



AGROCAMPUS OUEST
CFR Rennes

65, rue de Saint-Brieuc
CS 84215 35042 Rennes
Cedex



Unité Mixte de Recherche
AMAP

Bd de la Lironde TA A-51/PS2
34398 Montpellier Cedex 5



TELA BOTANICA

Institut de Botanique
163, Rue Auguste Broussonnet
34000 Montpellier

Mémoire de Fin d'Études

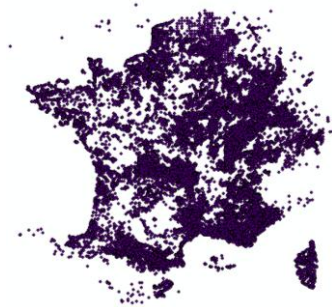
**Diplôme d'Ingénieur de l'Institut Supérieur des Sciences
Agronomiques, Agroalimentaires, Horticoles et du Paysage**

Année universitaire : 2011-2012

Spécialisation : Statistique appliquée

**Analyse écologique de la répartition de la végétation à partir d'une base de données
phytosociologiques : exemple de la végétation méditerranéenne**

Par : Marianne GERARD



Volet à renseigner par l'enseignant responsable de l'option/spécialisation*
Date : .../.../... Signature :

Bon pour dépôt (version définitive)
Autorisation de diffusion : Oui Non

Devant le jury :

Soutenu à Rennes le 5 septembre 2012

Sous la présidence de : Jérôme Pagès

Enseignant référent : David Causeur

Maître de stage: François Munoz

"Les analyses et les conclusions de ce travail d'étudiant n'engagent que la responsabilité de son auteur et non celle d'AGROCAMPUS OUEST".

Avant-propos :

L'étude ci-dessus a été initiée par l'association Tela Botanica, qui souhaitait valoriser la banque de données botaniques SOPHY qu'Henri Brisse, son créateur avait accepté de lui transmettre. François Munoz, maître de conférences à l'Unité Mixte de Recherche en botanique et bioinformatique de l'Architecture des Plantes (UMR AMAP) a accepté dans le cadre de ses recherches d'encadrer cette étude. C'est ce qui a donné lieu au stage de six mois réalisé de début mars à fin août dans les locaux de l'UMR AMAP et financé par Tela Botanica.

Ce stage va être poursuivi par un travail de thèse, le présent rapport constitue donc une base de travail pour le futur candidat qui prendra la suite de ce stage.

Remerciements :

Je tiens à remercier les personnes qui m'ont permis de mener à bien mon stage de fin d'étude. Celui-ci clôture mes trois ans de formation à Agrocampus Ouest.

En premier lieu, je remercie François Munoz pour m'avoir accordé une partie de son temps, pour son investissement sur le projet qui a débuté avec mon stage et enfin pour m'avoir permis de découvrir l'univers de la recherche en écologie végétale et les méthodologies statistiques associées.

Je tiens ensuite à remercier Cyrille Violle, chercheur au Centre d'Ecologie Fonctionnelle et Evolutive, pour les collaborations scientifiques qui ont eu lieu avec le projet DIVGRASS mais surtout pour sa confiance et ses conseils. Il a su me remotiver dans les moments de découragement et il a été un soutien tout au long de mon stage.

Je dois également beaucoup à l'équipe de Tela Botanica si sympathique, qui m'a chaleureusement accueillie à chacune de mes visites malgré ma délocalisation à l'UMR AMAP et qui m'a permis de sortir du monde de la recherche et de prendre du recul. Plus précisément, merci à Elise Mouysset, directrice de Tela Botanica, et Daniel Mathieu, son président, pour leur confiance. Merci à Delphine Cauquil pour son accueil et sa disponibilité pour mes questions sur la base SOPHY.

Je remercie également Arnaud Lalanne, chef de projet au Ministère de l'Ecologie, qui a porté un grand intérêt à la démarche et aux résultats obtenus au cours de mon stage et qui a souhaité contribuer à la poursuite des recherches initiées pendant ces six derniers mois.

Henri Brisse, fondateur de la base SOPHY, mérite que je le nomme car SOPHY a été mon sujet d'étude pendant six mois et grâce à elle, j'ai été plongée au cœur de la phytosociologie, des relevés botaniques et des noms de la végétation française.

Ma pensée finale va à toutes les personnes qui m'ont accompagnée durant mon séjour à Montpellier et qui l'on rendu agréable : mes voisines de bureau, Camille Barré et Diane-Laure Sorrel avec qui j'ai partagé de grandes discussions et bons moments, et Emilie Delétré, Marie Vasse, Simon Benateau et Christophe Plantamp qui ont toujours été présents et disponibles pour me conseiller mais aussi toujours prêts à partager des moments de détente !

Table des matières :

1. INTRODUCTION	1
1.1 Contexte général de l'étude	1
1.2 La recherche en écologie face à de nouveaux défis : développement de l'écoinformatique.....	1
1.3 La phytosociologie, une discipline oubliée au service de problématiques actuelles...	2
1.4 L'analyse de l'organisation des communautés végétales : ordination et analyse de modularité.....	3
1.5 Questions scientifiques de l'étude	3
2. MATERIELS ET METHODES	4
2.1 Préparation des données	4
2.1.1 Données floristiques	4
2.1.2 Harmonisation et sélection des données floristiques.....	6
2.1.3 Données environnementales.....	6
2.2 Détermination des grandes unités de végétation	6
2.2.1 Définition du réseau de végétation.....	7
2.2.1.1 Définition d'un graphe biparti et d'un graphe projeté	7
2.2.1.2 Prise en compte de l'abondance dans les graphes biparti et projeté	8
2.2.2 Détermination d'ensembles cohérents d'espèces et de stations	9
2.2.2.1 Analyse de modularité au sein du réseau projeté « espèce »	9
2.2.2.2 Spécificité des stations pour chaque module	10
2.3 Etude de la végétation dans le module « végétation méditerranéenne ».....	11
2.3.1 Ordinations non contraintes et typologie bioclimatique	11
2.3.1.1 Deux méthodes d'ordination non contrainte.....	11
2.3.1.2 Typologie bioclimatique des stations.....	11
2.3.2 Ordination contrainte par l'environnement bioclimatique et classification	12
3. RESULTATS	12
3.1 Caractérisation du jeu de données	12
3.1.1 Les données après sélection	12
3.1.2 Chiffres clés sur les espèces et les stations	13
3.2 Résultats de l'analyse de modularité	14
3.2.1 Le découpage des espèces en trois grandes classes d'abondance	14
3.2.2 Résultats de l'analyse de modularité.....	14
3.2.3 Stations caractéristiques des modules	15
3.2.4 Le module 5 : « végétation méditerranéenne »	15
3.3 Résultats des différentes méthodes d'ordination	17
3.3.1 Analyses d'ordination non contrainte	17
3.3.2 Typologie bioclimatique des stations	20
3.3.3 Ordination contrainte par l'environnement bioclimatique et classification	22
4. DISCUSSION	24
4.1 Organisation de la végétation méditerranéenne.....	24
4.2 L'apport des différentes analyses multivariées	25
4.3 L'apport de la théorie des graphes.....	26
4.4 Perspectives	26
4.4.1 Affinement des données	26
4.4.2 Analyses à continuer	27
5. CONCLUSION.....	28
BIBLIOGRAPHIE	29
ANNEXES	32

Table des illustrations :

Figure 1 : Schéma relationnel de la base de données SOPHY.....	5
Figure 2 : Schéma d'un graphe biparti théorique de végétation.....	8
Figure 3 : Schéma du passage d'un graphe biparti (a) à un graphe projeté (b).	8
Figure 4 : Schéma d'un graphe biparti théorique avec prise en compte de l'abondance..	9
Figure 5 : Schéma comparatif d'un graphe projeté théorique avec prise en compte de l'abondance (a) et d'un graphe projeté sans prise en compte de l'abondance (b).	9
Figure 6 : Visualisation des stations SOPHY avant (a) et après sélection des données (b).....	13
Figure 7 : Visualisation de la localisation de la diversité spécifique par station.....	14
Figure 8 : Projection des stations du module 5 avant (a) et après (b) filtrage des stations caractéristiques	16
Figure 9: Répartition des stations avec indication de leur diversité spécifique	16
Figure 10: Premier plan factoriel de l'AFC avec zoom sur l'axe 2.....	18
Figure 11 : Cartographie du deuxième axe factoriel de l'AFC	18
Figure 12 : Premier plan factoriel de l'ANSC (axe 1).....	19
Figure 13 : Cartographie du premier axe factoriel de l'ANSC	19
Figure 14 : Premier plan factoriel de l'ANSC (axe 2).....	20
Figure 15 : Cartographie du deuxième axe factoriel de l'ANSC	20
Figure 16 : Graphe des variables du premier plan factoriel de l'ACP	21
Figure 17 : Cartographie du premier (a) et deuxième (b) axe factoriel de l'ACP	21
Figure 19 : Premier plan factoriel de l'ANSCVI.....	23
Figure 18 : Cartographie du premier (a) et deuxième (b) axe factoriel de l'ANSCVI.....	23
Figure 20 : Localisation des stations du groupe 4 (a), le groupe 1 (b) et le groupe 6 (c)	24
Tableau 1 : Signification des classes d'abondance en phytosociologie	5
Tableau 2 : Nombre de stations en commun par module	15
Tableau 3 : Pourcentage de stations caractéristiques par module.....	15

Table des annexes :

Annexe I : Problèmes liés à la taxonomie et méthodes pour y palier.....	32
Annexe II : Données WorldClim.....	34
Annexe III: AFC et ANSC.....	35
Annexe IV : Résultats de l'AFC.....	36
Annexe V : Résultats de l'ANSC.....	39
Annexe VI : Résultats de l'ACP des variables bioclimatiques :.....	41
Annexe VII : Résultats de l'ANSCVI :.....	43

Glossaire :

Diversité β : mesure descriptive de la diversité des communautés d'espèces au sein d'un territoire qui prend en compte la manière dont des communautés d'espèces diffèrent les unes des autres et/ou de leur degré d'éloignement sur des gradients de composition spécifique. C'est aussi la probabilité de tirer au hasard dans deux communautés différentes, deux espèces différentes.

Géophyte : désigne un type de plante vivace possédant des organes lui permettant de passer la mauvaise saison enfouis dans le sol.

Halophyte : désigne des végétaux vivants sur des sols salés.

Mésophile : désigne des végétaux ou des formations végétales vivant dans des conditions moyennes d'humidité-sécheresse.

Végétation : terme désignant l'ensemble des végétaux peuplant un habitat ou une aire biogéographique donnée, aussi bien dans les formes végétales dominantes (végétation arborée, arbustive, herbacée) que prise au sens du ou des groupes taxonomiques dominants.

Xérophile : désigne des végétaux ou des formations végétales qui vivent en milieu sec.

Liste des abréviations :

AFC : Analyse Factorielle des Correspondances

ANSC : Analyse Non Symétrique des Correspondances

ANSCVI : Analyse Non Symétrique des Correspondances sur Variables Instrumentales

BDTFX : Base de données des Trachéophytes de France Métropolitaine

MNT : Modèle numérique de Terrain

SIG : Système d'Information Géographique

Liste des taxons végétaux cités :

Anthyllis barba-jovis: Anthyllis barbe-de-Jupiter

Anthyllis cytisoides: Anthyllis faux cytise

Aphyllanthes monspeliensis: Aphyllanthe de Montpellier

Asparagus acutifolius: Asperge sauvage

Asplenium sagittatum: Scolopendre sagittée

Atractylis humilis: Atractyle humble

Ballota frutescens: Ballote buissonnante

Brachypodium retusum: Brachypode rameux

Brassica montana: Chou des montagnes

Bromus erectus: Brome dressé

Calicotome spinosa: Calicotome épineux

Calicotome villosa: Calicotome velu

Centaurea jordaniana subsp. *Balbisiana*: Centaurée de Balbis

Cheirolophus intybaceus: Centaurée fausse chicorée

Cistus creticus var. *corsicus*: pas nom vernaculaire

Cistus creticus var. *eriocephalus*: pas nom vernaculaire

Cistus crispus: Ciste crépu

Cistus halimifolius: Ciste jaune

Cistus ladanifer: Ciste ladanifère

Clematis cirrhosa: Clématite à vrilles

Clematis flammula: Clématite odorante

Convolvulus lanuginosus: Liseron laineux

Coriaria myrtifolia: Corroyère à feuilles de myrte

Cyclamen repandum: Cyclamen du printemps

Cytisus ardoini: Cytise d'Ardoino

Cytisus triflorus: Cytise à longues grappes

Daucus carotta: Carotte sauvage

Eryngium campestre: Panicaut champêtre

Euphorbia biumbellata: Euphorbe à double ombelle

Euphorbia hyberna subsp. *Canutii*: Euphorbe de Canut

Festuca rubra: Fétuque diffuse

Fumana ericoides subsp. *Montana*:
Fumana procumbens: Fumana à tiges
retombantes
Galatella sedifolia: Aster à feuilles d'orpin
Galium obliquum: Gaillet oblique
Galium setaceum: Gaillet sétacé
Genista linifolia: Genêt à feuilles de lin
Hedera helix: Lierre
Helianthemum marifolium: Hélianthème à
feuilles de marum
Hieracium latifolium: Épervière à larges
feuilles
Hiracium pillosella: Piloselle,
Hormathophylla halimifolia: Alysson à
feuilles d'halimium
Koeleria vallesiana: Koélerie du Valais
Lavandula latifolia: Grande Lavande
Linum austriacum: Lin d'Autriche
Lonicera implexa: Chèvrefeuille des
Baléares
Moehringia intermedia: Sabline de
Provence
Myrtus communis: Myrtes commun

Ononis minutissima: Bugrane très grêle
Phillyrea angustifolia: Filaire à feuille
étroite
Pinus halepensis: Pin blanc de Provence
ou Pin d'Alep
Pistacia lentiscus: Pistachier lentisque
Potentilla crantzii: Potentille de Crantz,
Potentilla saxifraga: Potentille saxifrage
Potentilla zimmereri: Potentille du Midi
Primula allionii: Primevère d'Allioni
Quercus ilex: Chêne vert
Rhinanthus burnatii: Rhinante de Burnat
Rhus coriaria: Sumac des corroyeurs
Rosmarinus officinalis: Romarin
Rubia peregrine: Guarance voyageuse
Sedum fragrans: Orpin à odeur suave
Smilax aspera: Salsepareille
Stachys heraclea: Épiaire d'Héraclée
Teucrium montanum: Germandrée des
montagnes
Thymus vulgaris: Thym vulgaire
Viola arborescens: Violette ligneuse

Correspondance nom scientifique- nom vernaculaire. Source: www.tela-botanica.org.

1. INTRODUCTION

1.1 Contexte général de l'étude

La répartition spatiale actuelle d'une espèce végétale est le résultat de différents facteurs: environnementaux (conditions climatique, édaphique, topographique...), historiques (processus passés qui ont agi sur les populations antérieures) et biotiques (capacité intrinsèque de l'espèce et processus d'interactions interspécifiques) (Lacoste and Salanon, 1978). Les espèces végétales qui coexistent forment des assemblages caractéristiques appelés communautés (Looijen and van Andel, 1999). Celles-ci peuvent être étudiées et caractérisées en relation avec l'environnement sans évaluer l'écologie de chacune des espèces constitutives indépendamment: c'est le champ de l'écologie des communautés. Cependant, les limites spatiales des communautés végétales sont plus ou moins nettes en fonction des caractéristiques du milieu. Cela a alimenté un débat pendant de nombreuses années en écologie végétale, la végétation*, étant soit vue comme un assemblage de communautés végétales bien identifiées que l'on peut séparer et classer, soit comme un continuum d'espèces le long de gradients environnementaux (Austin, 2005). Ces deux approches sont en fait complémentaires et selon l'échelle et le contexte de l'étude, l'une ou l'autre sont privilégiées.

Une typologie des gradients susceptibles d'influencer la répartition de la végétation a été proposée par Austin, (Austin, 1980), en distinguant les gradients qui ont un effet direct sur la physiologie de la plante mais ne sont pas consommés par la plante comme la température ou le pH, les gradients de ressources, fondés sur une ressource qui est consommée par les plantes (lumière, eau, nutriments) et les gradients indirects qui n'ont pas d'effet physiologique sur la croissance des plantes ou sur la compétition (par exemple l'altitude). La corrélation de ce dernier type de gradient avec la distribution des espèces dépend de leur corrélation avec les deux autres types de gradients, tels que la pluviométrie ou la température.

Comprendre la répartition des espèces et des communautés d'espèces à grande échelle, identifier les conditions environnementales qui contraignent cette répartition et modéliser les aires de distribution des espèces ou communautés d'espèces sont au cœur des problématiques scientifiques pour la mise en place de programmes de préservation de la biodiversité, l'orientation des politiques d'aménagement du territoire mais aussi pour évaluer l'impact du changement climatique. A l'échelle du territoire français, par exemple, l'accompagnement de Natura 2000 et de la politique Trame Verte et Bleue requiert une connaissance fine de la structure des formations végétales afin d'identifier les cœurs de biodiversité et leur connectivité. Les études menées dans ce cadre sur de grands territoires deviennent possibles avec le développement des systèmes d'information géographique (SIG) et l'augmentation de la capacité calculatoire des outils informatiques. Cependant, il est difficile de mobiliser une connaissance fine de la composition des communautés sur de larges étendues. Ainsi, ces études restent encore peu nombreuses et amènent par ailleurs des questionnements sur la mise en œuvre: comment analyser l'organisation des communautés végétales à cette échelle? Quelles sont les difficultés rencontrées dans ce genre d'études et quels sont les verrous à lever? C'est dans ce cadre que se place la présente étude qui a pour but d'analyser la répartition de la végétation à l'échelle française. Elle amène une réflexion sur ces questions.

1.2 La recherche en écologie face à de nouveaux défis: développement de l'éco-informatique

Face aux problématiques actuelles citées plus haut, les études réalisées en écologie ont rapidement évolué ces dix dernières années, passant d'études conduites par quelques

personnes à petite échelle sur un court terme, à des projets interdisciplinaires (études écologiques associées à de la génomique, géologie, climatologie...) conduits à plus grande échelle sur un plus long terme et utilisant de nouveaux outils comme les outils de télédétection (Kelling et al., 2009, Michener and Jones, 2012). On assiste ainsi à l'émergence de nombreuses bases de données en écologie, recueillant de massives quantités d'informations, complexes et hétérogènes par nature car souvent multi-sources et issues de protocoles différents (Jones et al., 2006, Kelling et al., 2009, Michener and Jones, 2012, Reichman et al., 2011). Ces bases de données posent un certain nombre de défis, d'une part en termes de nettoyage et d'harmonisation des données et d'autre part en termes d'analyses statistiques capables de gérer de telles quantités d'informations (Jones et al., 2006, Kelling et al., 2009, Michener and Jones, 2012, Reichman et al., 2011). La taxonomie, par exemple est l'un des points délicats, dans l'intégration de données et peut être une source de biais important dans les analyses écologiques (Jansen and Dengler, 2010, Patterson et al., 2010). L'écoinformatique, science émergente située à l'interface entre l'écologie, l'informatique et les technologies de l'information (Jones et al. 2006), développe des outils et des approches innovantes pour gérer, intégrer et analyser les données pertinentes en biologie et en écologie (Michener and Jones 2012). La présente étude s'appuyant sur une base de données compilant une très grande quantité d'informations issues de sources différentes, a été confrontée à ces problématiques émergentes et se rattache à ces thématiques.

