



HAL
open science

Définition d'un indice de typicité des relevés en vue de classer automatiquement les relevés floristiques dans les systèmes phytosociologique et habitats

Jean-Claude Gégout, Benoît Renaux

► To cite this version:

Jean-Claude Gégout, Benoît Renaux. Définition d'un indice de typicité des relevés en vue de classer automatiquement les relevés floristiques dans les systèmes phytosociologique et habitats. *Revue forestière française*, 2010, 62 (3-4), pp.449-462. 10.4267/2042/38959 . hal-03449699

HAL Id: hal-03449699

<https://hal.science/hal-03449699>

Submitted on 25 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DÉFINITION D'UN INDICE DE TYPICITÉ DES RELEVÉS EN VUE DE CLASSER AUTOMATIQUEMENT LES RELEVÉS FLORISTIQUES DANS LES SYSTÈMES PHYTOSOCIOLOGIQUE ET HABITATS

JEAN-CLAUDE GÉGOUT – BENOÎT RENAUX

CONTEXTE

Le classement des communautés végétales est nécessaire pour leur reconnaissance, leur exploitation ou leur protection. Le système de classement le plus répandu en Europe, basé sur la composition floristique des communautés végétales, est le classement phytosociologique. Il a été développé dans la première partie du XX^e siècle par Flahault et Braun-Blanquet (Braun, 1913 ; Flahault et Schroeter, 1910) et a largement été utilisé ensuite (Guinochet, 1973). Le classement des relevés floristiques dans ce système est traditionnellement réalisé sur la base de l'expertise des opérateurs à partir des espèces caractéristiques des unités phytosociologiques. Le nombre de spécialistes en phytosociologie est actuellement réduit relativement aux besoins élevés de classement et de cartographie des communautés végétales induits par le programme de conservation Natura 2000 qui repose sur une définition phytosociologique des communautés à sauvegarder à l'échelle européenne. Ces besoins importants, couplés au développement à travers la France et l'Europe d'immenses bases de données comprenant plus de 1 800 000 relevés floristiques (Schaminée *et al.*, 2009), poussent au développement de méthodes formalisées et informatisées du rattachement des relevés au système phytosociologique.

Plusieurs études récentes ont amélioré la définition quantitative du caractère diagnostique des espèces (Chytrý *et al.*, 2002a ; Tichý et Chytrý, 2006 ; de Caceres *et al.*, 2008 ; Willner *et al.*, 2009). La méthode générale consiste actuellement à définir le caractère diagnostique de toutes les espèces vis-à-vis de chaque unité sur la base de leur fréquence dans l'unité par rapport à leur fréquence en dehors de l'unité. Une étude récente montre que l'indice de fidélité sélectionné est moins important que le jeu de données utilisé comme référence (quelques associations végétales ou des milliers de placettes couvrant une vaste diversité de conditions) pour définir les espèces diagnostiques d'une unité (Willner *et al.*, 2009). Les indices de fidélité des espèces définis, le rattachement des relevés est réalisé à partir des espèces les plus diagnostiques du relevé ou de l'ensemble de son cortège floristique (Guinochet, 1973).

Le classement automatisé à large échelle des relevés, dans les associations végétales définies tout au long du siècle dernier, se heurte à deux problèmes principaux. Le manque de numérisation des relevés classés dans les associations dans une base de données exhaustive empêche le calcul de la fidélité des espèces dans les associations à l'échelle de vastes territoires et de conditions variées. Les travaux actuels à l'échelle internationale utilisent de vastes bases de relevés non classés dans le système phytosociologique et recréent des nouvelles unités au détri-

ment d'une cohérence avec les milliers de travaux phytosociologiques réalisés auparavant et d'une application au système Natura 2000. Le second problème, plus fondamental, repose sur la sélection de l'emplacement des relevés dans l'espace et sur l'absence de prise en compte de l'évolution lente des communautés sous l'effet des changements environnementaux. La façon traditionnelle consiste à positionner les relevés dans des endroits homogènes floristiquement et le plus souvent "typiques", au sens où les sites possèdent l'ensemble floristique caractéristique attendu (Guinochet, 1973). Les sites à flore banale ou pauvre en espèces caractéristiques sont généralement délaissés. Certaines approches actuelles de prise de données, basées sur un échantillonnage aléatoire ou systématique du territoire (fréquent dans le cas de cartographies) conduisent à la réalisation de relevés dans des conditions de milieu hétérogènes, artificialisées ou à flore très pauvre en espèces diagnostiques d'associations végétales. De plus, les changements environnementaux actuels font dériver lentement la composition des communautés d'un syntaxon à l'autre ou hors des syntaxons connus. C'est pourquoi les jeux de relevés réalisés récemment contiennent généralement une partie de relevés typiques des associations définies traditionnellement et une partie de relevés banaux ou intermédiaires entre plusieurs associations qui sont difficiles à distinguer rapidement des relevés typiques.

L'objectif de cet article est de proposer une méthode de classement automatique des communautés végétales des forêts françaises dans les associations du système phytosociologique traditionnel qui permette de distinguer les relevés typiques, faciles à classer dans les associations, des relevés intermédiaires faisant transition entre plusieurs associations ou des relevés à flore banale délicats à classer. Les objectifs spécifiques sont les suivants :

- formaliser un indice de typicité/qualité de rattachement d'un relevé à un syntaxon comme la probabilité qu'il soit classé automatiquement dans la même unité que celle choisie par un expert en phytosociologie. Cet indice est faible pour les relevés de transition ou pauvres en espèces diagnostiques et élevé pour les relevés typiques à nombreuses espèces caractéristiques d'une seule association ;

- modéliser l'indice de typicité proposé à partir de milliers de relevés issus de 100 associations phytosociologiques forestières parmi les plus fréquentes en France dont les relevés ont été classés préalablement dans le système phytosociologique ;

- valider l'indice de typicité par le classement automatique de 800 relevés tests classés également par des experts phytosociologues ;

- montrer l'intérêt de l'indice créé pour isoler les relevés typiques des relevés banaux d'une base de 6 444 relevés réalisés de façon systématique dans les forêts françaises par l'Inventaire forestier national.

L'enjeu est de relier les méthodes récentes de réalisation, de stockage et d'analyse des relevés floristiques avec l'approche phytosociologique traditionnelle pour mieux connaître la distribution spatiale et temporelle des habitats européens, localiser les plus rares et favoriser leur protection.

