



20964-I RM

Agence de l'Eau
Rhin-Meuse

**IDENTIFICATION des ECOREGIONS
HYDROLOGIQUEMENT HOMOGENES
du TERRITOIRE NATIONAL**

*Etude Inter-Agences
Agence de l'Eau Rhin-Meuse*

Dernière phase :

**DELIMITATION MULTICRITERES DES ECOREGIONS
VOLUME I.**



CENTRE D'ECOLOGIE DES SYSTEMES FLUVIAUX DU C.N.R.S.

29, RUE JEANNE MARVIG,

31055 TOULOUSE CEDEX

ERIC TABACCHI, DECEMBRE 1995.

Sommaire



1. Avant-propos	5
2. Concepts	6
2.1. <i>Bref historique</i>	5
2.2. <i>Définitions et avertissements</i>	6
3. Méthodologie générale	7
4. Choix des plans d'information	8
5. Choix de la maille de travail et conséquences sur la démarche	9
6. Réalisation et codage des plans d'information	10
7. Etapes du travail	11
8. Modèles testés	12
8.1. <i>Introduction</i>	12
8.2. <i>Modèle 1. : Partition par arbres de régression</i>	12
8.2.1. <i>Concepts et principes</i>	12
8.2.2. <i>Critères de qualité du modèle</i>	14
8.3. Modèle 2. : Classification hiérarchique ascendante	14
8.3.1. <i>Concepts et principes</i>	14
8.3.2. <i>Critères de qualité du modèle</i>	15
9. Résultats	15
9.1. Sélection du modèle optimal	15
9.2. Hiérarchie d'assemblage des régions élémentaires	15
9.3. Niveaux finaux du découpage de la hiérarchie.	16
9.4. Comparaison avec Dupias et Rey	17
9.5. Hiérarchie d'intervention des classes de facteurs	17
9.6. Critique des résultats	18
10. Mise en place d'une base " Ecorégions "	18
10.1. <i>Principes</i>	18
10.2. <i>Utilisation (suggestions)</i>	19
10.3. <i>Conclusions</i>	20
11. Références	21

1. Avant-propos

En décembre 1991, le Centre d'Ecologie des Ressources Renouvelables du C.N.R.S. contractait avec le groupe Inter-Agences (sous la **responsabilité** de l'Agence Rhin-Meuse) une **opération intitulée "Définition des écorégions hydrologiquement homogènes du Territoire National"**. Ce découpage avait pour objet de mettre en **évidence** les **discontinuités** majeures **traduisant** les changements **simultanés** de plusieurs caractéristiques environnementales, susceptibles de contrôler le fonctionnement **écologique** des bassins **versants**. Le travail réalisé dans ce même laboratoire par G. Dupias et P. Rey en 1985 nous a assuré des fondements essentiels.

Ce premier volume du rapport **final** est l'aboutissement d'une **opération** technique **stimulante** qui nous a occupés pendant quatre **années**. Ce **travail** n'aurait pu aboutir sans le soutien **permanent** de l'agence pilote (Agence de l'Eau Rhin-Meuse), et en particulier sans le soutien de Jean-Luc Saleron et de Jean-François Zumstein. Notre partenaire local, l'Agence de l'Eau Adour-Garonne, nous a **également** prêté un secours important dans le transfert d'information entre les plates-formes **APIC™** et **ArcINFO™** (J.M Clotes) et dans la critique de notre maquette **Adour-Garonne** (F. Simonet).

James Sedell et Frederick Swanson (Forestry Sciences **Laboratory**, Oregon State University, Corvallis) nous ont fourni tous les moyens, techniques, conceptuels et humains, **afin** que nous soyons au fait des derniers développements en **matière** d'application des **écorégions** aux systèmes fluviaux du Nord-Ouest des Etats-Unis. Jim Omernik et Robert Hughes nous ont **également** accueilli sans **réserve** dans leur laboratoire de l'**Environmental Protection Agency** à Corvallis. Nous avons eu de **fructueuses** conversations sur les Ecorégions.

Sous la responsabilité de M. Henri **Décamps**, directeur du Centre, les tâches de ce travail ont **été** réparties comme il suit:

- **Documentation** : M. Izard et G. Guiraud.
- Elaboration du **modèle SIG** : E. Tabacchi et I-L Casanova
- **Pré-étude** (Département des Pyrénées Atlantiques) : J. Thibault
- **Maquette Adour-Garonne** : E. Tabacchi.
- **Réalisation des plans "climat"** : M. Izard, Cl. Druilhet, M. Lacassin, G. Guiraud.
- **Réalisation des plans "substrat"** : Cl. Druilhet et E. Tabacchi.
- **Réalisation du plan "végétation"** : J. Sajus
- Codage de la base **"agriculture"** : A.M. Planty-Tabacchi et E. Tabacchi.
- **Gestion du SIG et algorithmes d'extraction des données** : II. Casanova
- **Digitalisation des plans factoriels** : Cl. Druilhet.
- Validation des plans : H. Casanova et Cl. Dmilbet.
- Mise **en place** et tests des **modèles** statistiques : E. Tabacchi et AM Planty-Tabacchi.
- **Rédaction des rapports** : E. Tabacchi

2. Concepts

2.1. Brehistorique

Un des problèmes majeurs de l'aménagement du territoire réside dans l'intégration spatiale des principaux gradients environnementaux dans les schémas **opérationnels**. Pour la gestion globale de l'environnement, l'optique de la seule **géographie** physique ne peut **intégrer** de telles dimensions, dont beaucoup relèvent de l'écologie. C'est cependant à partir des disciplines proches de la géographie physique (climatologie, par exemple) qu'ont été dérivées les **premières** approches **écorégionales**.

C'est à l'échelle planétaire qu'a d'abord **été** abordé le problème. Parmi les pionniers en la matière, Herbertson (1905) a pris, sans doute le premier, conscience de l'importance d'intégrer les impacts humains dans la définition des "régions naturelles" du Globe. Aux Etats-Unis, cette démarche a connu un essor rapide à partir des travaux de Bayley (1976), puis d'**Omernik et al.** (1982). Ces travaux sont les fondements de diverses bases de données (Loveland *et al.*, 1991 ; Johnson *et al.*, 1991 ; Clarke *et al.*, 1994) visant à i) implémenter le fond documentaire existant et ii) augmenter la résolution de l'étude. Actuellement, de nombreux projets orientés vers l'aménagement des rivières sont en cours aux Etats-Unis (*cf.* Clarke *et al.*, 1994). Dérivés des grandes écorégions des USA au 1 : 7 500 000^{ème} et 1 : 2 500 000^{ème} (Omernik, 1987), ces zonages éco-régionaux sont établis aux échelles du paysage (1 : 100 000^{ème}), voire supérieures. Les écorégions issues de ces travaux relèvent invariablement de deux types de facteurs dominants : substrat et végétation.

En France, la démarche des écorégions s'est **confirmée** par des classifications biogéographiques du climat (*cf.* Emberger, 1954) et de la végétation (Gausson 1954) dont l'application a été envisagée au niveau mondial. La mise en place de la carte de la végétation de la France au 200 000^{ème} éditée par le CNRS à partir de 1947 a révélé la nécessité d'associer au thème majeur des facteurs explicatifs, tels le **relief**, le sol, le climat ou l'agriculture. Ces facteurs ont été intégrés à la carte de végétation sous forme de cartons au 1 : 2 500 000^{ème}. En 1985, le document rédigé par **Dupias** et Rey sous le titre "Régions phytogéographiques de la France" intègre en fait bien plus qu'une zonation fondée sur la seule végétation. Y sont inclus d'autres facteurs, tels le type d'occupation du sol, la géologie ou la pédologie, qui ne sont malheureusement pas explicités en détail dans le document. Cette carte résulte en fait de la grande connaissance du terrain par les auteurs, et fournit un découpage pertinent sur l'ensemble du territoire, mais non renseigné. En 1990, Gallant et al. ont proposé un découpage écorégional du bassin de la Seine. En 1991, le groupe Inter-Agences prenait contact avec le Centre **d'Ecologie** des Ressources Renouvelables du C.N.R.S. (devenu depuis le Centre **d'Etude** des Systèmes Fluviaux, puis le Centre **d'Etude** des Systèmes Aquatiques et Continentaux) afin de mettre en place la présente opération. Durant ce projet, de nombreuses études locales (e.g. **Wasson et al.**, 1993) ou intéressant la Communauté Européenne (e.g. programme **CORINE** landcover) ont été effectuées. A notre connaissance, aucun de ces travaux ne réalise une synthèse multi-thèmes à l'échelle nationale.

2.2. Définitions et avertissements

Le terme **d'écoringion** (ou *éco-zone*, ou *région écologique*) représente **actuellement** une source non **négligeable** de polémiques, qui **résultent** surtout des multiples disciplines **impliquées** dans ce concept, ainsi que des nombreuses échelles de travail adoptées. **Il** semble **évident** que la terminologie doit être **fixée** le plus rapidement **possible** (G. Long, *comm. pers.*). **Entrer dans** le débat sémantique n'est pas l'objet de ce travail. Je tiens **néanmoins à** présenter notre point de vue sur la question' afin de **délimiter** clairement notre approche. Si l'on s'en tient au strict vocabulaire' le terme "**écoringion**" désigne une région englobant des **systemes écologiques homogènes**, autrement dit deux notions mal **définies**. Par système écologique, on peut en effet entendre un biome, un écosystème, une communauté d'organismes, un habitat, une population etc.. . **Il** est évident que le degré d'homogénéité est aussi une notion relative. C'est pourquoi les écorégions embrassent des objets dont la **définition** relève **avant tout de la problématique sous-jacente**.