1.3 La phytosociologie, une discipline oubliée au service de problématiques actuelles

La phytosociologie est une discipline apparue au début du XX^{ième} siècle, en Europe, et se divise alors en plusieurs écoles dont la plus célèbre est l'école Zurich-Montpellier, portée par M. Braun-Blanquet. Celui-ci publie en 1928, l'ouvrage de référence *Pflanzensoziologie* (Braun-Blanquet, 1928) qui permettra à cette école de s'imposer en Europe (Becking, 1957). En se fondant sur des inventaires floristiques, la phytosociologie permet de décrire précisément une communauté végétale en fonction de sa composition spécifique, reflet des caractéristiques du milieu. Cette discipline repose sur une méthodologie précise d'inventaire floristique (choix du lieu du relevé, surface échantillonnée, critères pour décrire l'abondance, la strate,..) et de classification des communautés végétales (Lacoste et Salanon 1978). Cette classification s'appuie notamment sur la notion d'espèces caractéristiques. Ce sont les espèces qui se trouvent exclusivement dans un type de communauté végétale (Guinochet, 1955). Ce système de classification a notamment servi d'appui à la typologie du programme CORINE Biotopes¹. Cependant, dans le milieu de la recherche scientifique, la phytosociologie est encore utilisée comme méthode d'inventaire floristique et de description de la végétation mais elle est actuellement délaissée comme méthode de classification, au profit de méthodes statistiques introduites en écologie plus récemment (Peet and Roberts, 2012). La phytosociologie comme méthodologie de terrain assure une cohérence au sein des bases de données de relevés phytosociologiques et notamment la base de données qui a fait l'objet de cette étude. La phytosociologie dans la création de base de données d'inventaires floristiques s'avère donc être un outil précieux.

¹« La typologie CORINE Biotopes est un système hiérarchisé de classification des habitats européens élaboré dans le cadre du programme CORINE (Coordination of Information on the Environment). L'objectif était d'identifier et de décrire les biotopes (habitats naturels et semi-naturels) d'importance majeure pour la conservation de la nature au sein de la Communauté européenne. » source : <http://inpn.mnhn.fr/programme/referentiels-habitats/corine-biotopes>

1.4 L'analyse de l'organisation des communautés végétales: ordination et analyse de modularité

L'analyse (Becking, 1957) de grosses bases de données d'inventaires floristiques qui couvrent une très large gamme de végétation à une large échelle (échelle nationale par exemple), suppose de bien définir le cadre méthodologique dans lequel on se place. L'analyse des gradients en écologie se fait principalement par des techniques statistiques d'ordination représentées par les analyses factorielles. L'ordination est le fait d'agencer des unités dans un ordre (Legendre and Legendre, 1998a). En écologie, il s'agit d'ordonner des espèces et/ou des relevés le long de gradients ou d'axes (Jongman et al., 1995) dans le but d'étudier les caractéristiques spatiales de la végétation (Whittaker, 1967). On distingue deux types d'ordination: l'ordination indirecte et l'ordination directe (Austin, 2005). La première ordonne les stations le long d'axes en fonction de leur composition floristique ou les espèces en fonction de là où elles sont présentes (Austin, 2005). Une des analyses les plus répandues pour ce type d'ordination est l'Analyse Factorielle des Correspondances (AFC) qui réalise une ordination simultanée des espèces et des stations (Dray and Chessel, 2009). Cependant, d'autres analyses se sont développées en écologie plus récemment comme l'Analyse Non Symétrique des Correspondances (Chessel and Gimaret, 1997, Gimaret-Carpentier et al., 1998, Kroonenberg and Lombardo, 1999, Lauro and D'Ambra, 1984, Péliissier et al., 2003) . L'ordination directe correspond à l'analyse de la distribution des espèces en relation avec des variables environnementales (Austin, 2005). L'ordination directe consiste pour les méthodes multivariées à contraindre les axes de l'analyse à être des fonctions des variables environnementales, d'où le nom aussi trouvé dans la littérature d'ordination contrainte (« *constrained ordination* » en anglais). Ce type d'analyse permet typiquement d'analyser la relation entre un tableau floristique et un tableau contenant des descripteurs environnementaux mesurés aux mêmes endroits (Legendre and Legendre, 1998b).

Une méthode encore peu explorée dans le domaine de l'écologie végétale pour étudier l'organisation des espèces, consiste à étudier les communautés végétales sous l'angle d'un graphe. Un graphe, qui est la représentation mathématique d'un réseau, est défini comme un ensemble de nœuds connectés par des liens ou des arrêtes. En considérant les espèces et les stations comme des nœuds reliés si l'espèce est présente sur la station, on forme un graphe biparti. Cette approche ouvre alors sur les méthodes d'analyses issues de la théorie des graphes, utilisées notamment en écologie pour étudier les réseaux mutualistes et trophiques entre espèces (Dormann et al., 2009, Thébault and Fontaine, 2010). La théorie des graphes est une branche des mathématiques qui permet d'étudier les réseaux. Les réseaux existants dans de nombreux domaines (réseaux de distribution, les réseaux trophiques, les réseaux sociaux..), de nombreuses branches de recherche (la physique, la sociologie, l'informatique, la biochimie..) se sont intéressées à la théorie des graphes (May, 2006, Newman, 2003). De nouvelles approches pour l'analyse de la structure des réseaux se sont développées dans la dernière décennie comme l'analyse de modularité (Blondel et al., 2008, Newman, 2006, Newman and Girvan, 2004). Cette analyse indique s'il existe au sein du réseau, des communautés de nœuds ou modules. Ceux-ci se définissent comme un groupe de nœuds ayant une densité de liens intra module plus importante que la densité de liens inter module (Newman and Girvan, 2004). L'application de la modularité à un réseau d'espèces végétales n'a jamais été développée et offre une perspective de détecter les espèces qui s'associent le plus.

1.5 Questions scientifiques de l'étude

La présente étude a eu pour but d'analyser la répartition de la végétation à l'échelle française parallèlement à l'élaboration d'une stratégie d'analyse de la végétation sur un jeu de

données très lourde et à l'exploration des possibilités de la stratégie adoptée. Plus précisément, les questions à laquelle nous souhaitons répondre sont les suivantes:

- Comment s'organisent les communautés végétales au sein d'un territoire ? Quelle est l'influence des variables bioclimatiques sur les communautés étudiées?
- Que mettent en avant les différentes méthodes d'ordination indirecte et directe?
- La théorie des graphes est-elle une technique pertinente pour définir des grands groupes d'espèces cohérents, c'est à dire qui correspondent à un même contexte écologique?

Les objectifs ont donc été premièrement d'harmoniser les données de la base de relevés phytosociologiques provenant de sources différentes, ensuite de définir au sein du jeu de données les grands types de végétation par une analyse de modularité (forêt, prairie, lande, zone humide...) et finalement, d'étudier la variation floristique au sein d'un de ces grands types en lien avec des variables bioclimatiques par deux méthodes d'ordination indirecte et une méthode d'ordination contrainte par les variables environnementales.

Remarque: le terme « espèce » sera employé de manière générique désignant un taxon végétal identifié soit au niveau taxonomique de l'espèce ou du niveau infra-espèce.

2. MATERIELS ET METHODES

2.1 Préparation des données

2.1.1 Données floristiques

Les données floristiques sont issues de la banque de donnée SOPHY (PHYtoSOciologie). Celle-ci est une banque de données botaniques, créée en 1978 par trois ingénieurs de recherche du CNRS, Gilles Grandjouan, Patrice de Ruffay et Henry Brisse. Il s'agit d'une compilation bibliographique de relevés phytosociologiques. Le plus vieux date de 1913 et les relevés s'échelonnent entre cette date et le début des années 2000. Cependant, 90% des relevés ont été échantillonnés entre 1965 et 2010. Ce qui assure la cohérence des relevés au sein de la base est le fait qu'ils ont tous été réalisés en suivant la méthodologie phytosociologique. Ainsi chaque relevé est l'inventaire floristique d'une communauté végétale se trouvant dans un lieu précis, appelé station. Une espèce répertoriée dans un relevé représente une observation.

Les informations de la banque de données ont été transformées en base de données SQL (fig.1) par Tela Botanica (association de loi 1901, www.tela-botanica.org) pour faciliter son utilisation. La base de données renseigne cinq grands types d'informations:

- Un tableau « publication » où sont consignées les références des publications d'où sont extraits les relevés floristiques.
- Un tableau « relevé » où sont consignés le numéro de la publication, le numéro du tableau phytosociologique, le numéro du relevé dans le tableau, et l'identifiant station correspondant.
- Un tableau « station » où sont consignées les informations relatives à chaque station: nom du lieu, coordonnées géographiques dans différents systèmes de coordonnées,

précision de la localisation dans la publication initiale et indication du pays et du département qui ne sont pas toujours renseignées.

- Un tableau « taxon » où sont consignés un identifiant taxon, le nom latin de l'espèce correspondant interprété par les fondateurs de la base, le numéro dans la flore où il a été identifié, et parfois le nom original dans la publication (l'identifiant taxon peut être différent pour un même nom, selon la flore où il a été identifié).
- Un tableau « observation », où sont consignés pour chaque relevé, l'identifiant du taxon, son abondance, et la strate où il se trouve.

Table publication
Id publication
Auteur
Date
Référence

Table relevé
Id publication
Numéro tableau
Numéro relevé
Id station

203 622 relevés
194 174 relevés datés
(1913-2010)

Table station
Id station
Localisation
Latitude
Longitude
Précision
Pays
Département

203 555 stations
161 226 stations géo référencées (système
WGS 84)

Table taxon
Id taxon
Nom latin
Numéro Flore de Fournier
Numéro Flore de Coste
...
Nom original

37666 identifiants taxon
11308 noms latins différents/36 621

Table Observation
Id station
Id taxon
Abondance
Strate

4 369 904 observations
4 369 378 observations avec
l'abondance renseignée

Figure 1: Schéma relationnel de la base de données SOPHY. Id= Identifiant

L'abondance se décompose en sept classes comme il est fait classiquement en phytosociologie:

Tableau 1: Signification des classes d'abondance en phytosociologie (Braun-Blanquet, 1928)

5	3/4 de la surface occupée par l'espèce
4	1/2 - 3/4 de la surface occupée par l'espèce
3	1/4- 1/2 de la surface occupée par l'espèce
2	1/20 - 1/4 de la surface occupée par l'espèce
1	<1/20 de la surface occupée par l'espèce
+	petit nombre- recouvrement négligeable
r	rare

2.1.2 Harmonisation et sélection des données floristiques

Pour cette étude, seuls les relevés situés en France métropolitaine et en Corse, localisés avec certitude et précision dans le système WGS 84 ont été retenus. La date du relevé et la proximité géographique des différents relevés n'ont pas été prises en compte dans la sélection des données. De même, toutes les espèces ont été gardées, y compris les espèces rares. En revanche, les relevés ne possédant qu'une seule espèce et les observations n'ayant pas d'information sur l'abondance ont été écartés. Par ailleurs, les données ont été sélectionnées pour qu'un taxon n'apparaisse qu'une seule fois dans un relevé. Si celui-ci était présent dans deux strates différentes, seule la strate où il était le plus abondant a été gardée. En amont de la sélection des données, un long travail de corrections, et d'ajout d'informations, sur les données a été réalisé pour obtenir des données propres à l'analyse (Annexe I). La taxonomie a été le point le plus délicat à traiter (Annexe I). Après harmonisation de la base de données, un tableau croisant les stations et les espèces a été formé avec à l'intersection la classe d'abondance.

2.1.3 Données environnementales

Afin d'étudier la variation de la composition floristique en fonction des gradients environnementaux, les caractéristiques environnementales correspondant à la localisation géographique des stations de la base SOPHY ont été récupérées. Elles sont issues de la jointure entre des couches cartographiques géo-référencées (couches SIG) renseignant une information environnementale à l'échelle de la France et les coordonnées des stations. La jointure de ces données a été réalisée sur ARCGIS version 10 et R 2.14.2. Les couches environnementales utilisées sont toutes des données qui étaient en libre accès. Les informations suivantes ont été récupérées:

- Données d'altitude de résolution 90m x 90m (Jarvis et al., 2008)
- 19 variables bioclimatiques WorldClim² de résolution 1 km x 1 km, (Hijmans et al., 2005)

Toutes les stations pour lesquelles il manquait une information et les espèces qui n'étaient présentes que sur ces stations ont été écartées des analyses.

2.2 Détermination des grandes unités de végétation

La détermination des grands types de végétation au sein de la base SOPHY a été effectuée par une approche d'analyse de réseaux, qui mathématiquement fait référence à la théorie des graphes.

² Les variables WorldClim représentent les tendances annuelles (moyennes annuelles de température et de précipitation), la saisonnalité (moyenne de l'écart des extrema de températures mensuel, différence des extrema de température sur l'année, iso thermalité, saisonnalité des précipitations et des températures) et les extrema ou les facteurs environnementaux limitant (minimum de température pour le mois le plus froid, maximum de température pour le mois le plus chaud, moyennes des températures pour le trimestre le plus chaud, le plus froid, le plus sec et le plus humide et quantité de précipitations pour le mois le plus chaud et le plus sec, quantité de précipitations pour le trimestre le plus chaud, le plus froid, le plus sec et le plus humide) avec les précipitations en mm et les températures en °C x 10 (Cf annexe II pour plus de précisions).

2.2.1 Définition du réseau de végétation

2.2.1.1 Définition d'un graphe biparti et d'un graphe projeté

A partir d'un tableau de présence-absence croisant les stations et les espèces, on peut définir un réseau de végétation tel qu'une espèce et une station sont liées si l'espèce est présente dans la station. De manière plus formelle, on peut définir le graphe G_b suivant:

$G_b = (S \cup E, U)$ où $U = \{\{e_i, s_j\}, e_i \in E, s_j \in S\}$ **Équation 1: Définition d'un graphe biparti**
avec $S \cup E$ l'ensemble des nœuds du graphe, S l'ensemble des stations, E l'ensemble des espèces, U l'ensemble des liens entre les espèces et les stations, un lien $\{e_i, s_j\}$ n'existant que si l'espèce e_i est présente sur la station s_j .

Le graphe G_b peut être décrit de manière algébrique par une matrice A de dimension $(p \times n)$ où p est le nombre de stations et n le nombre d'espèces. Cette matrice correspond au tableau de présence-absence. A se définit de la manière suivante:

$$a_{ij} = \begin{cases} 1 & \text{si } (e_i, s_j) \in U, \text{ c'est à dire si l'espèce } e_i \text{ est présente sur la station } s_j \\ 0 & \text{sinon} \end{cases}$$

Équation 2: Définition de la matrice A

Le graphe obtenu est un graphe biparti, non pondéré et non orienté (fig.2). Deux graphes projetés peuvent être formés à partir du graphe biparti: un graphe projeté « espèce » et un graphe projeté « station ». Le graphe projeté « espèce » G_e (fig.3b) est défini tel que

$G_e = (E, V)$ où $V = \{\{e_i, e_j\}, (e_i, e_j) \in V^2\}$ **Équation 3: Définition d'un graphe projeté**
avec E l'ensemble des nœuds formé par les espèces et V l'ensemble des liens entre espèces.

Un lien $\{e_i, e_j\}$ existe si les deux espèces e_i et e_j partagent au moins une station en commun et il a un poids égal au nombre de stations que partagent les deux espèces. La matrice A_e de dimension $n \times n$ décrit le graphe projeté « espèce »: $A_e = t(A) \times A$. Le réseau projeté entre stations se définit de manière symétrique.

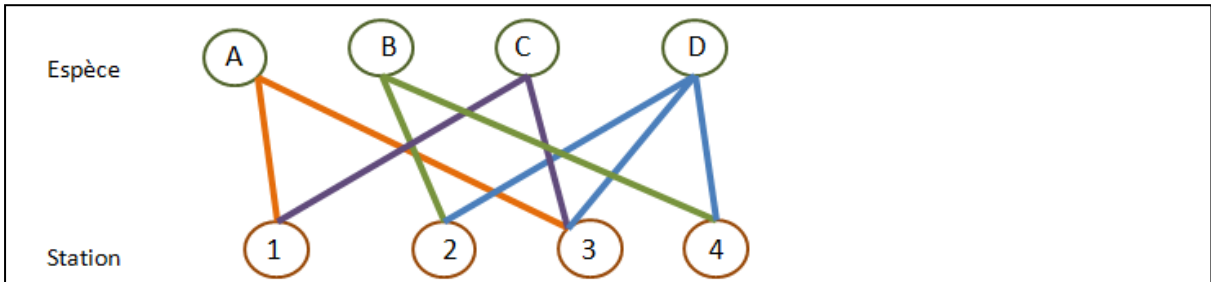


Figure 2 : Schéma d'un graphe biparti théorique de végétation. Le réseau étudié est formé de quatre espèces (A, B, C, D) et de quatre stations (1, 2, 3, 4). Un lien signifie la présence de l'espèce sur la station. Par exemple, l'espèce A est présente sur les stations 1 et 3.