DÉFINITION D'UN INDICE DE TYPICITÉ D'UN RELEVÉ VIS-À-VIS D'UN SYNTAXON

La démarche repose d'abord sur le classement des relevés dans les syntaxons sur la base des espèces diagnostiques qu'ils contiennent. La typicité d'un relevé vis-à-vis d'un syntaxon est ensuite analysée en fonction de l'importance des espèces diagnostiques qu'il contient et de son caractère exclusif d'une unité ou intergrade entre plusieurs unités.

Sélection d'un indice de fidélité des espèces pour le classement des relevés

Le classement des relevés dans les syntaxons repose sur la définition et le calcul d'un indice de fidélité des espèces. Un relevé est classé dans l'association pour laquelle il contient le plus d'espèces fidèles à cette association. L'indice de fidélité retenu est le coefficient d'association ϕ (Sokal et Rohlf, 1995, p. 741, 743 dans Chytrý *et al.*, 2002a) qui permet de déterminer la fidélité d'une espèce à une unité à partir des notes de présence absence dans cette unité et dans le reste du jeu de données :

$$\Phi = \frac{N \times n_p - n \times N_p}{\sqrt{n \times N_p \times (N - n) \times (N - N_p)}}$$

où : N = nombre de relevés dans le jeu de données ;
 N_p = nombre de relevés dans l'association étudiée ;
 n = nombre d'occurrences de l'espèce dans le jeu de données ;
 n_p = nombre d'occurrences de l'espèce dans l'association étudiée.

Ce coefficient, largement utilisé depuis les années 2000 (Chytrý *et al.*, 2002b ; Willner *et al.*, 2009), varie de 1, lorsque l'espèce est présente dans tous les relevés de l'association et absente dans tous les autres, à - 1, lorsque l'espèce est absente de tous les relevés de l'association et présente dans tous les autres. Les valeurs positives indiquent une fréquence de l'espèce dans l'association plus grande qu'attendue par le hasard (espèce caractéristique de l'association). Les valeurs proches de 0 indiquent que l'espèce est distribuée de façon aléatoire dans l'unité et le reste du jeu de données (espèce ubiquiste à large amplitude écologique). Les valeurs négatives indiquent que l'espèce est moins fréquente qu'attendu par le hasard dans l'association (espèce dite accidentelle dans les tableaux phytosociologiques). Le calcul du coefficient ϕ à partir de vastes bases de données conduit à la définition d'espèces diagnostiques généralement différentes de celles identifiées de façon traditionnelle sur la base de quelques associations voisines de l'association étudiée. Dans le premier cas, les espèces diagnostiques sont définies par rapport à un vaste ensemble d'associations sans référence particulière aux associations voisines alors que, dans l'approche traditionnelle, les espèces diagnostiques sont définies par rapport aux syntaxons voisins. Dans ce cas, elles ont souvent leur optimum en dehors des associations proches de référence et peuvent de ce fait être peu diagnostiques à vaste échelle (Chytrý *et al.*, 2002a ; Kuzelova et Chytrý, 2004).

Définition d'un indice de typicité des relevés

La démarche employée repose sur l'hypothèse que plus un relevé est caractéristique d'une association avec de nombreuses espèces diagnostiques de cette association et peu d'autres espèces, plus les différentes approches de classement convergeront vers une association unique. À l'inverse, plus le relevé contiendra des espèces peu caractéristiques ou correspondra à un relevé intermédiaire entre plusieurs syntaxons, plus les différentes approches de classement conduiront à des résultats divergents. Deux paramètres peuvent être définis pour caractériser ces critères : le niveau d'affinité (A) du relevé relié au nombre et au caractère diagnostique des espèces caractéristiques qu'il contient et le niveau de différenciation (D) du relevé d'autant plus grand que le risque de confusion entre l'habitat sélectionné et d'autres habitats sera faible. Un relevé sera considéré comme "typique" lorsqu'il contiendra beaucoup d'espèces diagnostiques (A élevé) d'un seul syntaxon (D élevé).

Le niveau d'affinité (A) du relevé sera d'autant plus grand que le relevé possède un grand nombre d'espèces très diagnostiques par rapport aux espèces des relevés de référence utilisés par les experts pour définir l'association phytosociologique (figure 1, p. 452). Ce critère peut être formalisé par l'écart entre l'indice de fidélité du relevé (égal à la moyenne des ϕ des espèces

qu'il contient) et la moyenne des indices de fidélité des relevés classés par les experts dans l'unité sélectionnée :

$$A = (\phi_{ru} - m\phi_u) / s\phi_u$$

- où
- A est le niveau de fidélité du relevé ;
 - ϕ_{ru} est l'indice de fidélité ϕ du relevé à classer r dans l'unité u ;
 - $m\phi_u$ et $s\phi_u$ sont respectivement la moyenne et l'écart-type des indices de fidélité des relevés classés dans l'unité par les experts.

Un écart positif indique que le relevé est plus caractéristique que les relevés sélectionnés par les experts pour définir l'unité. Un fort écart négatif signifie que le relevé contient des espèces banales par rapport aux relevés de référence. Le critère non pondéré par l'écart-type des relevés dans l'unité a été testé et montre une efficacité un peu moindre que celle du critère retenu.

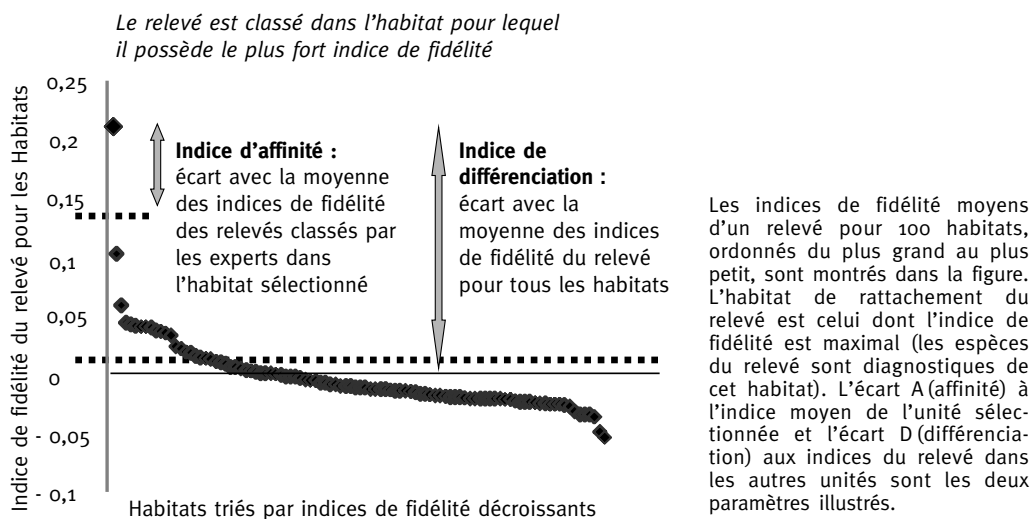
Le niveau de différenciation sera d'autant plus grand que le risque de confusion entre plusieurs unités sera faible (figure 1, ci-dessous). Les relevés intergrades contenant des espèces caractéristiques de plusieurs unités auront un indice faible. Ce critère peut être formalisé par l'écart entre l'indice de fidélité du relevé dans l'unité sélectionnée et l'indice de fidélité du relevé dans les autres unités forestières :

