ensemble des espèces (HR ou NR pour l'engorgement et la fertilité respectivement), HR50, HR100 et HR150 pour celles obtenues avec 50, 100 ou 150 espèces pour l'engorgement et NR50, NR100 et NR150 pour la fertilité. Nous avons ensuite comparé les évaluations à l'échelle des types de zones humides, des sites et des relevés par type de zones humides.

### 4.4.1 Impacts sur les informations disponibles

Les répercussions sur le nombre d'observations disponibles en fonction des niveaux de filtre et du type de zone humide sont montrées figure 12.



**Figure 12** : nombre d'observations disponibles selon le niveau de filtre (couleurs) et selon le type de zone humide.

Globalement, le filtrage à 50 espèces supprime près de 50% des observations, celui à 100 espèces en supprime 31,9%, et enfin celui à 150 espèces 21,7%. La plupart des types de zones humides sont touchés, mais à des degrés divers, les plus impactés étant les types 5\_6 (milieux alluviaux), 10 (marais) et 7.3\_9.2 (tourbières alcalines). La suppression d'espèces entraîne pour certains relevés la disparition de toutes les espèces de ces relevés ; ils ne peuvent donc plus être utilisés, ni pour les comparaisons méthodologiques, ni pour les évaluations des sites. Ce phénomène est toutefois marginal.

Concernant le nombre de taxons, le filtre à 50 espèces par type de zone humide aboutit à conserver au total 275 taxons, celui à 100, 421 taxons et celui à 150, 510 taxons. Les réductions des pourcentages de taxons correspondantes sont de 80,7%, 70,5% et 64,2% par rapport à l'ensemble des 1.422 taxons présents dans la base pour l'ensemble des sites analysés.

#### 4.4.2 Comparaisons des valeurs diagnostiques selon le niveau de filtre et le type de zone humide à l'échelle des sites

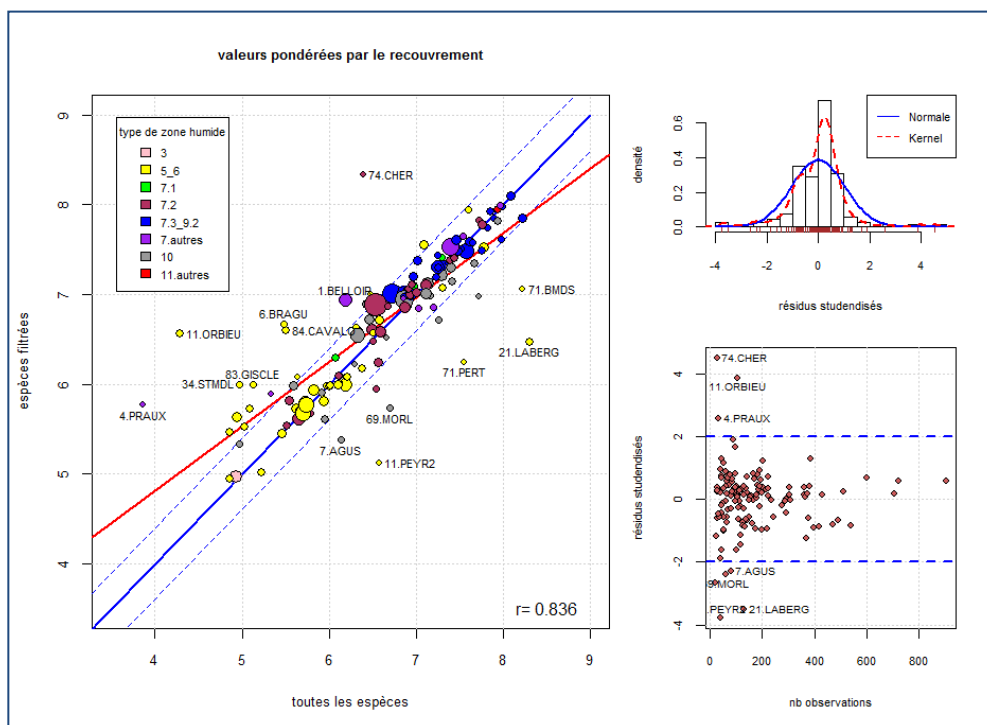
A l'échelle d'un site, nous avons calculé la valeur diagnostique médiane des placettes de ce site, et ce pour chaque niveau de filtre (non filtré, filtré à 50, 100 ou 150). Puis, nous avons étudié la relation entre la valeur non filtrée et les différentes valeurs issues des filtres. Les graphiques comportent les mêmes informations que celles présentées pour le filtre sur le caractère sténoèce des espèces.

Une autre approche aurait consisté à calculer des régressions pondérées, en donnant comme poids aux sites les nombres d'observations après filtrage. Nous avons effectué des calculs avec ces régressions pondérées, mais les écarts constatés sont très faibles avec les régressions non pondérées.

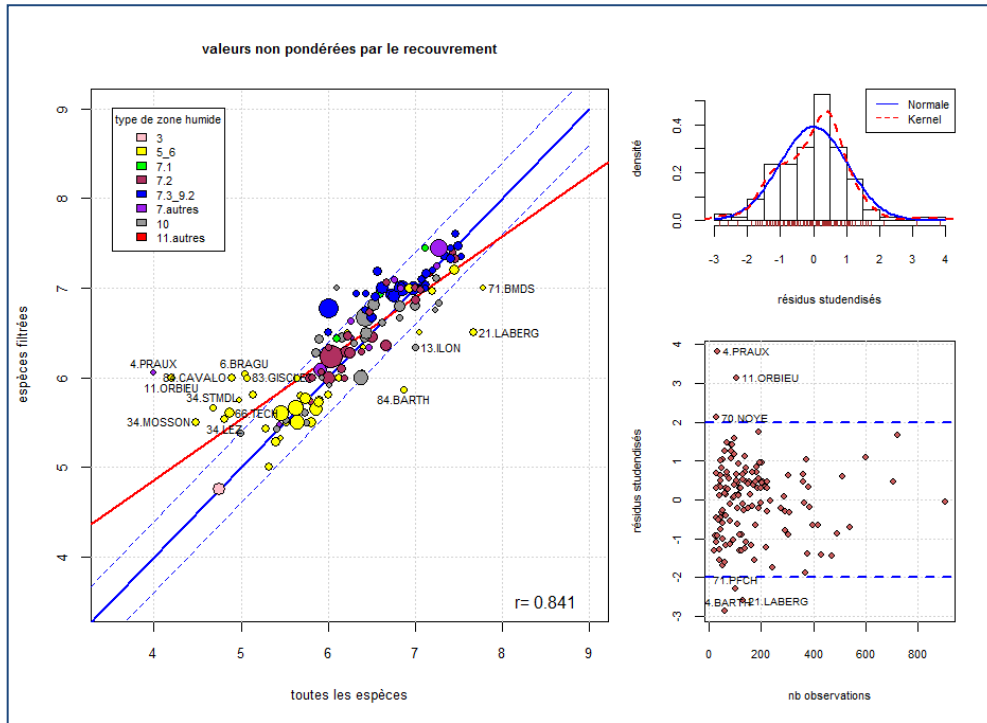
##### 4.4.2.1 Résultats pour l'engorgement

Les figures 13 et 14 illustrent les relations obtenues entre les valeurs non filtrées d'une part, et, d'autre part, celles filtrées à 50, avec et sans recouvrement, respectivement. La corrélation avec le niveau de filtre à 50 est assez bonne ( $r = 0,836$ ). Toutefois, 26 sites ont des valeurs médianes présentant un écart supérieur à  $\pm 0.4$ . Il s'agit pour l'essentiel de sites du type 5\_6 (milieux alluviaux), surtout du sud ; ces écarts sont dus soit à la présence d'assez nombreux relevés pauvres soit à celle d'espèces spécifiques limitées au sud. Dans une moindre mesure, les sites de type 10 (marais) sont touchés. Par ailleurs, la relation est nettement biaisée.

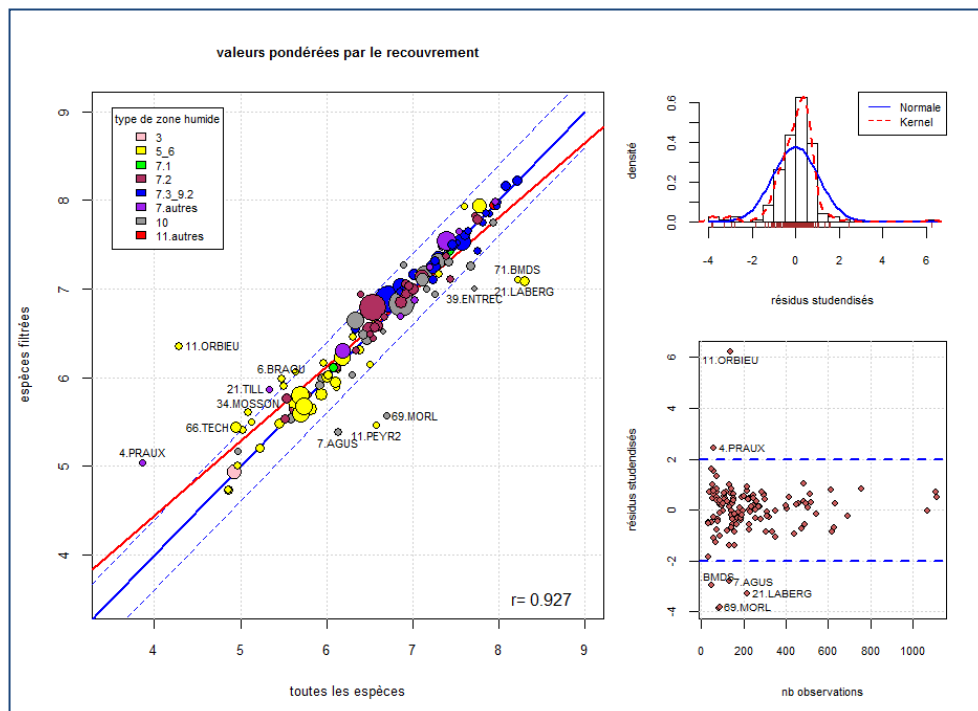
**Figure 13** : relation entre les valeurs observées pour l'engorgement avec recouvrement avec l'ensemble des données sur chaque site (en abscisses) et les valeurs prédites avec les seules données filtrées sur les 50 espèces les plus fréquentes par type de zone humide (en ordonnées) ; la taille des points est proportionnelle au nombre d'observations : l'histogramme du haut est celui des résidus studentisés de la régression ; le diagramme du bas donne la relation entre ces mêmes résidus et le nombre d'observations disponibles après filtrage.



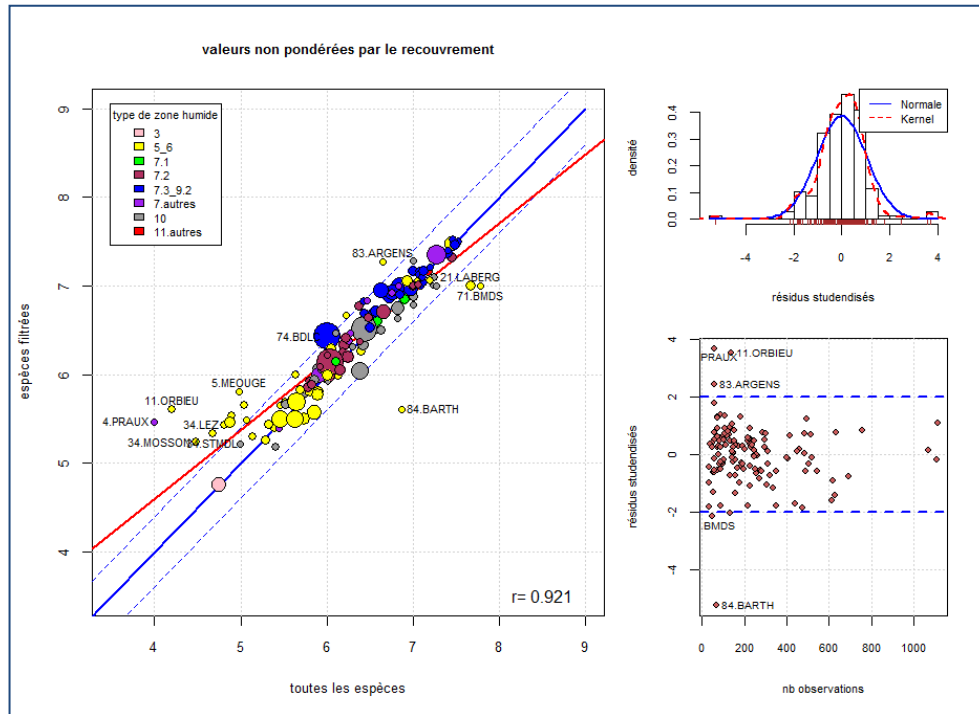
**Figure 14** : relation entre les valeurs observées pour l'engorgement sans recouvrement avec l'ensemble des données sur chaque site (en abscisses) et les valeurs prédites avec les seules données filtrées sur les 50 espèces les plus fréquentes par type de zone humide (en ordonnées) ; la taille des points est proportionnelle au nombre d'observations : l'histogramme du haut est celui des résidus studentisés de la régression ; le diagramme du bas donne la relation entre ces mêmes résidus et le nombre d'observations disponibles après filtrage.



**Figure 15** : relation entre les valeurs observées pour l'engorgement avec recouvrement avec l'ensemble des données sur chaque site (en abscisses) et les valeurs prédites avec les seules données filtrées sur les 100 espèces les plus fréquentes par type de zone humide (en ordonnées) ; la taille des points est proportionnelle au nombre d'observations : l'histogramme du haut est celui des résidus studentisés de la régression ; le diagramme du bas donne la relation entre ces mêmes résidus et le nombre d'observations disponibles après filtrage.



**Figure 16** : relation entre les valeurs observées pour l'engorgement avec recouvrement avec l'ensemble des données sur chaque site (en abscisses) et les valeurs prédites avec les seules données filtrées sur les 100 espèces les plus fréquentes par type de zone humide (en ordonnées) ; la taille des points est proportionnelle au nombre d'observations : l'histogramme du haut est celui des résidus studentisés de la régression ; le diagramme du bas donne la relation entre ces mêmes résidus et le nombre d'observations disponibles après filtrage.

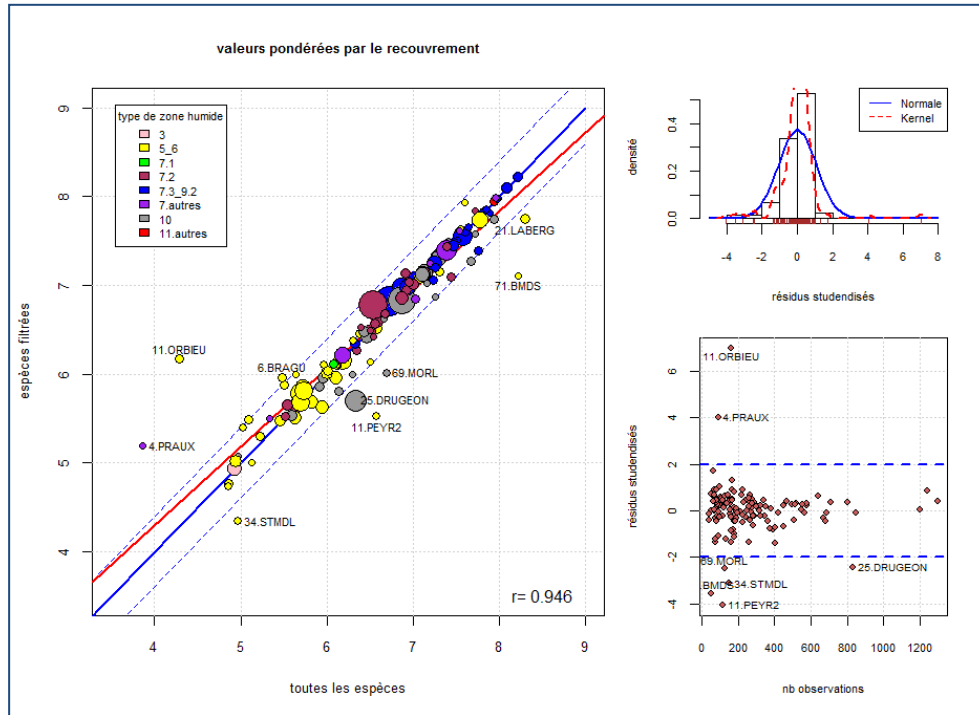


La relation obtenue sans recouvrement est d'une qualité semblable, avec une corrélation correcte mais 28 sites situés à plus de  $\pm 0.4$  et un biais assez net. Pour les deux cas, le nombre de sites avec des résidus studentisés supérieurs à  $\pm 2$  (6) est proche du seuil maximum des 5% des sites (7), et il s'agit de sites avec peu d'observations.

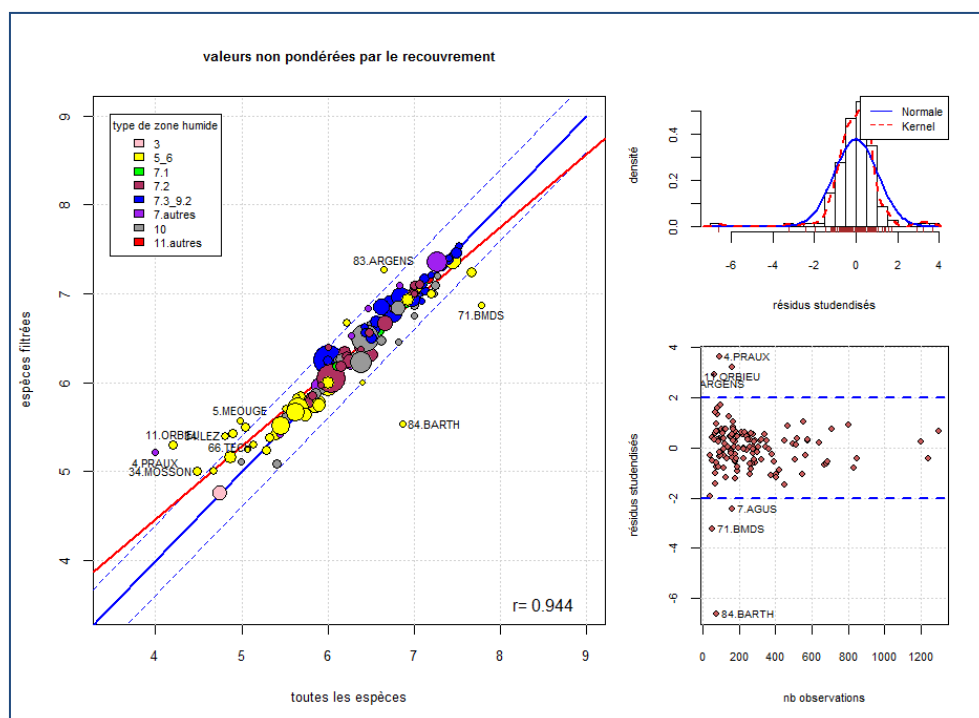
Avec un filtre de rareté à 100 espèces (figures 15 et 16), les corrélations deviennent bonnes ( $r = 0,927$  et  $r = 0,921$  avec et sans recouvrement). Seulement 14 sites sont situés en dehors de l'enveloppe des  $\pm 0.4$ , la plupart avec des écarts plus faibles que précédemment. Les types de zones humides touchées sont globalement les mêmes. Les sites les moins bien évalués sont la Ripisyle de l'Orbieu (ORBIEU), les Prairies humides de Praux (PRAUX), la Prairie de fauche de Labergement (LABERG), le Bras mort du Doubs (BMDS) et le domaine de Peyremale (PEYR2). Il s'agit globalement de sites assez peu riches et sont en grande partie communs avec les sites mal classés avec le filtre à 50 espèces. Les deux relations sont encore biaisées, mais de manière moindre qu'avec le filtre à 50 espèces. Les résidus ne sont pas parfaitement symétriques, mais sans être problématique. Les résidus studentisés supérieurs à 2 sont un peu moins nombreux, mais quelques sites sont vraiment mal prédits et exercent une influence forte sur la pente de la droite de régression (BARTH, ORBIEU, PRAUX, MORLIN, LABERG).

Avec un filtre de rareté à 150 espèces, la situation s'améliore encore avec une corrélation de 0,946 (figure 17) avec recouvrement et 0,944 sans recouvrement (figure 18). Seuls 9 sites sortent de l'enveloppe des  $\pm 0.4$ . Les sites les plus problématiques restent la Ripisyle de l'Orbieu (ORBIEU), les Prairies humides de Praux (PRAUX), le Bras mort du Doubs (BMDS) et le domaine de Peyremale (PEYR2), mais les écarts sont plus faibles.

**Figure 17** : relation entre les valeurs observées pour l'engorgement avec recouvrement avec l'ensemble des données sur chaque site (en abscisses) et les valeurs prédites avec les seules données filtrées sur les 150 espèces les plus fréquentes par type de zone humide (en ordonnées) ; la taille des points est proportionnelle au nombre d'observations : l'histogramme du haut est celui des résidus studentisés de la régression ; le diagramme du bas donne la relation entre ces mêmes résidus et le nombre d'observations disponibles après filtrage.



**Figure 18** : relation entre les valeurs observées pour l'engorgement sans recouvrement avec l'ensemble des données sur chaque site (en abscisses) et les valeurs prédites avec les seules données filtrées sur les 150 espèces les plus fréquentes par type de zone humide (en ordonnées) ; la taille des points est proportionnelle au nombre d'observations : l'histogramme du haut est celui des résidus studentisés de la régression ; le diagramme du bas donne la relation entre ces mêmes résidus et le nombre d'observations disponibles après filtrage.

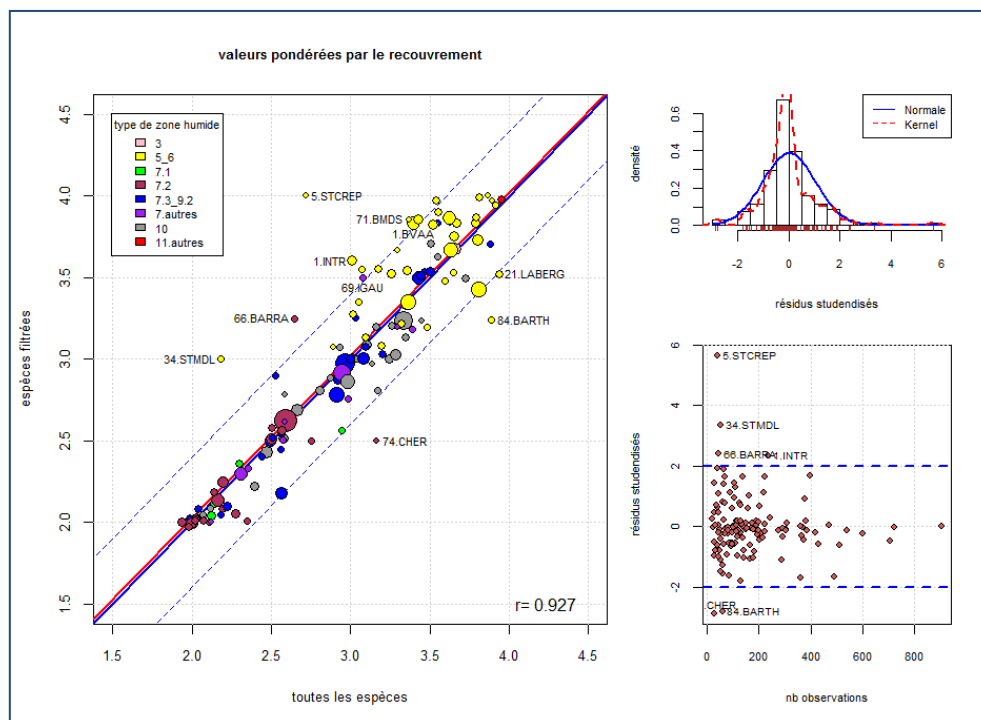


Sans recouvrement, le site de la forêt alluviale de la Barthelasse BARTH a un résidu studentisé très élevé, dû surtout à son effet levier (c'est-à-dire que la régression serait assez différente si ce site n'était pas dans le jeu de données), ce qui n'était pas autant le cas avec recouvrement ; ceci est dû à la présence d'assez nombreuses espèces plus ou moins particulières à ce site, mais qui ont des recouvrements faibles. A contrario, le site du Drugeon (DRUG) est mal prédit avec recouvrement alors qu'il l'est mieux sans, ceci étant dû à la présence d'espèces particulières ayant des recouvrements assez élevés. Malgré ces améliorations, la droite de régression continue de fournir des estimations biaisées, même avec 150 espèces par type de zone humide.

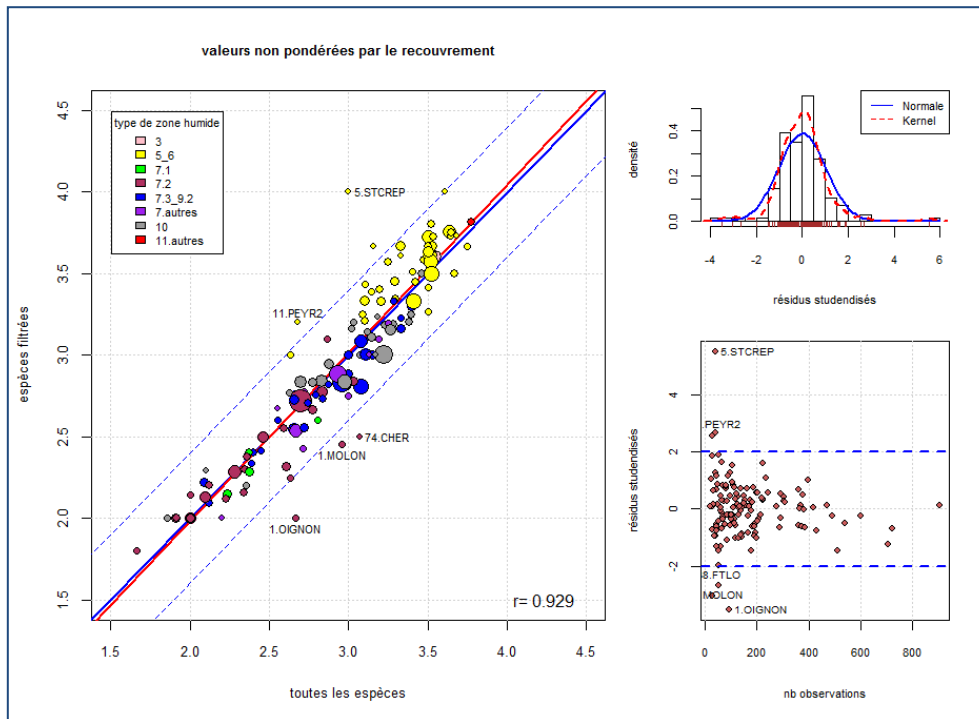
#### 4.4.2.2 Résultats pour la fertilité

La corrélation avec le niveau de filtre à 50 est bonne ( $r = 0,927$ ) avec recouvrement (figure 19). Toutefois, 13 sites ont des valeurs médianes présentant un écart supérieur à  $\pm 0.4$ . Il s'agit pour l'essentiel de sites du type 5\_6 (milieux alluviaux), surtout du sud ; ces écarts sont dus soit à la présence d'assez nombreux relevés pauvres soit à celle d'espèces spécifiques limitées au sud. On note également une légère surreprésentation de résidus positifs, et l'influence assez forte des Saint-Crépin (STCREP) de de Saint-martin-de-Londes (STML); les autres sites mal prédits sont Barraux (BARR), le Roc de Chère (CHER) et la forêt alluviale de la Barthelasse (BARTH). La droite de régression est quasiment superposée à la bissectrice, indiquant l'absence de biais. Sans recouvrement (figure 20), la situation est encore légèrement meilleure, avec moins de sites en dehors de l'intervalle de confiance empirique et moins de sites avec des résidus studentisés élevés. Dans les deux cas, le site de Saint-Crépin pèse beaucoup sur la droite de régression (effet levier élevé).

