

Indicateurs utiles au management de la recherche issus d'un corpus multi-sources

MULTON Jean-Louis*, BRANCA-LACOMBE Geneviève**,
DOUSSET Bernard***, DAGUILLANES Cécile****

* INRA, Mission Relations Internationales, 147 rue de l'Université, 75338 Paris cedex 07

** INRA, DISI, BP 2078, 06606 Antibes cedex

*** IRIT, SIG, Université Paul Sabatier, 118, route de Narbonne, 31062 Toulouse cedex 04

**** Avant-Veille, Toulouse

multon@paris.inra.fr, lacombe@antibes.inra.fr, dousset@irit.fr, c.daguillanes@wanadoo.fr

Mots-clés :

Veille scientifique et technique ; Bibliométrie ; Collaborations internationales.

Keywords :

Science and technology watch; Bibliometry; International collaboration

Résumé :

Ce travail est à la fois une réactualisation de deux études similaires réalisées il y a respectivement deux et quatre ans sur les collaborations internationales de l'INRA et la réutilisation des corpus constitués à ces occasions pour des analyses ciblées de laboratoires, de régions spécifiques ou de domaines de recherche particuliers. Pour cette dernière analyse, nous avons complété le corpus qui avait été constitué à partir des bases bibliographiques sur CD/Rom : Current-Contents, Pascal et Sci. La version CD/Rom de Sci qui ne disposait ni du résumé, ni de mots-clés a été remplacée cette fois-ci par la version en ligne du Web of Science qui est beaucoup plus complète mais assez difficile à télécharger. Nous disposons maintenant de plus de six années de production scientifique (janvier 1998 à juillet 2004) ce qui nous permet une détection beaucoup plus complète des acteurs et de leurs relations ainsi que des thématiques abordées. Après avoir analysé les collaborations internationales de l'INRA, nous avons pu réaliser un zoom sur trois entités de cet institut : le département TPV (Traitement des Produits Végétaux), le positionnement de l'INRA aux Caraïbes (centres INRA des Caraïbes et pays de la région qui collaborent avec l'INRA local ou de métropole) ainsi qu'une étude plus ciblée sur le thème de l'écologie. Ce complément d'étude montre que le corpus initialement constitué pour les collaborations internationales renferme bien d'autres informations stratégiques. Chaque département, chaque laboratoire, chaque région ou chaque expert d'un thème donné peut ainsi prendre connaissance de son environnement, des forces et des faiblesses de son entité, des relations avec les autres organismes de recherche. Une politique de développement peut en être déduite en prenant en compte de façon rigoureuse des éléments objectifs mesurant la lisibilité de l'activité de recherche et de son évolution. Nous présentons, ici, les principaux résultats de ces différentes analyses et les perspectives à moyen et long terme de ce type d'étude.

1 INTRODUCTION

A la suite de l'étude présentée au Colloque VSST'2001 (Veille Stratégique Scientifique et Technologique, 15-19 octobre 2001 – Barcelone) sur la période 1998-1999, une nouvelle étude bibliométrique d'évaluation quantitative des collaborations internationales de l'INRA a été effectuée sur la période 1998-04, complétée par deux zooms : le département « Transformation des Produits Végétaux », et la zone Caraïbe. Ces études ont été récemment complétées, après une réactualisation du corpus, par un zoom sur le thème de l'écologie et une mise à jour des liens qu'entretient l'INRA à l'international.

Comme précédemment :

- Le logiciel retenu pour effectuer cette étude est « Tétralogie » (IRIT Toulouse) identifié comme étant celui qui répond le mieux au cahier des charges fixé. Cependant, la présente étude a bénéficié des nouvelles avancées de « Tétralogie » en matière de cartographie géostratégique et d'analyse relationnelle évolutive.
- Le choix des bases bibliographiques (Current-Contents, Pascal et Science Citation Index,) à partir desquelles a été établi le corpus INRA, a été strictement imposé par la double nécessité d'avoir les noms et adresses complètes de tous les auteurs et de recenser toutes les activités de l'INRA.
- Après extraction des données brutes utiles, un travail considérable de normalisation (notamment de traitement des synonymies) a dû être réalisé au niveau des adresses, des noms de pays, de villes et d'organismes, travail pour lequel « Tétralogie » est particulièrement bien adapté.

2 LA NATURE DES PROBLEMES POSES

2.1 Etudier l'évolution des collaborations internationales

La démarche que nous avons utilisée pour les deux premières études et que nous avons reconduite pour cette actualisation se décompose en cinq points :

- Rechercher des bases consignant l'ensemble des adresses de co-signataires
- Retrouver la totalité des publications dans lesquelles est intervenu l'INRA depuis 1998
- Standardiser l'information: pays, villes, organismes, adresses, auteurs, journaux, ...
- Découper le temps en périodes significatives par rapport aux données disponibles
- Croiser les informations nécessaires à l'étude.

Les évolutions sur trois périodes de deux ans sont alors déduites de ces croisements : nombre de citations dans les bases retenues, nombre de publications nationales, nombre de publications internationales, nombre de co-signatures avec des pays étrangers, stratégies de coopérations, réorientations thématiques, stratégies de publication.