$$D = (\phi_{ru} - m\phi_r) / s\phi_r$$

- où
- D est le niveau de différenciation du relevé ;
 - ϕ_{ru} est l'indice de fidélité du relevé r dans l'unité u ;
 - $m\phi_r$ et $s\phi_r$ sont respectivement la moyenne et l'écart-type des indices de fidélité du relevé pour toutes les unités phytosociologiques du jeu de données.

Un écart faible, proche de 1 ou 2, signifie que le relevé présente un indice de fidélité semblable pour plusieurs unités conduisant à une plus forte probabilité d'erreur de classement. Un écart élevé, largement supérieur à 2, signifie que le risque de confusion entre plusieurs associations

FIGURE 1 ILLUSTRATION DES DEUX CRITÈRES PERMETTANT D'ÉVALUER LE NIVEAU DE CERTITUDE DE CLASSEMENT IDENTIQUE ENTRE EXPERT ET PROGRAMME DE CLASSEMENT AUTOMATIQUE



est faible. L'écart de l'indice de fidélité entre les 2 unités aux indices les plus élevés a la même signification que l'indice choisi et présente une efficacité prédictive comparable.

L'intérêt de ces deux critères comme indices de typicité ou de facilité/sûreté de classement des relevés dans les unités du système phytosociologique peut être testé en observant si leurs variations expliquent la différence ou l'identité de classement dans une même association entre expert et classement automatique. Ce test passe par la modélisation par les critères d'affinité et de différenciation de la variable binaire 1 = « classement identique entre expert et programme » – 0 = « classement différent entre expert et programme ». Un modèle logistique convient parfaitement à la modélisation de cette variable booléenne. Il fournit une probabilité de classement identique entre expert et programme qui varie en fonction de la valeur des paramètres du modèle. Ce modèle peut s'écrire :

$$p_{id} = \exp(a+b.A+c.D)/(1+ \exp(a+b.A+c.D))$$

où p_{id} est la probabilité que les classements expert et automatique des relevés soient identiques, A et D sont respectivement les critères d'affinité et de différenciation calculés pour les relevés ; a, b et c sont les coefficients du modèle logistique ajustés sur le jeu de données des 4 080 relevés de calibration.

p_{id} , qui combine en un seul indice les critères d'affinité et de différenciation, apparaît comme une méthode formelle de définir la typicité d'un relevé. Cet indice de typicité varie de 1 en cas de certitude de classement identique d'un relevé par les experts et le programme (le relevé contient de nombreuses espèces diagnostiques d'un seul syntaxon) à 0 en cas d'incertitudes très fortes sur le classement d'un relevé (espèces très banales caractéristiques de plusieurs syntaxons).

CLASSEMENT PHYTOSOCIOLOGIQUE ET CALCUL DE L'INDICE DE TYPICITÉ POUR LES RELEVÉS DE 100 ASSOCIATIONS FORESTIÈRES FRANÇAISES

Données de calibration et de validation

Les données sont constituées de 4 880 relevés floristiques des forêts françaises issus des bases de données EcoPlant et Sophy (Gégout *et al.*, 2005 ; Brisse *et al.*, 1995) et utilisés pour réaliser le document caractérisant la flore et l'écologie des habitats forestiers français de la France tempérée (Gégout *et al.*, 2007). Ils ont été classés jusqu'au niveau de l'association par des experts en phytosociologie. Les relevés sélectionnés correspondent aux relevés des 100 associations les plus abondantes dans la base de données et qui contiennent au minimum 15 relevés. Ces 100 associations sont incluses dans 30 alliances et 10 classes qui couvrent toutes les conditions écologiques et floristiques de la France collinéenne et montagnarde. Les 10 classes forestières présentes en France et définies dans le Prodrôme des végétations de France (Bardat *et al.*, 2004) sont représentées par au moins une association. Il s'agit des classes concernant les saulaies arbustives alluviales (*Salicetea purpureae*), les aulnaies et saulaies marécageuses (*Alnetea glutinosae*), les forêts tempérées caducifoliées ou mixtes, collinéennes, montagnardes ou plus rarement subalpines (*Quercu roboris-Fagetea sylvaticae*), les chênaies vertes (*Quercetea ilicis*) et les pineraies à genévrier (*Pino sylvestris-Juniperetea sabinae*) pour les milieux méditerranéens, et enfin les pessières et sapinières (*Vaccinio myrtilli-Piceetea abietis*), les pineraies et les mélézins (*Loiseleurio procumbentis-Vaccinietea microphylli*, *Erico carnea-Pinetea sylvestris*) montagnards et subalpins.

Vingt relevés de chacune des 40 associations contenant le plus de relevés ont été sélectionnés aléatoirement pour valider la méthode de classement des relevés et l'indice de typicité. Les

4 080 relevés restants ont été utilisés pour calculer les indices de fidélité des espèces vis-à-vis des associations, rattacher les relevés aux différentes associations et pour modéliser l'indice de typicité.

