



HAL
open science

Cartographie de l'écologie scientifique et analyse diachronique

Marie-Laure Masquillier, Pascal Cuxac

► **To cite this version:**

Marie-Laure Masquillier, Pascal Cuxac. Cartographie de l'écologie scientifique et analyse diachronique. 1st. International Symposium ISKO-Maghreb, May 2011, Hammamet, Tunisie. hal-00955513

HAL Id: hal-00955513

<https://hal.science/hal-00955513>

Submitted on 4 Mar 2014

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.



Distributed under a Creative Commons Attribution 4.0 International License

Cartographie de l'écologie scientifique et analyse diachronique

Masquillier M.-L. ; Cuxac P.

INIST-CNRS, 2 allée du Parc de Brabois, CS 10310, 54519 Vandoeuvre les Nancy Cedex, FRANCE

marie-laure.masquillier@inist.fr ; pascal.cuxac@inist.fr

Résumé :

Cette étude offre un mode de représentation original de la littérature scientifique sous forme de cartes thématiques dans le but d'analyser de façon synthétique un domaine scientifique. Nous montrons comment la cartographie permet à un expert, ici un écologue, de visualiser le paysage de l'écologie scientifique tel qu'il apparaît aujourd'hui et il y a dix ans dans les publications scientifiques et d'identifier les relations entre les thèmes de recherche. Le résultat de la classification d'un corpus bibliographique issu de la base de données PASCAL de l'INIST-CNRS a fait l'objet d'une méta-analyse avec regroupement des thématiques et étude de leurs liens. Dans une perspective de veille scientifique, une analyse diachronique a également permis d'identifier les problématiques émergentes en écologie, à partir des thématiques nouvelles ou en évolution depuis ces dix dernières années.

Mots-clés : Ecologie, Publication scientifique, Infométrie, Cartographie, Diachronie

appuyant sur des données bibliographiques.

Dans un premier temps, nous avons comparé et analysé les deux cartographies obtenues à partir des articles en écologie publiés en 2009-2010, l'une représentant les thématiques de la recherche française, l'autre de la recherche mondiale, pour tenter d'appréhender les grandes tendances qui s'en dégagent et leurs spécificités.

Dans un second temps, nous avons comparé les deux cartographies obtenues à partir des articles en écologie publiés en 1999-2000 et 2009-2010 dans le monde pour visualiser l'évolution de ce champ de recherche.

Afin de mieux comprendre la progression et le devenir des thématiques de l'écologie, nous nous sommes finalement appuyés sur une analyse diachronique portant sur ces deux périodes d'étude.

Pour obtenir ces résultats, nous avons appliqué des méthodes à base de classification automatique de textes et à base de règles d'associations.

I. INTRODUCTION

Dans un contexte de production scientifique croissante, l'analyse infométrique des publications scientifiques, notamment signalées dans les bases de données bibliographiques, permet d'analyser leur contenu et d'en identifier les acteurs, leurs relations, et leurs moyens de communication [1].

Dans cette perspective, nous nous sommes intéressés à l'analyse de la littérature scientifique dans le domaine de l'écologie, partie des sciences biologiques définie comme "la science qui étudie les conditions d'existence des êtres vivants et les interactions de toutes natures qui existent entre ces êtres vivants d'une part, entre ces être vivants et leur milieu d'autre part" [2].

Notre objectif est de proposer à une communauté scientifique donnée, dans l'exemple choisi celle des écologues¹, une représentation du contenu de la production scientifique sous forme de cartes thématiques en nous

II. LES DONNÉES

Pour appliquer une méthodologie de type bibliométrique, nous avons utilisé la base de données bibliographique multidisciplinaire PASCAL de l'INIST-CNRS, dont l'information est structurée et organisée par l'intermédiaire d'une indexation et d'un ou plusieurs codes de classement.

La pluridisciplinarité de la base de données nous permet de couvrir tous les champs thématiques de l'écologie, de l'écologie fondamentale à l'écologie appliquée.