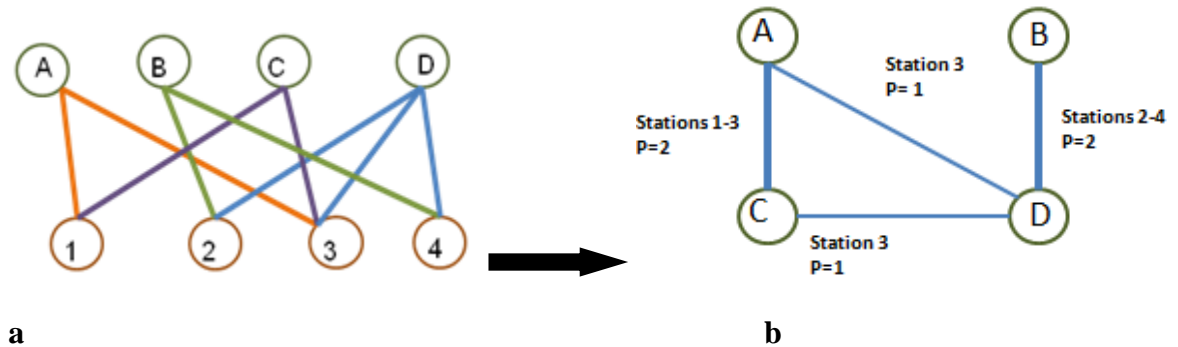


Figure 3 : Schéma du passage d'un graphe biparti (a) à un graphe projeté (b). Ce schéma illustre le passage du graphe biparti de végétation (a), étudié à la figure 2, au graphe projeté espèces (b). Les nœuds de ce dernier sont formés par les espèces et ceux-ci sont interconnectés si les espèces qu'ils représentent se trouvent ensemble sur au moins une station. Ainsi l'espèce A est connectée avec l'espèce C avec un lien pondéré de 2 car l'espèce A est présente sur deux stations (1 et 3) avec l'espèce C. p = poids du lien

2.2.1.2 Prise en compte de l'abondance dans les graphes biparti et projeté

La prise en compte des classes d'abondance des espèces dans le graphe projeté n'est pas une question aisée. Il est en effet possible de pondérer les liens espèce-station du graphe biparti par l'abondance mais son application au réseau projeté déjà pondéré rend la double pondération résultante peu significative écologiquement.

La solution envisagée dans cette étude est de découper chaque espèce en 3 niveaux d'abondance: niveau 1 (classes d'abondance + - 1), niveau 2 (classes d'abondance 2-3), niveau 3 (classes d'abondance 4-5). L'individu statistique n'est alors plus une espèce mais une espèce caractérisée par son niveau d'abondance. On obtient dans les cas les plus complets trois individus « espèces-abondances » pour une espèce. Une nouvelle matrice binaire, s'apparentant à un tableau de présence-absence est obtenue avec les nouveaux individus statistiques (stations x « espèces –abondances »). Les graphes biparti (fig.4) et projeté (fig.5a) sont définis de la même manière que ce qui est décrit dans le paragraphe précédent à la différence près que les nœuds ne sont plus les espèces mais les « espèces-abondances ». Ainsi un lien, dans le graphe biparti, est défini entre une station et des espèces très abondantes, abondantes, peu abondantes et dans le graphe projeté, entre des espèces très abondantes, abondantes, peu abondantes, pondéré par le nombre de stations que les deux « espèces-abondances » partagent. Par ailleurs, notre but étant de découper la végétation en groupe, cette

solution laisse la possibilité qu'une espèce appartienne à plusieurs groupes avec des niveaux d'abondance différents.

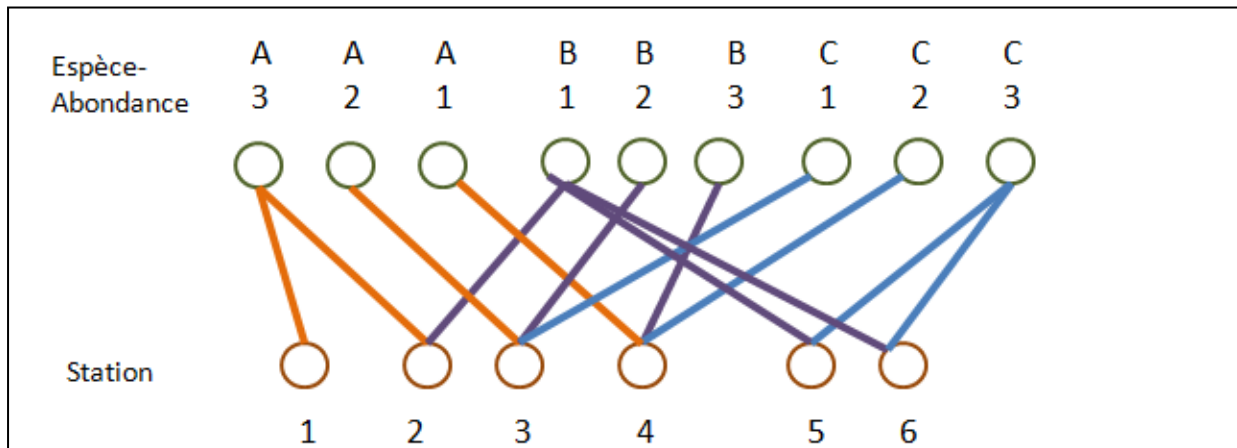


Figure 4 : Schéma d'un graphe biparti théorique avec prise en compte de l'abondance. Le réseau étudié est formé de trois espèces (A, B, C,) découpées en trois classes d'abondance comme précisé dans le texte (ex : A3 = espèce A très abondante) et de six stations (1, 2, 3, 4, 5, 6). Un lien signifie la présence de « espèce-abondance » sur la station. Par exemple, l'espèce A très abondante (A3) est présente sur les deux stations 1 et 2. D'un point de vue écologique, le graphe ci-dessous a été construit de manière à représenter la répartition des espèces le long d'un gradient. L'espèce A et l'espèce C sont les deux espèces qui vivent aux extrémités du gradient de manière très abondantes et leur abondance décroît en s'écartant des extrémités. B au contraire occupe la niche au centre du gradient.

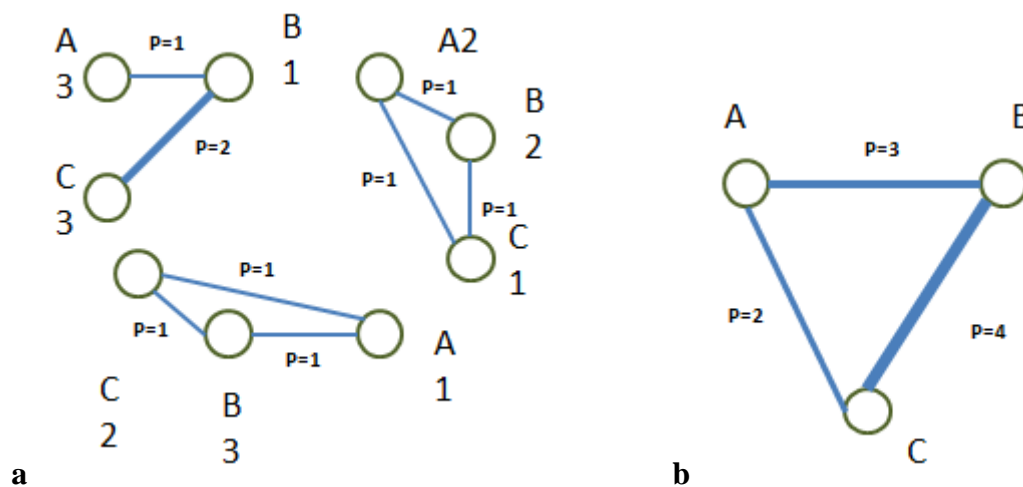


Figure 5 : Schéma comparatif d'un graphe projeté théorique avec prise en compte de l'abondance (a) et d'un graphe projeté sans prise en compte de l'abondance (b). On s'aperçoit de la différence de structure qu'engendre la prise en compte de l'abondance par les « espèces-abondances ». En effet des ensembles de nœuds non connexes apparaissent. p =poids du lien

2.2.2 Détermination d'ensembles cohérents d'espèces et de stations

2.2.2.1 Analyse de modularité au sein du réseau projeté « espèce »

A partir du graphe projeté « espèce-abondance », on recherche les « espèces-abondances » qui sont densément connectées entre elles, donc qui sont présentes ensemble sur de nombreuses stations, de manière à établir les grands types de végétation.

L'analyse de modularité d'un réseau permet de mettre en évidence l'existence de groupes densément connectés de nœuds (modules) au sein d'un réseau. La modularité Q d'une partition est une mesure de la qualité de la partition au sein du réseau. Elle a été définie comme la fraction des liens dans le réseau qui se trouve au sein des modules, moins la fraction que l'on aurait si les connexions (liens) étaient réparties au hasard (Newman and Girvan, 2004):

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{i,j} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad \text{Équation 4: Définition de la modularité}$$

avec $A_{i,j}$ le poids du lien entre le nœud i et le nœud j , $k_i = \sum_j A_{i,j}$ est la somme des poids des liens attachés au nœud i , c_i est la communauté à laquelle le nœud i est assigné, et la fonction δ est définie ainsi: $\delta(u, v)$ vaut 1 si $u=v$, 0 sinon et $= \frac{1}{2} \sum_{i,j} A_{ij}$.

La modularité est comprise entre -1 et 1. Il existe des modules au sein du réseau si la modularité est positive. Dans le cadre d'une analyse de modularité, la recherche d'une partition optimale dans les réseaux correspond à rechercher la partition de modularité maximale. L'algorithme utilisé dans cette étude pour la détection des modules est l'algorithme développé par Blondel et al (Blondel et al., 2008). Cet algorithme a été appliqué sous Matlab 7.3.

Remarque: Des tests d'analyse de modularité sur un graphe de même nature avec permutations aléatoires des liens entre les nœuds n'ont pas pu être réalisés du fait de limitation calculatoire. La significativité de la modularité n'a donc pas pu être calculée.

2.2.2.2 Spécificité des stations pour chaque module

L'analyse de modularité permet de déterminer des groupes d'« espèces-abondances » mais il est possible, pour chaque module, de lister les stations où sont présentes les « espèces-abondances » de ce module et de former ainsi des groupes de stations. Cependant, une station n'a pas forcément toutes ses espèces dans le même module. Une station peut être rattachée à un ou plusieurs modules selon la distribution des espèces de cette station à travers les modules. Le degré de spécificité des stations à un module dépend de l'isolement du module dans le réseau projeté: certains modules partagent plus de liens avec les autres modules que d'autres. Ces différences de structure permettent de discriminer les propriétés écologiques des grands types de végétation mis en évidence (Guimera and Amaral, 2005).

Afin de retenir les stations les plus spécifiques d'un module, un critère a été mis au point: les stations sélectionnées doivent avoir 80% et plus de leur couverture végétale associée au module considéré. Le pourcentage de couverture est calculé à partir des classes d'abondance initiales. Pour une station, chaque classe d'abondance (associée à une espèce) est attribuée à un taux de couverture selon la méthode phytosociologique (+ = 0.01, 1=0.1, 2=0.3, 3=0.5, 4=0.7, 5=0.9, cf tableau 1). On calcule le pourcentage de couverture atteint dans le module j , pour une station possédant n espèces, dont n_j dans le module j de la manière suivante:

$$\frac{\sum_{i=1}^{n_j} c_i}{\sum_{i=1}^n c_i} = \frac{\text{taux de couverture de la station dans le module } j}{\text{taux de couverture total de la station}}$$

avec c_i le taux de couverture de l'espèce i sur la station

Équation 5: pourcentage de couverture atteint par une station dans un module

Le taux de couverture total de la population végétale sur une station peut ne pas être de 100%, d'où la division par ce taux. Cette analyse modulaire et le filtrage des stations nous ont permis de définir des modules précis qui peuvent être rattachés à un type de végétation bien défini. Le module « végétation méditerranéenne », regroupant les espèces herbacées et ligneuses préférentiellement associées au climat méditerranéen, a fait l'objet de la suite de l'étude.

2.3 Etude de la végétation dans le module « végétation méditerranéenne »

2.3.1 Ordinations non contraintes et typologie bioclimatique

2.3.1.1 Deux méthodes d'ordination non contrainte

Dans le but d'étudier la structure des communautés végétales au sein de la végétation méditerranéenne, une Analyse Factorielle des Correspondances (AFC) a été réalisée avec le package FactoMineR, (version R.14.2.2) sur les données du module 5, soit un tableau de présence/absence avec en colonne les « espèces-abondances » de ce module et en ligne les stations caractéristiques associées. En réalisant cette analyse, on cherche à étudier la liaison entre les « espèces-abondances » et les stations. En termes de profils, les questions se posent ainsi: peut-on considérer que toutes les stations du module 5 ont le même profil floristique ou au contraire, certaines stations accueillent des communautés d'espèces végétales particulières ? De manière symétrique, peut-on considérer que toutes les « espèces-abondances » ont un profil spatial similaire, ou existe-il une répartition particulière des « espèces-abondances » au sein de l'espace méditerranéen? Si oui, quelles sont les « espèces-abondances » qui se ressemblent en termes de profil, c'est à dire quelles sont les communautés d'espèces au sein du jeu de données ?

Remarque: Le choix a été fait ici de garder les individus « espèces-abondances » (c'est à dire la catégorisation à 3 niveaux utilisée pour l'analyse de graphe) et de travailler sur un tableau de présence-absence, plutôt que recréer un tableau espèces x stations. Ce choix a été pour des soucis de cohérence avec les analyses précédentes mais également car les abondances sont en classes et non pas en effectif.

L'AFC est une méthode très classique pour analyser des données d'inventaire écologique du type stations x espèces. Cependant, une autre méthode dérivant de l'AFC, est proposée par certains écologues (Chessel and Gimaret, 1997, Gimaret-Carpentier et al., 1998, Péliissier et al., 2003) pour analyser ce genre de données: l'Analyse Non Symétrique des Correspondances (ANSC) (Annexe III). Cette analyse présente plusieurs avantages, notamment d'interpréter l'inertie totale comme étant la diversité β^* des données floristiques.

Dans cette étude, une ANSC a été réalisée sur le même tableau que l'AFC afin d'identifier éventuellement d'autres tendances du jeu de données qui ne sont pas mises en évidence dans l'AFC. Cette analyse a été réalisée sur Matlab 7.3, avec un script développé par P. Couteron, directeur de l'UMR AMAP.

2.3.1.2 Typologie bioclimatique des stations

Les dix-neuf variables bioclimatiques (Annexe II) issues des données WorldClim et la variable topographique, caractérisant les stations, ont permis d'étudier les différentes tendances bioclimatiques au sein du module « végétation méditerranéenne ». Une Analyse en

Composantes Principales (ACP) sur ces données environnementales centrées-réduites a été réalisée avec le package FactoMineR, dans R 14.2.2. Les composantes principales de l'AFC et de l'ANSC ont été placées en variables supplémentaires afin d'apporter un éclairage sur l'ordination des espèces et des stations en lien avec les gradients environnementaux.

2.3.2 Ordination contrainte par l'environnement bioclimatique et classification

Une analyse non symétrique des correspondances sur variables instrumentales (ANSCVI) a été réalisée avec Matlab 7.3. L'ordination a été contrainte par les 19 variables bioclimatiques de WorldClim, la variable topographie et ces mêmes variables au carré pour mettre en évidence leurs éventuels effets quadratiques.

Une classification des stations a été réalisée à partir des coordonnées factorielles de l'ensemble des axes de l'analyse précédente. La classification utilisée est une classification par partition fondée sur la méthode des k-medoids. L'approche PAM (Partition around medoids) n'étant pas appropriée pour les gros jeux de donnée car trop coûteuse en termes de calculs, c'est l'approche CLARA (Clustering LARge Application (Kaufman and Rousseeuw, 1990)) qui a été utilisée avec le package cluster dans R 2.14.2.

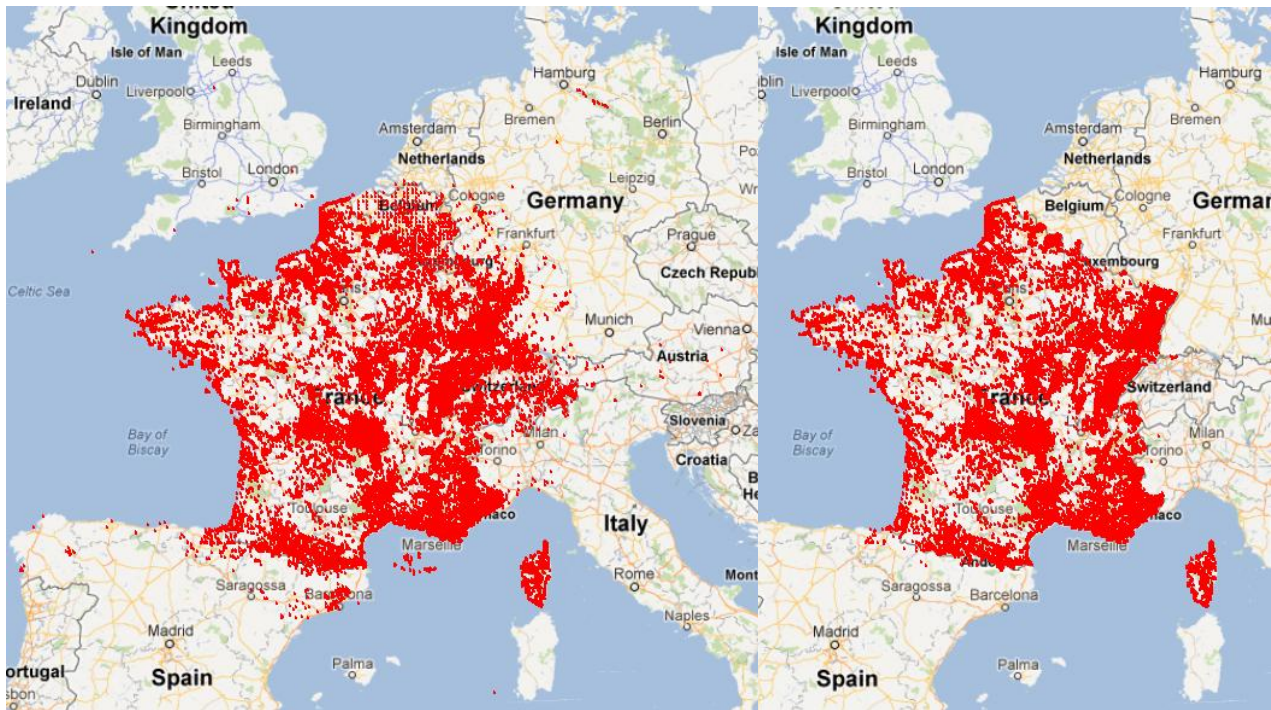
A l'issue de la classification, les « espèces-abondances » présentes exclusivement un groupe ont été identifiées. Ce sont les espèces caractéristiques de chaque groupe au sens phytosociologique du terme.