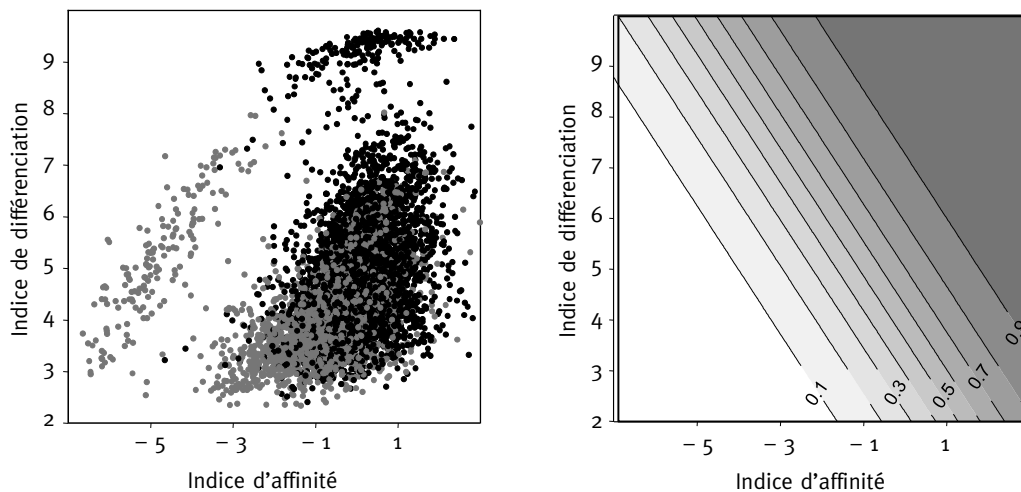
Le coefficient ϕ de toutes les espèces du jeu de données a été calculé pour chacune des 100 associations forestières étudiées à partir des 4 080 relevés du jeu de calibration. Un indice de fidélité a ensuite été calculé par relevé pour chacune des 100 associations étudiées à partir de la moyenne des indices de fidélité des espèces qu'il contient. Des tests utilisant seulement les espèces les plus caractéristiques pour calculer l'indice de fidélité des relevés ont été réalisés. Ils fournissent des taux de classements identiques à celui d'experts comparables à celui utilisant l'ensemble de la flore du relevé. Chaque relevé est classé dans l'association pour laquelle son indice de fidélité est le plus élevé, ce qui signifie que les espèces qu'il contient sont les plus caractéristiques de cette association.

**Classement automatique des relevés
et prédiction d'un classement identique entre expert et programme automatique**

Le classement automatique des relevés fournit un résultat identique à celui des experts pour 60 % des 4 080 relevés du jeu de calibration. Parmi les 100 associations de départ, 12 ne sont jamais prédites. Pour 7 d'entre elles, les relevés sont classés dans des vicariantes géographiques localisées sur une autre partie du territoire français. Les 5 autres sont des associations de chênaies dont les relevés sont classés dans des associations de hêtraies à flore proche. Dans les 88 autres associations, les pourcentages de relevés classés par le programme rattachés dans la même association par les experts varient de 0 à 100 % selon l'association avec une moyenne égale à 84 %. Le niveau d'affinité (variable A) des relevés prédits dans les associations varie de + 3 écart-types par rapport à la moyenne des relevés classés par les experts dans l'association à des valeurs très faibles atteignant jusqu'à - 6 écart-types (figure 2a), soit des relevés qui

FIGURE 2 LIEN ENTRE LA SIMILITUDE DE CLASSEMENT ENTRE EXPERT ET PROGRAMME ET LES NIVEAUX D’AFFINITÉ ET DE DIFFÉRENCIATION DES RELEVÉS

- a) Présentation des 4 080 relevés du jeu de calibration en gris lorsque les classements expert et programme diffèrent et en noir lorsqu'ils sont identiques.
- b) Évolution de la probabilité de classement identique entre expert et programme issue du modèle logistique intégrant ces deux variables.



contiennent des espèces beaucoup moins caractéristiques que les relevés classés dans l'association par les experts. L'écart d'indice de fidélité des relevés entre l'association de rattachement et la moyenne des indices de fidélité pour les 100 associations étudiées (variable D) varie d'environ + 2 à + 10 écart-types par rapport à la moyenne des indices de fidélité des relevés pour les 100 associations étudiées (figure 2a). Les écarts les plus forts correspondent aux classements sans risque de confusion entre plusieurs unités.

La fréquence de classement identique entre experts et programme varie clairement avec les niveaux d'affinité et de différenciation du relevé. Les relevés contenant plus d'espèces caractéristiques qu'en moyenne pour les relevés de référence (niveau d'affinité supérieur à 0) sont en général classés de façon identique par les experts et le programme (figure 2a). Les relevés avec un indice de fidélité pour une association très supérieur à celui calculé pour les autres associations (niveau de différenciation élevé, par exemple supérieur à 5) sont également classés en général de façon identique par le programme et les experts. Ces deux variables ont un effet significatif lorsqu'elles sont introduites seules ou ensemble dans le modèle logistique de prédiction de classement identique entre experts et programme. Les modèles incluant chacun des paramètres considérés isolément sont également présentés (tableau I).

TABEAU I **Déviations et coefficients des modèles logistiques de prédiction d'un classement identique entre expert et programme des relevés floristiques dans les associations du système phytosociologique.**

Dans toutes les équations, les paramètres ont un effet hautement significatif d'un point de vue statistique.

Variable	Déviante	constante	Coefficient Var. 1	Coefficient Var. 2
Déviante nulle.	5 501			
Indice d'affinité (A)	4 538	0,681	0,937	
Indice de différenciation (D) . .	4 771	- 3,243	0,779	
A + D	4 190	- 2,140	0,751	0,594

Le modèle isole très clairement les relevés peu typiques qui ont de fortes chances d'être classés différemment par le programme et les experts. Un relevé contenant à la fois peu d'espèces caractéristiques et des espèces caractéristiques de plusieurs associations comme le relevé 3 du tableau II, p. 456 (indice d'affinité : - 2,8, indice de différenciation : 2,7) présente 6 % de chance d'être classé de façon identique par les experts et le programme. Par contre, cette probabilité atteint 50 % si le niveau d'affinité atteint 0,5 ou le niveau de différenciation atteint 8 (figure 2b, ci-contre). La probabilité de classement identique dépasse 90 % si les deux paramètres ont simultanément des valeurs élevées comme dans le cas du relevé 1 du tableau II (indice d'affinité : 2,8, indice de différenciation : 6,4). Dans ce cas, le rattachement sans ambiguïté du relevé dans le *Festuco altissimae-Abietetum albae* Issler 1924 em. Hubert 1986 est lié à de nombreuses espèces caractéristiques uniquement de cette association qui induisent des valeurs élevées pour les indices d'affinité et de différenciation du relevé.