**Figure 19** : relation entre les valeurs observées **pour la fertilité avec recouvrement** avec l'ensemble des données sur chaque site (en abscisses) et les valeurs prédites avec les seules données filtrées sur **les 50 espèces les plus fréquentes** par type de zone humide (en ordonnées) ; la taille des points est proportionnelle au nombre d'observations : l'histogramme du haut est celui des résidus studentisés de la régression ; le diagramme du bas donne la relation entre ces mêmes résidus et le nombre d'observations disponibles après filtrage.



**Figure 20** : relation entre les valeurs observées pour la fertilité sans recouvrement avec l'ensemble des données sur chaque site (en abscisses) et les valeurs prédites avec les seules données filtrées sur les 50 espèces les plus fréquentes par type de zone humide (en ordonnées) ; la taille des points est proportionnelle au nombre d'observations : l'histogramme du haut est celui des résidus studentisés de la régression ; le diagramme du bas donne la relation entre ces mêmes résidus et le nombre d'observations disponibles après filtrage

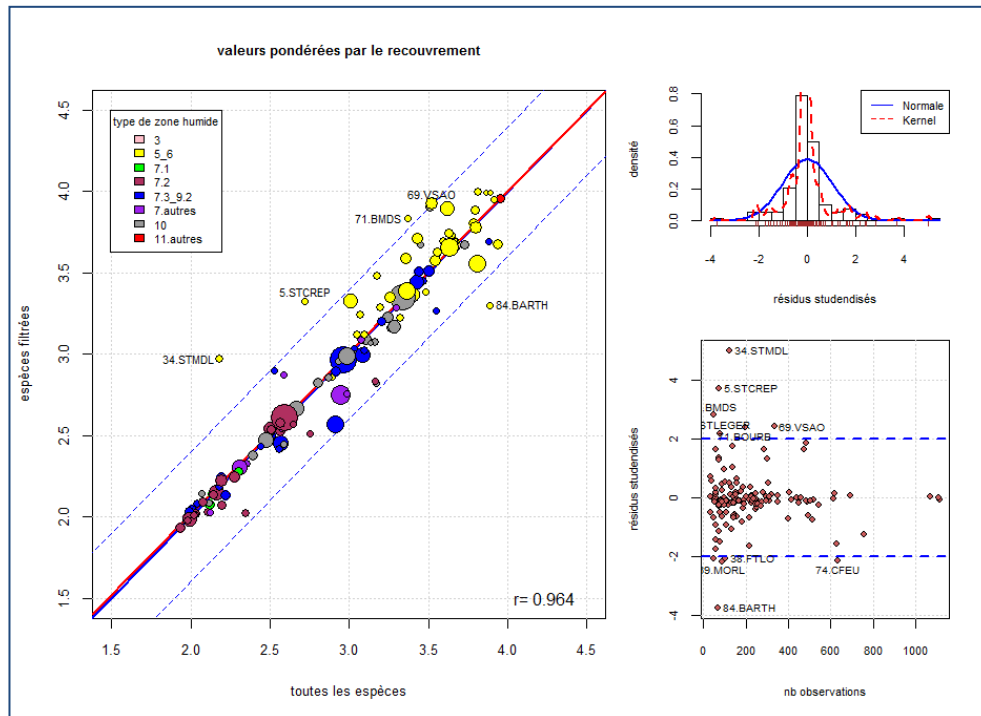


Avec un filtre de rareté à 100 espèces (figure 21), la corrélation devient très bonne ( $r = 0,964$ ) avec recouvrement. Seulement 5 sites sont situés en dehors de l'enveloppe des  $\pm 0.4$ , la plupart avec des écarts très faibles ; il s'agit globalement des mêmes sites qu'avec le filtre à 50. La droite de régression est quasiment confondue avec la bissectrice. Les types de zones humides touchées sont globalement les mêmes. Les sites mal représentés sont, pour 100 espèces, la Cuvette de Saint-Martin-de-Londres (STMDL), la Durance à Saint-Crépin (STCREP), les Forêts alluviales du Rhône à la Barthelasse (BARTH), et le Bras mort du Doubs, ce dernier site étant un des rares qui soit mal représenté tant pour l'engorgement que pour la fertilité. En revanche, les résidus studentisés élevés sont nombreux (11).

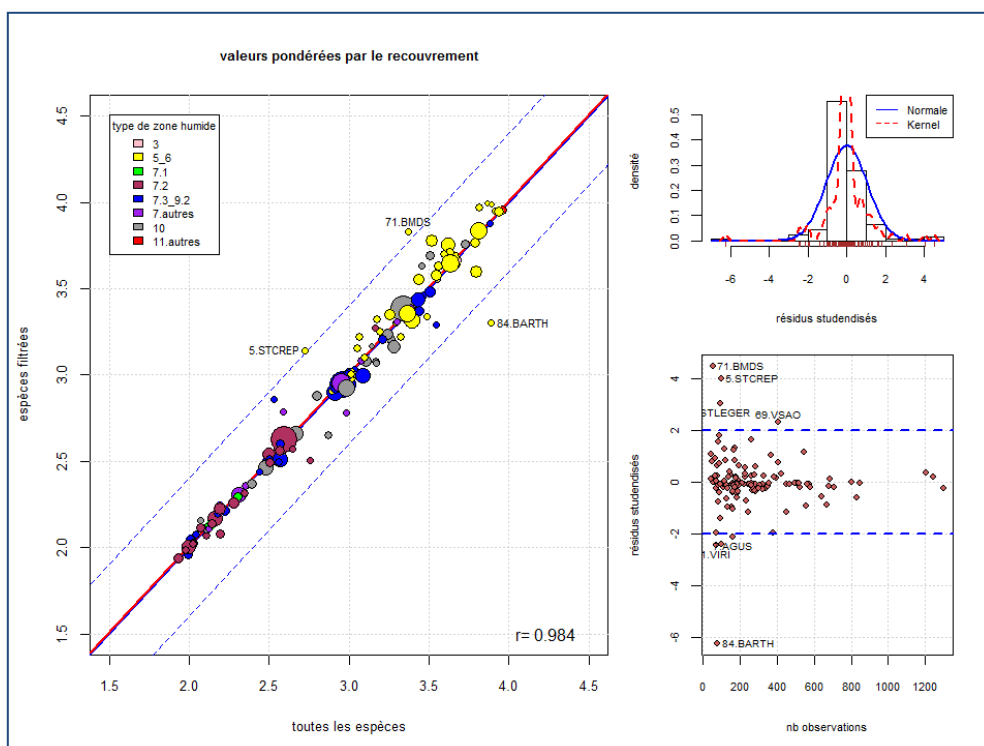
Sans recouvrement (figure non montrée), la situation est encore légèrement meilleure ( $r = 0,969$ ), les résidus studentisés sont légèrement plus faibles et un seul site est situé en dehors de « l'intervalle de confiance empirique ».

Avec un filtre de rareté à 150 espèces avec recouvrement (figure 22), la situation s'améliore encore avec une corrélation de 0,984. Seuls 3 sites sortent de l'enveloppe des  $\pm 0.4$ , BARTH, le Bras mort du Doubs (BMDS) et Saint-Crépin (STCREP) ; ce sont surtout ces sites qui ont des résidus studentisés élevés, notamment par leur effet levier important. Le biais est quasiment nul. La situation est quasiment identique pour les données sans recouvrement (résultats non montrés).

**Figure 21** : relation entre les valeurs observées pour la fertilité avec recouvrement avec l'ensemble des données sur chaque site (en abscisses) et les valeurs prédites avec les seules données filtrées sur les 100 espèces les plus fréquentes par type de zone humide (en ordonnées) ; la taille des points est proportionnelle au nombre d'observations : l'histogramme du haut est celui des résidus studentisés de la régression ; le diagramme du bas donne la relation entre ces mêmes résidus et le nombre d'observations disponibles après filtrage



**Figure 22** : relation entre les valeurs observées pour la fertilité avec recouvrement avec l'ensemble des données sur chaque site (en abscisses) et les valeurs prédites avec les seules données filtrées sur les 150 espèces les plus fréquentes par type de zone humide (en ordonnées) ; la taille des points est proportionnelle au nombre d'observations : l'histogramme du haut est celui des résidus studentisés de la régression ; le diagramme du bas donne la relation entre ces mêmes résidus et le nombre d'observations disponibles après filtrage



Au vu de ces éléments, le filtre sur les 50 espèces les plus fréquentes par type de zone humide ne semble pas totalement opérant ; un minimum de 100 espèces est nécessaire. Le passage à 150 espèces permet encore d'améliorer la situation, notamment par une réduction des écarts par rapport aux évaluations obtenues avec l'ensemble ces données. Globalement, les écarts sont un peu plus faibles pour la fertilité que pour l'engorgement. Un point important est qu'à l'échelle des sites, le processus de filtrage des espèces n'introduit pas de biais (sauf pour l'engorgement avec 100 espèces où un biais est persistant) entre les valeurs filtrées et les valeurs avec données complètes.

Certains sites sont mal prédits de manière récurrente. Le nombre de données disponibles après filtrage influe globalement sur la qualité des prédictions, mais il est évident que certaines spécificités ne sont pas prises en compte avec une approche globale par type de zone humide, comme la présence dans un site d'un contingent important d'espèces originales et/ou ayant des recouvrements élevés.

La prise en compte du recouvrement (versus la seule présence des espèces) influe finalement peu sur la qualité des résultats.

Un avantage majeur de cette approche est que la liste des espèces est la même pour l'engorgement que pour la fertilité. Les résultats montrent que ces listes uniques pour les deux indicateurs fournissent des résultats de même niveau de qualité (même s'ils sont légèrement supérieurs pour la fertilité). Pour la fertilité, le filtre à 100 taxons pourrait être suffisant, alors que pour l'engorgement, le filtre à 150 semble apporter un plus.

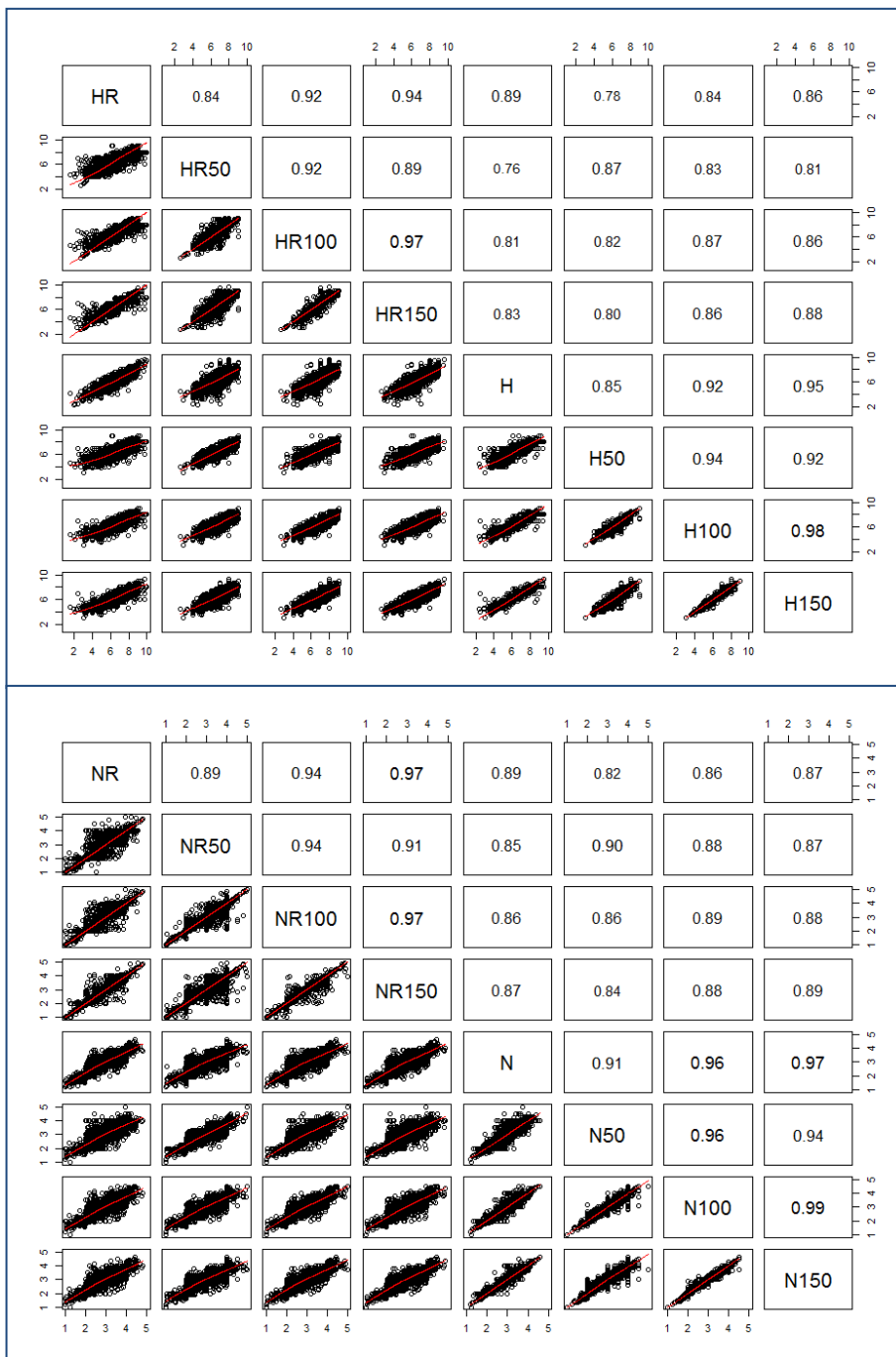
Dans la suite des analyses, seuls les filtres à 100 et 150 espèces seront étudiés et un focus particulier portera sur les types 5\_6 (milieux alluviaux) et 10 (marais) où se situent les sites les plus problématiques.

#### *4.4.3 Comparaisons des valeurs diagnostiques selon le niveau de filtre à l'échelle des relevés*

Les corrélations entre valeurs diagnostiques des relevés issues des données « complètes » et des valeurs issues des données filtrées sont très bonnes, tant pour l'engorgement (figure 23) que pour la fertilité (figure 24). On remarque néanmoins que les corrélations sont un peu plus faibles avec les listes filtrées à 50 espèces.

On constate également que les corrélations sont équivalentes à l'intérieur du bloc des valeurs pondérées par le recouvrement et à l'intérieur du bloc des valeurs non filtrées (respectivement comprises entre 0,84 et 0,97 et 0,85 et 0,98 pour les valeurs pondérées et non pondérées) ; en revanche, à niveau de filtre égal, les corrélations entre les valeurs avec ou sans recouvrement sont légèrement moins bonnes (entre 0,76 et 0,89). On constate également que la dispersion des points est globalement plus faible pour les valeurs non pondérées par le recouvrement. A titre d'exemple, la régression linéaire entre HR et HR100 donne une erreur standard résiduelle de 0,45 ; pour la régression entre H et H100, elle est légèrement plus faible (0,34). Ceci se traduit toutefois très faiblement sur les valeurs des corrélations car le nombre de points est très élevé.

Pour la fertilité, les résultats sont cohérents avec ceux obtenus pour l'engorgement, les corrélations étant encore meilleures (figure 24).



**Figure 23 :** relations entre les valeurs calculées avec l'ensemble des données (HR et H), ou filtrées à 50 espèces (HR50 et H50), 100 (HR100 et H100) ou 150 espèces (HR150 et H150), pour l'engorgement. Les valeurs dans le triangle supérieur sont les coefficients de corrélations ; les variables « HR » correspondent aux valeurs calculées avec pondération par le recouvrement, les variables « H » à celles sans pondération.

**Figure 24 :** relations entre les valeurs calculées avec l'ensemble des données (NR et N), en ne retenant que 50 espèces (NR50 et N50), 100 (NR100 et N100) ou 150 espèces (NR150 et N100), pour l'engorgement. Les valeurs dans le triangle supérieur sont les coefficients de corrélations ; les variables « NR » correspondent aux valeurs calculées avec pondération par le recouvrement, les variables « N » à celles sans pondération.

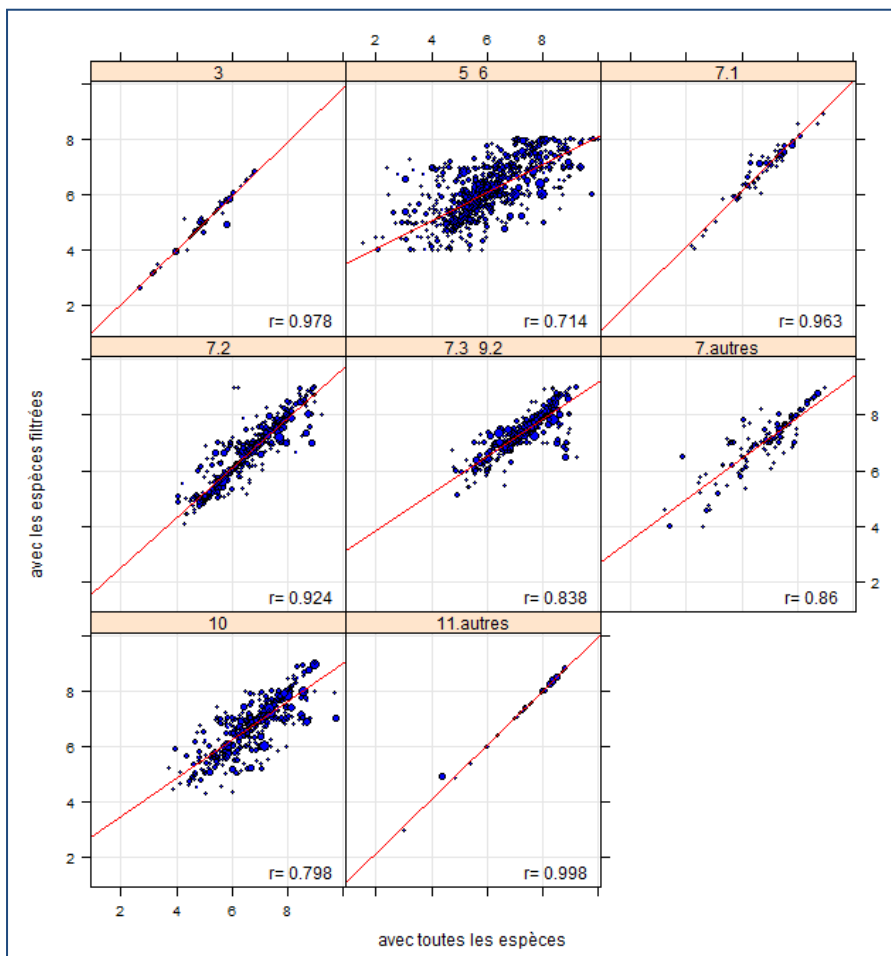
Dans le détail également, certaines placettes ont des valeurs très différentes (écarts de valeurs diagnostiques supérieurs à 2 voire 3) ; ceci correspond la plupart du temps à des placettes pauvres en espèces. De plus, ces corrélations quoique rassurantes, sont en partie flatteuses, comme le sont la plupart des corrélations obtenues sur un nombre élevé de données. Il convient au moins de s'assurer d'un éventuel effet type de zone humide.

Afin d'aller plus loin dans l'analyse, les relations entre valeurs filtrées et non filtrées ont été étudiées par type de zone humide. Pour chaque relevé, les valeurs calculées avec toutes les espèces ("valeur complète") sont confrontées avec les valeurs filtrées ("valeur filtrée"). Pour chaque niveau de filtre et par type de zone humide, un ajustement linéaire a été calculé (droites rouges sur les figures). En cas de bonne prédiction, ces droites

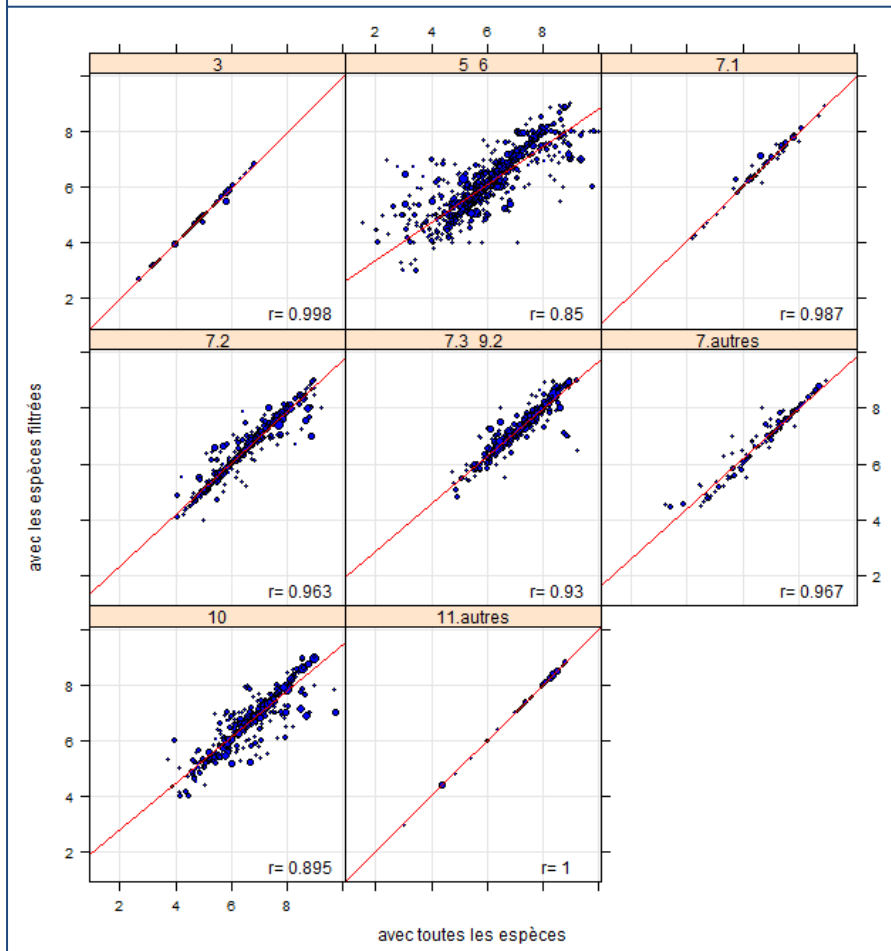
doivent se situer près de la bissectrice des graphiques car les 2 axes ont la même amplitude). **L'identification des cas où les estimations sont biaisées est importante car des prédictions biaisées peuvent empêcher (ou atténuer) la capacité à détecter une évolution des sites** : par exemple, si les valeurs faibles d'engorgement sont surestimées avec les valeurs filtrées, cela veut dire qu'une évolution à la baisse du niveau de la nappe sera mal prédite avec les données filtrées (car, *in fine*, les espèces à faibles valeurs indicatrices d'engorgement sont trop filtrées par rapport aux espèces à fortes valeurs indicatrices).

La figure 25 présente le cas de l'engorgement filtré à 50 espèces. Pour les types pauvres en relevés (3, 7.1, 11.autres), les corrélations semblent très bonnes, les sites sont assez étroitement alignés le long de la bissectrice ; ils ne seront pas discutés plus avant. Pour le type 5\_6, on constate que la dispersion des points est grande et la corrélation assez faible ; on remarque également que la courbe de tendance linéaire ajustée entre valeurs complètes et valeurs filtrées n'est pas parallèle à la bissectrice ; dans ce cas particulier, les prédictions obtenues avec 50 espèces surestiment les valeurs réelles pour des valeurs réelles faibles (inférieures en gros à 5), mais elles sous-estiment les valeurs réelles pour les valeurs complètes élevées (en gros supérieures à 5). Un schéma semblable émerge pour le type 10, mais de manière moins marquée, tant du point de vue de la dispersion des points que du biais. Enfin, pour les types 7.autres et 7.3\_9.2, les mêmes défauts se retrouvent mais de manière encore plus atténuée. En revanche, pour les tourbières alcalines (7.2), la relation est bonne, tant du point de vue de la dispersion que de l'absence de biais, et ce malgré un nombre important de relevés. Le passage à 100 espèces (figure 26) permet d'améliorer la situation et de rendre les prédictions acceptables pour la quasi-totalité des types de zones humides, à l'exception du type 5\_6 (milieux alluviaux) ; pour ce dernier, une dispersion encore assez importante persiste de même qu'un biais dans les valeurs prédites, ce qui est conforme avec les résultats à l'échelle des sites (figure 17). Le passage à 150 espèces (résultats non montrés) permet d'obtenir des relations satisfaisantes pour l'ensemble des types de zones humides ( $r$  compris entre 0,919 et 1,000 et quasi absence de biais, même pour le type 5\_6). **Il semble donc nécessaire d'avoir entre 100 et 150 espèces pour l'engorgement, selon le type de zone humide, pour prédire de manière assez précise et non biaisée les estimations obtenues avec l'ensemble des données à l'échelle des placettes.**

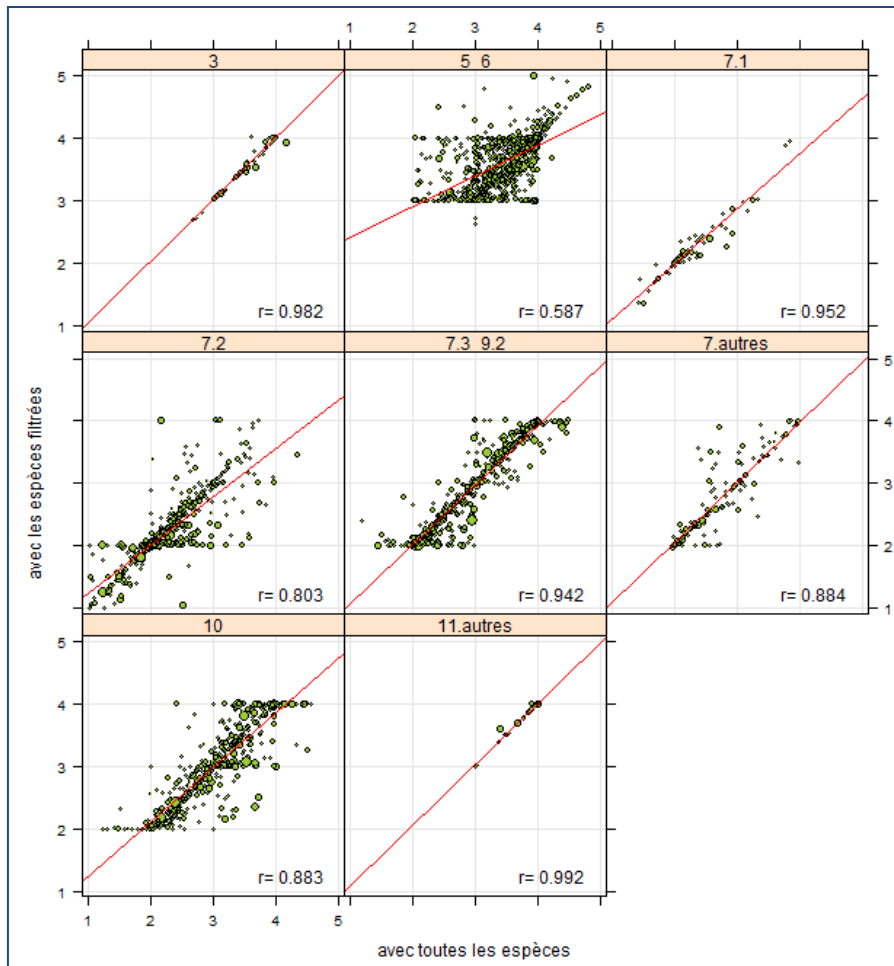
Pour les nutriments, la situation est assez semblable (figures 27 et 28). Les types 3 et 11.autres, peu représentés, semblent bien prédits. Les zones humides d'altitude (7.1) sont relativement bien prédites, malgré un léger biais. Les tourbières alcalines (7.3\_9.2) sont assez bien prédites ( $r = 0,942$ ) et sans biais. Les tourbières acides (7.2) sont assez bien (ou mal selon les points de vues) prédites, avec une forte dispersion des points et un léger biais. Le type 10 (marais) présente des points très étalés mais quasiment sans biais. Enfin, le type 5\_6 (milieux alluviaux) présente des points très dispersés et avec un biais très net. De plus, on constate, comme pour les marais (10) ou les tourbières acides (7.2), de manière beaucoup moins marquée, la présence de nombreuses valeurs prédites à 3 ou 4. Ceci s'explique par l'effet filtre qui a sélectionné beaucoup d'espèces ayant des valeurs indicatrices de 2, 3 ou 4 (selon les milieux), entraînant fréquemment pour un relevé filtré la présence d'espèces n'ayant qu'une seule et même valeur indicatrice. Pour le type 5\_6, très diversifié, d'assez nombreuses espèces rares ont des valeurs indicatrices de fertilité faible (1 ou 2) ; le filtre sur les 50 espèces les plus fréquentes les élimine de fait de données et entraîne une surestimation de la valeur diagnostique de fertilité pour les valeurs « vraies » faibles (« vraies » étant à prendre dans le sens de obtenues avec l'ensemble des données). La situation s'améliore sensiblement avec le filtre à 100 espèces (figure 28) : les corrélations augmentent, les biais diminuent voire disparaissent, pour l'ensemble des types de zones humides, même si la situation reste médiocre pour les milieux alluviaux. Ces tendances sont toutefois totalement gommées avec un filtre à 150 espèces (résultats non montrés) : **il semble donc nécessaire d'avoir entre 100 et 150 espèces pour la fertilité, selon le type de zone humide, pour prédire de manière assez précise et non biaisée les estimations obtenues avec l'ensemble des données.**