Notre démarche s'appuie d'abord sur une échelle de travail, qui a été **définie** en fonction du document de départ (Dupias et Rey, 1985), des **données** disponibles au niveau national et enfin de notre capacité à les **synthétiser** dans un laps de temps restreint. Cette échelle du un millionième limite, du moins en pratique, notre résolution. Du point de vue du géographe, on peut dire que les régions recherchées sont des "pays", des "terroirs". Du point de vue de l'écologiste, nous nous plaçons à un niveau supérieur à celui de **l'écosystème**, et même supérieur à celui du paysage. Notre objectif n'est pas d'identifier des objets physiques composant le paysage, mais plutôt **des conjonctions de variables environnementales localisées** caractérisant une zone spatiale donnée. Selon une approche des plus logiques, une région se distinguera de sa voisine par des valeurs différentes de ces variables. Autrement dit, la **variance** inter-régions sera, sauf rares exceptions, plus importante que la **variance** intra-région. Un tel découpage implique quelques remarques importantes.

La **délimitation des** régions constitue le problème méthodologique majeur. Classiquement, l'observateur se trouve en présence de deux types de variables : les variables discrètes (types de substrat ou occupation du sol) et les variables continues représentées par des gradients (climat, relief). Les premières définissent des frontières précises, alors que les secondes obligent l'observateur à découper les variables en classes. Par simple superposition' la délimitation de nouvelles frontières **à partir des frontières** existantes pose de délicats problèmes géométriques et thématiques. La plupart des auteurs proposent des zones de transition, qui ne résolvent pas le problème pour autant, et ont elles-mêmes des **frontières** imprécises. Ce problème est amoindri d'un point de vue strictement géométrique par l'adoption d'une maille polygonale unitaire supportant l'ensemble des données. C'est la technique que nous avons **adoptée**. Cependant, **la frontière d'une écorégion n'a pas de réalité physique**, à de rares exceptions près. **Il** serait de ce fait dangereux de considérer une telle limite au pied de la lettre.

A l'échelle du document sont liées la résolution géométrique et la résolution thématique. Si l'on peut envisager de subdiviser une **écoringion** en sous-unités (un paysage en écosystèmes, un écosystème en habitat, etc...) par un travail réalisé **à** une échelle plus fine, **il**

est illusoire de rechercher l'adéquation spatiale entre des structures mises en évidence à des échelles différentes.

Enfin, si notre démarche vise à définir des régions “hydrologiquement homogènes”, l'échelle de travail qui est la notre et les facteurs intégrateurs pris en compte n'impliquent nullement la reconnaissance de structures physiographiques définies par ailleurs (bassins versants ou secteurs de cours d'eau). Qu'une éco-zone puisse intégrer une ligne de crête sans que celle-ci constitue sa frontière n'est, *a priori*, pas choquant.

En outre, de notre point de vue comme de celui d'autres auteurs (cf. Omemik, 1982), les écorégions sont avant tout **un modèle géographique**, qu'il convient de valider ou d'infirmier en fonction de problématiques précises. En effet, le découpage proposé ne part d'aucune hypothèse de travail sinon du choix des thèmes impliqués. Il correspond donc à des coïncidences de facteurs, sans modèle biologique sous-jacent. Il serait par conséquent **injustifié d'attendre d'un tel modèle qu'il réponde à l'ensemble des attentes des gestionnaires du territoire**. Ceci étant posé, le fait que les régions ainsi définies ne correspondent pas à des entités physiographiques ne doit donc pas étonner.

Pour résumer notre approche, le travail proposé dans ce projet doit clairement être défini comme **un modèle à tester**. Nous avons inclus dans ce modèle les dix facteurs qui nous semblent les plus importants. Cependant, si ce premier volume fait l'objet d'une synthèse simultanée sur ces dix facteurs, il nous a semblé indispensable de fournir des informations moins synthétiques, incluant un jeu limité de facteurs intéressant un seul thème. Ceci fera l'objet du second volume.

3. Méthodologie générale

Notre objectif est de fournir un découpage du territoire national au 1. 000 000ème, reflétant la variabilité “écologique” régionale, et de quantifier la définition de chaque région au travers de variables susceptibles d'influencer l'organisation des systèmes fluviaux (bassins versants ou rivières). Le projet s'articule autour d'un Système d'Information Géographique construit à partir du progiciel orienté objets APIC™.

La démarche suivie au cours du projet se décompose en trois phases essentielles :

- **Phase 1. : Choix des facteurs de classification.** Ces facteurs, susceptibles de contrôler ou d'affecter la qualité des eaux au niveau régional ou le comportement des bassins versants, ont été sélectionnés selon un critère d'homogénéité sur l'ensemble du territoire national. Notamment nous nous sommes appliqués à éliminer les données trop fragmentaires ou incompatibles avec l'échelle de travail du 1 : 1000 000^{ème}.
- **Phase 2. : Choix de la maille de travail.** La seconde phase a consisté à tester l'effet du maillage sur la discrimination des régions. Une pré-étude de classification manuelle a été réalisée d'abord au niveau d'un département à partir de six facteurs

(Thibault 1993). La **pertinence du maillage** communal a ensuite été testée sur le Bassin Adour-Garonne pour l'ensemble des facteurs retenus. Une comparaison avec les Régions phytogéographiques de Dupias et Rey a **été réalisée**.

- **Phase 3. : Choix d'un modèle optimal.** La dernière phase du travail a **consisté en** l'application des facteurs et du maillage retenus **à** l'ensemble du territoire national. A l'issue de la seconde phase, nous avons **été** conduits à **reconsidérer** la **démarche** initiale, et notamment les **méthodes** statistiques **à** appliquer. Deux types de modèles ont été testés lors de la **classification** des polygones élémentaires en fonction de l'ensemble des facteurs retenus. Il a également **été** envisagé de réaliser des **classifications** plus thématiques, en utilisant des facteurs isolés ou regroupés en ensembles réduits.

4. Choix des plans d'information

Ont été éliminées les données **fragmentaires** au niveau national, les données dont la référence spatiale est trop ponctuelle, ainsi que les données sujettes à des variations temporelles trop importantes. Dans un premier temps, les plans d'information peuvent être regroupés en deux **sous-ensembles**.

Le premier constitue le support de l'information' et ne participe pas à la classification en tant que variable explicative. Il s'agit du "fond de carte", constitué des contours administratifs du territoire national et du réseau hydrographique. Ces documents sont les fonds de l'Institut Géographique National au 1: 1 000 000^{ème}, actualisés au format **ArcInfo™** par l'Atelier de Cartogrammie du Ministère de **l'Environnement**. Ce fond comprend les frontières nationales et communales et la majeure partie du réseau hydrographique. Le plan des régions phytogéographiques de Dupias & Rey (1985) est également inclus dans cet ensemble. Son intégration au S.I.G. devait à l'origine fournir une référence de découpage pour comparaison.

Le second ensemble constitue les plans d'information actifs (facteurs). Il comprend :

- **Les facteurs liés au substrat.**
 - **Hypsométrie (HPS).** Le plan du relief a été obtenu à partir de la carte 1. 000. 000^{ème} de l'IGN. Ce document s'est avéré plus précis que sa version numérisée au 500 000^{ème}, en particulier entre les isohypses 0 et 500 m.
 - **Pédologie (PEDO).** La carte au 1 000 000^{ème} de Dupuis (1967) a **été** retenue comme base de travail.
 - **Hydrolithologie (I-ILS).** La réalisation de ce plan s'est appuyée sur l'Atlas des Eaux souterraines de la France (**DATAR-BRGM, 1970**), sur les cartes géologiques du BRGM à diverses échelles (80 000^{ème}, 320 000^{ème} et 1 000 000^{ème}) et sur la **classification** proposée par Zumstein *et al.* (1989).
- **Les facteurs liés au climat.**
 - **Informations** thermométriques.
 - **Température moyenne de Janvier (TMJ).** Le plan est construit sur la base de l'Atlas de France (Comité National de Géographie, **1936-1958**), et **réactualisé** d'après Garnier (1964), et les données de la Météorologie Nationale (**Météo** France, période 1951-1993).

- **Température** moyenne de **Juillet (TMJT)**. La démarche est identique à celle adoptée pour les **températures** de Janvier.
- **Informations pluviométriques.**
 - Volume des précipitations annuelles (PLU). La base cartographique est celle de **Sanson (1945)**, qui s'est **avérée** plus **précise** que celle fournie par le modèle **Aurelhy (Direction** de la Météorologie Nationale, 1988). Les données ont été actualisées sur la base des normales longue **période** (1951-1993).
 - Durée des précipitations **> à 1 mm (NJP)**. Ce critère a été jugé plus fiable que son équivalent pour **P > 0,1 mm**. Le plan est **établi** sur la base de la Carte des Précipitations en France au **2 000 000^{ème}** (Météorologie Nationale, **1988**), et **réactualisée à** partir des valeurs 1951-1993.
 - Régimes pluviométriques (**REGPL**). La typologie est **réalisée de novo à** partir des données longue période de **France** et de diverses publications locales (Choisnel 1986, Choisnel et al. 1987, Coutagne 193 5).
- Les facteurs **liés à l'occupation du sol.**
 - Types de végétation potentielle (**VEG**). **Une** synthèse de la Carte de la Végétation de la France au **200 000^{ème}** (**CNRS** 1947-92) constitue la base de ce document.
 - Types d'agriculture dominants (**AGR**). Ce plan s'appuie sur une synthèse communale de la base **AGRESTE (SCEES, 1989)**.

5. Choix de la maille de travail et conséquences sur la démarche

Un des moyens techniques permettant d'éviter une complexité excessive du traitement du jeu de données spatiales consiste à discrétiser ces données sous forme d'unités élémentaires qui deviennent un support commun aux divers **plans** d'information. Ceci revient à recoder les plans d'origine (variables discrètes ou continues) par une opération booléenne simple (intersection) entre le plan d'information et une grille. Classiquement, cette grille est régulière et à maille carrée. Nous avons cependant adopté une démarche différente, en testant une grille irrégulière par la taille et la forme des ses mailles.