VALIDATION DE L'INDICE DE TYPICITÉ SUR UN JEU DE DONNÉES INDÉPENDANT

Le classement des 800 relevés du jeu de validation, non utilisés pour calculer les paramètres de l'indice de typicité des relevés, aboutit à un classement identique à celui des experts pour 47 % des relevés. Le lien entre la probabilité de classement identique expert-programme et la propor-

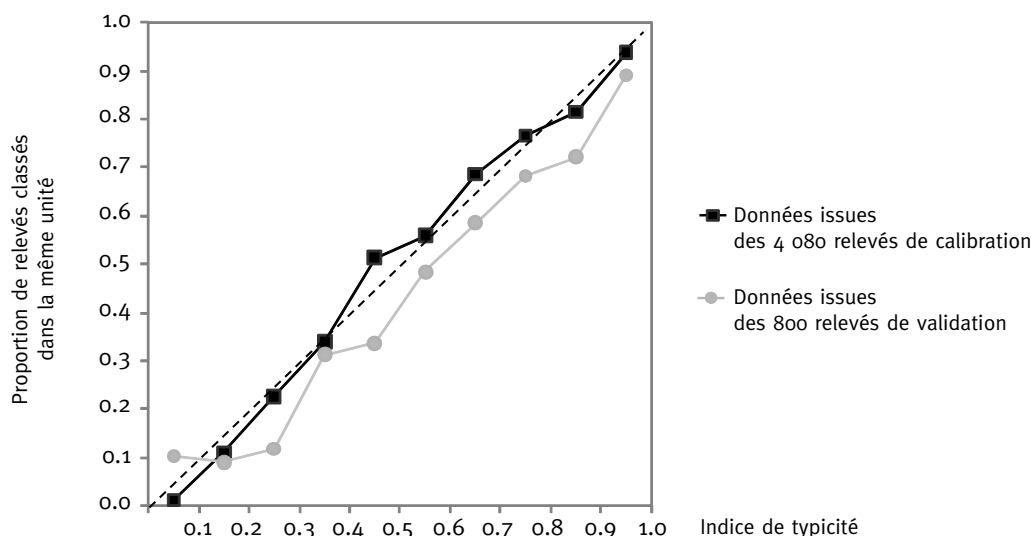
TABLEAU II **Exemple de trois relevés classés de façon automatique dans la sapinière-hêtraie à Fétuque des bois (*Festuco altissimae-Abietetum albae*) avec des indices de typicité (probabilité de classement identique entre expert et programme) décroissants.**

Pour chaque relevé, les indices de fidélité ainsi que les indices d'affinité, différenciation et typicité sont fournis pour le *Festuco altissimae-Abietetum albae*. L'association correspondant au deuxième choix du classement automatique et l'indice de fidélité associé sont indiqués en fin de tableau. Les indices Φ supérieurs à 0,1 des espèces sont également indiqués à droite du tableau pour les 3 associations correspondant au premier ou au deuxième choix du classement automatique. Abréviations : F-A : *Festuco altissimae-Abietetum albae* Issler 1924 em. Hubert 1986 ; G-F : *Carici flaccae-Fagetum sylvaticae* Thill 1964 = *Galio odorati-Fagetum sylvaticae* Rübél 1930 ex. Sougnez et Thull 1959 p.p. ; L-F : *Luzulo luzuloidis-Fagetum sylvaticae* Meusel 1937.

Espèces	Relevés			Φ espèces		
	R1	R2	R3	F-A	G-F	L-F
<i>Senecio nemor. subsp. fuchsii</i>	1			0,45		
<i>Rubus idaeus</i>	1			0,29		
<i>Impatiens noli-tangere</i>		1		0,26		
<i>Athyrium filix-femina</i>	1			0,26		
<i>Dryopteris filix-mas</i>	1	1		0,21		
<i>Mycelis muralis</i>			1	0,11		
<i>Festuca altissima</i>	1	1		0,57		0,20
<i>Dryopteris dilatata</i>	1	1		0,32		0,32
<i>Abies alba</i>	1	1	1	0,31		0,32
<i>Oxalis acetosella</i>	1			0,28		0,18
<i>Luzula luzuloides</i>	1			0,26		0,42
<i>Picea abies</i>			1	0,16		0,23
<i>Galium odoratum</i>	1	1		0,14	0,23	
<i>Fagus sylvatica</i>	1	1	1	0,12	0,10	0,13
<i>Lamium galeobdolon</i>		1			0,24	
<i>Mercurialis perennis</i>		1			0,20	
<i>Tilia platyphyllos</i>		1			0,20	
<i>Acer pseudoplatanus</i>		1	1		0,17	
<i>Corylus avellana</i>		1			0,16	
<i>Prunus avium</i>		1			0,16	
<i>Fraxinus excelsior</i>			1		0,16	
<i>Acer platanoides</i>		1			0,12	
<i>Deschampsia flexuosa</i>			1			0,18
<i>Veronica officinalis</i>			1			0,12
<i>Polypodium vulgare</i>			1			
<i>Rubus fruticosus</i>	1					
<i>Cardamine heptaphylla</i>		1				
<i>Helleborus foetidus</i>		1				
<i>Asplenium trichomanes</i>			1			
<i>Salix caprea</i>			1			
<i>Sorbus aria</i>			1			
<i>Populus tremula</i>			1			
<i>Betula pendula</i>			1			
Classement	F-A	F-A	F-A			
Ind. fidélité <i>Festuco alt.-Abietetum alb.</i>	0,27	0,12	0,05			
Indice d'affinité (A)	2,83	- 1,11	- 2,84			
Indice de différenciation (D)	6,39	3,69	2,66			
Typicité (probabilité de classement identique expert-programme)	0,98	0,31	0,06			
Classement 2 ^e choix	L-F	G-F	L-F			
Indice fidélité association 2 ^e choix	0,14	0,09	0,04			

tion des relevés effectivement classés de façon identique a été examiné par classes de 0,1 unité de probabilité. Le nombre de relevés varie de 29 à 128 relevés dans chaque classe. La relation est très étroite entre probabilité de bon classement et proportion réelle de relevés classés de façon identique (figure 3, ci-dessous). Les écarts, négatifs ou positifs, sont très faibles avec le jeu de calibration. Le jeu de validation montre que le taux de relevés réellement classés de façon identique est 5 à 10 % inférieur au taux prédit par le modèle. La proportion de relevés classés avec une probabilité de classement élevée, supérieure à 0,8, atteint 20 %.