A partir de l'interrogation de la base de données PASCAL combinant les champs mots-clés, ISSN, codes de classement, date de publication et affiliation des auteurs, nous avons construit les corpus suivants :

- un corpus sur la recherche en écologie dans le monde en 2009-2010, avec 22 145 références bibliographiques
- un corpus sur la recherche en écologie en France 2009-2010 (un auteur au moins d'affiliation française), avec 1 796 références bibliographiques
- un corpus basé sur les publications dans le monde en 1999

¹ Les résultats de cette étude ont fait l'objet d'une présentation orale à la communauté des écologues français [12]

et 2000, avec 27 826 références.

La figure 1 montre un exemple de notice bibliographique issu de la base PASCAL. Nous utiliserons par la suite les champs descripteurs (FD = French Descriptor et FG = French Generic) pour représenter le document.

NO : PASCAL 09-0482193 INIST
ET : Fertilization-Effects on deciduous tree growth and browsing by moose
AU : MANSSON (Johan); BERGSTRÖM (Roger); DANELL (Kjell)
AF : Department of Wildlife, Fish, and Environmental Studies, Swedish University of Agricultural Sciences/907 83 Umed/Suède (1 aut., 2 aut., 3 aut.)
DT : Publication en série; Niveau analytique
SO : Forest ecology and management; ISSN 0378-1127; Coden FECMDW; Pays-Bas; Da. 2009; Vol. 258; No. 11; Pp. 2450-2455; Bibl. 3/4 p.
LA : Anglais
EA : Knowledge of plant-herbivore interactions is fundamental for understanding foraging patterns of herbivores and their effects on the ecosystem as well as the human use of natural resources. The nutrient resources available for plants may determine not only growth rate, but also the amounts of secondary compounds in the plant tissues unfertilized ...
CC : 002A33; 002A14B04B; 002A36C03
FD : Fertilisation; Croissance; Broutage animal; Forêt boréale; Herbivore; Feuille végétal; Foresterie; Plante feuillage caduc; Arbre; Alces alces; Betula; Suède; Ecologie forestière
FG : Artiodactyla; Ungulata; Mammalia; Vertebrata; Betulaceae; Dicotyledones; Angiospermae; Spermatophyta; Europe; Plante ligneuse; Cervidae; Aménagement sol; Appareil végétatif; Végétal; Animal ruminant; Pays Scandinaves; Fennoscandie; Pays nordiques; Europe du Nord
ED : Fertilization; Growth; Browsing; Boreal forest...
EG : Artiodactyla; Ungulata; Mammalia; Vertebrata; Betulaceae; ...

Fig. 1 : Exemple notice

Chaque corpus est ensuite mis sous la forme d'une matrice creuse booléenne, ou chaque ligne représente un vecteur document et chaque colonne un mot.

III. UNE MÉTHODE DE CLUSTERING APPLIQUÉE AU CORPUS

Les méthodes de classification non supervisées (clustering) peuvent être appliquées sur des corpus documentaires afin de détecter des groupes de documents assimilables à des thématiques particulières du corpus étudié. Elles ont la particularité de ne faire aucune hypothèse sur le nombre de classes à trouver. Le but est de trouver des classes regroupant des individus (les documents) très proches entre eux (les plus semblables possibles) et faire en sorte que les classes soient les plus dissemblables possibles.

Sur le corpus obtenu, mis sous forme de matrice, nous appliquons la méthode de clustering Neurodoc, méthode inspirée des K-means [3], afin de construire des classes et produire des cartes thématiques.

L'algorithme des K-means consiste à créer K groupes de documents de façon à ce que chaque groupe soit le plus dense possible et que ces groupes soient les plus distants possibles. L'utilisateur choisit le nombre de classes maximum désiré (K) et va initialiser K classes par un point représentant le centre de chaque classe. Les documents sont ensuite affectés à la classe dont ils sont le plus proche et le centre de la classe est recalculé jusqu'à stabilisation.