3. RESULTATS

3.1 Caractérisation du jeu de données

3.1.1 Les données après sélection

Après l'ensemble des traitements réalisés et de filtres de sélection sur les données floristiques, le jeu de données se compose de 131 534 stations (fig.6a), 26 003 identifiants taxon soit 5 259 taxons différents et 2 597 945 observations. Après suppression des stations pour lesquelles il y avait des données environnementales manquantes, on obtient un jeu de données de 126 335 stations (fig.6b), 5223 taxons et 2 534 242 observations.



a **b**
Figure 6: Visualisation des stations SOPHY avant (a) et après sélection des données (b)

3.1.2 Chiffres clés sur les espèces et les stations

En moyenne chaque relevé comprend 20 espèces avec une médiane à 17. Deux taxons sont présents au minimum (par sélection des données) et le relevé avec le plus de taxons comprend 139 taxons. Inversement, en moyenne, une espèce est présente dans 494 relevés avec une médiane à 66. Cela témoigne d'une disparité forte avec un groupe d'espèces qui sont très fréquentes et des espèces très peu fréquentes. Une espèce au minimum est dans un seul relevé (382 espèces dans ce cas) et l'espèce la plus présente (*Hedera helix* ou Lierre) est présente sur 18 930 stations. Par ailleurs, on observe sur la figure 7, que la côte atlantique est caractérisée globalement par des relevés avec un nombre d'espèces inférieur à la moyenne alors que les relevés qui ont une richesse spécifique élevée se situent dans l'est de la France avec des zones remarquables en Alsace et en Méditerranée avec des relevés de plus de soixante espèces (fig.7, points bleus sur la carte).

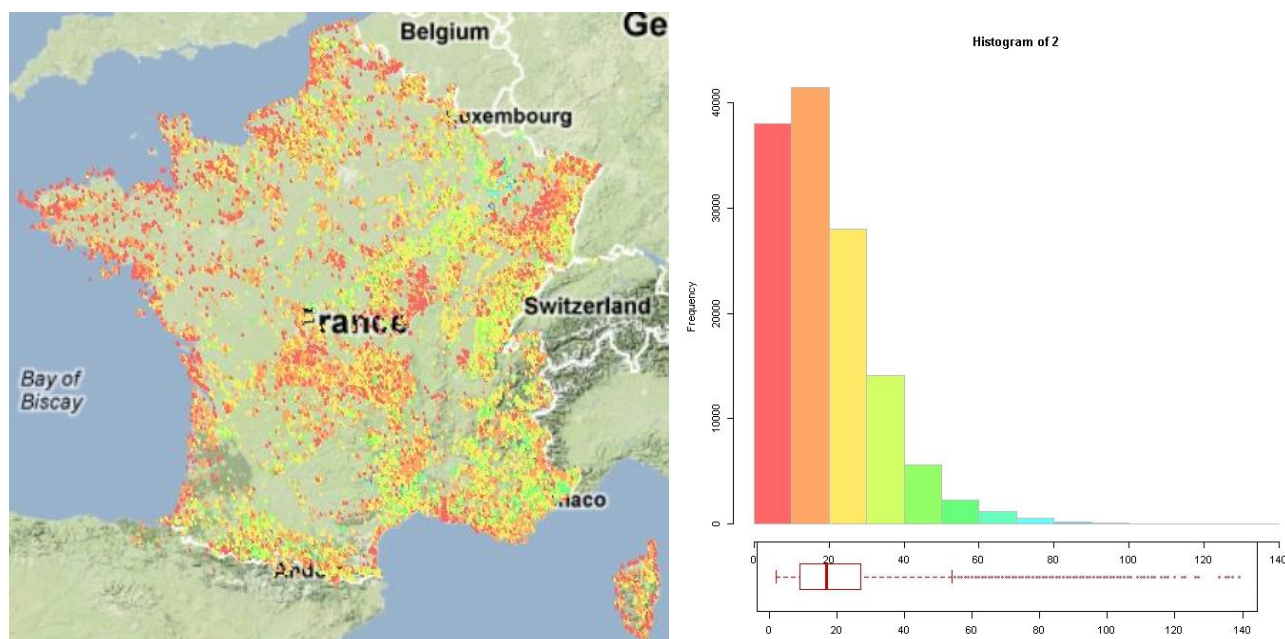


Figure 7 : Visualisation de la localisation de la diversité spécifique par station

3.2 Résultats de l'analyse de modularité

3.2.1 Le découpage des espèces en trois grandes classes d'abondance

Le découpage de chaque espèce en trois catégories selon l'abondance donne 10 829 individus différents, ce qui révèle que de nombreux taxons ne sont pas présents dans les trois catégories d'abondance dans le jeu de données. En effet 1399 taxons (sur un total de 5 223 taxons) ne sont présents que dans une seule catégorie d'abondance, et sur ces 1399 taxons, 1335 sont des taxons présents uniquement dans la catégorie 1 soit la catégorie de faible abondance. Nous avons donc des taxons qui sont toujours faiblement abondants, quel que soit les relevés présents dans la base.

Inversement, sur 126 335 relevés ou stations, 8 185 relevés possèdent des taxons que dans une seule catégorie d'abondance dont 94% sont des relevés avec des taxons présents dans la catégorie 1. Il y a donc des milieux qui sont peu couverts par des espèces végétales ou avec beaucoup d'espèces faiblement recouvrantes.

Au total, 2 073 120 observations sont attribuées à la classe 1, 394 790 sont attribuées à la classe 2 et 66 332 sont attribuées à la classe 3. Il y a donc sur l'ensemble du jeu de données, beaucoup plus d'espèces peu abondantes dans un milieu que d'espèces très abondantes.

3.2.2 Résultats de l'analyse de modularité

Après 12 itérations, la modularité trouvée est de $Q=0,3765$. Il y a donc bien des communautés présentes dans le graphe projeté « espèce ». Huit modules ont été formés, chacun ayant respectivement 2975, 1989, 1704, 1 516, 1 316, 734, 376 et 219 « espèces-abondances » et avec respectivement 41 899, 32 769, 74 344, 59 614, 58 166, 61 320 et 8 270 stations comprenant les « espèces-abondances » des modules.

D'après la liste de ces espèces par module, il est possible pour un expert en botanique de discerner les grandes tendances suivantes au sein des modules:

Module 1: Milieux ouverts à nombreuses annuelles et géophytes*.

Module 2: Pelouses écorchées et rocailles/rochers d'altitude.

Module 3: Communautés herbacées, arbustives et arborées de forêts, lisières et clairières (hors méditerranéen).

Module 4: Prairies et landes humides, tourbières et bords d'étangs.

Module 5: Communautés de forêts, garrigues et milieux ouverts méditerranéens.

Module 6: Prairies mésophiles*.

Module 7: Pelouses sèches et landes calcicoles.

Module 8: Groupements halophytes*.

Cependant, l'analyse de la répartition des 126 335 stations dans les modules a montré que les stations sont présentes dans plusieurs modules (tableau 2).

Tableau 2: Nombre de stations en commun par module

	Module 1	Module 2	Module 3	Module 4	Module 5	Module 6	Module 7	Module 8
Module 1	41899							
Module 2	9681	32769						
Module 3	18485	23075	74344					
Module 4	16094	12430	33321	59614				
Module 5	25699	22201	38953	19232	58166			
Module 6	25473	21451	41742	31425	37260	61320		
Module 7	15181	18305	26909	13133	30595	27645	36036	
Module 8	4584	39	208	3069	1133	1535	179	8270

3.2.3 Stations caractéristiques des modules

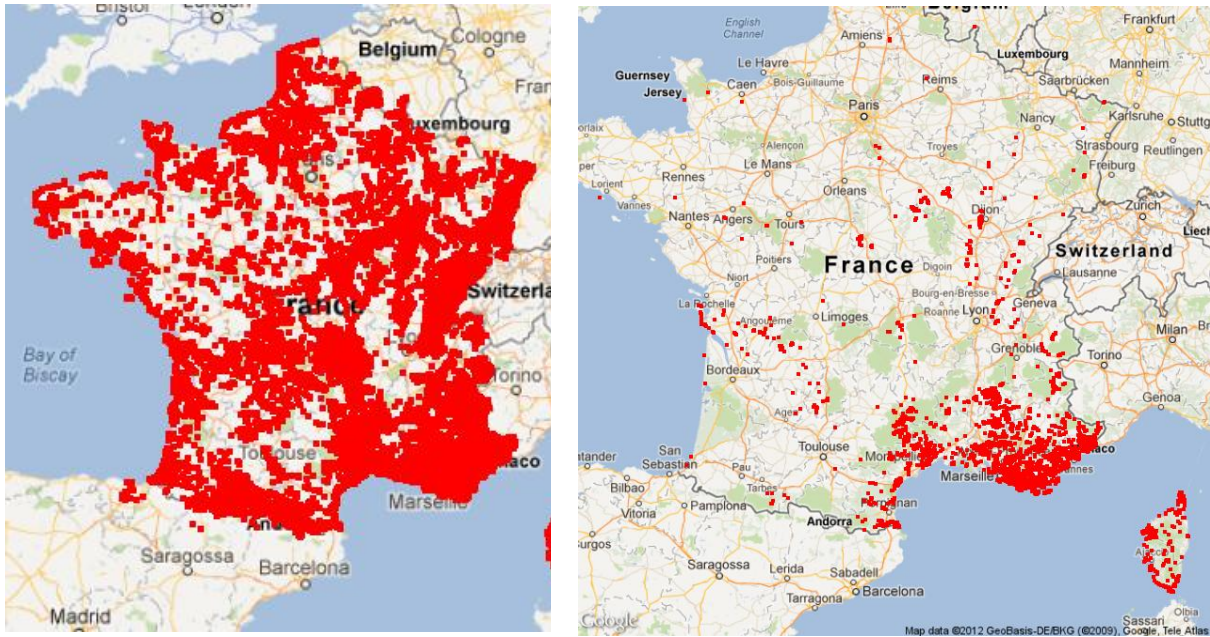
Le critère d'attribution des stations caractéristiques des modules a montré que 61,7% des stations avaient 80% de leur couverture dans un module précis. En décomposant par module, on peut observer que certains modules ont très peu de stations caractéristiques (tableau 3), comme le module 7. Il y a par ailleurs 38,3 % de stations ne sont pas associées de manière caractéristiques à un module. Le module 8 se détache nettement des autres avec 51,7% de stations caractéristiques.

Tableau 3: Pourcentage de stations caractéristiques par module

Module	1	2	3	4	5	6	7	8
% de stations	19,0	15,1	36,7	33,1	14,2	8,1	0,4	51,7

3.2.4 Le module 5: « végétation méditerranéenne »

Le module 5 a fait l'objet d'une analyse plus détaillée dans la suite de cette étude. Une étude similaire aurait pu être réalisée sur les autres modules. Le module 5, initialement composé de 1316 « espèces-abondances » et 58 166 stations (fig.8a), se compose après l'étude des stations caractéristiques à 8 287 stations (fig.8b) et 1176 « espèces-abondances » car certaines « espèces-abondances » n'étaient que sur les stations supprimées.



a **b.**
Figure 8: Projection des stations du module 5 avant (a) et après filtrage des stations caractéristiques (b)

Les stations caractéristiques du module 5 ont au minimum une « espèce-abondance » présente dans le module 5, cela signifie que celle-ci couvre 80% de la station. Certaines stations peuvent atteindre 89 « espèce-abondances » et en moyenne, 27 « espèce-abondance » co-occurrent sur les stations de ce module. La richesse spécifique moyenne est donc un peu au-dessus de la moyenne sur la France. Les stations avec le plus de diversité sont situées sur la côte est méditerranéenne avec notamment une zone remarquable en région PACA (fig.9).

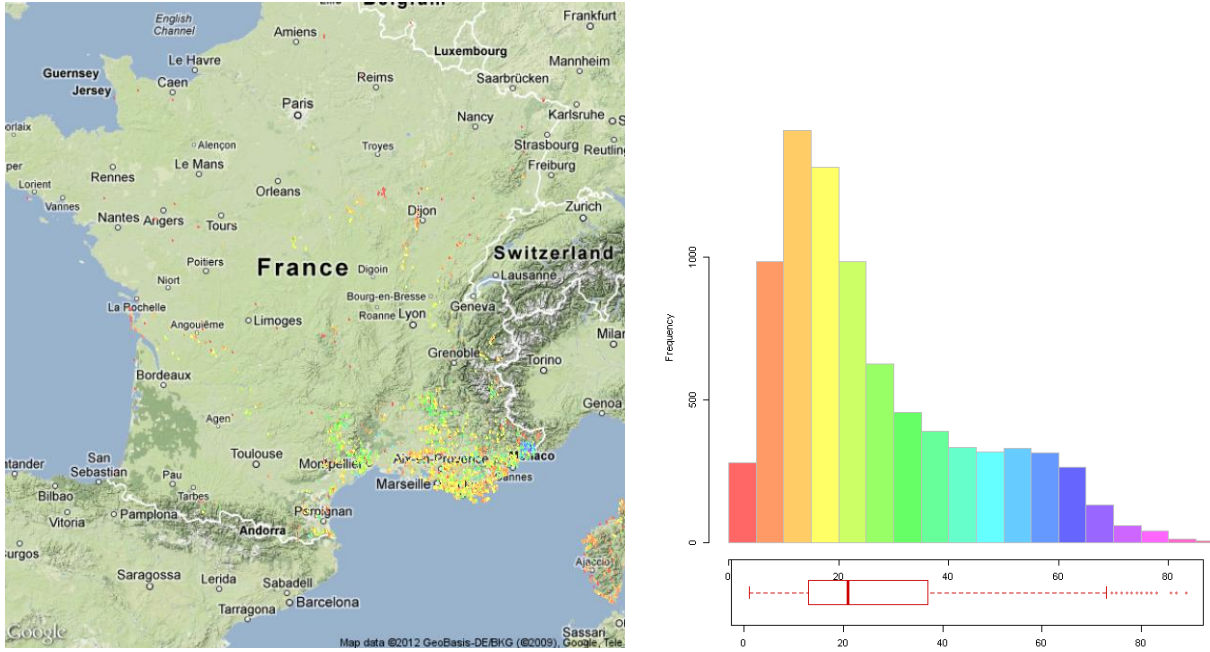


Figure 9: Répartition des stations avec indication de leur diversité spécifique

Dans les espèces les plus abondantes dans ce module, on note bien des espèces spécifiques, ou très présents dans la région méditerranéenne: le thym vulgaire (*Thymus vulgaris*), présent sur 50,8% des stations tous niveaux d'abondance confondus, la Garance voyageuse (*Rubia peregrina*) (49,7), Chêne vert (*Quercus ilex*) (44%), Brachypode rameux (*Brachypodium retusum*) (38,7%), Pin blanc de Provence ou Pin d'Alep (*Pinus halepensis*) (36,4%), Aphyllanthe de Montpellier (*Aphyllanthes monspeliensis*) (35%), Asperge sauvage (*Asparagus acutifolius*) (34%), Romarin (*Rosmarinus officinalis*) (32%), Salsepareille (*Smilax aspera*) (31,7%), Pistachier lentisque (*Pistacia lentiscus*) (31,6%).

Au niveau bioclimatique, les stations du module 5 sont caractérisées par des températures moyennes annuelles plus chaudes que pour l'ensemble des stations du jeu de données: 12,6°C en moyenne pour le module 5 contre 10°C pour le jeu de données, avec un écart de 3°C plus chaud sur les trimestres les plus chauds et les plus froids. Les précipitations au niveau des stations du module 5 sont inférieures à la moyenne (moyenne: 864 mm) avec un écart de 83 mm. La répartition des précipitations est différente de la moyenne avec une pluviométrie un peu supérieure à la moyenne pour le mois le plus humide (3 mm d'écart, moyenne: 96 mm), mais une pluviométrie inférieure à la moyenne pour le mois le plus sec (23 mm d'écart, moyenne: 49 mm). Ces données correspondent bien au climat méditerranéen avec des températures plutôt douces et une pluviométrie très contrastée avec une période estivale très sèche et des pluies intenses aux intersaisons notamment en automne.

3.3 Résultats des différentes méthodes d'ordination

3.3.1 Analyses d'ordination non contrainte

L'inertie associée à l'AFC est de 61.3. Le diagramme des valeurs propres (Annexe IV, fig.b) nous conduit à considérer, en premier lieu les deux premiers axes dont l'inertie est nettement plus importante que les autres. Ces deux axes ont une inertie relativement élevée (0,56 et 0,55) et des associations entre les stations et les « espèces-abondances » devraient apparaître. Après analyse, il s'avère que l'axe 1 (Annexe IV, fig.d et e) est le reflet de communautés végétales très spécifiques de certains lieux en Alpes-Maritimes et Alpes de Haute Provence. Quinze « espèces-abondances » seulement ont une contribution de plus de 1 (Annexe IV, fig.f), avec le maximum atteint par *Sedum fragrans* niveau 1 et *Moehringia intermedia* niveau 1 avec une contribution de 11. La première espèce est une espèce des roches et grottes humides spécifique de la zone décrite et la deuxième est une espèce menacée des montagnes calcaires du Sud-Est de la France. Dans les espèces qui contribuent le plus à cet axe, On trouve par exemple trois espèces spécifiques des communautés des falaises liguro-apennines (*Primula allionii*, *Ballota frutescens*, *Potentilla saxifraga*).

En revanche, l'axe 2 met en évidence un gradient de végétation nord-sud comme en témoigne la figure 10 et 11. Les stations situées au nord ont dans l'ensemble des coordonnées positives et les stations situées sur le littoral méditerranéen et en Corse ont dans l'ensemble des coordonnées négatives. En analysant les espèces qui ont une forte contribution à l'axe 2 (contribution > 1, maximum de 1,8 pour *Asparagus acutifolius* ou l'asperge sauvage), on met en évidence des espèces spécifiques du milieu méditerranéen (*Smilax aspera* ou Salsepareille, *Myrtus communis* ou Myrtes commun, *Asparagus acutifolius* ou l'asperge sauvage, *Phillyrea angustifolia* ou Filaire à feuille étroite (proche de l'olivier, *Lonicera implexa* ou chèvrefeuille des Baléares, *Rubia peregrina*, *Pistacia lentiscus*, *Calicotome spinosa* ou Calicotome épineux, *Clematis flammula* ou clématite odorante...) qui s'oppose à des espèces thermophiles mais qui peuvent être présentes au-delà de la zone méditerranéenne, (*Fumana procumbens* ou Fumana à tiges retombantes, *Teucrium montanum* ou Germandrée des

montagnes, *Potentilla crantzii* ou Potentille de Crantz, *Koeleria vallesiana* ou Koelérie du Valais, *Festuca rubra* ou fétuque diffuse, *Bromus erectus* ou Brome dressé, *Hieracium pilosella* ou Piloselle, *Eryngium campestre* ou Panicaut champêtre..).