Compte tenu de l'importance numérique des polygones à analyser, la mise en place d'un maillage de **référence** s'est avérée nécessaire. Plusieurs solutions ont été envisagées (**cf.** rapport de seconde phase). Nous nous sommes particulièrement attachés **à** analyser le biais induit par un maillage communal, qui présentait l'avantage d'être directement compatible avec les bases administratives regroupant les données techniques (**Insee, Scees, Agences de l'Eau**). Un premier travail (Thibault, 1993) a permis de préciser les modalités d'application d'un tel maillage, **à** partir de synthèses manuelles au **500 000^{ème}**. Les résultats obtenus nous ont encouragés **à** appliquer ce maillage à l'analyse au **1 000 000^{ème}** du Bassin Adour-Garonne (**cf.** rapport de seconde phase et base informatique associée). Ce bassin a été choisi en raison de la grande hétérogénéité de son territoire, qui le désignait naturellement comme zone atelier. La **classification** des **8000 communes** du Bassin **à** partir d'une AFCM réalisée sur 10 facteurs a abouti à des résultats très similaires à ceux **établis** par **Dupias & Rey (1985)**. De ce fait, le maillage communal a été abandonné pour deux raisons :

- il augmente considérablement le temps de calcul. Nous avons **préfé**rer consacrer les ressources calcul à tester différents algorithmes.

- il donne des résultats peu **différents** de ceux obtenus par Dupias et Rey. Compte tenu de l'incertitude géométrique apportée par le découpage **communal**, il nous a semblé **préférable** de **conserver** le découpage le plus fin des **régions** phytogéographiques, qui semble donc robuste.

Ce test de maillage a radicalement réorienté notre démarche. Ses résultats ont **confirmé** la pertinence du découpage Dupias et Rey (**1985**), qui **était** pressentie au **cours** de ce projet et au **cours** de démarches similaires menées par d'autres auteurs (**Gallant** et al. 1990, **Wasson** et al. 1993). Cependant, la hiérarchisation des régions élémentaires de Dupias et Rey (régions d'ordre 3) ne semble pas claire (niveau 2) ou devient trop triviale (niveau 3). De ce fait nous avons choisi :

- de conserver le maillage phytogéographique Dupias & Rey le plus fin (niveau 3).

Celui-ci est apparu raisonnablement détaillé **compte** tenu de l'échelle de travail.

- de procéder à sa classification à partir des 10 facteurs retenus, **afin** d'en préciser la structure.

- de procéder à une première critique des **résultants**, en attendant de véritables tests du modèle sur des variables biologiques.

6. Réalisation et codage des plans d'information

Comme nous l'avons indiqué lors de leur définition, les plans factoriels (plans d'information élémentaires) ont dû **subir** des remaniements par rapport aux documents dont ils s'inspirent. Ces remaniements sont de deux ordres :

- **géométriques** : l'homogénéisation des cartes à l'échelle du 1 000 000^{ème} et la réactualisation des données a nécessité la création de nouveaux contours. Cette création manuelle intéresse tous les plans, à l'exception de l'hypsométrie et de l'agriculture. Cette dernière n'étant disponible que sous une forme discrétisée (niveau communal), nous avons **préfé**rer conserver le maillage initial.

- **thématiques** : la création des cartes manuelles s'est accompagnée d'un nouveau découpage des thèmes. Ce découpage a été réalisé sous une double contrainte : minimiser le nombre de classes et rendre **compte** des contrastes à l'échelle nationale. **Ainsi**, certaines classes ont été éliminées car trop localisées, certaines ont été fusionnées car exprimant des subtilités jugées inutiles, et d'autres encore ont **été** subdivisées afin de rendre compte des gradients existants.

Nous avons privilégié le détail thématique (nombre de classes important) pour les facteurs de type "**intégrateur**", comme l'agriculture et la végétation potentielle (**cf.** Tab. 1.). Au contraire, nous avons simplifié au maximum les facteurs reflétant les gradients majeurs, **comme** l'altitude ou la température. La notion de régimes de précipitations fait appel à des critères complexes, et il nous a paru important de conserver une résolution importante pour ce facteur. L'agriculture a **été codée** selon l'importance surfacique relative des rubriques majeures (cultures + types d'élevage) de la base SCEEES. Une simplification récursive a ensuite été réalisée afin de ne retenir que le nombre minimal

de **combinaisons**. De même, de nouvelles combinaisons de classes de **végétation** potentielle ont été mises en place afin de refléter les contrastes **inter-régionaux**. Enfin, à de rares exceptions près nous avons **conservé** la **nomenclature** et la géométrie de **Dupias & Rey** pour la désignation des polygones **à classifier**. Ceci permet de faciliter l'**interprétation** de la **classification**, ainsi que les **comparaisons** avec la hiérarchie à trois niveaux proposée par ces auteurs.

Tableau I. Caractéristiques des plans d'information.

Plan	Acronyme	Nombre d'éléments géométriques*	Nombre de classes
Hypsométrie	HPS	2317	11
Hydrologie	HLS	14%	20
Pédologie	PED	2715	19
Volume des précipitations	PLU	258	10
Durée des précipitations	NJP	142	14
Régimes des précipitations	RPL	98	38
Températures de Janvier	TJN	247	14
Températures de Juillet	TJT	172	11
Agriculture	AGR	-36569	51
Végétation potentielle	VEG	1476	52
Régions phytoécologiques	PHYCOS	859	660

* ces nombres correspondent à des éléments homogènes (objets APIC liés par une relation thématique ou géométrique).

La nomenclature de codage adoptée pour chacun des plans est fournie en annexe A1.

7. Etapes du travail

La troisième **phase** du travail s'appuie sur les étapes suivantes.

- **Chaque** plan manuel est numérisé sur un digitaliseur **OCE 68000** en mode "**fil de fer**".
- Chaque polygone est **surfacé** et codé individuellement. Seule la base du plan AGR est importée de façon automatique après recodage de la base AGRESTE du SCEES sur le tableur **Excel™**.
- Le codage et la pertinence géométrique (unicité, jointure, relations topologiques) sont validés par une routine **APIC™**.
- **Chaque** classe *i* de chaque plan *P* est croisée avec PHYCOS par une routine de **type Polygonal Processing**. Pour chaque région de PHYCOS (un ou plusieurs polygones), le pourcentage de la surface occupée par la classe *i* du plan *P* est **calculé**.
- La pertinence du calcul est validée (somme des pourcentages partiels).
- La matrice régions x classes (660 x 240) constitue la base d'application des modèles testés sur les logiciels de statistiques. Deux types de calculs sont réalisés pour **chaque** modèle testé :
 - un calcul global, où les 660 régions sont classées en fonction des 240 classes, pondérées ou non (présent rapport).
 - des calculs partiels, ne faisant intervenir que certaines classes (**cf.** rapport final, vol. 2) :

- l'ensemble des classes d'un même facteur.
- l'ensemble des classes d'un même **thème** (climat, substrat' températures...).
- Les modèles sont critiques et compares.
- L'interprétation finale des résultats a lieu **à** partir du modèle retenu.

8. Modèles testés

8.1. Introduction

Divers modèles d'analyses exploratoire et typologique d'un ensemble de données ont **été** proposes dans la littérature, à partir des travaux originaux de Fisher (1958). Deux d'entre eux sont apparus **comme** bien adaptés à la démarche de ce travail. Ces deux modèles sont très **différents** tant par leur mathématique que par les informations qu'ils procurent. Ils ont été testés en parallèle, **afin** de **déterminer** celui qui serait le mieux adapté à la mise en place des écorégions.

Le principe des deux méthodes retenues relève des typologies (ou **classifications**) hiérarchiques.

8.2. Modèle 1. : **Partition par arbres de régression**

8.2.1. Concepts et principes

La méthode des arbres de régression permet d'établir une partition d'un ensemble de données à partir de critères multiples, qualitatifs ou quantitatif... Il s'agit d'une méthode de partition descendante. Cette méthode a été appliquée dans sa version primitive à des données **socio-économiques** en 1964 (Sonquist et Morgan). Son évolution a suivi celle des systèmes experts AID (idem), **THAID** (Morgan et Messenger, 1973) et **CHAID** (**Kass**, 1980). Les algorithmes testés bénéficient des améliorations apportées successivement par Breiman *et al.* (1984), Ciampi *et al.* (1987), et **Hastie** et **Pregibon** (1990).

Le modèle est ajusté en utilisant la méthode **BRP (binary recursive partitionning)** grâce à laquelle les données sont **partitionnées** successivement sur les axes des **prédicteurs** de telle sorte que chaque noeud de la partition **permette** de maximiser les différences entre les deux **sous-ensembles** créés. La partition se poursuit jusqu'à obtention de noeuds purs (individus de l'ensemble à partitionner), ou jusqu'à ce que les données soient déterminées comme identiques. Le critère d'identité est appelé le paramètre "coût-complexité" (**k**). L'importance d'un sous-graphe (**T**) de l'arbre T est calculée par :

$$Dk(T) = D(T) + k \cdot \text{card}(T)$$

où :

D(T) est la **déviante** du sous-graphe

card(T) est son cardinal (nombre d'individus le composant).

La déviance est une fonction de la fonction log-seuil, qui correspond à la fonction de probabilité de masse $f(y_i ; \mu_i)$ pour l'observation y_i pour un μ_i donné. Pour un ensemble de n observations, la fonction log-seuil est définie par :

$$l(\mu; y) = \sum_{i=1}^n \ln (f(y_i ; \mu_i))$$

la déviance $D(y ; \mu)$ est alors :

$$\frac{D(y; \mu)}{\phi} = 2l(\mu^* ; y) - 2l(\mu ; y)$$

où μ^* maximise le log-seuil sur μ sans contraintes,

et ϕ est le paramètre de dispersion de la distribution.

Dans le cas d'une distribution Gaussienne, ϕ est la variance et la déviance se réduit à la somme résiduelle des carrés des écarts.