FIGURE 3
LIEN ENTRE LA PROBABILITÉ DE CLASSEMENT IDENTIQUE
ENTRE EXPERT ET PROGRAMME FOURNIE PAR LE MODÈLE
ET LA PROPORTION DE RELEVÉS EFFECTIVEMENT CLASSÉS DE FAÇON IDENTIQUE
PAR LES PHYTOSOCIOLOGUES ET LE PROGRAMME DE CLASSEMENT



APPLICATION AU CLASSEMENT DE RELEVÉS DE L'INVENTAIRE FORESTIER NATIONAL DANS LES SYSTÈMES PHYTOSOCIOLOGIQUE ET HABITATS

L'Inventaire forestier national mesure chaque année 6 000 à 7 000 placettes incluant un inventaire floristique dans les forêts françaises. Ces placettes sont sélectionnées de façon aléatoire sur une maille de 500 m de côté disposée sur les zones forestières et sont réalisées toute l'année. La qualité des relevés et leur représentativité des associations phytosociologiques sont donc variables du fait de l'absence de sélection de sites représentatifs sur le terrain, de la pauvreté de la flore identifiable jusqu'à l'espèce en dehors de la saison de végétation ou de la situation des placettes dans des conditions de peuplements particulières (par exemple plantations). Les 6 444 relevés effectués en 2008 ont été classés automatiquement dans l'unité pour laquelle la moyenne des indices ϕ des espèces était la plus élevée. Ils ont été classés dans 71 habitats sur les 100 habitats possibles. Les habitats non rencontrés sont 8 habitats non prédits par le programme dans la phase d'élaboration du modèle et 20 habitats rares pour la plupart dont plusieurs liés aux bords de cours d'eau.

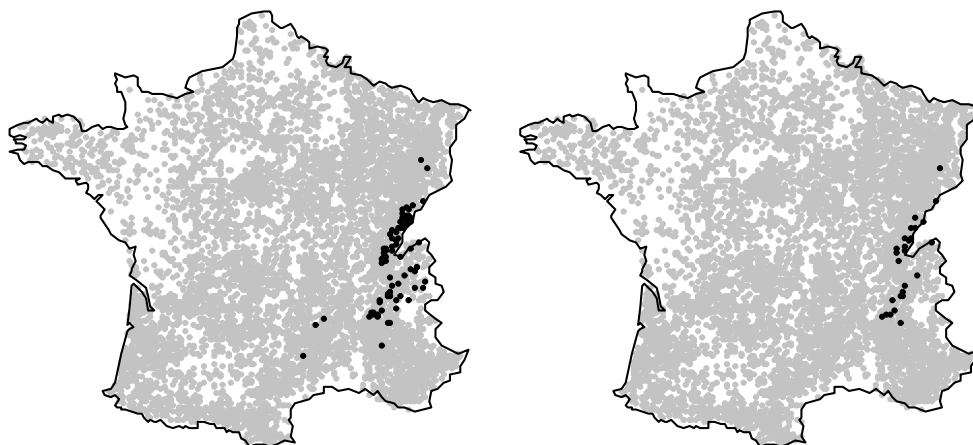
Les relevés classés dans les associations sont en général moins typiques que les relevés sélectionnés par les experts pour définir les associations : 17 % ont un indice de fidélité supérieur à

l'indice de fidélité moyen des relevés classés dans l'association par les experts. 32 % des relevés ont un indice de fidélité très inférieur à celui des relevés classés dans l'unité par les experts. Ces relevés possèdent une flore banale peu caractéristique d'une unité. La méthode d'échantillonnage et la prise de données tout au long de l'année expliquent cette différence. Il reste cependant remarquable que 17 % des placettes, soit 1 114 réalisées en 2008, restent plus typiques que la moyenne des relevés retenus pour définir les associations (en général une dizaine de relevés suffisent pour définir une association). D'autre part, plus de 50 % des relevés possèdent une quantité d'espèces caractéristiques semblable à ceux des relevés utilisés par les experts.

Un exemple d'utilisation du classement réalisé est utilisé pour étudier la répartition des relevés de hêtraies, hêtraies-sapinières calciclinales à Orge d'Europe (*Hordelymo europae-Fagetum sylvaticae* (Tüxen 1937) Kuhn 1937 em. Jahn 1972, code CORINE : 41-13, EUR27 : 9130-9). Les 81 relevés classés dans cette association sont prédits essentiellement dans le Jura et les Alpes du Nord. Quelques relevés sont prédits dans le Massif central et les Vosges, régions où cet habitat est normalement absent. Si on ne conserve que les 37 relevés comprenant une probabilité de classement identique à celui d'un expert > 0,3, ils sont tous localisés (à une exception près) dans le Haut Jura et les Alpes du Nord externes qui correspondent au cœur de la distribution de cet habitat (figure 4).

FIGURE 4 LOCALISATION DES RELEVÉS DE LA HÊTRAIE À ORGE D'EUROPE (*Hordelymo europae-Fagetum sylvaticae* (Tüxen 1937) Kuhn 1937 em. Jahn 1972) EN FRANCE À PARTIR DES RELEVÉS DE L'INVENTAIRE FORESTIER NATIONAL.

Les relevés de l'*Hordelymo europae-Fagetum sylvaticae* correspondent aux points noirs, les autres relevés aux points gris. À gauche, sans seuil de typicité. À droite, relevés dont le niveau de typicité est supérieur à 0,3.



DISCUSSION ET CONCLUSIONS

Le rattachement automatique des relevés aux associations phytosociologiques a déjà été tenté avec succès par différentes méthodes (Koci *et al.*, 2003 ; van Tongeren *et al.*, 2008 ; De Caceres *et al.*, 2009). L'indice de fidélité des espèces ϕ utilisé ici, validé par plusieurs études internationales (De Caceres *et al.*, 2008), permet de reproduire la démarche des experts en accordant une forte importance aux espèces diagnostiques des associations (combinaison caractéristique d'es-

pèces). La proportion de relevés classés avec cet indice de façon identique par les experts et le programme, de l'ordre de 50 %, est proche de ceux trouvés par Koci *et al.* (2003) pour 16 associations de milieu ouvert et par De Caceres *et al.* (2009) pour les associations des *Fagetalia sylvaticae* semblables à la plupart des groupements étudiés. Les différences de classement sont liées en partie à des relevés peu typiques et à la prise en compte d'un vaste territoire (la France) couvrant plusieurs zones biogéographiques associées chacune à une série d'associations parfois très proches floristiquement d'un territoire à l'autre (vicariance géographique). Les différentielles géographiques qui permettent normalement de les distinguer ne sont de plus pas systématiquement présentes dans les relevés, notamment dans les relevés pauvres en espèces et à flore banale.