La méthode des K-means axiales (KMA) [4] diffère en

représentant les classes non plus par leur centre mais par un demi-axe. Les documents se projettent alors plus ou moins loin sur ces axes. Quand un document se projette sur un axe avec une valeur supérieure à un certain seuil fixé par l'utilisateur, le document est affecté à la classe. De la même façon on peut projeter les mots-clés sur ces axes. Les documents et mots-clés sont ordonnés selon un degré de ressemblance au type idéal de la classe et sont affectés d'un poids plus ou moins important suivant qu'ils se projettent plus ou moins haut sur ces axes.

IV. ANALYSE DIACHRONIQUE

Des travaux précédents [5] ont montré qu'on pouvait utiliser des règles d'associations pour mettre en évidence des liens, et donc une évolution, entre classes de classifications différentes.

Une règle d'association $A \rightarrow B$ extraite d'une base de données, représente un lien entre les deux ensembles de propriétés A et B. Pour mesurer la qualité de cette règle, on dispose de nombreux indices [6] fonctions de ces effectifs, dont les plus courants sont le support, qui est le nombre d'objets vérifiant les propriétés de A et de B, et la confiance, qui est le quotient de ce support et du nombre d'objets vérifiant les propriétés de A, c'est-à-dire du support de A. On peut aisément transposer ce formalisme de la fouille de données à la classification d'un corpus bibliographique, en prenant pour propriétés les classes et pour objets les mots-clés. Habituellement les règles d'association sont utilisées pour des données binaires (0/1). Ici, la méthode de clustering utilisée (KMA) permet d'affecter aux mots de chaque classe un poids. Utilisant ce poids, nous avons alors des données qui ne sont plus binaires et les règles d'association habituelles ne sont plus utilisables. On les remplace alors par des règles d'association floues [7]. Du point de vue pratique, on redéfinit l'appartenance aux classes et le support d'une classe de la façon suivante : pour chaque mot-clé, sa valeur d'appartenance à la classe qu'il caractérise correspond à son poids dans la classe. Ainsi, si le mot-clé i a le poids a_i pour la classe A, et b_i pour la classe B, sa valeur d'appartenance aux deux classes est la plus petite des deux valeurs, $\min(a_i, b_i)$. Si les valeurs d'appartenance de tous les mots-clés à une classe ne dépassent pas 0,5 le support de la classe est nul, sinon il est égal à la somme des valeurs d'appartenance. Une fois les supports calculés, la confiance de la règle d'association floue est alors définie comme celle de la règle d'association classique. Pour plus de détail on se rapportera à l'article de Cuxac et al. (2005) [5].

V. APPLICATION AUX CORPUS « ECOLOGIE »

Les résultats sur les corpus « Ecologie » sont présentés sous formes de cartes (figures 2, 3 et 4) développées à l'aide du module Neurodoc de la plateforme Stanalyst <http://stanalyst.inist.fr/>. Ils ont été obtenus par la méthode des K-means axiales en initialisant à cinquante classes. La lecture

des cartes peut se faire selon deux axes, horizontal et vertical.

L'axe horizontal est considéré comme représentant le milieu d'étude, du milieu naturel au milieu plus anthropisé (milieu pollué, ou d'élevage).

L'axe vertical est vu comme représentant l'objet d'étude, avec un niveau de complexité croissante de bas en haut qui va du niveau communauté/écosystème/biosphère au niveau organisme/individu/espèce. Cela revient à considérer des études à large échelle (synécologie et macroécologie) dans la partie inférieure de la carte et des études à l'échelle de l'individu (écophysiologie et autoécologie) en haut de la carte.

Chacune des cinquante classes est un indicateur de thèmes ou de centres d'intérêt autour desquels s'agrège l'information.

Le regroupement manuel des thèmes tient compte de leur proximité, mais aussi de la force des liens qui les relie.

A. La recherche en écologie dans le monde en 2010

La première carte obtenue (figure 2) représente l'écologie dans le monde telle qu'elle apparaît dans les publications en 2009-2010. Elle met en évidence les grandes tendances de la recherche durant cette période.