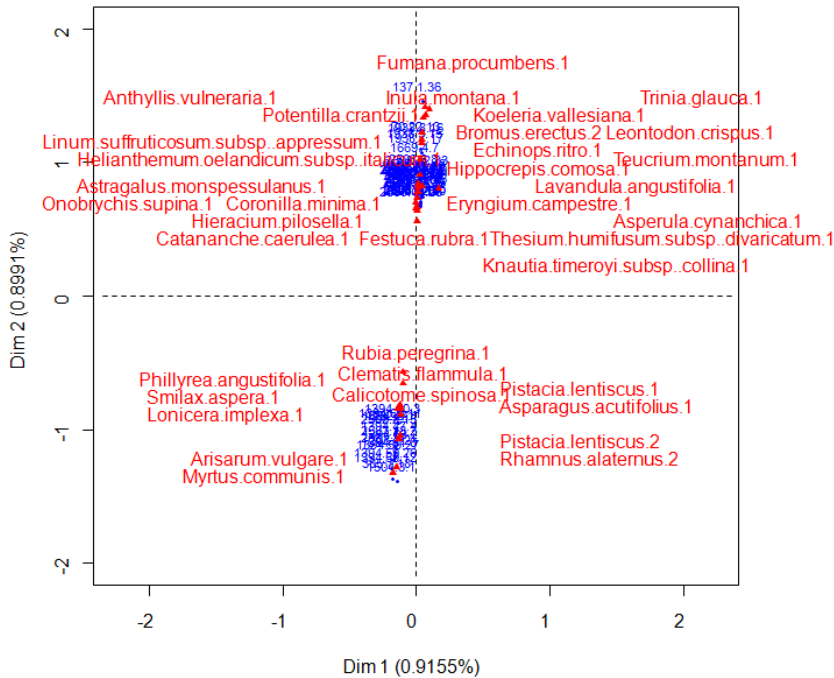


Figure 10: Premier plan factoriel de l'AFC avec zoom sur l'axe 2

Les individus de cette figure ont été sélectionnés sur la qualité de leur représentation sur l'axe 2 ($\cos^2 > 0,1$)

Figure 11 : Cartographie du deuxième axe factoriel de l'AFC

Les points représentent les stations et leur couleur indique leur position sur l'axe 2 (Rouge : coordonnée factorielle positive, blanc : coordonnée proche de zéro, bleu : coordonnée négative. Couleur intense rouge ou bleu : coordonnée élevée en valeur absolue.

Dans l'ANSC, l'inertie expliquée par la partition selon les relevés est de 0,032. La différence entre les relevés est principalement décrite par l'axe 1 qui a une valeur propre associée nettement supérieure aux autres axes (Annexe V, fig.c). Les deux premiers axes ayant des valeurs propres plus élevées que les autres, nous nous limiterons à leur étude. L'axe 1 montre dans l'ensemble une opposition entre les stations situées au nord et certaines stations situées sur le littoral méditerranéen particulièrement sur la Côte-d'Azur, en Camargue, en Corse et avec quelques points dans la région de Montpellier et de Perpignan (fig.13).

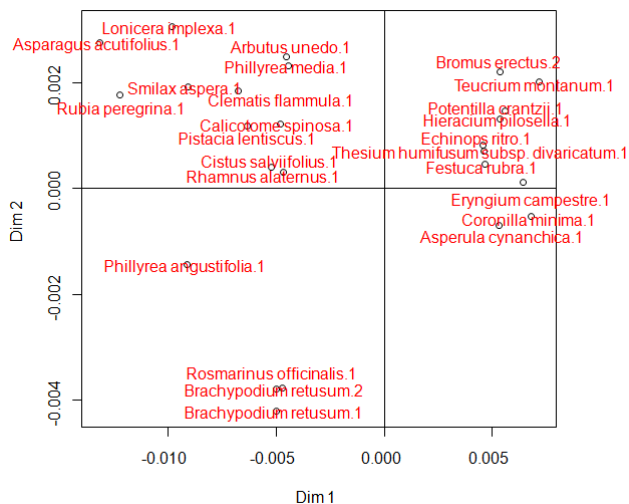


Figure 12: Premier plan factoriel de l'ANSC (axe 1)

Seules les « espèces-abondances » contribuant à plus de 0,01 à l'axe 1 ont été représentées



Figure 13 : Cartographie du premier axe factoriel de l'ANSC

Les points représentent les stations et leur couleur indique leur position sur l'axe 1 (Rouge : coordonnée factorielle positive, blanc : coordonnée proche de zéro, bleu : coordonnée négative. Couleur intense rouge ou bleu : coordonnée élevée en valeur absolue.

L'axe 2 montre lui une opposition entre des stations situées sur littoral marseillais, en Provence et autour de Montpellier avec des stations situées dans l'espace méditerranéen et un peu partout en France (fig.15). Sur les deux premiers axes les contributions des espèces et des stations s'échelonnent relativement régulièrement, confirmant que les tendances observées ne sont pas induites par quelques individus particuliers comme en AFC. En regardant les espèces qui contribuent le plus à l'axe 1 (fig.12), on retrouve les espèces qui contribuent le plus à l'axe 2 de l'AFC, avec la même séparation d'espèces très spécifique à l'espace méditerranéen et d'autres espèces qui ne le sont pas spécialement. En regardant les espèces qui contribuent le plus à l'axe 2 (fig.14), il y a une opposition entre les espèces présentes dans les garrigues provençales calcaires, -plus précisément des espèces thermophiles poussant sur calcaire basique- (*Brachypodium retusum*, *Fumana ericoides subsp. Montana*, *Thymus vulgaris*, *Rosmarinus vulgaris*, *Lavandula latifolia*, *Pinus halepensis*, *Ononis minutissima*...) et les autres espèces qui peuvent être présentes dans toute la France (ex: *Daucus carotta*, ie la carotte sauvage) ou des espèces méditerranéennes qui ne sont pas spécifiques des garrigues (ex: *Smilax aspera*, qui pousse plutôt dans les haies et les bois).

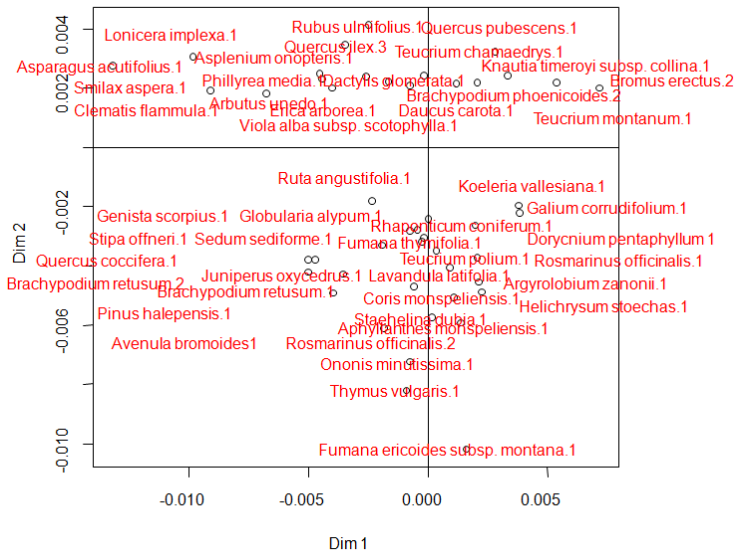


Figure 14: Premier plan factoriel de l'ANSC (axe 2)

Seules les « espèces-abondances » contribuant à plus de 0,01 à l'axe 2 ont été représentées

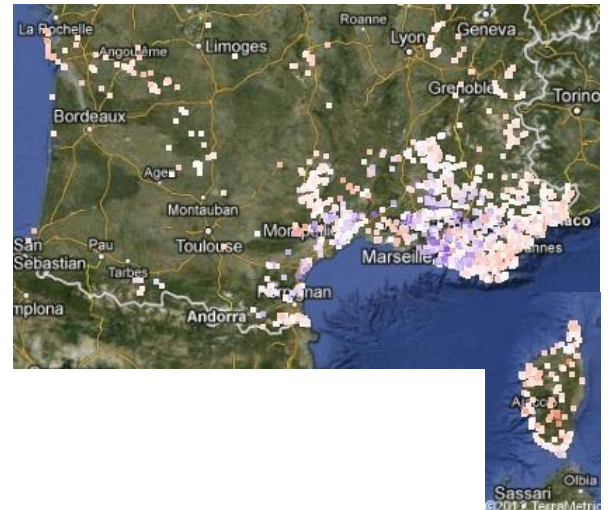


Figure 15 : Cartographie du deuxième axe factoriel de l'ANSC

Les points représentent les stations et leur couleur indique leur position sur l'axe 2 (Rouge : coordonnée factorielle positive, blanc : coordonnée proche de zéro, bleu : coordonnée négative. Couleur intense rouge ou bleu : coordonnée élevée en valeur absolue.

3.3.2 Typologie bioclimatique des stations

L'inertie expliquée par le premier axe est de 58,77% et le deuxième axe de 24,87%. Ainsi ces deux facteurs expriment 83,4% de l'inertie totale (Annexe VI). On se limitera donc à ces deux facteurs. A partir de deux variables synthétiques, on résume presque toute l'information apportée par les 14 variables bioclimatiques (description des variables WorldClim cf Annexe II) et la variable topographique (variable « mnt » pour MNT). On peut vérifier que les qualités de projections sont bonnes pour l'ensemble des variables et des individus.

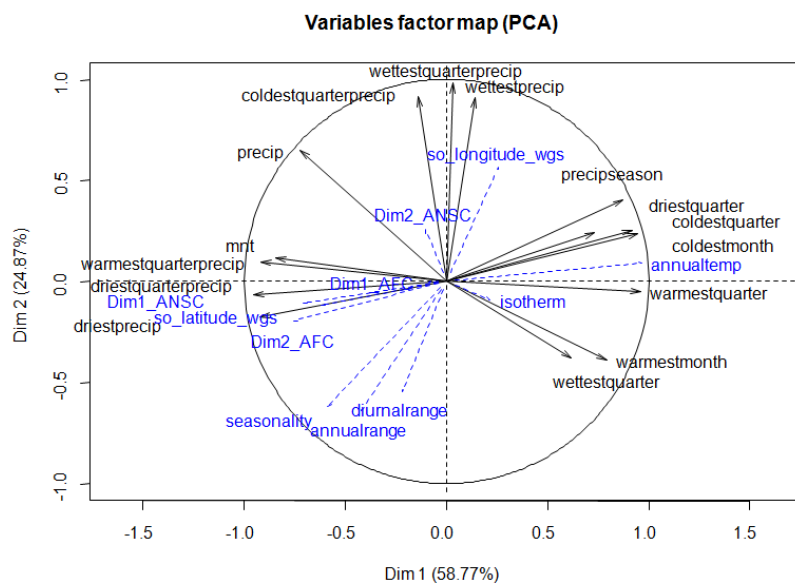
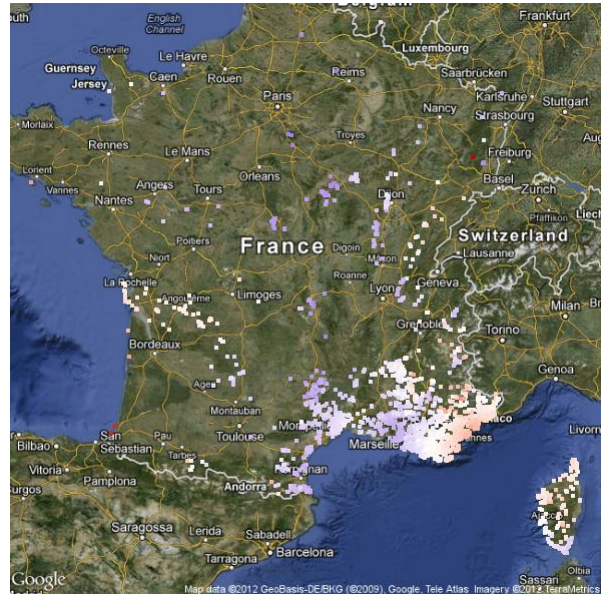
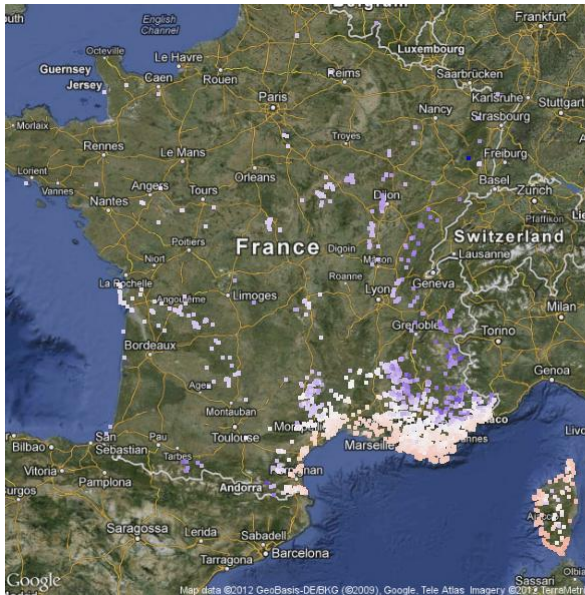


Figure 16: Graphe des variables du premier plan factoriel de l'ACP

Les variables de température sont toutes bien corrélées positivement à l'axe 1 (fig.16), à l'exception de la maximale de température pour le mois le plus chaud et la moyenne de température pour le trimestre le plus humide qui le sont un peu moins. La variable supplémentaire « moyenne annuelle de la température (annualtemp) » vient confirmer la corrélation positive de la température à l'axe 1. Par ailleurs, la quantité de précipitation pour les trimestres les plus chauds et les plus secs, la quantité de précipitation pour le mois le plus sec et la topographie sont inversement corrélées à l'axe 1 (fig.16). Les stations aux coordonnées factorielles positives sur l'axe 1 ont donc des températures plus élevées que la moyenne sur l'ensemble de l'année et des précipitations plus faibles que la moyenne sur les périodes les plus chaudes et les plus sèches. Au contraire, les stations aux coordonnées négatives sont des stations « froides », plus hautes en altitude, avec une pluviométrie plus importante que la moyenne sur leurs périodes les plus chaudes et les plus sèches. Cet axe traduit l'opposition des stations au climat méditerranéen, marquées par des températures douces toute l'année et des périodes de très faibles précipitations durant la période estivale et les stations sur le reste de la France ou en altitude où il fait plus froid avec des quantités de précipitations plus importantes. La latitude en variable supplémentaire et la visualisation des coordonnées factorielles (fig.17 a) sur la carte du premier axe factoriel confirment cette opposition. On remarque que la dimension 1 de l'ANSC et la dimension 2 de l'AFC sont corrélées négativement à l'axe 1 de cette ACP, montrant le lien entre le climat méditerranéen et la répartition de la végétation.

La quantité de précipitation sur les périodes les plus humides et les plus froides est corrélée positivement à l'axe 2, qui montre que ces précipitations ne sont pas liées aux précipitations de la période estivale. La visualisation des coordonnées factorielles sur la carte de ce deuxième axe factoriel met en évidence des différences entre les stations (fig.17 b) au sein de l'espace méditerranéen qui oppose notamment les Alpes-Maritimes aux coordonnées plutôt positives à la région de Montpellier et Marseille aux coordonnées plutôt négatives. Ainsi les Alpes-Maritimes, sans doute du fait des massifs montagneux de cette région, reçoivent plus de précipitation pendant les intersaisons et l'hiver contrairement au littoral provençal et languedocien.



a **b**
Figure 17: Cartographie du premier (a) et deuxième (b) axe factoriel de l'ANSCVI

Les points représentent les stations et leur couleur indique leur position sur les axes (Rouge: coordonnée factorielle positive, blanc: coordonnée proche de zéro, bleu: coordonnée négative. Couleur intense rouge ou bleu: coordonnée élevée en valeur absolue.

3.3.3 Ordination contrainte par l'environnement bioclimatique et classification

L'inertie de l'ANSCVI est de 0.0038. Les variables bioclimatiques expliquent donc 11,7% de la variabilité de la composition floristique au sein du jeu de données. Le premier plan factoriel explique 54% de la variabilité des données (Annexe VII). Dans ce premier plan, deux unités apparaissent dans le nuage des stations. En visualisant les coordonnées factorielles des deux axes sur une carte (fig.18), on met en évidence que l'axe 1 oppose des stations du littoral méditerranéen (coordonnées négatives) aux stations de l'arrière pays méditerranéen et les stations plus septentrionales. L'axe 2 oppose les stations des Alpes-Maritimes (coordonnées positives) avec les autres stations (coordonnées négatives).



a

b

Figure 18 : Cartographie du premier (a) et deuxième (b) axe factoriel de l'ANSCVI

Les points représentent les stations et leur couleur indique leur position sur les axes (Rouge : coordonnée factorielle positive, blanc : coordonnée proche de zéro, bleu : coordonnée négative. Couleur intense rouge ou bleu : coordonnée élevée en valeur absolue.

La classification CLARA sur l'ensemble des axes factoriels de l'ANSCVI montre plusieurs paliers où l'inertie intra-groupe ne diminue plus beaucoup avec augmentation du nombre de groupes (Annexe VII). On identifie un premier palier entre 5 et 6 groupes puis entre 9 et 10 groupes, suggérant une hiérarchie possible des groupements végétaux. On choisit ici d'analyser le découpage à 10 groupes (fig.19). L'étude des espèces caractéristiques de chaque groupe de stations montre, d'après les experts, des espèces caractéristiques associées très cohérentes au regard des communautés végétales décrites dans la région.

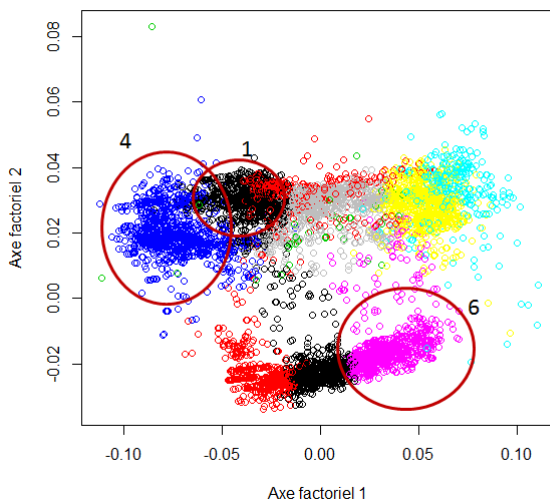


Figure 19: Premier plan factoriel de l'ANSCVI

Les couleurs correspondent aux groupes issus de la classification

Ainsi par exemple, le groupe 4, groupe de coordonnées négatives sur l'axe 1 et positif sur l'axe 2, est associée à des espèces présentes sur la Côte-d'Azur, en Corse et dans les Albères (fig.20 a): *Calicotome villosa*, *Cyclamen repandum*, *Anthyllis barba-jovis*, *Cistus ladanifer*, *Cistus creticus* var. *corsicus*, *Genista linifolia*, *Cistus ladanifer*, *Cistus halimifolius*, *Cistus creticus* var. *eriocephalus*, *Euphorbia biumbellata*, *Cistus crispus*, *Clematis cirrhosa*. C'est un groupe subissant l'influence littorale et un climat plus doux, sur sol non calcaire. Le groupe 1 (fig 20.b), groupe proche du groupe 4 sur le plan factoriel, est également associé à un climat doux, mais sur substrat: *Helianthemum marifolium*, *Atractylis humilis*, *Convolvulus lanuginosus*, *Anthyllis cytisoides*, *Viola arborescens*, *Rhus coriaria*, *Cheirolophus intybaceus*, *Asplenium sagittatum*, *Hieracium latifolium*, *Galium setaceum*, *Brassica montana*. Enfin, le dernier groupe que nous présenterons, le groupe 6 (fig.20 c) a des coordonnées positives sur l'axe 1 et négatives sur l'axe 2, est composé des espèces suivantes: *Rhinanthus burnatii*, *Linum austriacum*, *Hormathophylla halimifolia*, *Primula allionii*, *Cytisus ardoini*, *Cytisus triflorus*, *Sedum fragrans*, *Stachys heraclea*, *Galium obliquum*, *Galatella sedifolia*, *Potentilla saxifraga*, *Potentilla zimmereri*, *Centaurea jordaniana* subsp. *Balbisiana*, *Coriaria myrtifolia*, *Euphorbia hyberna* subsp. *Canutii*. Toutes ces espèces sont typiques et pour la plupart endémiques de la zone ligure.