Après les études sur le caractère diagnostique des espèces et les approches sur le rattachement automatique des relevés aux associations, l'indice de typicité proposé conduit à l'analyse des similitudes et différences de classement entre experts et programme automatique. Les différences peuvent provenir d'une erreur de l'expert ou du programme, mais aussi de la présence de relevés intermédiaires ou de relevés sans espèces caractéristiques qui sont difficiles à classer et que les experts préfèrent souvent ne pas classer. L'indice de typicité est un second indice caractérisant la communauté analysée qui s'ajoute à son classement dans le système phytosociologique. Il fournit un niveau de certitude de classement identique à celui d'un expert qui s'appuie, d'une part, sur le nombre et le caractère exclusif des espèces caractéristiques de l'association présentes dans le relevé (concept d'affinité) et, d'autre part, sur le caractère intermédiaire du relevé par rapport à plusieurs associations (concept de différenciation). Ces deux éléments sont combinés dans un seul indice qui varie de 0 à 1 et qui fournit la probabilité pour un relevé d'être classé de façon identique par un expert et un programme automatique. Il est probable que cet indice fournisse également des indications sur les possibilités de classements divergents entre experts. Le niveau de différenciation D, lié au caractère intermédiaire du relevé, s'apparente au classement par logique floue qui reconnaît des limites non nettes entre unités et accepte l'appartenance d'un relevé en proportion variable à plusieurs types. Il en diffère par le fait que l'importance des espèces est généralement égale dans l'approche par logique floue, alors que notre approche s'appuie sur les espèces diagnostiques comme l'approche phytosociologique traditionnelle.

L'indice de typicité proposé peut s'appliquer aux autres méthodes de rattachement à des groupes prédéfinis basées sur des espèces diagnostiques dont la fidélité s'exprime sous forme quantitative. Les coefficients calculés pour les principales associations des forêts françaises sont valides pour les milieux considérés et à l'échelle étudiée. Ils devront être recalculés pour d'autres régions et d'autres milieux et seront probablement différents pour des jeux de calibration limités à de petits territoires ou à un petit nombre d'unités pour lesquels le caractère diagnostique des espèces peut être plus élevé.

Près de deux millions de relevés floristiques ont été inventoriés en Europe dont la plupart ne sont pas classés dans le système phytosociologique (Schaminée *et al.*, 2009). Des milliers de placettes sont réalisées chaque année, souvent par des non phytosociologues et à partir d'échantillonnages systématiques ou aléatoires qui favorisent la sélection de relevés non typiques d'un point de vue phytosociologique. Parmi les relevés sélectionnés et classés par des phytosociologues en Catalogne, De Caceres *et al.* identifient 61 % de relevés intermédiaires. Le positionnement de ces relevés sur le terrain étant choisi dans des conditions floristiques et écologiques homogènes (Guinochet, 1973), la proportion de relevés non typiques échantillonnés de façon aléatoire ou systématique (dans le cadre de cartographies par exemple) est potentiellement très importante. En plus des outils de classement automatique créés récemment, il est donc utile de disposer d'un indice pour distinguer, parmi l'ensemble des données collectées, les relevés typiques d'unités des relevés incomplets, intermédiaires, ou pauvres en espèces diagnostiques. L'indice proposé présente l'intérêt d'être : opérationnel pour le système phytosociologique actuel

à la base du système Natura 2000, simple à implémenter à partir des caractéristiques des relevés et des communautés et basé sur les critères de doute utilisés par les experts lors du classement des relevés (absence d'espèces diagnostiques dans le relevé ou présence d'espèces diagnostiques de plusieurs unités). Un tel outil peut aider les phytosociologues à classer rapidement et automatiquement les relevés typiques, à s'attacher au classement des relevés peu typiques et à écarter si nécessaire les relevés banaux ou intermédiaires. Il ouvre également la voie du classement des communautés aux nombreux non phytosociologues qui peuvent fixer un seuil de typicité leur permettant de retenir les relevés classés avec une quasi-certitude de classement identique à celui que réaliserait un expert.

Les relevés classés comme typiques d'associations peuvent être utilisés individuellement pour identifier et localiser les sites à fort intérêt patrimonial comprenant par exemple des habitats de la directive "Habitats". Ils peuvent être utilisés collectivement pour évaluer la rareté des types d'habitats, connaître leur répartition géographique, ou déterminer les régions où leur flore est complètement exprimée. Le niveau de typicité des relevés classés dans une association peut également être utilisé pour détecter sa présence dans des zones biogéographiques où elle n'a pas été définie. Dans un contexte de changements environnementaux à long terme, il est important d'utiliser un système de classement formalisé pour détecter les modifications graduelles de flore. L'indice de typicité, sensible à l'appauvrissement en espèces diagnostiques comme à l'arrivée d'espèces caractéristiques d'autres unités, pourrait constituer un marqueur précoce du passage progressif d'une association à une autre association mieux adaptée aux nouvelles conditions climatiques. Son utilisation, conjointement à celle d'une méthode de classement formalisée de référence, permettrait une gestion fine des écosystèmes naturels à l'échelle des zones spéciales de conservation comme à celle du territoire national ou de l'Europe par tous les acteurs de la gestion des espaces naturels.

Jean-Claude GÉGOUT
UMR1092 Laboratoire d'étude des Ressources Forêt-Bois
(LERFoB)
AgroParisTech-ENGREF
14 rue Girardet CS 14216
F-54042 NANCY CEDEX
(jean-claude.gegout@engref.agroparistech.fr)

Benoît RENAUX
UMR1092 Laboratoire d'étude des Ressources Forêt-Bois
(LERFoB)
AgroParisTech-ENGREF
14 rue Girardet CS 14216
F-54042 NANCY CEDEX

— Actuellement —

CONSERVATOIRE BOTANIQUE NATIONAL
DU MASSIF CENTRAL
Le Bourg
F-43230 CHAVANIAC LAFAYETTE
(benoit_renaux@hotmail.com)