2.2 Analyser l'activité du département TPV

Ce département n'ayant pas de localisation propre, nous sommes partis de la liste de ses chercheurs pour établir un filtre permettant de récupérer, dans le corpus INRA, l'ensemble de sa production scientifique. Toutes les variations orthographiques des auteurs (plus de 550) ont été détectées et elles ont été synonymées à TPV. Un « retour aux notices » sur ce terme en mode multi-bases nous a permis d'extraire en une seule fois tout le corpus TPV. Celui-ci a ensuite été analysé en détail : détection de tous les acteurs, de leurs relations et des éléments sémantiques associés. Une analyse de l'évolution nous a ensuite permis de dégager, pour ce département, les composantes stratégiques et les signaux faibles.

2.3 Analyser l'influence de l'INRA aux Caraïbes

Pour extraire l'activité des laboratoires INRA des Caraïbes (Guadeloupe, Martinique, Guyane), nous avons travaillé dans un premier temps sur les adresses : recherche des ces trois noms et de tous leurs synonymes dans le corpus INRA. Comme certains organismes non INRA de ces départements travaillent avec l'INRA de métropole, nous avons ensuite éliminé manuellement une quinzaine de

publications où l'INRA local n'intervenait pas explicitement. Le corpus Caraïbe ainsi constitué a ensuite été analysé suivant les mêmes principes que celui de TPV. Mais l'influence de l'INRA s'exerce aux Caraïbes essentiellement depuis la métropole, nous avons donc fait un zoom sur les collaborations internationales de l'INRA avec les Pays de la région. Nous avons détecté les laboratoires concernés des deux cotés et les sujets de recherche abordés. Cet inventaire devrait permettre de promouvoir une meilleure coopération locale (entre INRA Caraïbe et pays de la région) puisque la métropole peut servir de pivot pour l'initialisation de nouveaux contacts.

2.4 Analyser le domaine de l'écologie

Cette analyse est beaucoup plus orientée évaluation de la recherche. Elle vise à connaître les pôles d'excellence en fonction de plusieurs critères *l'impact factor* des revues, le nombre de revues où publie un auteur, le nombre de collaborations internationales, la présence de signaux faibles.

3 DESCRIPTION DES BASES RETENUES

3.1 Base Current-Contents sur le Web of Science

Principales caractéristiques de cette base :

- Deux formats avant et après janvier 98.
- Plusieurs adresses depuis janvier 98.
- En avance sur l'indexation (articles sous presse)
- Propose un résumé et des mots-clés.

Format de l'adresse :

RP: de Lajudie, P; IRD; INRA,AGRO M,CIRAD; Campus Baillarguet,BP 5035; F-34032 Montpellier; France

IN: IRD, INRA,AGRO M,CIRAD, Lab Symbioses Trop & Mediterraneennes, F-34032 Montpellier, France; State Univ Ghent, Microbiol Lab, Ghent, Belgium; IRD, Lab Microbiol Sols, Dakar, Senegal

Périodes télé-déchargées et retenues : de 1998 à septembre 2004.

Certaines publications parues en 96 et 97 sont indexées sur ces CD/Rom.

3.2 Base Pascal

Principales caractéristiques de cette base :

- Deux formats avant et après janvier 96.
- Sciences exactes et appliquées.
- Mots-clés en français, anglais et espagnol.
- Présence d'un résumé.

Format de l'adresse :

CS- GIE Cartisol, 20 rue Bachaumont, 75001 Paris, France^Unite associee a l'INRA Organisation et variabilite des genomes vegetaux, Universite Blaise Pascal, Campus des Cezeaux, 63177 Aubiere, France^INRA, Station d'amelioration des plantes et de pathologie vegetale, Domaine de Crouelle, 234 avenue du Brezet, 63039 Clermont-Ferrand|

Périodes télé-déchargées et retenues : CD de 1998 à juillet 2004.

Certaines publications parues en 96 et 97 sont indexées sur ces CD.

3.3 Base SCI sur le Web of Science

Principales caractéristiques de cette base :

- Plus de 3500 journaux
- Sciences exactes et appliquées
- Pas de descripteur ni de résumé sur le CD/Rom (étude précédente)
- Descripteurs et résumé disponibles sur le Web of science (réactualisation)
- Les citations n'ont pas été récupérées dans le corpus INRA (problèmes techniques à répétition lors du déchargement en ligne).

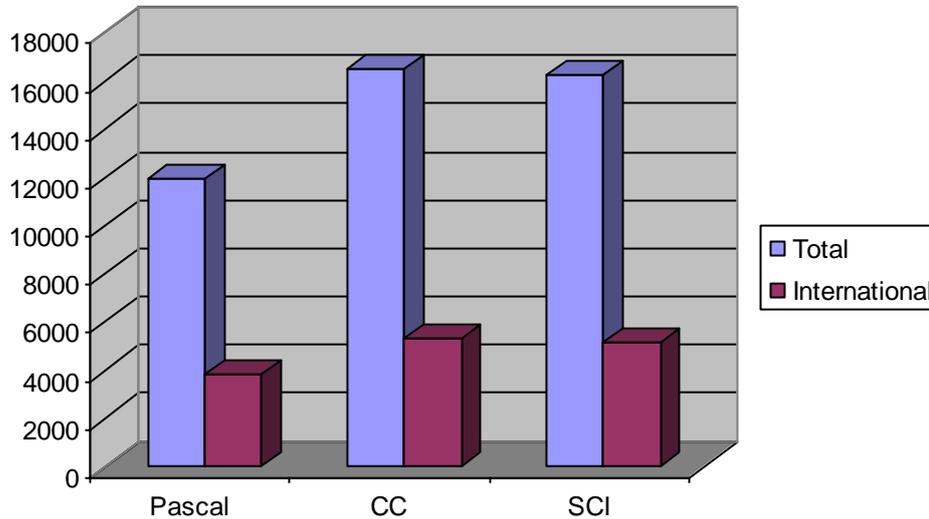
Format de l'adresse :