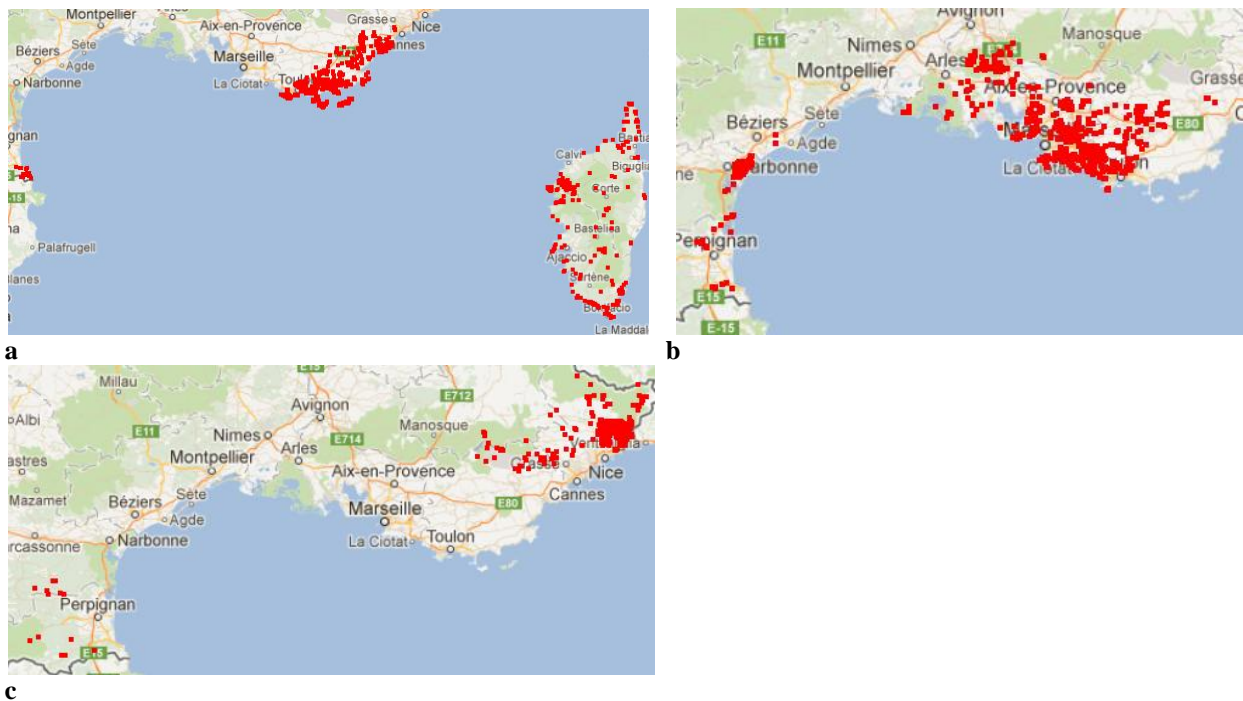


Figure 20: Localisation des stations du groupe 4 (a), le groupe 1 (b) et le groupe 6 (c)

4. DISCUSSION

4.1 Organisation de la végétation méditerranéenne

Il ressort des différentes analyses d'ordination, l'existence au sein de ce module d'une variation des données floristiques reliée aux conditions bioclimatiques, comme en témoigne la corrélation entre l'axe 2 de l'AFC et l'axe 1 de l'ANSC avec la dimension 1 de l'ACP (respectivement de -0,76 et de -0,71). Il existe un gradient global depuis les stations les plus chaudes du pourtour méditerranéen jusqu'aux stations plus septentrionales. En termes d'espèces, cela oppose des espèces caractéristiques de l'espace méditerranéen et des espèces

qui ne le sont pas mais qui caractérisent une influence méditerranéenne diffuse dans les habitats thermophiles et xérophiles* en remontant vers le nord.

Différents auteurs ont décrit à partir d'observations de terrain, l'organisation bioclimatique de l'espace méditerranéen et son influence sur la végétation méditerranéenne (Gausсен, 1938, Quezel, 1979, Rivas-Martinez, 1981, Unesco-FAO, 1963). La terminologie diffère selon les sources mais dans l'ensemble, les auteurs distinguent différents étages bioclimatiques qui diffèrent par leur température plus ou moins fraîche et leur sécheresse (caractéristiques de l'influence méditerranéenne): l'étage méso-méditerranéen, le plus chaud et le plus sec, caractérisée par des « Salsepareilles et Calycotome [...] Cistes, Lentisques, Romarin, Lavande » et par le pin d'Alep et le chêne vert (Gausсен, 1938) ; l'étage subméditerranéen ou supra-méditerranéen caractérisée notamment par l'apparition du chêne pubescent et l'étage oro ou laté méditerranéen, varié dans sa composition mais qui comprend des espèces xérophiles comme *Fumana procumbens*. Le gradient observé, dans notre étude, par les méthodes d'analyses factorielles, révèle ce gradient. En effet celui-ci oppose des espèces strictement inféodées à l'étage mésoméditerranéen à des espèces qui ne le sont pas et que l'on retrouve à l'étage subméditerranéen et latéméditerranéen.

La deuxième dimension de variabilité présente dans les données bioclimatiques et topographique met en évidence un gradient est-ouest au sein de la zone méditerranéenne qui oppose le pourtour du golfe du Lion avec une zone particulière, les Alpes-Maritimes. On observe au niveau de cette zone, une pluviométrie plus importante pendant les périodes plus froides et plus humides, tout en ayant de chaudes températures comme dans le reste de la méditerranée. Ce gradient se retrouve également dans les données floristiques, comme l'illustre le deuxième axe de l'ANSCVI mais également le premier axe de l'AFC pour des espèces très particulières. Cette zone est connue pour être un hotspot de biodiversité avec de nombreuses espèces endémiques (Casazza et al., 2008, Médail and Diadema, 2009, Médail and Quézel, 1999). En effet, les précipitations plus importantes sur cette zone, et sa latitude plus faible que les autres massifs alpins en ont fait une zone refuge pendant les périodes de glaciation (Médail and Diadema, 2009). De plus, la classification réalisée dans l'étude, permet de distinguer des catégories de communautés végétales sur les gradients bioclimatiques avec des particularités marquées. Or on retrouve dans le groupe 6 qui caractérise les Alpes-Maritimes, quatre espèces endémiques des Alpes-Maritimes (*Primula allionii*, *Cytisus ardoini*, *Potentilla saxifraga*, *Centaurea jordaniana subsp. Balbisiana*) (Casazza et al, 2008).

Le deuxième axe de l'ANSC n'est pas lié aux variables bioclimatiques mais reflète différentes communautés végétales au sein de l'espace méditerranéen: les espèces thermophiles poussant sur calcaire basique de l'étage mésoméditerranéenne et les autres espèces d'autres part. Le lien avec des données de sols auraient peut-être pu permettre de mettre en lumière une cause environnementale de cette répartition.

4.2 L'apport des différentes analyses multivariées

La comparaison des différentes méthodes d'ordination (AFC, ANSC, ANSCVI) montre les spécificités de chacune. Elles ont toutes mis en évidence sur les deux premiers axes, un des aspects très structurants du jeu de données: le gradient des étages de la végétation méditerranéenne. Ainsi lorsqu'un effet est très fort, avec en plus une correspondance bioclimatique, elles le détectent toutes. L'AFC a cependant montré, comme à l'usage, sa tendance à mettre en avant des modalités rares, ce qui est le cas pour l'axe 1 qui repose sur une quinzaine d'espèces seulement. Les premiers axes de l'ANSC, en revanche, reposent sur un grand nombre d'espèces et mettent en évidence des groupes pertinents d'espèces. Cette analyse apporte un éclairage intéressant sur ce qui structure globalement les communautés d'espèces au sein d'un territoire. Enfin, l'ANSCVI est intéressante pour interpréter

conjointement l'influence des variables environnementales et la répartition de la végétation. Cependant cela limite l'étude à la variabilité expliquée par ces variables environnementales et si celles-ci ne sont pas pertinentes, il est possible de passer à côté de structures qui dans le jeu de donnée expliquent, de manière plus forte, l'organisation des communautés végétales. Allier une analyse d'ordination non contrainte avec une autre analyse des variables environnementales, comme ici l'ACP, apparaît comme une approche alternative et complémentaire pour analyser conjointement la variabilité des variables environnementales et de la composition floristique. L'Analyse Factorielle Multiple est aussi dans ce cas une analyse qui serait intéressant à tester, sous réserve dans le cas d'un gros jeu de données, de ne pas être limité par la capacité calculatoire.

4.3 L'apport de la théorie des graphes

L'analyse de modularité appliquée au graphe projeté « espèce-abondance » a montré ses capacités à former des groupes d'espèces par grand types de végétation. C'est donc une approche intéressante et innovante qui permet de se détacher des algorithmes de classification habituels et du choix de mesure de similarité/dissimilarité entre espèces. Cela apporte aussi l'avantage d'être moins lourd calculatoirement. On peut cependant faire plusieurs remarques sur ces résultats qui mériteraient d'être étudiés plus en détails.

On observe, avant sélection des stations caractéristiques, que les modules regroupant le plus d' « espèces-abondances » ne sont pas les modules qui regroupent le plus de stations. On peut se demander dès lors si cela ne souligne pas que certains modules regroupent des espèces avec des faibles exigences écologiques et qui se trouvent ensemble sur un grand nombre de sites, comme le module 3, et des modules regroupant des espèces plus spécifiques d'un type de milieu qui favorisent par ailleurs la diversité floristique comme le module 2.

Un des résultats surprenant de cette analyse, est la distribution des stations à travers les modules. En effet non seulement, en moyenne, les stations se divisent entre 2 et 3 modules, mais en plus 38,3% des stations n'ont pas 80% de leur couverture dans un seul module. Ce qui veut dire qu'elles ont des espèces dans des modules différents. Une hypothèse est qu'il peut s'agir de milieux de transition où des espèces de différents types co-occurrent, et/ou que cela reflète le caractère relativement continu des gradients floristiques. Certains modules ont très peu de stations spécifiques (ex: module 7: Pelouses sèches et landes calcicoles). Une conclusion possible est que certaines espèces qui co-occurrent très souvent ne sont pas des espèces qui dominent. Par ailleurs, ces espèces qui dominent s'associent plus fortement à d'autres espèces et appartiennent donc à un autre module. Au contraire, ce qui expliquerait le nombre important de stations spécifiques du module 8 est la particularité des espèces halophytes. En effet les milieux salés induisent une spécialisation écologique forte et les espèces halophytes sont des espèces adaptées aux milieux salés comme les zones littorales notamment. Il paraît raisonnable de penser que les stations où elles se trouvent n'hébergent pas de manière importante d'autres espèces classées dans d'autres modules. Il serait intéressant de poursuivre l'étude de ce graphe pour valider ces hypothèses et comprendre de plus près les mécanismes en jeu.

4.4 Perspectives

4.4.1 Affinement des données

Comme on a pu le voir, les données de la base SOPHY sont très riches, complexes et hétérogènes. Malgré un long travail préliminaire sur les données avant la réalisation de cette étude, de nombreux paramètres n'ont pas été pris en compte et ceux-ci peuvent être des

sources de biais. Sur le long terme, pour aboutir à des analyses plus précises, il sera important de lever ces ambiguïtés dans le jeu de données floristiques. Le point le plus important est la question de la sélection des relevés. Dans cette étude, aucune sélection n'a été faite sur ceux-ci, mise à part l'exclusion des relevés en pays étrangers. Or de nombreux relevés au sein d'un même tableau ont été faits quasiment aux mêmes endroits, même si ce n'est pas le cas pour tous les tableaux. L'AFC met en évidence cette proximité des relevés sur son premier axe. On a donc une redondance de l'information d'autant plus forte que le nombre de relevés dans un tableau est important. Dans de nombreuses études sur les habitats potentiels de la végétation une distance minimale entre relevés est fixée. Par ailleurs, des zones en France ont été plus inventoriées que d'autres et certains milieux spécifiques et rares ont pu faire l'objet de plus nombreux inventaires que des milieux plus communs. Un rééquilibrage entre ces différents aspects sera sans doute à trouver en ré échantillonnant par exemple les stations. La question de la surface échantillonnée lors d'un inventaire est aussi une question délicate. En effet, en phytosociologie, un inventaire floristique se fait normalement sur une communauté végétale homogène avec selon le type de milieu une surface précise à inventorier. En s'appuyant sur cette méthodologie, on pourrait s'abstraire de l'information de surface comme dans cette étude. Mais on a pu voir que certains relevés pouvaient posséder jusqu'à 139 espèces, ce qui est forcément dû à une grande surface d'échantillonnage. Par ailleurs, une grande diversité spécifique en Alsace semble étrange. Ce que montre pourtant la carte de diversité spécifique (fig.9) avec de la diversité spécifique. Une réflexion est à mener pour limiter ce genre de biais. De plus, des choix différents auraient pu être faits comme notamment la suppression des taxons qui ne possédaient pas de nom valide. Certains relevés ont donc des espèces en moins.

4.4.2 Analyses à continuer

Les résultats de cette étude n'ont pas pu être tous explorés en détails et appellent d'autres analyses. L'étude de ce qu'apporte réellement le découpage en « espèce-abondance » pour les perspectives de l'étude mériterait d'être explorée. En effet, ce découpage en trois catégories multiplie le nombre d'individus statistiques, ce qui, tant au niveau calculatoire qu'au niveau de l'interprétation, complexifie la tâche. Par ailleurs, seulement l'analyse du module « végétation méditerranéenne » a été réalisée. Il en reste sept autres qui ne l'ont pas été. Mais pour cela, il faudra résoudre les problèmes calculatoires que certains modules engendrent. De plus, seules les variables bioclimatiques ont été utilisées en tant que variables environnementales alors que des données d'occupation du sol, du type de terrain géologiques et d'exposition ont été récupérées. La question de l'intégration de variables catégorielles et continues dans l'ANSCVI n'a pas été résolue. Enfin, l'analyse modulaire puis l'ANSCVI ont été réalisées dans une optique de cartographie des habitats potentielles des espèces et communautés d'espèces. En effet les axes de l'ANSCVI étant des combinaisons linéaires des variables environnementales, il est possible de calculer les coordonnées factorielles d'une station dont on ne connaîtrait que les données environnementales. En déterminant quel est le médoid de la classification CLARA le plus proche de cette station, il est possible d'attribuer à cette station la composition floristique caractéristique de cette classe. Par les techniques SIG actuelles, on peut imaginer étendre ce processus de cartographie potentielle à la France entière pixelisée, en déterminant pour chaque pixel (caractérisé par des variables environnementales) la composition floristique qui s'y rattache.

5. CONCLUSION

D'un point de vue méthodologique, l'étude de l'organisation de la végétation à partir de la base de données SOPHY a mis en évidence, d'une part, les problèmes soulevés par le traitement d'une base de données de cette ampleur (harmonisation de la taxonomie, répartition des données d'inventaires) et d'autre part, la pertinence de la théorie des graphes et les méthodes d'ordination pour étudier l'organisation de la végétation à cette échelle. D'un point de vue écologique, l'étude de la végétation méditerranéenne, prise comme exemple pour tester les différentes approches d'ordination, a mis en lumière l'organisation en gradient des différentes espèces végétales sous l'influence du climat. Sur ces gradients, des communautés végétales particulières ont pu être identifiées renouant avec la conception phytosociologique qui établit que la végétation se compose d'un assemblage de communautés d'espèces au sein d'un territoire. Cette étude a amené de nombreuses questions qui n'ont pas encore été approfondies, notamment dans le but d'une cartographie de la végétation potentielle. Par ailleurs les données de la base SOPHY sont très riches et de nombreuses informations attendent d'être intégrées et analysées. Cette étude laisse ainsi de belles perspectives d'analyses pour les années à venir.

BIBLIOGRAPHIE

- AUSTIN, M. P. (1980) Searching for a model for use in vegetation analysis. *Plant Ecology*, 42, 11-21.
- AUSTIN, M. P. (2005) Chapter 2: Vegetation and environment: discontinuities and continuities. IN VAN DER MAAREL, E. (Ed.) *Vegetation ecology*. Etats-Unis, BLACKWELL PUBLISHING.
- BECKING, R. W. (1957) The zürich-montpellier school of phytosociology. *Botanical Review*, 23, 411-488.
- BLONDEL, V., GUILLAUME, J.-L., LAMBIOTTE, R. & LEFEBVRE, E. (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, P10008.
- BRAUN-BLANQUET, J. (1928) *Pflanzensoziologie: Grundzüge der Vegetationskunde*, Berlin, Springer.
- CASAZZA, G., ZAPPA, E., MARIOTTI, M. G., MÉDAIL, F. & MINUTO, L. (2008) Ecological and historical factors affecting distribution pattern and richness of endemic plant species: the case of the Maritime and Ligurian Alps hotspot. *Diversity and Distributions*, 14, 47-58.
- CHESEL, D. & GIMARET, C. (1997) Analyse non symétrique des correspondances. *Ade4-Fiche thématique 2.8*. Lyon, Université Lyon 1.
- DORMANN, C., FRÜND, J., BLÜTHGEN, N. & GRUBER, B. (2009) Indices, Graphs and Null Models: Analyzing Bipartite Ecological Networks *The Open Ecology Journal*, 2, 7-24.
- DRAY, S. & CHESEL, D. (2009) L'ordination des tableaux écologiques. *Fiche TD avec le logiciel R: ter2*. Lyon, Université Lyon 1.
- GAUSSEN, H. (1938) Étages et zones de végétation de la France. *Annales de Géographie* 47, 463-487.
- GIMARET-CARPENTIER, C., CHESEL, D. & PASCAL, J. P. (1998) Non-symmetric correspondence analysis: an alternative for species occurrences data. *Plant Ecology*, 138, 97-112.
- GUIMERA, R. & AMARAL, L. A. N. (2005) Functional cartography of complex metabolic networks. *Nature*, 433, 895-900.
- GUINOCHET, M. (1955) *Logique et dynamique du peuplement végétal.*, Paris, Masson édition.
- HIJMANS, R. J., CAMERON, S. E., PARRA, J. L., JONES, P. G. & JARVIS, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-1978.
- JANSEN, F. & DENGLER, J. (2010) Plant names in vegetation databases – a neglected source of bias. *Journal of Vegetation Science*, 21, 1179-1186.
- JARVIS, A., REUTER, H. I., NELSON, A. & GUEVARA, E. (2008) Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database: <http://srtm.csi.cgiar.org>. Consultative Group on International Agricultural Research-Consortium for Spatial Information (CGIAR-CSI).
- JONES, M., SCHILDHAUER, M., REICHMAN, O. J. & BOWERS, S. (2006) The new bioinformatics: Integrating ecological data from the gene to the biosphere. *Annual Review of Ecology, Evolution, and Systematics*, 37, 519-544.