Le modèle proposé s'appuie sur une distribution **multinomiale**. Le vecteur $\mu = (\mu_1, \mu_2, \mu_3, \mu_4)$ tel que $\sum \mu_k = 1$ représente la probabilité qu'une variable dépendante soit prédite dans une des quatre classes possibles d'un **prédicteur** μ . Par exemple, si la variable y (ici, les polygones à hiérarchiser) répond au troisième niveau d'une variable test (la **probabilité** de présence d'une classe d'un plan d'information donné) sur quatre niveaux possibles, on peut noter cette distribution :

$$Y = (0, 0, 1, 0)$$

Le modèle **gaussien** s'écrit :

$$(\text{composante stochastique}) : y \sim N(\mu_i, d), i = 1, \dots, N$$

et

$$(\text{composante structurale}) : \mu_i = \tau(x_i)$$

La déviance s'écrivant alors :

$$D(\mu_i, y_i) = (y_i - \mu_i)^2$$

En un noeud donné du graphe, la déviance est définie comme la somme des déviances de toutes les déviances observées en aval du noeud :

$$D(\mu ; y) = \sum D(\mu ; y_i)$$

En ce noeud' on peut décomposer la déviance en deux éléments (branches droite et gauche de l'arbre) :

$$D(\mu_d, \mu_g ; y) = \sum_g D(\mu_g ; y_i) + \sum_d D(\mu_d ; y_i)$$

La maximisation de la **différence** ΔD à un noeud ("goodness of split")

$$\Delta D = D(\mu ; y) - D(\mu_g, \mu_d ; y)$$

devient le critère optimal permettant de choisir le noeud pour la classification.

Le calcul de tels modèles a été réalisé sur une station de travail Sun **Sparc™ 20/61** (système **Solaris™/Unix™**) à l'aide du langage de programmation orienté objets **S-plus™** (objets de classe "tree").

8.2.2. Critères de qualité du modèle

Le modèle de typologie par arbres de régression est optimal dans le sens où les noeuds du graphe sont **sélectionnés** selon une méthode probabiliste. La croissance du graphe établi peut **être** arrêtée en imposant un critère minimal de **déviante** (k , ci-dessus) ou un cardinal minimal aux groupes ($card(T)$). En outre, l'algorithme utilise permet de **définir** le seuil de la variable **responsable** au premier niveau d'une bifurcation. Par exemple, la bifurcation au noeud N° 127 du graphe permet de séparer deux sous-graphes (droit et gauche) à partir du critère : "altitude > ou < à 1700 m").

Nous avons retenu comme critères de qualité du modèle :

- la possibilité, -en faisant varier k , d'aboutir à un graphe complet (présentant les observations isolées **comme** noeuds terminaux).

- le niveau de symétrie de la partition (comparaisons à chaque niveau d e $\sum_g D(\mu_g; y_i)$ et de $\sum_d D(\mu_d; y_i)$).

8.3. Modèle 2. : Classification hiérarchique ascendante

8.3.1. Concepts et principes

La famille de modèles de classification hiérarchique ascendante fait désormais partie des outils classiques en analyse multivariée. Les observations initiales (ici les polygones élémentaires) sont agglomérées hiérarchiquement selon un critère de distance n-dimensionnelle (à partir de n variables), métrique ou non. Ceci impose le choix de deux critères de classification : un indice de distance et une méthode d'agglomération des groupes d'observations. La méthode d'agglomération correspond aux règles qui permettent de décider, à partir des valeurs de l'indice, de l'agrégation d'un élément (observation initiale ou groupe d'observations déjà formé) à un autre élément.

A chaque noeud de la hiérarchie se forme un nouveau groupe constitué de la réunion d'un *ainé* (groupe déjà formé) et d'un *benjamin* (nouvel élément du groupe); dans la hiérarchie, ce nouveau groupe est inclus dans un *père* de niveau supérieur. Le noeud est caractérisé par un *cardinal* (nombre d'éléments), une *masse* (inertie absolue du groupe), une *distance de création* (sur l'axe de l'indice), et par son *excentricité* (degré d'inertie par rapport aux autres groupes).

Les méthodes dites probabilistes sont fondées sur un modèle supposant que les **données** d'origine sont générées par une **combinaison** de lois de distributions de probabilités. Il est donc supposé que la population étudiée (ici, l'ensemble des polygones écorégionaux) est constituée de G sous-populations distinctes, et que la densité d'une observation x de la $k^{\text{ième}}$ sous-population est $f_k(x; \theta)$ pour un vecteur donné de paramètres θ . Pour une série d'observations $D = (x_1, \dots, x_n)$, soit $\gamma = (\gamma_1, \dots, \gamma_n)$ les **identifiants** des objets étudiés, avec $i = k$ si x_i appartient à la $k^{\text{ième}}$ sous-population. Dans la procédure du maximum de vraisemblance, θ et γ sont choisis de **façon** à maximiser la fonction de vraisemblance (likelihood) :

$$L(D; \theta, \gamma) = \prod_{i=1}^n f_{\gamma_i}(x_i; \theta)$$

Dans le cas où $f_k(\mathbf{x}; 0)$ est une distribution multivariée normale, de moyenne μ_k et de **variance** la matrice $\Sigma_k = \sigma^2 * I$, avec $I =$ matrice **identité**, alors la **maximisation** de la fonction de vraisemblance décrite ci-dessus revient à **minimiser la somme des carrés des écarts à la moyenne**, ce qui revient à appliquer la **méthode** de Ward (**Ward, 1963**). Dans le cas où les sous-graphes ne sont pas hypersphériques avec une même **variance** interne, la méthode de Ward tend à créer des ruptures hypersphériques dans l'ensemble d'origine.

Plusieurs indices et méthodes d'agglomération de la **littérature** (*cf. Mardia et al. 1979, Everitt 1980, Gordon 1981, Legendre et Legendre 1984, Banfield et Raftery 1993*) ont été testés. Les calculs ont été réalisés sur une plate-forme Sun **Sparc™ 20/61** (système **Solaris™/Unix™**) à l'aide du langage de programmation orienté objets **S-plus™** (objets de classes **hclust, mclust**) et sur une plate-forme Intel **Pentium™** (système MS-Windows **95™**) à l'aide des logiciels **Systat™** et **Statlab™**.

8.3.2. Critères de qualité du modèle

Les modèles de type 2 sont plus simples à mettre en oeuvre que les modèles de type 1. Cependant, ils imposent aussi des règles davantage empiriques lorsqu'il s'agit de limiter la parution à un niveau hiérarchique considéré comme optimal. De ce fait, le seul véritable critère de qualité du modèle est la pertinence thématique des groupes créés. **Globalement**, la qualité des méthodes a été évaluée i) par comparaison avec une méthode d'ordination en espace réduit (Analyse Factorielle des Correspondances Multiples), et ii) par comparaison avec les autres méthodes testées, (analyse des indices de **connexité**, Estabrook 1966, Day 1977, et discrimination des groupes; **Perruchet 1983**).

9. Résultats

9.1. Sélection du modèle optimal

Les tests effectués sur des plans isolés ou sur l'ensemble des données ont indiqué une tendance à une classification trop sévère et dissymétrique pour les modèles d'arbres de régression (modèles de type 1). Nous ne sommes pas parvenus à des classifications satisfaisantes sur l'ensemble des groupes de variables testés en forçant les valeurs du coefficient coût-complexité k . Les critères de qualité se sont révélés meilleurs pour les modèles de type 2 qui se sont avérés plus robustes vis-à-vis du biais lié à la taille des polygones et à la distribution des variables testées. Le modèle **finale**ment retenu est une méthode à liens hypersphériques (**Ward, 1963**) s'appuyant sur une distance du χ^2 , distribution sur laquelle reposait nos tests lors de la maquette de seconde phase. La pertinence statistique de découpage a été définie selon un critère **bayésien** (**Banfield et Raftery, 1993**).

9.2. Hiérarchie d'assemblage des régions élémentaires

Le détail de la hiérarchie complète est **fourni** dans le tableau **A2** en annexe. La **hiérarchie** est présentée de façon à obtenir, pour un groupe donné, l'origine de sa formation (aîné et benjamin) et sa situation par rapport au niveau hiérarchique supérieur (père). Le principe de notation est le suivant :

796 = **8C21** u 740 € 925

signifie :

- que le groupe n° 796 **résulte** de la capture de la région **8C21** (nomenclature **Dupias** et **Rey**, 1985) par le groupe n° 740 précédemment formé,
- et que le groupe n° 796 est une des sous parties du groupe n° 925.

La désignation "**root**" correspond à la racine de l'arborescente, noeud où tous les polygones sont agglomérés. Le tableau A2 (*cf.* annexes) donne, pour chaque noeud, ses aînés, benjamin et père, ainsi que l'effectif du groupe, son niveau de **création** sur l'axe des distances, et le pourcentage d'inertie exprimé par le groupe (**% de variance**). Bien que moins **facile** à lire qu'un dendrogramme, un tel tableau permet d'avoir accès précisément à l'ensemble de l'information disponible.

9.3. Niveaux finaux du découpage de la hiérarchie

Le graphe complet résultant de la **classification** des 660 par les 240 classes produit 13 19 groupes possibles de polygones (soit les 660 polygones d'origine et 659 groupements possibles). L'exploitation cartographique idéale d'un tel graphe, *qui est le seul exact*, est la production dynamique, niveau hiérarchique par niveau hiérarchique, de cartes d'agglomération. Si cela est aisé sur un support informatique de type S.I.G., l'exploitation de nos résultats via des supports papier nous contraint à réaliser un découpage arbitraire de l'arbre. En se reportant aux distributions des distances, des **variances** ou des effectifs, il est **difficile de définir** des ruptures isolant des niveaux évidents dans la structure de l'arborescence (*cf.* figure 2. en annexes).

Après analyse de la structure, nous avons opté pour un découpage du graphe en réalisant une progression géométrique d'ordre 2 dans le nombre de groupes. Cette **méthode** dichotomique permet une visualisation progressive de la hiérarchie sur les cartes. **Afin** de préserver une signification à ce découpage, nous avons finalement sélectionné les partitions **définissant** successivement 4, 8, 16, 32, 64, 128, 256 et **finalement** 660 régions sur le territoire national. On obtient ainsi huit niveaux hiérarchiques emboîtés. Pour des raisons de présentation, les documents cartographiques présentés dans ce rapport ne vont pas au-delà de 64 niveaux (*cf.* planches couleur en annexe).