C1 Univ Munster, Dept Dermatol, Von Esmarchstr 56, D-48149

Munster, Germany
 INRA, CEA, Lab Radiobiol Appl, Jouy En Josas, France
 Périodes télé-déchargées: 1998 à fin juillet 2004.

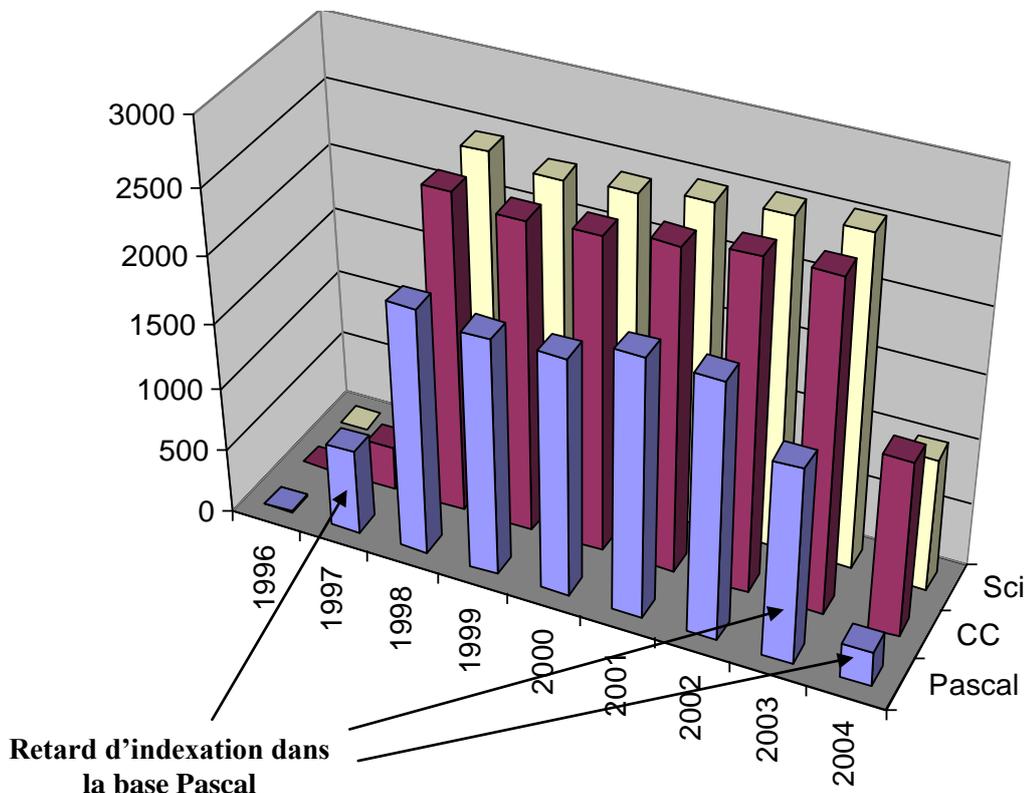
3.4 Part des publications internationales par base

Ce graphique montre bien la différence de contenu entre Pascal et les deux autres bases. Le petit **plus** de CC vient du fait qu'elle a été téléchargée avec un décalage de 2 mois par rapport à SCI. Par contre, la proportion de publications internationales est exactement la même dans les 3 bases : 32% pour l'ensemble de la période étudiée.



3.5 Répartition par bases et par années

L'année 2003 dépasse 2002 pour SCI et CC, par contre pour Pascal, le retard d'indexation est tel, qu'un déficit apparaît nettement pour 2003. Par contre coup, on retrouve un excédent en début de période (rattrapage de 1997).