- JONGMAN, R., TER BRAAK, C. & VAN TONGEREN, O. (1995) *Data Analysis in Community and Landscape Ecology*, Cambridge, Cambridge University Press.
- KAUFMAN, L. & ROUSSEEUW, P. J. (1990) *Finding Groups in Data: An introduction to cluster analysis*. New York, John Wiley & Sons, Inc.
- KELLING, S., HOCHACHKA, W., FINK, D., RIEDEWALD, M., CARUANA, R., BALLARD, G. & HOOKER, G. (2009) Data-Intensive Science: A New Paradigm for Biodiversity Studies. *BioScience*, 59, 613-620.
- KROONENBERG, P. M. & LOMBARDO, R. (1999) Nonsymmetric Correspondence Analysis: A Tool for Analysing Contingency Tables With a Dependence Structure. *Multivariate Behavioral Research*, 34, 367-396.
- LACOSTE, A. & SALANON, R. (1978) *Eléments de biogéographie et d'écologie*, Paris.
- LAURO, N. & D'AMBRA, L. (1984) L'analyse non symétrique des correspondances. IN DIDAY, E. (Ed.) *Data analysis and informatics III*. Amsterdam, Elsevier.
- LEGENDRE, P. & LEGENDRE, L. (1998a) Chapter 9 Ordination in reduced space. *Developments in Environmental Modelling*. Elsevier.
- LEGENDRE, P. & LEGENDRE, L. (1998b) Chapter 11 Canonical analysis. *Developments in Environmental Modelling*. Elsevier.
- LOOIJEN, R. C. & VAN ANDEL, J. (1999) Ecological communities: conceptual problems and definitions. *Perspectives in Plant Ecology, Evolution and Systematics*, 2, 210-222.
- MAY, R. M. (2006) Network structure and the biology of populations. *Trends in ecology & evolution (Personal edition)*, 21, 394-399.
- MÉDAIL, F. & DIADEMA, K. (2009) Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography*, 36, 1333-1345.
- MÉDAIL, F. & QUÉZEL, P. (1999) Biodiversity Hotspots in the Mediterranean Basin: Setting Global Conservation Priorities. *Conservation Biology*, 13, 1510-1513.
- MICHENER, W. K. & JONES, M. B. (2012) Ecoinformatics: supporting ecology as a data-intensive science. *Trends in ecology & evolution (Personal edition)*, 27, 85-93.
- NEWMAN, M. E. J. (2003) The Structure and Function of Complex Networks. *SIAM Review*, 45, 167-256.
- NEWMAN, M. E. J. (2006) Modularity and community structure in networks. *PNAS*, 103, 8577-8582.
- NEWMAN, M. E. J. & GIRVAN, M. (2004) Finding and evaluating community structure in networks. *Physical Review E*, 69, 026113.
- PATTERSON, D. J., COOPER, J., KIRK, P. M., PYLE, R. L. & REMSEN, D. P. (2010) Names are key to the big new biology. *Trends in Ecology & Evolution*, 25, 686-691.
- PEET, R. K. & ROBERTS, D. W. (2012) Chapter 2: Classification of natural and semi-natural vegetation. IN FRANKLIN, J. & VAN DER MAAREL, E. (Eds.) *Vegetation Ecology*. Second ed. New York, Oxford University Press.
- PÉLISSIER, R., COUTERON, P., DRAY, S. & SABATIER, D. (2003) Consistency between ordination techniques and diversity measurements: two strategies for species occurrence data. *Ecology*, 84, 242-251.
- QUEZEL, P. (1979) La Région Méditerranéenne française et ses essences forestières. Signification écologique dans le contexte circum-méditerranéen. *Forêt méditerranéenne*, 7-18.
- REICHMAN, O. J., JONES, M. B. & SCHILDHAUER, M. P. (2011) Challenges and Opportunities of Open Data in Ecology. *Science*, 331, 703-705.
- RIVAS-MARTINEZ, S. (1981) Les étages bioclimatiques de la végétation de la Péninsule ibérique. *ACTAS III CONGR. ÓPTIMA. ANALES JARD BOT MADRID*.

- THÉBAULT, E. & FONTAINE, C. (2010) Stability of Ecological Communities and the Architecture of Mutualistic and Trophic Networks. *Science*, 329, 853-856.
- UNESCO-FAO (1963) Bioclimatic map of the Mediterranean zone. *Explanatory notes*. Paris.
- WHITTAKER, R. H. (1967) Gradient analysis of vegetation. *Biological Reviews*, 42, 207-264.

ANNEXES

Annexe I: Problèmes liés à la taxonomie et méthodes pour y palier

La base SOPHY présente l'avantage d'avoir un très grand nombre de données mais étant issue d'un recueil de très nombreuses sources différentes et sur une longue période de temps, elle possède un certain nombre d'incohérences et d'imprécisions.

Le point le plus problématique a concerné la taxonomie, notamment avec les problèmes liés à la synonymie et les autonymes.

- Synonymes

Pour les non-botanistes, il peut paraître évident que chaque espèce ou sous-espèce possède un seul nom latin bien défini et inversement qu'un nom latin désignant une espèce, correspondent qu'à une seule espèce. Or ce n'est malheureusement pas le cas en France. Seul un binôme et trinôme latin avec la précision du nom de l'auteur qui l'a nommé permet, dans la majorité, des cas de remonter à un taxon précis. Ainsi il existe des problèmes de synonymie et des problèmes de niveaux taxonomiques flous (rang espèces et infra-espèces).

Par exemple, le bleuet *Centaurea cyanus* L. qui est le nom retenu par les scientifiques, peut-être également désigné dans la littérature par 9 binômes ou trinômes latins différents: *Cyanus vulgaris* Delarbre, *Centaurea umbrosa* A.L.P.Huet ex Reut, *Centaurea pulchra* DC., *Centaurea lanata* Roxb., *Centaurea hortorum* Pau, *Centaurea cynarocephala* Velen., *Centaurea cyanus* var. *denudata* Suksd., *Centaurea cyanus* subsp. *coa* Rech.f., *Centaurea cyanocephala* Velen. Et pour mesurer l'importance du nom d'auteur, si on prend un des synonymes du bleuet *Centaurea lanata* Roxb. sans préciser le nom de l'auteur, dans ce cas *Centaurea lanata*, on peut aussi bien désigner le bleuet que le carthame laineux dont un des synonymes est *Centaurea lanata* (L.) DC (nom retenu: *Carthamus lanatus* subsp. *lanatus*).

Cet exemple illustrant la complexité de la taxonomie, il a été nécessaire pour la cohérence des analyses du jeu de donnée, de régler au maximum ces problèmes dans la base SOPHY et de se référer à un référentiel taxonomique, qui recense les synonymes et donne le nom valide qui a été retenu par les scientifiques pour un taxon donné.

- Règle de décision sur l'attribution du nom valide

Dans notre cas, le référentiel choisi est la Base de données des Trachéophytes de France Métropolitaine (BDTFX v1_01). Pour le croisement entre le fichier SOPHY et la BDTFX, une routine a été construite sur R (v 2.14.2), pour récupérer le nom valide quand une seule possibilité se présentait. Il y a évidemment des cas, où plusieurs rattachements étaient possibles. Dans ce cas, il a fallu regarder à la main si le nom d'auteur n'était pas précisé dans les colonnes supplémentaires ou si la correspondance faite par les fondateurs de SOPHY, avec la flore de Fournier ne pouvait pas régler les cas litigieux. Enfin il y avait aussi des cas avec des erreurs orthographiques qui ont été également corrigées à la main. Les taxons pour lequel il a été impossible de déterminer le nom valide, soit parce qu'il pouvait être rattaché à plusieurs taxons, soit parce qu'il était inconnu dans la BDTFX ou enfin parce qu'aucun nom valide n'était référencé dans la BDTFX, ont été écartés.

- Niveau taxonomique et gestion des autonymes

Par ailleurs, il a fallu faire un choix quand au niveau taxonomique à garder. En effet, selon les taxons, le niveau taxonomique renseigné se situait au niveau du genre ou de l'espèce ou du niveau infra-espèce (comprenant les sous-espèces, variétés, formes, croisement). Nous avons

fait le choix de retirer les taxons renseignés au niveau du genre, l'information n'étant pas assez précise pour être pertinente. Enfin nous avons été confrontés au problème que pose les autonymes. Un autonyme est un taxon pour lequel le nom de l'espèce et de la sous-espèce se répète par exemple *Quercus pubescens ssp pubescens*. Ceux-ci désignent en fait le spécimen qui a servi de référence pour la première description de l'espèce et pour lequel il a été ensuite découvert des sous-espèces. Mais cela peut générer des ambiguïtés car selon les auteurs des relevés, la sous espèce pour un autonyme peut être précisé mais peut aussi ne pas l'être. Ainsi pour le taxon nommé *Quercus pubescens*, nous ne savons pas dans la base SOPHY s'il désigne l'autonyme *Quercus pubescens ssp pubescens* ou un individu de l'espèce *Quercus pubescens* mais pour lequel la sous-espèce n'a pas été identifiée. Pour palier à ces ambiguïtés, le choix de tout ramener au niveau de l'espèce a été envisagée, mais il signifiait fusionner dans certain cas des sous-espèces qui vivent dans des contextes écologiques très différents comme les sous-espèces littorales. Nous avons donc émis l'hypothèse que le niveau espèce désignait les autonymes. Ainsi, tous les autonymes présents ont été rattachés au nom du binôme valide associé et nous avons laissé les indications des infra-espèces quand elles étaient indiquées.

Annexe II: Données WorldClim

Les données WorldClim codées de la manière suivante:

Annualtemp	moyenne annuelle de températures
diurnalrange	moyenne de l'écart des extrema de températures mensuel (Moyenne de l'écart (max temp - min temp) par mois)
isotherm	iso thermalité (diurnalrange / annualrange) (* 100)
seasonality	saisonnalité des températures (écart type des températures*100)
warmestmonth	Température maximum du mois le plus chaud
coldestmonth	Température minimum du mois le plus froids
annualrange	Ecart des extrema de température annuelle (warmestmonth - coldestmonth)
wettestquarter	Température moyenne sur le trimestre le plus humide
driestquarter	Température moyenne sur le trimestre le plus sec
warmestquarter	Température moyenne sur le trimestre le plus chaud
coldestquarter	Température moyenne sur le trimestre le plus froid
precip	Précipitation annuelle
wettestprecip	Précipitation du mois le plus humide
driestprecip	Précipitation mois le plus sec
precipseason	Saisonnalité des précipitations (écart type des températures)
wettestquarterprecip	Précipitation du trimestre le plus humide
driestquarterprecip	Précipitation du trimestre le plus sec
warmestquarterprecip	Précipitation du trimestre le plus chaud
coldestquarterprecip	Précipitation du trimestre le plus froid

Les températures sont en °C * 10 et les précipitations en mm.

Annexe III: AFC et ANSC

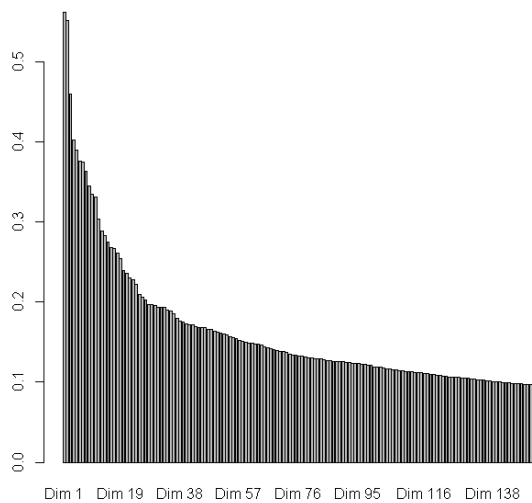
Cette méthode développée initialement par Lauro et D'Ambra (Lauro and D'Ambra, 1984) ne considère plus que les deux variables étudiées jouent un rôle symétrique mais qu'il y a un lien causal entre les deux variables et donc une dissymétrie (Kroonenberg and Lombardo, 1999). On distingue ainsi une variable de contrôle ou variable prédictrice (variable relevé dans notre cas en ligne) et une variable dépendante (ici, variable espèces en colonne). La principale différence entre l'ANSC et l'AFC est la métrique, induite par une pondération différente des modalités des deux variables. En réalisant une ANSC, dans l'espace du nuage des modalités de la variable de contrôle (R^I ici), les points représentant les modalités de la variable de contrôle (les I relevés) sont munis du poids de leur marge respective (f_i), comme en AFC. En revanche, les axes représentant les modalités de la variable dépendante (les J espèces), sont munis d'un poids uniforme contrairement à la métrique du χ^2 , où le poids de ces axes correspond à l'inverse de la marge des modalités représentées ($1/f_j$). La double propriété barycentrique n'est pas valable en ANSC et chaque nuage est représenté séparément. Les modalités de la variable de contrôle, ici les relevés, se trouvent au barycentre des espèces (modalités de la variable aléatoire) dans l'espace des composantes principales normées à 1. Du fait de la pondération, l'ANSC limite l'influence des espèces rares et favorise les espèces fréquentes.

La justification écologique que l'on trouve dans la littérature Chessel and Gimaret, 1997, Gimaret-Carpentier et al., 1998, Pélissier et al., 2003) à l'utilisation de cette analyse pour étudier les relevés floristiques repose sur différents arguments. Un des arguments est d'ordre conceptuel et se fonde, selon les écologues, sur le fait qu'il existe une dissymétrie entre les relevés et les espèces car un relevé est une unité défini par l'expérimentateur et la liste des relevés (ce qui revient à l'ensemble des modalités de la variable relevé) est également fixée par celui-ci. En revanche les différents taxons rencontrés (donc les modalités de la variable « espèce ») sont une conséquence de la définition des relevés et de la liste choisie et sont incontrôlés. Les arguments suivants portent sur les différences qu'implique le changement de métrique et de pondération par rapport à l'AFC dans ce que l'on souhaite mettre en évidence dans le jeu de donnée et dans l'interprétation des résultats. Si la mise en avant des espèces rares dans le jeu de donnée n'est pas souhaitée, l'ANSC propose une alternative à l'AFC. La prise en compte des données d'absence est aussi un critère de choix car la métrique du χ^2 , selon les écologues, permet justement de ne pas en tenir compte. Enfin Pélissier et al (Pélissier et al., 2003) ont pu montrer que l'ANSC décompose la diversité β de Simpson. Ce que ne fait pas l'AFC.

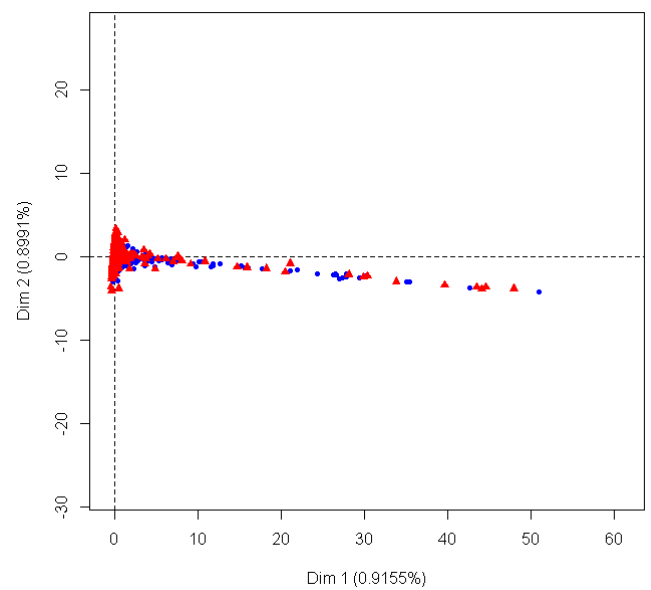
Annexe IV: Résultats de l'AFC

(a) Valeurs propres et pourcentage d'inertie expliquée pour les cinq premiers axes

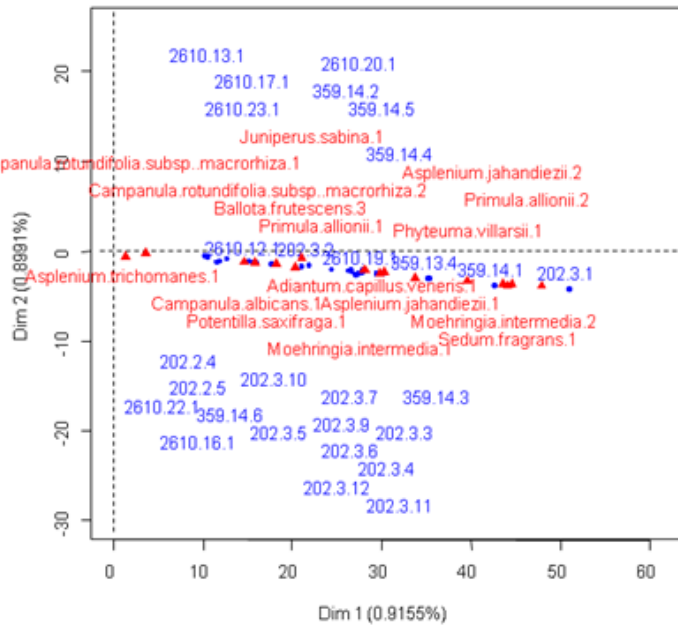
	Valeur propre	% de variance	% cumulé de variance
dim 1	0.561	0.916	0.916
dim 2	0.551	0.899	1.815
dim 3	0.460	0.750	2.565
dim 4	0.402	0.655	3.220
dim 5	0.389	0.635	3.854