BIBLIOGRAPHIE

- BRAUN (J.), FURRER (E.). — Sur l'étude des associations. — *Bulletin de la société languedocienne de Géographie*, 26, 1913, pp. 2-22.
- BARDAT (J.), BIORET (F.), BOTINEAU (M.), BOULLET (V.), DELPECH (R.), GÉHU (J.-M.), HAURY (J.), LACOSTE (A.), RAMEAU (J.-C.), ROYER (J.-M.), ROUX (G.), TOUFFET (J.). — Prodrôme des végétations de France. — Paris : Muséum national d'histoire naturelle, 2004. — 171 p. (Patrimoines naturels, vol. 61).
- BRISSE (H.), de RUFFRAY (P.), GRANDJOUAN (G.), HOFF (M.). — European vegetation survey. The phytosociological database "Sophy". — *Annali di botanica*, vol. 53, 1995, pp. 191-223.
- CHYTRÝ (M.), TICHÝ (L.), HOLT (J.), BOTTA-DUKAT (Z.). — Determination of diagnostic species with statistical fidelity measures. — *Journal of Vegetation Science*, vol. 13, n° 1, 2002a, pp. 79-90.
- CHYTRÝ (M.), EXNER (A.), HRIVNAK (R.), UJHAZY (K.), VALAKOVI (M.), *et al.* — Context-dependence of diagnostic species: A case study of the Central European spruce forests. — *Folia Geobotanica*, vol. 37, n° 4, 2002b, pp. 403-417.
- DE CACERES (M.), FONT (X.), OLIVA (F.). — Assessing species diagnostic value in large data sets: A comparison between phi-coefficient and Ochiai index. — *Journal of Vegetation Science*, 19, 2008, pp. 779-810.
- DE CACERES (M.), FONT (X.), VICENTE (P.), OLIVA (F.). — Numerical reproduction of traditional classifications and automatic vegetation identification. — *Journal of Vegetation Science*, vol. 20, n° 4, 2009, pp. 620-628.
- FLAHAULT (C.), SCHROETER (C.J.). — Rapport sur la nomenclature phytogéographique. In : Actes du III^e congrès international de botanique. Volume 1, comptes rendus des séances du 20 mai 1910, pp. 131-142. — Bruxelles : Albert de Boeck éditeur, 1910. — 121 p.
- GÉGOUT (J.-C.), COUDUN (C.), BAILLY (G.), JABIOL (B.). — EcoPlant: a forest site database linking floristic data with soil and climate variables. — *Journal of Vegetation Science*, vol. 16, n° 2, 2005, pp. 257-260.
- GÉGOUT (J.-C.), RAMEAU (J.-C.), RENAUX (B.), JABIOL (B.), BAR (M.). — Les habitats forestiers de la France tempérée ; typologie et caractérisation phytoécologique. — Nancy : AgroParisTech-ENGREF, 2007. — 716 p., 6 annexes.
- GUINOCHET (M.). — Phytosociologie. — Paris : Masson, 1973. — VI-228 pages (Écologie n° 1).
- KOCI (M.), CHYTRÝ (M.), TICHÝ (L.). — Formalized reproduction of an expert-based phytosociological classification: A case study of subalpine tall-forb vegetation. — *Journal of Vegetation Science*, vol. 14, n° 4, 2003, pp. 601-610.
- KUZELOVA (I.), CHYTRÝ (M.). — Interspecific associations in phytosociological data sets: how do they change between local and regional scale? — *Plant Ecology*, 173, 2004, pp. 247-257.
- SCHAMINÉE (J.H.J.), HENNEKENS (S.M.), CHYTRÝ (M.), RODWELL (J.S.). — Vegetation-plot data and databases in Europe: an overview. — *Preslia*, 81, 2009, pp. 173-181.
- SOKAL (R.R.), ROHLF (F.J.). — Biometry. — 3^e édition. — New York (NY, US) : Freeman, 1995.
- TICHÝ (L.), CHYTRÝ (M.). — Statistical determination of diagnostic species for site groups of unequal size. — *Journal of Vegetation Science*, 17, 2006, pp. 809-818.
- VAN TONGEREN (O.), GREMMEN (N.), HENNEKENS (S.). — Assignment of relevés to pre-defined classes by supervised clustering of plant communities using a new composite index. — *Journal of Vegetation Science*, vol. 19, n° 4, 2008, pp. 525-536.
- WILLNER (W.), TICHÝ (L.), CHYTRÝ (M.). — Effects of different fidelity measures and contexts on the determination of diagnostic species. — *Journal of Vegetation Science*, vol. 20, n° 1, 2009, pp. 130-137.

DÉFINITION D'UN INDICE DE TYPICITÉ DES RELEVÉS EN VUE DE CLASSER AUTOMATIQUEMENT LES RELEVÉS FLORISTIQUES DANS LES SYSTÈMES PHYTOSOCIOLOGIQUE ET HABITATS [Résumé]

La conservation des habitats à fort intérêt patrimonial repose sur leur identification dans les conditions naturelles. Ce travail vise à présenter une méthode de classement automatique et de hiérarchisation du caractère typique des relevés floristiques vis-à-vis de la classification phytosociologique à la base de la typologie Habitats. La méthode développée repose sur l'utilisation de 4 880 relevés classés par des experts dans 100 associations incluses dans 30 alliances et 10 classes phytosociologiques forestières présentes en France. La méthode de classement automatique permet de classer environ 50 % des relevés dans la même association que celle choisie par l'expert. Un indice nouveau de typicité est présenté qui permet de séparer les relevés banaux ou intermédiaires des relevés typiques d'association. Cet indice a permis d'isoler dans le jeu de données un quart de relevés très typiques d'associations qui sont classés dans 86 % des cas de façon identique entre experts et programme de classement automatique. Cet indice de typicité, qui repose sur la quantité d'espèces diagnostiques et sur l'absence de caractère intermédiaire du relevé entre plusieurs associations, est basé sur les mêmes critères de rattachement et d'incertitude que ceux utilisés par les phytosociologues de terrain. Il pourra être utilisé en complément de la méthode de classement automatique comme aide au classement des milliers de relevés réalisés annuellement par les gestionnaires des milieux naturels spécialistes ou pas en phytosociologie.

DEFINITION OF AN INDEX OF TYPICALITY FOR *RELEVÉS* FOR AUTOMATICALLY CLASSIFYING FLORISTIC *RELEVÉS* IN PHYTO-SOCIOLOGICAL SYSTEMS AND HABITATS [Abstract]

Conservation of forests that have high heritage value relies on their identification under natural conditions. This work aims at presenting a method for automatically classifying and ranking the typical character of floristic *relevés* for the purposes of the phytosociological classification that underlies the Habitats typology. The method developed is based on the use of 4880 *relevés* classified by experts among 100 associations included in 30 alliances and 10 phytosociological classes of forests present in France. The automated classification method classifies approximately 50% of the *relevés* in the same association as the expert. A new typicality index is described that distinguishes between general or intermediate *relevés* and those which are typical of an association. By using this index, one fourth of the set of data was distinguishable as being *relevés* highly typical of associations, which in 86% of cases are classified identically by the experts and by the automated classification software. This typicality index relies on the quantity of characteristic species and on the fact that the *relevé* is not intermediate between several associations. It calls on the same linking and uncertainty criteria as those used by field phytosociologists. It can be used to supplement the automatic classification method as an aid to classifying the thousands of *relevés* performed every year by managers of natural environments regardless of whether or not they are specialists in phytosociology.