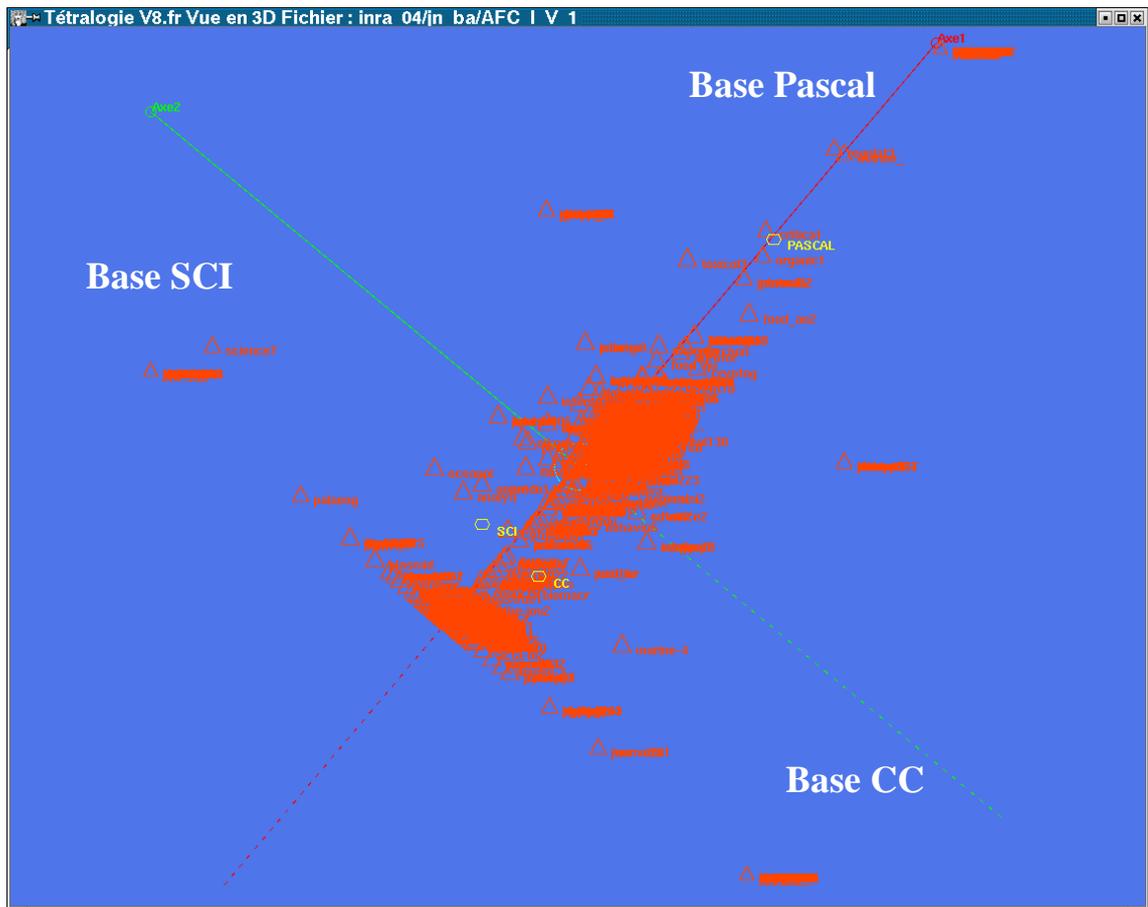


3.6 Répartition des journaux dans les 3 bases

Afin d'étayer notre hypothèse sur la complémentarité des 3 bases retenues pour couvrir au mieux la production scientifique de l'INRA, nous avons effectué une AFC (analyse factorielle des correspondances) sur la matrice de contingence croisant les journaux et les périodes.

Il ressort de cette analyse plusieurs éléments significatifs :

- De très nombreux journaux ne sont indexés que par une seule base,
- Certains journaux (indexés 2 ou 3 fois) sont mieux indexés par une des bases,
- SCI n'indexe donc pas complètement tous les journaux qu'elle prend en compte,
- Une majorité des journaux est indexée par les trois bases.
- Les apports respectifs de CC, Pascal et SCI sont respectivement de 101, 284 et 38 journaux originaux (donc Pascal est incontournable pour bien étudier l'INRA).
- Par contre, c'est SCI qui indexe le plus de journaux dans l'absolu.



- Début de la liste des journaux qui ne sont indexés que par CC :

Il y a 101 éléments dans cette liste

research-policy

sociologie-du-travail

economics-letters

environmental-and-resource-economics

applied-economics

revue economique

international-economic-review

jasss-the-journal-of-artificial-societies-and-social-simulation

applied-economics-letters

homme-

environment-and-planning-a

journal-of-productivity-analysis
journal-of-institutional-and-theoretical-economics-zeitschrift-fur-die-gesamte-staatswissenschaft
journal-of-public-economics
journal of risk and uncertainty

- Début de la liste des journaux qui ne sont indexés que par PASCAL :

Il y a 284 éléments dans cette liste

rencontres-recherches-ruminants
le-lait-lyon
phytoma-la-defense-des-vegetaux
plant-physiology-bethesda
fourrages-
revue-forestiere-francaise
les-dossiers-de-l'environnement-de-l'inra
les-colloques-inra
recents-progres-en-genie-des-procedes
microbiology-reading
ingenieries-antony
viandes-et-produits-carnes-aubiere
cahiers-de-nutrition-et-de-dietetique
industries-alimentaires-et-agricoles
revue-francaise-d'oenologie
courrier-de-l'environnement-de-l'inra
fruits-

- Début de la liste des journaux qui ne sont indexés que par SCI :

Il y a 38 éléments dans cette liste

journal of chromatography b-analytical technologies in the biomedical and life sciences
avian and poultry biology reviews
developmental neuroscience
journal-of-irrigation-and-drainage-engineering-asce
cellular and molecular neurobiology
immunity
gait & posture
journal of the royal statistical society series c-applied statistics
journal of mathematical imaging and vision
molecules and cells
physiological research
zoological journal of the linnean society