(b) Diagramme des 100 premières valeurs propres

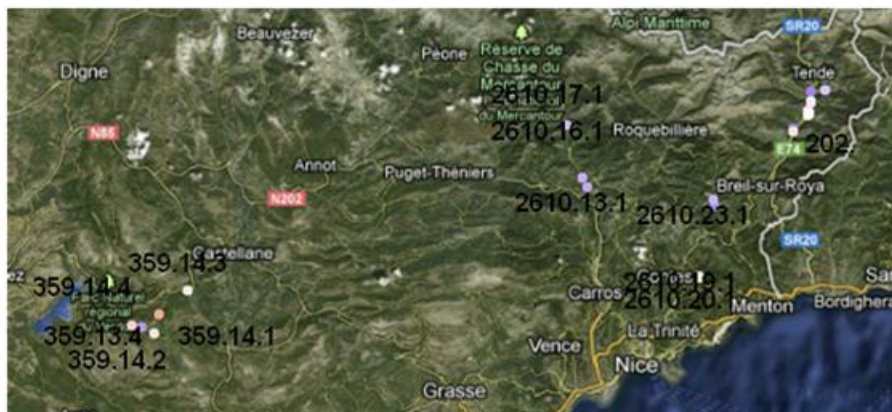


(c) Premier plan factoriel



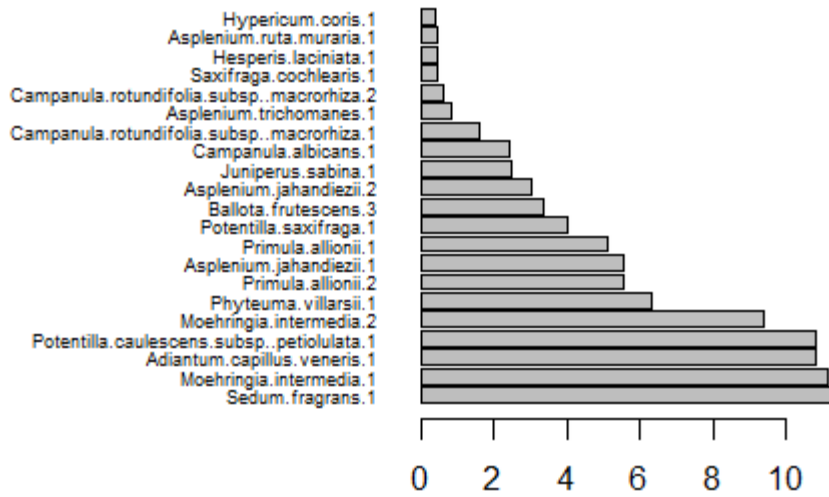
(d) Premier plan factoriel avec zoom sur l'axe 1

Les individus de cette figure ont une contribution à l'axe 1 supérieure à 0.5

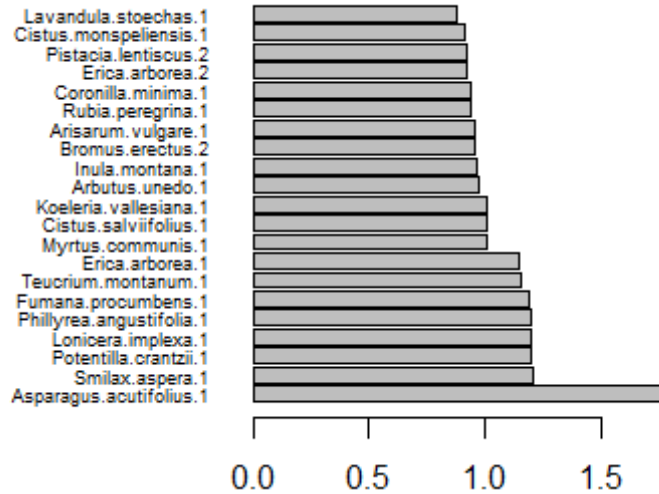


(e) Cartographie du premier axe factoriel

Seules les stations ayant une contribution supérieure à 0.5 sont représentées. Les points représentent les stations et leur couleur indique leur position sur l'axe 1 (Rouge : coordonnée factorielle positive, blanc : coordonnée proche de zéro, bleu : coordonnée négative. Couleur intense rouge ou bleu : coordonnée élevée en valeur absolue.



(f) « Espèces-abondances) contribuant le plus à l'axe 1



(g) « Espèces-abondances » contribuant le plus à l'axe 2

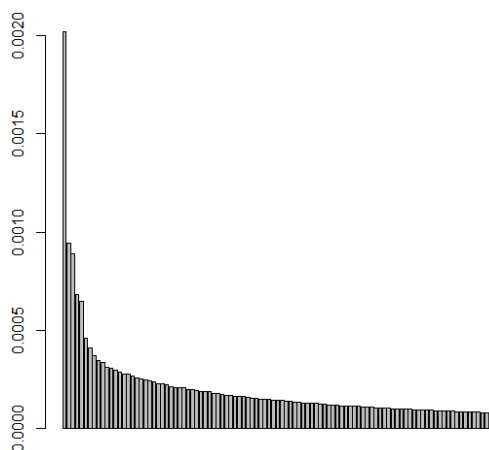
Annexe V: Résultats de l'ANSC

(a) Décomposition de l'inertie:

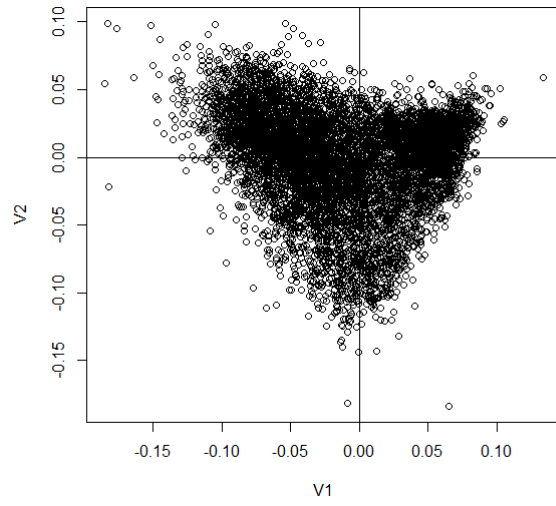
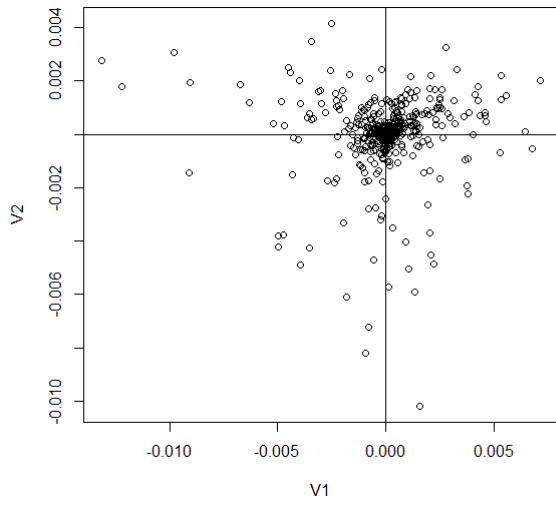
Inertie totale	0.9950
Inertie de la partition selon les lignes	0.0324
Pourcentage d'inertie de la partition	0.0326

(b) Valeurs propres et pourcentages d'inertie expliquée pour les cinq premiers axe

Valeur propre	% de variance	% cumulé de variance
0,0020	6,2	6,2
0,0009	2,9	9,1
0,0009	2,7	11,9
0,0007	2,1	14,0
0,0006	2,0	16,0



(c) Diagramme des 100 premières valeurs propres

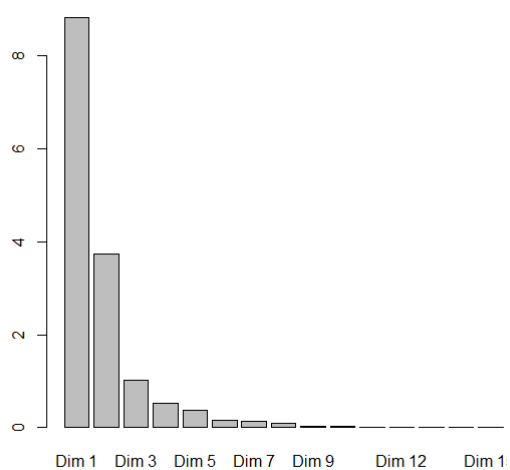


(d) Premier plan factoriel : nuage de points des espèces (droite) et nuage de points des stations (gauche)

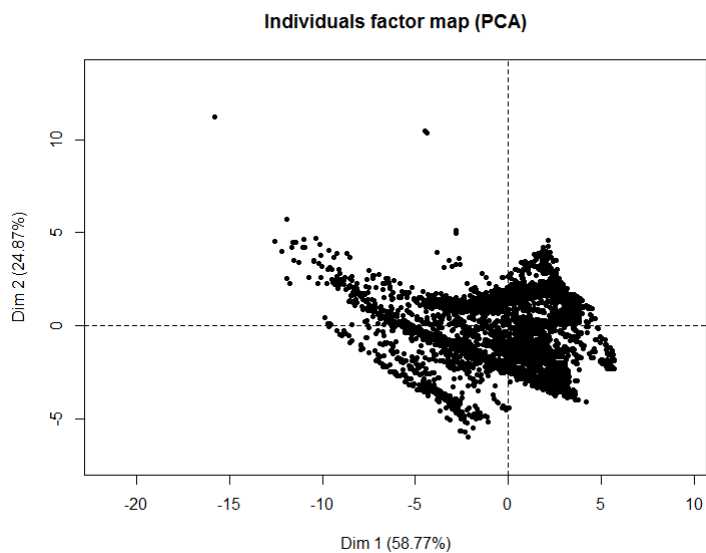
Annexe VI: Résultats de l'ACP des variables bioclimatiques:

(a) Valeurs propres et pourcentages d'inertie expliquée

	Valeur propre	% de variance	% cumulé de variance
comp 1	8,8	58,8	58,8
comp 2	3,7	24,9	83,6
comp 3	1,0	6,9	90,5
comp 4	0,5	3,6	94,1
comp 5	0,4	2,5	96,6
comp 6	0,2	1,1	97,8
comp 7	0,1	1,0	98,7
comp 8	0,1	0,6	99,4
comp 9	0,0	0,2	99,6
comp 10	0,0	0,2	99,8
comp 11	0,0	0,1	99,9
comp 12	0,0	0,0	99,9
comp 13	0,0	0,0	100,0
comp 14	0,0	0,0	100,0
comp 15	0,0	0,0	100,0



(b) Diagramme des valeurs propres



(c) Nuage des individus dans le premier plan factoriel

(d) Corrélation entre les deux premières composantes de l'ACP et les variables

	Dim.1	Dim.2
longitude_wgs	0.26	0.56
latitude_wgs	-0.62	-0.12
annualtemp	0.97	0.09
diurnalrange	-0.22	-0.54
isotherm	0.22	-0.09
seasonality	-0.59	-0.63
annualrange	-0.42	-0.64
Dim1_AFC	-0.09	0.00
Dim2_AFC	-0.76	-0.20
Dim1_ANSC	-0.71	-0.11
Dim2_ANSC	-0.10	0.26

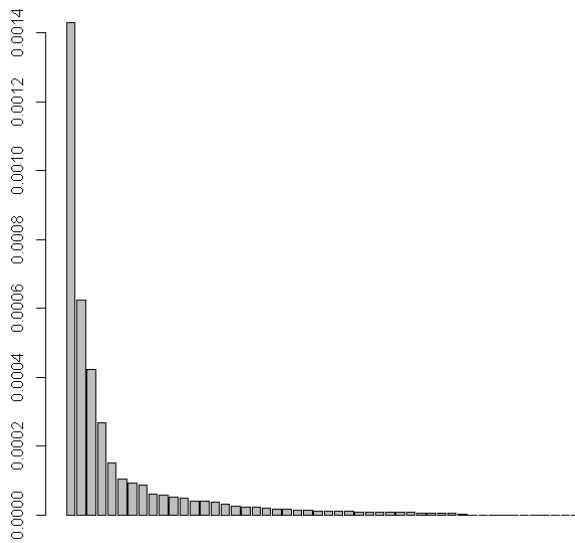
(e) Corrélation entre les deux premières composantes de l'AFC et l'ANSC et les variables

	Dim1_AFC	Dim2_AFC	Dim1_ANSC	Dim2_ANSC
longitude_wgs	0.02	-0.32	-0.23	0.17
latitude_wgs	0.05	0.59	0.56	0.09
annualtemp	-0.09	-0.72	-0.66	-0.04
diurnalrange	-0.02	0.35	0.24	-0.25
isotherm	-0.07	-0.05	-0.08	-0.09
seasonality	0.05	0.61	0.50	-0.23
warmestmonth	-0.09	-0.47	-0.49	-0.23
coldestmonth	-0.08	-0.75	-0.66	0.07
annualrange	0.02	0.49	0.37	-0.28
wettestquarter	-0.05	-0.38	-0.35	-0.08
driestquarter	-0.05	-0.59	-0.53	-0.16
warmestquarter	-0.09	-0.68	-0.64	-0.10
coldestquarter	-0.08	-0.75	-0.67	0.02
precip	0.06	0.45	0.47	0.23
wettestprecip	-0.01	-0.23	-0.13	0.18
driestprecip	0.07	0.76	0.71	0.09
precipseason	-0.07	-0.78	-0.70	0.05
wettestquarterprecip	0.00	-0.19	-0.10	0.25
driestquarterprecip	0.08	0.77	0.72	0.07
warmestquarterprecip	0.07	0.70	0.68	0.20
coldestquarterprecip	-0.02	-0.13	-0.07	0.31
mnt	0.10	0.62	0.59	0.08
Dim1_AFC	1.00	-0.10	0.08	0.02
Dim2_AFC	-0.10	1.00	0.89	-0.12
Dim1_ANSC	0.08	0.89	1.00	-0.14
Dim2_ANSC	0.02	-0.12	-0.14	1.00

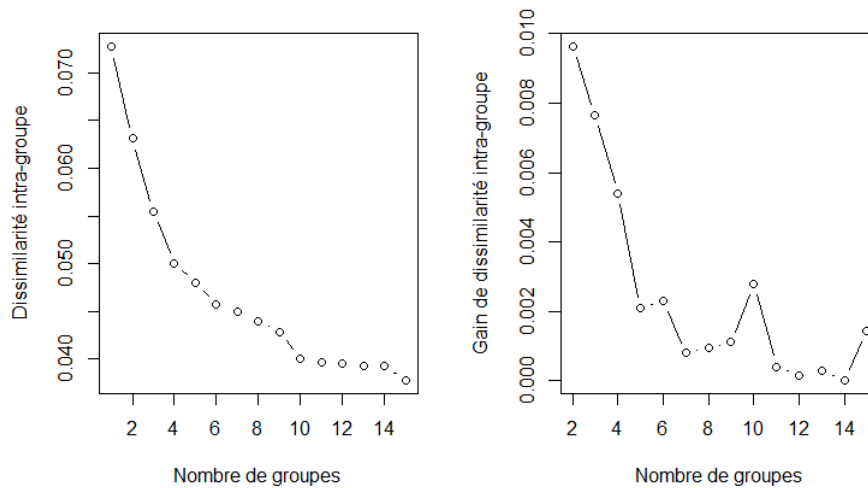
Annexe VII: Résultats de l'ANSCVI:

(a) Valeurs propres et pourcentages d'inertie expliquée

valeurs propres	% de variance	% de variance cumulé
0,00143	0,38	0,38
0,00062	0,16	0,54
0,00042	0,11	0,65
0,00027	0,07	0,72
0,00015	0,04	0,76



(b) Diagramme des 50 premières valeurs propres



(c) Evolution de la dissimilarité intra-groupe en fonction du nombre de groupes choisis



Laboratoire de mathématiques appliquées

Spécialité: Statistique Appliquée

Responsable: François Husson

Auteur: Marianne Gérard

Etablissement d'accueil: UMR AMAP

Nombre de pages: 27

Adresse: Bd de la Lironde TA A-51/
PS2 34398 Montpellier Cedex 5

Annexe (s): 12

Année de soutenance: 2012

Maître de stage: François Munoz

TITRE du MÉMOIRE en français: Analyse écologique de la répartition de la végétation à partir d'une base de données phytosociologiques: exemple de la végétation méditerranéenne

TITRE du MÉMOIRE en anglais: Study of the vegetation distribution from a phytosociologic database: example of Mediterranean vegetation

Résumé: Comprendre la répartition des espèces et des communautés d'espèces à grande échelle est au cœur des problématiques scientifiques pour la mise en place de programmes de préservation de la biodiversité et pour évaluer l'impact du changement climatique. La présente étude s'appuie sur une base de données de 200 000 relevés phytosociologiques sur la France entière. Elle a pour but de comprendre l'organisation de la végétation en lien avec l'environnement bioclimatique. Ce travail met en œuvre des notions et outils issus de la théorie des graphes afin de définir des groupes cohérents d'espèces végétales. Sur l'exemple du groupe « végétation méditerranéenne », des méthodes d'ordination directe et indirecte sont réalisées afin d'étudier la structuration de la végétation en rapport avec l'environnement bioclimatique des relevés. Nous avons pu montrer que l'approche de la théorie des graphes et les méthodes d'ordinations sont des méthodes pertinentes à l'étude de l'organisation de la végétation à cette échelle. Par ailleurs, l'étude de la végétation méditerranéenne a mis en lumière l'organisation en gradient des différentes espèces en lien avec les gradients climatiques. Sur ces gradients de végétation, des communautés végétales particulières ont pu être identifiées. Les spécificités de chaque analyse et les perspectives de cartographie de la végétation potentielle des communautés d'espèces sont discutées.

Abstract: The understanding of plant species and plant community distribution over large area is an important current scientific issue for biodiversity conservation planning and assessing the impacts of environmental change on vegetation distribution. The present study is based on a database of 200 000 phytosociological relevés. It aims to understand the vegetation organization in relation to bioclimatic variables. This work use tools and concepts from graph theory in order to define coherent groups of plant species in the data. On the example of Mediterranean vegetation, indirect and direct ordination analysis was performed in order to study the structuration of groups of plant species in relation to bioclimatic variables. We have shown that graph theory approach and ordinations methods are relevant methods to study the vegetation organization at national scale. The study of Mediterranean vegetation highlighted the gradient organization of plant in relation to the bioclimatic gradients. From theses vegetation gradient, specific plant communities have been identified. The specificity of each method and points to further, especially in order to predictive vegetation mapping are discussed.

Mots clés: Végétation, Ordination, Phytosociologie, Théorie des graphes, Biogéographie.