En partant du haut de la carte et dans le sens horaire inversé, le regroupement des classes proches par couleur permet de faire ressortir les principales thématiques du corpus en 2010 :

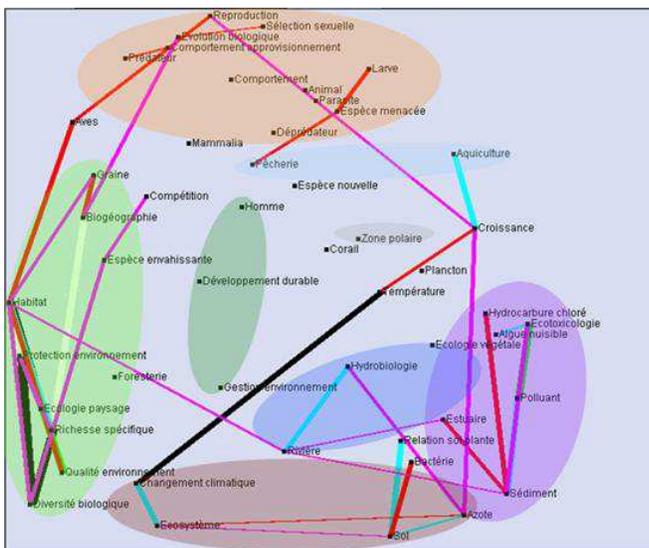


Fig. 2 : Carte « Ecologie 2010, recherche mondiale »

- l'écologie animale et évolutive rassemblant des études de comportement, évolution interactions et génétique d'organismes tels que parasites, prédateurs et déprédateurs ;
- l'écologie de la conservation regroupant biodiversité, habitat, paysage, et biogéographie ;
- l'étude des écosystèmes terrestres avec notamment les aspects de biogéochimie, écologie microbienne, services écosystémiques, impact du réchauffement climatique et

bilan carbone ;

- l'écotoxicologie aquatique réunissant bloom d'algues toxiques, eutrophisation par les nitrates, polluants chimiques tels que les PCB (polychlorobiphényles) et POP (polluants organiques persistants) ;
- l'écologie polaire, à mettre en relation avec la quatrième API (Année Polaire Internationale) qui a suscité de nombreuses études et publications ;
- l'écologie appliquée aux espèces exploitées par la pêche et l'aquaculture ;
- la gestion de l'environnement et le développement durable, avec la place centrale de l'homme.

Les classes les plus importantes en nombre de documents sont les classes « Ecotoxicologie » et « Foresterie », suivies de « Gestion environnement » « Protection environnement » et « Espèce menacée ». Les autres classes remarquables, par le nombre de documents dont les coordonnées sont supérieures au seuil, sont les classes « Habitat » et « Ecosystème ».

Une analyse bibliométrique sur les mots-clés nous permet de mettre en avant les notions les plus fréquentes durant cette période. En faisant abstraction des vingt premiers mots clés géographiques et taxonomiques, les mots-clés les plus fréquents sont toxicité, biodiversité et changement climatique.

B. La recherche en écologie en France en 2010

Cette carte a été réalisée à partir d'une requête sur les publications 2009-2010 dont un auteur au moins est d'affiliation française. Il s'en dégage différents grands thèmes (figure 3) :

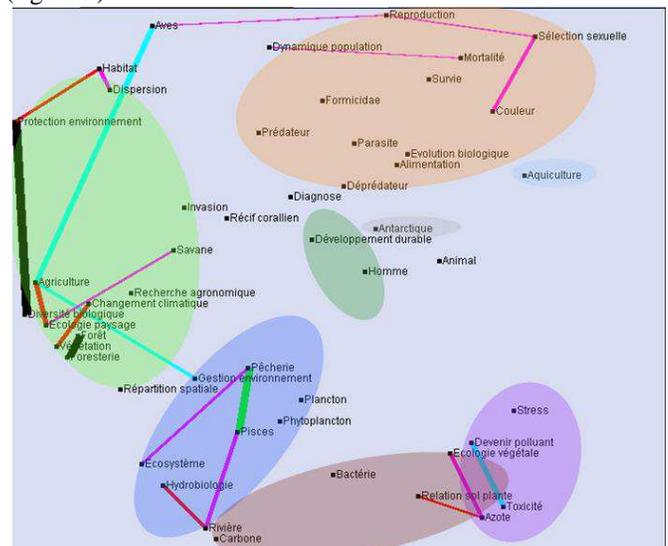


Fig. 3 : Carte « Ecologie 2010, recherche française »