4 ANALYSE DES COLLABORATIONS INTERNATIONALES

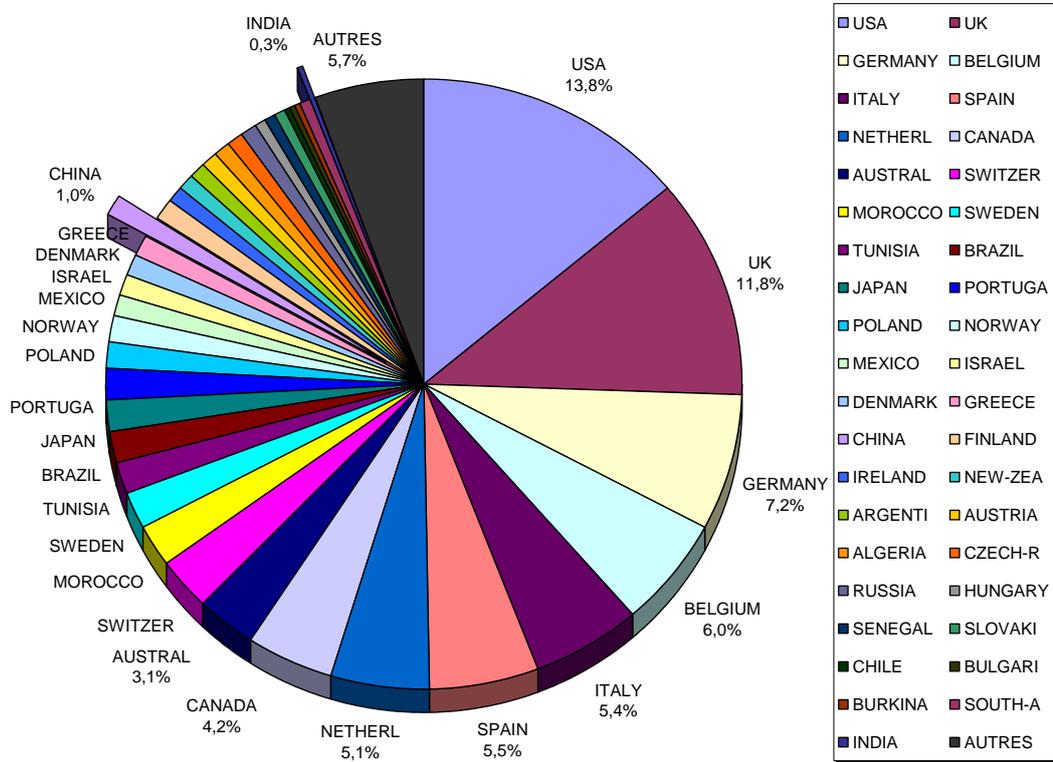
4.1 Avantages et inconvénients de l'analyse multi-bases

Cette analyse conduite simultanément sur les trois bases présente les caractéristiques suivantes :

- Réponse simultanée des trois bases.
- Etude de l'évolution la plus récente (avant 98, les formats diffèrent et pas d'accès à CC).
- Assez bonne homogénéité des quatre sous périodes : 1998-99, 2000-01 et 2002-04.
- Léger décalage de CC et SCI avec PASCAL (retard d'indexation pour PASCAL).
- La dernière sous période (02-04) a donc un déficit sur PASCAL.
- La dernière sous période (02-04) est sous représentée (indexation encore incomplète).
- Les valeurs brutes ne sont pas vraiment significatives.
- Il faut travailler si possible en évolution relative

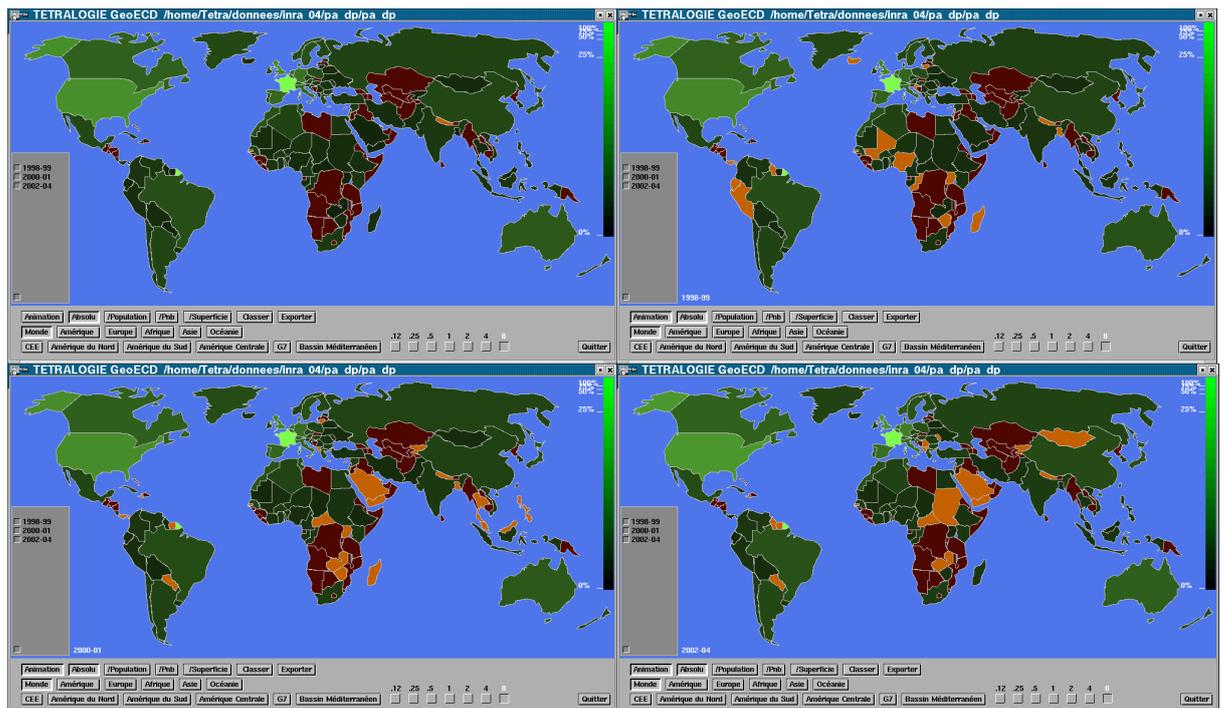
4.2 Répartition des pays pour l'ensemble du corpus