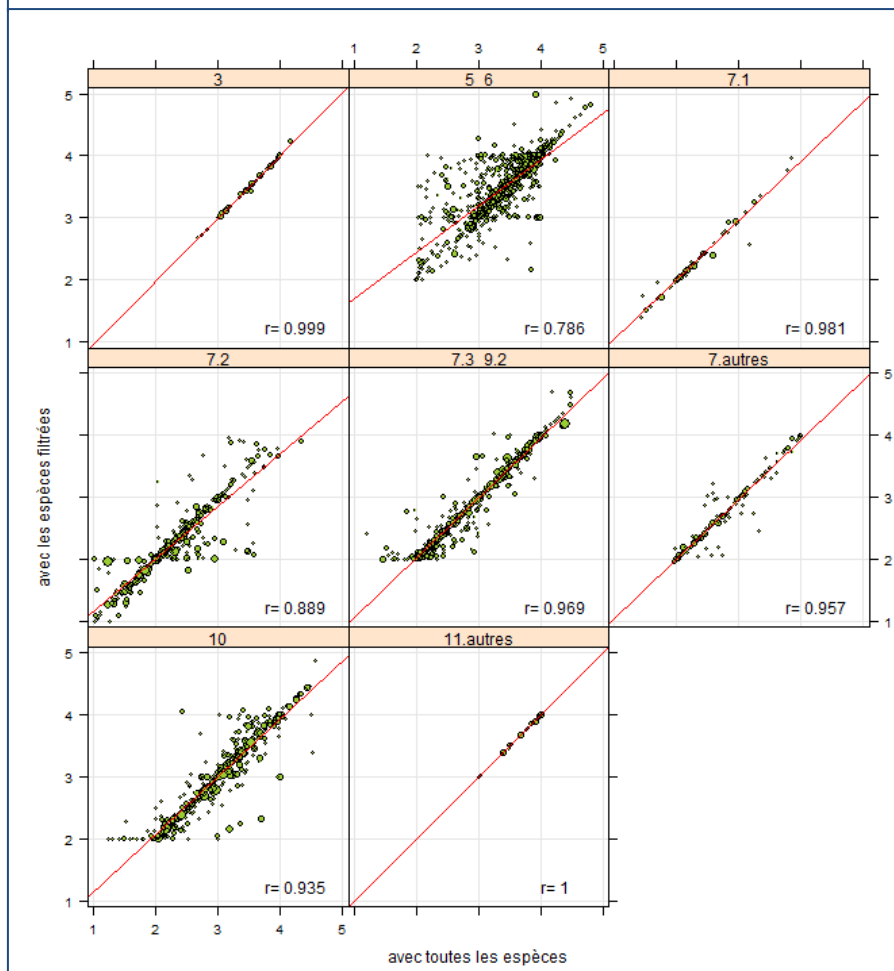
**Figure 25** : relations entre valeurs d'engorgement obtenues avec l'ensemble des données (abscisses) et celles obtenues avec les 50 espèces les plus fréquentes par type de zone humide (ordonnées). Chaque point est un relevé dont la taille est proportionnelle au nombre d'observations utilisées pour calculer la valeur diagnostique filtrée.



**Figure 26** : relations entre valeurs d'engorgement obtenues avec l'ensemble des données (abscisses) et celles obtenues avec les 100 espèces les plus fréquentes par type de zone humide (ordonnées). Chaque point est un relevé dont la taille est proportionnelle au nombre d'observations utilisées pour calculer la valeur diagnostique filtrée.



**Figure 27** : relations entre valeurs de fertilité obtenues avec l'ensemble des données (abscisses) et celles obtenues avec les 50 espèces les plus fréquentes par type de zone humide (ordonnées). Chaque point est un relevé dont la taille est proportionnelle au nombre d'observations utilisées pour calculer la valeur diagnostique filtrée.



**Figure 28** : relations entre valeurs de fertilité obtenues avec l'ensemble des données (abscisses) et celles obtenues avec les 100 espèces les plus fréquentes par type de zone humide (ordonnées). Chaque point est un relevé dont la taille est proportionnelle au nombre d'observations utilisées pour calculer la valeur diagnostique filtrée.

#### 4.4.4 Discussion et conclusions

Les filtres basés sur la fréquence des espèces par type de zone humide permettent d'assez bien à bien prédire les valeurs d'engorgement et de fertilité. A l'échelle des sites, des listes d'environ 100 espèces par type de zone humide donnent des résultats assez satisfaisants. Les erreurs de prédictions sont, dans la très grande majorité des cas, inférieures à  $\pm 0.4$ , ce qui est une indication favorable pour maintenir cette erreur empirique comme critère d'évaluation pour juger de l'évolution d'un site. Mais cette situation cache des disparités entre types de zones humides : les milieux alluviaux (type 5\_6) sont assez mal prédits, sauf avec 150 espèces, et pour d'autres types peu représentés, les prédictions sont probablement fragiles et à prendre avec prudence. De plus, si on veut s'intéresser à une échelle plus petite que le site (typiquement un transect, pour des sites en ayant plusieurs), la situation est moins satisfaisante et dépend notamment beaucoup des types de zones humides ; seul le passage au filtre à 150 espèces (particulièrement pour la fertilité) permet d'atténuer voire d'annuler les biais constatés. Toutefois, les valeurs de sites particuliers, notamment ceux riches en espèces très rares dans le jeu de données, restent difficiles à prédire.

Une difficulté persistante tient dans le classement, parfois contestable, des sites dans des types de zones humides. En effet, malgré un arbre d'aide à la décision, le classement final s'appuie dans des cas ambigus sur le dire d'expert ; or, ce classement, s'il influe peu sur la liste des espèces les plus fréquentes (au moins pour les types bien représentés dans l'échantillon), joue sur les espèces qui vont être utilisées pour un site donné : un exemple intéressant est celui des marais de Chautagne en Savoie, classé dans le type 10 (marais), un autre classement possible aurait été 7.3\_9.2 (tourbières alcalines) ; comme les transects ont été surtout mis en place dans les secteurs ouverts, qui sont essentiellement tourbeux, la flore contactée est essentiellement proche de celles des tourbières alcalines ; les filtres basés sur la rareté pour les marais dans leur ensemble écartent donc tout un ensemble d'espèces fréquentes dans l'échantillon de Chautagne. Les valeurs diagnostiques sont alors globalement assez mal prédites, même si à l'échelle du site, les erreurs se compensent et la prédiction globale reste acceptable. Un autre exemple est celui du Roc de Chère, constitué de deux parties nettement différentes (une tourbière acide et un marais) et classé en tourbières acides (7.2) ; avec des filtres à 50 espèces, les espèces des marais sont éliminées et les valeurs diagnostiques biaisées vers les valeurs faibles. Ce type de processus est globalement plus fréquent dans les milieux alluviaux, particulièrement dans les sites où plusieurs relevés pauvres en espèces existent.

**Une piste de travail serait de fournir des listes d'espèces guides, qui viendraient en complément de l'arbre de décision, pour l'attribution d'un site à un type.**

Une autre difficulté, inattendue, est que pour la fertilité, le nombre de classes des valeurs indicatrices (5 contre 10 pour l'engorgement), le filtrage conduit à ne garder dans d'assez nombreux relevés que des espèces ayant la même valeur indicatrice (3 ou 4 le plus souvent). Ce fait nuit à la précision des estimations à l'échelle des placettes ou de transects (dans le cas où un gestionnaire voudrait pouvoir analyser séparément les transects d'un même site). Une solution serait de répartir les valeurs indicatrices de fertilité sur une échelle de 1 à 10.

## 4.5 Filtres sur la valeur indicatrice (IndVal) ou le coefficient Phi

---

Une autre approche consiste à rechercher, via un indice statistique, le niveau de liaison entre la présence et/ou l'abondance d'une ou plusieurs espèces et, des types de zones humides. Deux indices seront testés. DUFRENE & LEGENDRE (1997) proposent l'indice « indVal », notamment pour étudier la valeur prédictive de certaines espèces par rapport à des unités de végétation, qui est lui-même le produit de deux quantités :

- A, qui est une mesure de la concentration de l'abondance d'une espèce dans une unité donnée ;
- B, qui est la fréquence relative d'une espèce dans une unité ;

Plus la valeur « IndVal » d'une espèce est élevée, mieux cette espèce permet de prédire que l'on se trouve dans un type de végétation donné ou un type de zone humide. La significativité statistique des liaisons est évaluée par des tests statistiques de permutations : on évalue, par rapport à des permutations aléatoires des unités de végétation, dans quelle mesure la valeur mesurée est supérieure à celles obtenues en répétant un grand nombre de fois des permutations. Classiquement, on ne retient que les espèces pour lesquelles la probabilité d'obtenir la valeur observée sachant qu'elle n'est en réalité pas liée à un type est inférieure à 5%. Ce seuil est arbitraire, le plus important étant de le définir au préalable pour ne pas rendre les résultats trop conformes à nos attentes. Au vu des résultats produits par les analyses IndVal publiées, les listes tendent à être très courtes (quelques espèces), ce qui est un des objectifs de la méthode, mais probablement trop courtes au vu des nôtres ; nous avons donc choisi de lister toutes les espèces ayant une probabilité inférieure à 10%. La valeur de l'indice est calculée pour chaque espèce et par rapport à chaque unité de végétation (ou type de zone humide dans notre cas) et les auteurs préconisent de retenir comme indicatrices d'un type les espèces ayant leur valeur IndVal maximale pour ce type et ayant une liaison statistiquement significative.

Dans notre problématique, il faut calculer les valeurs « IndVal » des espèces non pas par rapport à des unités de végétation, mais par rapport à des valeurs diagnostiques d'humidité ou de fertilité (rappelons, ci-besoin en est, qu'une valeur diagnostique est la valeur moyenne des valeurs indicatrices des espèces présentes sur un relevé, pondéré par le recouvrement). Ces valeurs diagnostiques sont des valeurs continues, qu'il faut découper au préalable en classes de valeurs discrètes, car l'indice « IndVal » n'évalue des liaisons qu'avec des variables qualitatives. Ce découpage permet en outre de mieux prendre en compte, *a priori*, la gamme des valeurs indicatrices des espèces que l'approche basée uniquement sur la fréquence (§ 4.4). On peut ensuite choisir les espèces qui permettent de représenter au mieux les différentes classes de valeurs diagnostiques. Ce processus est à répéter pour chaque type de zone humide.

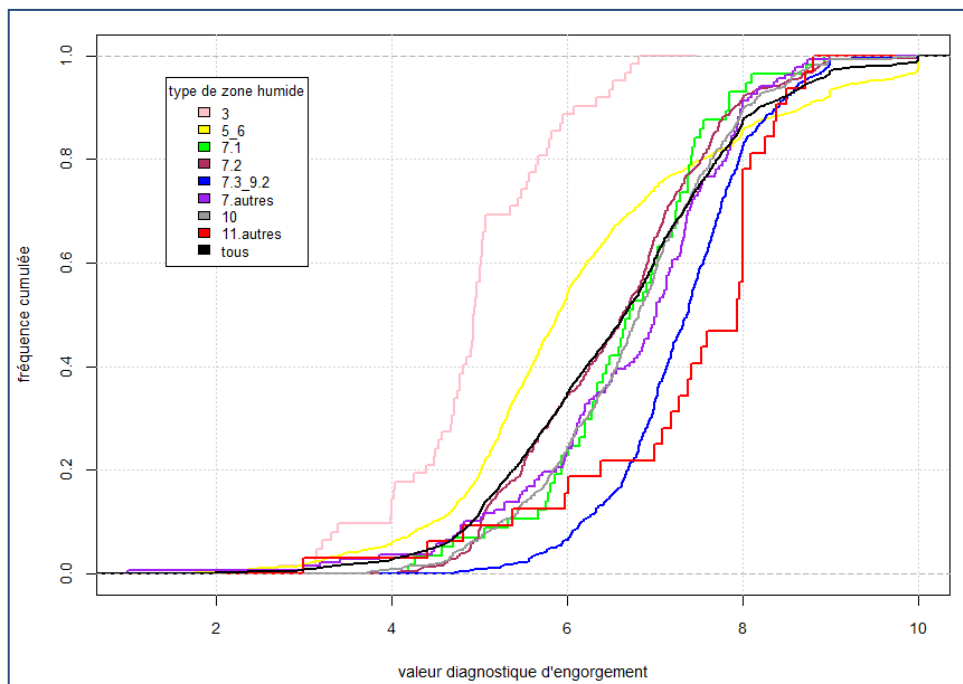
L'autre indice classique testé est le coefficient phi d'association de PEARSON (CHYTRY *et al.*, 2002). C'est une mesure basée uniquement sur des données d'occurrences des espèces dans les groupes ; les valeurs sont comprises entre -1, si une espèce est absente d'un groupe donné, et +1, si une espèce n'est présente que dans un groupe et dans toutes les placettes de ce groupe. Si les listes d'espèces issues de cet indice permettent de produire des valeurs diagnostiques de bonne qualité, un avantage majeur est qu'il sera également possible de ne plus recueillir les recouvrements des espèces, ce qui constitue un gain de temps non négligeable sur le terrain.

Par rapport à l'approche basée sur la fréquence des espèces, ces méthodes présentent l'inconvénient de devoir calculer les liaisons séparément pour l'engorgement et pour la fertilité ; il conviendra donc d'essayer d'harmoniser et de réduire les listes issues de ces deux analyses *a posteriori* pour limiter le nombre d'espèces. Une autre limite est que les espèces vraiment trop rares ne peuvent pas être utilisées ou même si elles le sont, sont souvent écartées de l'analyse car la significativité de la liaison d'une espèce est influencée positivement par le nombre d'occurrence (la significativité augmente, toutes choses égales par ailleurs, avec le nombre

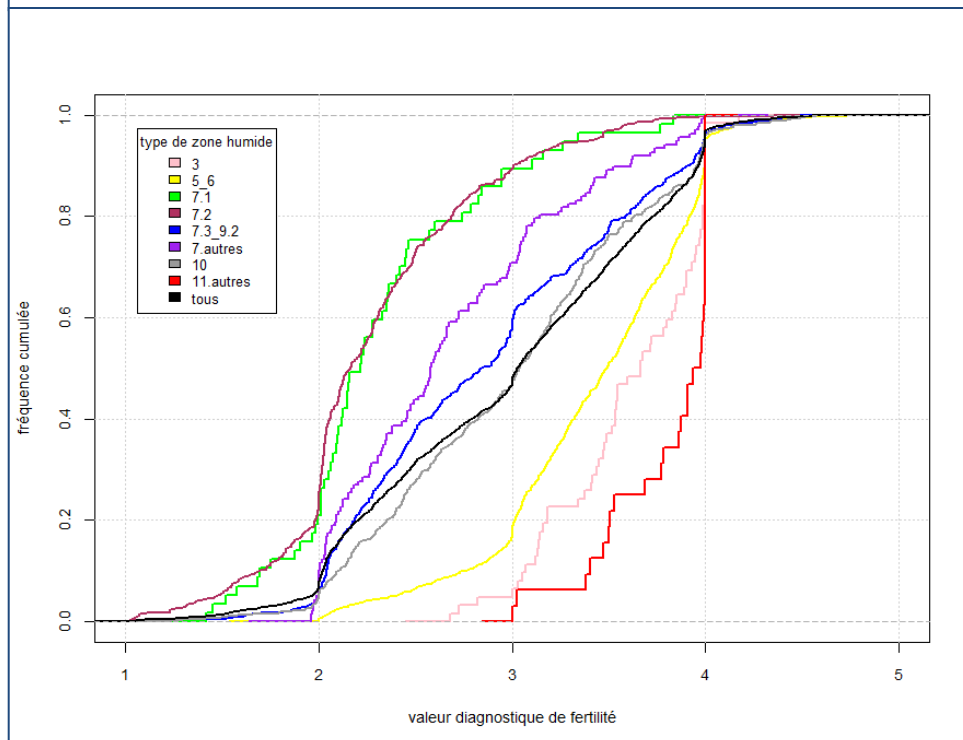
d'occurrences). En revanche, il est possible de trouver de meilleures espèces indicatrices qu'avec l'approche basée sur la fréquence.

#### 4.5.1 Découpage des valeurs diagnostiques en classes

Le découpage préalable en classes doit essayer de produire des effectifs équilibrés. La question qui se pose est : est-il possible de définir des valeurs seuils homogènes pour l'ensemble des sites ? Pour cela, nous avons calculé les courbes empiriques cumulées de distribution (figures 29 et 30) respectivement pour l'engorgement et la fertilité.



**Figure 29** : courbes des distributions empiriques cumulées des valeurs diagnostiques d'engorgement par type de zone humide ; la courbe noire représente l'ensemble des données, indépendamment du type de zone humide.



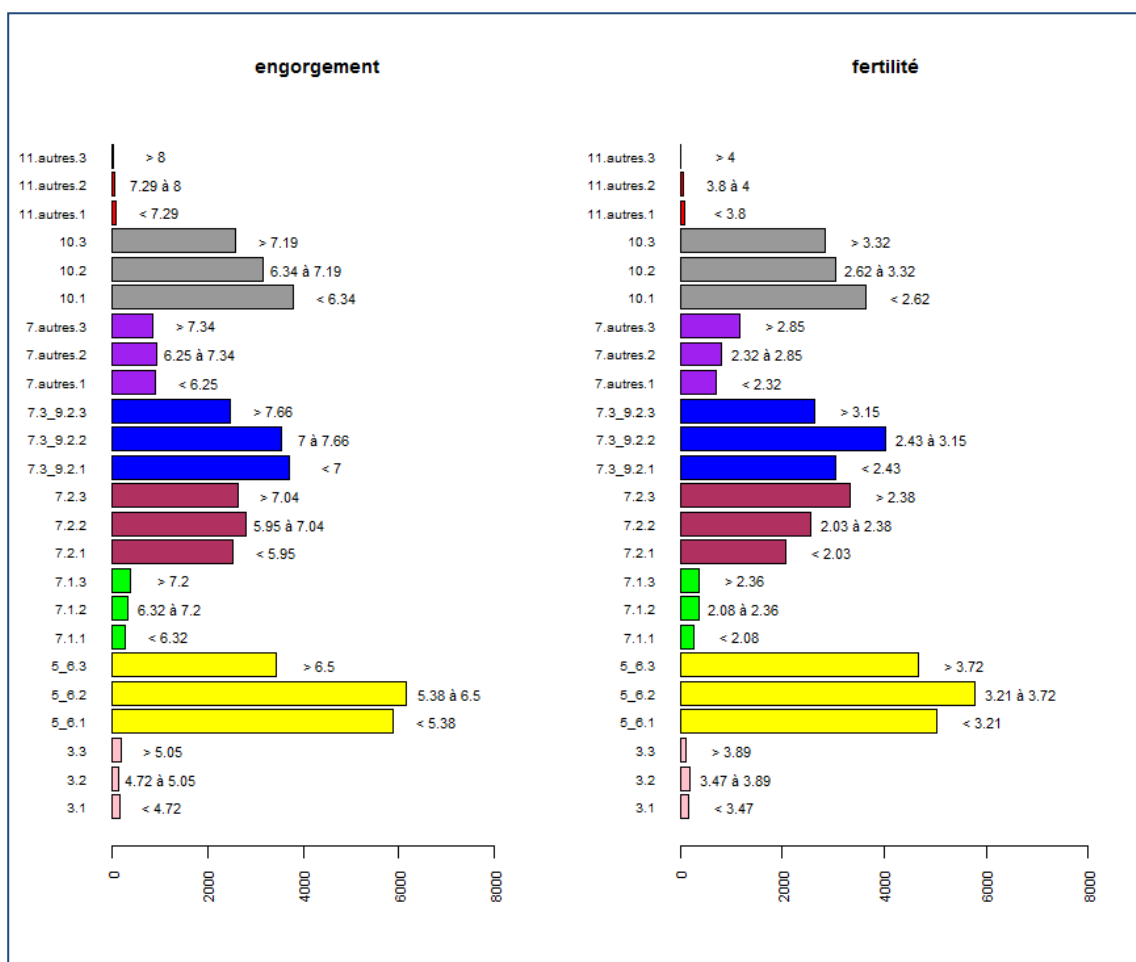
**Figure 30** : courbes des distributions empiriques cumulées des valeurs diagnostiques de fertilité par type de zone humide ; la courbe noire représente l'ensemble des données, indépendamment du type de zone humide.

Concernant l'engorgement, (figure 29), les courbes des différents types de zones humides sont assez regroupées, avec toutefois 4 exceptions notables. Le type 3, et dans une beaucoup plus faible mesure, le type 5\_6, présentent beaucoup plus de valeurs diagnostiques faibles que la moyenne. Les types 11.autres et 7.3\_9.2 présentent, eux beaucoup plus de valeurs diagnostiques élevées que la moyenne. Pour la fertilité (figure 30), on constate que les profils sont plus hétérogènes, mais que l'on peut les regrouper plus ou moins en trois ensembles : les zones humides de type 3, 5\_6 et 13 présentent peu de relevés avec des valeurs diagnostiques faibles ; les zones humides de type 7.3\_9.2, 7.autres, et 10 qui présentent des profils relativement équilibrés ; enfin, les zones humides de type 7.1 et 7.2, qui présentent beaucoup de placettes avec des valeurs diagnostiques faibles.

Le nombre optimal de classes doit être défini. Des règles empiriques plus ou moins reconnues (règle de STURGES, règle de SCOTT, règle de YULE, ...) existent, mais leur application pour l'engorgement sur l'ensemble des données suggère des nombres de classes de 26 (!) et 11 pour les deux premières. Une deuxième approche, basée sur les arbres de régression, suggère toujours pour l'engorgement, un nombre de classes de l'ordre de 8. Ces nombres élevés de classes aboutiraient à disposer de peu de relevés pour calculer les valeurs « IndVal » ou Phi, ce qui est pour le moins gênant.

Nous avons donc choisi de définir des valeurs seuils pour chaque type de zone humide. Afin de garder suffisamment de cas par niveau de seuil, nous avons également choisi de retenir 3 niveaux par type de zone humide avec des effectifs égaux en nombres de relevés. Les seuils choisis sont illustrés figure 31.

**Figure 31** : cumuls des nombres d'observations par sous-catégories (croisements des types de zones humides et des seuils de valeurs diagnostiques). A droite des barres donnant ces cumuls, les valeurs utilisées pour définir les seuils sont indiquées ; par exemple, la sous-catégorie 3.1 rassemble toutes les observations des relevés appartenant au type « 3 » ayant des valeurs diagnostiques d'engorgement < 4.72.



La catégorie 1 représente les relevés avec des valeurs diagnostiques faibles, la catégorie 3 ceux avec des valeurs fortes et la catégorie 2 ceux avec des valeurs moyennes. Avec 8 types de zones humides, séparées chacune en trois catégories, on a donc 24 sous-catégories ; chaque sous-catégorie est identifiée par un nom composé du nom du type de zone humide auquel on ajoute « .1 », « .2 » ou « .3 » selon le niveau de seuil. Les nombres d'observations par catégorie ne sont toutefois pas égaux car la richesse spécifique n'est pas homogène selon les valeurs diagnostiques en général.

Enfin, pour le calcul des valeurs indicatrices des espèces, nous avons écarté les espèces présentes moins de 5 fois sur le jeu de donnée ; 3 relevés ont été ainsi supprimés car toutes les espèces qu'ils contenaient étaient présentes moins de 5 fois. Après application du filtre, la table à analyser comporte donc 2.934 relevés pour 737 espèces, les relevés étant répartis en 24 catégories.

#### 4.5.2 Résultats pour l'engorgement

Les résultats sont rassemblés dans le tableau 9. En remarque préalable, signalons qu'il est en théorie possible de tester la significativité des valeurs Phi par des tests de permutations, mais la bibliothèque de fonctions utilisée pour cela (package 'indicspecies' pour R, DE CACERES & LEGENDRE, 2009) n'a pu mener les calculs à terme au bout de plus de 2 heures. De même, ce coefficient étant assez sensible au nombre de données, une version permettant de comparer entre eux des groupes avec des tailles différentes a été proposée par TICHY & CHYTRY (2006). Cette fonctionnalité implémentée dans le même package 'indicspecies' n'a elle aussi pas pu être mise en œuvre ; toutefois, ceci est moins problématique dans notre cas car nous cherchons avant tout des espèces indicatrices par sous-types.

Globalement, pour IndVal, avec un seuil de significativité à 10%, 266 espèces sont données comme indicatrices d'au moins 1 des 24 sous-groupes ; avec un seuil de significativité de 5%, ce nombre chute à 175. Toujours globalement, avec le coefficient Phi, ces nombres s'établissent à 324 et 144 pour des valeurs de Phi supérieures à 0,15 et 0,20, respectivement. Contrairement à IndVal où une espèce est considérée comme indicatrice du sous-type pour lequel sa valeur d'indice est la plus élevée, une espèce peut être considérée comme liée à plusieurs sous-types avec Phi ; si on ne prend en compte que la liaison la plus forte pour chaque espèce, ces nombres tombent à 200 avec Phi supérieur à 0,15 et 104 pour Phi supérieur à 0,2.

Si on regarde par type de zone humide, IndVal trouve entre 10 et 49 espèces (moyenne de 33,3) et entre 3 et 34 (moyenne de 21,9) pour des seuils de significativité à 10% et 5%, respectivement. Le type « 7.3\_9.2 » est, de loin, celui pour lequel le moins d'espèces indicatrices sont trouvées. La réduction du nombre d'espèces entraînée par la baisse du niveau de significativité est comprise entre 25% à 50%. Pour Phi, les nombres d'espèces s'échelonnent entre 7 à 92 (moyenne de 35,4) et 2 et à 39 (moyenne de 16,0) pour un seuil à 0,15 et 0,20, respectivement. De manière cohérente avec IndVal, le type de zone humide qui présente le moins d'espèces liées est le « 11.autres ». Le passage d'une limite de 0,15 à 0,20 entraîne une réduction du nombre d'espèces liées comprise entre 15% et 90%.