- l'écologie animale et évolutive qui regroupe les traits d'histoire de vie, interactions et évolution d'organismes tels que les prédateurs, déprédateurs et parasites ;
- l'écologie de la conservation regroupant les thèmes de

biodiversité, de protection de l'environnement et d'agoécosystèmes ;

- l'étude des écosystèmes aquatiques, de leurs communautés, de leur exploitation et de leur productivité biologique ;
- l'étude des écosystèmes terrestres avec notamment les aspects de biogéochimie, écologie végétale et microbienne
- l'écotoxicologie où se trouvent les thèmes sur les polluants, le stress, et la toxicité ;
- le développement durable et l'homme au centre de la carte ;
- l'écologie polaire à mettre en relation avec la quatrième API (Année polaire internationale) qui a suscité de nombreuses études et publications en France également.

L'utilisation d'un même code couleur pour chaque métaclasse permet aisément de comparer les deux cartes, mondiale et française. Les cartes présentent de grandes similarités mais font apparaître aussi quelques particularités propres à la recherche française :

- en écologie animale, figure notamment une classe portant sur les Formicidae, une famille qui paraît relativement bien étudiée par les chercheurs français, en particulier en Guyane française ;
- en écologie de la conservation apparaît nettement une approche agroécologique avec les classes agriculture et recherche agronomique fortement liées à la protection de l'environnement ;
- en écologie polaire, les études portent majoritairement sur l'écosystème antarctique, ce qui est à mettre en relation avec l'implantation des bases scientifiques françaises au pôle Sud.

Les classes les plus importantes de ce corpus, en nombre de documents sont la foresterie, la toxicité et la biodiversité. Les autres classes remarquables par le nombre de documents dont les coordonnées sont supérieures au seuil, sont les classes des taxons Pisces et Aves, et la classe « Changement climatique ».

L'analyse bibliométrique sur les mots clés met en avant les notions importantes de ce corpus telles que changement climatique, reproduction, habitat, écotoxicologie et écologie forestière.

C. La recherche en écologie dans le monde en 2000

La carte des thèmes pour l'année 1999-2000 présente les caractéristiques suivantes (figure 4)

- l'écologie animale et évolutive et les interactions biotiques de type relations intraspécifiques, entre animal et végétal ou encore entre prédateur et proie ;
- l'écotoxicologie aquatique avec le devenir des polluants et leur accumulation biologique dans les organismes, la pollution des zones littorales, et l'eutrophisation ;

- l'étude des écosystèmes aquatiques avec l'exploitation, la productivité des communautés dulçaquicoles et saumâtres ;
- l'écologie de la conservation avec notamment la protection et gestion de l'environnement et du paysage, la diversité des espèces.

A droite de la carte, on remarque des métaclasses un peu différentes par rapport à la carte 2010 (figure 2) :

- la surveillance écologique des écosystèmes dégradés avec l'étude des espèces menacées, la fragmentation de l'habitat et la biorépartition ;
- la démécologie regroupant des études de populations et métapopulations, de recrutement et de densité dépendance.