Le codage des groupes est réalisé en faisant précéder le code de la région d'un niveau donné par le ou les codes des régions de niveau hiérarchique supérieur auxquelles elle appartient. Ainsi les régions de premier niveau étant codées arbitrairement de 1 à 4, les régions de niveau **inférieur** seront codées de 1 à 8 et notées : 1 + 1, 1 + 2, . . . **4+8**. Par exemple, le code **4+8** signifie que la région n° 8 du second niveau appartient à la région n°4 du premier niveau. La région codée **4+7+14+27+56** est la **56^{ème}** région du cinquième niveau de la partition. (64 régions au **total**). Cette région appartient :

à la **27^{ème}** région du quatrième niveau (32 régions).

à la **14^{ème}** région du troisième niveau (16 régions)

à la **7^{ème}** région du second niveau (8 régions).

à la **4^{ème}** région du premier niveau (4 régions).

9.4. Comparaison avec Dupias et Rey

Nous limiterons la comparaison de nos résultats avec ceux de Dupias & Rey (1985) à un exemple concernant le premier niveau hiérarchique défini par ces auteurs. Ce premier niveau comportait 14 grandes régions notées de 1 à XIV. Afin d'établir la comparaison sur des bases similaires, nous avons donc extrait 14 classes de notre hiérarchie (cf. planche couleur en annexe). Nous avons considéré le nombre de polygones trouvés dans cette classe qui appartiennent à une classe donnée de Dupias et Rey (cf. annexe A4.).

Un **coefficient** d'homogénéité interne de nos classes a été calculé sur la base :

$$H = - \left[\sum_c (p_i * \log(p_i)) \right] / \log(C)$$

où : p_i est la proportion de polygones de la classe C de D. & R. compris dans une classe donnée de la hiérarchie, et où C est le nombre total de classes D. & R rencontrées au sein de cette classe hiérarchique.

Les résultats indiquent que, pour un découpage grossier, l'adéquation Dupias & Rey avec ce travail est très variable (adéquation parfaite si $C=1$, ou si $H \rightarrow 100$). Ces résultats sont toutefois moins pessimistes si l'on considère non plus les effectifs, mais les surfaces. L'examen des cartes montre que les divergences résultent souvent de captures d'éléments limitrophes des écorégions. Cependant les dynamiques de classification divergent (les 14 niveaux de notre classification ne semblent pas correspondre au même niveau d'organisation que les grandes régions D. & R.).

9.5. Hiérarchie d'intervention des classes de facteurs

Le type de modèle utilisé ne permet pas de classer simultanément les observations et les variables. Un des moyens pour parvenir à l'interprétation des groupes est de réaliser des statistiques simples par groupe étudié (moyenne, écart-type, pondérées ou non par la surface unitaire des polygones) sur les données de base de chaque classe. Les figures x en annexe, à titre d'exemples, les distributions observées pour les dix facteurs au sein des huit régions retenues pour le découpage cartographique.

On peut également appliquer le modèle de **classification** sur la matrice des données transposée (les classes de facteurs sont alors classées par les polygones). Il en résulte une **classification** hiérarchique des classes de paramètres. Les groupes de classes sont définis par ordre décroissant **d'affinité** selon le tableau A5 en annexe, qui est analogue à celui présenté pour les régions. La figure 7. (annexes) illustre cette hiérarchie sous forme graphique.

9.6. Critique des **résultats**.

Nous ne pouvons tirer guère de conclusions pertinentes de la comparaison avec Dupias & Rey (1985). En effet, les **différences** observées ne permettent pas de valider ou **d'infirmes** l'une ou l'autre des **classifications**. Par contre, il semble **nécessaire**, à l'issue de cette première **étape**, de procéder à une critique sévère de notre **modèle** avant de le livrer à validation sur variables indépendantes.

D'une manière **générale**, nos résultats **offrent** une bonne **cohérence** spatiale, compte tenu de la nature des données de départ (précision des classes et **zonage** des régions phytogéographiques). Cependant, un certain nombre de discontinuités spatiales dans les groupements sont apparues. Il convient de préciser si ces discontinuités expriment des similitudes **réelles** entre régions éloignées, ou si elles résultent **d'artefacts** liés à l'extraction des données (taille des polygones d'extraction) ou **liés** au modèle finalement retenu.

Cette critique sera réalisée selon deux approches :

- d'une part, un travail d'expertise sur la **classification** globale.
- d'autre part, l'application et l'expertise du modèle à chacun des plans factoriels, puis à chacun des thèmes dégagés (occupation du sol, substrat, températures, précipitations, climat général).

Ce sera l'objet du second volume de ce rapport **final**.

10. Mise en place d'une base "**Ecorégions**"

10.1. **Principes**

Le présent rapport est avant toute chose une introduction à notre travail. L'essentiel de **celui-ci** réside dans la base de données géographique informatisée mise à disposition des Agences de **l'Eau**. Il est prévu que cette base fonctionne sous le système **ArcInfo™**.

Le plan cartographique de travail de la base est le maillage Dupias & Rey transcodé de **APIC™** vers **ArcInfo™**, comportant 660 unités* codées "**xx_A_yyy**" où "xx", "A" et "**yy**" correspondent respectivement aux premier, second et troisième niveau des régions phytogéographiques. Les fichiers fournis avec ce plan peuvent être utilisés sous forme de tables attributaires (au sein du S.I.G.) ou traités indépendamment. Ces fichiers sont les suivants :

- fichier 660 régions x 240 classes, précisant, pour chaque classe, le pourcentage de représentation surfacique au sein d'une région donnée.
- fichier de légende des classes.
- fichier des caractéristiques de la hiérarchie des régions.
- fichier des caractéristiques de la hiérarchie des classes.
- fichier de l'appartenance d'une région D. & R à une classe cartographique d'un des sept niveaux retenus (partitions en **4, 8, 16, 32, 64, 128** et 256 régions cartographiques)
- fichier des correspondances entre régions cartographiques et noeuds de la hiérarchie.

- fichier des classes les plus **représentées** dans chaque région cartographique (pour illustration).

● eafkitk466ounihAPIc(objdsglm#iqw)œ correspondent pas à 660 enregistrements ARCINFO, ce dernier ne pouvant considérer plusieurs polygones dissociés dans l'espace comme une seule entité.

10.2. Utilisation (suggestions).

L'utilisation de la base est, bien entendu, sans contrainte. Nous proposons simplement ici quelques façons de l'exploiter.

- **Mise** en évidence de la hiérarchie.
 - ⇒ La façon la plus simple est d'exploiter la classification en sept niveaux fourme, et d'affecter des couleurs / trames proches pour une même famille de groupes (*cf.* cartes jointes à ce rapport).
 - ⇒ Si les groupements préétablis ne semblent pas satisfaisants, une autre solution consiste à reconstruire une partition, à partir du fichier de description de la hiérarchie. **Il suffit** pour cela de relier les champs **Groupe/Aîné/Benjamin/Père** et de créer la parution sur les effectifs, le niveau de distance, la **variance** relative ou le nombre de groupes désiré. L'idéal serait de construire un dendrogramme qui permette, par sélection, de visualiser interactivement les relations hiérarchiques.
- Explication des régions
 - ⇒ Il suffit de créer un état sur les données d'origine, fournissant la tendance moyenne et la variabilité au sein de l'ensemble de régions constituant le groupe hiérarchique étudié. Il est recommandé de pondérer ces calculs par la surface unitaire des polygones.
 - ⇒ D'une manière plus générale, les corrélations et les contingences entre facteurs et classes peuvent aussi être testées à partir du même fichier.
- Tests des régions
 - ⇒ Un premier **test** est la validation de **l'affinité** des régions à partir des données environnementales d'origine, ou encore à partir de variables complexes supposées intégratives des premières. En effet, la pertinence de la hiérarchie peut être entachée de biais lié à la référence **surfacique** adoptée. Ce biais est en grande partie l'objet de l'analyse plus détaillée faisant l'objet du second volume de ce rapport.
 - ⇒ Le véritable test des régions sera effectué au travers de variables **supposées** indépendantes de notre hiérarchie. Ce sont notamment des variables de type biologiques. Il faut cependant garder à l'esprit que les variables et l'échelle adoptées pour ce travail limitent la sensibilité des régions. Une liaison avec la qualité de l'eau dans un système soumis à l'anthropisation relèverait du miracle . . . Cependant' il est possible (et recommandé) de tester les régions

sur d'autres critères. Il **suffit** pour cela **d'affecter** à chaque polygone une **estimation** de la variable testée, et d'appliquer une ordination simultanée en espace réduit (**AFC**, ACP, AFD, **AFCa**) ou de procéder à des tests de type "tableaux croisés multidimensionnels" (**modèles log-linéaires** du carré de contingence). Ceci revient à reproduire les méthodes **développées** par Hugues et al. (**1987**), ou par Rohm et al. (**1987**), **afin** de tester les régions américaines sur les populations de poissons ou sur la qualité **physico-chimique** de **certains** bassins.

10.3. Conclusions.

En guise de conclusion' nous insisterons une dernière fois sur les conditions d'application de cette base.

Il serait vain de chercher dans notre travail un quelconque outil universel pour l'aide à la décision. Ceci suggère que l'on garde sans cesse à l'esprit la nature des variables utilisées, leur échelle d'application et la méthode employée. En particulier, il serait tentant de superposer l'information cartographique contenue dans notre document à d'autres types de cartes (**B.D.Carto** IGN, cartes détaillées, MNT, données recueillies très ponctuellement) pour en extraire de nouvelles limites. Nous ne pouvons que recommander de la prudence dans un tel cas. **Il** est clair que nous dégageons toute responsabilité dans une utilisation maladroite (et peut-être dangereuse) de ce travail.

Par contre, nous recommandons le développement conceptuel et technique d'un tel outil. A notre avis, ce développement devrait se faire simultanément selon deux axes :

- ⇒ un **axe thématique**, avec l'intégration **et/ou** la sélection de variables liées à un problème précis.
- ⇒ un **axe géographique**, avec une augmentation de la résolution spatiale de l'outil, qui devrait permettre de l'orienter plus directement sur la gestion des cours d'eau.