Par sous-types, le nombre d'espèces IndVal est compris entre 0 et 26 (pour une moyenne de 11,5) et entre 0 et 19 (pour une moyenne de 8,0) avec un seuil de significativité à 10% ou 5%, respectivement. Avec un seuil à 10%, 11 sous-types ont moins de 10 espèces indicatrices ; ce chiffre monte à 15 avec un seuil à 5%. Pour Phi, ces nombres s'établissent entre 1 et 45 (valeur moyenne de 13,5) et entre 0 et 17 (valeur moyenne de 6,2) pour des valeurs supérieures à 0,15 et 0,20, respectivement. La chute du nombre d'espèces liées est importante, et plus encore, beaucoup de sous-types n'ont que très peu d'espèces liées (<10 dans 17 cas sur 24) pour les valeurs supérieures à 0,20 (contre 11 cas sur 24 pour les valeurs supérieures à 0,15).

**Tableau 9 :** nombres d'espèces liées, selon la méthode IndVal ou selon les valeurs du coefficient Phi de PEARSON d'association à des sous-types ou à des types de zones humides, et ce pour 2 seuils pour chaque méthode (significativité à 10% et 5 % pour IndVal et Phi >0,15 et Phi>0,20 pour Phi), pour l'engorgement.

sous-type	Nb taxons IndVal 10%	Nb taxons IndVal 5%	Nb taxon par type IndVal 10%	Nb taxon par type IndVal 5%	Nb taxons Phi >0,15	Nb taxons Phi >0,2	Nb taxons par type Phi >0,15	Nb taxons par type Phi >0,2	Choix final IndVal	Choix final Phi
3.1	12	10	36	29	8	5	15	13	36 espèces - IndVal 10%	26 espèces - Phi >0.1
3.2	8	8			8	7				
3.3	16	11			5	5				
5_6.1	23	13	47	27	45	17	92	39	47 espèces - IndVal 10%	39 espèces - Phi >0.2
5_6.2	11	7			33	12				
5_6.3	13	7			24	13				
7.1.1	17	9	49	34	8	4	22	11	49 espèces - IndVal 10%	45 espèces - Phi >0.1
7.1.2	15	12			6	2				
7.1.3	17	13			11	6				
7.2.1	14	10	29	13	26	11	54	32	40 espèces - IndVal 10% et 15 espèces par sous-type	41 espèces - Phi >0.18
7.2.2	5	2			20	14				
7.2.3	10	1			21	13				
7.3_9.2.1	4	1	10	3	12	2	30	8	32 espèces - IndVal 10% et 15 espèces par sous-type	30 espèces - Phi >0.15
7.3_9.2.2	0	0			13	1				
7.3_9.2.3	6	2			11	5				
7.autres.1	24	19	34	24	13	6	21	9	43 espèces - IndVal 10% et 15 espèces par sous-type	42 espèces - Phi >0.11
7.autres.2	8	4			7	2				
7.autres.3	2	1			3	2				
10.1	26	14	35	18	27	11	42	14	38 espèces - IndVal 10% et 15 espèces par sous-type	42 espèces - Phi >0.15
10.2	3	1			4	0				
10.3	6	3			11	3				
11.autres.1	19	15	24	20	6	1	7	2	33 espèces - IndVal 10% et 15 espèces par sous-type	31 espèces - Phi >0.05
11.autres.2	2	2			1	1				
11.autres.3	3	3			1	1				

Pour certains sous-types, le nombre d'espèces liées (sensu IndVal ou Phi) est toujours très faible, et il semble très peu probable de pouvoir, avec des listes aussi restreintes, reproduire les valeurs diagnostiques issues de l'ensemble des données : c'est le cas de des sous-types « 7.autres.2 », « 7.autres.3 », « 10.2 », « 10.3 », « 11.autre.2 » et « 11.autres.3 ». Les sous-types « 7.2.2 » et « 7.2.3 » sont mal pris en compte avec IndVal mais bien avec Phi ; la situation est inversée pour les sous-types « 3.1.1 », « 3.1.2 », « 3.1.3 », « 7.1.1 » et « 7.1.2 ». Enfin, cas les plus rares, pour certains sous-types, le nombre d'espèces liées semble suffisant dans tous les cas, comme pour les sous-types « 5\_6.1 », « 10.1 » et dans une moindre mesure « 5\_6.3 », « 7.2.1 », « 7.2.3 » et « 7.autres.1 ».

Au final, on peut tirer les conclusions suivantes :

- la méthode avec Phi >0,2 est globalement à écarter car elle fournit trop peu d'espèces ; de plus, seuls 7 sous-types ont plus de 10 espèces liées ;

- la méthode IndVal avec un seuil de significativité à 5% doit également l'être, pour les mêmes raisons ;
- les deux autres méthodes sont assez proches, chacune fournissant des listes avec plus de 10 taxons par sous-types pour un nombre équivalent de sous-types ; toutefois, la méthode IndVal avec un seuil à 10% présente l'avantage **d'avoir des nombres d'espèces par type de zone humide plus équilibrés** : seul le type « 7.3\_9.2 » et ses sous-types seront probablement très mal ajustés ; avec Phi, deux types sont mal représentés (« 3 » et « 11.autres »), alors que le type « 5\_6 » possède une liste très allongée.

Un des reproches que l'on peut formuler à l'encontre d'IndVal est qu'une espèce n'est indicatrice que d'un seul sous-type, celui où IndVal est maximum. Si cette pratique permet de bien caractériser des sous-types, notre problématique est légèrement différente car il s'agit avant tout de reproduire aux mieux les données obtenues avec l'ensemble du jeu de données. Or, beaucoup d'espèces sont communes à différents **types** de zones humides (car beaucoup de zones humides ont un fond floristique commun assez étoffé), et un cas assez fréquent est qu'une espèce est liée à plusieurs sous-types appartenant à des types différents (par exemple, les sous-types riches (.3) en nutriments des bas-marais alcalins et des marais ont beaucoup d'espèces communes) ; en ne retenant que la liaison maximale, on se prive pour certains sous-types d'espèces qui seraient utiles ; c'est particulièrement vrai pour les sous-types où finalement, en partie à cause de ce mode de sélection final, peu d'espèces sont retenues. Un autre reproche, dans notre application, est qu'IndVal est le produit de la concentration des recouvrements dans le sous-type par la fréquence relative ; or, le premier terme est souvent faible par rapport au second, notamment car **nous avons choisi d'analyser l'ensemble des données**. Cela pénalise tout particulièrement un ensemble d'espèces peu abondantes mais assez fréquentes dans des sous-types, et qui pourraient être très utiles pour prédire l'engorgement. En revanche, ce choix d'une analyse globale, par rapport à des analyses par type de zone humide, présente l'avantage de limiter la mise en exergue d'espèces qui ne sont pas vraiment indicatrices mais simplement mise en avant à cause des hasards de l'échantillonnage (ce phénomène est d'autant plus marqué que la taille de l'échantillon est faible, donc dans le cas d'analyses séparées où l'on fractionne les données).

A titre d'exemples, prenons les cas des types « 7.3\_9.2 » et « 10 ». Pour le premier, seules 8 espèces sont retenues avec IndVal avec un seuil de significativité à 10%. Or, si on prend les 30 espèces ayant les valeurs IndVal les plus élevées, par exemple pour le sous-type « 7.3\_9.2.1 », on constate qu'une seule a une concentration supérieure à 0.01 alors que 5 espèces ont des fréquences supérieures à 0.3 (maximum de 0.62) et 14 pour lesquelles cette fréquence est supérieure à 0.2. Donc, même si dans ce sous-type peu d'espèces sont vraiment fréquentes, ce qui les pénalise surtout ce sont les valeurs faibles des concentrations ; la prise en compte d'un panel plus large d'espèces permettrait (probablement) d'améliorer sensiblement la qualité des prédictions obtenues avec les listes filtrées. Le même constat se répète pour le sous-type « 7.3\_9.2.2 », avec quelques nuances : 1 espèce (*Phragmites australis*) possède une concentration non négligeable (0.17), mais 12 espèces ont des fréquences relatives supérieures à 0.3 (maximum 0.60). Globalement, le même constat est valable pour le sous-type « 7.3\_9.2.3 » ou l'ensemble des sous-types du type « 10 ».

Nous proposons donc de retenir, pour la méthode IndVal, les listes d'espèces avec un seuil de significativité à 10%. Nous proposons également d'amender les listes de manière à assurer qu'au moins trente espèces par type de zone humide soient identifiées tout en essayant d'assurer un certain équilibre entre les différents sous-types par type. Pour cela, nous proposons de retenir en premier lieu, par sous-type, les 15 espèces ayant les valeurs IndVal les plus élevées. Cette procédure permet de conserver les espèces ayant les valeurs IndVal les plus élevées (à de rares exceptions près, lorsque le nombre de taxons par sous-type est supérieur à 15) par sous-types et proposées par la méthode, tout en permettant un bien meilleur équilibre des nombres d'espèces, tant au niveau des types de zones humides que des sous-types. Globalement, une cinquantaine d'espèces sont sélectionnées pour la plupart des types de zones humides et l'équilibre entre les sous-types est respecté.

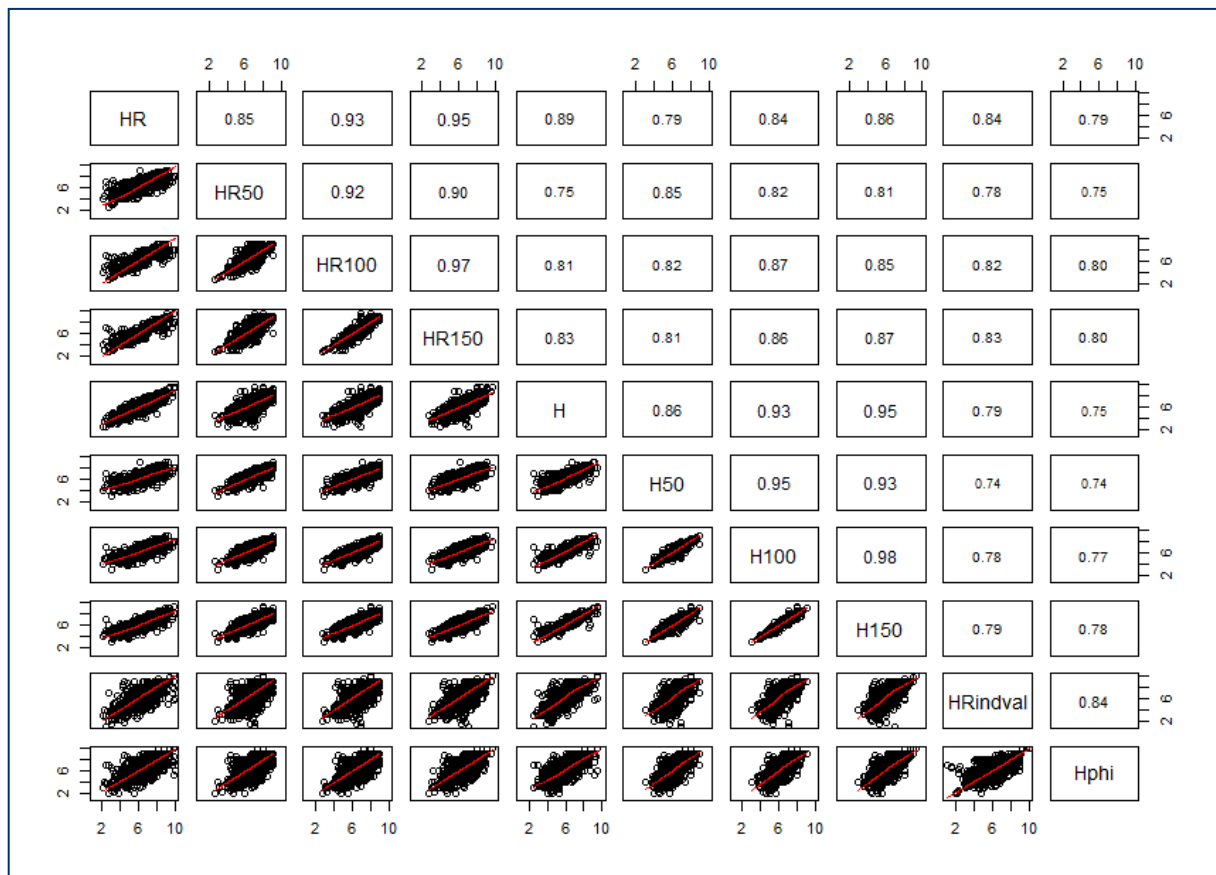
Au final, ce sont entre 32 et 49 espèces qui sont retenues par type de zone humide, soit un total de 290 sur l'ensemble des types testés.

Avec la méthode utilisant Phi, les mêmes types de problèmes sont posés, avec la difficulté supplémentaire que la valeur de Phi est assez conditionnée par le nombre de relevés par sous-type. Afin de confronter, malgré tout, les résultats avec les données complètes (car, rappelons-le, avec Phi on pourrait s'affranchir du besoin des recouvrements des espèces), nous avons gardé cette méthode. Il faut toutefois choisir des seuils par type de zone humide (les valeurs retenues sont fournies tableau 9). Ces valeurs de seuils ont été choisies de manière à fournir des nombres d'espèces comparables à ceux d'IndVal tout en assurant environ 10 à 15 espèces par sous-type ; toutefois, cette dernière condition n'a pas pu être toujours respectée, faute de quoi le nombre global d'espèces par type aurait été trop élevé (c'est le cas surtout pour les types 10 et 11).

### 4.5.3 Validation des résultats pour l'engorgement

La figure 32 donne les corrélations sur l'ensemble des relevés entre les valeurs calculées avec les différentes méthodes de filtre ou avec les valeurs non filtrées.

**Figure 32** : relations entre les valeurs calculées avec l'ensemble des données (HR et H), en ne retenant que 50 espèces (HR50 et H50), 100 (HR100 et H100) ou 150 espèces (HR150 et H100), et celles calculées avec Indval (HRindval) et avec Phi (Hphi) pour l'engorgement. Les valeurs dans le triangle supérieur sont les coefficients de corrélations ; les variables « HR » correspondent aux valeurs calculées avec pondération par le recouvrement, les variables « H » à celles sans pondération.

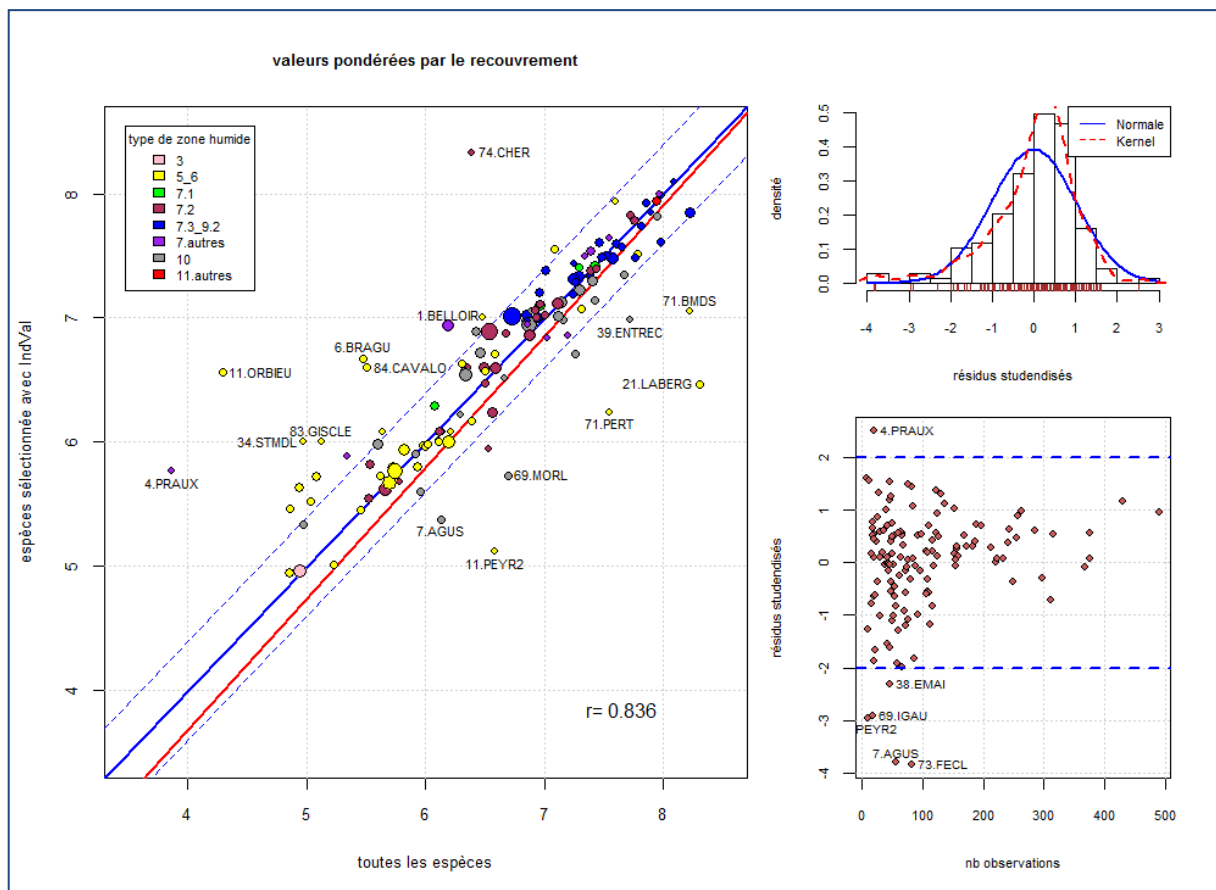


On constate que les corrélations entre HRindval et l'ensemble des valeurs sont moyennes à bonnes ; le minimum est obtenu pour le cas H50, c'est-à-dire en filtrant sur les 50 espèces les plus fréquentes et sans prendre en compte de recouvrement ; le maximum est lui obtenu avec HR, c'est-à-dire avec l'ensemble des espèces et prise en compte du recouvrement. Toutefois, les corrélations sont moins bonnes qu'avec HR100 et HR150 et équivalentes à celle entre HR et HR50.

Les corrélations entre Hphi et l'ensemble des autre valeurs sont dans l'ensemble les plus faibles.

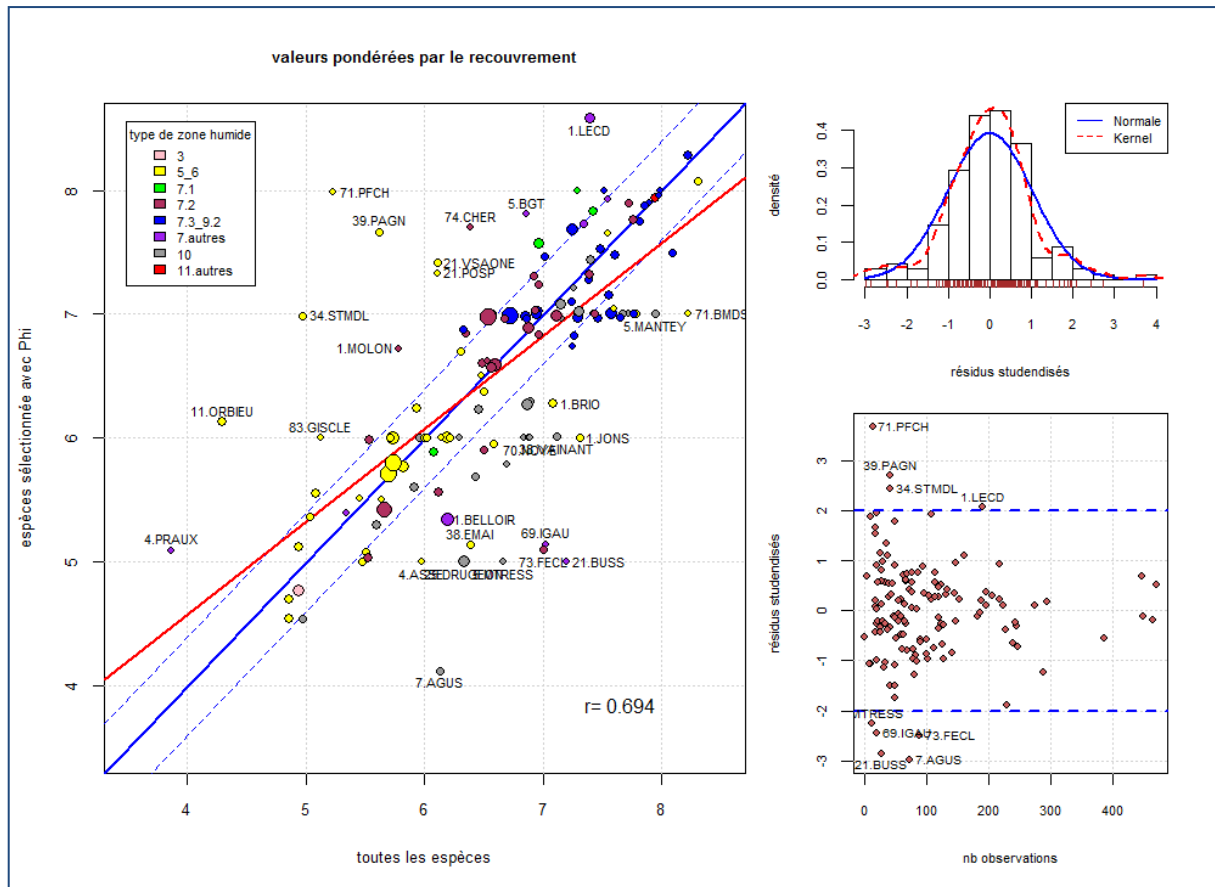
A l'échelle des sites (figure 33), la corrélation est assez bonne entre les valeurs médianes obtenues avec le filtre IndVal et celles obtenues avec l'ensemble des données ( $r = 0,836$ ) ; par ailleurs, un nombre important de sites sont situés en dehors des lignes matérialisant l'erreur empirique de  $\pm 0.4$ . Les résidus studentisés sont légèrement dissymétriques (quelques valeurs négatives fortes mais beaucoup de valeurs positives). Un très léger biais est visible dans les prédictions.

**Figure 33** : relation entre les valeurs observées pour l'engorgement avec l'ensemble des données sur chaque site (en abscisses) et les valeurs prédites avec les seules données filtrées sur les espèces retenues avec IndVal par type de zone humide (en ordonnées) ; la taille des points est proportionnelle au nombre d'observations, la droite rouge est la droite de régression, la droite bleue pleine matérialise le cas où les valeurs prédites sont égales aux valeurs observées, les deux droites bleues en tirets représentant l'enveloppe de l'erreur empirique ( $\pm 0.4$ ).



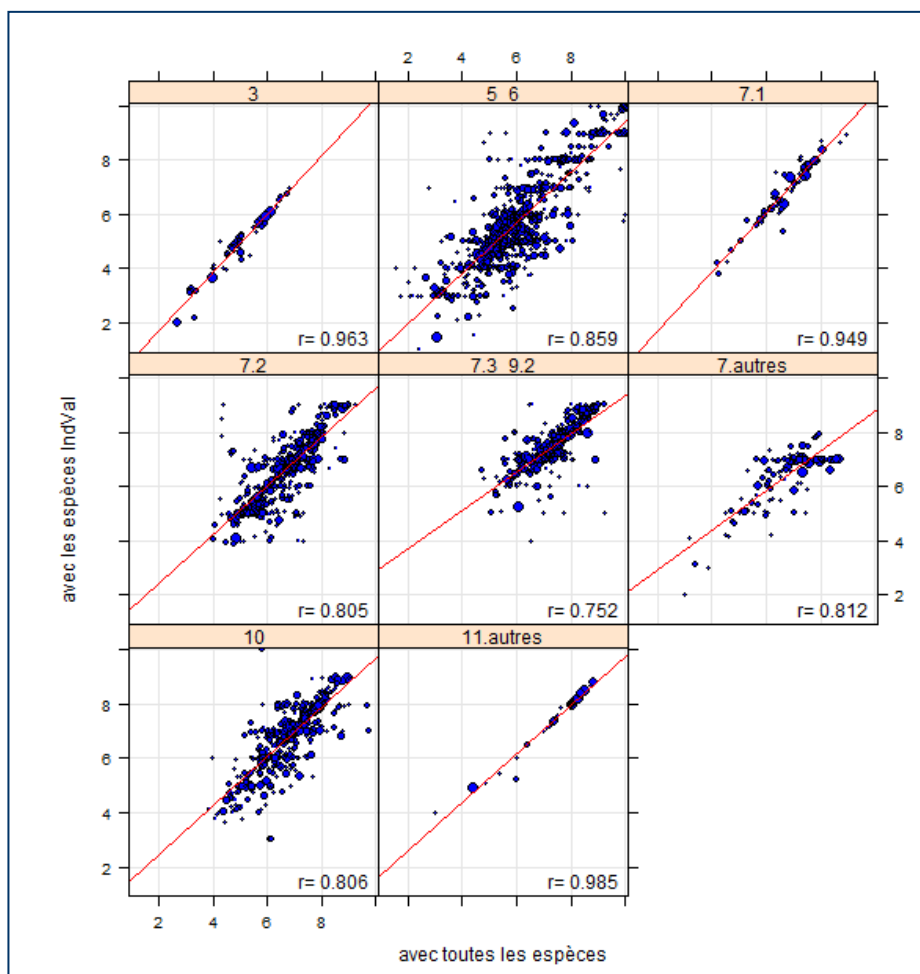
Globalement, avec des nombres d'espèces par type de zone humide inférieurs, les performances d'IndVal sont comparables avec le filtre de rareté à 50 et inférieures à celles des filtres de rareté à 100 ou 150 espèces. Les prédictions sont faiblement biaisées mais beaucoup de sites des milieux alluviaux sont sous-estimés (contrairement au filtre à 50 qui a tendance à les surestimer) et en dehors de l'enveloppe de l'erreur empirique des  $\pm 0.4$ .

**Figure 34** : relation entre les valeurs observées pour l'engorgement avec l'ensemble des données sur chaque site (en abscisses) et les valeurs prédites avec les seules données filtrées sur les espèces retenues avec Phi par type de zone humide (en ordonnées); la taille des points est proportionnelle au nombre d'observations, la droite rouge est la droite de régression, la droite bleue pleine matérialise le cas où les valeurs prédites sont égales aux valeurs observées, les deux droites bleues en tirets représentant l'enveloppe de l'erreur empirique ( $\pm 0.4$ ).



Avec Phi (figure 34), la corrélation est très médiocre ( $r = 0,694$ ), les sites étant très dispersés. Près du tiers d'entre eux sont situés en dehors des courbes des erreurs empiriques. En outre, on remarque que la plupart des sites sont représentés par peu de données ; en effet, le filtre avec le coefficient Phi entraîne la suppression de 33.093 observations sur les 46.926 disponibles (soit 70,5%). De plus, les valeurs prédites sont assez biaisées et 10 sites présentent des résidus studentisés supérieurs à  $\pm 2$ . **Un tel modèle ne peut être retenu.**

La figure 35 détaille un peu plus la situation par type de zone humide à l'échelle des relevés, pour l'engorgement. Pour les types avec peu de relevés (3, 11.autres et 7.1 dans une moindre mesure), les corrélations sont bonnes mais il ne faut pas sur-interpréter ces corrélations étant donné qu'elles sont assises sur peu de valeurs. Les valeurs diagnostiques des relevés avec données complètes sont assez bien à médiocrement prédites avec les espèces filtrées par IndVal. Les types 5\_6 (milieux alluviaux), 10 (marais) le sont malgré le grand nombre de relevés, et malgré un certain étalement des valeurs prédites. On remarque toutefois pour les milieux alluviaux quelques relevés mal prédits, alors qu'il s'agit de relevés assez riches en espèces. Les types 7.3\_92 (tourbières alcalines) et 7.autre (sources et suintements) se distinguent également car les prédictions sont assez biaisées.