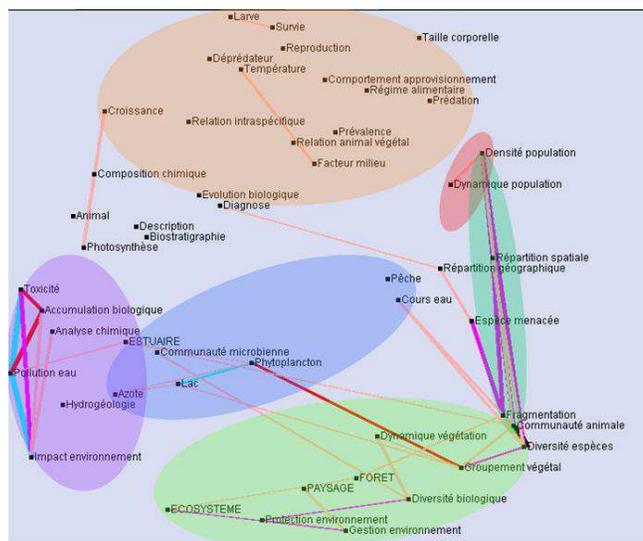


Fig. 4 : Carte « Ecologie 2000, recherche mondiale »

Les classes remarquables en nombre de documents pour cette période concernent les thématiques de pollution (toxicité, accumulation biologique), de dynamique et densité de population, et de fragmentation du paysage.

Ces tendances de la recherche sont corroborées par d'autres études bibliométriques sur les mots associés qui mettent en avant l'approche métapopulation et de fragmentation du paysage dans les publications à cette époque [8].

Les notions les plus fréquentes durant cette période mises en évidence par bibliométrie des mots clés sont celles portant sur la pollution et la toxicité, et la dynamique de population.

L'étude comparative des deux cartes 2000 et 2010 fait apparaître des différences dans l'interconnexion des thèmes. Ainsi, les classes présentent peu de liens entre elles dans la carte 2000 (fig 4) alors qu'elles sont beaucoup plus interconnectées en 2010 (fig 2), en particulier dans la partie inférieure de la carte. Cela traduirait le fait que les thématiques de recherche évoluent vers des études transversales et à plus large échelle : biogéographie de la biodiversité, macroécologie, services écosystémiques, changement global et effets du changement climatique.

Cette tendance de la recherche vers la macroécologie est aussi soulignée par la montée du facteur d'impact de revues développant ce thème comme *Global ecology and biogeography* : a journal of macroecology ou encore la revue *Ecography*.

Il est par ailleurs intéressant d'examiner le devenir des classes, notamment celles situées à droite de la carte 2000, et de suivre leur transformation dans le temps. Pour cela nous pouvons nous appuyer sur une analyse diachronique pour comparer les deux classifications (figure 2 et figure 4) et tenter de comprendre comment les classes ont évolué durant ces dix années.

D. Résultats de diachronie et analyse

L'étude diachronique met en évidence l'évolution temporelle des classes sur la période 2000 - 2010. Elle indique l'évolution du vocabulaire et permet de repérer les différents comportements des classes dont la stabilité peut varier dans le temps (figure 5).

L'analyse de ces résultats fait apparaître de nouvelles thématiques dans les publications scientifiques comme le développement durable, l'écologie polaire, ou encore l'étude d'espèces envahissantes telles que les algues nuisibles. Cela signifie que ces thématiques se sont développées durant ces dix dernières années, alors qu'elles n'étaient pas suffisamment importantes pour être individualisées en cluster dix ans auparavant.

Certaines classes, au contraire, vont disparaître logiquement comme « Hydrogéologie » et « Biostratigraphie » qui étaient liées à des artefacts d'indexation automatique corrigés par la suite.

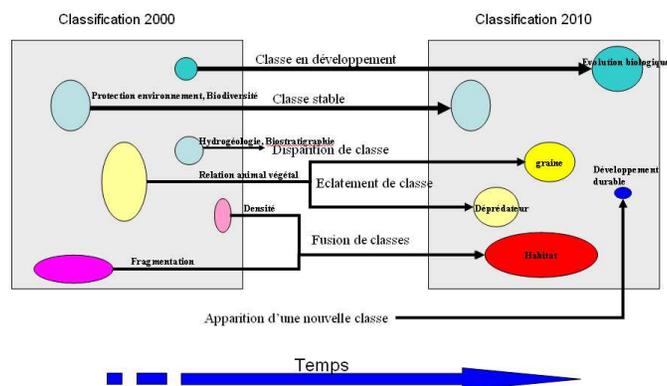


Fig. 5 : Diachronie

Il apparaît aussi des nouveaux noms de classe autour desquels s'agrègent l'information comme «Écologie du paysage», «Biogéographie» ou encore «Changement climatique».