Dans ce sens, toute optimisation du présent modèle sera la bienvenue, afin d'assurer des bases solides aux **futurs** développements. Nous ne pouvons donc qu'encourager les critiques futures de ce travail.

11. Références.

- Banfield** J.D. et **Raftery** AE. 1993 Model-Based Gaussian and non Gaussian clustering. *Biometrics*, **49(3)** : 803-22.
- Bayley RG. 1976 Ecoregions of the United States Map 1: 7 500 000 USDA **Forest** Service Intermountain **Region**, Ogden, UT.
- Bayley RG. 1989 Explanatory **supplement** to ecoregion map of the continents **Environmental Conservation**, 16 : 307-09.
- Bessemoulin J. 1969 Atlas climatique de la France. Direction de la Météorologie Nationale.
- Breiman L., Friedman J.H., Olsen R, et Stone C.J. 1984 Classification and Regression Trees. Wadworth International Group, Behnont, **California**.
- Chambers J.M. et Hastie T.J. (éds.) 1992 **Statistical** Moels in S. Wadsworth & Brooks, **Pacific Grove**, Califomia., 608 pp.
- Choisnel E. 1986 Les climats régionaux français. Presse thermale et climatique, 123 : 34-39.
- Choisnel E., Payen D. et **Lamarque** P. 1987 Climatologie de la zone du projet Hapex-Mobilhy. Direction de la Météorologie Nationale, Paris.
- Chou P.A, Lookabouth T. et Gray R.M. 1989 Optimal Pruning with applications to **Tree-structured** Source Coding and Modelling. **IEEE Trans. Inf. Theory**, 35 : 299-3 15.
- Ciampi A, Chang C.H., Hoggs S., et **McKinney** S. 1987 **Recursive Partitionning** : a Versatile Method for Explanatory Data **Analysis, Analysis** in Biostatistics, in : Biostatistics, (LB. Mc **Neil** et G.J. Humphrey, éds.), Reidel Publishing, New-York.
- Clarke S. ER, Garner M.W., **McIntosh B.A.**, et **Sedell J.R.**, 1994 **Landscape-Level** Ecoregions for Seven Contiguous Watersheds NE Oregon and SE Washington. BPA **Project** 89-104, non publié, 104 pp.
- Clarke S., Thiele S., **McIntosh** B., et Sedell, J 1993 Multiscale **Regional Analysis** of Stream **Physical** Habitat. Présentation et Poster au Eight **Annual** US Landscape Ecology Symposium, Patters and Processes in Landscape Ecology, Oak Ridge, TN, 24-27 Mars 1993.
- CNRS-Service de la Carte de la Végétation, Centre **d'Ecologie** des Ressources Renouvelables. 1947-92. Carte de la végétation de la France au 200.000 ^{ième}.
- Comité National de Géographie 1936-1958 Atlas de France. Editions Géographiques de France, Paris.
- Coutagne 1935 Des régimes pluviométriques en France et de leur classification. *La Météorologie*, 542-47.
- DATAR/BRGM** 1970 Atlas des eaux souterraines de la France. BRGM éds. Orléans.
- Day W.H.E. 1977 **Validity** of clusters **formed** by graph-theoretic cluster methods. *Math. Biosci.* 36 : 299-3 17.
- Dupias G et Rey P. 1985 Document pour un zonage des régions phyto-écologiques et carte des régions écologiques de la France. CNRS **éd.**
- Dupuis** J. 1967 Carte pédologique de la France à l'échelle du millionième. **INRA** éd., Paris.

- Emberger L. 1954 Projet d'une **classification biogéographique** des climats. pp. 5-11 in : Les divisions écologiques du Monde, **C.N.R.S., éd.** Paris.
- Escourroux G. 1980 Climat et Environnement. Les facteurs locaux du climat. Collection Géographie, Masson Paris.
- Estabrook **G.F.** 1966 A **mathematical model** in graph theory for biological **classification**. J. theor. Biol. 12 : 297-310.
- Everitt B. 1980 **Cluster** Analysis (seconde édition). Halsted, New-York.
- Fisher W.D. 1958 **On** Grouping for Maximum Homogeneity. J. Am. Statist. **Assoc.** 53 : 789-98.
- Gallant A.L., Omernik J.M., et Hughes R.M. 1990 Ecoregions of the Seine River **Basin**, France. A **status** report Map, U.S. E.P.A, **Corvallis**, OR 27 p. Non Publié
- Garnier M. 1964 Valeurs normales des températures en France (1921-1950). Monographies de la Météorologie Nationale.
- Gausson H. 1954 Expression des milieux par des formules écologiques. Leur représentation cartographique. pp. 13-25 in : Les divisions écologiques du Monde, C.N.R.S., éd. Paris.
- Gausson H. 1954 Rapport général sur la cartographie écologique. Pp. 221-23 in : Les divisions écologiques du Monde, **C.N.R.S.**, éd. Paris.
- Gordon A.D. 1981 Classification: Methods for the Explanatory Analysis of Multivariate Data. Chapman and **Hall**, Londres.
- Haggett P., **Cliff A.D.** et Frey A 1977 Locational Models. Edward Arnold, Bristol.
- Hartigan J.A 1975 Clustering algorithms. **Wiley**, New-York
- Hastie T. et Pregibon D. 1990 Shrinking Trees. **AT&T** Bell Laboratories **Tech.** Report.
- Herbertson, A.J. 1905 The Major **Natural** Regions : An essay in systematic geography, Geographical Journal, 25 : 300-12.
- Hughes R.M., et Larsen D.P., 1988 Ecoregions: an Approach to Surface Water Protection. Jour. Wat. **Poll.** Control, 487-89.
- Hughes R.M., Larsen D.P., Omernik J.M. 1986 **Regional References** Sites : A Method for **Assessing** Stream Potentials. Environmental Management, 10 (5) 629-35.
- Hughes R.M., Rexstad, E., et Bond C.E. 1987 The relationships of Aquatic Ecoregions, River Basins and Physiographical Provinces to the Ichthyogeographic Regions of Oregon. Copeia, 2 : 423-432.
- Institut Géographique National 1971 Carte de la France au 1/1000000.
- Johnson K.N., Franklin J.F., Thomas, J.W., **Gordon**, J. 1991 Alternatives for Management of **Late** Successional Forests of the **Pacific** Northwest. Rapport non publié, 59 pp.
- Kass** G.V. 1980 An explanatory technique for investigating large quantities of categorical data. Applied statistics 29 : 119-27.
- Larsen D.P., **Omernik** J.M., Hughes R.M., Rohm C.M., Whittier **T.R.**, Kinney A.J., Gallant A.L., et Dudley, **D.R.** 1986. Correspondance between Spatial Patterns in Fish Assemblages in Ohio Streams and Aquatic Ecoregions. Environmental Management 10 (6) 815-828.
- Legendre** P. et **Legendre** P. 1984. Ecologie numérique. Masson, Paris.