Remarque : les USA sont de peu en tête, d'où un potentiel peu utilisé pour les collaborations franco-américaines. Par contre, de petits pays (population et /ou surface agricole) sont très bien placés comme la Belgique, les Pays Bas et la Suisse. 70 pays (les moins cités) ne représentent que 6% du total.



4.3 Cartes cumulant puis comparant les trois périodes retenues

Nous utilisons, ici, le logiciel graphique GéoECD qui permet de générer automatiquement des cartes géographiques interactives à partir des tableaux de données produits par Tétralogie. Ces cartes sont entièrement manipulables par tout utilisateur local ou distant et permettent donc, dans des temps très courts, de communiquer des informations géostratégiques via le réseau avec la possibilité de les retravailler à distance.

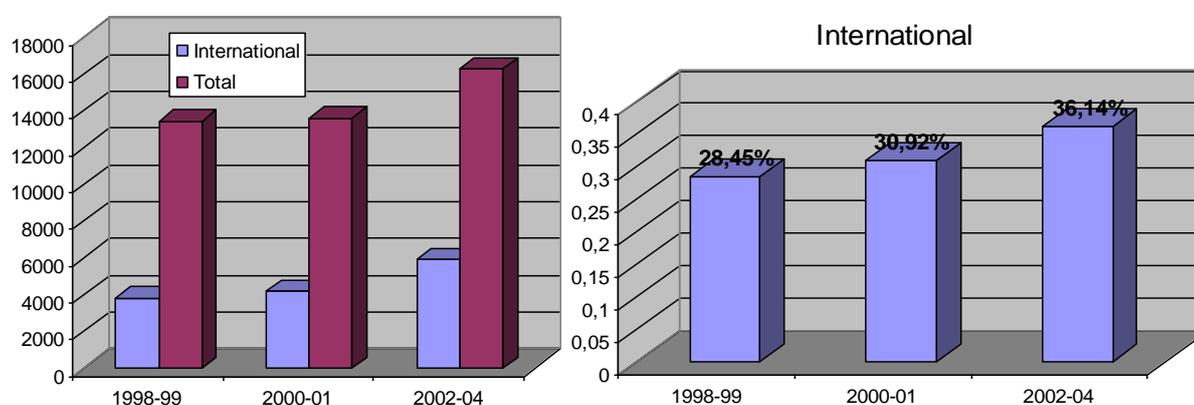


Il ressort de cette analyse que les collaborations internationales sont sporadiques avec les pays Africains et du moyen Orient et dans une moins grande mesure avec l'Amérique latine. La part des grands pays augmente : USA, Russie, Chine, Australie, Brésil, ... Les collaborations au sein de l'Europe restent très soutenues. Certains grands pays n'ont pas de collaboration avec l'INRA notamment en Afrique et au moyen Orient. Le sud-est asiatique vient d'être réintégré après une période de désaffection en 2000 et 2001.

4.4 Evolution des collaborations internationales de 1998 à 2004

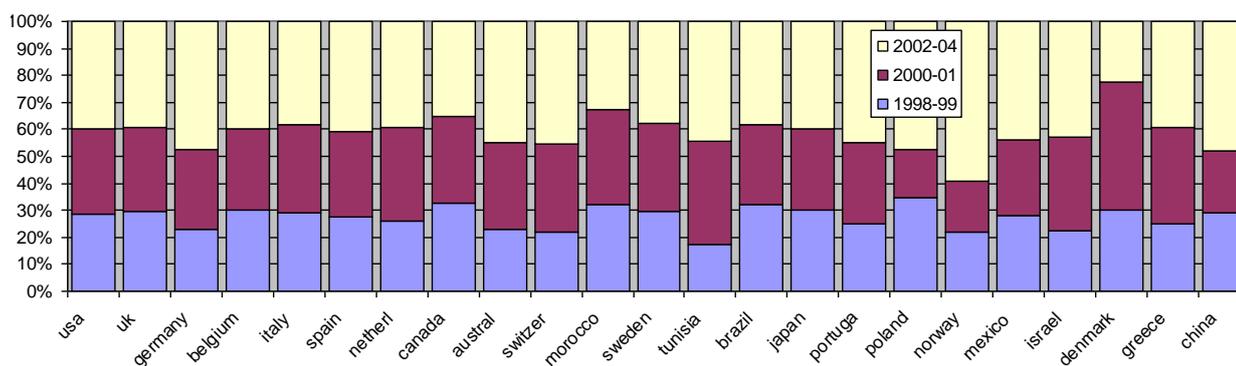
4.4.1 Répartition des références dans le temps