Des classes vont éclater au cours du temps, comme «Relation animal végétal» qui va se retrouver dans la classe «Graine» qui sous tend des publications sur la dispersion par zoochorie et la classe «Déprédateur» liée à la lutte biologique.

D'autres vont fusionner, comme «Fragmentation», «Densité», « Répartition » qui vont se situer dans la classe «Habitat» représentant en 2010 un nouvel enjeu de conservation. Les classes «Relation intraspécifique» et «Reproduction» vont fusionner dans la classe «Sélection sexuelle» liée aux études de biologie évolutive.

Par ailleurs, l'analyse diachronique met en évidence le développement de la classe «Evolution biologique». Les mots clés ayant constitué cette classe sont relatifs aux études d'écologie évolutive et de génétique et des techniques qui s'y appliquent (microsatellite, DNA mitochondrial) comme a pu le montrer également M.W. Neff [9].

Enfin certaines classes ne changent pas, comme «Protection environnement» et «Biodiversité» qui restent stables et prépondérantes dans les publications durant ces dix dernières années.

VI. CONCLUSIONS ET PERSPECTIVES

L'intérêt de notre démarche est de montrer que les méthodes que nous utilisons, bien connues dans le milieu de l'analyse de données ou de l'intelligence artificielle, ont des applications diverses et notamment dans le cadre de la veille scientifique. Notre expérience montre que ces méthodes, peu utilisées par des non spécialistes pour des raisons de méconnaissance, trouvent un grand intérêt auprès d'un large public à condition de les accompagner dans leur démarche d'analyse

Les cartes thématiques développées à partir de la station d'analyse de l'information STANALYST (<http://recherche.inist.fr/>), permettent une représentation synthétique de la littérature scientifique quelque soit le champ disciplinaire, et constituent un outil de veille et d'aide à la décision par la mise en évidence de thématiques émergentes.

Toutefois, les résultats obtenus ne reflètent pas la production scientifique totale car les données bibliographiques utilisées ne sont pas exhaustives. Nous obtenons une représentation cartographique du contenu de la base de données PASCAL dans le temps, reflet de sa couverture documentaire, multidisciplinaire, et de son évolution.

Dans notre exemple, cette caractéristique de multidisciplinarité de la base PASCAL nous a permis de mettre en évidence toutes les sous-disciplines qui nourrissent l'écologie, désignée par D. Worster comme une « cacophonie de disciplines » [10], telles que l'écotoxicologie, l'agroécologie, la biogéographie, ou encore l'écologie de la conservation et le développement durable.

L'apport de l'infométrie dans un contexte de veille nous semble intéressant. En écologie, la cartographie des publications 2009-2010 a ainsi mis en exergue une spécificité de la recherche française (figure 3) concernant les problématiques d'agroécologie et de protection de l'environnement étroitement liée au domaine agricole, ainsi qu'une tendance au renforcement des études macroécologiques dans la recherche scientifique mondiale.

Les résultats de diachronie ont fait apparaître des thématiques nouvelles comme le développement durable et

l'écologie polaire, ou qui se développent comme l'écologie évolutive. On note également la récurrence et l'importance des thématiques de biodiversité et de changement climatique durant cette décennie.

Certaines de ces tendances de la recherche en écologie, mises en avant dans cette étude par les cartes et la diachronie, sont corroborées par d'autres auteurs [8], [9].

D'un point de vue méthodologique, la comparaison de classifications temporelles à l'aide de règles d'association floues basée sur les descripteurs, permettant de visualiser des relations orientées entre classes, apparaît comme un bon outil d'analyse d'évolution temporelle. Cependant, cette méthode, comme toutes celles faisant appel à des règles d'associations, produit un nombre important de règles qu'il peut être difficile de filtrer et d'analyser. Il apparaît également que telle quelle, la méthode peut parfois s'avérer aveugle à de fines évolutions. Pour pallier ces inconvénients nous travaillons au développement d'une méthode de comparaison par inférence bayésienne [11] et également sur une nouvelle méthode de clustering incrémental.