**Figure 35** : relation entre les valeurs observées **pour l'engorgement** avec l'ensemble des données sur chaque relevé (en abscisses) et les valeurs prédites avec les seules données filtrées sur les espèces retenues avec IndVal par type de zone humide (en ordonnées) ; la taille des points est proportionnelle au nombre d'observations, la droite rouge est la droite de régression.

Avec un filtre basé sur IndVal, on obtient pour l'engorgement des résultats similaires à ceux obtenus avec le filtre basé sur les 50 espèces les plus courantes, et moins bons qu'avec ceux basés sur 100 ou 150 espèces. Ces résultats sont toutefois obtenus avec un peu moins d'espèces (une quarantaine), ce qui est encourageant. Il faut également les prendre particulièrement prudemment pour les types 3, 11 et dans une moindre mesure 7.1 car peu d'observations sont disponibles pour ces types. Enfin, les prédictions sont quasiment systématiquement sous-estimées pour le type 5\_6 (milieux alluviaux). Une autre limite non analysée dans le détail est que certaines espèces données comme indicatrices pour certains sous-types soient en fait des espèces plutôt occasionnelles, rudérales, ..., dont les liens identifiés reflètent plus les hasards de l'échantillonnage qu'une véritable liaison ; toutefois, nous avons essayé de limiter ces cas en supprimant les espèces présentes moins de 5 fois.

En revanche, les résultats obtenus avec le coefficient Phi sont décevants. Avec un nombre restreint d'espèces sélectionnées par cet indice, les performances sont globalement nettement moins bonnes qu'avec le filtre de rareté à 50 et qu'avec IndVal. Si on veut donc ne noter que la seule présence des espèces, il vaut mieux s'en tenir au filtre de rareté.

#### 4.5.4 Résultats pour la fertilité

Les résultats sont rassemblés tableau 10.

**Tableau 10** : nombres d'espèces liées, selon la méthode IndVal ou selon les valeurs du coefficient Phi de PEARSON d'association à des sous-types ou à des types de zones humides, et ce pour 2 seuils pour chaque méthode (significativité à 10% et 5 % pour IndVal et Phi >0,15 et Phi>0,20 pour Phi), pour la fertilité.

sous-type	Nb taxons IndVal 10%	Nb taxons IndVal 5%	Nb taxon par type IndVal 10%	Nb taxon par type IndVal 5%	Nb taxons Phi >0,15	Nb taxons Phi >0,2	Nb taxons par type Phi >0,15	Nb taxons par type Phi >0,2	Choix final IndVal	Choix final Phi
3.1	12	8	39	32	8	5	19	11	39 espèces - IndVal 10%	26 espèces - Phi >0.1
3.2	18	15			14	8				
3.3	9	8			7	2				
5_6.1	14	8	37	21	17	9	46	21	40 espèces - IndVal 10% et 10 espèces par sous-type	46 espèces - Phi >0.15
5_6.2	3	1			8	0				
5_6.3	20	12			25	12				
7.1.1	15	13	49		8	4	20	8	49 espèces - IndVal 10%	36 espèces - Phi >0.1
7.1.2	14	19			12	2				
7.1.3	13	18			3	3				
7.2.1	14	11	33	24	26	20	50	33	38 espèces - IndVal 10% et 10 espèces par sous-type	33 espèces - Phi >0.2
7.2.2	7	4			19	14				
7.2.3	12	9			17	7				
7.3_9.2.1	8		20	10	23	12	40	17	37 espèces - IndVal 10% et 10 espèces par sous-type	40 espèces - Phi >0.15
7.3_9.2.2	8				14	3				
7.3_9.2.3	4				6	3				
7.autres.1	14		34	27	6	2	18	7	34 espèces - IndVal 10% et 10 espèces par sous-type	46 espèces - Phi >0.1
7.autres.2	13				10	4				
7.autres.3	11				6	2				
10.1	13	9	29	17	22	12	30	15	40 espèces - IndVal 10% et 10 espèces par sous-type	40 espèces - Phi >0.14
10.2	8	4			5	2				
10.3	8	4			3	1				
11.autres.1	20	16	23	19	4	2	5	3	30 espèces - IndVal 10% et 10 espèces par sous-type	21 espèces - Phi >0.07
11.autres.2	3	3			1	1				
11.autres.3										

Comme pour l'engorgement, un seuil Indval à 5% ne permet pas un bon équilibre des nombres d'espèces entre sous-types ; la situation est améliorée avec un seuil à 10% mais ces nombres restent très faibles pour les types 10 et 11.autres tout particulièrement. Nous avons donc, comme pour l'engorgement rallongé ces listes de manière à les équilibrer en essayant d'assurer au moins 10 espèces par sous-type ; notre intention initiale était d'atteindre, comme pour l'engorgement, 15 espèces, mais cela s'est avéré impossible sans conduire à une inflation très importante du nombre de taxons, sauf à repêcher pour des sous-types des espèces ayant des valeurs Indval très faibles. Au final, entre 30 et 49 espèces ont été sélectionnées par sous-type. Comme pour l'engorgement, nous avons calculé le coefficient Phi entre l'ensemble des espèces et les sous-types et sélectionné par sous-type celles ayant les valeurs Phi les plus élevées. Au seuil Phi >0,2, aucun type hormis le

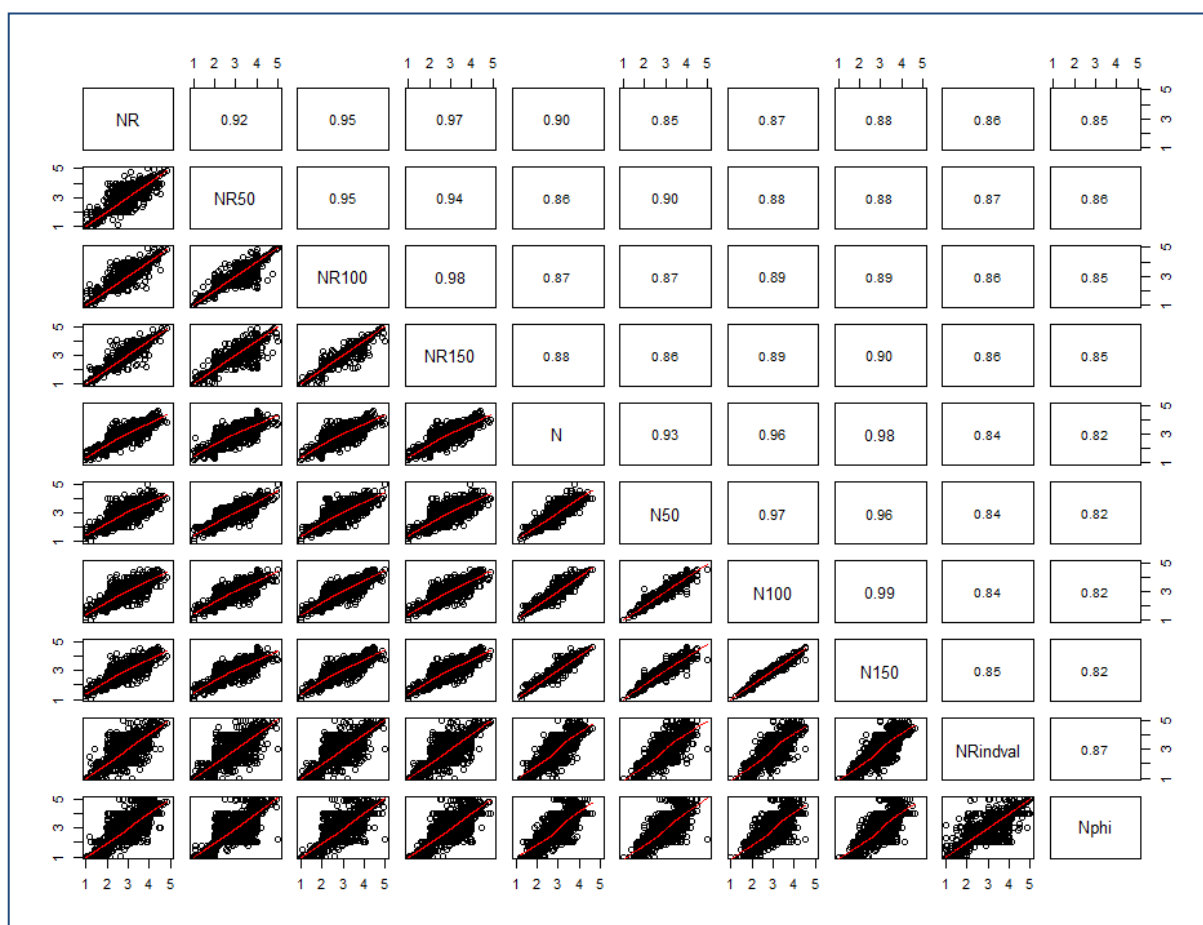
7.2 (tourbières acides) n'est suffisamment pourvu en espèces. Le constat est quasiment le même avec un seuil Phi à 0,1. Pour obtenir des espèces par sous-type en nombres suffisants et relativement équilibrés, il a fallu encore abaisser les valeurs de Phi (0,1 voire 0,7).

Avec un seuil Indval à 10%, 274 taxons différents sont retenus, mais l'équilibre entre les sous-types n'est pas toujours bon. La somme des valeurs Indval correspondante est de 32,46 dont 24,34 pour les seuls taxons significativement liés au seuil 10%. Pour l'engorgement, ces chiffres sont de 35,83 et 25,90 pour un nombre de taxons légèrement inférieur. **Il y a donc une meilleure discrimination des sous-types pour l'engorgement que pour la fertilité.**

#### 4.5.5 Validation des résultats pour la fertilité

Les corrélations entre les différentes méthodes de filtre sont illustrées figure 36. On constate que, comme ce que l'on observe pour l'engorgement, celles avec les données complètes avec prise en compte du recouvrement (NR) sont les plus faibles pour Indval et Phi. Sans être mauvaises, elles sont globalement comprises entre 0,84 et 0,87 pour Indval et 0,82 et 0,87 pour Phi. Avec la valeur de référence (NR), les deux corrélations sont de 0,86 et 0,85 respectivement.

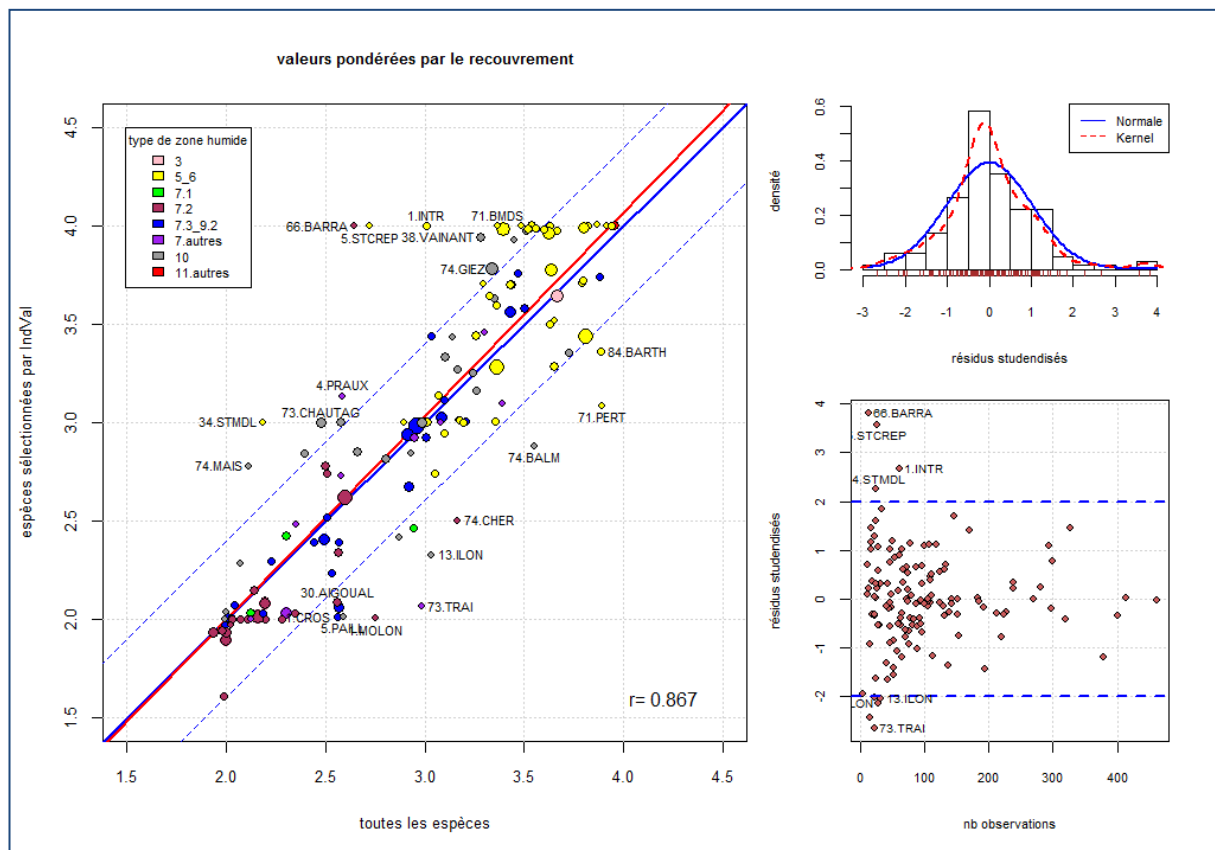
**Figure 36** : relations entre les valeurs calculées avec l'ensemble des données (NR et N), en ne retenant que 50 espèces (NR50 et N50), 100 (NR100 et N100) ou 150 espèces (NR150 et N100), et celles calculées avec Indval (NRindval) et avec Phi (Nphi) pour la fertilité. Les valeurs dans le triangle supérieur sont les coefficients de corrélations ; les variables « NR » correspondent aux valeurs calculées avec pondération par le recouvrement, les variables « N » à celles sans pondération.



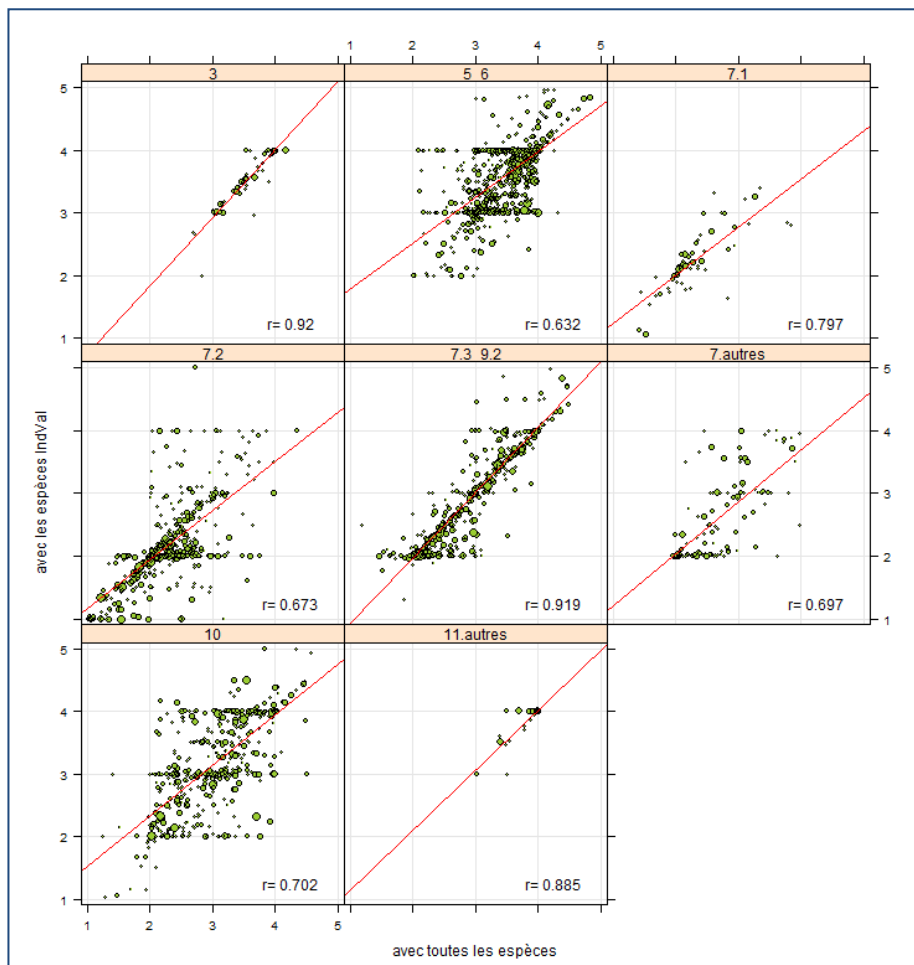
A l'échelle des sites (figure 37), la corrélation entre les valeurs calculées avec le filtre IndVal et les valeurs obtenues avec l'ensemble des espèces et en prenant en compte le recouvrement, est bonne sans plus ( $r = 0,867$ ). La droite de régression ne montre pas de biais net. Les résidus de la régression sont à peu près distribués normalement, malgré quelques résidus positifs très élevés. On note 9 résidus studentisés supérieurs en valeur absolue à 2, ce qui est déjà beaucoup. Ces sont surtout des sites des types 5\_6 (milieux alluviaux) et 10 (marais) qui sont mal prédits ; toutefois, on ne retrouve pas comme pour l'engorgement une tendance aussi nette à sur-ou-sous-estimer les sites alluviaux. Les sites de Barraux et Saint-Crépin s'écartent le plus du modèle. Les prédictions sont en revanche bonnes pour les tourbières au sens large (type 7). On note, comme c'est la règle générale, que le nombre d'observations disponibles (une fois le filtre IndVal mis en place) par site a un effet favorable sur les prédictions : tous les sites mal prédits ne disposent que de moins de 80 observations.

Les résultats avec Phi (non montrés) sont du même ordre de grandeur ( $r = 0,866$ ), avec toutefois des résidus plus étalés et quelques valeurs vraiment mal prédites (notamment les sites du Pertuis, de Barraux à nouveau). On retrouve également une tendance pour les sites de type 5\_6 (milieux alluviaux) et 10 (marais) à être les moins bien prédits. On trouve un léger biais, un peu plus marqué qu'avec IndVal, traduisant une tendance à surestimer les valeurs fortes de fertilité, particulièrement pour les sites de type 5\_6 (milieux alluviaux). L'effet du nombre d'observations sur la qualité des prédictions se retrouve également, les sites mal prédits ayant en général moins de 80 observations ; signalons toutefois l'exception du site du Dugeon qui rest mal prédit (avec une sous-estimation) malgré 200 observations disponibles.

**Figure 37** : relation entre les valeurs observées **pour la fertilité** avec l'ensemble des données sur chaque site (en abscisses) et les valeurs prédites avec les seules données filtrées sur les espèces retenues avec IndVal par type de zone humide (en ordonnées) ; la taille des points est proportionnelle au nombre d'observations, la droite rouge est la droite de régression, la droite bleue pleine matérialise le cas où les valeurs prédites sont égales aux valeurs observées, les deux droites bleues en tirets représentant l'enveloppe de l'erreur empirique ( $\pm 0.4$ ).



Concernant la fertilité à l'échelle des relevés et toujours par type de zone humide, la figure 38 montre des corrélations nettement moins bonnes, hormis pour les types peu représentés : 3 (péri-lagunaire), 7.1 (tourbières acides) et 11.autres (zones humides ponctuelles), pour lesquels la pression d'échantillonnage est probablement trop faible pour tirer des conclusions solides. Pour les types avec beaucoup de relevés disponibles, la situation est la moins bonne pour le type 5\_6 (milieux alluviaux), 7.2 (tourbières alcalines) et 10 (marais). Pour le type 5\_6 (milieux alluviaux), les relevés sont assez dispersés sans structuration nette ; pour les tourbières alcalines, on remarque un gros ensemble de relevés bien alignés (plus ou moins) le long de la bissectrice et un ensemble de points très dispersés ; beaucoup de ces relevés (mais pas tous !) sont des relevés pauvres en espèces.

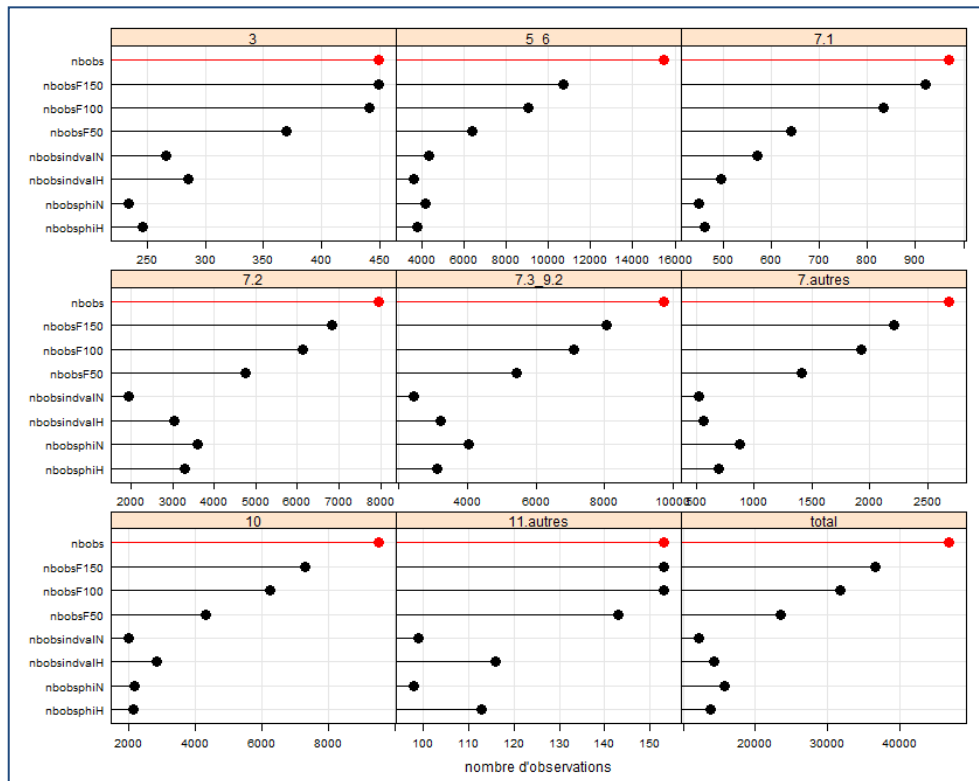


**Figure 38** : relation entre les valeurs observées pour la fertilité avec l'ensemble des données sur chaque relevé (en abscisses) et les valeurs prédites avec les seules données filtrées sur les espèces retenues avec IndVal par type de zone humide (en ordonnées) ; la taille des points est proportionnelle au nombre d'observations, la droite rouge est la droite de régression.

On remarque également, particulièrement pour les types 5\_6 et 10, une tendance des valeurs prédites à être concentrées sur des valeurs diagnostiques entières. Cela s'explique probablement par le fait que peu d'espèces ont été retenues et par la contrainte (via le découpage préalable des valeurs diagnostiques en classes ordonnées) de retenir des espèces ayant des valeurs indicatrices proches des centres des classes : ainsi, par exemple, pour le type 5\_6 (milieux alluviaux), le premier seuil pour la fertilité est à 3,26 ; IndVal a donc eu tendance à sélectionner des espèces ayant des valeurs indicatrices de 1, 2 ou 3, or, les espèces ayant une valeur de 3 sont beaucoup plus fréquentes que les deux autres, et in fine, ce sont essentiellement des espèces ayant une valeurs de 3 qui sont retenues ; il arrive souvent, comme peu d'espèces sont retenues par sous-type, que l'ensemble des espèces d'un relevés aient donc la même valeur (de 3 dans notre exemple), ce qui entraîne un nombre important de valeurs diagnostiques prédites à 3. Ce phénomène est moins sensible pour l'engorgement car les valeurs indicatrices sont étalées sur 10 niveaux.

## 4.6 Analyse des listes d'espèces

L'ensemble des données relatives aux effets des filtres et aux degrés de cohérences des listes produites par les méthodes est donné en annexe 1, les listes des espèces en elles-mêmes étant données en annexe 2.

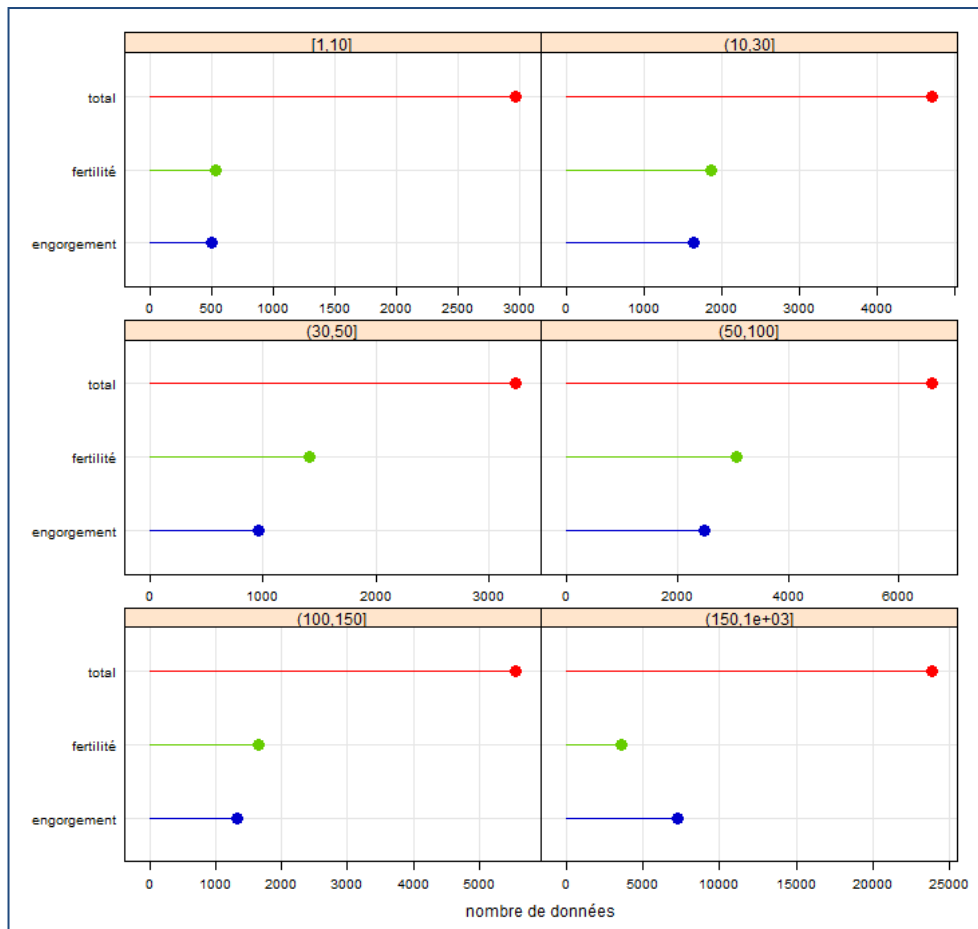


**Figure 39 :** nombres d'observations par type de zone humide et au total, selon les méthodes de filtre (F50, F100 et F150 = filtre par la rareté ; IndVal H et N = filtre par l'engorgement et la fertilité ; phi H et N = filtre avec Phi sur l'engorgement et la rareté).

Concernant les nombres d'observations (figure 39), l'effet global des filtres (panel d'en bas et à droite) sur la fréquence des espèces entraîne la suppression d'entre la moitié (pour le filtre à 50) à un cinquième des données (pour le filtre à 150), soit entre 24.000 et 36.000 observations utilisées environ. Les filtres avec IndVal ou Phi ne laissent que moins de 15.000 observations pour l'engorgement ; les nombres d'observations restantes sont encore légèrement plus faibles pour la fertilité avec ces deux méthodes. Par type de zone humide, les types les moins fournis en données sont globalement moins touchés (3 et 11.autres surtout). Sinon, dans l'ensemble, le patron général se retrouve dans les autres types. Le filtre le plus drastique est souvent IndVal pour la fertilité dans les milieux tourbeux (7.2, 7.3\_9.2 et 7.autres).