En valeur absolue le nombre des publications internationales augmente. Mais cette mesure n'est pas fiable car les chiffres de chaque période ne sont pas obtenus dans les mêmes conditions. Les années 1998 à 2001 sont normalement indexées à 100%, pour 2002 il subsiste un léger déficit d'indexation surtout pour Pascal, mais la période 2002-04 couvre plus de deux ans, en fait elle couvre virtuellement 30 mois avec un fort déficit d'indexation pour 2003 et une indexation très incomplète des 6 premiers mois de 2004. Par contre, **la part relative de publications internationales** est parfaitement exploitable et nous pouvons constater qu'elle **augmente régulièrement** en passant, en quatre ans (de début 1999 à début 2003), de **28,45%** à **36,14%**.



4.4.2 Evolution relative du nombre de publications par pays sur les 3 périodes

Ce comparatif essaie de mettre en évidence l'évolution dans les 24 pays les plus présents. Il peut être efficacement remplacé, pour la dernière période, par la carte en relatif ci dessous.



4.4.3 Evolution relative entre les périodes 1998-2001 et 2002-2004.

Nous avons calculé ici l'évolution du volume relatif des collaborations internationales par pays ramené à deux périodes (une de 4 ans : de 1998 à 2001, l'autre de 2 ans et demi : de 2002 aux 6 premiers mois de 2004). Nous ne visualisons que la part relative de la seconde période : au-dessus de 50% le pays voit sa part augmenter, sinon elle diminue. On peut ainsi immédiatement juger des différences de comportement de l'INRA entre ces deux périodes.

Codes des couleurs utilisées dans les cartes géographiques :

En **grenat** : pays absents sur l'ensemble des périodes.

En **orange** : pays perdus après 2001 mais présents avant.

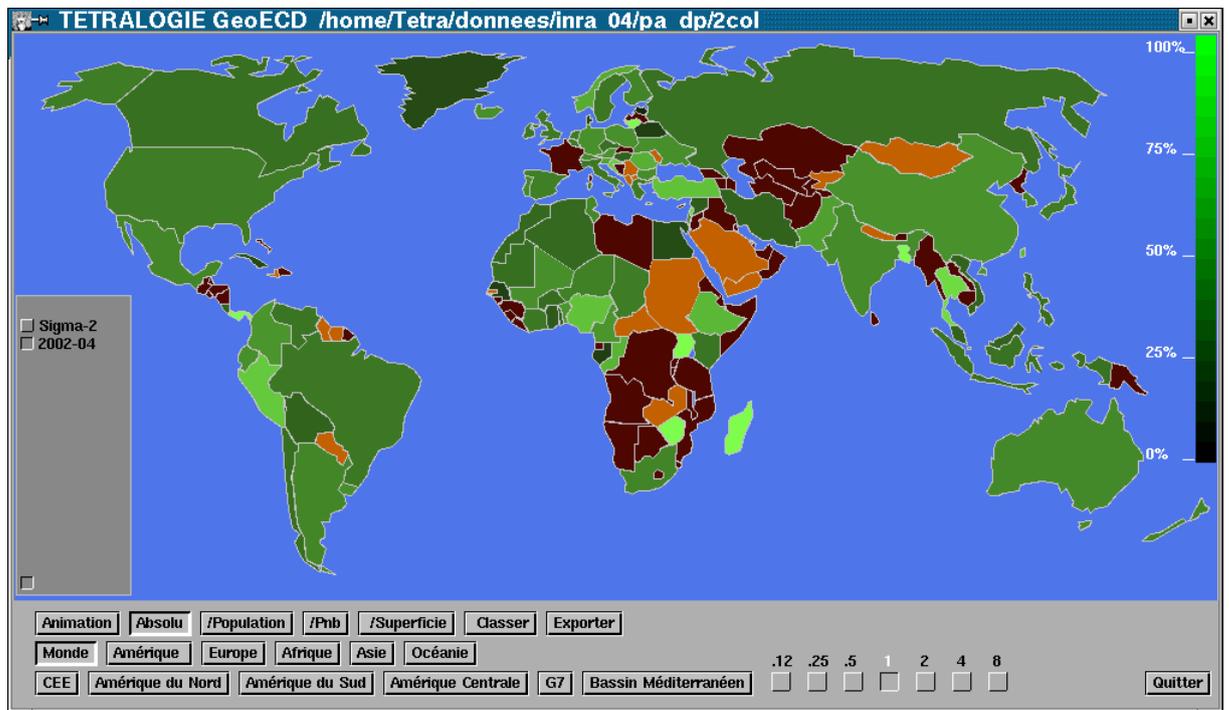
En **vert** part de la période 2002-04 sur la totalité.

Vert sombre diminution des collaborations

Vert moyen (50%) maintient

Vert clair augmentation des collaborations.

Dans l'exemple ci-dessous : Madagascar augmente, le Canada diminue.



4.5 Conclusions globales de cette nouvelle étude

Le nouveau système de cartes géographiques interactives propre à « Tétralogie » permet ainsi de formuler les principales conclusions suivantes :

Les publications internationales de l'INRA

- Se concentrent sur moins de pays,
- Leur nombre augmente,
- Leur proportion augmente (de 28 à 36%) par rapport aux publications franco-françaises. Plus même si l'on compare 1998 et 2003.