VII. BIBLIOGRAPHIE

- [1] Grivel L., François C. (1995) : Une station de travail pour classer, cartographier et analyser l'information bibliographique dans une perspective de veille scientifique). *Solaris*, 1995, n.2, p 81- 112
- [2] Dajoz E. (1975) : *Précis d'écologie*, Collection : Ecologie fondamentale et appliquée, 3e édition Paris, Dunod.
- [3] MacQueen J. B. (1967) : Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1: 281-297.
- [4] Lelu A. (1993) : *Modèles neuronaux pour l'analyse de données documentaires et textuelles*, Thèse de doctorat de l'université de Paris VI, 4 mars 1993, 238 pages.
- [5] Cuxac P., Cadot M., François C. (2005) : Analyse comparative de classifications : apport des règles d'association floues. *Revue des Nouvelles Technologies de l'Information RNTI-E-3 Vol. II, Pp.519-530 ; 5èmes journées d'Extraction et Gestion de Connaissances EGC*, Paris, Université Paris 5, 19-21 janvier 2005
- [6] Kodratoff Y. (2001) : Rating the Interest of Rules Induced from Data and within texts, *12th IEEE -International Conference on Database and Expert Systems Applications-Dexa 2001*, Munich, sept 2001, p. 265-269.
- [7] Cadot M., Napoli A. (2004) : Règles d'association et codage flou des données, *11èmes Rencontres de la Société Francophone de Classification, SFC 2004* (Bordeaux), pp. 130-133.
- [8] Nobis M., Wohlgemuth T. (2004) : Trend words in ecological core journals over the last 25 years (1978-2002). *Oikos*, 2004, vol 106, n.2, p 411-420
- [9] Neff M.W., A. Corley E.A. (2010) : 35 years and 160 000 articles : a bibliometric exploration of the evolution of ecology. *Scientometrics*, 2009, vol 80, n.3, p.657-682
- [10] Worster D., (1994) : *Nature's economy : a history of ecological ideas*. Second edition. Cambridge University Press, 1994
- [11] Lamirel J.C., Cuxac P., Haydar C., François C. (2010) : A new general paradigm for mining research evolving over time. *SIIE*, 18-20 janvier 2010, Sousse, Tunisie.
- [12] Masquillier M.L., Cuxac P. 2010. Etude infométrique diachronique de l'écologie scientifique, *Ecologie 2010*, Montpellier, France.

Marie-Laure MASQUILIER est Ingénieure d'Etudes au département Ingénierie de l'Information Scientifique à l'INIST/CNRS à Nancy (France). Après une maîtrise de Sciences naturelles et une année spéciale en documentation et gestion automatisée de l'information, elle intègre le CNRS en 1990, comme Ingénieure d'Etudes. Actuellement chargée des ressources

documentaires au service Sciences de la vie, elle veille à la constitution de la base de données PASCAL et à l'enrichissement des ressources terminologiques intégrées au système documentaire en écologie. Elle est responsable éditoriale du site d'actualité Recherches Polaires de l'INIST-CNRS (<http://recherchespolaires.veille.inist.fr/>).

Pascal CUXAC est Ingénieur de Recherche et Développement en Science de l'Information à l'INIST/CNRS à Nancy (France).

Docteur en "Génie Géologique et Minier", (thèse soutenue à l'Ecole de géologie de Nancy en 1991) il travaille sur le comportement mécanique de roches anisotropes. En 1993, il intègre le CNRS comme Ingénieur de Recherche. Actuellement, au sein du service Recherche Développement en Ingénierie de l'INIST (Institut de l'Information Scientifique et Technique), il participe aux programmes de recherche sur les méthodes de classifications de corpus bibliographiques et notamment sur les algorithmes incrémentaux de classification non supervisée. Dernièrement ses travaux abordent la qualité des résultats de clustering et l'analyse de graphes.