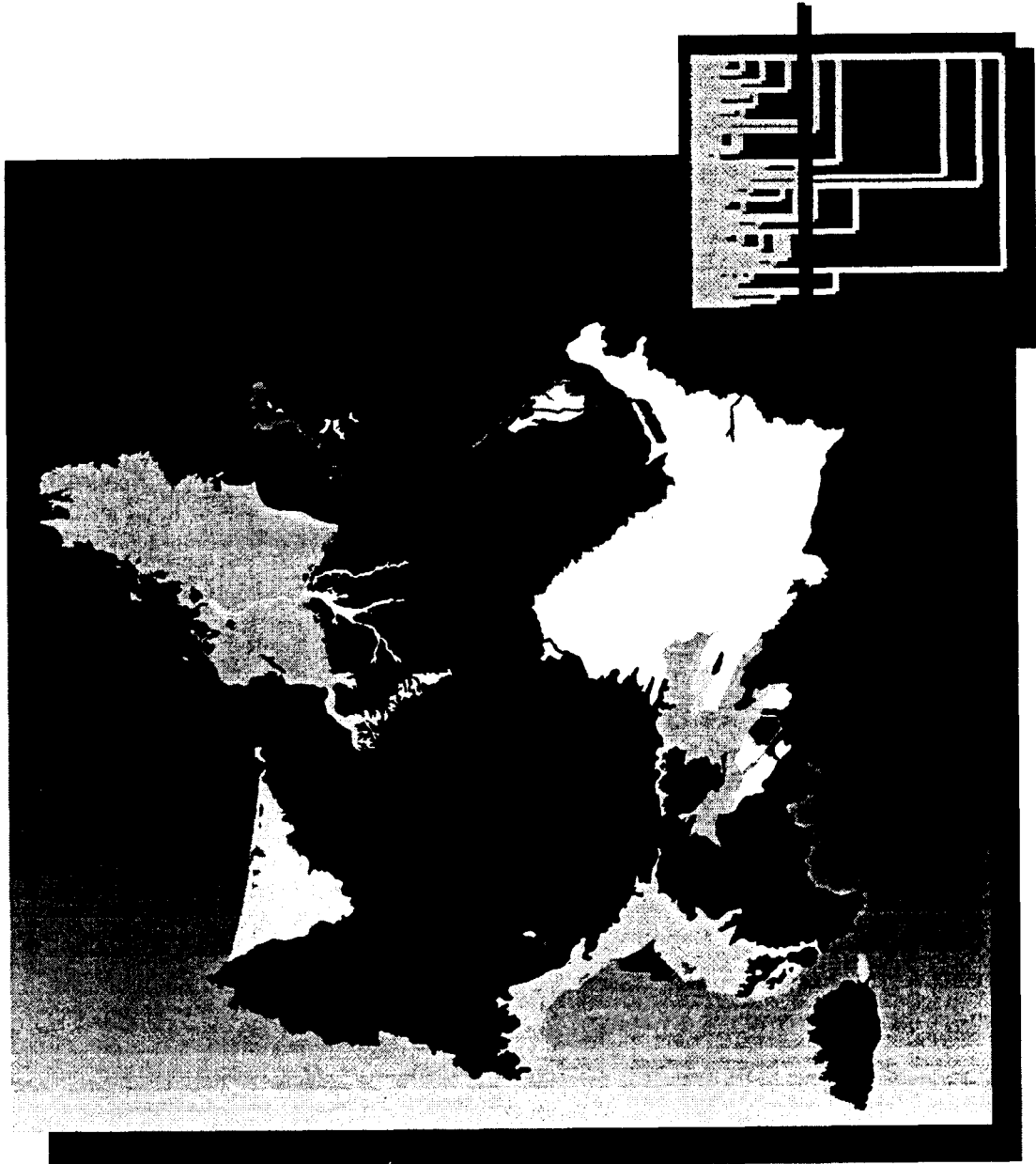
- Loveland TR., Merchant J.W.,** Ohlen D.O., Brown JF., 1991 Development of a Land-Cover Characteristics **Database** for the Conterminous US. **Photogrammetric Engineering and Remote Sensing**, 57 (11) 1453-63.
- Mardia K.V.,** Kent J.T., et Bibby J.M 1979 Muhivariate **Analysis**. Academic **Press**, London.
- Ministère de l'Agriculture et de la Forêt - INSEE SCEES. 1989 AGRESTE **Recensement** agricole 1988.
- Morgan J.N. et Messenger RC. 1973 A sequential **analysis** program for the analysis of nominal **scale** dependent variables. **Survey** Research Center, University of Michigan.
- Murtagh F. 1985 Multidimensionnal Clustering Algorithms, vol. **4.**, Physica-Verlag, Heidelberg.
- Omernik J.M. 1987 Ecoregions of the Conterminous United States. **Annl. Assoc. Am. Geogr.**, **77(1)** : 118-25.
- Omernik J.M. et **Gallant AL.** 1989 **Defining** Regions for Evaluating Environment Resources. pp 936-947 in : Proceedings of the International **Conference** and Workshop on Global **Natural** Resources Monitoring and Assessments. Am. Soc. Phot. And **Remote Sensing Publ.**
- Omernik J.M. et Larsen D.P. 1988 Summer Total Phosphorus in **Lakes**: A Map of Minnesota, Wisconsin, and Michigan' USA. *Environmental Management* 12 (6) : 815-826.
- Omernik J.M., et **Griffith G.E.** 1991 Ecological regions versus **hydrologic** units : Frameworks for managing water quality. *Journ. Soil Wat. Conservation* 46 (5) **334-40**.
- Pagny P.** 1988 Climats et cours d'eau de France. Collection Géographie, Masson éd. Paris.
- Perruchet C.** 1983 **Significance** tests for **clusters** : overview and **comments**. 109-208 in : J. Felsenstein éd. Numerical taxonomy : Proceedings of a NATO Advanced Study Institute., Série G (Ecological Sciences), 1, Springer Verlag, Berlin' x + 644 p.
- Rohm C.M., et Giese J.W. 1987 Evaluation of an Aquatic Ecoregion Classification of Streams in Arkansas. *Journ. Freshwat. Ecol.*, 4 (1) **127-40**.
- Scott AJ. et Symons M.J. 1971 Clustering Methods Based on Likelihood Ratio Criteria. *Biometrics*, 27 : 387-97.
- Sonquist J.N. et Morgan J.N. 1964 The **detection** of interaction **effects**. Monograph 35, Survey Research Center, University of Michigan.
- Thibault J. 1993 Contribution à l'élaboration d'une carte des écorégions de la France pour une application à l'étude des systèmes fluviaux. **DEA** Ecologie des Eaux Continentales, Toulouse **III**, **48 pp. + x**.
- Ward J.H. 1963 Hierarchical groupings to **optimize** an objective **function**. *Journ. Am. Statistical. Assoc.* 83 : **394-405**.
- Wasson J.G., Bethemont J., **Degorce J.N., Dupuis B., Joliveau T.**, 1993 Approche Ecosystémique du Bassin de la Loire, Phase 1: Etat initial-Problématique. 102 pp. (non publié).
- Zumstein J.F., Gille E., Decloux J.P.** et Paris P. 1989 Atlas de la lithologie et de la perméabilité du Bassin Rhin-Meuse. Publ. Agence de **l'Eau** Rhin Meuse.

**IDENTIFICATION des ECOREGIONS
HYDROLOGIQUEMENT HOMOGENES
du TERRITOIRE NATIONAL**



Dernière phase :

**DELIMITATION MULTICRITERE DE SE COREGIONS
VOLUME II**



*Etude Inter-Agences
Agence de l'EAU Rhin-Meuse*

CENTRE D'ECOLOGIE DES SYSTEMES AQUATIQUES CONTINENTAUX,
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE PAUL SABATIER TOULOUSE III
UMR C5576
29, RUE JEANNE MARVIG
31055 TOULOUSE CEDEX 04

ERIC TABACCHI, SEPTEMBRE 1996

SOMMAIRE

Avant-propos

1. Approche initiale

2. Rappels méthodologiques

2.1. Méthode de conception de la hiérarchie.

2.2. Plans d'information initiaux.

3. Approche récurrente de la CAH initiale

3.1. Méthode

3.2. Plans et C.A.H. dérivés

4. Résultats

4.1. Formats

4.2. Commentaires des résultats graphiques

4.2.1. Cartes dérivées.

4.2.2. Exemples de contingence entre plans dérivés et plan écorégional

4.3. Corrélations entre hiérarchies.

5. Utilisation de la base *ECOREGIONS*

5.1. Composition de la base

5.2. Proposition d'applicatifs dérivés de la base (rappels et compléments)

Annexe 1 : Vues générales des arbres de classification

Annexe 2 : Corrélations de Pearson sur les paramètres des classifications

Annexe 3 : Guide d'utilisation de la base *ECOREGIONS*

AVANT-PROPOS

Le premier volume de la dernière phase de la présente opération intitulée “Définition des écorégions hydrologiquement homogènes du Territoire National” avait pour but de fournir un découpage hiérarchisé de la France. Ce découpage, réalisé à l’échelle du 1: 1 000 000 visait à mettre en évidence les discontinuités majeures de notre Territoire National en fonction d’un certain nombre de paramètres susceptibles d’influencer le fonctionnement des écorégions ainsi définies. Le présent volume en constitue l’achèvement.

De 1992 à 1996, ce travail réalisé sous la coordination de *l’Agence de l’Eau Rhin-Meuse* a impliqué 13 personnes appartenant au *Centre d’Ecologie des Ressources Renouvelables du CNRS*, devenu depuis *Centre d’Ecologie des Systèmes Fluviaux*, puis *Centre d’Ecologie des Systèmes Aquatiques Continentaux*. Ce travail a été réalisé sous la responsabilité de M. Henri Décamps.

Le second volume de cette opération vise à préciser les correspondances de la classification des écorégions et les plans d’information ayant servi à l’établir. L’achèvement de cette opération doit beaucoup à l’aide et à la compréhension de l’Agence Rhin-Meuse, et en particulier à J.L. Salleron et à J.F. Zumstein. *L’Agence de l’Eau Adour-Garonne*, en la personne de J.L. Clotes, a largement contribué à la mise au point de l’aspect informatique des documents finaux.

1. APPROCHE INITIALE

Les écorégions ont été définies dans le premier volume sous un format très souple. Il s'agit d'une classification hiérarchique de 660 sous unités, dont le regroupement résulte d'un processus statistique établi sur 240 classes réparties en 10 variables environnementales. Un tel format se justifie dans la mesure où le modèle écorégional a été mis en place sans problématique-cible particulière, et se doit donc de rester adaptable à un problème précis. Cela ne veut pas dire pour autant que ce modèle soit universel, et seul un test ultérieur pourra confirmer sa pertinence dans tel ou tel cas.

En l'état actuel, les écorégions, quel que soit leur niveau d'agrégation, constituent une coïncidence de facteurs environnementaux. Leurs limites ne constituent pas des structures géographiquement stables, mais plutôt un gradient fort au sein du champ de facteurs. De ce fait, il ne nous paraît pas inutile de rappeler ici que ces régions n'ont pas d'adéquation obligatoire avec des structures géographiques **et/ou** écologiques reconnaissables. Cette adéquation n'apparaît i) qu'à un certain niveau d'agglomération et ii) qu'en fonction d'une problématique ayant initié le processus de classification. Ce phénomène est amplifié par la quantité importante de variables mises en oeuvre.

Ceci justifie l'approche de ce second volume, qui vise à établir une série de nouvelles classifications, faisant référence à un thème précis, formé d'un nombre réduit de plans d'information.

2. RAPPELS METHODOLOGIQUES

2.1. Méthode de conception de la hiérarchie.

Les méthodes ont été détaillées et justifiées dans le premier volume. Rappelons simplement que les unités géographiques élémentaires correspondent, à quelques modifications mineures près, aux petites régions de Dupias et Rey (1985). Ces polygones élémentaires ont été regroupés selon une classification hiérarchique ascendante (C.A.H) établie à partir des surfaces relatives de chaque classe (valeur qualitative ou **semi-quantitative**) d'un plan d'information (facteur environnemental). Ces données élémentaires ont été obtenues via un Système d'Information Géographique en croisant chacun des dix plans fondamentaux avec le plan des polygones élémentaires.

La classification hiérarchique établit les liens existant entre les 13 19 groupes possibles, depuis la région unité isolée jusqu'au regroupement des 660 régions unités. Le critère de ressemblance adopté est une distance du χ^2 .

De façon arbitraire, cette classification a été découpée en ensembles de 2'' (de 4 à 256) groupes afin de fournir une illustration de la hiérarchie. Ces groupes ont été appelés niveaux.

Accessoirement, une classification des facteurs permet d'apprécier la hiérarchie de regroupement des classes des divers plans d'information.