La qualité et le niveau des publications augmentent (publications acceptées dans les grandes revues référencées dans les bases).

Si nous comparons cette étude avec l'étude précédente (analyse 1998-2001), la période 2002-04 montre comme dynamique des collaborations internationales de l'INRA :

- Reprise en Afrique, mais absence de collaboration avec environ un tiers des pays africains.
- Déclin confirmé en Asie mineure et centrale,
- Reprise avec la Chine et le Japon,
- Forte reprise en Asie du sud-est,
- Emergence de certains pays d'Amérique latine,
- Emergence de nouveaux sujets de recherche à l'international.

CONCLUSION

Cette troisième étude complétée par trois zooms spécifiques a permis de montrer l'étendue des possibilités de la bibliométrie pour extraire d'un corpus conséquent (près de 45 000 références) des

informations stratégiques pour la conduite et le développement d'un grand institut de recherche comme l'INRA. L'identification précise des acteurs et la détection de toutes leurs relations permet de mieux comprendre le fonctionnement d'une telle entité. Le suivi de ses relations internationales dévoile les forces et les faiblesses en terme de collaboration et de coopération ainsi que des stratégies peu visibles autrement. Chaque département peut être analysé séparément notamment sur le plan sémantique (domaine plus ciblé que l'agronomie dans son ensemble), ses caractéristiques structurelles et son fonctionnement sont alors mieux compris, ses orientations mieux évaluées, sa lisibilité mesurée, son influence délimitée. Pour une région comme celle des Caraïbes, le but d'une telle étude est de détecter tous les partenaires locaux potentiels, les contacts qu'ils ont déjà avec l'INRA et les domaines dans lesquels ils collaborent. Une extension naturelle de cette étude est la mise en ligne de l'analyse du corpus INRA sur un portail (Internet ou Intranet). Les dictionnaires et les matrices sont alors compilés dans une base de données afin de permettre à un utilisateur de faire sa propre analyse (micro analyse) en scrutant son environnement (acteurs et thèmes de recherche). Des statistiques en ligne doivent l'aider à appréhender les éléments stratégiques qui le concernent, à se situer dans le contexte national et international et à détecter à temps toute innovation pouvant l'intéresser. La macro analyse, que nous venons de produire, lui servant de référentiel tout au long sa navigation et de ses investigations.

BIBLIOGRAPHIE

J. Mothe, C. Chrisment, T. Dkaki, B. Dousset, D. Egret

"Information mining: use of the document dimensions to analyse interactively a document set." " 23rd BCS European Colloquium on IR Research: ECIR, Darmstadt. BCS IRSG, pp 66-77, 4-6 avril 2001.

J.-L. Multon, G. Lacombe, B. Dousset

"Analyse bibliométrique des collaborations internationales de l'INRA". Veille stratégique, scientifique et technologique : VSST'01, Vol 1, pp 261-270, (Barcelone, Espagne), octobre 2001.

D. Sosson, M. Vassard, B. Dousset

"Portail pour la navigation en ligne dans les analyses stratégiques." " Veille stratégique, scientifique et technologique : VSST'01, Vol 1, pp 347-358, (Barcelone, Espagne), octobre 2001.

B. Dousset, S. Karouach

"Collaboration interactive entre classifications et cartes thématiques ou géographiques." " 9^{èmes} rencontres de la société francophone de classification, (Toulouse France), 16-18 septembre 2002.

J.-L. Multon, G. Lacombe, B. Dousset

"Analyse bibliométrique des collaborations internationales de l'INRA". 9^{èmes} journées d'études sur les systèmes d'information élaborée: Bibliométrie - Informatique stratégique - Veille technologique, (Ile Rousee Corse France), CD-ROM, 14-18 octobre 2002.

J. Mothe, C. Chrisment, B. Dousset, S. Karouach,

"Représentation des documents textuels : étude d'un domaine à travers des publications". 5^{ème} Congrès de la société française de recherche opérationnelle et d'aide à la décision ROADEF. (Avignon France), pp 130-131, 26-28 février 2003.

S. Karouach, B. Dousset

"Les graphes comme représentation synthétique et naturelle de l'information relationnelle de grande taille". Workshop sur la recherche d'information : un nouveau passage à l'échelle, associé à INFORSID'2003, (Nancy France), 3-6 juin 2003.

J. Mothe, C. Chrisment, B. Dousset, J. Alaux

"DocCube : multi-dimensional visualization and exploration of large document sets". Journal of the American Society for Information Science and Technology JASIST, Special topic section: "Web Retrieval and Mining". Guest Editor: Hsinchun Chen, 54(7), pp 650-659, March 2003.

S. Karouach, B. Dousset

"Visualisation de relations par des graphes interactifs de grande taille". ISDM Journal of Information Sciences of Decision Making, vol. 6, Mars 2003.

S. Karouach, B. Dousset

"Analyse d'information relationnelle par des graphes interactifs de grandes tailles". 4^{èmes} journées d'EGC (Extraction et Gestion de Connaissances), Clermont Ferrand, 20-23 janvier 2004.

C. Chrisment, B. Dousset, S. Karouach, J. Mothe

"Information mining : extracting, exploring and visualising geo-referenced information". Workshop on Geographic Information Retrieval, SIGIR 2004.