Pour Phi, il aurait été intéressant de pouvoir le calculer avec la formule modifiée de TICHY & CHYTRY (2006) permettant de prendre en compte les différences de données par types. L'effet contexte a également probablement joué : le fait d'évaluer le caractère indicateur, avec IndVal et surtout avec Phi (qui rappelons-le ne prend que les présences des espèces en compte), sur l'ensemble du jeu de données, et non pas par type de zone humide, a pesé sur la sélection d'espèces et favorisant trop des espèces globalement assez rares. Toutefois, ce choix a été fait pour essayer de trouver des espèces indicatrices ayant la portée la plus générale possible. Il ne fallait donc pas que les listes produites soient trop sensibles à l'échantillon de sites que nous avons, pour espérer pouvoir les transposer sur d'autres sites : **c'est d'ailleurs là une piste de travail et d'amélioration, qui consisterait à évaluer de nouveaux sites non pris en compte dans cette calibration. Une autre piste de travail consisterait également à ré-échelonner les valeurs indicatrices pour la fertilité sur 10 niveaux comme pour l'engorgement.**

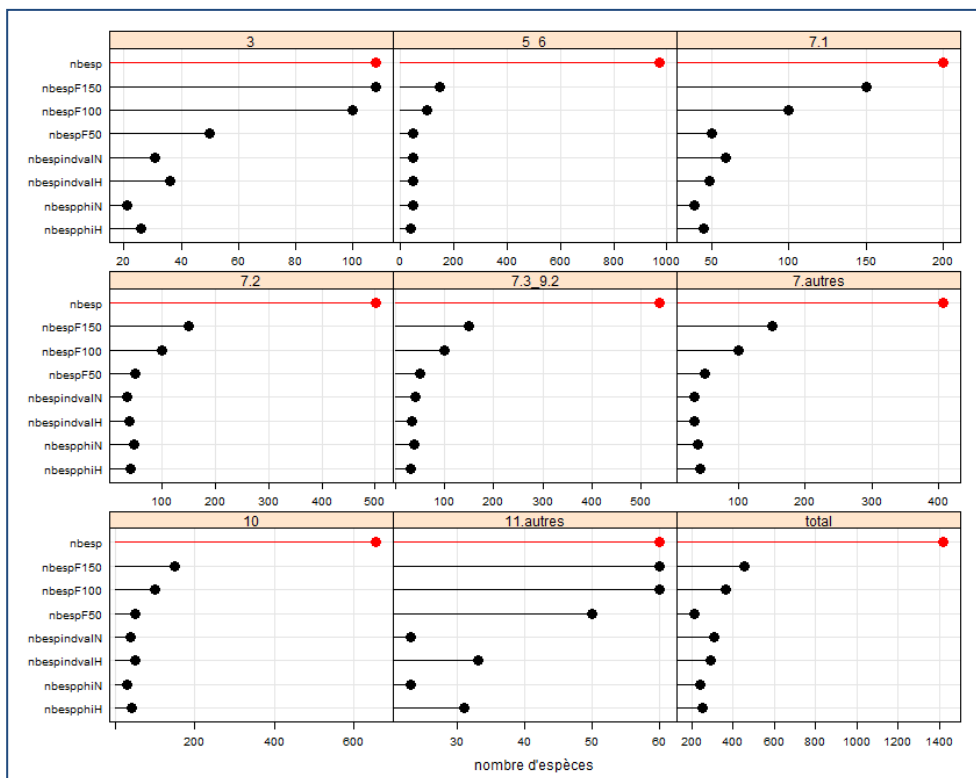
La question d'éventuels biais de sélection des espèces avec les patrons généraux de rareté des espèces a été envisagée et est présentée figure 40 pour IndVal à titre d'illustration. Les espèces ont été classées en 5 classes de rareté sur l'ensemble du jeu de données (présentes de 1 à 10 fois, de 10 à 50, ..., plus de 150 fois) et le nombre d'occurrences dans cette classe de rareté (en rouge) ou d'occurrences sélectionnées avec les espèces retenues par IndVal pour l'engorgement (en bleu) ou la fertilité (en vert) ont été comptés.



**Figure 40 :** nombres d'observations par classe de rareté globale pour les lots sélectionnés avec indVal pour l'engorgement (en bleu), pour la fertilité (en vert) et pour l'ensemble des données (en rouge).

Pour les classes de rareté intermédiaires (30 à 100 présences), IndVal a sélectionné relativement plus d'espèces (et donc d'observations) pour la fertilité que pour l'engorgement. Cette hiérarchie est inversée pour les espèces les plus fréquentes (présentes plus de 150 fois). Un autre fait intéressant est que, en proportion, IndVal a sélectionné plus d'observations parmi les espèces de rareté plutôt intermédiaire, mais beaucoup moins pour les espèces rares (ce qui n'est pas nécessairement un problème) et surtout pour les espèces les plus courantes (ce qui peut l'être).

Concernant les nombres d'espèces (figure 41), les réductions globales sont comprises entre un facteur 3 et un facteur 6 par rapport aux 1.420 espèces du jeu de données analysé. Au niveau global, entre 210 et 290 espèces sont sélectionnées, hormis dans le cas des filtres sur la rareté à 100 et 150 espèces où ils sont plus élevés. Les réductions sont très marquées pour les types 5\_6 (milieux alluviaux), 7.2 (tourbières acides), 7.3\_9.2 (tourbières alcalines), soit globalement tous les types riches en espèces. A contrario, les types pauvres sont moins touchés. Les filtres les plus drastiques sont obtenus avec Phi, tant pour l'engorgement que pour la fertilité ; suivent en général, les filtres avec IndVal et le filtre sur la rareté à 50 espèces.



**Figure 41 :** nombres d'espèces par type de zone humide et au total, selon les méthodes de filtre (F50, F100 et F150 = filtre par la rareté ; IndVal H et N = filtre par IndVal sur l'engorgement et la fertilité ; phi H et N = filtre avec Phi sur l'engorgement et la rareté).

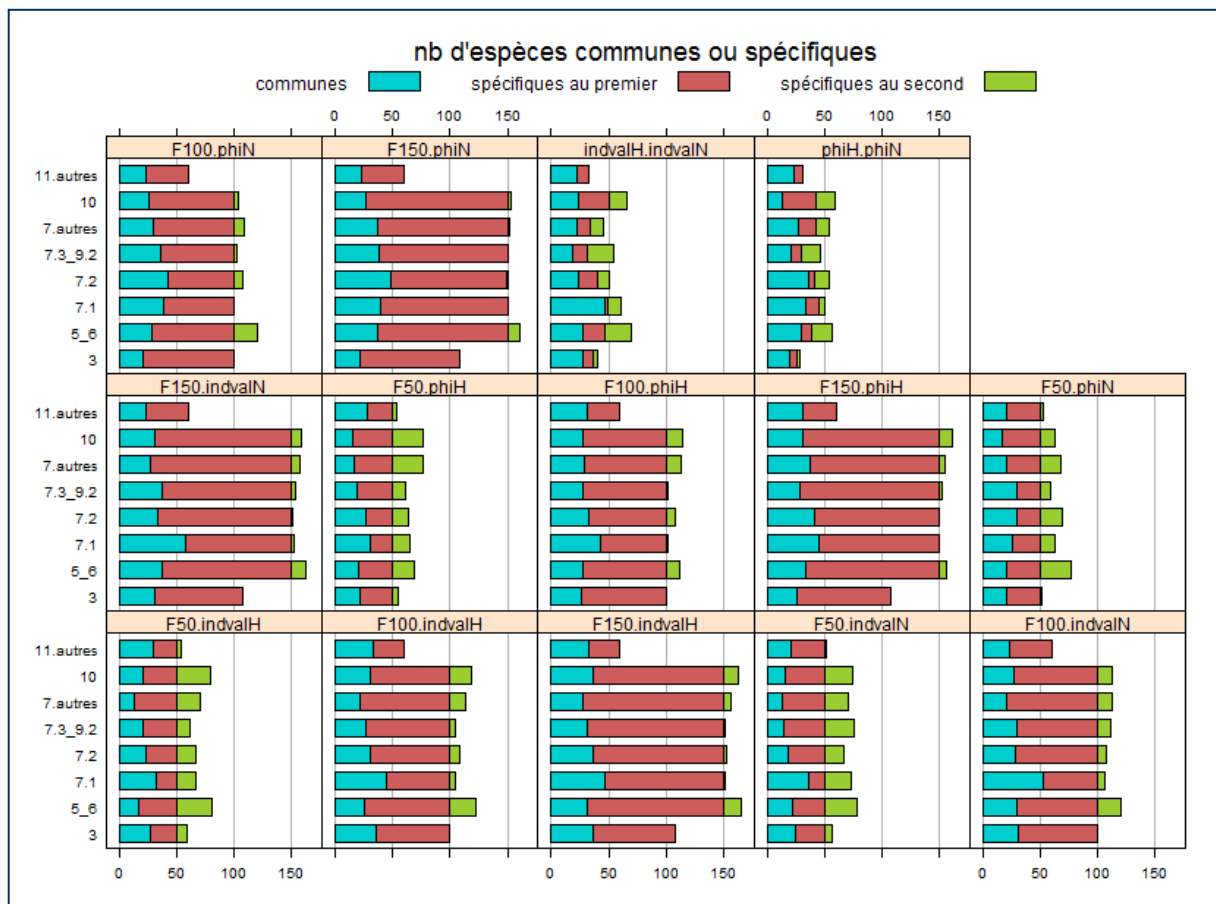
Plusieurs éléments d'explication peuvent être avancés pour expliquer ces réductions importantes pour IndVal et Phi. Tout d'abord, c'est la contrainte de limiter le nombre d'espèces à 40 ou 50 par type de zone humide. Un autre élément d'explication tient au choix d'évaluer le caractère indicateur sur l'ensemble du jeu de données plutôt que séparément pour chaque type de zone humide (même si la valeur indicatrice est calculée pour un type de zone humide, cette valeur est contextualisée par rapport à l'ensemble des valeurs indicatrices pour l'ensemble des types, alors qu'avec des analyses strictement séparées, on évalue la valeur indicatrice uniquement au sein d'un type, sans comparaisons avec les valeurs des vis-à-vis des autres types). Rappelons que ce choix a été fait pour éviter, *a priori*, la mise en exergue de trop d'espèces rares pour un type de zone humide par l'effet des hasards de l'échantillonnage : fréquemment, une espèce accidentelle à occasionnelle dans un type de zone humide est échantillonnée ; son évaluation par IndVal par exemple, sans prise en compte du contexte global de ses fréquences dans l'ensemble des zones humides, peut amener à la considérer indicatrice, alors qu'elle ne l'est pas. A titres d'illustrations, nous avons repris les calculs pour les types 5\_6 et 7.2, analysés séparément. Pour le type 5\_6 analysé seul, IndVal donne 109 espèces liées au seuil de significativité de 5% (28, 40 et 41 par sous-type), contre 37 pour l'analyse globale (14,3, et 20 par sous-type). Pour le type 7.2, les chiffres correspondants sont de 77 (9,49, et 19 par sous-type), contre 24 (11, 4 et 9 par sous-type) si l'analyse est globale. En général, les espèces données liées dans le contexte global sont également liées dans le contexte d'un seul type de zone humide, mais de nombreuses espèces viennent s'y ajouter, tant des espèces globalement rares que fréquentes. Il y aurait là un moyen d'améliorer les performances obtenues avec IndVal, mais ceci se ferait également au prix d'un nombre important d'espèces indicatrices, ce qui va à l'encontre de l'objectif recherché de restriction des listes.

Si on s'intéresse à la cohésion globale des listes, 214 taxons sont sélectionnés avec le filtre de rareté à 50, 290 avec la méthode IndVal dont 154 sont communs aux deux listes (représentant 71,9% et 53,1% des deux listes respectivement). La liste produite avec IndVal reprend donc l'essentiel des espèces sélectionnées avec le filtre à 50, mais elle est plus longue globalement, notamment à cause de la contrainte d'introduire des espèces par niveau de valeur diagnostique. Avec la liste filtrée sur la rareté à 100 espèces, 364 espèces sont retenues dont 219 en commun avec IndVal (soit 60,2% et 75,5%). On trouve une bonne cohérence réciproque même si dans

ce cas c'est le filtre à 100 espèces qui retient l'essentiel des espèces sélectionnées avec IndVal plutôt que l'inverse ; par ailleurs, les prédictions avec IndVal ne sont pas biaisées, contrairement à celles avec le filtre de rareté à 100. Avec le filtre de rareté à 150, on a 458 espèces sélectionnées, dont 250 communes avec les 290 sélectionnée par IndVal (soit 54,6% et 86,2%). La tendance perçue avec le filtre à 100 augmente donc, mais c'est aussi la conséquence de la volonté de limiter les nombres d'espèces à 50 (au maximum) par type de zone humide avec IndVal.

Avec Phi, on a au total 253 taxons retenus, dont 144 sont communs avec les 214 du filtre à 50 (soit 67,3% et 56,9% respectivement), et 204 sont communs avec les 364 espèces issues du filtre à 100 taxons (soit 56,0% et 80,6% respectivement). L'indice Phi, basé rappelons-le uniquement sur des présences, a également tendance à sélectionner des espèces globalement rares et à écarter les espèces qui sont réparties dans plusieurs sous-types. C'est également vrai pour IndVal, mais dans une bien moindre mesure, notamment car la partie A de l'indice, la concentration, vient récupérer certaines espèces fréquentes dans plusieurs sous-types mais avec des recouvrements plus élevés dans certains sous-types.

**Figure 42** : nombres d'espèces en commun (en bleu), propres à la première méthode de chaque panel (en rouge) et propres à la deuxième méthode de chaque panel, par type de zone humide et selon les méthodes de filtre (F50, F100 et F150 = filtre par la rareté ; IndVal H et N = filtre par IndVal sur l'engorgement et la fertilité ; phi H et N = filtre avec Phi sur l'engorgement et la rareté) ; par exemple, le panel du haut à gauche intitulé "F100.phiN", les longueurs des barres donne les nombres d'espèces communes aux deux méthodes (filtre sur rareté à 100 espèces et méthode avec Phi pour la fertilité) en bleu, le nombre d'espèces présentes uniquement avec F100 (en rouge) et celui des espèces présentes uniquement avec PhiN (en vert).



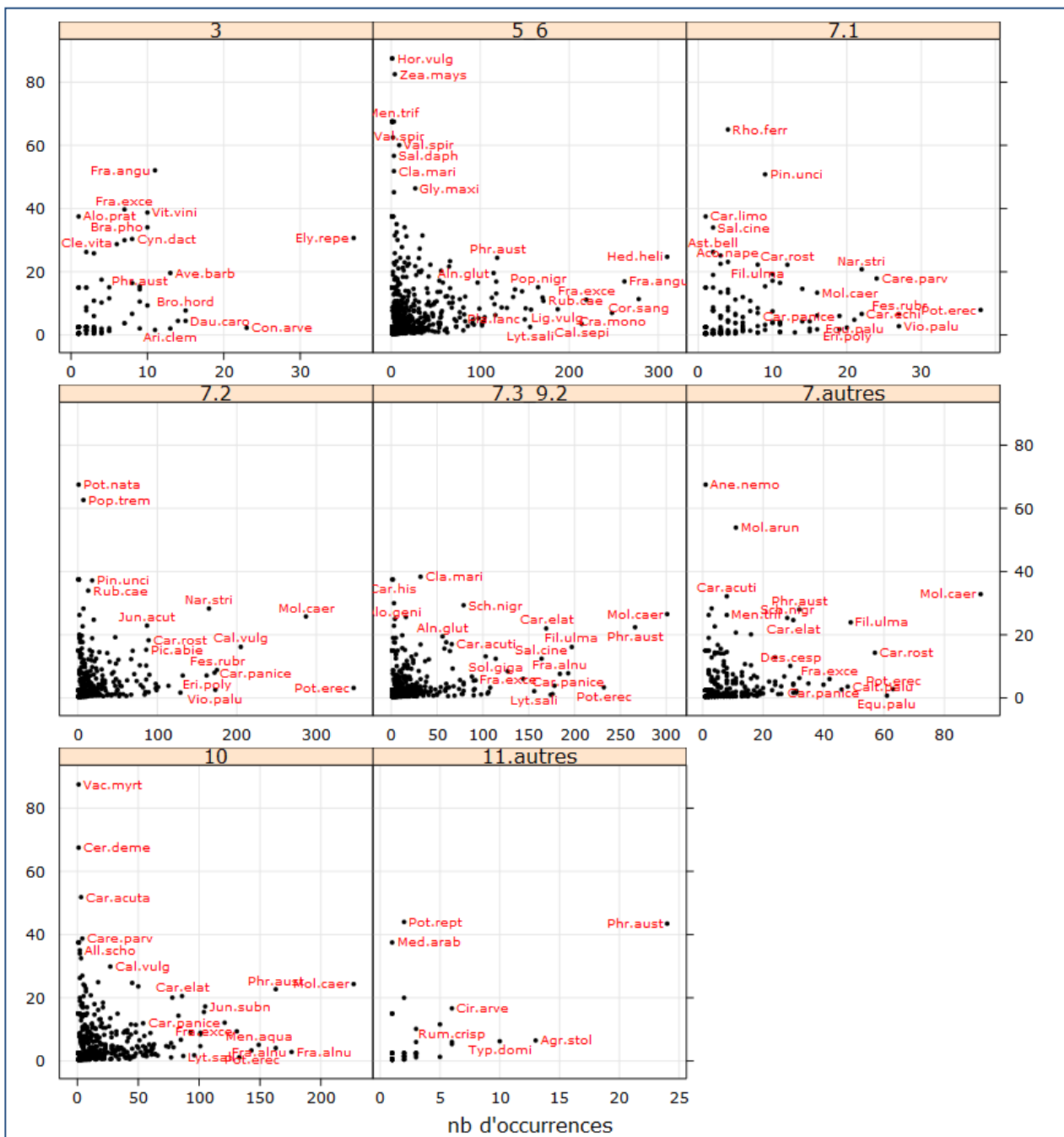
Concernant les listes d'espèces, la figure 42 résume l'essentiel des résultats. Elle donne les nombres d'espèces communes, selon différentes combinaisons de filtres (par exemple, le panel « F50.indval » donne le nombre d'espèces communes présentes dans les 50 espèces les plus courantes et sélectionnées avec IndVal), et ce pour chaque type de zone humide (en ordonnées).

Par type de zone humide, le nombre moyen d'espèce en commun varie peu : il est minimum pour les types 3 et 11.autres, maximum pour les types 10 et 5\_6. En relatif par rapport au nombre d'espèces par type de zone humide, on a au contraire un gradient : les types les moins riches (11.autre, 7.1 et 3) donnent les listes en moyennes les plus cohérentes (% moyen de cohérence de plus de 40%), à l'exception du type 7.autre qui bien que pauvre donne des listes peu cohérentes par les différentes méthodes de filtre. Les types riches en espèces donnent des listes peu à moyennement cohérentes entre elles (types 7.3\_9.2 et 10 surtout), les zones humides de type 7.2 étant en situation intermédiaire. Par type de comparaisons entre méthodes de filtre, comme attendu la cohérence relative est la moins bonne entre les listes filtrées à 150 et IndVal ou Phi, simplement car le nombre d'espèces est beaucoup plus élevé avec les filtres à 150 (pourcentages moyens d'espèces en communs compris entre 25 et 28,5%). Ensuite, viennent les listes filtrées à 100 avec IndVal ou Phi (pourcentages communs de l'ordre de 30%), puis enfin les listes les plus cohérentes entre elles sont celles obtenues par Phi ou par IndVal (cohérence de l'ordre de 50% en moyenne). Quelques surprises contrecarrent ce schéma général, comme la faible cohérence entre F50 et IndVal pour la fertilité.

L'annexe 2 fournit pour les 515 taxons ayant été sélectionnés au moins une fois par une des 7 méthodes principales employées (filtre sur la rareté à 50, 100 ou 150 espèces par type de zone humide, méthode IndVal pour l'engorgement et la fertilité, méthode Phi pour l'engorgement et la fertilité), par type de zone humide, le nombre de fois où un taxon a été sélectionné. Par exemple, *Acer negundo* a été sélectionné par les 7 méthodes pour le type 5\_6 (milieux alluviaux), par 2 méthodes pour le type 10 (marais), et aucune autre fois pour les autres types de zones humides.

La figure 43 illustre, pour chaque type de zone humide, la relation entre la fréquence des espèces et leur recouvrement moyen (calculé uniquement sur les relevés où une espèce est présente), et permet grâce aux noms abrégés, d'identifier les espèces fréquentes (vers l'extrémité droite de l'axe des abscisses) et les espèces ayant, sur les placettes où on les trouve, des recouvrements élevés (vers l'extrémité haute de l'axe des ordonnées).

**Figure 43 :** relation, par type de zone humide, entre la fréquence des espèces (axes des abscisses) et leur recouvrement moyen quand elles sont présentes (axe des ordonnées) ; les espèces sont symbolisées par des points noirs et certaines sont identifiées avec un nom abrégé en rouge (voir annexe 2 pour les noms abrégés).



## 4.7 Conséquences possibles sur l'interprétation des évolutions

---

Pour mémoire, l'interprétation des différences de valeurs avec le protocole flore dans le temps pour un site donné est basé sur 3 critères : l'écart des médianes des placettes (doit être supérieur à 0.4 si on prend en compte le recouvrement), le test des rangs signés de WILCOXON pour échantillons appariés doit être significatif ou le V de KRAMER doit être supérieur à environ 0.2. Pour évaluer l'effet des filtres sur la capacité de cette méthode a pouvoir être employée, il faudrait disposer d'un panel de sites ayant fait l'objet d'au moins deux évaluations avec l'ensemble des espèces, afin de comparer si des évolutions sont détectées avec les données complètes ou avec les données filtrées. Nous ne disposons pas de telles données, sauf pour deux sites savoyards. Les résultats pour ces deux sites pour le test de WILCOXON sont présentés tableau 11, tout en insistant sur le caractère illustratif plus que démonstratif de ce tableau.

Pour chaque site et les deux indicateurs, nous avons effectué le test de WILCOXON entre les deux années 2010 et 2014 pour lesquelles nous disposons de données. Comme d'habitude, HR désigne les valeurs obtenues avec le recouvrement avec l'ensemble des espèces, HR50 la même chose avec uniquement les 50 espèces les plus fréquentes, etc.... Outre les comparaisons entre 2010 et 2014, certaines comparaisons entre méthodes de filtre pour une année donnée sont fournies, en limitant les comparaisons entre IndVal et une des méthodes de filtre sur la rareté ; les calculs comparant entre elles les méthodes basées sur la rareté ont été faits mais ne sont pas fournis (les résultats seront indiqués le moment voulu).

Pour le site de la Chautagne (classé en type 10 – marais), les comparaisons entre 2010 et 2014 pour l'engorgement sont toutes non significatives (probabilité  $\gg 0.05$ ) ; c'est également vrai pour la fertilité, même si les probabilités sont globalement plus faibles (c'est-à-dire que les différences entre les deux années sont un peu plus élevées mais non significatives) ; le comportement d'IndVal est donc cohérent avec celui des filtres basés sur la rareté et avec l'absence de tendance observée avec l'ensemble des données. Pour l'engorgement, au sein d'une année, certaines comparaisons entre niveaux de filtre et IndVal apparaissent significatives, tant pour 2010 que pour 2014 ; signalons que c'est également le cas pour certaines comparaisons entre filtres basés sur la rareté entre eux. Pour la fertilité, toutes les comparaisons entre IndVal et les filtres basés sur la rareté, ainsi qu'entre une bonne partie des filtres basés sur la rareté entre eux (résultats non montrés), donnent des différences significatives.

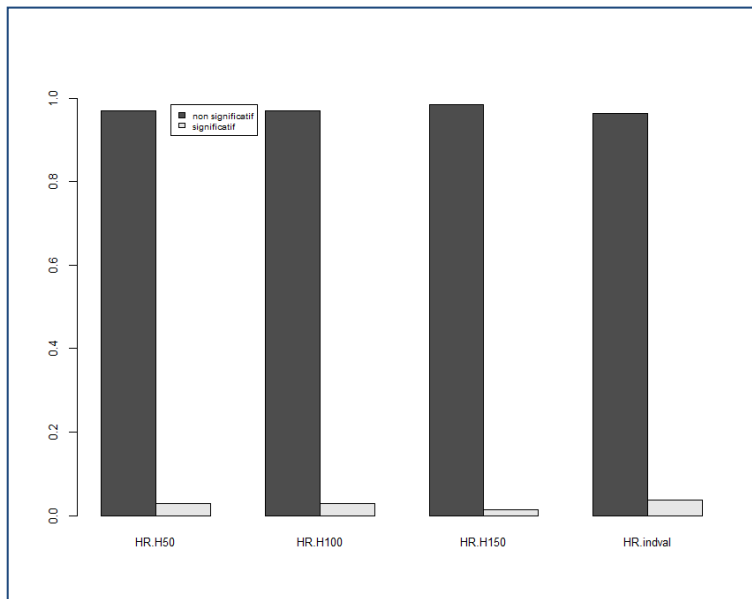
Pour le site des Creusates, la situation est différente : une évolution significative est détectée par le test de WILCOXON pour l'engorgement (mais comme globalement la différence médiane est inférieure à 0.4, elle n'est toutefois globalement pas considérée significative), mais pas pour la fertilité ; ceci est vrai tant avec l'ensemble des données qu'avec les différentes méthodes de filtre. Au sein d'une année donnée, les comparaisons de méthodes de filtres basés sur la rareté avec IndVal sont globalement non significatives et cohérentes pour l'engorgement (il en est de même pour les comparaisons entre filtres basés sur la rareté entre eux) ; autrement dit, les deux années diffèrent entre elles, et ceci est montré par toutes les méthodes mais dans l'ensemble, les résultats pour une année donnée sont cohérents entre eux. On peut toutefois remarque que les probabilités sont globalement assez faibles (et donc assez proches de la significativité à 5%). Concernant la fertilité, aucune tendance n'est détectée par aucune des méthodes, mais les différences entre les deux années est toutefois perceptible (les probabilités sont assez faibles).

**Tableau 11 :** valeurs diagnostiques d'engorgement et de fertilité pour deux sites évalués 2 années (2010 et 2014) et comparaisons des résultats obtenus avec les différentes méthodes de filtre.

Site	Indicateur	Année(s)	Comparaison	Probabilité
Chautagne	engorgement	2010-2014	HR 2010 et HR 2014	0.8132
			HR50 2010 et HR50 2014	0.8582
			HR100 2010 et HR100 2014	0.9596
			HR150 2010 et HR150 2014	0.9865
			indval 2010 et indval 2014	0.8986
		2010	indval 2010 et HR 2010	0.0020
			indval 2010 et HR50 2010	0.2798
			indval 2010 et HR100 2010	0.06796
		2014	indval 2010 et HR150 2010	0.02412
			indval 2014 et HR 2014	0.00167
			indval 2014 et HR50 2014	0.2778
			indval 2014 et HR100 2014	0.05477
	fertilité	2010-2014	indval 2014 et HR150 2014	0.03582
			NR 2010 et NR 2014	0.3258
			NR50 2010 et NR50 2014	0.2313
			NR100 2010 et NR100 2014	0.5656
			NR150 2010 et NR150 2014	0.3174
		2010	indval 2010 et indval 2014	0.9453
			indval 2010 et NR 2010	0.000365
			indval 2010 et NR50 2010	0.0000006
2014	indval 2010 et NR100 2010	0.00028		
	indval 2010 et NR150 2010	0.000256		
	indval 2014 et NR 2014	0.0032		
	indval 2014 et NR50 2014	0.000438		
Les Creusates	engorgement	2010-2014	indval 2014 et NR100 2014	0.005827
			indval 2014 et NR150 2014	0.00485
			HR 2010 et HR 2014	0.0105
			HR50 2010 et HR50 2014	0.03028
			HR100 2010 et HR100 2014	0.006714
		2010	HR150 2010 et HR150 2014	0.006714
			indval 2010 et indval 2014	0.005249
			indval 2010 et HR 2010	0.2412
		2014	indval 2010 et HR50 2010	0.8077
			indval 2010 et HR100 2010	0.3575
			indval 2010 et HR150 2010	0.2412
			indval 2014 et HR 2014	0.08032
	fertilité	2010-2014	indval 2014 et HR50 2014	0.02454
			indval 2014 et HR100 2014	0.1353
			indval 2014 et HR150 2014	0.07849
			NR 2010 et NR 2014	0.086
			NR50 2010 et NR50 2014	0.09168
		2010	NR100 2010 et NR100 2014	0.1842
			NR150 2010 et NR150 2014	0.09349
			indval 2010 et indval 2014	0.359
2014	indval 2010 et NR 2010	0.02148		
	indval 2010 et NR50 2010	0.02148		
	indval 2010 et NR100 2010	0.02148		
	indval 2010 et NR150 2010	0.02148		
2014	indval 2014 et NR 2014	0.008057		
	indval 2014 et NR50 2014	0.01343		
	indval 2014 et NR100 2014	0.008057		
	indval 2014 et NR150 2014	0.008057		

Par année, toutes les comparaisons entre IndVal et les filtres basés sur la rareté donnent des différences importantes. Il n'en est pas de même pour les comparaisons de données filtrées sur la rareté entre elles.