2.2. Plans d'information initiaux.

Les plans actifs ayant présidé à la mise en place de la classification générale des écorégions sont au nombre de dix :

- Relief (HPS). Le plan a été obtenu à partir de la carte 1. 000. 000 ème de l'IGN. Ce document s'est avéré plus précis que sa version numérisée au 500 000^{ème}, en particulier entre les isohypses 0 et 500 m.
- Pédologie (PED). La carte au 1 000 000^{ème} de Dupuis (1967) a été retenue.
- Hydrolithologie (HLS). La réalisation de ce plan s'est appuyée sur l'Atlas des Eaux souterraines de la France (DATAR-BRGM, 1970), sur les cartes géologiques du BRGM à diverses échelles (80 000^{ème}, 320 000^{ème} et 1 000 000^{ème}) et sur la classification proposée par Zumstein *et al.* (1989).
- Température moyenne de Janvier (TJN). Le plan est construit sur la base de l'Atlas de France (Comité National de Géographie, 1936-1958), et réactualisé d'après Garnier (1964), et les données de la Météorologie Nationale (1951 - 1993).
- Température moyenne de Juillet (TJT). La démarche est identique à celle adoptée pour les températures de Janvier.
- Volume des précipitations annuelles (PLU). La base cartographique est celle de Sanson (1945), qui s'est avérée plus précise que celle fournie par le modèle Aurelhy (Direction de la Météorologie Nationale, 1988). Les données ont été actualisées sur la base des normales longue période (1951- 1993).
- Durée des précipitations > à 1 mm (NJP). Ce critère a été jugé plus fiable que son équivalent pour $P > 0,1$ mm. Le plan est établi sur la base de la Carte des Précipitations en France au 2 000 000^{ème} (Météorologie Nationale, 1988), et réactualisée à partir des valeurs 1951-1993.
- Régimes pluviométriques (RPL). La typologie est réalisée *de novo* à partir des données longue période de Météo France et de diverses publications locales (Choisnel 1986, Choisnel et al. 1987, Coutagne 1935).
- Types de végétation potentielle (VEG). Une synthèse de la Carte de la Végétation de la France au 200 000^{ème} (CNRS 1947-92) constitue la base de ce document.
- Types d'agriculture dominants (AGR). Ce plan s'appuie sur une synthèse communale de la base AGRESTE (SCEES, 1989).

La classification proposée dans le premier volume résulte donc du traitement simultané de la matrice 660 polygones x 240 classes. D'un point de vue thématique, cette classification associe les dix plans d'information sans discrimination.

3. APPROCHE RECURRENTE DE LA C.A.H. GENERALE.

3.1. Méthode

Parmi les différentes méthodes permettant d'analyser la classification générale des écorégions, nous avons opté pour une méthode récurrente. Elle consiste à considérer non plus l'ensemble des classes et plans factoriels, mais des sous-ensembles cohérents d'un point de vue thématique. Une telle approche ne se justifie pas d'un point de vue mathématique : il n'existe pas d'effet additif entre des classifications "partielles" et une classification globale. La pondération des variables dans la classification peut être approchée par les paramètres de l'arbre. Cependant, nous avons privilégié la méthode "au cas par cas" dans la mesure où elle fournit simultanément i) des informations sur l'impact des facteurs sur la CAH globale et ii) une vision plus thématique des écorégions.

Il est important de rappeler que la définition analytique d'une écorégion à partir des données initiales reste le meilleur moyen de définir une telle entité. L'approche proposée ici peut donc paraître surprenante, dans la mesure où une redondance est attendue entre objets **classifiants** (plans) et objets classifiés (écorégions). Cependant, il convient de préciser que l'objectif n'est pas de définir de nouvelles écorégions sur tel ou tel plan, mais de préciser les mécanismes de construction de la hiérarchie générale. De même, il ne faut pas confondre cette approche avec un test des écorégions. Un tel test devra être réalisé avec des données externes aux plans d'information utilisés dans l'élaboration de notre classification.

La correspondance entre les CAH partielles et la CAH générale a été analysée au travers des contingences existant entre les deux ensembles pour un niveau de partition donné (par exemple, 32 ou 64 niveaux). Cette contingence est une façon particulière de considérer le problème. En effet, rien n'indique qu'un facteur (ensemble de classes) fournira une partition cohérente au même niveau qu'un autre facteur. Simplement, notre hypothèse ici est que si ce cas de figure se vérifie, alors l'influence du facteur testé s'avère dominante dans les deux classifications.

Cette approche s'accompagne d'une **lourdeur** évidente. De même que la caractérisation exacte d'une région d'un niveau donné doit nécessairement passer par un retour aux données et par l'extraction des informations via un SGBD, les analyses de contingences sont données à titre indicatif. Une analyse rationnelle du système relève également d'une exploitation a posteriori des données initiales via un système de traitement annexe. Un chapitre dans ce volume sera consacré aux possibilités de mise en place d'un système d'exploitation des divers arbres de classification.

3.2. Plans et C.A.H. dérivés

Afin de juger de la contribution géographique des différents thèmes impliqués dans la classification générale, le présent volume se propose de proposer une série de quinze classifications accessoires des 660 polygones élémentaires. Chacune de ces classifications décrit l'agglomération hiérarchique des polygones-unité soit en fonction des classes appartenant à un plan d'information unique, soit en fonction des classes appartenant à un nombre limité de plans regroupés selon une thématique forte.

Les classifications proposées résultent donc de deux types de facteurs :

- Classifications selon chaque plan d'information (facteurs simples):
 - Relief (HPS) : 11 classes

- Pédologie (PED) : 19 classes
 - Hydrolithologie (HLS) : 20 classes
 - Température moyenne de Janvier (**TJN**) : 14 classes
 - Température moyenne de Juillet (**TJT**) : 11 classes
 - Volume des précipitations annuelles (PLU) : 10 classes
 - Durée des précipitations > à 1 mm (**NJP**) : 14 classes
 - Régimes pluviométriques (**RPL**) : 38 classes
 - Types de végétation potentielle (**VEG**) : 52 classes
 - Types d'agriculture dominants (AGR) : 51 classes
- Classifications selon des thèmes (combinaison de facteurs) :
 - Substrat (SBT) : HPS + PED + HLS
 - Occupation du sol (OCS) : AGR + VEG
 - Synthèse climatique (CLI) : TJN + TJT + PLU + NJP + RPL
 - Synthèse des températures (TMP) : TJN + TJT
 - Synthèse des précipitations (PRC) : PLU + NJP + RPL

Les méthodes mathématiques utilisées sont identiques à celles de la C.A.H. générale (*Cf.* Vol. 1).

4. RESULTATS

4.1. Formats

Les résultats sont présentés sous trois formes :

- 0 pour synthèse, le codage des polygones élémentaires pour chacun des quinze plans analysés dans ce volume. Le codage a été réalisé de la même façon que pour la hiérarchie générale des écorégions (cf. vol. 1).
- 0 pour illustration, des cartes de niveau 32. Cette résolution a été retenue en raison d'un compromis entre la lisibilité des cartes et leur fidélité par rapport au plan d'origine. Les couleurs et trames ont été choisies arbitrairement ; elles visent à représenter un gradient dans le cas de plans référant à des variables quantitatives.
- 0 pour analyse, les arbres de classification. Leur présentation est cohérente avec celle de la classification générale du volume 1, à ceci près que les polygones élémentaires sont codés de 1 à 660 (voir table de correspondance avec la nomenclature type "**Dupias** et Rey" en annexe). Chaque tableau fournit les informations suivantes :
 - * numéro du groupe
 - * numéro du aîné
 - * numéro du benjamin
 - * numéro du père
 - * effectif du groupe
 - * distance de création
 - * pourcentage d'inertie du groupe

L'Annexe 1 fournit une illustration des agglomérations hiérarchiques.

4.2. Commentaires des résultats graphiques

Les classifications de niveau 32 ont été évaluées au travers de deux critères :

- la sévérité de la partition entre les niveaux 4 et 32.
- le niveau de contingence existant entre ces classifications et la classification des écorégions de niveau 32.

4.2.1. Cartes dérivées.

Pour les plans simples, la dégradation de l'information par rapport aux plans initiaux ne correspond pas seulement à l'interception de ces plans par le maillage écorégional le plus fin (660 régions). Les différentes hiérarchies intègrent en effet le codage initial et permettent l'analyse de l'agglomération des polygones selon ce codage. Les cartes de synthèse (occupation du sol, substrat, climat, précipitations, températures) peuvent être considérées comme une analyse de contingence des plans initiaux, mais nullement comme une addition de l'information concernant ces plans.

Les quinze cartes de niveau 32 sont un compromis acceptable pour une analyse du comportement de la hiérarchie et pour une représentation graphique. Ces cartes sont proposées ci-après, la carte écorégionale de niveau 32 leur étant associée comme référence.

4.2.2. Exemples de contingence entre plans dérivés et plan écorégional

En première approximation, la contingence entre plans dérivés et plans écorégionaux a été évaluée par comparaison des hiérarchies de niveau 32. Le but de cette analyse est de proposer un exemple d'évaluation de l'implication des divers plans dans la construction de la hiérarchie des écorégions. D'un point de vue méthodologique, la mesure employée peut être substituée par d'autres coefficients (carré moyen, chi carré, coefficient de régression non paramétrique...). Les résultats sont du même ordre. Un second objectif est de souligner l'importance des variations d'appréciation de cette implication selon le niveau retenu. Aussi a-t-on comparé la mesure établie sur 32 niveaux à celle obtenue sur 64 niveaux.

Le niveau de contingence général a été évalué de la façon suivante :

- le code de chacun des polygones élémentaires a été mis en correspondance entre les deux classifications (écorégions et plan thématique). La contingence a été évaluée par le pourcentage de concordance pour chacune des 32 classes.
- le nombre de coïncidences entre les deux plans a été comptabilisé pour des classes de concordance (1-10%, 11-20%, 21-30% ... 91-100%).
- Pour le cumul de chaque classe de 10% (plus de 10%, plus de 20%, ...etc.). les concordances des divers plans thématiques (cartes) ont été comparées et classées par ordre croissant. Un rang (de 1 à 15) a été attribué pour chaque classe.
- La moyenne des rangs de l'ensemble des classes a été calculé pour chaque carte. Cette valeur donne une indication de la contingence générale du plan considéré avec celui des écorégions, les valeurs faibles signifiant une contingence élevée.