Hormis ces deux sites, on peut considérer que si le test de WILCOXON est applicable aux données filtrées, les tests de comparaisons des données filtrées et non filtrées d'un même site (à la même date donc mais avec plus ou moins de données), ne devraient pas conclure à des différences significatives, ou du moins pas trop fréquemment. On sait en effet qu'en effectuant des nombreuses comparaisons 2 à 2, il y a statistiquement un certain nombre de différences qui seront détectées significatives alors qu'elles ne le sont pas ; on admet classiquement que le risque de première espèce, alpha, doit être diminué en fonction du nombre de comparaisons : pour 137 sites, ce risque nominal doit en gros être abaissé à  $137 \times 136 / 2$ . La figure 44 illustre une application au test de WILCOXON pour l'engorgement dans le cas des données avec recouvrement.



**Figure 44** : pourcentages des tests des rangs signés de WILCOXON détectant une différence significative au seuil 5% (en gris clair) ou non significative (en gris foncé) à l'échelle des sites, pour différentes comparaisons de filtres avec la référence non filtrée (HR) : HR.HR50 avec le filtre de rareté à 50, HR.HR100 pour le filtre de rareté à 100, HR.HR150 pour le filtre de rareté à 150 et HR.indval pour le filtre avec IndVal. Les comparaisons portent sur 137 sites.

Le nombre de tests déclarant les différences significatives est faible à très faible une fois la correction du niveau du test effectuée (compris entre 1,4% et 3,6%). Le taux maximum est atteint pour la comparaison entre les données totales et le filtre avec IndVal (HR.indval). Il semble donc que le test des rangs signés puisse être utilisé pour comparer des données issues du protocole complet avec d'autres issues d'un protocole avec liste d'espèces filtrée pour l'engorgement. Toutefois, il est évident que le test doit d'abord et avant tout être appliqué uniquement pour comparer des résultats issus du même protocole (donc soit non filtré avec non filtré soit une méthode de filtre avec cette même méthode de filtre). Un autre problème difficile à évaluer avec les données dont nous disposons est l'effet de la réduction du nombre de données disponibles pour réaliser les tests : toutes choses égales par ailleurs, plus la taille de l'échantillon est petite, moins un test est puissant (c'est-à-dire qu'il est capable de mettre en évidence un effet quand il y en a un) ; même si dans notre cas, la taille de l'échantillon est le nombre de placette (qui lui n'est pas affecté) et non pas le nombre de données, la réduction de celles-ci joue sur la précision de l'estimation des valeurs diagnostiques des placettes.

Concernant le V de KRAMER, nous n'avons pas fait le même exercice car il s'agit d'un coefficient « semi-statistique ». Il est toutefois évident que la réduction du nombre de données va rendre plus délicat son emploi (en particulier, de plus nombreuses cellules risquent d'avoir des effectifs inférieurs à 5, ce qui est une limite à la statistique du Khi deux sous-jacente), d'autant que ce nombre de données est directement lié à la taille de l'échantillon dans ce cas.

Concernant l'erreur empirique, évaluée à  $\pm 0.4$  dans le cas de données avec recouvrement, nous ne savons pas non plus dans quelle mesure elle reste valable pour les données filtrées : les données utilisées pour calibrer l'erreur sont en effet difficilement remobilisables : nous ne disposons pas d'une partie d'entre elles (celles

issues du sud du bassin) et le temps nous a manqué pour en conduire l'analyse sur les quelques sites rhônalpins utilisés initialement.

Nous disposons de très peu d'éléments pour évaluer dans quelle mesure les éléments d'interprétations d'évolution fournis pour le protocole avec l'ensemble des espèces est applicable avec des listes filtrées d'espèces. Sur les deux sites pour lesquels nous disposons de deux campagnes de mesures complètes, il s'avère que les résultats des tests de WILCOXON basés sur les données complètes, les données filtrées à 50, 100 et 150 espèces ainsi que celles basées sur IndVal, sont cohérents ; c'est une indication positive à l'utilisation de ce test sur les données filtrées. Il est toutefois préférable, sinon obligatoire, de comparer entre elles des valeurs obtenues avec la même méthode (complète, filtre sur la rareté ou IndVal) aux différentes dates. En revanche, nous ne disposons pas d'indications sur la pertinence des autres critères utilisés pour l'évaluation des tendances (erreur empirique et V de KRAMER). Il semble probable que la méthode globale d'interprétation (combinant les trois critères) soit à revoir.

## 5 Conclusion générale

---

Un cadre typologique commun, simplifié et cohérent, basé sur le respect des types de zones humide du programme RhoMéO a pu être défini. Il s'appuie sur les parentés floristiques des sites entre eux, vu au travers du prisme de l'échantillonnage de chacun d'entre eux. A partir des 19 types initiaux, 8 types plus ou moins synthétiques ont été identifiés, regroupant 16 des types initiaux. Pour les trois derniers, le nombre de sites ou de données disponibles n'ont pas été jugés suffisant pour poursuivre les investigations ; il serait possible d'y adjoindre deux autres types synthétiques, maintenus dans les analyses mais pour lesquels des données manquent.

Nous avons choisi de respecter le cadre typologique initial et l'assignement des sites aux types effectués par les opérateurs. Toutefois, à l'analyse, il s'avère que certains types sont très hétérogènes ; cette hétérogénéité peut être intrinsèque (à cause de la grande diversité possible des zones humides et des milieux associés) ou liée au manque de caractères basés sur la flore et les habitats pour l'assignement. **Fournir des listes d'espèces et d'habitats de référence pour l'assignement de sites, en plus de la clé fournie dans la boîte à outils, est une piste majeure de travail pour le futur** ; ce n'était pas l'objet de ce travail mais quelques éléments sont fournis. Cela permettrait une meilleure qualité globale des données et pourrait fournir un cadre plus cohérent pour l'analyse de l'évolutions d'ensembles de sites.

Pour les types synthétiques retenus et suffisamment bien pourvus en sites et en données, différentes pistes ont été creusées pour établir des listes d'espèces restreintes et permettant de reproduire au mieux les résultats obtenus avec l'ensemble des données.

Trois méthodes de filtrage des espèces donnent des résultats plus ou moins précis :

- Le filtre sur la rareté des espèces, avec les 50, 100 ou 150 espèces les plus fréquentes par type de zone humide, donne des résultats moyens à bons (dans cet ordre). Toutefois, ce résultat n'est réellement bons qu'avec 100 ou 150 espèces, ce qui fait trop dans l'optique d'un déploiement sur le terrain : en effet, il est très difficile d'identifier toutes les espèces contenues dans une liste de 150 noms, probablement plus difficile que de réaliser un relevé complet. Pour l'engorgement, avec 50 espèces les résultats sont moyens, et un peu biaisés. Ce sont surtout les milieux alluviaux qui sont mal prédits. Pour la fertilité, les résultats sont globalement meilleurs et déjà satisfaisants avec 50 espèces (absence de biais, écarts plus faibles). Certains sites sont assez systématiquement mals prédits, ceci s'expliquant soit par le faible nombre de données disponibles, soit par la présence de contingents importants d'espèces que l'on ne retrouve pas (ou peu) ailleurs, soit enfin, et notamment pour certains grands sites, par un échantillonnage trop faible. A l'échelle des relevés, les valeurs sont assez mal prédites avec les listes filtrées ; ceci n'est pas rédhibitoire car l'analyse est prévue à l'échelle des sites, mais rend des analyses à l'échelle de transect difficile. **Globalement, la prise en compte du recouvrement ne permet pas d'améliorer les prédictions par comparaison avec la seule présence des espèces.**
- Le filtre issu sur l'identification d'espèces liées à des types de zones humides selon la valeur IndVal. Cette méthode combine des données en présence/absence et avec recouvrement. Pour s'assurer que des espèces indicatrices de bon et de mauvais état de conservation soient sélectionnées, une stratification préalable des placettes selon leurs valeurs diagnostiques a été réalisée. Les listes d'espèces produites sont en général plus courtes (entre 35 et 45 espèces le plus souvent). Les résultats pour l'engorgement sont de qualité semblable à celle obtenue avec le filtre de rareté à 50 espèces, mais dépendent du type de zone humide. Pour la fertilité, les résultats sont moins bons que ceux obtenus avec le filtre de rareté à 50 espèces.

- Le filtre basé sur l'utilisation du coefficient Phi de PEARSON n'utilise que la présence/absence des espèces. Il s'est avéré globalement moins précis que celui basé sur la rareté à 50 et du même niveau que celui avec IndVal.

En revanche, les filtres basés sur le caractère sténocène des espèces ou sur l'inscription sur la liste des espèces caractéristiques des zones humides ne sont pas opérants. Une autre piste de travail, commencée mais non aboutie dans le programme RhoMéO, **serait de travailler sur les habitats**. Une liste d'habitats caractéristiques des zones humides existe également et pourrait être utilisée pour l'évaluation de l'état de conservation ou de fonctionnalités particulières des zones humides. **Cette approche par habitats permettrait également de mieux lier les nécessités d'évaluations découlant de la DCE et de Natura 2000.**

Sur l'ensemble des types de zones humides testées (16 types initiaux plus ou moins regroupés pour 137 sites), 214 espèces sont sélectionnées avec le filtre à 50, 290 avec IndVal et 253 avec Phi. La cohérence entre celles des listes filtrées produites dépend de leurs tailles, mais est maximale pour celles produites par Phi ou IndVal, et ce pour les deux indicateurs I2 et I6. Malgré le peu d'espèces, ces deux méthodes produisent des listes cohérentes pour les 2 indicateurs, ce qui est un point positif indéniable.

La méthode d'interprétation proposée par le protocole flore utilise trois critères d'analyse des tendances. Deux d'entre eux (erreur « empirique » et test des rangs signés de WILCOXON) semblent utilisables globalement mais sont certainement affaiblis par la réduction du nombre de données sur lesquelles ils sont assis. Pour le troisième, nous n'avons pas d'éléments. Il faut donc être prudent concernant la transposabilité du processus d'interprétation des données filtrées.

Globalement, les méthodes de filtrage pourraient être une alternative au protocole complet, mais plusieurs points importants seraient à travailler ou améliorer :

- réduire encore les listes d'espèces (20-30 par type) pour les rendre pleinement utilisables sur le terrain ; au-delà de ces valeurs, il devient en effet difficile de détecter toutes les espèces présentes sur une liste ;
- proposer des amendements ou produire une nouvelle méthodologie d'interprétation ; si on part de l'hypothèse que l'état « vrai » est celui évalué avec l'ensemble des données, il conviendra de s'assurer sur des sites évalués à plusieurs reprises avec le protocole complet que les résultats sont effectivement convergents avec ceux obtenus avec les listes filtrées, les deux seuls exemples dont nous disposons ne permettant pas de se faire une idée fiable ;
- une réflexion est également à mener sur l'échantillonnage ; en effet, avec des listes très restreintes d'espèces, et même si elles sont bien choisies, le nombre de données sur lesquelles on peut s'appuyer pour évaluer une évolution devient faible ; une option pour y pallier serait d'augmenter le nombre de placettes par site, ce qui permettrait également d'améliorer la couverture spatiale.

Ces évolutions nécessiteraient une refonte assez profonde, sinon complète des fiches protocole-indicateur-interprétation, ainsi qu'un gros travail de mise au point des listes. De plus, des listes aussi limitées en ampleur risqueraient de perdre leur applicabilité à des types de zones humides assez larges dans leur conception, et un risque serait d'avoir à ajuster de nombreuses listes, certes courtes, mais adaptées à des contextes géographiques étroits et des types de zones humides étroitement définies.

## BIBLIOGRAPHIE

- **AHO K., ROBERTS D.W. & WEAVER T.,**2008. *Using geometric and non-geometric internal evaluators to compare eight vegetation classification methods*. Journal of Vegetation Science, 19 : 549-562.
- **BORCARD D., GILLET F. & LEGENDRE P.,** 2011. *Numerical Ecology with R*. Springer, 2e eds., 319p.
- **DE CACERES M. & LEGENDRE P.,** 2009. *Associations between species and groups of sites: indices and statistical inference*. Ecology, 90 : 3566-3574.
- **DRAY S. & DUFOUR A.B.,** 2007: *The 'ade4' package: implementing the duality diagram for ecologists*. Journal of Statistical Software. 22(4): 1-20.
- **GILLET F.,** 2000. *Cours de statistiques*. Cours de EPFL, Lausanne, 30p.
- **JULVE Ph.,** 2016. *Baseflor*. Document internet EXCEL. [philippe.julve.pagesperso-orange.fr/baseflor.xlsx](http://philippe.julve.pagesperso-orange.fr/baseflor.xlsx)
- **LANDOLT E. et al.,** 2010. *Flora indicativa*. CJB Genève, Haupt, Berne, 376p.
- **MacQUEEN J.,** 1967. *Some methodes for classification of multivariate observations*. in L.M. LE CAM & J. NEYMAN (eds.). *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, University of California Press, Berkeley, Vol.1, pp. 281-297.
- **OKSANEN J., F. GUILLAUME BLANCHET F., KINDT R., LEGENDRE P., MINCHIN P.R., O'HARA R.B., SIMPSON G.L, SOLYMOS P., M. STEVENS M.H.H & WAGNER H.,** 2015. *'vegan' : community ecology package*. R package version 2.3-2. <http://CRAN.R-project.org/package=vegan>
- **R Core Team,** 2015. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- **REYNOLDS A. et al.,** 1992. *Clustering rules: A comparison of partitioning and hierarchical clustering algorithms*. Journal of Mathematical Modelling and Algorithms, 5 : 475–504.
- **ROBERTS D.W.,** 2016. *'optpart': optimal partitioning of similarity relations*. R package version 2.2-0. <http://CRAN.R-project.org/package=optpart>
- **ROBERTS D.W.,** 2015. *'labdsv': ordination and multivariate analysis for ecology*. R package version 1.7-0. <http://CRAN.R-project.org/package=labdsv>
- **ROUSSEUW P.J.,**1987. *"Silhouettes" : a graphical aid to the interpretation and validation of cluster analysis*. Journal of Computation an Applied Mathematics, 20 : 53-65.
- **SARKAR D.,** 2008. *'Lattice': multivariate data visualization with R*. Springer, New York. 265p.

- **TICHY L. & CHYTRY M.**, 2006. *Statistical determination of diagnostic species for site groups of unequal size*. Journal of Vegetation Science, 17 : 809-818.
- **WICKHAM H.**, 2011. *The split-apply-combine strategy for data analysis*. Journal of Statistical Software, 40(1), 1-29. URL <http://www.jstatsoft.org/v40/i01/>.
- **WICKHAM H.**, 2007. *Reshaping data with the 'reshape' package*. Journal of Statistical Software, 21(12), 1-20. URL <http://www.jstatsoft.org/v21/i12/>.

ANNEXE 1 : données relatives aux nombres d'observations, nombre d'espèces et nombre d'espèces en commun selon les principales méthodes (filtre sur la richesse à 50, 100 ou 150 espèces, filtre avec IndVal, filtre avec Phi), par type de zone humide et globalement.

	type de zone humide / variable	3	5_6	7.1	7.2	7.3_9.2	7.autres	10	11.autres	total
Nombres d'observations	nb.obs.	449	15.476	970	7.954	9.723	2.685	9.516	153	46.926
	nb.obs.F50	370	6.400	642	4.748	5.434	1.415	4.308	143	23.460
	nb.obs.F100	441	9.018	833	6.120	7.088	1.923	6.227	153	31.803
	nb.obs.F150	449	10.688	920	6.840	8.062	2.213	7.297	153	36.622
	nb.obs.IndvalH	285	3.602	496	3.043	3.231	559	2.840	116	14.172
	nb.obs.PhiH	246	3.780	461	3.303	3.118	695	2.117	113	13.833
	nb.obs.IndValN	266	4.331	570	1.936	2.430	524	1.978	99	12.134
nb.obs.PhiN	234	4.160	449	3.605	4.044	875	2.165	98	15.630	
Nombres d'espèces	nb.esp.	108	972	200	501	539	407	655	60	1.420
	nb.esp.F50	50	50	50	50	50	50	50	50	214
	nb.esp.F100	100	100	100	100	100	100	100	60	364
	nb.esp.F150	108	150	150	150	150	150	150	60	458
	nb.esp.IndvalH	36	47	49	40	32	34	50	33	290
	nb.esp.PhiH	26	39	45	41	30	42	42	31	253
	nb.esp.IndValN	31	50	59	35	40	34	40	23	310
nb.esp.PhiN	21	48	39	49	38	39	30	23	240	
Nombre d'espèces en commun	F50.indvalH	27	17	32	23	20	13	21	29	154
	F100.indvalH	36	25	44	31	27	21	31	33	219
	F150.indvalH	36	32	47	37	31	28	37	33	250
	F50.indvalN	25	22	36	18	14	13	16	21	163
	F100.indvalN	31	30	53	28	29	21	27	23	240
	F150.indvalN	31	37	57	34	37	27	31	23	275
	F50.phiH	21	20	30	27	19	16	15	28	144
	F100.phiH	26	27	43	33	28	37	31	31	204
	F150.phiH	26	33	45	41	28	37	31	31	229
	F50.phiN	20	21	26	30	29	21	17	21	144
	F100.phiN	21	28	39	42	36	30	26	23	199
	F150.phiN	21	37	39	48	38	37	27	23	224
	IndValH.IndValN	27	27	47	24	18	23	24	23	213
	phiH.phiN	19	30	34	36	21	27	13	23	209
phiH.IndValH	21	33	36	31	20	23	34	26	202	
phiN.IndValN	18	40	37	30	17	20	20	20	181	

## ANNEXE 2 : liste des espèces données indicatrices par l'une des sept méthodes de filtre, par type de zone humide.

Pour 515 taxons ayant été sélectionnés au moins une fois par une des 7 méthodes principales employées (filtre sur la rareté à 50, 100 ou 150 espèces par type de zone humide, méthode IndVal pour l'engorgement et la fertilité, méthode Phi pour l'engorgement et la fertilité), le tableau fournit, par type de zone humide, les nombres de méthodes l'ayant sélectionnés. Par exemple, *Acer negundo* a été sélectionné par les 7 méthodes pour le type 5\_6 (milieux alluviaux), par 2 méthodes pour le type 10 (marais), et aucune autre fois pour les autres types de zones humides

espèce	3	5_6	7.1	7.2	7.3_9.2	7.autres	10	11.autres
<i>Abies alba</i> Mill.	0	0	0	1	0	2	0	0
<i>Acer campestre</i> L.	0	1	0	0	0	0	1	0
<i>Acer negundo</i> L.	0	7	0	0	0	0	1	0
<i>Acer pseudoplatanus</i> L.	0	0	0	1	1	1	2	0
<i>Achillea millefolium</i> L.	0	0	0	1	0	0	3	0
<i>Achillea ptarmica</i> L.	0	1	0	0	0	2	0	0
<i>Achillea roseo-alba</i> Ehrend.	0	0	0	0	0	0	2	0
<i>Aconitum napellus</i> L.	0	0	6	0	0	0	0	0
<i>Agrostis canina</i> L.	1	0	0	7	0	0	0	0
<i>Agrostis capillaris</i> L.	2	0	2	7	0	2	1	0
<i>Agrostis gigantea</i> Roth	0	0	0	0	2	0	4	0
<i>Agrostis stolonifera</i> L.	1	3	2	2	3	2	2	7
<i>Aira caryophylla</i> L.	0	0	0	1	0	0	0	0
<i>Ajuga reptans</i> L.	0	1	1	2	2	1	2	0
<i>Alchemilla glabra</i> Neygenf.	0	0	0	0	0	3	0	0
<i>Alchemilla monticola</i> Opiz	0	0	0	0	0	0	2	0
<i>Alisma plantago-aquatica</i> L.	0	0	0	0	0	0	0	4
<i>Alliaria petiolata</i> (M.Bieb.) Cavara & Grande	0	6	0	0	0	0	0	0
<i>Allium schoenoprasum</i> L.	0	0	0	0	0	3	0	0
<i>Alnus glutinosa</i> (L.) Gaertn.	0	3	1	1	4	2	2	0
<i>Alnus incana</i> (L.) Moench	0	0	0	0	3	0	1	0
<i>Alopecurus bulbosus</i> Gouan	5	0	0	0	0	0	1	0
<i>Alopecurus myosuroides</i> Huds.	1	0	0	0	0	0	0	0
<i>Alopecurus pratensis</i> L.	1	2	0	0	0	0	0	0
<i>Althaea officinalis</i> L.	0	6	0	0	0	0	0	0
<i>Ambrosia artemisiifolia</i> L.	0	3	0	0	0	0	0	0
<i>Anacamptis palustris</i> (Jacq.) Bateman, Pridgeon & Chase	0	0	0	0	1	0	0	0
<i>Anagallis arvensis</i> L.	3	0	0	0	0	0	0	6
<i>Angelica sylvestris</i> L.	0	2	0	3	3	5	3	0
<i>Anthoxanthum odoratum</i> L.	0	2	4	3	2	2	3	0
<i>Anthoxanthum odoratum</i> L. subsp. <i>nipponicum</i> (Honda)	0	0	2	0	0	0	0	0



espèce	3	5_6	7.1	7.2	7.3_9.2	7.autres	10	11.autres
Carex acuta L.	0	4	0	0	0	0	0	0
Carex acutiformis Ehrh.	0	2	0	0	4	2	6	0
Carex appropinquata Schumach.	0	0	0	0	5	0	0	0
Carex cuprina (Sandor ex Heuff.) Nendtv. ex A.Kern.	2	0	0	0	0	0	0	7
Carex curta Gooden.	0	0	5	0	0	0	0	0
Carex davalliana Sm.	0	0	4	2	5	5	1	0
Carex diandra Schrank	0	0	0	0	1	0	0	0
Carex distans L.	0	0	0	0	0	0	1	3
Carex disticha Huds.	0	2	0	0	0	0	0	0
Carex divisa Huds.	7	0	0	0	0	0	0	0
Carex echinata Murray	0	0	7	5	1	2	0	0
Carex elata All.	0	2	1	1	7	3	4	0
Carex ferruginea Scop.	0	0	3	0	0	0	0	0
Carex flacca Schreb.	0	2	1	0	6	6	6	0
Carex flava L.	0	0	7	1	2	1	0	0
Carex hirta L.	0	3	0	0	1	2	3	0
Carex hispida Willd.	0	0	5	5	0	0	0	0
Carex hostiana DC.	0	0	1	1	7	3	3	0
Carex lasiocarpa Ehrh.	0	0	0	2	1	0	2	0
Carex limosa L.	0	0	0	4	0	0	0	0
Carex magellanica Lam. subsp. irrigua (Wahlenb.) Hiitonen	0	0	1	0	0	0	0	0
Carex nigra (L.) Reichard	0	0	6	3	4	3	1	0
Carex ovalis Gooden.	0	0	5	2	0	0	0	0
Carex pallescens L.	0	0	1	1	0	0	0	0
Carex panicea L.	0	2	3	3	5	3	6	0
Carex paniculata L.	0	0	0	0	2	1	0	0
Carex parviflora Host	0	0	7	5	0	0	0	0
Carex pauciflora Lightf.	0	0	0	5	0	0	0	0
Carex pendula Huds.	0	1	0	0	0	0	0	0
Carex pilulifera L.	0	0	0	1	0	0	0	0
Carex pulcaris L.	0	0	0	0	4	0	0	0
Carex riparia Curtis	0	6	0	0	0	0	0	0
Carex rostrata Stokes	0	0	6	6	3	5	1	0
Carex sempervirens Vill.	0	0	1	0	0	0	0	0
Carex tomentosa L.	0	0	0	0	0	0	3	0
Carex vesicaria L.	0	0	0	0	1	0	0	0
Carex viridula Michx. subsp. brachyrrhyncha (Celak.) B.Schmid	0	0	5	2	4	2	2	0
Carex viridula Michx. subsp. oedocarpa (Andersson) B.Schmid	0	0	0	1	0	0	0	0
Carum carvi L.	0	0	0	0	0	1	2	0
Carum verticillatum (L.) W.D.J.Koch	0	0	0	7	0	0	0	0
Catapodium rigidum (L.) C.E.Hubb.	0	0	0	0	0	0	0	2
Celtis australis L.	0	1	0	0	0	0	0	0
Centaurea jacea L.	0	2	0	0	1	0	7	0

espèce	3	5_6	7.1	7.2	7.3_9.2	7.autres	10	11.autres
Centaurea nigra L.	0	3	0	1	0	0	0	0
Cerastium fontanum Baumg. subsp. vulgare (Hartm.) Greuter & Burdet	0	1	1	0	2	2	2	0
Ceratophyllum demersum L.	0	4	0	0	0	0	0	0
Chelidonium majus L.	0	2	0	0	0	0	0	0
Chenopodium album L.	5	0	0	0	0	0	0	0
Cichorium intybus L.	7	0	0	0	0	0	0	0
Circaea lutetiana L.	0	2	0	0	2	0	2	0
Cirsium arvense (L.) Scop.	5	2	0	0	0	0	3	7
Cirsium eriophorum (L.) Scop.	2	0	0	0	0	0	0	0
Cirsium palustre (L.) Scop.	0	0	4	4	6	3	3	0
Cirsium vulgare (Savi) Ten.	0	0	0	0	0	3	0	0
Cladium mariscus (L.) Pohl	0	0	0	0	5	0	5	0
Clematis vitalba L.	3	5	0	0	0	0	0	0
Colchicum autumnale L.	0	0	0	0	1	3	1	0
Convolvulus arvensis L.	7	2	0	0	0	0	0	4
Conyza canadensis (L.) Cronquist	4	0	0	0	0	0	0	0
Conyza sumatrensis (Retz.) E.Walker	0	0	0	0	0	0	0	7
Cornus sanguinea L.	0	7	0	0	2	2	3	0
Corylus avellana L.	0	2	0	0	1	5	2	0
Crataegus monogyna Jacq.	0	7	0	0	2	2	3	0
Crepis aurea (L.) Tausch	0	0	2	0	0	0	0	0
Crepis foetida L.	0	0	0	0	0	0	0	2
Crepis paludosa (L.) Moench	0	0	7	2	2	5	0	0
Cruciata laevipes Opiz	0	0	0	0	0	6	0	0
Cucubalus baccifer L.	0	1	0	0	0	0	0	0
Cynodon dactylon (L.) Pers.	7	0	0	0	0	0	0	2
Cynosurus cristatus L.	0	1	0	2	1	1	2	0
Cyperus eragrostis Lam.	2	0	0	0	0	0	0	0
Cytisus scoparius (L.) Link	0	0	0	1	0	0	0	0
Dactylis glomerata L.	0	4	0	0	2	2	4	0
Dactylorhiza fistulosa (Moench) Baumann & Künkele	0	0	2	0	1	3	1	0
Dactylorhiza incarnata (L.) Soó	0	0	0	0	5	1	1	0
Dactylorhiza maculata (L.) Soó	0	0	7	5	0	0	0	0
Dactylorhiza savogiensis D.Tyteca & Gathoye	0	0	7	0	0	0	0	0
Danthonia decumbens (L.) DC.	0	0	0	2	0	0	0	0
Daucus carota L.	7	2	0	0	1	0	1	0
Deschampsia cespitosa (L.) P.Beauv.	0	2	6	3	3	3	3	0
Deschampsia flexuosa (L.) Trin.	0	0	0	6	0	0	0	0
Diplotaxis eruroides (L.) DC.	7	0	0	0	0	0	0	0
Diplotaxis tenuifolia (L.) DC.	5	0	0	0	0	0	0	0
Dipsacus fullonum L.	0	0	0	0	0	0	0	4
Dittrichia viscosa (L.) Greuter	6	0	0	0	0	0	0	0
Drosera rotundifolia L.	0	0	2	7	0	0	0	0

espèce	3	5_6	7.1	7.2	7.3_9.2	7.autres	10	11.autres
Drosera x obovata Mert. & W.D.J.Koch	0	0	0	5	0	0	0	0
Dryopteris carthusiana (Vill.) H.P.Fuchs	0	0	2	0	0	0	1	0
Echinochloa crus-galli (L.) P.Beauv.	0	2	0	0	0	0	0	0
Echium vulgare L.	0	3	0	0	0	0	0	0
Eleocharis palustris (L.) Roem. & Schult.	0	0	0	0	0	0	0	4
Eleocharis quinqueflora (Hartmann) O.Schwarz	0	0	7	0	0	0	0	0
Eleocharis uniglumis (Link) Schult.	0	0	0	0	2	0	4	0
Elodea nuttallii (Planch.) H.St.John	0	3	0	0	0	0	0	0
Elytrigia repens (L.) Desv. ex Nevski	7	2	0	0	0	0	1	0
Epilobium hirsutum L.	0	1	0	0	3	1	1	5
Epilobium nutans F.W.Schmidt	0	0	2	0	0	0	0	0
Epilobium palustre L.	0	0	7	2	2	3	0	0
Epilobium parviflorum Schreb.	0	0	0	0	6	0	0	5
Epilobium tetragonum L.	0	0	0	0	0	6	0	0
Epipactis helleborine (L.) Crantz	0	0	0	0	0	5	0	0
Epipactis palustris (L.) Crantz	0	0	2	0	4	4	5	0
Equisetum arvense L.	0	2	0	0	0	1	1	0
Equisetum fluviatile L.	0	2	2	1	4	3	1	0
Equisetum hyemale L.	0	0	0	0	0	1	0	0
Equisetum palustre L.	0	2	7	2	3	5	2	0
Equisetum ramosissimum Desf.	2	0	2	0	0	0	0	5
Equisetum sylvaticum L.	0	0	6	0	0	0	0	0
Equisetum telmateia Ehrh.	0	0	0	0	0	4	0	0
Equisetum variegatum Schleich.	0	0	0	0	0	5	0	0
Eriophorum latifolium Hoppe	0	0	5	2	4	2	0	0
Eriophorum polystachion L.	0	0	5	7	1	0	1	0
Eriophorum vaginatum L.	0	0	1	7	0	0	0	0
Eupatorium cannabinum L.	0	2	0	0	3	7	3	0
Euphorbia palustris L.	0	2	0	0	0	0	0	0
Euphorbia segetalis L.	3	0	0	0	0	0	0	0
Euphorbia serrata L.	2	0	0	0	0	0	0	0
Euphrasia officinalis L.	0	0	3	0	0	0	0	0
Evonymus europaeus L.	0	5	0	0	1	1	3	0
Fagus sylvatica L.	0	0	1	1	0	2	0	0
Festuca arundinacea Schreb.	3	2	0	0	2	1	5	0
Festuca filiformis Pourr.	0	0	0	1	0	0	1	0
Festuca nigrescens Lam.	0	0	0	2	0	3	0	0
Festuca pratensis Huds.	0	0	0	0	0	1	4	0
Festuca rubra L.	0	0	7	6	1	0	1	0
Festuca rubra L. subsp. rubra	0	0	0	0	0	3	0	0
Festuca trichophylla (Ducros ex Gaudin) K.Richt.	0	0	0	0	6	0	6	0
Filipendula ulmaria (L.) Maxim.	0	3	3	2	7	5	3	0
Filipendula vulgaris Moench	0	0	0	0	0	0	1	0
Foeniculum vulgare Mill.	7	0	0	0	0	0	0	0

espèce	3	5_6	7.1	7.2	7.3_9.2	7.autres	10	11.autres
Frangula dodonei Ard.	0	1	2	2	7	4	6	0
Fraxinus angustifolia Vahl	4	6	0	0	0	0	2	0
Fraxinus excelsior L.	4	3	2	0	3	4	3	0
Galactites elegans (All.) Soldano	3	0	0	0	0	0	0	0
Galeopsis tetrahit L.	0	1	1	1	3	1	1	0
Galium aparine L.	3	5	0	0	2	0	2	3
Galium boreale L.	0	0	0	0	2	0	6	0
Galium mollugo L.	4	1	0	0	1	0	1	0
Galium mollugo L. subsp. mollugo	0	1	0	0	2	0	4	0
Galium palustre L.	0	3	3	2	6	3	5	0
Galium saxatile L.	0	0	0	4	0	0	0	0
Galium uliginosum L.	0	0	4	3	3	3	3	0
Galium verum L.	2	2	0	1	2	2	5	0
Gaudinia fragilis (L.) P.Beauv.	7	0	0	0	0	0	0	0
Genista anglica L.	0	0	0	5	0	0	0	0
Genista tinctoria L.	0	0	0	0	0	0	4	0
Gentiana bavarica L.	0	0	1	0	0	0	0	0
Gentiana pneumonanthe L.	0	0	0	1	2	2	0	0
Geranium dissectum L.	4	1	0	0	0	0	0	0
Geranium robertianum L.	0	1	0	0	0	0	1	0
Geranium rotundifolium L.	0	0	0	0	0	0	0	5
Geranium sylvaticum L.	0	0	2	0	0	4	0	0
Geum rivale L.	0	0	0	1	0	4	0	0
Geum urbanum L.	0	2	0	0	1	0	3	0
Gladiolus palustris Gaudin	0	0	0	0	0	0	2	0
Glechoma hederacea L.	0	2	0	0	0	0	0	0
Glyceria maxima (Hartm.) Holmb.	0	5	0	0	0	0	0	0
Gymnadenia conopsea (L.) R.Br.	0	0	1	0	2	2	1	0
Gymnadenia odoratissima (L.) Rich.	0	0	0	0	0	0	1	0
Hedera helix L.	0	7	0	0	2	2	2	0
Hieracium lactucella Wallr.	0	0	5	1	0	0	0	0
Hieracium pilosella L.	0	0	0	2	0	0	0	0
Hippocrepis emerus (L.) Lassen	0	4	0	0	0	0	0	0
Hippuris vulgaris L.	0	1	0	0	0	0	0	0
Holcus lanatus L.	0	3	0	4	3	2	5	0
Holcus mollis L.	0	0	0	2	0	0	0	0
Homogyne alpina (L.) Cass.	0	0	4	0	0	0	0	0
Hordeum secalinum Schreb.	0	1	0	0	0	0	0	0
Humulus lupulus L.	3	3	0	0	0	0	2	0
Hydrocotyle vulgaris L.	0	0	0	0	1	0	1	0
Hypericum maculatum Crantz	0	0	4	0	0	1	0	0
Hypericum perforatum L.	0	1	0	0	0	0	0	3
Hypericum tetrapterum Fr.	0	0	0	0	3	0	2	0
Hypochaeris radicata L.	0	1	0	0	0	0	0	0

espèce	3	5_6	7.1	7.2	7.3_9.2	7.autres	10	11.autres
Imperatoria ostruthium L.	0	0	2	0	0	0	0	0
Inula helvetica Weber	0	0	0	0	0	0	1	0
Inula salicina L.	0	0	0	0	0	0	1	0
Iris pseudacorus L.	2	5	0	0	3	0	2	0
Juglans regia L.	0	2	0	0	0	3	0	0
Juncus acutiflorus Ehrh. ex Hoffm.	0	0	2	7	0	2	1	0
Juncus acutus L.	0	0	0	2	0	0	0	0
Juncus alpinoarticulatus Chaix	0	0	7	0	0	0	0	0
Juncus articulatus L.	0	1	2	1	5	2	2	4
Juncus compressus Jacq.	0	0	0	0	0	0	0	7
Juncus conglomeratus L.	0	0	4	2	1	0	1	0
Juncus effusus L.	0	0	2	6	1	3	2	0
Juncus filiformis L.	0	0	2	5	0	0	0	0
Juncus inflexus L.	0	1	0	0	3	5	1	0
Juncus pyrenaicus Timb.-Lagr. & Jeanb.	0	0	0	1	0	0	0	0
Juncus squarrosus L.	0	0	0	5	0	0	0	0
Juncus subnodulosus Schrank	0	0	0	0	5	1	7	0
Juniperus communis L.	0	0	4	1	0	6	0	0
Juniperus sibirica Lodd. ex Burgsd.	0	0	1	0	0	0	0	0
Knautia maxima (Opiz) J.Ortmann	0	0	4	1	0	0	0	0
Lactuca serriola L.	6	0	0	0	0	0	0	6
Lamium maculatum (L.) L.	0	1	0	0	0	0	0	0
Laserpitium prutenicum L.	0	0	0	0	1	0	3	0
Lathyrus linifolius (Reichard) Bässler	0	0	0	1	0	0	0	0
Lathyrus palustris L.	0	0	0	0	0	0	2	0
Lathyrus pratensis L.	0	3	0	2	3	3	5	0
Laurus nobilis L.	0	6	0	0	0	0	0	0
Leersia oryzoides (L.) Sw.	0	1	0	0	0	0	0	0
Leontodon autumnalis L.	0	0	6	0	0	0	0	0
Leontodon autumnalis L. subsp. autumnalis	0	0	0	0	0	0	2	0
Leontodon hispidus L.	0	0	2	1	0	0	0	0
Lepidium heterophyllum Benth.	2	0	0	0	0	0	0	0
Leucanthemum vulgare Lam.	0	0	0	0	1	0	4	0
Ligustrum vulgare L.	0	5	0	0	0	5	2	0
Limonium narbonense Mill.	7	0	0	0	0	0	0	0
Linaria repens (L.) Mill.	2	0	0	0	0	0	0	0
Linum bienne Mill.	4	0	0	0	0	0	0	0
Linum catharticum L.	0	0	0	0	6	0	3	0
Lolium perenne L.	2	5	0	0	0	0	1	0
Lolium rigidum Gaudin	4	0	0	0	0	0	0	3
Lonicera periclymenum L.	0	0	0	1	0	0	0	0
Lonicera xylosteum L.	0	1	0	0	0	7	2	0
Lotus alpinus (DC.) Schleich. ex Ramond	0	0	2	0	0	0	0	0
Lotus corniculatus L.	3	3	0	2	1	0	4	6

espèce	3	5_6	7.1	7.2	7.3_9.2	7.autres	10	11.autres
Lotus glaber Mill.	0	0	0	0	3	0	3	0
Lotus maritimus L.	0	0	0	0	3	3	0	0
Lotus pedunculatus Cav.	0	2	3	7	1	1	1	0
Ludwigia peploides (Kunth) P.H.Raven	0	3	0	0	0	0	0	0
Luzula campestris (L.) DC.	0	0	0	2	0	0	0	0
Luzula multiflora (Ehrh.) Lej. subsp. multiflora	0	0	1	6	1	2	0	0
Luzula sudetica (Willd.) Schult.	0	0	5	3	0	0	0	0
Lycopus europaeus L.	2	4	0	0	2	0	1	3
Lysimachia nummularia L.	0	2	0	0	0	1	0	0
Lysimachia vulgaris L.	2	3	3	1	7	2	5	0
Lythrum salicaria L.	0	3	0	0	6	1	5	0
Malus sylvestris Mill.	0	0	0	0	0	0	0	3
Malva sylvestris L.	5	0	0	0	0	0	0	7
Medicago arabica (L.) Huds.	2	0	0	0	0	0	0	3
Medicago lupulina L.	2	0	0	0	1	0	1	3
Medicago minima (L.) L.	0	0	0	0	0	0	0	3
Medicago polymorpha L.	5	0	0	0	0	0	0	0
Melampyrum pratense L.	0	0	0	1	0	0	0	0
Melica ciliata L.	0	0	0	0	0	0	0	7
Melilotus albus Medik.	0	1	0	0	0	0	0	7
Mentha aquatica L.	0	3	0	0	4	3	7	0
Mentha longifolia (L.) Huds.	0	1	0	0	0	1	0	0
Mentha spicata L.	4	0	0	0	0	0	0	0
Menyanthes trifoliata L.	0	0	6	5	6	2	0	0
Meum athamanticum Jacq.	0	0	0	5	0	0	0	0
Molinia caerulea (L.) Moench	0	1	3	5	6	6	5	0
Molinia caerulea (L.) Moench subsp. arundinacea (Schrank) K.Richt.	0	0	0	0	0	6	0	0
Mutellina purpurea (Poir.) Reduron, Charpin & Pimenov subsp. purpurea	0	0	7	0	0	2	0	0
Myosotis scorpioides L.	0	0	1	2	0	2	1	0
Myriophyllum spicatum L.	0	3	0	0	0	0	0	0
Narcissus poeticus L. subsp. radiiflorus (Salisb.) Baker	0	0	0	0	0	3	0	0
Nardus stricta L.	0	0	7	6	0	0	0	0
Oenanthe lachenalii C.C.Gmel.	0	0	0	0	1	0	5	0
Osyris alba L.	0	2	0	0	0	0	0	0
Parietaria judaica L.	2	0	0	0	0	0	0	0
Parietaria officinalis L.	0	6	0	0	0	0	0	0
Parnassia palustris L.	0	0	7	3	4	2	2	0
Parthenocissus inserta (A.Kern.) Fritsch	0	4	0	0	0	0	0	0
Pedicularis palustris L.	0	0	6	0	0	0	0	0
Pedicularis sylvatica L.	0	0	1	5	0	0	0	0
Pedicularis verticillata L.	0	0	0	0	0	1	0	0
Phalaris arundinacea L.	2	7	0	0	0	0	0	0
Phleum alpinum L.	0	0	3	2	0	0	0	0

espèce	3	5_6	7.1	7.2	7.3_9.2	7.autres	10	11.autres
Phleum pratense L.	0	3	0	0	0	2	2	0
Phragmites australis (Cav.) Steud.	3	3	0	0	6	3	3	7
Picea abies (L.) H.Karst.	0	0	2	5	1	6	0	0
Picris echioides L.	7	0	0	0	0	0	0	4
Picris hieracioides L.	5	0	0	0	0	0	0	7
Pinguicula alpina L.	0	0	2	0	0	0	0	0
Pinguicula vulgaris L.	0	0	7	2	0	1	0	0
Pinus sylvestris L.	0	0	0	2	0	7	0	0
Pinus uncinata Ramond ex DC.	0	0	7	2	0	0	0	0
Plantago lanceolata L.	3	4	0	2	2	1	2	0
Plantago major L.	2	1	0	0	0	0	1	0
Plantago media L.	0	0	0	0	0	0	2	0
Platanthera bifolia (L.) Rich.	0	0	0	0	4	2	0	0
Poa alpina L.	0	0	1	0	0	0	0	0
Poa bulbosa L.	0	0	0	0	0	0	0	4
Poa pratensis L.	0	2	2	0	0	1	1	0
Poa trivialis L.	0	3	0	3	3	3	4	3
Polygala amarella Crantz	0	0	0	0	3	0	0	0
Polygala vulgaris L.	0	0	0	1	0	0	0	0
Polygonum amphibium L.	0	1	0	0	0	0	0	0
Polygonum aviculare L.	2	1	0	0	0	0	0	0
Polygonum bistorta L.	0	0	3	6	1	3	0	0
Polygonum mite Schrank	0	2	0	0	0	0	0	0
Polygonum persicaria L.	0	3	0	0	0	0	0	0
Polygonum viviparum L.	0	0	7	0	0	0	0	0
Polypogon maritimus Willd.	3	0	0	0	0	0	0	0
Polypogon monspeliensis (L.) Desf.	2	0	0	0	0	0	0	3
Populus alba L.	3	6	0	0	0	0	0	0
Populus nigra L.	0	7	0	0	0	0	0	0
Populus tremula L.	0	0	0	0	0	1	1	0
Potentilla aurea L.	0	0	1	0	0	0	0	0
Potentilla erecta (L.) Rausch.	0	0	6	6	5	4	4	0
Potentilla palustris (L.) Scop.	0	0	0	5	0	0	1	0
Potentilla reptans L.	0	3	0	0	0	0	3	5
Primula elatior (L.) Hill subsp. elatior	0	0	0	0	0	2	0	0
Primula farinosa L.	0	0	2	1	2	0	0	0
Prunella vulgaris L.	0	1	3	2	2	1	5	0
Prunus padus L. subsp. padus	0	0	0	0	0	0	6	0
Prunus spinosa L.	0	1	0	0	0	1	2	0
Quercus petraea Liebl.	0	0	1	0	0	1	0	0
Quercus pubescens Willd.	0	4	0	0	0	0	0	0
Quercus robur L.	0	2	0	2	1	3	2	0
Ranunculus aconitifolius L.	0	0	1	0	0	7	0	0
Ranunculus acris L.	0	4	3	3	1	3	3	0

espèce	3	5_6	7.1	7.2	7.3_9.2	7.autres	10	11.autres
Ranunculus angustifolius DC.	0	0	5	0	0	0	0	0
Ranunculus bulbosus L.	0	0	0	0	0	0	3	0
Ranunculus flammula L.	0	0	1	2	2	0	1	0
Ranunculus grenierianus Jord.	0	0	1	0	0	0	0	0
Ranunculus montanus Willd.	0	0	1	0	0	0	0	0
Ranunculus repens L.	0	3	0	2	2	3	3	0
Ranunculus sardous Crantz	5	0	0	0	0	0	0	0
Ranunculus tuberosus Lapeyr.	0	0	0	0	5	0	0	0
Reynoutria japonica Houtt.	0	2	0	0	0	0	0	0
Reynoutria sachalinensis (F.Schmidt) Nakai	0	2	0	0	0	0	0	0
Rhamnus cathartica L.	0	0	0	0	2	0	4	0
Rhinanthus alectorolophus (Scop.) Pollich	0	0	0	0	1	0	1	0
Rhinanthus minor L.	0	0	0	3	0	0	0	0
Rhododendron ferrugineum L.	0	0	2	0	0	0	0	0
Rhynchospora alba (L.) Vahl	0	0	0	6	0	0	0	0
Robinia pseudoacacia L.	2	7	0	0	0	0	0	0
Rosa arvensis Huds.	0	0	0	0	0	1	0	0
Rosa canina L.	0	1	0	0	0	0	0	0
Rubia peregrina L.	0	4	0	0	0	0	0	0
Rubia tinctorum L.	2	0	0	0	0	0	0	0
Rubus caesius L.	0	7	0	1	2	1	2	0
Rubus fruticosus L.	5	3	1	1	1	2	2	0
Rubus idaeus L.	0	0	3	4	0	0	0	0
Rubus ulmifolius Schott	0	2	0	0	0	0	0	3
Rumex acetosa L.	0	1	2	4	0	3	1	0
Rumex acetosa L. subsp. acetosa	0	1	0	0	0	0	0	0
Rumex arifolius All.	0	0	1	0	0	0	0	0
Rumex crispus L.	3	3	0	0	0	0	0	7
Rumex obtusifolius L.	0	0	0	0	0	0	2	0
Rumex pulcher L.	2	0	0	0	0	0	0	0
Rumex sanguineus L.	0	1	0	0	0	0	0	0
Ruscus aculeatus L.	0	5	0	0	0	0	0	0
Salix alba L.	0	5	0	0	0	0	0	0
Salix aurita L.	0	0	4	2	0	0	0	0
Salix caesia Vill.	0	0	0	0	0	3	0	0
Salix caprea L.	0	0	1	0	0	6	0	0
Salix cinerea L.	0	1	1	1	7	2	4	0
Salix daphnoides Vill.	0	0	0	0	0	1	0	0
Salix eleagnos Scop. subsp. angustifolia (Cariot & St.-Lag.) Rech.f.	0	3	0	0	0	0	0	0
Salix foetida Schleich. ex DC.	0	0	1	0	0	0	0	0
Salix myrsinifolia Salisb.	0	0	0	0	1	0	0	0
Salix pentandra L.	0	0	0	0	1	0	0	0
Salix purpurea L.	0	2	0	0	0	1	0	0

espèce	3	5_6	7.1	7.2	7.3_9.2	7.autres	10	11.autres
Salix repens L. subsp. repens	0	0	0	4	2	0	0	0
Sambucus nigra L.	0	7	0	0	0	0	0	0
Samolus valerandi L.	0	0	0	0	0	0	0	7
Sanguisorba minor Scop.	0	4	0	0	0	0	0	0
Sanguisorba officinalis L.	0	0	0	2	3	3	4	0
Saponaria officinalis L.	2	7	0	0	0	0	0	0
Scabiosa atropurpurea L.	5	0	0	0	0	0	0	0
Schoenoplectus tabernaemontani (C.C.Gmel.) Palla	0	0	0	0	0	0	3	0
Schoenus ferrugineus L.	0	0	0	0	3	0	0	0
Schoenus nigricans L.	0	0	0	0	5	7	4	0
Scirpoides holoschoenus (L.) Soják	2	0	0	0	0	0	0	0
Scirpus sylvaticus L.	0	0	1	0	0	4	0	0
Scorzonera humilis L.	0	1	0	4	0	0	0	0
Scutellaria galericulata L.	2	0	0	0	2	0	0	0
Scutellaria minor Huds.	0	0	0	1	0	0	0	0
Selaginella selaginoides (L.) P.Beauv. ex Schrank & Mart.	0	0	6	0	0	0	0	0
Selinum carvifolia (L.) L.	0	0	0	0	6	0	1	0
Senecio inaequidens DC.	2	0	0	0	0	0	0	0
Serratula tinctoria L.	0	0	0	0	0	0	1	0
Silaum silaus (L.) Schinz & Thell.	0	2	0	0	3	1	3	0
Silene armeria L.	3	0	0	0	0	0	0	0
Silene flos-cuculi (L.) Clairv.	0	1	1	2	3	2	1	0
Silene latifolia Poir. subsp. alba (Mill.) Greuter & Burdet	0	0	0	0	0	0	0	7
Solanum dulcamara L.	2	1	2	0	1	0	0	0
Soldanella alpina L.	0	0	2	0	0	0	0	0
Solidago gigantea Aiton	0	2	0	0	5	1	7	0
Solidago virgaurea L.	0	0	1	0	0	0	0	0
Sonchus arvensis L.	5	0	0	0	0	0	0	0
Sonchus asper (L.) Hill subsp. asper	0	0	0	0	0	0	0	3
Sonchus tenerrimus L.	2	0	0	0	0	0	0	0
Sorbus aria (L.) Crantz	0	0	0	0	0	3	0	0
Sorbus aucuparia L.	0	0	2	6	0	0	0	0
Sparganium emersum Rehmman	0	1	0	0	0	0	0	0
Stachys officinalis (L.) Trevis.	0	0	0	0	2	0	6	0
Stachys palustris L.	0	1	0	0	0	0	0	0
Stellaria alsine Grimm	0	0	0	0	0	1	0	0
Stellaria graminea L.	0	0	5	2	0	5	0	0
Succisa pratensis Moench	0	2	1	4	4	3	5	0
Swertia perennis L.	0	0	2	2	0	6	0	0
Tamus communis L.	0	4	0	0	0	0	0	0
Taraxacum officinale gr.	2	3	0	0	1	0	5	0
Thalictrum flavum L.	0	2	0	0	0	0	0	0
Thelypteris palustris Schott	0	0	0	0	6	0	0	0
Thyselinum palustre (L.) Hoffm.	0	0	0	0	5	0	3	0

espèce	3	5_6	7.1	7.2	7.3_9.2	7.autres	10	11.autres
Tofieldia calyculata (L.) Wahlenb.	0	0	1	1	0	2	0	0
Torilis arvensis (Huds.) Link	3	0	0	0	0	0	0	0
Torilis arvensis (Huds.) Link subsp. purpurea (Ten.) Hayek	0	0	0	0	0	0	0	3
Tragopogon porrifolius L.	3	0	0	0	0	0	0	0
Trichophorum alpinum (L.) Pers.	0	0	1	6	0	3	0	0
Trichophorum cespitosum (L.) Hartm.	0	0	7	6	1	1	0	0
Trifolium angustifolium L.	2	0	0	0	0	0	0	0
Trifolium aureum Pollich	4	0	0	0	0	0	0	0
Trifolium badium Schreb.	0	0	7	0	0	1	0	0
Trifolium campestre Schreb.	0	0	0	0	0	0	0	5
Trifolium maritimum Huds.	3	0	0	0	0	0	0	0
Trifolium medium L.	0	0	3	0	0	0	0	0
Trifolium montanum L.	0	0	0	0	0	5	0	0
Trifolium ochroleucon Huds.	0	0	0	0	1	0	0	0
Trifolium pratense L.	0	3	3	2	2	2	5	0
Trifolium repens L.	2	2	2	3	0	2	0	3
Trifolium resupinatum L.	3	0	0	0	0	0	0	0
Trifolium spadiceum L.	0	0	5	1	0	0	0	0
Trigonella esculenta Willd.	0	0	0	0	0	0	0	3
Trisetum flavescens (L.) P.Beauv.	0	0	0	1	0	0	2	0
Trollius europaeus L.	0	0	1	2	1	1	0	0
Typha domingensis (Pers.) Steud.	0	0	0	0	0	0	0	7
Typha latifolia L.	0	0	0	0	0	0	0	7
Ulmus minor Mill.	3	7	0	0	0	0	0	0
Urtica dioica L.	3	7	0	1	2	1	2	0
Vaccinium microcarpum (Turcz. ex Rupr.) Schmalh.	0	0	0	6	0	0	0	0
Vaccinium myrtillus L.	0	0	4	7	0	0	0	0
Vaccinium oxycoccus L.	0	0	0	0	0	0	4	0
Vaccinium uliginosum L. subsp. microphyllum (Lange) Tolm.	0	0	1	0	0	0	0	0
Vaccinium uliginosum L. subsp. uliginosum	0	0	0	3	0	0	0	0
Vaccinium vitis-idaea L.	0	0	0	7	0	0	0	0
Valeriana dioica L.	0	0	5	3	4	3	3	0
Valeriana officinalis L.	0	1	0	0	1	0	0	0
Veratrum lobelianum Bernh.	0	0	1	7	0	0	0	0
Verbena officinalis L.	0	0	0	0	0	0	0	7
Veronica anagalloides Guss.	0	0	0	0	0	0	1	0
Veronica arvensis L.	0	0	0	0	0	0	0	7
Veronica chamaedrys L.	0	0	0	2	0	1	0	0
Veronica officinalis L.	0	0	6	3	0	0	0	0
Veronica persica Poir.	2	0	0	0	0	0	0	0
Veronica scutellata L.	0	0	1	0	0	0	0	0
Viburnum lantana L.	0	0	0	0	0	7	0	0
Viburnum opulus L.	0	1	1	0	4	2	2	0

